UNIVERSITÄT
LEIPZIG

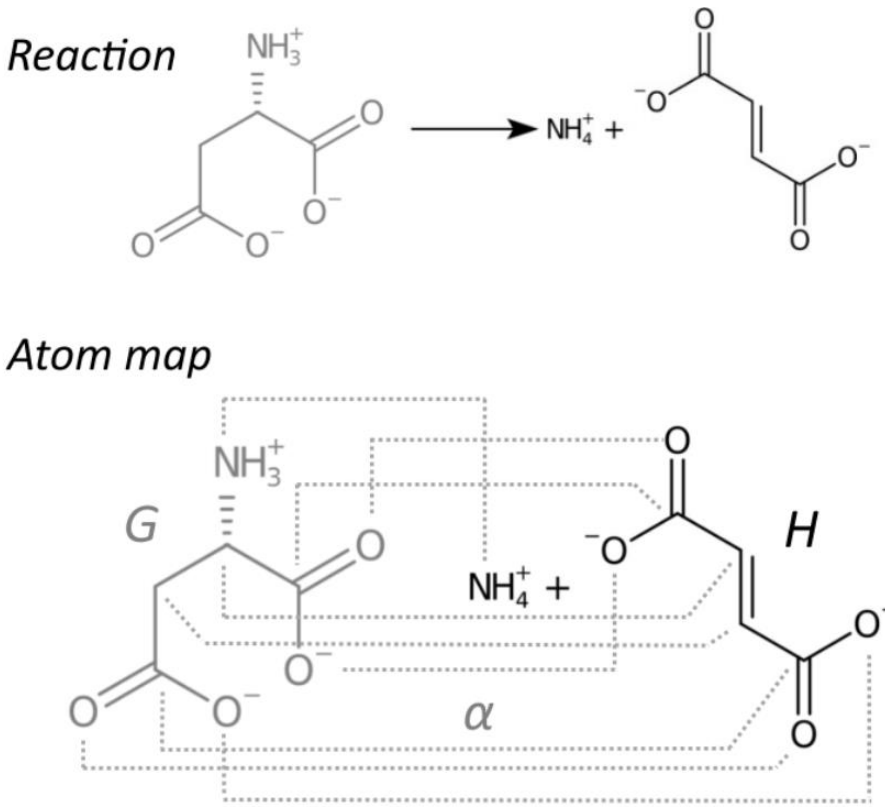
MATOMIC Mathematical Modelling for Microbial Community Induced Metabolic Diseases

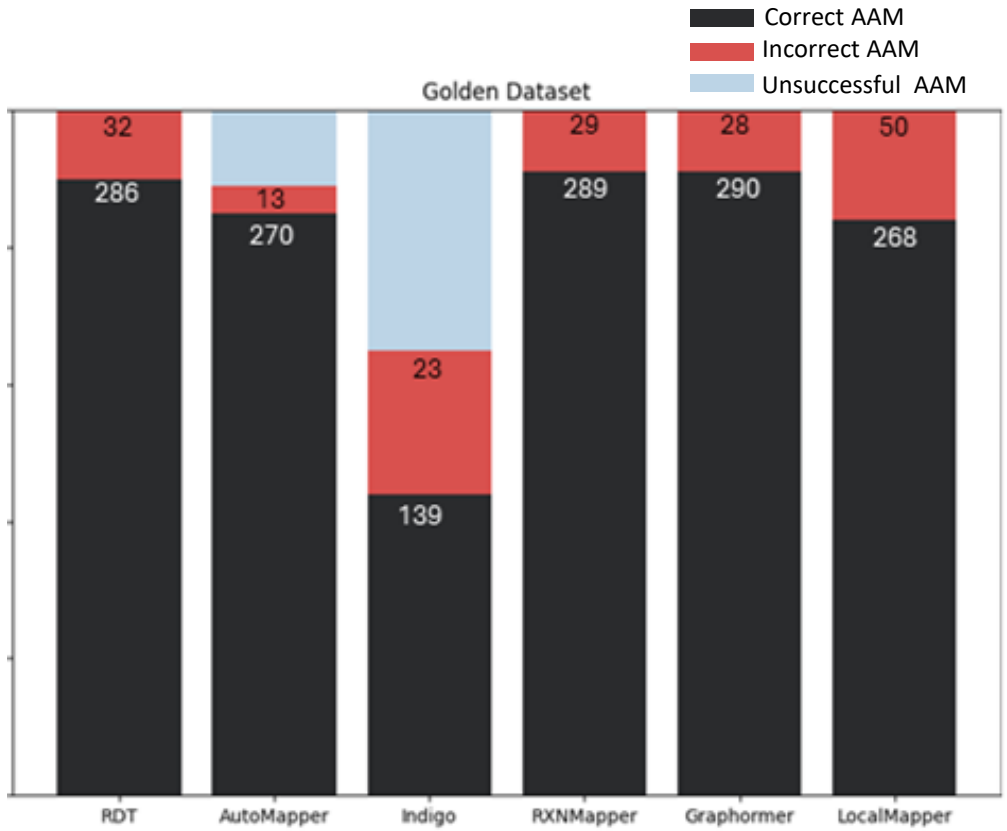
Not Perfect, just Right – Suboptimal Solutions as a Secret Ingredient for Better Atom-to-Atom Mappers

Nora Beier, 10.02.2025

UNIVERSITÄT LEIPZIG

Atom-to-Atom Maps (AAMs)



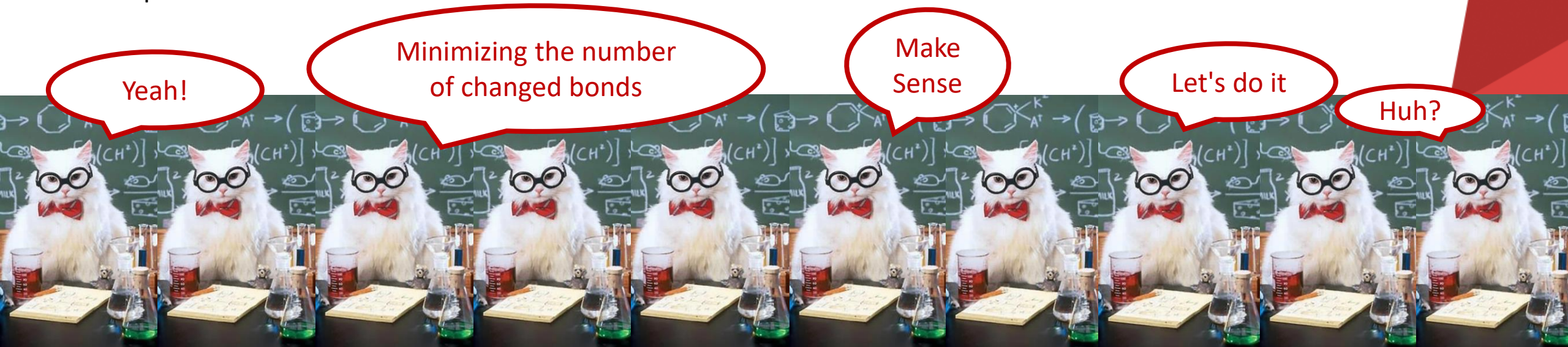


AAMs are a bijection between the atoms in the reactant molecules and the atoms in the product molecules. Their purpose is to identify the position of each atom in the molecules before and after a chemical reaction.

