CLASSIFICATION AND PREDICTION OF DISEASE-RELATED GENES USING BIOLOGICAL NETWORK-BASED ALGORITHMS

Thi Hue Bui

University of Economics – Technology for Industries
Nguyen Tuong Lan

Vietnam Academy of Social Sciences

Le Nguyen Ha Vy
Foreign Language Specialized Schoolemail: lenguyenhavy4466@gmail.com

Corresponding author: Thi Hue Bui, *University of Economics – Technology for Industries*

Email: bthue@uneti.edu.vn

Abstract:

The goal of the study was to evaluate the overall methods of classifying and predicting disease-related genes using algorithms based on biological networks. The study conducts a review of material from SSCI and SCIE articles on related topics, then conducts analysis synthesis and finds new points for future research is the segmentation and prediction of disease-related genes using algorithms based on biological networks.

Keywords: Classification, gene prediction, biological network

DOI Number: 10.48047/NQ.2022.20.20.NQ109218 **NeuroQuantology2022;20(20): 2170-2182**

1. Introduction

Predicting disease-causing genes is one of the key goals in biomedical research. There are now quite a few methods built to predict the genes associated with certain diseases. However, due to the complex relationship between genes and diseases, many of the genes that cause some genetic diseases have yet to be discovered. The problem of gene classification to find disease genes is one of the problems that many scientists are interested in researching. To find a good method with the goal of predicting disease-causing genes with high performance, we investigated a number of existing gene classification methods based on the biological network, then proposed a prediction method using the Boolean Network model. In the biological network, defects caused by mutations in genes/proteins can cause a certain disease in humans.

Identifying disease-causing genes is an important problem in biomedicine and molecular biology. To predict disease genes, several methods have been proposed (Kann,

2010). In the past, the identification of disease-causing genes was carried out mainly by biological experiments. This method is performed for hundreds of candidate genes located on a suspicious chromosome region, so it requires a lot of time and a very high cost. Gene classification is the use of computational methods to arrange candidate genes so that genes that are likely to be associated with the disease receive higher rankings. After classification, a small group of genes with high rankings are then selected for experimental testing. The proposed candidate gene classification methods can be divided into 3 main directions: i) Based on functional label marking; ii) network-based and iii) machine learning-based. In particular, methods based on functional labeling classify candidate genes by measuring the similarity of each candidate gene to a set of known pathogenic genes based on profiles constructed from multiple data sources (Aerts, 2006). Therefore, these methods mainly focus on integrating various biological datasets to get a more accurate analogy to cover the entire human genome. Besides methods based on functional labeling, machine learning-based methods with binary layering to identify corresponding disease genes have also been studied.

In the early days, machine learning-based studies often approached disease gene prediction as a binary layering problem. Several binary layering techniques have been proposed for this problem such as: decision tree (Adie, 2005), k-nearest neighbor (Li, 2006), Naïve Baysian classification (Calvo, 2006), artificial neural networks (Sun, 2009) and vector machines (Keerthikumar, 2009). In these studies, the learning patterns included positive and negative coaching patterns. In particular, the positive training sample is built from known pathogenic genes, the negative training sample is an unidentified gene associated with the disease. This is the limitation of binary layering solutions to the disease gene prediction problem because negative training is not really genes that are not associated with the disease. However, building this dataset is nearly impossible in biomedical studies because in biomedical cases where no link is observed does not mean that the link does not exist. Thus, to reduce this uncertainty of previous methods, a semi-supervised approach has been proposed for the problem, in which the classifier is learned from both: labeled (e.g., known disease genes) and unlabeled (e.g., unknown genes) data. However, negative patterns still have to be identified in these studies. To overcome the limitations of both methods, network-based methods for identifying disease genes have been proposed (Wang, 2011). These methods are mainly based on biological networks such as protein interaction networks, which are quite commonly used due to the increasingly complete and diverse protein/gene interaction data. In addition, this approach is superior to the previous two approaches because it is based on the "disease module" principle (i.e. genes/proteins that bind to the

eISSN 1303-5150

same disease or similar diseases tend to be located in close proximity to each other in the gene/protein interaction network). Furthermore, this network-based approach targets the nature of the disease gene prediction problem, classifying rather than classifying candidate genes (e.g., labeling a candidate gene as a disease gene or not) as machine learning-based methods do.

