Combinatorial optimization methods to complete and analyse a metabolic network

Julie Laniau
Director thesis : Anne Siegel and Damien Eveillard

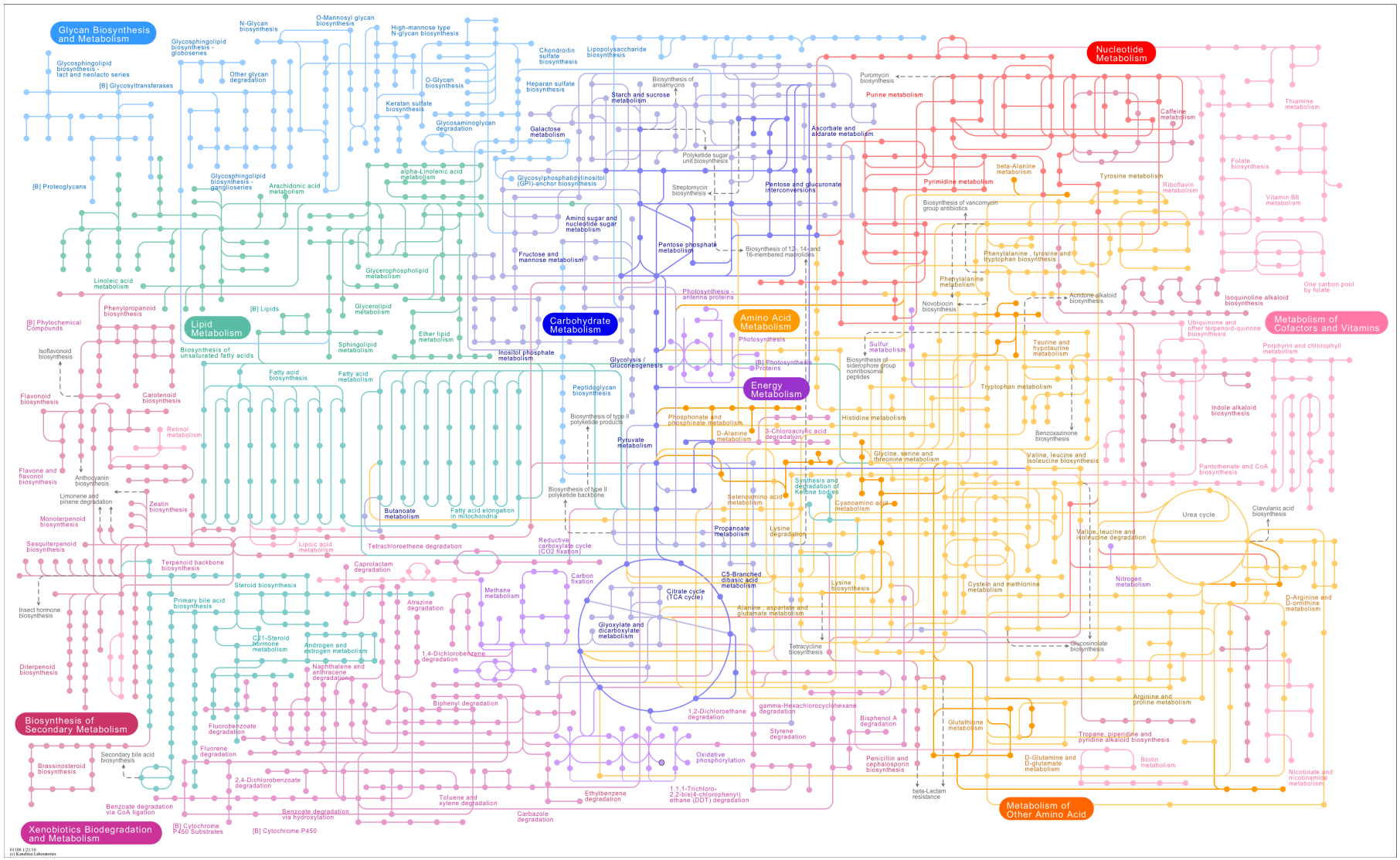
INRIA Dyliss

13/10/2015

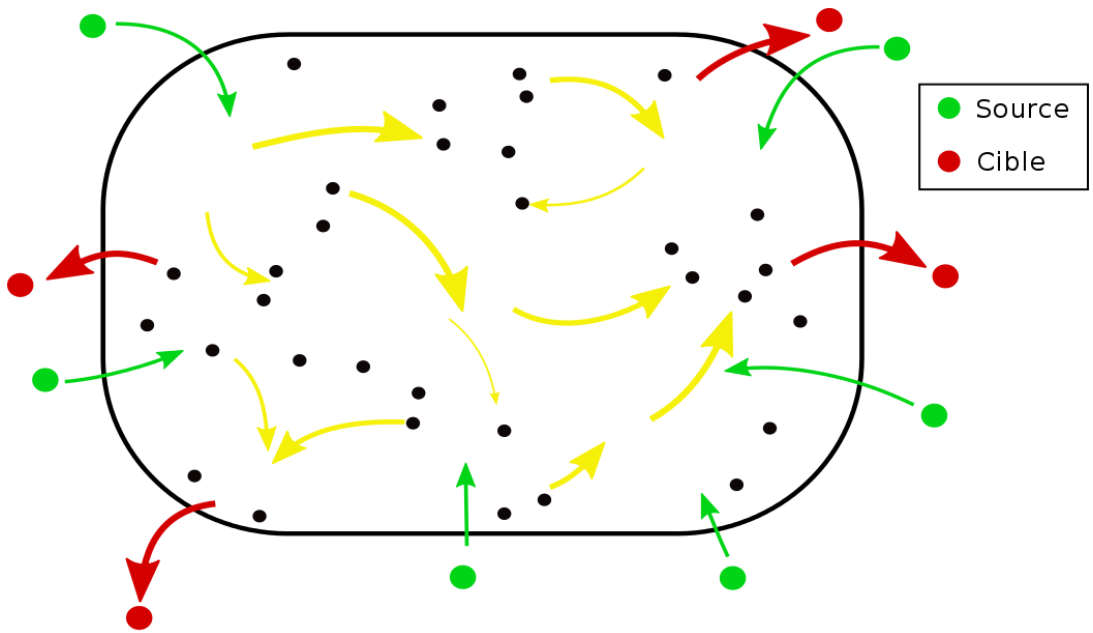
Tables of content

Tables of content

Metabolic map



Scheme of a metabolic map



- Seed : metabolites of the environment
- Target : characteristic metabolites of the organism studied

Graph definition

Based on paper "*Extending the Metabolic Network of Ectocarpus Siliculosus using Answer Set Programming*", Gebser et al.

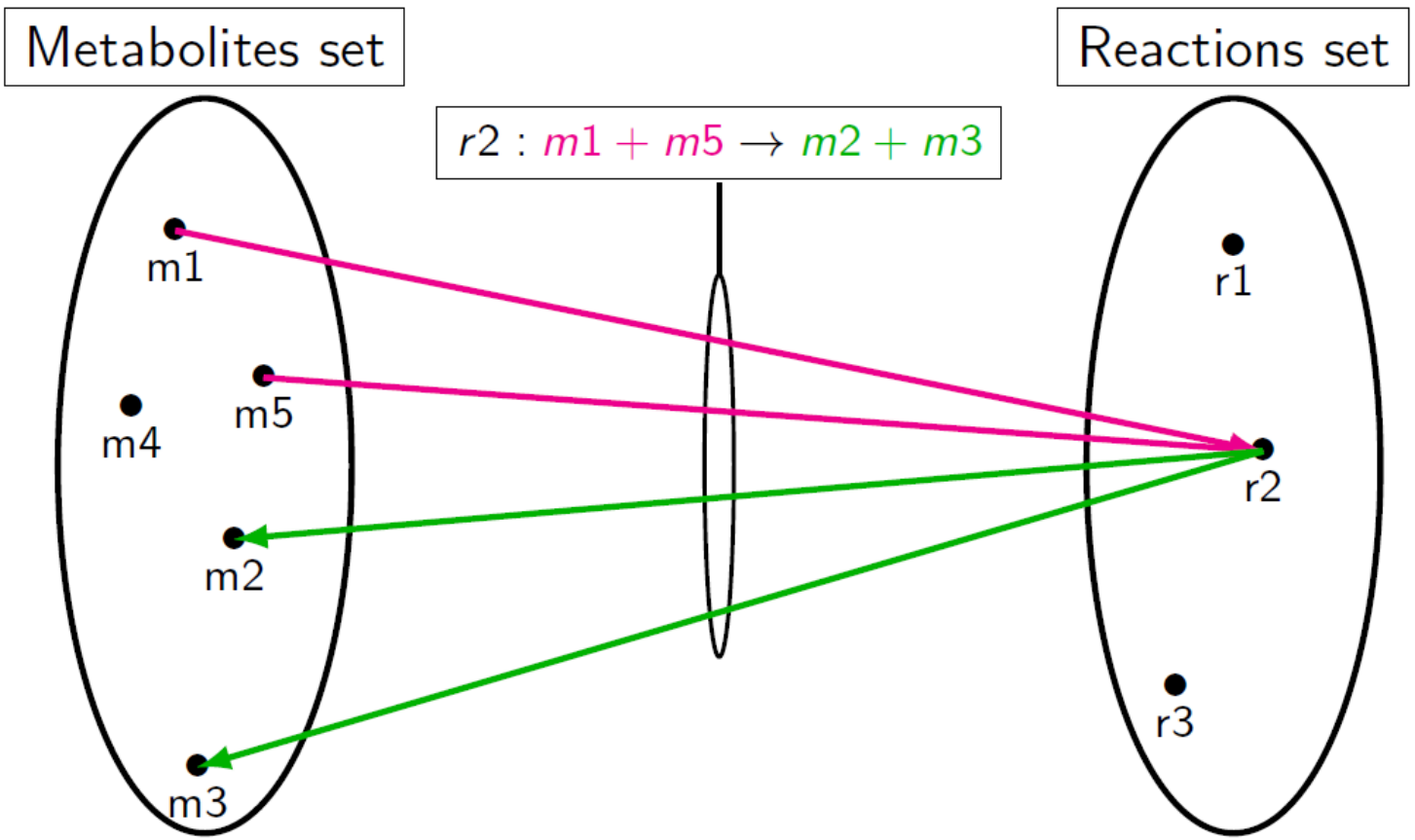
Graph network

$$G = (\mathcal{R} \cup M, E)$$

$\mathcal{R} \rightarrow$ set of *reaction* nodes
$M \rightarrow$ set of *metabolite* nodes
$E \rightarrow$ set of edges representing the relationship between metabolite and reaction

Example of a reaction in a graph-map

Representation of the reaction r2 : m1 + m5 → m2 + m3

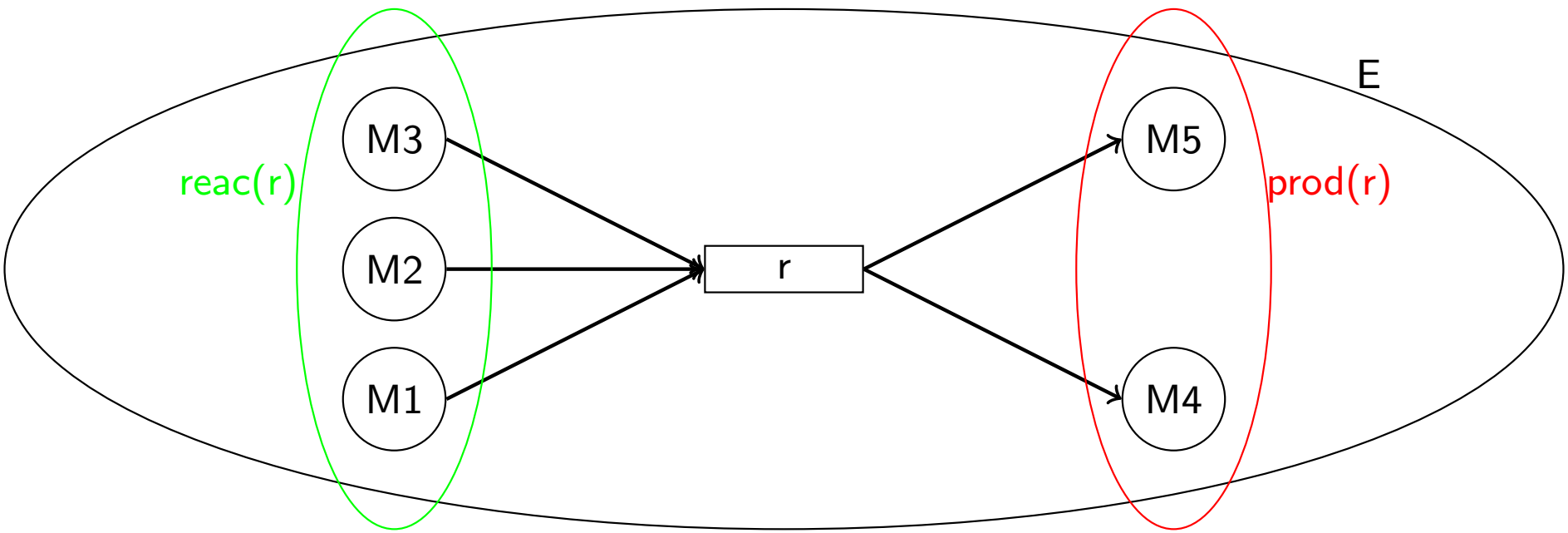


Subsets associated to a reaction

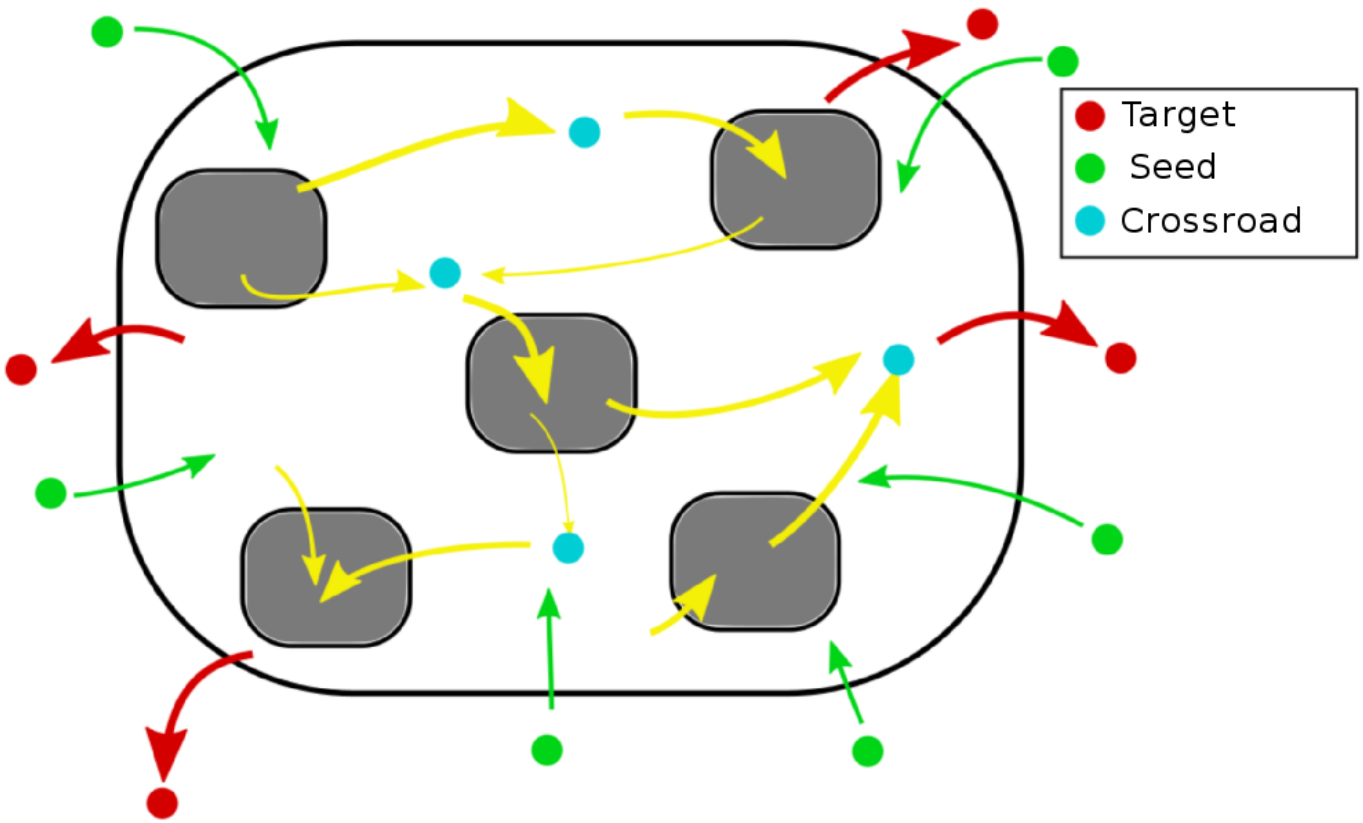
Reaction elements

Substrate	$reac(r)$	$=$	$\{m \in M \mid (m, r) \in E\}$
Product	$prod(r)$	$=$	$\{m \in M \mid (r, m) \in E\}$

Sets for a reaction



If we work on a metabolic map with compartments, we can see crossroad metabolites between it. Main question in my work is :
what is the impact of crossroad metabolites on the analysis of large-scale metabolic networks ?

