Stoichiometric Constraints-Based Modeling of Large-Scale Complex Biochemical Networks

Will Heuett

University of Washington

Department of Applied Mathematics

March 11th and 13th, 2003

Flux Balance Analysis

$$\xrightarrow{J^e} \nu_1^j X_1 + \nu_2^j X_2 + \cdots + \nu_N^j X_N \underset{k_-^j}{\overset{k_+^j}{\rightleftharpoons}} \kappa_1^j X_1 + \kappa_2^j X_2 + \cdots + \kappa_N^j X_N \xrightarrow{J^e}$$

$$\Downarrow$$

$$\frac{dx_i}{dt} = \sum_{j=1}^{M} S_i^j J^j \quad \text{or} \quad \frac{dx_i}{dt} = \sum_{j=1}^{M} S_i^j J^j + J_i^e$$

$$\Downarrow$$

$$\mathbf{SJ} = \mathbf{0} \quad \text{or} \quad \mathbf{SJ} = -\mathbf{J^e}$$

Flexibility of the Model

It is possible to incorporate as much, or as little, as is known about the individual reactions.

$$2A + B + C \rightleftharpoons 2Y + Z$$

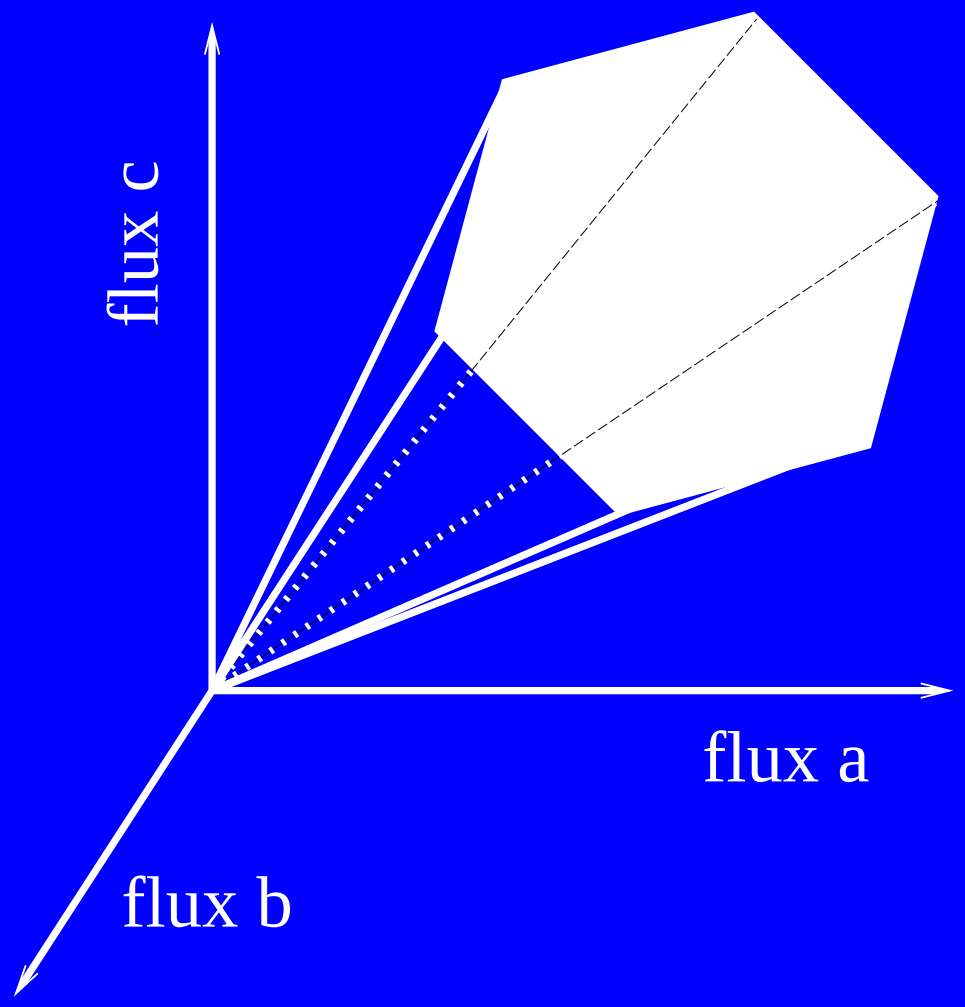
$$\Downarrow$$

$$A + A \rightleftharpoons (2A)$$

$$(2A) + B \rightleftharpoons (2AB)$$

$$(2AB) + C \rightleftharpoons 2Y + Z$$

Feasible J



Easy Example

$$A + 2B \rightleftharpoons C$$
$$C + D \rightleftharpoons 2A + B$$
$$A + B \rightleftharpoons 2D$$
$$A + C \rightleftharpoons B + 2D$$
$$B \rightleftharpoons D$$

$$\Rightarrow \quad \mathbf{S} = \begin{pmatrix} -1 & 2 & -1 & -1 & 0 \\ -2 & 1 & -1 & 1 & -1 \\ 1 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & 2 & 1 \end{pmatrix}$$

