V13 Graph connectivity – Metabolic networks

In the first half of this lecture section, we use the theory of network flows to give constructive proofs of Menger‘s theorem.

These proofs lead directly to algorithms for determining the edge-connectivity and vertex-connectivity of a graph.

The strategy to prove Menger‘s theorems is based on properties of certain **networks** whose arcs all have **unit capacity**.

These **0-1 networks** are constructed from the original graph.

Determining the connectivity of a graph

Lemma 12.3.1. Let *N* be an *s-t* network such that

outdegree(s) > *indegree(s),*

indegree(t) > *outdegree (t),* and

outdegree(v) = *indegree(v)* for all other vertices *v.*

Then, there exists a directed *s-t* path in network *N.*

Proof. Let *W* be a longest directed trail (trail = walk without repeated edges; path = trail without repeated vertices) in network *N* that starts at source *s*, and let *z* be its terminal vertex.

If vertex *z* were not the sink *t,* then there would be an arc not in trail *W* that is directed from *z* (since *indegree(z)* = *outdegree(z)).*

But this would contradict the maximality of trail *W.*

Thus, *W* is a directed trail from source *s* to sink *t.*

If *W* has a repeated vertex, then a part of *W* determines a directed cycle, which can be deleted from *W* to obtain a shorter directed *s-t* trail.

This deletion step can be repeated until no repeated vertices remain, at which point, the resulting directed trail is an *s-t* path. □

Determining the connectivity of a graph

Proposition 12.3.2. Let *N* be an *s-t* network such that

outdegree(s) – indegree(s) = m = indegree(t) – outdegree (t),

and *outdegree(v) = indegree(v)* for all vertices $v \neq s,t$.

Then, there exist *m* disjoint directed *s-t* path in network *N.*

Proof. If *m* = 1, then there exists an open eulerian directed trail *T* from source *s* to sink *t* by Theorem 6.1.3.

Review: An eulerian trail in a graph is a trail that visits every edge of that graph exactly once.

Theorem 6.1.3. A connected digraph *D* has an open eulerian trail from vertex *x* to vertex *y* if and only if *indegree(x)* + 1 = *outdegree(x), indegree(y)* = *outdegree(y)* + 1, and all vertices except *x* and *y* have equal indegree and outdegree.
Euler proved that a necessary condition for the existence of Eulerian circuits is that all vertices in the graph have an even degree.

Theorem 1.5.2. Every open *x-y* walk *W* is either an *x-y* path or can be reduced to an *x-y* path.

Therefore, trail *T* is either an *s-t* directed path or can be reduced to an *s-t* path.

Determining the connectivity of a graph

By way of induction, assume that the assertion is true for $m = k$, for some $k \geq 1$, and consider a network N for which the condition holds for $m = k + 1$.
There does exist at least one directed s-t path P by Lemma 12.3.1.

If the arcs of path P are deleted from network N, then the resulting network $N - P$ satisfies the condition of the proposition for $m = k$.

By the induction hypothesis, there exist k arc-disjoint directed s-t paths in network $N - P$. These k paths together with path P form a collection of $k + 1$ arc-disjoint directed s-t paths in network N. □

Basic properties of 0-1 networks

Definition A **0-1 network** is a capacitated network whose arc capacities are either 0 or 1.

Proposition 12.3.3. Let *N* be an *s-t* network such that *cap(e)* = 1 for every arc *e.* Then the value of a maximum flow in network *N* equals the maximum number of arc-disjoint directed *s-t* paths in *N.*

Proof: Let *f** be a maximum flow in network *N*, and let *r* be the maximum number of arc-disjoint directed *s-t* paths in *N.*
Consider the network *N** obtained by deleting from *N* all arcs *e* for which *f*(e)* = 0.
Then *f*(e)* = 1 for all arcs *e* in network *N*.*
It follows from the definition that for every vertex *v* in network *N*,*

$$\sum_{e \in Out(v)} f^*(e) = |Out(v)| = outdegree(v)$$

and

$$\sum_{e \in In(v)} f^*(e) = |In(v)| = indegree(v)$$

Basic properties of 0-1 networks

Thus by the definition of *val(f*)* and by the conservation-of-flow property,

$$outdegree(s) - indegree(s) = val(f^*) = indegree(t) - outdegree(t)$$

and $outdegree(v) = indegree(v)$, for all vertices $v \neq s,t$.

By Proposition 12.3.2., there are *val(f*)* arc-disjoint *s-t* paths in network *N*,* and hence, also in *N*, which implies that $val(f^*) \leq r$.

To obtain the reverse inequality, let $\{P_1, P_2, \ldots, P_r\}$ be the largest collection of arc-disjoint directed *s-t* paths in *N,* and consider the function $f: E_N \rightarrow R^+$ defined by

$$f(e) = \begin{cases} 1, & \text{if some path } P_i \text{ uses arc } e \\ 0, & \text{otherwise} \end{cases} .$$

Then *f* is a feasible flow in network *N,* with *val(f)* = *r.*
It follows that $val(f^*) \geq r$. □

Separating Sets and Cuts

Review from §5.3

Let *s* and *t* be distinct vertices in a graph *G.* An *s-t* **separating edge set** in *G* is a set of edges whose removal destroys all *s-t* paths in *G.*

Thus, an *s-t* separating edge set in *G* is an edge subset of E_G that contains at least one edge of every *s-t* path in *G.*

Definition: Let *s* and *t* be distinct vertices in a digraph *D.*
An *s-t* **separating arc set** in *D* is a set of arcs whose removal destroys all directed *s-t* paths in *D.*

Thus, an *s-t* separating arc set in *D* is an arc subset of E_D that contains at least one arc of every directed *s-t* path in digraph *D.*

Remark: For the degenerate case in which the original graph or digraph has no *s-t* paths, the empty set is regarded as an *s-t* separating set.

Separating Sets and Cuts

Proposition 12.3.4 Let N be an s-t network such that $cap(e) = 1$ for every arc e. Then the capacity of a minimum s-t cut in network N equals the minimum number of arcs in an s-t separating arc set in N.

Proof: Let $K^* = \langle V_s , V_t \rangle$ be a minimum s-t cut in network N, and let q be the minimum number of arcs in an s-t separating arc set in N.
Since K^* is an s-t cut, it is also an s-t separating arc set. Thus $cap(K^*) \geq q$.

To obtain the reverse inequality, let S be an s-t separating arc set in network N containing q arcs, and let R be the set of all vertices in N that are reachable from source s by a directed path that contains no arc from set S.

Then, by the definitions of arc set S and vertex set R, $t \notin R$, which means that $\langle R, V_N - R \rangle$ is an s-t cut.

