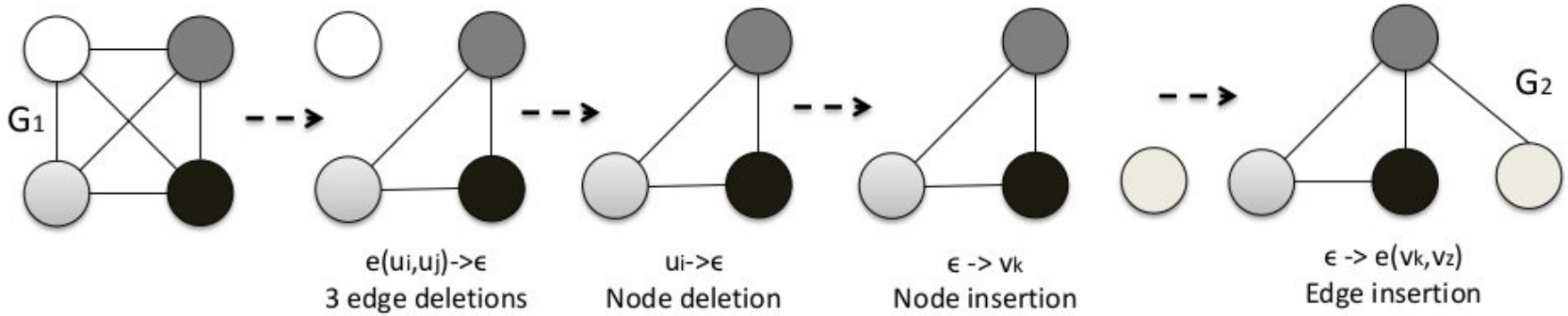
Clique-Based Encodings for Graph Edit Distance

Paulius Dilkas
Project Supervisor: Ciaran McCreesh

School of Computing Science

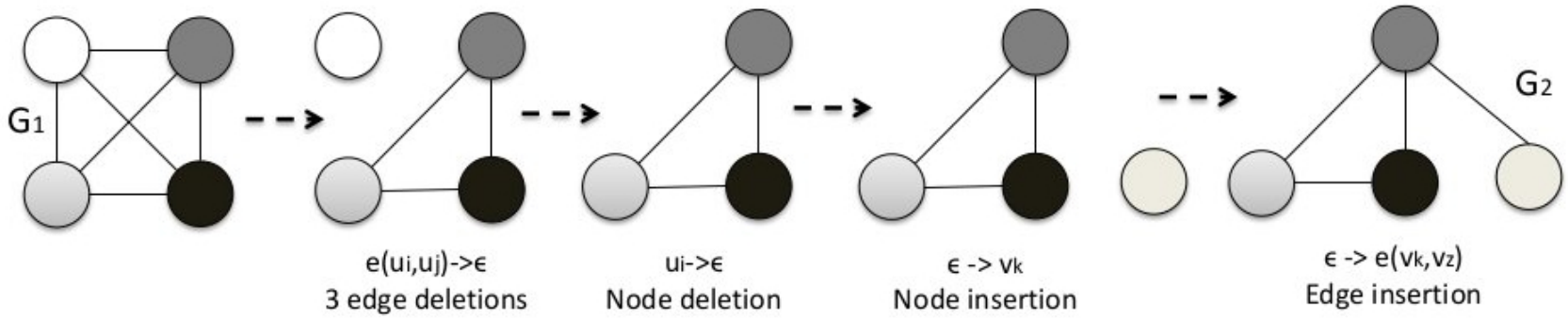
5th September 2017

The problem

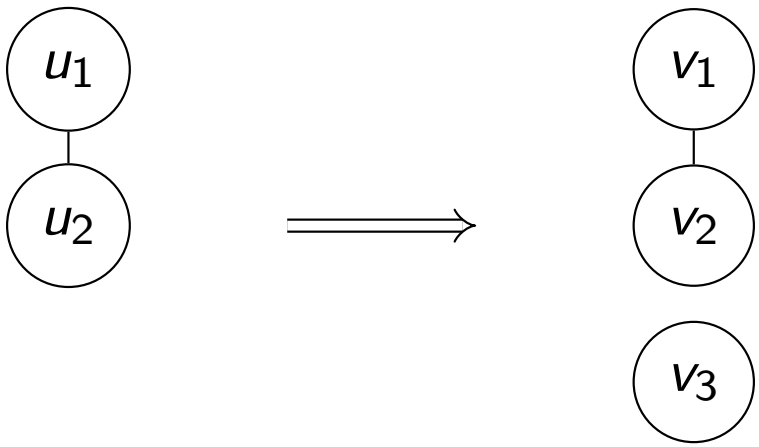


Source: [1]

