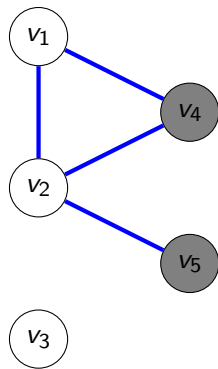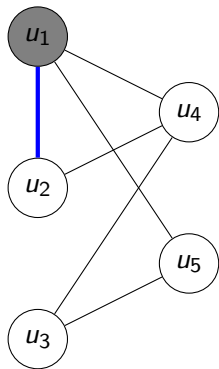# Algorithm Selection for Maximum Common Subgraph

Paulius Dilkas
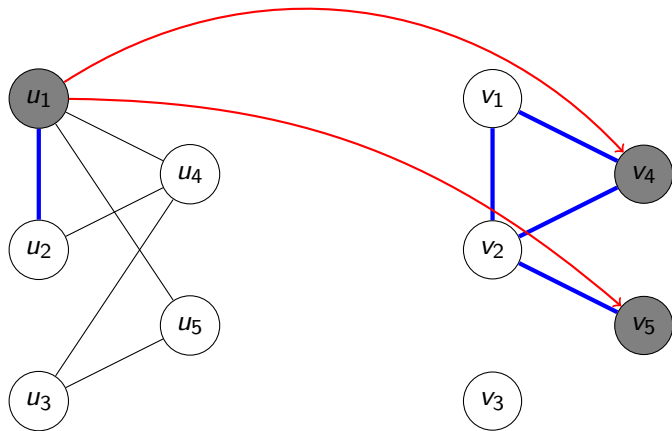
School of Computing Science
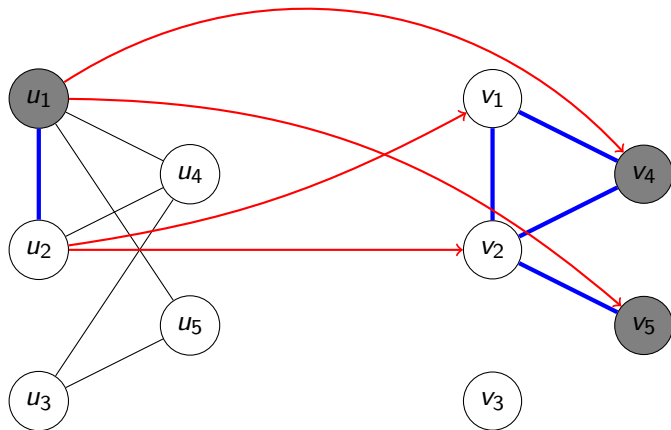University of Glasgow

23rd March 2018

Maximum Common Subgraph
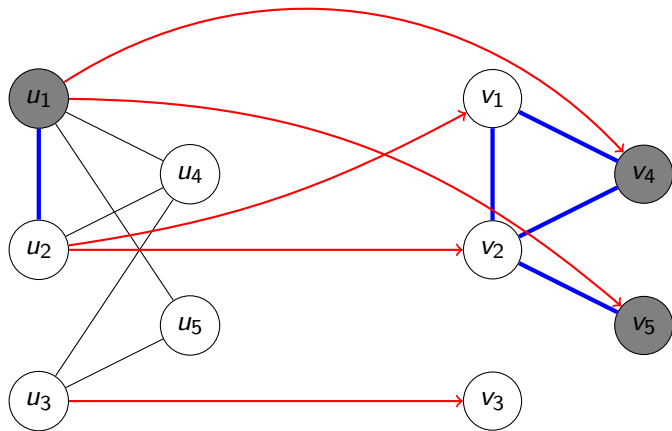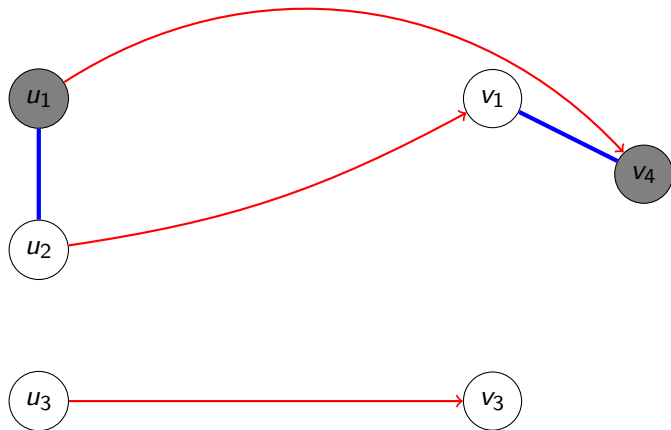
Maximum Common Subgraph

