enviPath | THE ENVIRONMENTAL CONTAMINANT BIOTRANSFORMATION PATHWAY RESOURCE

Eawag
Swiss Federal Institute of Aquatic Science and Technology

eawag aquatic research

Universität Zürich[UZH]

Generating suspect lists for transformation product screening

Athira Shankar

11 May 2025

Overview

1. Background and motivation
2. The algorithm
3. Hands on session

Background

- Suspect screening lists are predefined sets of TPs **expected** or **suspected** to form from the parent compounds.
- These lists help in priortizing features that match the suspect list among thousands of detected features, focusing efforts on the most relevant transformation products.
- Pathway prediction tools are popularly used for the generation of suspect screening lists (Trostel et al.,2023)

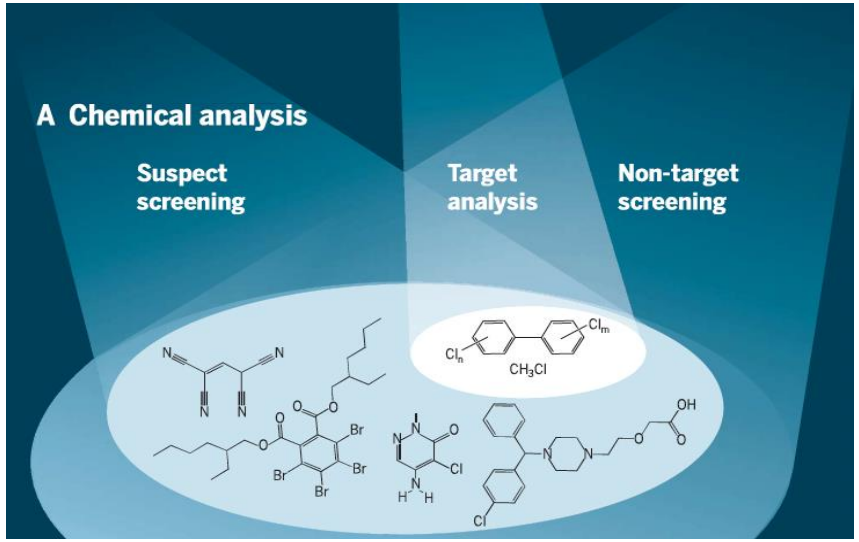


Figure adapted from Escher et al., 2020

Motivation- Need for Efficient Suspect list Generation

Background:

- Trostel et al.,2023 conducted a comprehensive literature review on suspect screening of transformation products (TPs) in wastewater and activated sludge systems.

Key Finding:

- Most studies relied heavily on EAWAG-PPS for TP prediction

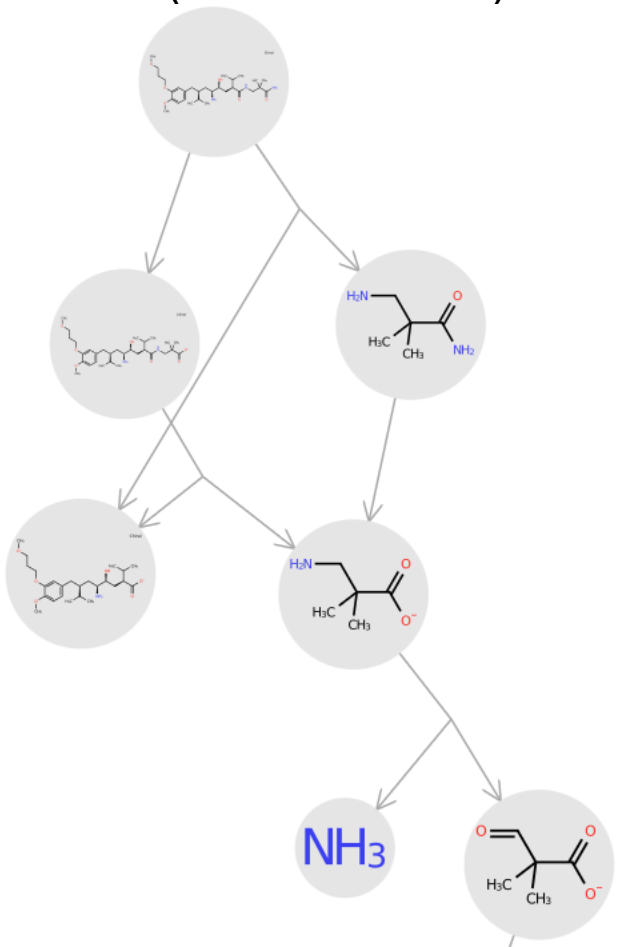
Problem identified:

- enviPath lacked batch prediction capabilities making it difficult to generate suspect lists for a large set of compounds.