$$\frac{d}{dt}\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} -1 & 2 & -1 & -1 & 0 \\ -2 & 1 & -1 & 1 & -1 \\ 1 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} J^1 \\ J^2 \\ J^3 \\ J^4 \\ J^5 \end{pmatrix}$$

Add External Fluxes

Add $\longrightarrow A$

Remove $B \longrightarrow$

$$\mathbf{S} = \begin{pmatrix} -1 & 2 & -1 & -1 & 0 & 1 & 0 \\ -2 & 1 & -1 & 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 2 & 1 & 0 & -1 \end{pmatrix}$$

$$\frac{d}{dt}\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} -1 & 2 & -1 & -1 & 0 & 1 & 0 \\ -2 & 1 & -1 & 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 2 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} J^1 \\ J^2 \\ J^3 \\ J^4 \\ J^5 \\ J^I \\ J^O \end{pmatrix}$$

Optimization using FBA

Biological networks are assumed to optimize a certain biologically meaningful objective function.

$$\max_{\mathbf{J}} \quad \mathbf{c}^{\mathbf{T}}\mathbf{J}$$

$$\text{s.t.} \quad \mathbf{SJ} = \mathbf{0} \quad \alpha^{\mathbf{j}} \leq \mathbf{J}^{\mathbf{j}} \leq \beta^{\mathbf{j}}$$

Sample Problem With D Output

$$\max_{\mathbf{J}} \quad \mathbf{J}^{\mathbf{O}}$$

$$\text{s.t.} \quad \mathbf{SJ} = \mathbf{0}$$

$$J^I = J^O = 1$$

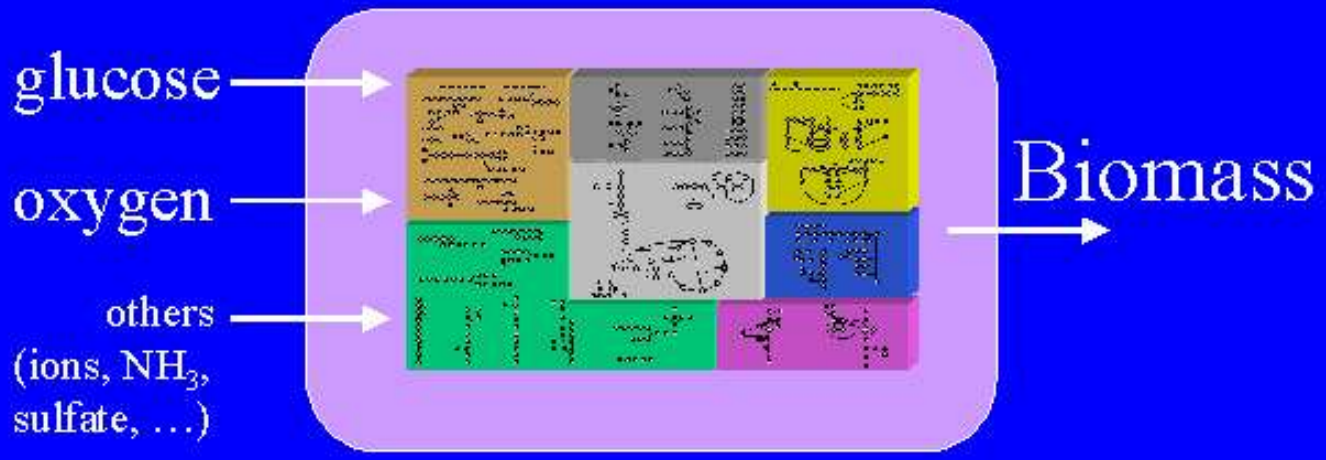
A input $\Rightarrow$ $\mathbf{J} = \begin{pmatrix} 0 \\ -0.3 \\ 0.1 \\ 0.3 \\ -0.1 \end{pmatrix}$

