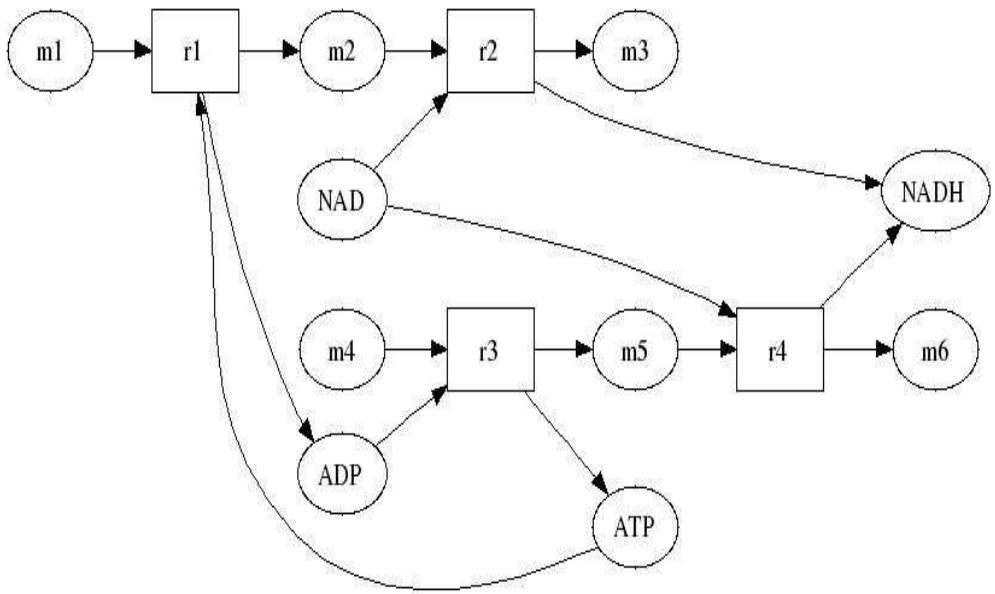
Pitfalls in substrate graph analysis: co-factors

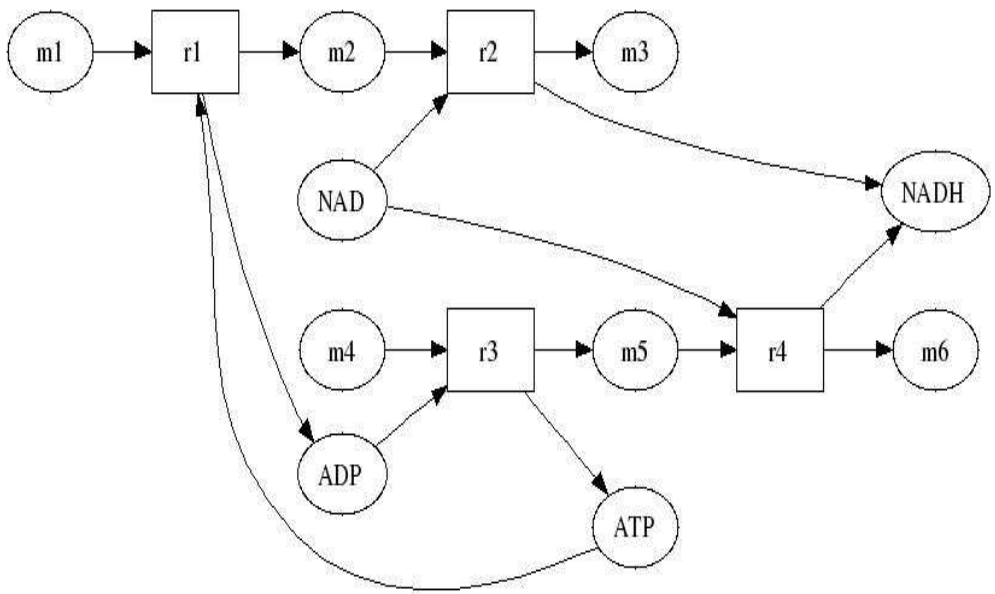
- Path length in substrate graphs may not be biologically relevant
- Shortest paths between metabolites in otherwise distant parts of metabolism tend to go through co-factor metabolites (NADP, NAPH, ATP, ADP).
- However, transfer of atoms occurs only between the co-factors



Pitfalls in substrate graph analysis: co-factors

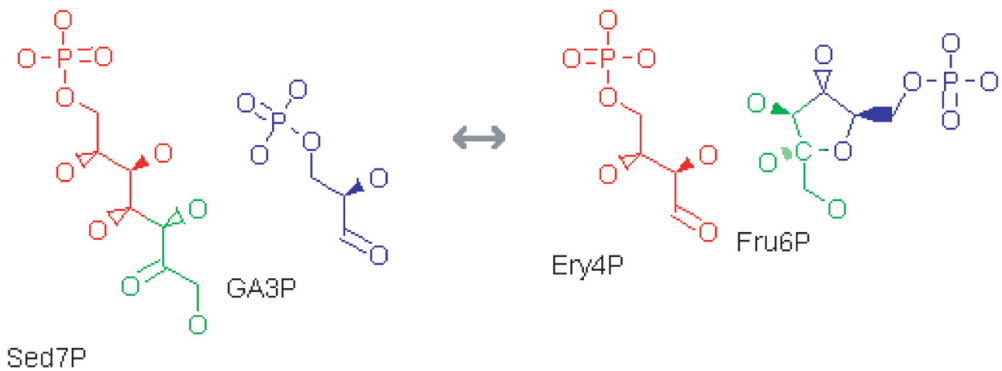
Quick remedy used in most studies:

- Remove co-factors from the graph
- But sometimes it is difficult to decide which ones should be removed and which ones to leave.



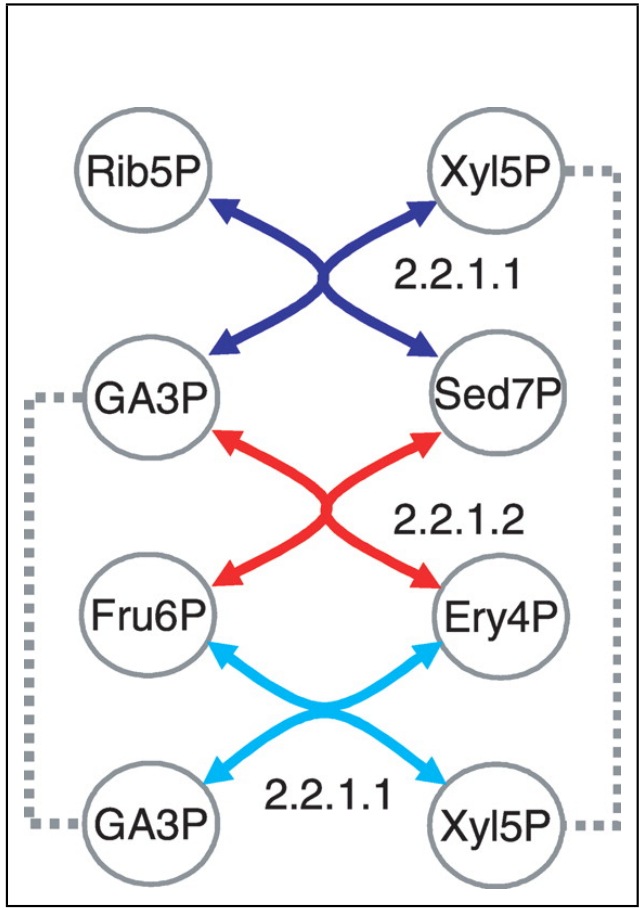
Atom-level representation

- Better solution is to trace the atoms accross pathways
- An acceptable path needs to involve transfer of atoms from source to target.
- Spurious pathways caused by the co-factor problem are filtered out
- This paradigm is used by Arita in his ARM software (www.metabolome.jp)



Example system: transketolase and transaldolase

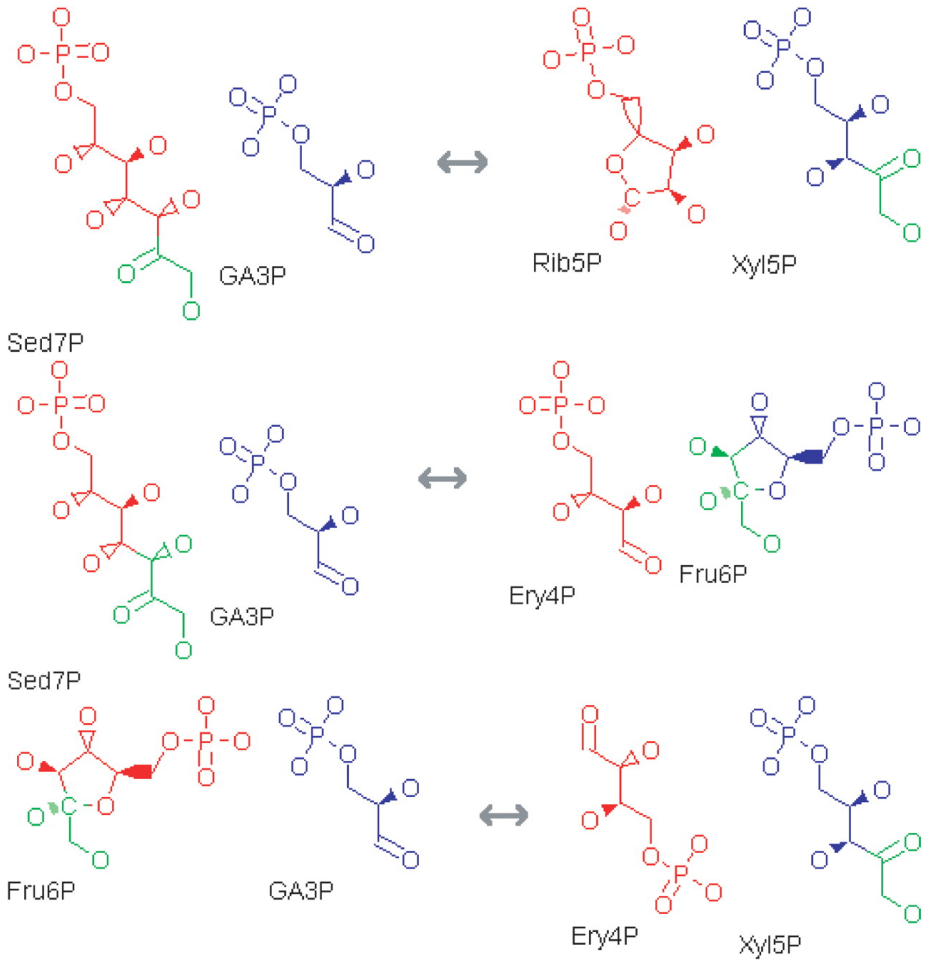
- Consider a system of three reactions, catalyzed by transketolase (tkt) and transaldolase (tal) enzymes:
- Inspecting the reaction equations, it would seem that it takes two reactions to make E4P out of R5P and X5P