The problem



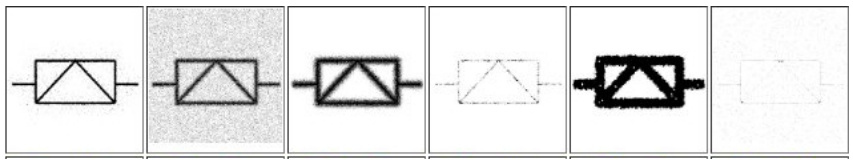
Source: [1]



Available operations

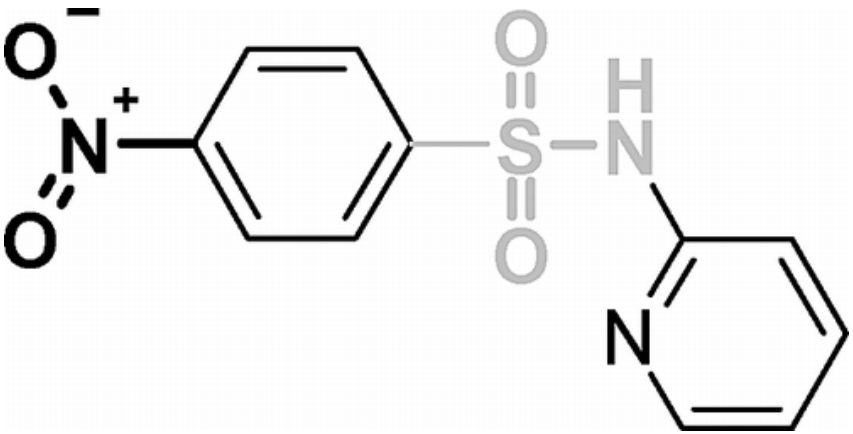
- node insertion $(\epsilon \rightarrow v_i)$,
- node deletion $(u_i \rightarrow \epsilon)$,
- node substitution $(u_i \rightarrow v_j)$,
- edge insertion $(\epsilon \rightarrow e_i)$,
- edge deletion $(e_i \rightarrow \epsilon)$,
- edge substitution $(e_i \rightarrow e_j)$.

