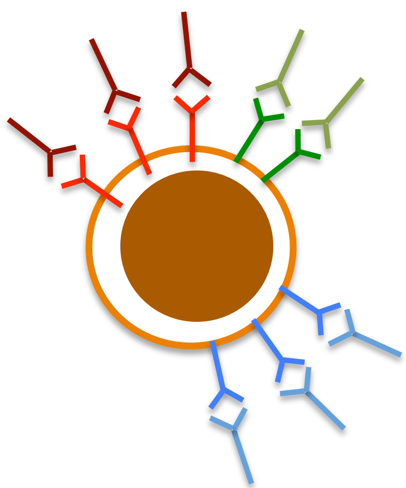
Dramatic Reduction of Dimensionality in Large Biochemical Networks Due to Strong Pair Correlations

Jayajit Das

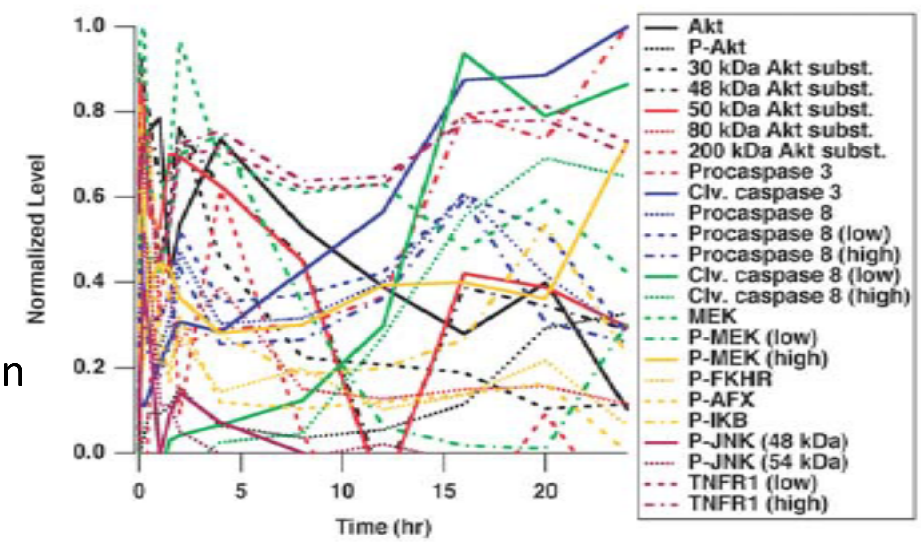
Battelle Center for Mathematical Medicine
The Research Institute at Nationwide Children's
Hospital and Ohio State University

Workshop on Systems Biology and Formal Methods, NYU, March 30, 2012

High-throughput Methods Reveal Cellular Complexity

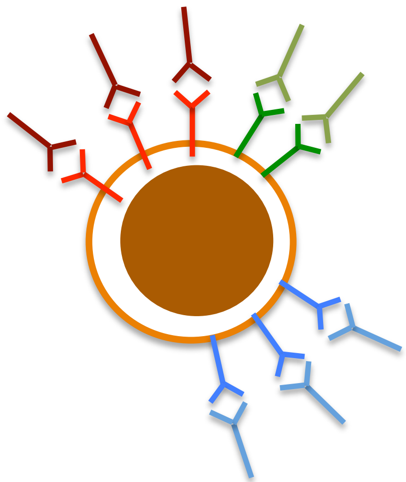


diverse responses
- cytokine secretion
- proliferation
- apoptosis

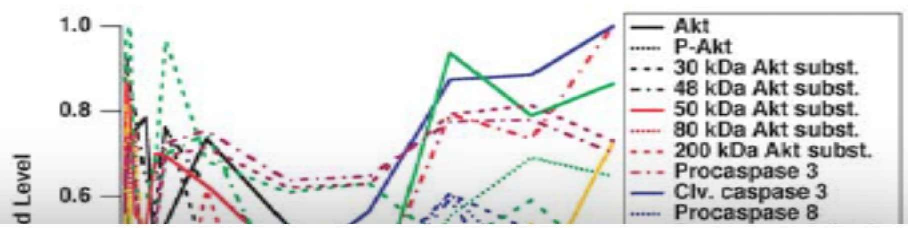


Janes et al. J. Comp. Biol. (2004)

High-throughput Methods Reveal Cellular Complexity



diverse response
- cytokine secr
- proliferation
- apoptosis



Difficulties in generating predictive models

Large number of variables

Unknown interactions

Difficult to generate mechanistic models

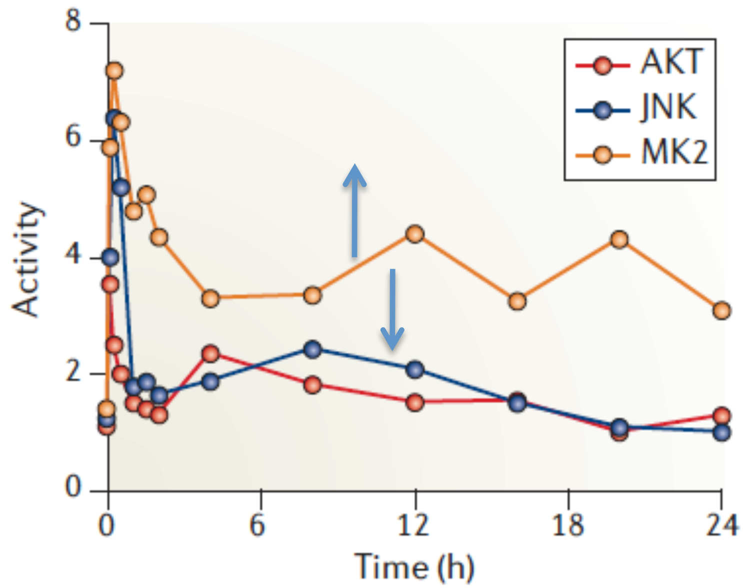
few variables
well defined interactions

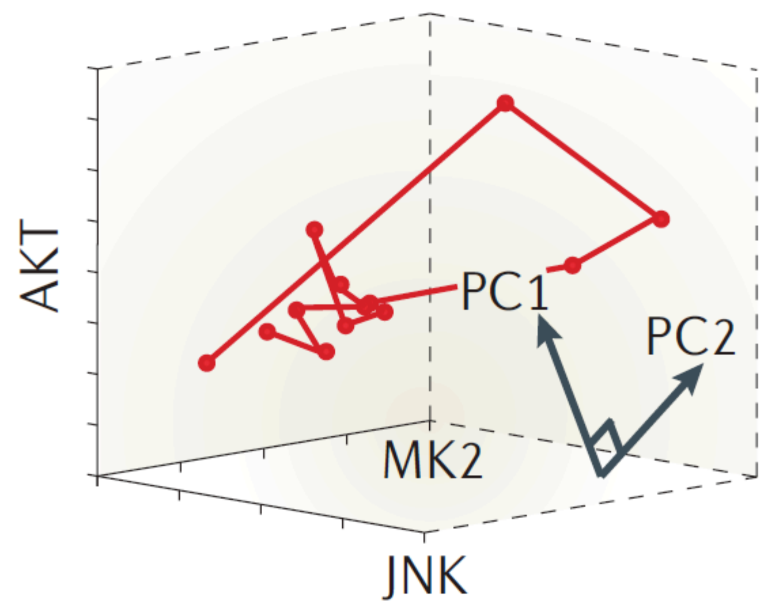
Success of Multivariate Statistical Methods

use pair-correlations

effective variables (principal components)

≡ Linear sum of variables in the high-dimensional dataset





Large reduction of dimensionality (hundreds to few ~5)

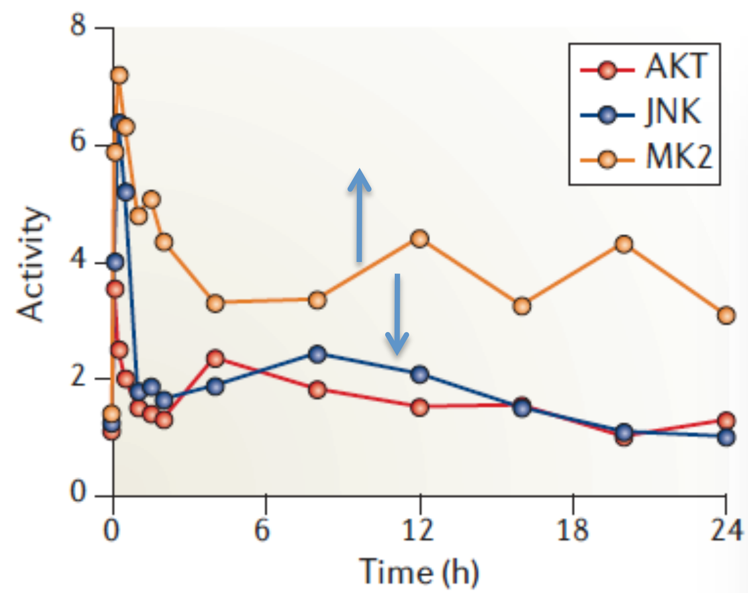
Janes and Yaffe, Nat. Rev. MCB (2006)

Success of Multivariate Statistical Methods

effective variables (principal components)

use pair-correlations

≡ linear sum of variables in
the high-dimensional dataset



Large reduction of dir

Issues not understood

Dramatic reduction in dimensionality
- Accidental or Generic?

Can this reduction be used to extract mechanisms and construct coarse grained variables for mechanistic models?