Example system: transketolase and transaldolase

- If we trace the atoms through the reactions, we notice that the atoms of X5P take the route X5P - G3P - F6P - E4P
- So after two steps (R1, R2) no atoms from X5P have been transferred to E4P

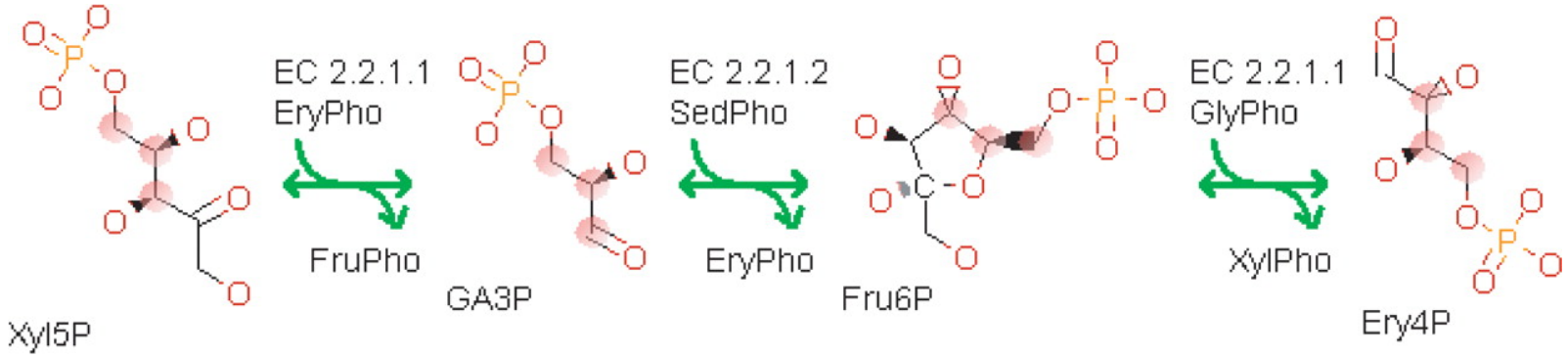
- R1: $R5P + X5P \overset{tkt}{\Rightarrow} G3P + S7P$
- R2: $G3P + S7P \overset{tal}{\Rightarrow} F6P + E4P$
- R3: $F6P + G3P \overset{tkt}{\Rightarrow} X5P + E4P$



Example system: transketolase and transaldolase

- It takes one further step (R3) to transfer atoms from F6P to E4P

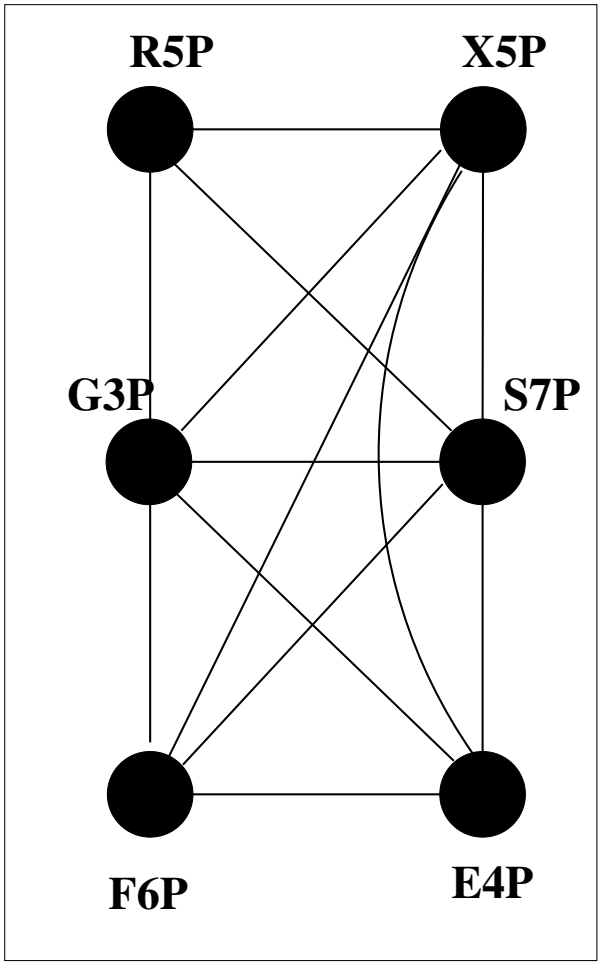
- R1: $R5P + X5P \overset{tkt}{\Rightarrow} G3P + S7P$
- R2: $G3P + S7P \overset{tal}{\Rightarrow} F6P + E4P$
- R3: $F6P + G3P \overset{tkt}{\Rightarrow} X5P + E4P$