Maximum Common Subgraph

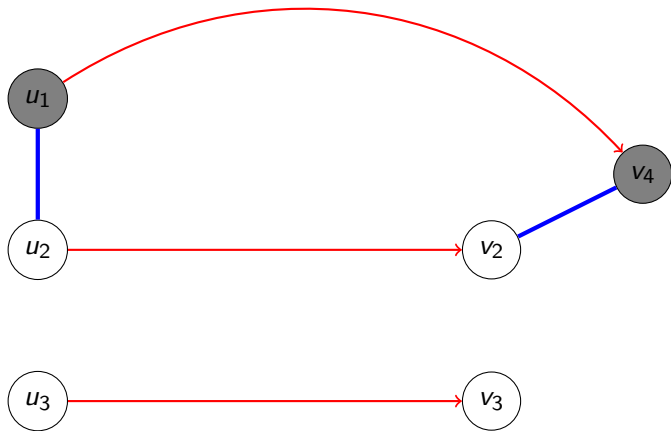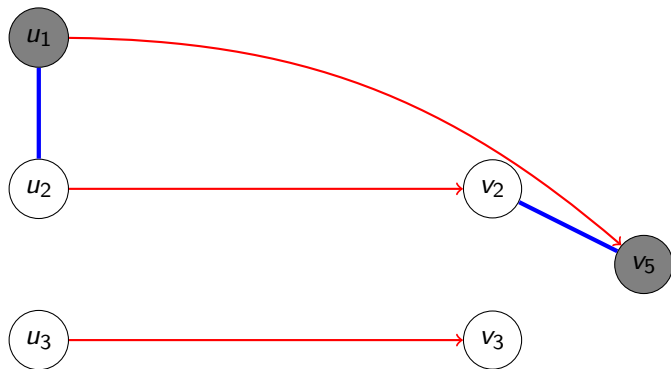# Maximum Common Subgraph

Maximum Common Subgraph

Maximum Common Subgraph

Maximum Common Subgraph

Algorithms

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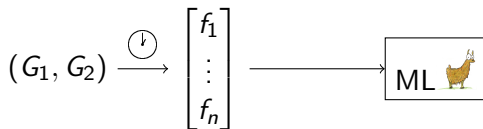- $k\downarrow$
  - Hoffmann, McCreesh and Reilly 2017

Algorithm Selection

(G_1, G_2)

Algorithm Selection

$$(G_1, G_2) \xrightarrow{\enspace\text{\textcircled{$\downarrow$}}\enspace} \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$$

Algorithm Selection

$$(G_1, G_2) \xrightarrow{\;\;\textcircled{\small\text{\textbackslash}}\;\;} \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} \longrightarrow \boxed{\text{ML} \;\text{🦙}}$$

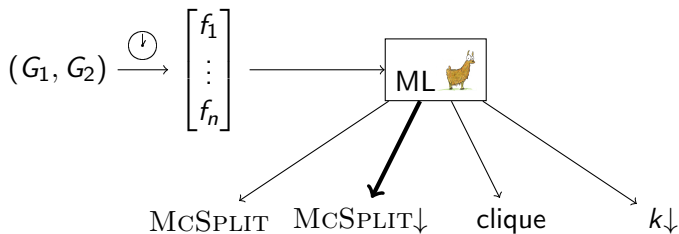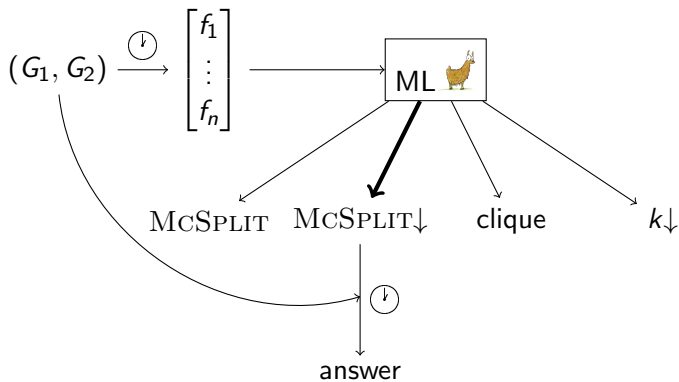# Algorithm Selection
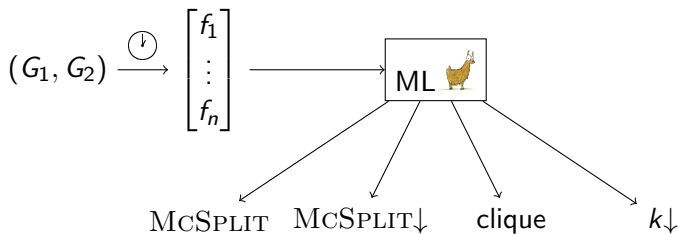
Algorithm Selection

