



**HAL**  
open science

## A graph matching method based on probes assignments

Romain Raveaux, Jean-Christophe Burie, Jean-Marc Ogier

► **To cite this version:**

Romain Raveaux, Jean-Christophe Burie, Jean-Marc Ogier. A graph matching method based on probes assignments. 2008. hal-00305232v1

**HAL Id: hal-00305232**

**<https://hal.science/hal-00305232v1>**

Preprint submitted on 23 Jul 2008 (v1), last revised 25 Aug 2008 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A graph matching method based on probes assignments

Romain Raveaux, Jean-christophe Burie and Jean-Marc Ogier.

L3I, University of La Rochelle, av M. Crépeau, 17042 La Rochelle Cedex 1, France, E-mail:  
{romain.raveaux01}@univ-lr.fr

**Abstract.** In this paper, a graph matching method and a distance between attributed graphs are defined. Both approaches are based on graph probes. Probes can be seen as features extracted from a given graph. They represent a local information. According to two graphs  $G_1, G_2$ , with respect to the cost function.

## 1 Probe Matching and Probe Matching Distance

### 1.1 Probes of graph

Let  $L_V$  and  $L_E$  denote the set of node and edge labels, respectively. A labeled, undirected graph  $G$  is a 4-tuple  $G = (V, E, \mu, \xi)$ , where

- $V$  is the set of nodes,
- $E \subseteq V \times V$  is the set of edges
- $\mu : V \rightarrow L_V$  is a function assigning labels to the nodes, and
- $\xi : E \rightarrow L_E$  is a function assigning labels to the edges.

From this definition of graph, probes of graph for the matching problem can be expressed as follow:

Let  $G$  be an attributed graph with edges labeled from the finite set  $\{l_1, l_2, \dots, a\}$ . Let  $P$  be a set of probes extracted from  $G$ . There is a probe  $p$  for each vertex of the graph  $G$ . A probe ( $p$ ) is defined as a pair  $\langle V_i, H_i \rangle$  where  $H_i$  is an edge structure for a given vertex ( $V_i$ ),  $H_i$  is a  $2a$ -tuple of non-negative integers  $\{x_1, x_2, \dots, x_a, y_1, y_2, \dots, y_a\}$  such that the vertex has exactly  $x_i$  incoming edges labeled  $l_i$ , and  $y_j$  outgoing edges labeled  $l_j$ .

### 1.2 Probe Matching

Let  $G_1(V_1, E_1)$  and  $G_2(V_2, E_2)$  be two attributed graphs. Without loss of generality, we assume that  $|P_1| \geq |P_2|$ . The complete bipartite graph  $G_{em}(V_{em} = P_1 \cup P_2 \cup \Delta, P_1 \times (P_2 \cup \Delta))$ , where  $\Delta$  represents an empty dummy probe, is called the probe matching graph of  $G_1$  and  $G_2$ . A probe matching between  $G_1$  and  $G_2$  is defined as a maximal matching in  $G_{em}$ . Let there be a non-negative

metric cost function  $c : P1 \times P2 \rightarrow \mathbb{R}_0^+$ . We define the matching distance between  $G1$  and  $G2$ , denoted by  $d_{match}(G1, G2)$ , as the cost of the minimum-weight probe matching between  $G1$  and  $G2$  with respect to the cost function  $c$ .

### 1.3 Cost function for probe matching

Let  $p_1$  and  $p_2$  be two probes. The cost function can be expressed as a distance between two probes :  $c(p_1, p_2) = |V_1 - V_2| + |H_1 - H_2|$

### 1.4 Time complexity analysis

The matching distance can be calculated in  $O(n^3)$  time in the worst case. To calculate the matching distance between two attributed graphs  $G1$  and  $G2$ , a minimum-weight probe matching between the two graphs has to be determined. This is equivalent to determining a minimum-weight maximal matching in the probe matching graph of  $G1$  and  $G2$ . To achieve this, the method of Kuhn [1] and Munkres [2] can be used. This algorithm, also known as the Hungarian method, has a worst case complexity of  $O(n^3)$ , where  $n$  is the number of probes in the larger one of the two graphs [3].

### 1.5 The probe matching distance for attributed graphs is a metric.

*Proof.* To show that the probe matching distance is a metric, we have to prove the three metric properties for this similarity measure.

- $d_{match}(G_1, G_2) \geq 0$   
The probe matching distance between two graphs is the sum of the cost for each probe matching. As the cost function is non-negative, any sum of cost values is also non-negative.
- $d_{match}(G_1, G_2) = d_{match}(G_2, G_1)$   
The minimum-weight maximal matching in a bipartite graph is symmetric, if the edges in the bipartite graph are undirected. This is equivalent to the cost function being symmetric. As the cost function is a metric, the cost for matching two probes is symmetric. Therefore, the probe matching distance is symmetric.
- $d_{match}(G_1, G_2) \leq d_{match}(G_1, G_2) + d_{match}(G_2, G_3)$   
As the cost function is a metric, the triangle inequality holds for each triple of probes in  $G1$ ,  $G2$  and  $G3$  and for those probes that are mapped to an empty probe. The probe matching distance is the sum of the cost of the matching of individual probes. Therefore, the triangle inequality also holds for the probe matching distance.

