

# Graph comparison : homo-home morphism mapping in Metabolic Pathways

Qiong Cheng, Robert Harrison, Alexander Zelikovsky

Department of Computer Science, Georgia State  
University, Atlanta, Georgia 30303  
<http://www.cs.gsu.edu/>

**Abstract.** Network mappings has been commonly applied on it to detect conserved subnetworks across species in biological network. Our previous work focused on finding best mapping of pattern graph to text graph in which vertices in the image of pattern graph are allowed to be deleted [1],[2]. Arising from the requirement of allowing deleting pattern vertices in mapping, we reformulate the graph comparison problem to be the one of a homo-home morphism mapping among arbitrary graphs and further analyze and compare our work with previous ones.

**Key words:** homomorphism, metabolic pathway mapping

## 1 Introduction

Network mappings are extensively used for comparing, exploring, and predicting biological networks, are essential for pathway database search and also helpful for finding and filling pathway holes. Existing mapping tools are mostly based on isomorphic and homeomorphic embedding effectively solving a problem that is NP-complete even when searching a match for a tree in acyclic networks.

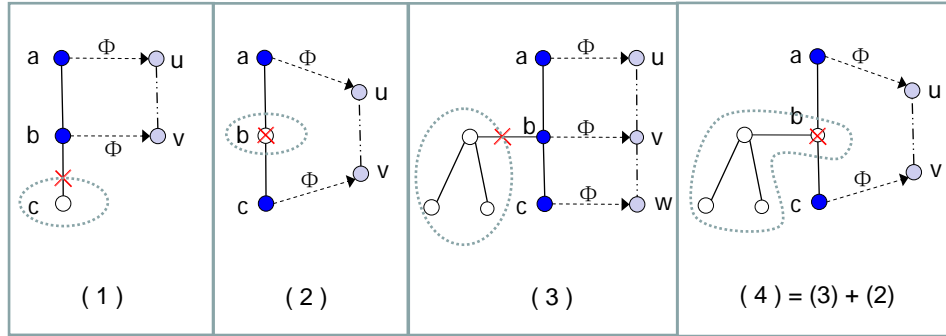
Our previous work focused on finding best mapping of pattern graph to text graph in which vertices in the image of pattern graph are allowed to be deleted [1],[2]. Arising from the requirement of allowing deleting pattern vertices in mapping [see fig. 1], we reformulate the graph comparison problem to be the one of a homo-home morphism mapping from a directed graph with restricted cyclic structure to an arbitrary network. Furthermore, we analyze and compare our work with previous ones. The reformulation revokes a new heuristic and interesting algorithm we are working on so that this problem could be solved in a limited run time.

## 2 Problem Formulation

### 2.1 NOTATION AND DEFINITIONS

Given graph  $G = \langle V_G, E_G \rangle$ , and graph  $H = \langle V_H, E_H \rangle$ , let us give the following definition.

A function  $f : V_G \rightarrow V_H$  is homomorphism  $h$  from  $G$  to  $H$  if  $\forall (u, v) \in E_G, u, v \in \text{Dom}(f), (f(u), f(v)) \in E_H$ .



**Fig. 1.** Samples for pattern vertex deletion constraints. Real line represents an edge in pattern; dashed line represents a path in text; dashed arrow represents a mapping of vertex-to-vertex; dashed circle represent the deletion group;  $\times$  happens on the associated vertex or edge related to the deletion; blue circle represents a vertex in pattern; white circle represents a deleted vertex in pattern; grey circle represents a vertex in text. (1) Deletion of patten vertex with degree 1; (2) Deletion of patten vertex with degree 2; (3) For vertex with degree large than 2, previously visited child branch has been deleted but the vertex will not be deleted; (4) If child branch has been deleted so that this vertex has only degree 2, the vertex can be allowed to delete.

A homomorphism  $h$  from  $G$  to  $H$  is called 1-to-1 homomorphism if  $h$  is an 1-to-1 function.

A partial homomorphism  $h'$  from  $G$  to  $H$  is a homomorphism from a subset of  $G$  to  $H$ .

The partial homomorphism is one to one such that if  $h_{v'}(v) = h_{v'}(u)$ , then  $v = u$ .

1-to-1 homomorphism is called embedding. 1-to-1 partial homomorphism is called partial embedding.

For any graph  $G$ , its transitive closure is  $G^* = \langle V, E_{G^*} \rangle$ , in which any  $e = (u, v) \in E_{G^*}$  is a path from  $u$  to  $v$  in  $G$ .

An embedding from  $G$  to the transitive closure of  $H$  is called a homeomorphic mapping  $hf : G \rightarrow H$ .

A necklace in graph from  $u$  to  $v$  is a specific  $u$ - $v$  path such that any vertex except  $u$  and  $v$  has degree 2. The intermediate vertices between  $u$  and  $v$  are called beads. The set of beads is denoted as  $B$ .

Bead deletion  $d : G \rightarrow H$  is a partial embedding from  $G$  to  $H$  which satisfies  $V_G - \text{Dom}(d) \in B$  and  $\forall u, v \in \text{Dom}(d)$ , if there exist a path from  $u$  to  $v$ , there is an edge  $(d(u), d(v))$  in  $H$ .

Partial homeomorphic embedding consists of the composition  $d \circ hf$  of bead deletion  $d$  and homeomorphism mapping  $hf$ .