UNIVERSITÄT LEIPZIG

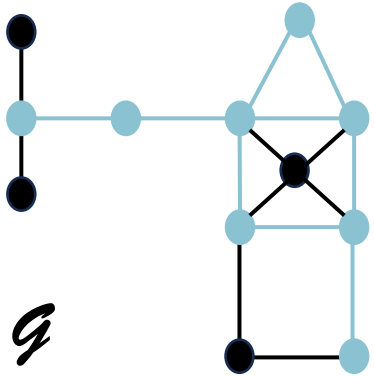
Machine Learning based on Atom-to-Atom Mapper

- Machine learning atom-to-atom mappers are always mapping on completely unknown reactions.
- The accuracy rates are below 90%
- Can we improve the results by giving the AI a pre-selection?
- 9 out of 10 chemists would recommend minimal bond change as a good marker for correct atom-to-atom maps.

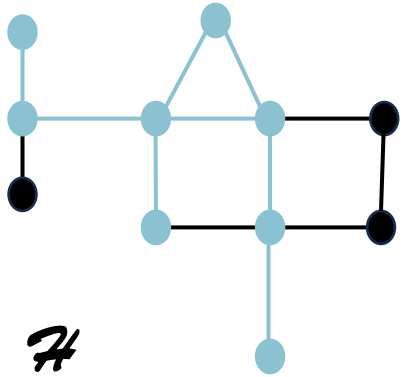


UNIVERSITÄT LEIPZIG

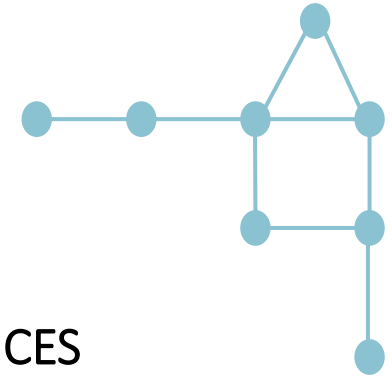
Maximum Common Edge Subgraph (MCES)



$\mathcal{G}$



$\mathcal{H}$



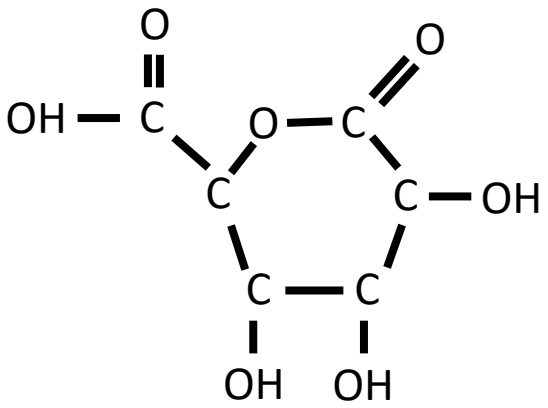
MCES

MCES is a subgraph with the maximum number of edges common to graph $\mathcal{G}$ and $\mathcal{H}$