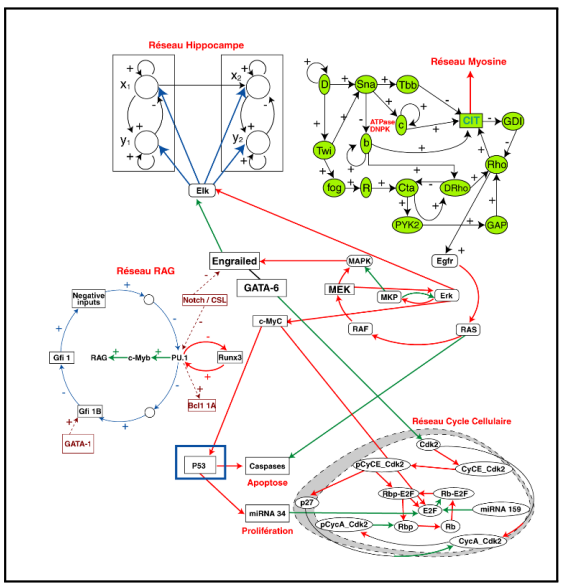


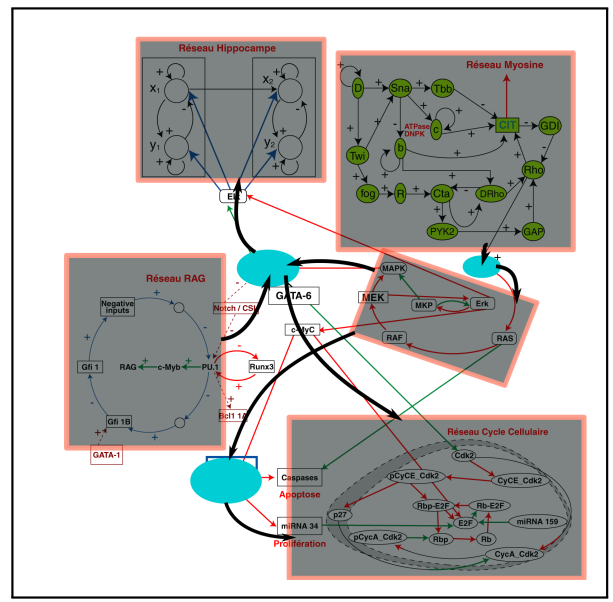
Tables of content

Tables of content

Functional Scale

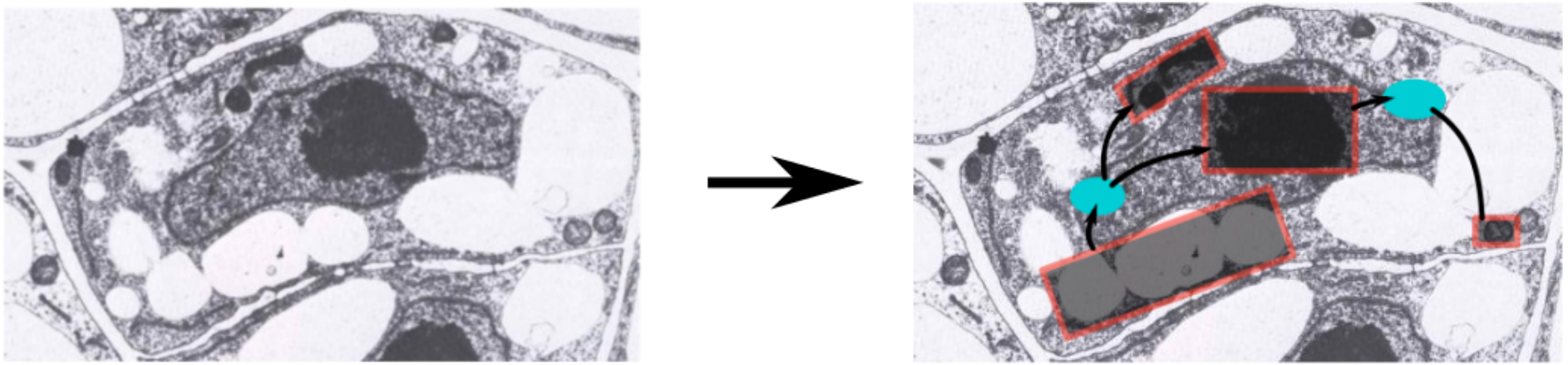
- **seed** : metabolites required to initiate the metabolite functions
- **target** : metabolites of interested within the cell or within the external media
- **compartment** : functional subnetworks





Cellular Scale

- **seed** : nutrients required for the cell survival
- **target** : metabolites of interested within the external media
- **subnetwork** : biological compartment

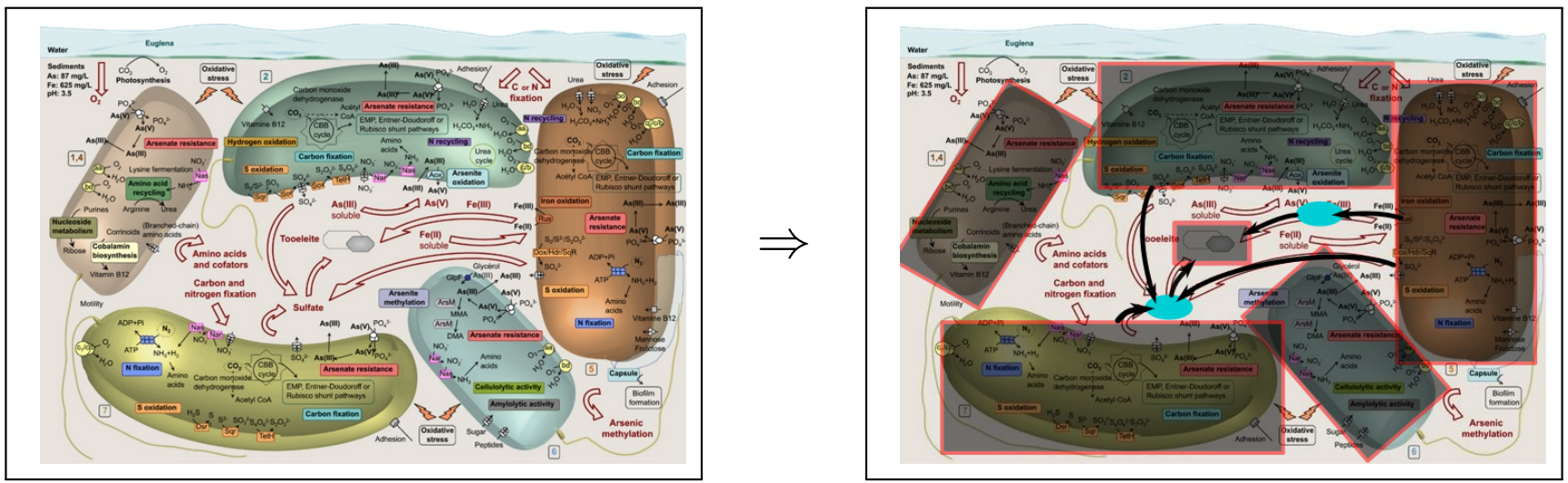


Metagenomic Scale

The network represents a community of organisms

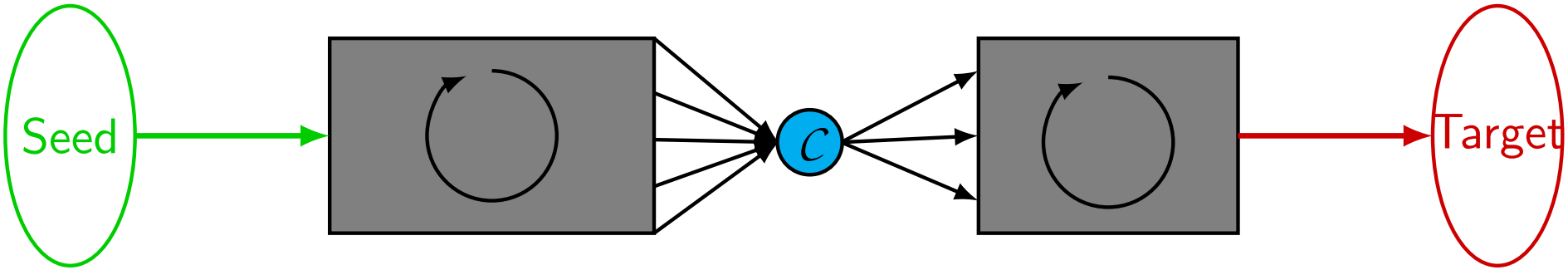
- **seeds** : media nutrients
- **targets** : metabolites we want to produce
- **one subnetwork** : one organism

Study of Euglena



Tables of content

General model



Crossroad

∗ Metabolite which is produced by the map then consumed (production and consumption can occur at different timescales)
∗ The metabolite is also required to produce at least one target

Problem

Will the reconstruction of metabolic map from genome-scale data be different if we require the production and consumption of crossroad metabolites ?

Tables of content

Tables of content

Definition of the problem

Why ?

- Creation of the metabolic map from genome-scale experimental data
- Genomic data are incomplete
- The metabolic map associated to a genome contains many gaps

How ?

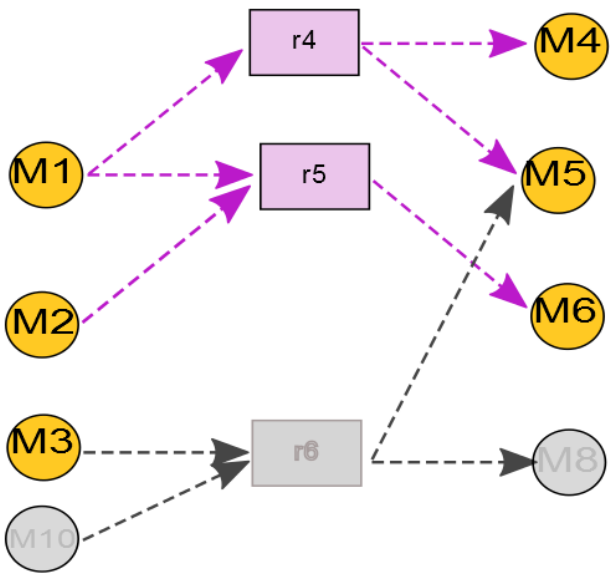
Gap filling of the metabolic map using a reference database of reactions to product targets from seeds.

Qualitative reachability in a metabolic map

Accessibility

$*r \in \mathcal{R}$ accessible by M' if $reac(r) \subseteq M'$

$*m \in M$ accessible by M' if $\begin{cases} m \in M' \\ \exists r \in \mathcal{R} \text{ accessible by } M' \ \& \ m \in prod(r) \end{cases}$



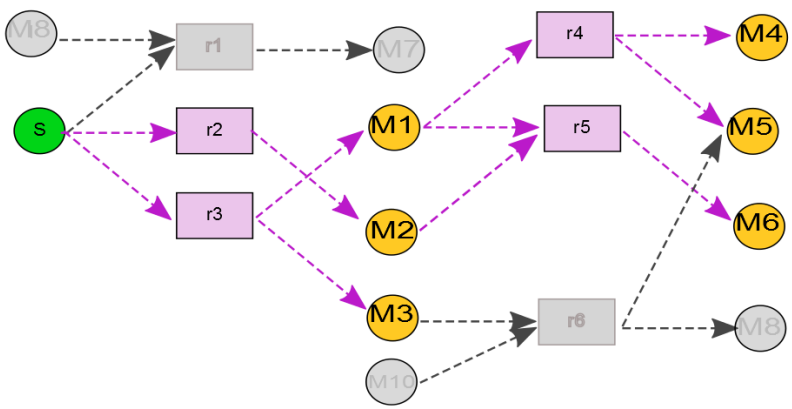
M' $= \{M1, M2, M3\}$
accessible reactions : $\{r4, r5\}$
accessible metabolites :
$\{M1, M2, M3, M4, M5, M6\}$

Scope

$\sum_G(M')$ *or* $\sum(M')$: Set of accessible metabolites by M' in graph G

$$\sum_G(M') = \cup_i M_i$$

$$M_{i+1} = M_i \cup \{m \text{ accesible by } M_i\}$$



$M' = \{S\}$

$M_0 = \{S\}$

$M_1 = M_0 \cup \{M1, M2, M3\} = \{S, M1, M2, M3\}$

$M_2 = M_1 \cup \{, M4, M5, M6\}$

$= \{S, M1, M2, M3, M4, M5, M6\}$

$$\sum_G(M') = \{S, M1, M2, M3, M4, M5, M6\}$$

Tables of content

Networks

Initial network

$$G = (\mathcal{R} \cup M, E)$$

$\mathcal{R}$, M : reactions and metabolites evidenced in the specie

Reference database of reactions

$$G' = (\mathcal{R}' \cup M', E')$$

Database : Metacyc

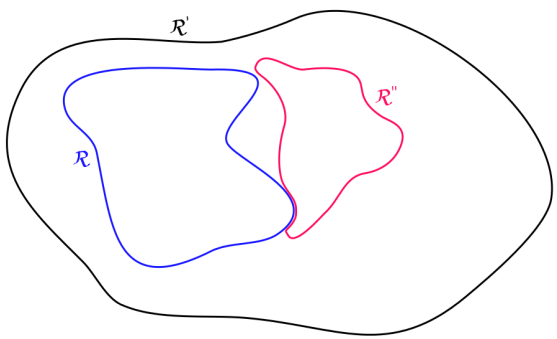
Functional gapfilling of a metabolic networkl

Goal : Find a minimum set of reactions $\mathcal{R}'' \subseteq \mathcal{R}' \setminus \mathcal{R}$ such that $T \subseteq \sum_{G_{res}}(S)$

Network after functional completion

$G_{res} = ((\mathcal{R} \cup \mathcal{R}'') \cup (M \cup M''), E \cup E'')$ where

- $\mathcal{R}'' \subseteq \mathcal{R}'$
- $M'' = \{m \in M' \mid r \in \mathcal{R}'', m \in reac(r) \cup prod(r)\}$
- $E'' = \{(m, r) \in E' \mid r \in \mathcal{R}'', m \in reac(r)\} \cup \{(r, m) \in E' \mid r \in \mathcal{R}'', m \in prod(r)\}$