Pair-Correlations in Chemical Reactions: A Simple Example

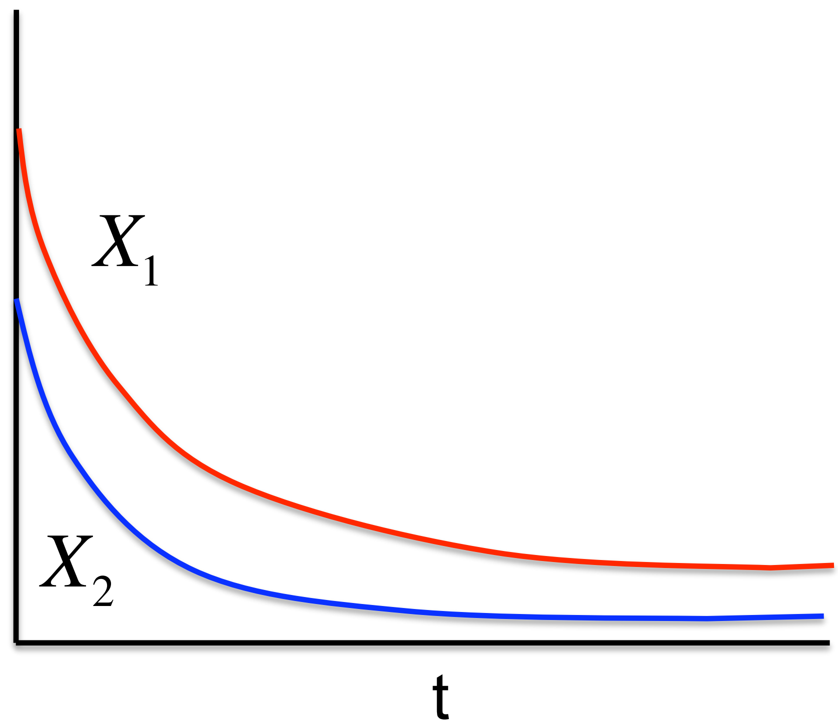
$$X_1 \underset{k_{1r}}{\overset{k_{1f}}{\rightleftharpoons}} X_2 \underset{k_{2r}}{\overset{k_{2f}}{\rightleftharpoons}} X_3$$

deterministic mass-action kinetics

$$\frac{dc_1}{dt} = -k_{1f}c_1 + k_{1r}c_2$$

$$\frac{dc_2}{dt} = -(k_{2f} + k_{1r})c_2 + k_{1f}c_1 + k_{2r}c_2$$

$$c_1 + c_2 + c_3 = c_0$$



Pair-Correlations in Chemical Reactions: A Simple Example

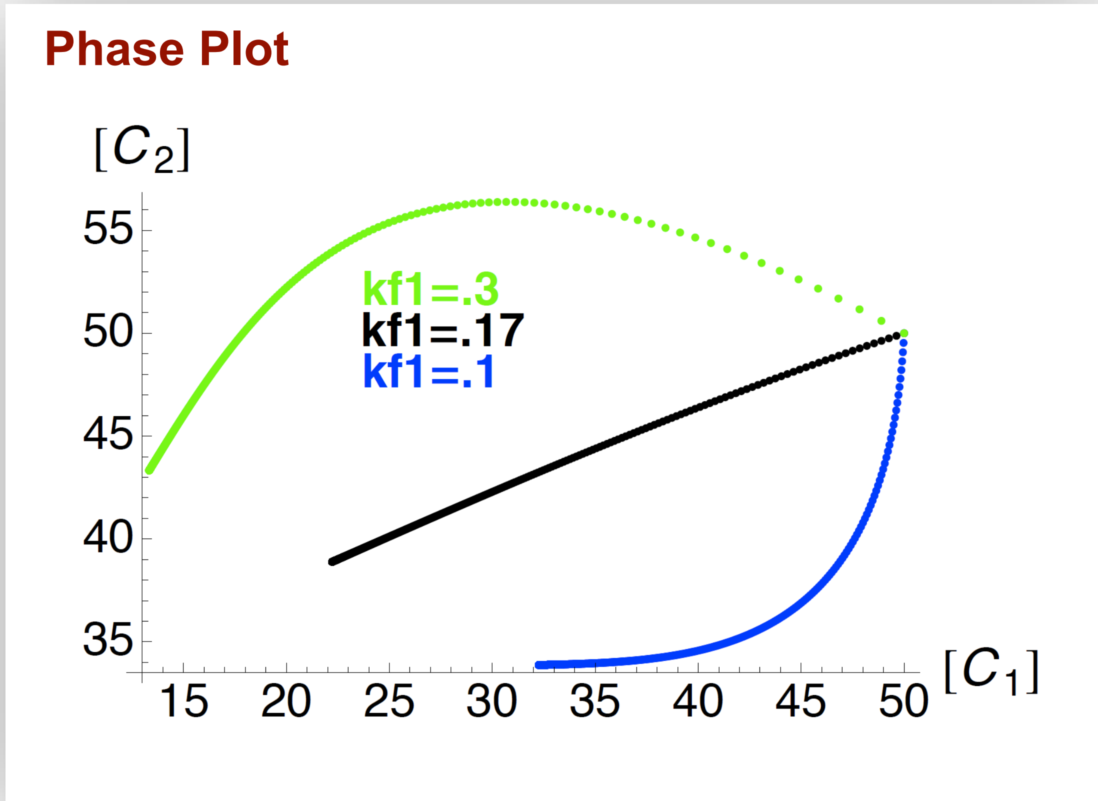
$$X_1 \underset{k_{1r}}{\overset{k_{1f}}{\rightleftharpoons}} X_2 \underset{k_{2r}}{\overset{k_{2f}}{\rightleftharpoons}} X_3$$

deterministic mass-action kinetics

$$\frac{dc_1}{dt} = -k_{1f}c_1 + k_{1r}c_2$$

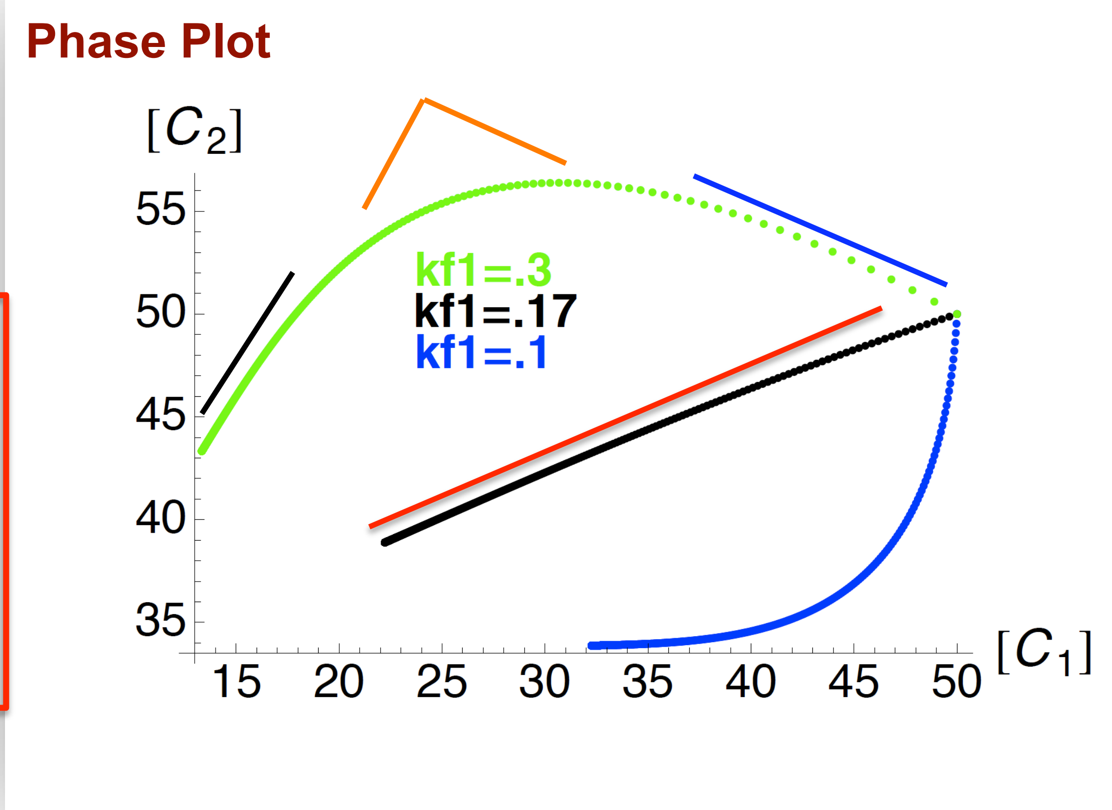
$$\frac{dc_2}{dt} = -(k_{2f} + k_{1r})c_2 + k_{1f}c_1 + k_{2r}c_2$$

$$c_1 + c_2 + c_3 = c_0$$



Pair-Correlations in Chemical Reactions: A Simple Example

$$X_1 \underset{k_{1r}}{\overset{k_{1f}}{\rightleftharpoons}} X_2 \underset{k_{2r}}{\overset{k_{2f}}{\rightleftharpoons}} X_3$$



Covariance

$$C_{ij}(t_n) = n^{-1}\sum_{\alpha=1}^{n}\left(c_i(t_\alpha)-\bar{c}_i\right)\left(c_j(t_\alpha)-\bar{c}_j\right)$$

$$\bar{c}_i = n^{-1}\sum_{\alpha=1}^{n} c_i(t_\alpha) \qquad i=1,2$$

Eigenvalues

$$(\mathbf{C}-\lambda_a \mathrm{I})\hat{\mathbf{p}}_a = 0$$

percent explained = $\lambda_a \,/\, tr(\mathbf{C}) \times 100\%$ →

~ 90% variance explained (1 PC is sufficient)