Use cases

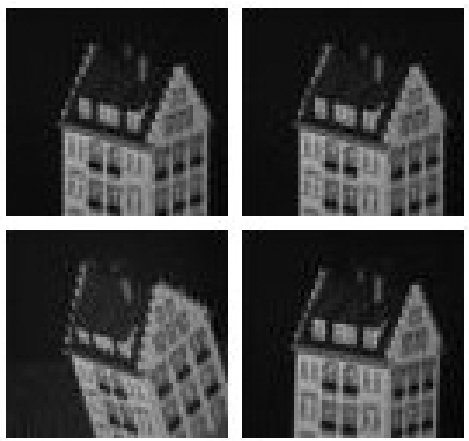


Source: International Symbol Recognition Contest GREC'2005

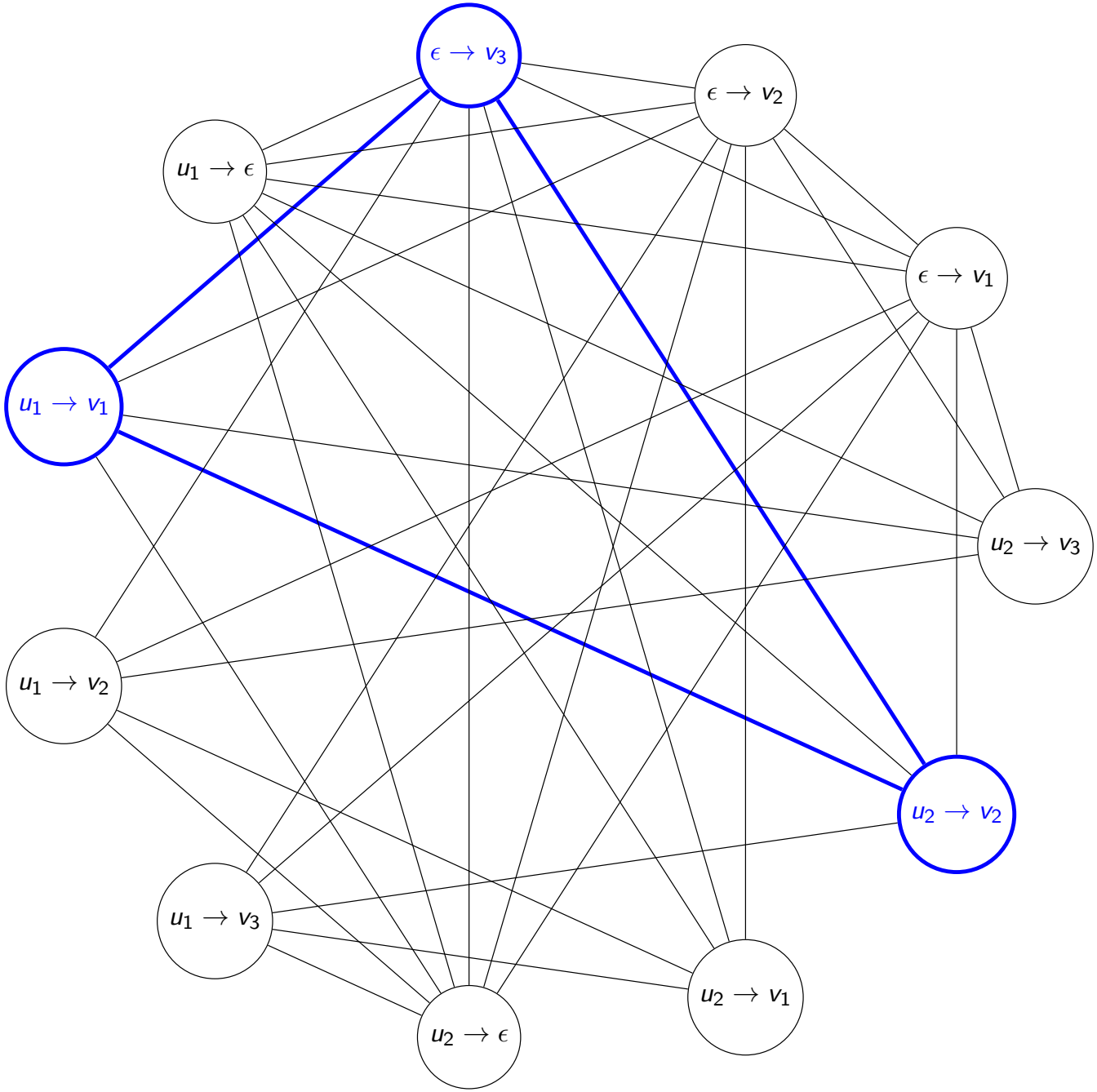
Source: RCSB Protein Data Bank



Source: [3]



