Finding Minimal Reaction Sets in Large Metabolic Pathways

Takehide Soh[*1] and Katsumi Inoue[*2*1]

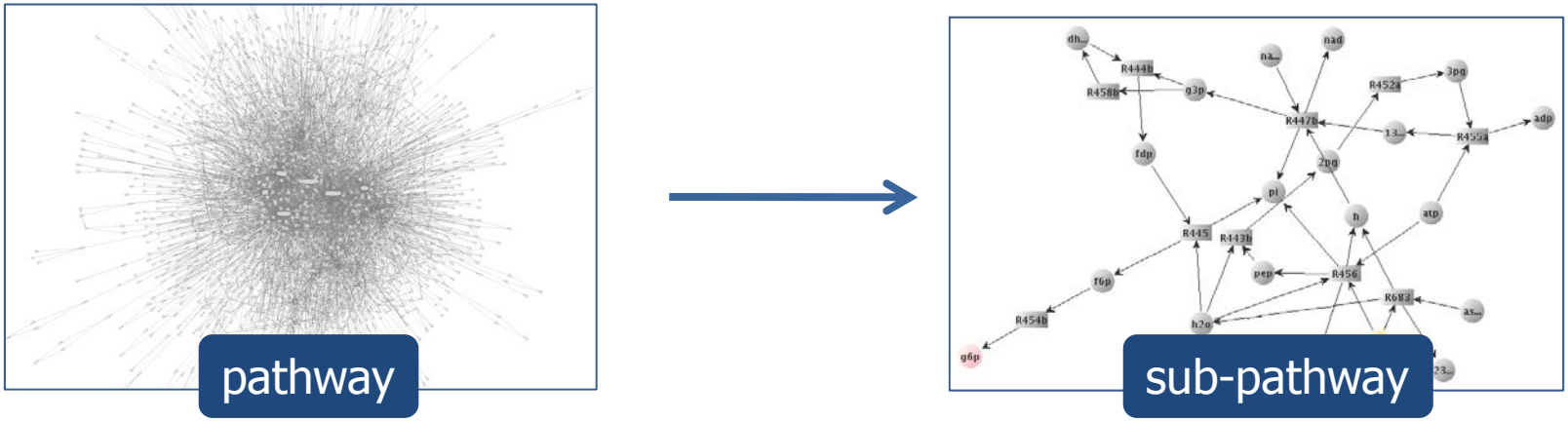
*1) The Graduate School for Advanced Studies
*2) National Institute of Informatics

WCB10@Edinburgh

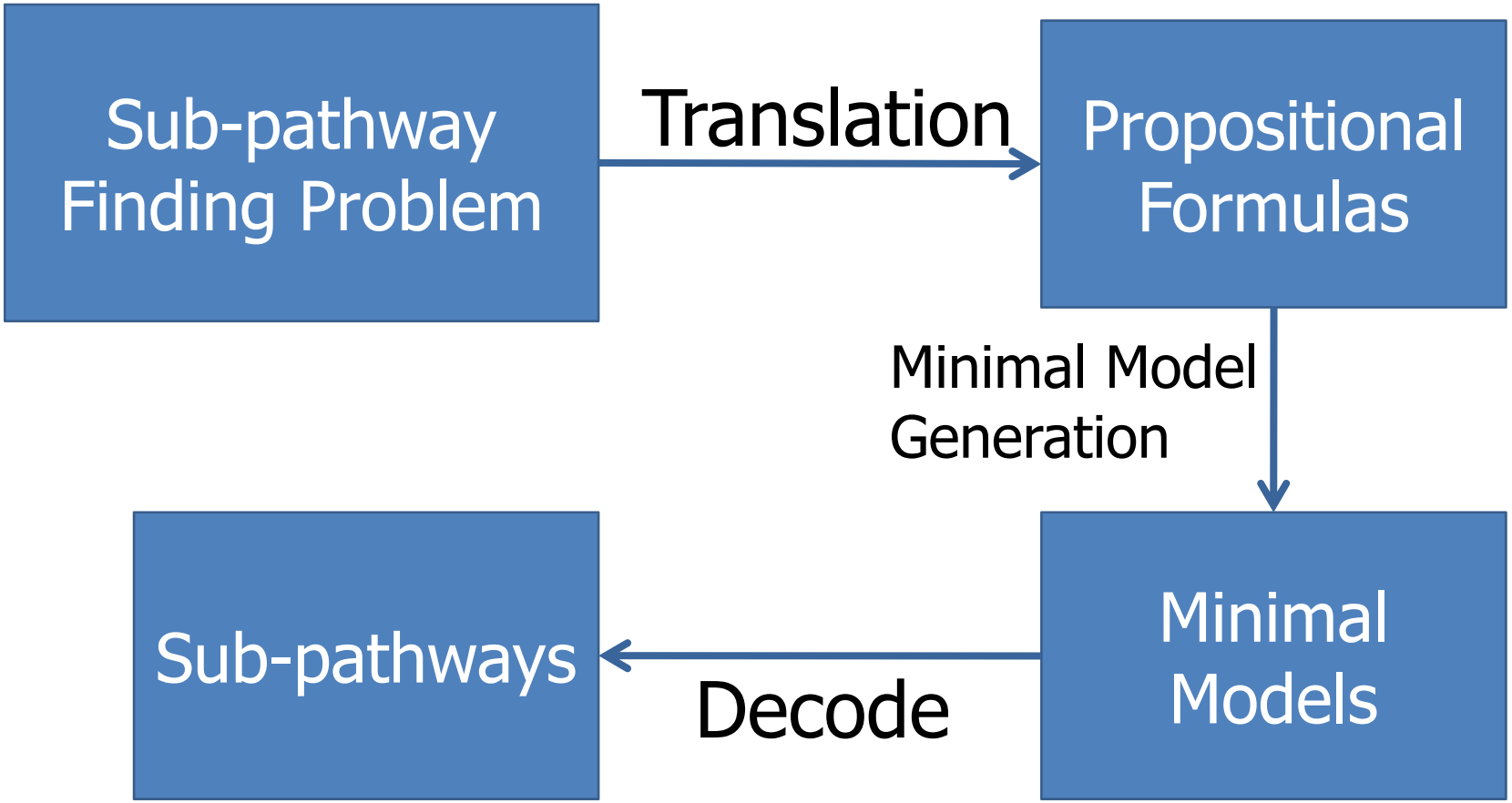
Background

- Interactions of metabolites are represented in a network called a (metabolic) pathway.

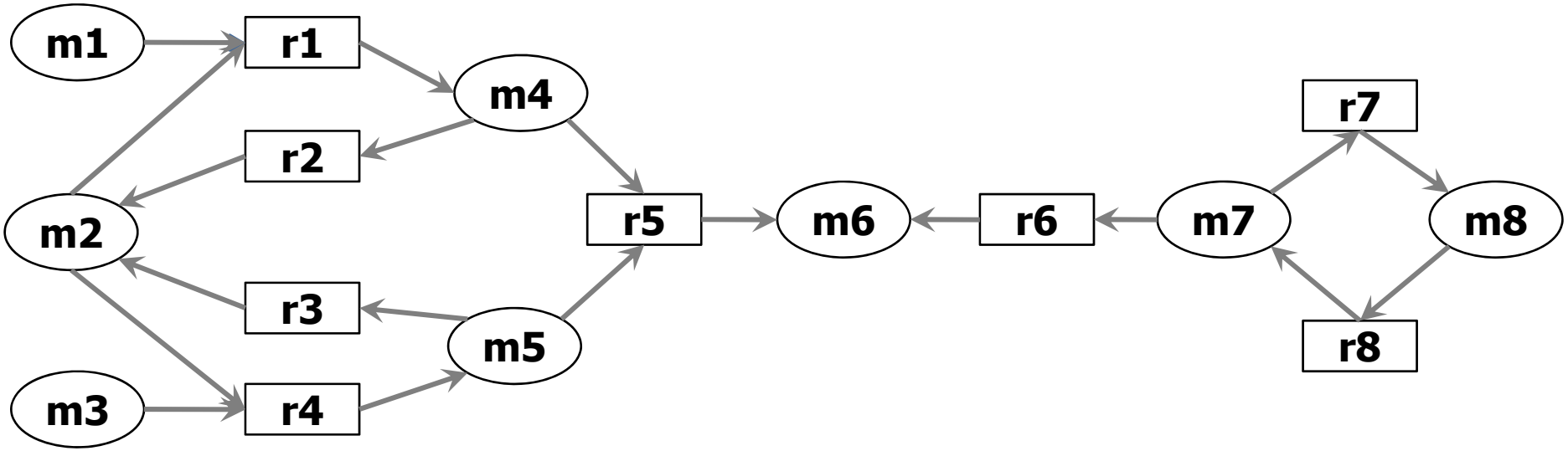
- It is important to identify a specific part (sub-pathway) of a given pathway, which provides a vital function.