B input $\Rightarrow$ $\mathbf{J} = \begin{pmatrix} -1 \\ -0.5 \\ 0.5 \\ -0.5 \\ 0.5 \end{pmatrix}$

E. Coli Growth Objective Function

Edwards and Palsson *PNAS. USA* **97**: 5528-5533, (2000)

- Precursors to cell growth.
  - Production of Biomass (growth is maximized).
- Components: amino acids, cell wall, DNA, RNA, energy storage cofactors, polyamines.
- Stoichiometry of growth factors determined experimentally.



In Silico Predictions of *E. Coli Growth*

Edwards and Palsson *PNAS. USA* **97**: 5528-5533, (2000)

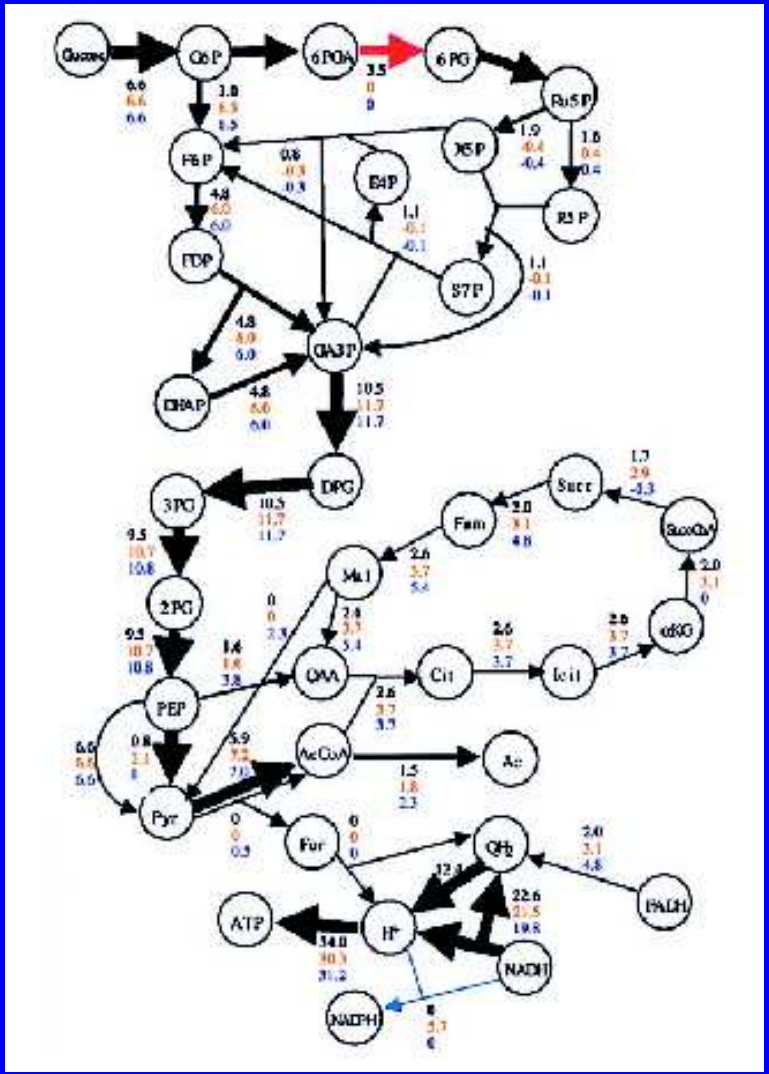
7 -/+ of 79 ***In silico*/Experimental**

