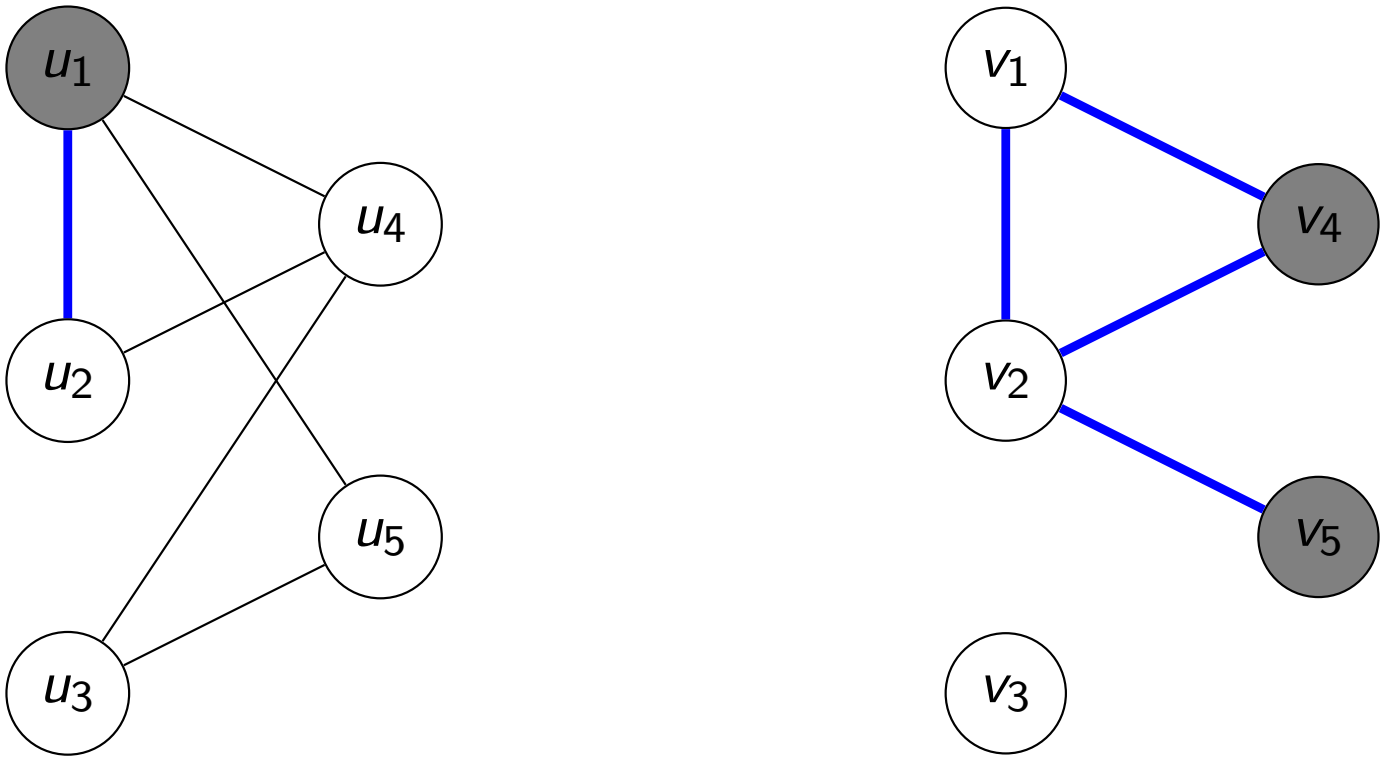
Algorithm Selection for Maximum Common Subgraph

Paulius Dilkas

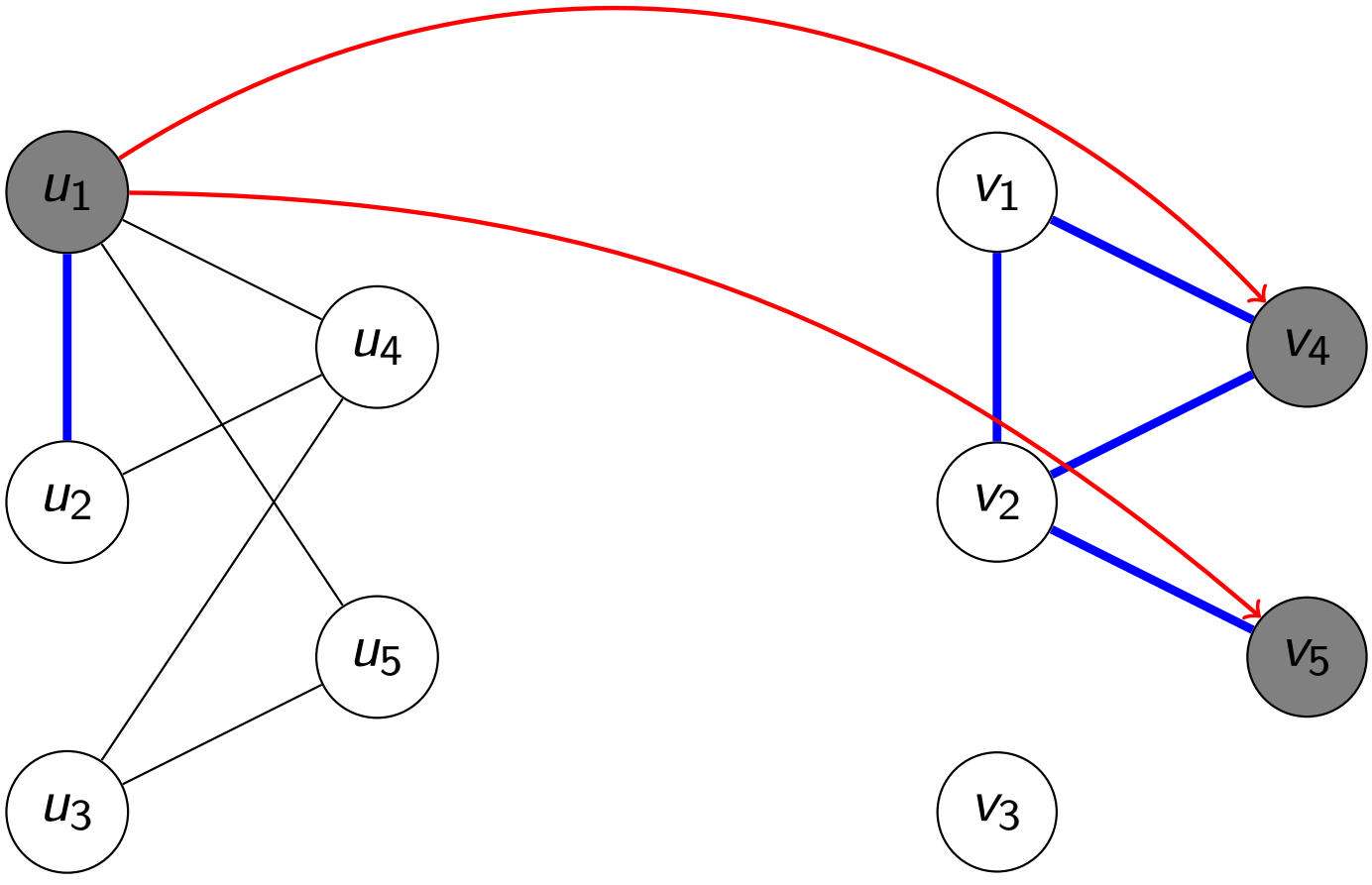
School of Computing Science
University of Glasgow