Meneco

Main characteristics of the reconstruction

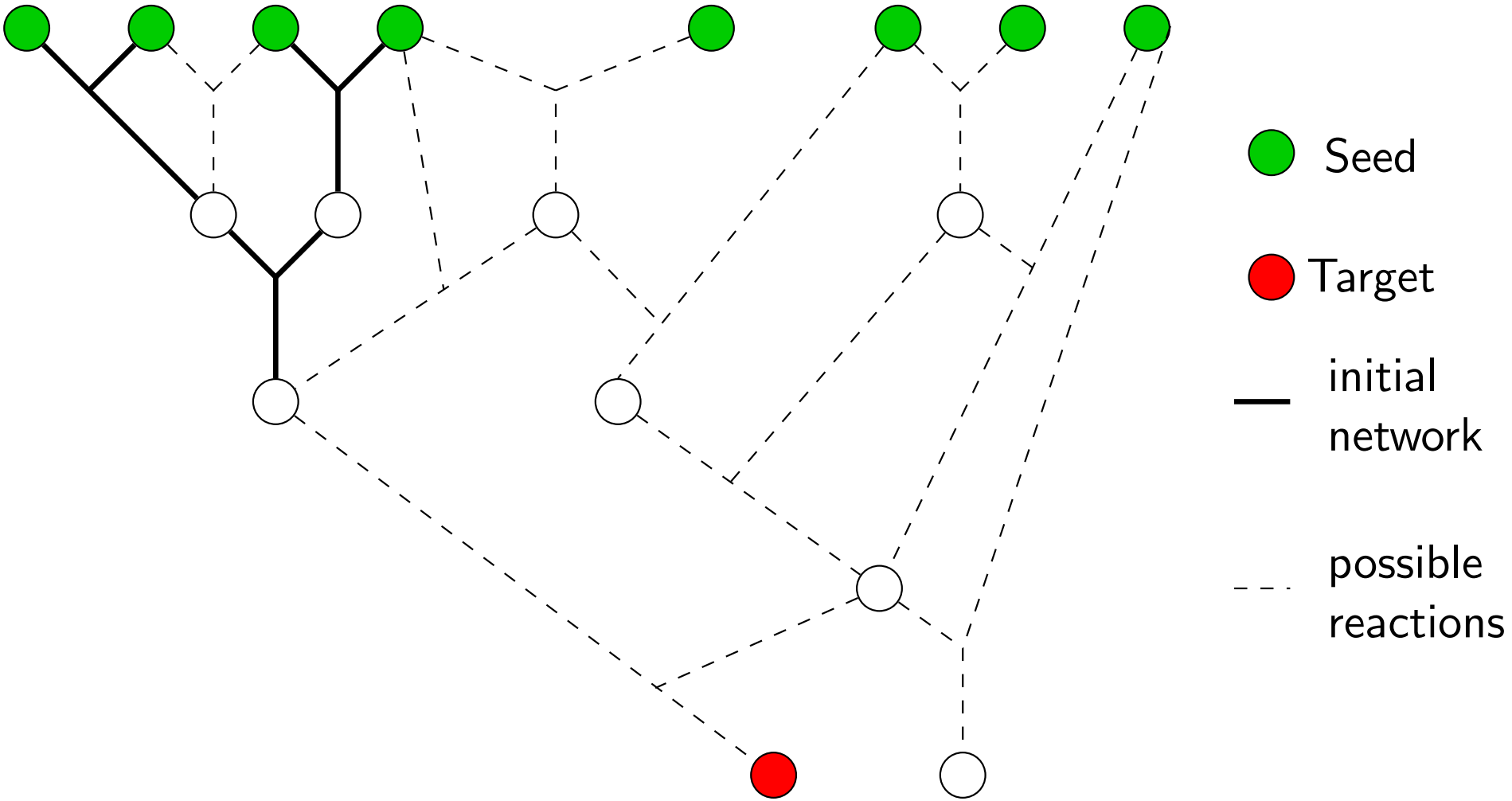
- Use of an initial network derived from experimental data and knowledge
- Modelling the gap-filling problem in a logic constraints formalism (ASP)
- Search for a minimal number of reactions to restore the producibility of targets

Meneco outputs

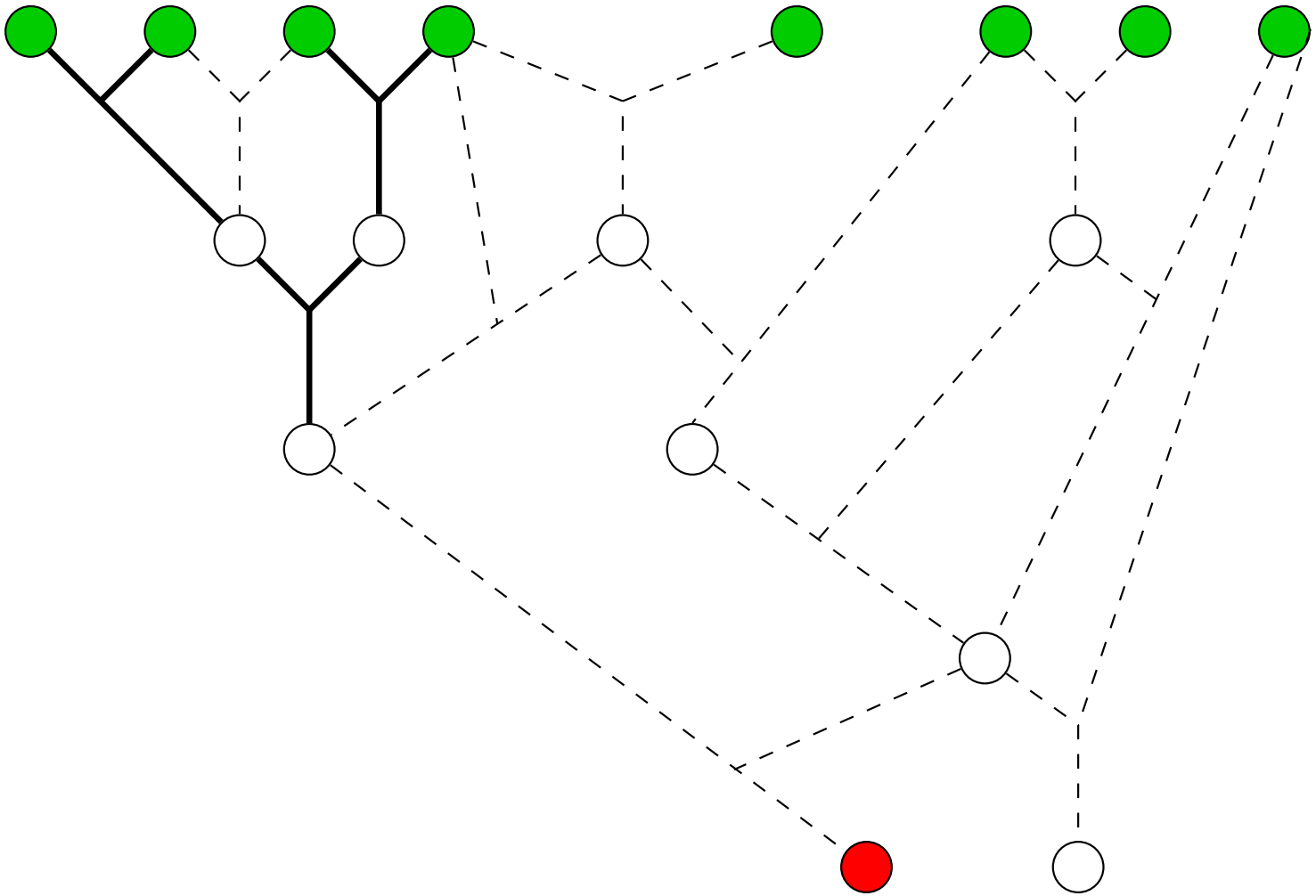
Three sets returned

- One minimal solution : a set (among others) of reactions which restore the production
- Union of all solutions : set of all reactions appearing in at least one functional completion
- Optional : every solutions

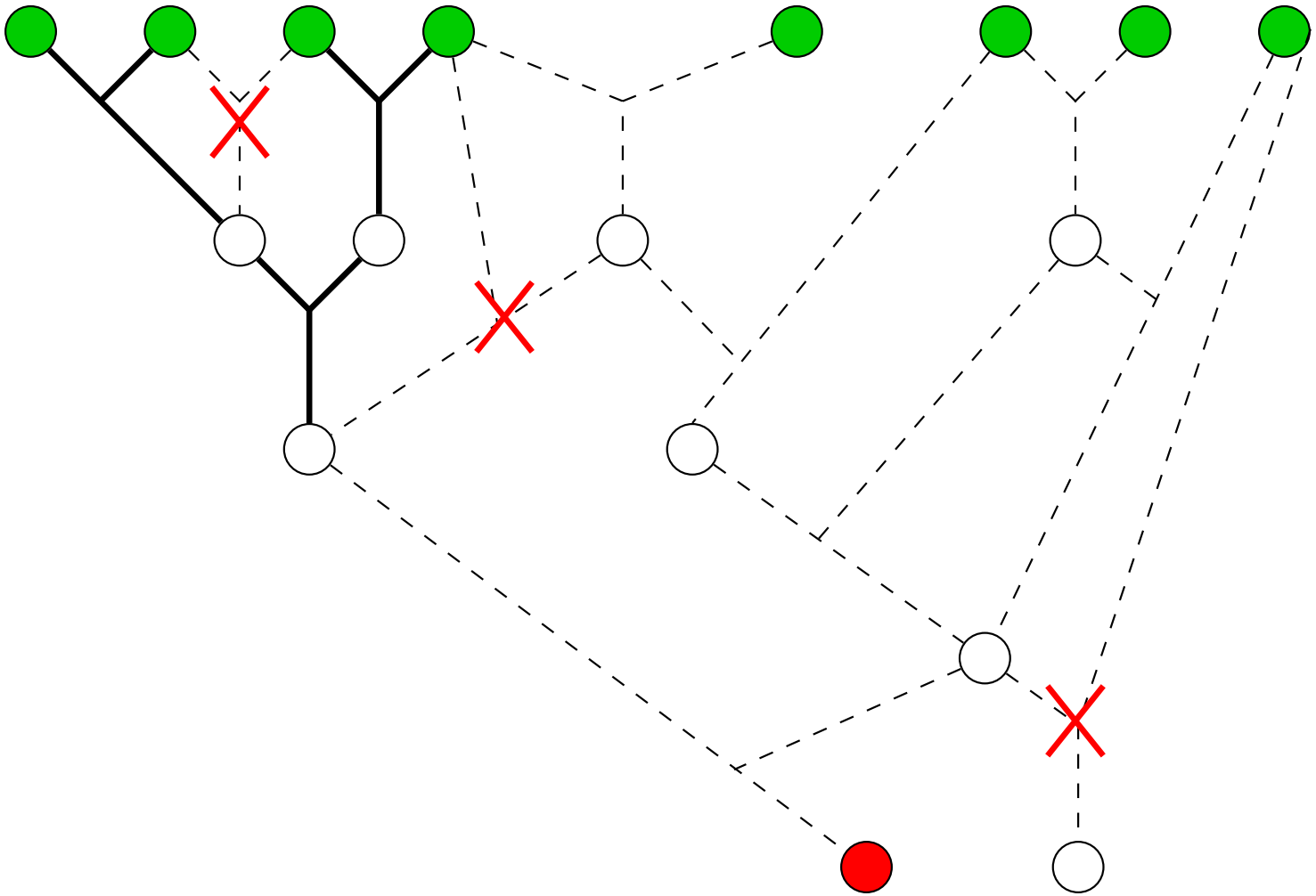
Example



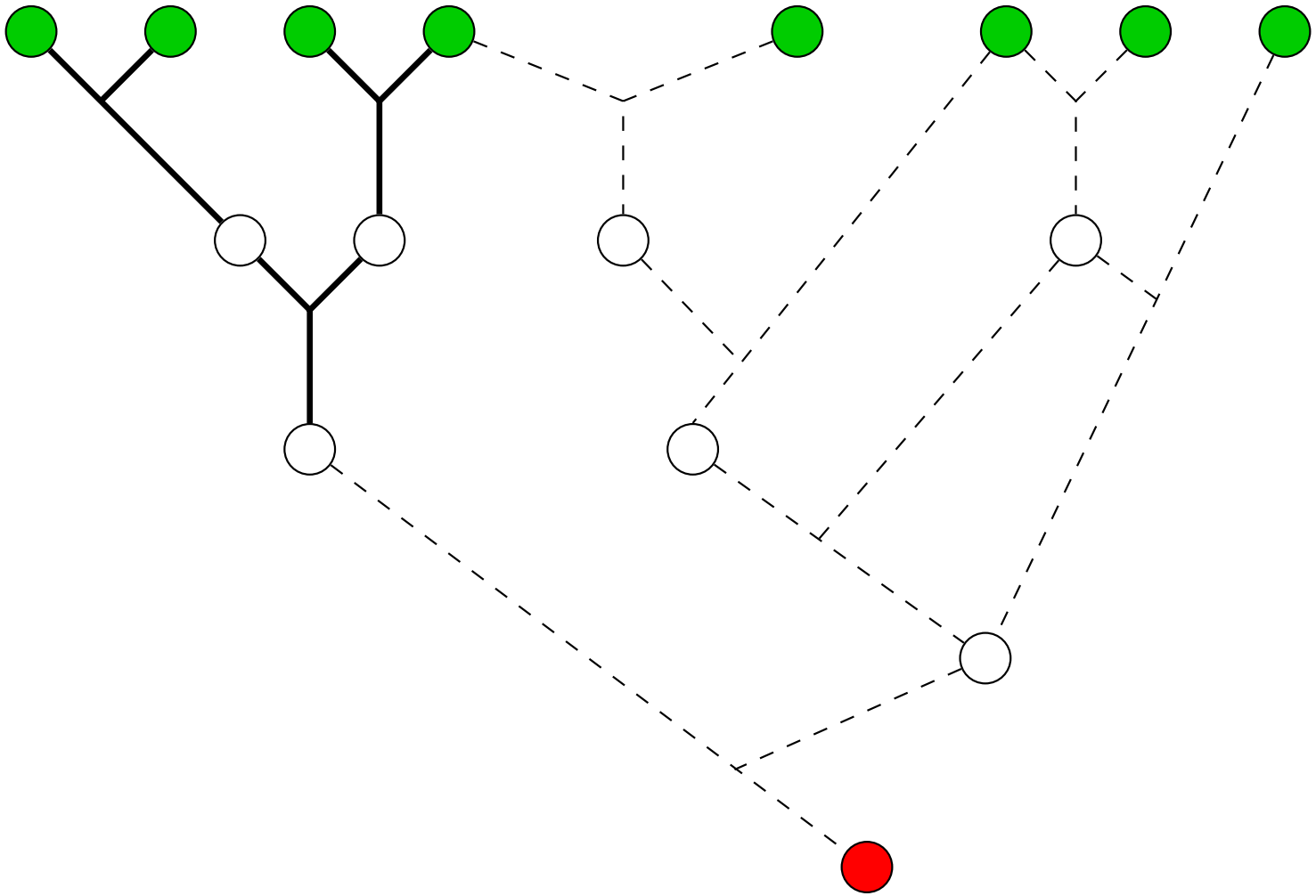
First step : remove not interesting reactions



