Bioinformatics: Network Analysis

Analyzing Stoichiometric Matrices

COMP 572 (BIOS 572 / BIOE 564) - Fall 2013

Luay Nakhleh, Rice University

Biological Components Have a Finite Turnover Time

- Most metabolites turn over within a minute in a cell
- mRNA molecules typically have 2-hour half-lives in human cells
- The renewal rate of skin is on the order of 5 days to a couple of weeks
- Therefore, most of the cells that are contained in an individual today were not there a few years ago
- However, we consider the individual to be the same

Biological Components Have a Finite Turnover Time

- Components come and go
- The interconnections between cells and cellular components define the essence of a living process

Components vs. Systems

- In systems biology, it is not so much the components themselves and their state that matters, but it is the state of the whole system that counts

Links and Functional States of a System

- Links between molecular components are basically given by chemical reactions or associations between chemical components

- These links are therefore characterized and constrained by basic chemical rules

- Multiple links between components form a network, and the network can have functional states

- Functional states of networks are constrained by various factors that are physiochemical, environmental, and biological in nature

Links and Functional States of a System

- The number of possible functional states of networks typically grows much faster than the number of components in the network
- The number of candidate functional states of a biological network far exceeds the number of biologically useful states to an organism
- Cells must select useful functional states by elaborate regulatory mechanisms

Elucidating Metabolic Pathways

- Metabolism is broadly defined as the complex physical and chemical processes involved in the maintenance of life
- It is comprised of a vast repertoire of enzymatic reactions and transport processes used to convert thousands of organic compounds into the various molecules necessary to support cellular life
- Metabolic objectives are achieved through a sophisticated control scheme that efficiently distributes and processes metabolic resources throughout the cell's metabolic network

Elucidating Metabolic Pathways

- The obvious functional unit in metabolic networks is the actual enzyme or gene product executing a particular chemical reaction or facilitating a transport process
- The cell controls its metabolic pathways in a switchboard-like fashion, directing the distribution and processing of metabolites throughout its extensive map of pathways
- To understand the regulatory logic implemented by the cell to control the network it is imperative to elucidate the cell's metabolic pathways

Elucidating Metabolic Pathways

- In this lecture, we will cover theoretical techniques, based on convex analysis, that have been used to identify metabolic pathways and analyze their properties
- The techniques have also been applied to analysis of regulatory networks (signal transduction networks)

Stoichiometry

- The set of chemical reactions that comprise a network can be represented as a set of chemical equations
- Embedded in these chemical equations is information about reaction stoichiometry (the quantitative relationships of the reaction's reactants and products)
- Stoichiometry is invariant between organisms for the same reactions and does not change with pressure, temperature, or other conditions
- All this stoichiometric information can be represented in a matrix form; the stoichiometric matrix, denoted by **S**

The Stoichiometric Matrix

- Mathematically, the stoichiometric matrix S is a linear transformation of the flux* vector

$$\mathbf{v}=(v_1,v_2,...,v_n)$$

to a vector of derivatives of the concentration vector

$$\mathbf{x}=(x_1,x_2,...,x_m)$$

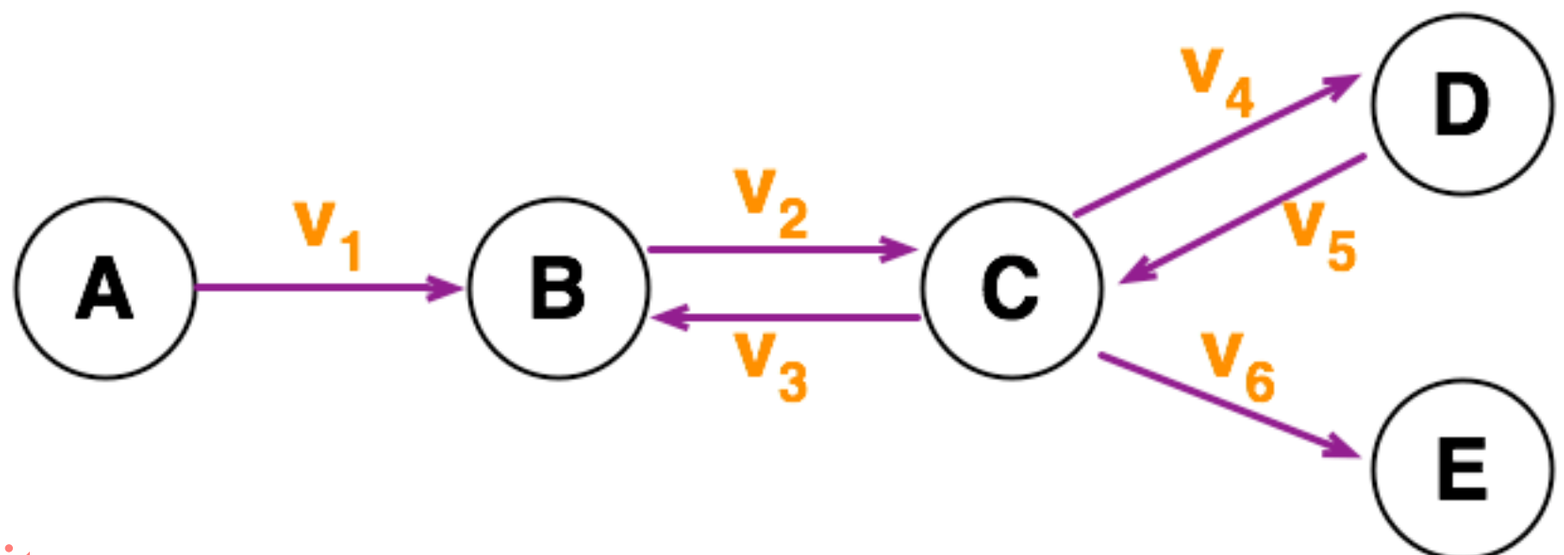
as

$$\frac{d\mathbf{x}}{dt} = \mathbf{S} \cdot \mathbf{v}$$

The dynamic mass balance equation

*Flux: the production or consumption of mass per unit area per unit time

The Stoichiometric Matrix



Five metabolites

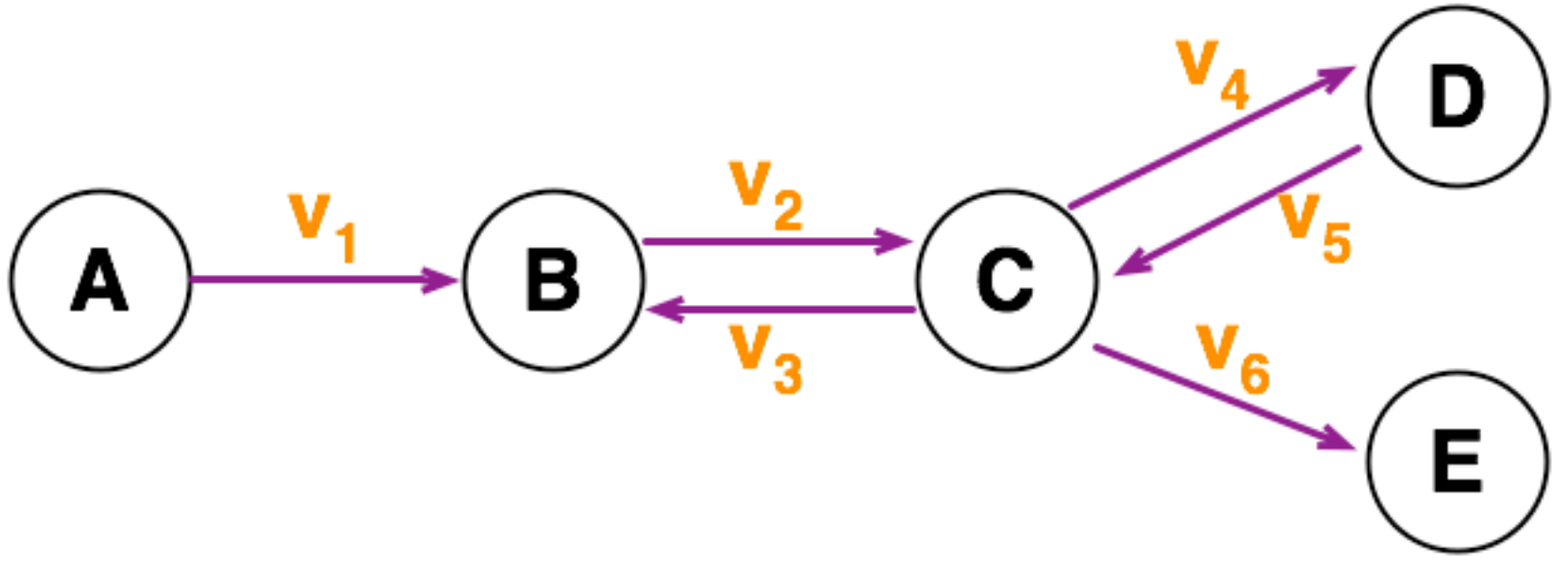
A,B,C,D,E

Four internal reactions, two of which are reversible, creating six internal fluxes

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} dA/dt \\ dB/dt \\ dC/dt \\ dD/dt \\ dE/dt \end{bmatrix}$$

S **v** **dx/dt**

The Stoichiometric Matrix



$$\frac{dx_i}{dt} = \sum_k s_{ik} v_k$$

$$\frac{dC}{dt} = 0v_1 + 1v_2 - 1v_3 - 1v_4 + 1v_5 - 1v_6$$

Fluxes that form C

Fluxes that degrade C

The Fundamental Subspaces of a Matrix

- Each matrix A defines four fundamental subspaces
- The column space: the set of all possible linear combinations of the columns of A
- The row space: the set of all possible linear combinations of the rows of A
- The null space: the set of all vectors v for which Av=0
- The left null space: the null space of A^T

The Column and Left Null Spaces of the Stoichiometric Matrix

- Writing the dynamic mass balance equation as

$$\frac{dx}{dt} = s_1 v_1 + s_2 v_2 + \cdots + s_n v_n$$

where s_i are the reaction vectors that form the columns of **S**, it is clear that dx/dt is in the column space of **S**

- The reaction vectors are structural features of the network and are fixed
- The fluxes v_i are scalar quantities and represent the flux through reaction i
- The fluxes are variables
- The vectors in the left null space are orthogonal to the column space; these vectors represent a mass conservation

The Row and Null Spaces of the Stoichiometric Matrix

- The flux vector can be decomposed into a dynamic component and a steady state component:

$$\mathbf{v} = \mathbf{v}_{dyn} + \mathbf{v}_{ss}$$

- The steady state component satisfies

$$\mathbf{S}\mathbf{v}_{ss} = 0$$

and $\mathbf{v}_{ss}$ is thus in the null space of **S**

- The dynamic component of the flux vector $\mathbf{v}_{dyn}$ is orthogonal to the null space and consequently it is in the row space of **S**

The Null Space of **S**

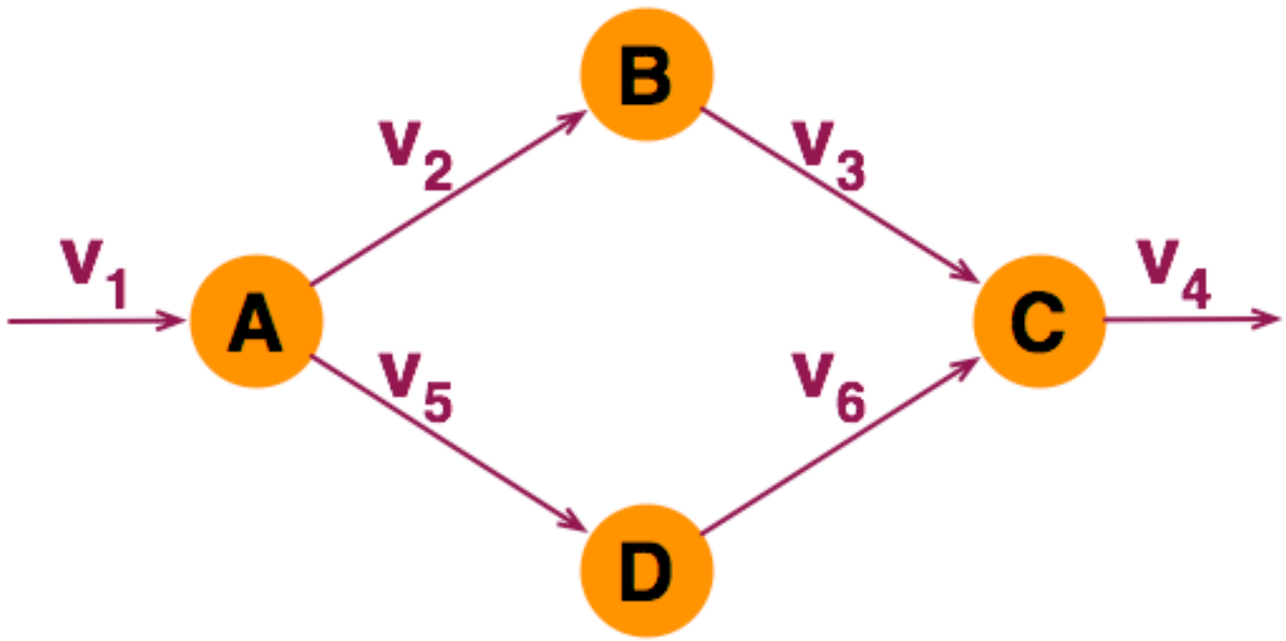
- The (right) null space of **S** is defined by $\mathbf{S}\mathbf{v}_{ss} = 0$

- Thus, all the steady-state flux distributions, $\mathbf{v}_{ss}$, are found in the null space

- The null space is spanned by a set of basis vectors that form the columns of matrix **R** that satisfies **SR**=0

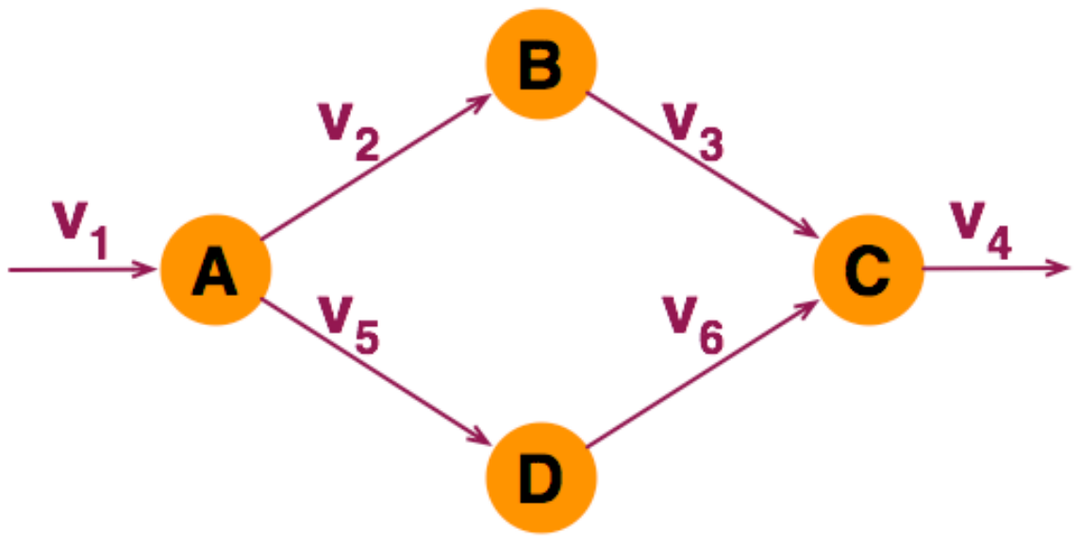
- A set of linear basis vectors is not unique, but once the set is chosen, the weights (w_i) for a particular $\mathbf{v}_{ss}$ are unique