Identifying Necessary Reactions by Minimal Model Generation

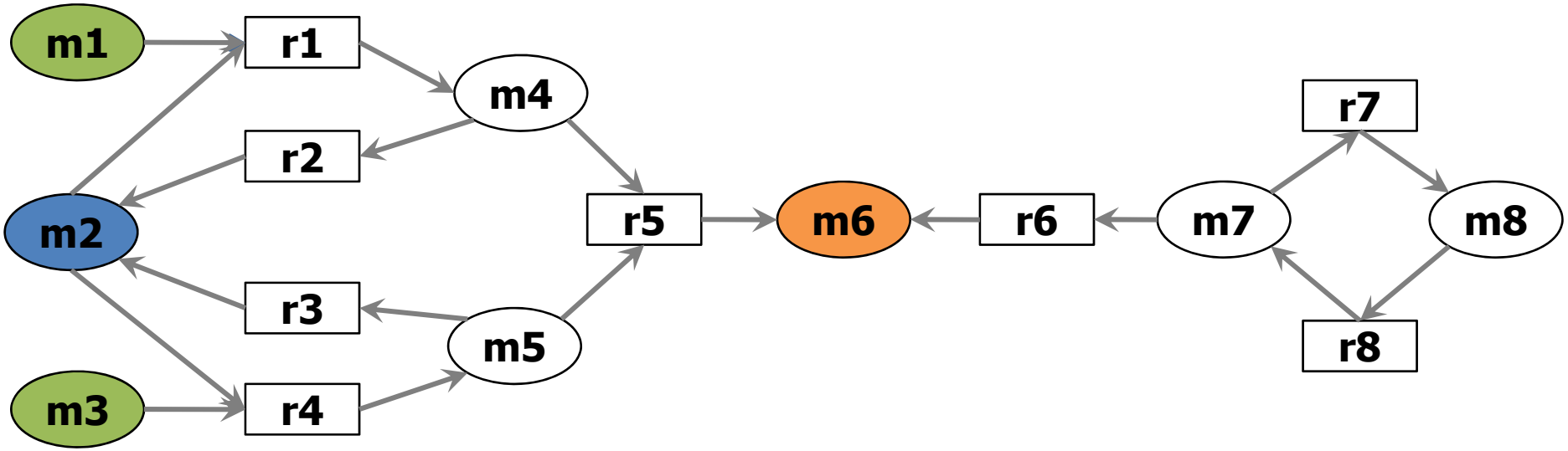


Pathway Instance (1/2)



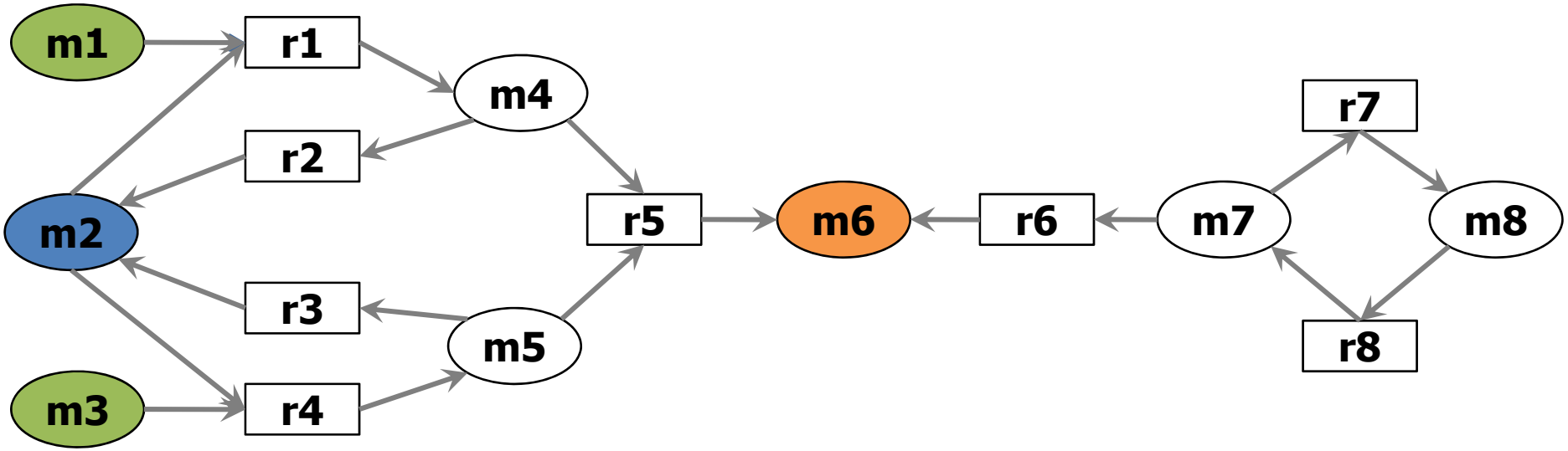
- A set of metabolites $M = \{m_1, m_2, \ldots, m_i\}$
- A set of reactions $R = \{r_1, r_2, \ldots, r_j\}$
- A set of arcs $A \subseteq (R \times M) \cup (M \times R)$
- A pathway is given by a bipartite directed graph $G = (M, R, A)$

Pathway Instance (2/2)



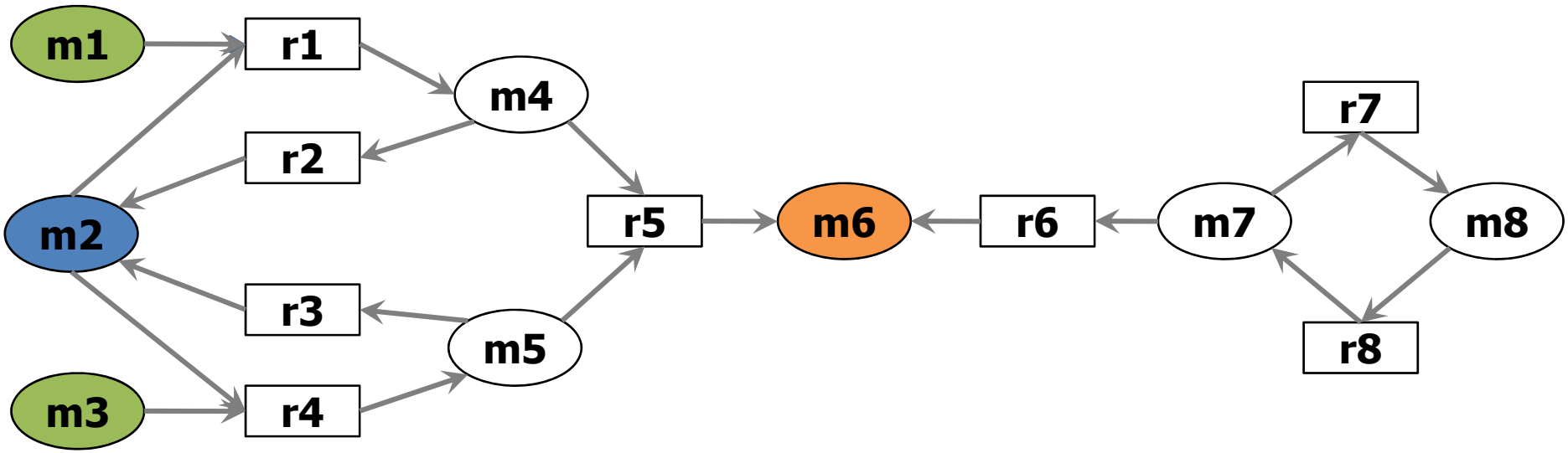
- $M_I \subset M$: a set of initial metabolites.
- $M_S \subset M$: a set of source metabolites.
- $M_T \subset M$: a set of target metabolites.
- M_I represents universal metabolites which are always producible in pathways.
- A pathway instance is given as $\pi = (M, R, A, M_I, M_S, M_T)$.