~ 50% variance explained (needs 2 PCs)

contains information about the variation of the phase trajectory

Pair-Correlations in Chemical Reactions: A Simple Example

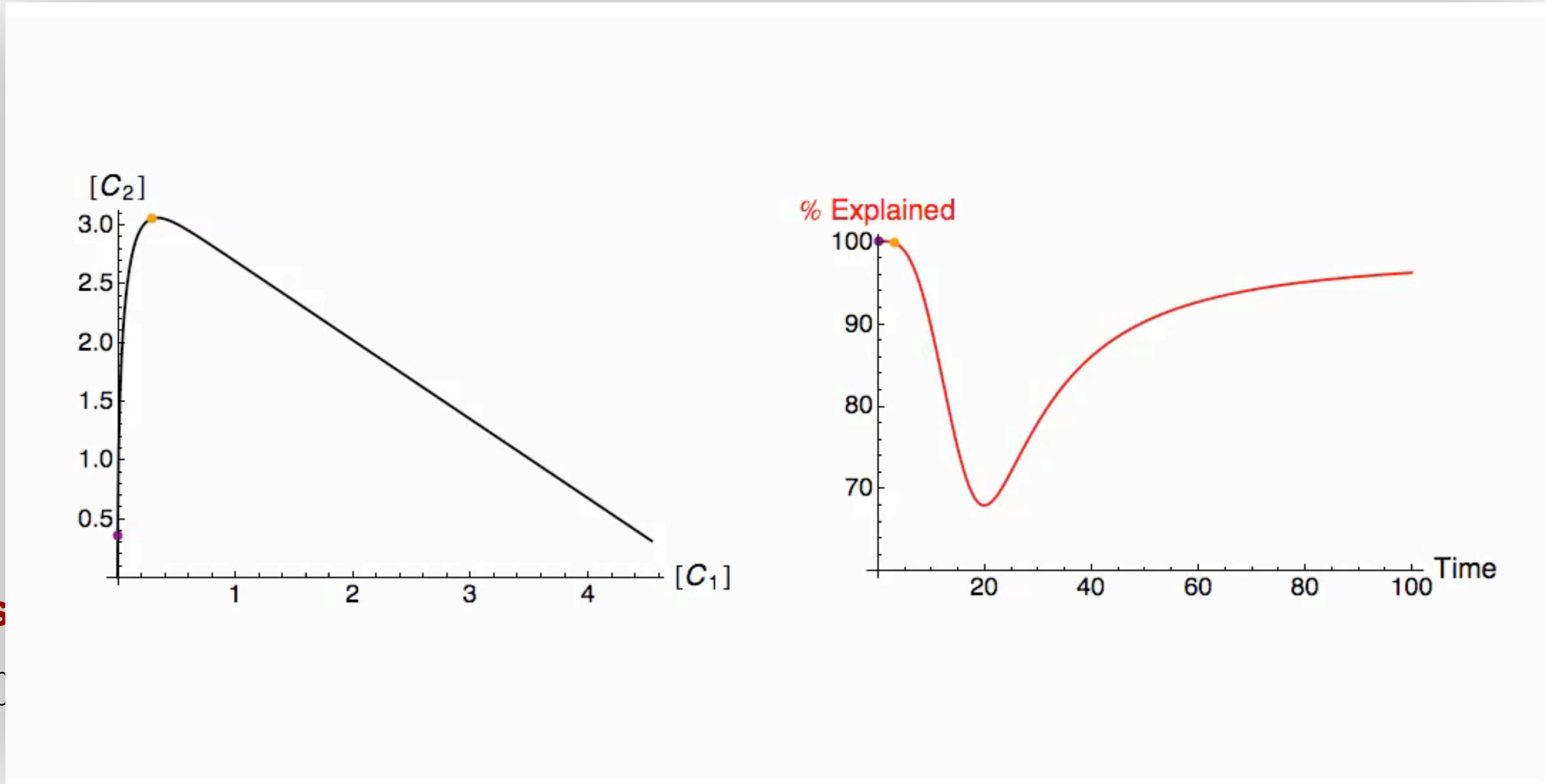
$$X_1 \underset{k_{1r}}{\overset{k_{1f}}{\rightleftharpoons}} X_2 \underset{k_{2r}}{\overset{k_{2f}}{\rightleftharpoons}} X_3$$

Covariance

$$\bar{c}_i = n^{-1}\sum_{\alpha=1}^{n} c_i$$

Eigenvalues

$$(\mathbf{C} - \lambda_a \mathbf{I})\hat{\mathbf{p}}_a = ($$



$\boxed{\text{percent explained} = \lambda_a \,/\, tr(\mathbf{C}) \times 100\%}$ →

~ 90% variance explained (1 PC is sufficient)

~ 50% variance explained (needs 2 PCs)

Rule Based Modeling

Linear Networks

-increase number of species

-change network topology

-introduce non-linear mass-action kinetics

-vary kinetic rates and initial concentrations over 100 times

Biological Networks

-vary kinetic rates and initial concentrations over 100 times

Can be systematically approached via Rule-Based Modeling

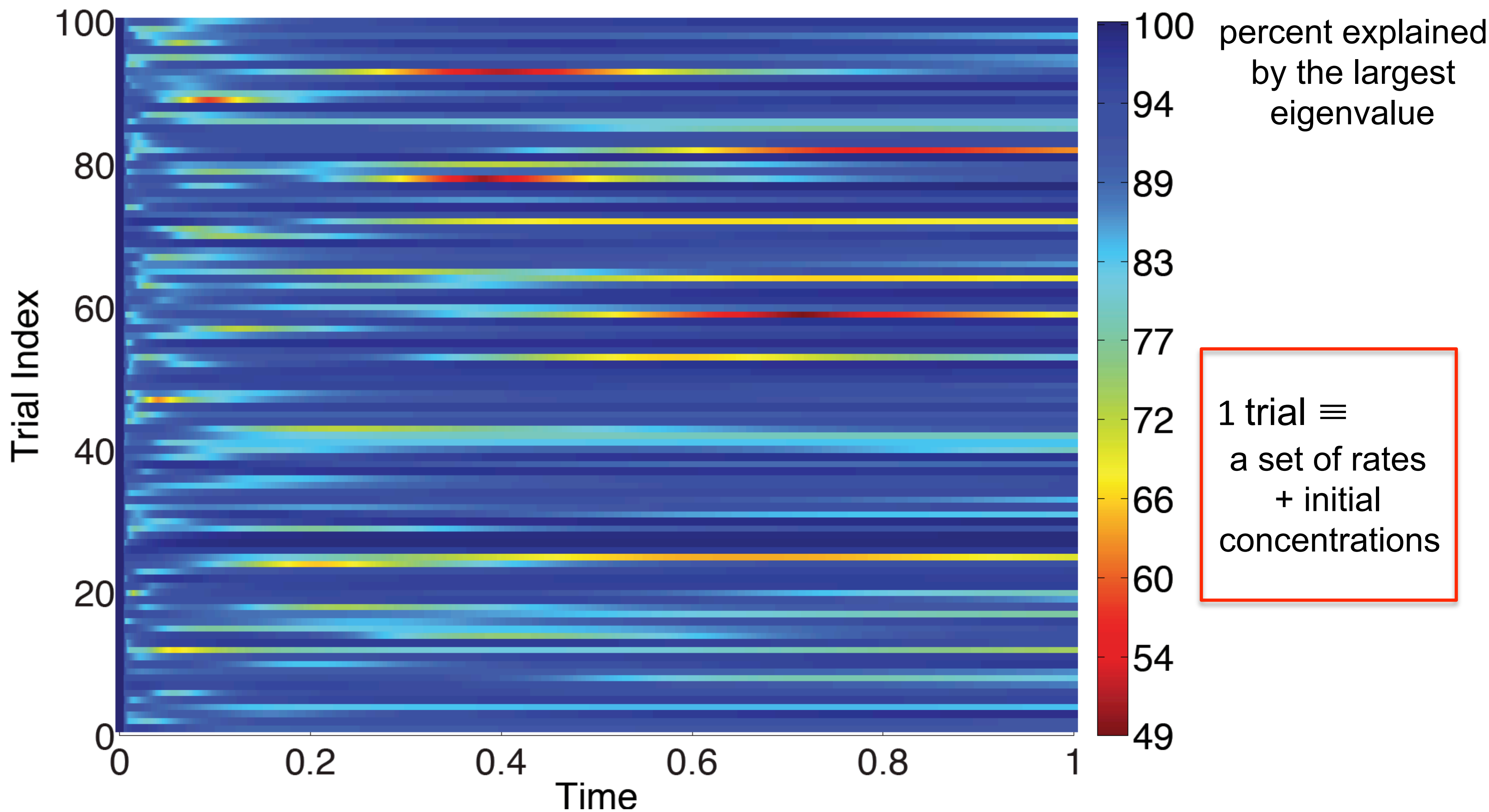
Hlavacek et al. Sci. Sig. (2006)

BioNetGen (bionetgen.org)

Linear Network with Linear Kinetics

N=64

$$X_1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} X_2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} X_3 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} X_4 \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} X_5 \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} X_6 \underset{k_{-6}}{\overset{k_6}{\rightleftharpoons}} \cdots$$