Example



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

Example



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

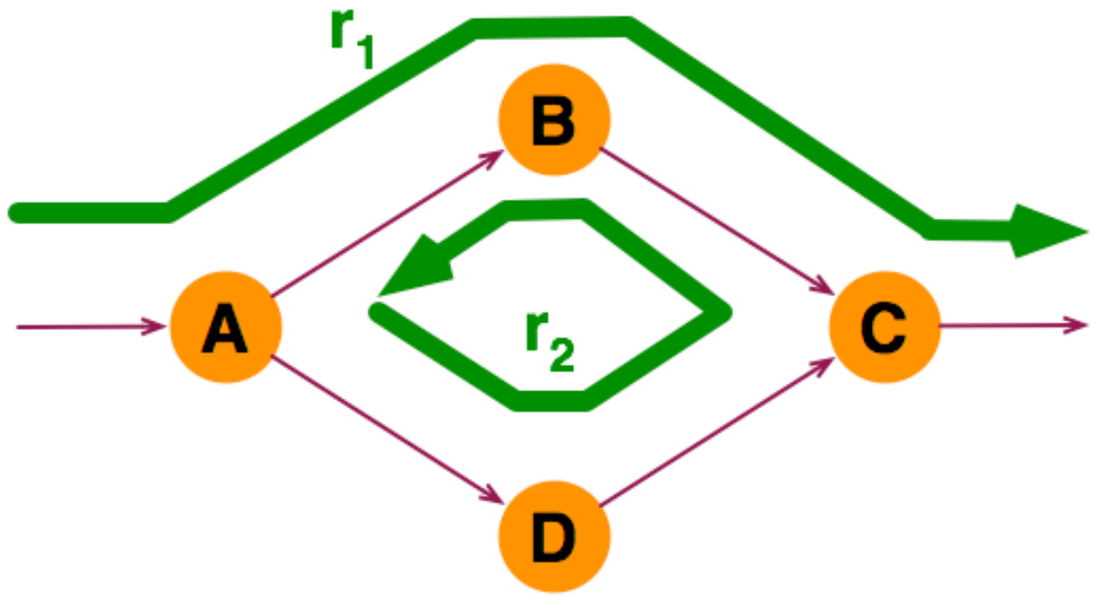
The set of linear equations can be solved using v_4 and v_6 as free variables to give

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} v_4 \\ v_4-v_6 \\ v_4-v_6 \\ v_4 \\ v_6 \\ v_6 \end{bmatrix} = v_4 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + v_6 \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = w_1 r_1 + w_2 r_2$$

r_1 and r_2 form a basis

Example

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} v_4 \\ v_4-v_6 \\ v_4-v_6 \\ v_4 \\ v_6 \\ v_6 \end{bmatrix} = v_4 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + v_6 \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = w_1 r_1 + w_2 r_2$$

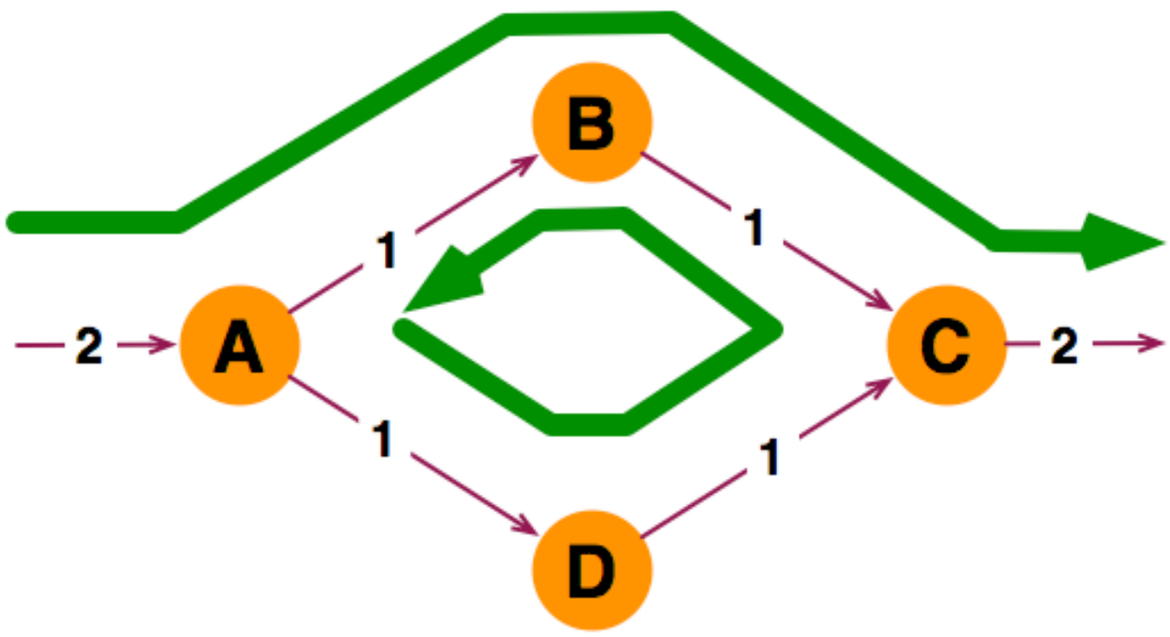


For any numerical values of v4 and v6, a flux vector will be computed that lies in the null space

Example

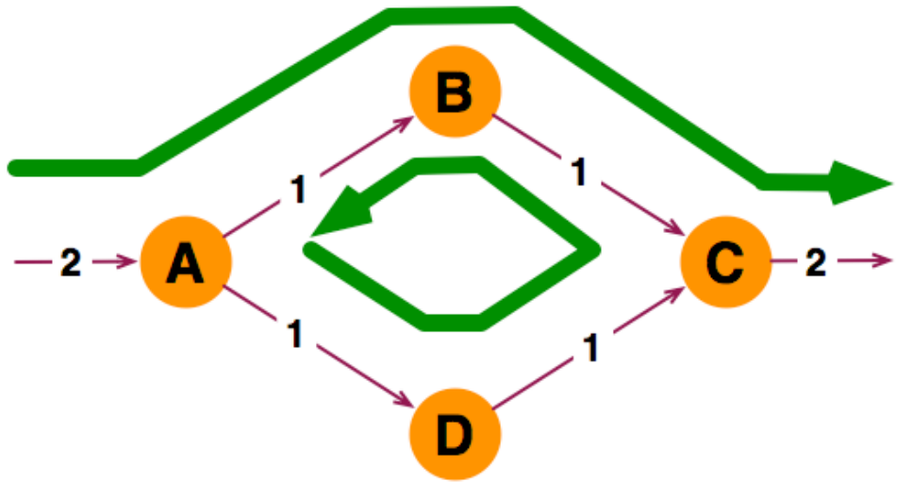
Any steady-state flux distribution is a unique linear combination of the two basis vectors. For example,

$$\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \\ 1 \\ 1 \end{bmatrix} = w_1 r_1 + w_2 r_2 = 2 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = 2r_1 + 1r_2$$



Example

$$v = \begin{bmatrix}2\\1\\1\\2\\1\\1\end{bmatrix} = w_1r_1+w_2r_2 = 2\begin{bmatrix}1\\1\\1\\1\\0\\0\end{bmatrix} + 1\begin{bmatrix}0\\-1\\-1\\0\\1\\1\end{bmatrix} = 2r_1+1r_2$$



This set of basis vectors, although mathematically valid, is chemically unsatisfactory. The reason is that the second basis vector, r_2, represents fluxes through irreversible elementary reactions, v_2 and v_3, in the reverse direction, and it thus represents a chemically unrealistic event

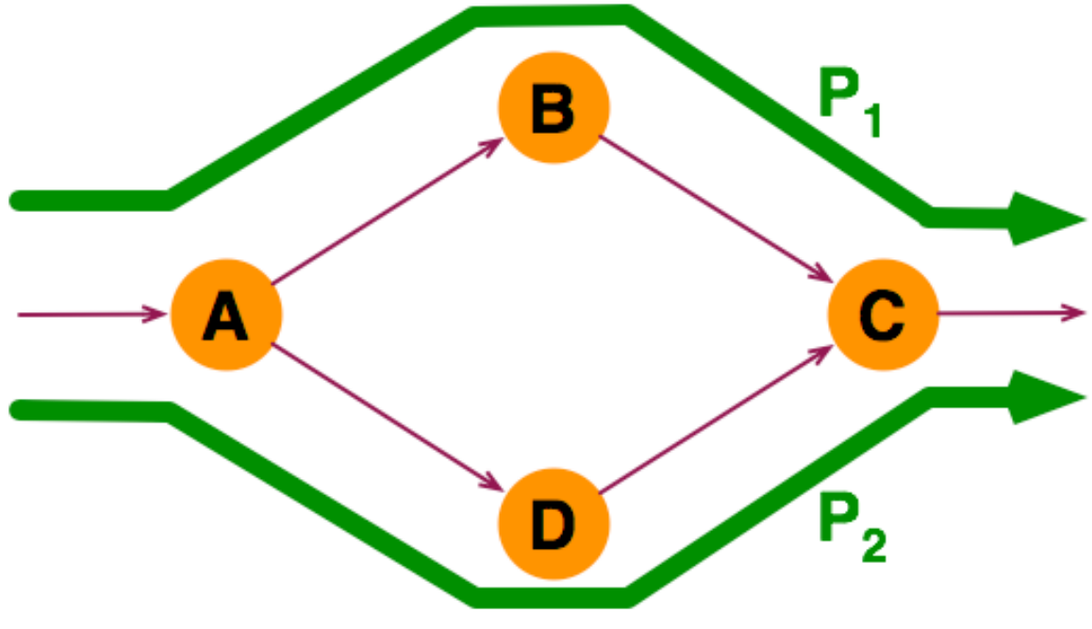
The problem with the acceptability of this basis stems from the fact that the flux through an elementary reaction can only be positive, i.e., $v_i \geq 0$. A negative coefficient in the corresponding row in the basis vector that multiplies the flux is thus undesirable

Example

We can combine the basis vectors to eliminate all negative elements in them. This combination is achieved by transforming the set of basis vectors by

$$(r_1,r_2) = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & -1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = (p_1,p_2)$$

In this new basis, $\mathbf{p}_1=\mathbf{r}_1$, whereas $\mathbf{p}_2=\mathbf{r}_1+\mathbf{r}_2$



Linear vs. Convex Bases

- The introduction of nonnegative basis vectors leads to convex analysis
- Convex analysis is based on equalities (in this case, $Sv=0$) and inequalities (in this case, $0 \leq v_i \leq v_{i,max}$)
- It leads to the definition of a set of nonnegative generating vectors

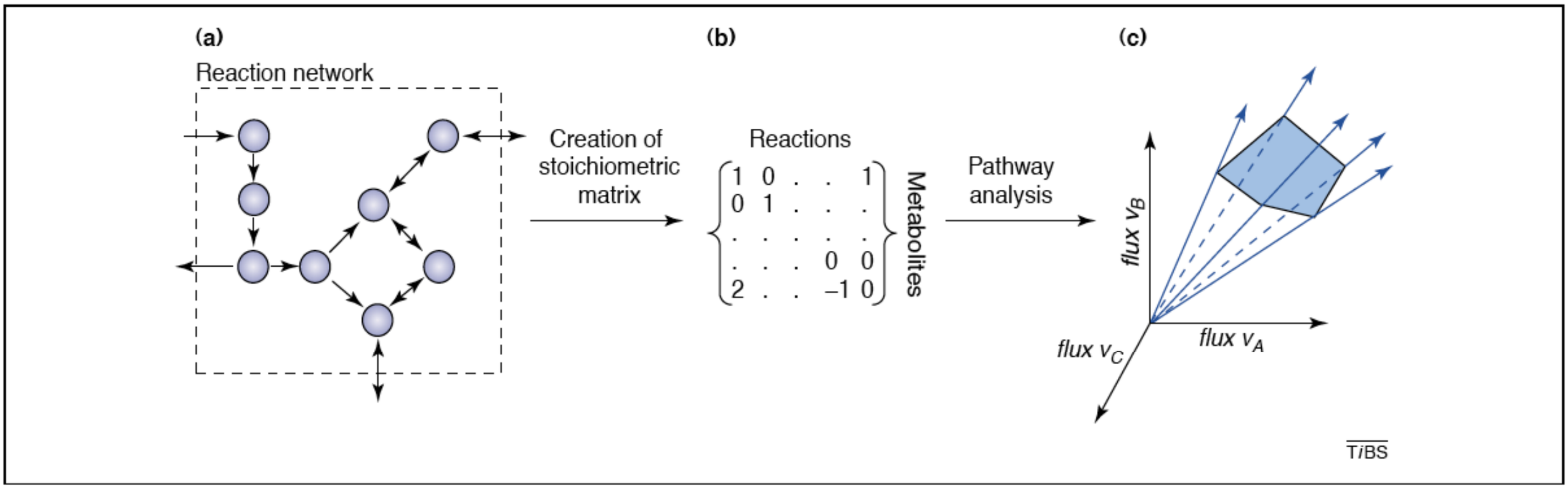
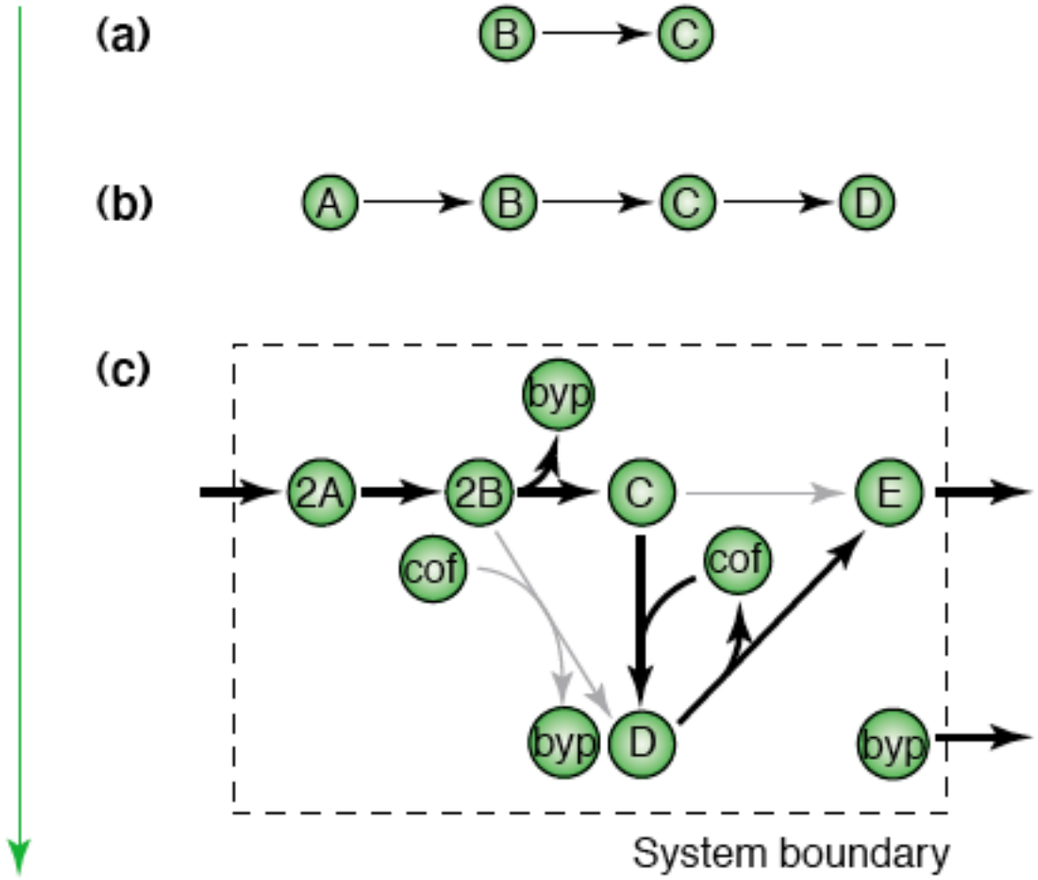


Fig. 2. Network-based pathways mathematically characterize the functions of a metabolic network. As a framework for understanding what these network-based pathways can characterize, we begin with a given reaction network (e.g. *E. coli* metabolism). Genome annotations, biochemical experiments and cell physiology data provide information to describe all reactions within a system. (a) A reaction network is created from diverse data sets by defining all the different reactions in an organism. EXCHANGE FLUXES cross the system boundary; INTERNAL FLUXES are within the system boundary. (b) These data are then used to create a stoichiometric matrix that relates all of the reactions within a network to all of the participating metabolites in a given organism. The stoichiometric matrix is an important part of the *in silico* model. With the matrix, extreme pathway and elementary mode analyses can be used to generate a unique set of pathways. (c) In a high-dimensional flux space in which each axis corresponds to the flux through a given reaction, the network-based pathways define the limits to the possible steady-state flux distributions that a network can achieve. All possible flux distributions of a metabolic network lie within the 'cone' circumscribed by the pathways.

