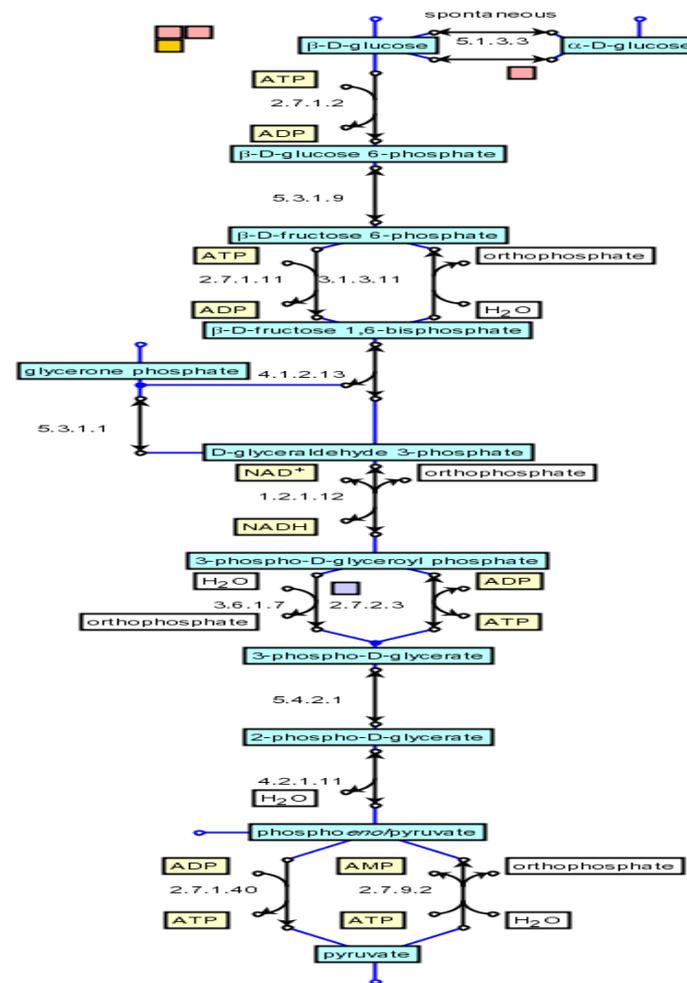


Representing metabolic networks

- Given a reconstructed network, how should we represent it?
- For computational and statistical analyses, we need to be exact, much more so than when communicating between humans
- Everything relevant should be included
- What is relevant depends on the questions that we want to solve

(picture: E. Coli glycolysis, EMP database,

www.empproject.com/)



Levels of abstraction

- Graph representations: Connectivity of reactions/metabolites, structure of the metabolic network
- Stoichiometric (reaction equation) representation: capabilities of the network, flow analysis, steady-state analyses
- Kinetic models: dynamic behaviour under changing conditions

Representing metabolic networks as graphs

For structural analysis of metabolic networks, the most frequently encountered representations are:

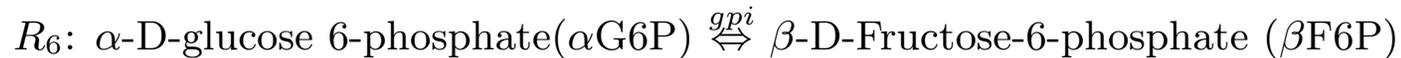
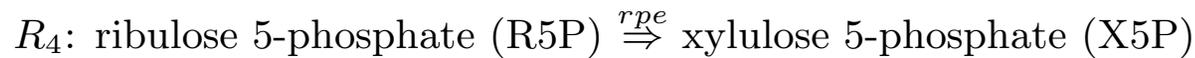
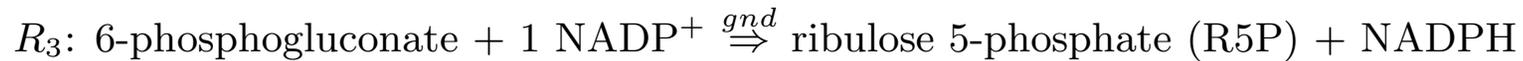
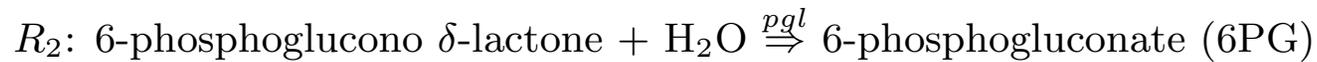
- Enzyme interaction network
- Reaction graph
- Substrate graph

