Fakultät für Informatik
Facoltà di Scienze e Tecnologie informatiche
Faculty of Computer Science

Exact Computation of Graph Edit Distance for Uniform and Non-Uniform Metric Edit Costs

David B. Blumenthal & Johann Gamper

GbRPR, Anacapri, 18 May 2017

Overview

- graph edit distance: flexible distance measure for labelled graphs
- supports uniform and non-uniform edit costs
- exact computation is NP-hard
- existing exact algorithms
  - A*-GED (Riesen, Fankhauser, and Bunke 2007)
  - BLP-GED (Lerouge et al. 2016)
  - DF-GED: node-based DFS, designed for non-uniform edit costs (Abu-Aisheh et al. 2015)
  - CSI_GED: edge-based DFS, supports uniform edit costs only (Gouda and Hassaan 2016)
- contributions
  (1) DF-GED^u: speed-up of DF-GED for uniform edit costs
  (2) CSI_GED^{nu}: generalised version of CSI_GED that supports non-uniform edit costs

Two Communities

- Pattern Recognition
- Database Technologies
  - a lot of work o graph edit distance exists
  - publications in venues such as VLDB, ICDE, SIGMOD, TKDE, CIKM
  - main focus: filtering and lower bounds
  - slightly different definitions
  - main difference: restriction on uniform edit costs

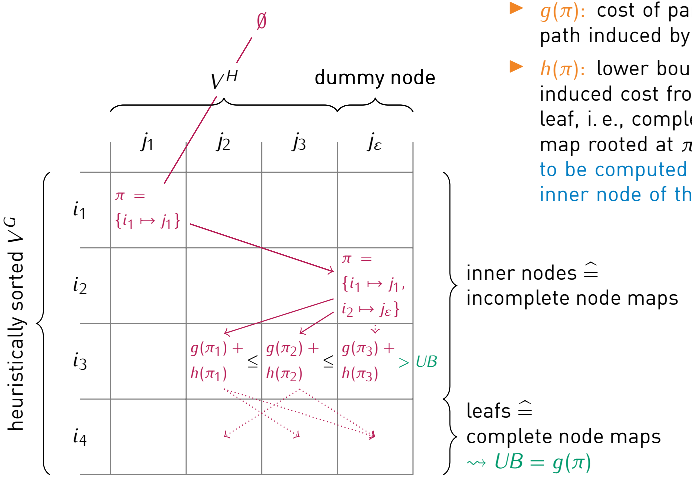
Graph Edit Distance

- labelled undirected graph: 4-tuple $G = (V^G, E^G, \ell_V^G, \ell_E^G)$
- label functions: $\ell_V^G : V^G \to \Sigma_V$ for nodes, $\ell_E^G : E^G \to \Sigma_E$ for edges
- edit path between G and H: sequence of edit operations starting at G and ending at $H' \simeq H$
- edit operations: deleting, inserting, relabelling
- edit costs: $c_V : \Sigma_V \times \Sigma_V \to \mathbb{R}$ for operations on nodes, $c_E : \Sigma_E \times \Sigma_E \to \mathbb{R}$ for operations on edges
- uniform edit costs: $c_V(\alpha, \beta), c_E(\alpha, \beta) = \begin{cases} 1 & \alpha \neq \beta \\ 0 & \alpha = \beta \end{cases}$
- graph edit distance $\lambda(G, H)$: minimum cost of edit path between G and H

Node Maps

- $V^{G+|H|}$: V^G plus $|V^H|$ isolated dummy nodes
- node map: injective partial function $\pi : V^{G+|H|} \rightarrow V^{H+|G|}$ with $V^G \subseteq dom(\pi)$ and $V^H \subseteq img(\pi)$
- edit path induced by node map: let $i \in V^G$, $k \in V^H$, $ij \in E^G$, $kl \in E^H$
  - $\pi(i) = k \rightsquigarrow$ change node label from $\ell_V^G(i)$ to $\ell_V^H(k)$
  - $\pi(i) = k_\varepsilon \rightsquigarrow$ delete node i
  - $\pi^{-1}(k) = i_\varepsilon \rightsquigarrow$ insert node k
  - $\pi(i)\pi(j) = kl \rightsquigarrow$ change edge label from $\ell_E^G(ij)$ to $\ell_E^H(kl)$
  - $\pi(i)\pi(j) \notin E^H \rightsquigarrow$ delete edge ij
  - $\pi^{-1}(k)\pi^{-1}(l) \notin E^G \rightsquigarrow$ insert edge kl
- alternative definition of $\lambda(G, H)$: minimum cost $g(\pi)$ of edit path induced by a node map π

DF-GED: Node-Based DFS



- $g(\pi)$: cost of partial edit path induced by π
- $h(\pi)$: lower bound for induced cost from π to a leaf, i. e., complete node map rooted at π $\rightsquigarrow$ has to be computed at each inner node of the DFS

Our Speed-Up DF-GED^u for Uniform Edit Costs

- $h(\pi)$: defined as $MLA(\overbrace{\ell_V^G(V^{G+|H|-\pi})}^{\text{multiset with unassigned labels from nodes in } V^{G+|H|}} \times \ell_V^H(V^{H+|G|-\pi}), c_V) + MLA(\underbrace{\ell_E^G(E^{G-\pi})}_{\text{multiset with unassigned labels from edges in } E^G} \times \ell_E^H(V^{H-\pi}), c_E)$
- computation for non-uniform edit costs requires cubic time

Lemma

For uniform edit costs, $h(\pi)$ can be computed in linear time.

