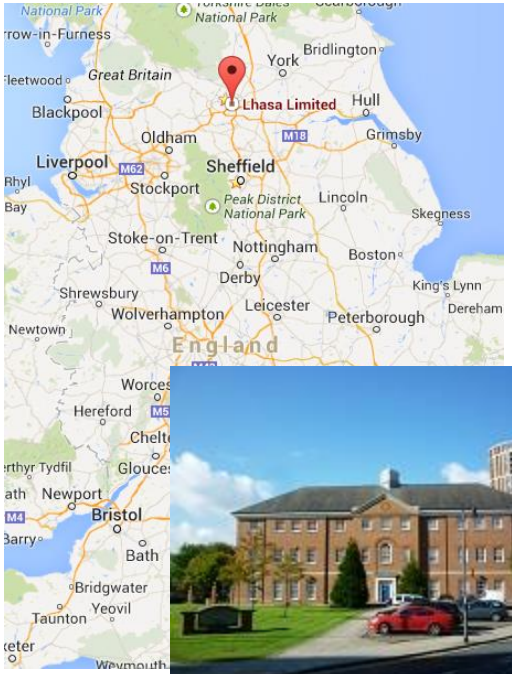
Feature combinations networks with statistical (Q)SAR models: interpretation and knowledge mining

Samuel Webb
samuel.webb@lhasalimited.org

Who is Lhasa

- Not-for-profit organisation
- Data sharing initiatives
- Software for *in silico* prediction of:
  - Toxicity: Derek & Sarah Nexus
  - Metabolism: Meteor Nexus
  - Degradation: Zeneth
- Data sharing initiatives:
  - Toxicity database: Vitic Nexus
  - Pre-competitive data sharing initiatives: aromatic amines, intermediates…



What is this presentation about?

- Interpretation of *in silico* predictions
  - What is an interpretation?
  - Why do we want an interpretation?
  - What are the current limitations?
- How do we get an interpretation
  - Literature published algorithms
  - Feature combination networks algorithm
- Interpretation to knowledge mining…

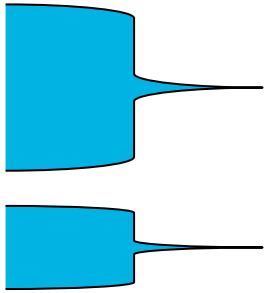
BACKGROUND

Model purpose

- To predict the target activity or property of a previously unseen structure
- In this presentation we will be considering the specific case of binary activity predictions

(Q)SAR / (Q)SPR modelling

- Take a dataset
- Build a model
- Make a prediction



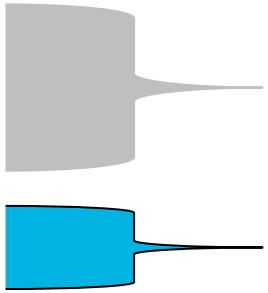
Focus of the modeller

Focus of the user

There can be a disconnect between the purpose the model is being built for and concerns of the modeller.

(Q)SAR / (Q)SPR modelling

- Take a dataset
- Build a model
- Make a prediction



Focus of the modeller

Focus of the user

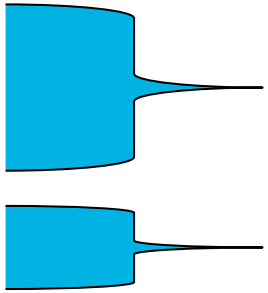
The requirements/interests of a user may differ depending on the use case the model is required for.

Two main points:

1) Prediction: accuracy is important
2) Understanding: accuracy and interpretation are important

(Q)SAR / (Q)SPR modelling

- Take a dataset
- Build a model
- Make a prediction



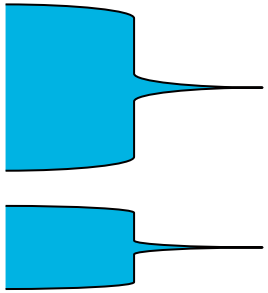
Focus of the modeller

Focus of the user

If accuracy is key and we may not require an interpretation (batch processing / screening / scoring) then the modeller may be able to limit the concerns over the interpretability of a particular learning algorithm.

(Q)SAR / (Q)SPR modelling

- Take a dataset
- Build a model
- Make a prediction



Focus of the modeller

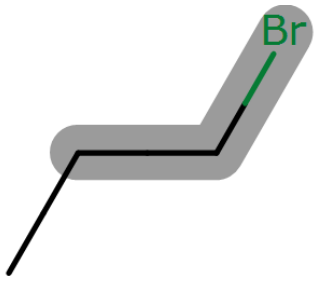
Focus of the user

If interpretation is important the modeller may be more constrained in terms of choice of descriptors and algorithm.

In some cases a poorer performing model may be chosen for its interpretation

What constitutes an interpretation?

- The model provides a prediction
  - This structure is predicted to be mutagenic
- The ability of the model to explain the prediction is the first step
  - This structure is predicted to be mutagenic due to the structural motif X
- The ability to link the explanation of the prediction to the domain being modelled
  - Structural motif X is an alkylating agent which is an electrophilic species capable of directly alkylating DNA



What limits an interpretation?

- The learning algorithm used:
  - White box models give an explanation for the prediction
  - Black box models do not provide an explanation for the prediction
- The descriptors used:
  - If the descriptors are not intelligible the prediction can't be either
  - If too many descriptors are used even interpretable learning algorithms become uninterpretable
- Machine learning generally does not provide a mechanistic reasoning
  - Given sufficient choice in descriptors and a method of identifying a cause in the prediction the user can be supported in forming a mechanistic reason

What limits an interpretation?

- Decision tree:
  - White box: path through the tree provides the cause of the prediction
    - Requires careful choice in descriptor!
- kNN:
  - White box (?): nearest neighbours can be shown and weights indicated
    - Subjective, doesn't indicate the impact of structural features
- Random Forest:
  - Black box: global importance measures do not provide sufficient detail to understand the cause of a specific prediction
- Neural network, Support Vector Machine:
  - Black box: requires post processing to elucidate the cause of the prediction

What limits an interpretation?

- Why not just use a white box learning algorithm?
  - In practice the performance of the white box algorithms is often lower than that of the black box algorithms
- Why not just use an expert system?
  - Time consuming to develop expert knowledge and encode in a predictive system – not always the desirable choice
  - Interpretation algorithms can be combined with machine learning algorithms to support the development of expert systems

Why do we want an interpretation?

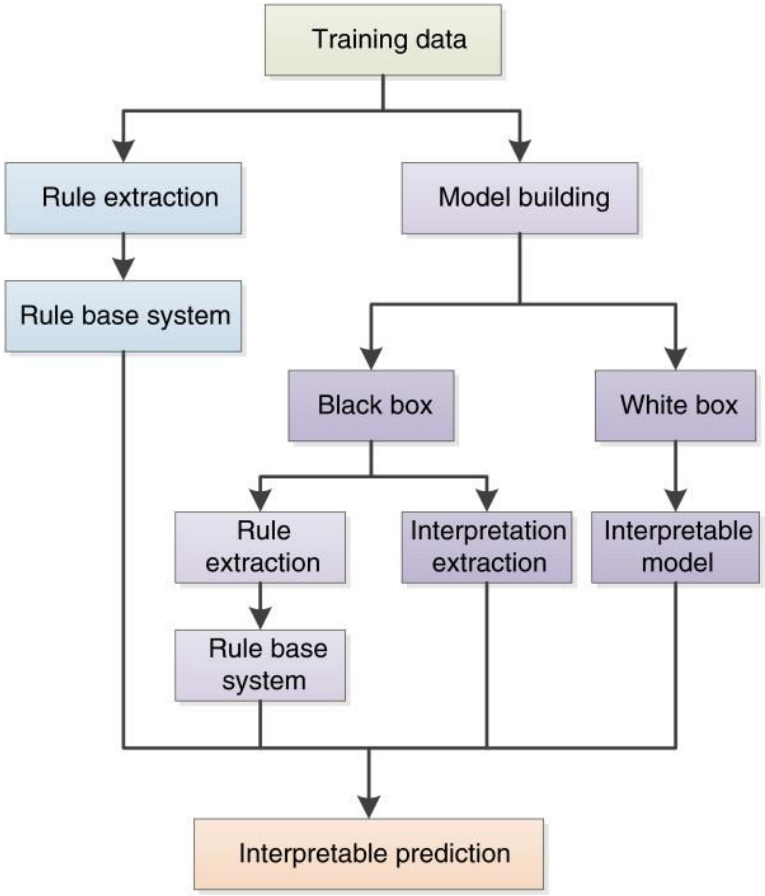
- When we know why a prediction has been made we are a step further to:
  - Mitigate negative effects
  - Improve positive effects
  - Can allow an expert to challenge or accept a prediction

- We want this interpretation without a loss in performance:
  - Add an interpretation to black box algorithms
  - Designed to be agnostic to the learning algorithm chosen

EXISTING ALGORITHMS

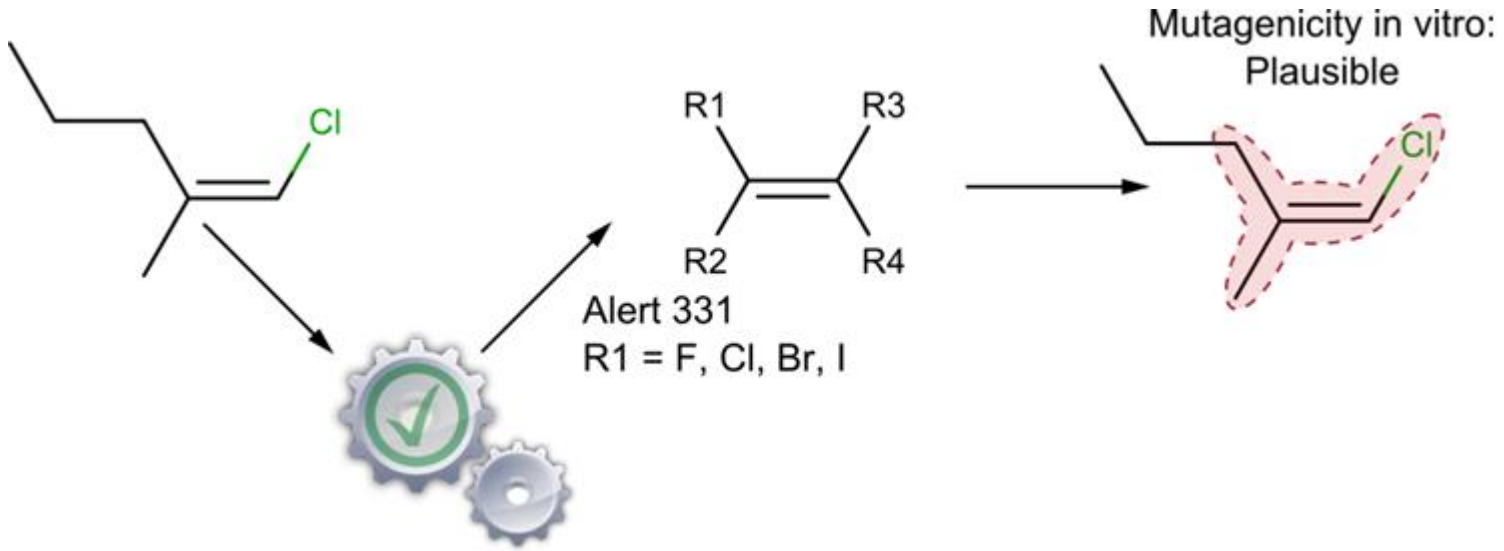
Existing algorithms

- A variety of approaches for interpretation have been developed
- Often specific to a learning algorithm
- Can be grouped into:
  - Visualisation of relevant training structures
  - Global importance measures
  - Identification of significance of features
  - Identification of behaviour of:
    - Fragments
    - Atoms and bonds
    - Physicochemical properties



Expert systems

- Tell you why a prediction was made
- May provide mechanistic / domain information for the prediction



Identifying the importance of features

- Global ranking of descriptors
  - Can be achieved through descriptor scrambling (such as random forest approach)
- Can be misleading:
  - Sparse features may make an insignificant global impact
  - May not account for combination effects
- Provides broad level information when performed globally
- Can provide fine level information when performed locally with interpretable descriptors

Visualising relevant training structures

- Coarse level interpretation
- Activity not ascribed to a feature, the user is responsible for identifying the cause from the provided examples
- Useful when combined with other approaches
  - Can provide structural analogues

Identifying the behaviour of atoms and fragments

- Aim to assign contribution to a given set of atoms and bonds or a fragment
- Some approaches define the contribution of an entity as the difference between the prediction with the entity and the prediction without

$$R_i = f\big(x(f_i = 1)\big) - f\big(x(f_i = 0)\big)$$

Franke *et al*. feature importance, where *x* is a fingerprint with the presence ($f_i = 1$) or absence ($f_i = 0$) of feature f_i.

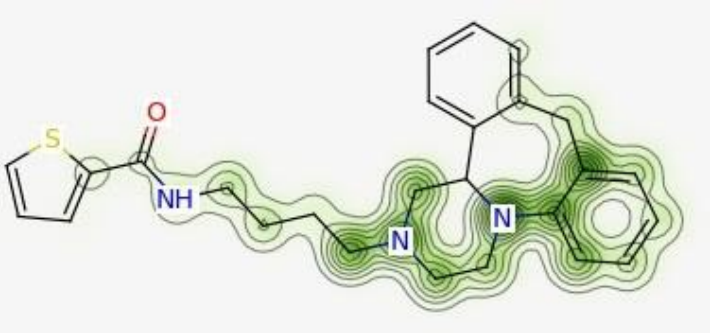
f_i = 3 point pharmacophore
Normalise within a given query

L. Franke, E. Byvatov, O. Werz, D. Steinhilber, P. Schneider, and G. Schneider, "Extraction and visualization of potential pharmacophore points using support vector machines: application to ligand-based virtual screening for COX-2 inhibitors.," J. Med. Chem., vol. 48, no. 22, pp. 6997–7004, Nov. 2005.

Identifying the behaviour of atoms and fragments

- Riniker and Landrum have used a similar approach:
  - The contribution of an atom is the difference between the full fingerprint probability for the active class and the probability of the active class with the contribution of the atom removed
  - http://www.jcheminf.com/content/5/1/43
  - Available in RDKit

```
this_fp = calculate_fingerprints(this_mol)
weights = []
orig_proba = predict_model_probabilities(this_fp)
for atom in this_mol.get_atoms():
            new_fp = calculate_fingerprint_without_atom(this_mol, atom)
            new_proba = predict_model_probability(new_fp)
            weight = orig_proba - new_proba
            weights.append(weight)
```



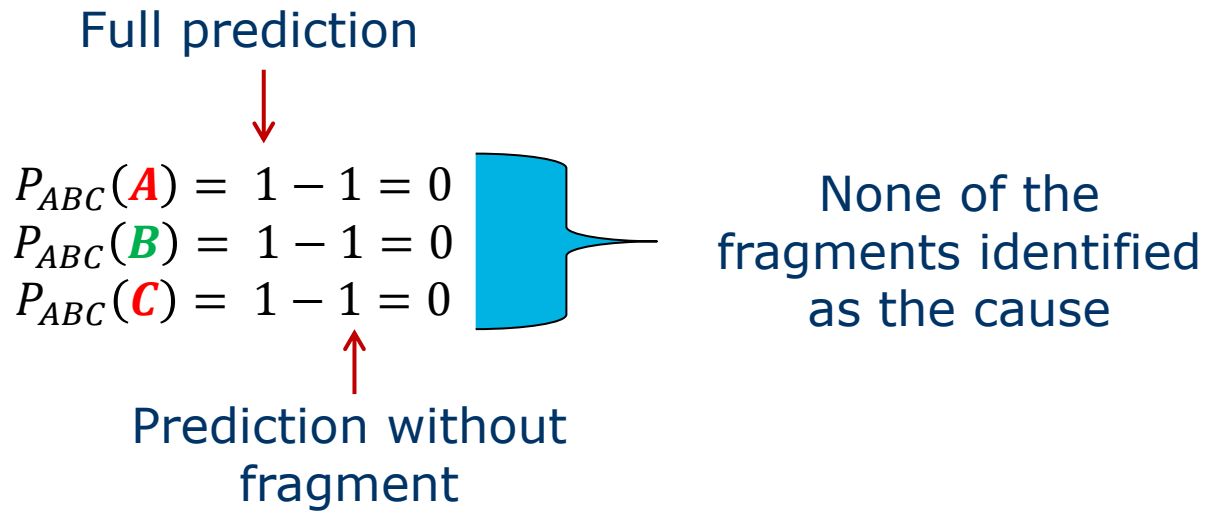
- This interpretation is investigating an atoms contribution to the active class probability

Identifying the behaviour of atoms and fragments

- Polishchuk *et al*. have developed a similar methodology using fragments instead of atoms

$$P_{AB}(B) = Ppred(AB) - Ppred(A)$$

- For binary endpoints a feature can be activating, no contribution or deactivating based on **crossing the class boundary**
- Multiple causes of activity are a problem



P. G. Polishchuk, V. E. Kuz'min, A. G. Artemenko, and E. N. Muratov, "Universal Approach for Structural Interpretation of QSAR/QSPR Models," Mol. Inform., vol. 32, no. 9–10, pp. 843–853, Oct. 2013.

Why does the model predict class x?

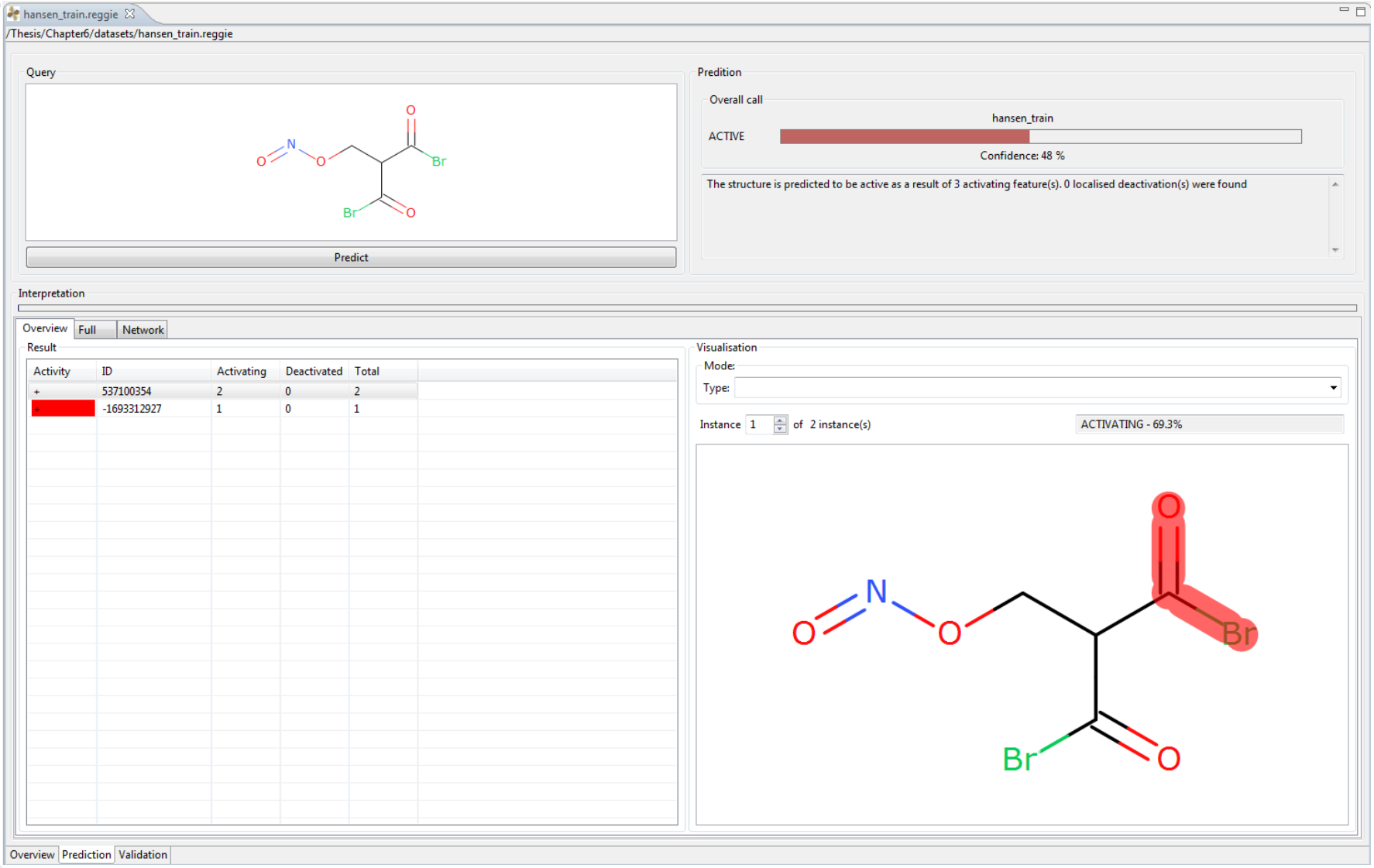
FEATURE COMBINATIONS NETWORKS

The goal

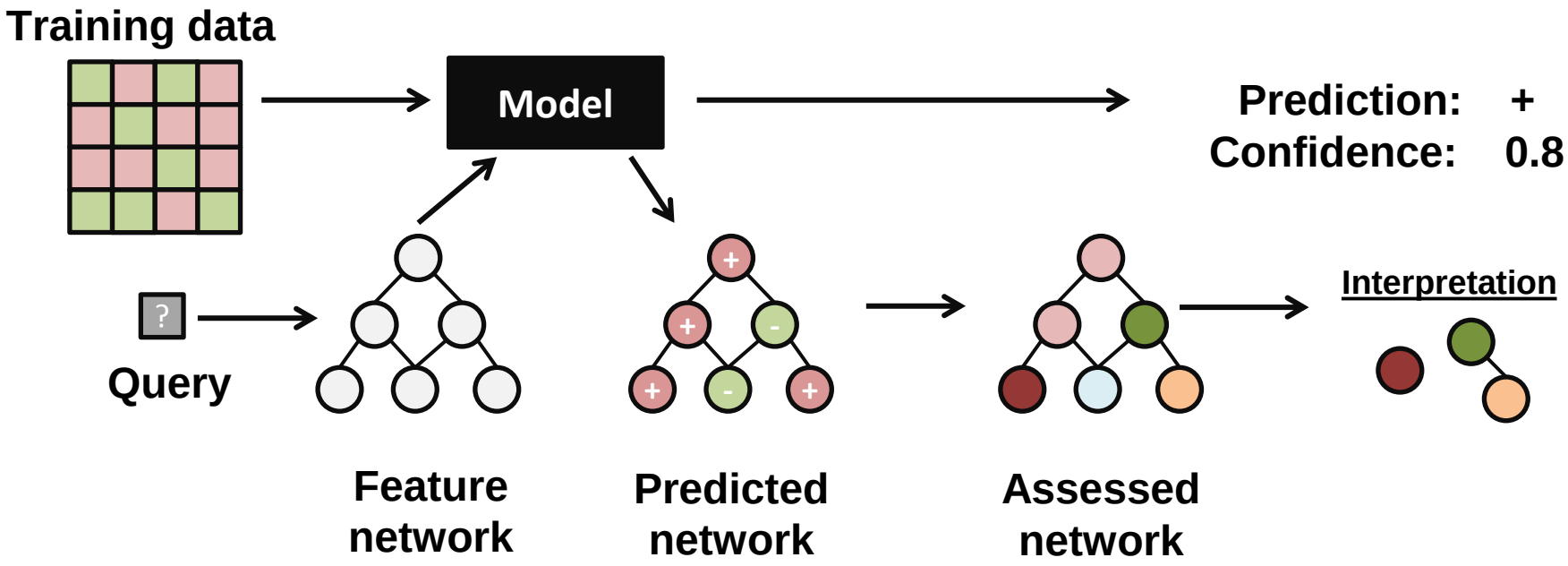
- We want to be able to identify why any model predicts the given class in a binary classification problem.
  - Not specific to any particular learning algorithm

- We want this to be meaningful in the context of chemical structures

Where we are heading…



How do we get there?