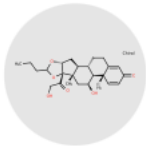
Limitations of the current threshold based prediction system

Single Threshold Setting Fails Across Different Compound Types

Aliskiren (Threshold=0.5)



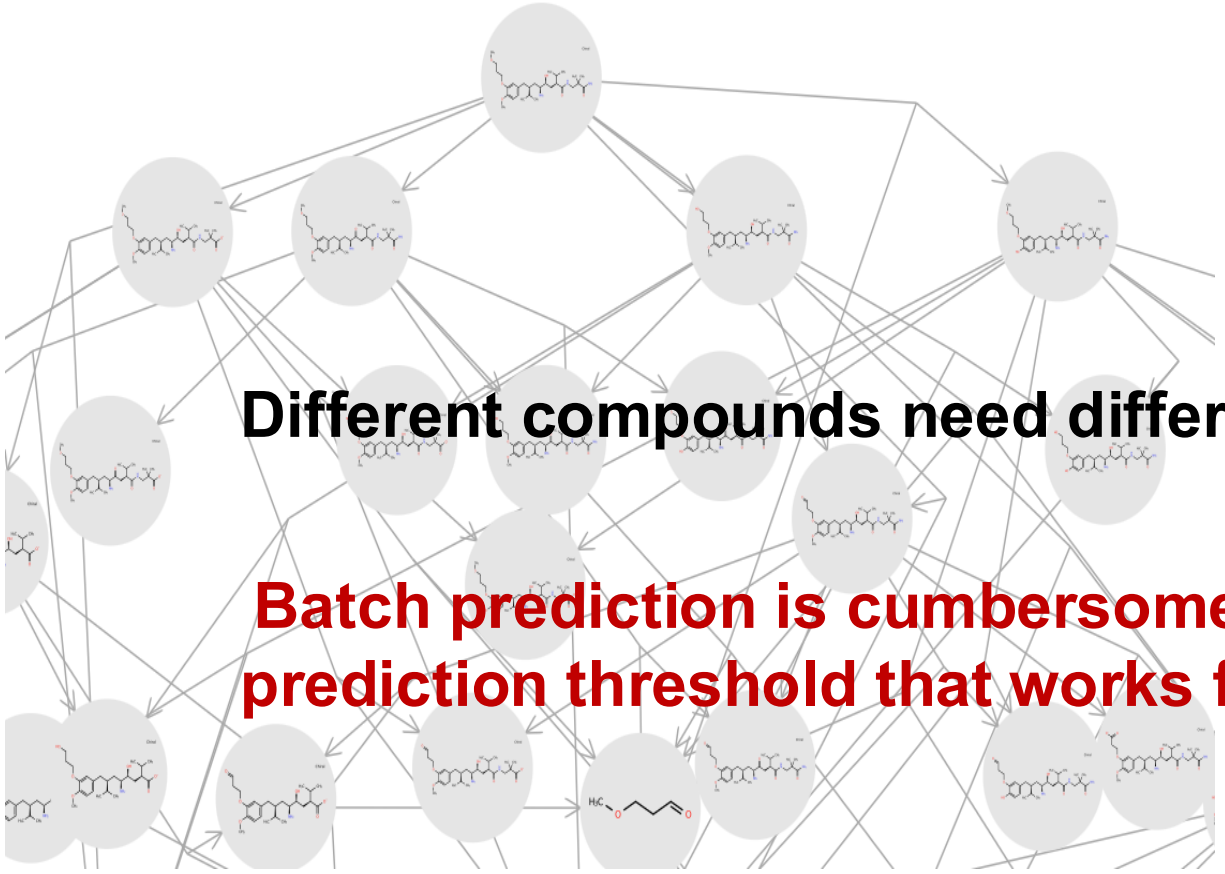
Budesonide (Threshold=0.5)



Limitations of the current threshold based prediction system

Single Threshold Setting Fails Across Different Compound Types

Aliskiren (Threshold=0.1)



Budesonide (Threshold=0.1)

Different compounds need different prediction thresholds!

Batch prediction is cumbersome without a magic prediction threshold that works for every compound!

Challenge of batch prediction in enviPath

- This issue was encountered by Trostel et al., 2023 while generating a suspect list for 42 pharmaceuticals.
- **Solution:**
  To address this, a Python-based batch prediction workflow was introduced using **enviPath-python.**
- Generated more correctly predicted TPs than EAWAG-PPS

Combining predictive and analytical methods to elucidate pharmaceutical biotransformation in activated sludge†

Leo Trostel, ‡[a] Claudia Coll, ‡[a] Kathrin Fenner *[ab] and Jasmin Hafner [ab]