23rd March 2018

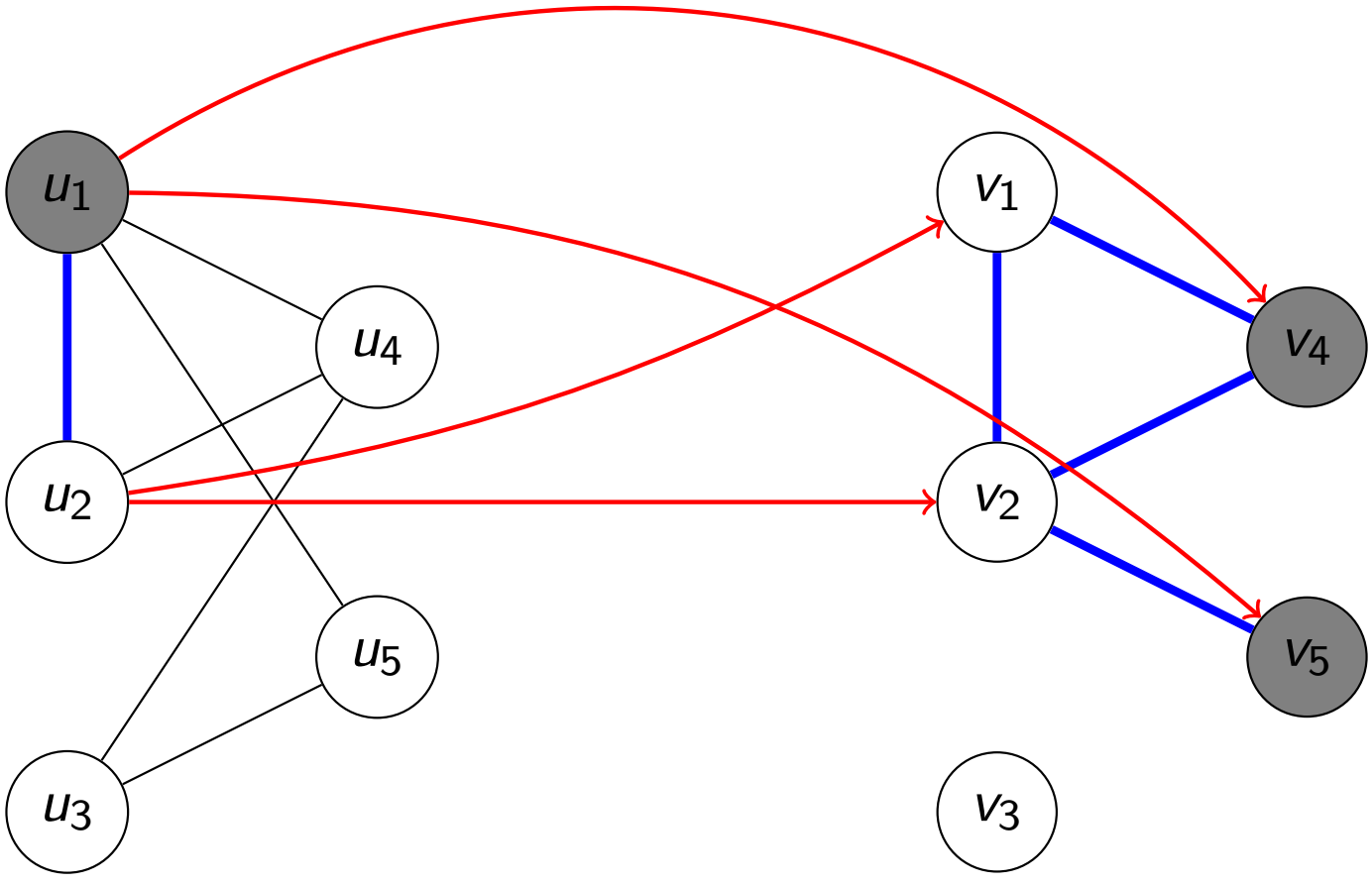
Maximum Common Subgraph



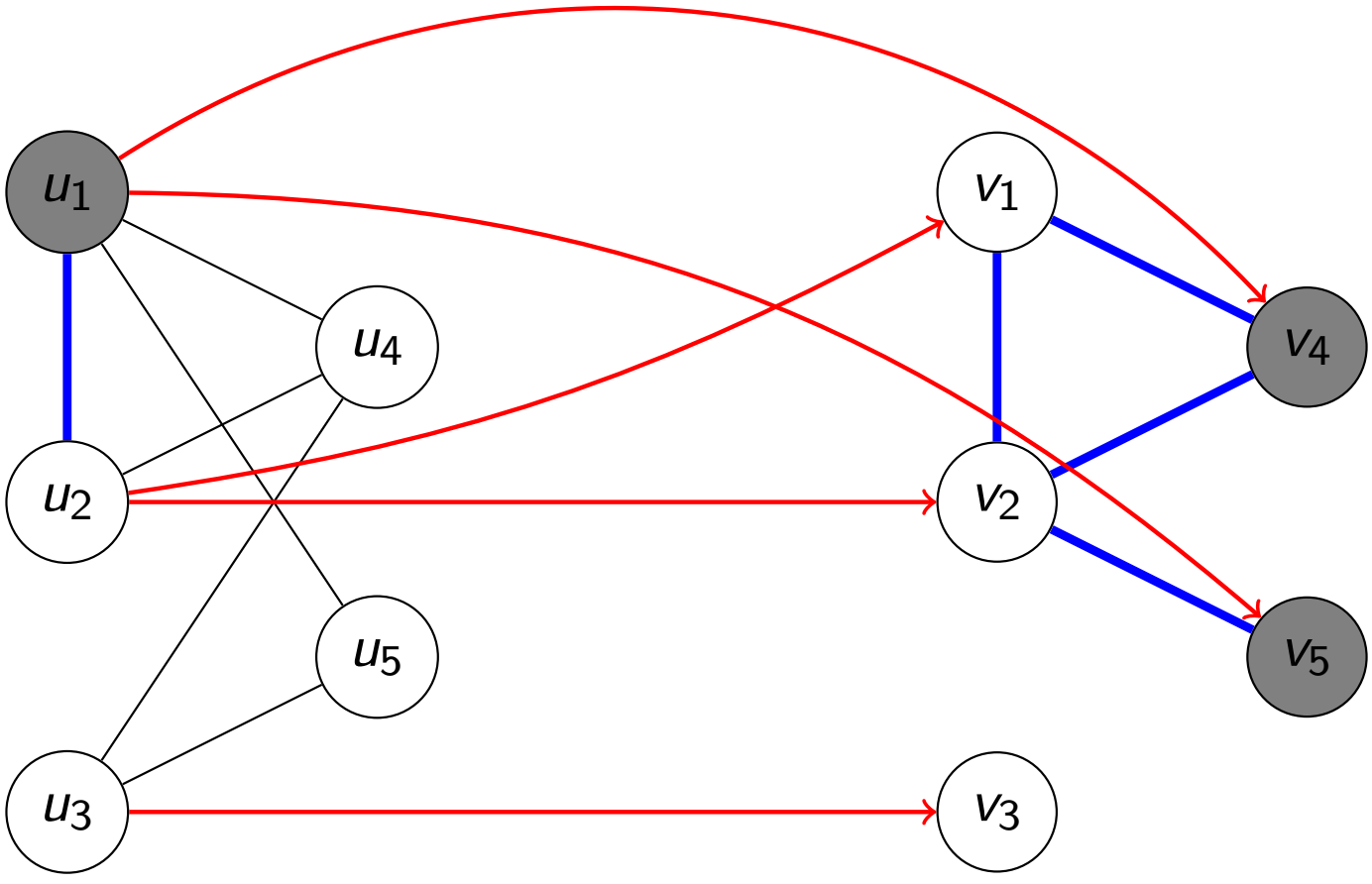
Maximum Common Subgraph



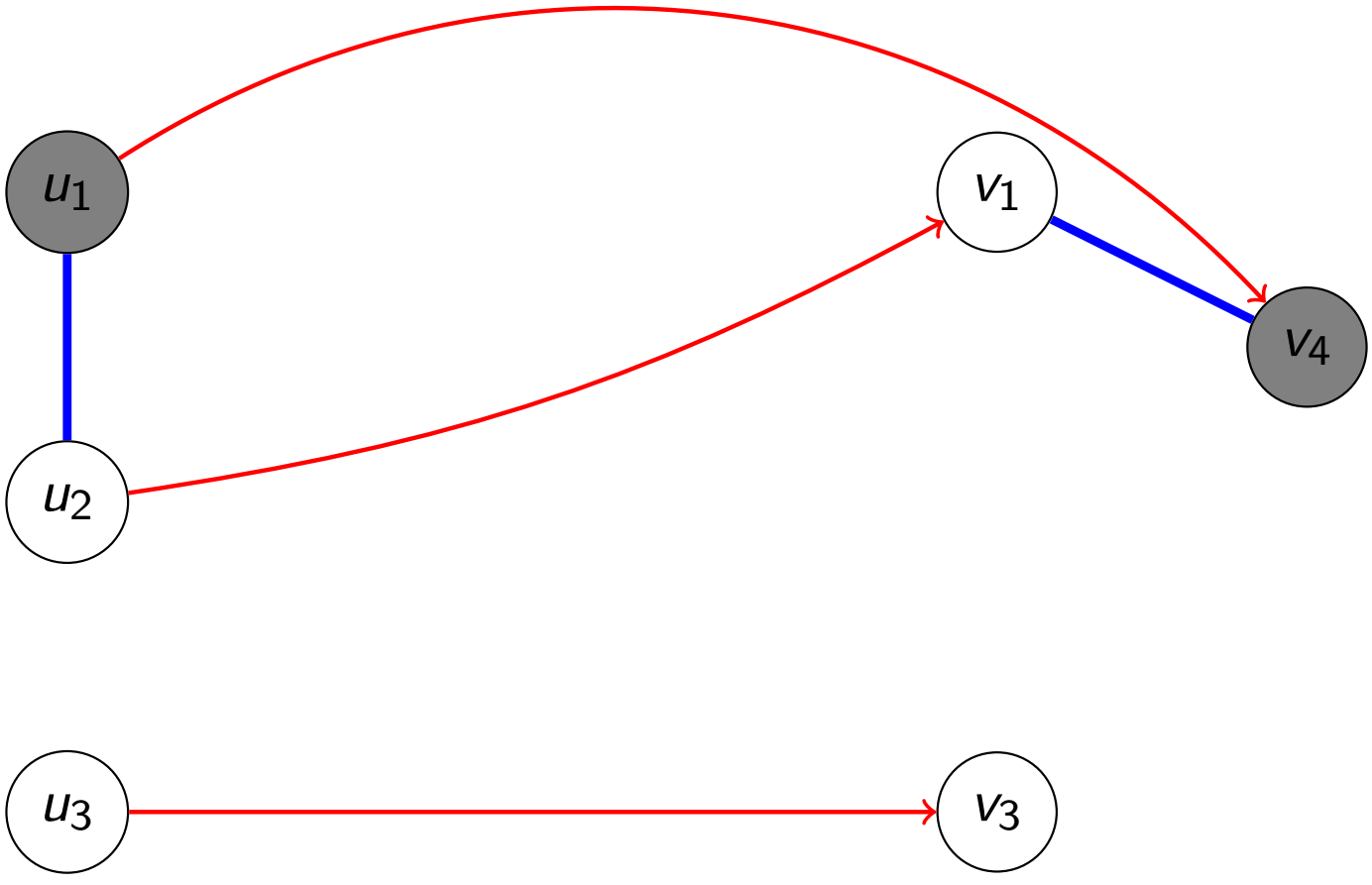
Maximum Common Subgraph



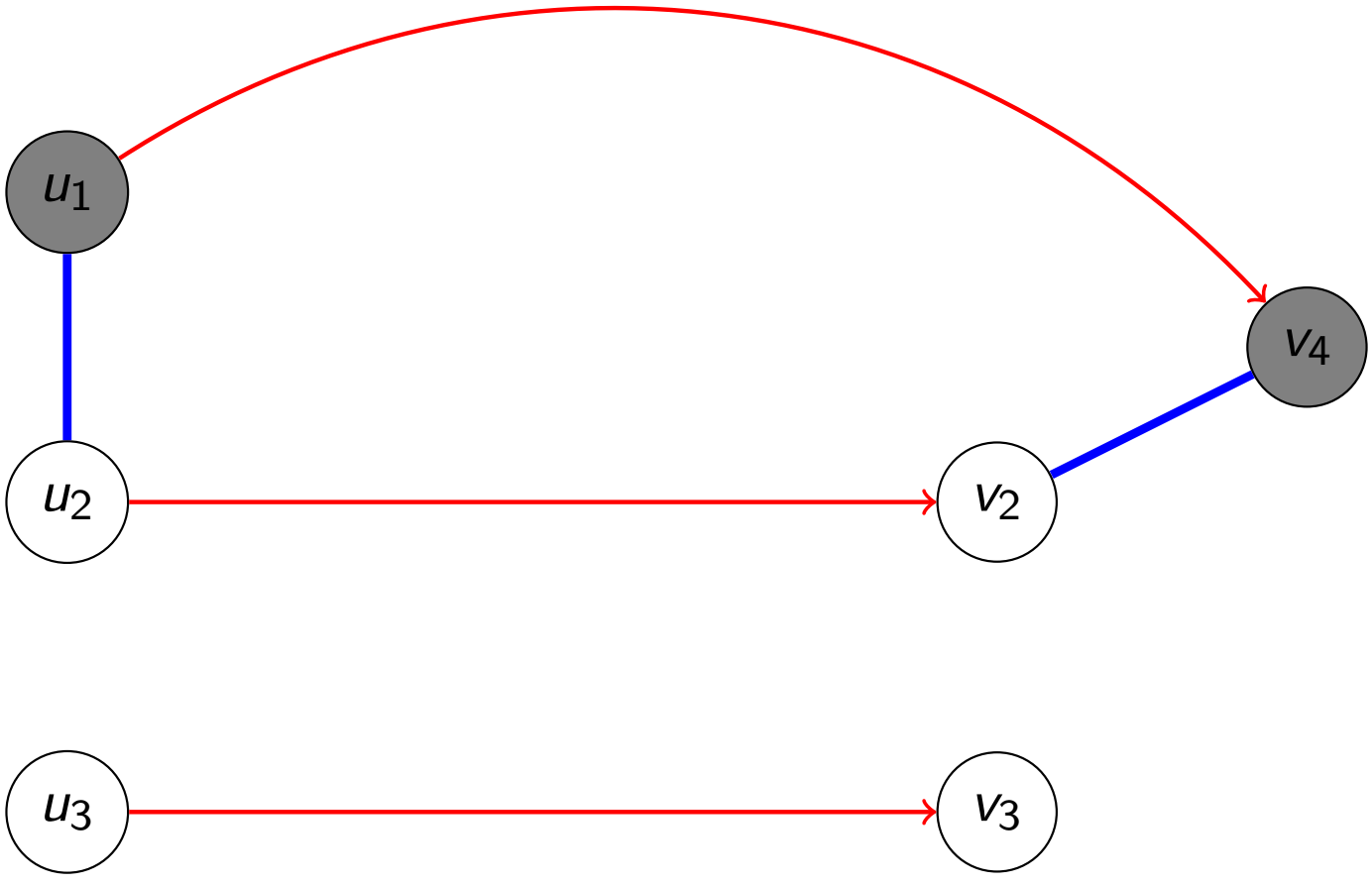
Maximum Common Subgraph



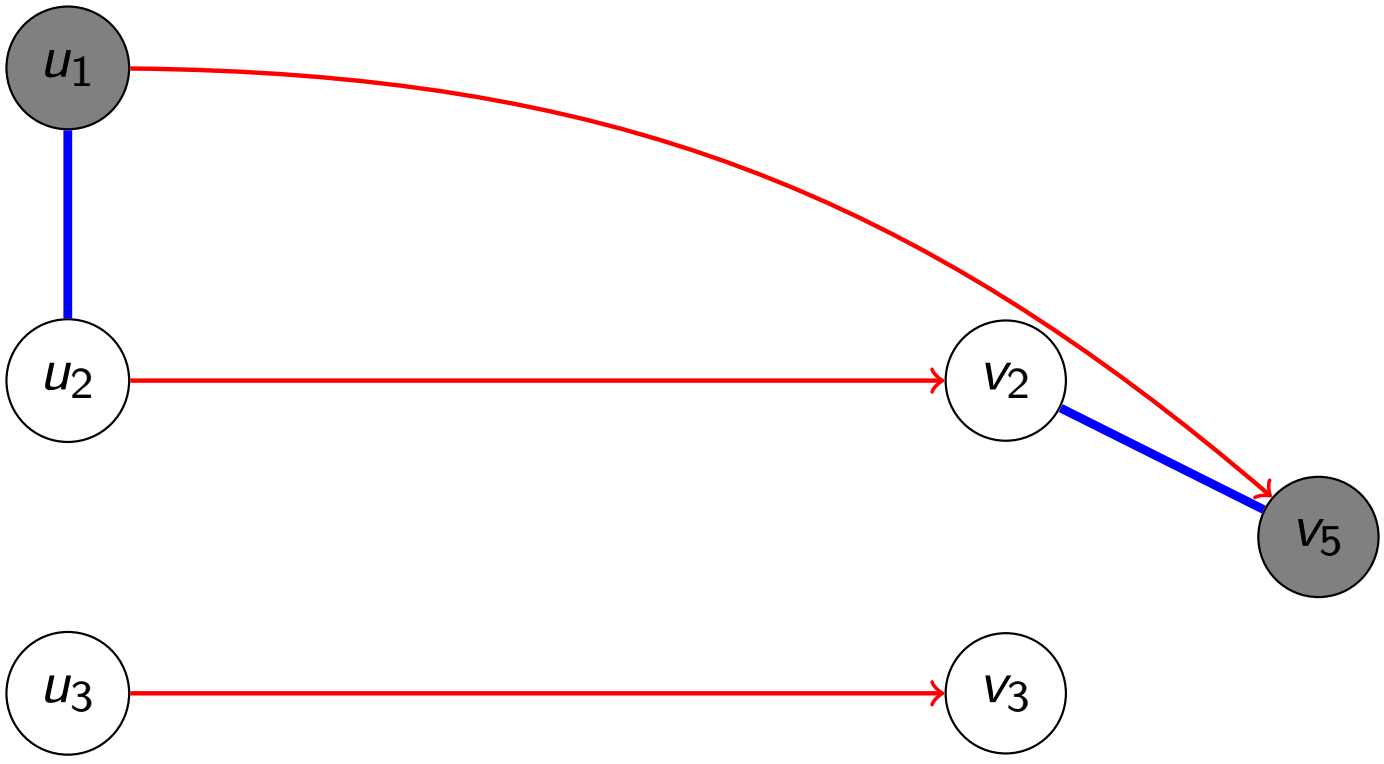
Maximum Common Subgraph