Endpoint definition

1) Activity is caused by the presence of a feature (structural or physicochemical)
2) Inactivity can be described by either:
    1) The lack of an activating feature
    2) The deactivation of all activating features

Descriptor choice

- Descriptors must be represented as a binary fingerprint
- Continuous values (such as logKp) are incorporated through the use of a discretisation algorithm

Continuous variable
0.0
5.0
6.5
...

Discretisation →

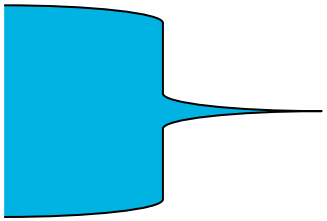
Discretised variable
(-inf, 1.0]
(2.4, 5.1]
(5.1, 7.8]
...

Discretisation fingerprint →

(-inf, 1.0]	(2.4, 5.1]	(5.1, 7.8]	...
1	0	0	...
0	1	0	...
0	0	1	...

Features

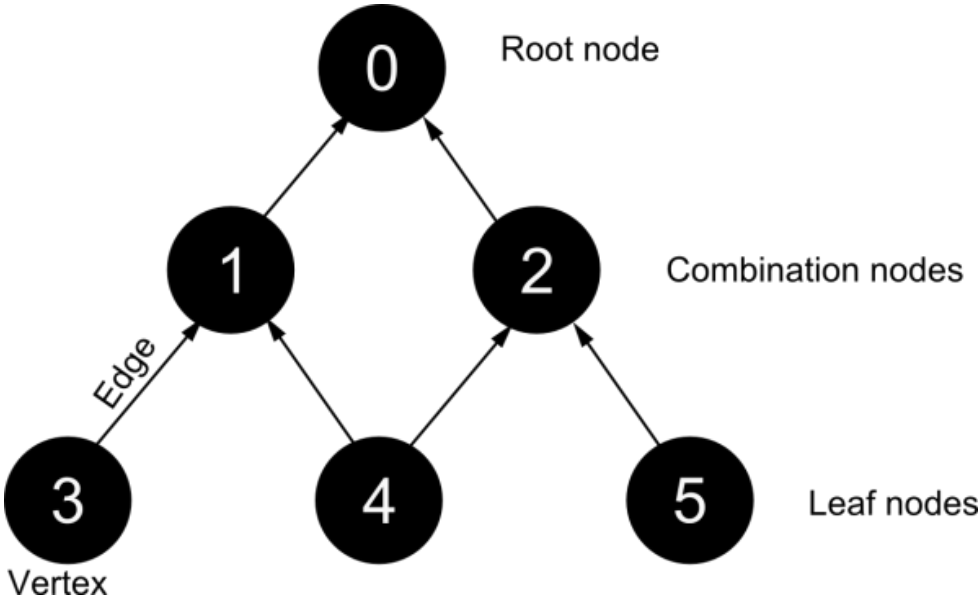
- Like with the approaches discussed earlier an entity is removed and a prediction is made on the altered descriptor set
- In this approach combinations of the bits in the fingerprint are generated
- A feature contains:
  - Identifier
  - Descriptor subset
  - Fragment
  - Atom list
  - Bond list



When using fragment features

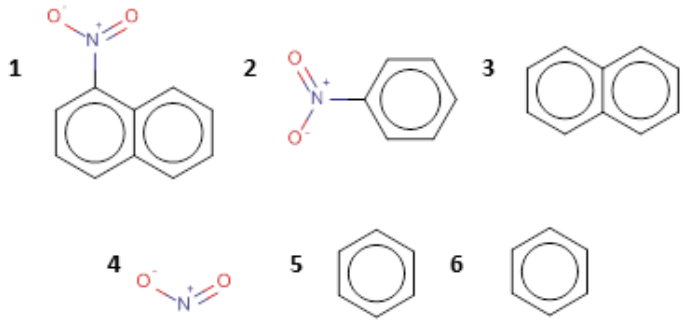
Feature organisation

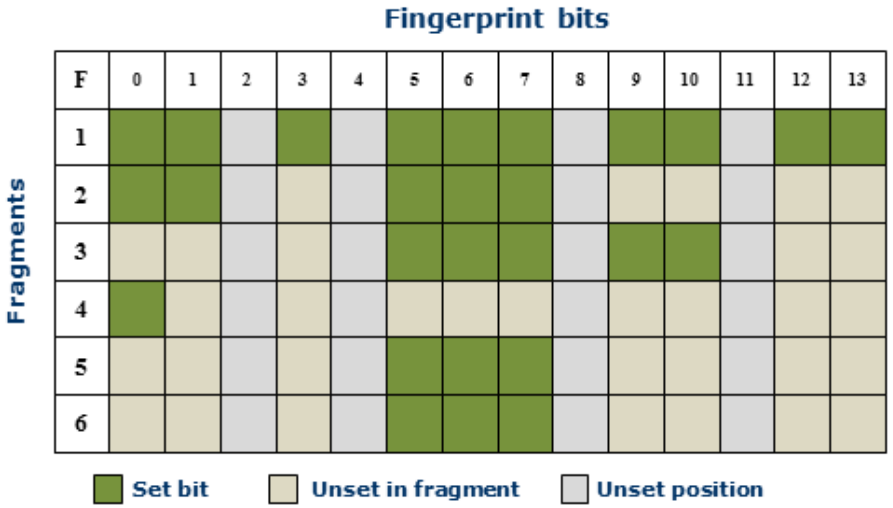
- Features are organised into hierarchies
- A parent represents the union of its children
- The root node represents the feature describing the full query
- The leaf nodes represent the smallest features



Fragment networks

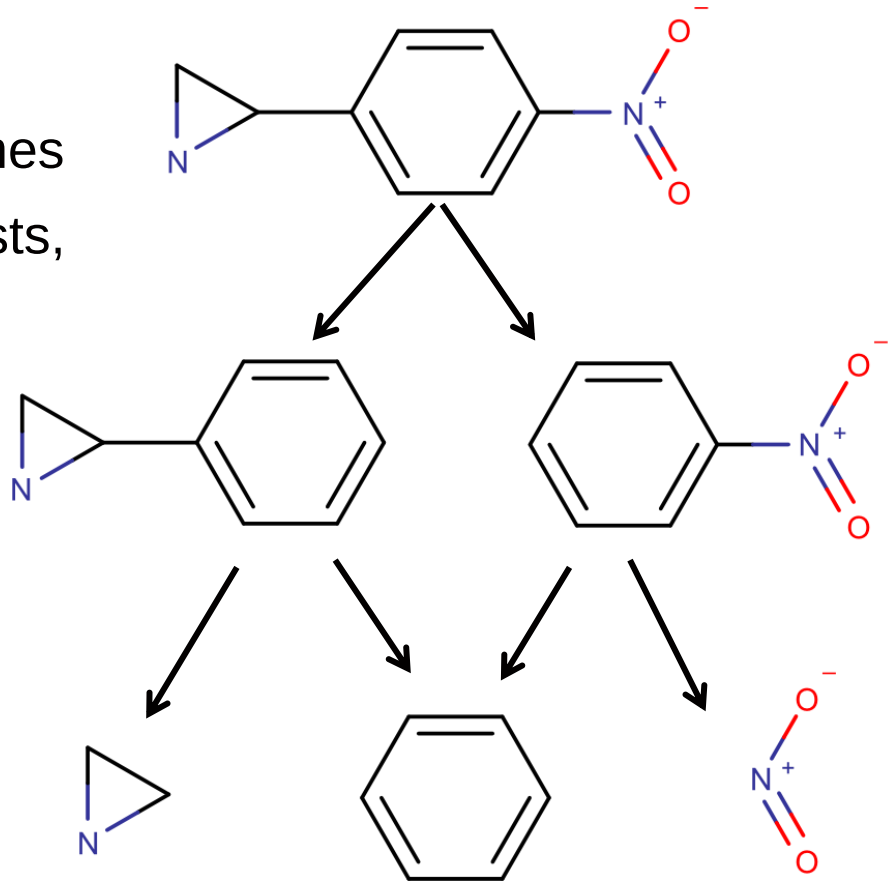
- We can use a fragmentation algorithm to map between a structural fingerprint and a set of atom and bonds
- The fragment can be organised based on the atoms and bonds they represent on the query
- We can generate the descriptors from the fragments:



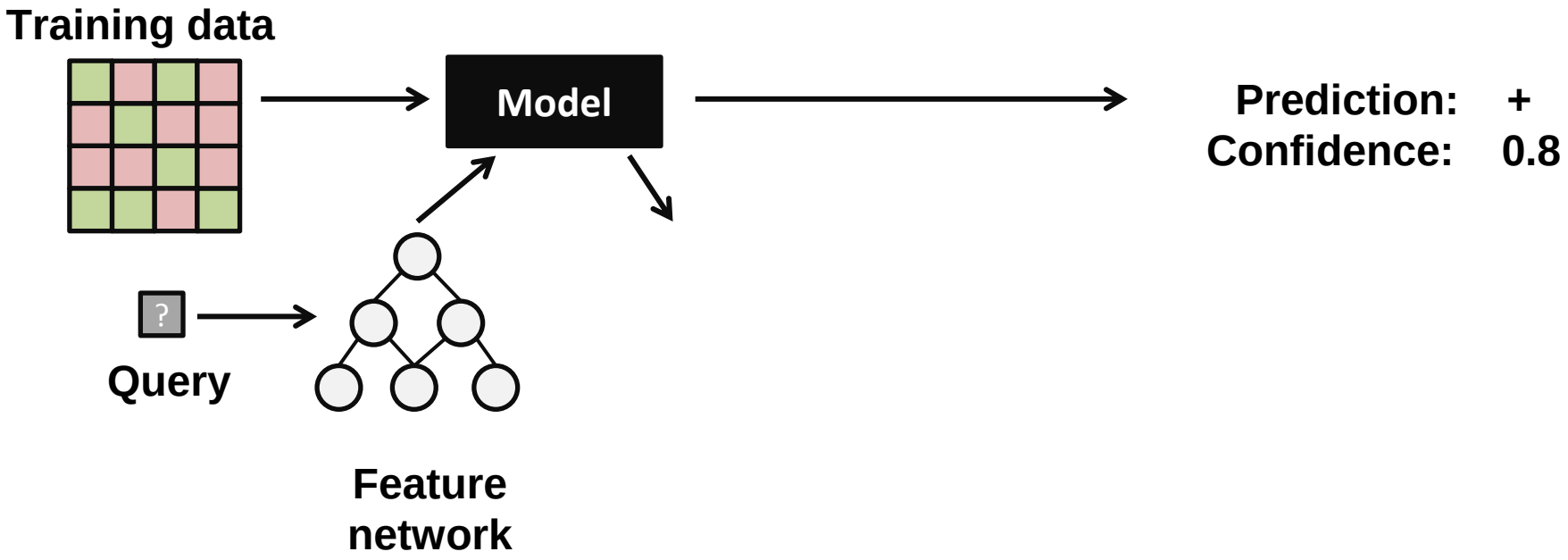


Fragment networks

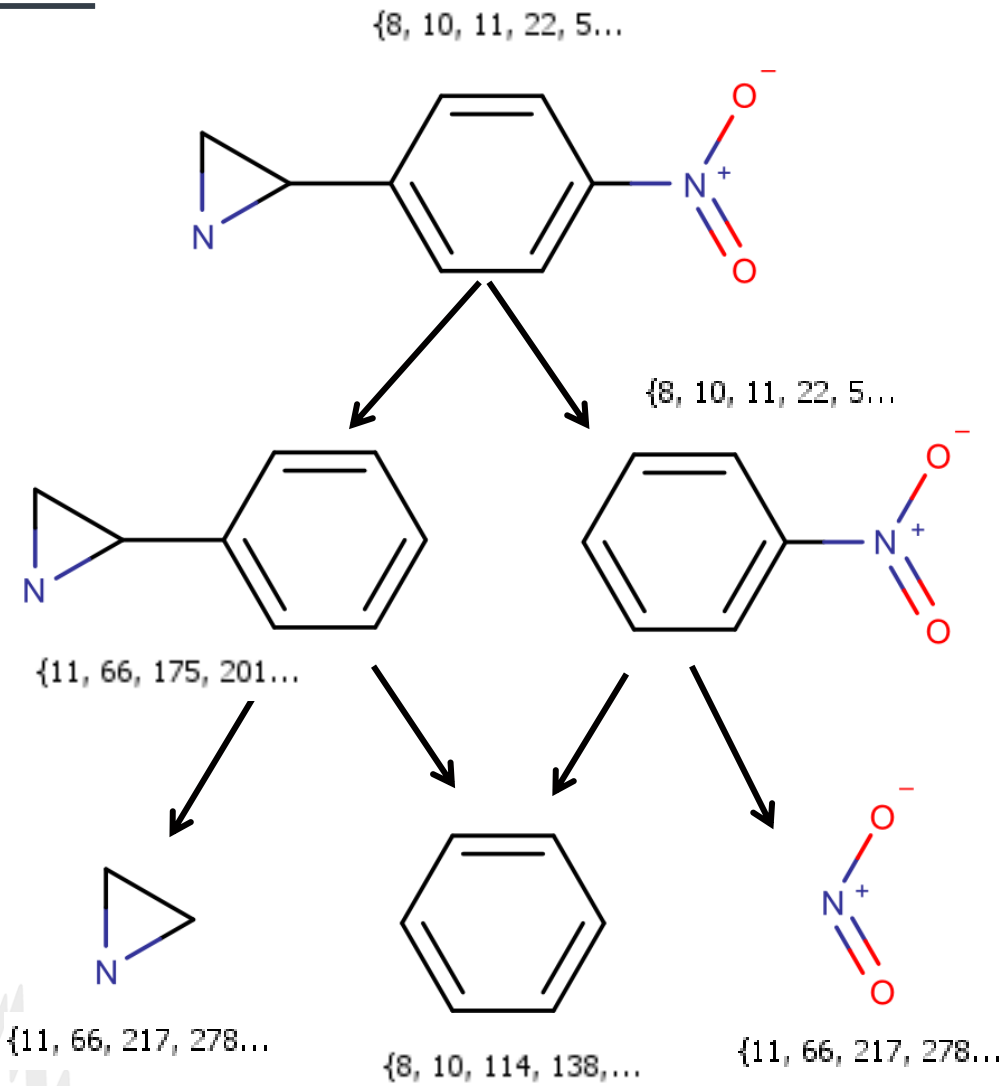
- Organise based on atoms and bonds
- Original query structure at the top
- Fragments may occur multiple times while varying in atom and bond lists, these are considered to be independent features



Network generation: descriptors and prediction



Calculate descriptors

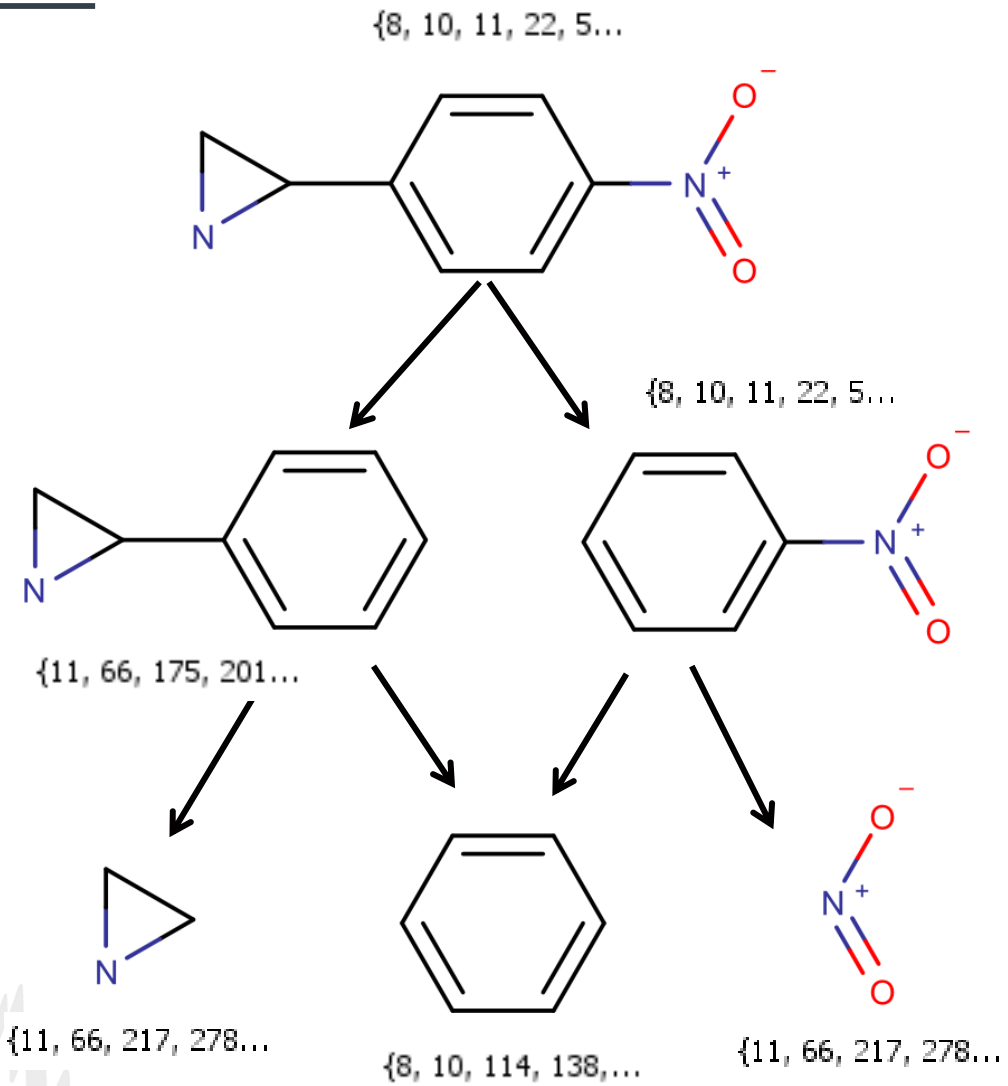


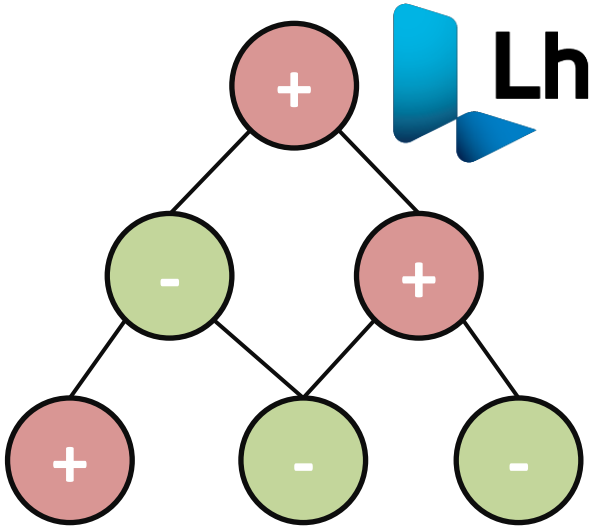
The descriptors are calculated from the fragments.

A work around may be required to handle novel bits to the fragment.

Structural fingerprints where the contributions of specific atoms can be accounted for work best.

