

# Pair wise Alignment of Pathways



Qiong Cheng; Alexander Zelikovsky  
Department of Computer Science, Georgia State University, Atlanta, GA 30303

## Outline

- Motivation
- Problem Formulation
- Solution
- Results From the First Stages
- Plans for Next Stage
- Related Works
- Reference

## Motivation

The Homeomorphism problem of pair wise graphs has been an open problem in the theory of graph. There have been lots of applications which can be reducible to the problem. We aims to propose a feasible solution to calculate the minimal cost of homeomorphism alignment in biological pathways so that we can

- Study the approximate solution to the homeomorphism problem in the system biology;
- Study algorithms to calculate the significant similarity of pathways in biological networks;
- Study the evolution of pathways in different species and different functions in the same species;
- Study the reconstruction problem of phylogenetic networks ;

## Problem Formulation

Given pattern tree  $T$  and text graph  $G$ , find a homeomorphic mapping that minimize the homeomorphic cost.

Def. A mapping from pattern  $T = \langle V^T, E^T, L^T \rangle$  to text  $G = \langle V^G, E^G, L^G \rangle$  ( $f: PatternT \rightarrow TextG$ ) is called homeomorphic if

1. every vertex  $v$  in  $V^T$  is mapped a vertex in  $V^G$  ( $f_v: V^T \rightarrow V^G$ );
2. every edge  $e = (v_i, v_j)$  in  $E^T$  is mapped to a path in  $P^G(f_i, f_j: E^T \rightarrow P^G)$ . The path is a ordered sequence  $u_0 = f(v_i), u_1, u_2, \dots, u_k = f(v_j)$  of vertices in  $P^G$ ;

Def. Let  $\Delta: L^T \times L^G \rightarrow \mathbb{R}$  denote the replace relation on the set of all ordered pairs of a vertex in  $V^T$  and a vertex in  $V^G$  to a real number. The relation represents a 2-dimension table of the vertex-to-vertex similarity cost scores.  $\delta: L^G \rightarrow \mathbb{R}$  denote the binary relation. Its domain is all vertices in  $V^G$ , and its range is the real number set.  $\delta$  represents the penalty score for deleting a vertex from a path. The cost of a homeomorphic mapping  $f: T \rightarrow G$  is

$$cost(f) = \sum_{v \in V^T} \Delta(v, f(v)) + \delta \times \sum_{e \in E^T} (|f(e)| - 1)$$

## Our Solution

- Employ dynamic program to get an exact result for the specified problem
- Build a two dimension table for dynamic program by the post order traversal;
- Recursive Function:

$$M(v_i^T, v_j^G) = \min \{ \Delta(v_i^T, v_j^G) + \sum_{v_k^G \in \text{Children}(v_j^G)} \max_{v_l^T \in \text{Children}(v_i^T)} M(v_l^T, v_k^G), \delta \times |v_j^G| + \max_{v_l^G \in \text{Children}(v_j^G)} M(v_i^T, v_l^G) \}$$

## Table for Dynamic Program

	Pattern Graph	Text Graph				
		4.2.1.-	2.3.1.9	6.3.2.4	2.-.-.-	2.3.1.46
Penalty		2	2	2	2	2
2.6.1.19		4	3	4	3	3
4.1.1.15		INF	INF	7 (+)	INF	7 (+)

## Results from the First Stage

- We did experiments on alignment of pathways of bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae*;
- It supports the alignment of directed cyclical and acyclical pattern graphs and directed cyclical and acyclical text graphs;
- It supports two types of calculation of significance based on random graph by reshuffling edges and nodes of the original text graph;
- Statistical Significant of alignments ( An example- chorismate biosynthesis in *E. Coli*)

	1	2	3
ReshufflingEdge	0.07	0.07	0.07
ReshufflingNode	0.18	0.19	0.19
Dr.Pinter's solution	0.29	0.29	0.29

## Plans for the Next Stage

- To further prove whether the homeomorphism problem on biological pathways is NP-completeness;
- To improve our approach to the node-disjoint path problem so that it has higher significant similarity of pathways;
- Try to employ linear programming to obtain the approximate solution;
- Accomplish more experiments to study the evolution of pathways;
- Try the clustering technology to study the evolution of pathways;

## Related Works

- **Kelley et al (2003)**
  - Goal: identify conserved pathways (chains)
  - Idea: can be done efficiently by reducing it to sequence alignments
- **Ron Pinter (2005)**
  - Goal: identify conserved pathways (directed acyclic graphs)
  - Idea: Dynamic program for mapping on multi source trees

## Reference

1. Ron Y.Pinter, Oleg Rokhlenko, Esti Yeger-Lotem and Michal Ziv-Ukelson (2005) Aignment of metabolic pathways Vol.21 no. 16 Bioinformatics
2. Ming Chen and Ralf Hofest, An algorithm for linear metabolic pathway alignment, *In silico biology ISSN, 1386-6338:111-128, (2005)*
3. Tohsata Y., Matsuda H. and Hashimoto A. A Multiple Alignment Algorithm for Metabolic Pathway Analysis using Enzyme Hierarchy, Proc. 8th International Conference on Intelligent Systems for Molecular Biology,(ISMB 2000): 376-383
4. M. Garey and D. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman, New York, 1979

For additional information, please contact:

Alexander Zelikovsky  
Qiong Cheng  
Computer Science Department  
Georgia State University

[alexz@cs.gsu.edu](mailto:alexz@cs.gsu.edu)  
[cscoqcx@cs.gsu.edu](mailto:cscoqcx@cs.gsu.edu)