Reactants and Products



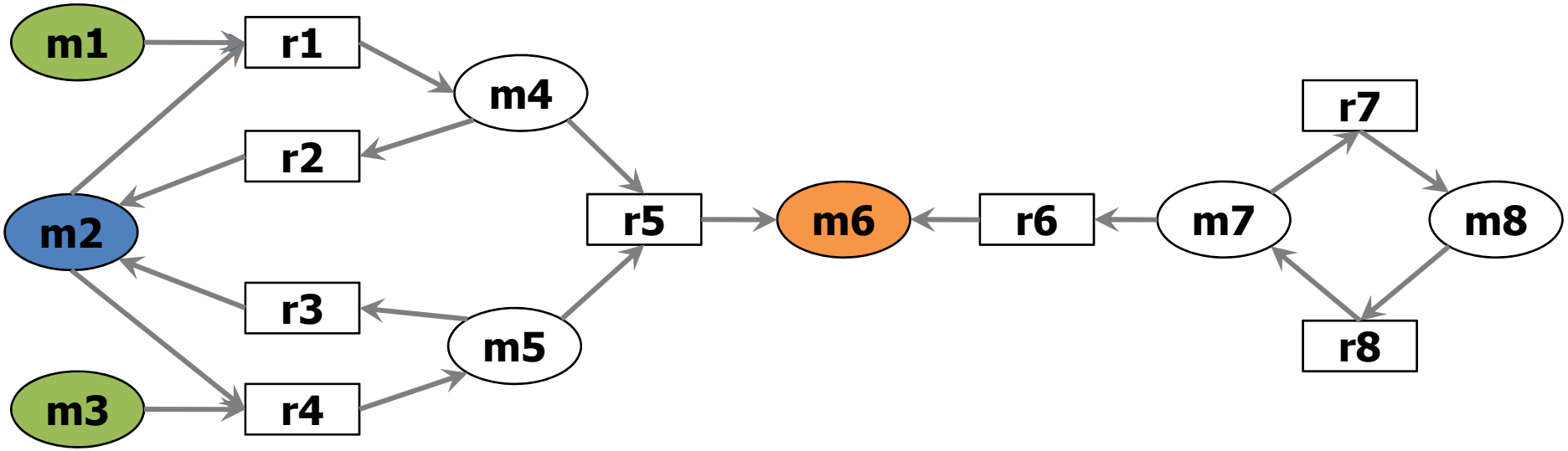
- $m \in M$ is called a reactant of $r \in R$ if $(m, r) \in A$.
- $m \in M$ is called a product of $r \in R$ if $(r, m) \in A$.
- $s : R \rightarrow 2^M$, $s(r) = \{m \in M | (m, r) \in A\}$
- $p : R \rightarrow 2^M$, $p(r) = \{m \in M | (r, m) \in A\}$
- $s^{-1}(m)$, $p^{-1}(m)$ represents the set of reactions which consume and produce m, respectively.

Producibility and Activatability



- $M' \subset M$
- $m \in M$ is producible at t=0 from M' if $m \in M'$ holds.
- $r \in R$ is activatable at t=e (0<e) from M' if $\forall m \in s(r)$, m is producible at a time t=e-1 from M'.
- $m \in M$ is producible at t=e (0<e) from M' if $m \in p(r)$ at least one reaction r which is activatable at t=e from M'.

Producibility and Activatability



- If r is activatable at t=e then r is activatable at t=e+1.
- If m is producible at t=e then r is producible at t=e+1.

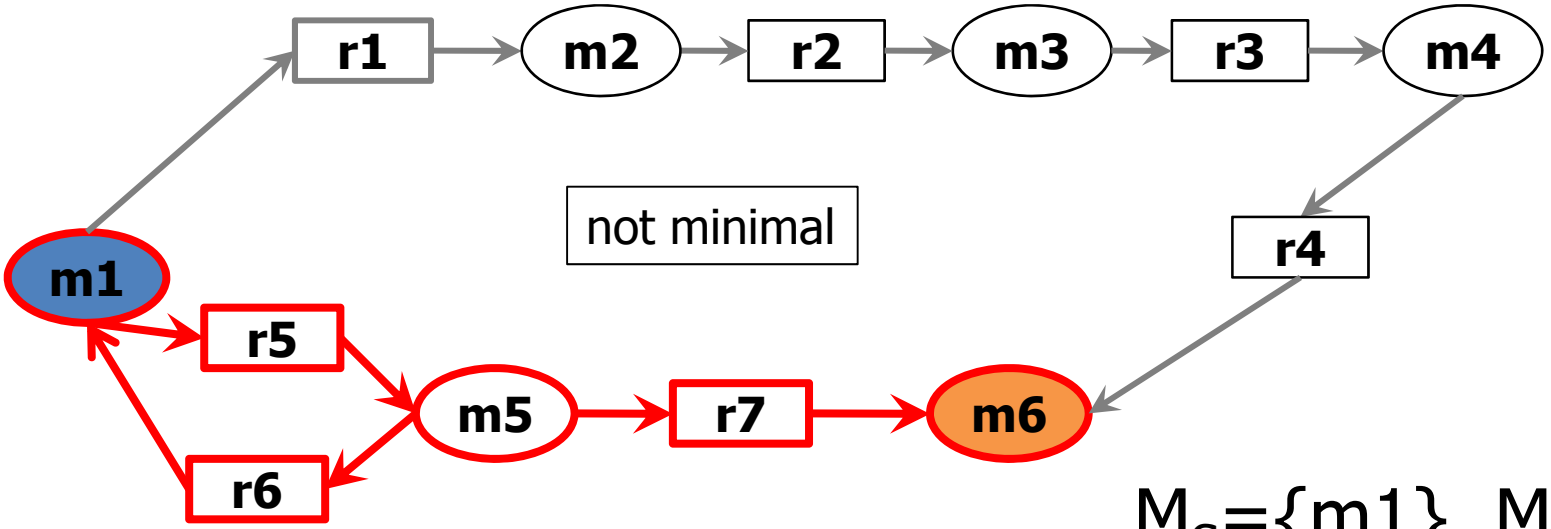
Minimal Sub-pathway

- π : a pathway instance (M, R, A, M_I, M_S, M_T)