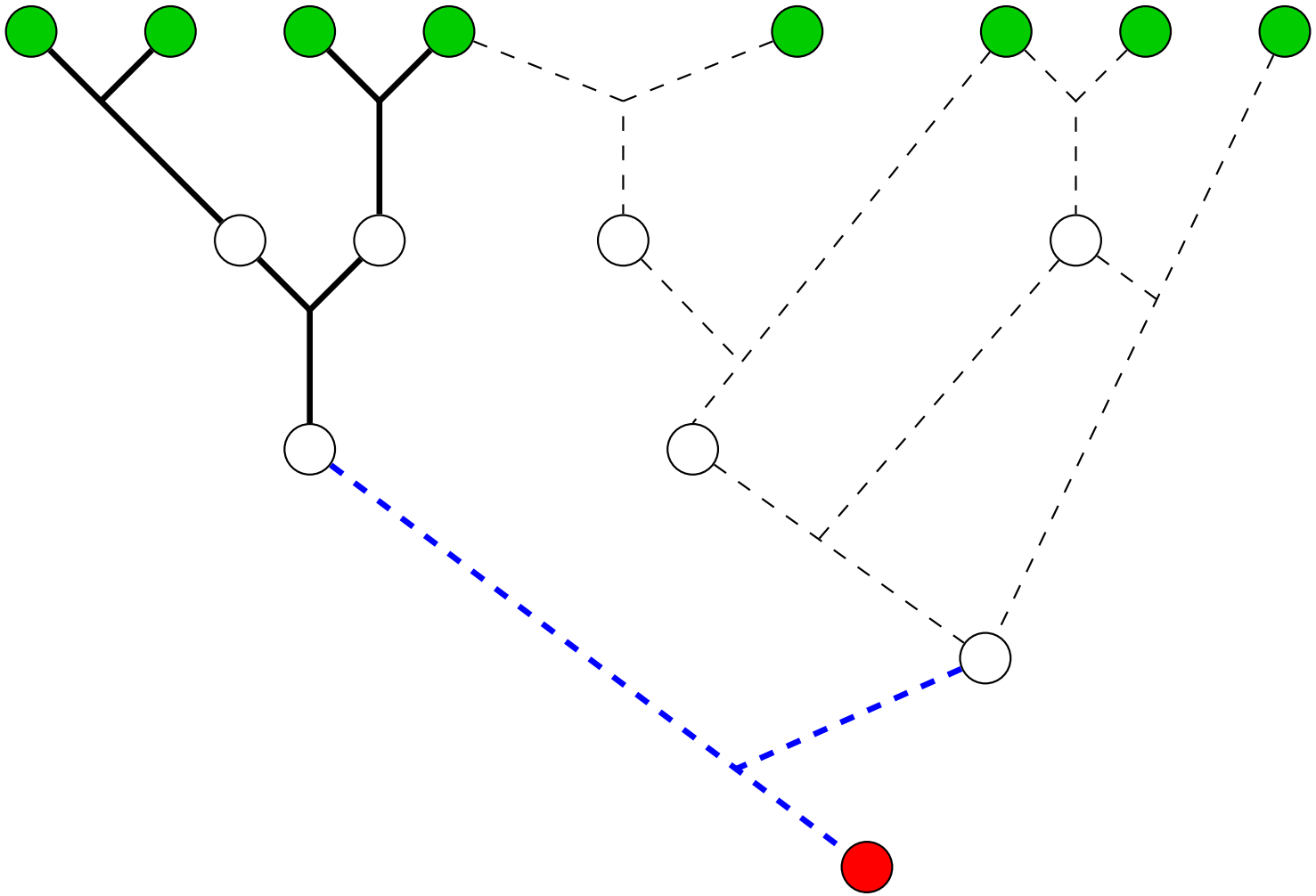
First step : remove not interesting reactions



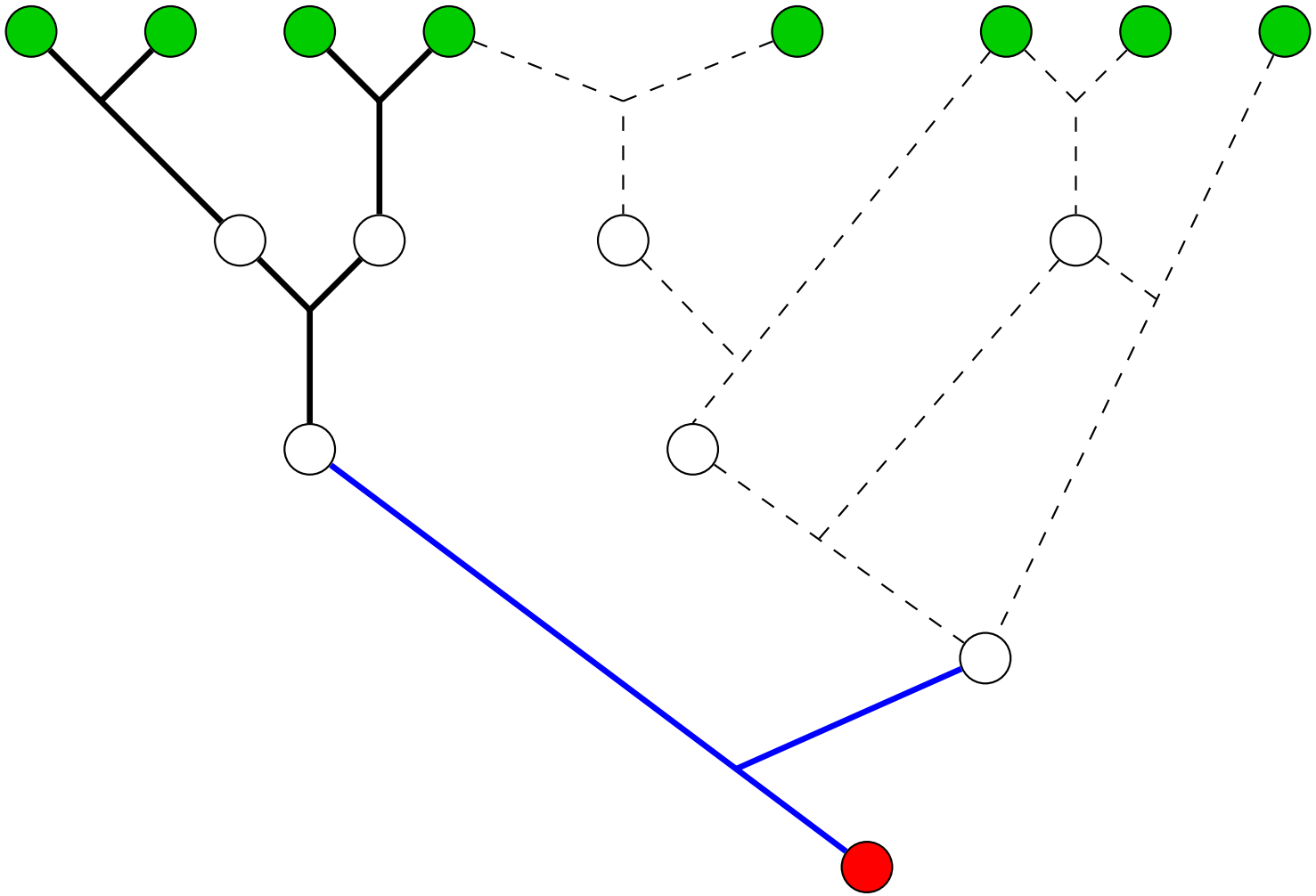
First step : remove not interesting reactions



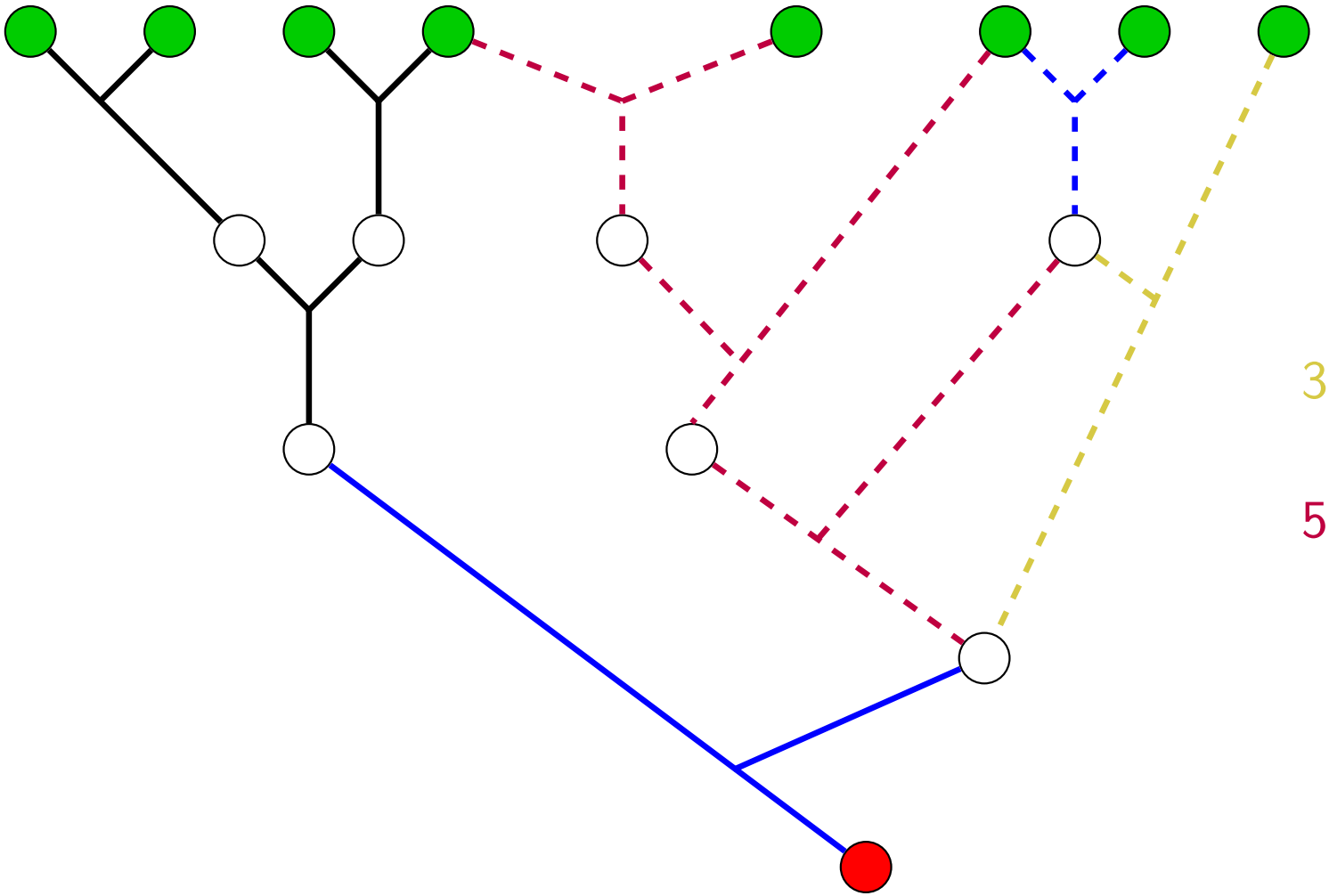
Second step : A minimal pathway from seeds to target



Second step : A minimal pathway from seeds to target



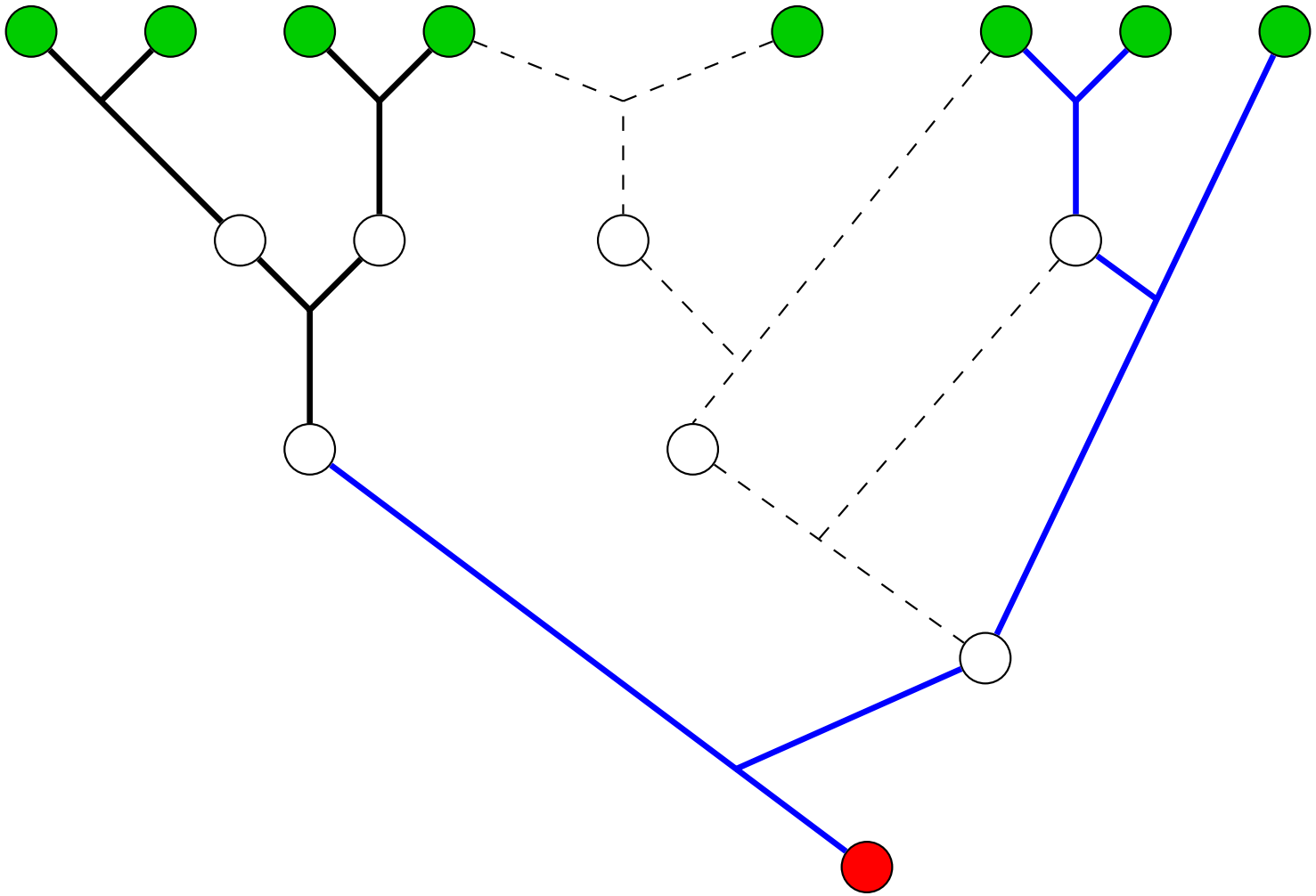
Second step : A minimal pathway from seeds to target



3 reactions

5 reactions

Second step : A minimal pathway from seeds to target



Tables of content

Change between Meneco and Menecowi

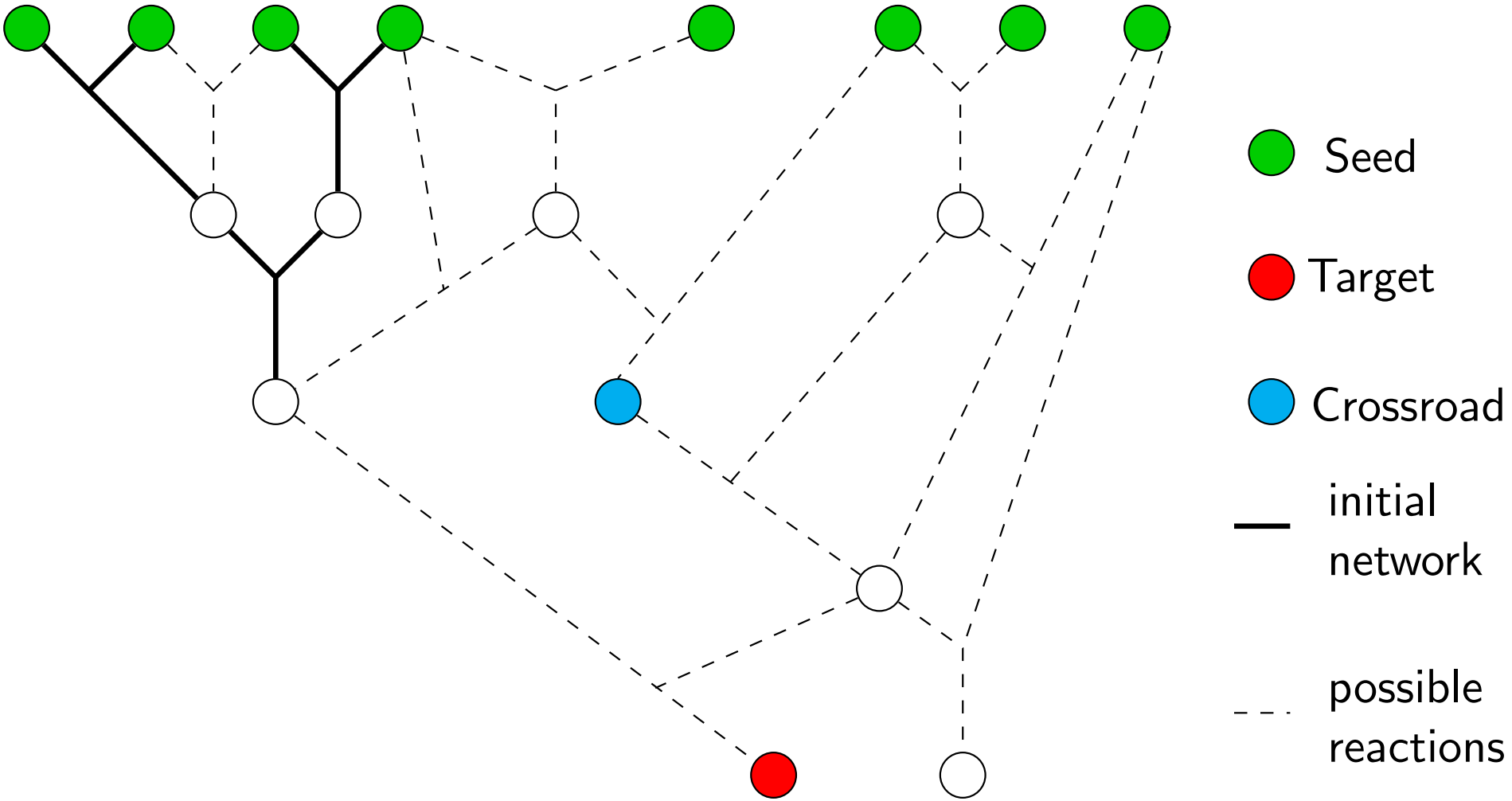
Menecowi : *Me*tabolic *ne*twork *w*ith *i*ntermediaries