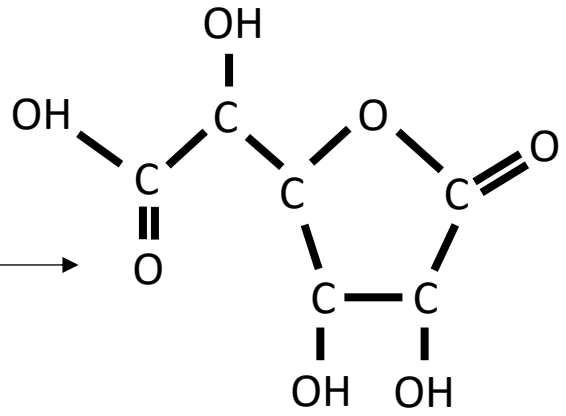
UNIVERSITÄT LEIPZIG

MCES does not always have to correspond to the reaction mechanism

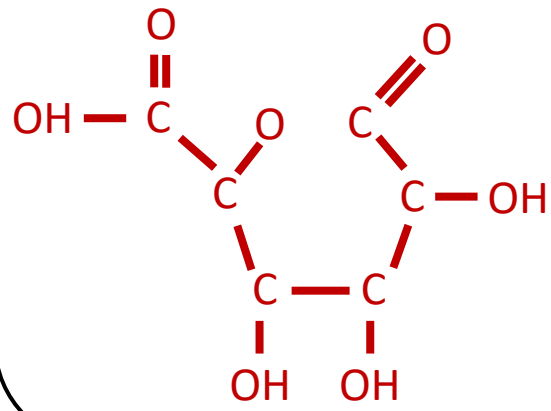
D-Galactaro-1,5-lactone

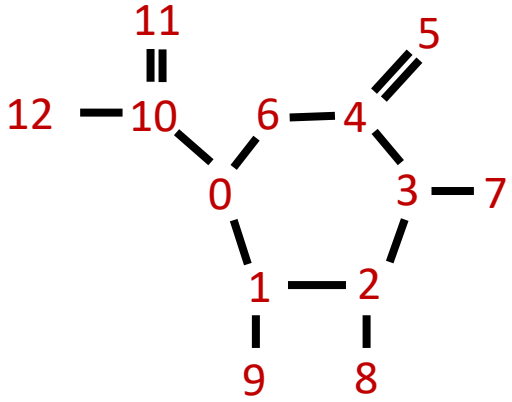


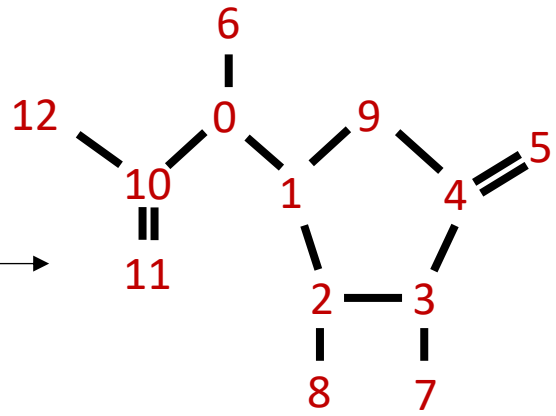
D-Galactaro-1,4-lactone



MCES

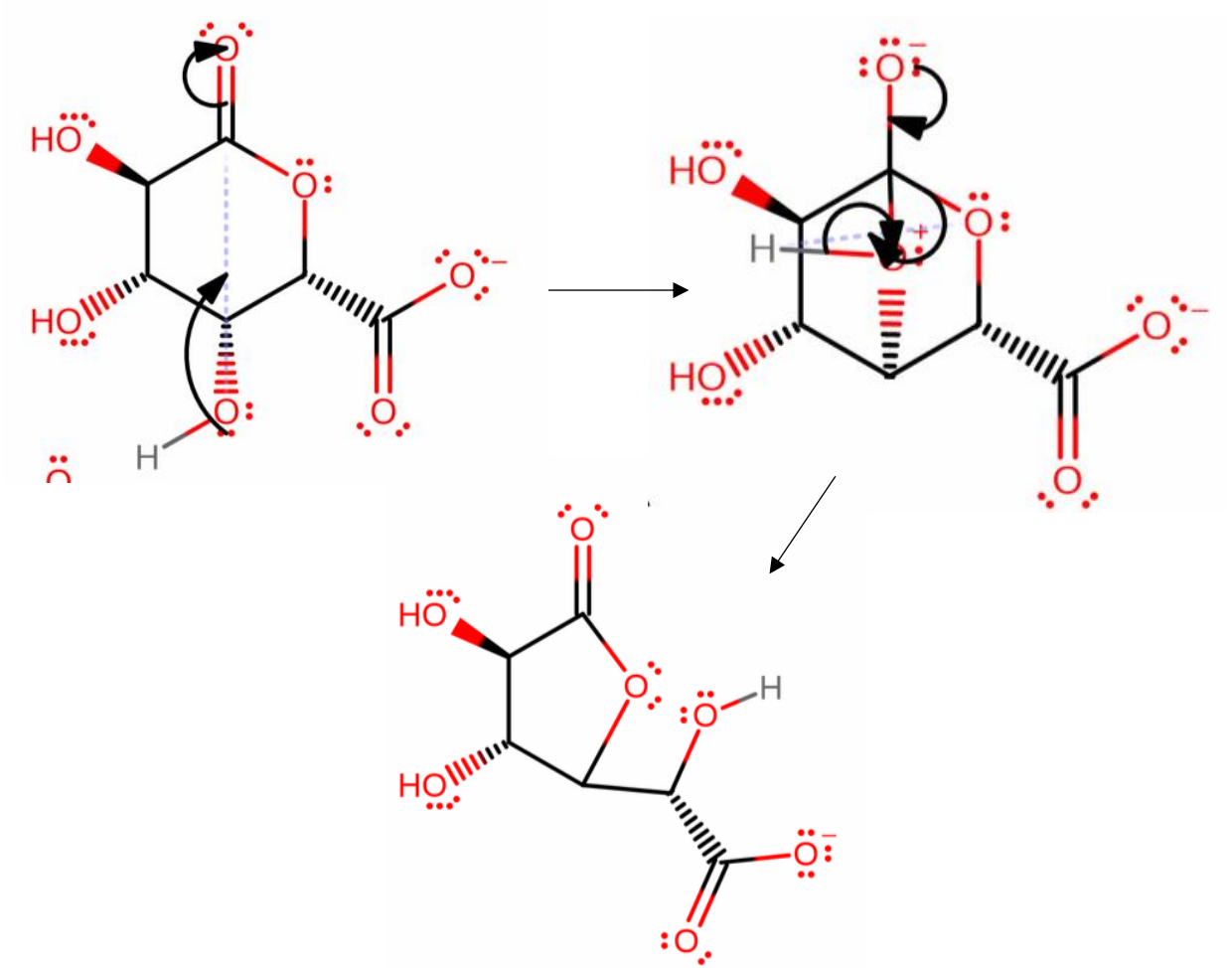




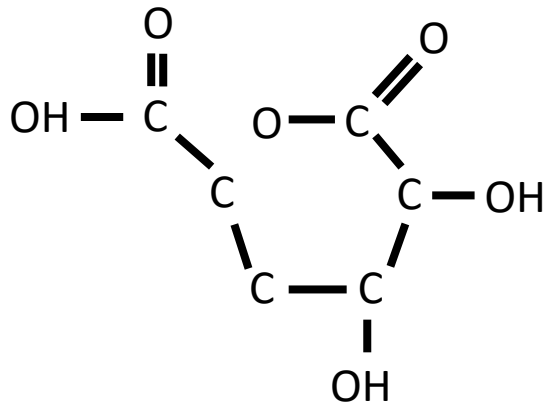


UNIVERSITÄT LEIPZIG

MCES does not always have to correspond to the reaction mechanism

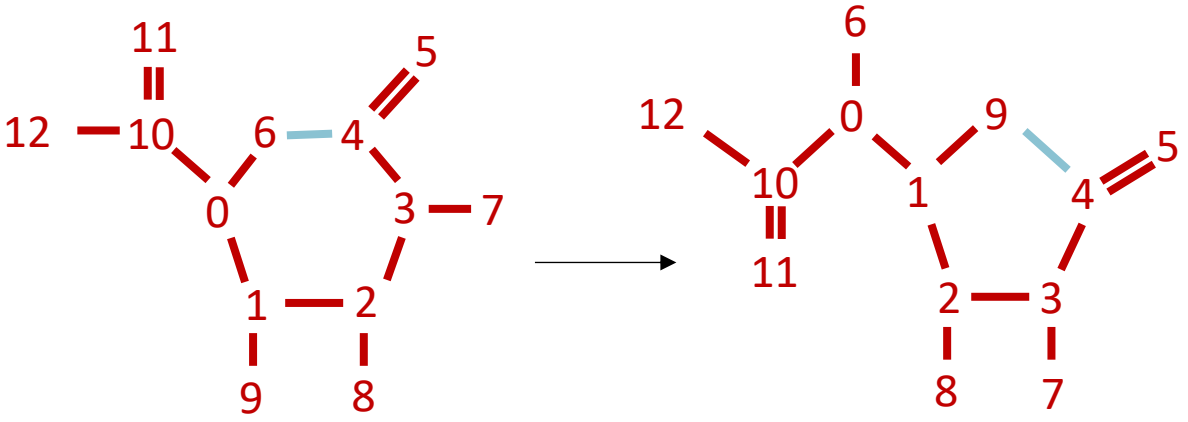


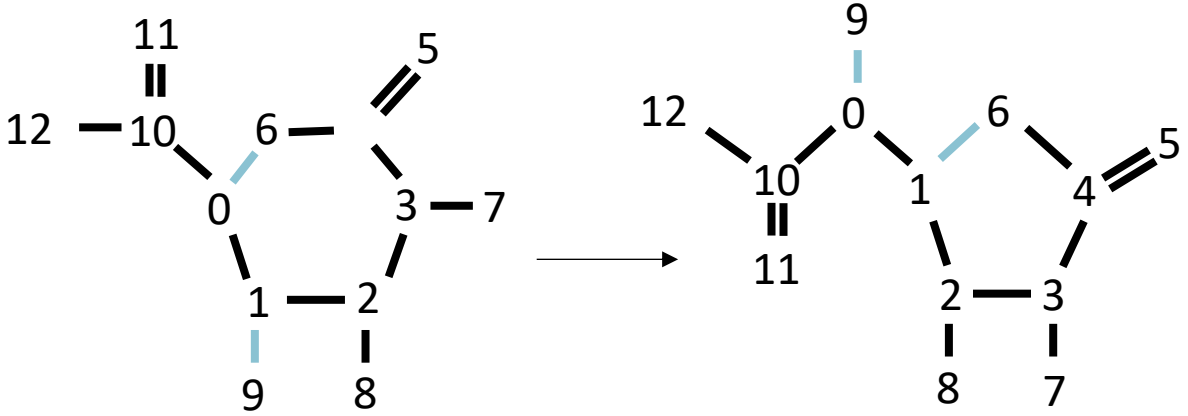
Suboptimal MCES