1. at initialisation, sort node and edge labels
2. compute $MLA(A \times B, c)$ as $\Gamma(A, B) = \max\{|A|, |B|\} - |A \cap B|$

Valid Edge Maps (I)

- $\overrightarrow{E^G}$: one oriented edge (i, j) for each undirected $ij \in E^G$
- $\overleftrightarrow{E^H}$: both (k, l) and (l, k) for each $kl \in E^H$
- edge map: mapping $\phi : \overrightarrow{E^G} \to \overleftrightarrow{E^H} \cup \{e_\varepsilon\}$
- induces relation π_ϕ on $V^G \times V^H$: if $\phi(i, j) = (k, l)$, then $(i, k) \in \pi_\phi$ and $(j, l) \in \pi_\phi$
- valid edge map: ϕ is valid iff π_ϕ is partial injective function

Valid Edge Maps (II)

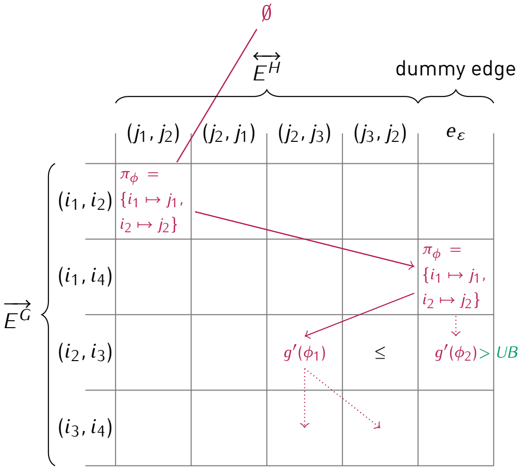
- partial edit path induced by valid edge map: let $i \in V^G$, $k \in V^H$, $(i,j) \in \overrightarrow{E^G}$, $(k,l),(l,k) \in \overleftrightarrow{E^H}$
  - $\phi(i,j) = (l,k) \rightsquigarrow$ change edge label from $\ell_E^G(ij)$ to $\ell_E^H(kl)$
  - $\phi(i,j) = e_\varepsilon \rightsquigarrow$ delete edge ij
  - $\phi^{-1}[\{(k,l),(l,k)\}] = \emptyset \rightsquigarrow$ insert edge kl
  - $\pi_\phi(i) = k \rightsquigarrow$ changed node label from $\ell_V^G(i)$ to $\ell_V^H(k)$

Theorem

$\lambda(G,H) = \min\{g(\phi) + \Gamma(V^{G-\pi_\phi}, V^{H-\pi_\phi}) \mid \phi \text{ is valid edge map}\}$ holds for uniform edit costs, where $g(\phi)$ is the cost of the partial edit path induced by edge map ϕ.

- can compute $\lambda(G,H)$ by traversing space of all valid edge maps

CSI_GED: Edge-Based DFS



- $g(\phi)$: cost of partial edit path induced by ϕ
- $g'(\phi)$: lower bound for induced cost of complete edge map rooted at ϕ

inner nodes $\widehat{=}$ incomplete edge maps

leafs $\widehat{=}$ complete edge maps
$\rightsquigarrow UB = g(\phi) + \Gamma(V^{G-\pi_\phi}, V^{H-\pi_\phi})$

Our Generalisation CSI_GED^{nu}

Theorem

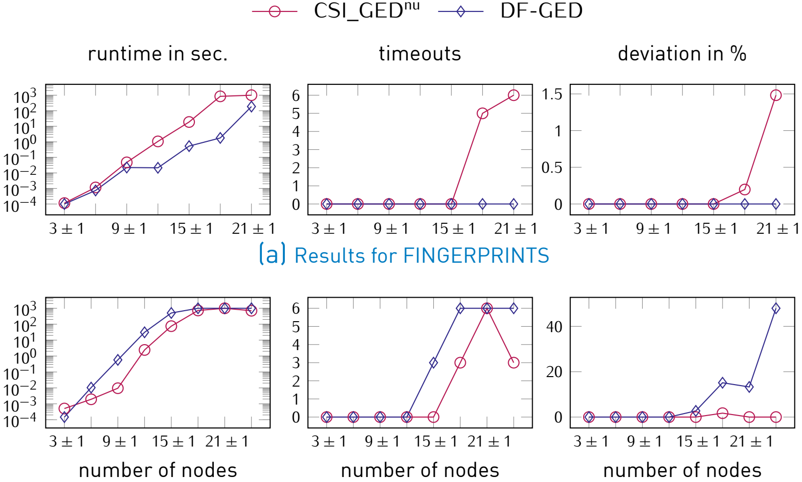
$\lambda(G,H) = \min\{g(\phi) + MLA(\ell_V^G(V^{G+|H|-\pi_\phi}) \times \ell_V^H(V^{H+|G|-\pi_\phi}), c_V) \mid \phi \text{ is valid edge map}\}$ holds for non-uniform metric edit costs.