Introduce the concept of crossroad metabolites in the gapfilling problem

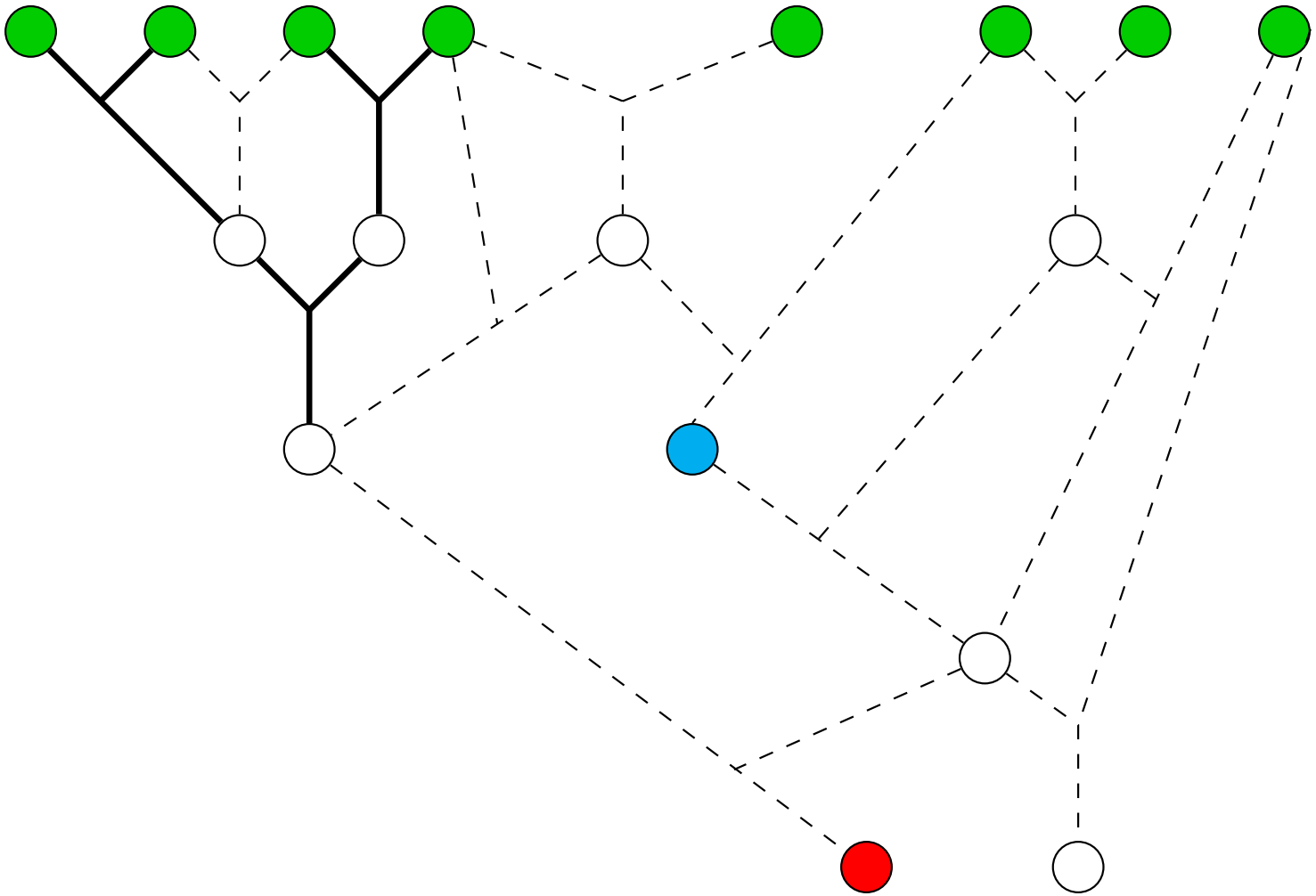
Add the minimal number of reactions taken in a database to product targets from seeds as Meneco but using crossroads metabolites which are :

- consumed
- produced
- necessary for at least one target production

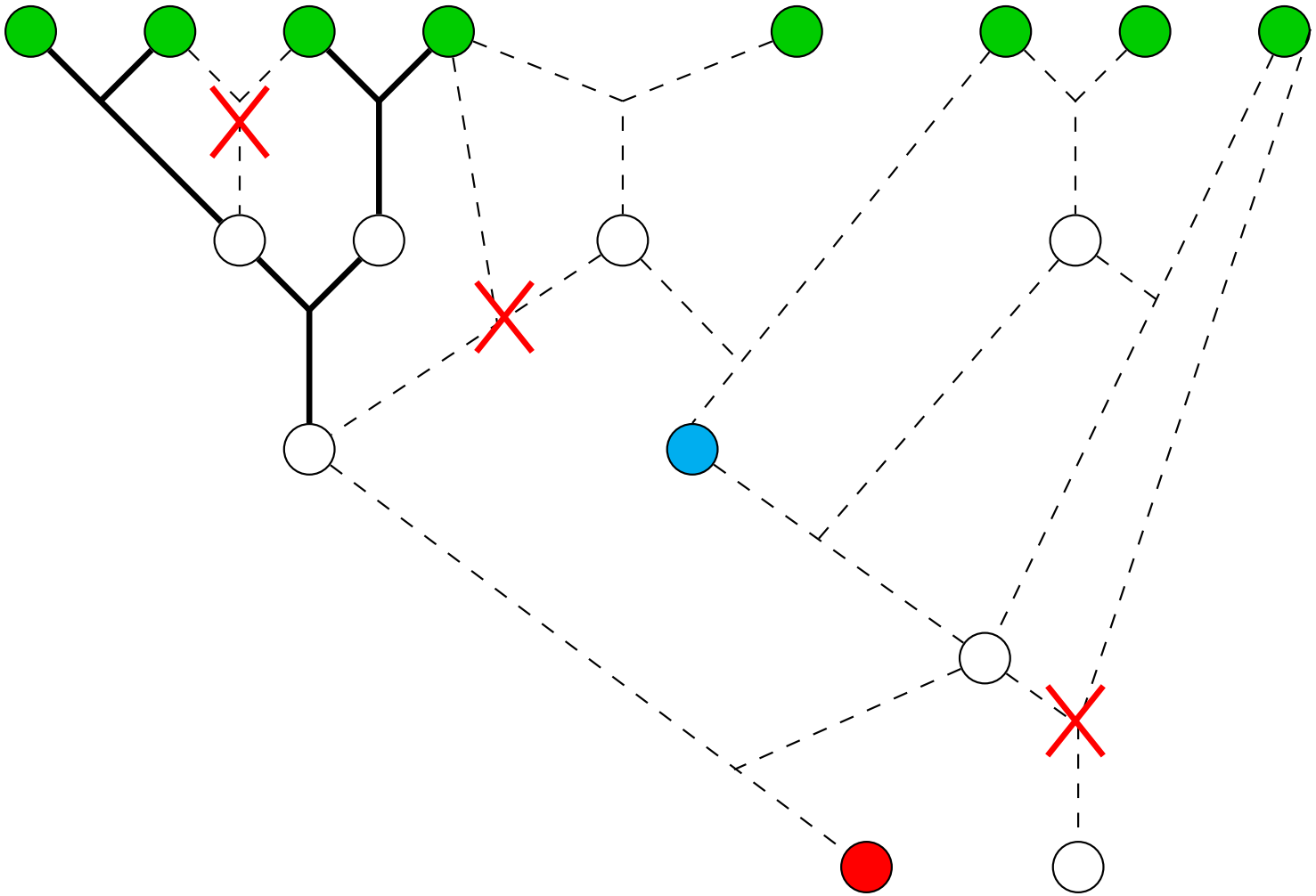
Example


Seed
Target
Crossroad
initial network
possible reactions

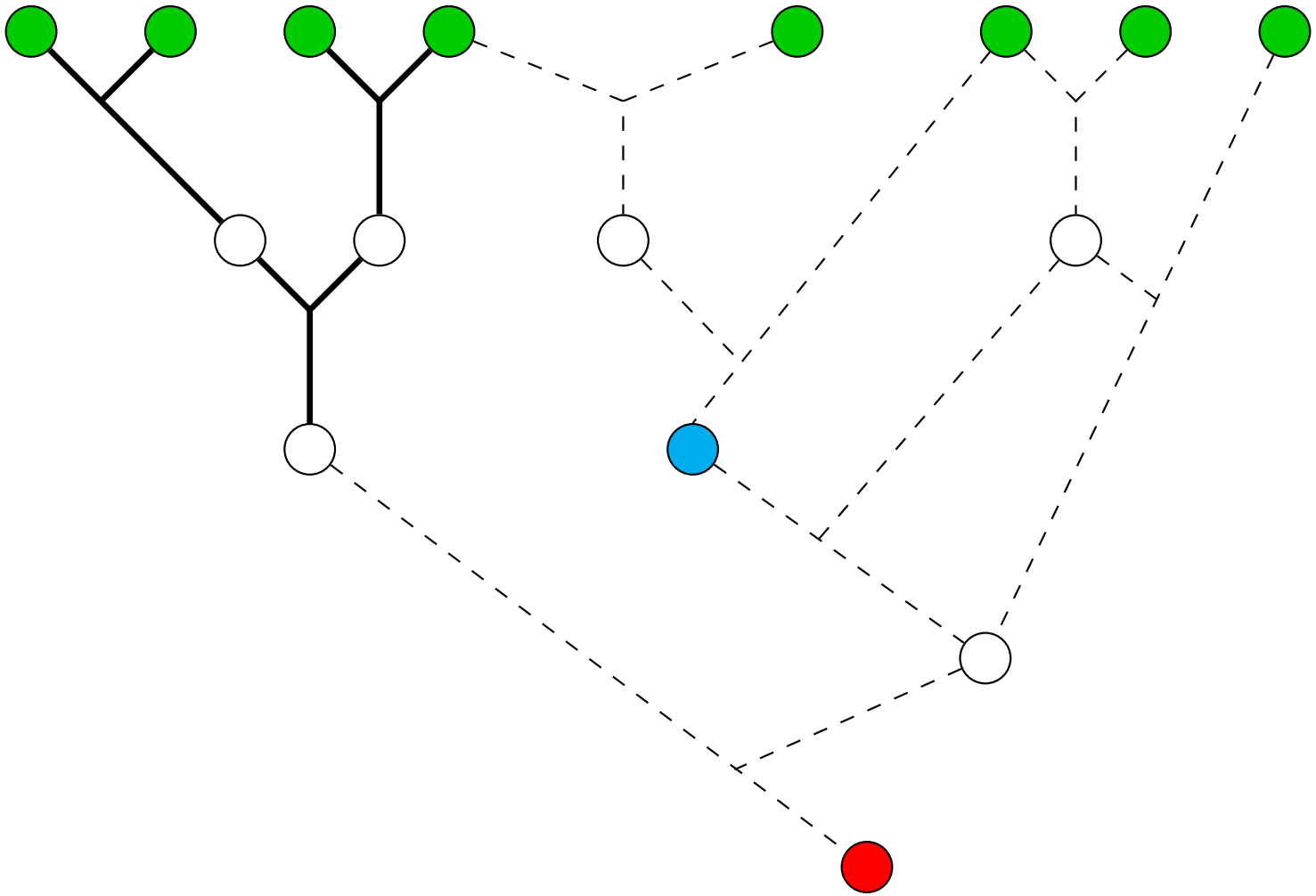
First step : remove not interesting reactions



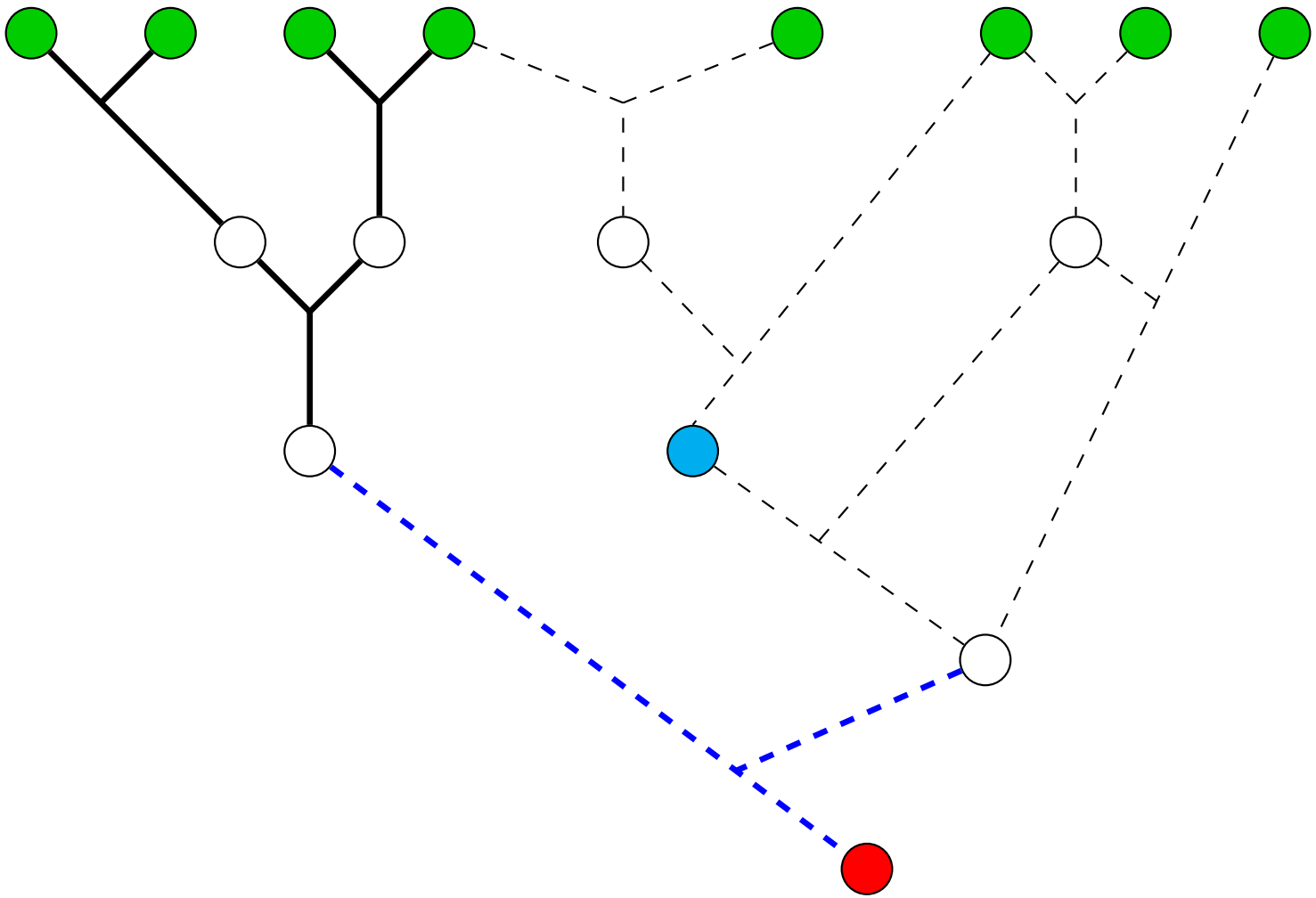
First step : remove not interesting reactions