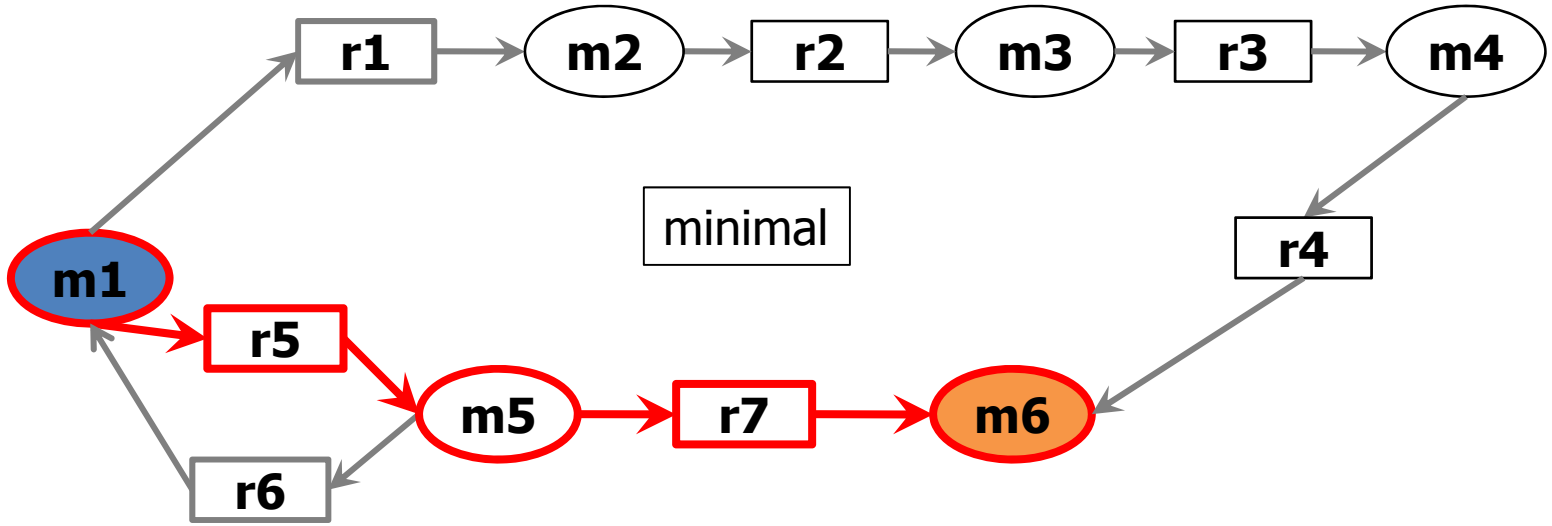
- A sub-graph $G'=(M', R', A')$ of $G=(M, R, A)$ is a sub-pathway of π if:
  i. $M_S \subset M'$ and $M_T \subset M'$
  ii. $\forall m \in M'$, m is producible from $M_I \cup M_S$ at $t \geqq e$ for some $e \in Z^+$
  iii. $\forall r \in R'$, r is activatable from $M_I \cup M_S$ at $t \geqq e$ for some $e \in Z^+$ and $p(r) \in M'$.

- A sub-pathway is called a minimal sub-pathway if:
  iv. There is no sub-pathway G" such that $G'' \subset G'$

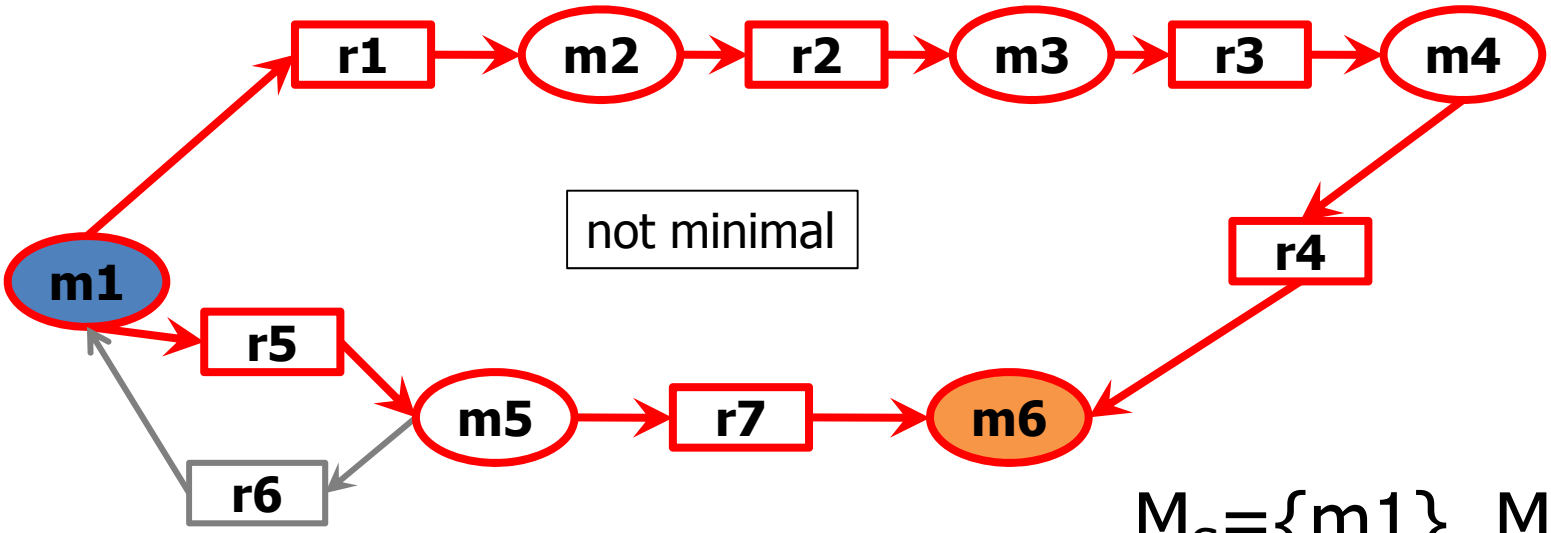
Minimal Sub-pathway (Example.)