We will also look briefly at

- Atom-level representations
- Feasible pathways

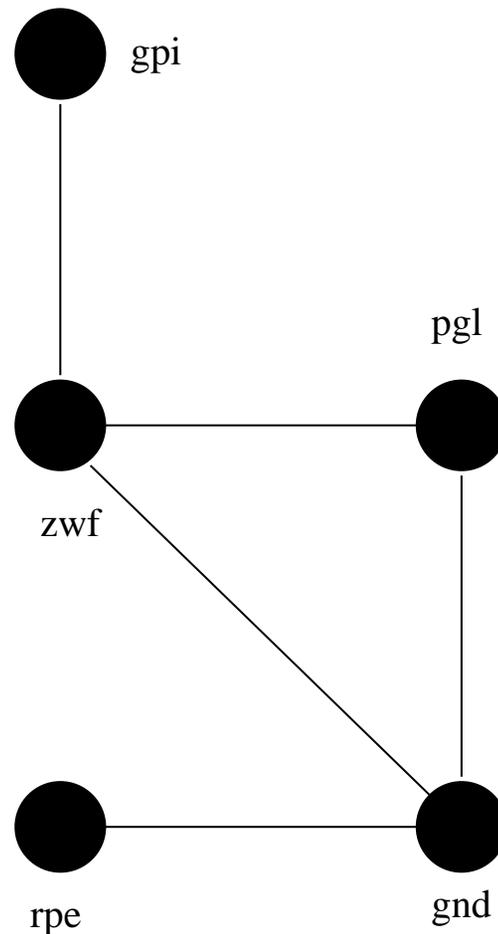
Example reaction List

Let us consider the following example list of reactions, implementing a part of pentose-phosphate pathway of E. Coli:

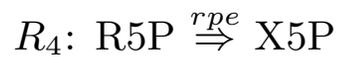
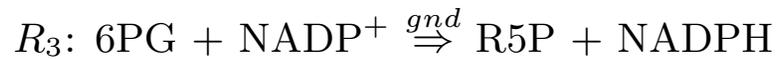
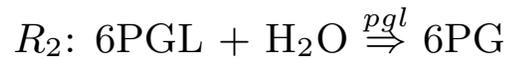
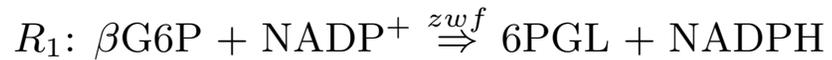


Enzyme interaction network (1/2)

- Enzymes as nodes
- Link between two enzymes if they catalyze reactions that have common metabolites
- A special kind of protein-protein interaction network



Enzyme interaction network (2/2)



Edge Supporting metabolites

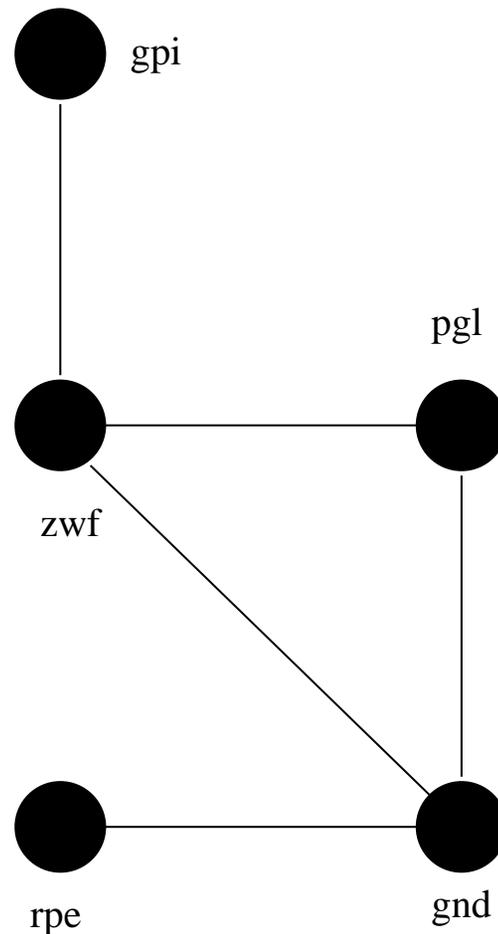
(gpi,zwf) βG6P

(zwf, pgl) 6PGL

(zwf, gnd) NADP, NADPH

(pgl,gnd) 6PG

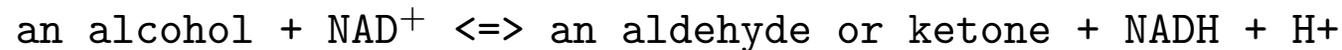
(gnd, rpe) R5P



Reaction clumping in Enzyme networks

As each enzyme is represented once in the network reactions catalyzed by the same enzyme will be clumped together:

- For example, an alcohol dehydrogenase enzyme (ADH) catalyzes a large group of reactions of the template:



- Mandelonitrile lyase catalyzes a single reaction:



- The interaction between the two is very specific, only via *benzaldehyde*, but this is not deducible from the enzyme network alone

Reaction graph

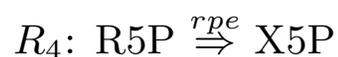
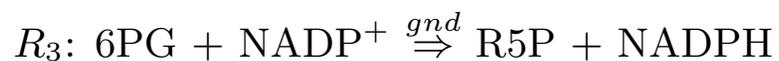
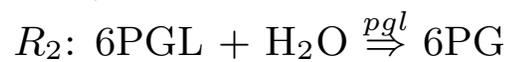
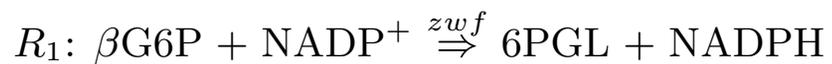
A reaction graph removes the reaction clumping property of enzyme networks.

- Nodes correspond to reactions
- A connecting edge between two reaction nodes R_1 and R_2 denotes that they share a metabolite

Difference to enzyme networks

- Each reaction catalyzed by an enzyme as a separate node
- A reaction is represented once, even if it has multiple catalyzing enzymes

Reaction graph example



Edge

Supporting metabolites

$(R_6, R_7): \beta\text{F6P}$

$(R_6, R_5): \alpha\text{G6P}$

$(R_5, R_7): \beta\text{G6P}$

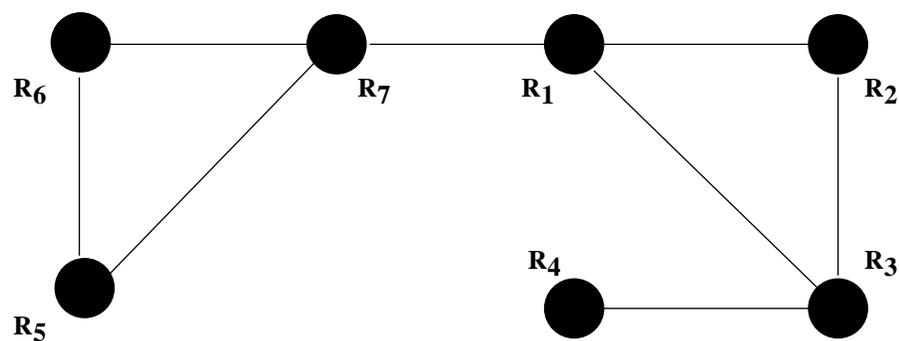
$(R_7, R_1): \beta\text{G6P}$

$(R_1, R_2): 6\text{PGL}$

$(R_1, R_3): \text{NADP}, \text{NADPH}$

$(R_2, R_3): 6\text{PG}$

$(R_3, R_4): \text{R5P}$

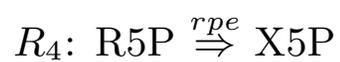
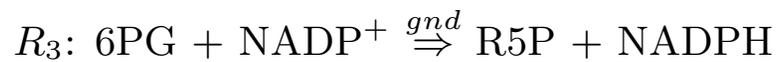
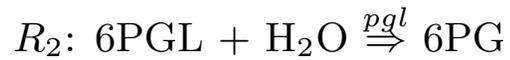
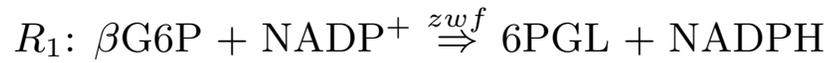


Substrate graph

A dual representation to a reaction graph is a substrate graph.