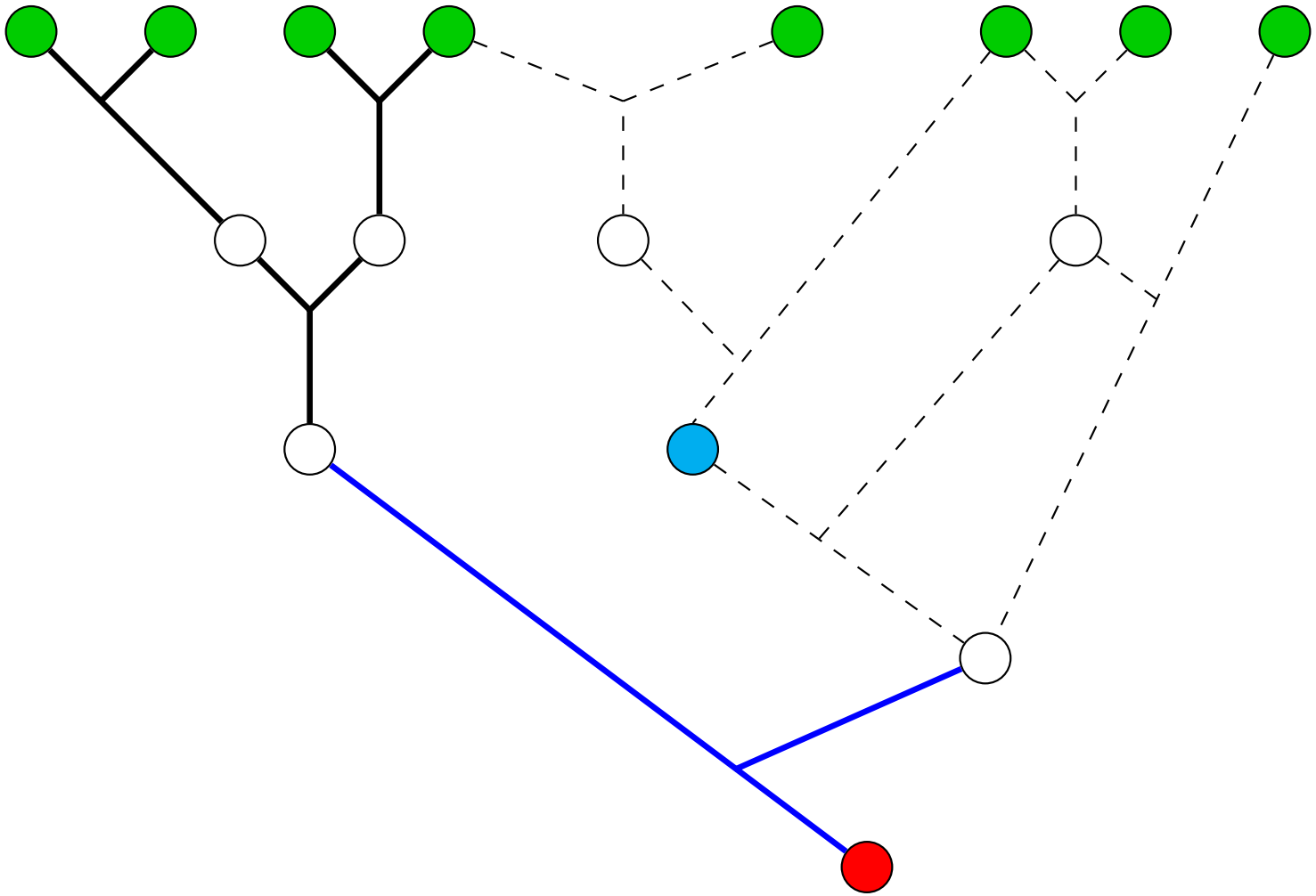
First step : remove not interesting reactions



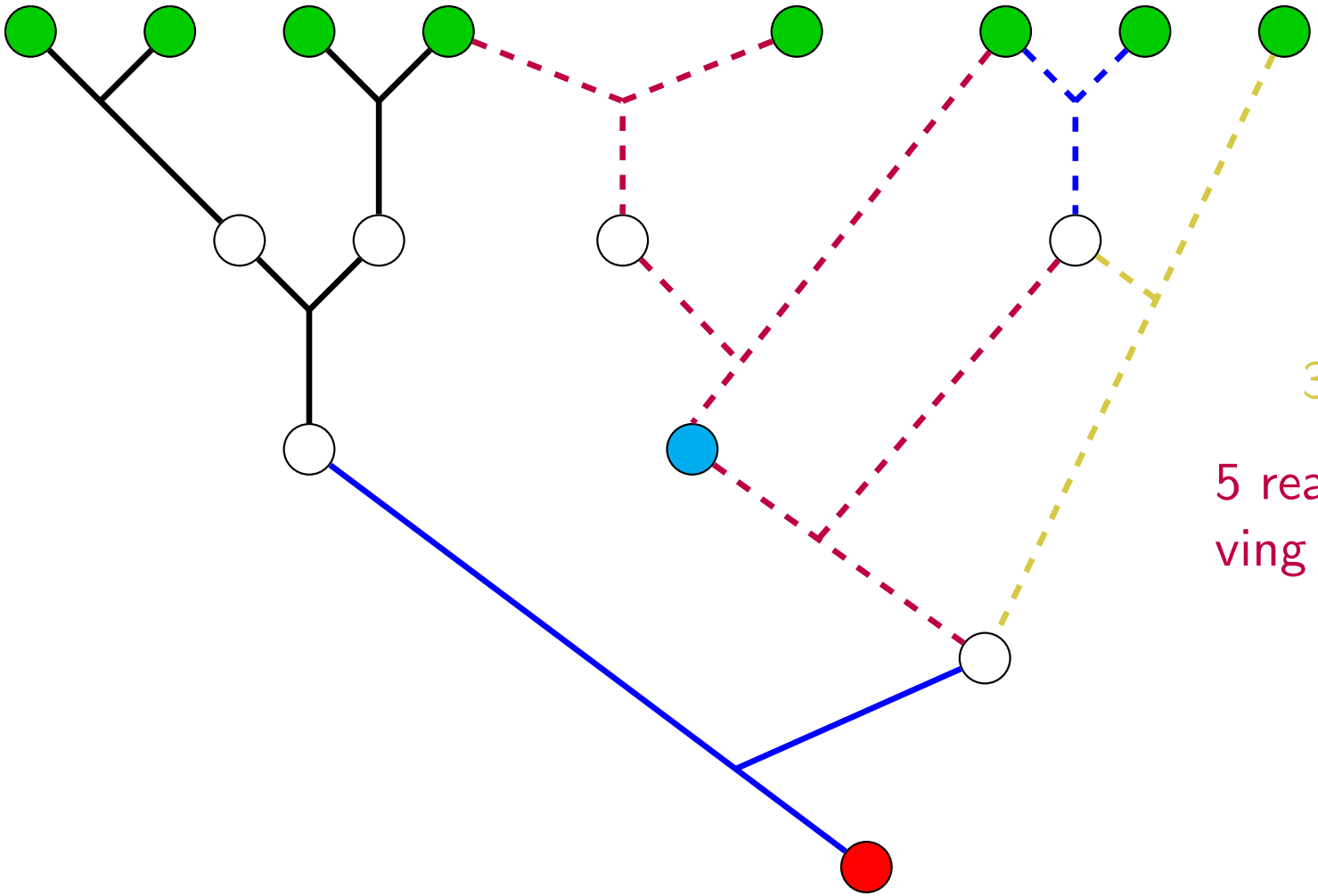
Second step : A minimal pathway from seeds to target through crossroad



Second step : A minimal pathway from seeds to target



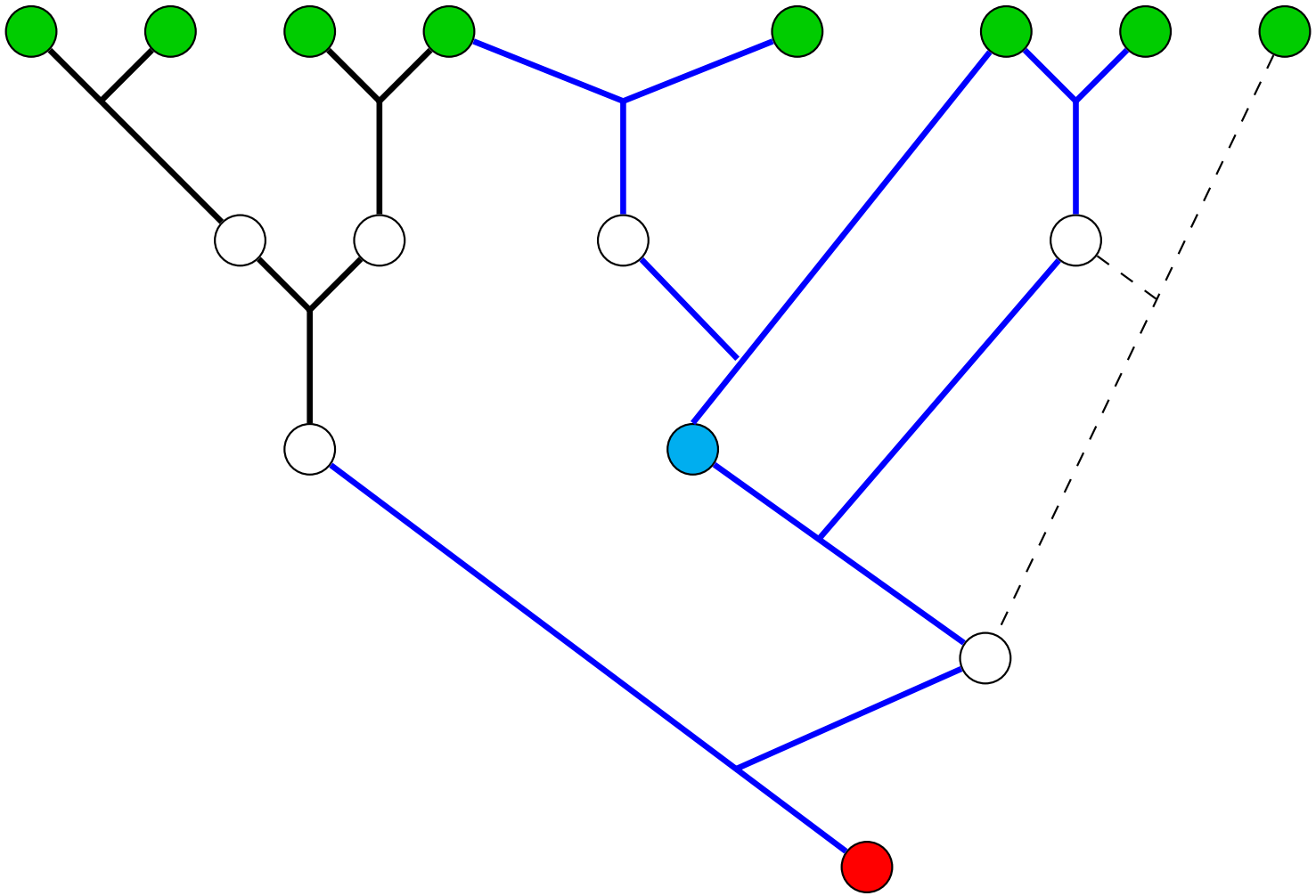
Second step : A minimal pathway from seeds to target



3 reactions

5 reactions involving crossroad

Second step : A minimal pathway from seeds to target



Tables of content

Tables of content

A crossroad metabolite

First characteristic of a crossroad metabolite

A crossroad metabolite must be :

- consumed
- produced

Mathematical definition

$$C = \{crossroads\}$$
$$\forall c \in C, \exists r \in \mathcal{R}'' \cup \mathcal{R} \mid c \in reac(r)$$
$$C \subseteq \sum_{G_{res}}(S)$$

Second characteristic of a crossroad metabolite

A crossroad metabolite must be necessary to produce at least one target

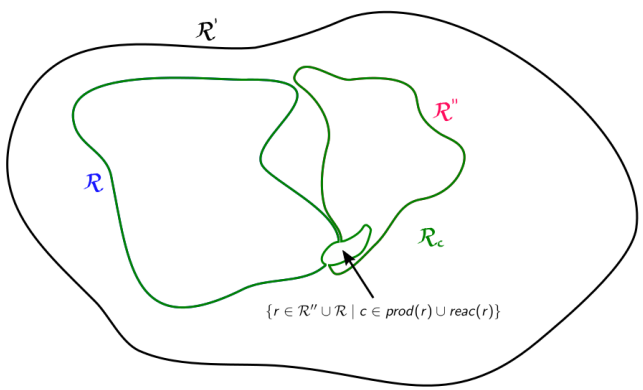
Mathematical definition

Creation of sub-graph according to a crossroad metabolite :

$G_c = (\mathcal{R}_c \cup M, E_c)$

$\mathcal{R}_c = (\mathcal{R}'' \cup \mathcal{R}) \setminus \{r \in \mathcal{R}'' \cup \mathcal{R} \mid c \in prod(r) \cup reac(r)\}$

No production of at least one target : $\forall c \in C, \exists t \in T \mid t \notin \sum_{G_c}(S)$



Tables of content

Reconstruction problem

Problem

Goal : Find a minimal set of reactions $\mathcal{R}'' \subseteq \mathcal{R}' \setminus \mathcal{R}$ such that :

- $T \subseteq \sum_{G_{res}}(S)$
- $C \subseteq \sum_{G_{res}}(S)$
- $\forall c \in C, \exists r \in \mathcal{R}'' \cup \mathcal{R} \mid c \in reac(r)$
- $\forall c \in C, \exists t \in T \mid t \notin \sum_{G_c}(S)$

with

$G_c = (\mathcal{R}_c \cup M, E_c)$

$\mathcal{R}_c = (\mathcal{R}'' \cup \mathcal{R}) \setminus \{r \in \mathcal{R}'' \cup \mathcal{R} \mid c \in prod(r) \cup reac(r)\}$

Tables of content

Resolution

Tools

- Constraints programming to encode the combinatorial optimization problem
- Solving with very efficient combinatorial solvers (ASP)

ASP : Answer Set Programming

- Writing and comprehension easy
- Enumeration of multiple solution and union
- Simple multi-objective

Complexity

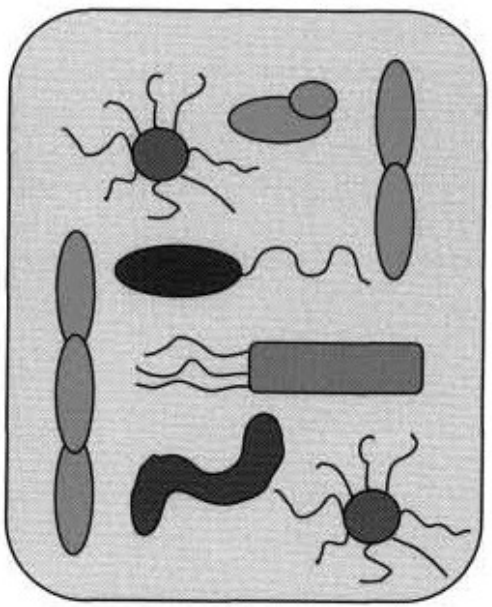
"*Metabolic networks are NP-hard to reconstruct*", Nikoloski et al.