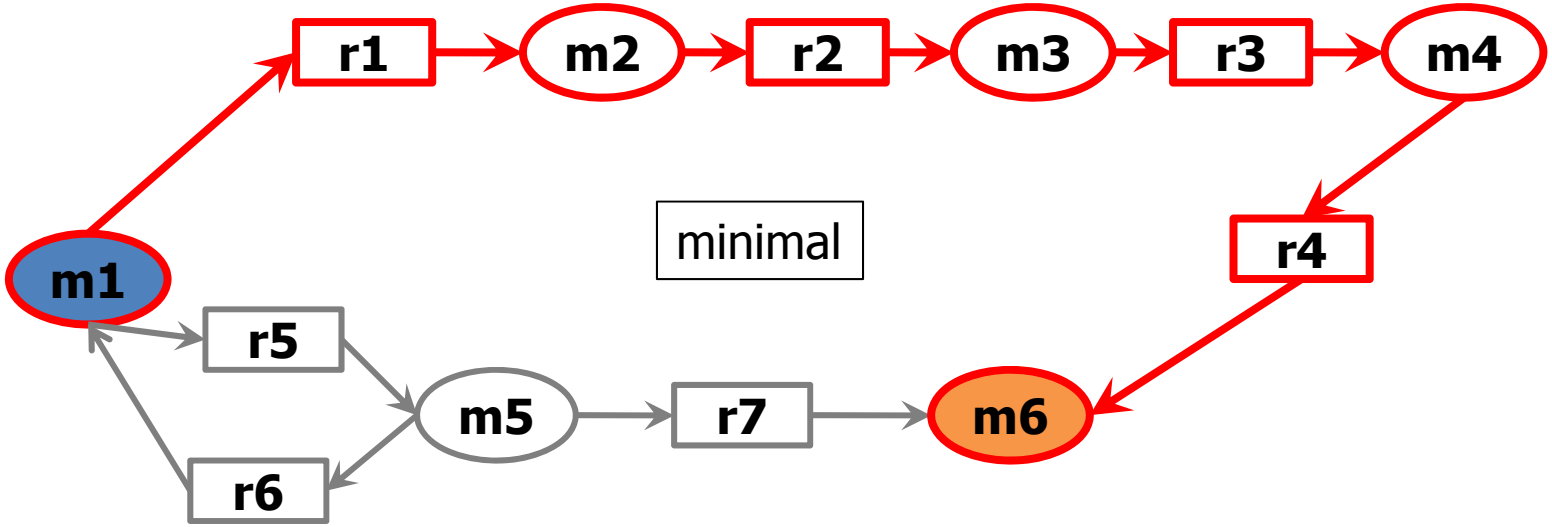
$M_S=\{m1\}$, $M_T=\{m6\}$



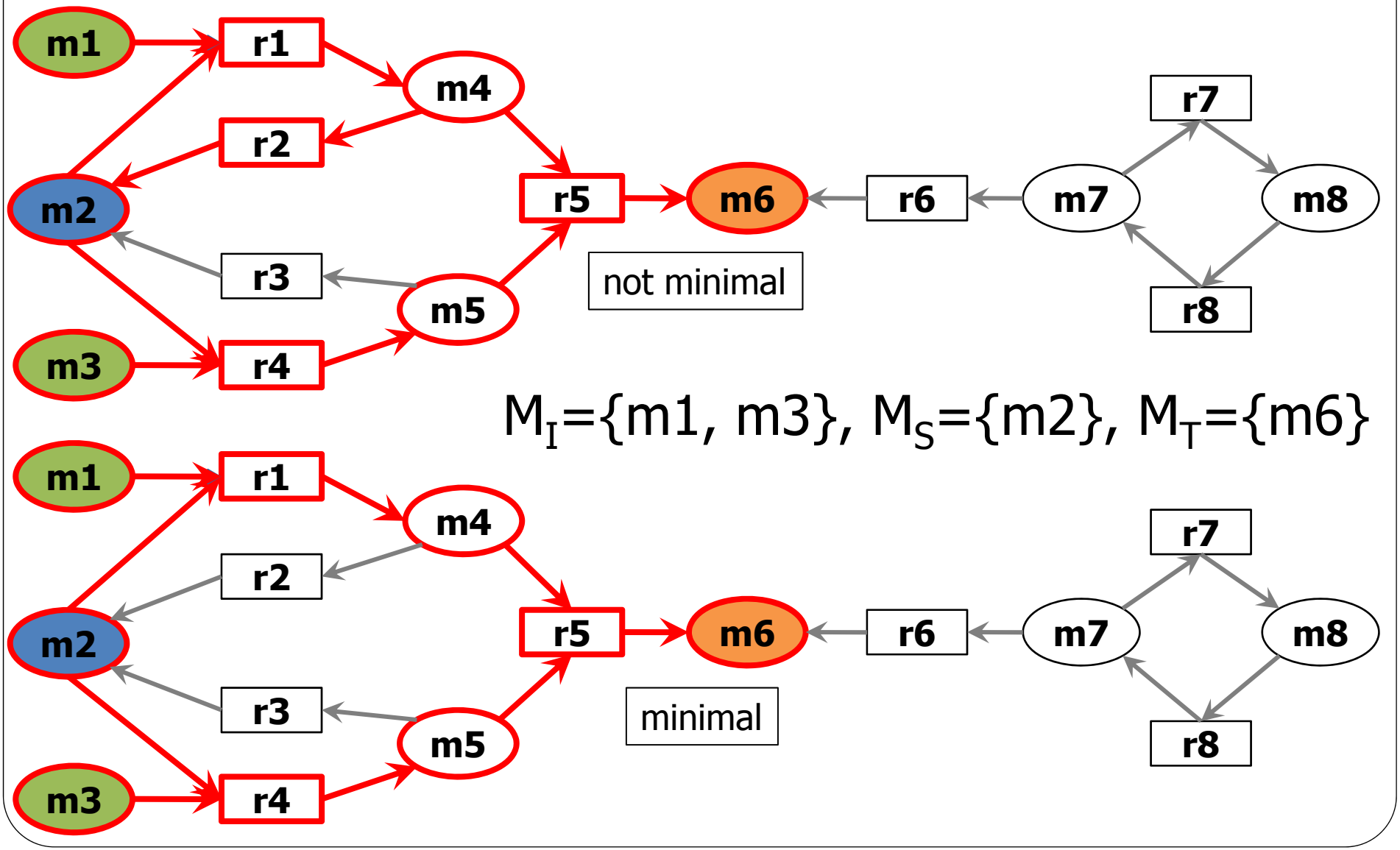
Minimal Sub-pathway (Example.)



$M_S=\{m1\}$, $M_T=\{m6\}$



Minimal Sub-pathway (Example.)



Problem Definition

Sub-pathway Finding Problem

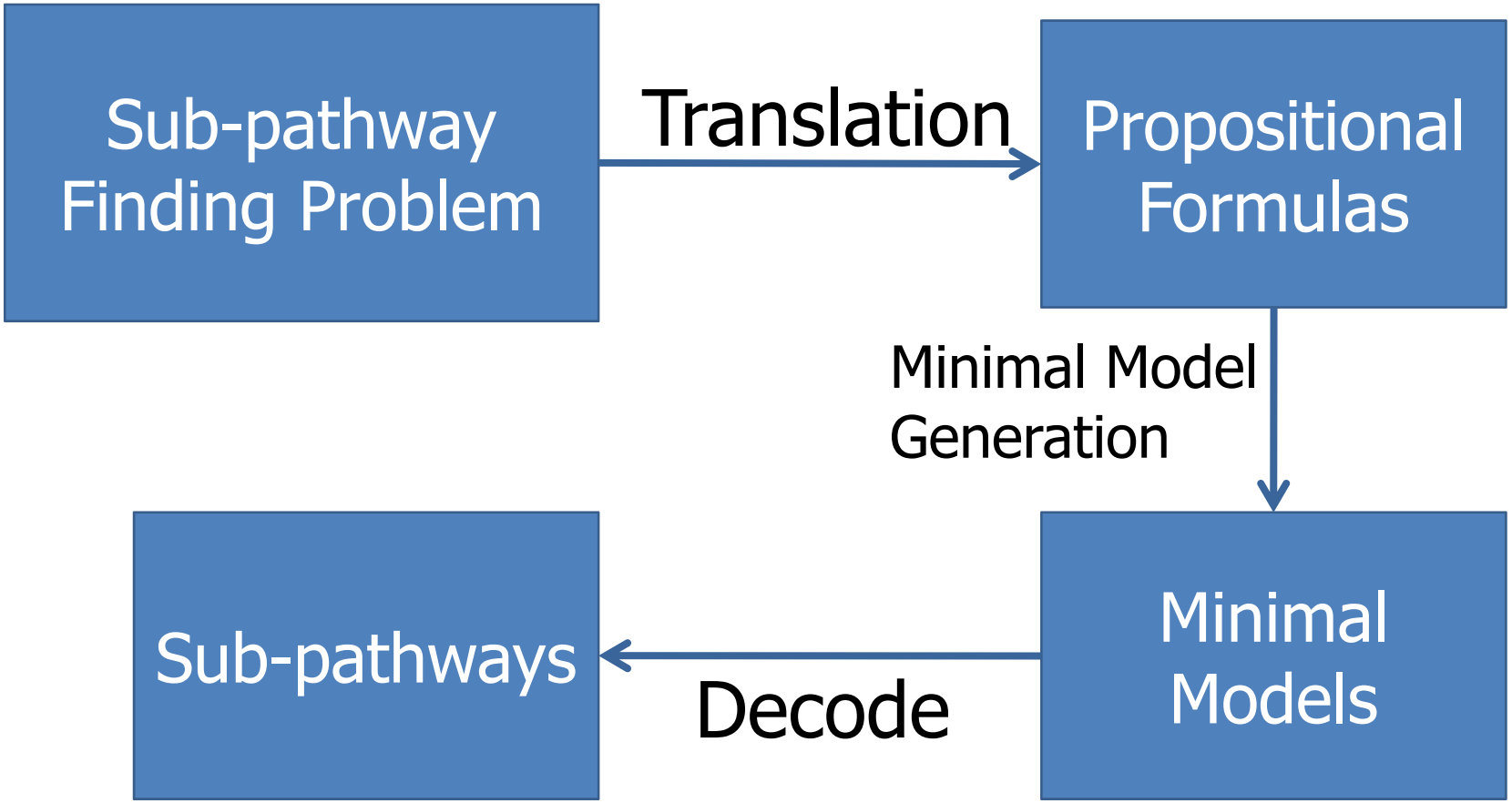
Input:

a pathway instance $\pi = (M, R, A, M_I, M_S, M_T)$

Output:

all minimal sub-pathways of π

Identifying Necessary Reactions by Minimal Model Generation



Translation: variables

- $rt_{n,t} \in V$: a propositional variable which is true if a reaction r_n is activatable at t=e and later.

$$rt_{n,e} \rightarrow rt_{n,e+1}$$

- $mt_{i,t} \in V$: a propositional variable which is true if a metabolite m_i is producible at t=e and later.

$$mt_{i,e} \rightarrow mt_{i,e+1}$$

Translation: formulas (1), (2)

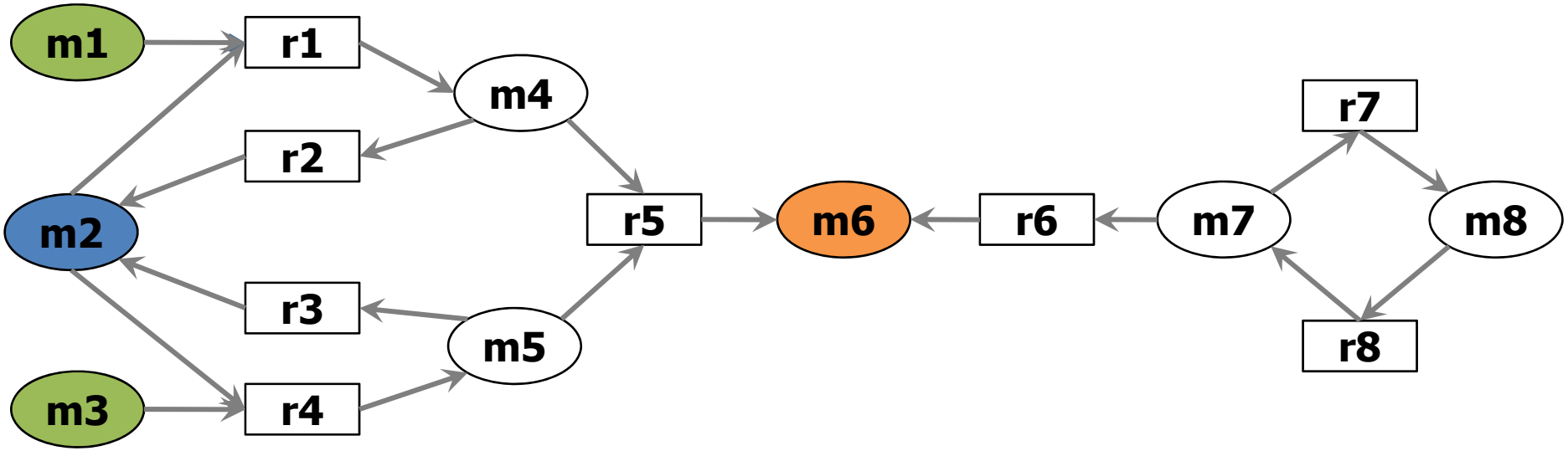
- For each reaction r_n:
  - If a reaction r_n is activatable at t=e then its reactants must be producible at t=e-1.

$$rt_{n,e} \rightarrow \bigwedge_{m_i \in s(r_n)} mt_{i,e-1} \qquad (1)$$

  - If a reaction r_n is activatable at t=e then its products must be producible at t=e

$$rt_{n,e} \rightarrow \bigwedge_{m_j \in p(r_n)} mt_{j,e} \qquad (2)$$

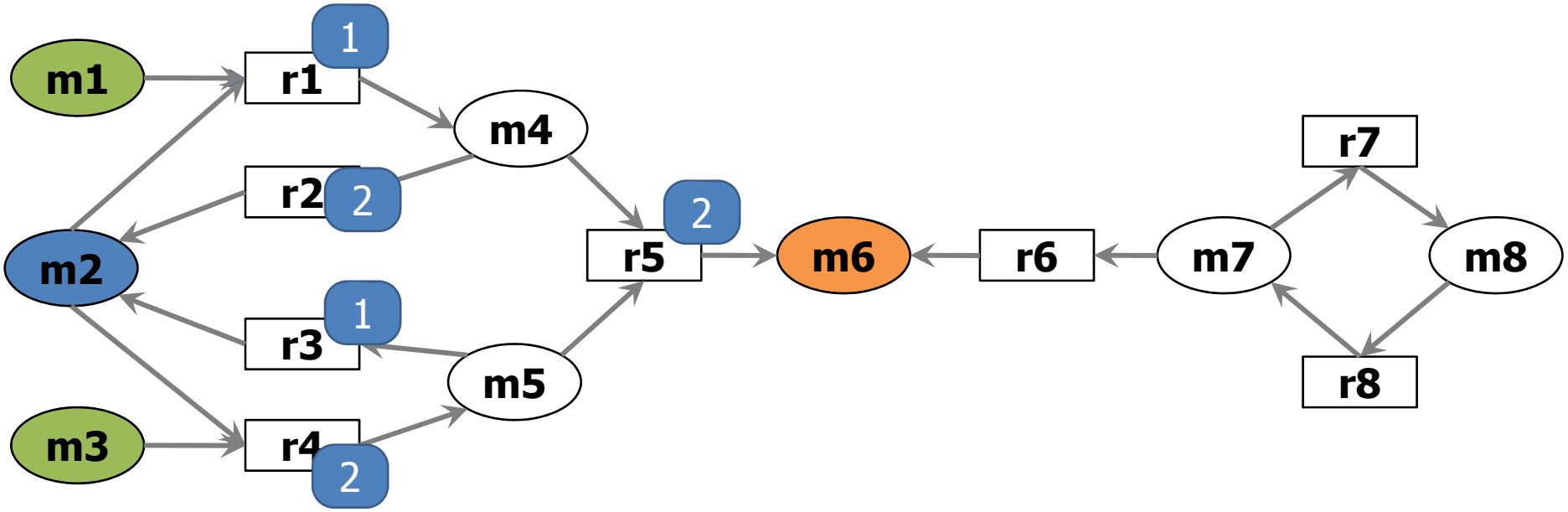
Earliest Activatable Time



- In standard way, above formulas are generated for every time t and every reaction.
- It results in the expansion of translated clauses.