- Nodes correspond to metabolites
- Connecting edge between two metabolites A and B denotes that there is a reaction where both occur as substrates, both occur as products or one as product and the other as substrate
- A reaction $A + B \Rightarrow C + D$ is spread among a set of edges $\{(A, C), (A, B), (A, D), (B, C), (B, D), (C, D)\}$

Substrate graph example



Edge

(G6P,NADPH)

(G6P,6PGL)

(G6P,NADP⁺)

(NADP⁺,NADPH)

(NADP⁺, 6PGL)

(NADP⁺, 6PG)

(NADP⁺, R5P)

...

Common reactions

R_1

R_1

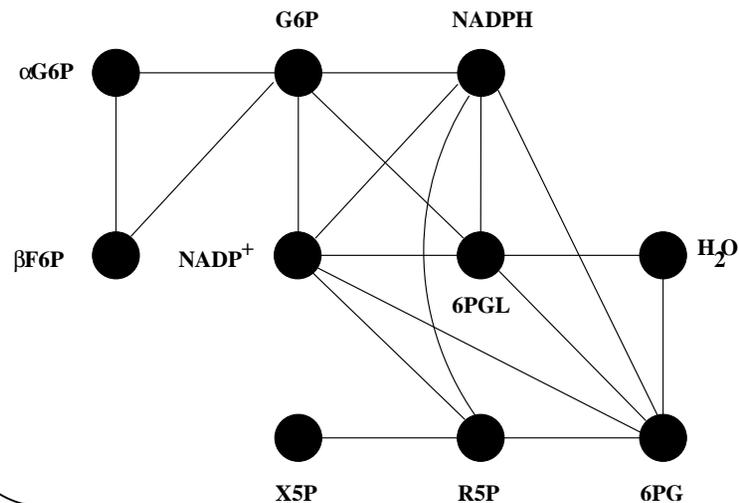
R_1

R_1, R_3

R_1

R_3

R_3



Graph analyses of metabolism

Enzyme interaction networks, reaction graphs and substrate graph can all be analysed in similar graph concepts and algorithms

We can compute basic statistics of the graphs:

- Connectivity of nodes: degree $k(v)$ of node v how many edges are attached to each node .
- Path length between pairs of nodes
- Clustering coefficient: how tightly connected the graph is

Clustering coefficient (1/2)

Clustering coefficient $C(v)$ for node v measures to what extent v is within a tight cluster

- Let $G = (V, E)$ be a graph with nodes V and edges E
- Let $\mathcal{N}(v)$ be the set of nodes adjacent to v
- The clustering coefficient is the relative number of edges between the nodes in $\mathcal{N}(v)$:

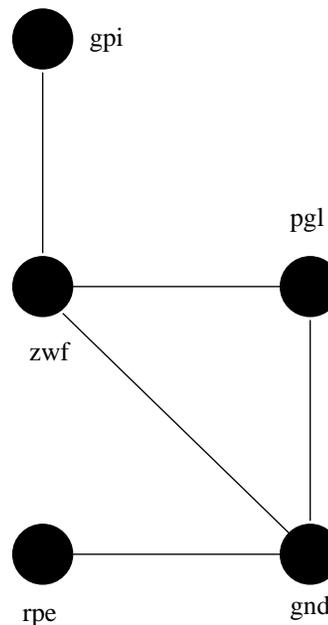
$$C(v) = \frac{|\{(v', v'') \in E \mid v', v'' \in \mathcal{N}(v)\}|}{N_{max}},$$

where $N_{max} = \max\{|\mathcal{N}(v)|(|\mathcal{N}(v)| - 1)/2, 1\}$

- Maximum $C(v) = 1$ occurs when $\mathcal{N}(v)$ is a fully connected graph
- Clustering coefficient of the whole graph is the node average:
 $C(G) = 1/n \sum_v (C(v))$

Clustering coefficient (2/2)

- $\mathcal{N}(gpi) = \{zwf\}$, $C(gpi) = 0$
- $\mathcal{N}(zwf) = \{gpi, pgl, gnd\}$,
 $C(zwf) = \frac{|\{(pgl, gnd)\}|}{3} = 1/3$
- $\mathcal{N}(pgl) = \{zwf, gnd\}$,
 $C(pgl) = \frac{|\{(zwf, gnd)\}|}{1} = 1/1$
- $\mathcal{N}(gnd) = \{rpe, zwf, pgl\}$,
 $C(gnd) = \frac{|\{(zwf, pgl)\}|}{3} = 1/3$
- $\mathcal{N}(rpe) = \{gnd\}$, $C(rpe) = 0$
- $C(G) = 1/3$



Comparison to graphs with known generating mechanism

One way to analyse our graphs is to compare the above statistics to graphs that we have generated ourselves and thus know the generating mechanism.