Moreover, $\langle R, V_N - R \rangle \subseteq S$. Therefore

Separating Sets and Cuts

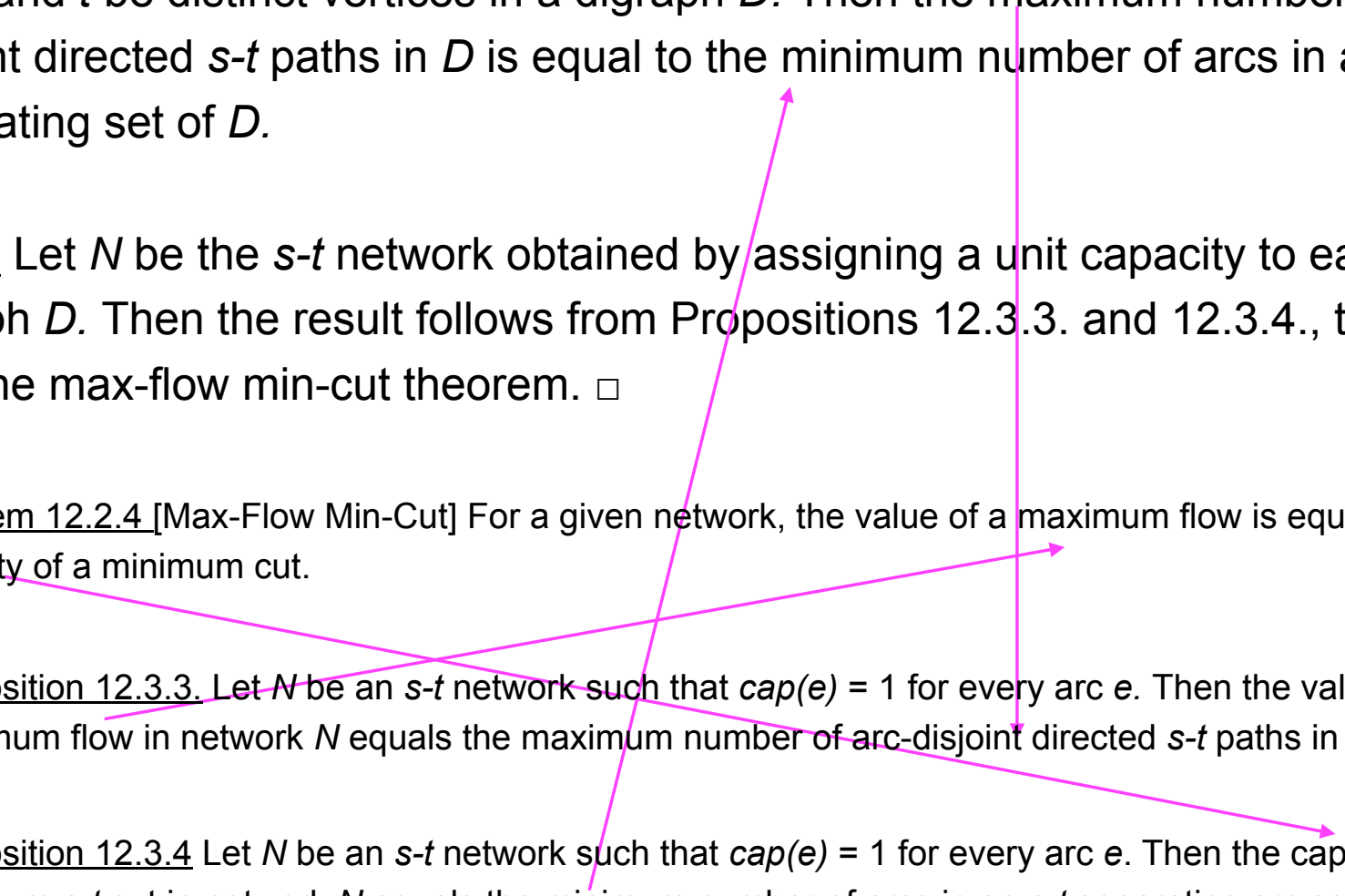
$$
\begin{aligned}
cap(K^*) &\leq cap\langle R, V_N - R\rangle && \text{since } K^* \text{ is a minimum } s-t \text{ cut} \\
&= \left|\langle R, V_N - R\rangle\right| && \text{since all capacities are } 1 \\
&\leq |S| && \text{since } \langle R, V_N - R\rangle \subseteq S \\
&= q
\end{aligned}
$$

which completes the proof. □

Arc and Edge Versions of Menger's Theorem Revisited

Theorem 12.3.5 [Arc form of Menger's theorem]
Let *s* and *t* be distinct vertices in a digraph *D.* Then the maximum number of arc-disjoint directed *s-t* paths in *D* is equal to the minimum number of arcs in an *s-t* separating set of *D.*

Proof: Let *N* be the *s-t* network obtained by assigning a unit capacity to each arc of digraph *D.* Then the result follows from Propositions 12.3.3. and 12.3.4., together with the max-flow min-cut theorem. □

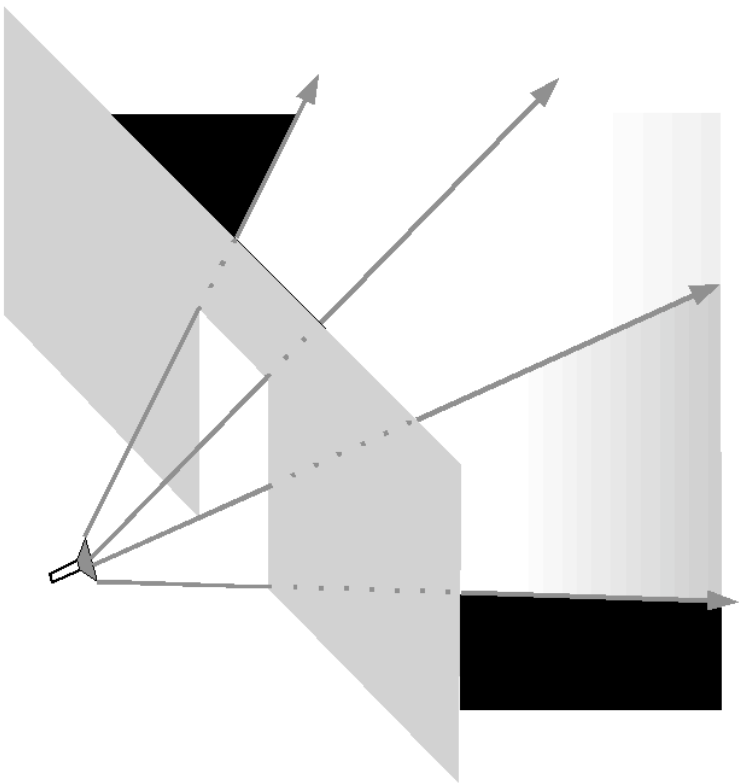


Theorem 12.2.4 [Max-Flow Min-Cut] For a given network, the value of a maximum flow is equal to the capacity of a minimum cut.

Proposition 12.3.3. Let *N* be an *s-t* network such that *cap(e)* = 1 for every arc *e.* Then the value of a maximum flow in network *N* equals the maximum number of arc-disjoint directed *s-t* paths in *N.*

Proposition 12.3.4 Let *N* be an *s-t* network such that *cap(e)* = 1 for every arc *e*. Then the capacity of a minimum *s-t* cut in network *N* equals the minimum number of arcs in an *s-t* separating arc set in *N.*

Idea – extreme pathways



A torch is directed at an open door and shines into a dark room ...

What area is lighted ?

Instead of marking all lighted points individually,
it would be sufficient to characterize the „extreme rays“ that go through the corners of the door.

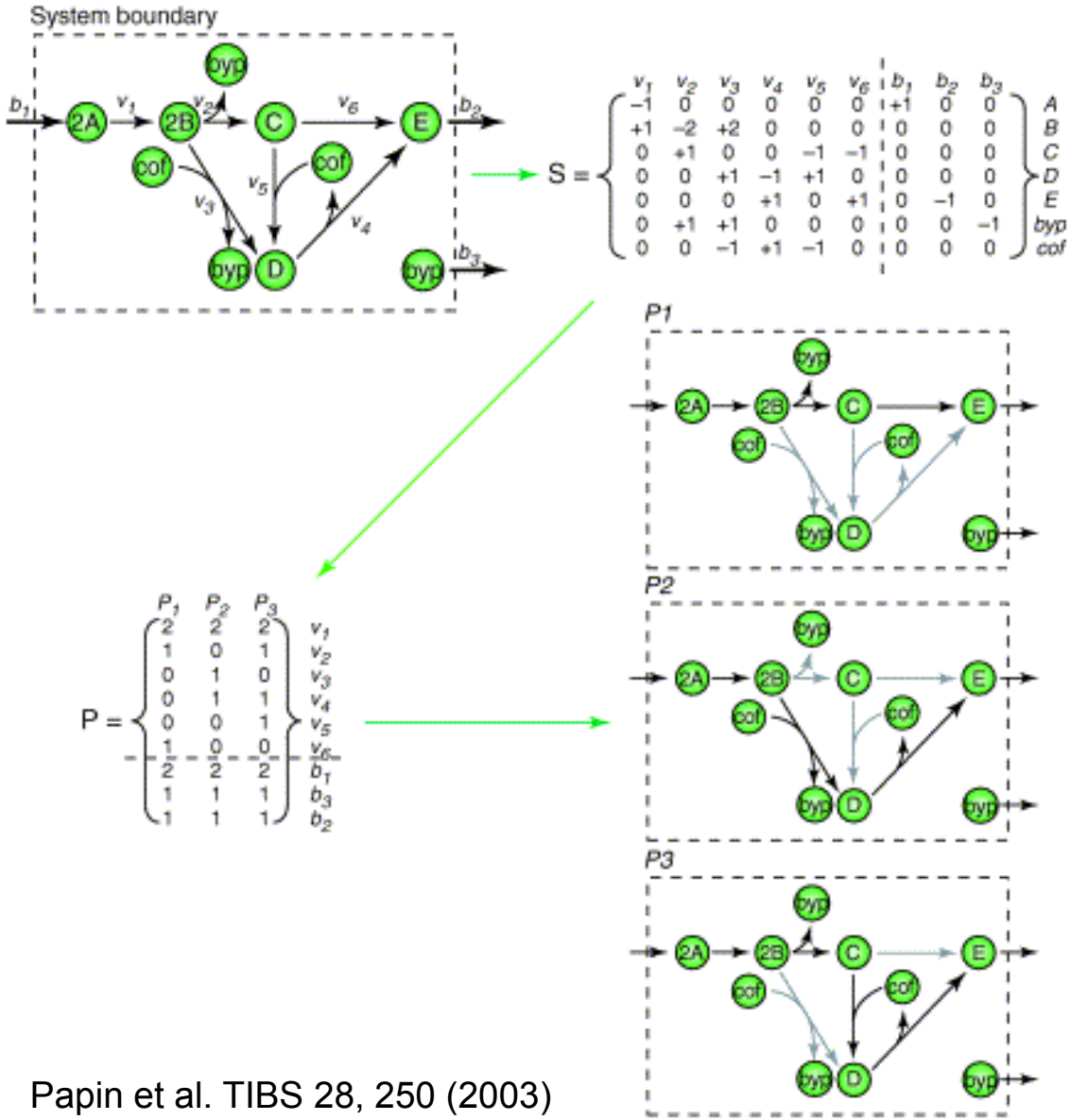
The lighted area is the area between the extreme rays = linear combinations of the extreme rays.