Maximum Common Subgraph



Maximum Common Subgraph



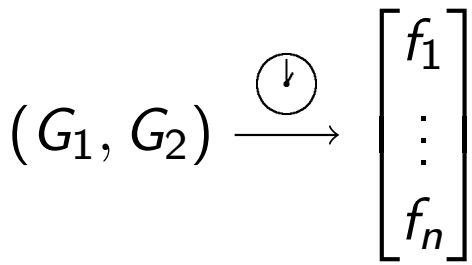
Algorithms

- McSplit, McSplit↓
  - McCreesh, Prosser and Trimble 2017
- clique encoding
  - McCreesh, Ndiaye et al. 2016
- $k\downarrow$
  - Hoffmann, McCreesh and Reilly 2017

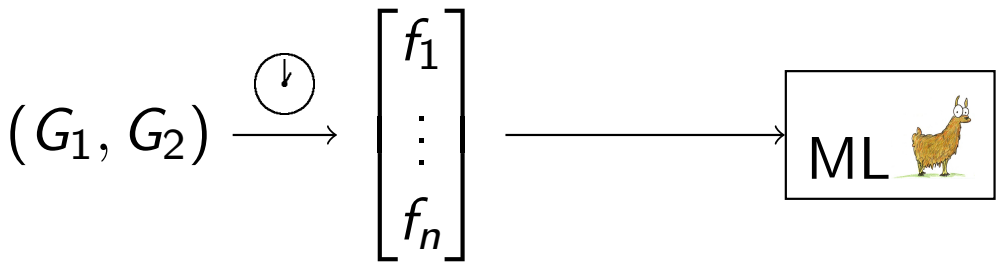
Algorithm Selection

(G_1, G_2)

