Novel Ranking Methods Applied to Complex Membership Determination Problems

Loc Hoang Tran[1], Linh Hoang Tran[2], and Hoang Trang[3]

[1] University of Minnesota/Computer Science Department, Minneapolis, USA
tran0398@umn.edu
[2] Portland State University/ECE Department, Portland, USA
linht@pdx.edu
[3] Ho Chi Minh City University of Technology-VNU HCM
Ho Chi Minh City, Vietnam
hoangtrang@hcmut.edu.vn

Abstract. The biological motivated problem that we want to solve in this paper is to predict the new members of a partially known protein complex (i.e. complex membership determination). In this problem, we are given a core set of proteins (i.e. the queries) making up a protein complex. However, the biologist experts do not know whether this core set is complete or not. Our objective is to find more potential members of the protein complex by ranking proteins in protein-protein interaction network. One of the solutions to this problem is a network reliability based method. Due to high time complexity of this method, the random walk on graphs method has been proposed to solve this complex membership determination problem. However, the random walk on graphs method is not the current state of the art network-based method solving bioinformatics problem. In this paper, the novel un-normalized graph (p-) Laplacian based ranking method will be developed based on the un-normalized graph p-Laplacian operator definitions such as the curvature operator of graph (i.e. the un-normalized graph 1-Laplacian operator) and will be used to solve the complex membership determination problem. The results from experiments shows that the un-normalized graph p-Laplacian ranking methods are at least as good as the current state of the art network-based ranking method ($p=2$) but often lead to better ranking accuracy performance measures.

Keywords: Ranking, graph, p-Laplacian, protein complex, membership.

1 Introduction

Recent development in genome projects has shown that the pair of genes or group of genes sharing an edge (or hyper-edge) in the protein (hyper-) networks [1, 2] tends to have similar complex biological function. Therefore, in recent years, the considerable amount of research effort has gone into discovering the complete set of interacting proteins in an organism. Protein-protein interaction networks have become certainty in recent years thanks to high throughput methods such as yeast-two-hybrid [3, 4] and affinity purification with mass spectrometry [5, 6]. Corresponding to the readiness of

protein networks, various graph analysis techniques have been proposed to mine these networks for protein function prediction [1, 2, 7, 8, 9, 10, 11] and complex/pathway membership determination [12, 13].

The biological motivated problem that we want to solve in this paper is to predict the new members of a partially known protein complex (i.e. complex membership determination). In this problem, we are given a core set of proteins (i.e. the queries) making up a protein complex. However, the biologist experts do not know whether this core set is complete or not. Our objective is to find more potential members of the protein complex by ranking proteins in protein-protein interaction network. Then the proteins with highest ranks (i.e. probability of membership in the partially known complex) will then be selected and checked by biologist experts to see if the extended genes in fact belong to the same complex/pathway.

One of the solutions to this problem is a network reliability based method proposed by [12]. This method finds the close proximity proteins. This method approximates the reliability between two nodes using Monte Carlo simulation. However, due to high time complexity of this method, the random walk on graphs method has been proposed to solve this complex membership determination problem [13]. The random walk method simulates a random walker that starts at a source node (i.e. the query or the member of the complex/pathway) and moves to other nodes through connecting edges. The random walker can also choose to teleport to the start node with a definite probability, the restart probability. The probability of finding the random walker at a certain node gives the similarity of that node to the starting node. To solve the complex membership determination problem, [13] simulates a random walker that starts from a set of nodes (i.e. the queries or the set of members of the complex/pathway) instead of a single node. Thus, given a set of proteins that are members of a partially known complex as the start set, the random walk on graphs method ranks the remaining proteins in the protein-protein interaction network with respect to their proximity to the queries' complex.

This ranking method [13] has also been employed by Google Company to exploit the global hyperlink structure of the Web and produce better rankings of search results [14]. Its idea [13] has also been employed in [1] to solve the protein function prediction problem (i.e. the classification problem). However, based on [1], the random walk on graphs method is not the best network-based method solving the classification bioinformatics problems such as protein function prediction [1, 11] and cancer classification [15]. To the best of my knowledge, the un-normalized graph (p-) Laplacian based semi-supervised learning method is considered the current state of the art network-based method solving protein function prediction problem [7, 11] and cancer classification problem [15]. However, the un-normalized graph (p-) Laplacian based ranking method has not yet been developed and obviously has not been applied to any practical applications. In this paper, the un-normalized graph (p-) Laplacian based ranking method will be developed based on the un-normalized graph p-Laplacian operator definitions such as the curvature operator of graph (i.e. the un-normalized graph 1-Laplacian operator). Finally, this proposed method will be used to solve the complex membership determination problem.

