A Coloureed Petri-Net Based Framework for Atom-Tracing in Chemical Networks

Casper Asbjørn Eriksen

Department of Mathematics and Computer Science,
University of Southern Denmark

Winterseminar Bled - February 2023

Introduction

CV:

- Physics 2016-2018
- Computer Science 2018-2022
- Ph.D. Student in Algorithmic Cheminformatics (MATOMIC project, SDU)
- *Still learning how to speak chemistry*

Current project:

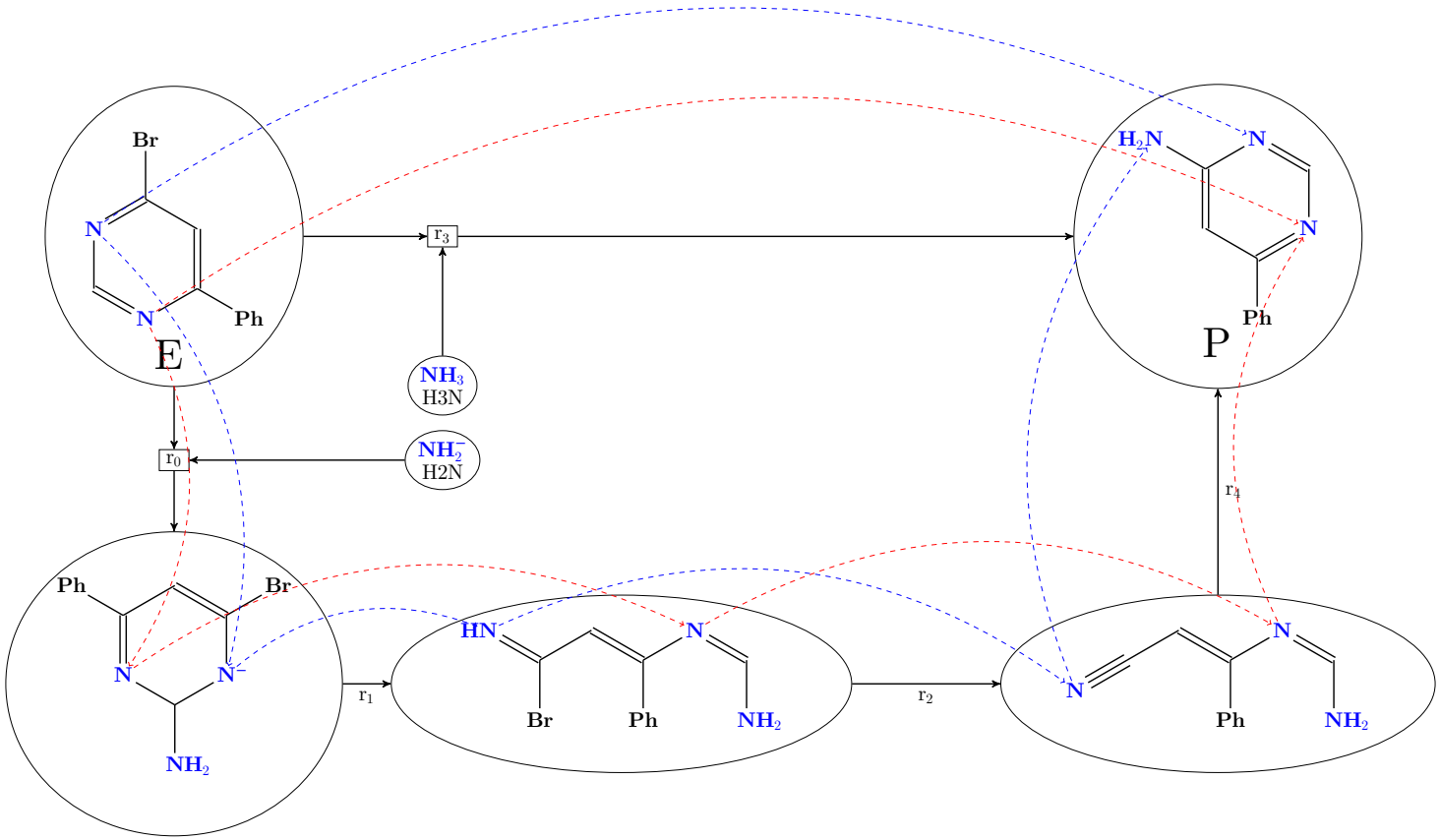
- Developing a framework for atom tracing in chemical networks