Example system: transketolase and transaldolase

- The metabolite graph of the example system is (almost) fully connected graph
- Suggests path length 1 for all metabolite pairs except R5P - F6P and R5P - E4P which have path length 2.

- R1: $\mathrm{R5P} + \mathrm{X5P} \overset{tkt}{\Rightarrow} \mathrm{G3P} + \mathrm{S7P}$
- R2: $\mathrm{G3P} + \mathrm{S7P} \overset{tal}{\Rightarrow} \mathrm{F6P} + \mathrm{E4P}$
- R3: $\mathrm{F6P} + \mathrm{G3P} \overset{tkt}{\Rightarrow} \mathrm{X5P} + \mathrm{E4P}$



Pitfalls in substrate graph analysis: self-suffiency

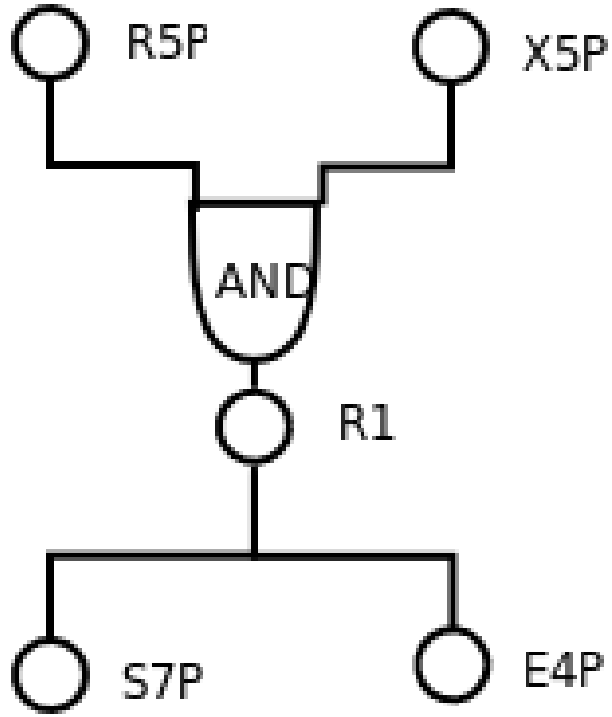
- The shortest path may not correlate well with the *e*ffort that the cell needs to make the conversion
- The conversions require other metabolites to be produced than the ones along the direct path.
- Arguably a feasible pathway should be self-sufficiently capable of performing the conversion from sources to target metabolites
- To make this notion precise, we will use Boolean circuits as the Metabolic representation

Metabolic networks as boolean circuits

- Each metabolite and reaction is either reachable (1) or not reachable (0)
- A metabolite is reachable if and only if
  - it is an external input substrate, or
  - there EXISTS a reachable reaction that produces it
- A reaction is reachable if and only if ALL of its substrates are reachable

Metabolic networks as boolean circuits

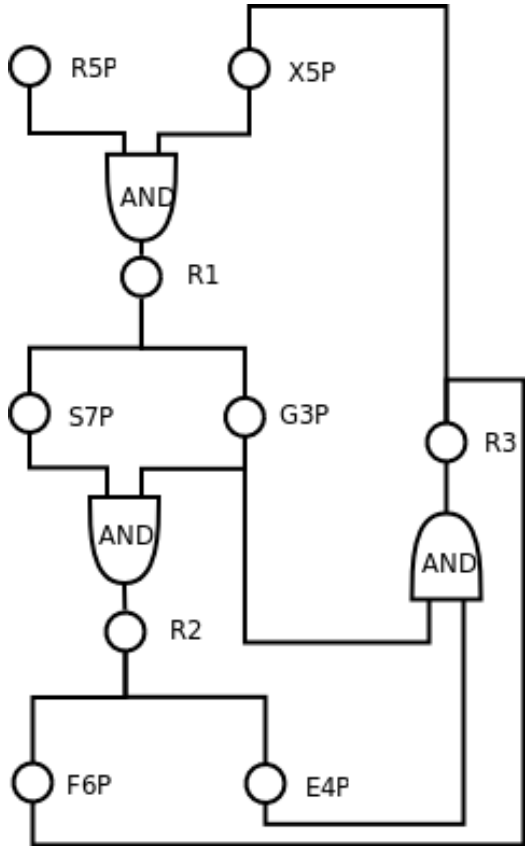
- Graphically a single reaction can be drawn as an AND gate that sends a 1 if all substrates have state 1, and 0 otherwise