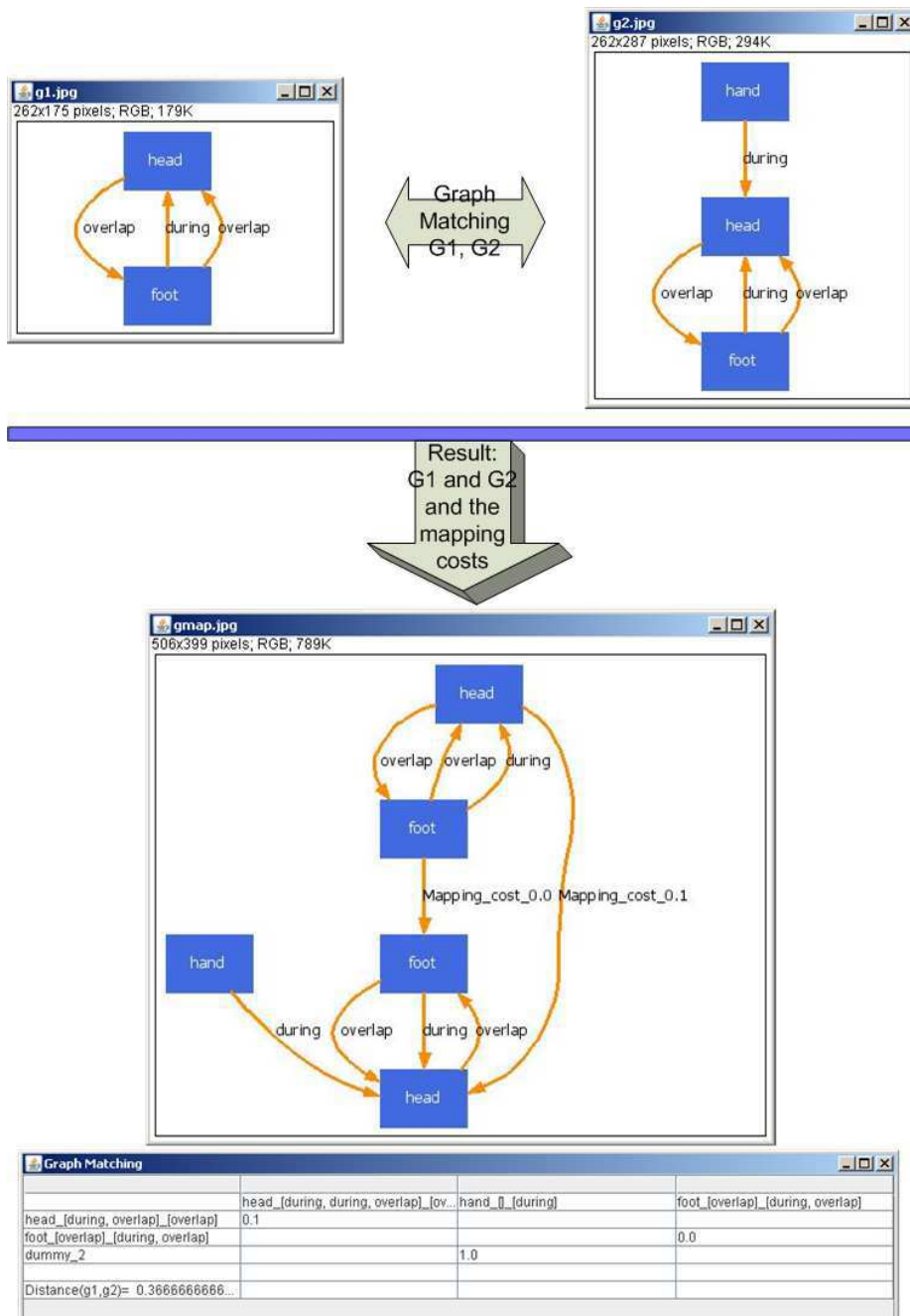
### 1.6 The probe matching distance is a lower bound for the edit distance.

Given a cost function for the edge matching which is always less than or equal to the cost for editing an probe, the matching distance between attributed graphs is a lower bound for the edit distance between attributed graphs:  $\forall G_1, G_2 : d_{match}(G_1, G_2) \leq d_{ED}(G_1, G_2)$

*Proof.* The edit distance between two graphs is the number of edit operations which are necessary to make those graphs isomorphic. To be isomorphic, the two graphs have to have identical probe sets. As the cost function for the probe matching distance is always less than or equal to the cost to transform two probes into each other through an edit operation, the probe matching distance is a lower bound for the number of edit operations, which are necessary to make the two probe sets identical. It follows that the edge matching distance is a lower bound for the edit distance between attributed graphs.

### References

1. H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistic Quarterly*, 2:83–97, 1955.
2. J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society of Industrial and Applied Mathematics*, 5(1):32–38, March 1957.
3. Hans-Peter Kriegel and Stefan Schonauer. Similarity search in structured data.



**Fig. 1.** Graph matching between two graphs  $g_1$ , and  $g_2$ .  $G_{map}$  represents  $g_1$  and  $g_2$  in a single graph with the addition of "cost" edges to map node of  $g_1$  in  $g_2$