Other projects:

- Boltzmann random sampling of RNA secondary structures with pseudoknots using Analytic Combinatorics
- Alternate solution to isomer generation and bond perception using ILP

Atom Tracing

Example – ANRORC mechanism:

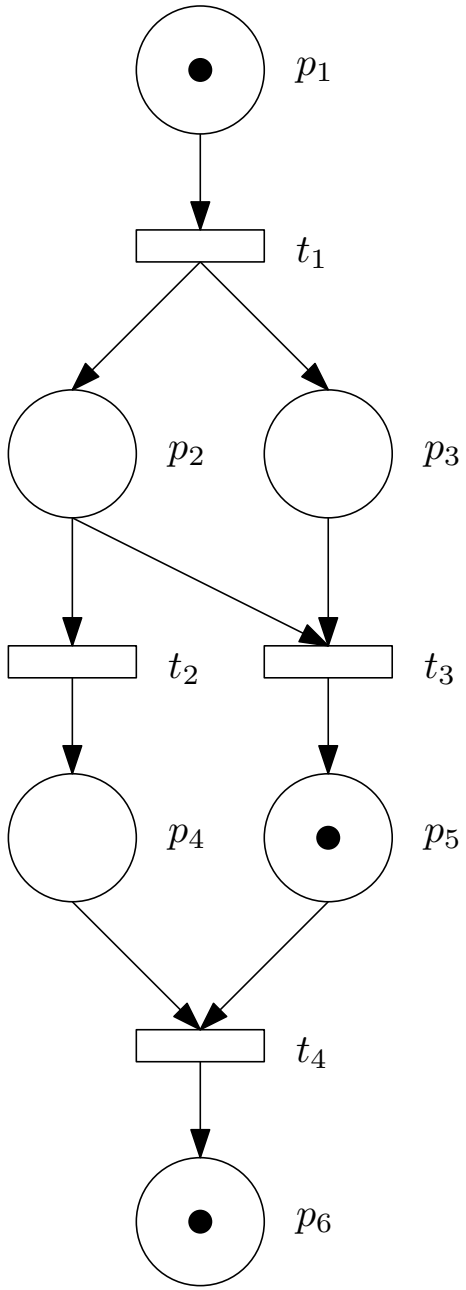


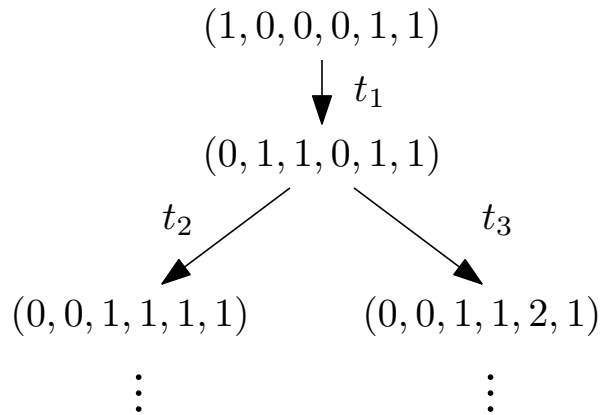
Motivation

Some potential use-cases:

- **Hypothesis generation:** Given marking of input compounds, by which mechanism is an observed marking of another compound created?
- **Isotope-marking Experiment design:** Given multiple candidate pathways, how to label input compounds such that the pathways can be distinguished by observing the marking of some other set of compounds?