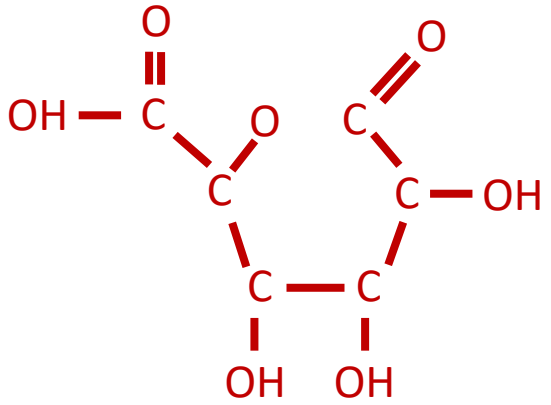


UNIVERSITÄT LEIPZIG

MCES does not always have to correspond to the reaction mechanism

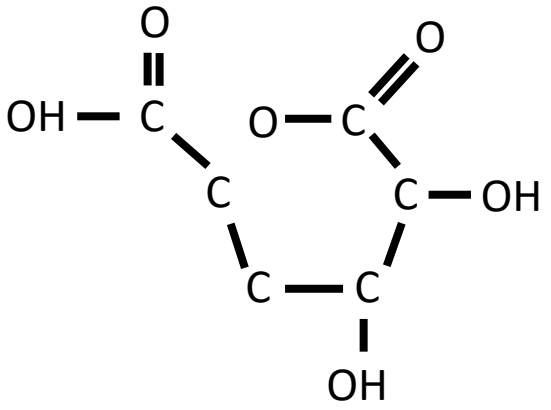






MCES

Edges: 12



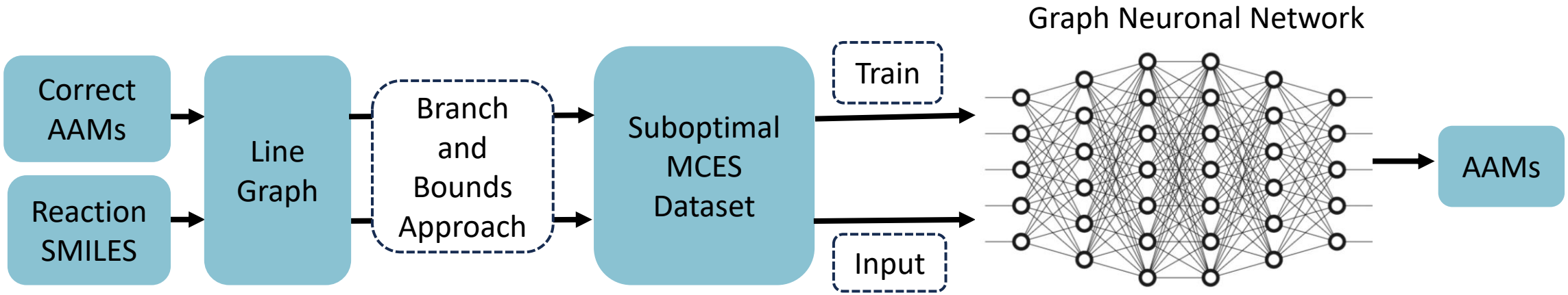
Suboptimal MCES

Edges: 11

UNIVERSITÄT LEIPZIG

Overview of the new Mapper

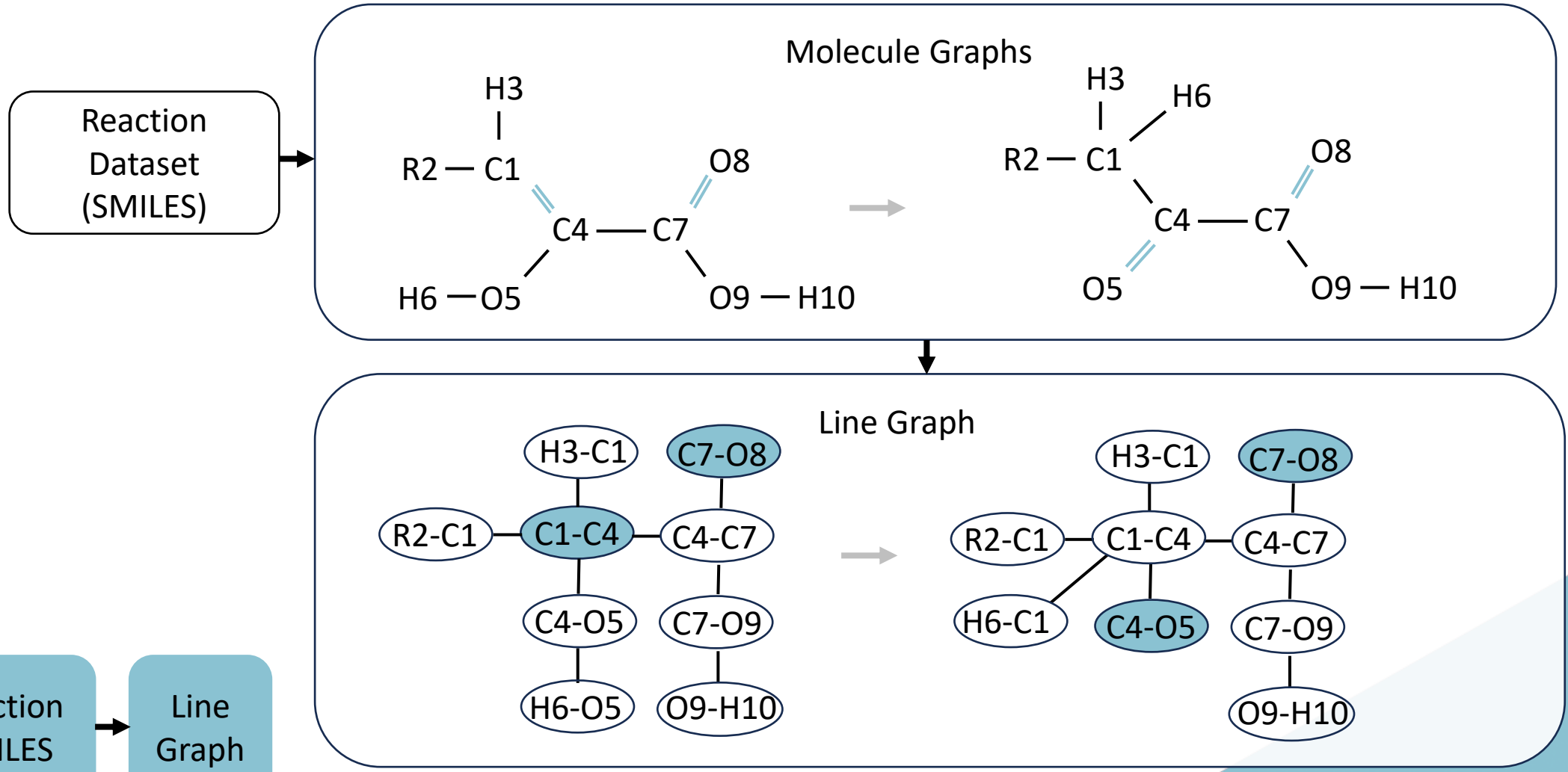
(I don't have a cool name yet)



We Are Here! ☺

UNIVERSITÄT LEIPZIG

Compiling of Suboptimal Solutions



UNIVERSITÄT LEIPZIG

Branch and Bound Approach: The McSplit Algorithm*

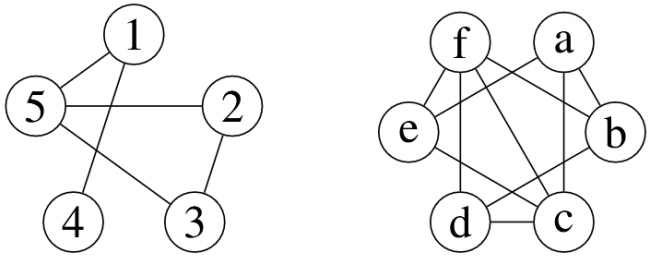


Figure 3.1: Example graphs G and H.