- can use CSI_GED's DFS framework for non-uniform edit costs
- at leafs, use *MLA* instead of Γ to compute *UB*
- increased complexity at leafs (cubic instead of linear)
- no increased complexity at inner nodes of search tree

Setup

- used the datasets AIDS and FINGERPRINTS (Riesen and Bunke 2008)
- formed groups of size four containing graphs of fixed size and ran all algorithms for all pairs of graphs in one test group
- set time limit of 1000 seconds
- recorded the runtime, the number of timeouts, and the deviation of an algorithm's upper bound after 1000 seconds from the best upper bound

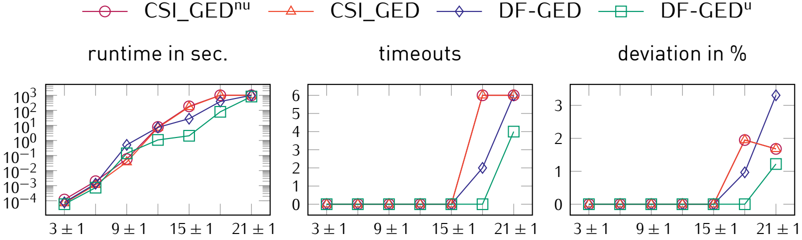
Results for Non-Uniform Metric Edit Costs



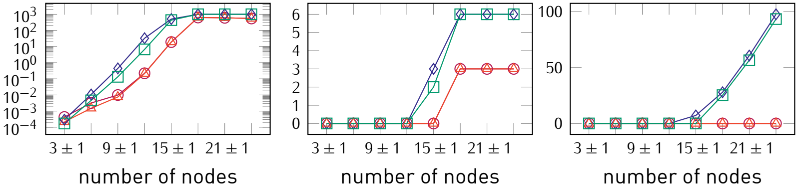
(a) Results for FINGERPRINTS

(b) Results for AIDS

Results for Uniform Edit Costs



(a) Results for FINGERPRINTS



(b) Results for AIDS

Upshot of the Results

- uniform edit costs
  - our speed-up DF-GED^u always outperforms DF-GED
  - CSI_GED and our generalisation CSI_GED^{nu} perform similarly
- general observation: no clear winner between node based and edge based algorithms
- FINGERPRINTS: DF-GED and DF-GED^u perform better
- AIDS: CSI_GED^{nu} and CSI_GED perform better
- CSI_GED and CSI_GED^{nu} are more stable than DF-GED and DF-GED^u: their deviation is small even if DF-GED and DF-GED^u perform better
- no prior knowledge about dataset and both uniform and non-uniform edit costs relevant $\rightsquigarrow$ CSI_GED^{nu} is algorithms of choice

Future Work

- individuate characteristics of datasets, for which the node based/edge based approaches perform better
- develop meta-algorithm based on these characteristics
- combine techniques from both communities in order to come up with significantly faster algorithm

References

Abu-Aisheh, Zeina et al. (2015). “An Exact Graph Edit Distance Algorithm for Solving Pattern Recognition Problems”. In: ICPRAM 2015. Ed. by Maria De Marsico, Mário A. T. Figueiredo, and Ana L. N. Fred. Vol. 1. SciTePress, pp. 271–278.

Gouda, Karam and Mosab Hassaan (2016). “CSI_GED: An Efficient Approach for Graph Edit Similarity Computation”. In: 32nd IEEE International Conference on Data Engineering. IEEE Computer Society, pp. 265–276.

Lerouge, Julien et al. (2016). “Exact Graph Edit Distance Computation Using a Binary Linear Program”. In: S+SSPR 2016. Vol. 10029. LNCS. Heidelberg: Springer, pp. 485–495.

Riesen, Kaspar and Horst Bunke (2008). “IAM Graph Database Repository for Graph Based Pattern Recognition and Machine Learning”. In: S+SSPR 2008. Ed. by Niels da Vitoria Lobo et al. Vol. 5342. LNCS. Springer, pp. 287–297.

Riesen, Kaspar, Stefan Fankhauser, and Horst Bunke (2007). “Speeding Up Graph Edit Distance Computation with a Bipartite Heuristic”. In: MLG 2007. Ed. by Paolo Frasconi, Kristian Kersting, and Koji Tsuda, pp. 21–24. URL: %7Bhttp://mlg07.dsi.unifi.it/pdf/02_Riesen.pdf%7D.