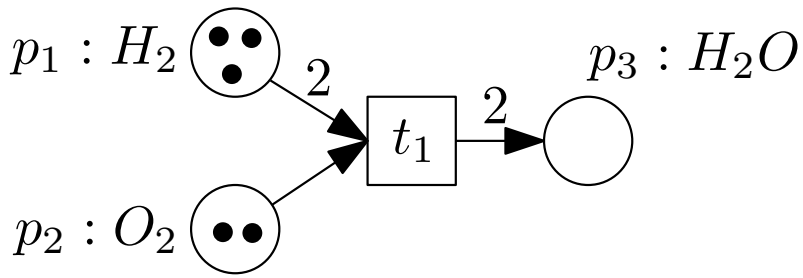
Petri Nets

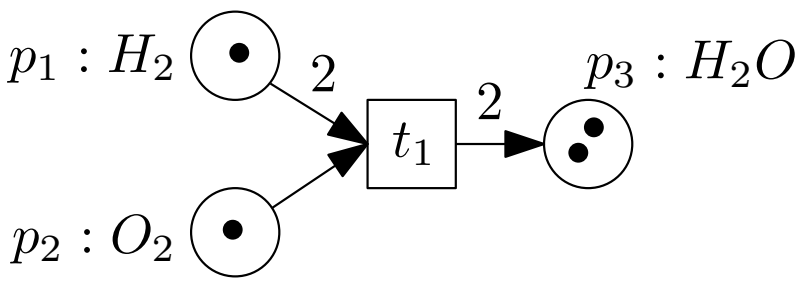

p_1
t_1
p_2
p_3
t_2
t_3
p_4
p_5
t_4
p_6

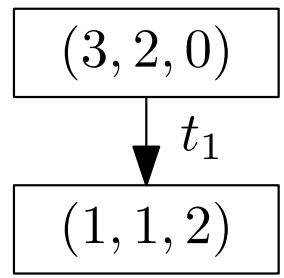

$(1, 0, 0, 0, 1, 1)$
t_1
$(0, 1, 1, 0, 1, 1)$
t_2
t_3
$(0, 0, 1, 1, 1, 1)$
$(0, 0, 1, 1, 2, 1)$

Petri Nets for Chemistry

Chemical Example: $2\,H_2 + O_2 \rightarrow 2\,H_2O$





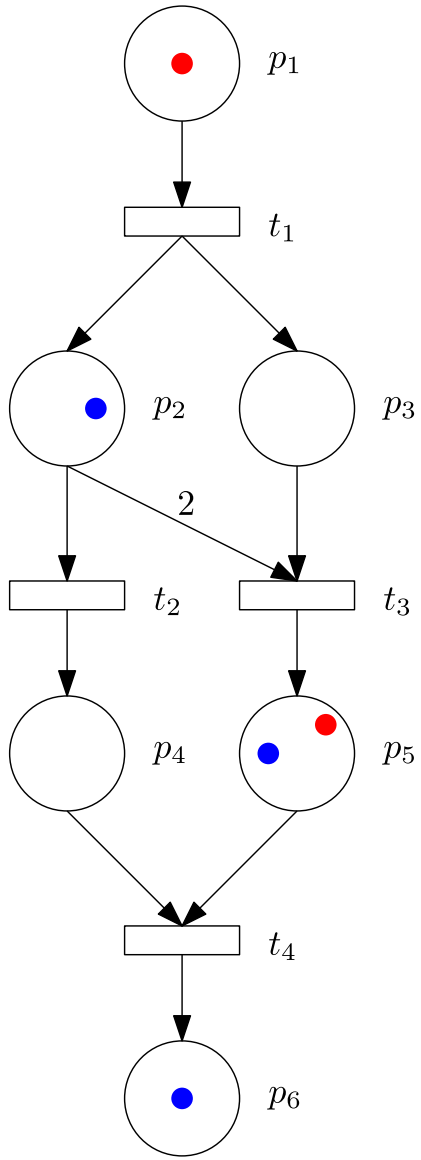


Coloured Petri Nets

For each place $p \in P$, a set of possible token colours $\Sigma(p)$.

For each transition $t \in T$, a function C which maps from the colour space of in-places to the colour space of out-places.