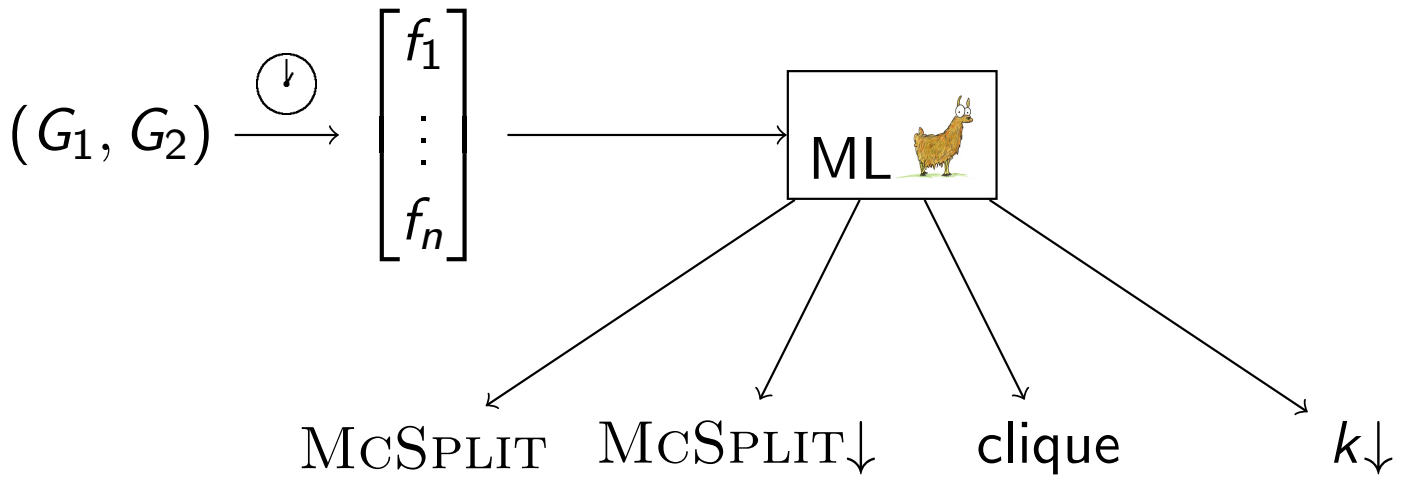
Algorithm Selection



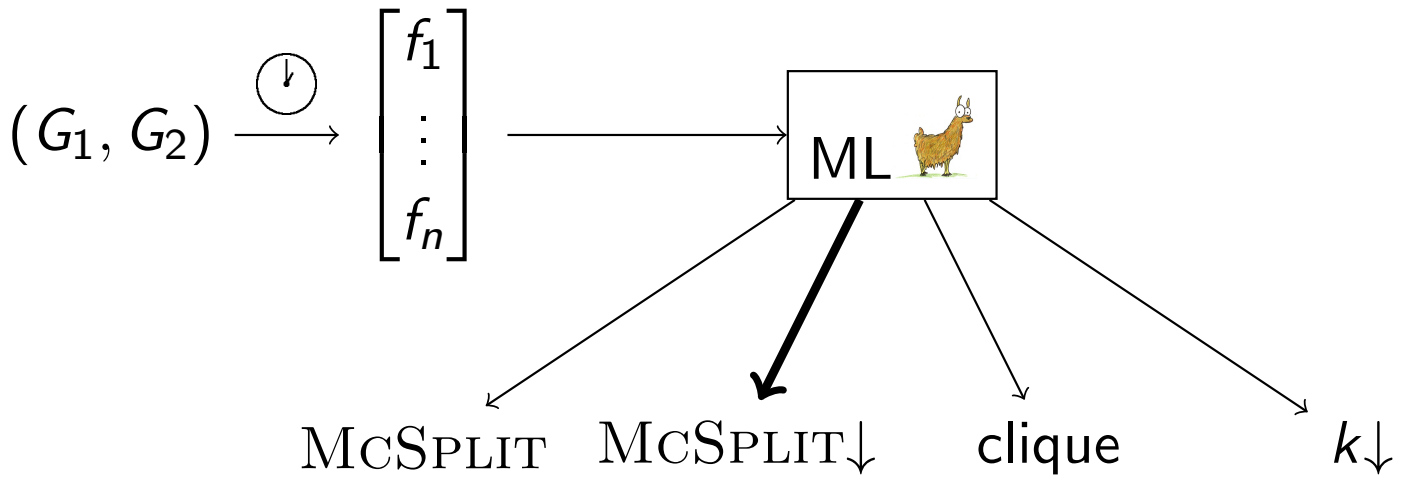
Algorithm Selection



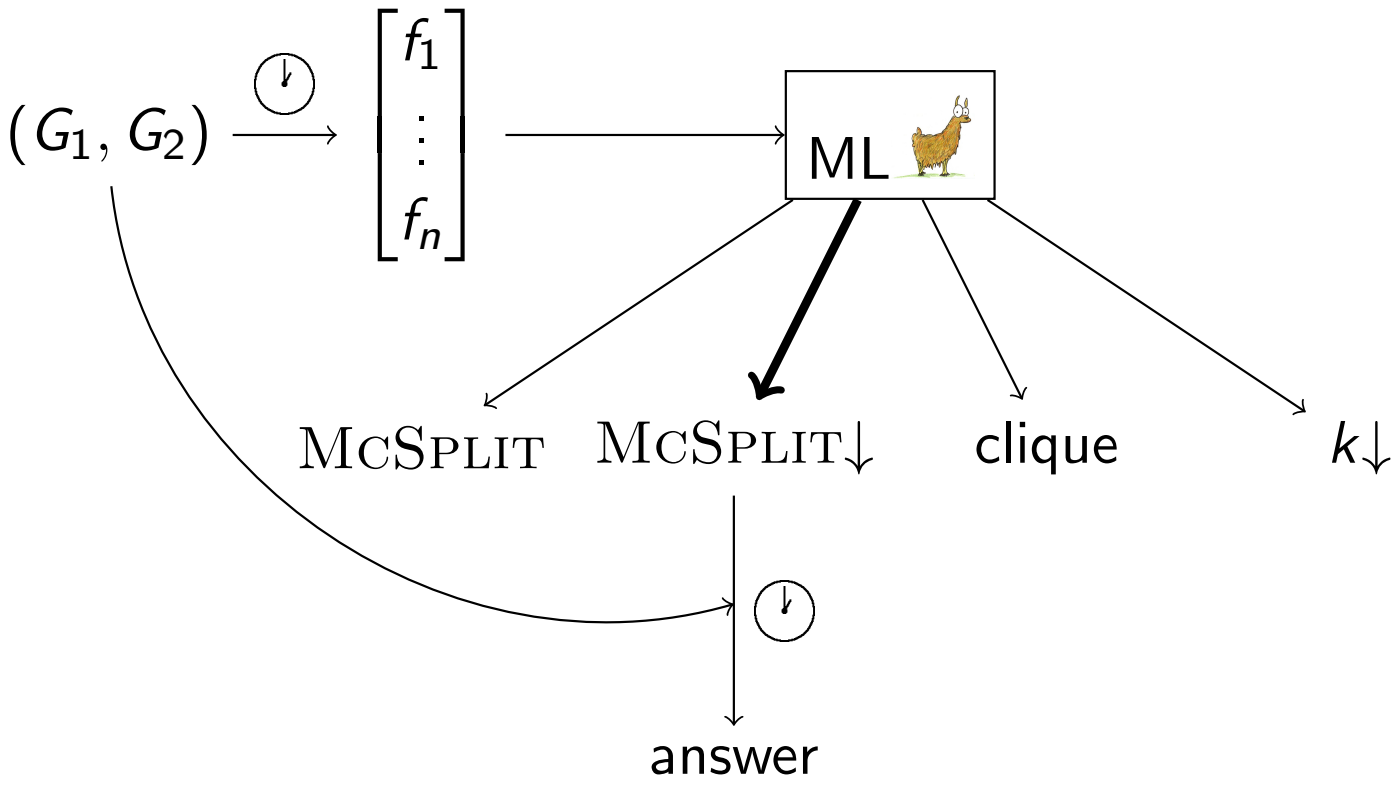
Algorithm Selection



Algorithm Selection



Algorithm Selection



Features (34 in total)

1–8 are from Kotthoff, McCreesh and Solnon 2016

1. number of vertices
2. number of edges
3. mean/max degree
4. density
5. mean/max distance between pairs of vertices
6. number of loops
7. proportion of vertex pairs with distance $\geq$ 2, 3, 4
8. connectedness

Features (34 in total)

1–8 are from Kotthoff, McCreesh and Solnon 2016

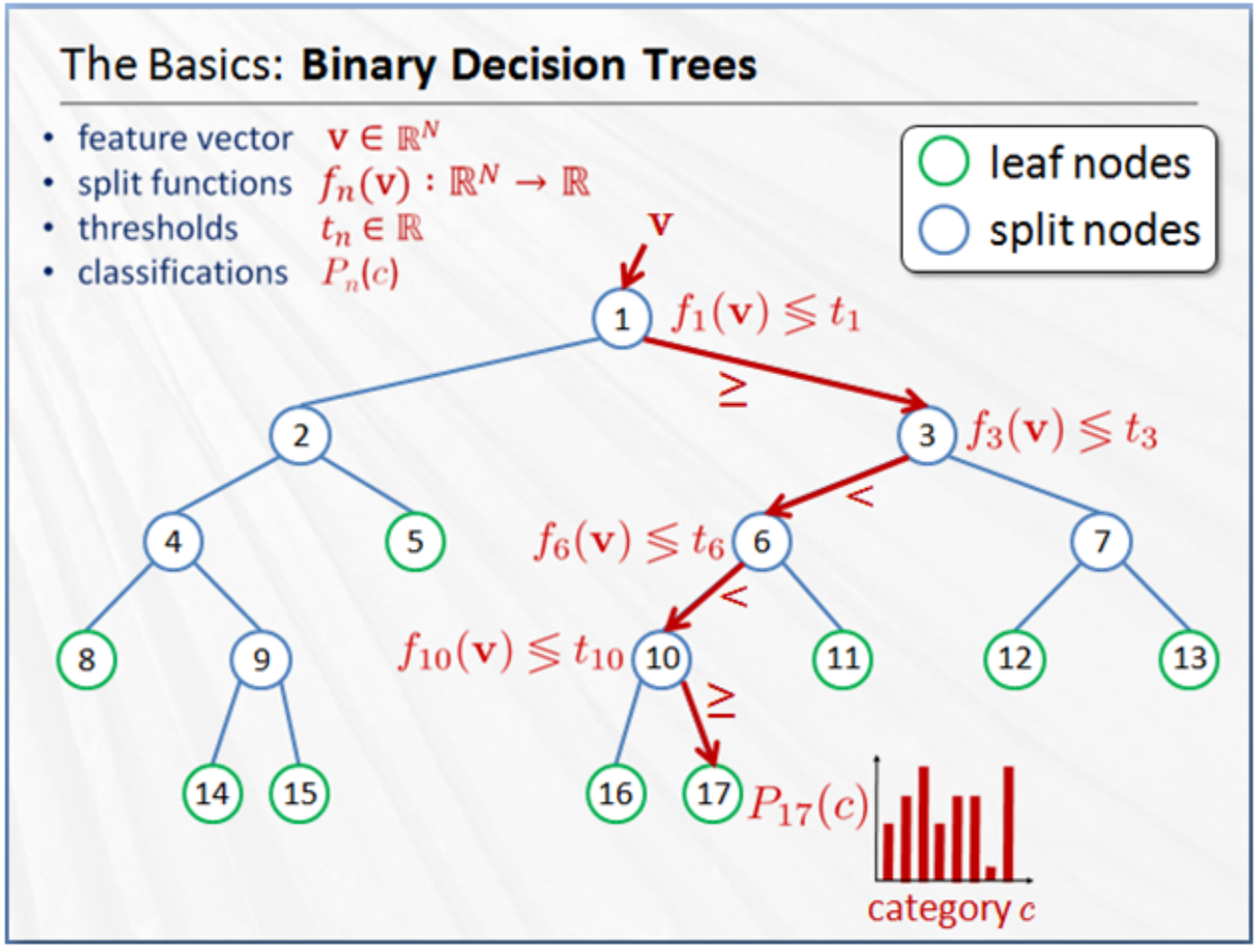
1. number of vertices
2. number of edges
3. mean/max degree
4. density
5. mean/max distance between pairs of vertices
6. number of loops
7. proportion of vertex pairs with distance $\geq$ 2, 3, 4
8. connectedness
9. standard deviation of degrees
10. labelling percentage