2. Literature review

2.1. Overview of disease-related gene classification and prediction

Cells are the basic building blocks of life, they provide structure to the body, make nutrients from food, convert nutrients into energy and perform specialized functions. In addition, the cell contains the genetic elements of living organisms and can replicate itself. Each cell consists of many components, each of which is specialized structures with different functions. Some components called organelles perform certain tasks inside the cell.

DNA (deoxyribonucleic acid) is a genetic element in most living organisms. Nearly every cell in a living organism has the same DNA. Most of the DNA is in the cell nucleus. Information in DNA is stored as a coding of four bases: adenine (A), cytosine (C), guanine (G) and thymine (T). DNA bases are paired together: A with T and C with G, forming base pairs. Each base pair binds to a sugar molecule and a phosphate molecule called a nucleotide. The nucleotides are arranged in two long strands that twist evenly around an imaginary axis forming a double helix.

An important characteristic of DNA is that it can duplicate itself, making multiple copies of itself. This is essential when the cell divides, then each new cell needs an exact copy of the DNA present in the old cell.

Genes are considered a base unit of genetic phenomena at the molecular level. Each gene is a piece of DNA that contains information that regulates the composition of functional molecules such as RNA (ribonucleic acid) and proteins through chemical reactions in the organism. Proteins interact with each other to perform the functions of the body, so it is possible to think of proteins as functional forms of genes.

In humans, gene length can vary from a few hundred to more than 2 million base DNA. The Human Genome Project estimates that humans have between 20,000 and 25,000 genes. Each individual has two copies of a gene, one from the father and the other from the mother. Most genes are the same in everyone, but there are a small number of genes (less than 1% of the total) that differ between each person. Alleles are forms of the same gene with slight differences in DNA sequence. These differences make up the characteristics of each person.

The process of controlling protein synthesis from genes consists of two main steps: transcription and translation.

Transcription: information stored in the DNA gene is transferred to a similar

molecule, RNA (ribonucleic acid) in the cell nucleus. Both RNA and DNA are formed from sequences of nucleotides, but they differ in chemical properties. The type of protein that contains protein-forming information is called messenger RNA (mRNA), because they carry information, or messages sent from the nucleus to the cytoplasm.

Translation: in the cytoplasm, the mRNAs interact with a special complex called a ribosome (base mRNA sequencer). Each unit of code usually encodes a separate amino acid (amino acids are the microstructure that make up proteins). Another type of RNA called transport RNA (tRNA) carries specific amino acids to ribosomes to assemble into proteins. The amino acids in turn are assembled based on information from the mRNA until the ribosome encounters an end code (stop codon, a 3-base sequence that encodes no amino acids).

2.2. Gene classification and resolution approaches

Gene classification is the use of computational methods to rank genes according to their relevance to the disease under consideration. The high-ranking genes are then confirmed by biological experiments to verify their association with the disease.

The concept of gene classification was first introduced in 2002 by Perez-Iratxeta et al. (2002). In the paper, Perez-Iratxeta et al. describe the first computational approach to solving this problem. Since then, many gene classification methods using different strategies, algorithms, and data sources have been developed.

The gene classification problem can be stated as follows: For a disease D, a genome C is a candidate to consider and training data T. After entering the data and calculating, the method calculates the score for each candidate gene, the genes with a high score are the genes that are most likely to be related to the disease.

The purpose of gene classification is to provide biomedical researchers with initial hints about genes that may be associated with the disease, helping to narrow down the list of candidate genes to predict and the mechanisms involved in the disease. These contributions are essential for identifying new genes associated with disease, especially for complex diseases.