R1: R5P + X5P $\stackrel{tkt}{\Rightarrow}$ G3P + S7P

Metabolic networks as boolean circuits

- A system of reactions induces a boolean circuit by joining together the circuits of the single reactions
- One could draw an OR gate above each metabolite so as to denote that the metabolite is on if one of the reactions producing it is on.
- These have been omitted for clarity.



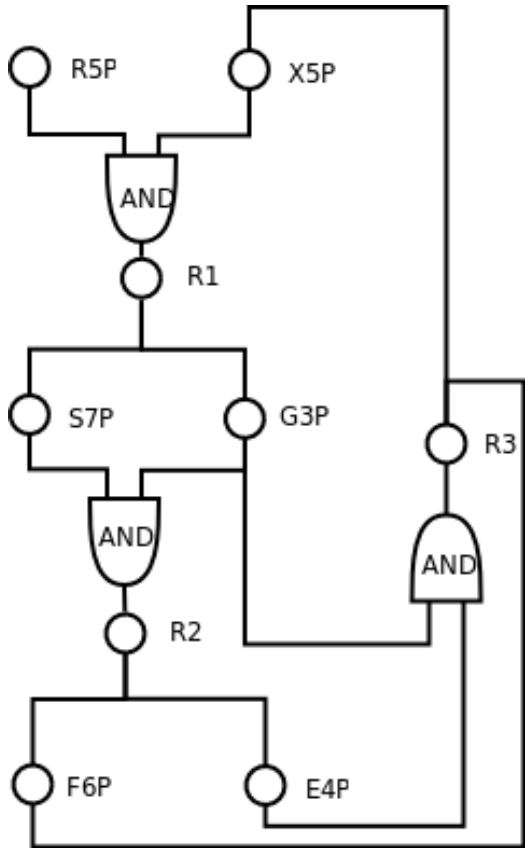
R1: $R5P + X5P \stackrel{tkt}{\Rightarrow} G3P + S7P$

R2: $G3P + S7P \stackrel{tal}{\Rightarrow} F6P + E4P$

R3: $F6P + G3P \stackrel{tkt}{\Rightarrow} X5P + E4P$

Metabolic networks as boolean circuits

- From the circuit below one can deduce that in order to produce E4P, both R5P and X5P are needed as substrates
- This is complementary kind of information to atom-level representation:
  - Atom-level representation tells us that only after R3 atoms from X5P are transferred to E4P
  - The need for R5P is not high-lighted



$$\text{R1: } \text{R5P} + \text{X5P} \overset{tkt}{\Rightarrow} \text{G3P} + \text{S7P}$$
$$\text{R2: } \text{G3P} + \text{S7P} \overset{tal}{\Rightarrow} \text{F6P} + \text{E4P}$$
$$\text{R3: } \text{F6P} + \text{G3P} \overset{tkt}{\Rightarrow} \text{X5P} + \text{E4P}$$

Applications of metabolic circuits

- The above described *metabolic circuit*, or *AND-OR graph*, representation has two major uses
- First, we can pose the question whether the reconstructed metabolic network is structurally consistent in the sense that all metabolites can be produced available nutrients
  - Interestingly, not all published metabolic reconstructions satisfy this property
- Second, we can analyze the difficulty of producing some metabolite from another in terms of how large a circuit needs to be activated.

Checking reachability in metabolic circuits

Given a metabolic circuit, it is easy to check the reachability using breadth-first search:

- Consider a set of nutrients, input metabolites that are marked reachable from the start
- Iterate the following, until no new metabolites are reached
  1. Mark reachable all reactions whose all substrates are marked reachable
  2. Mark reachable all products of the reachable reactions

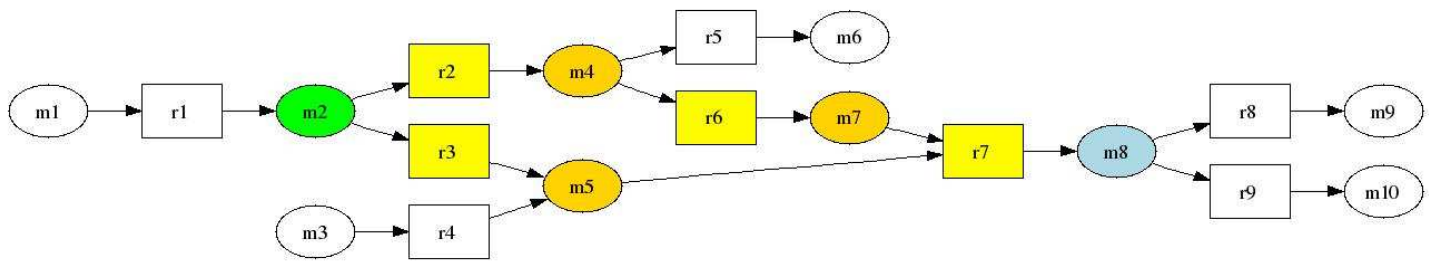
In the end all reachable metabolites and reactions have been found