Features (34 in total)

1–8 are from Kotthoff, McCreesh and Solnon 2016

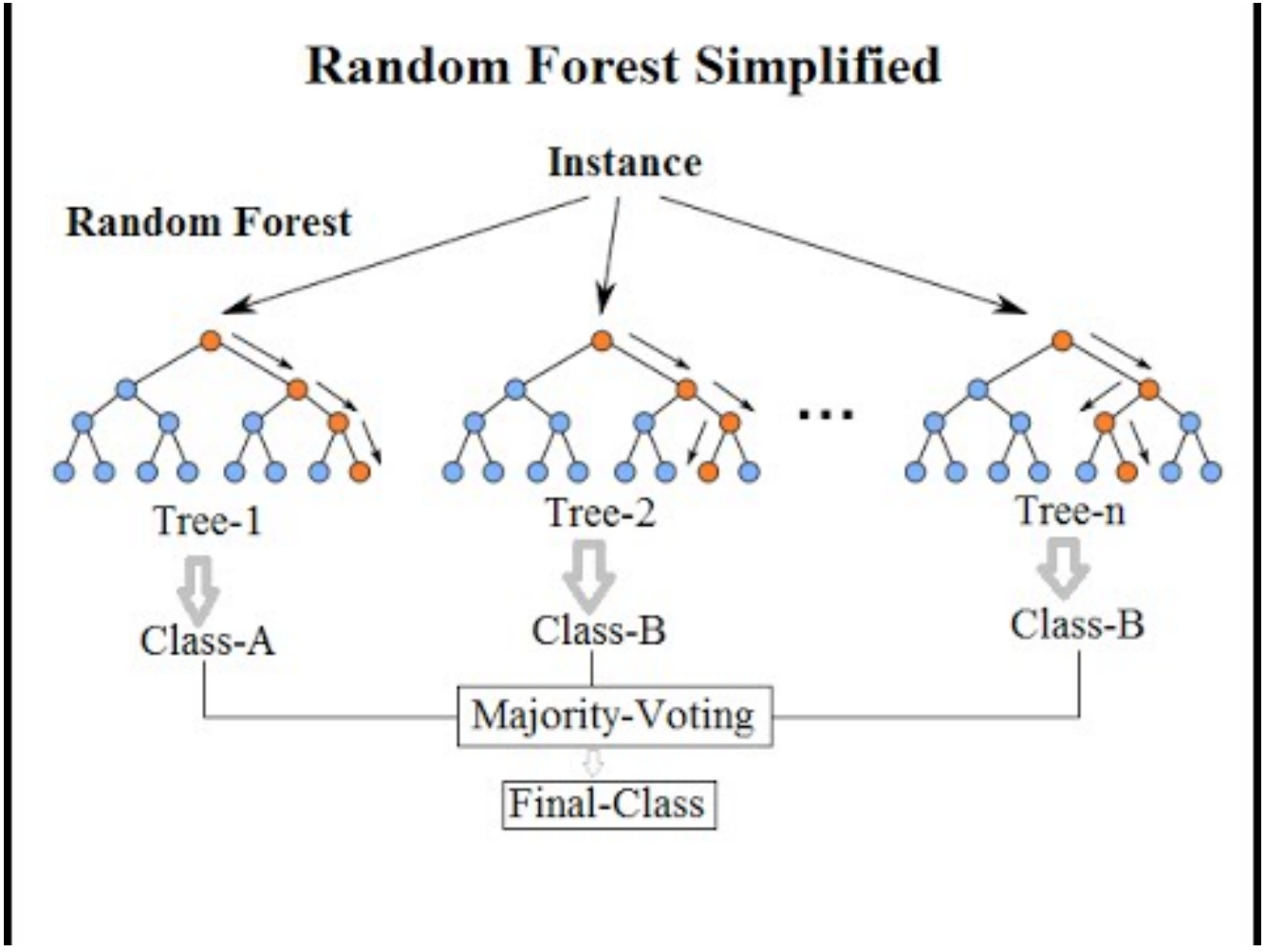
1. number of vertices
2. number of edges
3. mean/max degree
4. density
5. mean/max distance between pairs of vertices
6. number of loops
7. proportion of vertex pairs with distance $\geq$ 2, 3, 4
8. connectedness
9. standard deviation of degrees
10. labelling percentage
11. ratios of features 1–5

Random forests (Breiman 2001)



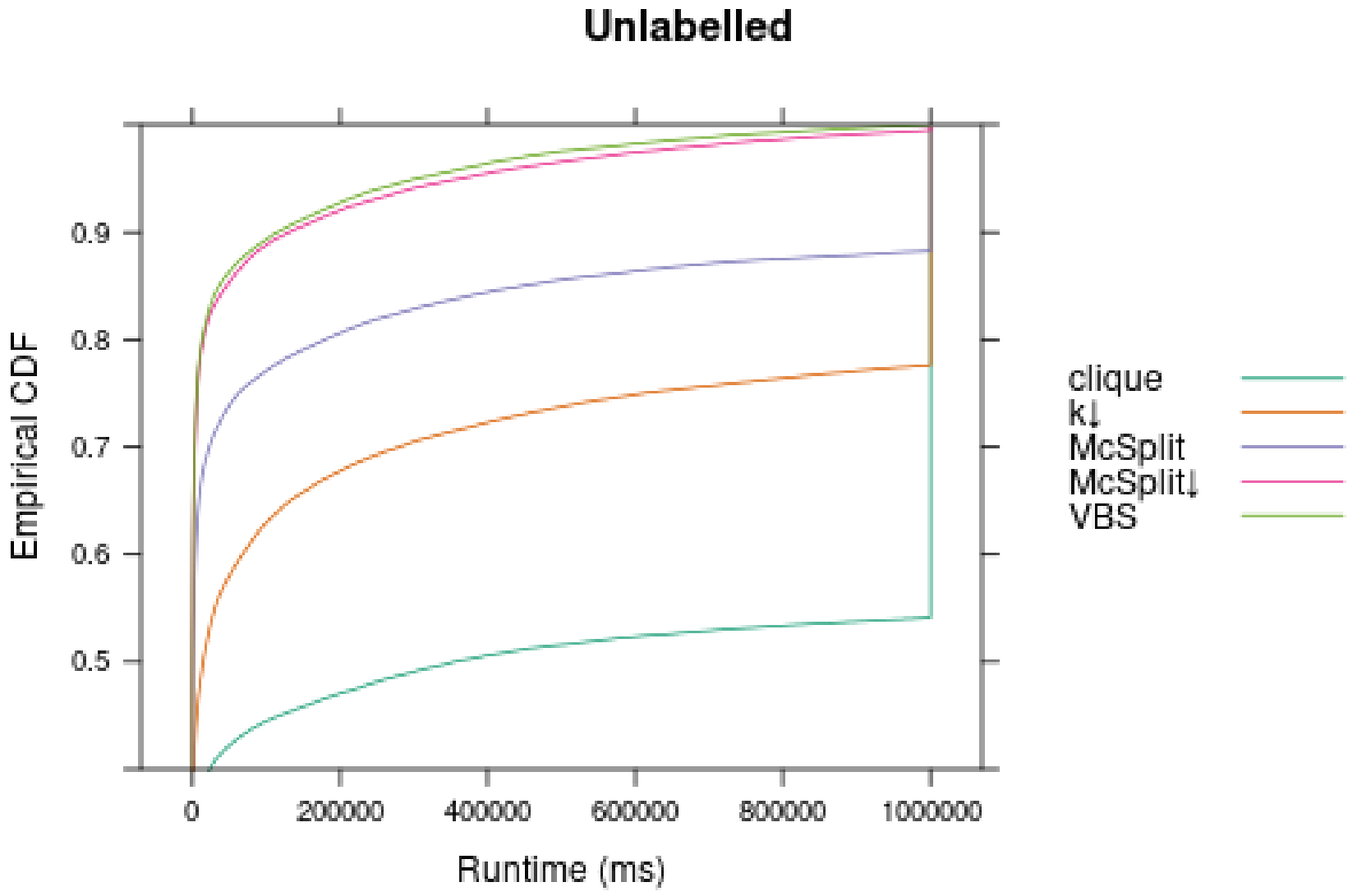
Source: Tae-Kyun Kim & Bjorn Stenger, Intelligent Systems and Networks (ISN) Research Group, Imperial College London

Random forests (Breiman 2001)