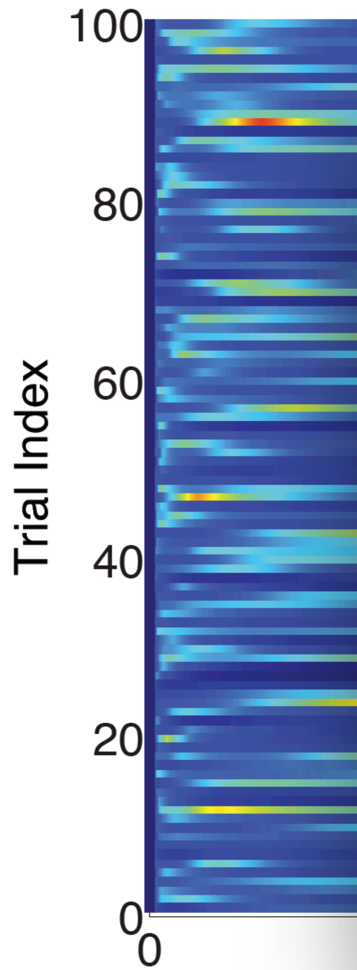


1 trial ≡ a set of rates + initial concentrations

percent explained decreases in short time intervals

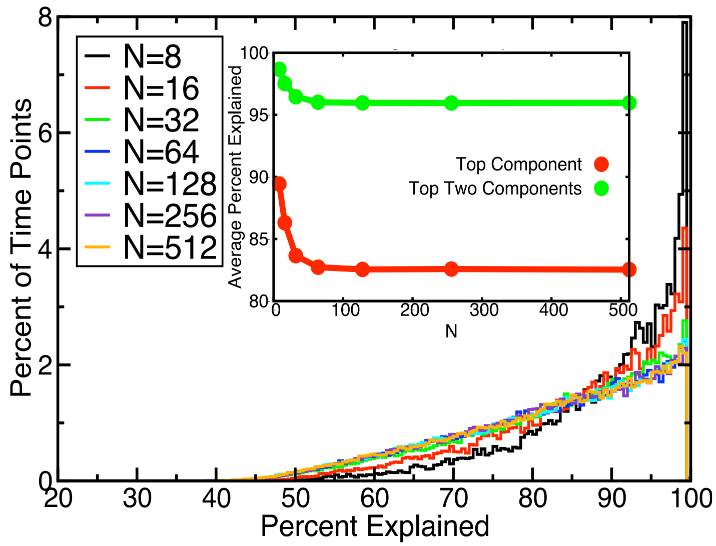
Linear Network with Linear Kinetics

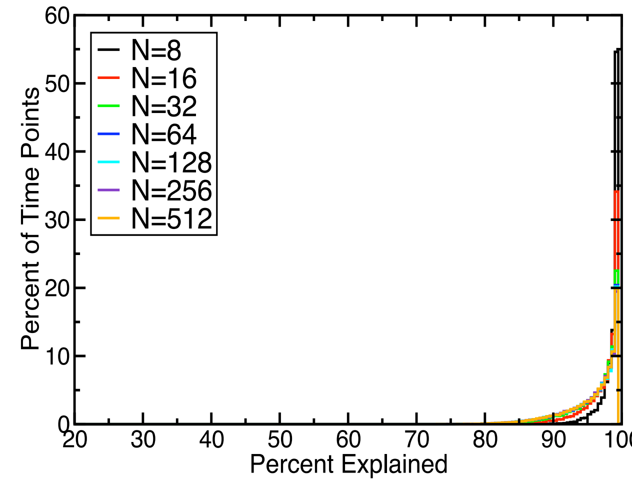
N=64 $X_1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} X_2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} X_3 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} X_4 \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} X_5 \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} X_6 \underset{k_{-6}}{\overset{k_6}{\rightleftharpoons}} \cdots$



Dependence on number of species

10,000 trials





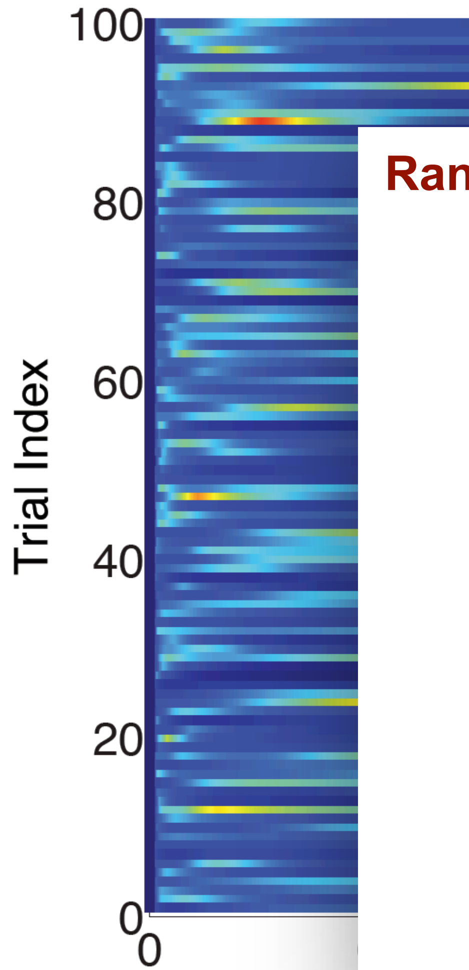
>90% variance captured by 2 components for all N's

reduction insensitive to number of species

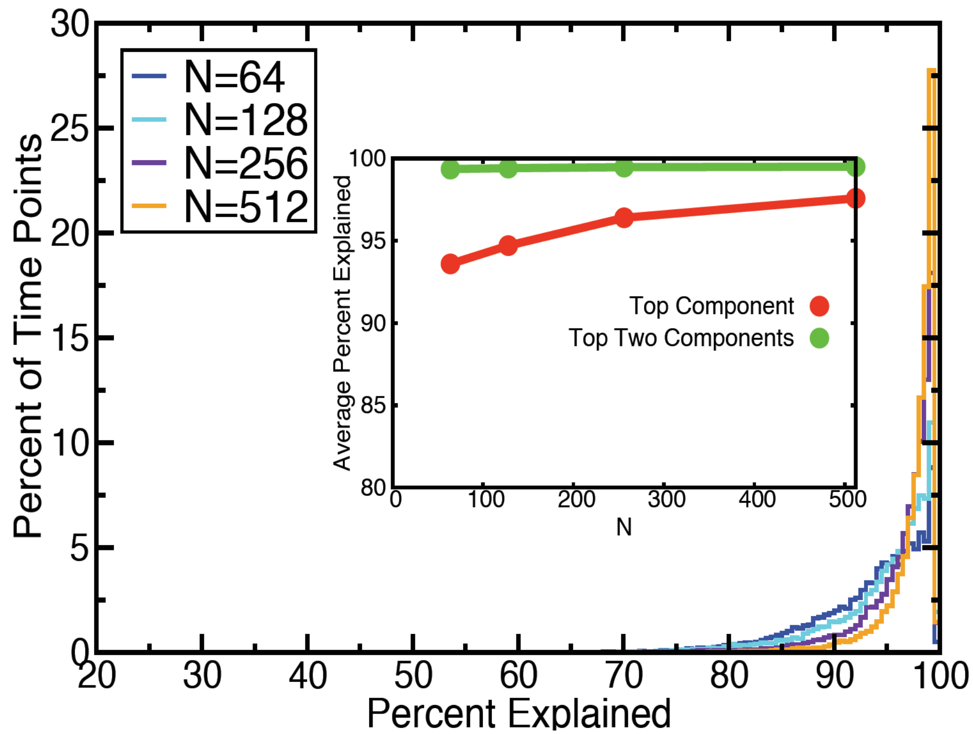
percent explained decreases in short time intervals

Linear Network with Linear Kinetics

N=64 $X_1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} X_2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} X_3 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} X_4 \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} X_5 \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} X_6 \underset{k_{-6}}{\overset{k_6}{\rightleftharpoons}} \cdots$



Random network with linear kinetics

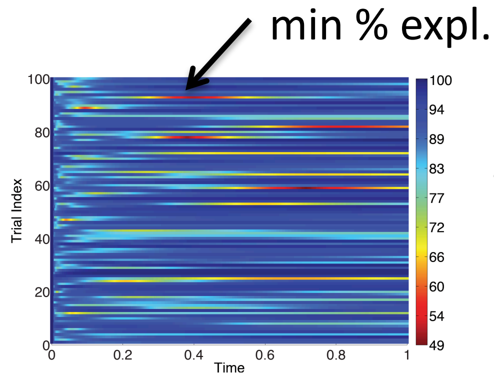


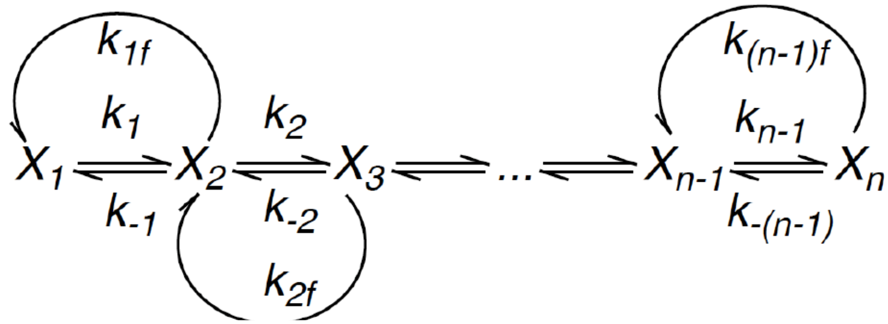
>95% variance captured by 2 components for all N's