Get predictions



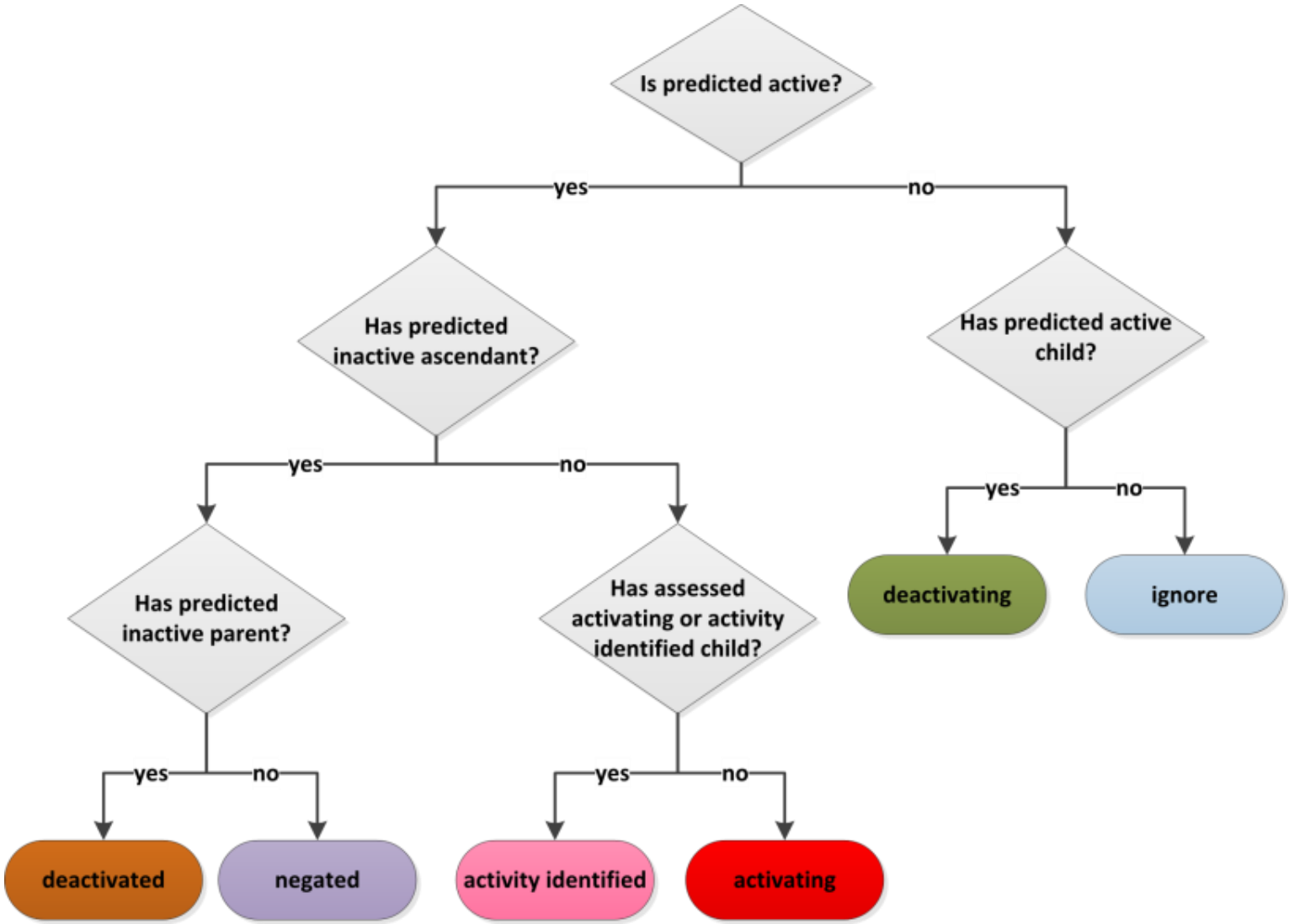


Predicted network

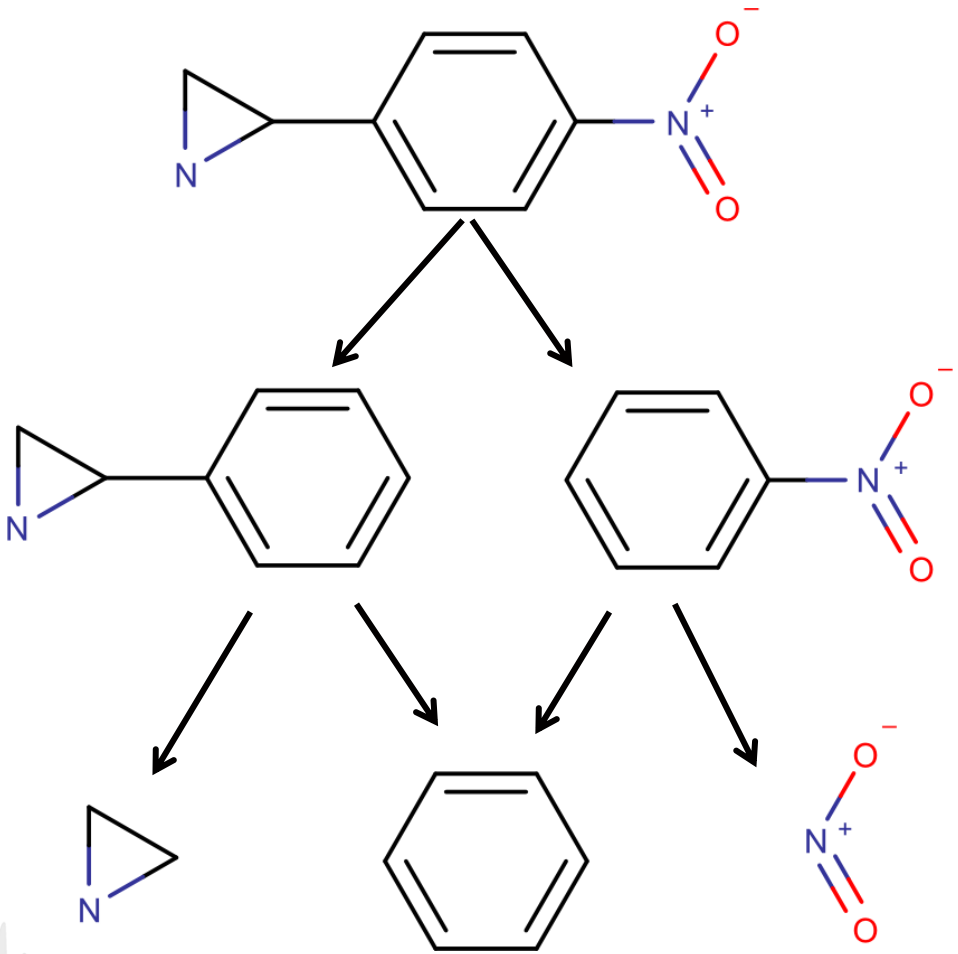
A prediction is made for each descriptor vector and associated with the node.

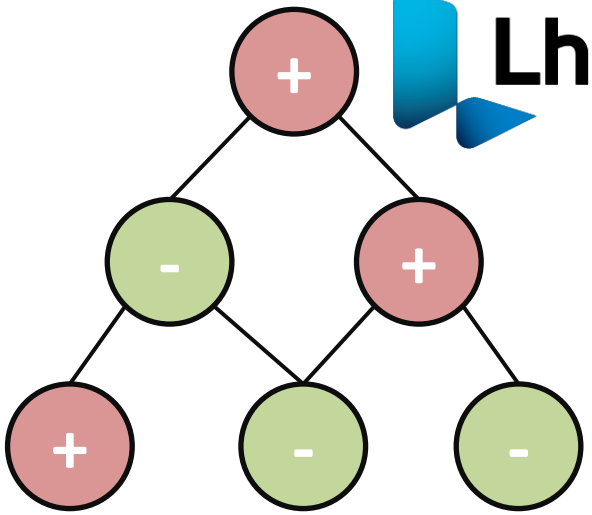
In the above image green is an inactive prediction and red an active prediction.

Network generation: assessment rules

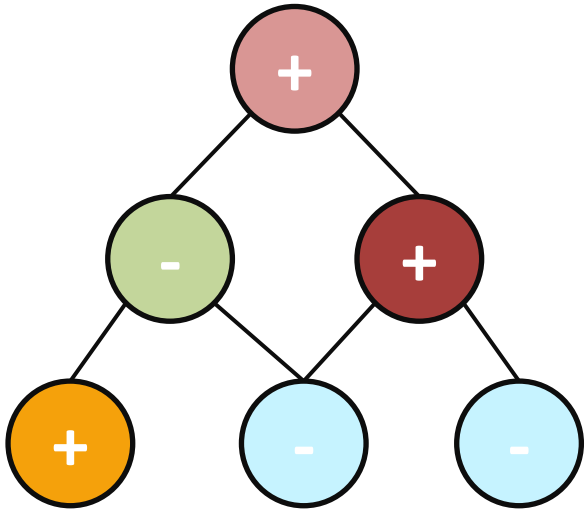


Assess the network



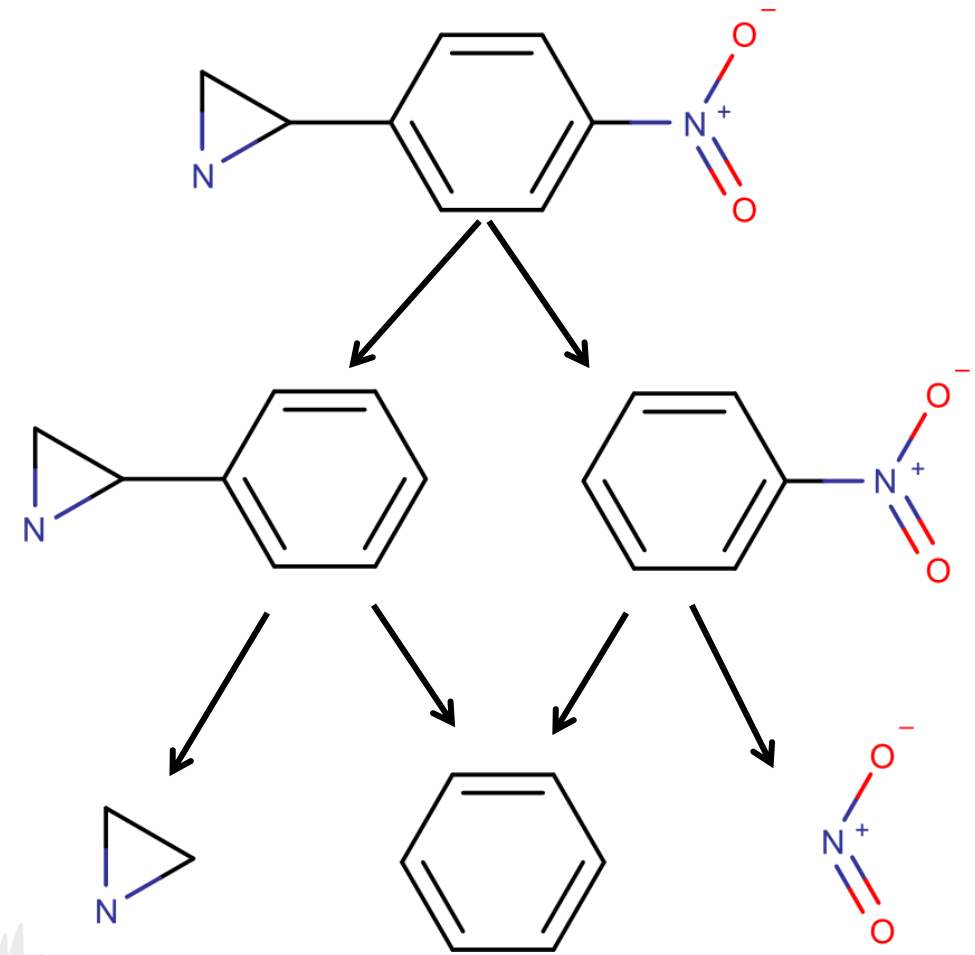


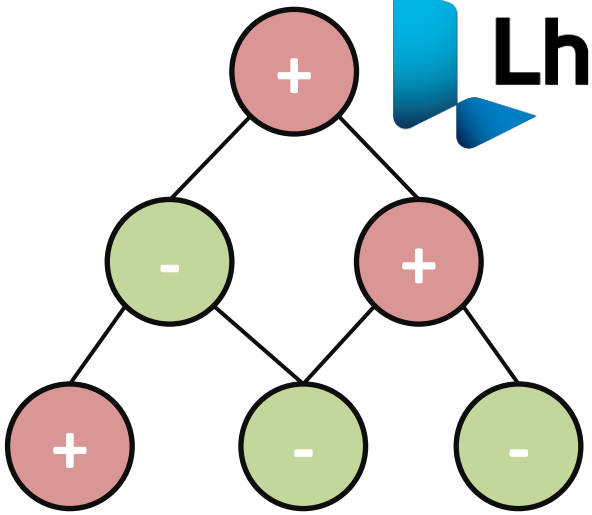
Predicted network



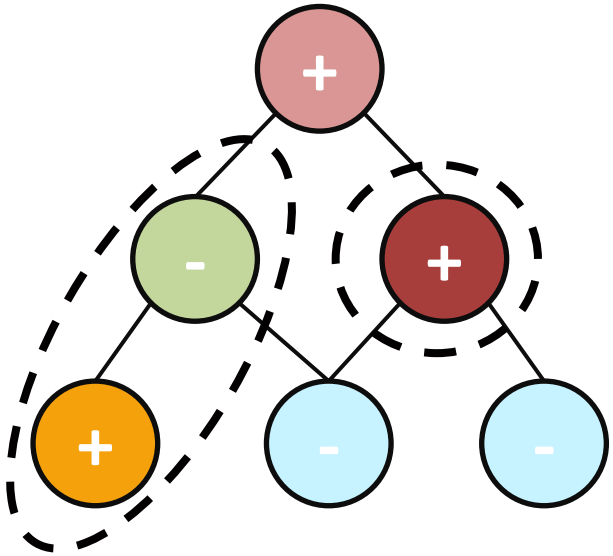
Assessed network

Summarise the network





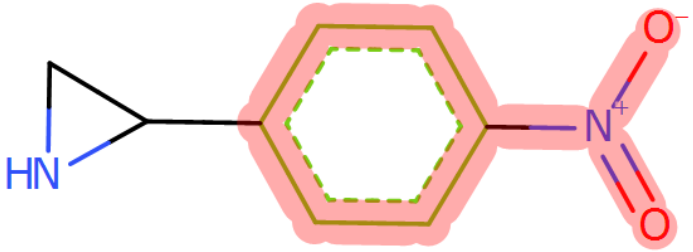
Predicted network



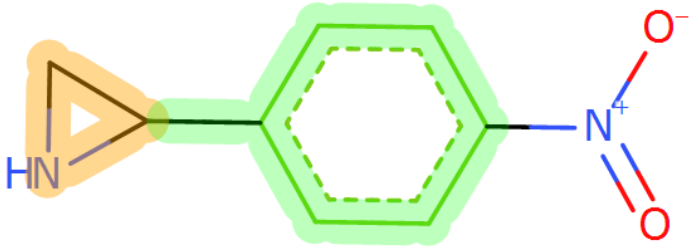
Assessed network

Project the summary

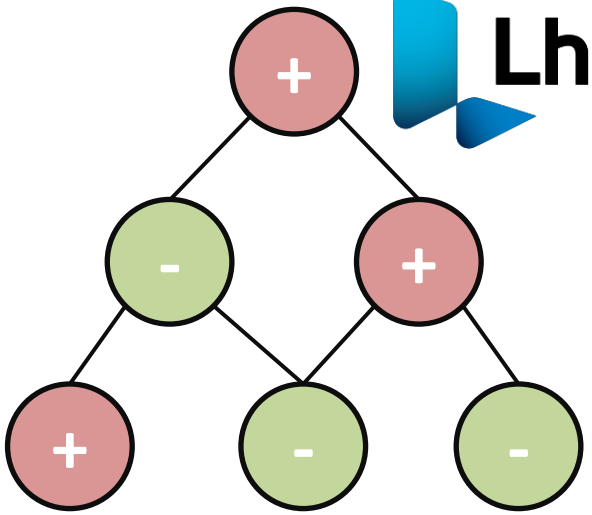
Activating nitrobenzene motif



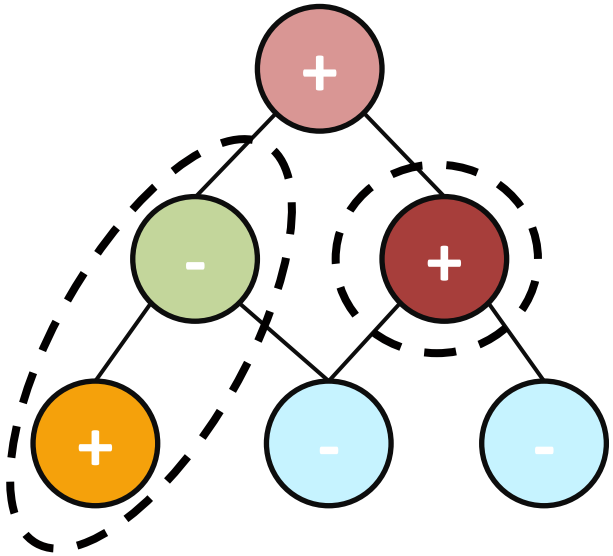
Aziridine motif deactivated by benzene ring attachment



This structure is **predicted to be active** due to the presence of the nitrobenzene fragment motif

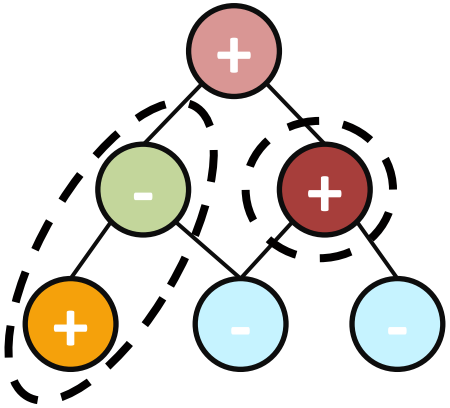


Predicted network



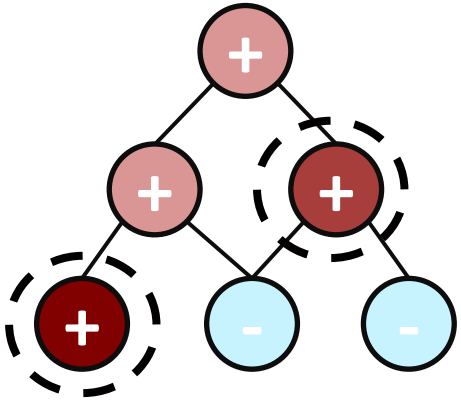
Assessed network

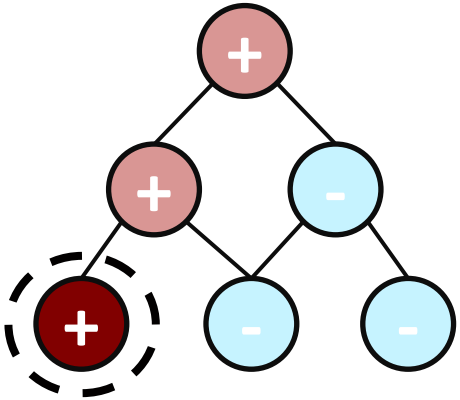
Summaries



- Activity due to an activating feature
- Localised deactivation found

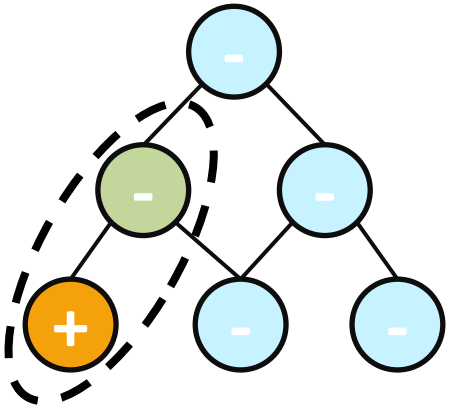
- Activity due to two independent features