We will organize the paper as follows: Section 2 will introduce the preliminary notations and definitions used in this paper. Section 3 will introduce the definition of the gradient and divergence operators of graphs. Section 4 will introduce the definition of

Laplace operator of graphs and its properties. Section 5 will introduce the definition of the curvature operator of graphs and its properties. Section 6 will introduce the definition of the p-Laplace operator of graphs and its properties. Section 7 will show how to derive the algorithm of the un-normalized graph p-Laplacian based ranking method from regularization framework. In section 8, we will compare the accuracy performance measures of the un-normalized graph Laplacian based ranking algorithm and the un-normalized graph p-Laplacian based ranking algorithms. Section 9 will conclude this paper and the future direction of researches of other practical applications in bioinformatics utilizing discrete operator of graph will be discussed.

2   Preliminary Notations and Definitions

Given a graph $G=(V,E,W)$ where V is a set of vertices with $|V| = n$, $E \subseteq V * V$ is a set of edges and W is a $n * n$ similarity matrix with elements $w_{ij} > 0$ $(1 \leq i,j \leq n)$.

Also, please note that $w_{ij} = w_{ji}$.

The degree function $d: V \rightarrow R^+$ is

$$d_i = \sum_{j \sim i} w_{ij}, \tag{1}$$

where $j \sim i$ is the set of vertices adjacent with i.

Define $D = diag(d_1, d_2, ..., d_n)$.

The inner product on the function space R^V is

$$< f, g >_V = \sum_{i \in V} f_i g_i \tag{2}$$

Also define an inner product on the space of functions R^E on the edges

$$< F, G >_E = \sum_{(i,j) \in E} F_{ij} G_{ij} \tag{3}$$

Here let $H(V) = (R^V, <.,.>_V)$ and $H(E) = (R^E, <.,.>_E)$ be the Hilbert space real-valued functions defined on the vertices of the graph G and the Hilbert space of real-valued functions defined in the edges of G respectively.

3   Gradient and Divergence Operators

We define the gradient operator $d: H(V) \rightarrow H(E)$ to be

$$(df)_{ij} = \sqrt{w_{ij}}(f_j - f_i), \tag{4}$$

where $f: V \rightarrow R$ be a function of $H(V)$.

We define the divergence operator $div: H(E) \rightarrow H(V)$ to be

$$< df, F >_{H(E)} = < f, -divF >_{H(V)}, \tag{5}$$

where $f \in H(V), F \in H(E)$

Next, we need to prove that

$$(divF)_j = \sum_{i \sim j} \sqrt{w_{ij}} \, (F_{ji} - F_{ij})$$

Proof:

$$< df, F >= \sum_{(i,j) \in E} df_{ij} F_{ij}$$

$$= \sum_{(i,j) \in E} \sqrt{w_{ij}} \, (f_j - f_i) F_{ij}$$

$$= \sum_{(i,j) \in E} \sqrt{w_{ij}} \, f_j F_{ij} - \sum_{(i,j) \in E} \sqrt{w_{ij}} \, f_i F_{ij}$$

$$= \sum_{k \in V} \sum_{i \sim k} \sqrt{w_{ik}} \, f_k F_{ik} - \sum_{k \in V} \sum_{j \sim k} \sqrt{w_{kj}} \, f_k F_{kj}$$

$$= \sum_{k \in V} f_k \left(\sum_{i \sim k} \sqrt{w_{ik}} \, F_{ik} - \sum_{i \sim k} \sqrt{w_{ki}} \, F_{ki} \right)$$

