Combinatorial Optimization Algorithms for Metabolic Networks Alignments and Their Applications

Qiong Cheng, University of Miami, FL
Alexander Zelikovsky, Georgia State University, GA

ABSTRACT

The accumulation of high-throughput genomic and proteomic data allows for reconstruction of large and complex metabolic networks. To analyze accumulated data and reconstructed networks, it is critical to identify network patterns and evolutionary relations between metabolic networks; finding similar networks is computationally challenging. Based on gene duplication and function sharing in biological networks, a network alignment problem is formulated that asks the optimal vertex-to-vertex mapping allowing path contraction, different types of vertex deletion, and vertex insertions. This paper presents fixed parameter tractable combinatorial optimization algorithms, which take into account the similarity of both the enzymes’ functions arbitrary network topologies. Results are evaluated by the randomized P-Value computation. The authors perform pairwise alignments of all pathways for four organisms and find a set of statistically significant pathway similarities. The network alignment is used to identify pathway holes that are the result of inconsistencies and missing enzymes. The authors propose a framework of filling pathway holes by including database searches for missing enzymes and proteins with the matching pro sites and further finding potential candidates with high sequence similarity.

Keywords: Feedback Vertex Sets, Graph Homeomorphism, Metabolic Networks, MetNetAligner, Network Alignment

1 INTRODUCTION

The wealth of the network information have been stored in several public collections of databases, including KEGG (Kanehisa et al., 2006) and BioCyc (2010). These databases provide tools for pathway visualization and for queries on pathway components such as substrates, products and reactions. However, computational tools are called for transferring the well studied knowledge in databases to unknown one (e.g. for searching for homologues to a query pathway in a collection of known pathways) and for aligning two pathways to locate conserved pathway fragments.

DOI: 10.4018/jkdb.2011010101
Most of the high-throughput genomic, proteomic data are not catalogued or processed directly by hand but by computational technologies. It means that the high-throughput datasets may be filled with technical and biological noise and have different technical biases and coverage (Han, 2008) or the usage of inconsistent data/tool versions. All of them requires data curation. However not all data are curated. For BioCyc (2010), only a part of pathway data have received person-decades of literature-based curation and are the most accurate. However the logical interpretation of the whole network is definitely not easily comprehensible to the human brain and therefore it is impossible to curate all the more and more increasing data only by hand. The computational tool is required for inferring the missing/inconsistent information in genome and pathway databases.

All of them require efficient computational methods for analyzing and comparing networks. In this paper, we focus on network alignment for comparing, exploring, and predicting these networks.

Let the pattern be a pathway for which one is searching for homologous pathways in the text, i.e., the known metabolic network of a different species. Alignment of two networks, pattern and text, is a basic task which can meet the requirement of a series of open questions such as network evolution and critical target search. This is a challenging research topic from both biological and computational perspective.

Existing approaches to subgraph iso- and homeomorphism restrict the size (Sharan et al., 2005; Yang & Sze, 2007) or topology of the pattern (Chen & Hofestaedt, 2004, 2005; Pinter, Rokhlenko, Yeger-Lotem, & Ziv-Ukelson, 2005; Cheng, Harrison, & Zelikovsky, 2007; Cheng, Kaur, Harrison, & Zelikovsky, 2007) or use heuristics and approximation algorithms. GraphMatch (Yang & Sze, 2007) allows to delete disassociated vertices or induced subnetwork in query network and then align its remainder to target network by subgraph isomorphism.

However, the widespread evolutionary machinery of gene duplication results in vertex copying (Sharan & Ideker, 2006). The results of network alignments can be enhanced when gene duplication and divergence are taken into account. If two enzymes in the pattern species are evolutionarily related, the corresponding enzymes can be mapped into a single enzyme. The mapping explores the characteristics of graph homomorphism.

Based on the property, we have formulated the network alignment problem which asks the optimal vertex-to-vertex mapping allowing path contraction, vertex deletion, and vertex insertions. In this paper we present combinatorial algorithms, which take into account the similarity of both the enzymes’ functions arbitrary network topologies such as trees and arbitrary graphs. The proposed algorithm is fixed parameter tractable in the liner or square of the size of feedback vertex set respectively for the case of disallowing or allowing the deletions.

We have developed the web service tool MetNetAligner which aligns metabolic networks. We evaluated our results by the randomized P-Value computation. In the computation, we followed two standard randomization procedures and further developed two other random graph generators which meet the more stringent and consistent topology constraints. By comparing their distribution of the significant alignment pairs, we observed that the more stringent constraints in the topology the random graph generator has, the more pairs of significant alignments there exist.

We applied network alignments (Figure 1) to identifying pathway holes and proposed a framework to find potential candidates for filling the holes. The pathway holes may be the result of ambiguity in identifying a gene and its product in an organism or when the gene encoding an enzyme is not identified in an organism’s genome (Green & Karp, 2004). Due to gaps in sequence motif research, several sequences may not get specific annotations. Specific function of a protein may not be known during annotation. Reactions catalyzed by those proteins result in metabolic pathway holes. An error in reading an ORF (open reading frame) may also lead to a pathway hole. With further
Chemosensitivity Prediction of Tumours Based on Expression, miRNA, and Proteomics Data
*International Journal of Systems Biology and Biomedical Technologies* (pp. 1-19).
www.igi-global.com/article/chemosensitivity-prediction-tumours-based-expression/67103?camid=4v1a