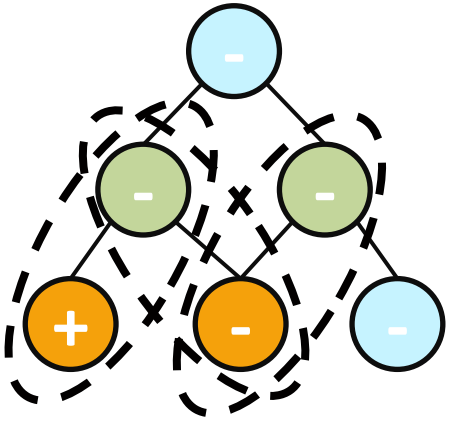


- Activity due to a single activating feature

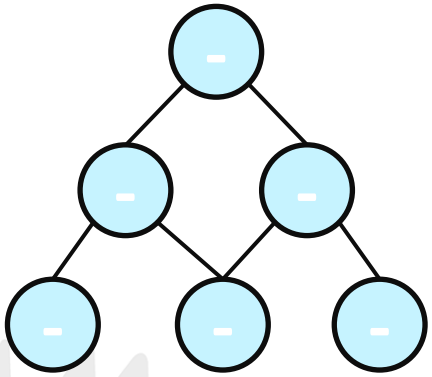
Summaries



- Inactivity due to deactivation of single feature



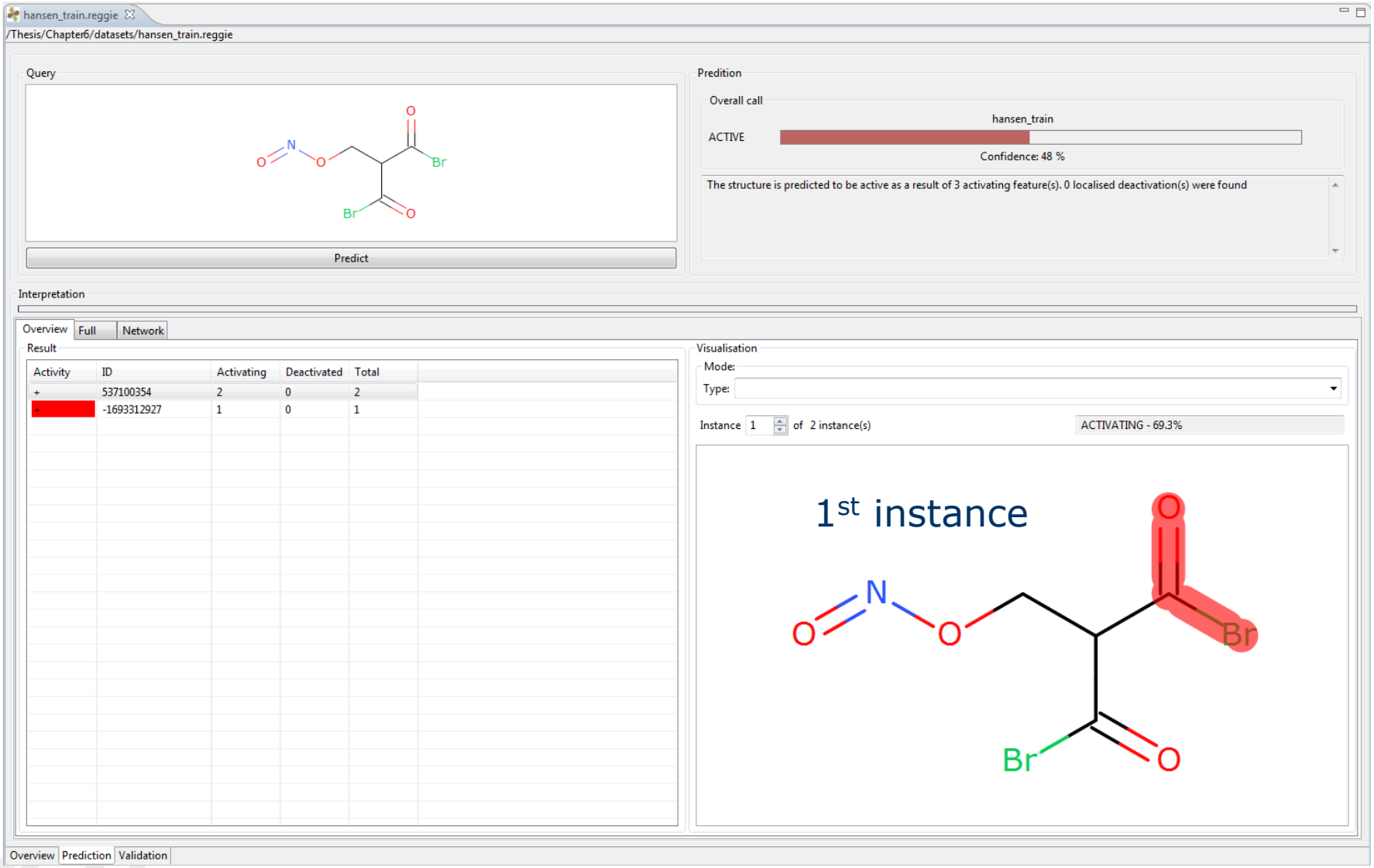
- Inactivity due to deactivation of single multiple features (multiple deactivating features)



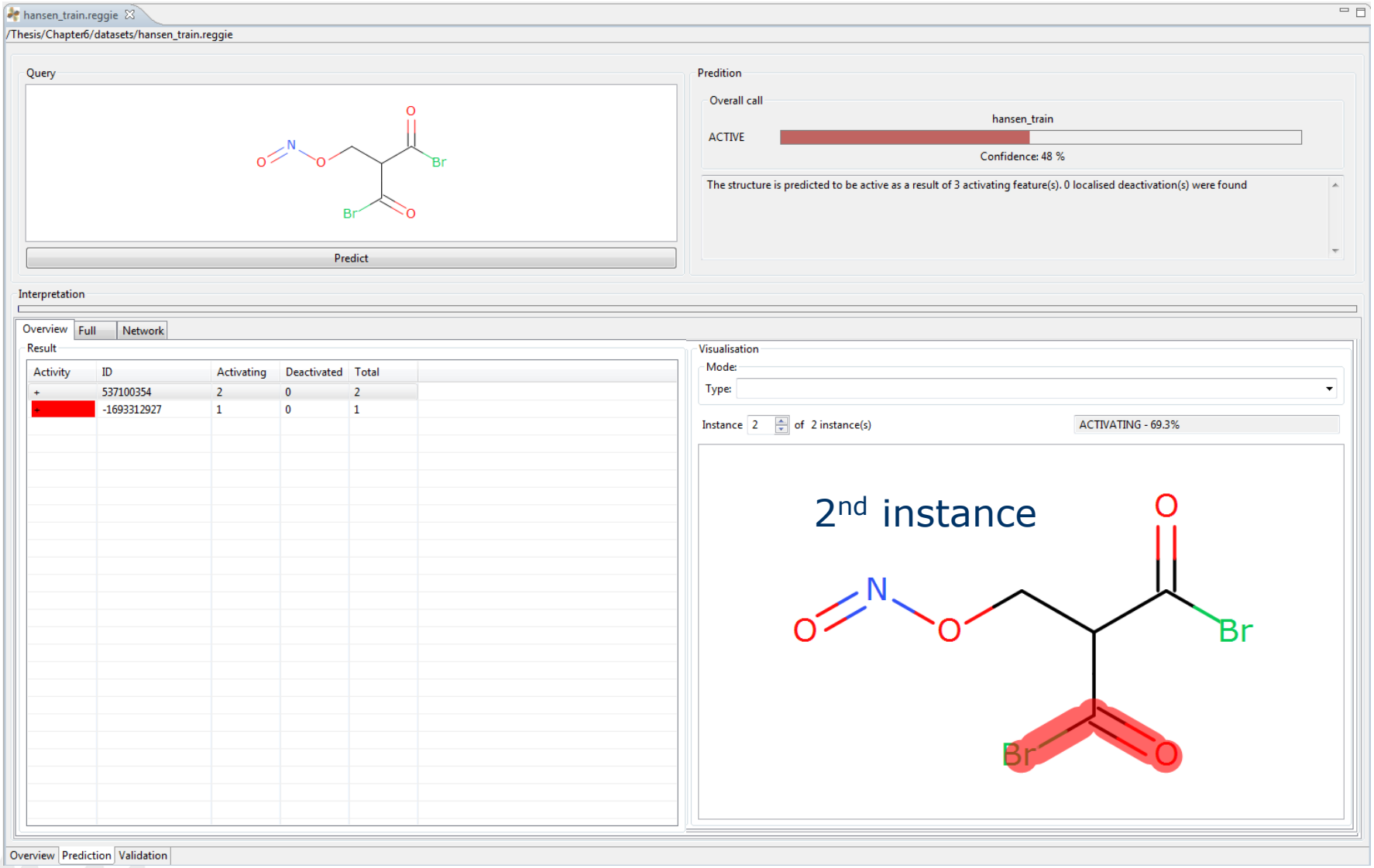
- Inactivity due to lack of an activating feature

IMPLEMENTATION

Eclipse RCP implementation



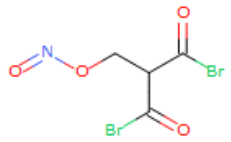
Eclipse RCP implementation



Eclipse RCP implementation

/Thesis/Chapter6/datasets/hansen_train.reggie

Query



Predict

Predition

Overall call

hansen_train

ACTIVE

Confidence: 48 %

The structure is predicted to be active as a result of 3 activating feature(s). 0 localised

Interpretation

Overview | Full | Network

Result

Activity	ID	Activating	Deactivated	Total
+	537100354	2	0	2
+	-1693312927	1	0	1

Visualisation

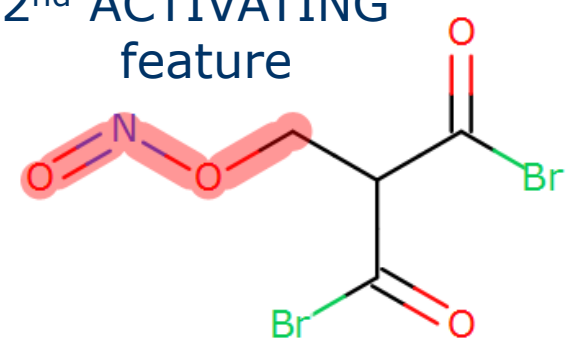
Mode:

Type: EXPLICIT_DEACTIVATION

Instance 1 of 1 instance(s)

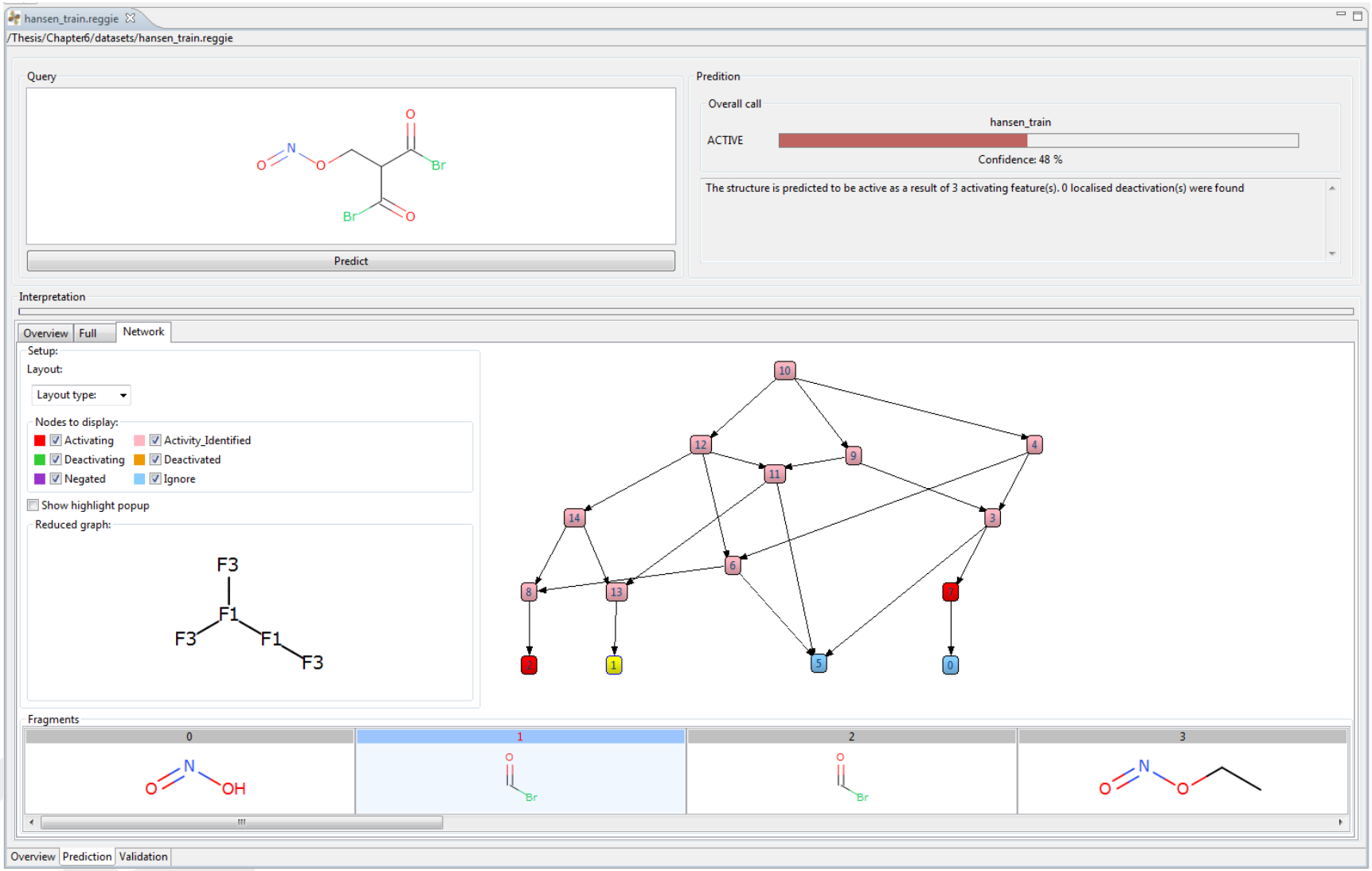
ACTIVATING - 36%

2nd ACTIVATING feature

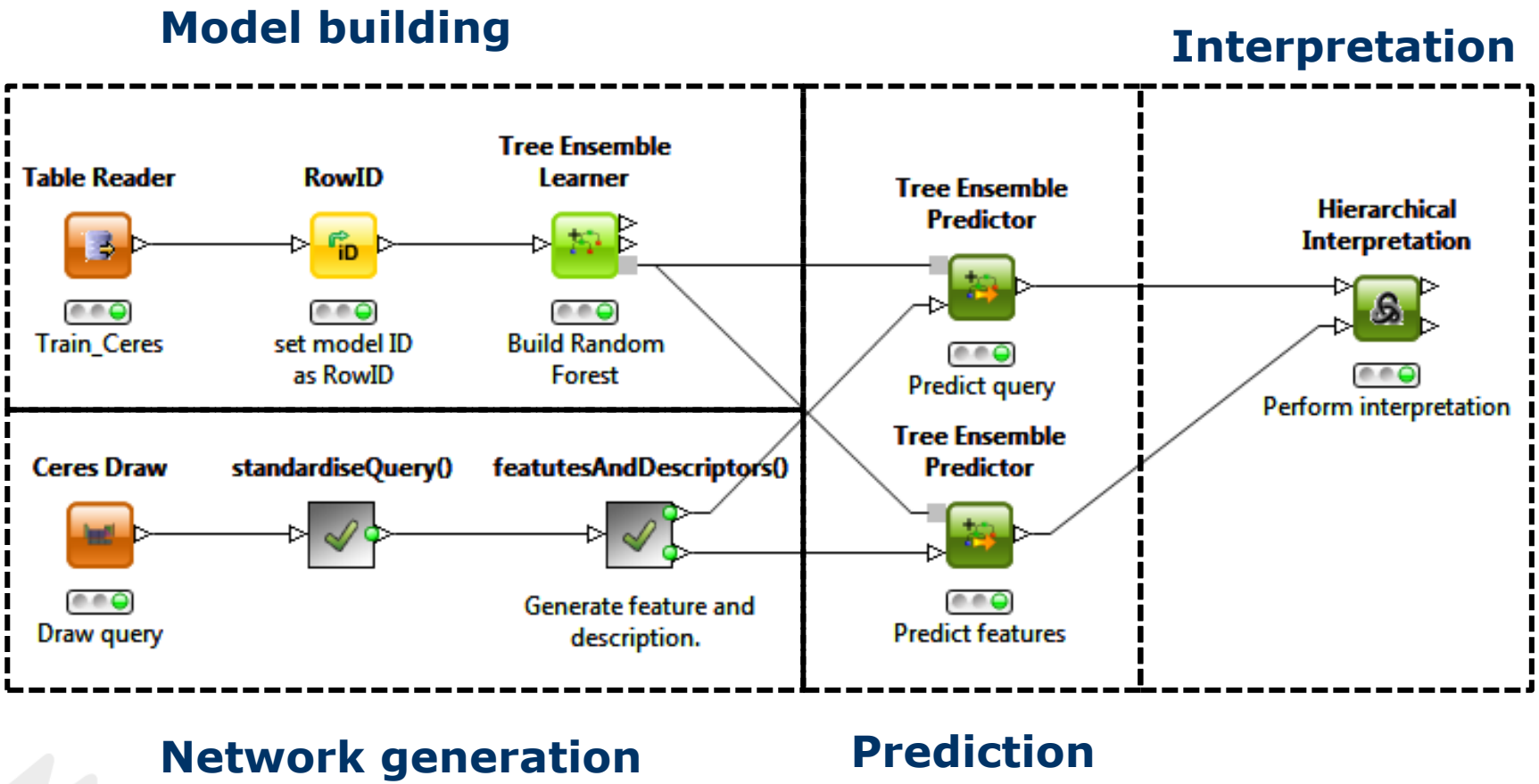


Overview | Prediction | Validation

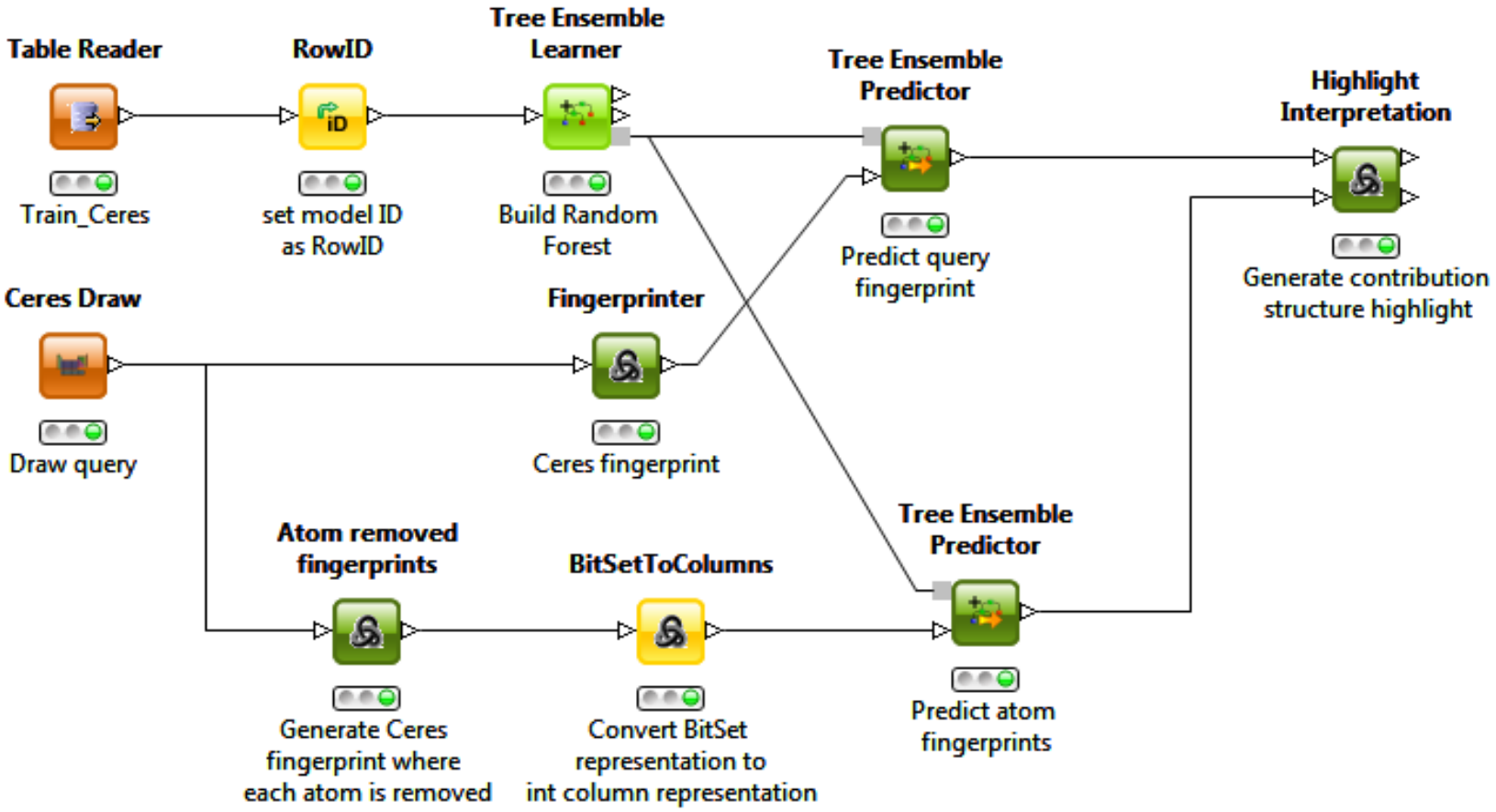
Eclipse RCP implementation



KNIME implementation



Similarity maps approach – KNIME implementation



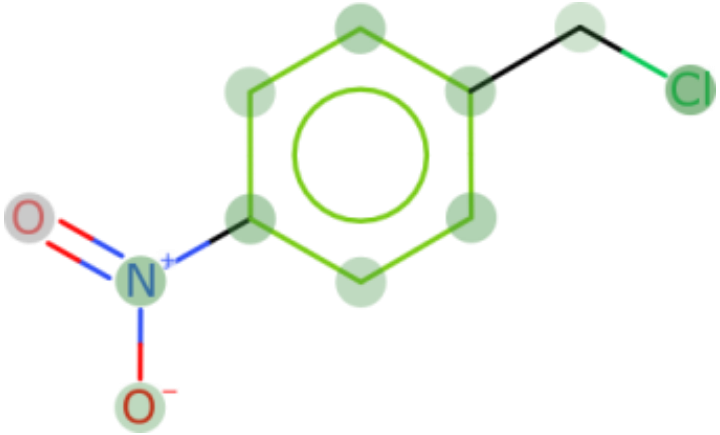
EXAMPLES

Models

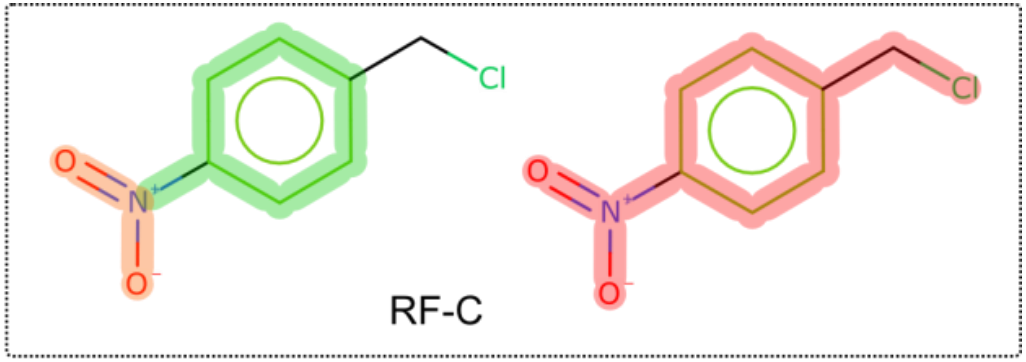
- Mutagenicity
  - Curated Hansen dataset
  - Split into train – test – validation
  - Examples shown from validation set
  - SVM, RF, DT, kNN models evaluated
  - Hashed fingerprint model shown