From Reactions To Pathways To Networks, and Back to Pathways

Evolution of network-based pathways



"Pathways are concepts, but networks are reality."

Extreme Pathways

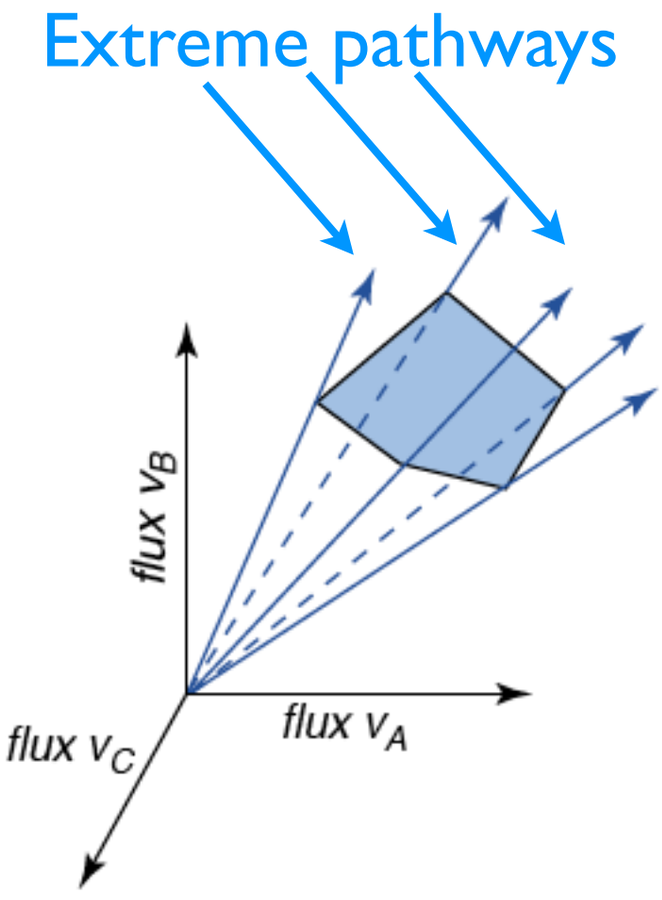
- Biochemically meaningful steady-state flux solutions can be represented by a nonnegative linear combination of convex basis vectors as

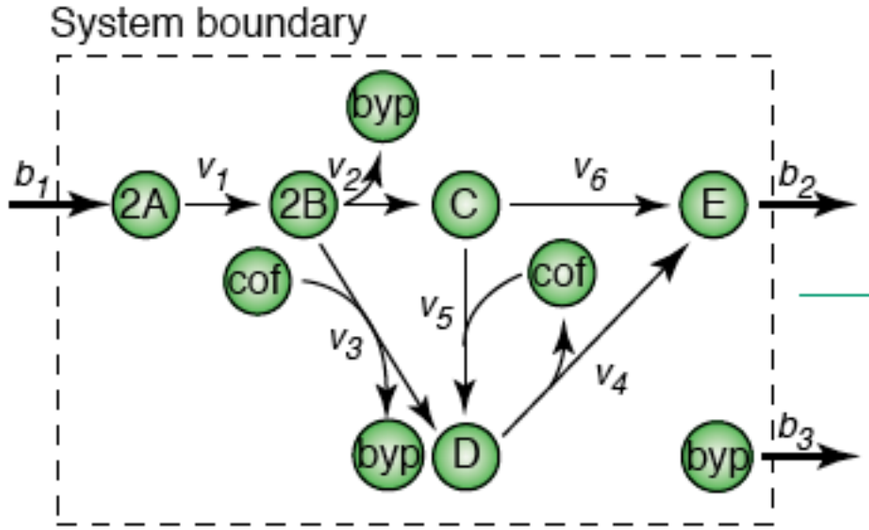
$$\mathbf{v}_{ss} = \sum \alpha_i \mathbf{p}_i \qquad \text{where} \quad 0 \leq \alpha_i \leq \alpha_{i,\max}$$

- The vectors $\mathbf{p}_i$ are a unique convex generating set, but α_i may not be unique for a given $\mathbf{v}_{ss}$

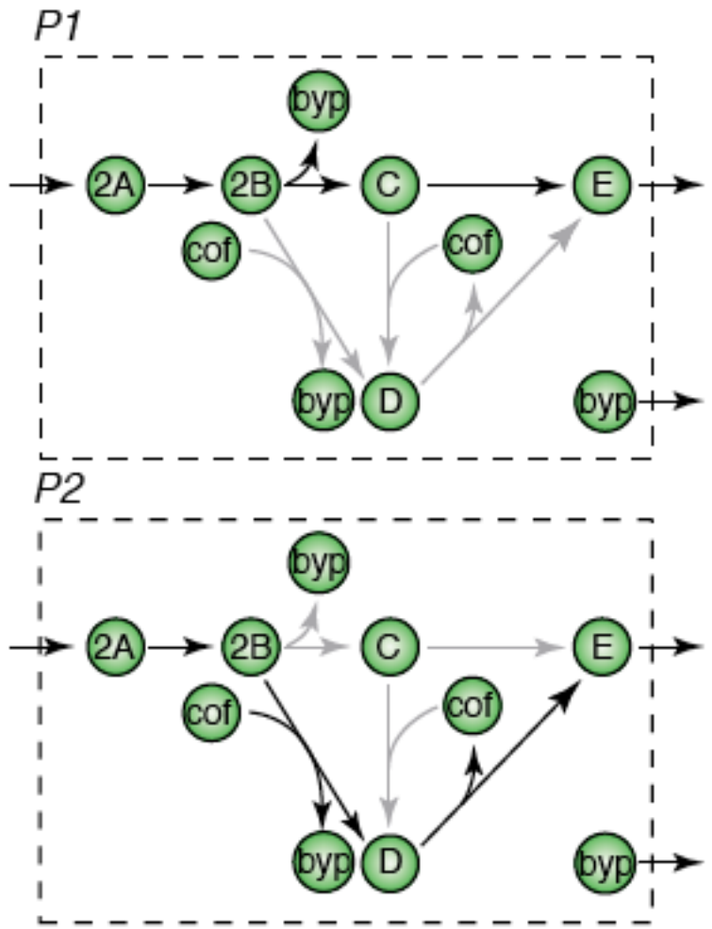
- These vectors correspond to the edges of a cone

- They also correspond to pathways when represented on a flux map and are called extreme pathways, since they lie at the edges of the bounded null space in its conical representation





$$S = \begin{array}{c} \begin{array}{cccccc|ccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & b_1 & b_2 & b_3 \end{array} \\ \left\{\begin{array}{cccccc|ccc} -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 \\ +1 & -2 & +2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & +1 & -1 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 & 0 & +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & +1 & -1 & 0 & 0 & 0 & 0 \end{array}\right. \end{array} \begin{array}{l} A \\ B \\ C \\ D \\ E \\ byp \\ cof \end{array}$$



$$P = \begin{array}{c} \begin{array}{ccc} P_1 & P_2 & P_3 \end{array} \\ \left\{\begin{array}{ccc} 2 & 2 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{array}\right. \\ \left\{\begin{array}{ccc} 2 & 2 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{array}\right. \end{array} \begin{array}{l} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ b_1 \\ b_3 \\ b_2 \end{array}$$

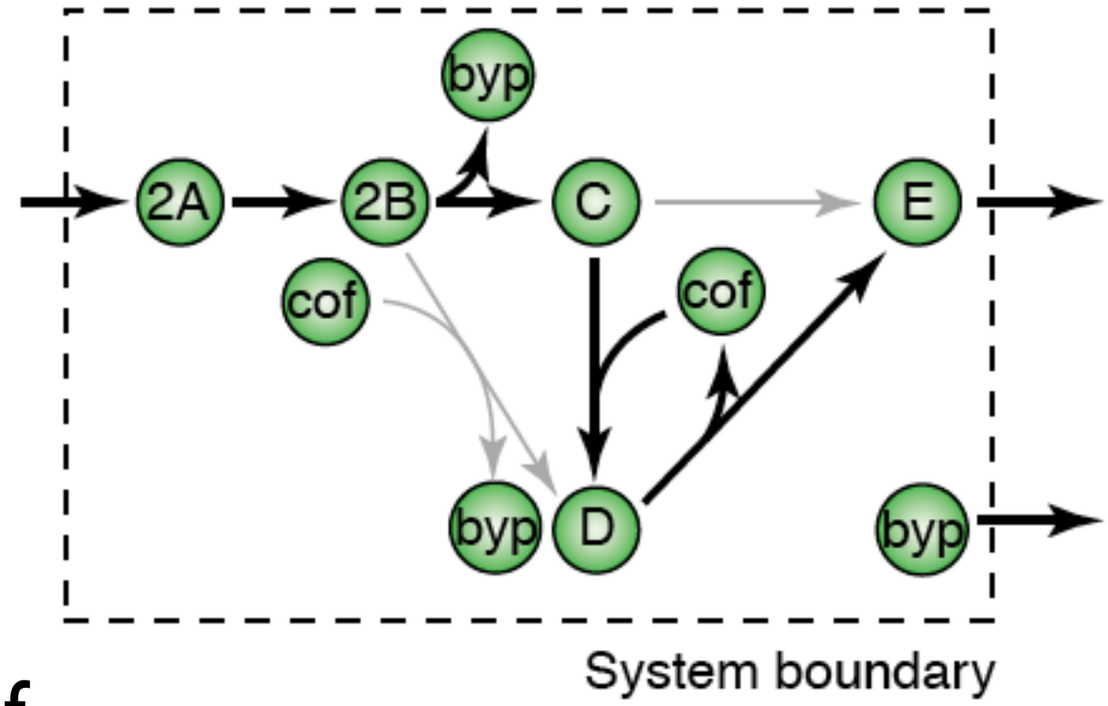
Extreme Pathways

- Every point within the cone (C) can be written as a nonnegative linear combination of the extreme pathways as

$$C = \left\{ \mathbf{v} : \mathbf{v} = \sum_{i=1}^{k} w_i \mathbf{p}_i, \quad w_i \geqslant 0 \ \forall i \right\}$$

Putting It All Together: Convex Analysis of Metabolic Networks

- A cellular metabolic reaction network is a collection of enzymatic reactions and transport processes that serve to replenish and drain the relative amounts of certain metabolites



- A system boundary can be drawn around all these types of physically occurring reactions, which constitute internal fluxes operating inside the network

Putting It All Together: Convex Analysis of Metabolic Networks

- The system is closed to the passage of certain metabolites while others are allowed to enter and/or exit they system based on external sources and/or sinks which are operating on the network as a whole
- The existence of an external source/sink on a metabolite necessitates the introduction of an exchange flux, which serves to allow a metabolite to enter or exit the theoretical system boundary

Putting It All Together: Convex Analysis of Metabolic Networks

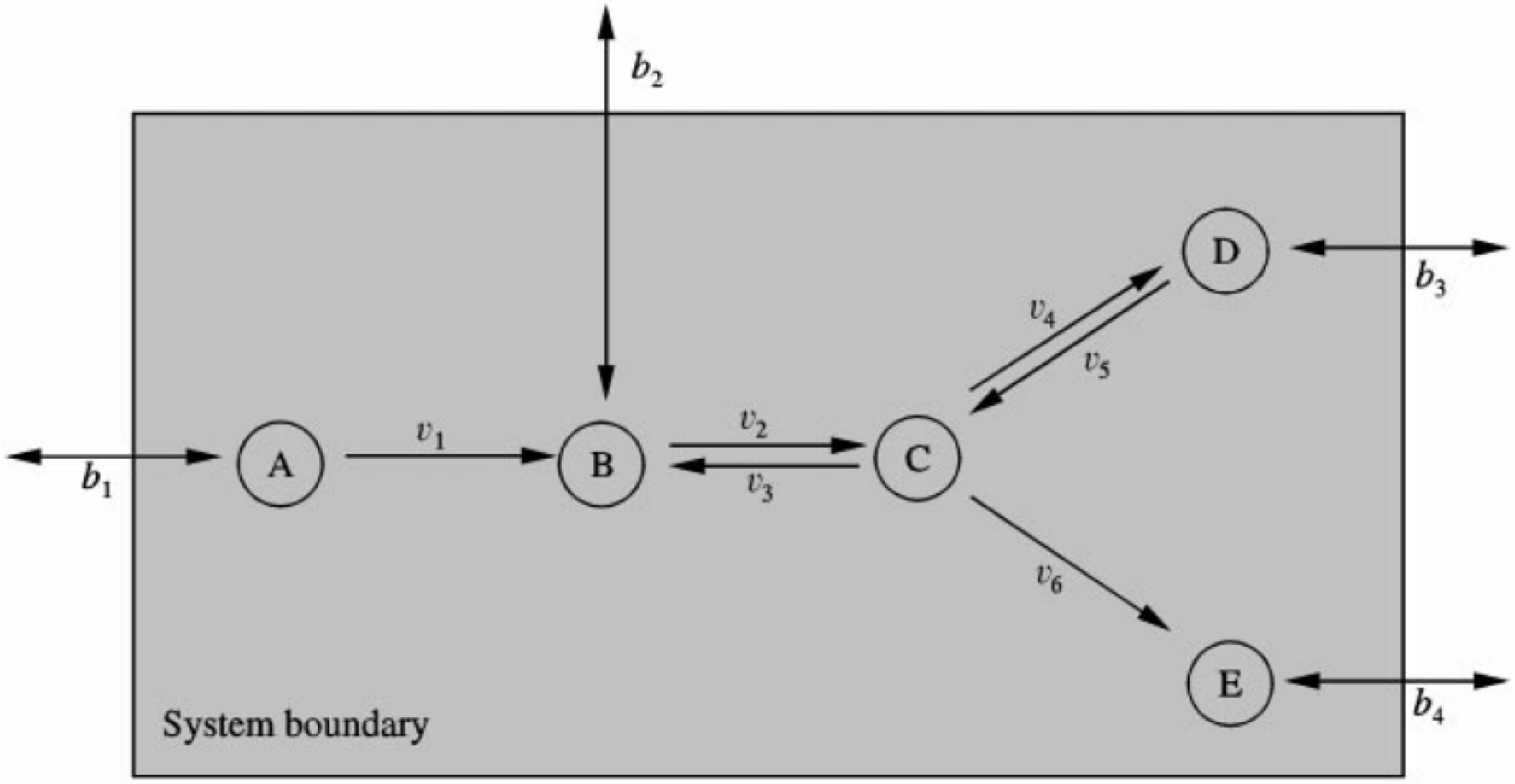
- All internal fluxes are denoted by v_i, for $i \in [1, n_I]$, where n_I is the number of internal fluxes
- All exchange fluxes are denoted by b_i, for $i \in [1, n_E]$, where n_E is the total number of exchange fluxes
- Thermodynamic information can be used to determine if a chemical reaction can proceed in the forward and reverse directions or it is irreversible thus physically constraining the direction of the reaction