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)

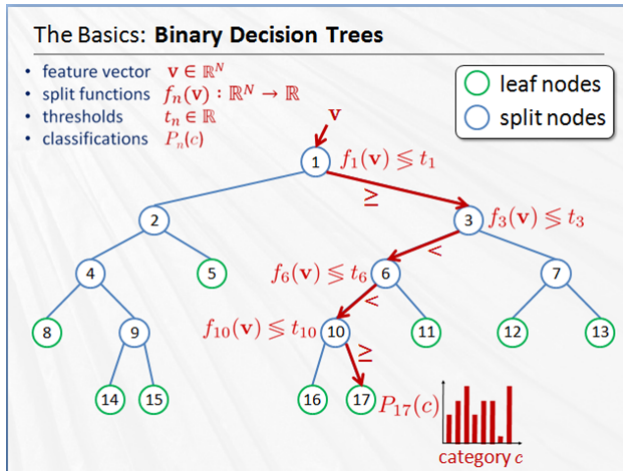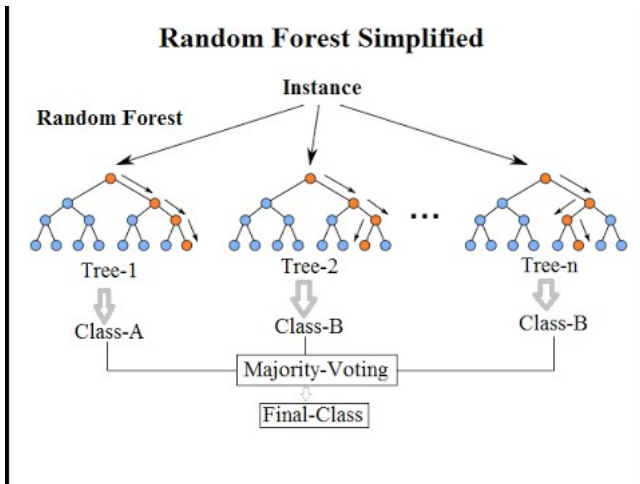1–8 are from Kotthoff, McCreesh and Solnon 2016

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
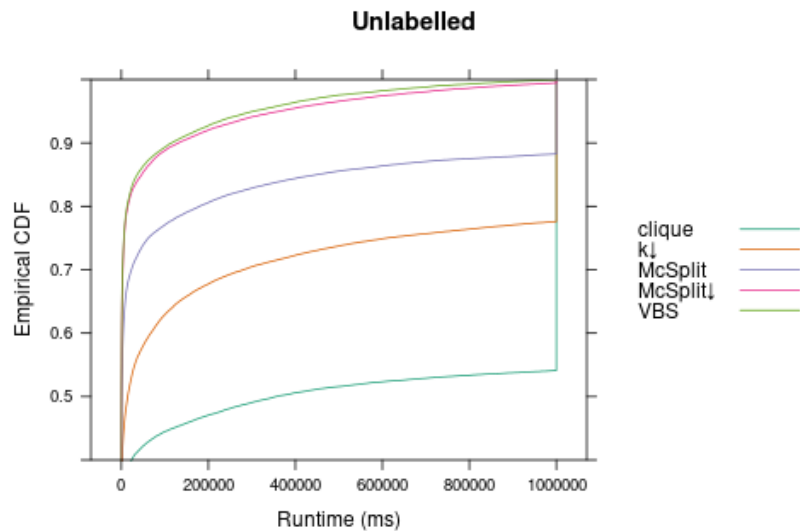11. ratios of features 1–5

Random forests (Breiman 2001)

Random forests (Breiman 2001)

Results

Unlabelled

Empirical CDF vs Runtime (ms)

Legend:
- clique
- k↓
- McSplit
- McSplit↓
- VBS

Results (27%)

Results

Vertex labels

Empirical CDF vs Runtime (ms)

clique
McSplit
McSplit↓
VBS

Results (86%)