2.3. Methods based on biological networks

To overcome the limitations of these two approaches, methods for predicting disease-related genes based on biological networks have been developed and show better results than methods based on functional annotation and machine learning.

These methods mainly use biological networks to carry out the classification process. Biological networks are built on different biomedical data, so are not limited by coverage as functional annotation data sources. In addition, these methods can be thought of as semi-supervised learning techniques using unlabeled data, and the

resulting ranking of candidate genes is estimated based on their relative association with genes associated with known diseases.

Classification methods based on biological networks can be divided into local method and overall method. The local method uses local information of seed genes. Basically, these methods classify based on proximity through direct testing of neighbors of seed genes. In other words, neighboring genes of genes associated with known diseases will be assigned higher scores than the rest of the genes in the network. Meanwhile, holistic methods spread disease-related signals over the network to provide scores of connections and the impact of seed genes on the remaining genes. Network-based methods using recent propagation algorithms have proved superior to local methods.

In summary, network-based approaches exploit the interconnected properties of a biological network to calculate a link score between candidate genes and genes associated with known disease. Although many methods have been proposed, there are still some limitations such as: most of the biological networks used are based on physical interactions that have not covered the entire human genome; methods that have not yet incorporated weighted interactions; In local methods, scores are calculated only for genes that interact directly with disease-related genes but ignore genes with other indirect interactions.

3. Databases and biological networks

3.1. Databases

In the field of disease-related gene classification and prediction, various data sources have been successfully exploited to predict the disease association of candidate genes. The data source used plays an important role, which is directly related to quality and predictability. Different data sources can be considered as different views of the same object as genes. A single data source may not guarantee the necessary accuracy, but additional data sources need to be used. This section introduces several data sources that can be used for disease-related gene classification and prediction.

Mining data from the medical literature is one of the first approaches to predicting the genes associated with the disease. To date, there have been millions of documents describing the association between diseases and genes provided by the PubMed database and thousands of disease-gene phenotypes contained in OMIM profiles. These data can be mined using specialized text mining and natural language processing techniques. However, these medical sources are not organized systematically and each concept can be described in a variety of approaches. Therefore, a number of tools for automatic processing of medical records and texts based on controlled vocabulary such as MeSH, UMLS or eVOC have been proposed to build organizational charts of medical literature, for the purpose of data mining.

Gene function exegesis in a broad sense includes not only biological processes and molecular functions but also metabolism and pathway signals. These are accurate,

valuable and widely used bases for predicting genes associated with disease. Relying on gene function annotations is a reasonable way to predict candidate genes, but it is only suitable for genes with clear trait expression. Moreover, in fact only a small percentage of gene function annotations are determined experimentally.

One of the most widely used functional annotation databases is gene ontology (GO). This is a controlled ontology that includes a standard set of words and phrases used to index and query information. In addition to defining standards, GO also defines relationships between terms making it a structured vocabulary.

In addition, many other gene function annotation databases are also of interest, used in predicting disease-related genes and annotating genes/proteins such as KEGG, MeSH, UMLS, eVOC, MPO [108] and more recently, the HPO human phenotype are typical examples. Particularly noteworthy is the HPO database, a controlled ontology of more than 8000 standards representing single anomalous phenotypes. In addition, HPO provides annotations for all items in OMIM with standards, thereby helping to standardize concepts. This is a very important source of data describing diseases.

3.2. Biological network

Biological networks include protein interaction networks, metabolic networks, gene regulation networks, genetic networks, and co-expression networks. These networks have now been built, tested and developed continuously to characterize material or functional interactions between biological molecules. Disturbances in these networks can create special phenotypes in monogenic, polygenic, and cancers. In a biological network, when a node mutates due to different types of etiology, it directly affects neighboring nodes. Disease A is formed as a result of edge removal, and disease B is formed as a result of node removal. These two diseases are not necessarily the same, but may share similar phenotypes. Deciphering the properties of biological networks will provide deeper insights into the relationships between genotypes and complex phenotypes.