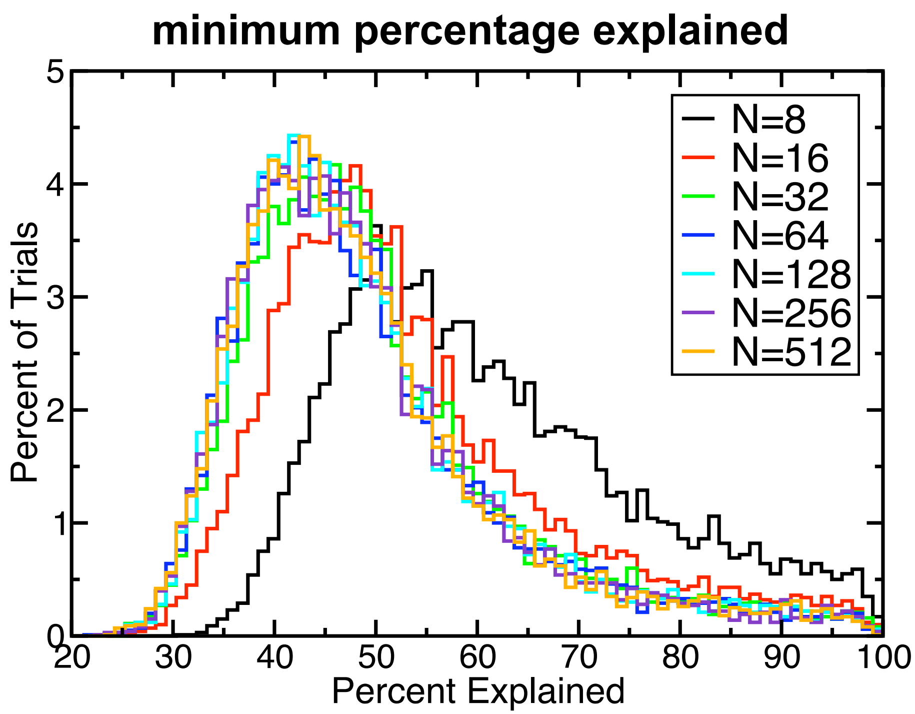
reduction insensitive to network architecture

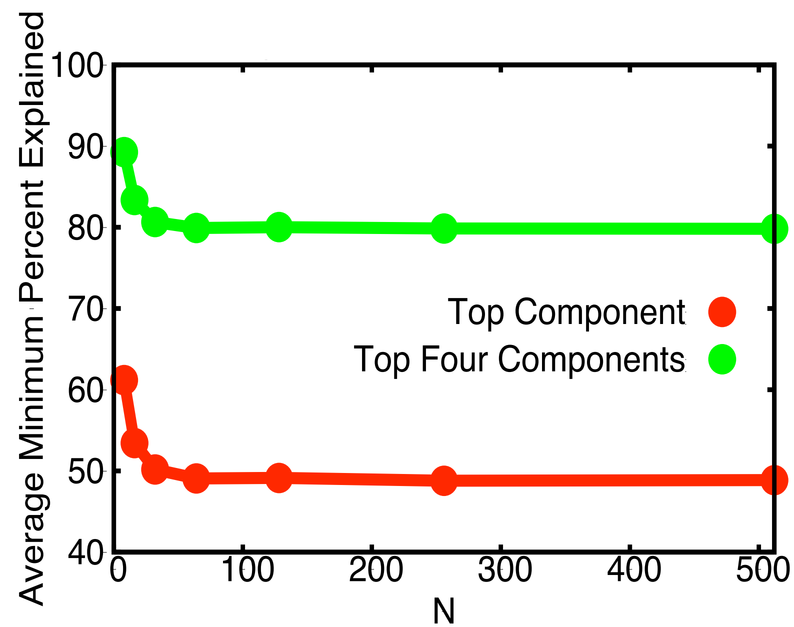
percent explained decreases in short time intervals

Branched Linear Network with Non-Linear Kinetics



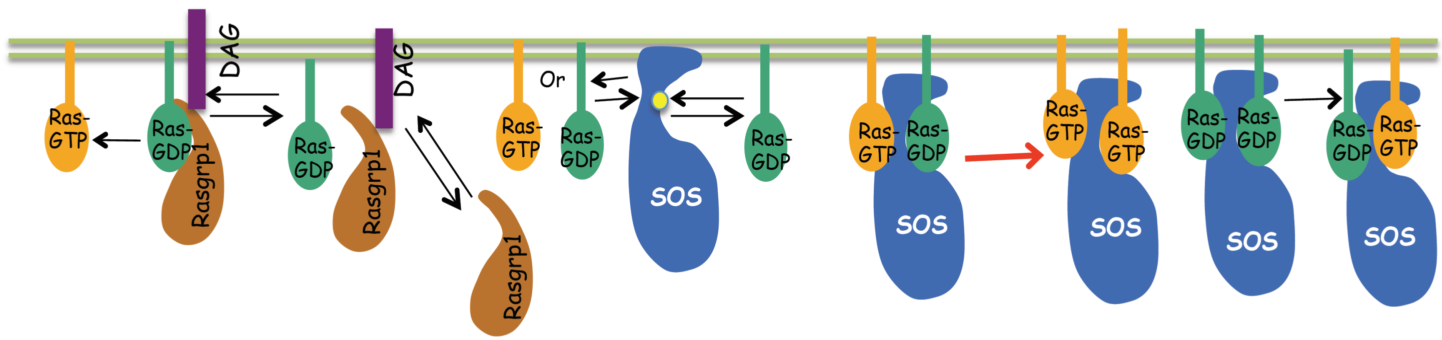




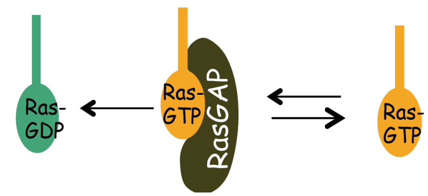


~90% variance captured by 4 components for all N's for 80% of the trials

Ras Activation Network

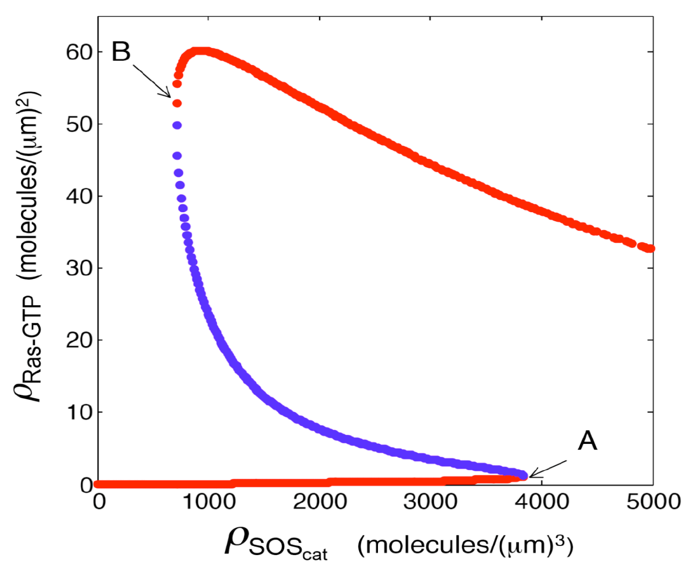


responsible for activation threshold in lymphocytes



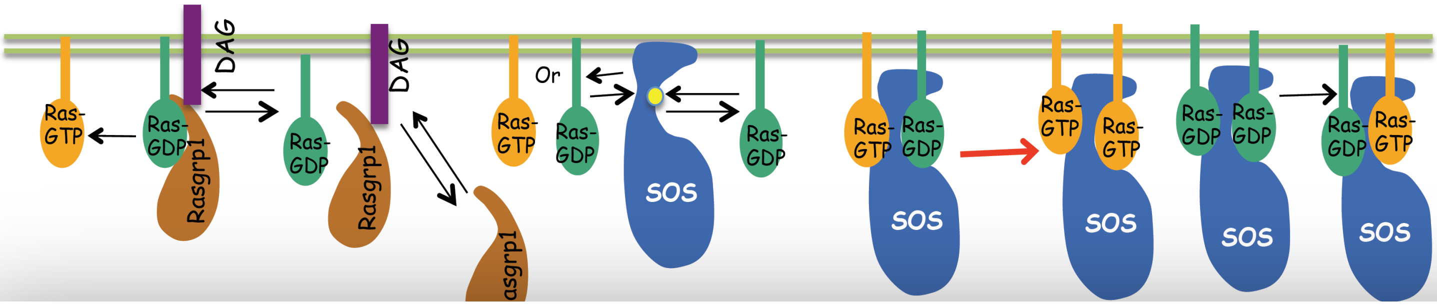
positive feedback
bistability

20 species, 23 kinetic rates



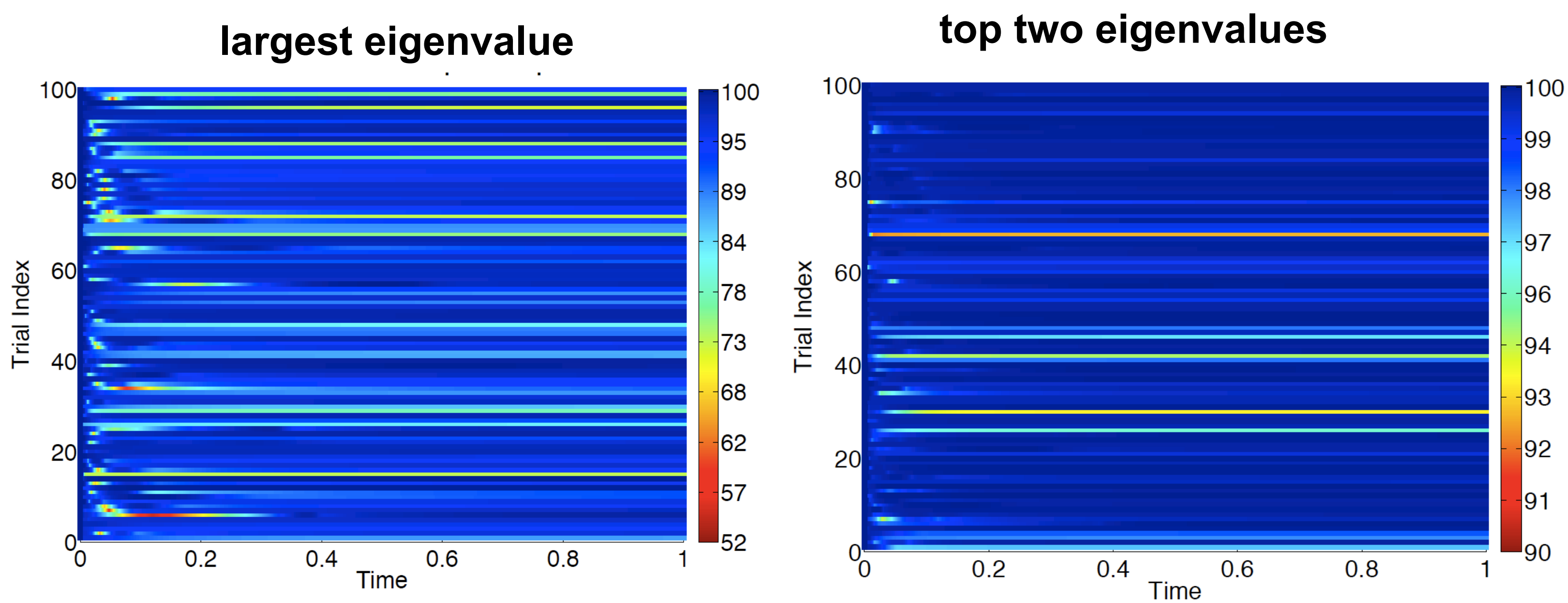
Das et al. Cell (2009)