Introducing Batch Mode in enviPath

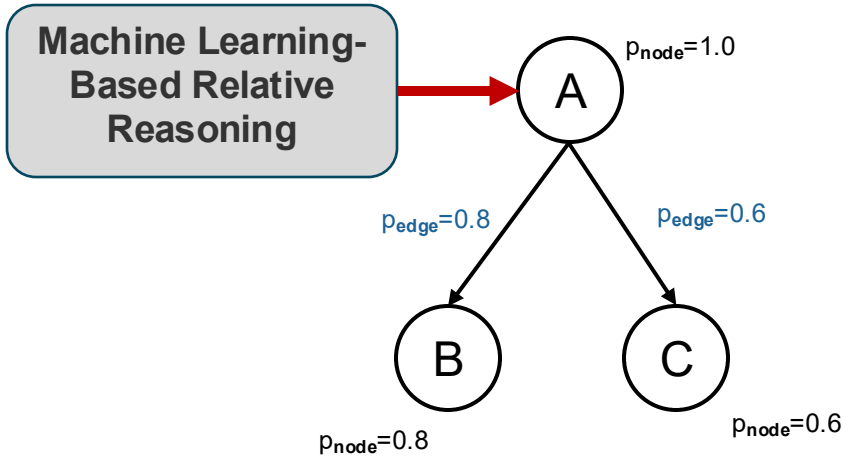
Single threshold predictions were inefficient for diverse compounds, especially in the context of suspect list generation.

Our solution:

- **No fixed threshold** – We consider all TPs with probability>0
- **Greedy Top-K search**: Expand highest-probability TPs first
- Pathway **expansion stops** when a **user-defined number of TPs** is reached or no more products exist to explore.
- A user friendly visual interface that enables scalable and efficient suspect list generation

Greedy search algorithm

- Greedy algorithm makes the local optimal choice at each stage.



A priority queue that shuffles TPs in descending order based on the calculated combined probability

B: 0.8	C: 0.6

For a child node n generated during pathway search, the **combined probability** is calculated as:

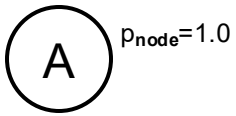
$$p_{node,n}=p_{node,n-1}\times p_{edge,n-1\rightarrow n}$$