Biological networks can be divided into two types: Interaction networks (metabolic networks, protein interaction networks and gene regulation networks, ...) represent material, biochemical interactions between molecules. Functional networks (transcription, phenotypic shaping, genetic interaction networks, ...) exhibit functional relationships or similarity between genes and gene products.––

Biological networks are usually represented by scalar or directional graphs with nodes being molecules and edges representing material or functional bonds between them. Protein interaction networks are usually scalar graphs whose edges represent material bonds between proteins, whereas gene regulatory networks are usually directed graphs, and edges that represent the physical link between one node (which is

the transcription element) and other nodes (which are DNA regulatory elements).

In metabolic networks, nodes are metabolites and edges (directional or scalar) are activators or enzymes that catalyze reactions to convert one node into another. Functional networks that shape transcription, phenotypic shaping, and genetic interaction networks all have nodes representing genes but edges that represent a high degree of co-representation of correlations between phenotypic shaping and corresponding known genetic interactions. Information and approaches to building these networks are presented in.

The gene/protein interaction network is a data source commonly used to predict disease-related genes, and each physical interaction between genes/proteins produces a basic function. Therefore, when a material interaction between genes / proteins changes, it leads to a disease phenotype. This fact was demonstrated by Brunner et al. (2010) when studying the complex bonds between proteins and disease.

Gene/protein interaction networks are often built on the method of collecting interactions from experiments. The two main technologies that have been developed and successfully applied in generating a large number of human protein interactions are: the method of developing a Y2H high-throughput system for selective direct binary interactions between protein pairs and the method of high-performance homomorphism cleaning by mass spectrometry to identify the above protein complexes. human being.

4. Methods for classifying and predicting disease-related genes based on biological networks

4.1. The method is based on how close the genes are/proteins

Most current methods of classifying and predicting disease-related genes are based on the "disease modulus" hypothesis, i.e. genes associated with the same disease are often juxtaposed on interacting networks. These genes tend to be involved in the same biological pathway and have a similar impact on disease phenotypes. These classification methods use different scoring strategies but are essentially measuring the distance between the gene associated with the known disease and the candidate gene on the protein interaction network. The measurements fall into three main categories: local distance measurement, overall distance measurement, and graph partition method for calculating how close each pair of proteins is in a network.

The simplest method of classifying and predicting disease-related genes is to consider whether two proteins in a network are interconnected (direct neighbor counting methods). For any pair of proteins, if they are directly connected by an edge, a value of 1 is assigned, whereas a value of 0 is assigned. If the candidate gene is directly connected to many disease-related genes, it is likely to be disease-related genes.

When two proteins join the biological pathway without physical or functional interaction, some studies have determined the distance between them in gene/protein interaction networks using the shortest path method. The genes involved in the known disease are treated as seed nodes and calculate the shortest path length between one node and another on the network. A node close to the seed node will have a higher score like a gene associated with the candidate disease. Krauthammer et al. (2015) evaluated methods for predicting genes associated with Alzheimer's disease and showed that the genes associated with the disease predicted by the shortest path method had the same results as the experimental determination method. However, the shortest path between two proteins cannot adequately represent how close they are but also considers the network structure around the two proteins. For example, two proteins connected by a center or between them with multiple pathways will exhibit different levels of proximity between them.

Overall distance measurements cover these factors by assuming the probability of a protein diffusing along the bonds of the gene/protein interaction network is equal. Kohler et al. (2017) conducted trials on 110 disease groups including 783 classified disease genes. The results showed that overall distance measurements (nuclear diffusion, random step with rollback) performed better than local distance measurements (direct neighbor, shortest path length).