Putting It All Together: Convex Analysis of Metabolic Networks

- All internal reactions that are considered to be capable of operating in a reversible fashion are considered (for mathematical purposes only) as two fluxes occurring in opposite directions, therefore constraining all internal fluxes to be nonnegative
- There can only be one exchange flux per metabolite, whose activity represents the net production and consumption of the metabolite by the system
- Thus, n_E can never exceed the number of metabolites in the system

Putting It All Together: Convex Analysis of Metabolic Networks

- The activity of these exchange fluxes is considered to be positive if the metabolite is exiting the system, and negative if the metabolite is entering the system or being consumed by the system
- For all metabolites in which a source or sink may be present, the exchange flux can operate in a bidirectional manner and is therefore unconstrained


b_2
D
b_3
v_4
v_5
b_1
A
v_1
B
v_2
v_3
C
v_6
E
b_4
System boundary

- A simple, yet informative, analysis of a metabolic system may involve studying the systems structural characteristics or invariant properties − those depending neither on the state of the environment nor on the internal state of the system, but only on its structure

- The stoichiometry of a biochemical reaction network is the primary invariant property that describes the architecture and topology of the network

Dynamic Mass Balance

S: Stoichiometric matrix

v=(v_1,v_2,...,v_n): flux vector

x=(x_1,x_2,...,x_m): vector of derivatives of the concentration vector

$$\frac{d\mathbf{x}}{dt} = \mathbf{S} \cdot \mathbf{v}$$

Steady-state Analysis

- The desired pathway structure should be an invariant property of the network (along with stoichiometry)
- This can be achieved by imposing a steady-state condition:

$$S \cdot v = 0$$

Constraints

- All internal fluxes must be nonnegative:

$$v_i \geq 0, \forall i$$

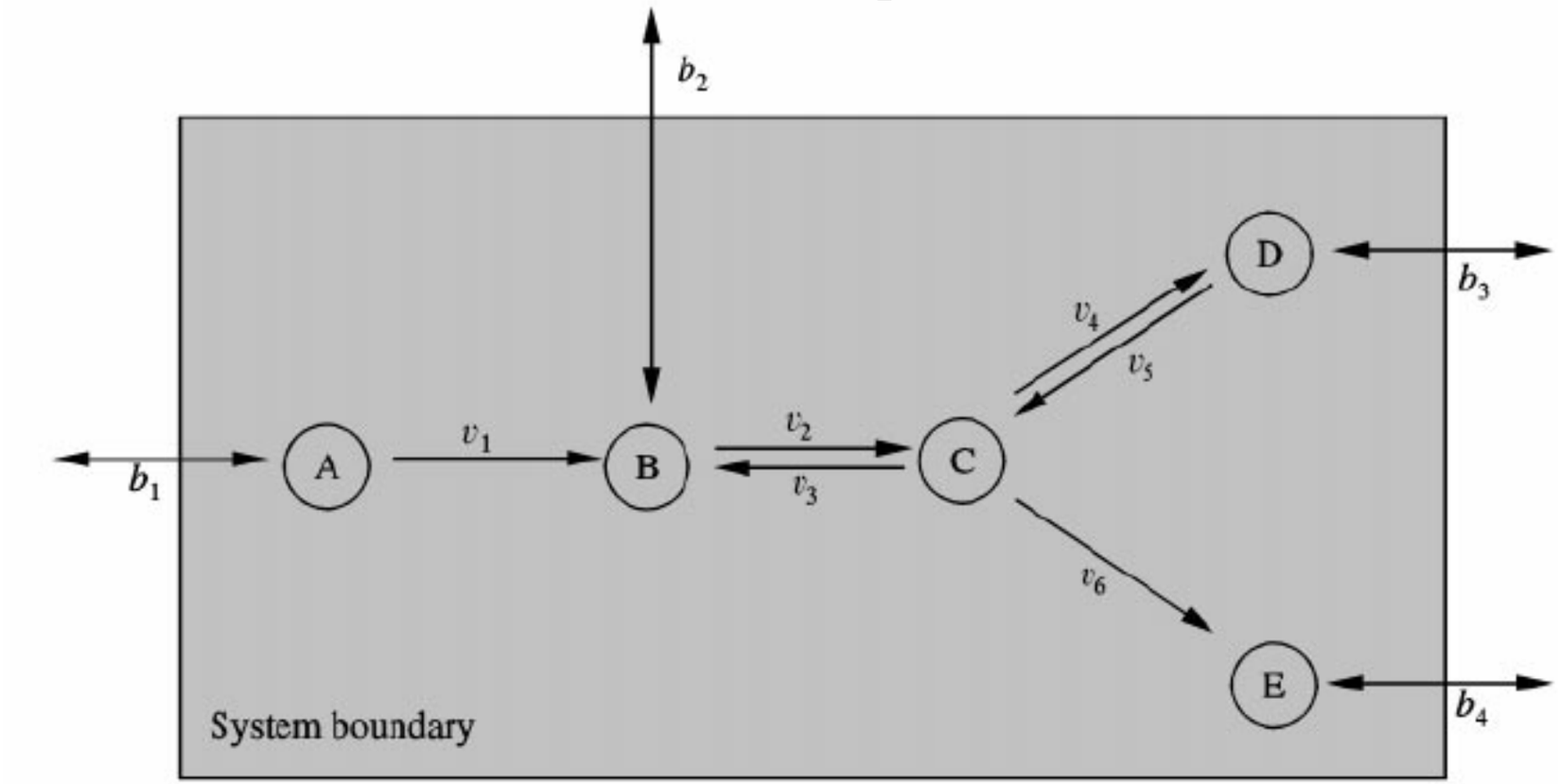
Constraints

- For external fluxes, we have:

$$\alpha_j \leq b_j \leq \beta_j, \quad \forall j$$

- If only a source (input) exists, only α_j is set to negative infinity and β_j is set to zero
- If only a sink (output) exists on the metabolite, α_j is set to zero and β_j is set to positive infinity
- If both a source and a sink are present on the metabolite, then the exchange flux is bidirectional with α_j set to negative infinity and β_j set to positive infinity, leaving the exchange flux unconstrained

Example



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

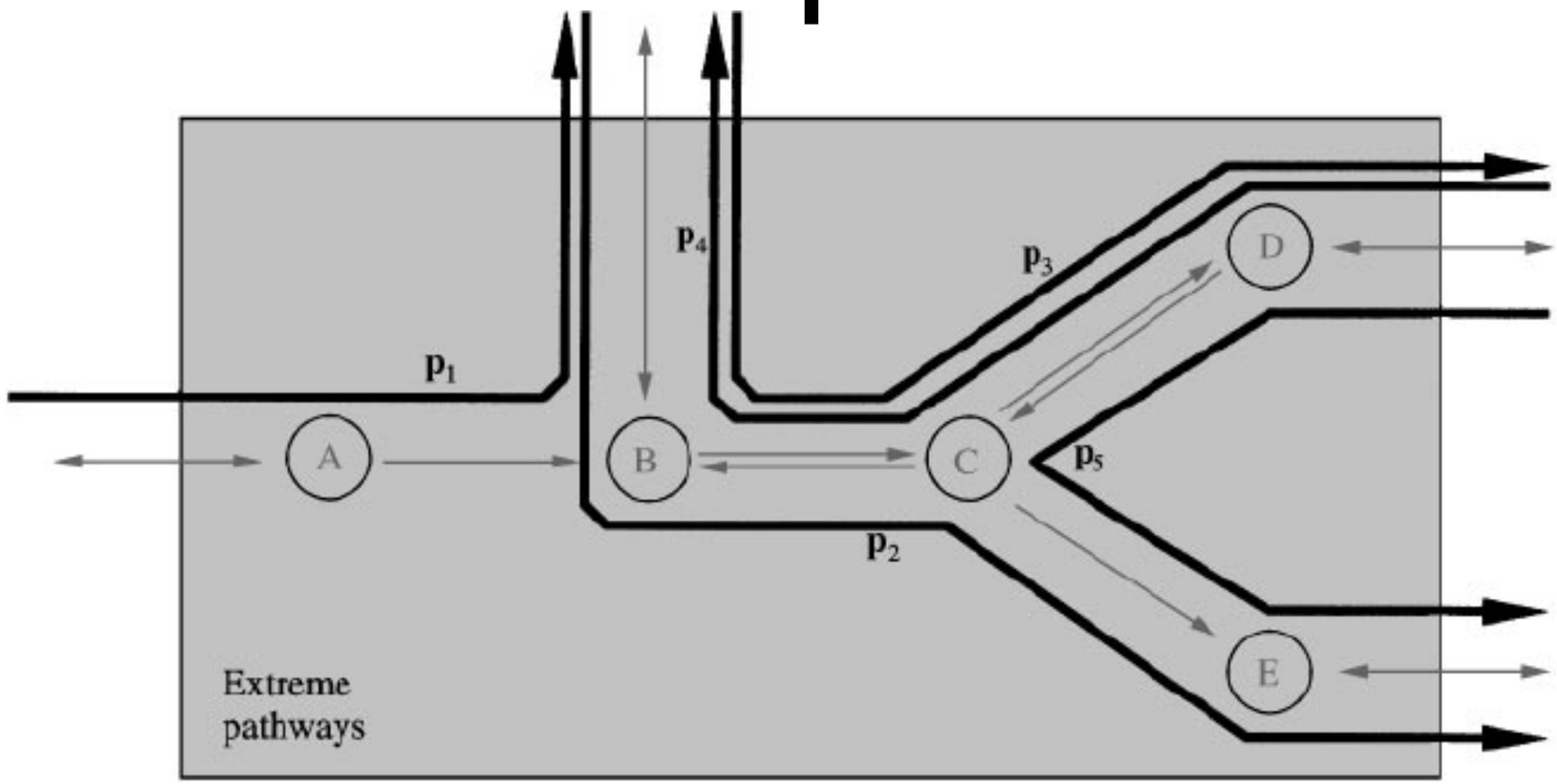
$$v_j \geq 0, \quad j = 1,\ldots,6$$

Exchange flux constraints

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

Example

(a)



(b)

$$\mathbf{P} = \begin{array}{c} \begin{matrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 & \mathbf{p}_5 & \mathbf{p}_6 & \mathbf{p}_7 \end{matrix} \\ \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ \hdashline -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{array}\right] \end{array} \begin{array}{l} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ \hdashline b_1 \\ b_2 \\ b_3 \\ b_4 \end{array}$$

(c) **Flux distribution:**

$$\mathbf{v}^T = [4\ 2\ 0\ 1\ 0\ 1 -4\ 2\ 1\ \ 1]$$

Subset pathways of v:

$$\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$$

Null space dimensions of $\mathbf{S}_{mod}$:

$$d(\mathbf{S}_{mod}) = n - r(\mathbf{S}_{mod}) = 8 - 5 = 3$$

Unique decomposition of v:

$$\mathbf{w}^T = [4\ 1\ 1\ 0\ 0\ 0\ 0]$$

or

$$\mathbf{v} = (4)\cdot\mathbf{p}_1 + (1)\cdot\mathbf{p}_2 + (1)\cdot\mathbf{p}_3$$

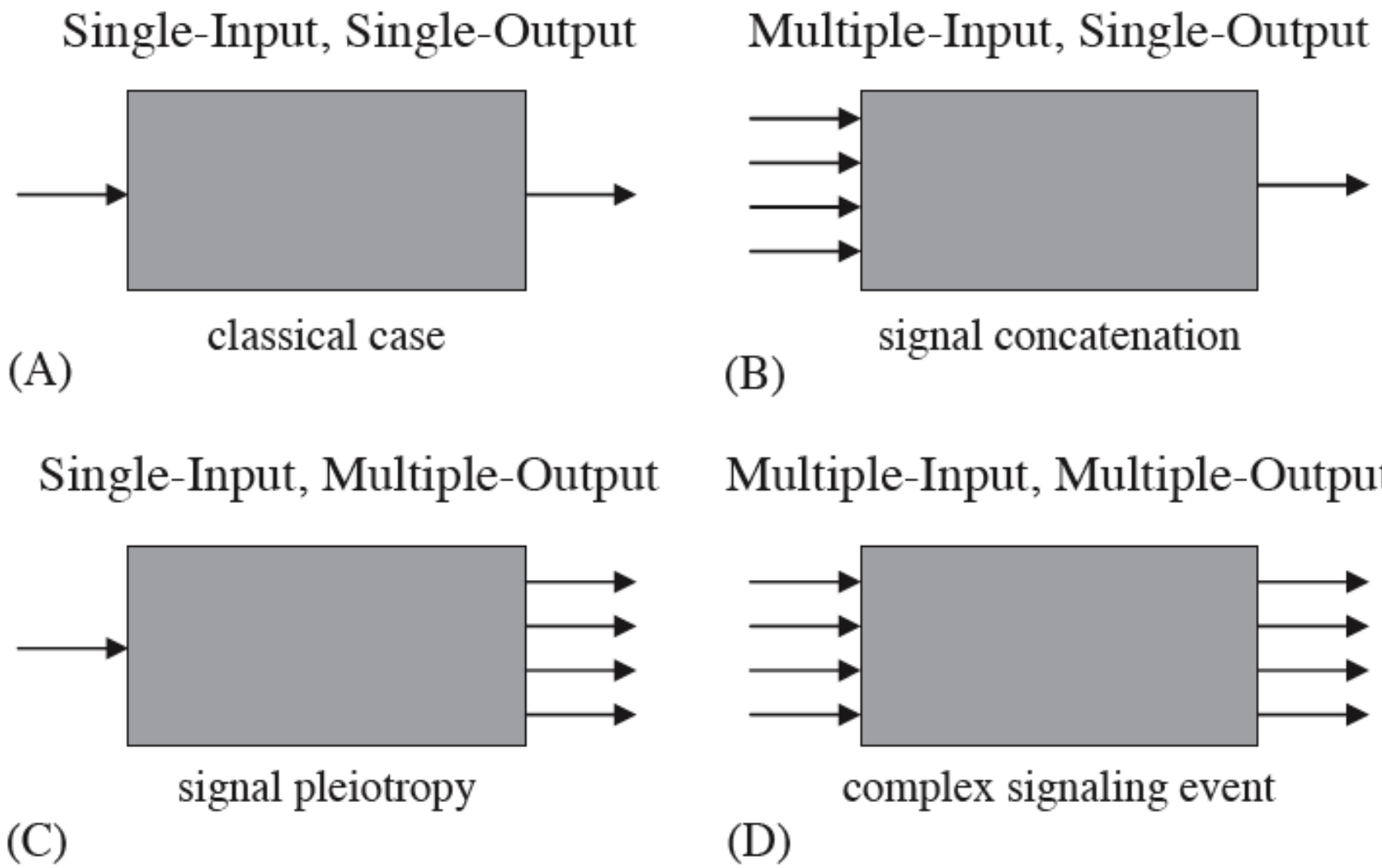
Topological Analysis of (mass-balanced) Regulatory Networks Using Extreme Pathways

- Various modeling approaches have been successfully used to investigate particular features of small-scale signaling networks

- However, large-scale analyses of signaling networks have been lacking, due in part to (1) a paucity of values for kinetic parameters, (2) concerns regarding the accuracy of existing values for kinetic data, (3) strong computational demands of kinetic analyses, and (4) limited scalability from small signaling modules using kinetic models