Source: CMU/VASC Image Database

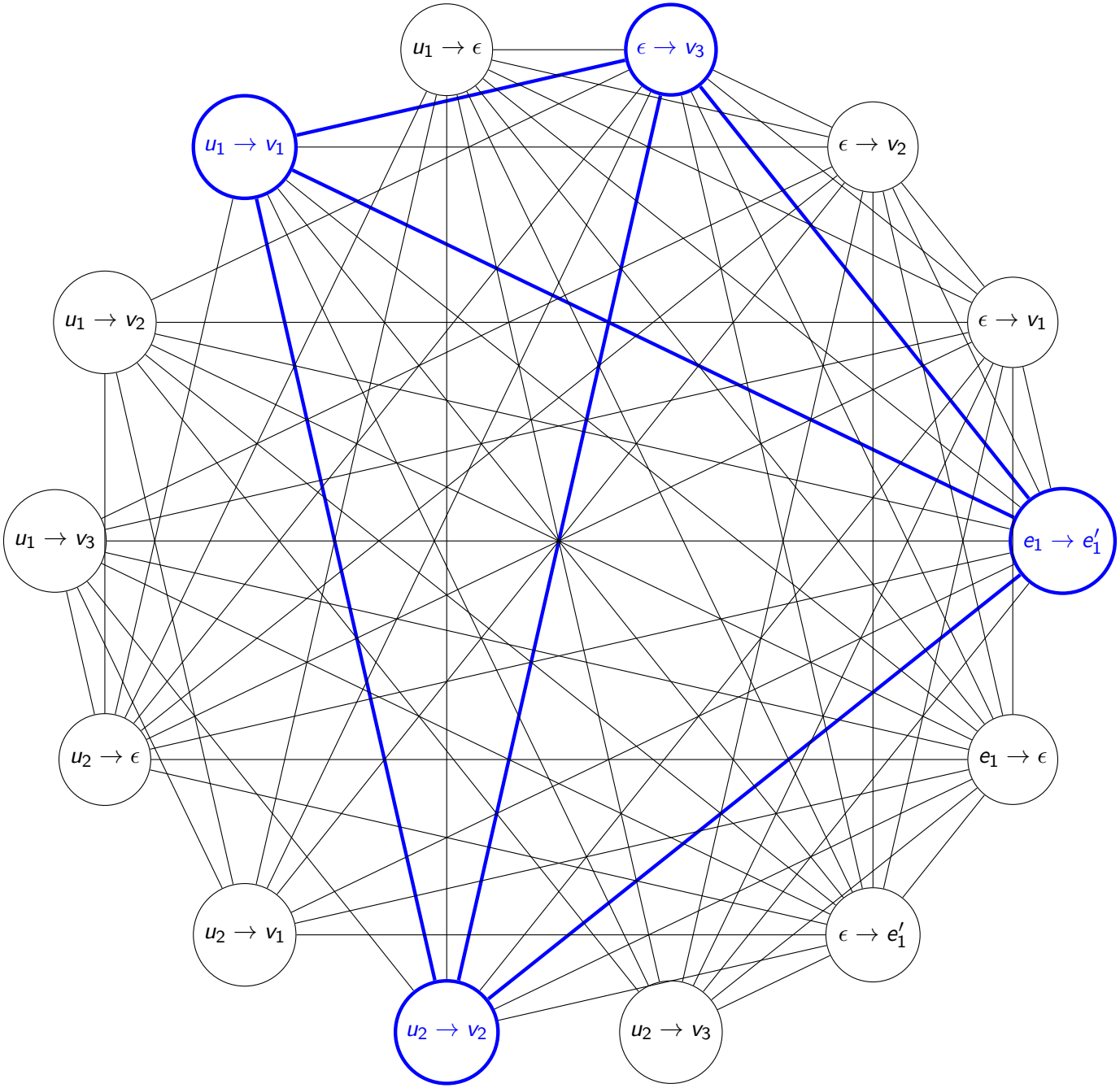

$\epsilon \to v_3$
$\epsilon \to v_2$
$u_1 \to \epsilon$
$\epsilon \to v_1$
$u_1 \to v_1$
$u_2 \to v_3$
$u_1 \to v_2$
$u_2 \to v_2$
$u_1 \to v_3$
$u_2 \to v_1$
$u_2 \to \epsilon$

