Quasimetric Graph Edit Distance as a Compact Quadratic Assignment Problem

D. B. Blumenthal[1], É. Daller[2], S. Bougleux[2], L. Brun[2], J. Gamper[1]

[1]Free University of Bozen-Bolzano, Faculty of Computer Science, Bolzano, Italy
{david.blumenthal,gamper}@inf.unibz.it

[2]Normandie Université, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France
{evariste.daller,bougleux}@unicaen.fr, luc.brun@ensicaen.fr

Contributions

- **graph edit distance (GED):** minimal cost of transforming one graph into another by substituting, removing, and inserting nodes and edges
- widely used in Pattern Recognition community but NP-hard to compute
- **one state of the art approach [1–3]:**
  1. transform GED to instance of **quadratic assignment problem (QAP)**
  2. use well-performing heuristics for QAP for approximating GED
- **our assumption:** edit costs are quasimetric, i. e., satisfy triangle inequality
- **our contributions:**
  1. reduce size of QAP-instance constructed by the transformation
  2. speed up QAP-based heuristics by using the smaller instances

Quadratic Assignment Problem (QAP)

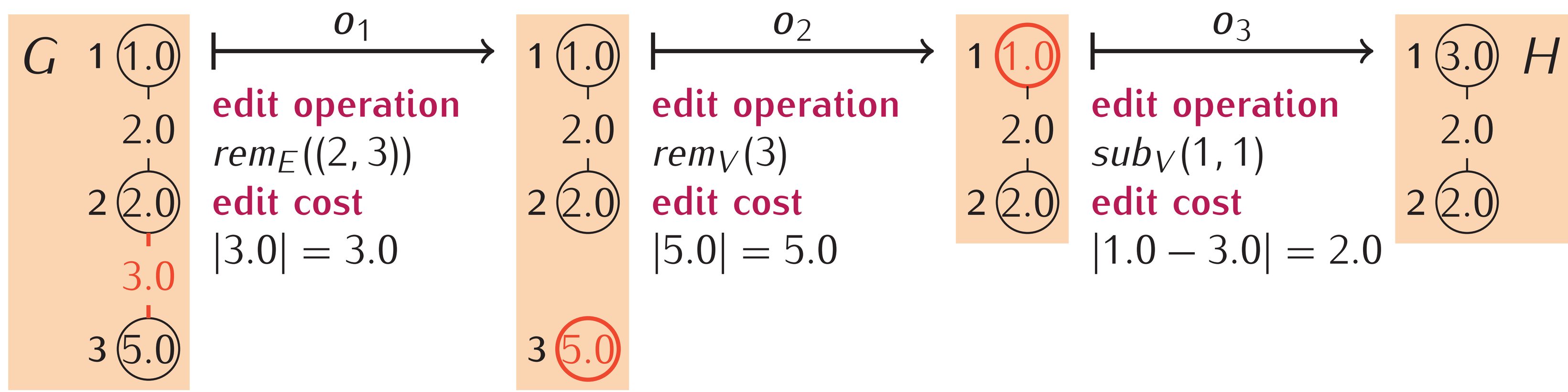
- $\mathrm{QAP}(\mathbf{C}) := \min\{\mathrm{vec}(\mathbf{X})^T\mathbf{C}\,\mathrm{vec}(\mathbf{X}) \mid \mathbf{X} \text{ max. partial permutation matrix}\}$
- cost matrix $\mathbf{C} \in \mathbb{R}^{(N\cdot M)\times(N\cdot M)}$, assignment matrix $\mathbf{X} \in \{0,1\}^{N\times M}$
- **sizes of instances constructed by transformations from GED:**
  - **baseline transformation [1]:** $N = M = |V^G| + |V^H|$
  - **first improvement [2]:** $N = |V^G| + 1$, $M = |V^H| + 1$; uses non-standard version of QAP $\rightsquigarrow$ QAP-based heuristics must be adapted
  - **our transformation:** $N = |V^G|$, $M = |V^H|$; uses standard version of QAP like baseline $\rightsquigarrow$ QAP-based heuristics can be used off-the-shelf

Graph Edit Distance (GED)

- $\mathrm{GED}(G,H) := \min\{c(P) \mid P \text{ edit path between } G \text{ and } H\}$
- $G = (V^G, E^G)$ and $H = (V^H, E^H)$ are attributed graphs
- $P = (o_1, \ldots, o_r)$ is sequence of edit operations transforming G into H
- **edit operations and edit costs ($i \in V^G$, $k \in V^H$, $(i,j) \in V^G$, $(k,l) \in E^H$):**