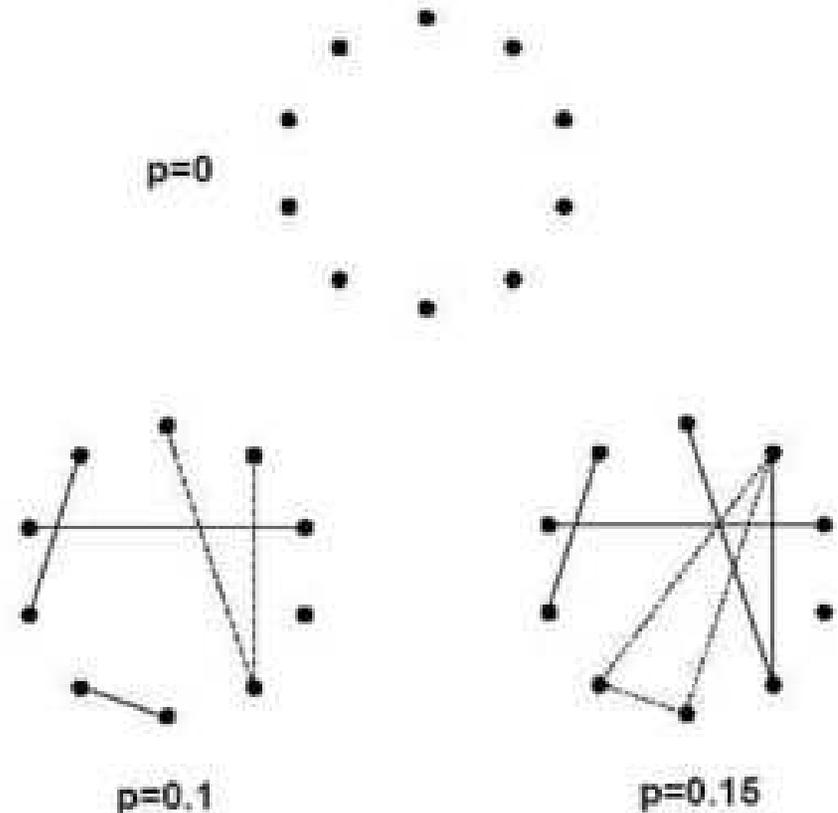
We will use the following comparison:

- Erdős-Renyi (ER) random graph
- Small-world graphs with preferential attachment

Erdős-Renyi random graph

Generation of ER graph:

- Start with a network with n nodes and no edges.
- Draw an edge between each pair of nodes is with probability p .



Preferential attachment

Preferential attachment (PA) is a mechanism that is proposed to generate many networks occurring in nature.

- Start with a small number n_0 of nodes and no edges.
- Iterate the following:
 - insert a new node v_j ,
 - draw $m \leq n_0$ edges from v_i existing nodes v_i with probability $p \sim \frac{k_i+1}{\sum_j (k_j+1)}$

When drawing new edges, nodes with many edges already are preferred over nodes with few or no edges.

Degree distributions of ER and PA graphs

$P(k)$ - The probability of encountering a node with degree k :

- Erdős-Renyi random graph:

$$P(k) \approx \binom{n}{k} p^k (1-p)^{n-k}.$$

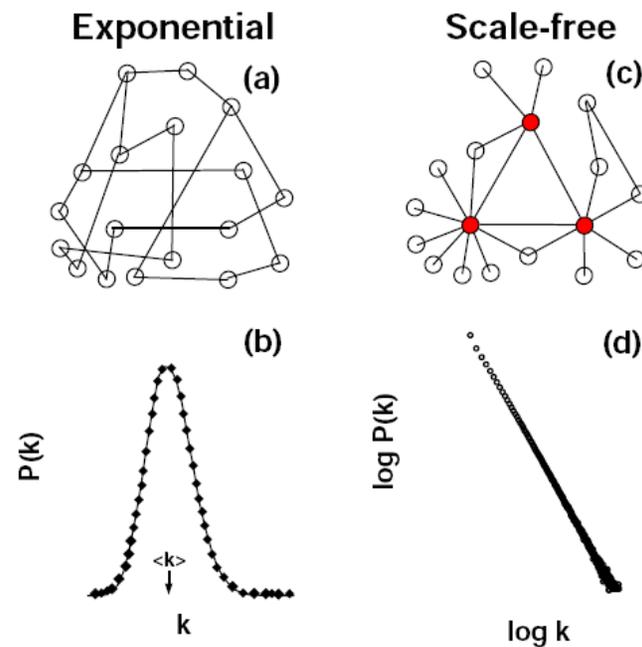
Distribution tightly peaked around the average degree: low variance.