First encoding: weighted vertices & edges

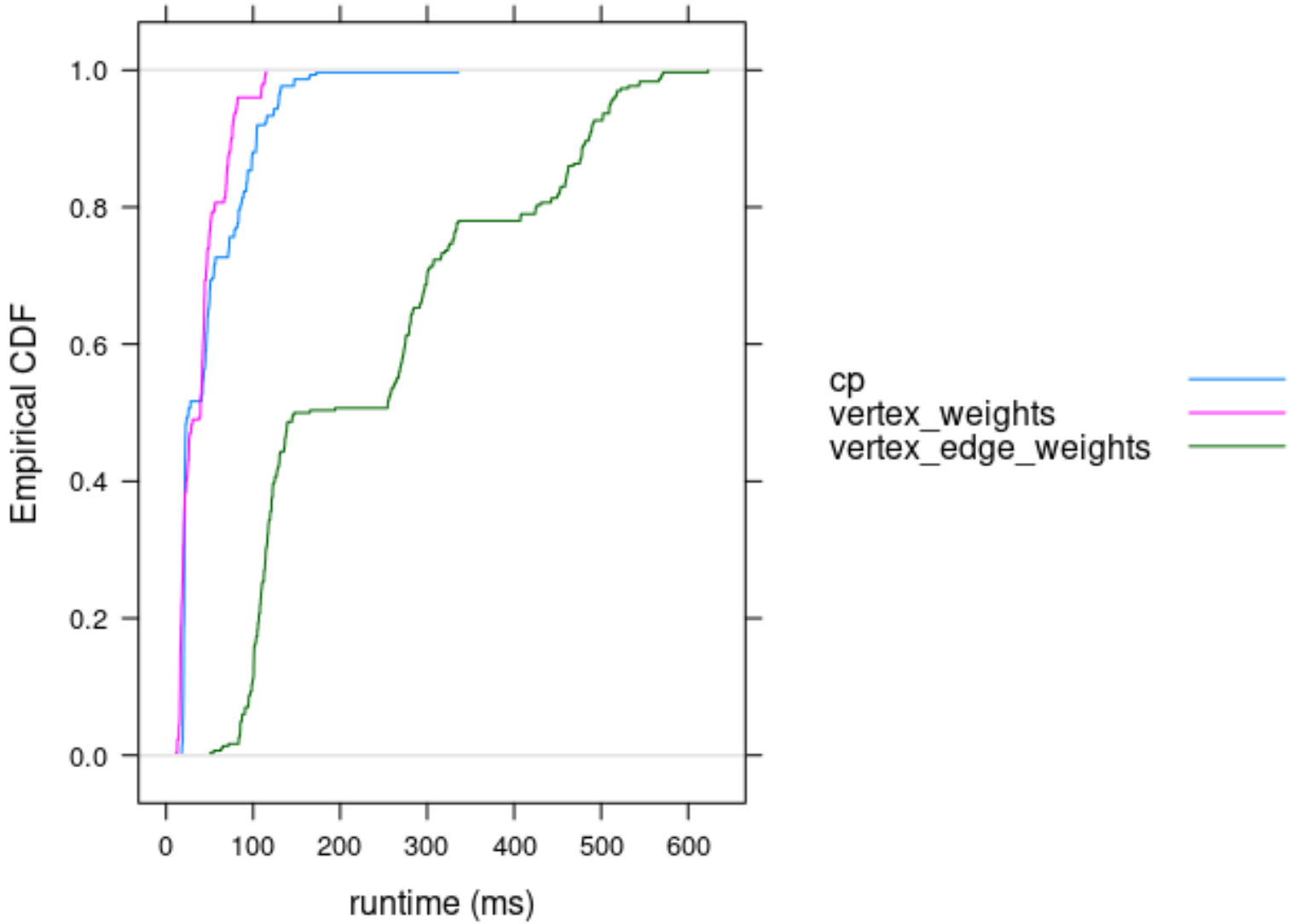
- Each vertex represents an operation on nodes
- Weights of a vertex is equal to cost of operation
- Additional constraint: every node must be involved in some operation

Second encoding: weighted vertices

- Vertices represent ALL possible operations
- Weights correspond to costs of operations
- Additional constraint: every node and edge must be involved in some operation
- The graph is quadratically bigger than the first encoding


$u_1 \to \epsilon$
$\epsilon \to v_3$
$u_1 \to v_1$
$\epsilon \to v_2$
$u_1 \to v_2$
$\epsilon \to v_1$
$u_1 \to v_3$
$e_1 \to e_1'$
$u_2 \to \epsilon$
$e_1 \to \epsilon$
$u_2 \to v_1$
$\epsilon \to e_1'$
$u_2 \to v_2$
$u_2 \to v_3$