Stoichiometric matrix - Flux Balance Analysis

Stoichiometric matrix S:
m × *n* matrix with stochiometries of the *n* reactions as columns and participations of *m* metabolites as rows.

The stochiometric matrix is an important part of the *in silico* model.

With the matrix, the methods of extreme pathway and elementary mode analyses can be used to generate a unique set of pathways P1, P2, and P3 that allow to express all steady-state fluxes as linear combinations of P1 – P3.

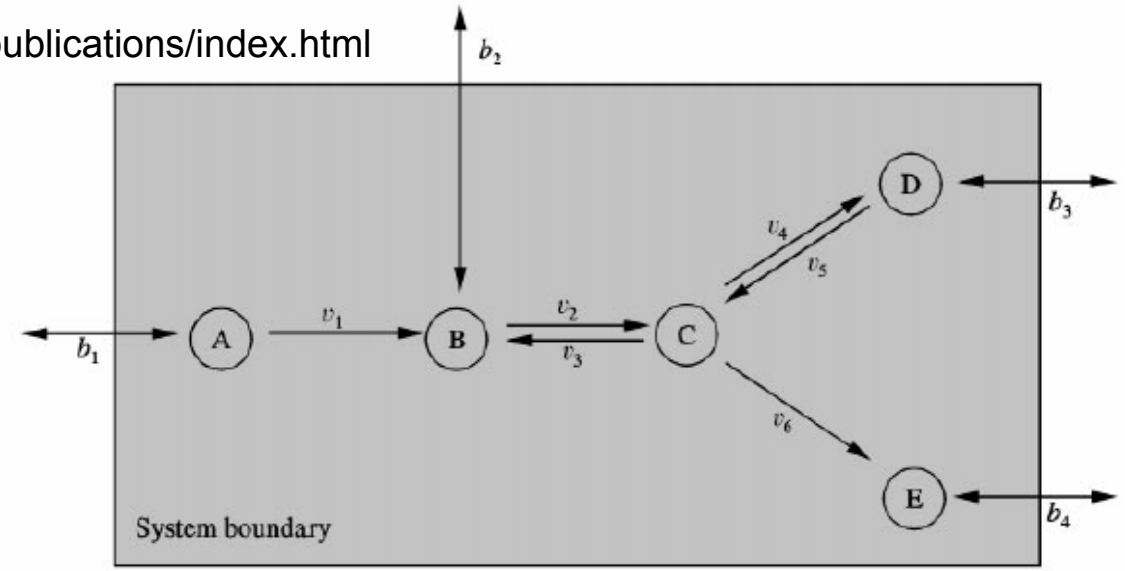


Papin et al. TIBS 28, 250 (2003)

Extreme Pathways

introduced into metabolic analysis by the lab of Bernard Palsson (Dept. of Bioengineering, UC San Diego). The publications of this lab are available at http://gcrg.ucsd.edu/publications/index.html

The extreme pathway technique is based on the stoichiometric matrix representation of metabolic networks.



All external fluxes are defined as pointing outwards.

Mass balance constraints

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$(\mathbf{S} \cdot \mathbf{v} = \mathbf{0})$$

Internal flux constraints

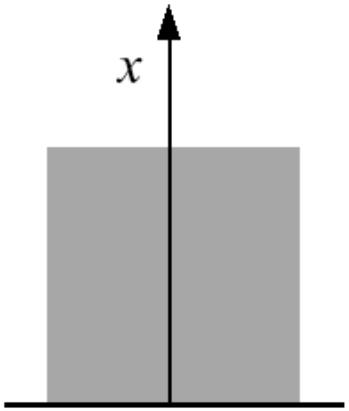
$$v_j \geq 0, \quad j = 1,\ldots,6$$

Exchange flux constraints

$$-\infty \leq b_j \leq +\infty, \quad j = 1,\ldots,4$$

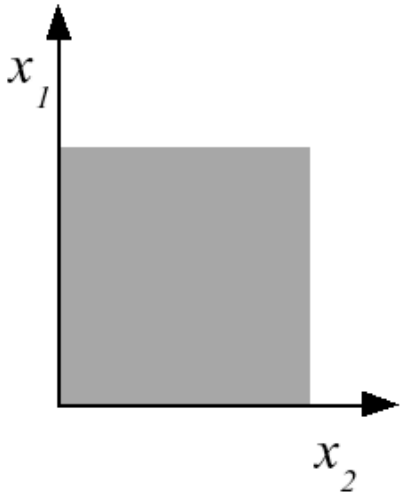
Schilling, Letscher, Palsson, J. theor. Biol. 203, 229 (2000)

Idea – extreme pathways



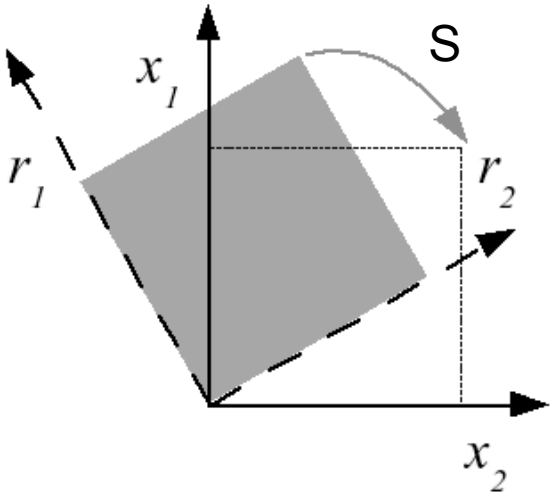
Shaded area:

$x \geq 0$



Shaded area:

$x_1 \geq 0 \wedge x_2 \geq 0$



Either **S** · **x** ≥ **0**

(**S** acts as rotation matrix)

or find optimal vectors

✹ change coordinate system

from x_1, x_2 to r_1, r_2.