Ras Activation Network



responsib

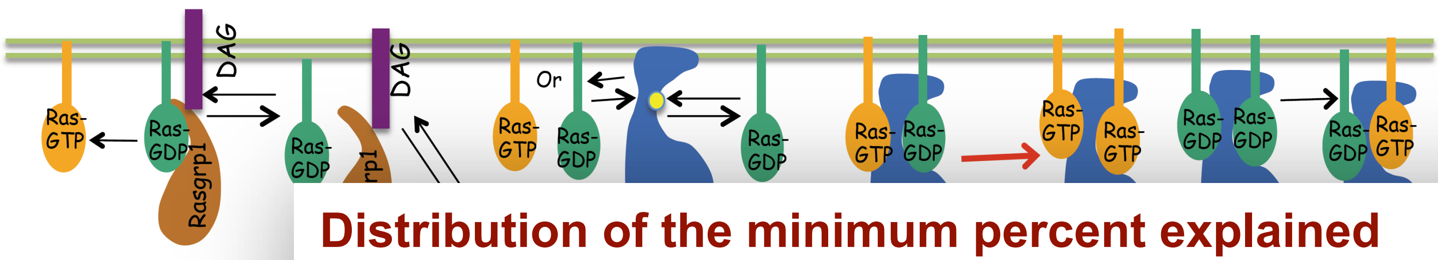
Kinetics of the percent explained



decrease of % explained in small time intervals as linear networks

Das et al. Cell (2009)

Ras Activation Network

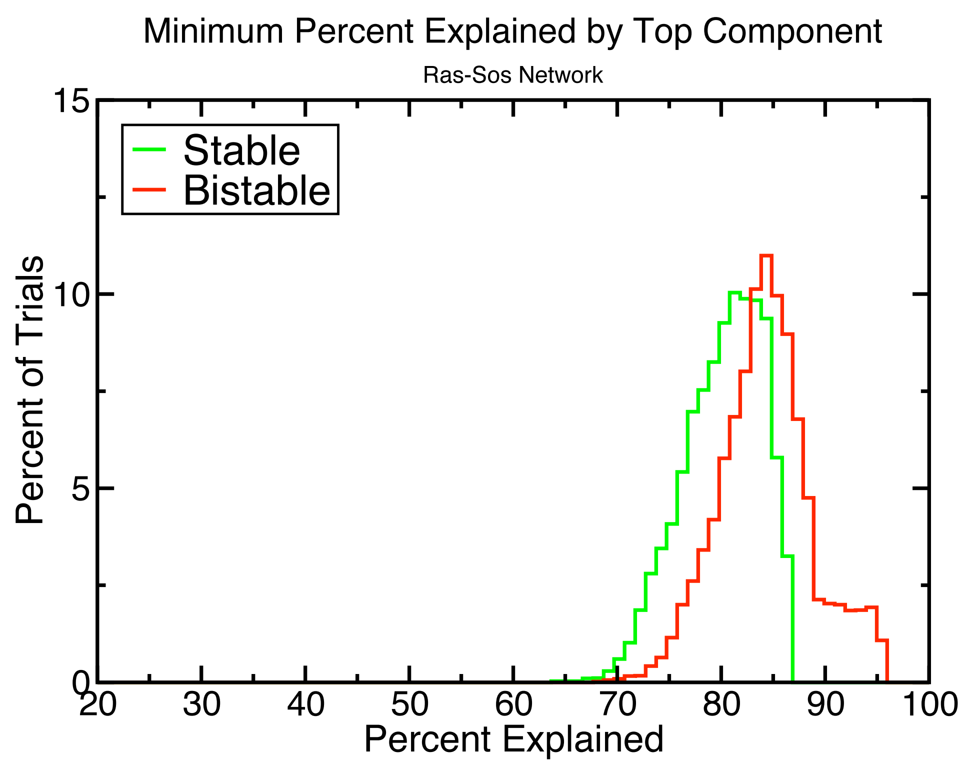


responsibl

positive feedb
bistability

20 species, 23 kinetic

Distribution of the minimum percent explained



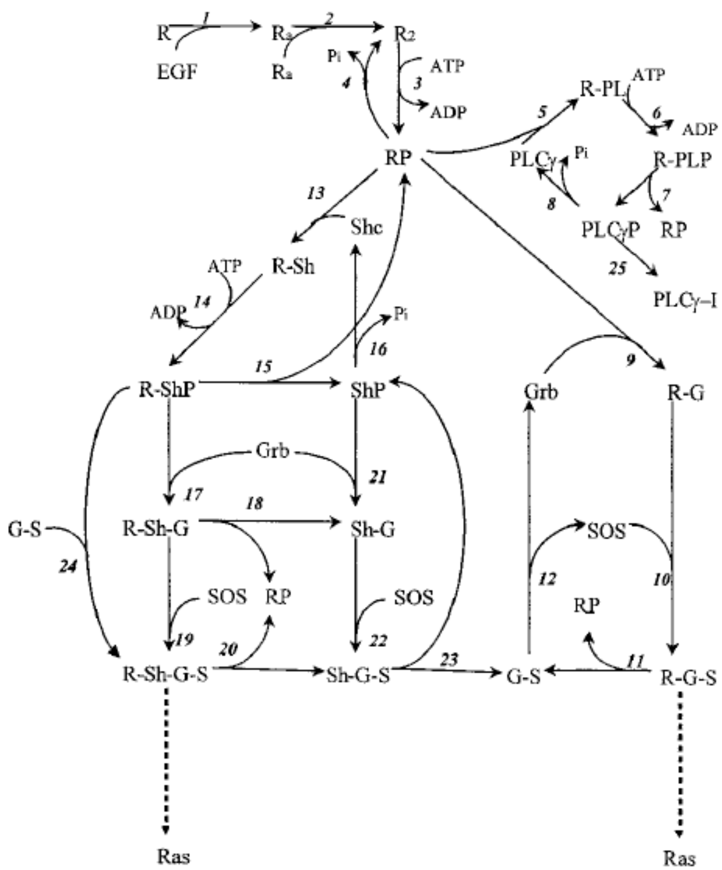
10,000 trials analyzed

reduction persists in network with bistability

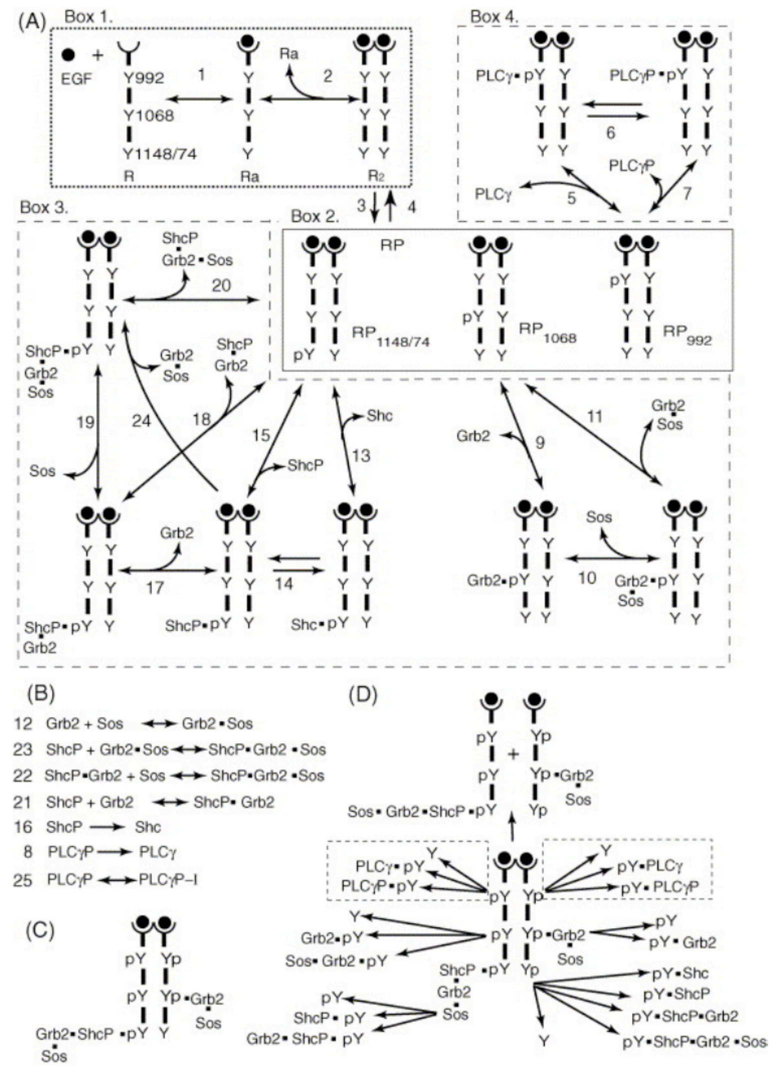
> 69% variance captured by the top eigenvalue