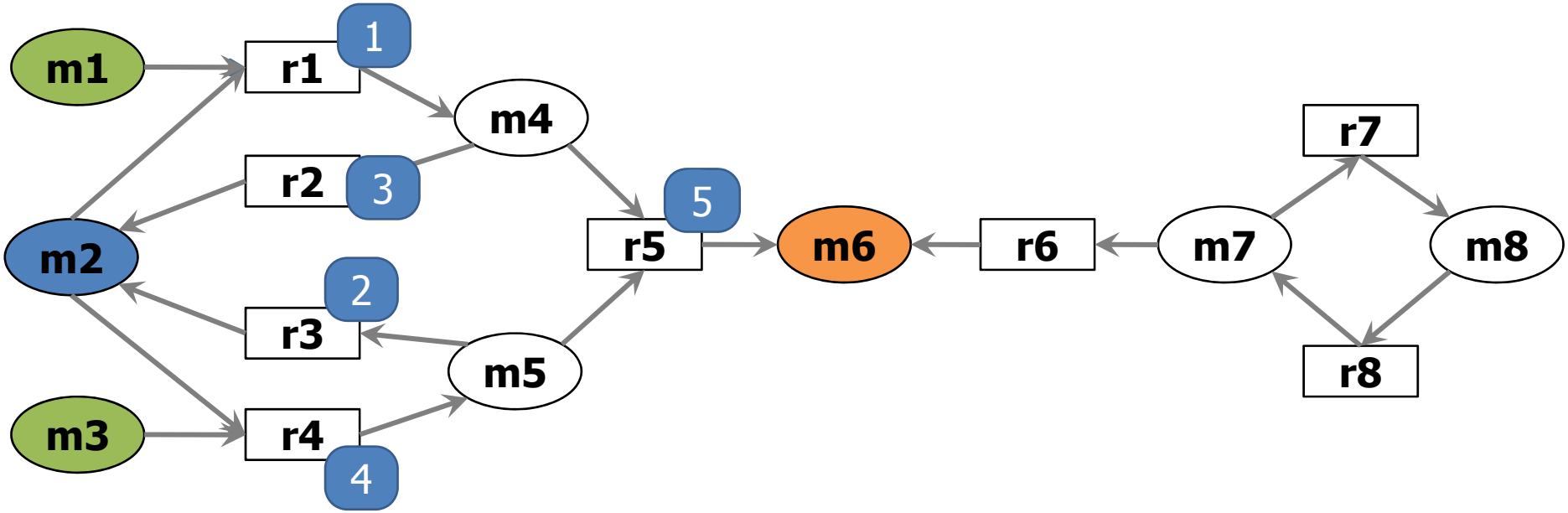
Assign Earliest Activatable Time (M')
begin
$d := 0$;
while $(M' \neq \emptyset)$
mark $\forall m_i \in M'$ as visited;
$M'' = \emptyset$;
$d := d + 1$;
loop for $m_i \in M'$
loop for unvisited $r_j \in s^{-1}(m_i)$
if $\forall m_k \in s(r_j)$ is visited **then**
mark r_j as visited;
$f_e := f_e \cup \{(r_j, d)\}$;
loop for unvisited $m_k \in p(r_j)$
$M'' := M'' \cup \{m_k\}$;
$M' := M''$;
return (f_e, d);
end

Earliest Activatable Time



- By the procedure, we do not need to translate
  - r_6, r_7, r_8, m_7, m_8 for any time.
  - r_2, r_4, r_5 for t=1.
- Let G′=(M′,R′,A′) be a reduced pathway.

Unique Time



- f_u: $R' \rightarrow Z^+$ such that $f_u(r)$ represents the unique time of $r \in R'$.
- We only need to translate
  - $r_i \in R'$ for its unique time $f_u(r)$.

Translation: formula (3)

- For each reaction r_n and its unique time $f_u(r_n)$:
  - If a reaction r_n is not activatable then metabolite $m_j \in p(r_n)$ keeps its state from $t=f_u(r_n)-1$

$$\neg rt_{n,f_u(r_n)} \rightarrow \bigwedge_{m_j \in p(r_n)} \left(\neg mt_{j,f_u(r_n)-1} \rightarrow \neg mt_{j,f_u(r_n)}\right) \quad (3)$$

Translation: Reaction

- z: integer denotes step, k: integer variable ($1 \leqq k \leqq z$)
- $o_{k,n} = n(R')*(k-1) + f_u(r_n)$
- The conjunction of the formula (1), (2) and (3) is as follows:

$$D^k_{r_n} = \left(rt_{n,o_{k,n}} \rightarrow \bigwedge_{m_i \in s(r_n)} mt_{i,o_{k,n}-1} \wedge \bigwedge_{m_j \in p(r_n)} mt_{j,o_{k,n}} \right) \wedge \left(\neg rt_{n,o_{k,n}} \rightarrow \bigwedge_{m_j \in p(r_n)} \left(\neg mt_{j,o_{k,n}-1} \rightarrow \neg mt_{j,o_{k,n}} \right) \right)$$

Translation: Condition

- Initial condition

$$C(0) = \bigwedge_{m_i \in M_s \cup M_i} mt_{i,0} \wedge \bigwedge_{m_j \in M \setminus (M_s \cup M_i)} \neg mt_{j,0}$$

- Target condition

$$C(n(R') * z) = \bigwedge_{m_i \in M_t} mt_{i,n(R')*z}$$

Translation

- The translated formula is as follows:

$$\Psi = C(0) \wedge C(n(R') * z) \wedge \psi_s \wedge \bigwedge_{k=1}^{z} \bigwedge_{n=1}^{n(R')} \left(D_{r_n}^{k}\right)$$

- The size of the translated clause is O(n(A')).

Minimal Model (1/2)

- [Koshimura '09] reports a method to compute a minimal model of a propositional formula Ψ with respect to a set of variables V', which is an extension of a work by [Niemela '96].

- We here represent a model of Ψ in the set of propositional variables to which it assigns true.

- A model I is a minimal model of a propositional formula Ψ with respect to a set of propositional variables V' iff I is a model of Ψ and there is no model I' of Ψ such that $I' \cap V' \subset I \cap V'$.

Minimal Model (2/2)

By [Koshimura '09]

A model I is a minimal model of Ψ with respect to V' iff a formula

$$\Psi' = \Psi \wedge \neg(x_1 \wedge x_2 \wedge \ \ldots \ \wedge x_i) \wedge \neg y_1 \wedge \neg y_2 \wedge \ \ldots \ \wedge \neg y_j$$

is unsatisfiable, where $I \cap V' = \{x_1, x_2, \ldots, x_i\}$, $\overline{I} \cap V' = \{y_1, y_2, \ldots, y_j\}$.

Minimal Model Generation Procedure $(\Psi,\ V_p)$

begin

$\Sigma := \emptyset$;

loop

$(\text{res},\ I) = \textbf{Solve}(\Psi)$;

if res $=$ `UNSAT` **then** return Σ ;

else

$V_x := I \cap V_p$;

$V_y := \overline{I} \cap V_p$;

$\Psi_c := \Psi \wedge \left(\bigvee_{x_i \in V_x} \neg x_i\right) \wedge \left(\bigwedge_{y_j \in V_y} \neg y_j\right)$;

$(\text{res},\ V_c) = \textbf{Solve}(\Psi_c)$;

if res $=$ `UNSAT` **then** $\Sigma := \Sigma \cup \{I\}$;

$\Psi := \Psi \wedge \left(\bigvee_{x_i \in V_x} \neg x_i\right)$;

end

Translation: Decode

- If I is a minimal model of a formula Ψ with respect to V′ then $G_{ms}=(M_{ms}, R_{ms}, A_{ms})$ is a minimal sub-pathway of $\pi=(M, R, A, M_I, M_S, M_T)$ where:
  - $M_{ms}=\{f_v(mt_{i,t})|mt_{i,t}\in I\cap V'\}$,
  - $R_{ms}=\{f_v(rt_{j,t})|rt_{j,t}\in I\cap V'\}$,
  - $A_{ms}=\{(m_j,r_i)|\ m_j\in s(r_i),\ r_i\in R_{ms}\}$
    $\cup\{(r_i,\ m_j,)|\ m_j\in p(r_i),\ r_i\in R_{ms}\}$
  - $\forall m\in M_s,\ rt_{i,t}\in I$ s.t. $f_v(rt_{j,t})\in s^{-1}(m)$

- $V' = \{mt_{i,t} \mid mt_{i,t}\in V,\ t=n(R')*z\}\cup\{rt_{j,t} \mid rt_{j,t}\in V,\ t=n(R')*z\}$
- $f_v : V \rightarrow M\cup R$ such that $f_v(mt_{i,t})=m_i$, $f_v(rt_{j,t}) = r_j$

Experiment 1

- Pathway Instance
  - 880 reactions
  - target, source, initial condition from [Beasley '07].
- Computational Environment:
  - CPU (Centrino 2.53GHz), RAM (2GB)
  - Each experiment has been done within a second.
- Comparison with [Beasley '07] and [Planes '09].