- Skin irritation
  - PaDEL GHS hazard codes
  - Highly biased dataset, utilised class weights and learning performed with Weka
  - DT, kNN and RF models evaluated
  - Hashed fingerprint + logKp binary cut-off

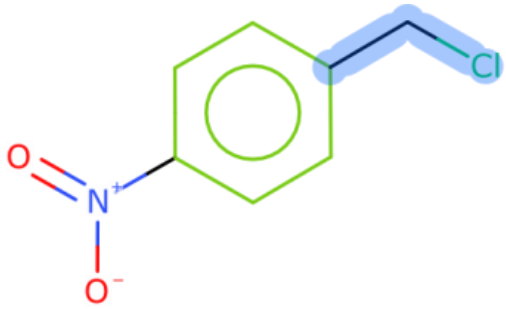
Mutagenicity 1



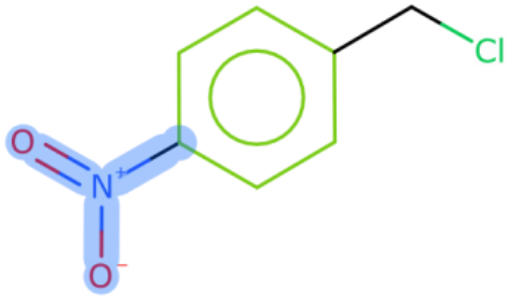
Similarity maps



Feature networks



Alert 027: Alkylating agent

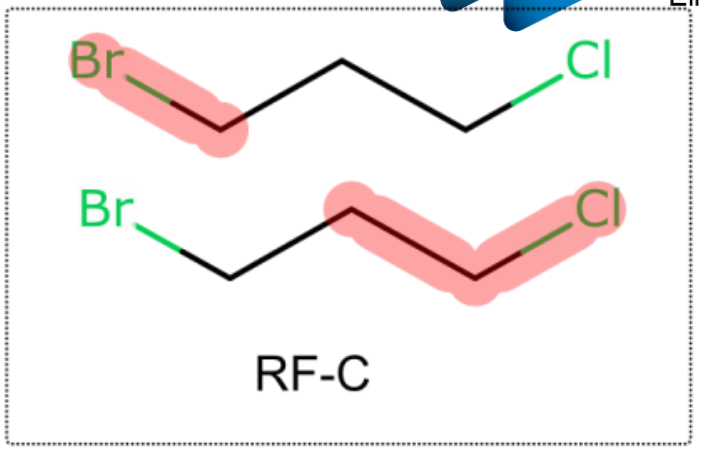


Alert 329: Aromatic nitro compound

Derek Nexus

Mutagenicity 2

Similarity maps



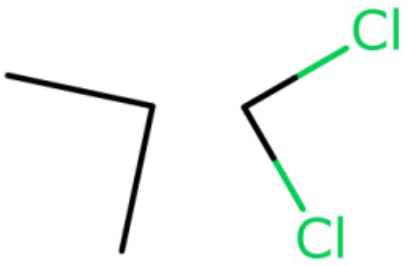
Feature networks

Alert 027: Alkylating agent

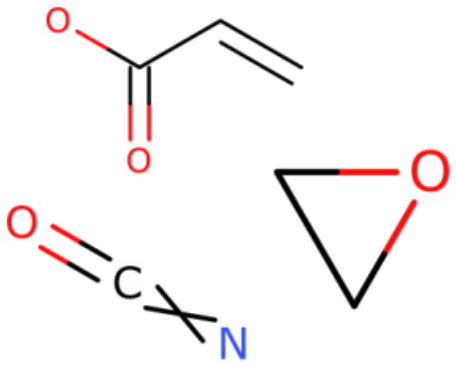
Alert 027: Alkylating agent

Derek Nexus

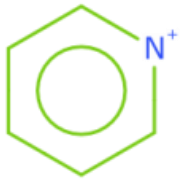
Skin irritation examples



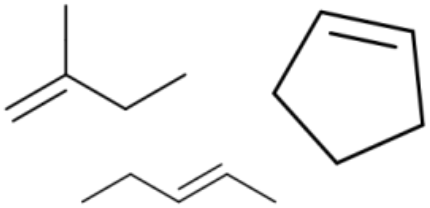
Combination features



Fragment features
(functional group)



Fragment features
(aromatic)

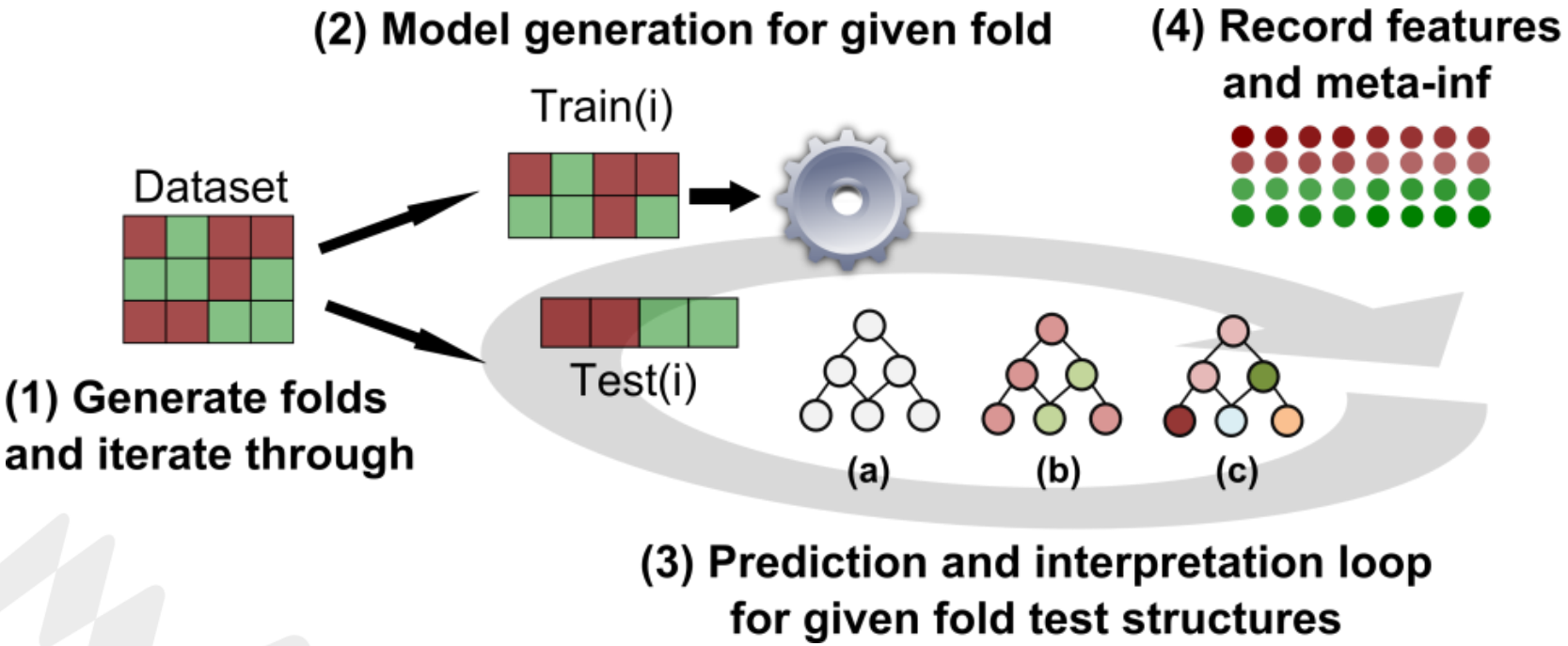


Fragment features
(saturated and unsaturated hydrocarbons)

KNOWLEDGE MINING

Knowledge mining

- The developed algorithm provides the interpretation in such a way that it easily lends itself to use for knowledge mining

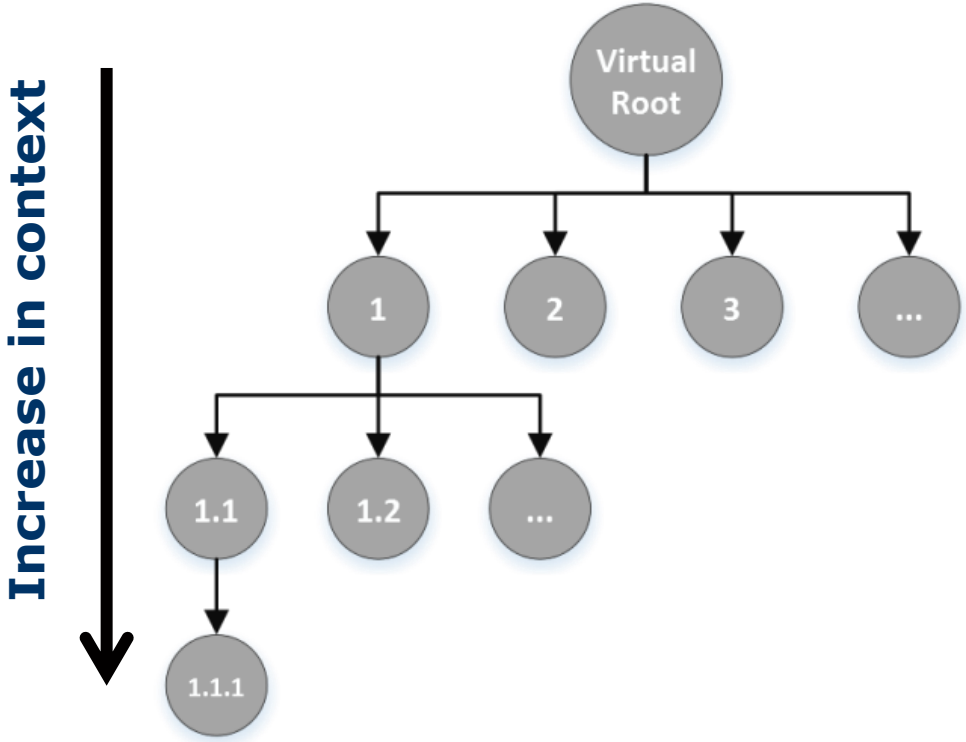


Use cases

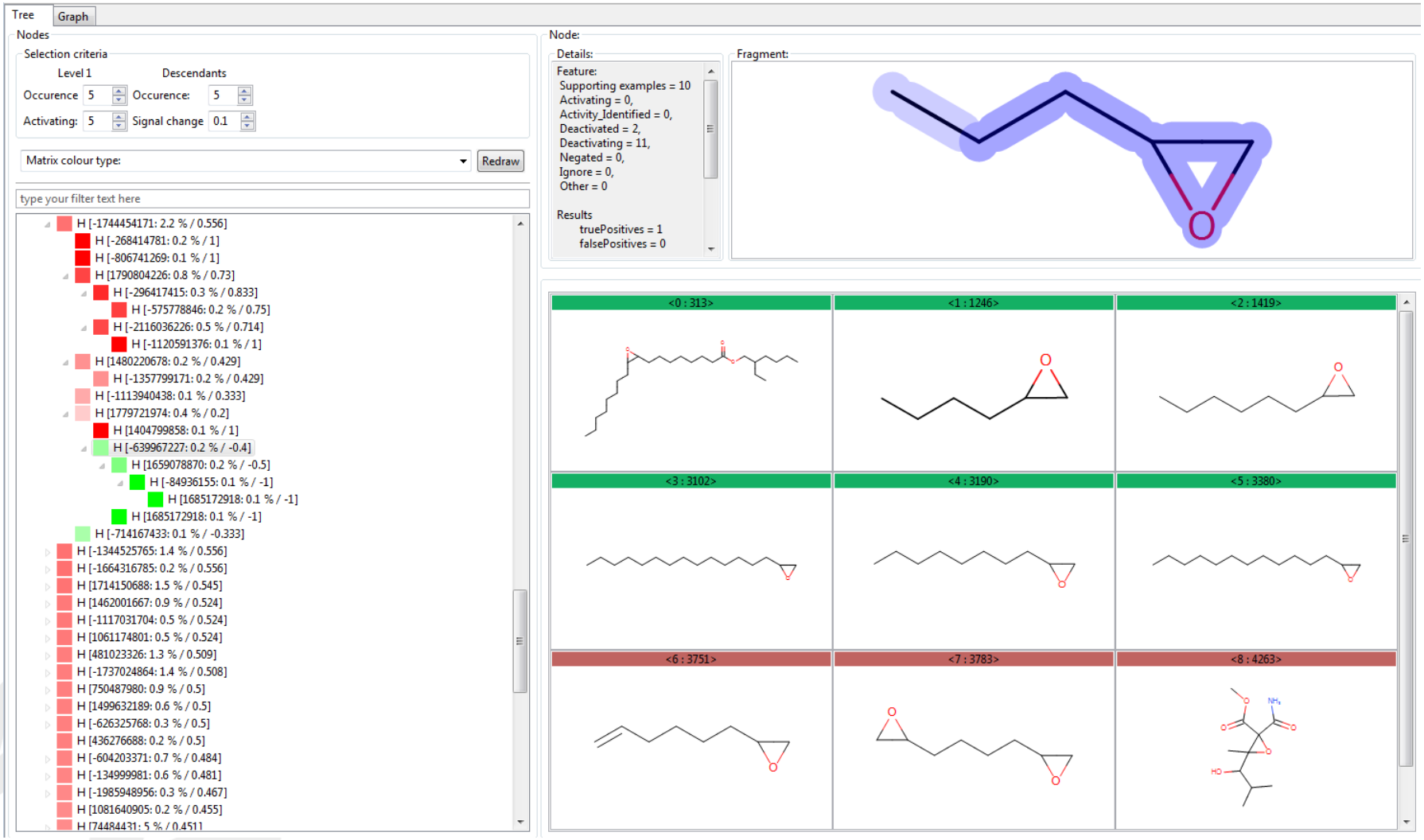
- Support the development of human expert knowledge
- Support the development of expert system knowledge base
  - Improve existing alerts
  - Identify new potential toxicophores

SAR pattern network

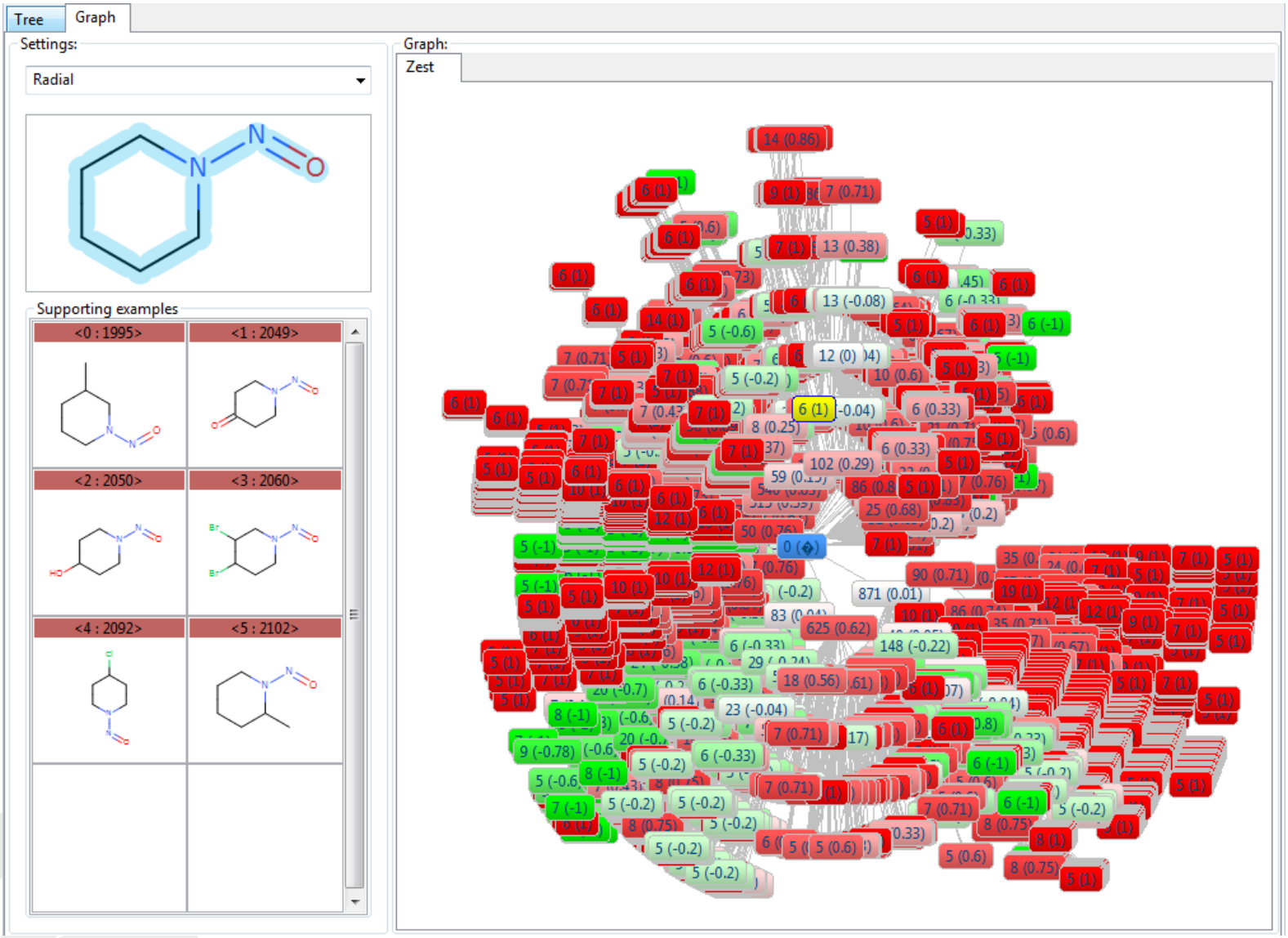
- The dictionary can be represented as a tree
- The tree represents the relationship between features across the whole dataset