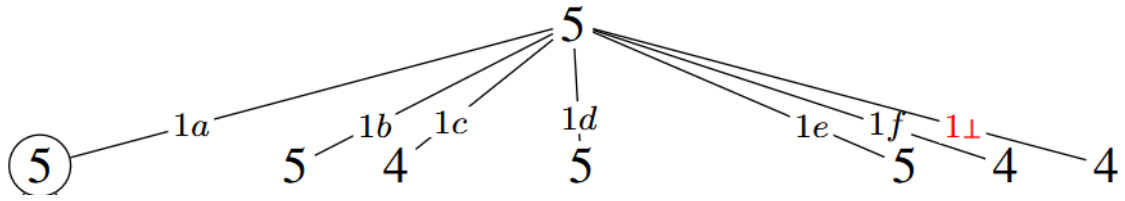


Figure 3.3: The search tree of McSplit on example graphs G and H

Reaction SMILES → Line Graph → Suboptimal MCES Dataset

*Adopted from James Trimble's PhD Thesis, 2023, Glasgow University.

UNIVERSITÄT LEIPZIG

Branch and Bound Approach: The McSplit Algorithm*

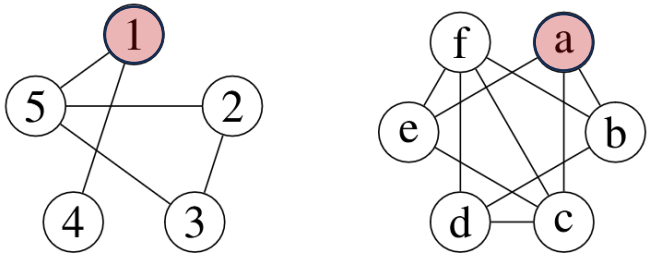


Figure 3.1: Example graphs G and H.

Labelling of G

Vertex	Label
2	0
3	0
4	1
5	1

Labelling of H

Vertex	Label
b	1
c	1
d	0
e	1
f	0

(a) After mapping 1 to a

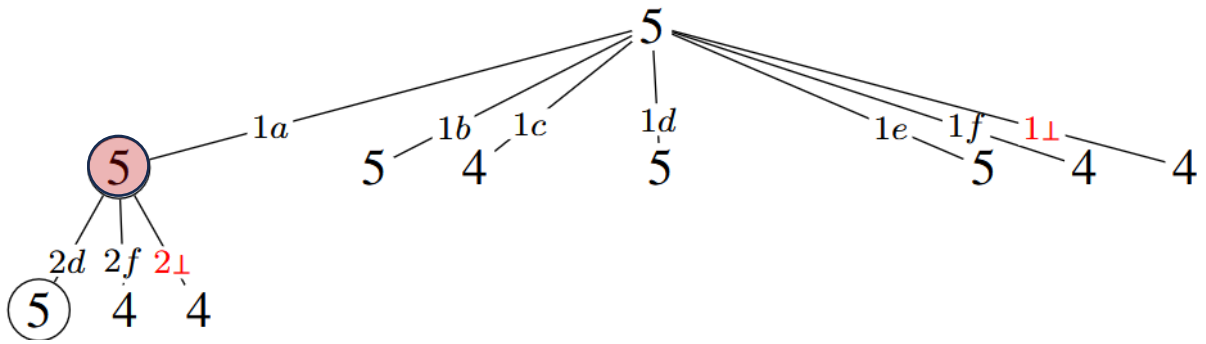


Figure 3.3: The search tree of McSplit on example graphs G and H

Reaction SMILES → Line Graph → Suboptimal MCES Dataset

*Adopted from James Trimble's PhD Thesis, 2023, Glasgow University.

UNIVERSITÄT LEIPZIG

Branch and Bound Approach: The McSplit Algorithm*

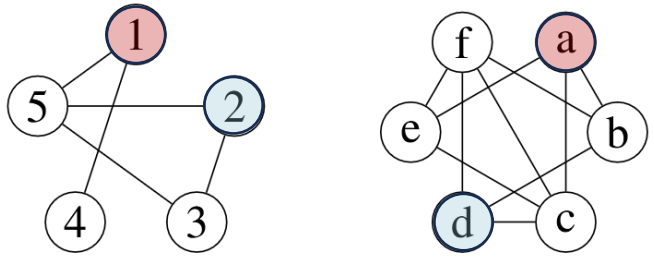


Figure 3.1: Example graphs G and H.

Labelling of G

Vertex	Label
3	01
4	10
5	11

Labelling of H

Vertex	Label
b	11
c	11
e	10
f	01

(b) After mapping 2 to d

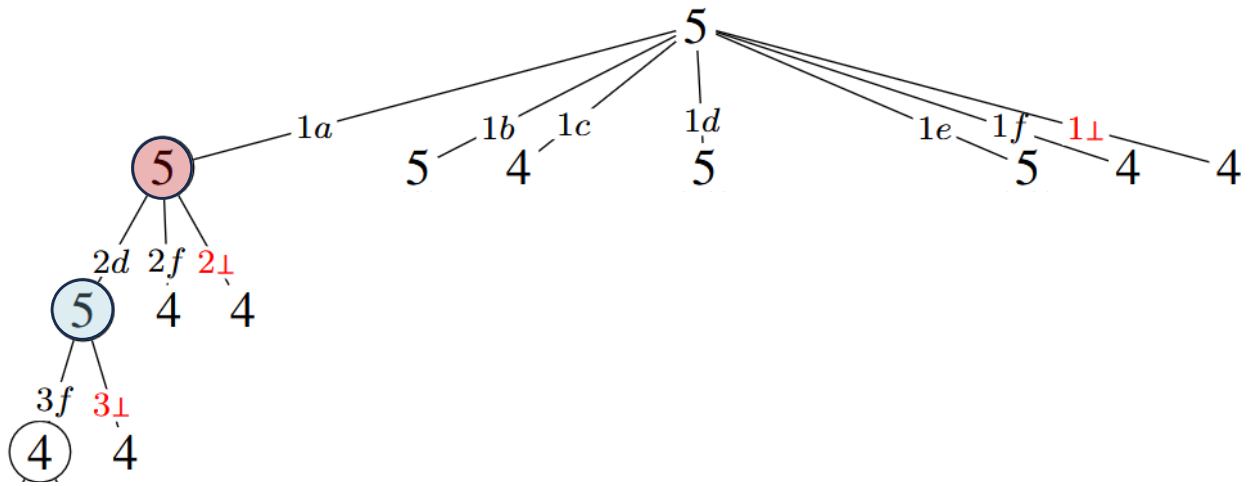


Figure 3.3: The search tree of McSplit on example graphs G and H

Reaction SMILES → Line Graph → Suboptimal MCES Dataset

*Adopted from James Trimble's PhD Thesis, 2023, Glasgow University.