Gene	Glucose	Glycerol	Succinate	Acetate	Gene	Glucose	Glycerol	Succinate	Acetate
aceEF	-/+				*pgl*	+/+			
aceA				-/-	*prtAB*	+/+	+/+	+/+	+/+
aceB				-/-	*glk*	+/+			
ackA				+/+	*ppc*	+/+	-/+	+/+	+/+
acs				+/+	*pta*				+/+
acn	-/-	-/-	-/-	-/-	*pts*	+/+			
cyd	+/+				*pyk*	+/+			
cyo	+/+				*rpi*	-/-	-/-	-/-	-/-
eno	-/+	-/+	-/-	-/-	*sdhABCD*	+/+			
fba	-/+				*tpi*	-/+	-/-	-/-	-/-
fbp	+/+	-/-	-/-	-/-	*unc*	+/+		+/+	-/-
gap	-/-	-/-	-/-	-/-	*zwf*	+/+			
gltA	-/-	-/-	-/-	-/-	*sucAD*	+/+			
gnd	+/+				*zwf, prt*	+/+			
idh	-/-	-/-	-/-	-/-	*pck, mez*			-/-	-/-
ndh	+/+	+/+			*pck, pps*			-/-	-/-
nuo	+/+	+/+			*pgi, zwf*	-/-			
pfk	-/+				*pgi, gnd*	-/-			
pgi	+/+	+/+			*pta,acs*				-/-
pgk	-/-	-/-	-/-	-/-	*tktA, tktB*	-/-			

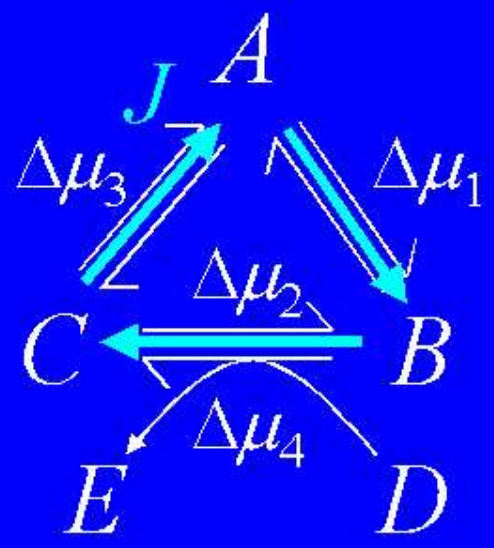
-/+ Essential gene not predicted by model.

In Silico Behavior of *E. Coli Growth*

Edwards and Palsson *PNAS. USA* **97**: 5528-5533, (2000)



Energy Balance for Simple Loops



$$\Delta\mu_1 + \Delta\mu_2 + \Delta\mu_3 = \Delta\mu_4$$

$$J\Delta\mu_1 + J\Delta\mu_2 + J\Delta\mu_3 = J\Delta\mu_4$$

Rate of heat dissipation = Rate of energy pumped in

Energy Balance Analysis

Using SVD: $\mathbf{SB} = \mathbf{D}\sigma$

$$\Downarrow$$

$$\mathbf{SK} = \mathbf{S}[\mathbf{B}(:,\mathbf{r}+\mathbf{1})\ldots\mathbf{B}(:,\mathbf{M})] = \mathbf{0}$$

$$\Downarrow$$

$$G^T SK = \Delta G^T K = 0$$

Easy Example

$$A + 2B \rightleftharpoons C$$
$$C + D \rightleftharpoons 2A + B$$
$$A + B \rightleftharpoons 2D$$
$$A + C \rightleftharpoons B + 2D$$
$$B \rightleftharpoons D$$

$$\Rightarrow \quad \mathbf{S} = \begin{pmatrix} -1 & 2 & -1 & -1 & 0 \\ -2 & 1 & -1 & 1 & -1 \\ 1 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & 2 & 1 \end{pmatrix}$$

$$\mathbf{SB} = \mathbf{U}\sigma \quad \Rightarrow \quad \mathbf{K} = \begin{pmatrix} 0 \\ -0.2236 \\ -0.6708 \\ 0.2236 \\ 0.6708 \end{pmatrix} \quad \Rightarrow \quad \mathbf{SK} = \mathbf{0}$$