Greedy search algorithm

Local expansions

Priority queue

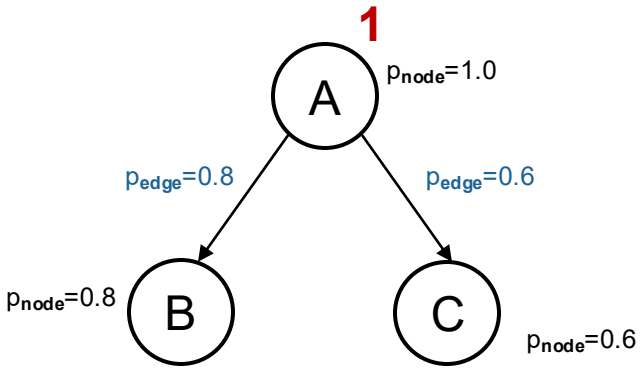
Parent compound



Empty queue

Greedy search algorithm

Local expansions

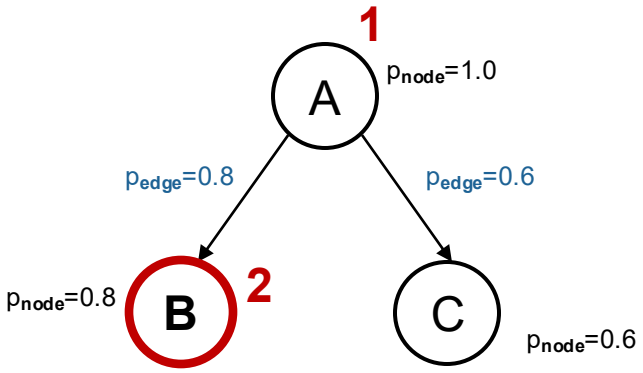


Priority queue

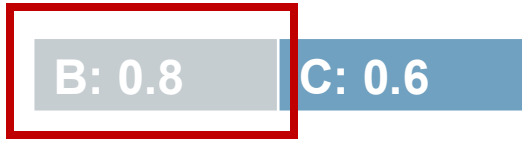
B: 0.8	C: 0.6

Greedy search algorithm

Local expansions

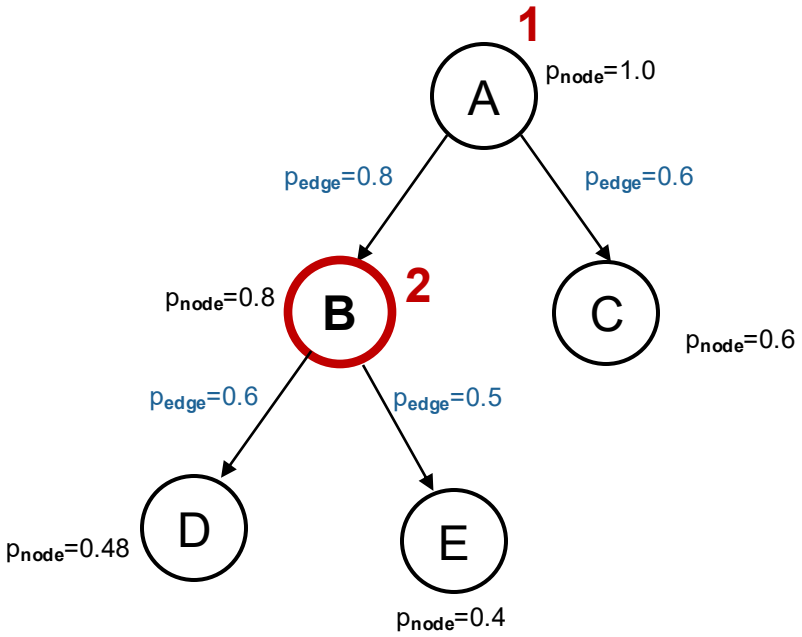


Priority queue

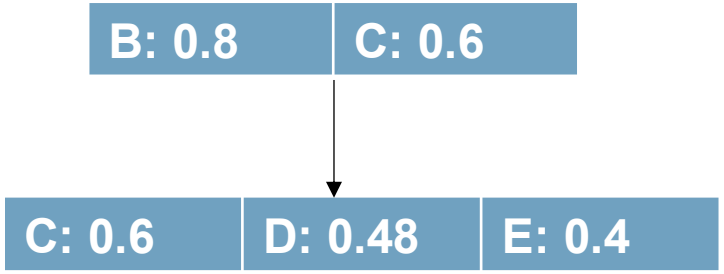


Greedy search algorithm

Local expansions

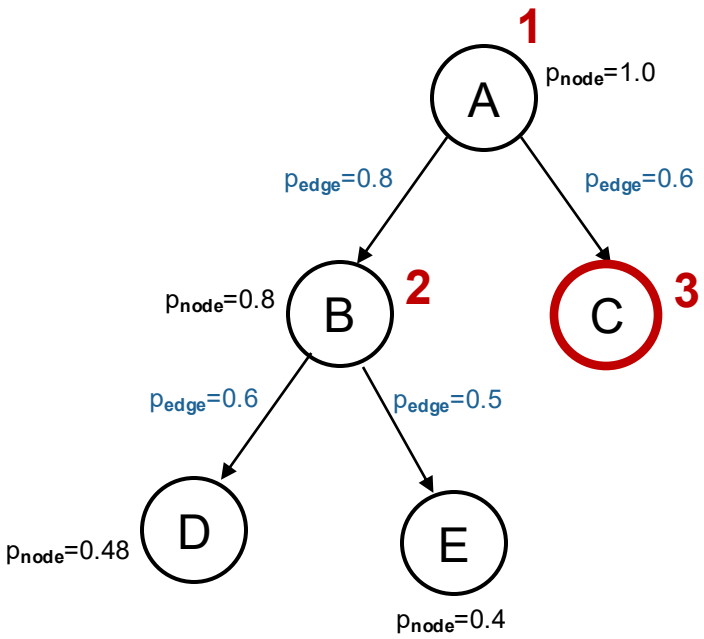


Priority queue

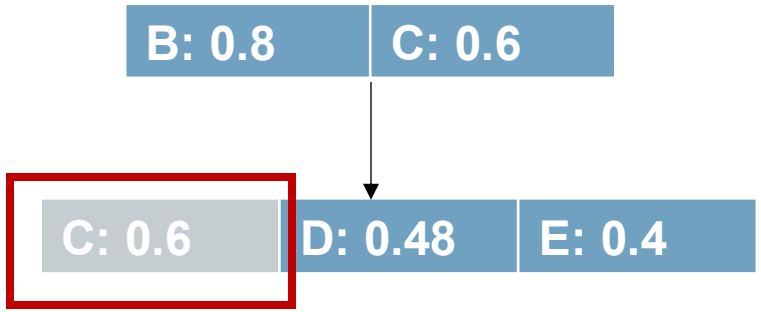