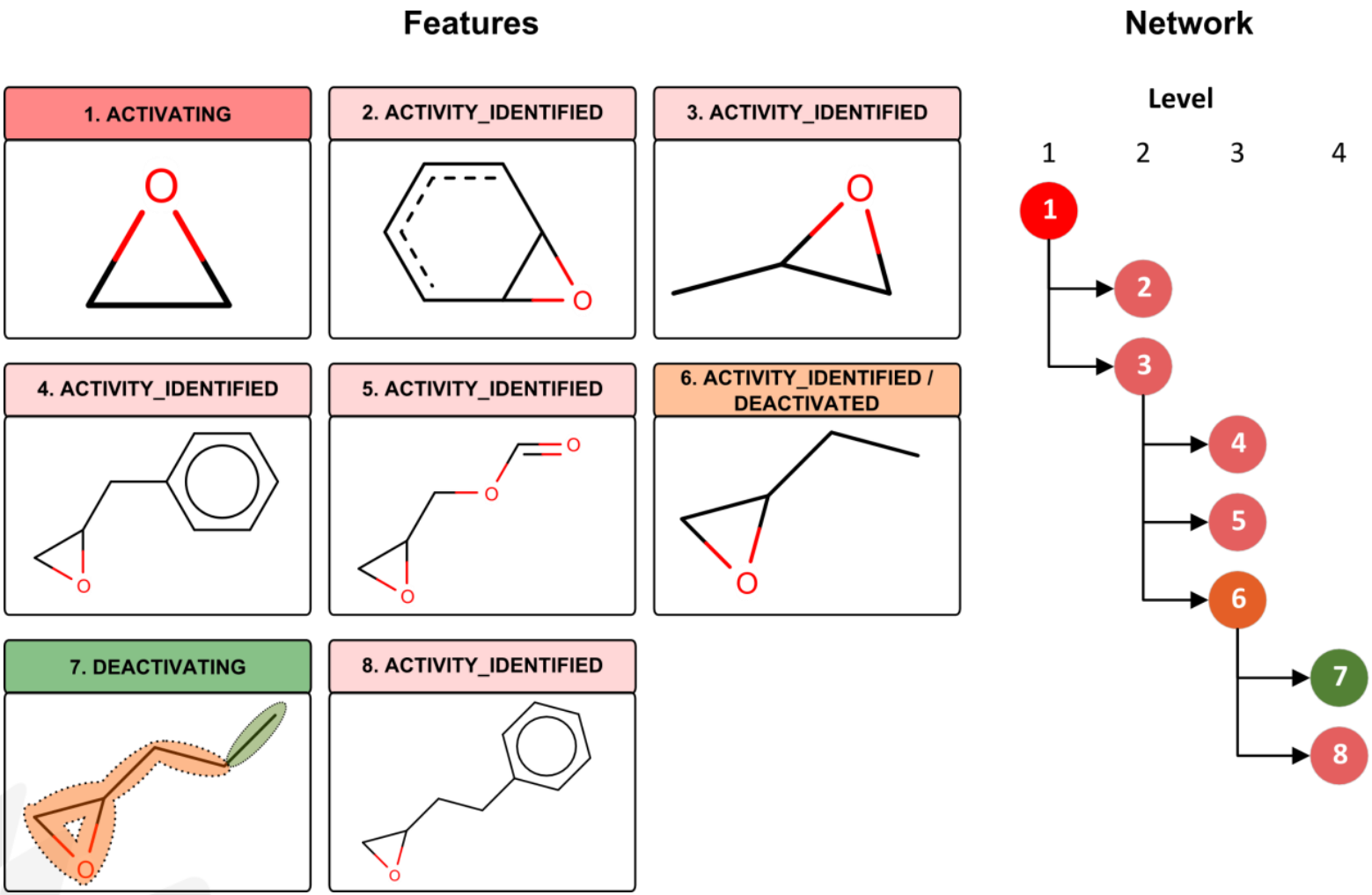
SAR patterns



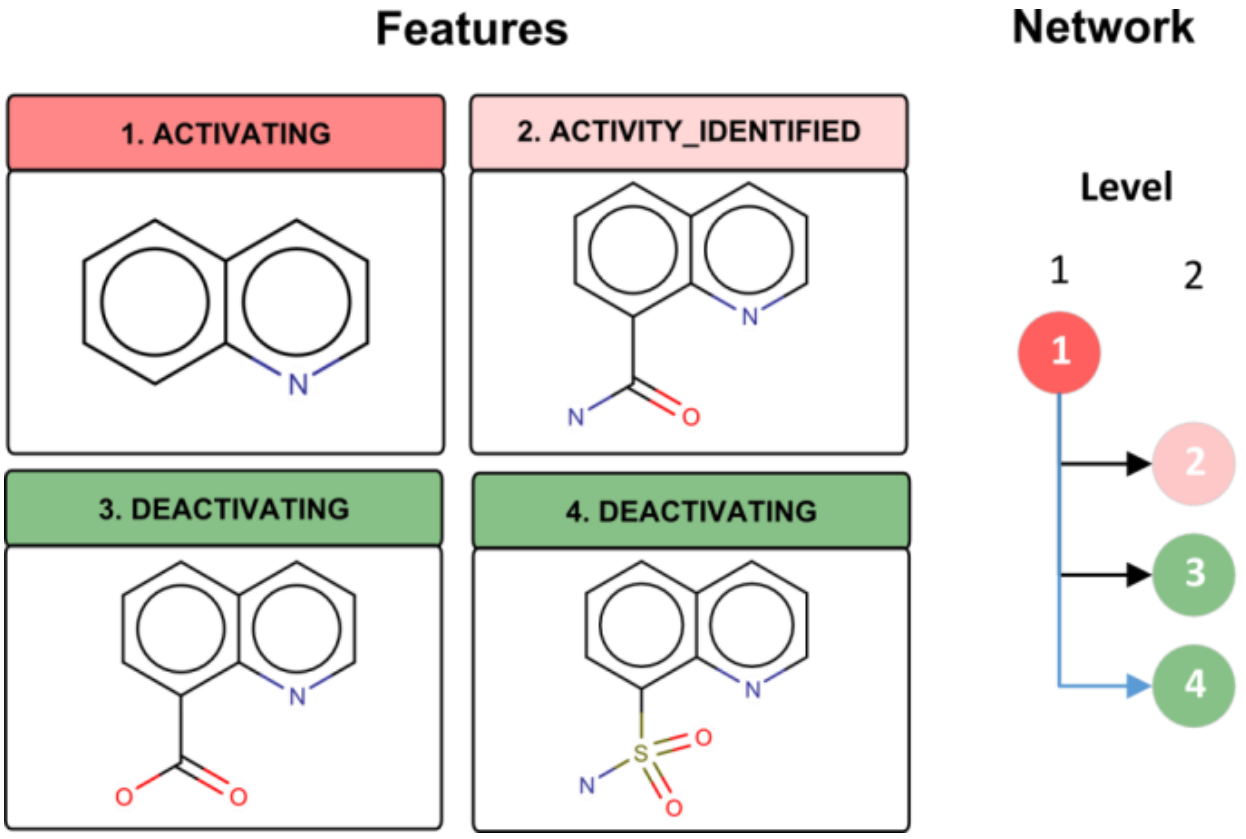
SAR patterns



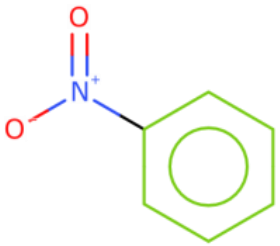
Mutagenic example: epoxide



Mutagenic example: quinoline

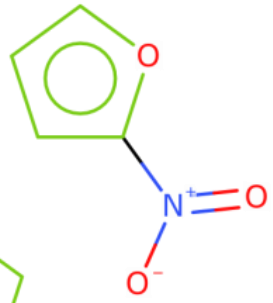


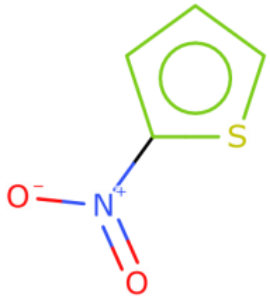
Mutagenic features



nitrobenzene motif

nitrofuran motif



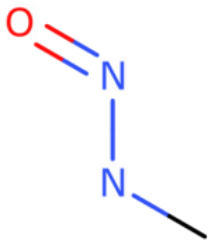


nitrothiophene motif

epoxide motif

anthracene motif

phenanthrene motif



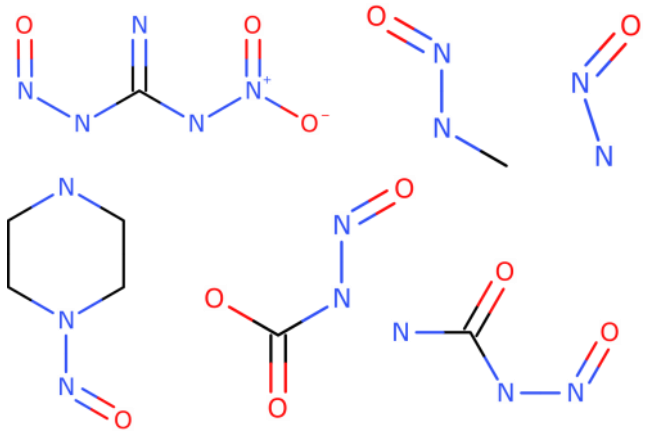
N-nitroso motif

chloro-alkyl motif

bromo-alkyl motif

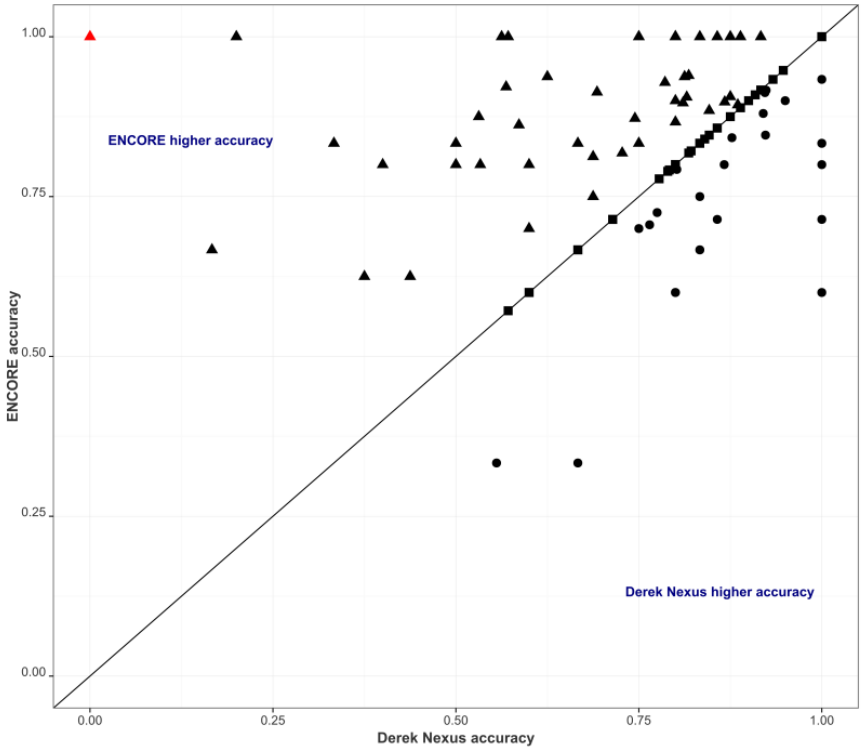
Summary

- Human expertise is required to abstract the information into a structural alert



Features similar to the N-Nitroso Derek Nexus alert

There is scope for comparing against existing alerts



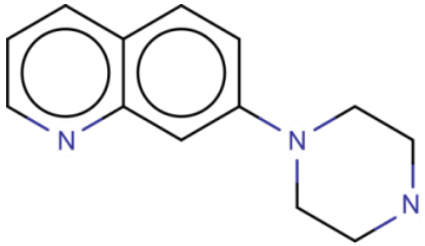
CONCLUSION

Interpretation: solved or unsolved?

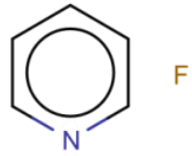
- Multiple interpretation algorithms exist
  - Different levels of granularity
- For some endpoints the feature networks algorithm can remove the accuracy vs interpretation trade-off
- Providing an interpretation can have negative consequences on the trust in the model
  - Right for the wrong reason: potentially rejected model
    - But this may help assess the quality of a model
  - Trend identified differs to experience: potentially rejected model

Knowledge mining

- It is useful to perform knowledge mining activities with more than one approach
  - Variation in cause
  - Variation in support set



Feature



Emerging pattern

- The method has been able to identify existing knowledge and identify potential new structural alerts for:
  - Mutagenicity
  - Skin sensitisation
  - Skin irritation

Top
Abstract
Background
Algorithm
Experimental ...
Results and discussion
Conclusion
Abbreviations
Competing interests
Authors' contributions
Acknowledgments
References

This article is part of the series
6th Joint Sheffield Conference on Chemoinformatics
.

Research article

Highly accessed Open Access

Feature combination networks for the interpretation of statistical machine learning models: application to Ames mutagenicity

Samuel J Webb[12*], **Thierry Hanser**[1], **Brendan Howlin**[2], **Paul Krause**[2] and **Jonathan D Vessey**[1]

* Corresponding author: Samuel J Webb samuel.webb@lhasalimited.org ▾ Author Affiliations

1 Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Holbeck, Leeds LS11 5PY UK

2 University of Surrey, Guildford, Surrey GU2 7XH, UK

Email: Samuel J Webb samuel.webb@lhasalimited.org - Thierry Hanser thierry.hanser@lhasalimited.org - Brendan Howlin b.howlin@surrey.ac.uk - Paul Krause p.krause@surrey.ac.uk - Jonathan D Vessey jonathan.vessey@lhasalimited.org

Journal of Cheminformatics 2014, **6**:8 doi:10.1186/1758-2946-6-8

The electronic version of this article is the complete one and can be found online at:
http://www.jcheminf.com/content/6/1/8

Received: 25 November 2013
Accepted: 18 March 2014
Published: 25 March 2014

Journal of Cheminformatics
Volume 6

Viewing options
Abstract
Full text
PDF (1.9MB)
ePUB (910KB)
Additional files

Associated material
PubMed record
Article metrics
Readers' comments

Related literature
Cited by
Google blog search
Other articles by authors
▸ on Google Scholar
▸ on PubMed
Related articles/pages
on Google
on Google Scholar
on PubMed

Tools
Download references
Download XML
Email to a friend
Order reprints
Post a comment
Download to ...

Thank you

Lhasa Limited

shared **knowledge** • shared **progress**

Lhasa Limited Registered Office
Granary Wharf House, 2 Canal Wharf, Leeds LS11 5PS
UK Registered Charity (290866)

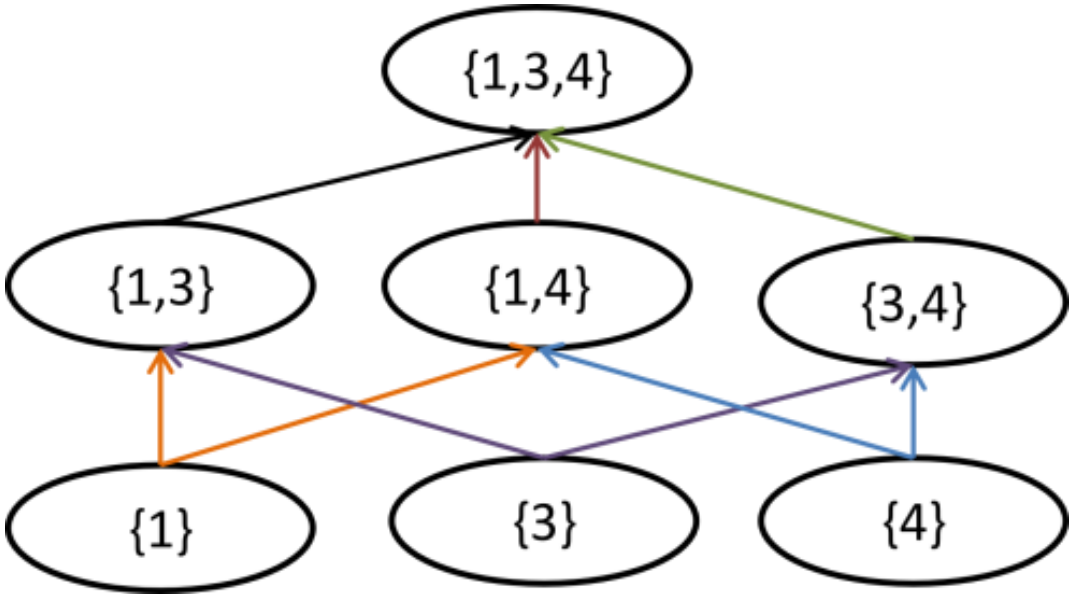
+44 (0)113 394 6020
info@lhasalimited.org
www.lhasalimited.org

Simple example

$$C(n,r) = {}^{n}C_k = {}_{n}C_k = \frac{n!}{k!\,(n-k)!}$$

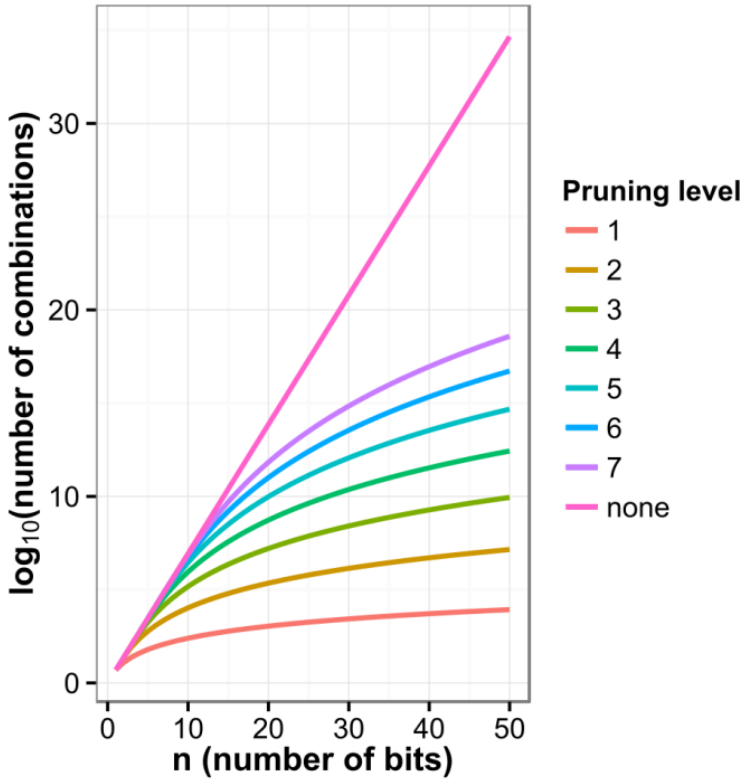
Lhasa Limited

- Given the fingerprint {1,3,4} we enumerate all combinations of set bits without repetition
- We can organise into hierarchies based on subset-superset relationships

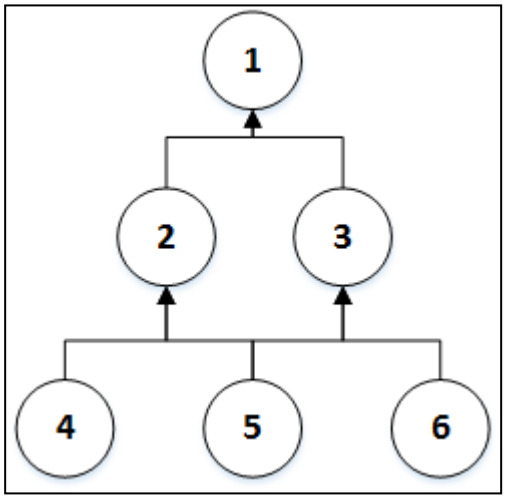


Limitations of direct enumeration

- Direct enumeration of the fingerprint has limitations
  - Disconnected features (may be undesirable)
  - Must use key based fingerprints to allow for mapping back to the structure
  - Combinatorial explosion



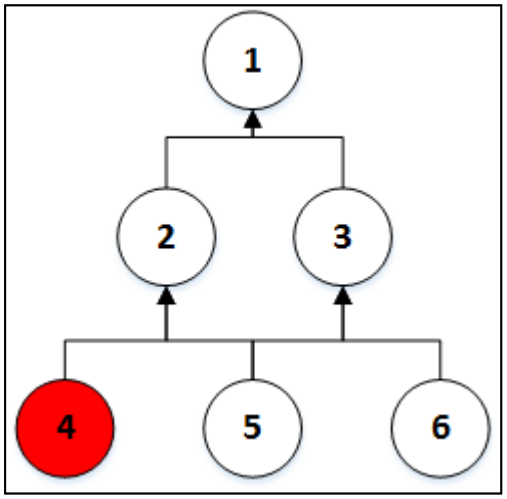
Network generation: process through model



For each node:

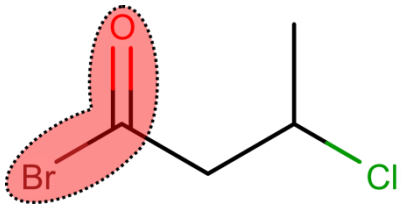
1) Generate descriptors
2) Process through model
3) Assign a prediction to the node

Network generation: process through model



We don’t need the information on the network at this point.

Start point is arbitrary.



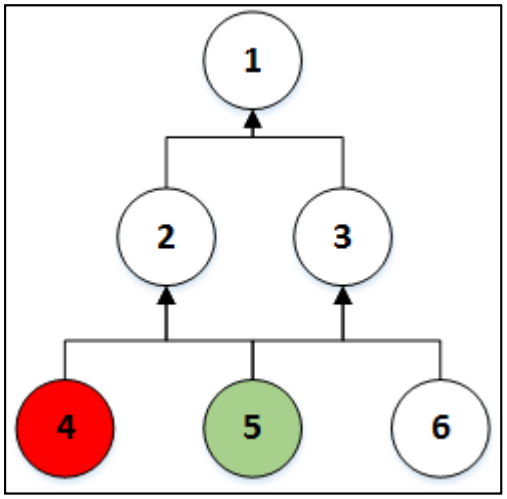
getFingerprint()

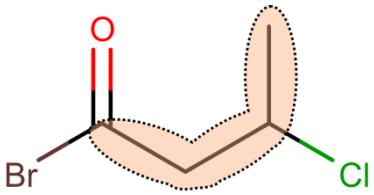
getPrediction()

active – 0.89

addPrediction()

Network generation: process through model





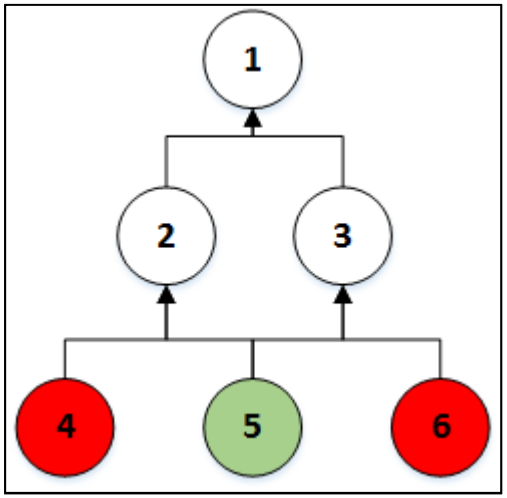
getFingerprint()

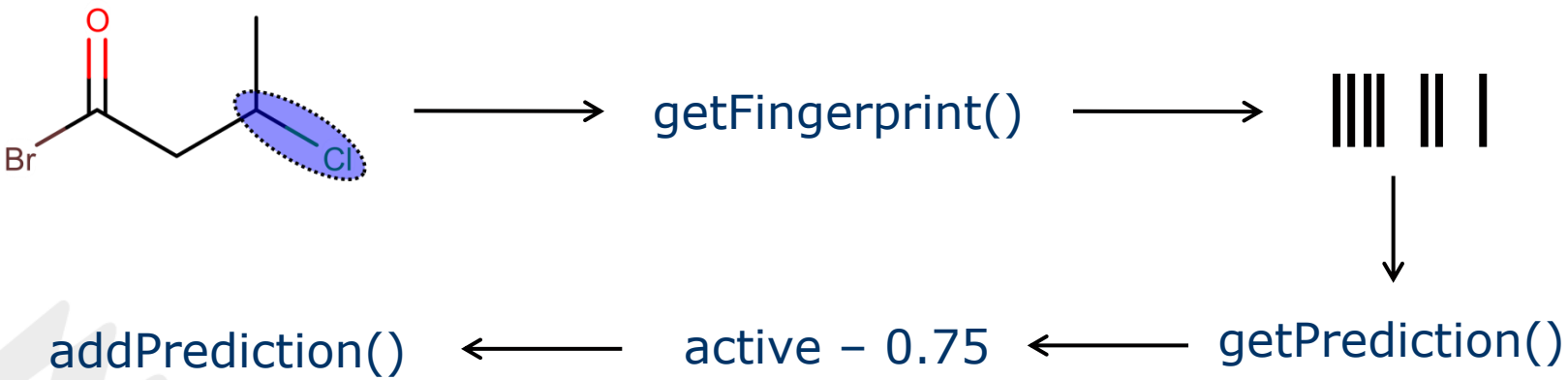
getPrediction()

inactive – 0.6

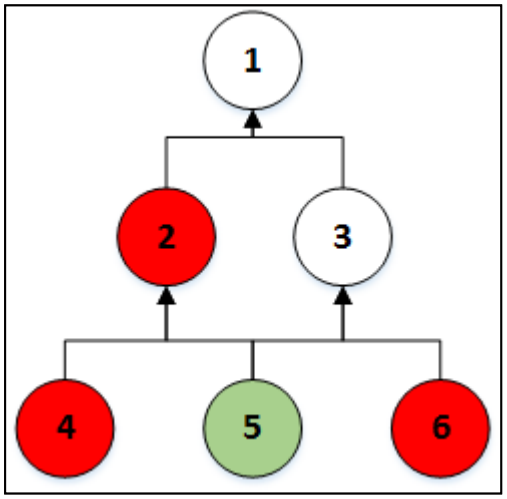
addPrediction()