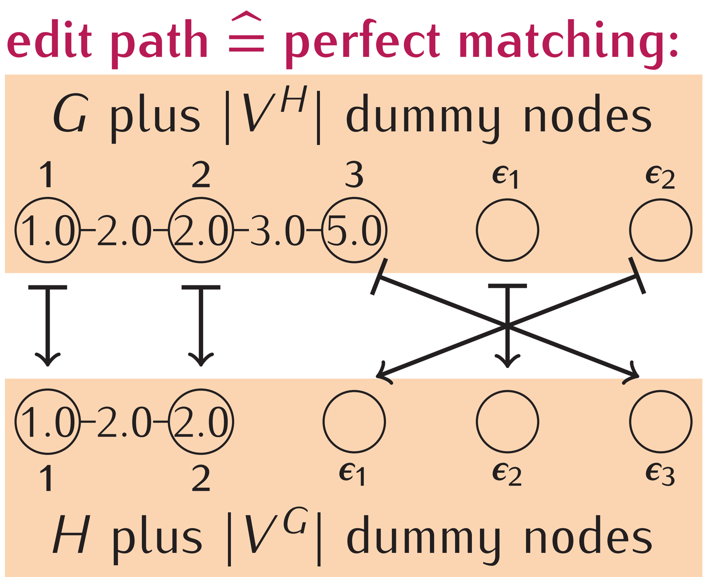
node edit operation	edit cost	edge edit operation	edit cost
$sub_V(i,k)$	$c_V(i,k)$	$sub_E((i,j),(k,l))$	$c_E((i,j),(k,l))$
$rem_V(i)$	$c_V(i,\epsilon)$	$rem_E((i,j))$	$c_E((i,j),\epsilon)$
$ins_V(k)$	$c_V(\epsilon,k)$	$ins_E((k,l))$	$c_E(\epsilon,(k,l))$

- **edit path cost:** $c(P) = \sum_{i=1}^{r} c(o_i)$
- **example with quasimetric edit costs and attributes $\alpha, \beta \in \mathbb{R}_{\geq 0}$:**
  - $sub_V(\cdot,\cdot)$, $sub_E(\cdot,\cdot)$: edit cost $= |\alpha - \beta|$
  - $rem_V(\cdot)$, $rem_E(\cdot)$, $ins_V(\cdot)$, $ins_E(\cdot)$: edit cost $= |\alpha|$
  - **edit path cost:** $c(P) = c(o_1) + c_(o_2) + c_(o_3) = 3.0 + 5.0 + 2.0 = 10.0$