Vertex labels

Results

Both labels

Empirical CDF vs Runtime (ms)
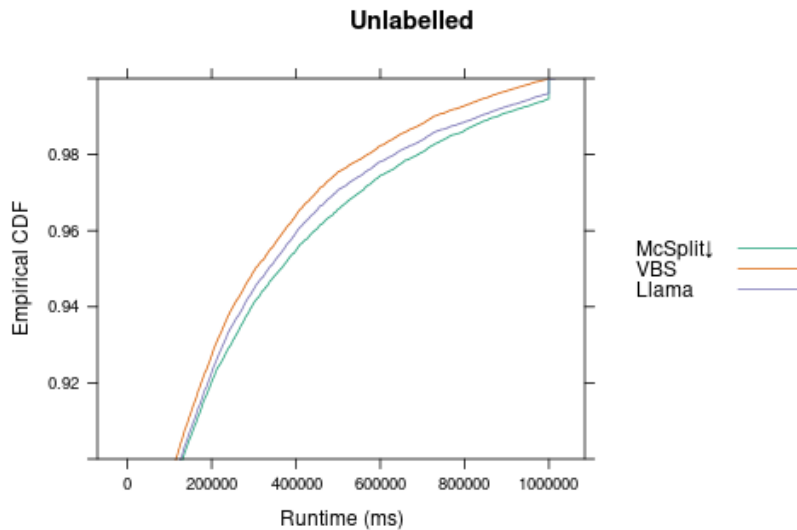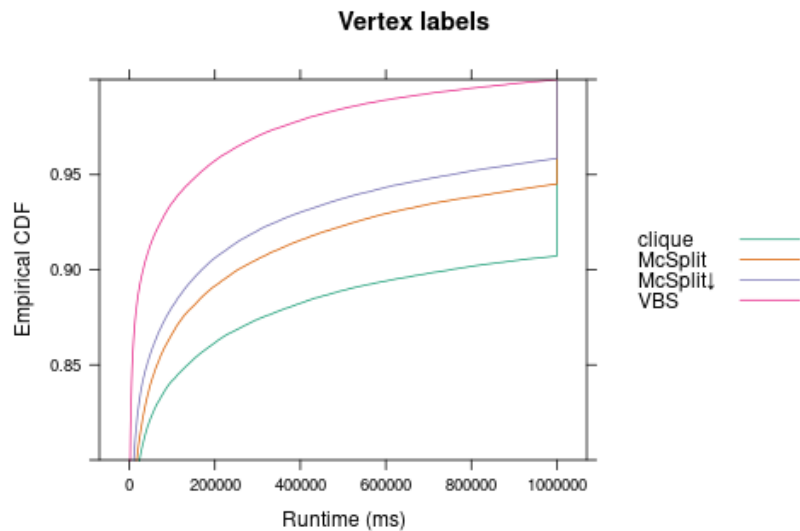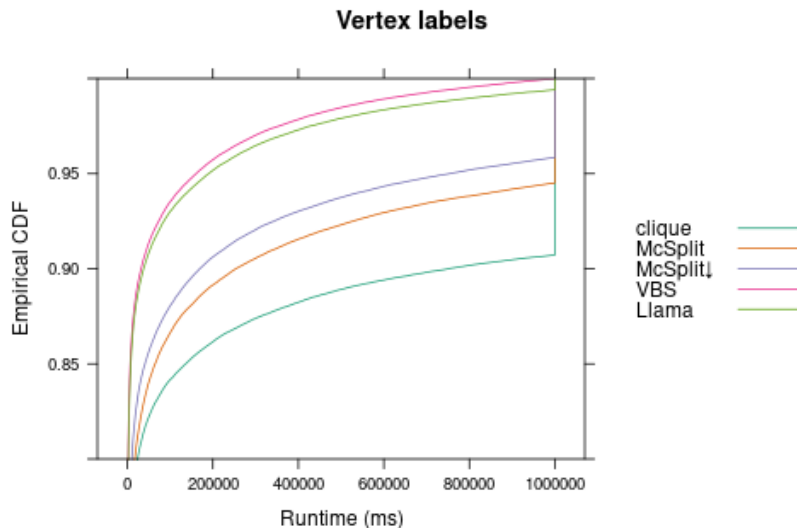
clique
McSplit
McSplit↓
VBS
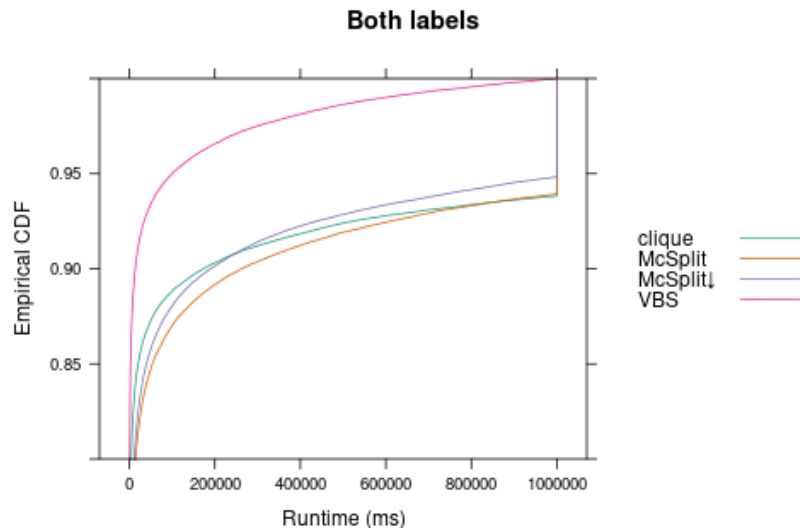
Results (88%)