Shaded area:

$r_1 \geq 0 \wedge r_2 \geq 0$

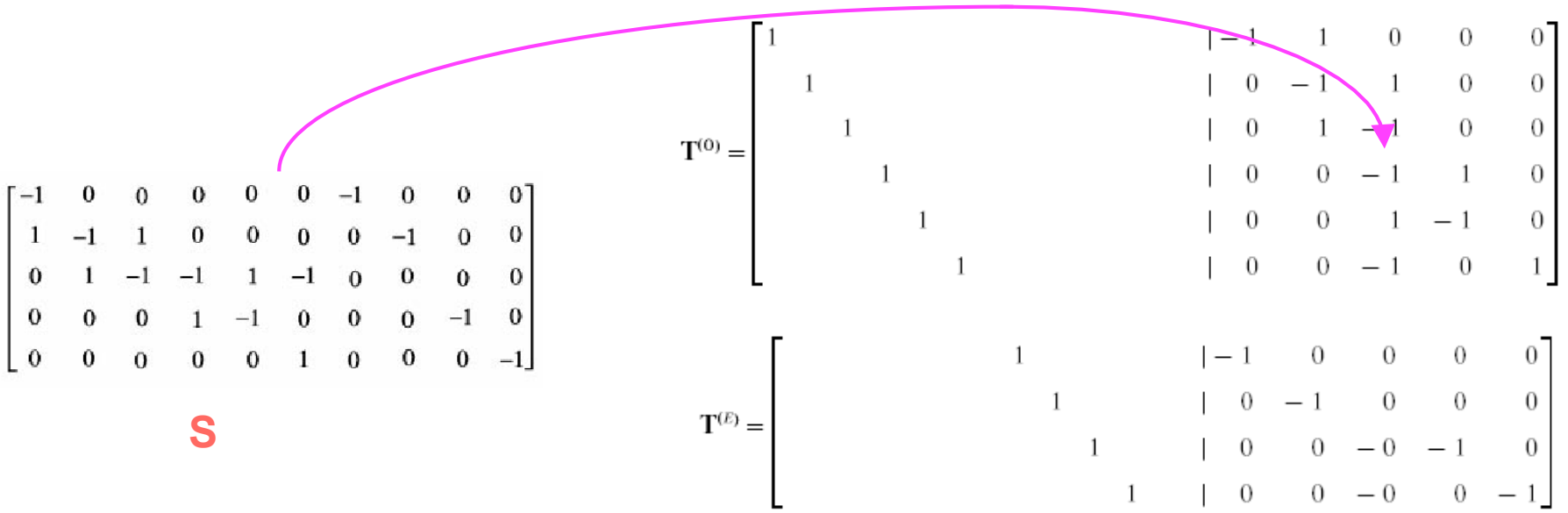
Duality of two matrices S and R.

Edwards & Palsson PNAS 97, 5528 (2000)

Extreme Pathways – algorithm - setup

The algorithm to determine the set of extreme pathways for a reaction network follows the pinciples of algorithms for finding the extremal rays/ generating vectors of convex polyhedral cones.

Combine $n \times n$ identity matrix (**I**) with the transpose of the stoichiometric matrix $\mathbf{S}^T$. **I** serves for bookkeeping.



Schilling, Letscher, Palsson,
J. theor. Biol. 203, 229 (2000)

separate internal and external fluxes

Examine constraints on each of the exchange fluxes as given by

$$\alpha_j \leq b_j \leq \beta_j$$

If the exchange flux is constrained to be positive → do nothing.

If the exchange flux is constrained to be negative → multiply the corresponding row of the initial matrix by -1.

If the exchange flux is unconstrained → move the entire row to a temporary matrix $\mathbf{T}^{(E)}$. This completes the first tableau $\mathbf{T}^{(0)}$.

$\mathbf{T}^{(0)}$ and $\mathbf{T}^{(E)}$ for the example reaction system are shown on the previous slide.

Each element of these matrices will be designated T_{ij}.

Starting with $i = 1$ and $T^{(0)} = T^{(i-1)}$ the next tableau is generated in the following way:

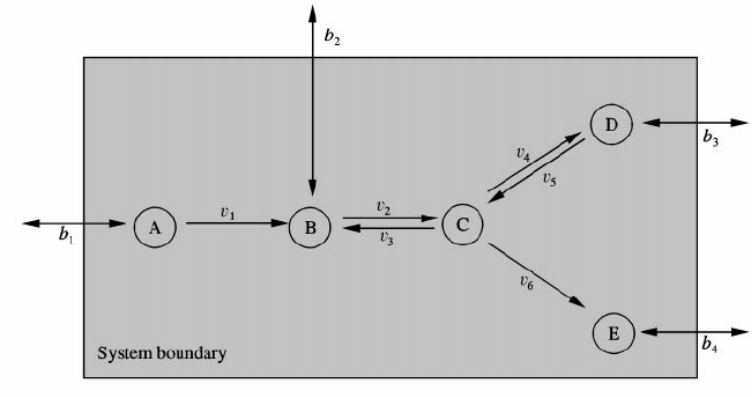
Schilling, Letscher, Palsson,
J. theor. Biol. 203, 229 (2000)

idea of algorithm

(1) Identify all metabolites that do not have an unconstrained exchange flux associated with them.

The total number of such metabolites is denoted by μ.

The example system contains only one such metabolite, namely C (μ = 1).



What is the main idea?

- We want to find balanced extreme pathways that don‘t change the concentrations of metabolites when flux flows through (input fluxes are channelled to products not to accumulation of intermediates).
- The stochiometrix matrix describes the coupling of each reaction to the concentration of metabolites X.
- Now we need to balance combinations of reactions that leave concentrations unchanged. Pathways applied to metabolites should not change their concentrations → the matrix entries need to be brought to 0.

Schilling, Letscher, Palsson, J. theor. Biol. 203, 229 (2000)

keep pathways that do not change concentrations of internal metabolites

(2) Begin forming the new matrix $\mathbf{T}^{(i)}$ by copying all rows from $\mathbf{T}^{(i-1)}$ which already contain a zero in the column of $\mathbf{S}^T$ that corresponds to the first metabolite identified in step 1, denoted by index *C*. (Here 3rd column of $\mathbf{S}^T$.)

$T^{(0)}$ =

						A	B	**C**	D	E
1						-1	1	0	0	0
	1					0	-1	1	0	0
		1				0	1	-1	0	0
			1			0	0	-1	1	0
				1		0	0	1	-1	0
					1	0	0	-1	0	1

$T^{(1)}$ =

1						-1	1	0	0	0

+

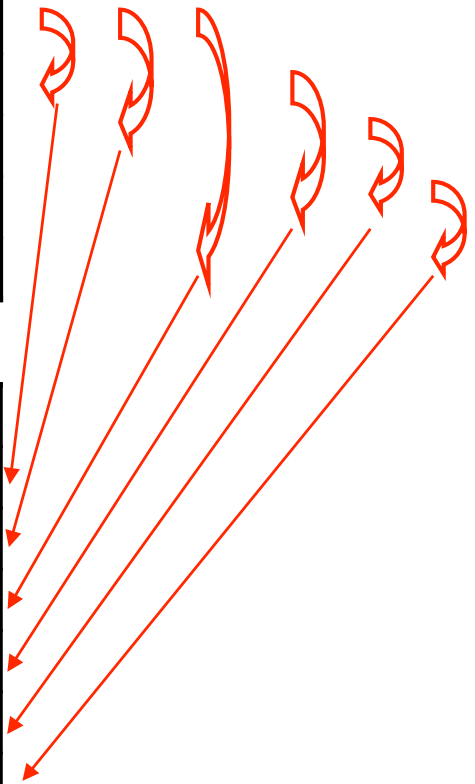
Schilling, Letscher, Palsson, J. theor. Biol. 203, 229 (2000)

balance combinations of other pathways

(3) Of the remaining rows in $\mathbf{T}^{(i-1)}$ add together all possible combinations of rows which contain values of the opposite sign in column C, such that the addition produces a zero in this column.

$T^{(0)}$ =

1						-1	1	**0**	0	0
	1					0	-1	1	0	0
		1				0	1	-1	0	0
			1			0	0	-1	1	0
				1		0	0	1	-1	0
					1	0	0	-1	0	1



$T^{(1)}$ =

1	0	0	0	0	0	-1	1	0	0	0
0	1	1	0	0	0	0	0	**0**	0	0
0	1	0	1	0	0	0	-1	**0**	1	0
0	1	0	0	0	1	0	-1	**0**	0	1
0	0	1	0	1	0	0	1	**0**	-1	0
0	0	0	1	1	0	0	0	**0**	0	0
0	0	0	0	1	1	0	0	**0**	-1	1
1	2	3	4	5	6	7	8	9	10	11

Schilling, et al.
JTB 203, 229

remove “non-orthogonal” pathways

(4) For all rows added to $\mathbf{T}^{(i)}$ in steps 2 and 3 check that no row exists that is a non-negative combination of any other rows in $\mathbf{T}^{(i)}$.

