Abstract

The research community is inundated with data such as the genome sequences of various organisms, microarray data and so on, of biological origin. This data-volume is rapidly increasing and the process of understanding the data is lagging behind the process of acquiring it. The sheer enormity calls for a systematic approach to understanding this using computational method. The rapid progress of biotechnology and bio-data analysis methods has led to the emergence and fast growth of a promising new field: bioinformatics. It is a field having a tremendous amount of bio-data which needs in-depth analysis. Bio-data is available as, Nucleotide sequences (DNA and RNA sequences), Protein sequences, Genomes and structures in the form of Biological networks (metabolic pathways, gene regulatory network, and protein interaction network).

A framework to discover frequent patterns and modules from biological networks is presented. From the study of different Biological networks, it can be concluded that the best way to analyze and extract the information (frequent functional module) from the biological network is through
graph mining since these networks can be modeled into different types of graphs according to the information needs to be extracted. But this graph-based mining approach often leads to the computationally hard problem due to their relation with subgraph isomorphism. Graph simplification technique is used that is suitable to biological networks, which makes the graph mining problem computationally tractable and scalable to large numbers of networks. So the detection of frequently occurring patterns and modules will be a computationally simpler task since the reduction in the effective graph size significantly.

References

14. Jiawei Han, “How can data mining help bio-data Analysis?” Workshop on data mining in Bioinformatics with SIGKDD02 Conferences” 2002.
15. Mehmet Koyuturk, Yohan Kim, Shankar Subramaniam, “Detecting Conserved Interaction Patterns in Biological Networks”.
17. Bioinformatics center and Institute of chemical research http://www.genome.ad.jp
18. Protein Data Bank (PDB) http://www.rcsb.org/pdb/

**Index Terms**

Computer Science  
Data Mining

**Keywords**

Data mining, Biological networks, graph mining, metabolic pathways.