Our specifical **homo-home morphism**  $HH : G \rightarrow H$  is the composition of several relations as follows

- (1)  $h' : G \rightarrow G'$ , partial homomorphism from  $G$  to  $G'$ ;
- (2)  $d : G' \rightarrow G''$ , bead deletion from  $G' \rightarrow G''$ ;
- (3)  $hf : G'' \rightarrow H$ , homeomorphism mapping from  $G'' \rightarrow H$ ;

So

$$HH : G \xrightarrow{h'} G' \xrightarrow{d} G'' \xrightarrow{hf} H$$

## 2.2 Problem Formulation

The **cost of homo-home morphism**  $HH : G \rightarrow H$  is

$$\text{cost}(f) = \sum_{v \in V_G \ \& \ HH(v) \neq \emptyset} \Delta(v, f(v)) + \sum_{v \in V_{G'} \ \& \ d(v) \equiv \emptyset} (\Omega(v)) + \lambda \sum_{e \in E_{G''} \ \& \ |hf(e)| \geq 1} (|hf(e)| - 1)$$

Given a graph  $P$  and a graph  $H$ , to find minimum cost of homo-home morphism  $HH : G \rightarrow H$ .

## 3 Previous Work

In the above relation composition, Pinter et al [7], [8] focused on homeomorphism mapping  $hf$  from multi-source tree to multi-source tree. Based on the formulation - finding minimum cost homeomorphism, they proposed a dynamic programming algorithm in polynomial runtime.

Kelly et al [3], [4], [5] worked on partial homeomorphic embedding of query path onto path. They reduced it to be a problem on finding a highest-scoring path of length  $L$  in acyclic graph. It can be solved in polynomial time. Authors randomly decomposed the text graph into linear pathways which are then aligned against the pattern path and then for every decomposed linear text graph path, they run algorithm to finding partial homeomorphic embedding  $d \circ hf$  of query path onto path. When effectively limiting the size of the query path to about six vertices, authors can solve the problem of aligning query path to arbitrary graph.

Sze et al [6] focused on two problems : path matching and graph matching. Their path matching is exactly a partial homeomorphic embedding  $d \circ hf$  from graph path to arbitrary graph. Authors reduced the path matching problem to finding a longest weighted path in a directed acyclic graph and showed that the problem of finding top  $k$  suboptimal paths can be solved in polynomial time.

Sze's graph matching is exactly the composition  $h \circ hf$  of partial homomorphism and homeomorphic mapping between two graphs. The first relation—partial homomorphism make it disappear the subgraph consisting of vertices in query graph which have no association in text graph. The second relation generates the edge-to-path mapping. Authors reduced the graph matching problem to finding highest scoring subgraphs in a graph and give an exact algorithm to solve the problem when the query graph  $G_0$  is of moderate size.

In both Sze's solutions individually for their two matching cases, authors omit that they allow different vertices in pattern can be associated with a vertex in text as well as a vertex in pattern can be associated multiple vertices in text graph. This permission is equivalent to allowing different vertices in query path can be mapped to a vertex in text in an optimal graph matching.

## 4 Conclusions

In this paper we have reformulated our graph comparison problem to be the one of a homo-home morphism mapping from a directed graph with restricted cyclic structure to an arbitrary network; we further analyze and compare previous works. The formulation allow to delete preimage vertex as well as image vertex in the mapping results. It also allows different vertices in preimage could be mapped to the same vertex in text. We are working on the implementation of a heuristic algorithm to solve this problem in a limited run time and the corresponding experiments.

## References

1. Q. Cheng, R. Harrison, and A. Zelikovsky: Homomorphisms of Multisource Trees into Networks with Applications to Metabolic Pathways. Proc. of IEEE 7-th International Symposium on BioInformatics and BioEngineering (BIBE'07)
2. Q. Cheng, D. Kaur, R. Harrison, and A. Zelikovsky, "Mapping and Filling Metabolic Pathways", RECOMB Satellite Conference on Systems Biology 2007
3. Brian P. Kelly, Roded Sharan, Richard M. Karp, Taylor Sittler, David E. Root, Brent R. Stockwell: Conserved pathways within bacteria and yeast as revealed by global protein network alignment. PNAS. (Sep. 30 2003) 11394-11399
4. Brian P. Kelly, Roded Sharan, Richard M. Karp, Taylor Sittler, David E. Root, Brent R. Stockwell: PathBLAST: a tool for alignment of protein interaction networks. Nucleic Acids Research. (2004) Vol.32 : W83-W88
5. Roded Sharan, Silpa Suthram, Ryan M. Kelley, Tanja Kuhn, Scott McCuine, Peter Uetz, Taylor Sittler, Richard M. Karp, Trey Ideker: Conserved patterns of protein interaction in multiple species. PNAS. Vol.102 : 1974-1979 (2005)
6. Qingwu Yang, Sing-Hoi Sze: Path Matching and Graph Matching in Biological Networks. Journal of Computational Biology. Vol. 14, No. 1: 56-67 : 5527-5530 (2007)
7. Pinter,R.Y., Rokhlenko,O., Tsur,D., Ziv-Ukelson,M.: Approximate labeled subtree homeomorphism. In Proceedings of 15th Annual Symposium of Combinatorial Pattern Matching, LNCS 3109. Springer-Verlag. (2004) 59-73
8. Ron Y Pinter, Oleg Rokhlenko, Esti Yeger-Lotem, Michal Ziv-Ukelson: Alignment of metabolic pathways. Bioinformatics. LNCS 3109. Springer-Verlag.(Aug 2005)21(16): 3401-8