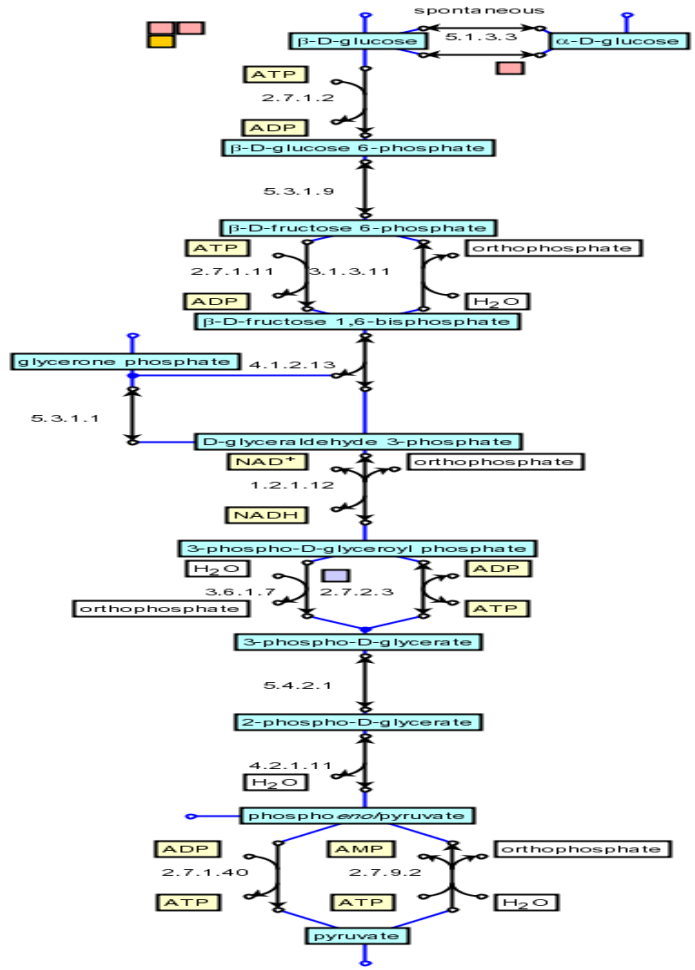
Representing metabolic networks

- In the following we assume that we posses a set of reactions composing our metabolic network, with catalyzing enzymes assigned
- How should we represent the network?
- For computational and statistical analyses, we need to be exact, much more so than when communicating between humans



(picture:E.Coli glycolysis, EMP database,

www.empproject.com/)

Levels of abstraction

- Everything relevant should be included in our representation
- What is relevant depends on the questions that we want to solve
- There are several levels of abstraction to choose from
  1. Graph representations: Connectivity of reactions/metabolites, structure of the metabolic network
  2. Stoichiometric (reaction equation) representation: capabilities of the network, flow analysis, steady-state analyses
  3. 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
- Boolean circuits (AND-OR graphs)

Example reaction List

- A set of reactions implementing a part of pentose-phosphate pathway of E. Coli
- Enzyme catalyzing the reaction annoted over the arrow symbol

R_1: β-D-glucose 6-phosphate (βG6P) + $NADP^+ \overset{zwf}{\Rightarrow}$ 6-phosphoglucono-lactone (6PGL) + NADPH

R_2: 6-phosphoglucono δ-lactone + $H_2O \overset{pgl}{\Rightarrow}$ 6-phosphogluconate (6PG)

R_3: 6-phosphogluconate + 1 $NADP^+ \overset{gnd}{\Rightarrow}$ ribulose 5-phosphate (R5P) + NADPH

R_4: ribulose 5-phosphate (R5P) $\overset{rpe}{\Rightarrow}$ xylulose 5-phosphate (X5P)

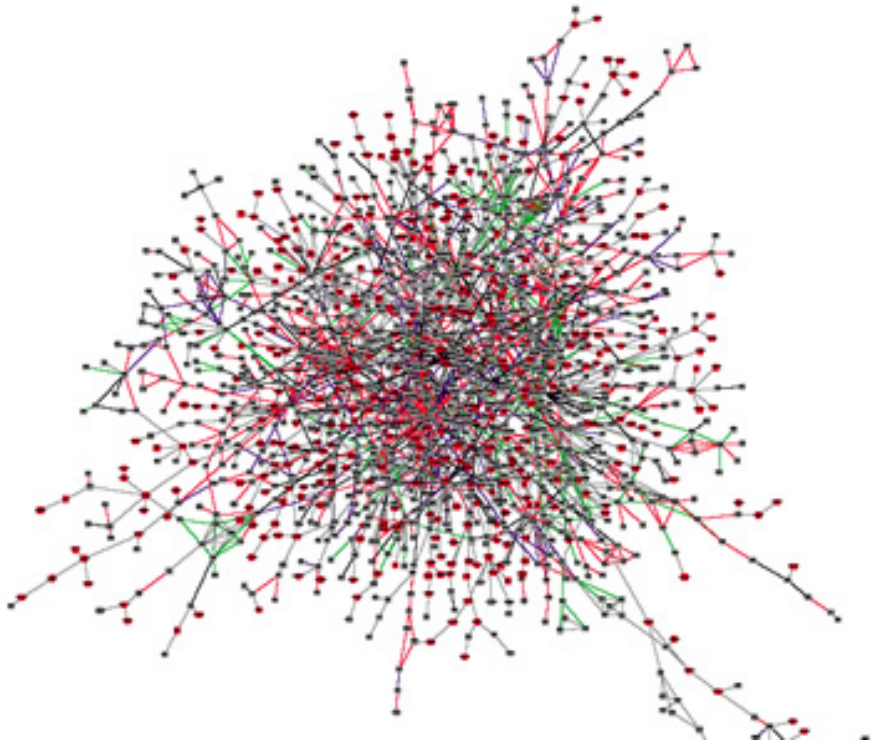
R_5: α-D-glucose 6-phosphate(αG6P) $\overset{gpi}{\Leftrightarrow}$ βG6P

R_6: α-D-glucose 6-phosphate(αG6P) $\overset{gpi}{\Leftrightarrow}$ β-D-Fructose-6-phosphate (βF6P)