UNIVERSITÄT LEIPZIG

Branch and Bound Approach: The McSplit Algorithm*

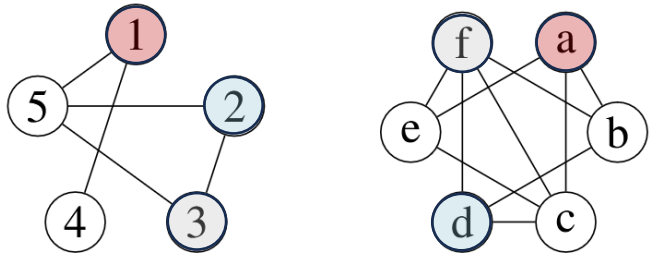


Figure 3.1: Example graphs G and H.

Labelling of G

Vertex	Label
4	100
5	111

Labelling of H

Vertex	Label
b	111
c	111
e	101

(c) After mapping 3 to f

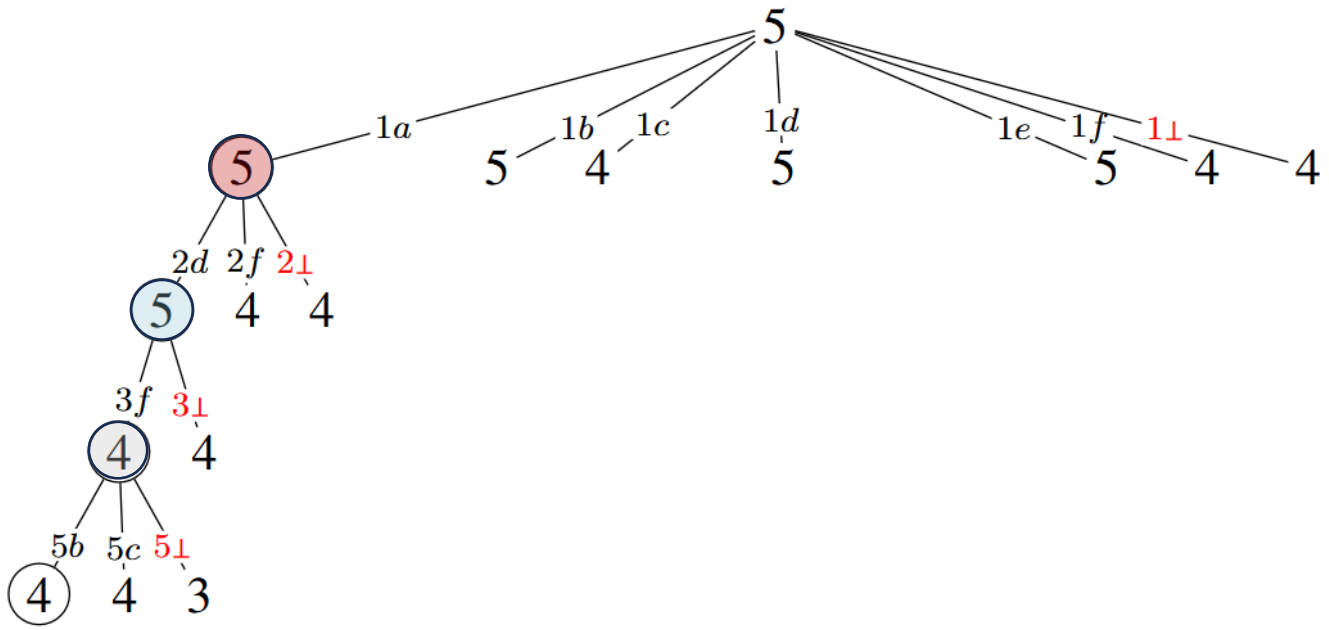


Figure 3.3: The search tree of McSplit on example graphs G and H

Reaction SMILES → Line Graph → Suboptimal MCES Dataset

*Adopted from James Trimble's PhD Thesis, 2023, Glasgow University.

UNIVERSITÄT LEIPZIG

Branch and Bound Approach: The McSplit Algorithm*

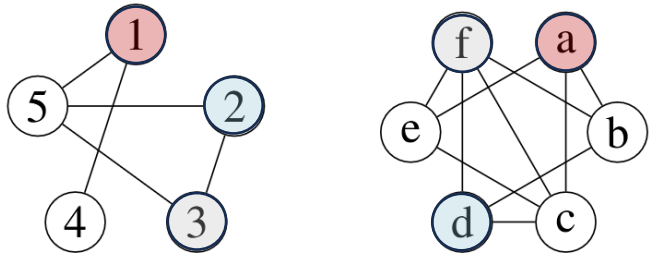


Figure 3.1: Example graphs G and H.

Labelling of G

Vertex	Label
4	100
5	111

Labelling of H

Vertex	Label
b	111
c	111
e	101

(c) After mapping 3 to f

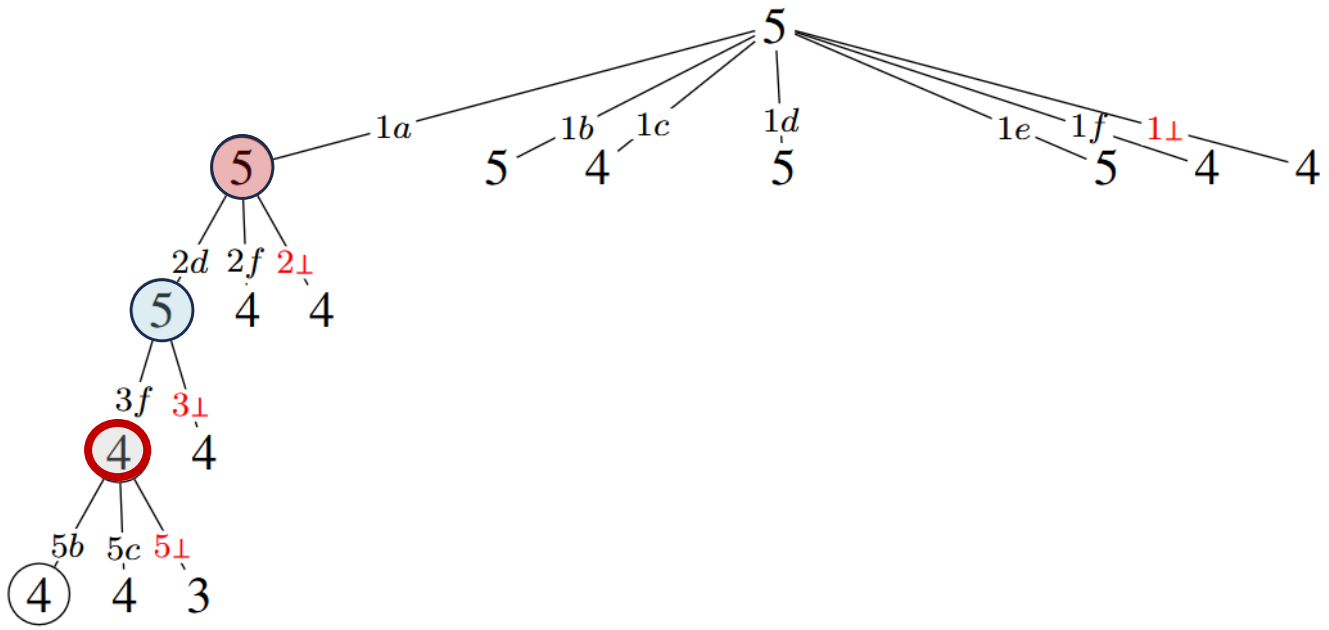


Figure 3.3: The search tree of McSplit on example graphs G and H

Reaction SMILES → Line Graph → Suboptimal MCES Dataset

*Adopted from James Trimble's PhD Thesis, 2023, Glasgow University.