Das et al. Cell (2009)

EGFR Signaling Network



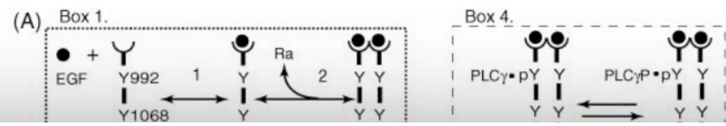
smaller network (19 species)
Kholodenko et al. JBC (1999)

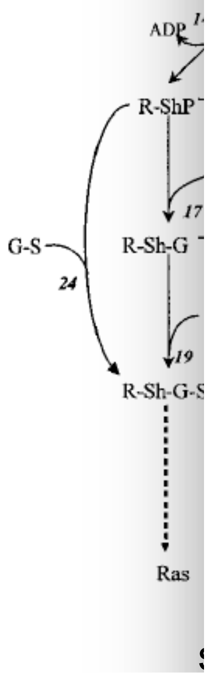


larger network (> 300 species)
Blinov et al. Biosys. (2006)

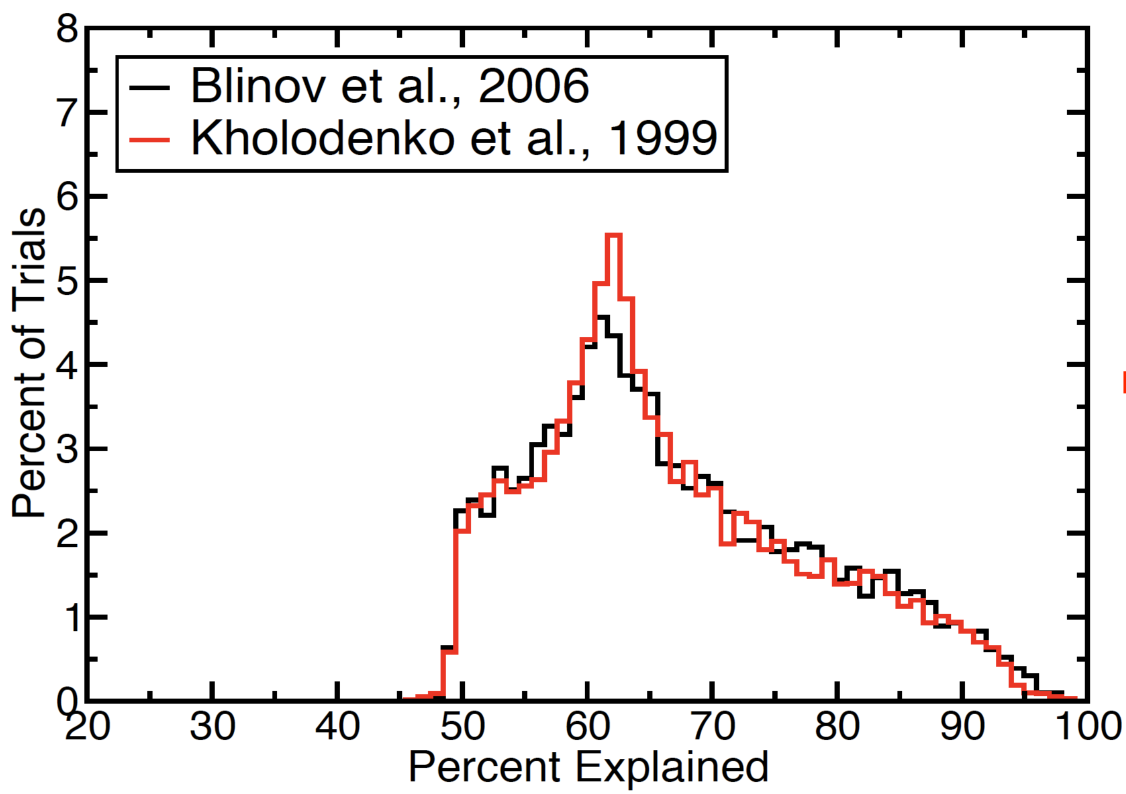
responsible for cell growth, differentiation

EGFR Signaling Network





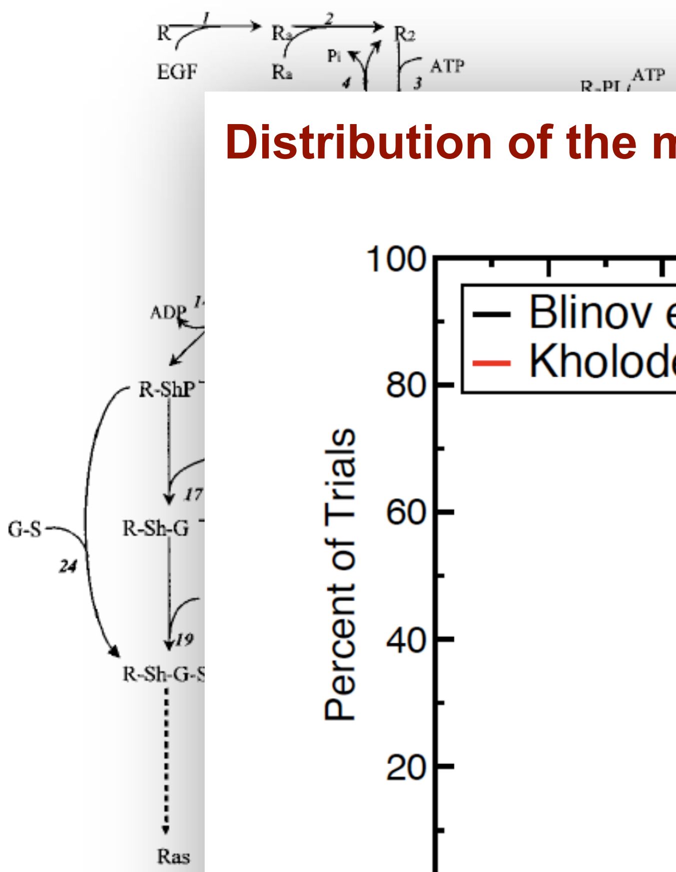
Distribution of the minimum percent explained