Metabolic circuits and the small-world property

- Using the metabolic circuit representation, one can redefine the concept of pathway
- We define a *feasible* pathway from metabolite A to B, as the minimal set of reactions F in the metabolic network so that in the metabolic circuit constructed from F, B is reachable whenever A is reachable.
- Intuitively, if we feed cell A as the sole nutrient, it can convert A to B only using the reactions given in F.

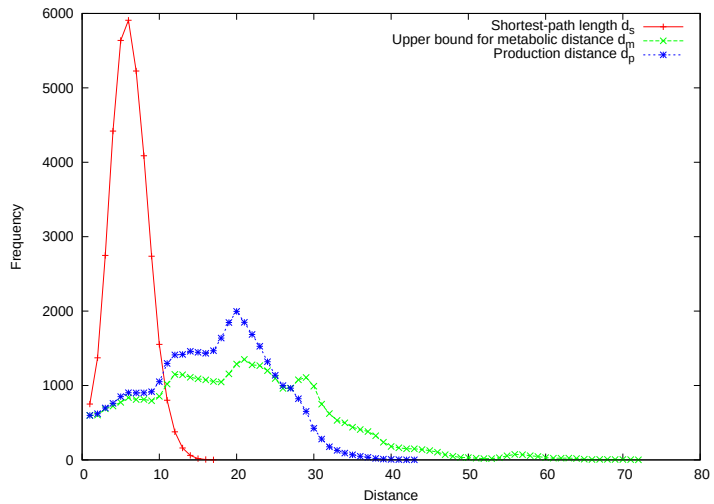
Feasible pathway vs. shortest simple path

- Feasible pathway contains the yellow reactions $r2$, $r3$, $r6$ and $r7$
- Shortest simple path has length 2, corresponding to the simple path through $r3$ and $r7$



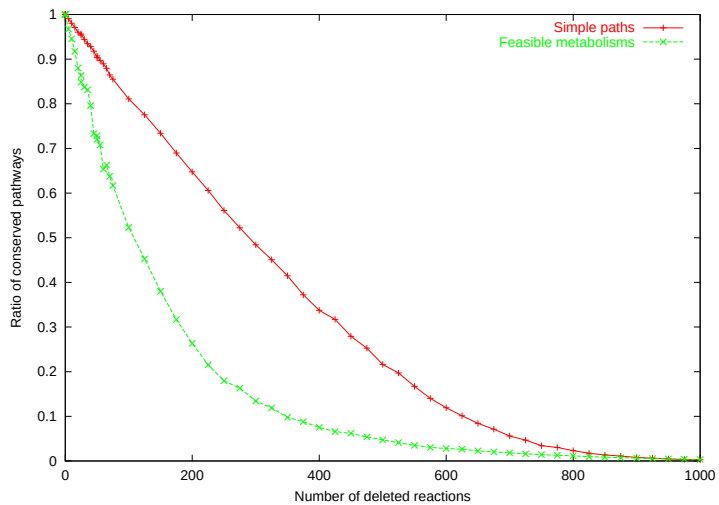
Feasible pathway vs. shortest simple path

- Simple path length distribution shows the small-world property: most paths are short
- Feasible pathway size (in the figure: green) shows no small world property
- Many conversions between two metabolites that involve a large number of enzymes



Robustness & small world property

- It has been claimed that the small-world property gives metabolic networks robustness towards random mutations.
- As evidence the conservation of short pathways under random gene deletions has been offered
- However, the smallest feasible pathways are not as robust, showing that even random mutations can quickly damage the cells capability to make conversions between metabolites (as easily).



Stoichiometric network analysis

In stoichiometric analysis of metabolic networks, one concerns the effect of the network structure on the behaviour and capabilities of metabolism.
Questions that can be tackled include:

- Discovery of pathways that carry a distinct biological function (e.g. glycolysis) from the network, discovery of dead ends and futile cycles, dependent subsets of enzymes
- Identification of optimal and suboptimal operating conditions for an organism
- Analysis of network flexibility and robustness, e.g. under gene knockouts

Stoichiometric coefficients

Soitchiometric coefficients denote the proportion of substrate and product molecules involved in a reaction. For example, for a reaction

$$r : A + B \mapsto 2C,$$

the *stoichiometric coefficients* for A, B and C are $-1, -1$ and 2, respectively.

- Assignment of the coeefficients is not unique: we could as well choose $-1/2, -1/2, 1$ as the coefficients
- However, the relative sizes of the coeefficients remain in any valid choice.
- Note! We will denote both the name of a metabolite and its concentration by the same symbol.