UNIVERSITÄT LEIPZIG

Branch and Bound Approach: The McSplit Algorithm*

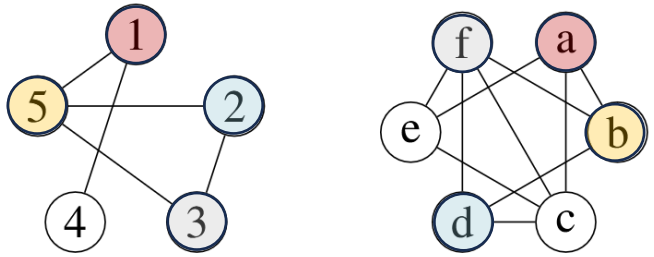


Figure 3.1: Example graphs G and H.

Labelling of G

Vertex	Label
4	100
5	111

Labelling of H

Vertex	Label
b	111
c	111
e	101

(c) After mapping 3 to f

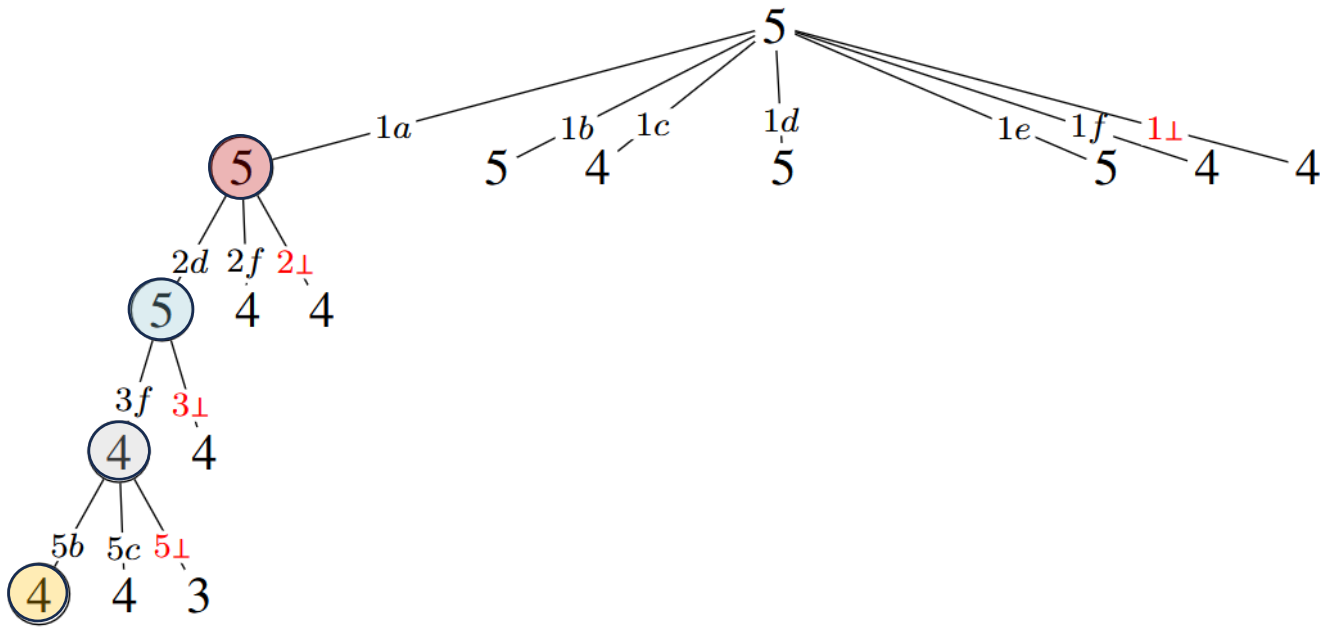


Figure 3.3: The search tree of McSplit on example graphs G and H

Reaction SMILES → Line Graph → Suboptimal MCES Dataset

*Adopted from James Trimble's PhD Thesis, 2023, Glasgow University.

UNIVERSITÄT LEIPZIG

Branch and Bound Approach: The McSplit Algorithm*

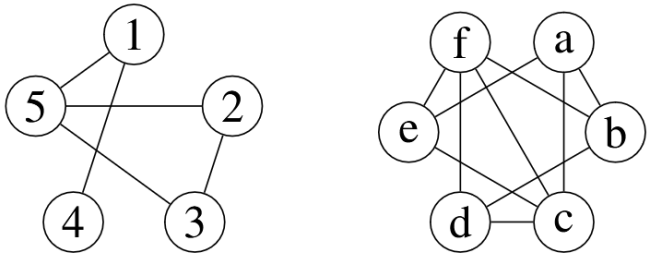


Figure 3.1: Example graphs G and H.