$$= \sum_{k \in V} f_k \sum_{i \sim k} \sqrt{w_{ik}} \, (F_{ik} - F_{ki})$$

Thus, we have

$$(divF)_j = \sum_{i \sim j} \sqrt{w_{ij}} \, (F_{ji} - F_{ij}) \tag{6}$$

4   Laplace Operator

We define the Laplace operator $\Delta: H(V) \to H(V)$ to be

$$\Delta f = -\frac{1}{2} div(df) \tag{7}$$

Next, we compute

$$(\Delta f)_j = \frac{1}{2} \sum_{i \sim j} \sqrt{w_{ij}} \, ((df)_{ij} - (df)_{ji})$$

$$= \frac{1}{2} \sum_{i \sim j} \sqrt{w_{ij}} \, (\sqrt{w_{ij}}(f_j - f_i) - \sqrt{w_{ij}}(f_i - f_j))$$

$$= \sum_{i \sim j} w_{ij} \, (f_j - f_i)$$

$$= \sum_{i \sim j} w_{ij} f_j - \sum_{i \sim j} w_{ij} f_i$$

$$= d_j f_j - \sum_{i \sim j} w_{ij} f_i$$

Thus, we have

$$(\Delta f)_j = d_j f_j - \sum_{i \sim j} w_{ij} f_i \tag{8}$$

The graph Laplacian is a linear operator. Furthermore, the graph Laplacian is self-adjoint and positive semi-definite.

Let $S_2(f) = < \Delta f, f >$, we have the following **theorem 1**

$$D_f S_2 = 2\Delta f \tag{9}$$

The proof of the above theorem can be found from [15, 16].

5   Curvature Operator

We define the curvature operator $\kappa: H(V) \to H(V)$ to be

$$\kappa f = -\frac{1}{2} div \left(\frac{df}{\|df\|}\right) \tag{10}$$

Next, we compute

$$(\kappa f)_j = \frac{1}{2} \sum_{i \sim j} \sqrt{w_{ij}} \left(\left(\frac{df}{\|df\|}\right)_{ij} - \left(\frac{df}{\|df\|}\right)_{ji}\right)$$

$$= \frac{1}{2} \sum_{i \sim j} \sqrt{w_{ij}} \left(\frac{1}{\|d_i f\|} \sqrt{w_{ij}}(f_j - f_i) - \frac{1}{\|d_j f\|} \sqrt{w_{ij}}(f_i - f_j)\right)$$

$$= \frac{1}{2} \sum_{i \sim j} w_{ij} \left(\frac{1}{\|d_i f\|} + \frac{1}{\|d_j f\|}\right)(f_j - f_i)$$

Thus, we have

$$(\kappa f)_j = \frac{1}{2} \sum_{i \sim j} w_{ij} \left(\frac{1}{\|d_i f\|} + \frac{1}{\|d_j f\|}\right)(f_j - f_i) \tag{11}$$

From the above formula, we have

$$d_i f = ((df)_{ij}: j \sim i)^T \tag{12}$$

The local variation of *f* at *i* is defined to be

$$\|d_i f\| = \sqrt{\sum_{j \sim i}(df)_{ij}^2} = \sqrt{\sum_{j \sim i} w_{ij}(f_j - f_i)^2} \tag{13}$$

To avoid the zero denominators in (11), the local variation of *f* at *i* is defined to be

$$\|d_i f\| = \sqrt{\sum_{j \sim i}(df)_{ij}^2 + \epsilon}, \tag{14}$$

where $\epsilon = 10^{-10}$.

The graph curvature is a non-linear operator.

Let $S_1(f) = \sum_i \|d_i f\|$, we have the following **theorem 2**

$$D_f S_1 = \kappa f \tag{15}$$

The proof of the above theorem can be found from [15, 16].

6   p-Laplace Operator

We define the p-Laplace operator $\Delta_p : H(V) \to H(V)$ to be

$$\Delta_p f = -\frac{1}{2} div(\|df\|^{p-2} df) \tag{16}$$

Clearly, $\Delta_1 = \kappa$ and $\Delta_2 = \Delta$. Next, we compute

$$\left(\Delta_p f\right)_j = \frac{1}{2}\sum_{i \sim j} \sqrt{w_{ij}} \left(\|df\|^{p-2} df_{ij} - \|df\|^{p-2} df_{ji}\right)$$

$$= \frac{1}{2}\sum_{i \sim j} \sqrt{w_{ij}} \left(\|d_i f\|^{p-2} \sqrt{w_{ij}}(f_j - f_i) - \|d_j f\|^{p-2} \sqrt{w_{ij}}(f_i - f_j)\right)$$

$$= \frac{1}{2}\sum_{i \sim j} w_{ij} \left(\|d_i f\|^{p-2} + \|d_j f\|^{p-2}\right)(f_j - f_i)$$