When comparing the performance of disease gene classification and prediction methods use different distance measurements such as: direct neighbors, random steps with backtracking, propagation flow, unsupervised graph partitioning, Markov clustering, and semi-supervised graph partitioning. Navlakha and Kingsford (2008) also showed that the random step has a rollback for the best performance in terms of accuracy and coverage, followed by the propagation flow, clustering, and direct neighbor method. Each method makes its own new predictions, and only a few are incorrect. Therefore, a method that logically combines the proposed proximity measurements will give the best classification efficiency due to its ability to recognize the different properties of the interactive network.

4.2. Large-scale integration of genomic data

Multiple data integration methods have been proposed to classify and predict disease-related genes. These methods are based on the assumption that disease-related genes will share common characteristics in the data on gene annotation, gene semantics, gene expression, gene sequencing, and protein domains. They are also capable of participating in biological pathways and functionally similar pathways. When better predictive efficiency can be achieved by combining multiple data sources, the

question is how heterogeneous data can be integrated.

The information about disease-related gene characteristics that are often considered are functional annotations, microarray expression, EST expression (which is a small part of a DNA sequence generated from one or two ends of a gene), medical literature, protein domains, protein interactions, metabolic pathways, Cis regulatory complexes, transcription motifs, sequence data, and other potential data sources are supplemented by the user. An overall ranking by combining the ratings of each trait is used to classify candidate genes. Endeavour (2001) is a classification method that uses a synthetic genomic dataset, integrating 12 traits to rank candidate genes based on their similarity to known disease-related genes based on each trait. The results show that the performance of the method using all data sources is much better than the method of using only separate data sources.

Functional linkage networks were also proposed for the classification of candidate genes by consolidating information from different data sources using a Bayesian subclass (2009). The Prioritizer Method (2012) built four types of functional networks by combining different data sources of gene semantics, gene expression, and protein interactions. Experiment with artificially sensitive loci containing between 50 -150 genes. In the 50th step, neighboring genes associated with the disease were created.

Linghu et al. (2012) conducted extensive genome classification by building a functional network linking evidence-weighted of 21657 genes based on 16 data sources. The functional bonds of each pair of genes at each trait are integrated into a single functional link network, weighted by the sum of functional bonds, using a Naive Bayesian subclass. For any given disease, scores for classifying candidate genes are assigned based on the total weighting of links to genes associated with known diseases. The algorithm has been tested to classify and predict disease-related genes for 110 diseases and has shown remarkable efficacy.

4.3. Methods based on the integration of phenotypic information

It has been proven that diseases with homologous phenotypes often share a set of genes that are potentially risky or functionally related. This observation has been used to build disease networks, in which two diseases are interconnected if they share at least one common gene. Several different methods have been developed to calculate similarities between diseases. Rzhetsky et al. (2015) conducted a study on 1.5 million patient records and 161 body dysfunctions using statistical modeling and showed that disease phenotypes form a highly connected network with strong correlations between each pair. A homologous disease phenotypic network has been built by connecting diseases based on their co-occurrence in a large number of patients. Based on the assumption that

overlapping disease phenotypes share potentially functionally similar genes, it is the purpose of closely incorporating these homologous phenotypic profiles in the candidate gene classification. Some studies suggest that the integration of disease phenotype networks and protein interaction networks gives better results than other approaches in the task of classifying and predicting disease genes. Wu et al. (2013) built the CIPHER algorithm using a simple linear regression method to model correlations between phenotypic homologous profiles and profiling closely spaced genes in the protein interaction network. The basic assumption of the algorithm is that the phenotypic similarity between the two diseases can be explained by the proximity of disease-related genes in the interacting network. Experimental results show that their disease-related candidate gene predictions are reliable in both associative and genome-wide genomic types. They also demonstrated that CIPHER's performance is comparable to Endeavor (2008), an integrated approach that uses more than 10 large-scale genomic data as mentioned above.