Frequency of nodes with very high degree is low.

- Preferential attachment:

$P(k) \approx k^{-\gamma}$. Frequency distribution is scale-free: $\log P(k)$ and $\log k$ are linearly correlated. Distribution has

a fat tail: high variance and high number of nodes with high degree.



Degree distributions in metabolism (Wagner & Fell, 2000)

Degree distributions of substrate and reaction graphs

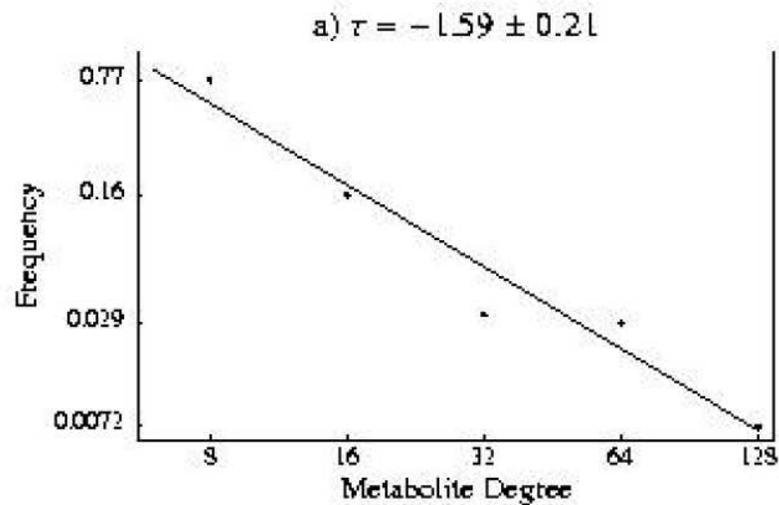
Table 1: Elementary Statistics of Substrate and Reaction Graphs. Shown are the number of nodes (n), the mean degree (k), and standard deviation in degree (σ_k) for reaction graph and substrate graph. For reference, standard deviation in degree is also shown for 100 numerically generated random graphs with the same n and k as those of the metabolic graphs. Two versions of each metabolic graph were analyzed, one in which the metabolites ATP, ADP, NAD, NADP, NADH, NADPH, CO₂, NH₃, SO₄, thioredoxin, phosphate and pyrophosphate were eliminated, and another one in which ATP, ADP, NAD, NADP, NADH, and NADPH were included. Upon removal of one or more metabolites, other vertices in the graph may become isolated. Such vertices were removed before analysis.

	n	k	σ_k	σ_k random graph
<i>Substrate Graph</i> <i>w/o ATP, ADP, NAD(P)(H)</i>	275	4.76	4.79	2.12±0.08
<i>Substrate Graph</i>	282	7.35	10.5	2.67±0.11
<i>Reaction Graph</i> <i>w/o ATP, ADP, NAD(P)(H)</i>		311	9.27	9.59 3.01± 0.12
<i>Reaction Graph</i>	315	28.3	29.1	5.04 ± 0.21

Degree distributions in metabolism

Substrate graph shows a fat-tailed distribution, consistent with a network generated via preferential attachment.