Network generation: process through model

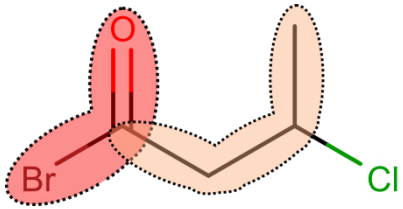




Network generation: process through model



Now we start the combination nodes



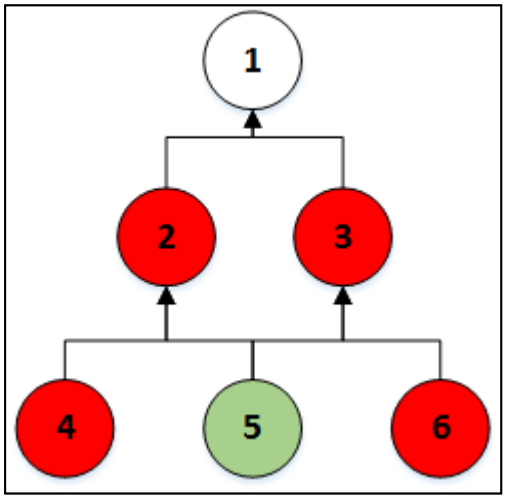
getFingerprint()

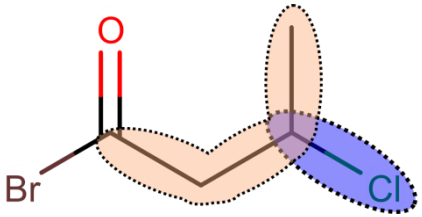
getPrediction()

active – 0.95

addPrediction()

Network generation: process through model





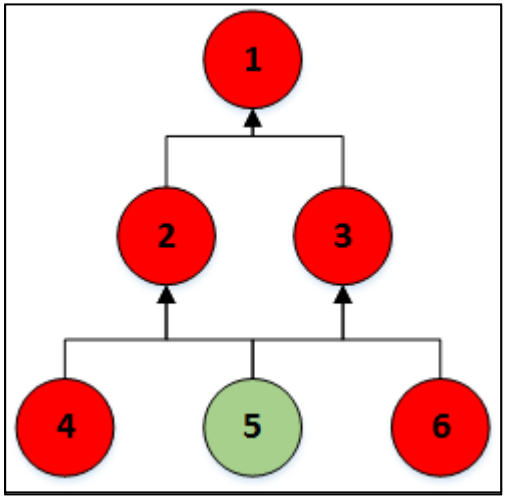
getFingerprint()

getPrediction()

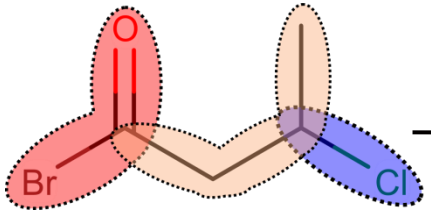
active – 0.72

addPrediction()

Network generation: process through model



Prediction for the full query structure



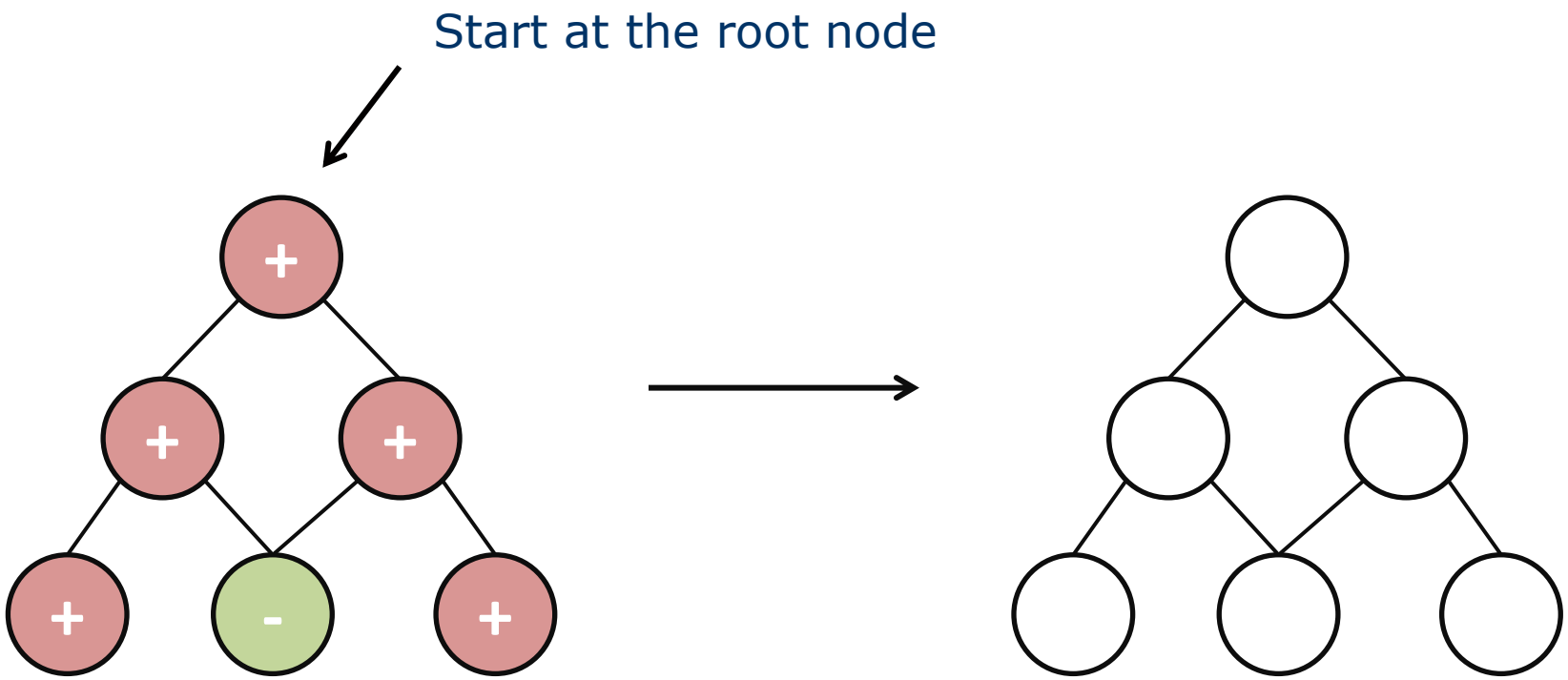
getFingerprint()

getPrediction()

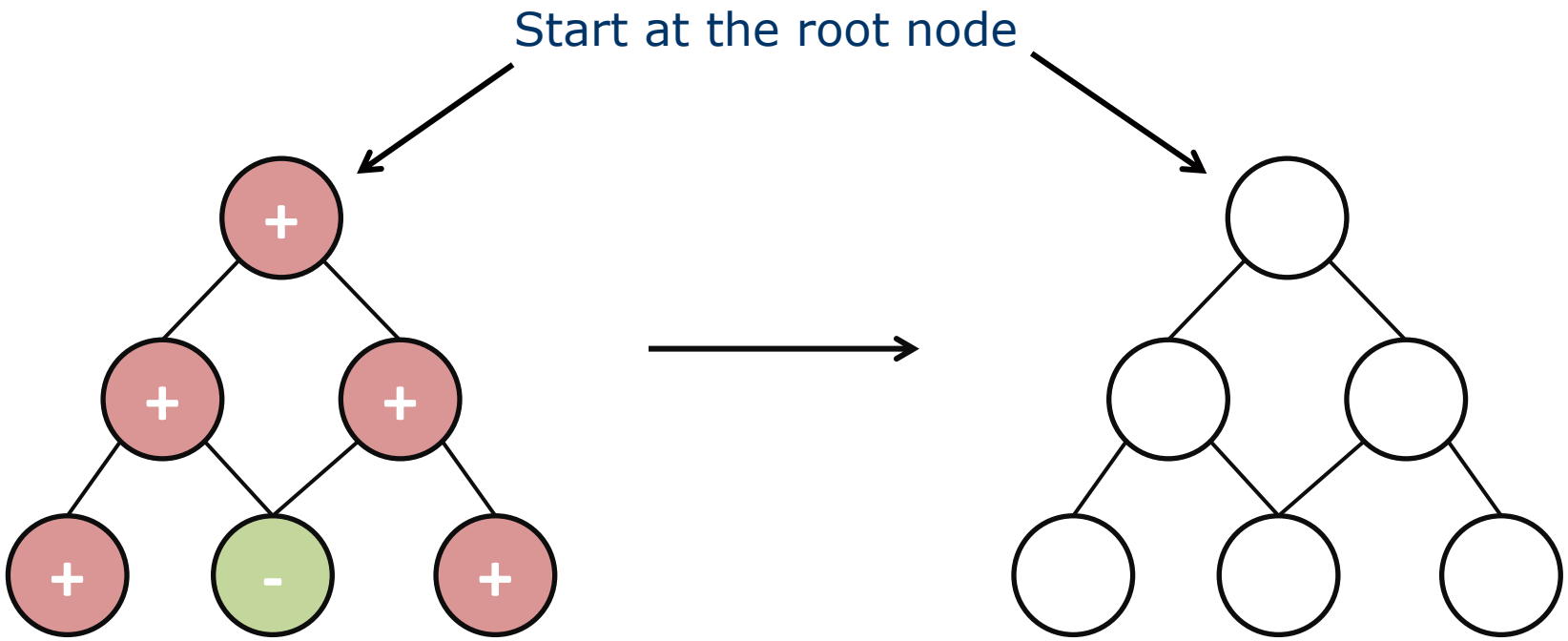
active – 0.97

addPrediction()