A similar approach is PRINCE developed by Vanunu et al. (2016). In their study, they calculated the association between a D query disease with a p protein and a known disease gene of a D0 disease that is phenotypically similar to D. Protein binding – this disease represents the association of the p protein with D disease and is used as a priori knowledge to construct classification functions. PRINCE has been shown to successfully predict not only genes, but also combinations of proteins associated with a disease.

Li and Patra (2016) built a mixed network by integrating protein interaction networks and phenotypic networks based on gene-disease relationships in OMIM. The authors developed a new algorithm by extending the random-step algorithm with rollback to heterogeneous networks. In this case, the random-step algorithm with rollback is no longer limited to the gene network, but is also allowed to pass to the phenotypic network to classify genes and phenotypes simultaneously. The inclusion of an improved phenotypic network and algorithm in both gene networks and phenotypes has greatly enhanced the disease-related gene classification effect.

4.4. Method of building disease modules

In addition to algorithms for classifying disease-related candidate genes for general diseases, significant efforts have also been made towards detecting disease-related genes for single, specific diseases by building disease modules. Network components in topological modules are thought to be functionally related, and a malfunction of a module will lead to a specific disease. Information about known disease-related genes is collected to build disease modules or subnets in which members will share homologous functions, expression patterns, or metabolic pathways. The concept of disease modules has been used in the study of many different diseases

such as cancer, type 2 diabetes, obesity, asthma, neurology. A modular approach to disease, especially for diseases that have not been studied much, often requires significant empirical efforts to identify interactions for the formulation of underlying disease modules. Liu et al. (2018) used a network-based approach and identified an insulin signaling module and a network of nuclear receptors that play an important role in type 2 diabetes. Along with a subnet of protein interactions, the authors proposed basic biological processes for this disorder. In a study of obesity, tissue-tissue co-expression networks in the hypothalamus, liver, or adipose tissue were built and allowed to identify specific disease-related genes. The study found that many genes in subnets were involved in biological functions associated with obesity such as circadian rhythm disturbances, energy imbalances, stress responses, or immune responses. Another approach was developed to classify and predict disease-specific genes by building disease subnets with specific conditions. Specific disease-related genes, such as those with distinct expression identified under disease conditions, are then mapped to the overall protein interaction network. Subnets consist of the shortest pathways built with nodes connecting in the shortest pathway between specific disease-related genes. Each node in this subnet is evaluated and assigned a topological score by comparing the number of shortest paths of pairs of nodes passing through it in this subnet with the number of shortest paths passing through it in the overall network.

4. CONCLUSIONS

In this article, we presented studies of potentially disease-related candidate gene classification methods, then proposed a network-based method for predicting disease-related genes. A review of classification methods such as Cross Validation (CV) is a method used to evaluate machine learning models on a given dataset. There are three commonly used cross-checking methods: Hold-out: As the simplest method, in this method, the data is randomly divided into two sets, one used for training and one used to test the machine learning model. K-fold cross validation: Is the upgrade method of Hold-out. In this method the data is divided into K subsets of the same size and performs K iterations. At each iteration, K-1 subset is used for training and the other subset is used to test machine learning models. The accuracy of the evaluation of this method increases as K increases, but the number of executions also increases accordingly, so in each case, the K value can be selected in accordance with the model to be evaluated. K values are usually chosen as: 3, 5, 10 .... Then we have the following methods: 3-fold cross validation, 5-fold cross validation or 10-fold cross validation. Leave-one-out cross validation (LOOCV): This method is similar to K-fold cross validation but maximizes the number of subsets. Here, K equals the number of elements in the initial dataset, and each subset has exactly 1 element. In the field of classifying and predicting disease-related genes, the LOOCV method is often used to

eISSN 1303-5150

evaluate classification algorithms. According to this method, with each iteration, one known disease-related gene is removed and treated as a normal candidate gene, the remaining disease-related genes are used as training sets as input data for the algorithm. This process is repeated for each gene associated with the known disease, which is the same as the method of cross-checking K times with K being the number of genes associated with the original known disease. As such, disease-related gene classification and prediction and bionetwork-based algorithms have enormous implications in the biomedical and life fields.