K Balances Internal Reactions

$$A + C \rightarrow B + 2D$$

$$B \rightarrow D$$

$$C + D \leftarrow 2A + B$$

$$A + B \leftarrow 2D$$

$$0.2236A + 0.6708B + 0.2236C \rightarrow .2236B + 1.118D$$

$$0.6708A + 0.6708B + 0.2236C + 0.2236D \leftarrow 0.4472A + 0.2236B + 1.3416D$$

$$0.6708A + 0.8944B + 0.2236C + 1.3416D \leftrightharpoons 0.6708A + 0.8944B + 0.2236C + 1.3416D$$

Second Law of Thermodynamics

$$-J^j \Delta G^j = -k_B T (J_+^j - J_-^j) \ln \frac{J_-^j}{J_+^j} \geq 0$$

Each reaction must dissipate energy (or heat). J^j is the turnover per unit time and ΔG^j is the chemical potential change per turnover. By summing over all reactions, this formula can be used to compute the total heat dissipation rate, which is necessarily equal to the entropy production rate of the isothermal biochemical network in a steady-state.

Duality

A sufficient and necessary condition that the system:

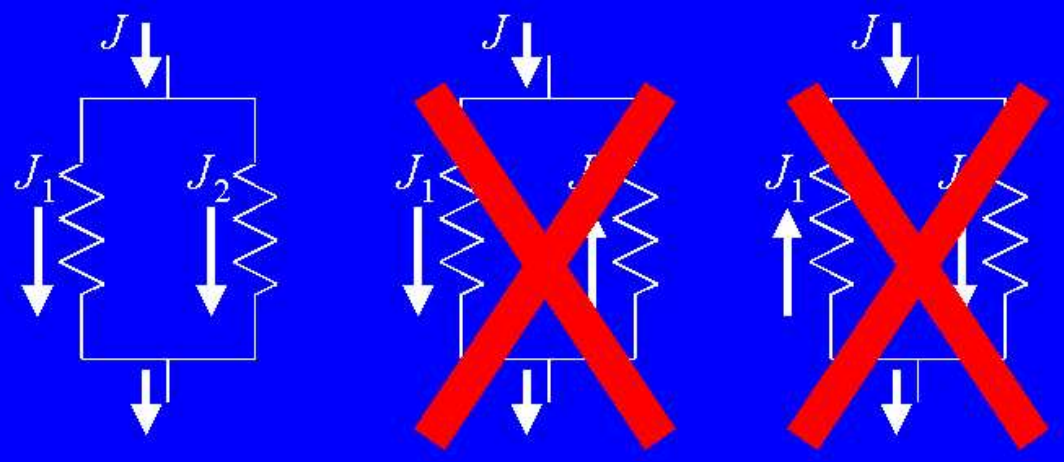
$$\sum_{i=1}^{N} G_i S_i^j J^j \leq 0$$

admits a nontrivial solution (i.e. at least one $G_i \neq 0$) is that it's dual problem:

$$\sum_{j=1}^{M} S_i^j J^j y^j \leq 0 \quad y^j \geq 0$$

admits only a trivial solution.

Ruling Out Unrealistic Scenarios



By only using FBA constraints to analyze complex networks, one allows some nonphysical solutions to be included in the feasible set of solutions. EBA and thermodynamic constraints need to be used in combination with FBA to rule out these unrealistic scenarios.