Stoichiometry and reaction rates

- The rate of change of concentration of metabolites is the most fundamental quantity in stoichiometric models
- Assume a reaction

$$r : A + B \mapsto 2C,$$

operates at some rate or velocity v (arbitrary units e.g. mol/hour)

- Then, the change of concentration of the reactants and the product are given by the reaction rate multiplied by the shoichiometric coefficients

$$\frac{dA}{dt} = -1 \cdot v, \frac{dA}{dt} = -1 \cdot v, \frac{dC}{dt} = 2 \cdot v$$

- Thus, A and B are consumed at the rate of the reaction, C is produced at the double rate.

Reversible reactions

- Many of metabolic reactions are reversible,

$$r : A + B \leftrightharpoons 2C,$$

  so they can work in either direction, depending on the conditions within the cell
- In stoichiometric models a reversible reaction can be modelled in two ways:
  - As a single reaction that can operate from left to right, indicated by positive reaction rate $v > 0$ or right to left, indicated by negative reaction rate $v < 0$.
  - As two separate reactions $r' : A + B \mapsto 2C$ and $r'' : 2C \mapsto A + B$, both with non-negative reaction rates $v', v'' \geq 0$.

Concentration and rate vectors

- Let us assume that our metabolic network has the reactions $\mathcal{R} = \{R_1, R_2, \ldots, R_r\}$ and the metabolites $\mathcal{M} = \{M_1, M_2, \ldots, M_m\}$
- Let the reaction R_i operate with rate v_i
- We collect the individual reaction rates to a *rate vector* $\mathbf{v} = (v_1, \ldots, v_r)^T$

Stoichiometric vector and matrix

- The stoichiometric coefficients of a reaction are collected to a vector s_r
- In s_r there is a one position for each metabolite in the metabolic system
- The stoichiometric co-efficient of the reaction are inserted to appropriate positions, e.g. for the reaction

$$r : A + B \mapsto 2C,$$

$$s_r = \begin{matrix} \cdot \\ \cdot \\ A \\ \cdot \\ \cdot \\ B \\ \cdot \\ \cdot \\ C \end{matrix} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

Stoichiometric matrix

- The stoichiometric vectors can be combined into the stoichiometric matrix S.
- In the matrix S, the is one row for each metabolite $M_1, dots, M_m$ and one column for each reaction $R_1, \ldots, R_r$.
- The coefficients s_{*j} along the j'th column are the stoichiometric coeefficients of of the reaction j.

$$\mathbf{S} = \begin{bmatrix} s_{11} & \cdots & s_{1j} & \cdots & s_{1r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{i1} & \cdots & s_{ij} & \cdots & s_{ir} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mj} & \cdots & s_{mr} \end{bmatrix}$$

Stoichiometric matrix

- The coefficients s_{*j} along the j'th column are the stoichiometric coeefficients of of the reaction j.

$$\mathbf{S} = \begin{bmatrix} s_{11} & \cdots & s_{1j} & \cdots & s_{1r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{i1} & \cdots & s_{ij} & \cdots & s_{ir} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mj} & \cdots & s_{mr} \end{bmatrix}$$

Stoichiometric matrix

- The coefficients along the i'th row denote the relationship between the concentration of metabolite M_i and the reactions consuming or producing it.

$$\mathbf{S} = \begin{bmatrix} s_{11} & \cdots & s_{1j} & \cdots & s_{1r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{i1} & \cdots & s_{ij} & \cdots & s_{ir} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mj} & \cdots & s_{mr} \end{bmatrix}$$

Example: stoichiometric matrix

- The stoichiometric matrix of our example system is a 10-by-7 matrix:

R_1: βG6P + NADP$^+$ $\overset{zwf}{\Rightarrow}$ 6PGL + NADPH

R_2: 6PGL + H_2O $\overset{pgl}{\Rightarrow}$ 6PG

R_3: 6PG + NADP$^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH

R_4: R5P $\overset{rpe}{\Rightarrow}$ X5P

R_5: αG6P $\overset{gpi}{\Leftrightarrow}$ βG6P

R_6: αG6P $\overset{gpi}{\Leftrightarrow}$ $\beta F6P$

R_7: βG6P $\overset{gpi}{\Leftrightarrow}$ $\beta F6P$

$$
S = \begin{array}{c} \\ \beta G6P \\ \alpha G6P \\ \beta F6P \\ 6PGL \\ 6PG \\ R5P \\ X5P \\ NADP^+ \\ NADPH \\ H_2O \end{array}
\begin{bmatrix}
-1 & 0 & 0 & 0 & 1 & 0 & -1 \\
0 & 0 & 0 & 0 & -1 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
-1 & 0 & -1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

Systems equations

- Suppose that reactions R_1, R_5 and R_7 operate at rates 2, 1 (left to right) and -2 (right to left), respectively
- Multiply the reaction rates with stoichiometric coefficients to obtain the rates of change of concentration of βG6P caused by each reaction: $R_1 : (-1) \cdot 2 = -2$, $R_5 : 1 \cdot 1 = 1$, $R_7 : (-1) \cdot (-2) = 2$