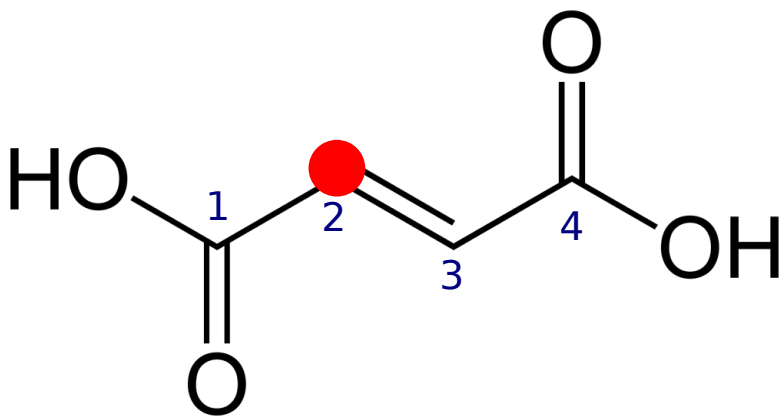
- Colours can be arbitrary mathematical objects
- Transition functions can be arbitrarily complex
- When firing, make a non-deterministic choice of token colours ('binding')



Coloured Petri Nets for Atom Tracing

The following model is used:

- **Token Colours** – Compound marking (list of booleans)
- **Transition function** – Atom map



Colour representation:

$$(0, 1, 0, 0)$$

Atom maps obtained from database or from other tools

- Potentially exponential increase in size of state space

Abstraction levels

More coarse-grained abstraction levels can be used, depending on the use-case. These differ on how markings are represented in the state space.

Level 1

Full simulation

Store a *multiset* of markings for each compound.

Abstraction levels

More coarse-grained abstraction levels can be used, depending on the use-case. These differ on how markings are represented in the state space.

Level 1

Full simulation

Store a *multiset* of markings for each compound.

Level 2

Full simulation with underlying unmarked network

Store a multiset of *marked* compounds, but do not track the count of unmarked compounds.

Abstraction levels

More coarse-grained abstraction levels can be used, depending on the use-case. These differ on how markings are represented in the state space.

Level 1

Full simulation

Store a *multiset* of markings for each compound.

Level 2

Full simulation with underlying unmarked network

Store a multiset of *marked* compounds, but do not track the count of unmarked compounds.

Level 3

Omega-marking

Store a set of markings for each compound, but do not store counts.

State Space Reduction

Two main phases:

Pre-computation (Static)

State space expansion (Dynamic) $\leftarrow$ Performance bottleneck!

The goal is to use the pre-computation to optimise the performance and memory usage of the dynamic phase.

- Atom Transition Network
- Canonicalise and compute automorphism groups

State Space Reduction

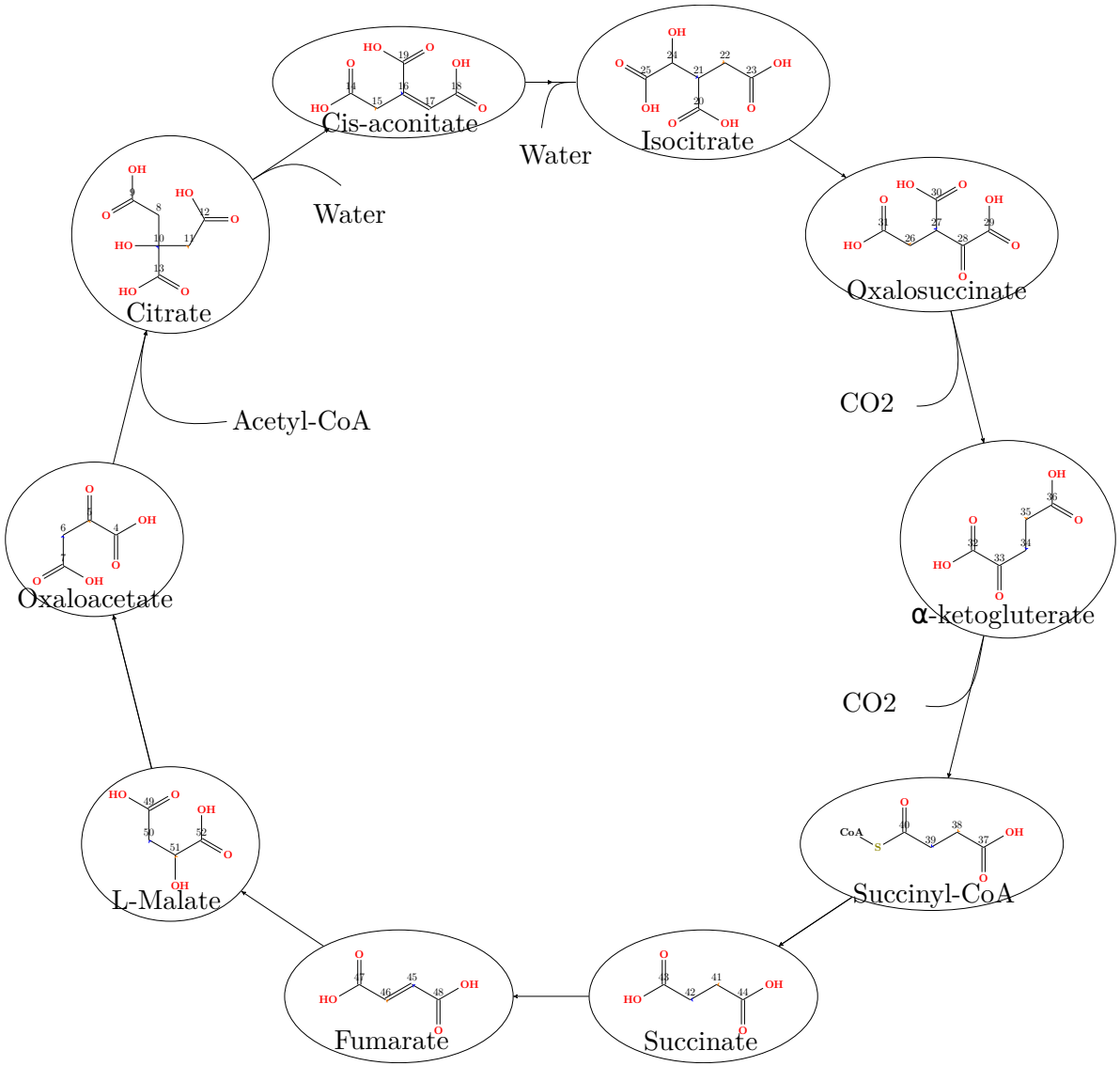
The corresponding *Atom Transition Network* (ATN) is used to optimise state space expansion.