Thus, we have

$$\left(\Delta_p f\right)_j = \frac{1}{2}\sum_{i \sim j} w_{ij} \left(\|d_i f\|^{p-2} + \|d_j f\|^{p-2}\right)(f_j - f_i) \tag{17}$$

Let $S_p(f) = \frac{1}{p}\sum_i \|d_i f\|^p$, we have the following **theorem 3**

$$D_f S_p = p\Delta_p f \tag{18}$$

The proof of the above theorem can be found from [15, 16].

7   Discrete Regularization on Graphs and Protein Function Classification Problems

Given a protein network $G=(V,E)$. V is the set of all proteins in the network and E is the set of all possible interactions between these proteins. Let y denote the initial function in $H(V)$. y_i can be defined as follows

$$y_i = \begin{cases} 1 \; if \; protein \; i \; is \; the \; query \\ 0 \; if \; protein \; i \; is \; not \; the \; query \end{cases}$$

Our goal is to look for an estimated function f in $H(V)$ such that f is not only smooth on G but also close enough to an initial function y. Then each protein i is ranked as value of f_i. This concept can be formulated as the following optimization problem

$$argmin_{f \in H(V)}\{S_p(f) + \tfrac{\mu}{2}\|f - y\|^2\} \tag{19}$$

The first term in (19) is the smoothness term. The second term is the fitting term. A positive parameter μ captures the trade-off between these two competing terms.

7.1   2-smoothness

When $p=2$, the optimization problem (19) is

$$argmin_{f \in H(V)}\{\tfrac{1}{2}\Sigma_i\|d_if\|^2 + \tfrac{\mu}{2}\|f - y\|^2\} \tag{20}$$

By theorem 1, we have

Theorem 4: The solution of (20) satisfies

$$\Delta f + \mu(f - y) = 0 \tag{21}$$

Since Δ is a linear operator, the closed form solution of (21) is

$$f = \mu(\Delta + \mu I)^{-1}y, \tag{22}$$

Where I is the identity operator and $\Delta = D - W$. (22) is the algorithm proposed by [1, 10].

7.2   1-smoothness

When $p=1$, the optimization problem (19) is

$$argmin_{f \in H(V)}\{\Sigma_i\|d_if\| + \tfrac{\mu}{2}\|f - y\|^2\}, \tag{23}$$

By theorem 2, we have

Theorem 5: The solution of (23) satisfies

$$\kappa f + \mu(f - y) = 0, \tag{24}$$

The curvature κ is a non-linear operator; hence we do not have the closed form solution of equation (24). Thus, we have to construct iterative algorithm to obtain the solution. From (24), we have

$$\tfrac{1}{2}\Sigma_{i \sim j} w_{ij}\left(\tfrac{1}{\|d_if\|} + \tfrac{1}{\|d_jf\|}\right)(f_j - f_i) + \mu(f_j - y_j) = 0 \tag{25}$$

Define the function $m: E \to R$ by

$$m_{ij} = \tfrac{1}{2}w_{ij}(\tfrac{1}{\|d_if\|} + \tfrac{1}{\|d_jf\|}) \tag{26}$$

Then (25)

$$\sum_{i \sim j} m_{ij}(f_j - f_i) + \mu(f_j - y_j) = 0$$

can be transformed into

$$\left(\textstyle\sum_{i\sim j} m_{ij} + \mu\right)f_j = \textstyle\sum_{i\sim j} m_{ij}\, f_i + \mu y_j \tag{27}$$

Define the function $p: E \to R$ by

$$p_{ij} = \begin{cases} \dfrac{m_{ij}}{\sum_{i\sim j} m_{ij}+\mu} & if\ i \neq j \\[2mm] \dfrac{\mu}{\sum_{i\sim j} m_{ij}+\mu} & if\ i = j \end{cases} \tag{28}$$

Then

$$f_j = \textstyle\sum_{i\sim j} p_{ij} f_i + p_{jj} y_j \tag{29}$$

Thus we can consider the iteration

$$f_j^{(t+1)} = \textstyle\sum_{i\sim j} p_{ij}^{(t)} f_i^{(t)} + p_{jj}^{(t)} y_j \ \forall j \in V$$

to obtain the solution of (23).