R_1: βG6P + NADP$^+$ $\overset{zwf}{\Rightarrow}$ 6PGL + NADPH
R_5: αG6P $\overset{gpi}{\Leftrightarrow}$ βG6P
R_7: βG6P $\overset{gpi}{\Leftrightarrow}$ βF6P

Stoichiometric coefficients from matrix S:

$$S_{\beta G6P} = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

Systems equations

- The *net rate* of change βG6P is therefore

$$\frac{d[\beta G6P]}{dt} = -2 + 1 + 2 = 1,$$

thus the system is accumulating βG6P

R_1: βG6P + NADP$^+$ $\overset{zwf}{\Rightarrow}$ 6PGL + NADPH
R_5: αG6P $\overset{gpi}{\Leftrightarrow}$ βG6P
R_7: βG6P $\overset{gpi}{\Leftrightarrow}$ βF6P

Stoichiometric coefficients from matrix S:

$$S_{\beta G6P} = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

Systems equations

In a network of n metabolites and r reactions, the dynamics of the system are characterized by the systems equations

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j, \text{ for } i = 1, \ldots, m$$

- X_i is the concentration of the ith metabolite
- v_j is the rate of the jth reaction and
- s_{ij} is the stoichiometric coefficient of ith metabolite in the jth reaction.

Intuitively, each system equation states that the rate of change of concentration of a is the sum of metabolite flows to and from the metabolite.

Systems equation example

- Assume our example metabolic network has the following rate vector $\mathbf{v} = (1, 1, 0, 0, 1, 0, 0)$
- Let us compute the rate of change for metabolites

R_1: βG6P + NADP$^+$ $\overset{zwf}{\Rightarrow}$ 6PGL + NADPH
R_2: 6PGL + H_2O $\overset{pgl}{\Rightarrow}$ 6PG
R_3: 6PG + NADP$^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH
R_4: R5P $\overset{rpe}{\Rightarrow}$ X5P
R_5: αG6P $\overset{gpi}{\Leftrightarrow}$ βG6P
R_6: αG6P $\overset{gpi}{\Leftrightarrow}$ βF6P
R_7: βG6P $\overset{gpi}{\Leftrightarrow}$ βF6P

$$\frac{d\beta G6P}{dt} = -1v_{R_1} + 1v_{R_5} - 1v_{R_7} = 0$$
$$\frac{d\alpha G6P}{dt} = -1v_{R_5} - 1v_{R_6} = -1$$
$$\Rightarrow \text{net consumption!}$$
$$\frac{d\beta F6P}{dt} = 1v_{R_6} + 1v_{R_7} = 0$$
$$\frac{d6GPL}{dt} = 1v_{R_1} - 1v_{R_2} = 0$$
$$\frac{d6PG}{dt} = 1v_{R_2} - 1v_{R_3} = 1$$
$$\Rightarrow \text{net production!}$$

Systems equation example

- Assume our example metabolic network has the following rate vector $\mathbf{v} = (1, 1, 0, 0, 1, 0, 0)$

R_1: βG6P + NADP$^+$ $\overset{zwf}{\Rightarrow}$ 6PGL + NADPH
R_2: 6PGL + H_2O $\overset{pgl}{\Rightarrow}$ 6PG
R_3: 6PG + NADP$^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH
R_4: R5P $\overset{rpe}{\Rightarrow}$ X5P
R_5: αG6P $\overset{gpi}{\Leftrightarrow}$ βG6P
R_6: αG6P $\overset{gpi}{\Leftrightarrow}$ βF6P
R_7: βG6P $\overset{gpi}{\Leftrightarrow}$ βF6P

$$\frac{dR5P}{dt} = 1v_{R_3} - 1v_{R_4} = 0$$

$$\frac{dX5P}{dt} = 1v_{R_4} = 0$$

$$\frac{dNADPH}{dt} = 1v_{R_1} + 1v_{R_3} = 1$$

$\Rightarrow$ net production!

$$\frac{dNADP^+}{dt} = -1v_{R_1} - 1v_{R_3} = -1$$

$\Rightarrow$ net consumption!

$$\frac{dH_20}{dt} = -1v_{R_2} = -1$$

$\Rightarrow$ net consumption!

Systems equations in matrix form

- The systems equation can be expressed in vector form as

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j = S_i^T \mathbf{v},$$

where S_i contains the stoichiometric coefficients of a single metabolite, that is a row of the stoichiometric matrix

- All the systems equations of different equations together can then be expressed by a matrix equation

$$\frac{d\mathbf{X}}{dt} = S\mathbf{v},$$

- Above, the vector

$$\frac{d\mathbf{X}}{dt} = \left(\frac{d\mathbf{X_1}}{dt}, \ldots, \frac{d\mathbf{X_n}}{dt} \right)^T$$

collects the rates of concentration changes of all metabolites