R_7: β-D-glucose 6-phosphate(αG6P) $\overset{gpi}{\Leftrightarrow}$ βF6P

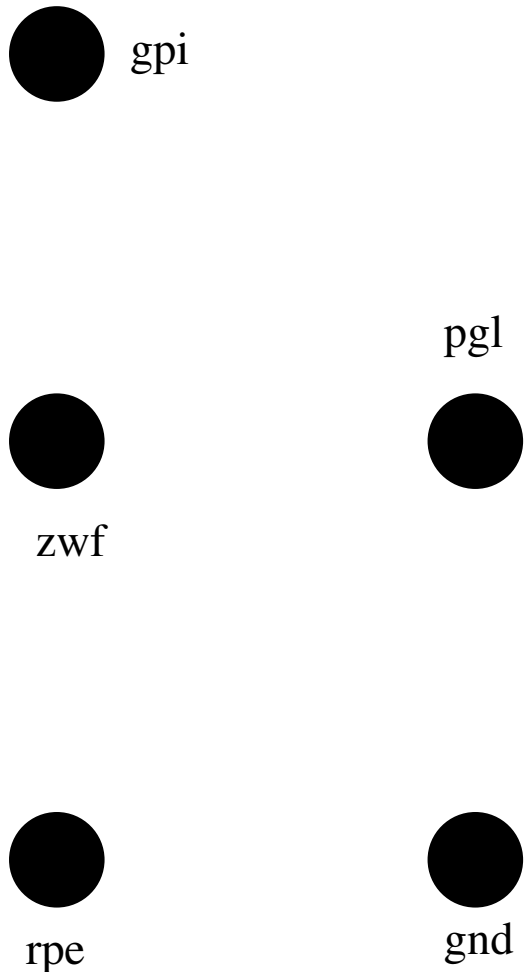
Enzyme interaction networks

- 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 construction

- In our pathway, we have 5 enzymes catalyzing a total of 7 reactions



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

Enzyme interaction network construction

- We take each pair of enzymes in turn
- Draw and edge between if they share metabolites
- (gpi,zwf) — βG6P

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

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

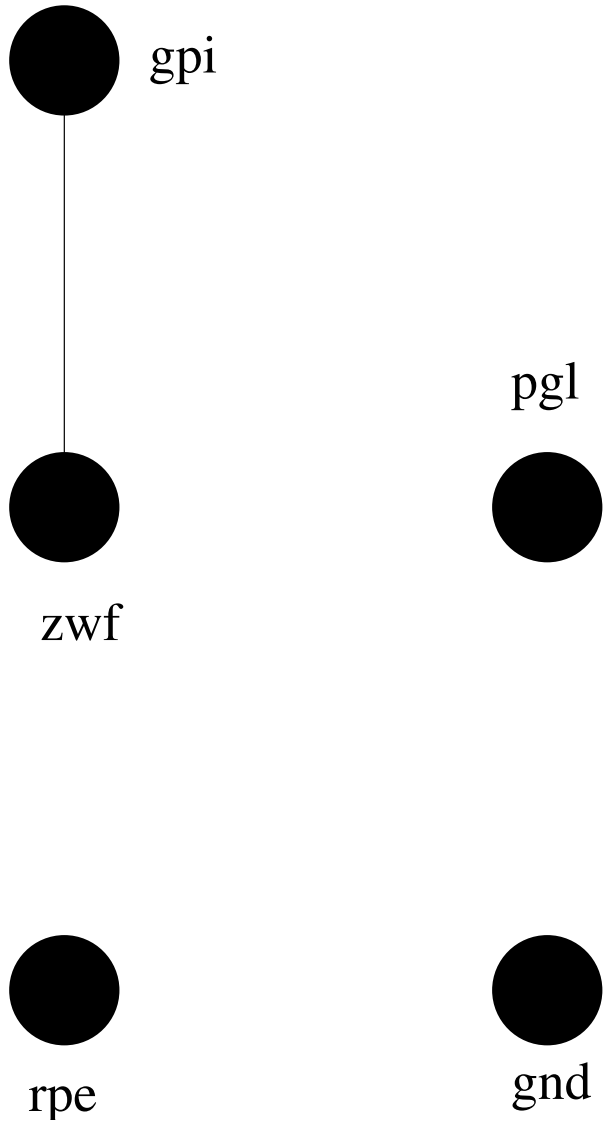
R_3: 6PG + NADP$^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH

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



Enzyme interaction network construction

- We take each pair of enzymes in turn
- Draw and edge between if they share metabolites
- (zwf, pgl) — 6PGL

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

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

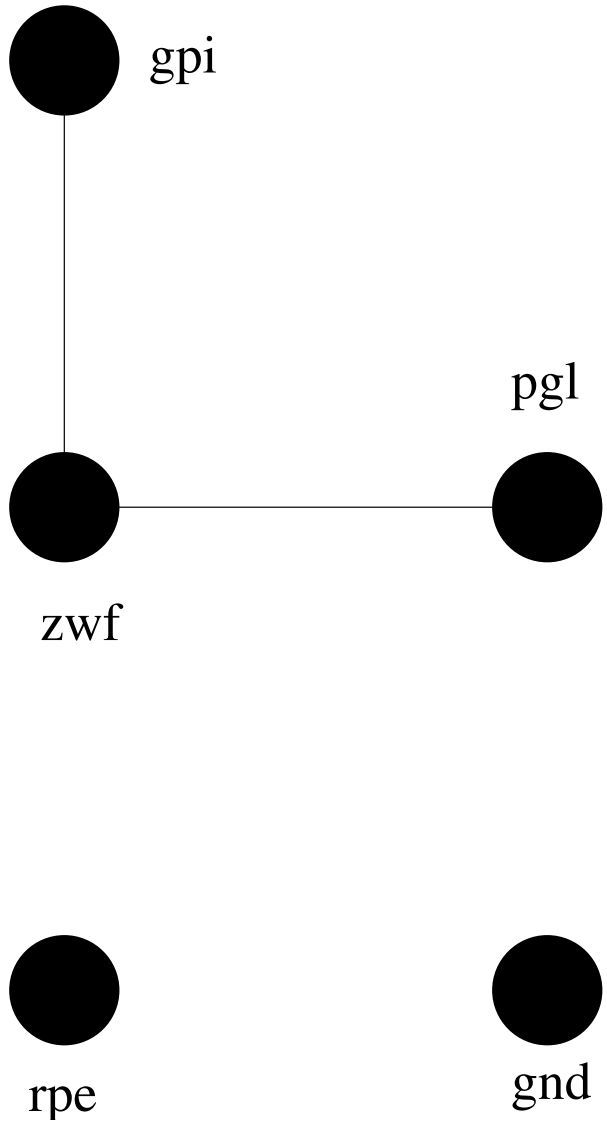
R_3: 6PG + NADP$^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH

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

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

R_6: αG6P $\overset{gpi}{\Leftrightarrow}$ βF6P

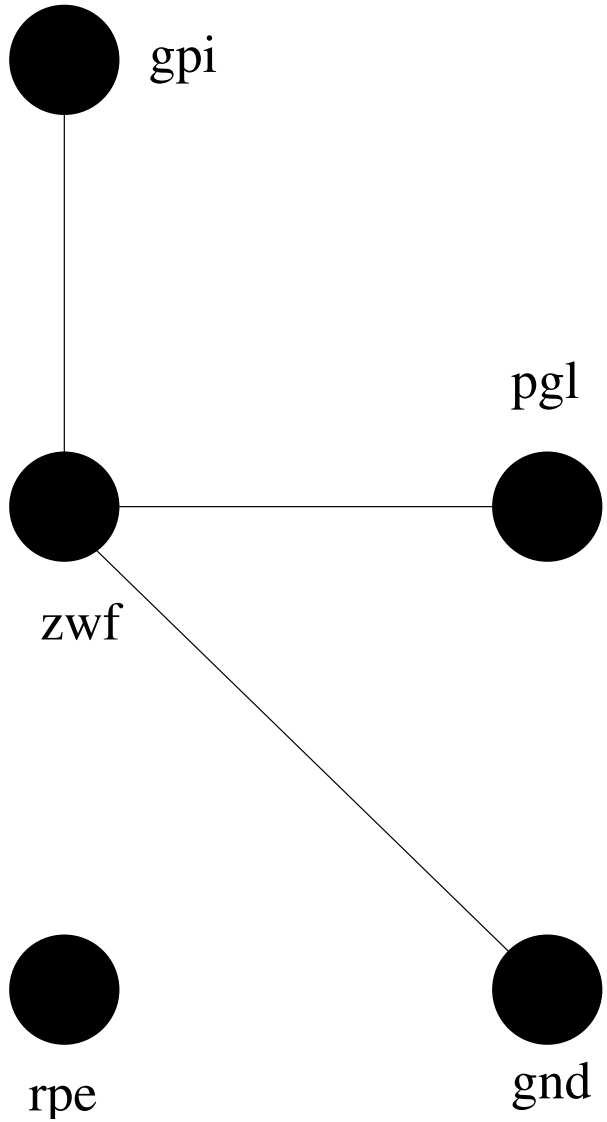
R_7: βG6P $\overset{gpi}{\Leftrightarrow}$ βF6P



Enzyme interaction network construction

- (zwf, gnd) — NADP, NADPH
- (zwf, rpe) — $\emptyset$

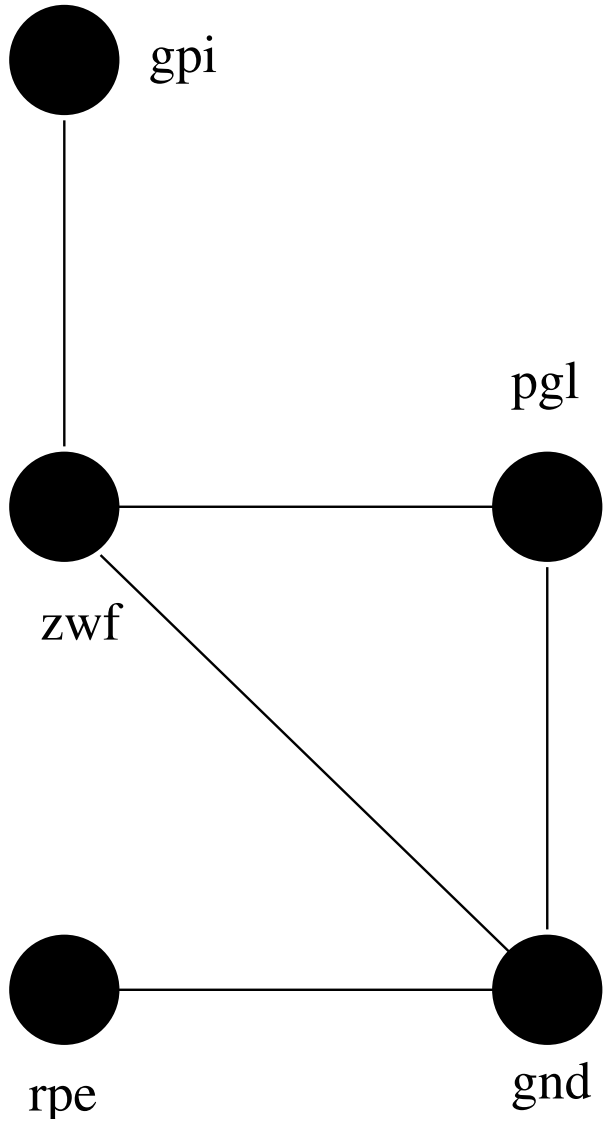
R_1: βG6P + NADP$^+$ $\overset{zwf}{\Rightarrow}$ 6PGL + NADPH
R_2: 6PGL + H_2O $\overset{pgl}{\Rightarrow}$ 6PG
R_3: 6PG + NADP$^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH
R_4: R5P $\overset{rpe}{\Rightarrow}$ X5P
R_5: αG6P $\overset{gpi}{\Leftrightarrow}$ βG6P
R_6: αG6P $\overset{gpi}{\Leftrightarrow}$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



Enzyme interaction network construction

- (pgl,gnd) — 6PG
- (gnd, rpe) — R5P

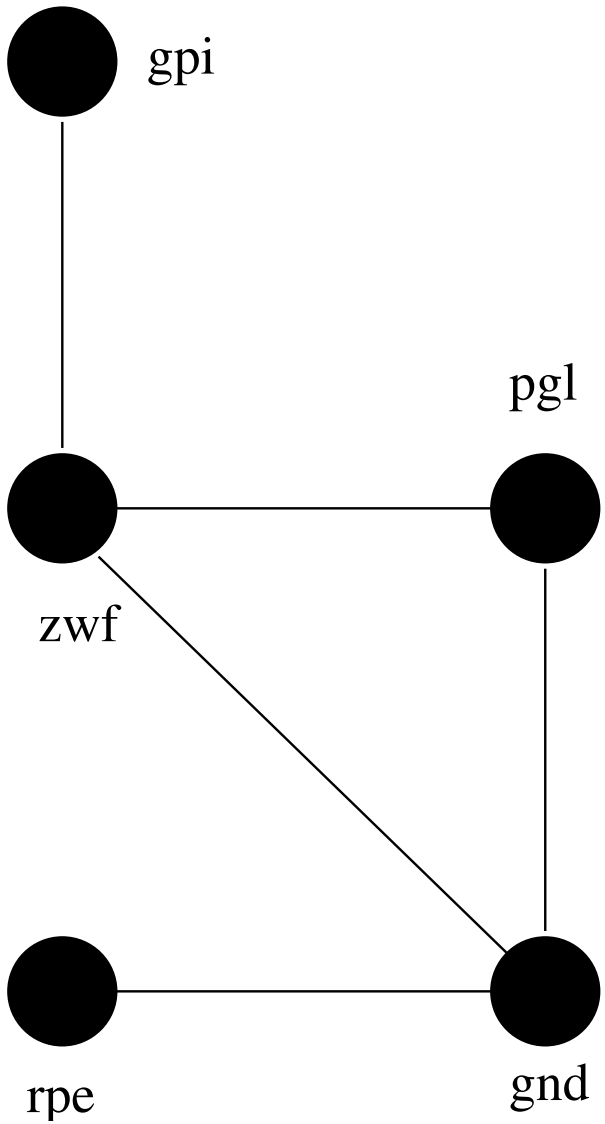
R_1: βG6P + NADP$^+$ $\overset{zwf}{\Rightarrow}$ 6PGL + NADPH
R_2: 6PGL + H_2O $\overset{pgl}{\Rightarrow}$ 6PG
R_3: 6PG + NADP$^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH
R_4: R5P $\overset{rpe}{\Rightarrow}$ X5P
R_5: αG6P $\overset{gpi}{\Leftrightarrow}$ βG6P
R_6: αG6P $\overset{gpi}{\Leftrightarrow}$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