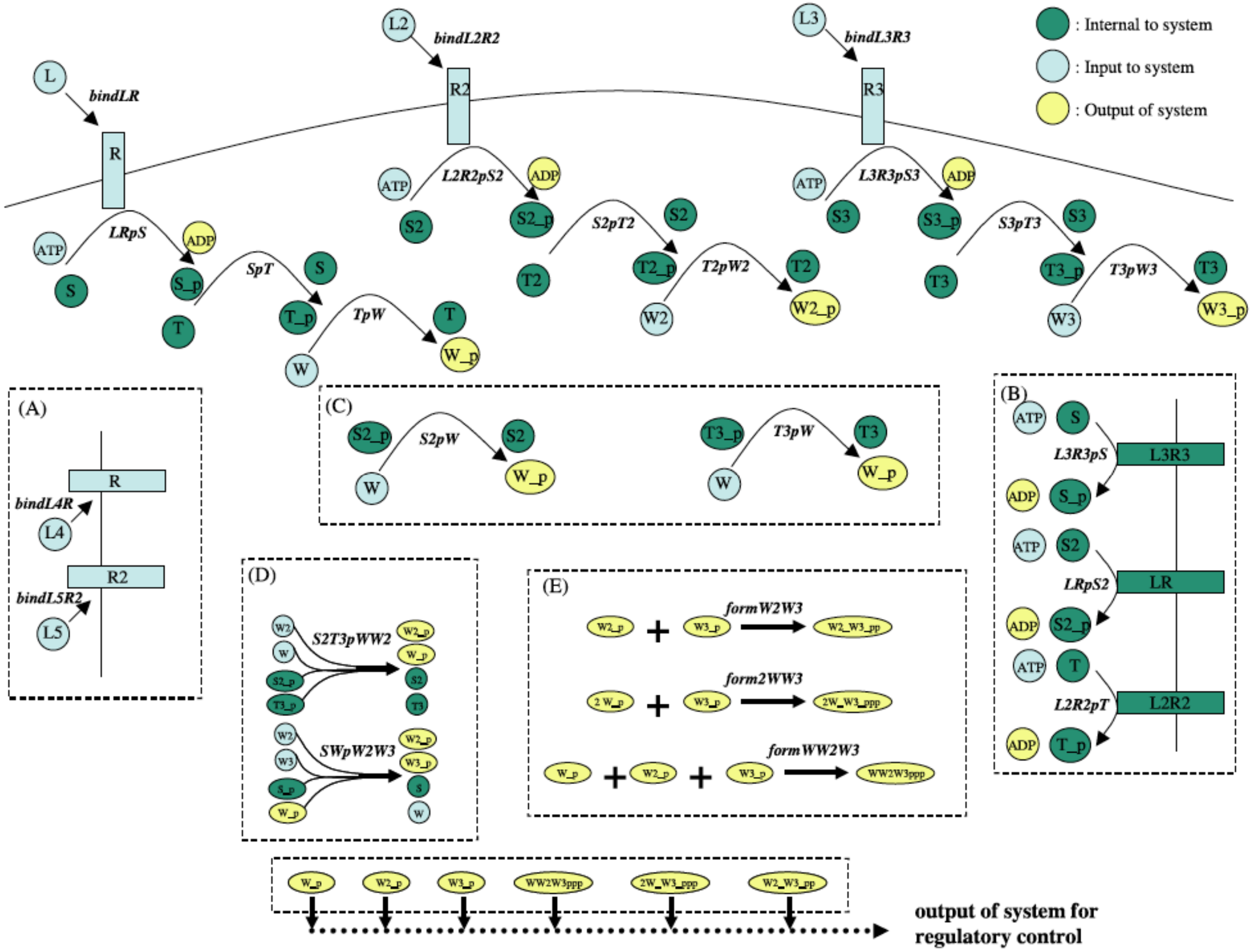
- Obtaining the extreme pathways of a mass-balanced signaling networks allows for analyses focused and based solely on the structure (topology, or connectivity) of a signaling network

Categories of Signal Transduction Events



Classification of signal transduction input–output relationships. The classical case of a transduced signal relates a single input to a single output (A). Some outputs require the concatenation of multiple inputs (B). Other signaling interactions occur in which the transduction of a single input generates multiple outputs, a type of signaling pleiotropy (C). Complex signaling events arise as multiple inputs trigger interacting signaling cascades that result in multiple outputs (D).

Prototypic Signaling Network



Reaction Listing

Name	Chemical equation
bindLR	$L + R \rightarrow LR$
LRpS	$LR + ATP + S \rightarrow ADP + S\_p + LR\_in$
SpT	$S\_p + T \rightarrow T\_p + S$
TpW	$T\_p + W \rightarrow W\_p + T$
bindL2R2	$L2 + R2 \rightarrow L2R2$
L2R2pS2	$L2R2 + ATP + S2 \rightarrow ADP + S2\_p + L2R2\_in$
S2pT2	$S2\_p + T2 \rightarrow S2 + T2\_p$
T2pW2	$T2\_p + W2 \rightarrow T2 + W2\_p$
bindL3R3	$L3 + R3 \rightarrow L3R3$
L3R3pS3	$L3R3 + ATP + S3 \rightarrow ADP + S3\_p + L3R3\_in$
S3pT3	$S3\_p + T3 \rightarrow S3 + T3\_p$
T3pW3	$T3\_p + W3 \rightarrow T3 + W3\_p$
bindL4R	$L4 + R \rightarrow LR$
bindL5R2	$L5 + R2 \rightarrow L2R2$
L3R3pS	$L3R3 + ATP + S \rightarrow ADP + S\_p + L3R3\_in$
LRpS2	$LR + ATP + S2 \rightarrow ADP + S2\_p + LR\_in$
L2R2pT	$L2R2 + ATP + T \rightarrow ADP + T\_p + L2R2\_in$
S2pW	$S2\_p + W \rightarrow S2 + W\_p$
T3pW	$T3\_p + W \rightarrow T3 + W\_p$
S2T3pWW2	$S2\_p + T3\_p + W + W2 \rightarrow S2 + T3 + W2\_p + W\_p$
SWpW2W3	$S\_p + W\_p + W2 + W3 \rightarrow W2\_p + W3\_p + W + S$
formW2W3	$W2\_p + W3\_p \rightarrow W2\_W3\_pp$
formWW2W3	$W\_p + W2\_p + W3\_p \rightarrow WW2W3\_ppp$
form2WW3	$2\ W\_p + W3\_p \rightarrow 2W\_W3\_ppp$

System Inputs and Outputs

Compound	Input/ Output	Group
L	Input	Signaling inputs
L2	Input	
L3	Input	
L4	Input	
L5	Input	
R	Input	Components of signaling network
R2	Input	
R3	Input	
W	Input	
W2	Input	
W3	Input	
ATP	Input	Interaction with energy metabolism
ADP	Output	
W_p	Output	Signaling outputs
W2_p	Output	
W3_p	Output	
W2_W3_pp	Output	
WW2W3_ppp	Output	
2W_W3_ppp	Output	
LR_in	Output	Degradation of inactive receptor-ligand complex
L2R2_in	Output	
L3R3_in	Output	

- Signaling reactions, like those just described, are subject to mass balance and thermodynamic constraints, and consequently can be analyzed using network-based pathways, including extreme currents, elementary modes, and extreme pathways

- We focus here on extreme pathways and their use in characterizing topological properties of the signaling network

- There are a total of 211 extreme pathways
- These extreme pathways can be studied for

1. feasibility of input/output relationships
2. quantitative analysis of crosstalk
3. pathway redundancy
4. participation of reactions in the extreme pathways
5. correlated reaction sets

Feasibility of Input/Output Relationships

- An assessment of the feasibility of input/output relationships can be performed with extreme pathway analysis because all possible routes through a network can be described by nonnegative combinations of the extreme pathways
- A feasible input/output relationship signifies that with the available signaling inputs there exists a valid combination of the extreme pathways that describes the given signaling output
- Analysis is represented as an "input/output feasibility matrix"

Inputs

Outputs

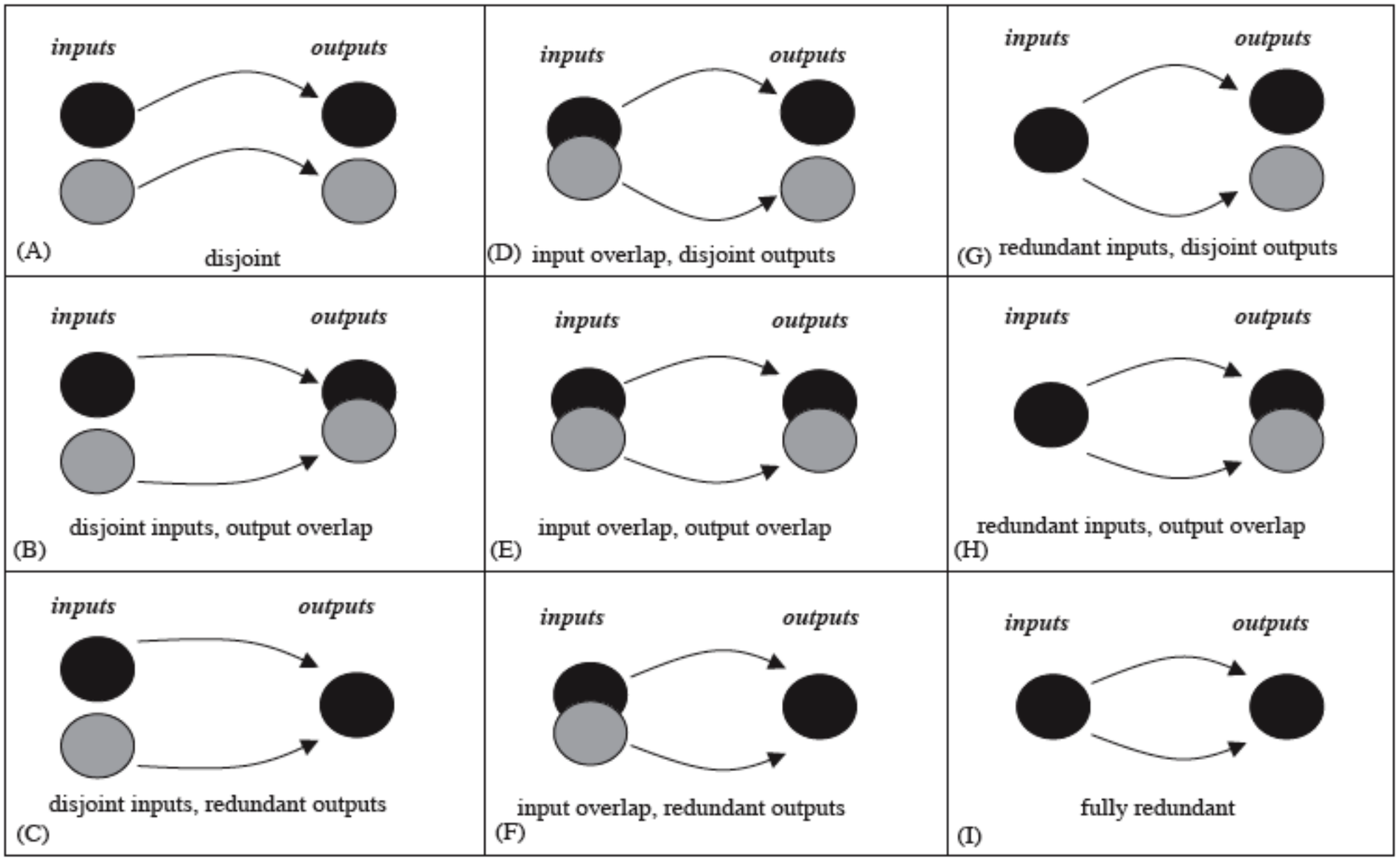
W_p	W2_p	W3_p	W2W3_pp	WW2W3_ppp	pppW2W3		L	L2	L3	L4	L5	Number of Inputs
					x		x					16
				x				x				28
				x	x		x					16
			x					x				28
			x		x		x					16
			x	x				x				28
			x	x	x		x					16
		x					x					16
		x			x		x					16
		x		x			x					16
		x		x	x		x					16
		x	x				x					16
		x	x		x		x					16
		x	x	x			x					16
		x	x	x	x		x					16
	x						x	x				30
	x				x		x					28
	x			x			x	x				27
	x			x	x		x					28
	x		x				x	x				27
	x		x		x		x					28
	x		x	x			x	x				27
	x		x	x	x		x					28
	x	x					x	x				28
	x	x			x		x					16
	x	x		x			x	x				28
	x	x		x	x		x					16
	x	x	x				x	x				28
	x	x	x			x	x					16
	x	x	x		x		x	x				28
	x	x	x		x	x	x					16
	x	x	x	x			x	x				28
	x	x	x	x		x	x					16
	x	x	x	x	x		x	x				28
	x	x	x	x	x	x	x					16
x							x	x				31
x					x		x					16
x				x			x	x				28
x				x	x		x					16
x			x				x	x				28
x			x		x		x					16
x			x	x			x	x				28
x			x	x	x		x					16
x		x					x	x				16
x		x			x		x					16
x		x		x			x	x				16
x		x		x	x		x					16
x		x	x				x	x				16
x		x	x		x		x					16
x		x	x	x			x	x				16
x		x	x	x	x		x					16
x	x						x	x				30
x	x				x		x					28
x	x			x			x	x				27
x	x			x	x		x					28
x	x		x				x	x				27
x	x		x		x		x					28
x	x		x	x			x	x				27
x	x		x	x	x		x					28
x	x	x					x	x				28
x	x	x			x		x					16
x	x	x		x			x	x				28
x	x	x		x	x		x					16
x	x	x	x				x	x				28
x	x	x	x		x		x					16
x	x	x	x	x			x	x				28
x	x	x	x	x	x		x					16
Number of Outputs	31	3	55	31	3	31	63	63	31	31	63	31	3	63	31	63	31	63	63	31	63	63	31	31	63	63	63	31	63	63	63	

Crosstalk Analysis

- Extreme pathway analysis can be used to quantitatively analyze the interconnection of multiple inputs and multiple outputs of signaling pathways, which has been called crosstalk
- Herein, crosstalk is the nonnegative linear combination of extreme pathways of a signaling network
- The pairwise combination of extreme pathways is the simplest form of crosstalk

Crosstalk Analysis

- As such, crosstalk can be classified into nine categories based on extreme pathways



