

Prediction of Functional Associated Proteins Using Topological Similarity in Protein-Protein Interaction Network

Yu-Wei Chiu¹, Hsueh-Fen Juan², Hsuan-Cheng Huang^{1,*}

1. Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan

2. Department of Life Science, National Taiwan University, Taipei, Taiwan

*email: hsuancheng@ym.edu.tw

Introduction

Many recent studies have demonstrated the topology of network offers effective information to predict protein function besides sequence similarity [1]. Since proteins which share more common neighbors are likely to share similar biological characteristics, neighbors counting method has been mostly studied and widely used in protein function prediction. However, considering that proteins share similar functions may also physically perform similar behavior in the biological process, functional associated protein pairs may share similar topology in the interaction network. Therefore, based on assumption that functional similar protein pairs share same network topology, we characterize a model, which can be used to determine topology similarity of protein pairs and also further to predict functional associated proteins.

Methods

The methods used in this study consisted of three steps: (1) Have a start protein node given in the protein interaction network; we then computed the probabilities of every protein in the network randomly walked by chance. Each step of the random walk taken was formalized as a probability matrix, and each element represented the probability of one protein was traversed from the started protein. (2) Until all the protein within the network was traversed, we summed up the probability matrix of each step. In the added matrix, each row showed the probability profiles of every protein traversed. (3) At last, since two different proteins were similarity in topology with similar probability profiles, we computed the Pearson correlation of each row within the added matrix. The values of the correlation profiles were used as the measurement of the similarity in topology of two different proteins within the same interaction network.

Results

After data collected, in statistical analysis, we compared our method with several existing methods applied to *Saccharomyces cerevisiae* protein interaction network, downloaded from Database of Interacting proteins [2]. Within the protein annotation downloaded from Functional Category Database (FunCat) [3], we defined protein pairs that share at least one common functional category within certain level of specificity as functional associated. In analyzing the similarity score against functional associated rate (Figure 1), our method performs better than functional similarity weight scoring scheme [4] and proven an effective model in competing to other topology-based function prediction methods. Our method also shows higher functional association in competing with sequence similarity score, which is traditionally thought as the most reliable method in predicting functional associated protein pairs. More advanced, we input these evaluation score as features of training sets into the support vector machine (SVM) [5]. The predictive models being built through the training process can be used to predict functional associated proteins. The ten-fold cross validation, used to test the accuracy of

the built predictive models, also shows the method we used in this study performs relatively better in the receiver operating characteristic (ROC) curve in comparison to those existing methods (Figure 2). Overall, judging from the data we collected shows the topological similarity applied in this research has much greater potential in predicting functional associated protein pairs. Our results imply that topological similarity in protein interaction network is strongly correlated with protein function association, and can be further exploited to infer functional associated protein pairs within the protein interaction network.

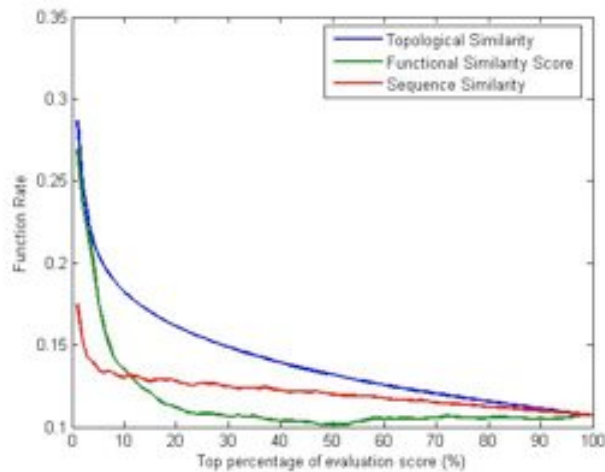


Figure 1. The functional association of different evaluation score.

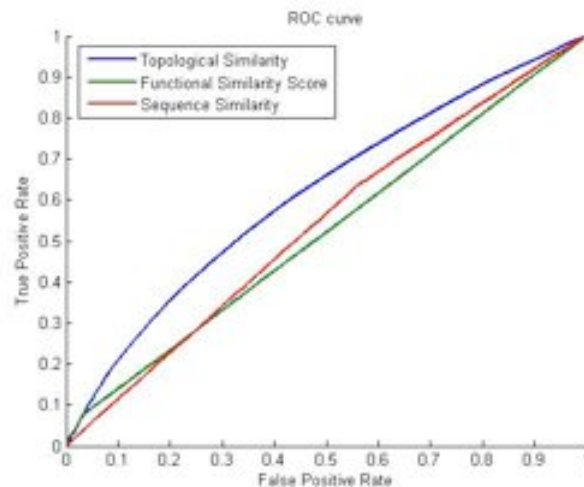


Figure 2. The ROC curve of different training set.

References

- [1] Sharan, R., Ulitsky, I., Shamir, R. (2007). Network-based prediction of protein function. *Mol. Syst. Biol.*, **3**, 88.
- [2] Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S. M., Eisenberg, D. (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.*, **30**, 303-305.
- [3] Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Guldener, U., Mannhaupt, G., Munsterkotter, M., *et al.* (2004) The FunCat, a

functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res.*, **32**, 5539-5545.

- [4] Chua, H. N., Sung, W., and Wong, L. (2006). Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. *Bioinformatics*, **22**, 1623-1630.
- [5] Chang, C. C. and Lin, C. J. (2001) LIBSVM : a library for support vector machines.