Tables of content

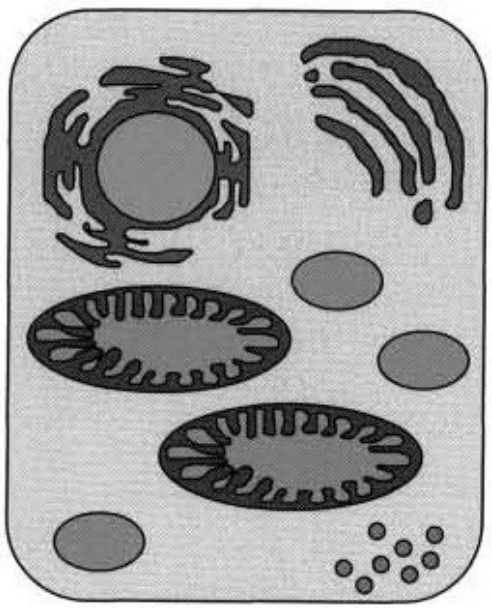
Tables of content

Creation of a benchmark dataset

yeastnet : community-driven reconstruction of the yeast metabolic network
`http://sourceforge.net/projects/yeast/?source=navbar`



Microbial community



Eukaryotic Cell

Size of the considered benchmark

Network

Reactions in initial network : 2122

Reactions in the data base : 3053

Sets of metabolites

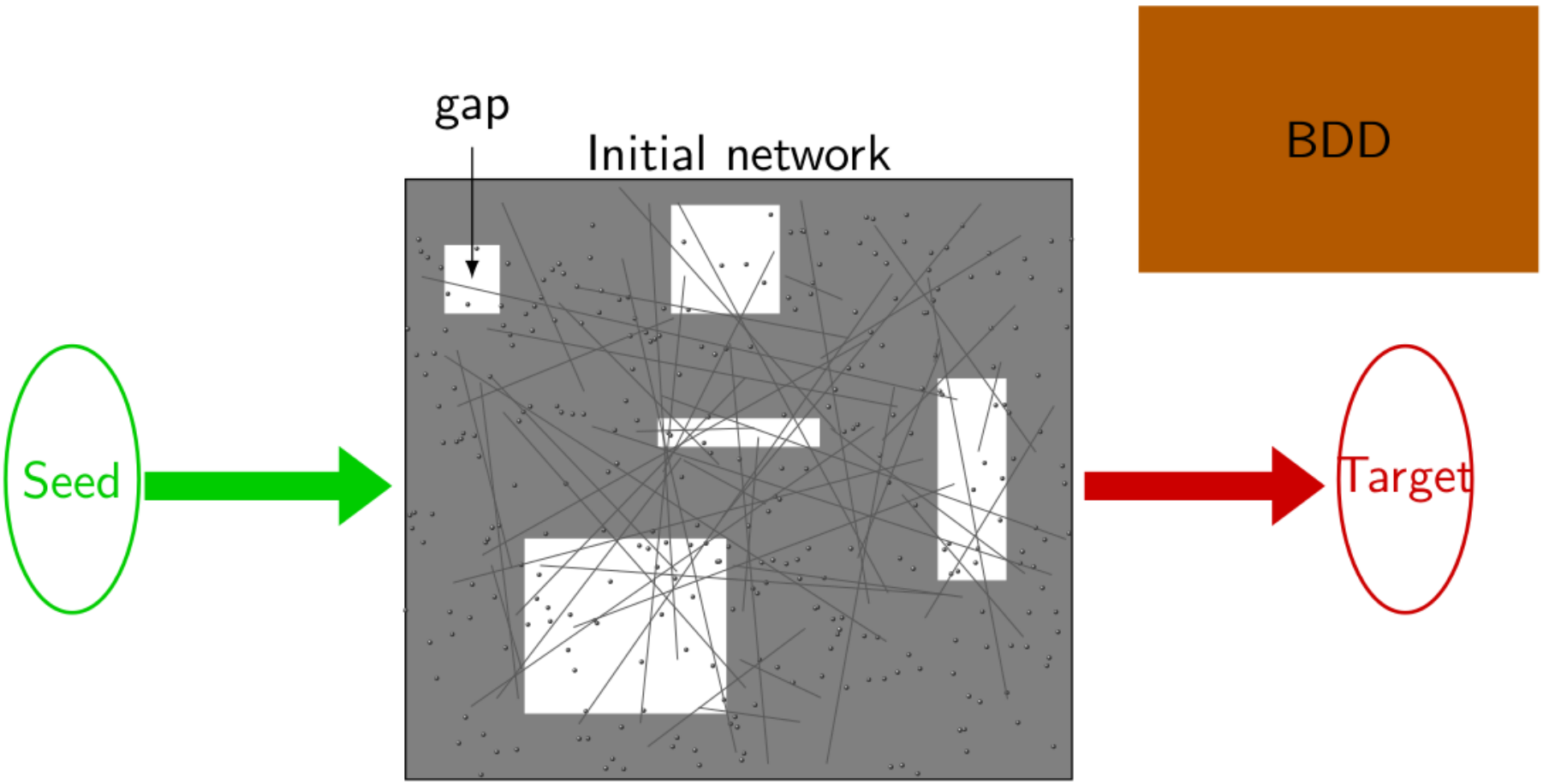
Seeds : 220

Crossroads : 223

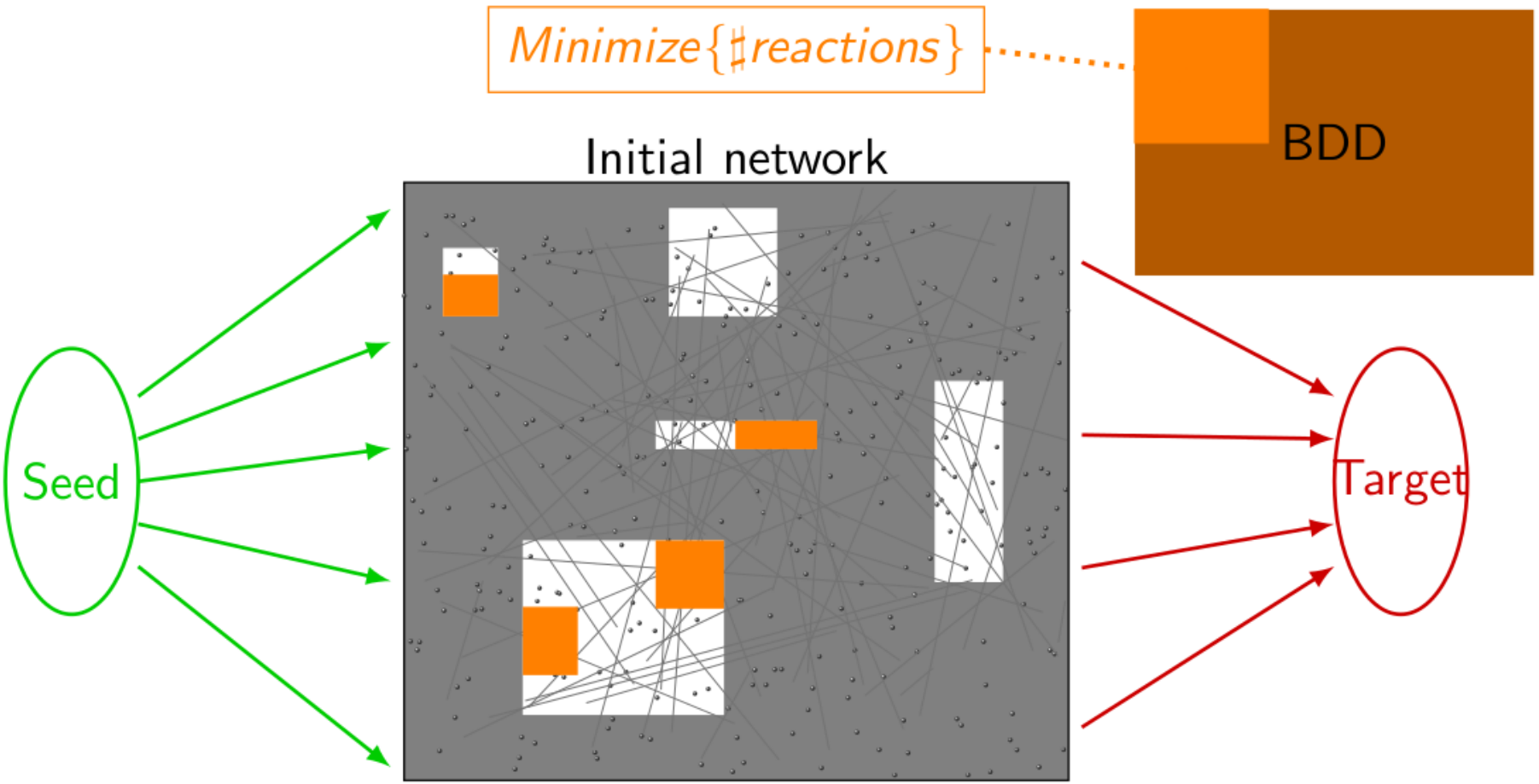
Targets : 32

Tables of content

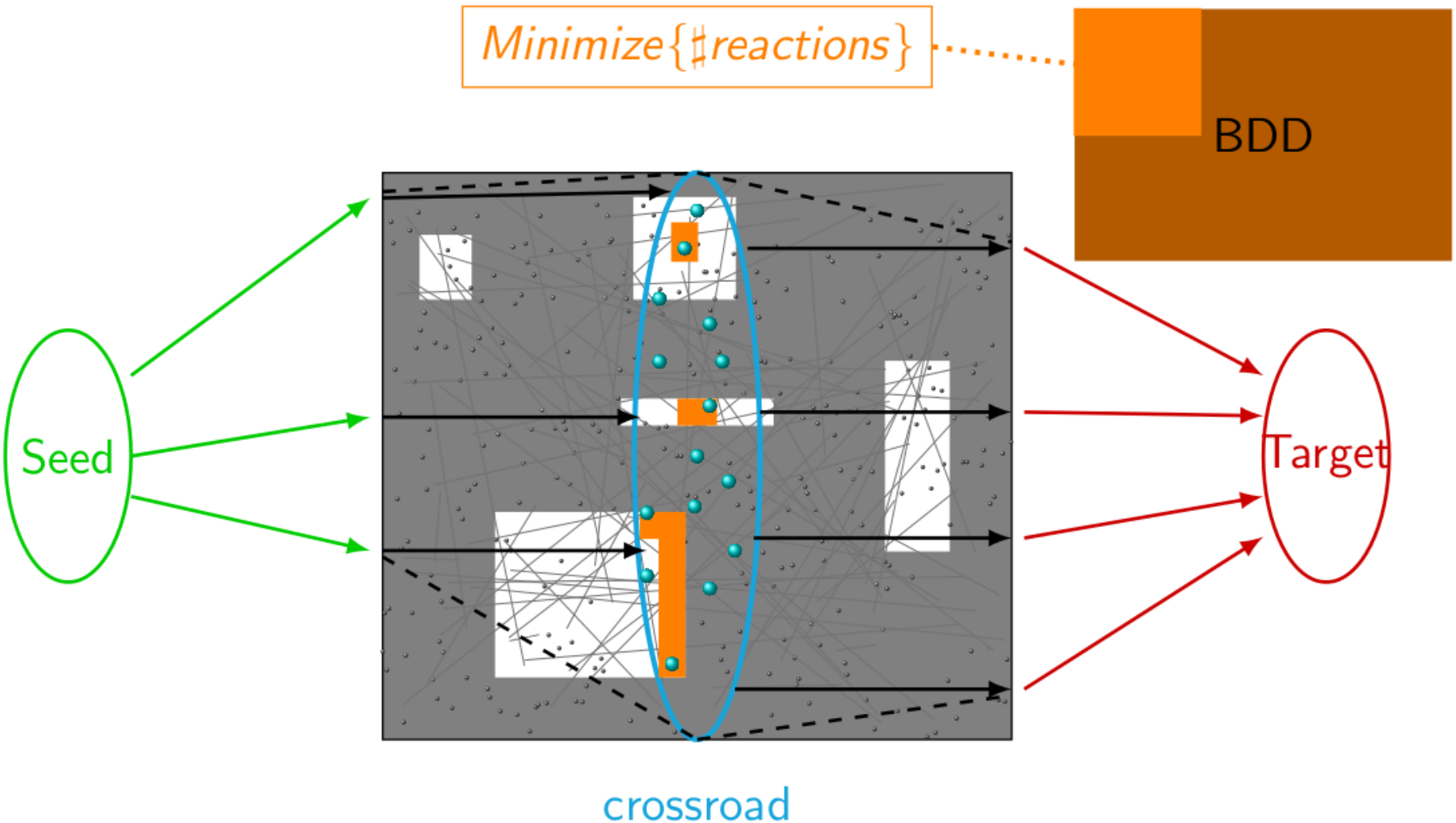
Reminder



Reminder



Reminder



First call of Menecowi

Program output

```
Computing one minimal completion to produce all targets ... done.
reconstruction impossible because of unsatisfiable model
```

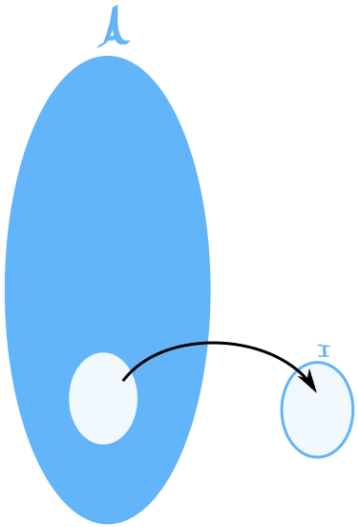
The model is unsatisfiable : there is no pathway from seeds to targets passing through ALL crossroad metabolites

Modification of the problem

Modification

Use a maximal number of crossroads to reconstruct the metabolic map taken in a set of potential metabolites.