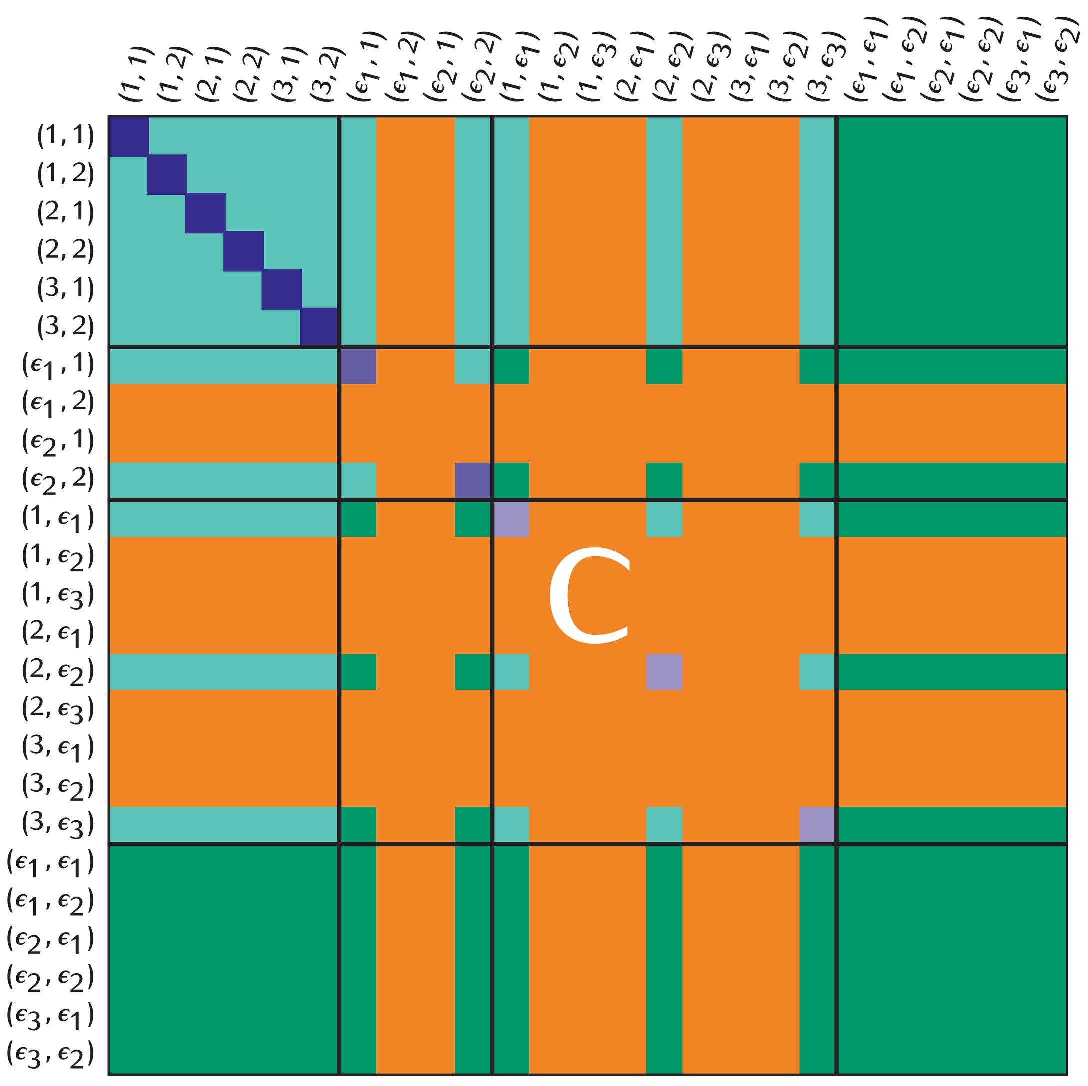
Baseline Transformation

- **edit path $\widehat{=}$ perfect matching:**



- **as assignment matrix:**

$$\mathbf{X} = \begin{array}{c|ccccc} & 1 & 2 & \epsilon_1 & \epsilon_2 & \epsilon_3 \\ \hline 1 & 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 \\ \epsilon_1 & 0 & 0 & 0 & 1 & 0 \\ \epsilon_2 & 0 & 0 & 1 & 0 & 0 \end{array}$$



- **node substitution cost:** $c_{(i,k),(i,k)} = c_V(i,k)$
- **node insertion cost:** $c_{(\epsilon_k,k),(\epsilon_k,k)} = c_V(\epsilon,k)$
- **node removal cost:** $c_{(i,\epsilon_i),(i,\epsilon_i)} = c_V(i,\epsilon_i)$
- **edge edit operation cost:** $c_{(i,k),(j,l)} = 0.5 \cdot (\delta_{(i,j)\in E^G}\delta_{(k,l)\in E^H} c_E((i,j),(k,l)) + \delta_{(i,j)\notin E^G}\delta_{(k,l)\in E^H} c_E(\epsilon,(k,l)) + \delta_{(i,j)\in E^G}\delta_{(k,l)\notin E^H} c_E((i,j),\epsilon))$
- **forbidden assignment cost:** $c_{(i,k),(j,l)} = \infty$
- **free assignment cost:** $c_{(i,k),(j,l)} = 0$

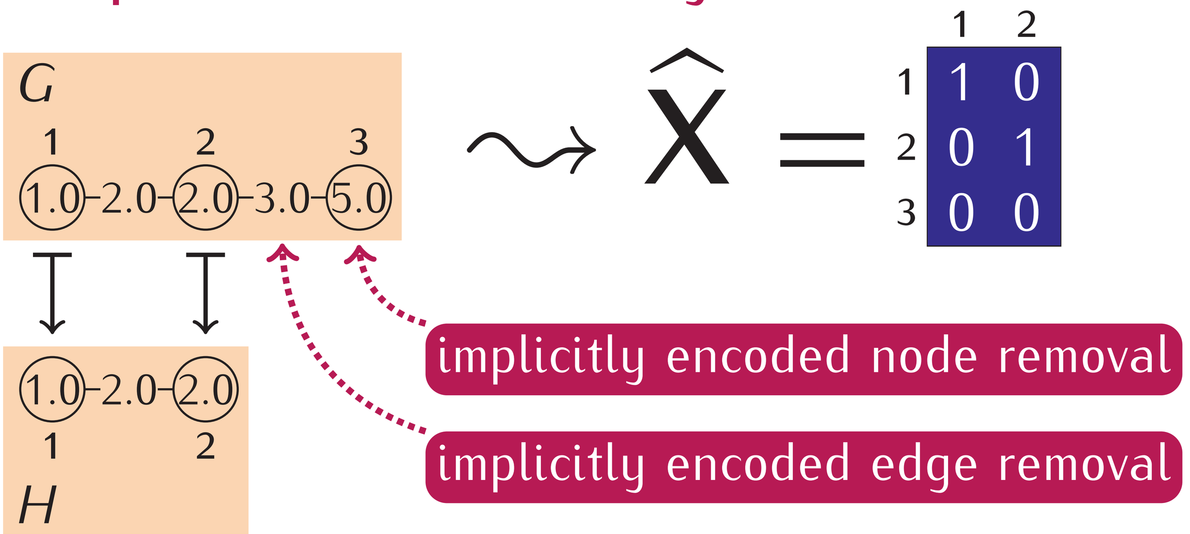
Experiments

- **tested QAP-based heuristic:** mIPFP (conditional gradient descent for QAP) [3]; best available QAP-based heuristic for GED
- **compared transformations:** baseline [1], non-standard [2], transformation proposed in this paper
- **metrics:** computed distance (d), error (e), runtime in seconds (t)