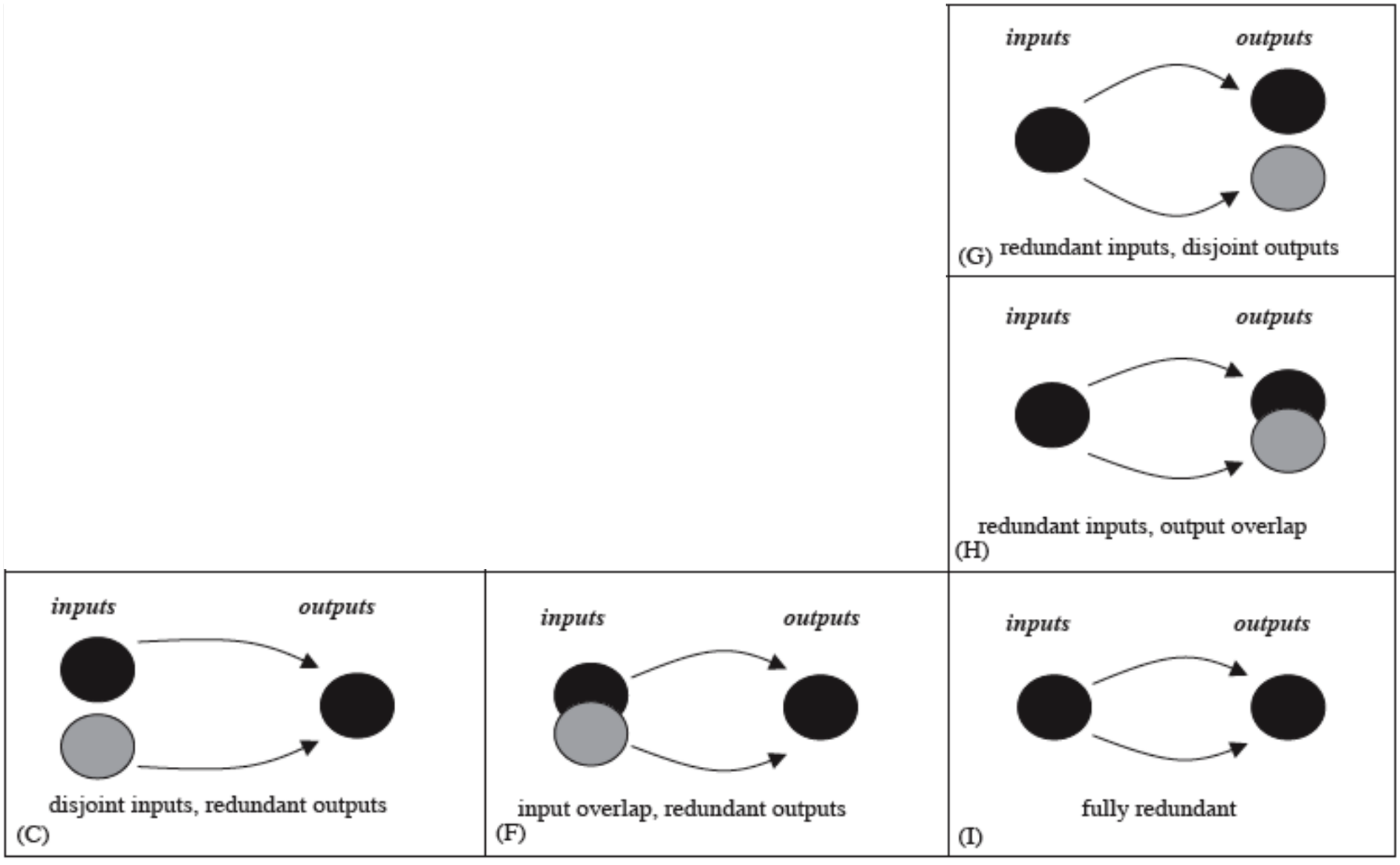
- These classifications are topological, and thus do not account for changes in the activity level of a reaction

Inputs / Outputs				Total
	1.5	58.9	1.6	62.0
	0.0	23.0	0.7	23.7
	0.0	13.4	1.0	14.4
Total	1.5	95.3	3.3	

Crosstalk analysis of the prototypic signaling network following the classification scheme on the previous slide. With 211 extreme pathways, there are a total of 22,155 [=(211^2-211)/2] pair-wise comparisons.

Pathway Redundancy

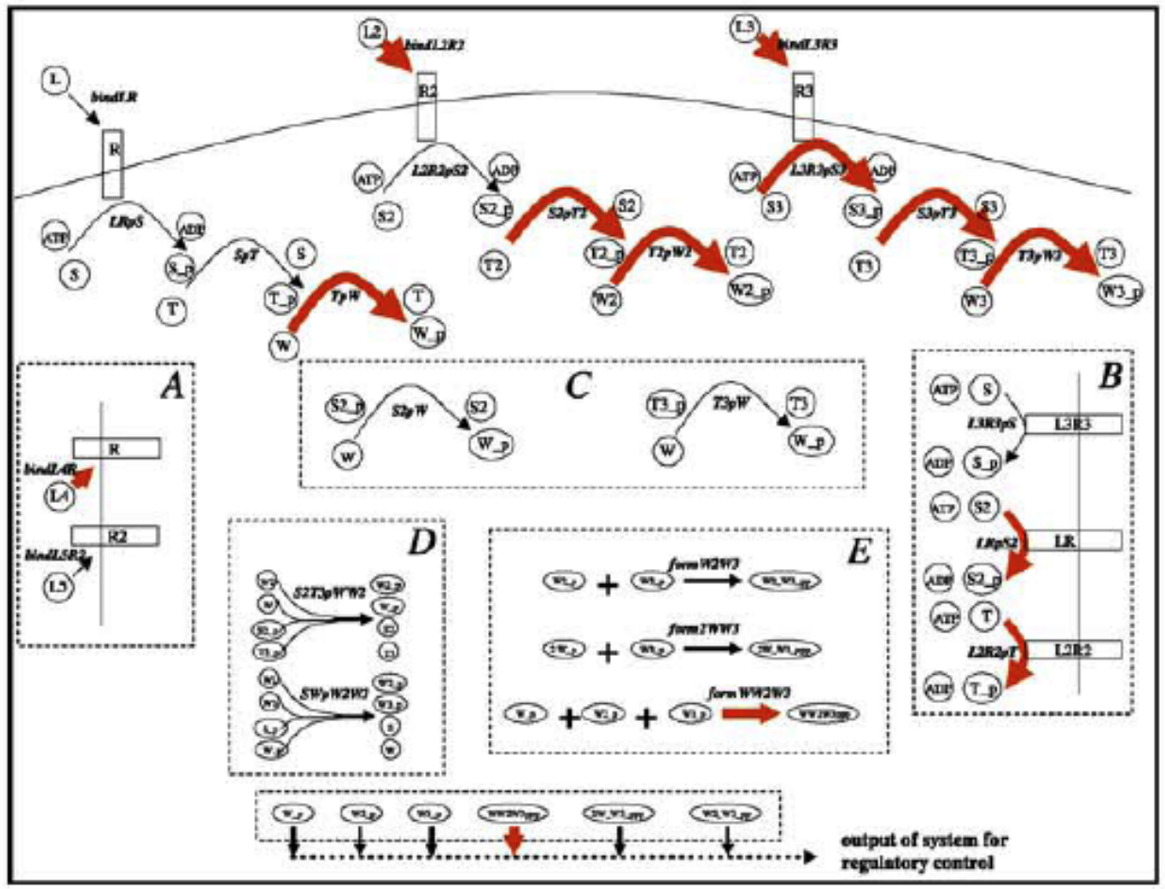
- Two extreme pathways with identical inputs and/or outputs represent two systemically independent routes by which a network can be utilized to reach the same objective



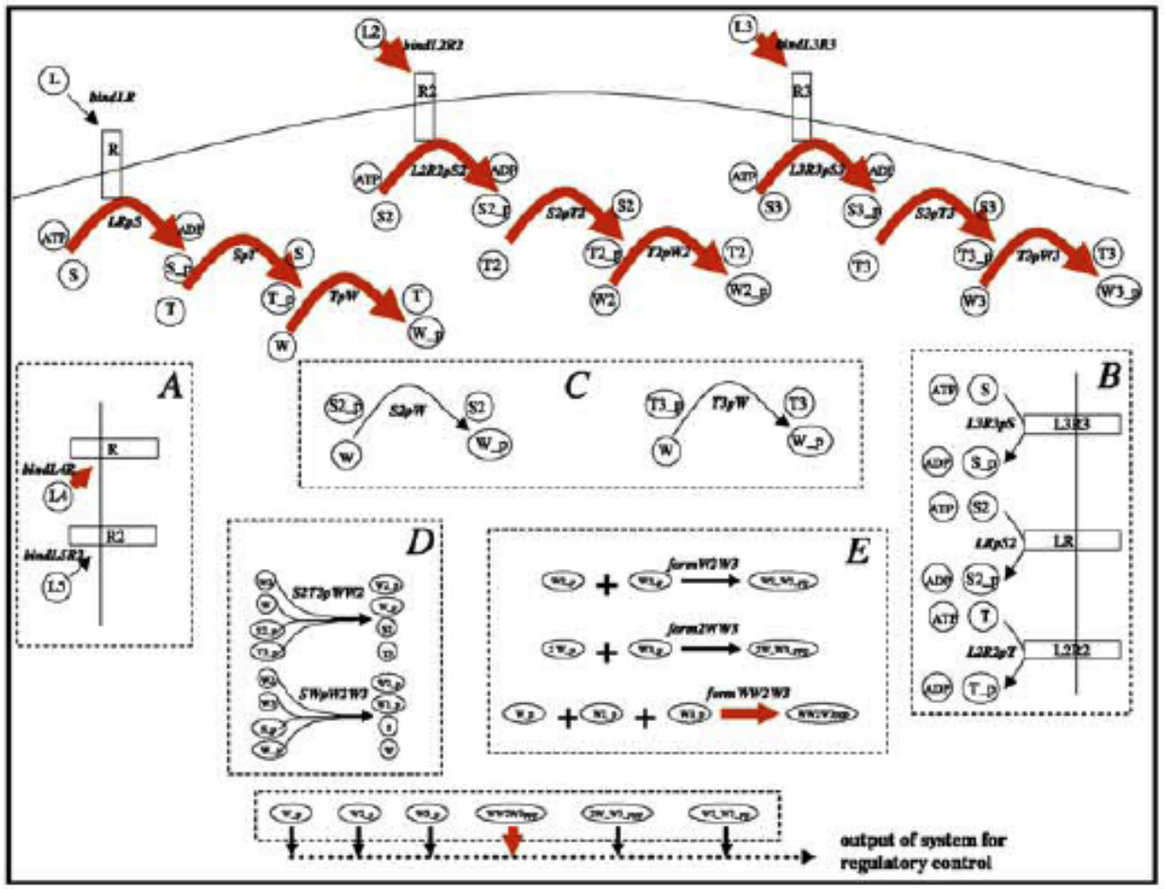
- There were 135 distinct input/output states of the 211 extreme pathways

- This result suggests that on average the prototypical signaling network can convert an identical set of inputs to an identical set of outputs using two systemically independent routes

P_{88}



P_{108}



Completely redundant extreme pathways. Pathways 88 and 108 have identical inputs and outputs and yet use different internal reactions.

- The number of redundant output states with different inputs was also calculated

- There were 17 distinct output states in the set of extreme pathways for the network

The number of extreme pathways with equivalent output states

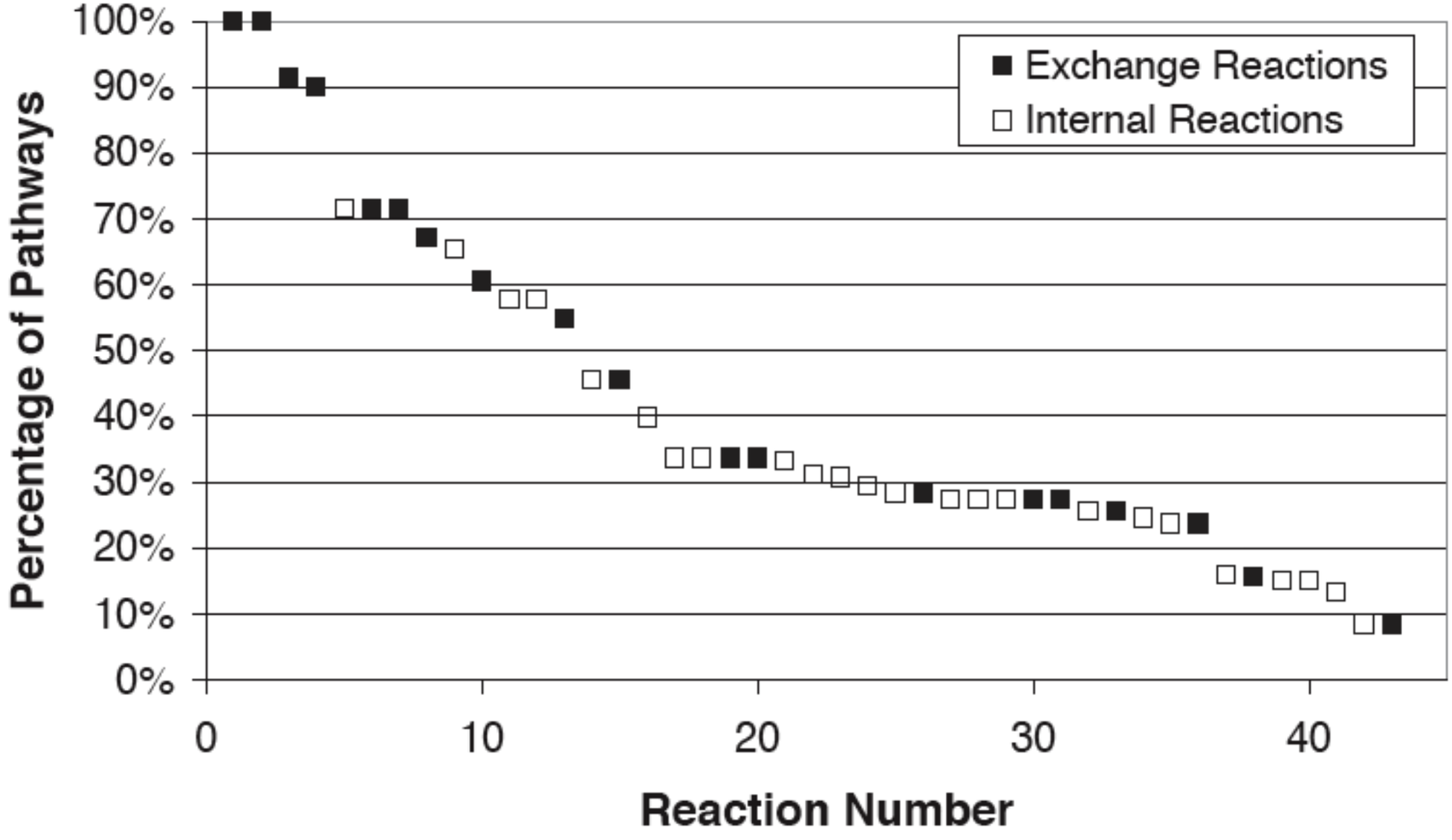
Distinct output states						Number of pathways
W_p	W2_p	W3_p	W2_W3_pp	WW2W3_ppp	2W_W3_ppp	
0	0	0	0	1	0	50
0	0	0	1	0	0	26
0	1	0	0	0	1	22
0	1	1	0	0	0	22
0	4	0	0	0	1	10
0	2	0	0	1	0	10
0	2	1	0	0	0	10
0	1	0	1	0	0	10
0	0	0	2	0	0	10
0	0	0	0	0	1	10
1	0	0	0	0	0	10
0	2	0	0	0	1	4
0	0	0	2	0	1	4
0	1	0	0	0	0	4
1	1	0	0	0	0	4
1	0	0	1	0	0	4
0	0	1	0	0	0	1

Because extreme pathways are systemically independent, the combinatorial effect of the multiple pathways that produce W_p, W2_p, and W3_p cannot explain the redundancy in the output of WW2W3_ppp. Rather, the redundancy is a result of emergent uses of the network to produce the particular transcription factor.