One method for this works as follows:
let *A(i)* = set of column indices *j* for which the elements of row *i* = 0.

For the example above
A(1) = {2,3,4,5,6,9,10,11}
A(2) = {1,4,5,6,7,8,9,10,11}
A(3) = {1,3,5,6,7,9,11}
A(4) = {1,3,4,5,7,9,10}
A(5) = {1,2,4,6,7,9,11}
A(6) = {1,2,3,6,7,8,9,10,11}
A(7) = {1,2,3,4,7,8,9}

Then check to determine if there exists another row *(h)* for which *A(i)* is a subset of A*(h)*.

If $A(i) \subseteq A(h),\ i \neq h$
where
$A(i) = \{ j : T_{i,j} = 0,\ 1 \leq j \leq (n+m) \}$
then row *i* must be eliminated from $\mathbf{T}^{(i)}$

Schilling et al.
JTB 203, 229

repeat steps for all internal metabolites

(5) With the formation of $\mathbf{T}^{(i)}$ complete steps 2 – 4 for all of the metabolites that do not have an unconstrained exchange flux operating on the metabolite, incrementing *i* by one up to μ. The final tableau will be $T^{(\mu)}$.

Note that the number of rows in $T^{(\mu)}$ will be equal to *k*, the number of extreme pathways.

Schilling et al.
JTB 203, 229

balance external fluxes

(6) Next we append $\mathbf{T}^{(E)}$ to the bottom of $\mathbf{T}^{(\mu)}$. (In the example here $\mu = 1$.)
This results in the following tableau:

$T^{(1/E)} =$

1										-1	1	0	0	0
	1	1								0	0	0	0	0
	1		1							0	-1	0	1	0
	1				1					0	-1	0	1	0
		1		1						0	1	0	-1	0
			1	1						0	0	0	0	0
				1	1					0	0	0	-1	1
						1				-1	0	0	0	0
							1			0	-1	0	0	0
								1		0	0	0	-1	0
									1	0	0	0	0	-1

Schilling et al.
JTB 203, 229

balance external fluxes

(7) Starting in the *n+1* column (or the first non-zero column on the right side), if $T_{i,(n+1)} \neq 0$ then add the corresponding non-zero row from $\mathbf{T}^{(E)}$ to row *i* so as to produce 0 in the *n+1*-th column.
This is done by simply multiplying the corresponding row in $\mathbf{T}^{(E)}$ by $T_{i,(n+1)}$ and adding this row to row *i* .

Repeat this procedure for each of the rows in the upper portion of the tableau so as to create zeros in the entire upper portion of the *(n+1)* column.

When finished, remove the row in $\mathbf{T}^{(E)}$ corresponding to the exchange flux for the metabolite just balanced.

Schilling et al.
JTB 203, 229

balance external fluxes

(8) Follow the same procedure as in step (7) for each of the columns on the right side of the tableau containing non-zero entries.

(In our example we need to perform step (7) for every column except the middle column of the right side which correponds to metabolite C.)

The final tableau $\mathbf{T}^{(final)}$ will contain the transpose of the matrix **P** containing the extreme pathways in place of the original identity matrix.

Schilling et al.
JTB 203, 229

pathway matrix

$\mathbf{T}^{(final)}$ =

1						-1	1			0	0	0	0	0	0
	1	1								0	0	0	0	0	0
	1		1				-1	1		0	0	0	0	0	0
	1				1		-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			-1	1	0	0	0	0	0	0

$\mathbf{P}^{T}$ =

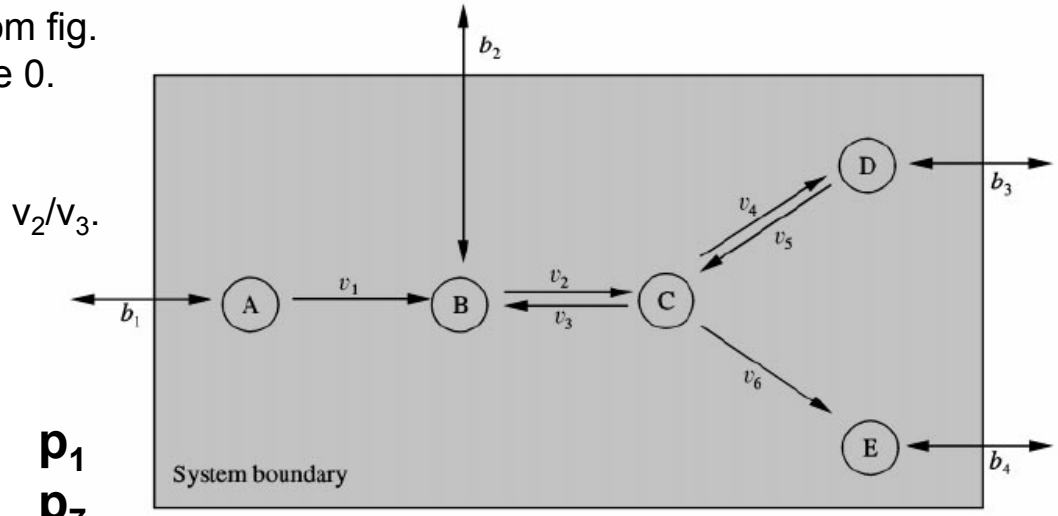
v_1	v_2	v_3	v_4	v_5	v_6	b_1	b_2	b_3	b_4	
1	0	0	0	0	0	-1	1	0	0	$\mathbf{p_1}$
0	1	1	0	0	0	0	0	0	0	$\mathbf{p_7}$
0	1	0	1	0	0	0	-1	1	0	$\mathbf{p_3}$
0	1	0	0	0	1	0	-1	0	1	$\mathbf{p_2}$
0	0	1	0	1	0	0	1	-1	0	$\mathbf{p_4}$
0	0	0	1	1	0	0	0	0	0	$\mathbf{p_6}$
0	0	0	0	1	1	0	0	-1	1	$\mathbf{p_5}$

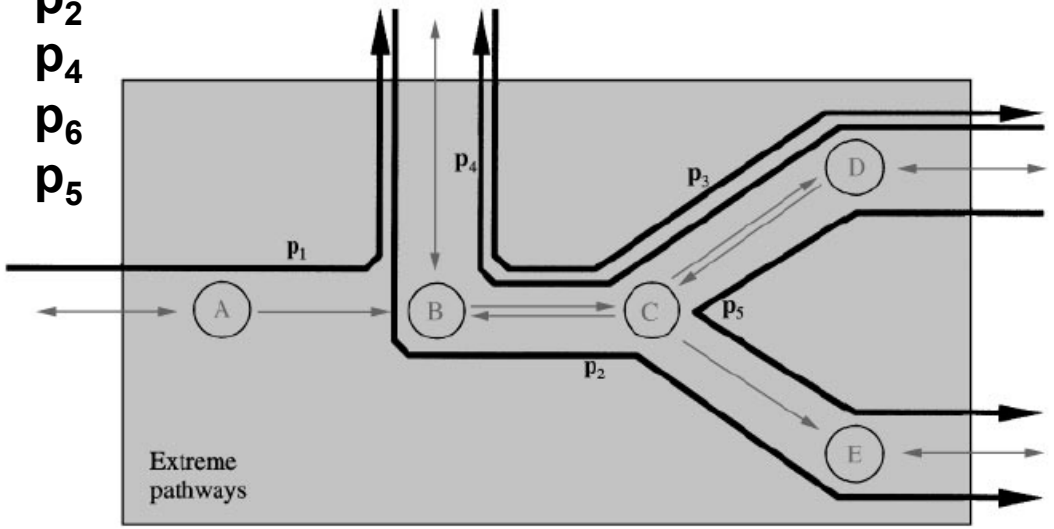
Schilling et al.
JTB 203, 229

Extreme Pathways for model system

2 pathways p_6 and p_7 are not shown in the bottom fig. because all exchange fluxes with the exterior are 0. Such pathways have no net overall effect on the functional capabilities of the network.
They belong to the cycling of reactions v_4/v_5 and v_2/v_3.