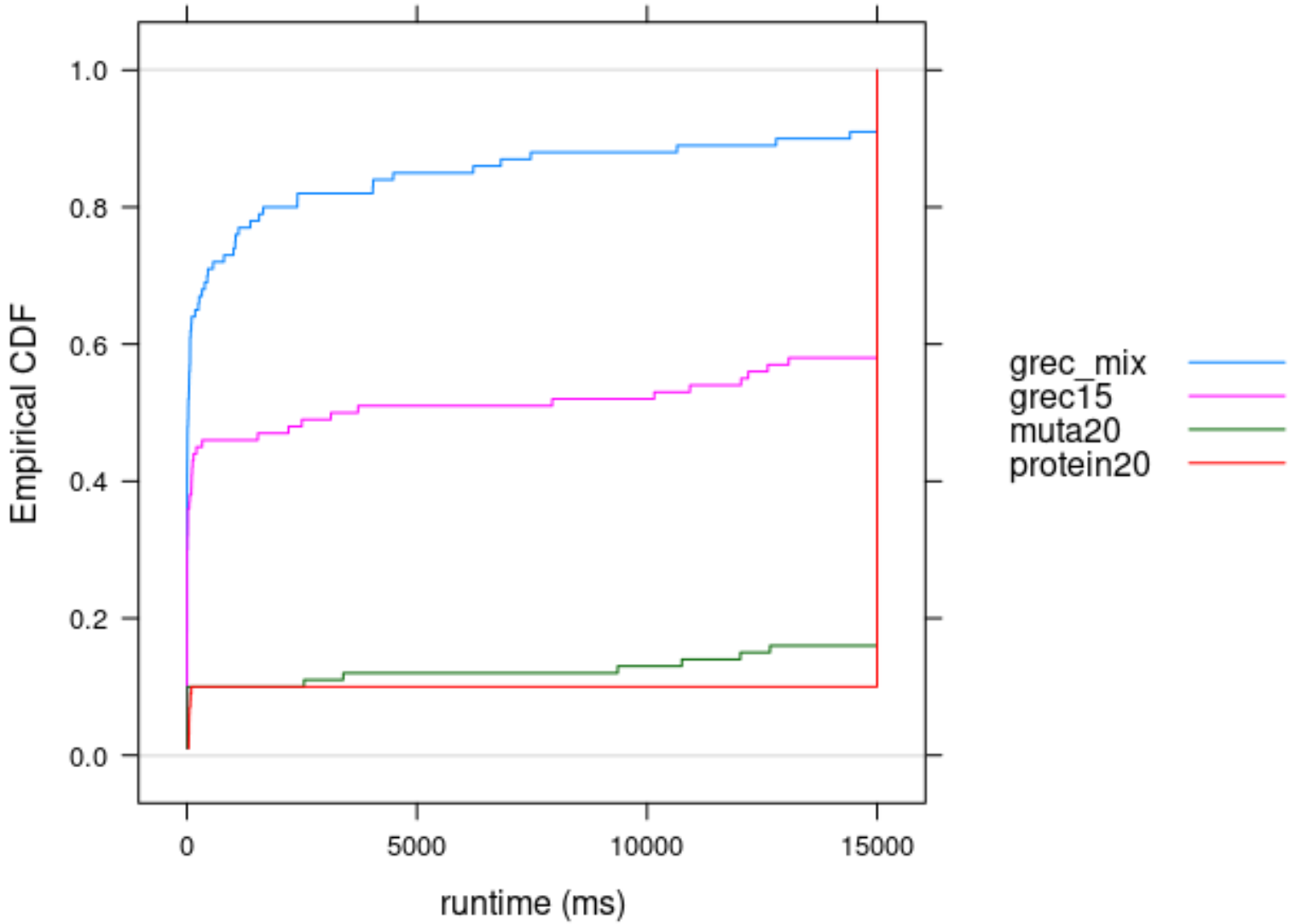
Performance of the CP models



The algorithm

- Branch and bound
- Lower bound function based on W. A. Tavares' colouring [4]
- Bitsets for keeping track of vertices that can still be part of the clique

Performance on different datasets



Compared to other algorithms [2]

- DF-GED: optimal answers in 350 ms with 20-vertex graphs
- My algorithm: answers after 15 s up to 5% higher than optimal for 15-vertex graphs

- A*GED: fails to finish in 350 ms with most of 10-vertex graphs
- My algorithm: maximum running time is 195 ms, mean running time is 34.73 ms

References I


Zeina Abu-Aisheh. "Anytime and Distributed Approaches for Graph Matching". PhD thesis. Université François-Rabelais de Tours, 2016.


Zeina Abu-Aisheh et al. "An Exact Graph Edit Distance Algorithm for Solving Pattern Recognition Problems". In: *ICPRAM 2015 - Proceedings of the International Conference on Pattern Recognition Applications and Methods, Volume 1, Lisbon, Portugal, 10-12 January, 2015.* Ed. by Maria De Marsico, Mário A. T. Figueiredo and Ana L. N. Fred. SciTePress, 2015, pp. 271–278. ISBN: 978-989-758-076-5.


Jeroen Kazius, Ross McGuire and Roberta Bursi. "Derivation and Validation of Toxicophores for Mutagenicity Prediction". In: *Journal of Medicinal Chemistry* 48.1 (2005). PMID: 15634026, pp. 312–320. DOI: 10.1021/jm040835a.

References II

Wladimir Araujo Tavares. "Algoritmos exatos para problema da clique maxima ponderada. (Exact algorithms for the maximum-weight clique problem / Algorithmes pour le problème de la clique de poids maximum)". PhD thesis. University of Avignon, France, 2016. URL: https://tel.archives-ouvertes.fr/tel-01401999.

Contact details

Paulius Dilkas
2146879d@student.gla.ac.uk
or paulius.dilkas@gmail.com

Code and paper avalable at

https://github.com/dilkas/graph-edit-distance