Greedy search algorithm

Local expansions

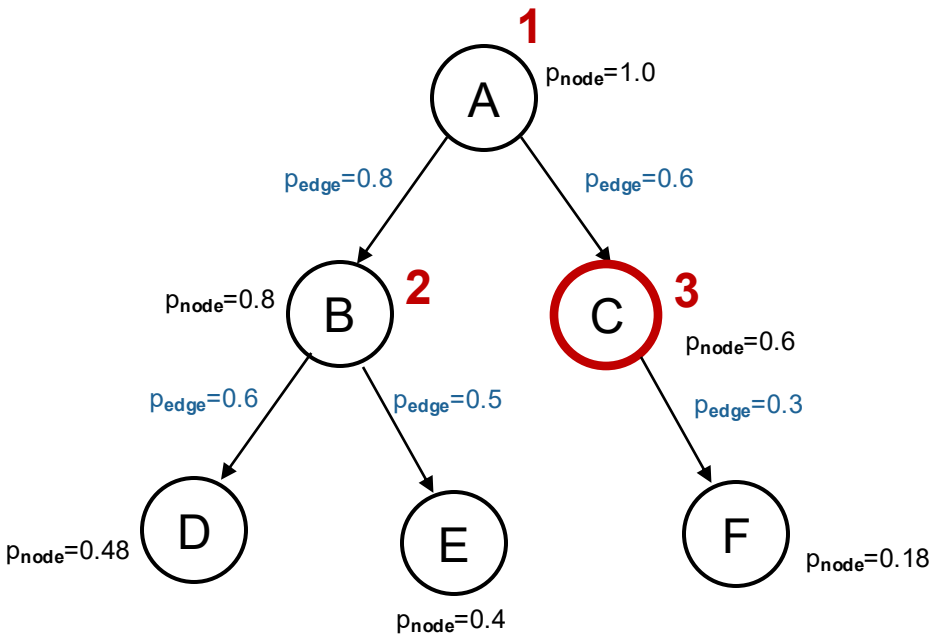


Priority queue

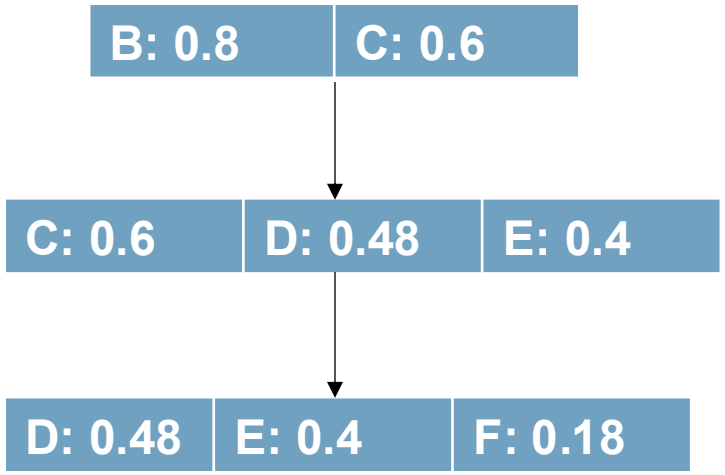


Greedy search algorithm

Local expansions

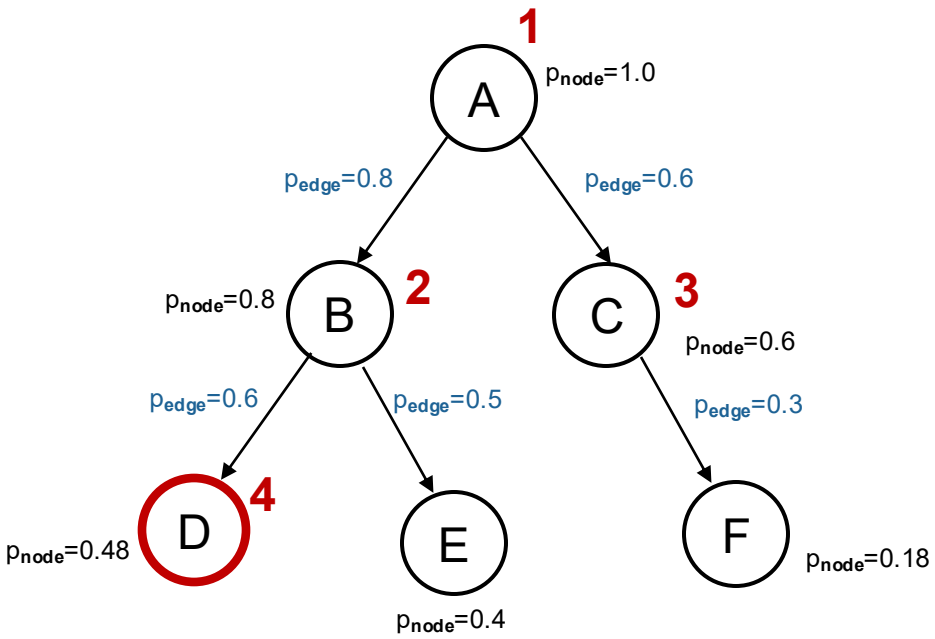


Priority queue

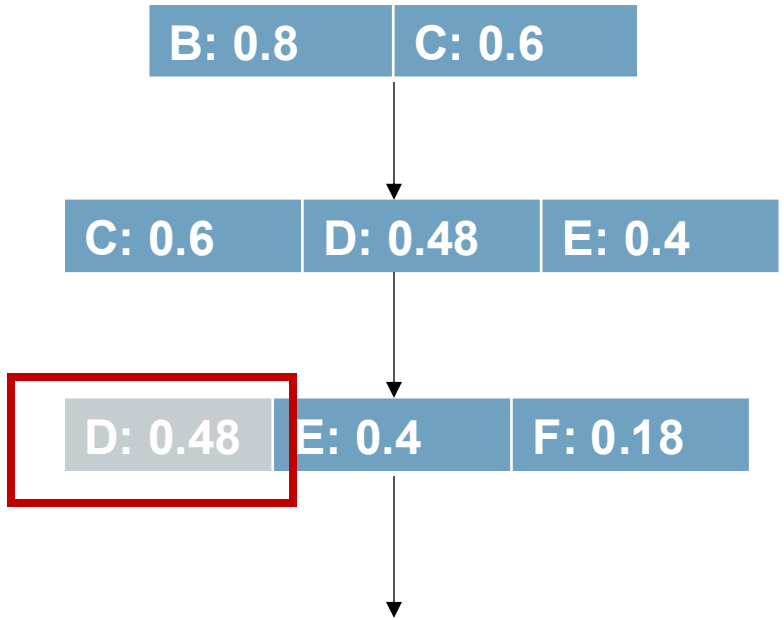


Greedy search algorithm

Local expansions

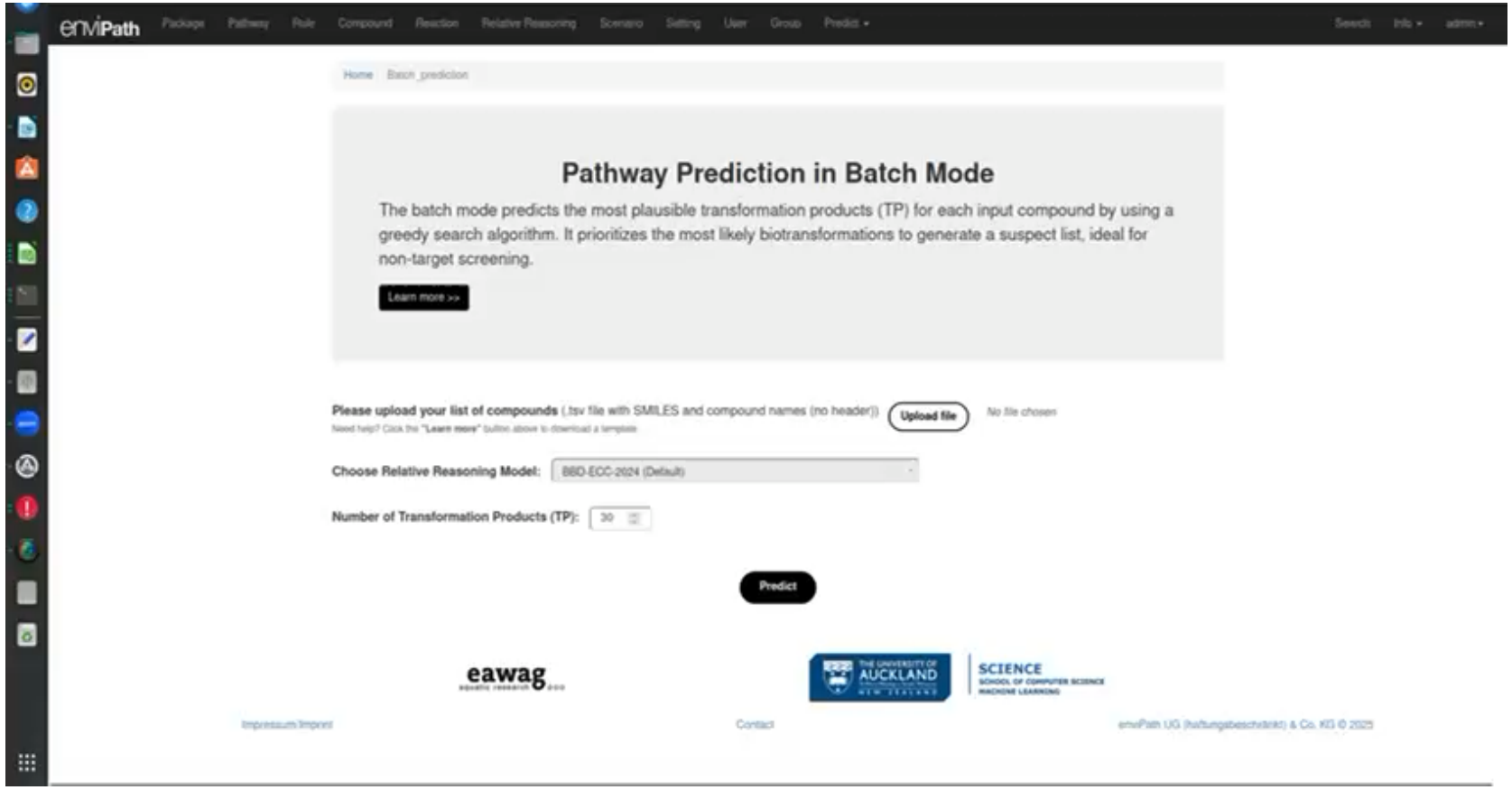


Priority queue



The expansion continues until the user-defined number of TPs is reached or until the queue becomes empty.

Hands on exercise



Output file

Output fields	Definition
SMILES	SMILES representation of the TP
name	“Parent” or “TP_n”
combined_probability	Node probability used to prioritize TPs
rule	bt rule number
rule_id	Link to the bt_rule in enviPath
generation	Generation of the TP
probability	Reaction probability generated by enviPath relative reasoning model
parent	SMILES of the parent of the TP
parent depth	Generation of the parent of the TP

References

1. Trostel, L., Coll, C., Fenner, K., & Hafner, J. (2023). *Combining predictive and analytical methods to elucidate pharmaceutical biotransformation in activated sludge*. **Environmental Science: Processes & Impacts**, 25, 1322–1336.