	v_1	v_2	v_3	v_4	v_5	v_6	b_1	b_2	b_3	b_4
p_1	1	0	0	0	0	0	-1	1	0	0
p_7	0	1	1	0	0	0	0	0	0	0
p_3	0	1	0	1	0	0	0	-1	1	0
p_2	0	1	0	0	0	1	0	-1	0	1
p_4	0	0	1	0	1	0	0	1	-1	0
p_6	0	0	0	1	1	0	0	0	0	0
p_5	0	0	0	0	1	1	0	0	-1	1



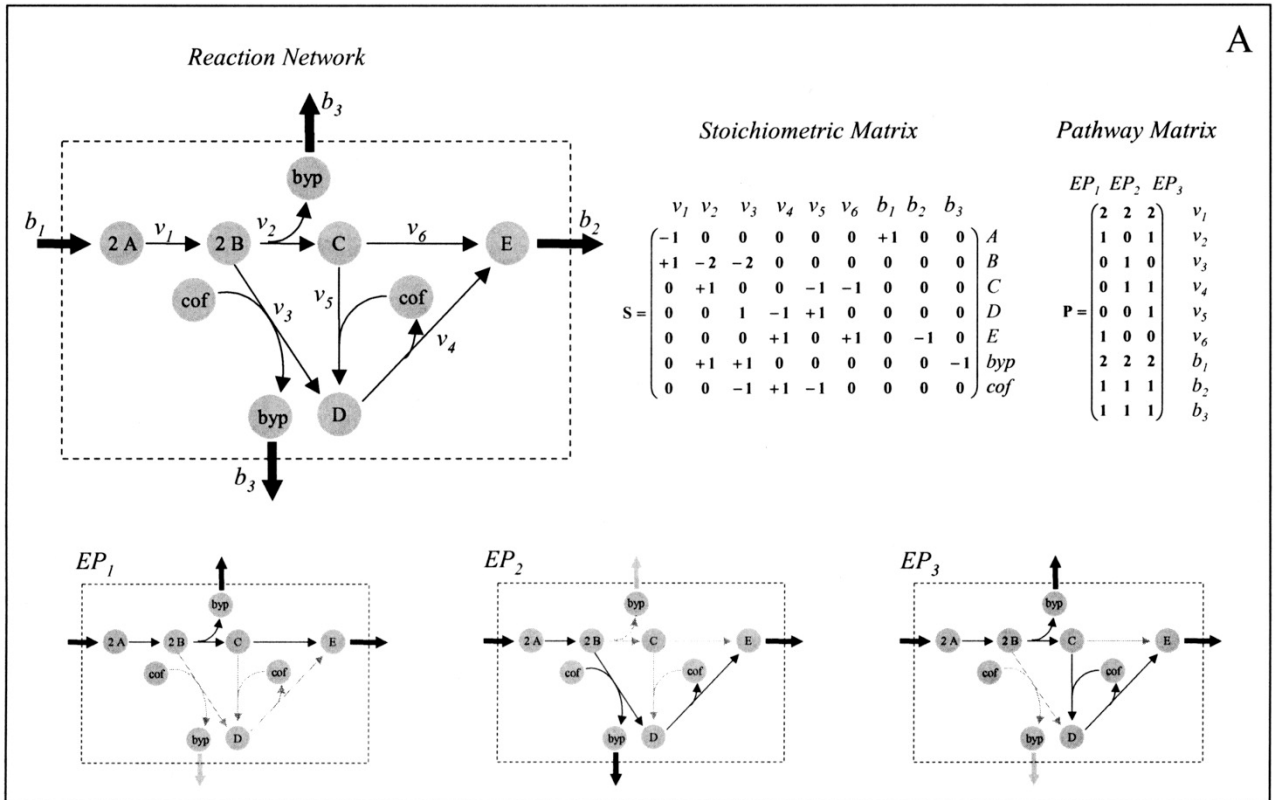


Schilling et al.
JTB 203, 229

How reactions appear in pathway matrix

In the matrix **P** of extreme pathways, each column is an EP and each row corresponds to a reaction in the network.

The numerical value of the *i,j*-th element corresponds to the relative flux level through the *i*-th reaction in the *j*-th EP.



Papin, Price, Palsson,
Genome Res. 12, 1889 (2002)

Properties of pathway matrix

After normalizing **P** to a matrix with entries 0 or 1,
the symmetric Pathway Length Matrix $\mathbf{P}_{LM}$ can be calculated:

$$\mathbf{P}_{LM} = \mathbf{P}^T \cdot \mathbf{P}$$

where the values along the diagonal correspond to the length of the EPs.

Pathway Length

$$\mathbf{P} = \begin{pmatrix} 2 & 2 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \rightarrow \tilde{\mathbf{P}} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \rightarrow \tilde{\mathbf{P}}^T \bullet \tilde{\mathbf{P}} = \begin{array}{c} \begin{matrix} EP_1 & EP_2 & EP_3 \end{matrix} \\ \begin{pmatrix} 6 & 4 & \textcircled{5} \\ & 6 & 5 \\ & & 7 \end{pmatrix} \end{array} \begin{matrix} EP_1 \\ EP_2 \\ EP_3 \end{matrix}$$

Comments:

1) The lengths of EP_1, EP_2,and EP_3 are 6, 6, and 7, respectively, the highlighted diagonal elements of the final matrix.

2) EP_2 and EP_3 have a shared length of 5 (indicated by the circle). As seen in the schematics above, they share reactions v_1, v_4, b_1, b_2, and b_3.

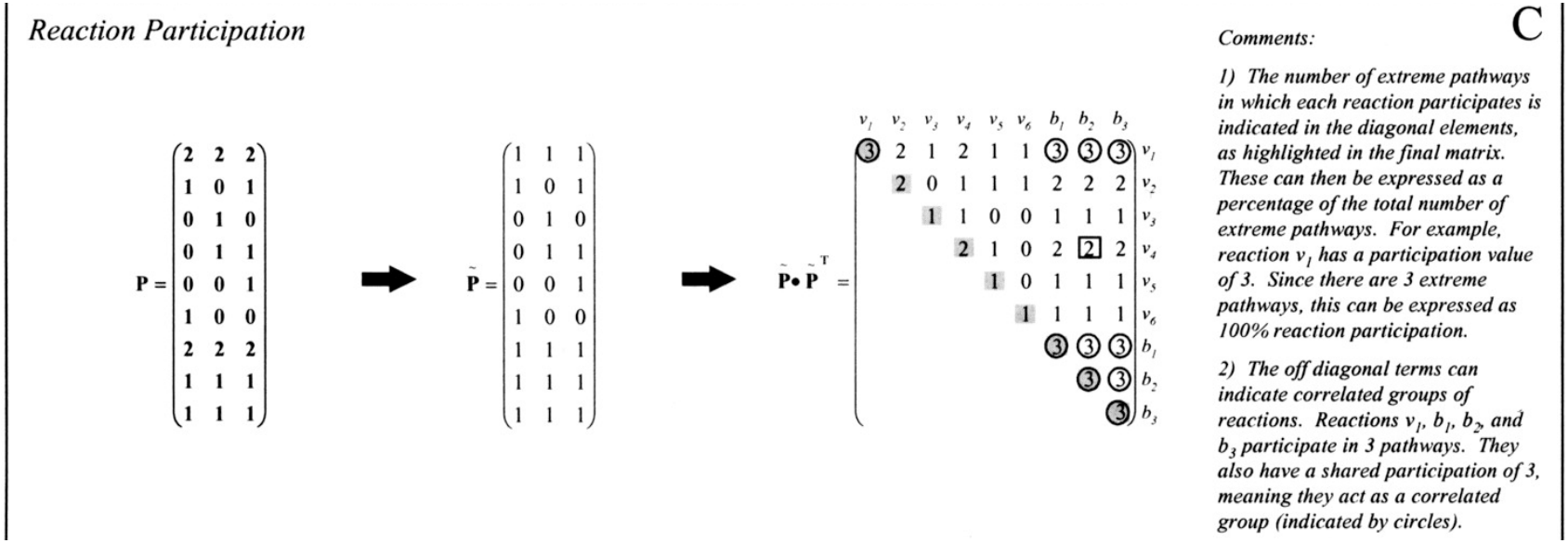
The off-diagonal terms of $\mathbf{P}_{LM}$ are the number of reactions that a pair of extreme pathways have in common.