Substrate Graph: Degree Distribution



Small-world graphs

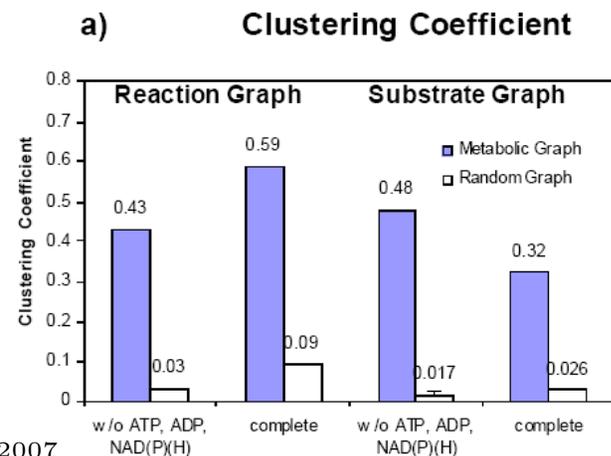
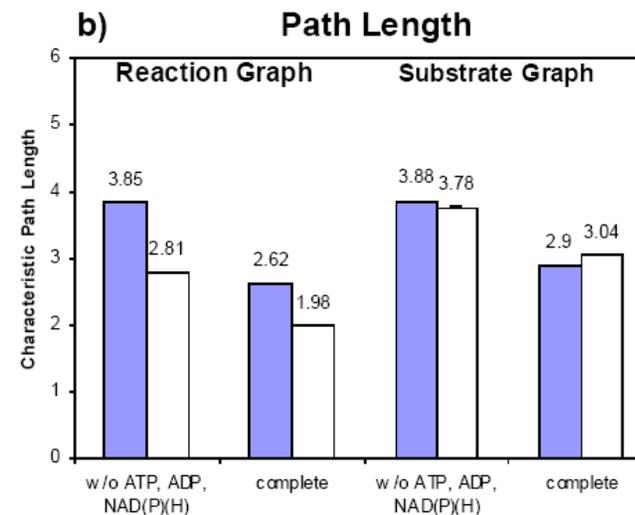
Graphs fulfilling the following two criteria are called small-world graphs

- Small average shortest path length between two nodes, the same level as ER graphs, lower than many regular graphs: shortcuts accross the graphs go via hubs
- High clustering coefficient compared to ER graph: the neighbors of nodes are more often linked than in ER graphs.

Graphs generated with preferential attachment are small-world graphs. However, small-world graphs can be generated with other mechanisms as well.

Metabolic graphs as small worlds

- Path lengths in reaction and substrate graphs are about the same as Erdős-Renyi random graph with the same average connectivity
- Clustering coefficients are much larger than in ER graphs
- Reaction and substrate graphs resemble small-world graphs

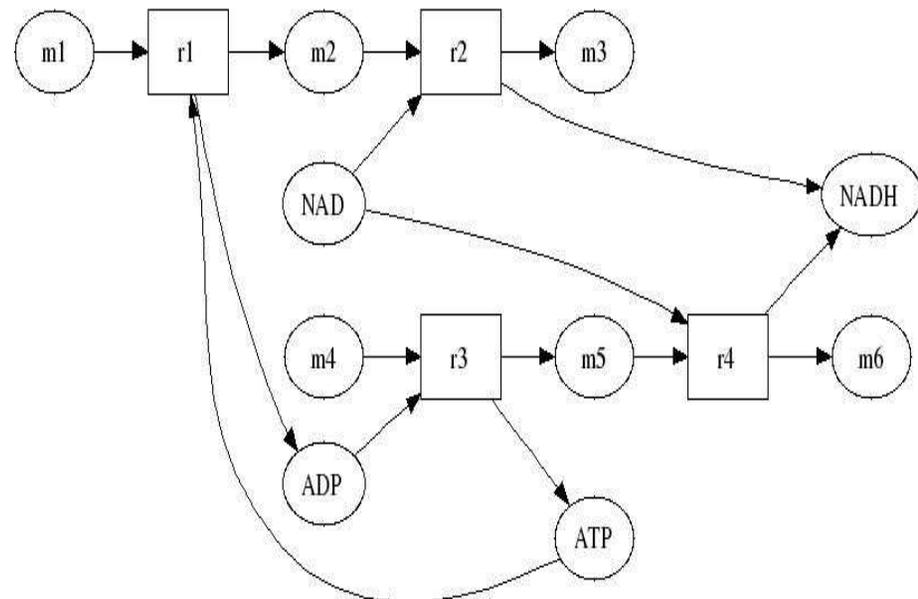


Pitfalls in substrate graph analysis: co-factors

Short paths in substrate graphs may not be biologically relevant

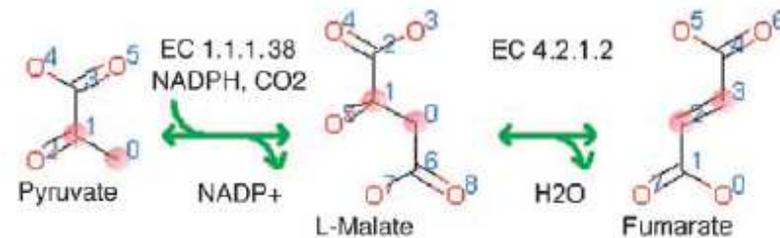
- Shortest paths between metabolites in otherwise distant parts of metabolism tend to go through co-factor metabolites (NADP, NAPH, ATP, ADP).
- However, transfer of atoms occurs only between the co-factors
- Quick remedy used in most studies: remove co-factors from the graph — but sometimes it is difficult 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)