Enzyme interaction network construction

- (zwf, gnd) — NADP, NADPH
- (pgl,gnd) — 6PG
- (gnd, rpe) — R5P

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



Reaction clumping in Enzyme networks

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

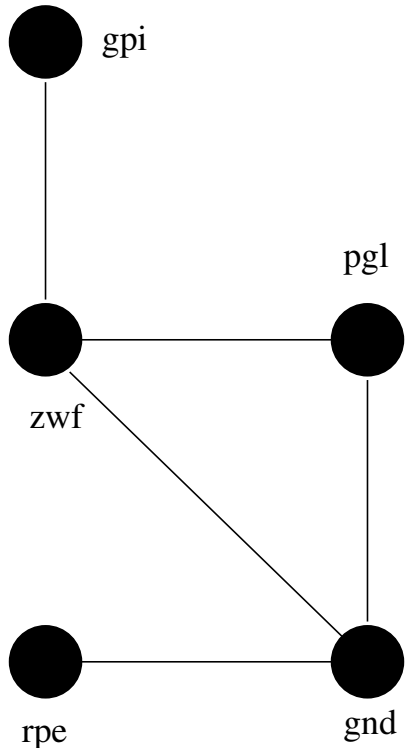
- For example, an alcohol dehydrogenase enzyme (ADH) catalyzes a large group of reactions of the template: `an alcohol + NAD+ <=> an aldehyde or ketone + NADH + H+`
- Mandelonitrile lyase catalyzes a single reaction: `Mandelonitrile <=> Cyanide + Benzaldehyde`
- The interaction between the two is very specific, only via *benzaldehyde*, but this is not deducible from the enzyme network alone

Co-factor effects

- A group of "currency molecules" (ATP,ADP,NAD,NADH, NADP, NADPH) act as co-factors in many reactions
- The reactions that share the co-factors may not otherwise have anything in common
- Sharing a co-factor induces an arc between reactions.
- This can be misleading, unless we are specifically interested in co-factors

zwf: $G6P + NADP^+ \Rightarrow 6PGL + NADPH$

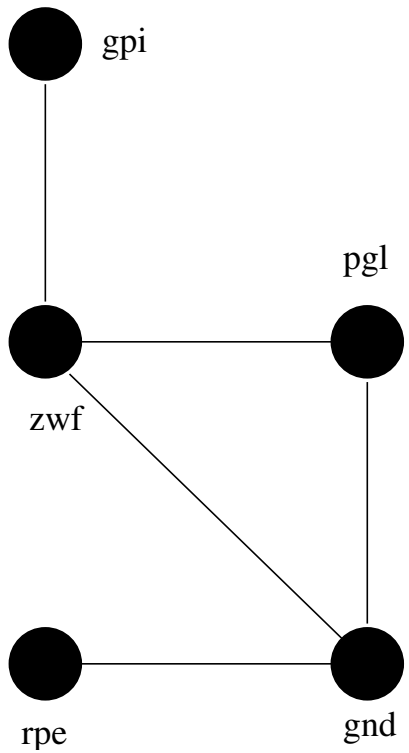
gnd: $6PG + NADP^+ \Rightarrow R5P + NADPH$



Co-factor effects

- For example, the edge (zwf,gnd) in our example network arises solely because of the co-factor molecules (NADP,NADPH)
- This fact cannot be decuded from the enzyme network
- Chance to be mislead?

zwf: G6P + $NADP^+$ $\Rightarrow$ 6PGL + NADPH
gnd: 6PG + $NADP^+$ $\Rightarrow$ R5P + NADPH



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

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

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

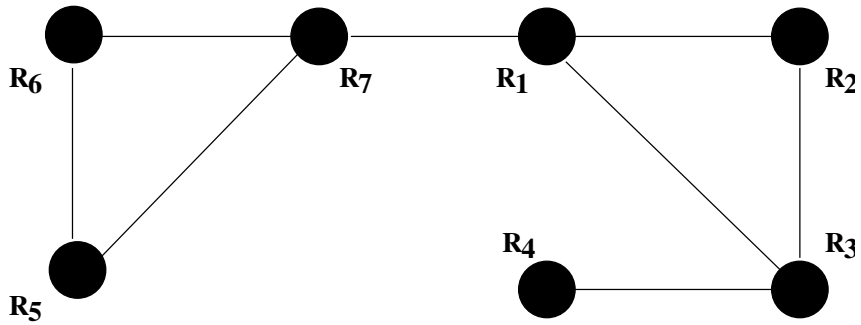
R_3: 6PG + NADP$^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH

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



Edge	Supporting metabolites
(R_6, R_7):	βF6P
(R_6, R_5):	αG6P
(R_5, R_7)	βG6P
(R_7, R_1):	βG6P
(R_1, R_2):	6PGL
(R_1, R_3):	NADP, NADPH
(R_2, R_3):	6PG
(R_3, R_4):	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

- Add and edge between all molecule pairs in R1
- (G6P,NADPH), (G6P,6PGL), (G6P,$NADP^+$), ($NADP^+$,NADPH), ($NADP^+$, 6PGL), (6PGL, NADPH)

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

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

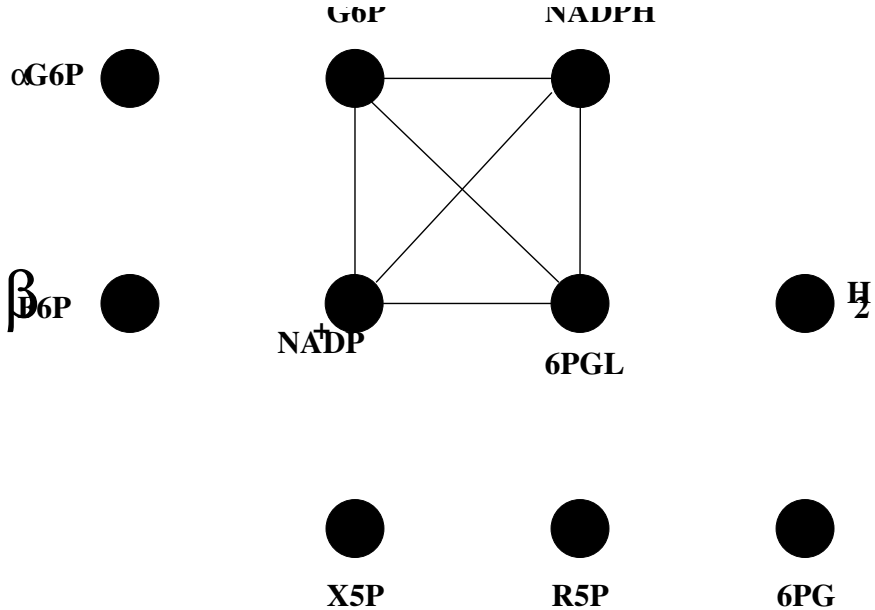
R_3: 6PG + $NADP^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH

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



Substrate graph example

- Add and edge between all molecule pairs in R2
- (6PGL,6PG), (6PGL,H_2O), (6PG,H_2O)