transformation	d	e	t
	alkane		
baseline	15.37	0.023	0.41
non-standard	15.34	0.009	0.22
this paper	15.39	0.062	0.15
	acyclic		
baseline	16.77	0.035	0.24
non-standard	16.73	0.0076	0.13
this paper	16.81	0.079	0.06
	mao		
baseline	33.4	–	2.9
non-standard	33.3	–	0.8
this paper	39.7	–	1.5
	pah		
baseline	36.7	–	3.14
non-standard	36.6	–	1.17
this paper	36.7	–	0.89

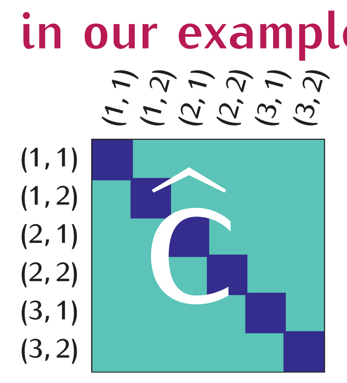
Compact Transformation for Quasimetric Edit Costs

- **key property for quasimetric edit costs:**
  - exactly $\max\{0, |V^G| - |V^H|\}$ node removals in optimal edit path
  - exactly $\max\{0, |V^H| - |V^G|\}$ node insertions in optimal edit path
- **edit path $\widehat{=}$ maximal matching:**



$$\rightsquigarrow \widehat{\mathbf{X}} = \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 1 & 0 \\ 2 & 0 & 1 \\ 3 & 0 & 0 \end{array}$$

- **possibly implicitly encoded edit operations:**
  - $|V^G| > |V^H|$: removals of nodes $k \in V^G$ and edges $(k,l) \in E^G$
  - $|V^G| < |V^H|$: insertions of nodes $k \in V^H$ and edges $(k,l) \in E^H$
- **strategy:** define compact QAP-instance $\widehat{\mathbf{C}}$ by including the costs of possibly implicitly encoded edit operations in top-left part of $\mathbf{C}$
- **in our example:**



- **node substitution cost:** $\widehat{c}_{(i,k),(i,k)} = c_{(i,k),(i,k)} - c_V(i,\epsilon)$
- **edge edit operation cost:** $\widehat{c}_{(i,k),(i,k)} = c_{(i,k),(i,k)} - 1.5 \cdot c_E((i,j),\epsilon)$
- **general formula for node substitution cost:**
  $\widehat{c}_{(i,k),(i,k)} = c_{(i,k),(i,k)} - \delta_{|V^G|>|V^H|} c_V(i,\epsilon) - \delta_{|V^G|<|V^H|} c_V(\epsilon,k)$
- **general formula for edge edit operation cost:**
  $\widehat{c}_{(i,k),(i,k)} = c_{(i,k),(i,k)} - 1.5 \cdot (\delta_{|V^G|>|V^H|} c_E((i,j),\epsilon) - \delta_{|V^G|<|V^H|} c_E(\epsilon,(k,l)))$

main theorem

$$\mathrm{GED}(G,H) = \mathrm{QAP}(\widehat{\mathbf{C}}) + \delta_{|V^G|>|V^H|}\Big(\sum_{(i,j)\in E^G} c_E((i,j),\epsilon) + \sum_{i\in V^G} c_V(i,\epsilon)\Big) + \delta_{|V^G|<|V^H|}\Big(\sum_{(k,l)\in E^H} c_E(\epsilon,(k,l)) + \sum_{k\in V^H} c_V(\epsilon,k)\Big)$$

References

[1] S. Bougleux, L. Brun, V. Carletti, P. Foggia, B. Gaüzère, and M. Vento, "Graph edit distance as a quadratic assignment problem," *Pattern Recogn. Lett.*, vol. 87, pp. 38–46, 2017.

[2] S. Bougleux, B. Gaüzère, and L. Brun, "Graph edit distance as a quadratic program," in *ICPR*, 2016, pp. 1701–1706.

[3] É. Daller, S. Bougleux, B. Gaüzère, and L. Brun, "Approximate graph edit distance by several local searches in parallel," in *ICPRAM*, 2018, pp. 149–158.

Presented at ICPR 2018, Beijing, China, August 20–24, 2018