Properties of pathway matrix

One can also compute a reaction participation matrix $\mathbf{P}_{PM}$ from **P**:

$$\mathbf{P}_{PM} = \mathbf{P} \cdot \mathbf{P}^{T}$$

where the diagonal correspond to the number of pathways in which the given reaction participates.



EP Analysis of *H. pylori* and *H. influenza*

Amino acid synthesis in *Heliobacter pylori* vs. *Heliobacter influenza* studied by EP analysis.

Table 4. Summary of the Statistical Analyses of Extreme Pathway Lengths

H. pylori Target product	Number of EPs	Pathway length average	maximum	minimum	coefficient of variation
Asparagine	340	44	54	28	15%
Aspartic Acid	491	43	52	24	14%
Cysteine	1022	59	71	45	10%
Glutamine	315	41	53	23	18%
Glutamic Acid	493	41	53	25	17%
Glycine	377	51	60	38	10%
Lysine	611	54	66	39	12%
Proline	867	43	56	15	16%
Serine	355	45	54	33	12%
Threonine	469	48	60	31	14%
Tryptophan	1958	64	73	51	6%
Tyrosine	1008	58	68	44	7%
Equimolar Amino Acids	6032	106	112	99	2%
E. coli Ratio Amino Acids	5553	106	112	99	2%
H. influenzae Target product	Number of EPs	Pathway length average	maximum	minimum	coefficient of variation
Alanine	1739	36	49	18	10%
Asparagine	445	39	52	29	13%
Aspartic Acid	690	35	49	27	14%
Glutamine	690	37	46	28	11%
Glycine	456	39	48	35	7%
Histidine	1507	65	74	61	3%
Isoleucine	1480	47	61	37	9%
Leucine	3884	42	55	31	10%
Lysine	1168	47	61	37	9%
Methionine	1343	48	63	40	8%
Phenylalanine	1758	51	64	43	7%
Proline	2624	38	51	25	11%
Serine	690	37	50	30	10%
Threonine	1318	42	55	32	10%
Tryptophan	3540	58	69	49	6%
Tyrosine	1758	51	64	43	7%
Valine	1739	39	52	23	9%

The coefficient of variation is the standard deviation normalized to the average (expressed as a percent). Equimolar amino acids refers to the set of amino acids in equimolar ratios. *E. coli* ratio amino acids refers to the set of amino acids in ratios analogous to those seen in *E. coli* biomass. EPs, extreme pathways.

Table 1. Number of Reactions Involved in the Production of the Indicated Target Product

H. pylori Target product	Essential reactions	Utilized reactions
Tryptophan	32	105
Tyrosine	28	101
Cysteine	25	102
Glycine	22	97
Lysine	22	102
Serine	16	91
Threonine	14	96
Asparagine	13	91
Aspartic Acid	12	91
Proline	10	91
Glutamic Acid	7	91
Glutamine	6	91
Equimolar Amino Acids	85	140
E. coli Ratio Amino Acids	85	140
H. influenzae Target product	Essential reactions	Utilized reactions
Histidine	51	112
Tryptophan	41	108
Phenylalanine	36	108
Tyrosine	36	108
Methionine	34	106
Isoleucine	31	108
Lysine	31	108
Glycine	29	82
Threonine	26	103
Asparagine	25	98
Serine	25	97
Leucine	23	105
Aspartic Acid	22	97
Glutamine	21	102
Proline	18	103
Valine	17	102
Alanine	12	99

See Fig. 3 for the indicated network inputs and outputs. Essential reactions refers to the number of reactions that were used in every extreme pathway (region I in Fig. 4). Utilized reactions refers to the number of reactions that were used at least once in the set of extreme pathways for the production of the associated product (region II in Fig. 4). The individual amino acids are sorted in descending order according to the number of essential reactions. Equimolar amino acids refers to the set of amino acids in equimolar ratios. *E. coli* ratio amino acids refers to the set of amino acids in ratios analogous to those seen in *E. coli* biomass.

Papin, Price, Palsson, Genome Res. 12, 1889 (2002)

Extreme Pathway Analysis

Calculation of EPs for increasingly large networks is computationally intensive and results in the generation of large data sets.

Even for integrated genome-scale models for microbes under simple conditions, EP analysis can generate thousands or even millions of vectors!

It turned out that the number of reactions that participate in EPs that produce a particular product is usually poorly correlated to the product yield and the molecular complexity of the product.

Possible way out?

Matrix diagonalisation – eigenvectors: only possible for quadratic $n \times n$ matrices with rank n.

Papin, Price, Palsson, Genome Res. 12, 1889 (2002)

Quasi-diagonalisation of pathway matrix by SVD

Suppose M is an $m \times n$ matrix with real or complex entries.

Then there exists a factorization of the form

$M = U \Sigma V^*$ where

U : $m \times m$ unitary matrix, ($U^*U = UU^* = I$)

Σ : is an $m \times n$ matrix with nonnegative numbers on the diagonal and zeros off the diagonal,

V^* : the transpose of V, is an $n \times n$ unitary matrix of real or complex numbers.

Such a factorization is called a **singular-value decomposition** of M.

U describes the rows of M with respect to the base vectors associated with the singular values.

V describes the columns of M with respect to the base vectors associated with the singular values. Σ contains the singular values.

One commonly insists that the values $\Sigma_{i,i}$ be ordered in non-increasing fashion.

Then, the diagonal matrix Σ is uniquely determined by M (but not U and V).

Single Value Decomposition of EP matrices

For a given EP matrix $\mathbf{P} \in \Re^{n \times p}$, SVD decomposes **P** into 3 matrices

$$\mathbf{P} = \mathbf{U} \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{n \times p} \mathbf{V}^T$$

where $\mathbf{U} \in \Re^{n \times n}$: orthonormal matrix of the left singular vectors,

$\mathbf{V} \in \Re^{p \times p}$: an analogous orthonormal matrix of the right singular vectors,

$\mathbf{\Sigma} \in \Re^{r \times r}$:a diagonal matrix containing the singular values $\sigma_{i=1..r}$ arranged in descending order where *r* is the rank of **P**.

The first *r* columns of **U** and **V**, referred to as the left and right singular vectors, or modes, are unique and form the orthonormal basis for the column space and row space of **P**.

The singular values are the square roots of the eigenvalues of $\mathbf{P}^T\mathbf{P}$.

The magnitudes of the singular values in Σ indicate the relative contribution of the singular vectors in **U** and **V** in reconstructing **P**.