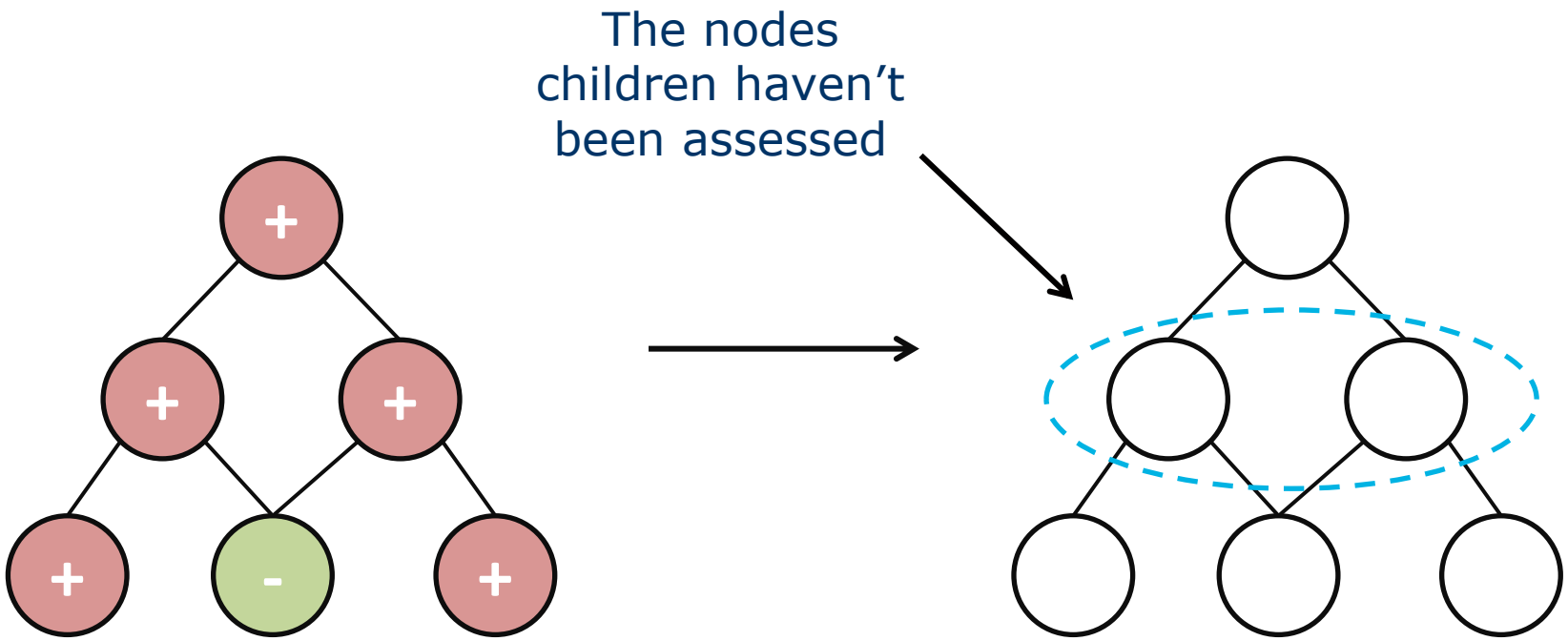
Assessment



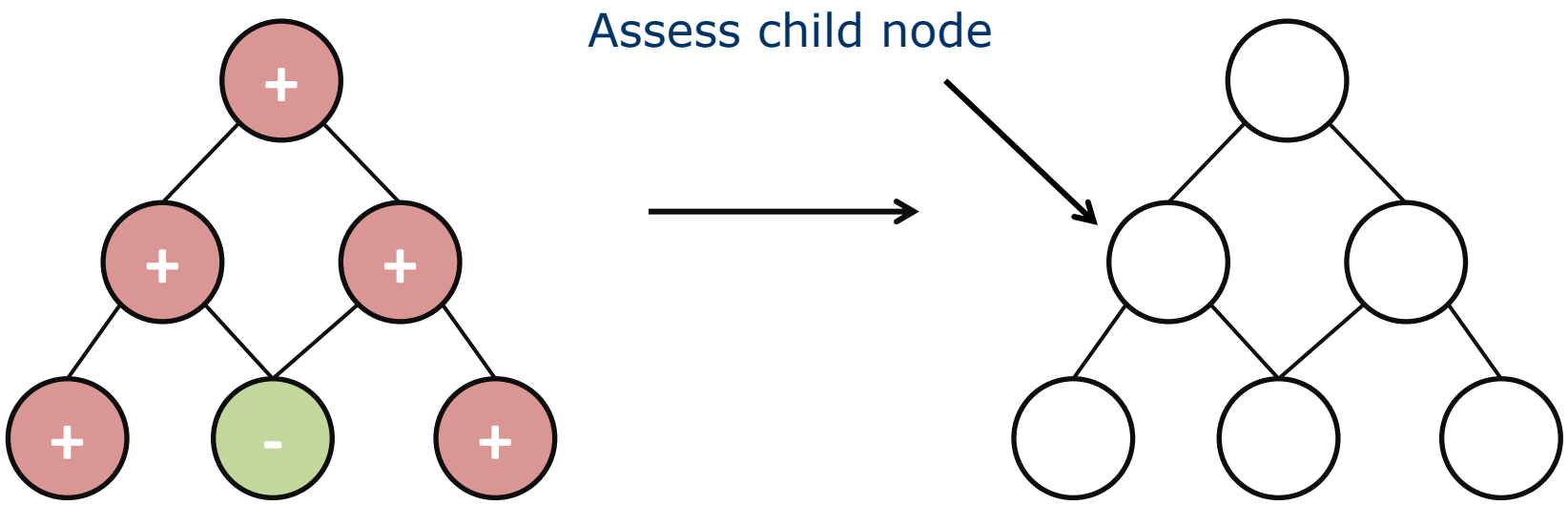
Assessment



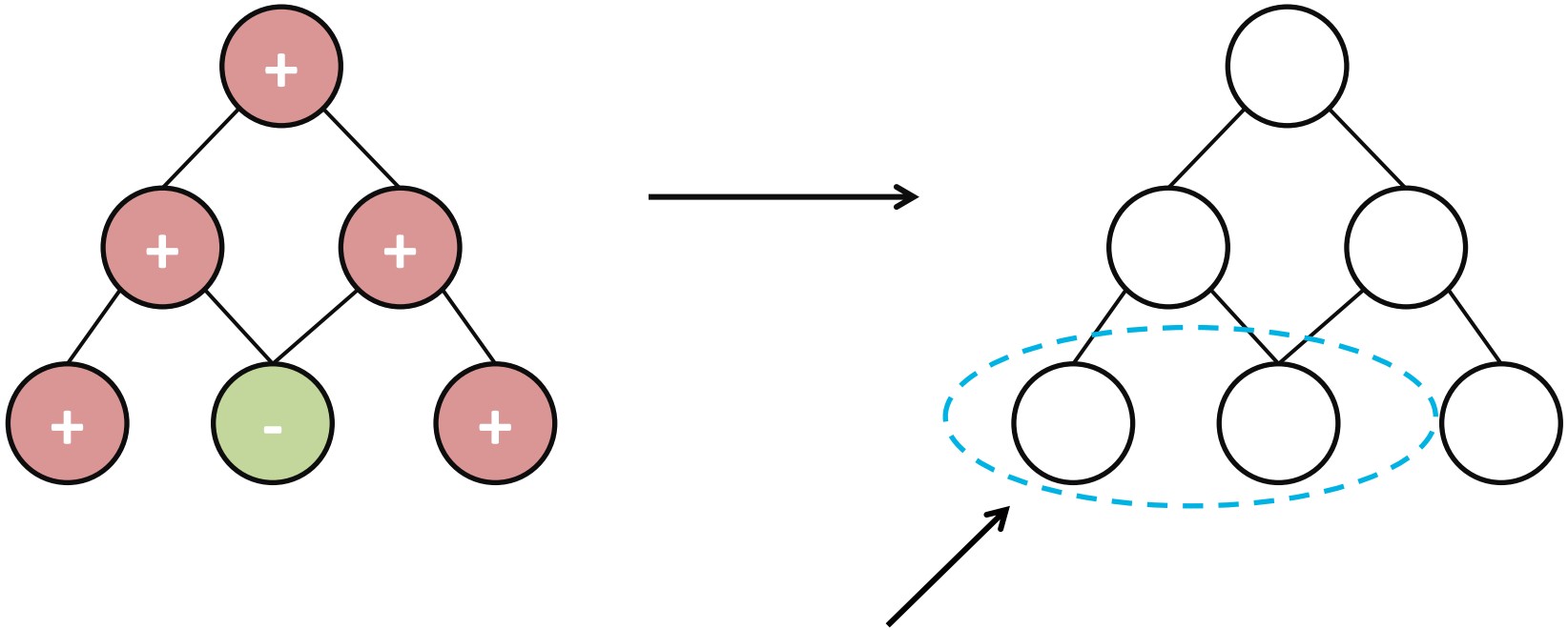
Assessment



Assessment

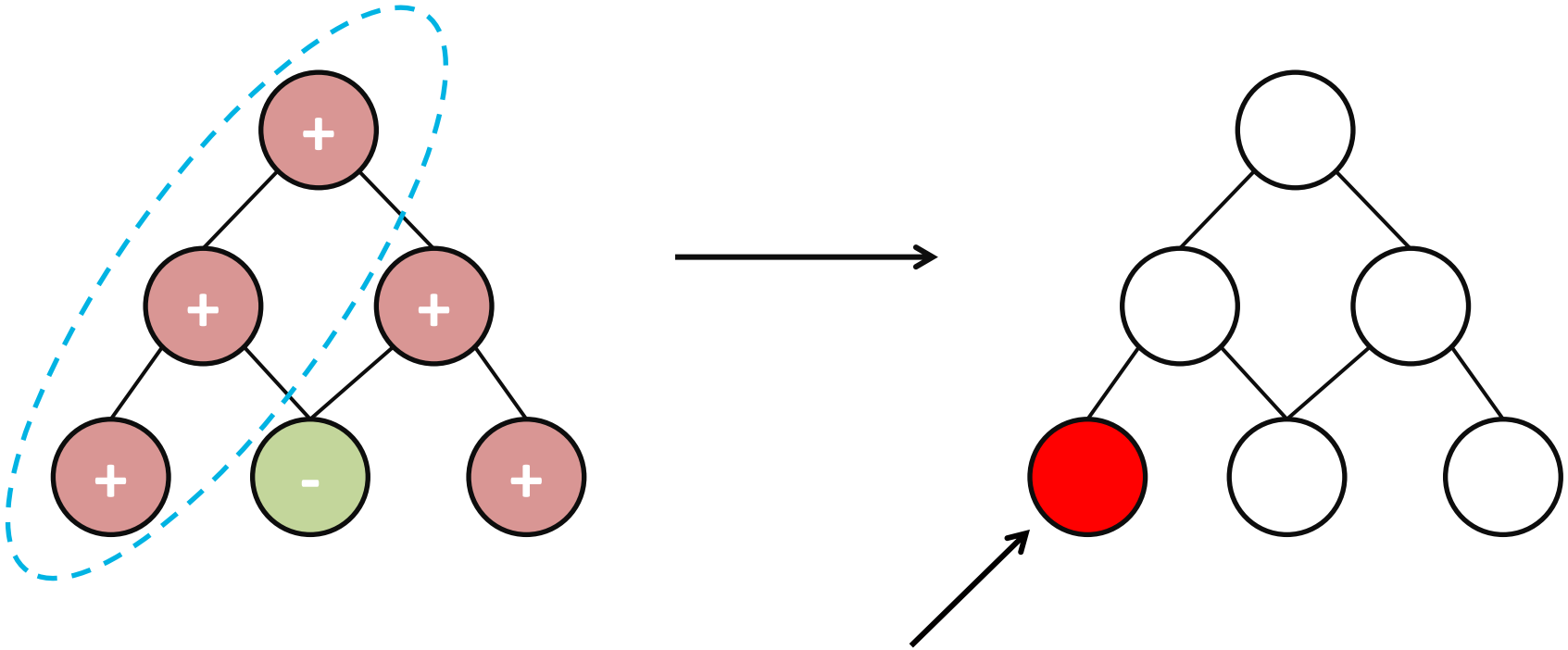


Assessment

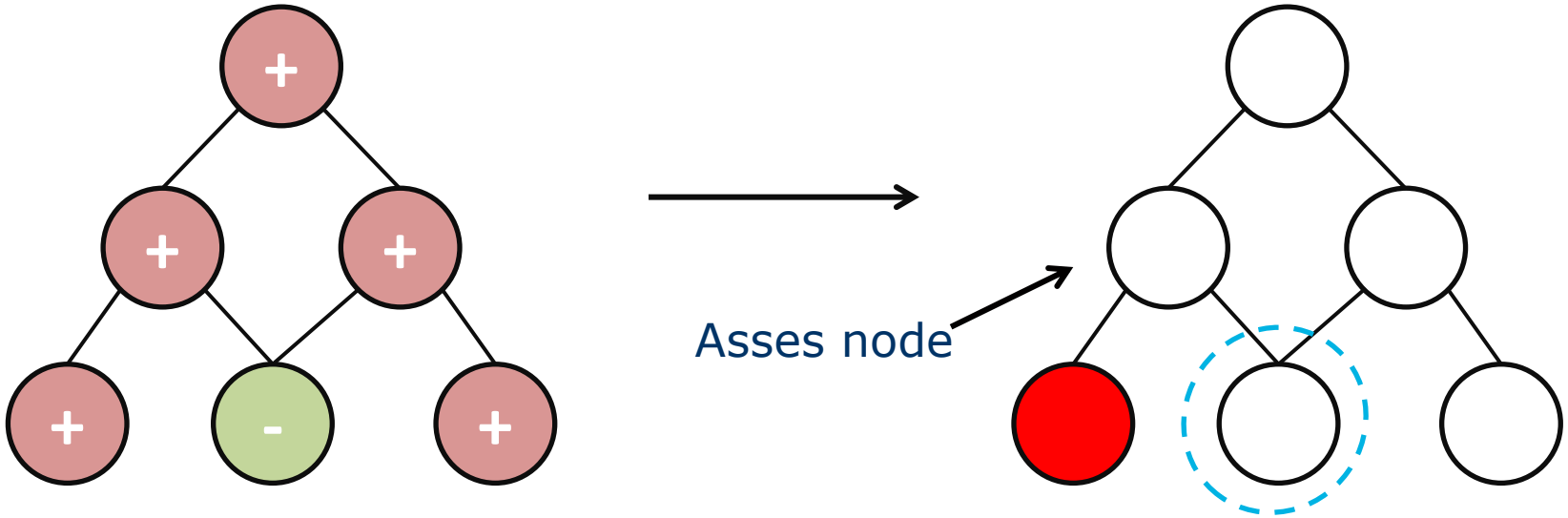


Assessment

The node is predicted active and the activity remains in the ascendant nodes
ACTIVATING



Assessment

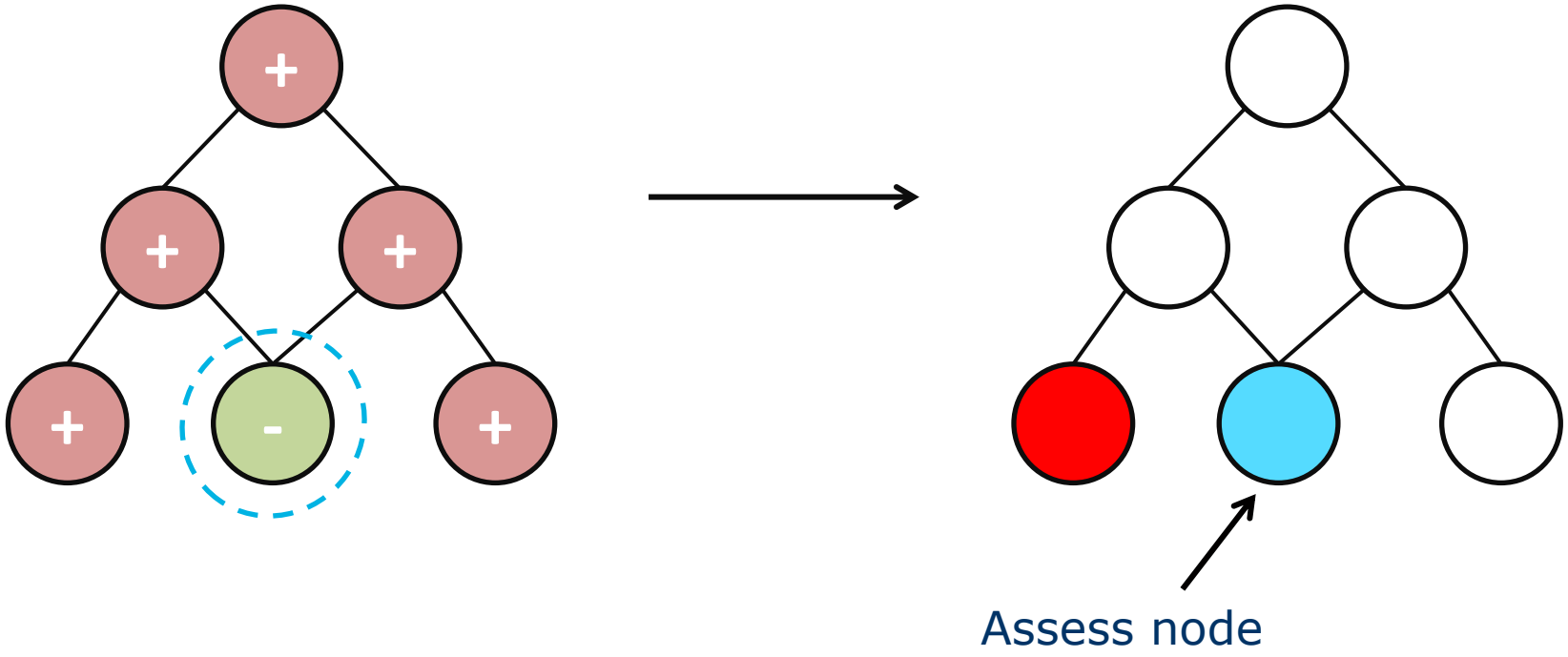


A child node still remains to be assessed

Assessment

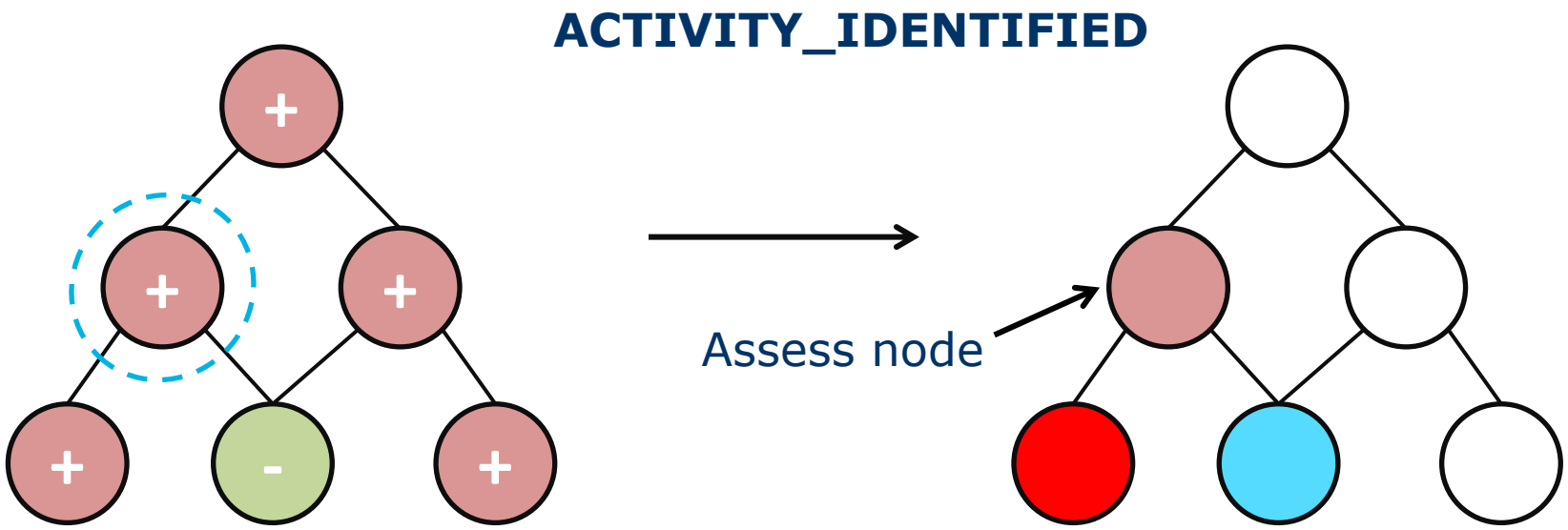
This node is predicted to be inactive and has no children.

IGNORE

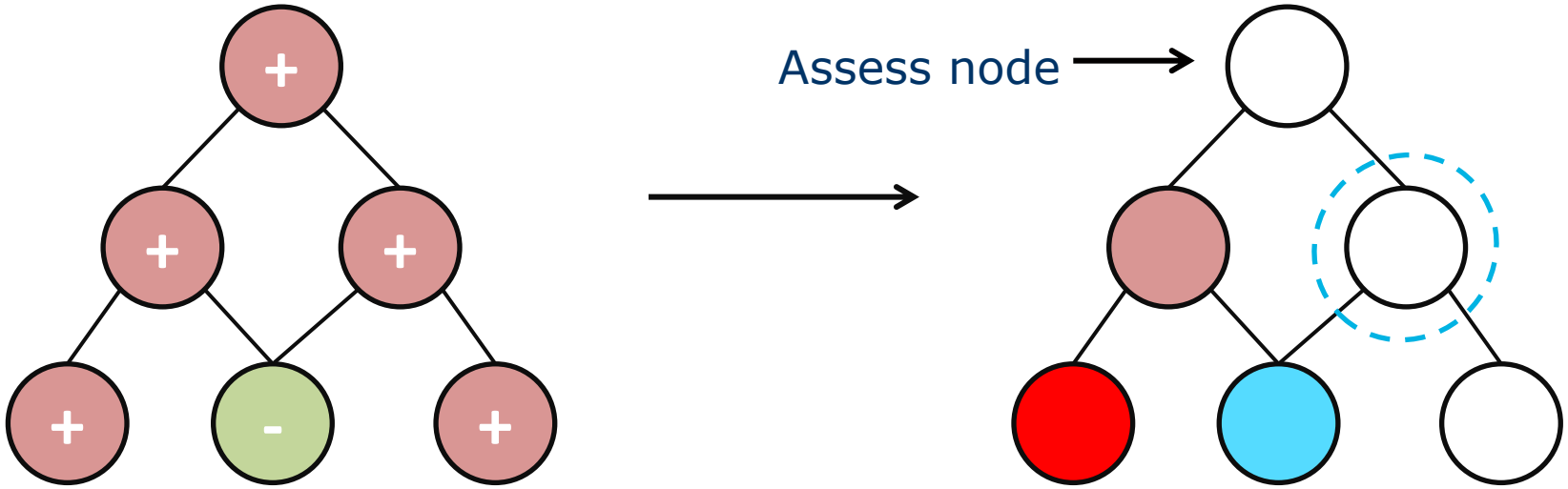


Assessment

This node is predicted to be active, all ascendants are active but it has an ACTIVATING descendant.

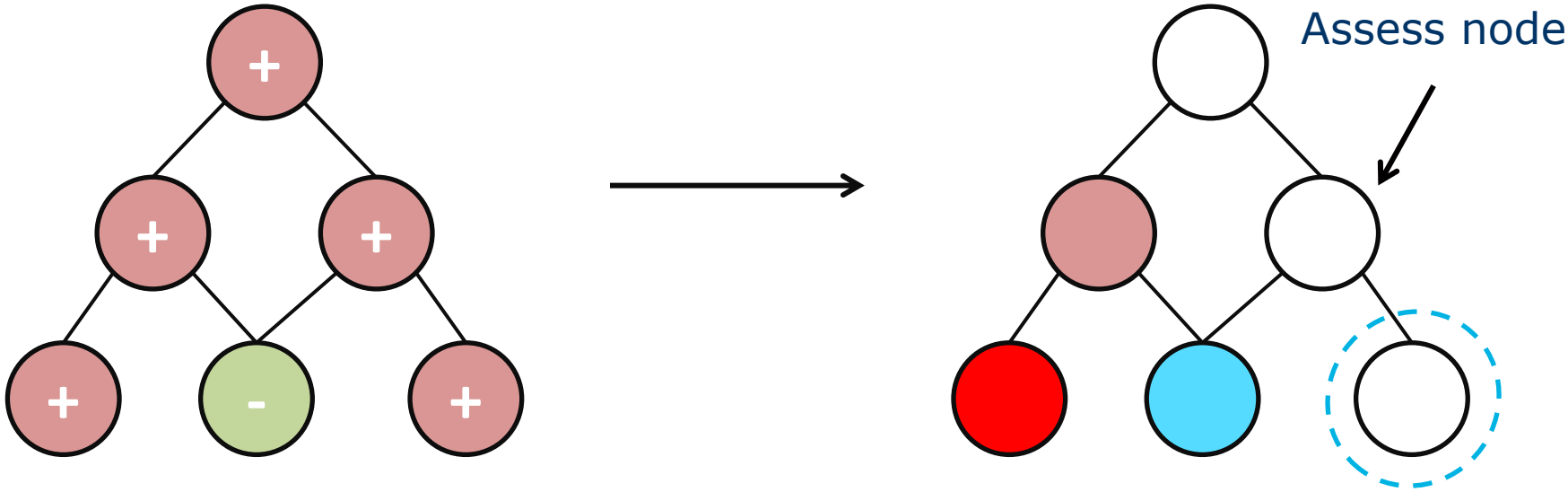


Assessment



We still have a child node
to assess

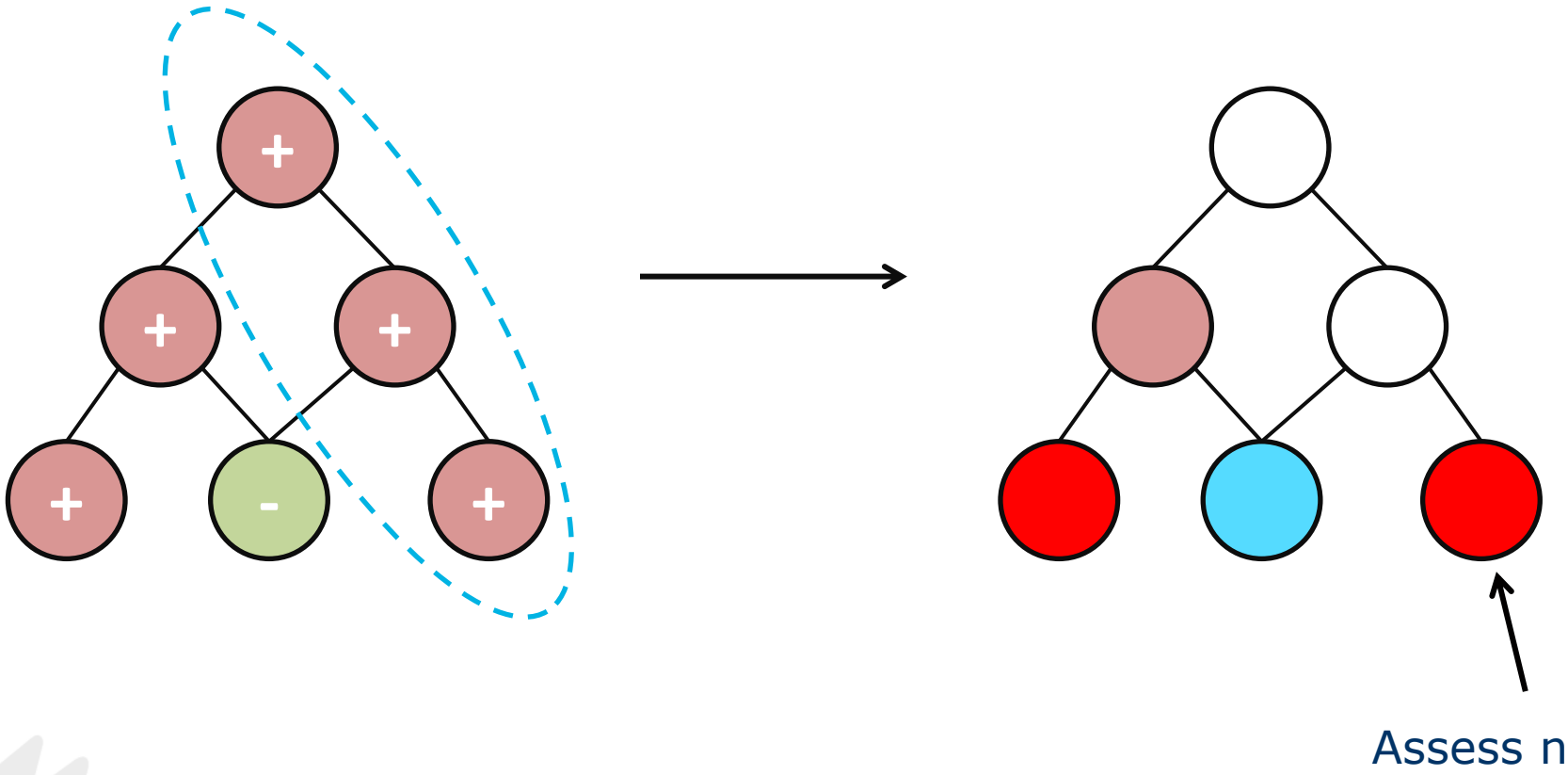
Assessment



We still have a child node
to assess

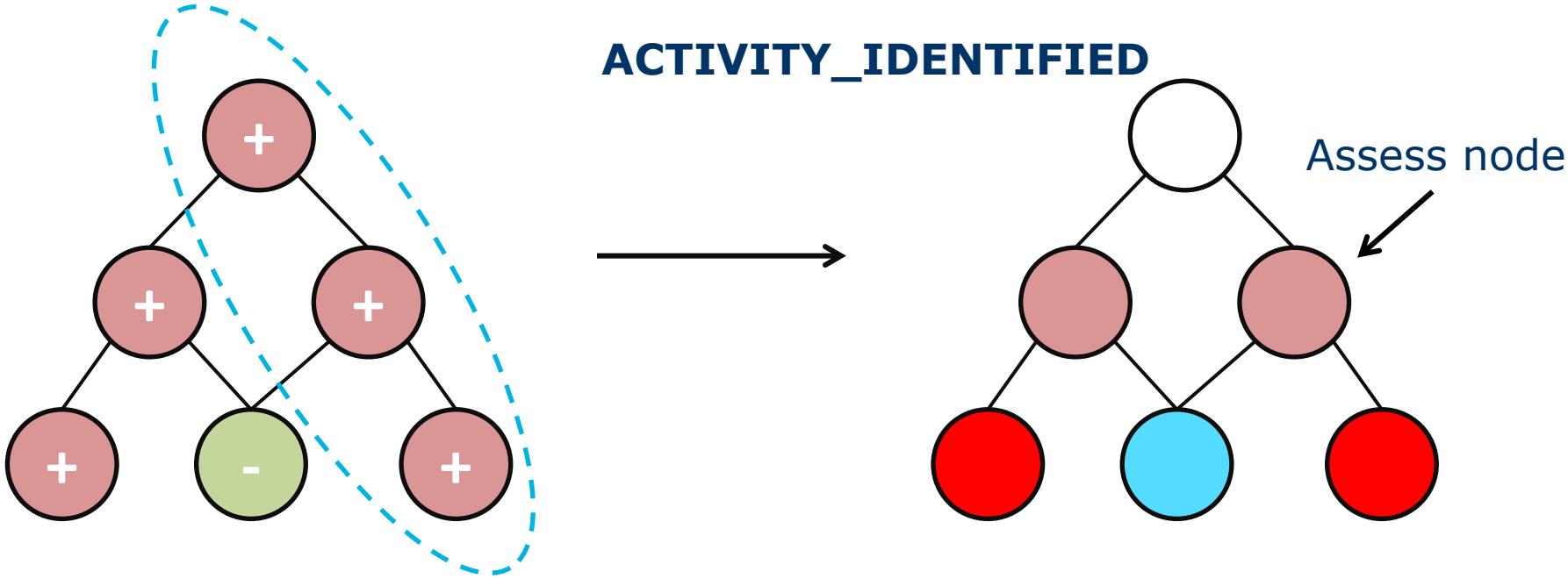
Assessment

The node is predicted active and the activity remains in the ascendant nodes
ACTIVATING



Assessment

This node is predicted to be active, all ascendants are active but it has an ACTIVATING descendant.



Assessment

This node is predicted to be active, it has no ascendants but it has multiple ACTIVATING descendants.