The atoms reached in the ATN represent a superset of reachable atoms.

- A trace only exists in the CPN if it exists in the ATN.
- We can greatly reduce the number of atoms we need to track explicitly, reducing memory use of the state space.

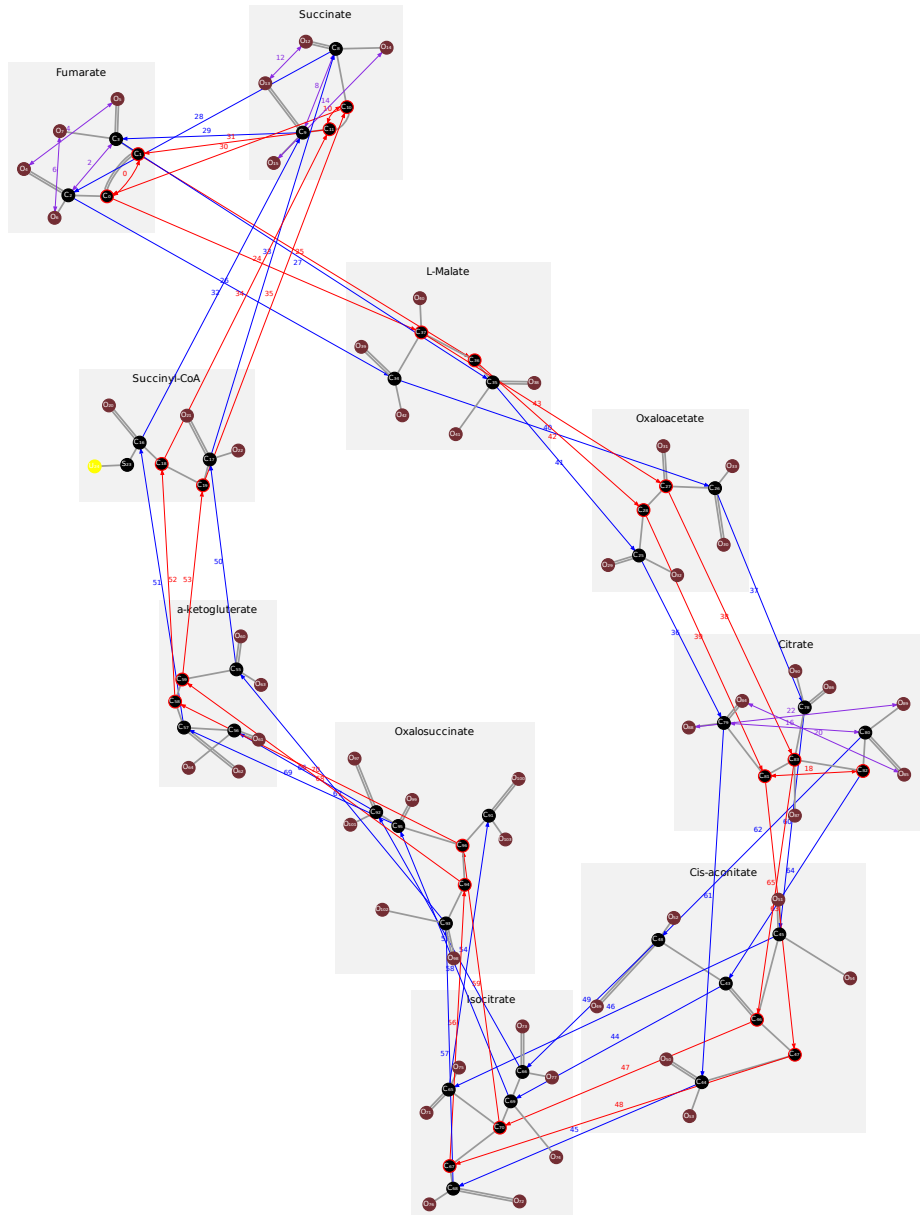
Atom Transition network

TCA cycle (simplified):



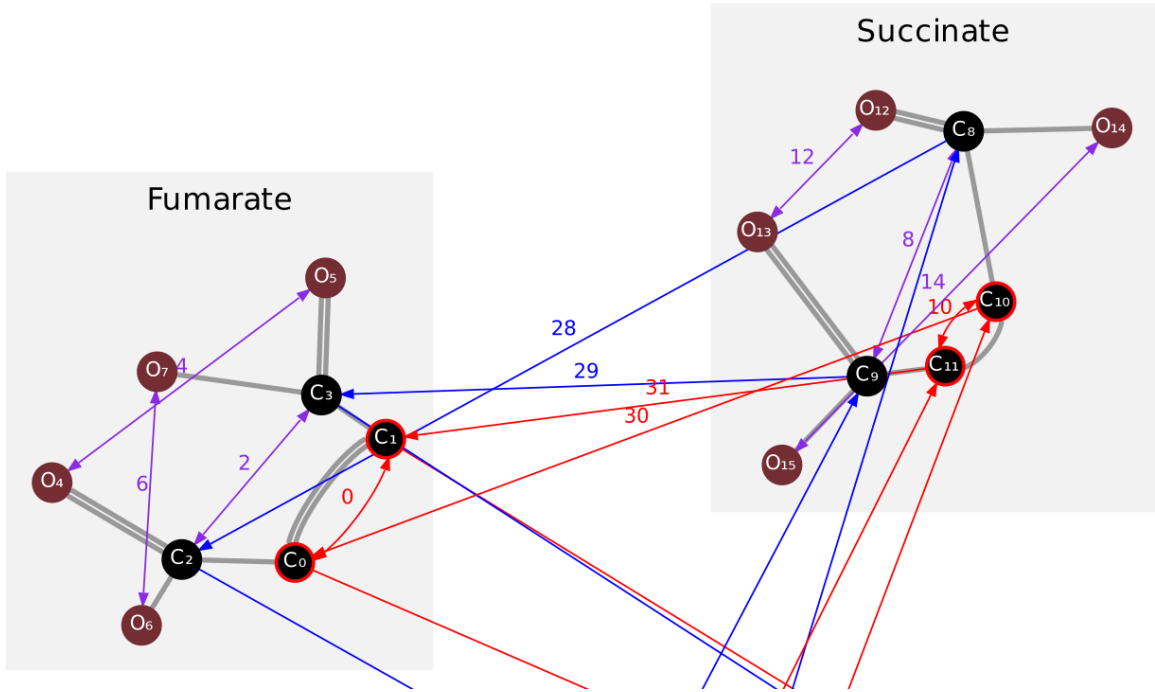
Atom Transition Network

- Reaction and automorphism edges
- Subnetwork generated based on source/sink atoms



Atom Transition Network

- Reaction and automorphism edges
- Subnetwork generated based on source/sink atoms



Automorphism edges generate the automorphism group of the compound. The orbit of a vertex $\mathcal{O}(v)$ is the set of vertices reahable by the automorphism edges.