BIBLIOGRAPHY

Adie E., R. A. (2005). Speeding disease gene discovery by sequence based candidate. BMCBioinformatics, 6: 55.

Aerts, S., D. L. (2006). Gene prioritization through genomic data fusion. Nature Biotechnology, 24: 537-544.

Albert, R. (2004). Boolean Modeling of GeneticRegulatory Networks. Lecture Notes in Physics., 650: 459-481.

Albert, R. (2004). Boolean Modeling of Genetic Regulatory Networks. Lect. Notes Phys., 650: 459-481.

Amberger, J., C. A. (2009). McKusick's Online Mendelian Inheritance in Man (OMIM®). Nucleic Acids Research, 37: D793-D796.

Calvo, S., M. J. (2006). Systematic identification of human. Nat Genet, 38: 576-582.

Cui Q., E. P. (2009). Protein evolution on a human. BMC Systems Biology, 3: 21.

Dang Vu Tung, D. A. (2015). Pathogenic gene classification uses increased learning in combination with pre-experimental probability. ICT Research, Development and Application Works, Vol. V-1, No. 13 (33).

Duc-Hau Le, Y.-K. K. (2012). A Cytoscape plug-in for random walk-based gene prioritization and biomedical evidence collection. Computational Biology and Chemistry, pp. 17-23.

Faure A., A. N. (2006). Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics, 22: e124- 131.

Hanley JA, M. B. (1982). The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve. Radiology, 143: 29-36.

Kann, M. G. (2010). Advances in translational bioinformatics: computational approaches for the hunting of disease genes. Briefings in Bioinformatics, 11: 96-110.

Kauffman, S., C. P. (2003). Random Boolean network models and the yeast transcriptional. Proceedings of the National Academy of Sciences, 100: 14796- 14799.

Keerthikumar, S., S. B. (2009). Prediction of candidate primary immunodeficiency disease genes using a support vector

eISSN 1303-5150

machine learning approach. DNA Research, 16: 345-351.

Kwon, D.-H. L.-K. (2011). The effects of feedback loops on disease comorbidity in human signaling networks. Bioinformatics, 27: 1113-1120.

Kwon, D.-H. L.-K. (2013). A coherent feedforward loop design principle to sustain robustness of biological networks. Bioinformatics, 29: 630-637.

Kwon, D.-H. Le and Y.-K. (2011). NetDS: a Cytoscape plugin to analyze the robustness of dynamics and feedforward/feedback loop structures of biological networks. Bioinformatics, 27: 2767-2768.

Li, J. X. (2006). Discovering disease-genes by topological features in human protein-protein interaction network. Bioinformatics, 22: 2800- 2805.

LOVASZ, L. (1996). Random walks on graphs: A survey. Combinatorics, Paul Erdos is Eighty, 2: 353-398.

Sun J., J. C. (2009). Functional link artificial neural. In Neural Networks.

Tran Thi Bich Phuong, N. V. (2013). A method of analyzing protein interaction networks to predict cancer-causing genes. Journal of Science of HNUE, 58: 38-46.

Trinh H.-C., D.-H. L.-K. (2014). PANET: A GPUBased Tool for Fast Parallel Analysis of Robustness Dynamics and Feed-Forward/Feedback Loop Structures in Large-Scale Biologica lNetworks. PLoS ONE, 9: e103010.

Derhami, E. K. (2013). Applying reinforcement learning for web pages ranking algorithms. Applied Soft Computing, 13: 1686-1692.

Vanunu O, M. O. (2010). Associating genes and protein complexes with disease via network propagation. PLoSComput Biol, 6: e1000641.

Wang X., N. G. (2011). Network-based methods for human disease gene prediction. Briefings in Functional, 10: 280-293.

eISSN 1303-5150