1.	gluconeogenesis	6.	pentose phosphate
2.	glycogen	7.	deoxythymidine phosphate
3.	glycolysis	8.	Kreb's cycle
4.	proline bio-synthesis	9.	NAD biosynthesis
5.	ketogluconate metabolism	10.	arginine biosynthesis

Experiment 1

Pathway#	Proposal			Beasley'07		Planes'09
	#Steps	#Sols.	res.	res. (a)	res. (b)	res.
1	3	1	yes	yes	no	no
2	1	1	yes	yes	no	yes
3	2	38	yes	yes	yes	no
4	1	1	yes	yes	no	no
5	3	4	yes	no	no	yes
6	2	7	yes	yes	no	yes
7	1	1	yes	yes	no	yes
8	3	28	yes	no	yes	no
9	1	3	yes	yes	no	yes
10	1	1	yes	yes	no	yes
Total # of yes in res.			10	8	2	6

Columns 4-7 (res.) show that found reactions correspond to the reactions described in the literature [Beasley '07]. Columns 5-6 represents the optimizations of (a) a number of reactions and (b) production of ATP, respectively.

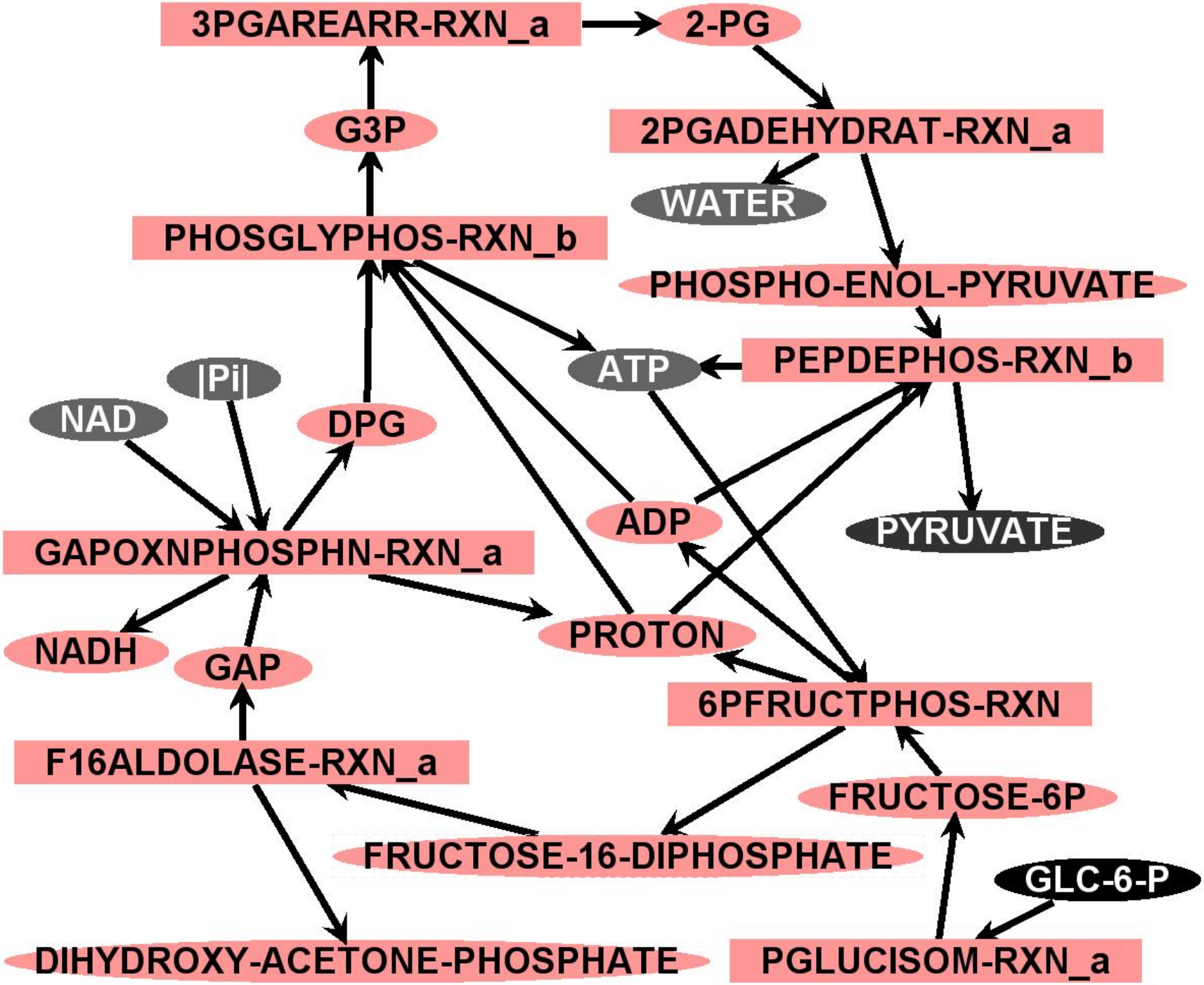
Experiment 2

- Pathway Instance:
  - *E. coli* consists of 1777 reactions and 1073 metabolites (from EcoCyc version 13.6).
  - M_I={PROTON, WATER, ATP, ADP, |pi|, NAD}
  - M_S={GLC-6-P}
  - M_T={PYRUVATE}

- Computational Environment:
  - CPU (Centrino 2.53GHz), RAM (2GB)
  - The experiment has been done within a minute.

Experiment 2

- We found 4880 minimal sub-pathways on the pathway (z=1) and ordered those sub-pathways according to the number of reactions.
- The figure shows a sub-pathway found in the best 10 solutions corresponding to the conventional glycolysis sub-pathway.


3PGAREARR-RXN_a
2-PG
G3P
2PGADEHYDRAT-RXN_a
WATER
PHOSGLYPHOS-RXN_b
PHOSPHO-ENOL-PYRUVATE
ATP
PEPDEPHOS-RXN_b
|Pi|
NAD
DPG
ADP
PYRUVATE
GAPOXNPHOSPHN-RXN_a
NADH
GAP
PROTON
6PFRUCTPHOS-RXN
F16ALDOLASE-RXN_a
FRUCTOSE-6P
FRUCTOSE-16-DIPHOSPHATE
GLC-6-P
DIHYDROXY-ACETONE-PHOSPHATE
PGLUCISOM-RXN_a

Conclusion

- We formalize the sub-pathway finding problem, to identify a set of reactions to produce target metabolites using source and initial metabolites.
- As a proof-of-concept, we apply our method to a whole *E. coli* pathway.
- Future work
  - To evaluate and rank models.
  - To add more biological rules to the translation.
  - To apply the method to other pathways including gene regulatory network etc...
- Please also see the conference paper "Identifying Necessary Reactions in Metabolic Pathways by Minimal Model Generation," PAIS (sub-conference of ECAI), Lisbon, 2010.