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

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

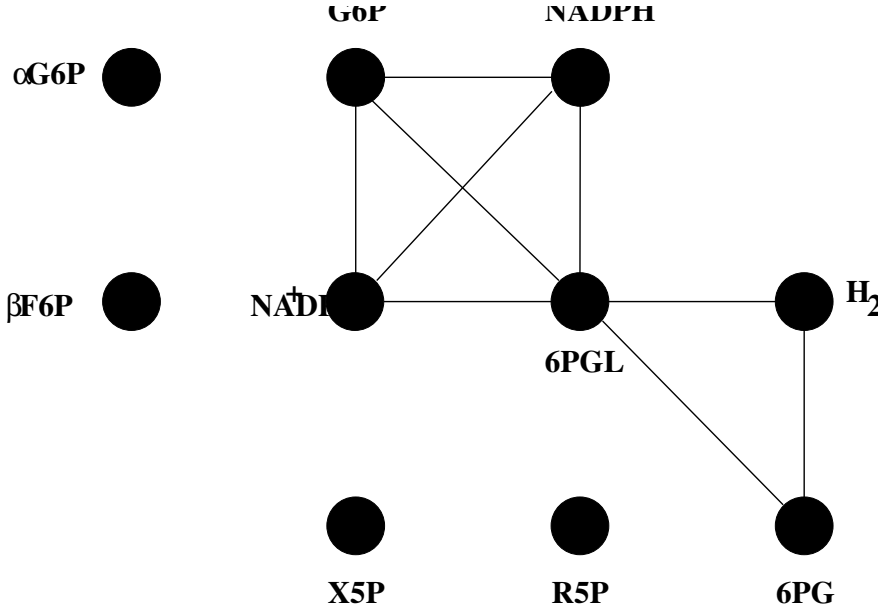
R_3: 6PG + $NADP^+ \overset{gnd}{\Rightarrow}$ R5P + NADPH

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



Substrate graph example

- Add and edge between all molecule pairs in R3
- (6PG,$NADP^+$),(6PG,R5P), (6PG,NADPH), ($NADP^+$, R5P), (R5P,NADPH)

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

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

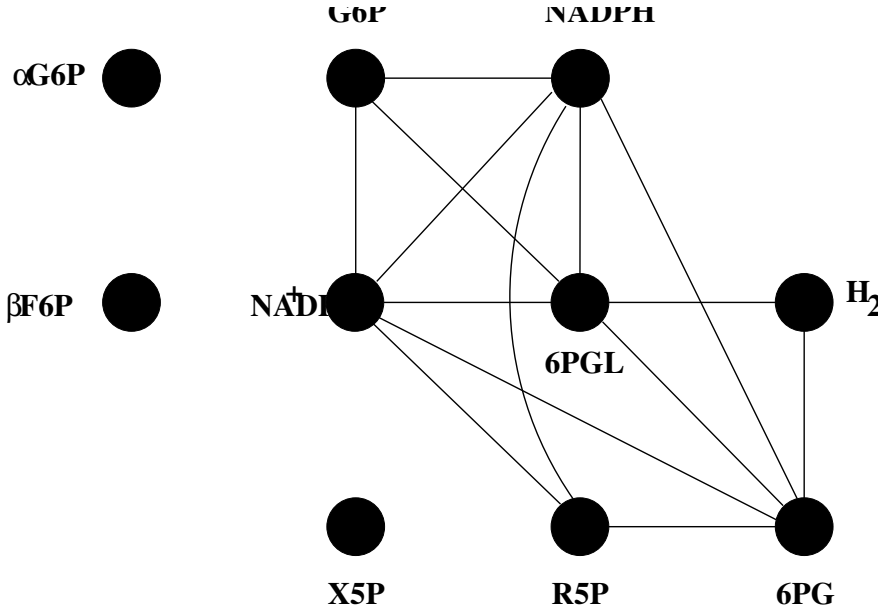
R_3: 6PG + $NADP^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH

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



Substrate graph example

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

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

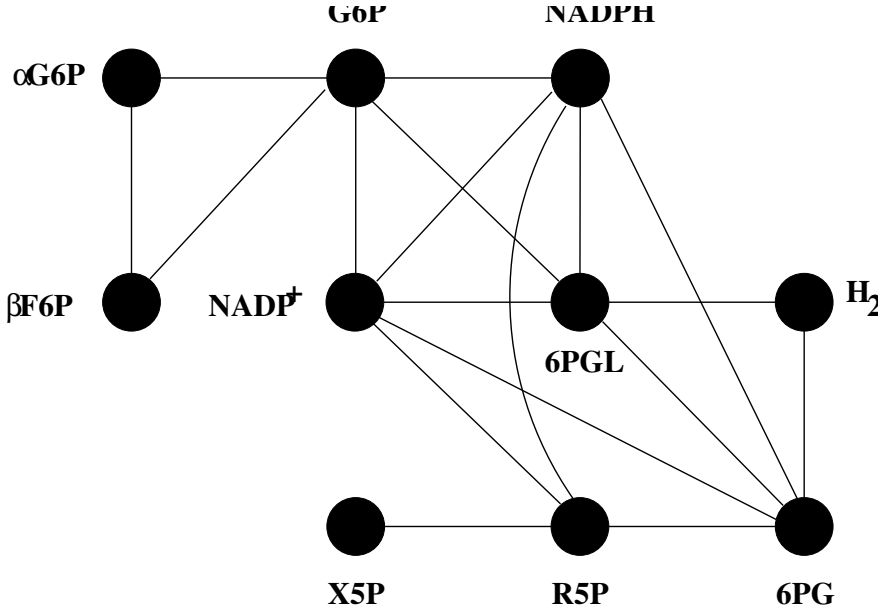
R_3: 6PG + NADP$^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH

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



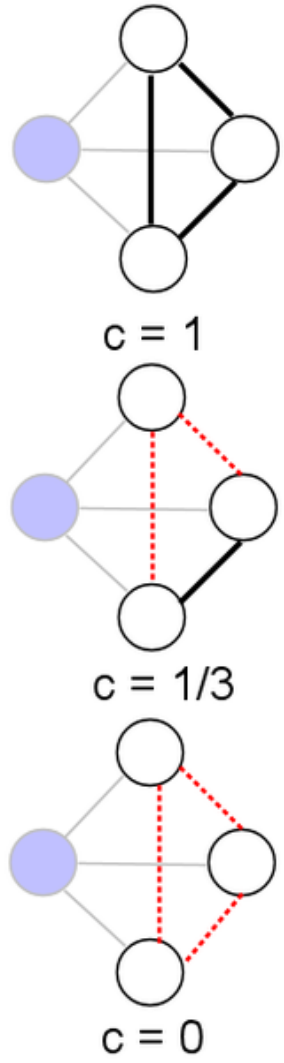
Graph analyses of metabolism

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 coeefficient: how tightly connected the graph is

Clustering coefficient

- Clustering coefficient measures the connectivity of graph around single nodes
- Informally: How close to a fully connected graph are the neighbors of given node v, if we remove the node v and all edges adjacent to it
- In the example on the right, the clustering coefficient of the blue node is given for three different neighborhoods



Clustering coefficient formally

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 $\mathcal{N}(v)$:

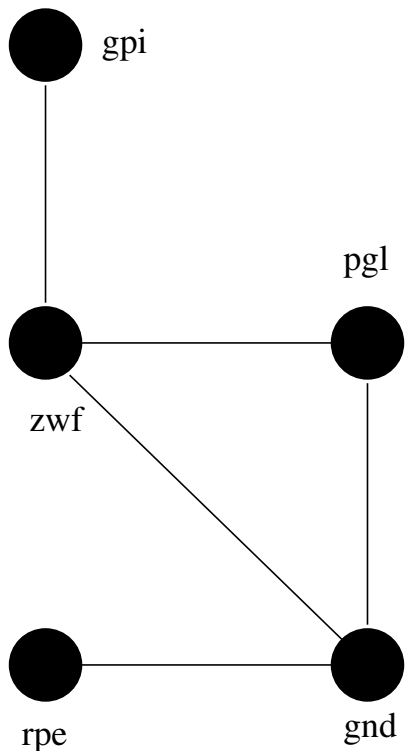
$$C(v) = \frac{|\{(v', v'') \in E | v', v'' \in \mathcal{N}(v)\}|}{N_{max}},$$

  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 in our enzyme network

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

- $C(G) = 1/3$



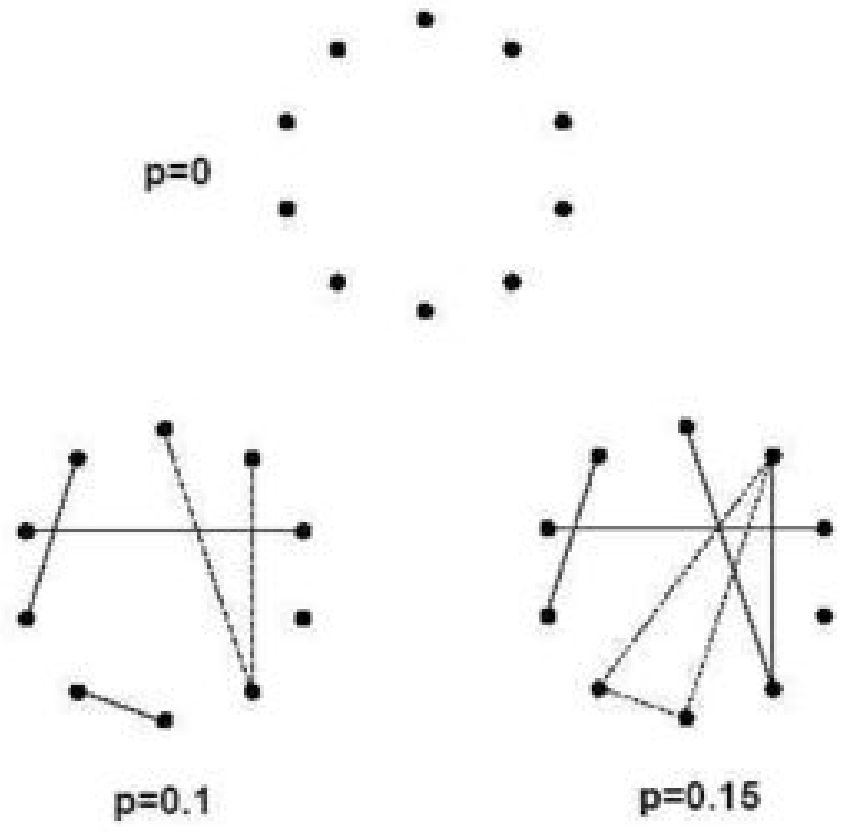
Comparison to graphs with known generating mechanism

One way to analyse our graphs is to compare the above statistics to graphs that we have generated ourself and thus know the generating mechanism.
We will use the following comparison:

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

Erdös-Renyi random graph

- Well studied model for random graphs proposed by Paul Erdös and Alfred Renyi in 1959
- 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.



Properties of ER graph

- ER graph of size n has on average $\binom{n}{2}p$ edges.
- Node degree distribution of is binomial

$$P(deg(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

- The connectivity of ER graph follows directly from the quantities n and p:
  - $p < \frac{(1-\epsilon)\ln n}{n}$: graph will almost surely be non-connected
  - $p > \frac{(1+\epsilon)\ln n}{n}$: graph will almost surely be connected
  - $np < 1$: graph will almost surely have no large connected components, otherwise almost surely will have one
- Due to its mathematical elegance, ER model has been very popular subject of study in graph theory

ER networks and biology

- It has been observed that the ER graph is not a good explanation for the generating mechanism of many biological networks
- Prime symptom is that the node degree distribution of biological networks does not fit to binomial distribution
- In particular, biological networks often have so called hubs, nodes with very high connectivity

Preferential attachment

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

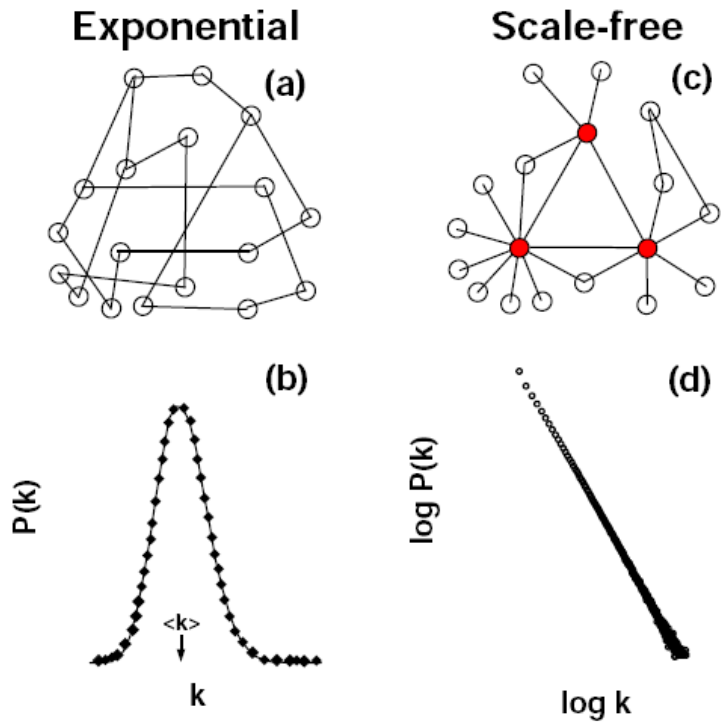
- Start with a small number n_0 of nodes and no edges.
- Iterate the following:
  - insert a new node v,
  - draw $m \leq n_0$ edges from v to existing nodes v_i with probability $p \sim \frac{k_i+1}{\sum_j (k_j+1)}$