Reaction Participation

- The number of extreme pathways that a particular reaction participates in can be computed efficiently
- Disrupting or regulating the activity of highly connected reactions would influence a large number of extreme pathways, or functional network states
- The percentage (of a total of 211) of extreme pathways that use each individual reaction in the prototypic signaling network was computed

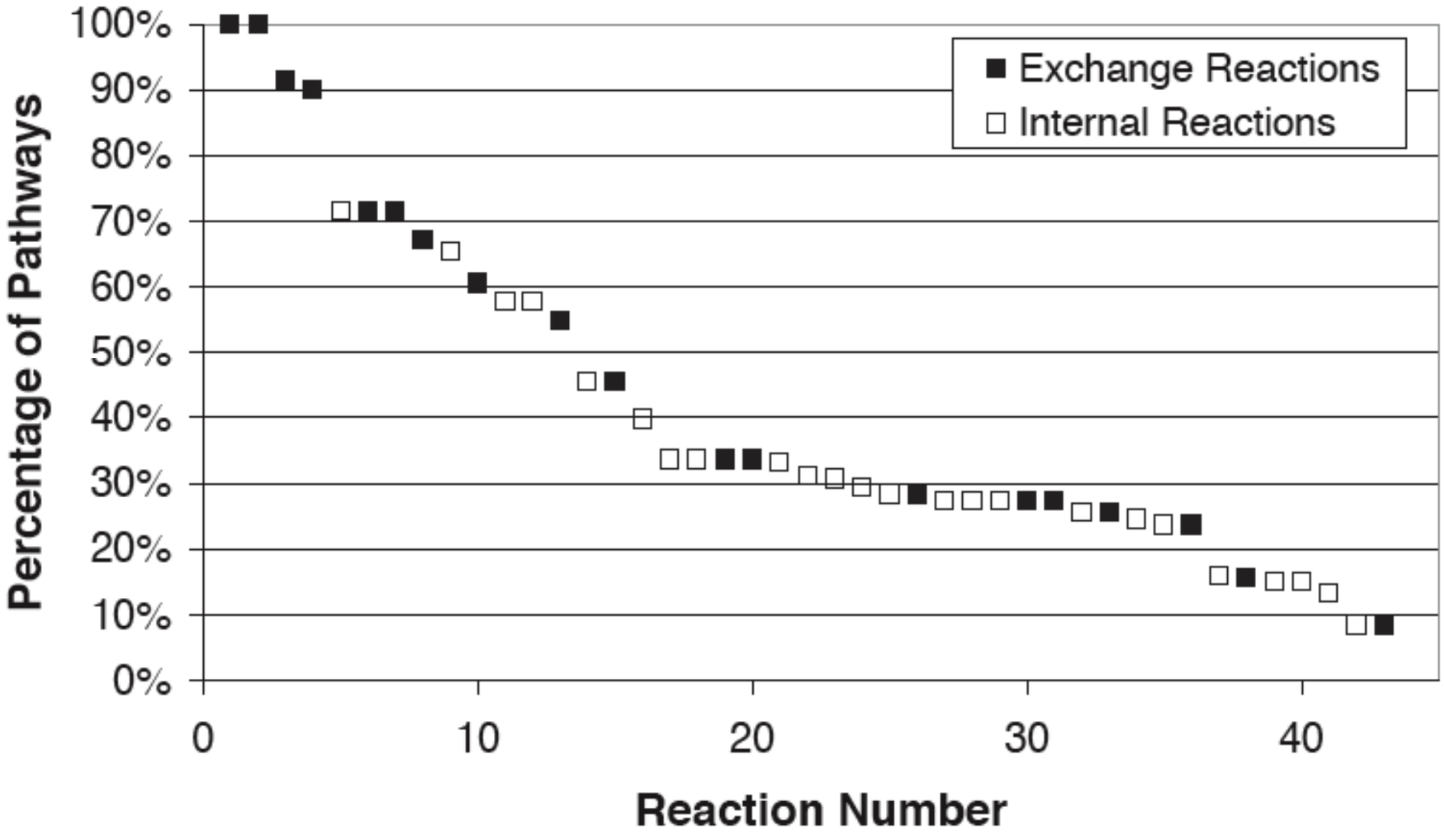
Reaction Participation



Internal Reaction	Percent of Pathways
bindL3R3	72%
SWpW2W3	65%
L3R3pS3	58%
S3pT3	58%
LRpS	46%
L2R2pS2	40%
bindLR	34%
bindL4R	34%
S2T3pWW2	33%
LRpS2	31%
T3pW3	31%
TpW	29%
formWW2W3	28%
bindL2R2	27%
bindL5R2	27%
L3R3pS	27%
formW2W3	26%
S2pW	25%
form2WW3	24%
L2R2pT	16%
S2pT2	15%
T2pW2	15%
SpT	13%
T3pW	9%

Exchange Reaction	Percent of Pathways
ATP	100%
ADP	100%
W3	91%
W2	90%
L3	72%
R3	72%
L3R3_in	72%
R	67%
LR_in	67%
W	61%
R2	55%
L2R2_in	55%
W2_p	45%
L	34%
L4	34%
WW2W3_ppp	28%
L2	27%
L5	27%
W2_W3_pp	26%
2W_W3_ppp	24%
W3_p	16%
W_p	9%

Reaction Participation

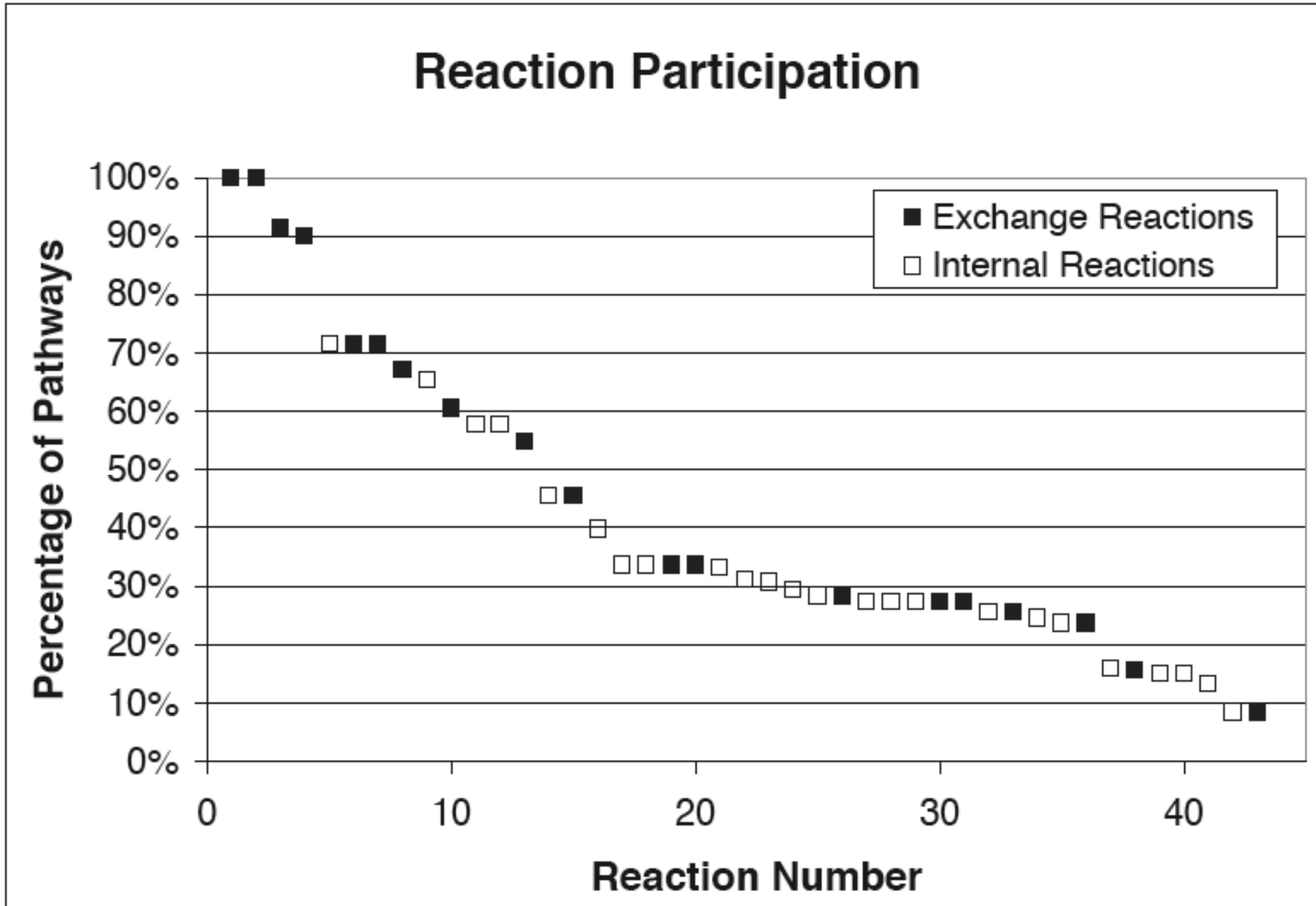


Functional significance

Internal Reaction	Percent of Pathways
bindL3R3	72%
SWpW2W3	65%
L3R3pS3	58%
S3pT3	58%
LRpS	46%
L2R2pS2	40%
bindLR	34%
bindL4R	34%
S2T3pWW2	33%
LRpS2	31%
T3pW3	31%
TpW	29%
formWW2W3	28%
bindL2R2	27%
bindL5R2	27%
L3R3pS	27%
formW2W3	26%
S2pW	25%
form2WW3	24%
L2R2pT	16%
S2pT2	15%
T2pW2	15%
SpT	13%
T3pW	9%

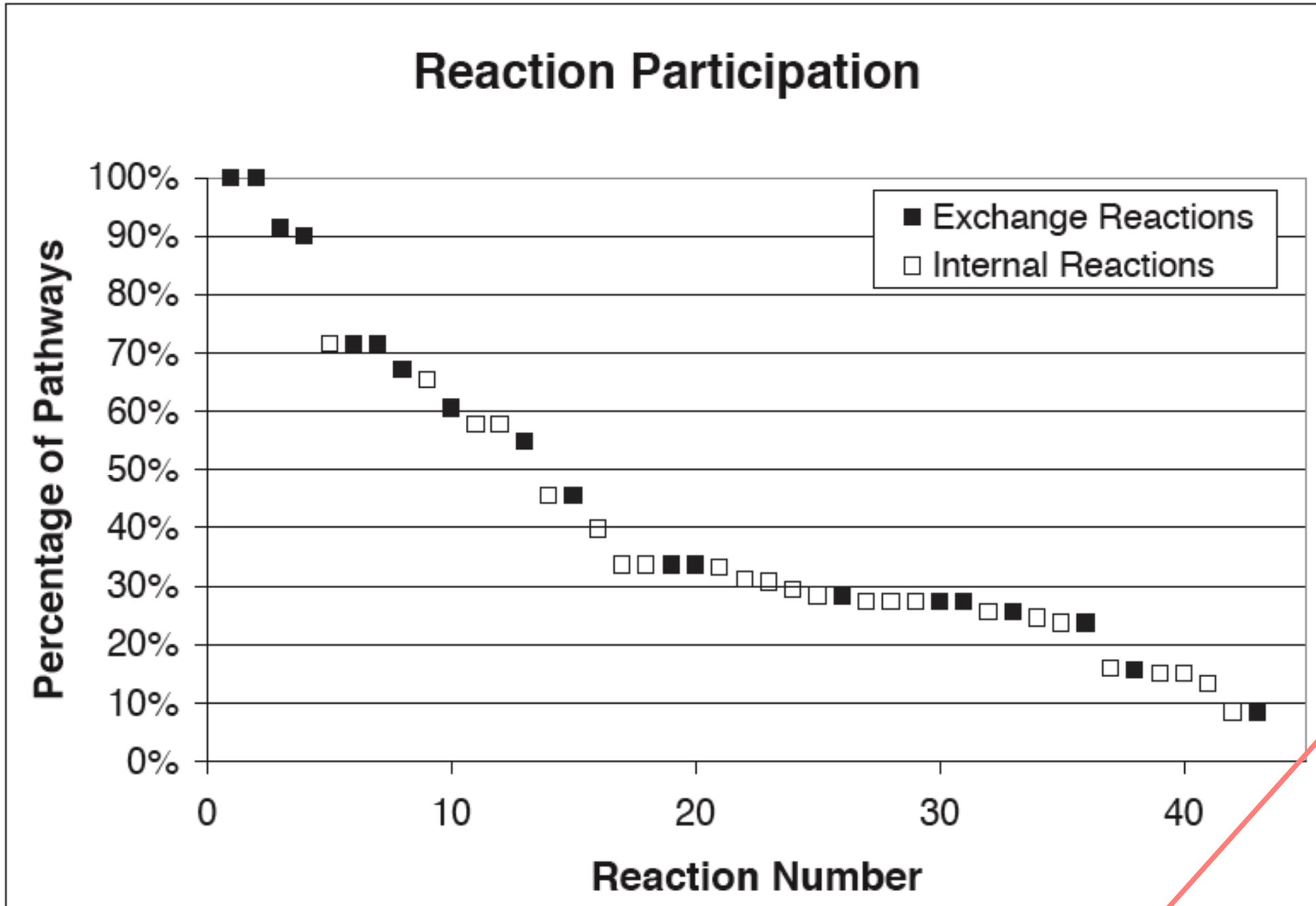
Functional significance

Exchange Reaction	Percent of Pathways
ATP	100%
ADP	100%
W3	91%
W2	90%
L3	72%
R3	72%
L3R3_in	72%
R	67%
LR_in	67%
W	61%
R2	55%
L2R2_in	55%
W2_p	45%
L	34%
L4	34%
WW2W3_ppp	28%
L2	27%
L5	27%
W2_W3_pp	26%
2W_W3_ppp	24%
W3_p	16%
W_p	9%



Internal Reaction	Percent of Pathways
bindL3R3	72%
SWpW2W3	65%
L3R3pS3	58%
S3pT3	58%
LRpS	46%
L2R2pS2	40%
bindLR	34%
bindL4R	34%
S2T3pWW2	33%
LRpS2	31%
T3pW3	31%
TpW	29%
formWW2W3	28%
bindL2R2	27%
bindL5R2	27%
L3R3pS	27%
formW2W3	26%
S2pW	25%
form2WW3	24%
L2R2pT	16%
S2pT2	15%
T2pW2	15%
SpT	13%
T3pW	9%

Exchange Reaction	Percent of Pathways
ATP	100%
ADP	100%
W3	91%
W2	90%
L3	72%
R3	72%
L3R3_in	72%
R	67%
LR_in	67%
W	61%
R2	55%
L2R2_in	55%
W2_p	45%
L	34%
L4	34%
WW2W3_ppp	28%
L2	27%
L5	27%
W2_W3_pp	26%
2W_W3_ppp	24%
W3_p	16%
W_p	9%