E. Coli Growth with EBA and Thermodynamics

Beard, Liang, and Qian *Biophysical Journal*, **83**, 79-86 (2002)

Enzyme	Gene(s)	Wild Type flux	Wild Type Δ%	Wild Type c	*pyk* knockout flux	*pyk* knockout Δ%	*pyk* knockout c	*pgi* knockout flux	*pgi* knockout Δ%	*pgi* knockout c	*eno* knockout flux	*eno* knockout Δ%	*eno* knockout c	*pfk* knockout flux	*pfk* knockout Δ%	*pfk* knockout c
Glucose phosphotransferase	*ptsG*	10	-4	2.5	10	-4	2.5	10	-4	2.5	0	-5.6449	0	1.3418	-0.5367	2.5
Glucokinase	*glk*	0	-7.0082	0	0	-4.3242	0	0	-4.1363	0	10	-4	2.5	5.1759	-2.8512	1.8154
PGI (G-6-P / F-6-P)	*pgi*	2.3309	-1.2083	1.9291	1.2058	-1.8777	0.6422	0	-7.3982	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PGI (bDG6-P / G-6-P)	*pgi*	-1.5104	0.6041	2.5	-2.3471	0.9388	2.5	0	3.6991	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PGI (bDG6-P / F-6-P)	*pgi*	1.5104	-0.6041	2.5	2.3471	-0.9388	2.5	0	-3.6991	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Glucose-1-phosphatase	*agp*	0	-1.2708	0	0	-0.951	0	0	-0.9552	0	0	-6.6044	0	0.0006	-1.8055	0.0004
Phosphofructokinase	*pfkA,B*	7.1478	-2.8591	2.5	7.0635	-2.8254	2.5	5.8813	-4.8729	1.2069	5.7591	-13.364	0.4309	0	-6.4622	0
Fructose 1,6-bisphosphatase	*fbp*	0	-6.6906	0	0	-3.4008	0	0	-1.1738	0	0	-3.8448	0	0	0	Eq.
Fructose 1,6-bp aldolase	*fba*	7.1478	-2.8591	2.5	7.0635	-2.8254	2.5	5.8813	-3.6991	1.5899	5.7591	-9.5194	0.605	0	-19.267	0
Triosphosphate isomerase	*tpia*	-7.0564	2.8226	2.5	-6.9735	2.7894	2.5	-5.7915	3.6991	1.5656	-5.6936	9.5194	0.5981	[illegible]	[illegible]	[illegible]
GAP-dehydrogenase	*gapA,C1,C2*	15.774	-6.3097	2.5	15.71	-6.2841	2.5	14.532	-5.8127	2.5	14.776	-5.9103	2.5	5.6279	-19.267	0.2921
Phosphoglycerate kinase	*pgk*	15.774	-6.3097	2.5	15.71	-6.2841	2.5	14.532	-5.8127	2.5	14.776	-5.9103	2.5	5.6279	-19.267	0.2921
Phosphoglycerate mutase	*gpmAB*	14.322	-5.7286	2.5	14.279	-5.7116	2.5	13.191	-5.2766	2.5	0	0	Eq.	5.3105	-9.3348	0.5689
Enolase	*eno*	14.322	-5.7286	2.5	14.279	-5.7116	2.5	13.191	-5.2766	2.5	0	9.9992	0	5.3105	-9.3348	0.5689
Phoshphoenolpyr. synthetase	*ppsA*	0	-4.3797	0	0	-2.828	0	0	-2.0828	0	2.1528	-0.8611	2.5	0.0008	-0.0003	2.5
Pyruvate kinase	*pykA,F*	1.3145	-0.8041	1.6348	0	-2.5288	0	0.2368	-2.7292	0.0868	0	-14.864	0	0.0023	-1.7574	0.0013
Pyruvate dehydrogenase	*lpdA, aceEF*	9.2368	-3.6947	2.5	9.2577	-3.7031	2.5	5.0462	-2.0185	2.5	0	0	Eq.	0	-11.287	0

Up Regulated
Down Regulated
Reverse Direction
Knock Out

$$c^j = \frac{-J^j}{\Delta G^j} = \text{measure of the activation level of the pathway}$$

Conclusions

- By using a Stoichiometric Constraints-Based Model, we are able to avoid the nasty details of the complex reaction kinetics. By doing so, we create models which depend only on invariant properties of the system.
- The complexity of the models can easily be controlled. We can add as much, or as little, detail as we would like, depending on what we are interested in studying.
- It is important to combine the Flux Balance, Energy Balance, and thermodynamic constraints in order to get physically realistic results. Such results have proven to be quite accurate in predicting the behavior of a network under various conditions.