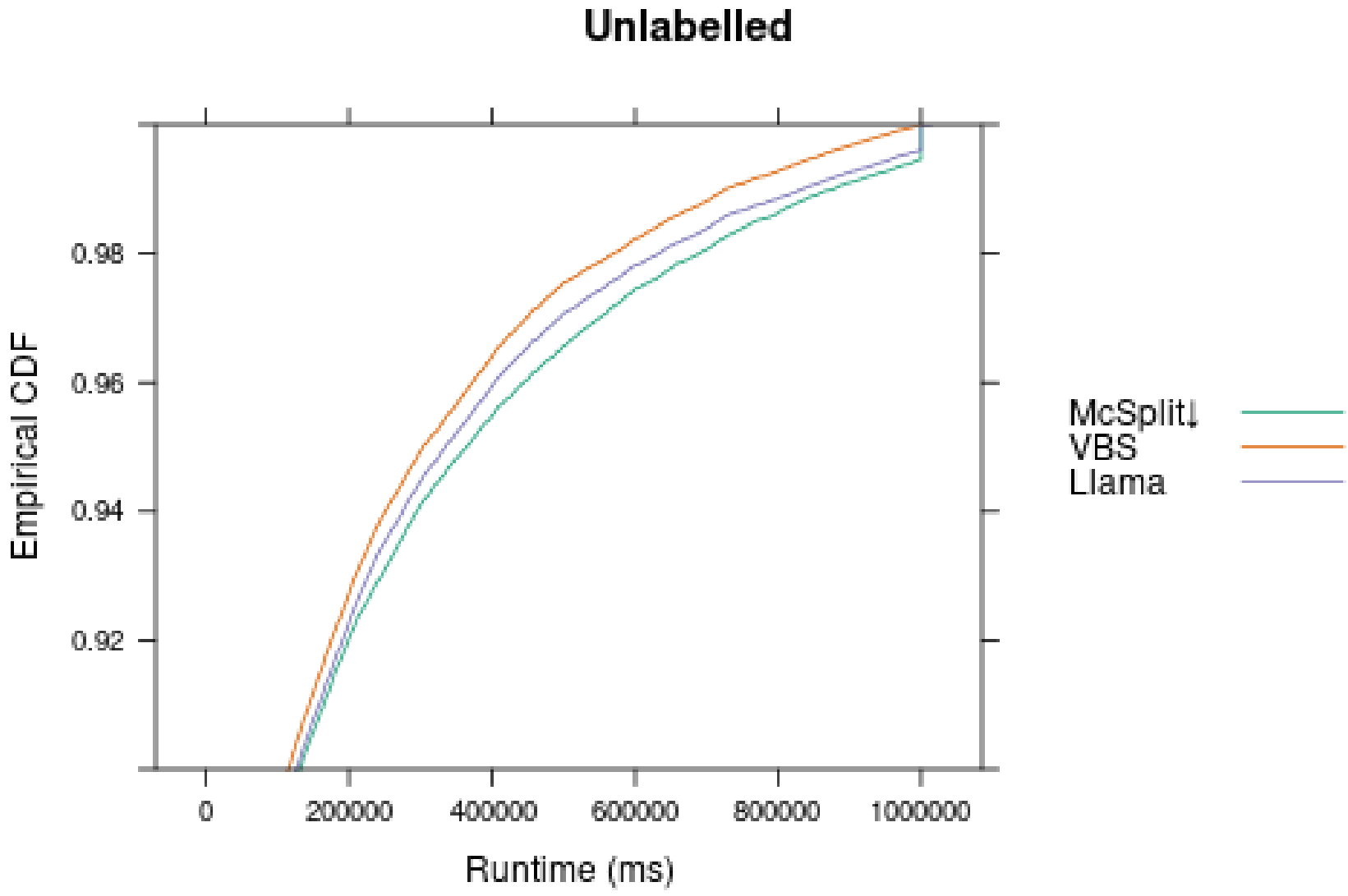


Source: Random Forests(r), Explained, Ilan Reinstein, KDnuggets

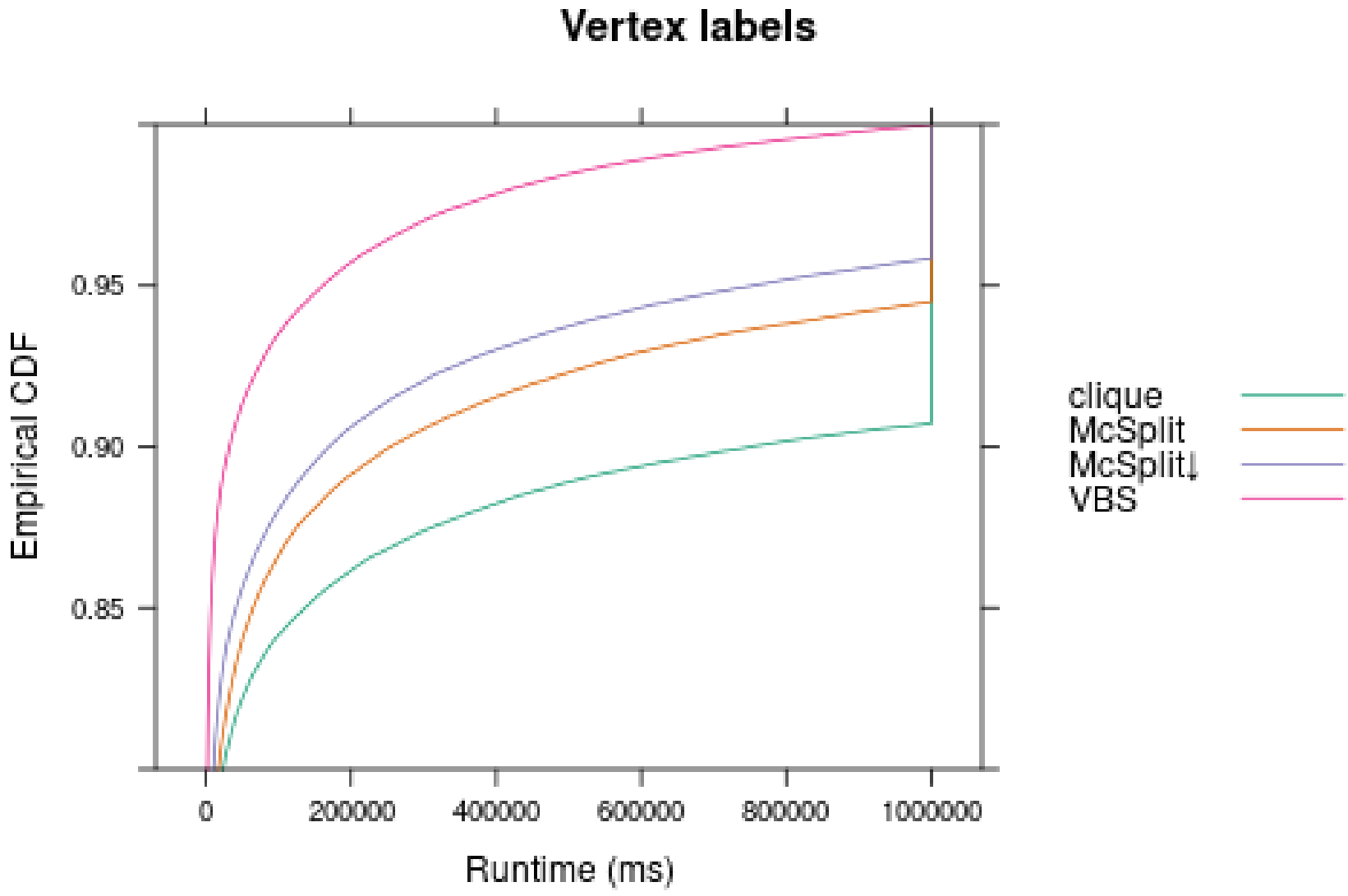
Results


Unlabelled
Empirical CDF
0.9
0.8
0.7
0.6
0.5
0
200000
400000
600000
800000
1000000
Runtime (ms)
clique
k↓
McSplit
McSplit↓
VBS

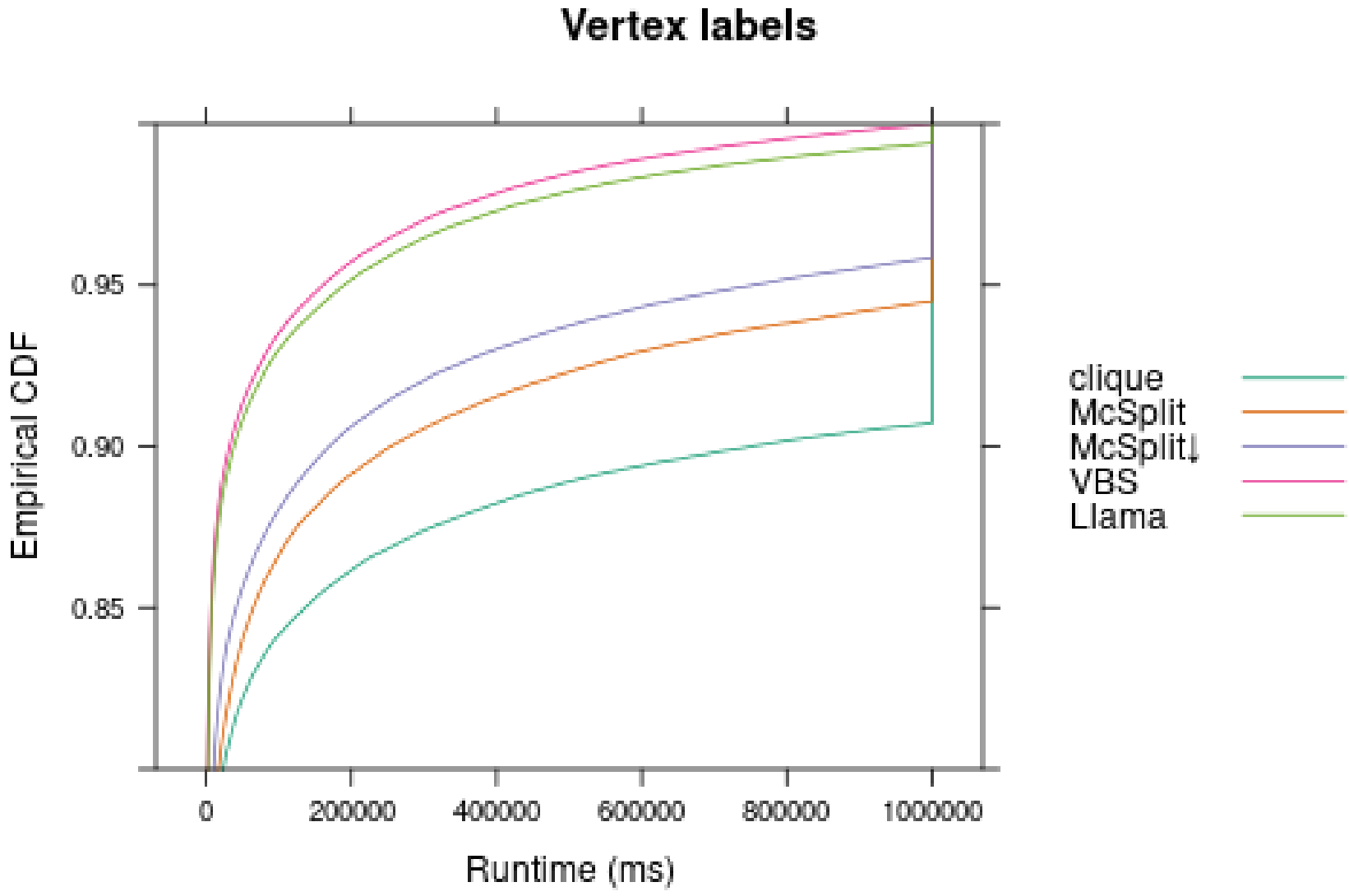
Results (27%)