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

- *Hubs*, i.e. highly connected nodes, will emerge from the generating process

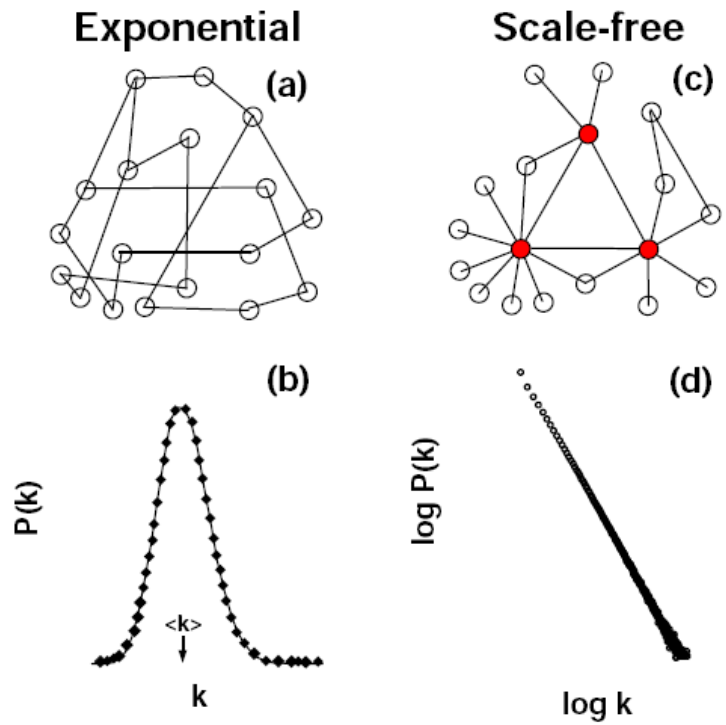
Degree distributions of ER and PA graphs

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



Degree distributions of ER and PA graphs

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

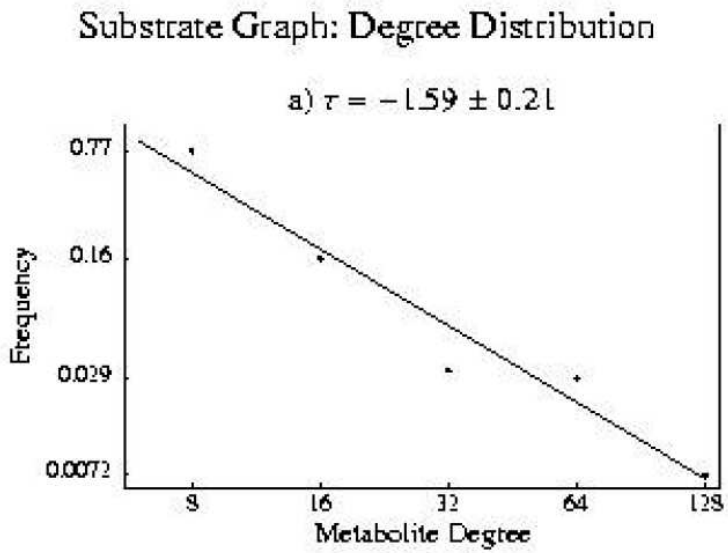
Degree distributions of substrate and reaction graphs

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, CO_2, NH_3, SO_4, 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
Substrate Graph w/o ATP, ADP, NAD(P)(H)	275	4.76	4.79	2.12±0.08
Substrate Graph	282	7.35	10.5	2.67±0.11
Reaction Graph w/o ATP, ADP, NAD(P)(H)	311	9.27	9.59	3.01± 0.12
Reaction Graph	315	28.3	29.1	5.04 ± 0.21

Degree distributions in metabolism (Wagner & Fell, 2000)

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



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.

Small-world graphs

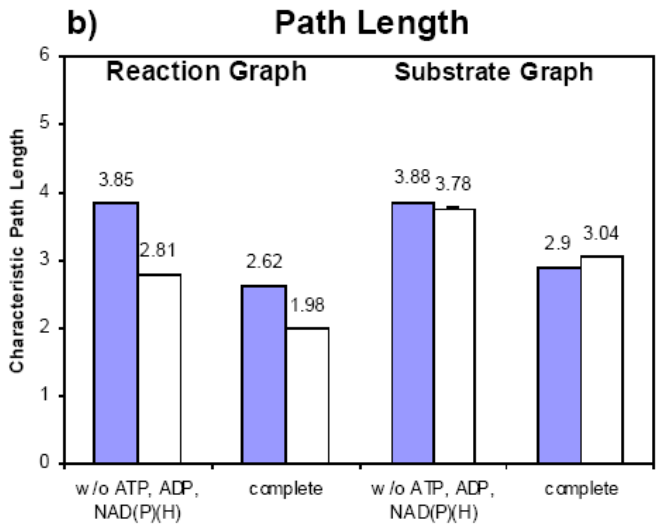
Graphs fulfilling the following two criteria are called small-world graphs

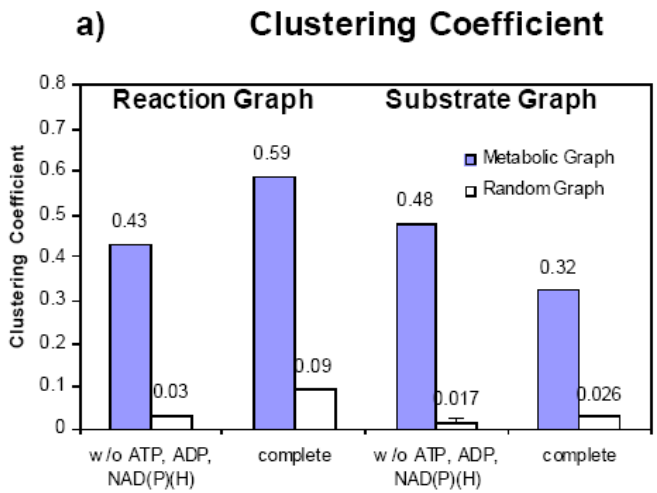
- Small average shortest path length between two nodes
  - The same level as ER graphs, lower than many regular graphs:
  - Shortcuts accross the graphs go via hubs
- High clustering coefficient compared to ER graph: the neighbors of nodes are more often linked than in ER graphs.

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

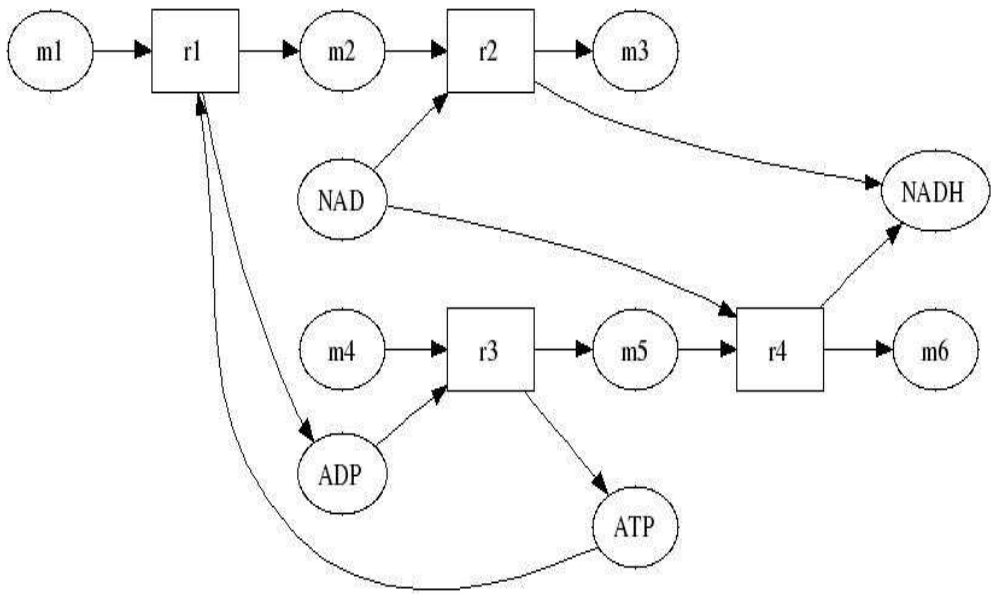
- Path lengths in reaction and substrate graphs are about the same as Erdös-Renyi random graph with the same average connectivity
- Clustering coefficients are much larger than in ER graphs
- The graphs resemble small-world graphs