Acknowledgements

Jasmin Hafner

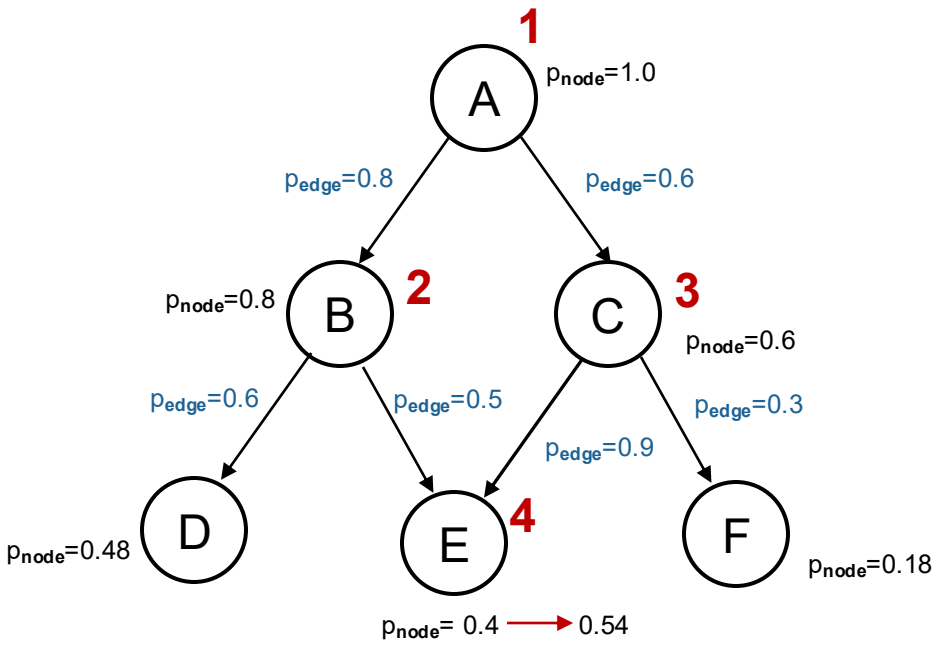
Kunyang Zhang

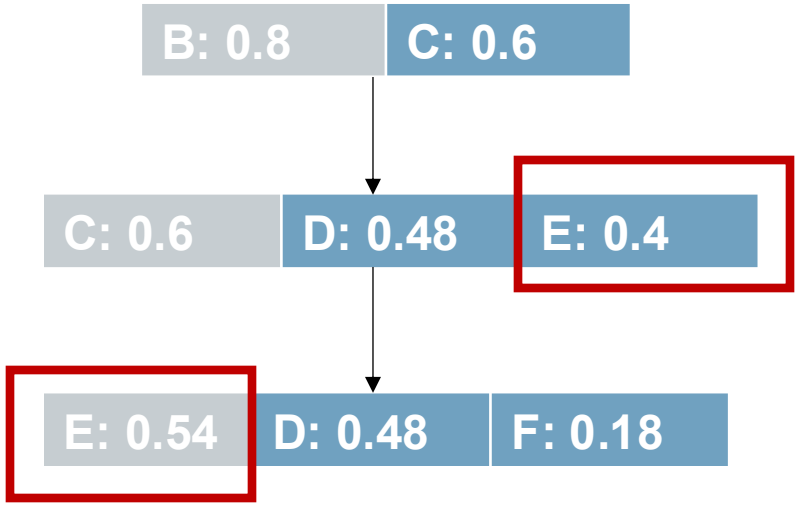
Tim Lorsbach

Kathrin Fenner

Supplementary slides

Probability of nodes with multiple parents



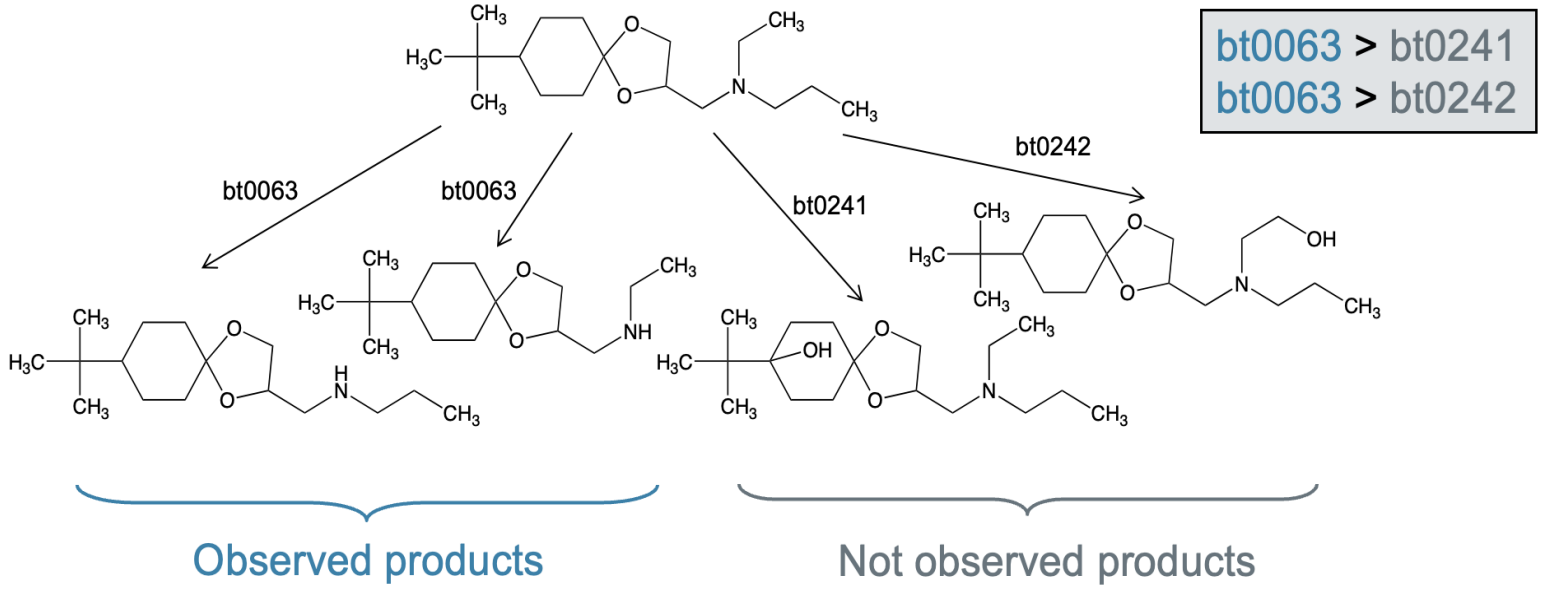


$P_{node}(E) = \max(0.8 * 0.5, 0.6 * 0.9)$

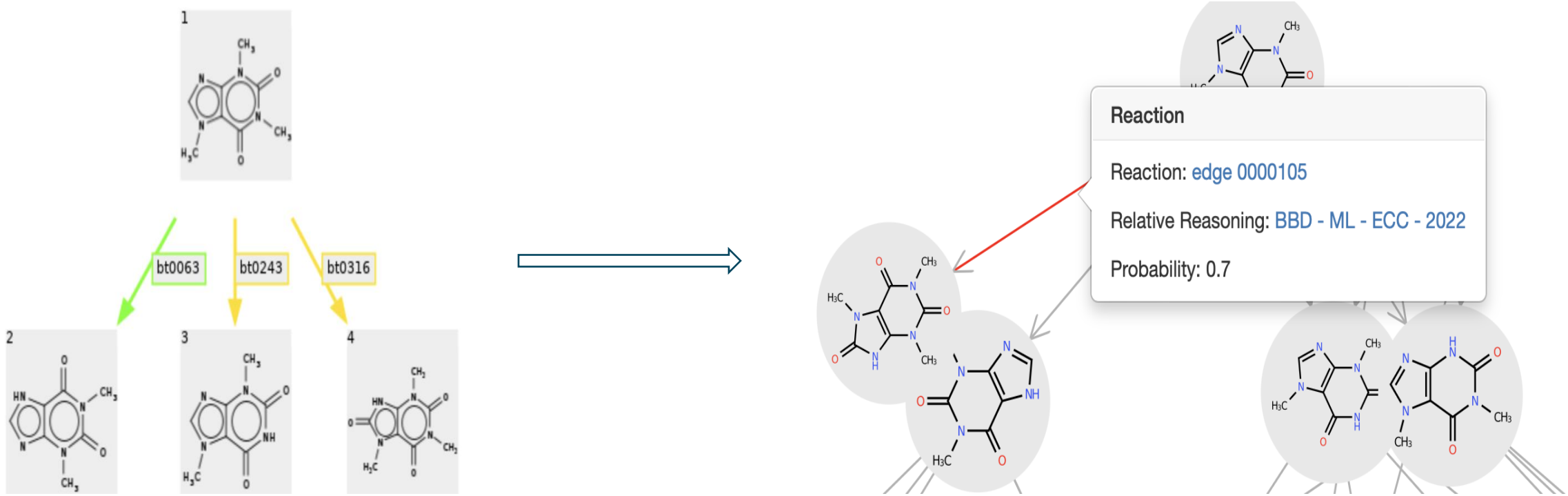
$P_{node}(E) = 0.54$

Bonus slide: Machine learning based relative reasoning models in enviPath

- Rule based relative reasoning was introduced in EAWAG-PPS (Fenner et al., 2008) to tackle combinatorial explosion.
- Manual relative reasoning rules requires expert knowledge, is time consuming to develop and hard to scale.
- Machine learning based relative reasoning (Wicker et al., 2010) was introduced that automatically rank rules for compounds.



Bonus slide: Machine learning based relative reasoning models in enviPath



- enviPath allows users to set probability thresholds, displaying only rules with probability above the selected threshold.