Systematic Identification of Weak Connections between Clusters in Biological Network

Hiroki Sakai [1] **Katsuhisa Horimoto** [2]
hiroki@ims.u-tokyo.ac.jp khorimot@ims.u-tokyo.ac.jp
Minoru Kanehisa [1]
kanehisa@kuicr.kyoto-u.ac.jp

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokashou, Uji, Kyoto 611-0011, Japan
[2] Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

Keywords: network, protein-protein interaction, scale-free, weak connection

1 Introduction

The knowledge of interactions between biological molecules has been accumulated rapidly. One of the remarkable features of biological network is that some vertices have a huge number of connections, but almost all vertices are connected only with few vertices [1]. In the network structure, the vertices are grouped into some clusters, and the connections are divided into 'strong connection' within clusters, and 'weak connection' between clusters. The 'weak connection' is considered to play a important role in many networks [4].

In this work, we consider the problem of finding the weak connections in the networks. We develop our method to find the weak connection between clusters, and then our method is applied to the observed protein-protein network. Finally we evaluate the weak connection found by our method in terms of biological knowledge.

2 Materials and Methods

In this work, we analyzed protein-protein interaction data of Saccharomyces cerevisiae. The interaction data were collected from DIP (Database of Interacting Protein [3]). Our method consisted of three steps. 1)At first we simplified input network structure. We excluded vertices whose degree is one in the network, recursively (Fig.1). Then, we combined vertices which have degree two and connected same pair of vertices (Fig.2). 2)Next we found all cliques of the network. In this work, we defined a clique as a cluster of network. We searched all cliques using the algorithm of Bron et.al [2]. 3)Finally, we counted the number of connections(vertices and edges) between two cliques

3 Results and Discussion

Figure 3 shows an example of the partial connections between cluster finding by our procedure. The bold arrow between translation initiation factor and ribosomal RNA processing protein indicates that there are 22 connections, between 2 vertices with less than 12 connections. In contrast, the bold arrow between ribosomal RNA processing protein and pre-mRNA splicing factor indicates that there are 11 connections between 2 vertices with more than 50 connections. In other words, most vertices

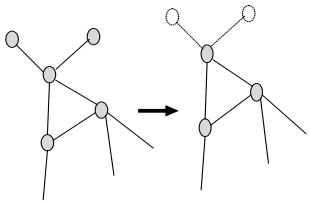


Figure 1: Branch Exclusion

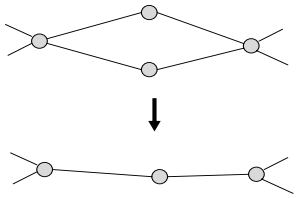
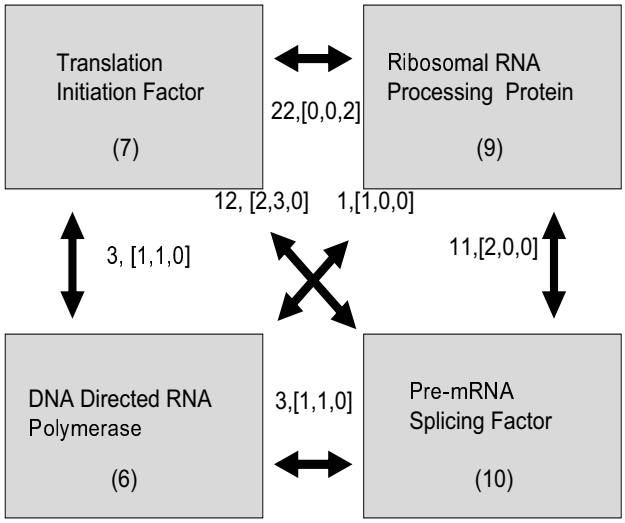


Figure 2: Combine Vertices

Figure 3: Partial connections between clusters by our procedure. A shaded box indicates a clique obtained by our method, in which the vertices are composed of proteins belonging to one functional category, and the number in parenthesis denotes the number of vertices. A bold arrow indicates the vertices between the two cliques, and the number of connections is shown at the side of the arrow. In addition, the vertices connected between the cliques are classified into 3 categories according to the degree; the number of vertices with the large (>50), the middle, and the small (<12) degrees are listed in order in parenthesis.



contribute to form the connection between the two cliques in the former case, and in the latter case, many vertices are connected with the vertices except for the vertices in the two cliques. Since translation initiation factor directly binds the ribosome, the difference in the above cases implies a biological specificity between the functional clusters. All connections and the detailed correspondence with biological functions will be discussed.

References

[1] Albert-Laszlo Barabasi and Eric Bonabeau. Scale-free networks. *Scientific American*, May 2003.

[2] Coen Bron and Joep Kerbosch. Finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973.

[3] http://dip.doe mbi.ucla.edu/.

[4] Mark Newman. The structure and function of complex networks. *Society for Industrial and Applied Mathematics*, 45:167–256, 2003.