TOTAL:
2 Solutions of Maximum Common Subgraphs

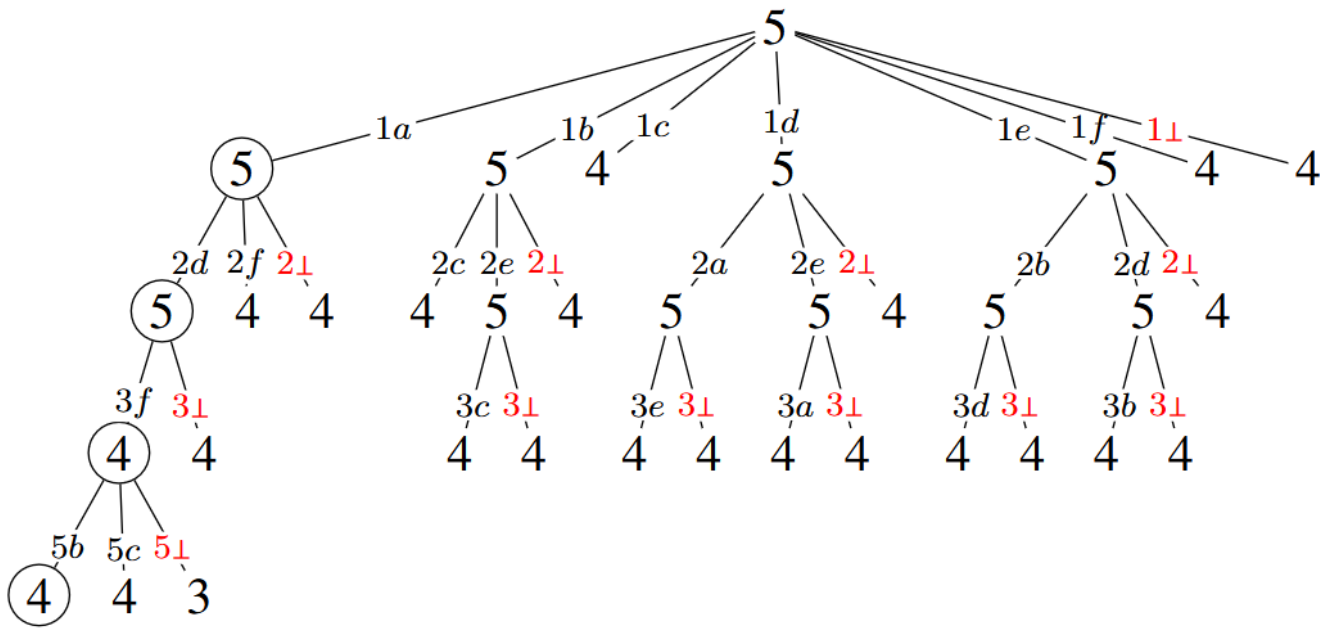


Figure 3.3: The search tree of McSplit on example graphs G and H

Reaction SMILES → Line Graph → Suboptimal MCES Dataset

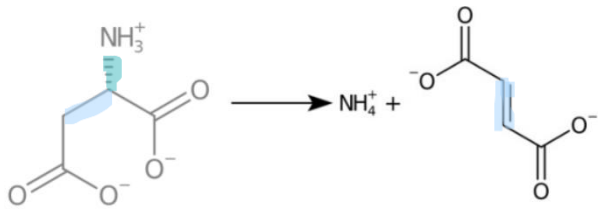
*Adopted from James Trimble's PhD Thesis, 2023, Glasgow University.

UNIVERSITÄT LEIPZIG

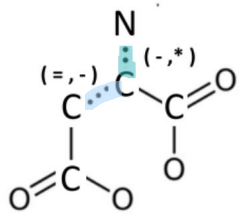
NEXT: The Machine Learning Approach

- Network Training on **ITS Graphs**
- **Graph Neuronal Networks (GNNs)** using Pytorch Geometrics → Processing of non-euclidean data
- GNNs can learn the spatial structure and chemical rules directly from the data, without manual feature engineering
- GNNs use a '**message passing**' mechanism in which information is exchanged iteratively between the nodes and their neighbours. This creates context-sensitive representations that capture local and global graph properties.

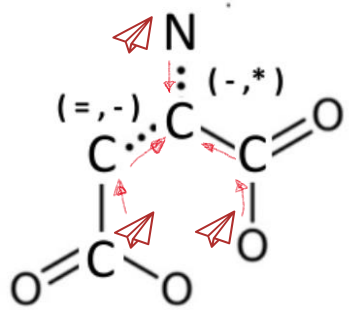
Reaction



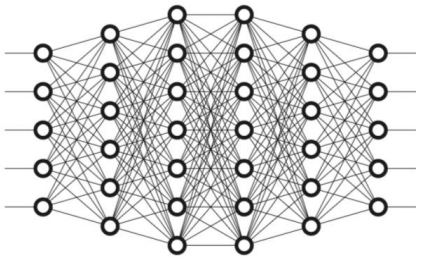
ITS



Message Passing

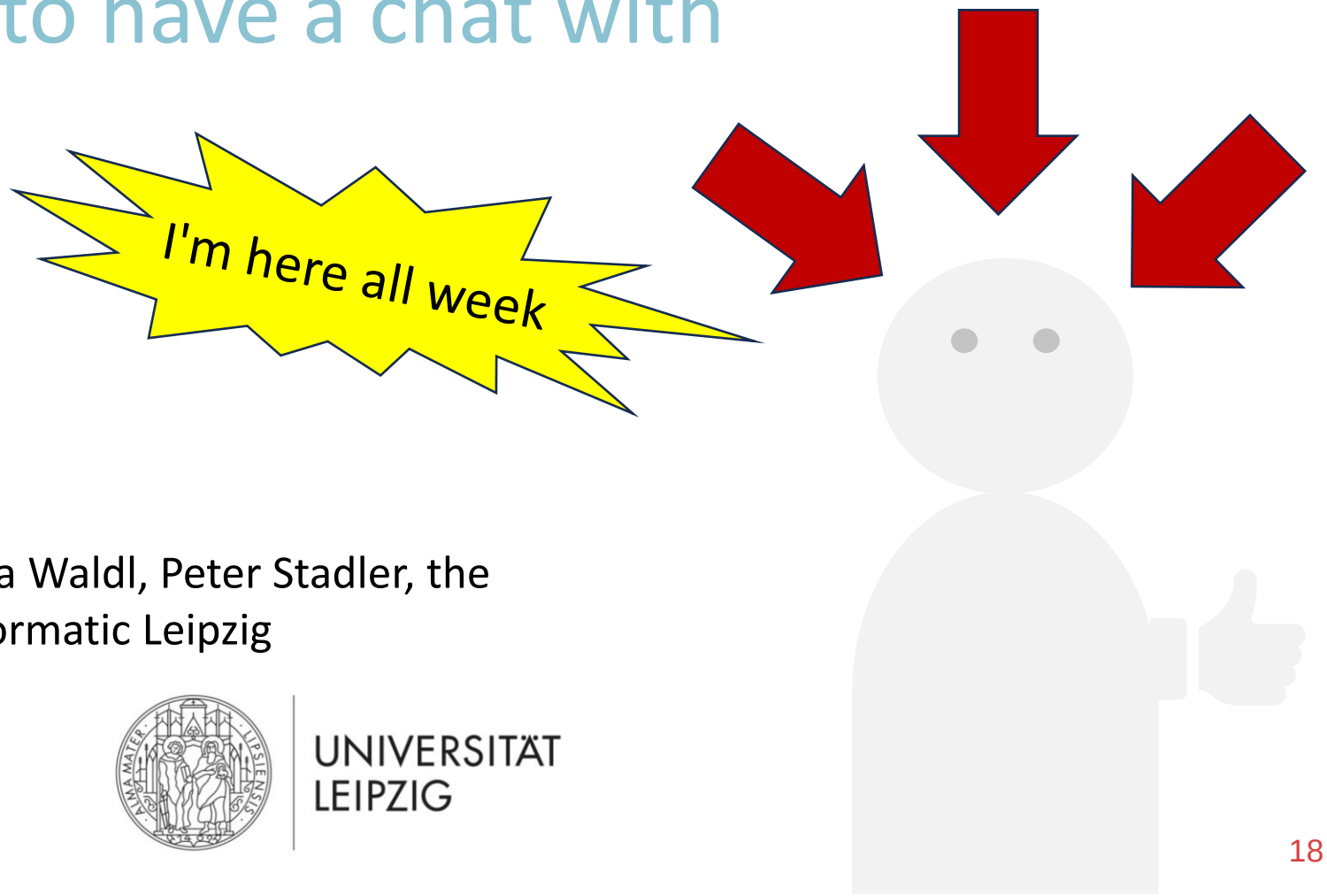


Reaction SMILES → Line Graph → Suboptimal MCES Dataset →



Thank you for the attention

And if nyone has any Experience or Ideas, I am happy to have a chat with you!



…And Thank to Thomas Gatter, Maria Waldl, Peter Stadler, the MATOMIC Team and the hole Bioinformatic Leipzig

MATOMIC by novo nordisk fonden

UNIVERSITÄT LEIPZIG