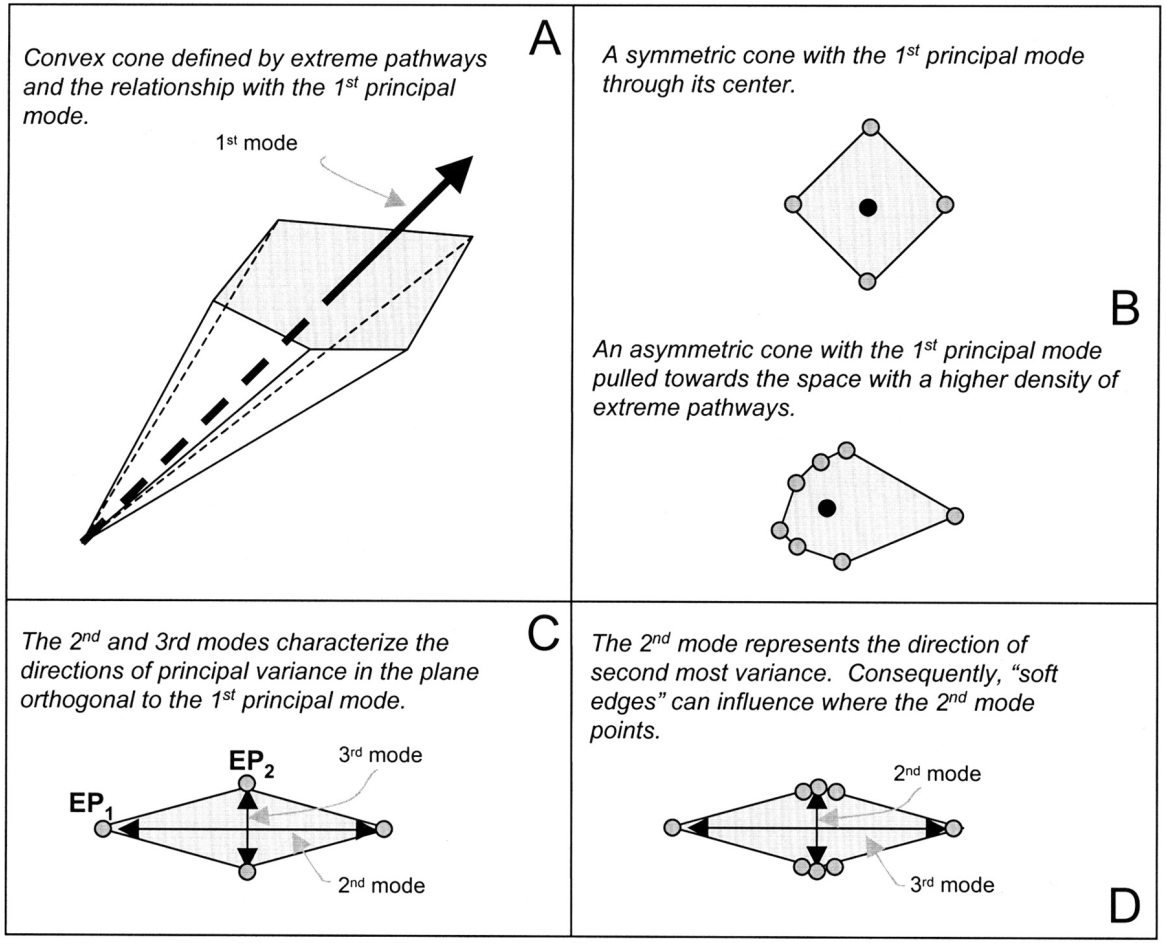
E.g. the second singular value contributes less to the construction of **P** than the first singular value etc.

Price *et al.* Biophys J 84, 794 (2003)

Single Value Decomposition of EP: Interpretation

The first mode (as the other modes) corresponds to a valid biochemical pathway through the network.

The first mode will point into the portions of the cone with highest density of EPs.



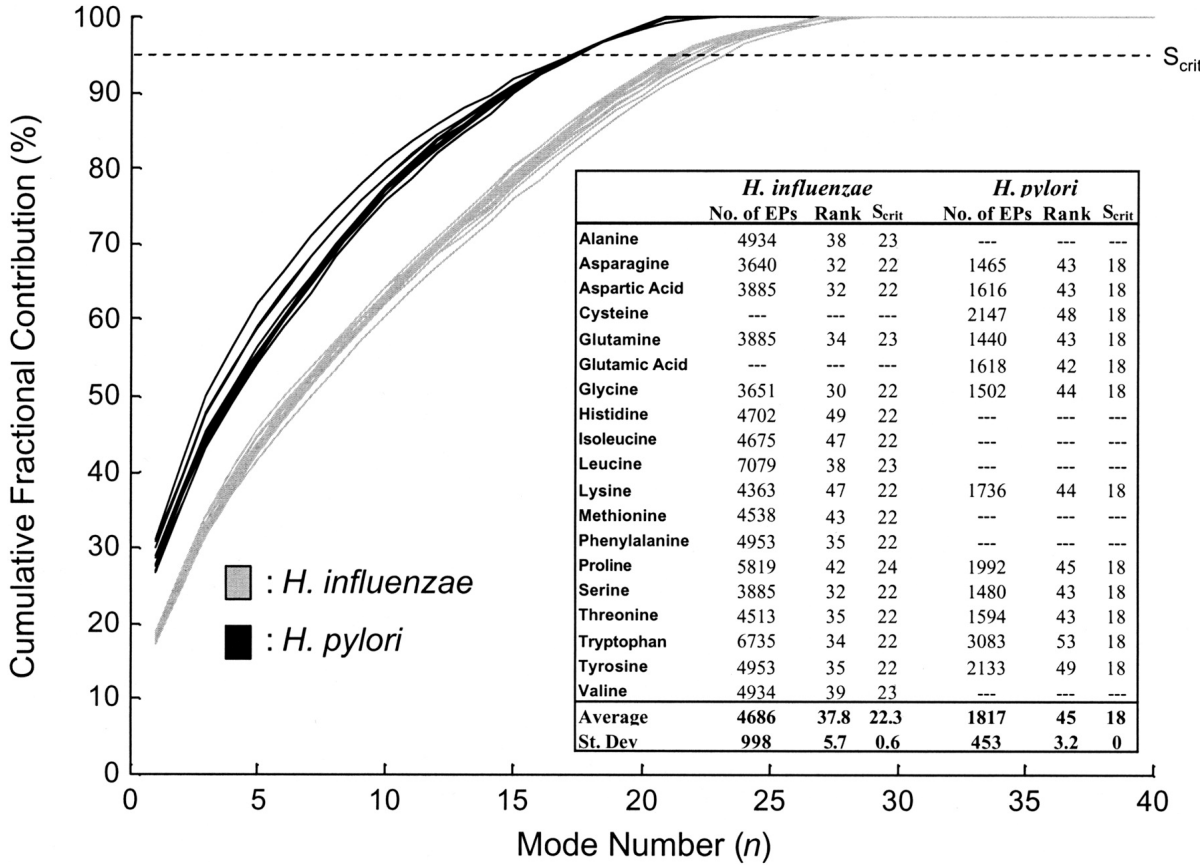
Price *et al.* Biophys J 84, 794 (2003)

SVD applied for *Heliobacter* systems

Cumulative fractional contributions for the SVD of the EP matrices of *H. influenza* and *H. pylori*.

This plot represents the contribution of the first *n* modes to the overall description of the system.

Ca. 20 modes allow describing most of the metabolic activity in the Network.



	H. influenzae			*H. pylori*		
	No. of EPs	Rank	S_{crit}	No. of EPs	Rank	S_{crit}
Alanine	4934	38	23	---	---	---
Asparagine	3640	32	22	1465	43	18
Aspartic Acid	3885	32	22	1616	43	18
Cysteine	---	---	---	2147	48	18
Glutamine	3885	34	23	1440	43	18
Glutamic Acid	---	---	---	1618	42	18
Glycine	3651	30	22	1502	44	18
Histidine	4702	49	22	---	---	---
Isoleucine	4675	47	22	---	---	---
Leucine	7079	38	23	---	---	---
Lysine	4363	47	22	1736	44	18
Methionine	4538	43	22	---	---	---
Phenylalanine	4953	35	22	---	---	---
Proline	5819	42	24	1992	45	18
Serine	3885	32	22	1480	43	18
Threonine	4513	35	22	1594	43	18
Tryptophan	6735	34	22	3083	53	18
Tyrosine	4953	35	22	2133	49	18
Valine	4934	39	23	---	---	---
Average	**4686**	**37.8**	**22.3**	**1817**	**45**	**18**
St. Dev	**998**	**5.7**	**0.6**	**453**	**3.2**	**0**

Cumulative fractional contribution : sum of the first n fractional singular values. This value represents the contribution of the first n modes to the overall description of the system. The rank of the respective extreme pathway matrix is shown for nonessential amino acids. S_{crit}: number of singular values that account for 95% of the variance in the matrices. Entries with "- - -" correspond to essential amino acids.

Price *et al.* Biophys J 84, 794 (2003)

Summary – Extreme Pathways

Extreme Pathway Analysis is a standard technique for analysis of metabolic networks.

Number of EPs can become extremely large – hard to interpret.

EP is an excellent basis for studying systematic effects of reaction cut sets.

SVD could facilitate analysis of EPs. Has not been widely used sofar.

It will be very important to consider the interplay of metabolic and regulatory networks.