State Space Reduction

Reducing the size of each state in the state space:

- ATN subnetwork specifies a set of atoms to track in each compound.
- Only track reachable atoms of desired type of element.

$$(0, 0, 1, 1, 0, 0, 1, 0)$$
$$\downarrow$$
$$(0, 0, 1, 1, 0, 0, 1, 0)$$
$$\downarrow$$
$$(1, 1, 0, 1)$$

Given graph $G = (V, E)$ and a subset of tracked vertices $V' \subset V$:

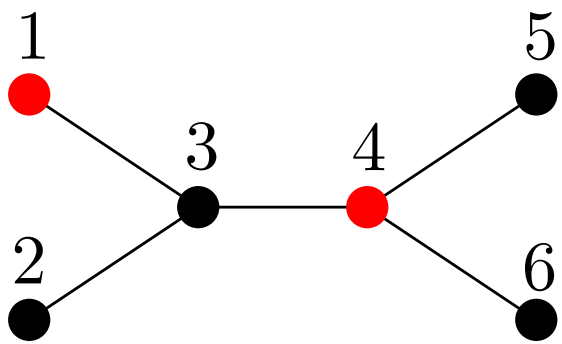
$$\forall v \in V' : \quad \mathcal{O}(v) \subseteq V',$$

so the automorphism group $\mathrm{Aut}(V')$ is trivially calculated due to the construction of V'.

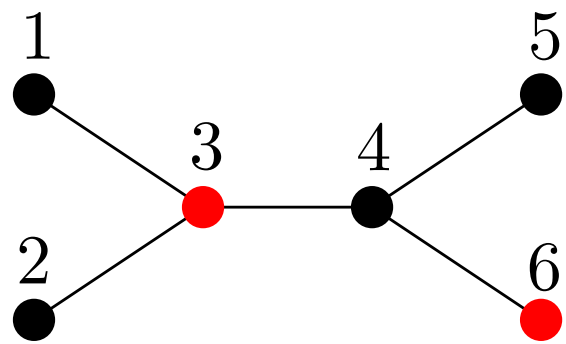
Labeled graph canonicalisation

State space is reduced by only storing canonical labelled compound:

- Canonicalisation can be optimised based on the automorphism group of the underlying graph.



$(1, 0, 0, 1, 0, 0)$



$(0, 0, 1, 0, 0, 1)$

Labeled graph canonicalisation

State space is reduced by only storing canonical labelled compound:

- Canonicalisation can be optimised based on the automorphism group of the underlying graph.

Problem: Let $G = (V, E)$ be a graph with labels $l : V \to \mathbb{Z}^+$, and automorphism group $\mathrm{Aut}(G, l)$. Given an additional set of labels $l' : V \to \mathbb{Z}^+$, canonicalise the graph according to the labelling $l^* : V \to (\mathbb{Z}^+ \times \mathbb{Z}^+)$:

$$l^*(x) = (l(x), l'(x)).$$

In this case, l is the labeling which is given statically (element, charge, ...), l' is a binary labeling representing whether a given atom is marked with an isotope.

Challenges

- Unreliable atom maps
  - Accuracy of atom mappers greatly depends on reaction type (60% to 99%)[1]
  - Compounding errors lead to incorrect traces
- Incomplete chemical networks

[1] Preciat Gonzalez, G. A. et al. *Comparative evaluation of atom mapping algorithms for balanced metabolic reactions*. J Cheminform 9, 1–15 (2017).

Progress & Next Steps

- Currently implementing state space search algorithms
- Experiment modelling
  - Model fragmentation patterns
- User interface

Thank you!

DEPARTMENT OF MATHEMATICS
AND COMPUTER SCIENCE