7.3   p-smoothness

For any number p, the optimization problem (19) is

$$argmin_{f\in H(V)}\{\tfrac{1}{p}\textstyle\sum_i \|d_i f\|^p + \tfrac{\mu}{2}\|f - y\|^2\}, \tag{30}$$

By theorem 3, we have

Theorem 6: The solution of (30) satisfies

$$\Delta_p f + \mu(f - y) = 0, \tag{31}$$

The *p-Laplace* operator is a non-linear operator; hence we do not have the closed form solution of equation (31). Thus, we have to construct iterative algorithm to obtain the solution. From (31), we have

$$\tfrac{1}{2}\textstyle\sum_{i\sim j} w_{ij}\left(\|d_i f\|^{p-2} + \|d_j f\|^{p-2}\right)(f_j - f_i) + \mu(f_j - y_j) = 0 \tag{32}$$

Define the function $m: E \to R$ by

$$m_{ij} = \tfrac{1}{2} w_{ij}(\|d_i f\|^{p-2} + \|d_j f\|^{p-2}) \tag{33}$$

Then equation (32) which is

$$\sum_{i\sim j} m_{ij}\left(f_j - f_i\right) + \mu(f_j - y_j) = 0$$

can be transformed into

$$\left(\textstyle\sum_{i\sim j} m_{ij} + \mu\right)f_j = \textstyle\sum_{i\sim j} m_{ij}\, f_i + \mu y_j \tag{34}$$

Define the function $p: E \to R$ by

$$p_{ij} = \begin{cases} \frac{m_{ij}}{\sum_{i \sim j} m_{ij} + \mu} & \text{if } i \neq j \\ \frac{\mu}{\sum_{i \sim j} m_{ij} + \mu} & \text{if } i = j \end{cases} \tag{35}$$

Then

$$f_j = \sum_{i \sim j} p_{ij} f_i + p_{jj} y_j \tag{36}$$

Thus we can consider the iteration

$$f_j^{(t+1)} = \sum_{i \sim j} p_{ij}^{(t)} f_i^{(t)} + p_{jj}^{(t)} y_j \; \forall j \in V$$

to obtain the solution of (30).

8   Experiments and Results

Datasets
In this paper, we use the dataset available from [12, 13] and references therein. This dataset contains the probabilistic yeast network containing 3112 genes and 25113 undirected interactions (i.e. edges). This probabilistic yeast network is called ProNet. In order to evaluate the performance of the un-normalized graph p-Laplacian based ranking algorithms, we used three protein complexes from the MIPS databases. Table 1 shows the three protein complexes used in our experiments.

Table 1. Description of three protein complexes

Complex ID	Number of genes in complex	List of genes in the complex
1	3	ydr148c, yfl018c, yil125w
2	4	ybr126c, ydr074w, yml100w, ymr261c
4	3	ybr195c, yml102w, ypr018w

Experiments
In this section, we experiment with the above proposed un-normalized graph p-Laplacian ranking methods with *p=1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9* and the current state of the art method (i.e. the un-normalized graph Laplacian based ranking method *p=2*) in terms of accuracy performance measure. All experiments were implemented in Matlab 6.5 on virtual machine. The leave-one-out testing strategy is used to compute the accuracy performance measures of all methods used in this paper. For each of the complexes, one member gene is left out and the remaining genes are used as the core complex in the membership query. Effective ranking methods should report the left out gene in top k ranks. The parameter μ is set to 1. The accuracy performance measures of the above proposed methods and the current state of the art method is given in the following table 2.

Table 2. The comparison of accuracies of proposed methods with different p-values

Top k ranks		k=1	k=5	k=10	k=20
Accuracy Performance Measures (%)	p=1	10	60	90	100
	p=1.1	10	60	90	100
	p=1.2	20	70	100	100
	p=1.3	30	70	100	100
	p=1.4	30	80	100	100
	p=1.5	30	80	100	100
	p=1.6	30	80	100	100
	p=1.7	30	80	100	100
	p=1.8	40	80	100	100
	p=1.9	40	80	100	100
	p=2 (i.e. the current state of the art method)	40	70	100	100

The results from the above table shows that the un-normalized graph p-Laplacian ranking methods are at least as good as the current state of the art method (*p=2*) but often lead to better ranking accuracy performance measures.