Pitfalls in substrate graph analysis: co-factors

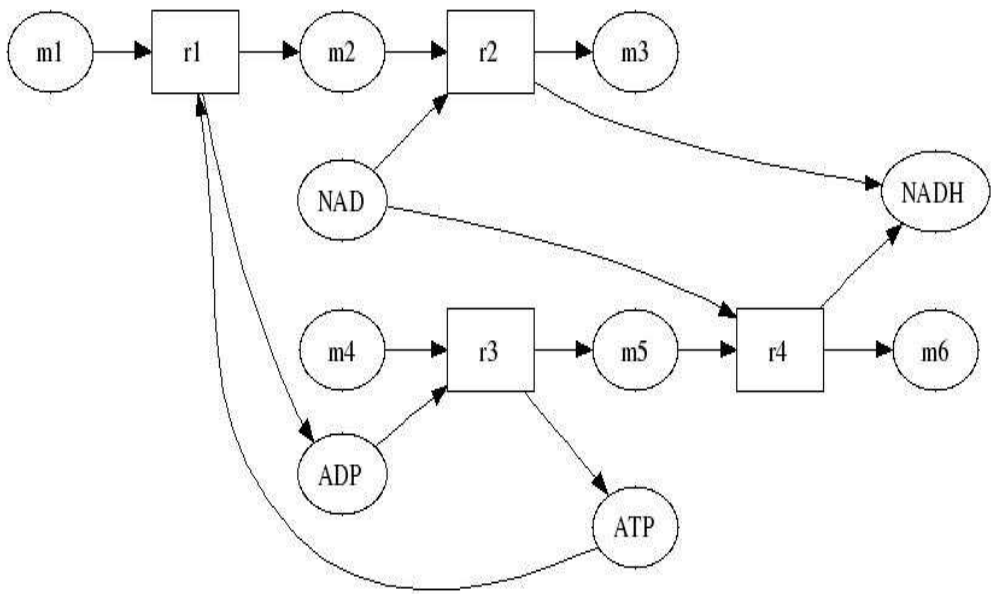
- Path length in substrate graphs may not be biologically relevant
- Shortest paths between metabolites in otherwise distant parts of metabolism tend to go through co-factor metabolites (NADP, NAPH, ATP, ADP).
- However, transfer of atoms occurs only between the co-factors



Pitfalls in substrate graph analysis: co-factors

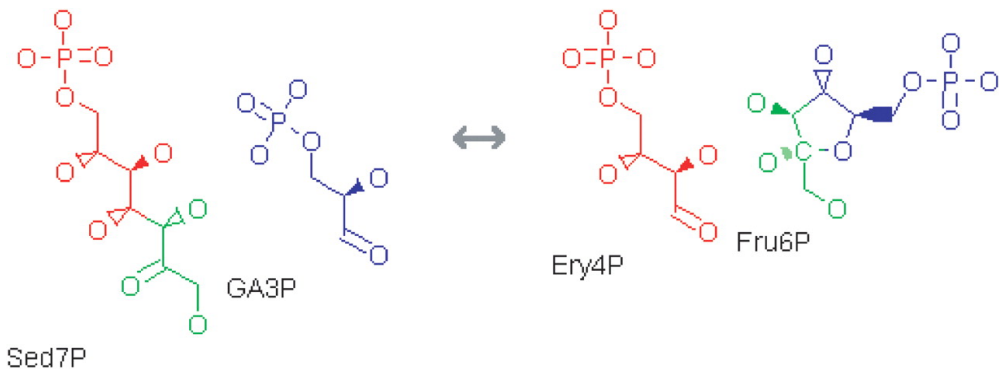
Quick remedy used in most studies:

- 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)

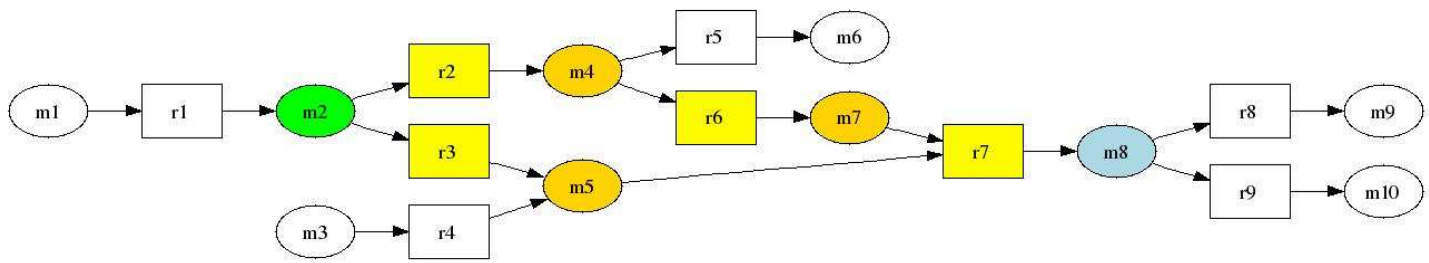


Pitfalls in substrate graph analysis: self-suffiency

- 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

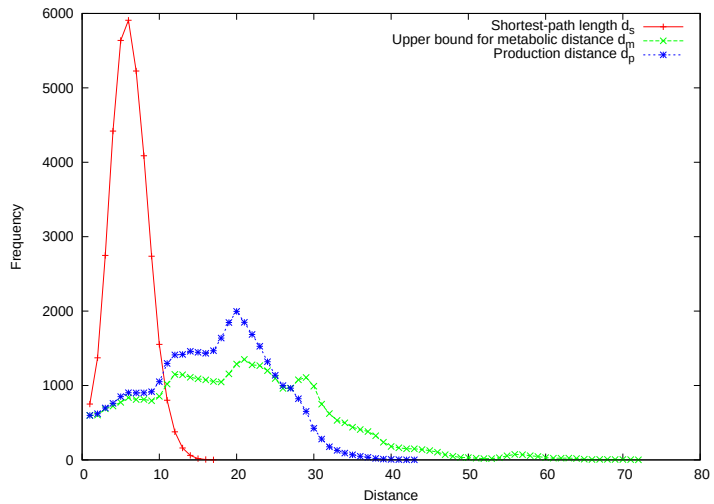
Feasible pathway vs. shortest simple path

- 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).