Vertex labels

Proportion solved

Labelling (%)

- clique
- McSplit
- McSplit↓

Both labels

Proportion solved

Labelling (%)

- clique
- McSplit
- McSplit↓

Vertex labels

Proportion solved vs Labelling (%)

clique — red
McSplit — green
McSplit↓ — blue

Both labels

Proportion solved vs Labelling (%)

clique — red
McSplit — green
McSplit↓ — blue

Vertex labels

Times won vs Labelling (%)

clique — red
McSplit — green
McSplit↓ — blue

Both labels

Times won vs Labelling (%)

clique — red
McSplit — green
McSplit↓ — blue
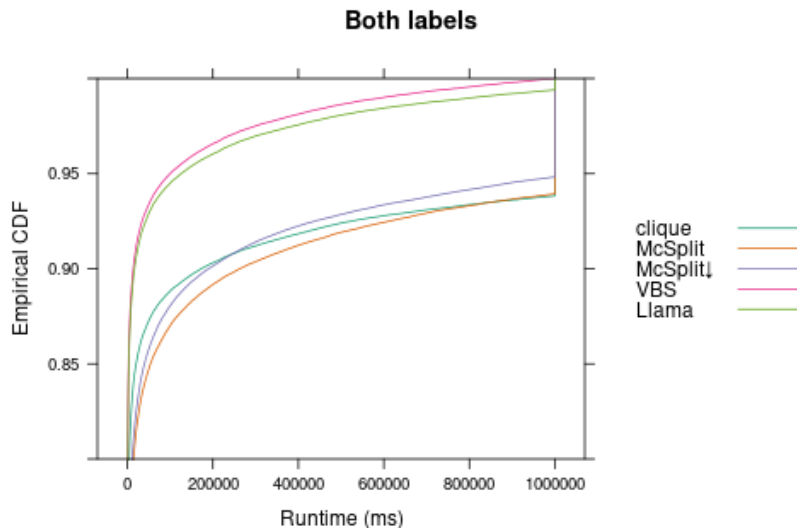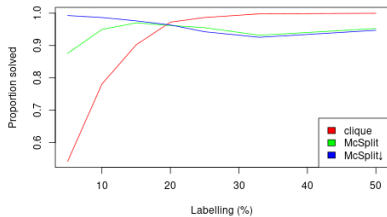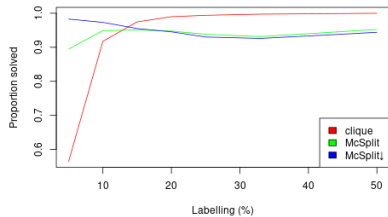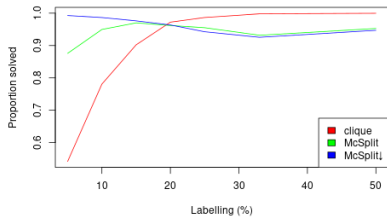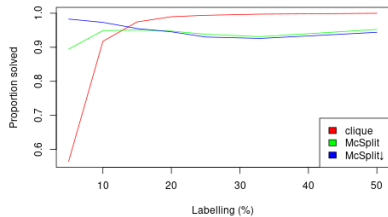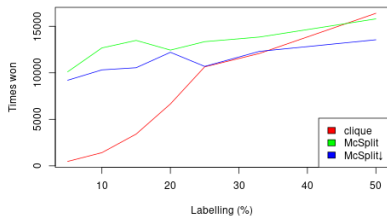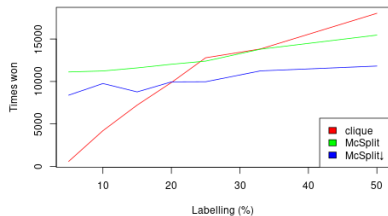
Other Accomplishments

- Identified important features:
  - labelling, standard deviation of degrees
- 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