9   Conclusions

We have developed the detailed regularization frameworks for the un-normalized graph p-Laplacian ranking methods applying to complex membership determination problem. Experiments show that the un-normalized graph p-Laplacian ranking methods are at least as good as the current state of the art method (i.e. *p=2*) but often lead to significant better ranking accuracy performance measures.

Moreover, these un-normalized graph p-Laplacian ranking methods can not only be used in complex membership determination problem but also in biomarker discovery problem in cancer classification. In specific, given a set of genes (i.e. the queries) involved in a specific disease (for e.g. leukemia), these methods can also be used to find more genes involved in the same disease by ranking genes in gene co-expression network (derived from gene expression data) or the protein-protein interaction network or the integrated network of them. The genes with the highest rank then will be selected and then checked by biologist experts to see if the extended genes in fact are involved in the same disease. This problem is called biomarker discovery in cancer classification.

Recently, to the best of my knowledge, the un-normalized directed graph p-Laplacian based ranking methods have not yet been developed and applied to any practical problems. This method is worth investigated because of its difficult nature and its close connection to partial differential equation on directed graph field.

References

1. Tran, L.: Application of three graph Laplacian based semi-supervised learning methods to protein function prediction problem. CoRR abs/1211.4289 (2012)
2. Tran, L.: Hypergraph and protein function prediction with gene ex-pression data CoRR abs/1212.0388 (2012)
3. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y.: A com-prehensive two-hybrid analysis to explore the yeast protein interac-tome. Proc. Natl. Acad. Sci. 98, 4569–4574 (2001)
4. Uetz, P., Cagney, G., Mansfield, T.A., Judson, R., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P.: A comprehensive analysis of protein-protein interactions in saccharomyces cerevisiae. Nature 403, 623–627 (2000)
5. Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M.: Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147 (2002)
6. Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K.: Systematic identification of protein complexes in saccharomyces cerevisiae by mass spectrometry. Nature 415, 180–183 (2002)
7. Tsuda K., Shin H.H., and Schoelkopf B. Fast protein classification with multiple networks. Bioinformatics (ECCB 2005) 21(suppl. 2), ii59–ii65 (2005)
8. Lanckriet, G.R.G., Deng, M., Cristianini, N., Jordan, M.I., Noble, W.S.: Kernel-based data fusion and its application to protein function prediction in yeast. In: Pacific Symposium on Biocomputing (PSB) (2004)
9. Letovsky, S., Kasif, S.: Predicting protein function from pro-tein/protein interaction data: a probabilistic approach Bioinformatics 19, i197–i204 (2003)
10. Tsuda, K., Noble, W.S.: Learning kernels from biological networks by maximizing entropy. Bioinformatics 20(S1), i326–i333 (2004)
11. Tran, L.: The un-normalized graph p-Laplacian based semi-supervised learning method and protein function prediction problem. In: The Fifth International Conference on Knowledge Systems and Engineer (2013)
12. Asthana, S., King, O.D., Gibbons, F.D., Roth, F.P.: Predicting protein complex member-ship using probabilistic network reliability. Genome Research 14, 1170–1175 (2004)
13. Can, T., Camoglu, O., Singh, A.K.: Analysis of protein-protein interaction networks using random walks. In: Proceedings of the 5th ACM SIGKDD Workshop on Data Mining in Bioinformatics, Chicago (August 2005)
14. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. Comput-er Networks and ISDN Systems 30, 107–117 (1998)
15. Tran, L., Tran, L.: Un-normalized graph p-Laplacian semi-supervised learning method applied to cancer classification problem. In: The Second International Conference on Intelligent and Automation Systems (2014)
16. Zhou, D., Schölkopf, B.: Discrete Regularization. In: Chapelle, O., Schölkopf, B., Zien, A. (eds.) Semi-Supervised Learning, pp. 221–232. MIT Press, Cambridge (2006)