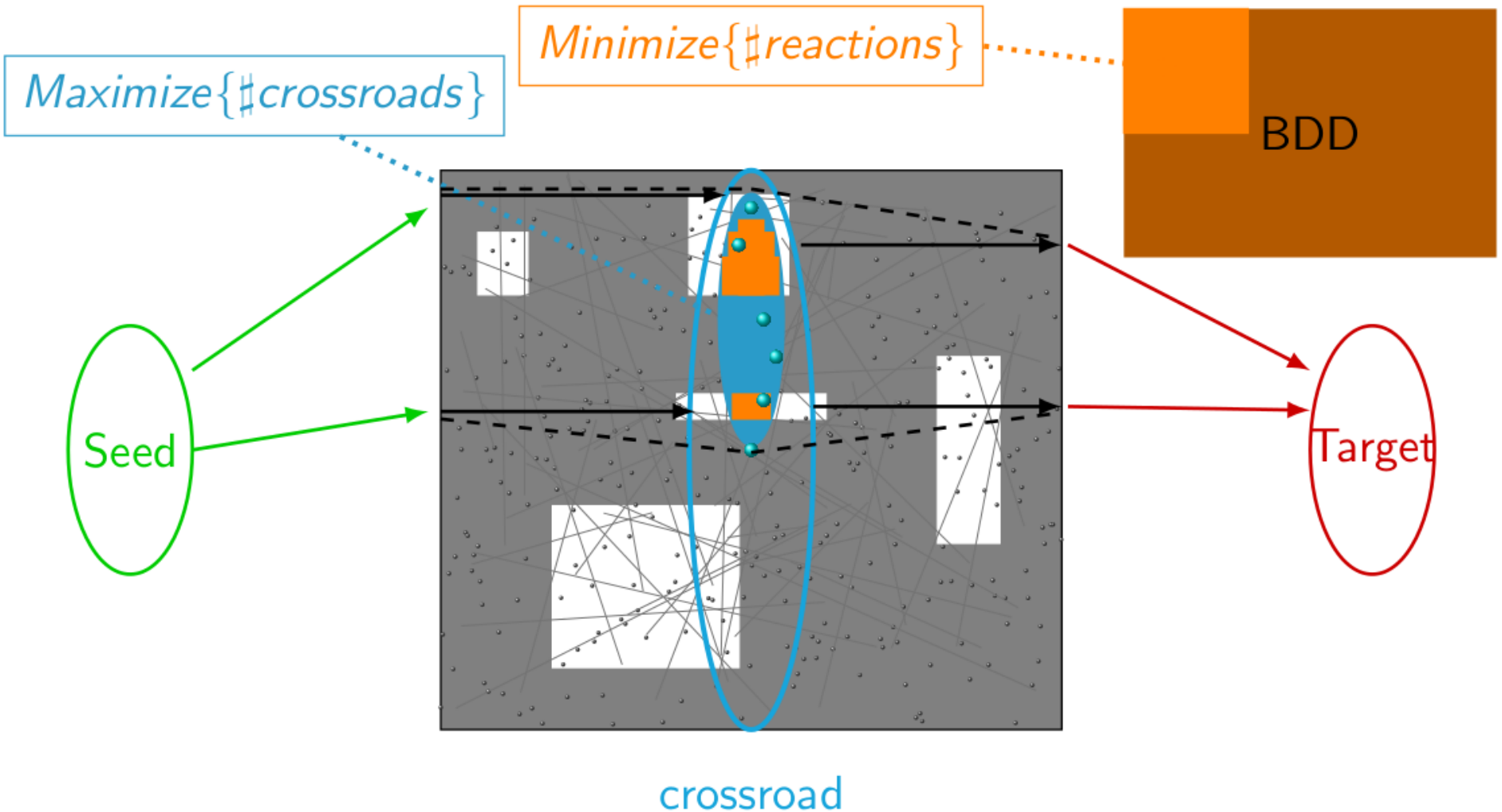
$$Set(Functional\ \ crossroads) \subset Set(Crossroads)$$



Two optimizations :

1°) Maximization of crossroads

2°) Minimization of added reactions


Maximize{♯crossroads}
Minimize{♯reactions}
BDD
Seed
Target
crossroad

New results

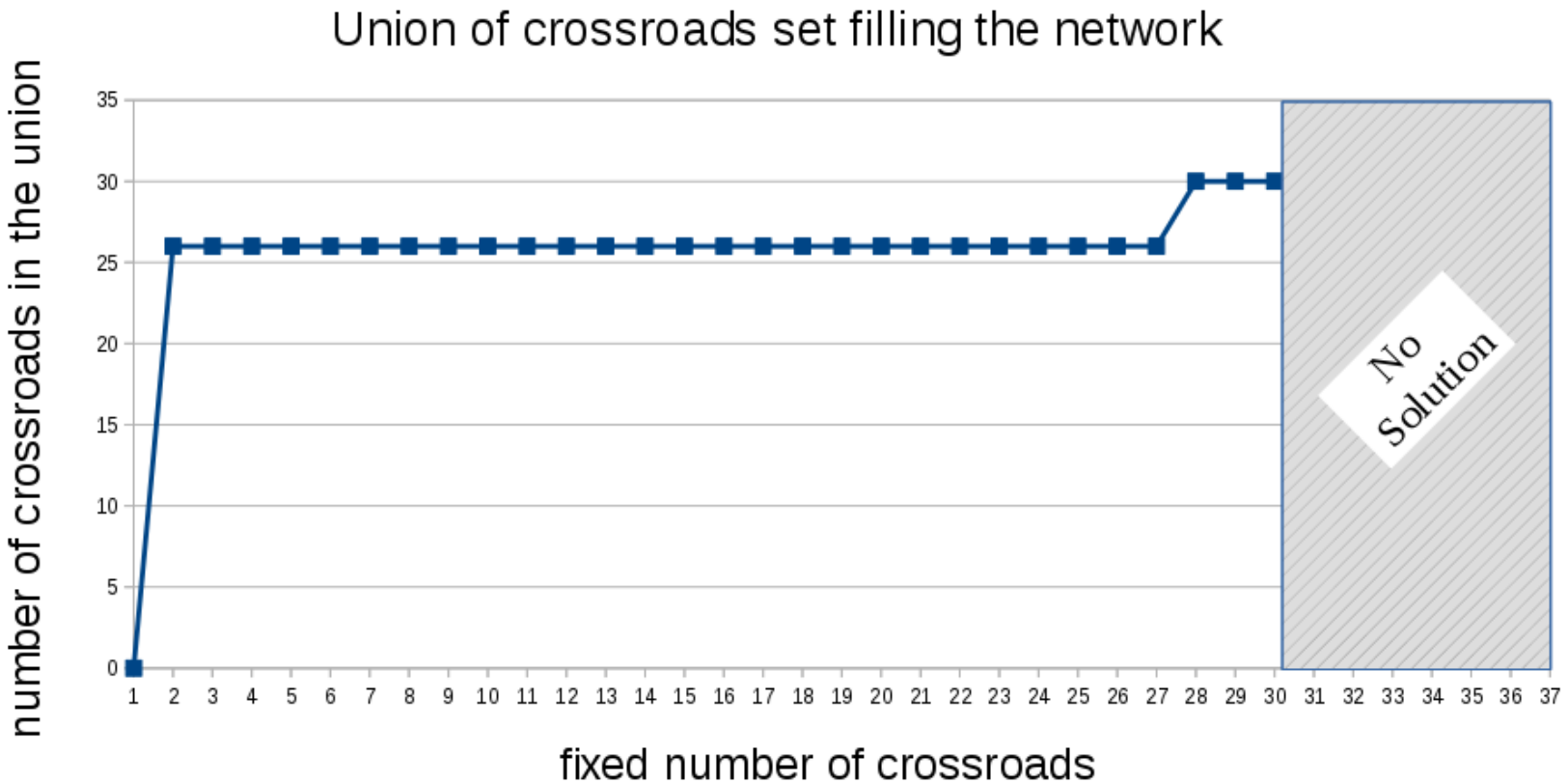
Shell

```
Computing one minimal completion to produce all targets ... done.
  minimal size = 10  with  1  accumulators ( 7 initial accumulators)
  "r_4036"
  "r_3971"
  "r_3981"
  "r_4003"
  "r_1363"
  "r_1729"
  "r_1616"
  "r_3965"
  "r_3978"
  "r_3995"
  "t_0338"
```

Answer

To restore the production of 32 targets, we can use at most 30 crossroads within the candidate crossroads. Then, at least 17 reactions shall be added to the model.

Number of crossroads evolution



Observation

Threshold at 26 crossroads before the maximal value of 30 crossroads

Impact of the crossroads number

Test

Study the solution according to number of crossroads used

Problem modification

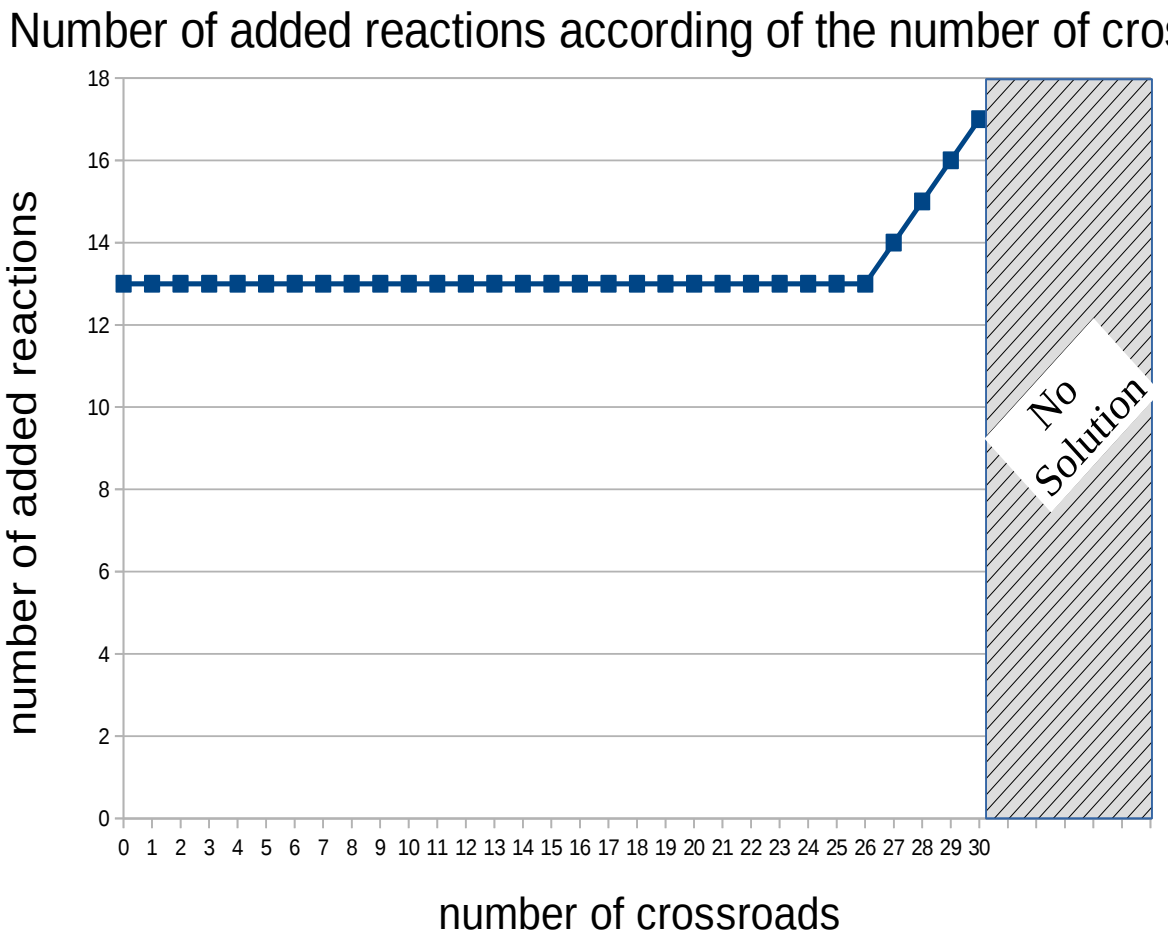
Fix the number of crossroads taken in the set of 223 candidate crossroads.

Output

Program output :

- number of reactions required to restore the production of the target while using a fixed number of crossroads
- union of all solutions : reactions appearing in at least one completion for at least one set of XX crossroad metabolites.

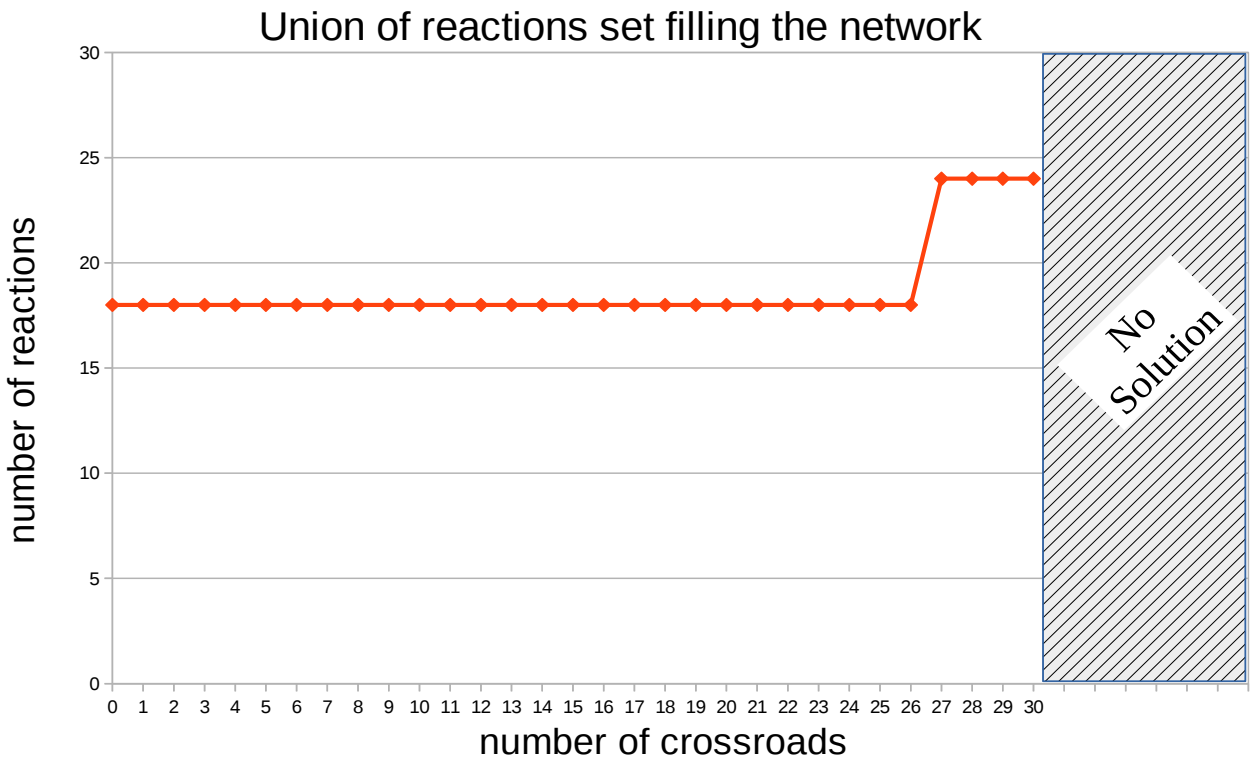
Number of reactions added according to crossroads



Observation

Threshold at 13 added reactions until we used 27 crossroads in the solution.

Union of all solutions



Observation

Several sets possible of added reactions to fill the network

Subsets analysis

Test

Solution according to a pair (reaction, crossroad) which must be in the solution.

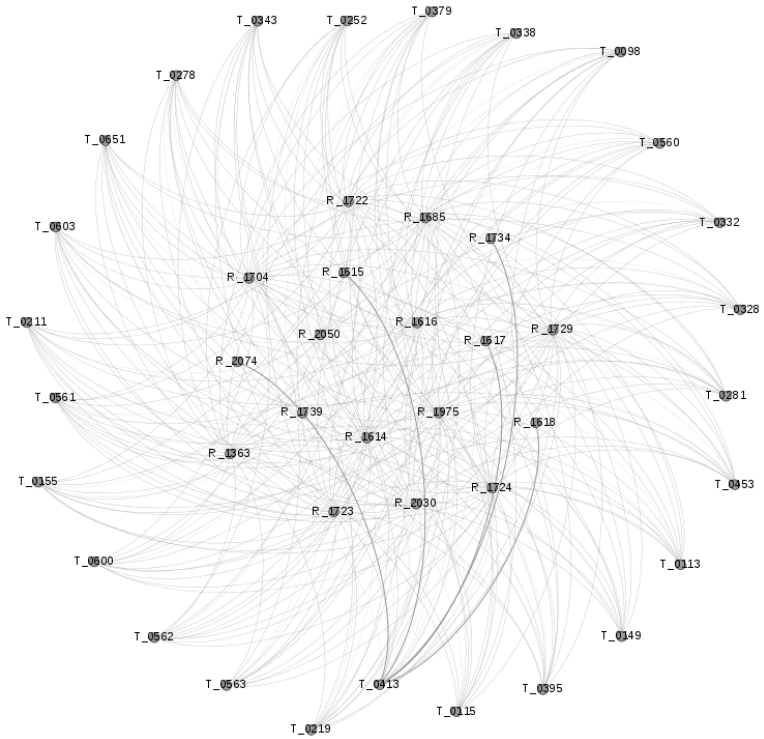
Problem modification

Fix the pair (reaction and crossroad) before reconstruction

Output

Program output :

- union of solutions
- number of solutions



An edge (R,C) = there exists a solution with the reaction R and the crossroads C

Intuition

Complete graph

Whatever the pair (reaction,crossroad) fixed, there is a solution using it.

Thanks for your attention

Questions ?