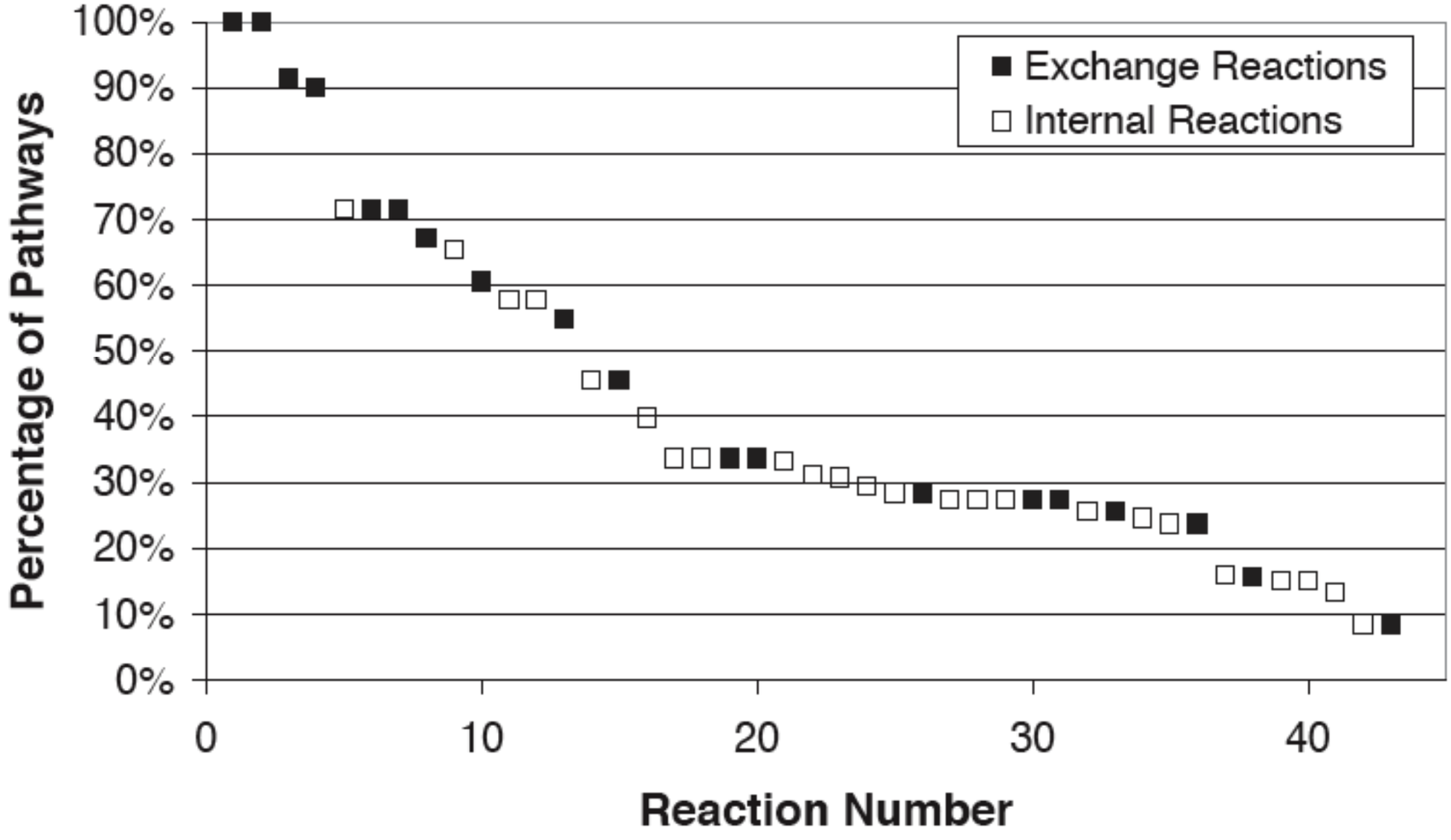


Internal Reaction	Percent of Pathways
bindL3R3	72%
SWpW2W3	65%
L3R3pS3	58%
S3pT3	58%
LRpS	46%
L2R2pS2	40%
bindLR	34%
bindL4R	34%
S2T3pWW2	33%
LRpS2	31%
T3pW3	31%
TpW	29%
formWW2W3	28%
bindL2R2	27%
bindL5R2	27%
L3R3pS	27%
formW2W3	26%
S2pW	25%
form2WW3	24%
L2R2pT	16%
S2pT2	15%
T2pW2	15%
SpT	13%
T3pW	9%

Exchange Reaction	Percent of Pathways
ATP	100%
ADP	100%
W3	91%
W2	90%
L3	72%
R3	72%
L3R3_in	72%
R	67%
LR_in	67%
W	61%
R2	55%
L2R2_in	55%
W2_p	45%
L	34%
L4	34%
WW2W3_ppp	28%
L2	27%
L5	27%
W2_W3_pp	26%
2W_W3_ppp	24%
W3_p	16%
W_p	9%

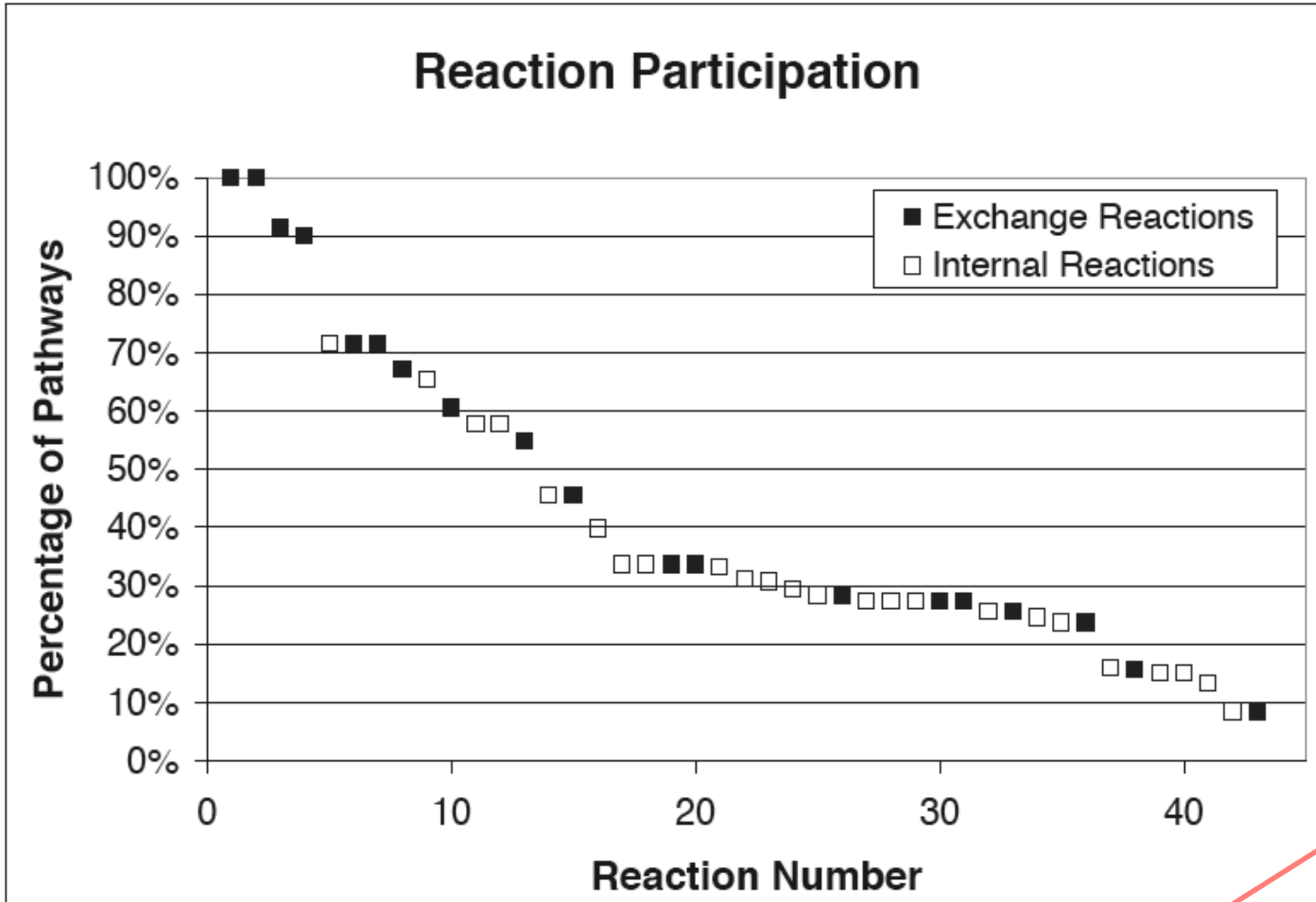
The prototypic signaling network is tightly coupled to energy metabolism

Reaction Participation



Internal Reaction	Percent of Pathways
bindL3R3	72%
SWpW2W3	65%
L3R3pS3	58%
S3pT3	58%
LRpS	46%
L2R2pS2	40%
bindLR	34%
bindL4R	34%
S2T3pWW2	33%
LRpS2	31%
T3pW3	31%
TpW	29%
formWW2W3	28%
bindL2R2	27%
bindL5R2	27%
L3R3pS	27%
formW2W3	26%
S2pW	25%
form2WW3	24%
L2R2pT	16%
S2pT2	15%
T2pW2	15%
SpT	13%
T3pW	9%

Exchange Reaction	Percent of Pathways
ATP	100%
ADP	100%
W3	91%
W2	90%
L3	72%
R3	72%
L3R3_in	72%
R	67%
LR_in	67%
W	61%
R2	55%
L2R2_in	55%
W2_p	45%
L	34%
L4	34%
WW2W3_ppp	28%
L2	27%
L5	27%
W2_W3_pp	26%
2W_W3_ppp	24%
W3_p	16%
W_p	9%



Internal Reaction	Percent of Pathways
bindL3R3	72%
SWpW2W3	65%
L3R3pS3	58%
S3pT3	58%
LRpS	46%
L2R2pS2	40%
bindLR	34%
bindL4R	34%
S2T3pWW2	33%
LRpS2	31%
T3pW3	31%
TpW	29%
formWW2W3	28%
bindL2R2	27%
bindL5R2	27%
L3R3pS	27%
formW2W3	26%
S2pW	25%
form2WW3	24%
L2R2pT	16%
S2pT2	15%
T2pW2	15%
SpT	13%
T3pW	9%

Exchange Reaction	Percent of Pathways
ATP	100%
ADP	100%
W3	91%
W2	90%
L3	72%
R3	72%
L3R3_in	72%
R	67%
LR_in	67%
W	61%
R2	55%
L2R2_in	55%
W2_p	45%
L	34%
L4	34%
WW2W3_ppp	28%
L2	27%
L5	27%
W2_W3_pp	26%
2W_W3_ppp	24%
W3_p	16%
W_p	9%

Greater degree of variability in the synthesis of the transcription factor W2_p

Correlated Reaction Sets

- From the set of extreme pathways for a network, correlated reaction sets can be calculated
- Correlated reaction sets are a collection of reactions that are either always present or always absent in all of the extreme pathways
- Effectively, these sets of reactions function together in a given network, although the reactions themselves may not be adjacent in a network map

Correlated Reaction Sets

- The correlated reactions for the prototypic signaling network were computed and are summarized in the following table

Reaction set	Reaction names
1	bindLR, L
2	bindL2R2, L2
3	S2pT2, T2pW2
4	bindL3R3, L3, R3, L3R3_in
5	L3R3pS3, S3pT3
6	bindL4R,L4
7	bindL5R2, L5
8	formW2W3, W2_W3_pp
9	form2WW2W3, WW2W3-ppp
10	form2WW3, 2W_W3_ppp
11	R, LR_in
12	R2, L2R2_ in
13	ATP, ADP

Correlated Reaction Sets

- The correlated reactions for the prototypic signaling network were computed and are summarized in the following table

Reaction set	Reaction names
1	bindLR, L
2	bindL2R2, L2
3	S2pT2, T2pW2
4	bindL3R3, L3, R3, L3R3_in
5	L3R3pS3, S3pT3
6	bindL4R,L4
7	bindL5R2, L5
8	formW2W3, W2_W3_pp
9	form2WW2W3, WW2W3-ppp
10	form2WW3, 2W_W3_ppp
11	R, LR_in
12	R2, L2R2_ in
13	ATP, ADP

Expected grouping of ATP and ADP

Correlated Reaction Sets

- The correlated reactions for the prototypic signaling network were computed and are summarized in the following table

Reaction set	Reaction names
1	bindLR, L
2	bindL2R2, L2
3	S2pT2, T2pW2
4	bindL3R3, L3, R3, L3R3_in
5	L3R3pS3, S3pT3
6	bindL4R,L4
7	bindL5R2, L5
8	formW2W3, W2_W3_pp
9	form2WW2W3, WW2W3-ppp
10	form2WW3, 2W_W3_ppp
11	R, LR_in
12	R2, L2R2_ in
13	ATP, ADP

Correlated Reaction Sets

- The correlated reactions for the prototypic signaling network were computed and are summarized in the following table

Reaction set	Reaction names
1	bindLR, L
2	bindL2R2, L2
3	S2pT2, T2pW2
4	bindL3R3, L3, R3, L3R3_in
5	L3R3pS3, S3pT3
6	bindL4R,L4
7	bindL5R2, L5
8	formW2W3, W2_W3_pp
9	form2WW2W3, WW2W3-ppp
10	form2WW3, 2W_W3_ppp
11	R, LR_in
12	R2, L2R2_ in
13	ATP, ADP

Input, receptor, reaction, and output

Correlated Reaction Sets

- The correlated reactions for the prototypic signaling network were computed and are summarized in the following table

Reaction set	Reaction names
1	bindLR, L
2	bindL2R2, L2
3	S2pT2, T2pW2
4	bindL3R3, L3, R3, L3R3_in
5	L3R3pS3, S3pT3
6	bindL4R,L4
7	bindL5R2, L5
8	formW2W3, W2_W3_pp
9	form2WW2W3, WW2W3-ppp
10	form2WW3, 2W_W3_ppp
11	R, LR_in
12	R2, L2R2_ in
13	ATP, ADP

Correlated Reaction Sets

- The correlated reactions for the prototypic signaling network were computed and are summarized in the following table

Reaction set	Reaction names
1	bindLR, L
2	bindL2R2, L2
3	S2pT2, T2pW2
4	bindL3R3, L3, R3, L3R3_in
5	L3R3pS3, S3pT3
6	bindL4R,L4
7	bindL5R2, L5
8	formW2W3, W2_W3_pp
9	form2WW2W3, WW2W3-ppp
10	form2WW3, 2W_W3_ppp
11	R, LR_in
12	R2, L2R2_ in
13	ATP, ADP

Input and reaction only
(receptors are not specific to the particular ligand)

Correlated Reaction Sets

- The correlated reactions for the prototypic signaling network were computed and are summarized in the following table

Reaction set	Reaction names
1	bindLR, L
2	bindL2R2, L2
3	S2pT2, T2pW2
4	bindL3R3, L3, R3, L3R3_in
5	L3R3pS3, S3pT3
6	bindL4R,L4
7	bindL5R2, L5
8	formW2W3, W2_W3_pp
9	form2WW2W3, WW2W3-ppp
10	form2WW3, 2W_W3_ppp
11	R, LR_in
12	R2, L2R2_ in
13	ATP, ADP

Correlated Reaction Sets

- The correlated reactions for the prototypic signaling network were computed and are summarized in the following table

Reaction set	Reaction names
1	bindLR, L
2	bindL2R2, L2
3	S2pT2, T2pW2
4	bindL3R3, L3, R3, L3R3_in
5	L3R3pS3, S3pT3
6	bindL4R,L4
7	bindL5R2, L5
8	formW2W3, W2_W3_pp
9	form2WW2W3, WW2W3-ppp
10	form2WW3, 2W_W3_ppp
11	R, LR_in
12	R2, L2R2_ in
13	ATP, ADP

The formation of the TF complexes

Correlated Reaction Sets

- The correlated reactions for the prototypic signaling network were computed and are summarized in the following table

Reaction set	Reaction names
1	bindLR, L
2	bindL2R2, L2
3	S2pT2, T2pW2
4	bindL3R3, L3, R3, L3R3_in
5	L3R3pS3, S3pT3
6	bindL4R,L4
7	bindL5R2, L5
8	formW2W3, W2_W3_pp
9	form2WW2W3, WW2W3-ppp
10	form2WW3, 2W_W3_ppp
11	R, LR_in
12	R2, L2R2_ in
13	ATP, ADP

Correlated Reaction Sets

- The correlated reactions for the prototypic signaling network were computed and are summarized in the following table

Reaction set	Reaction names
1	bindLR, L
2	bindL2R2, L2
3	S2pT2, T2pW2
4	bindL3R3, L3, R3, L3R3_in
5	L3R3pS3, S3pT3
6	bindL4R,L4
7	bindL5R2, L5
8	formW2W3, W2_W3_pp
9	form2WW2W3, WW2W3-ppp
10	form2WW3, 2W_W3_ppp
11	R, LR_in
12	R2, L2R2_ in
13	ATP, ADP

Non-obvious correlated reaction sets

Acknowledgments

- The material is mostly based on
  - Bernhard Palsson, "Systems Biology: Properties of Reconstructed Networks." Cambridge University Press, 2006.
  - Papin and Palsson, "Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk." Journal of Theoretical Biology, 227: 283-297, 2004.