Unlabelled
Empirical CDF
0.98
0.96
0.94
0.92
0
200000
400000
600000
800000
1000000
Runtime (ms)
McSplit↓
VBS
Llama

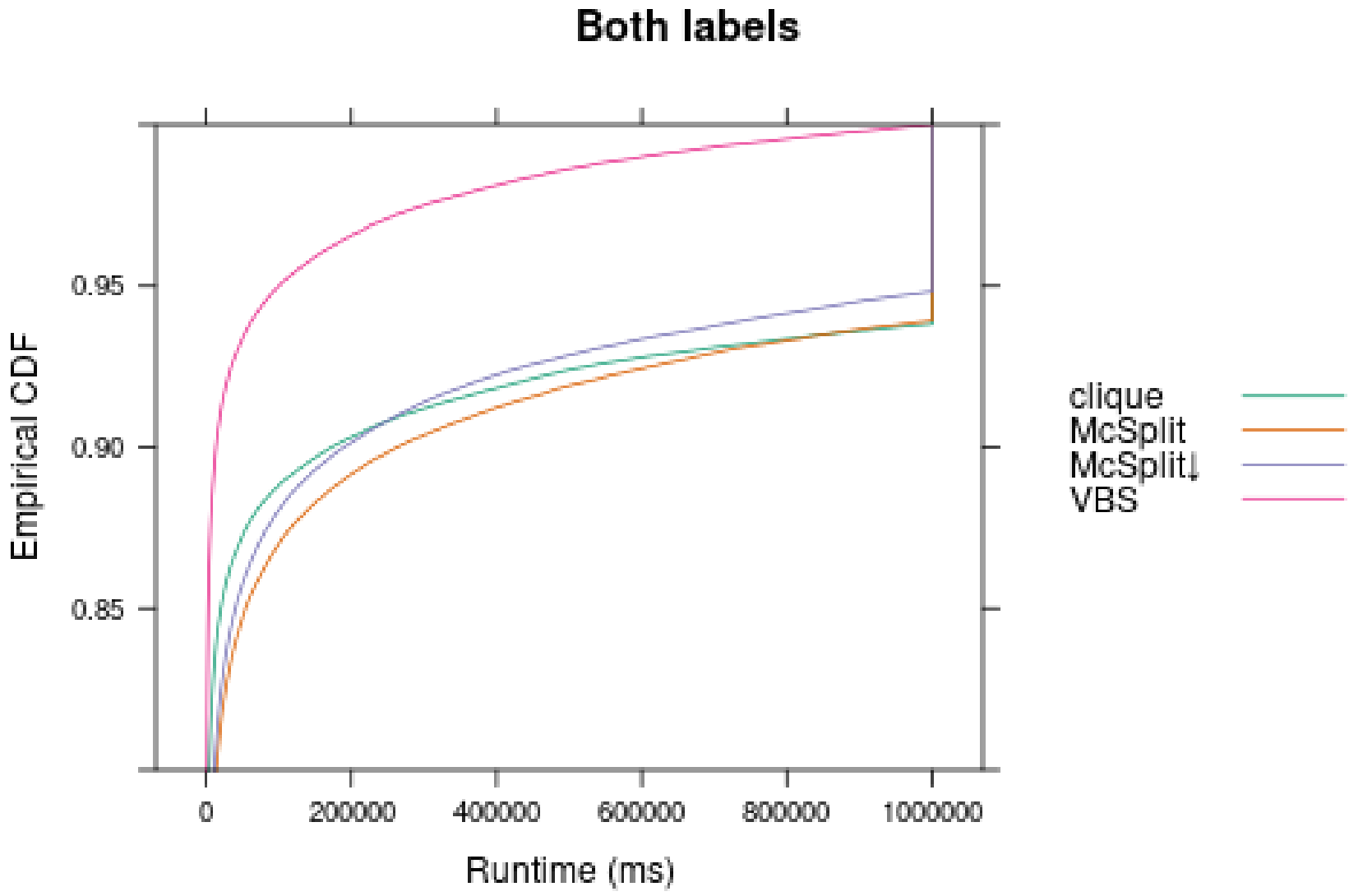
Results


Vertex labels
Empirical CDF
0.95
0.90
0.85
0
200000
400000
600000
800000
1000000
Runtime (ms)
clique
McSplit
McSplit↓
VBS

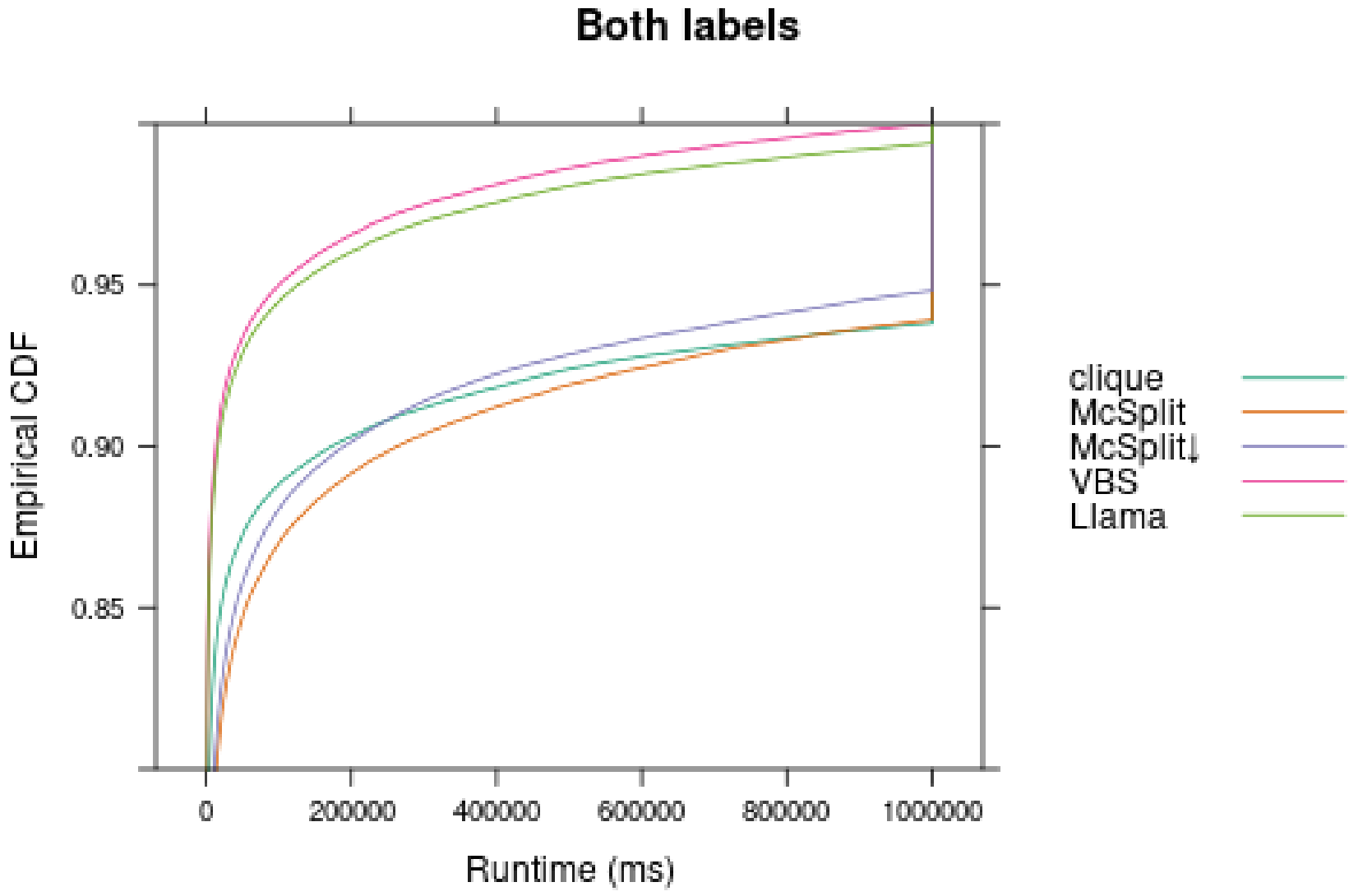
Results (86%)