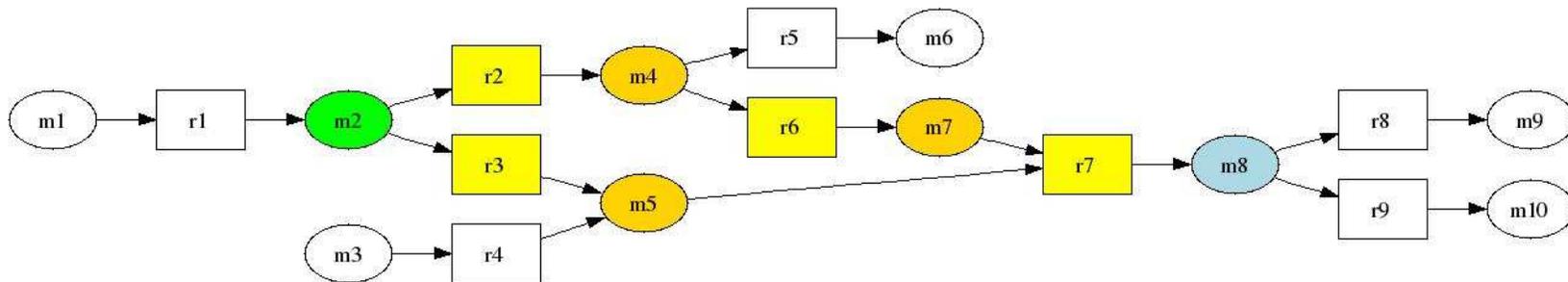


Pitfalls in substrate graph analysis: self-sufficiency

- The shortest path may not correlate well with the *effort* 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

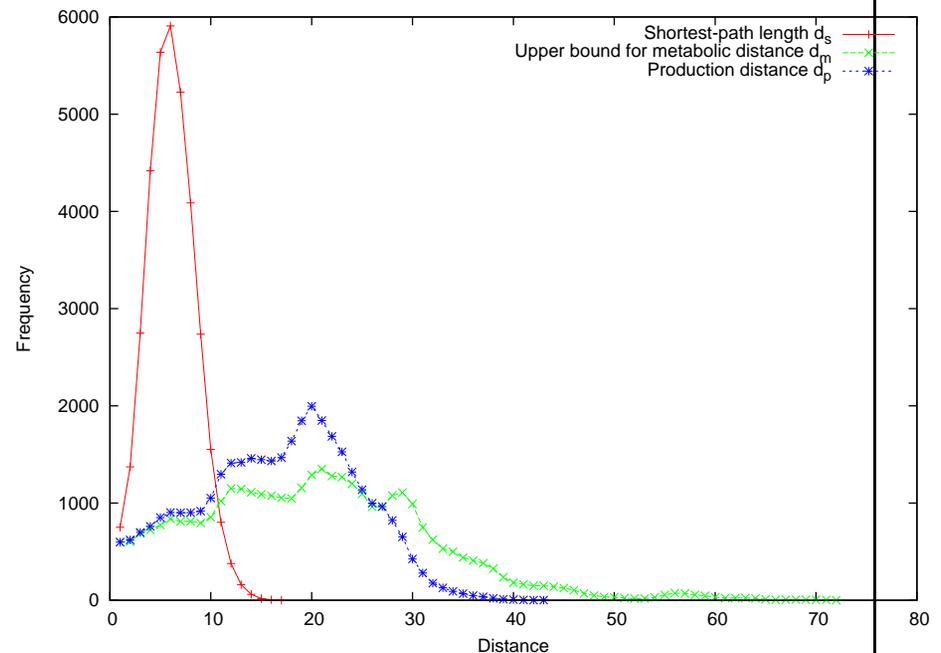
Feasible pathway vs. shortest simple path

- 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).