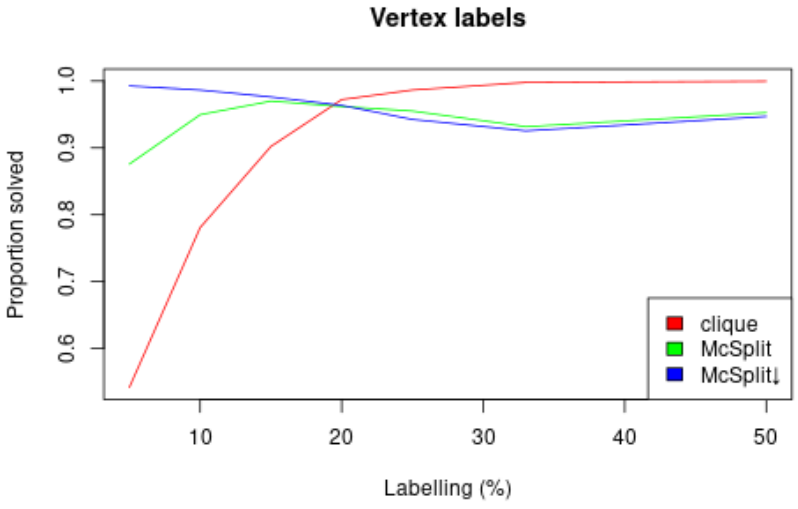

Vertex labels
Empirical CDF
0.95
0.90
0.85
0
200000
400000
600000
800000
1000000
Runtime (ms)
clique
McSplit
McSplit↓
VBS
Llama

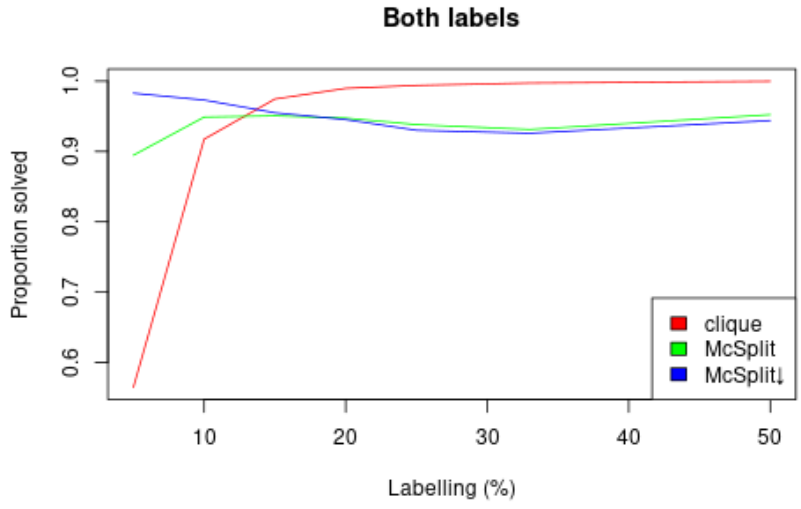
Results

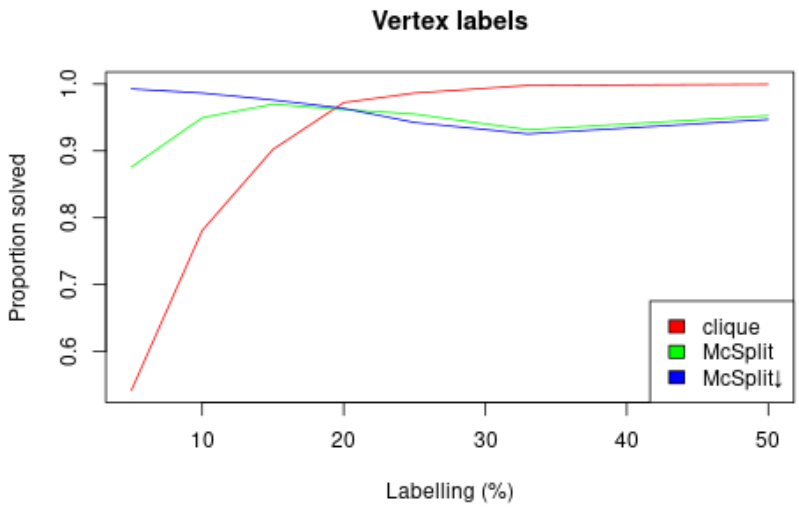

Both labels
Empirical CDF
0.95
0.90
0.85
0
200000
400000
600000
800000
1000000
Runtime (ms)
clique
McSplit
McSplit↓
VBS

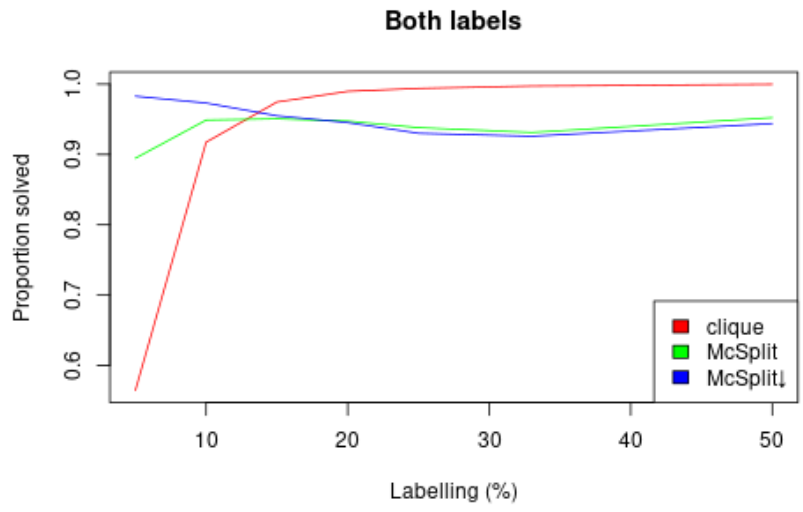
Results (88%)

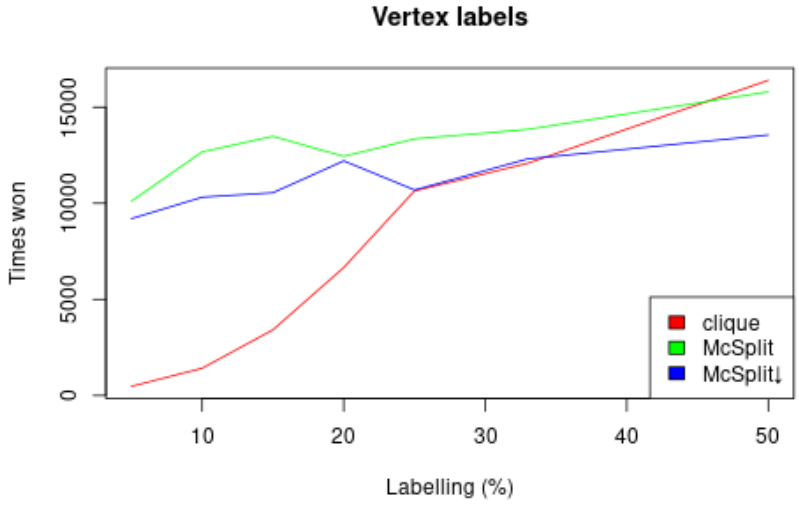

Both labels
Empirical CDF
0.95
0.90
0.85
0
200000
400000
600000
800000
1000000
Runtime (ms)
clique
McSplit
McSplit↓
VBS
Llama

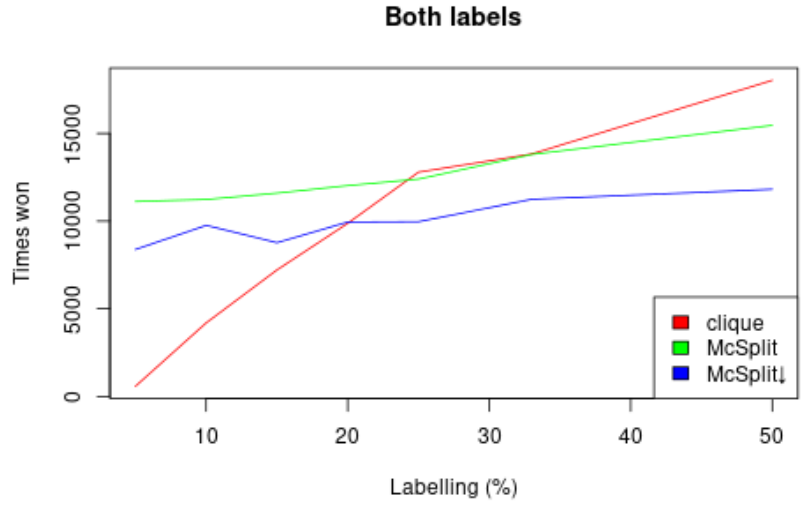

Vertex labels
Proportion solved
1.0
0.9
0.8
0.7
0.6
10
20
30
40
50
Labelling (%)
clique
McSplit
McSplit↓


Both labels
Proportion solved
1.0
0.9
0.8
0.7
0.6
10
20
30
40
50
Labelling (%)
clique
McSplit
McSplit↓


Vertex labels
Proportion solved
0.6
0.7
0.8
0.9
1.0
10
20
30
40
50
Labelling (%)
clique
McSplit
McSplit↓


Both labels
Proportion solved
0.6
0.7
0.8
0.9
1.0
10
20
30
40
50
Labelling (%)
clique
McSplit
McSplit↓


Vertex labels
Times won
0
5000
10000
15000
10
20
30
40
50
Labelling (%)
clique
McSplit
McSplit↓


Both labels
Times won
0
5000
10000
15000
10
20
30
40
50
Labelling (%)
clique
McSplit
McSplit↓

Other Accomplishments

- Identified important features (labelling, standard deviation of degrees),
  - and how they affect best choice of algorithm
- Extended $k\downarrow$ to support vertex labels, using neighbourhood degree sequences
- Defined and developed new algorithms capable of switching between McSplit and the clique encoding