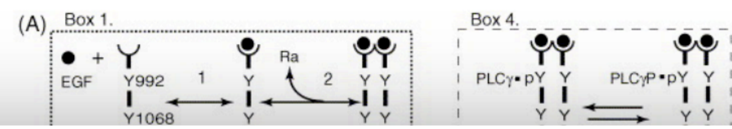


reduction persists in nonlinear network with multiple state activation

> 88% variance captured by the top three eigenvalues for all cases

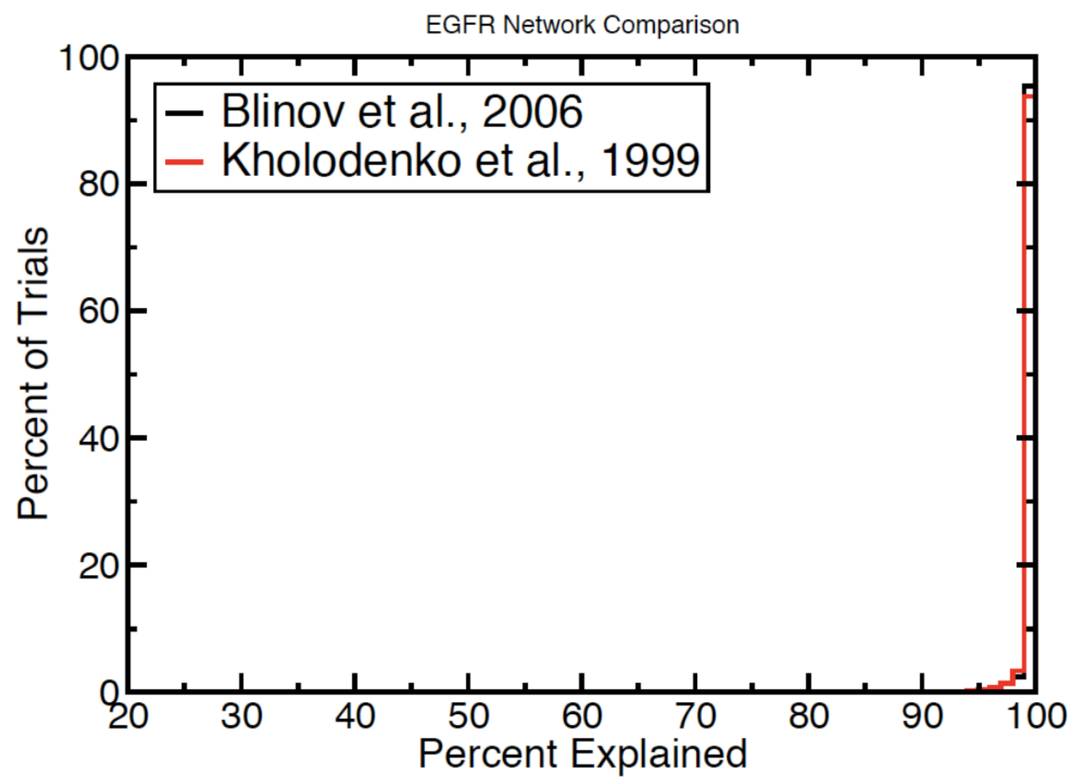
EGFR Signaling Network





Kholod

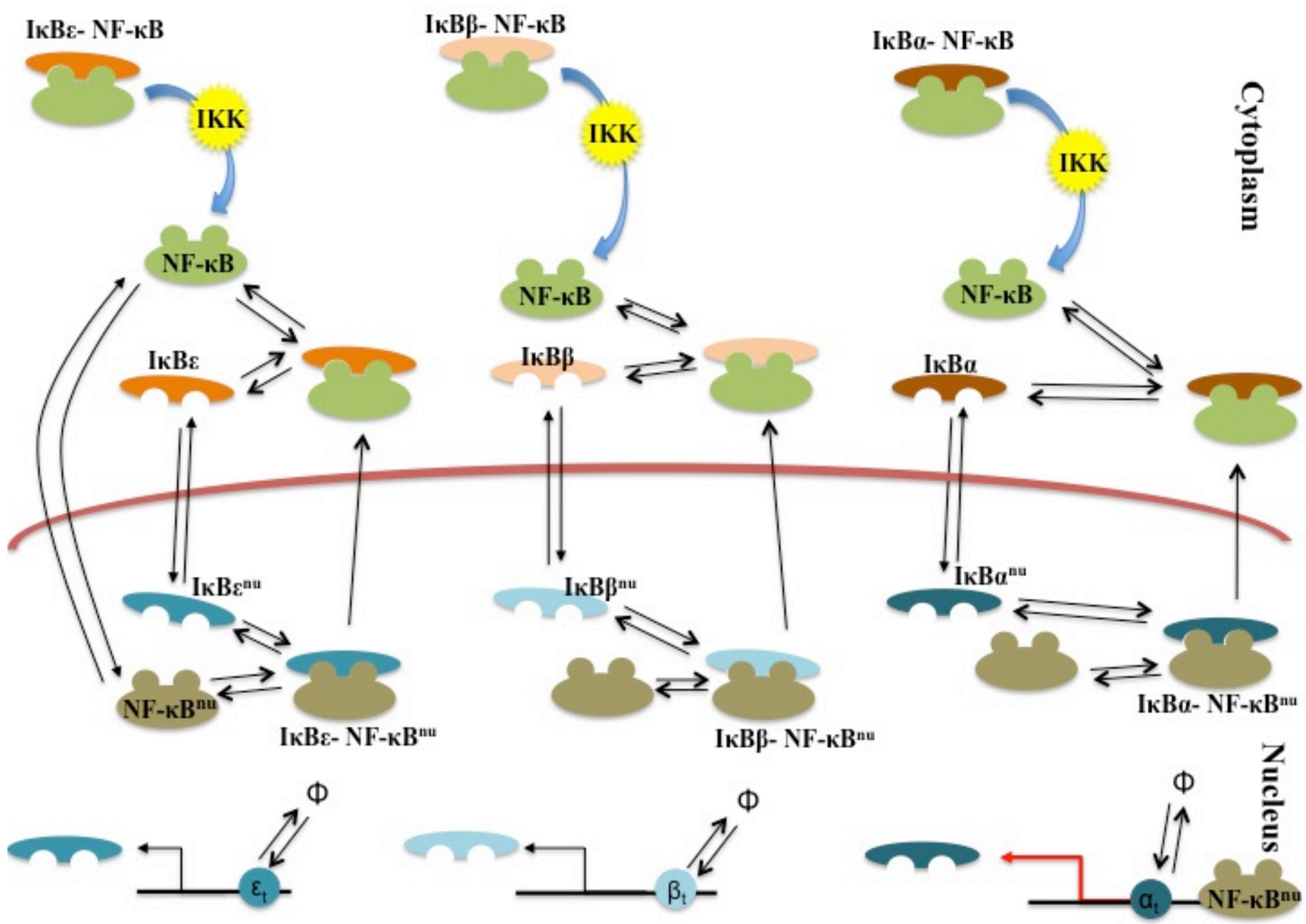
Distribution of the minimum percent explained



reduction persists in nonlinear network with multiple state activation

> 88% variance captured by the top three eigenvalues for all cases

NF-κB Signaling Network

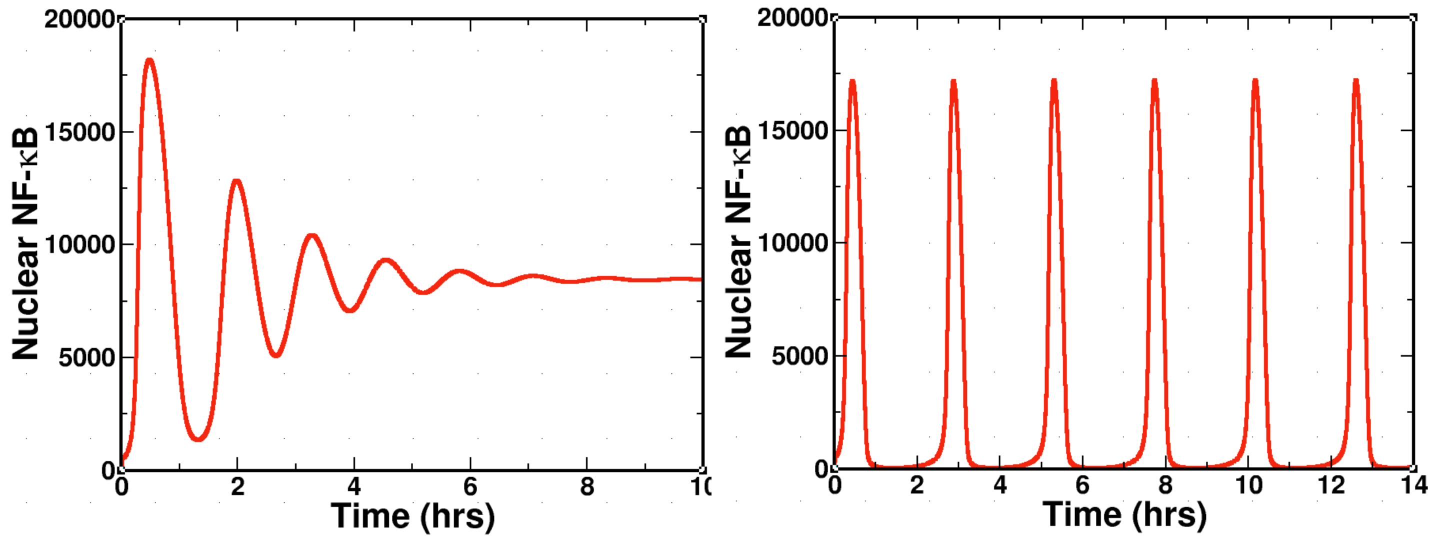


26 species, 38 kinetic rates

Hoffmann et al. Science (2002)

NF-κB Signaling Network

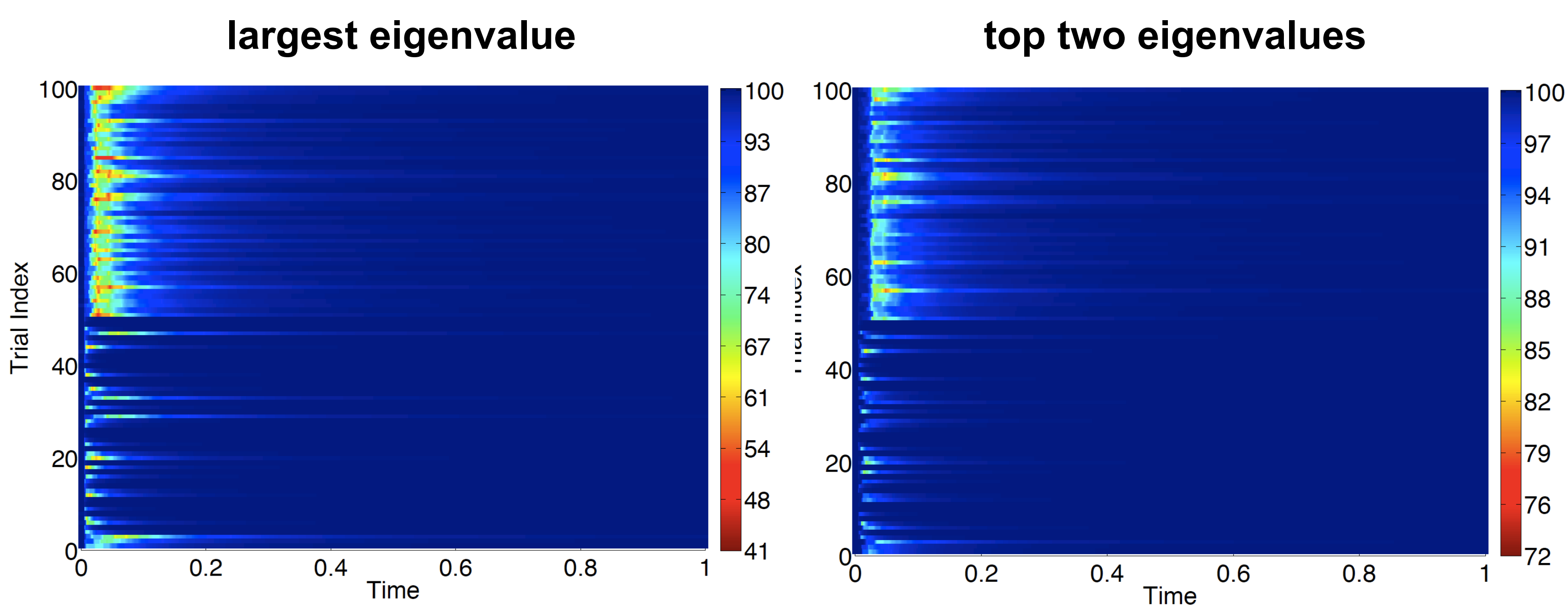
Sustained and damped oscillations in the kinetics



Hoffmann et al. Science (2002)

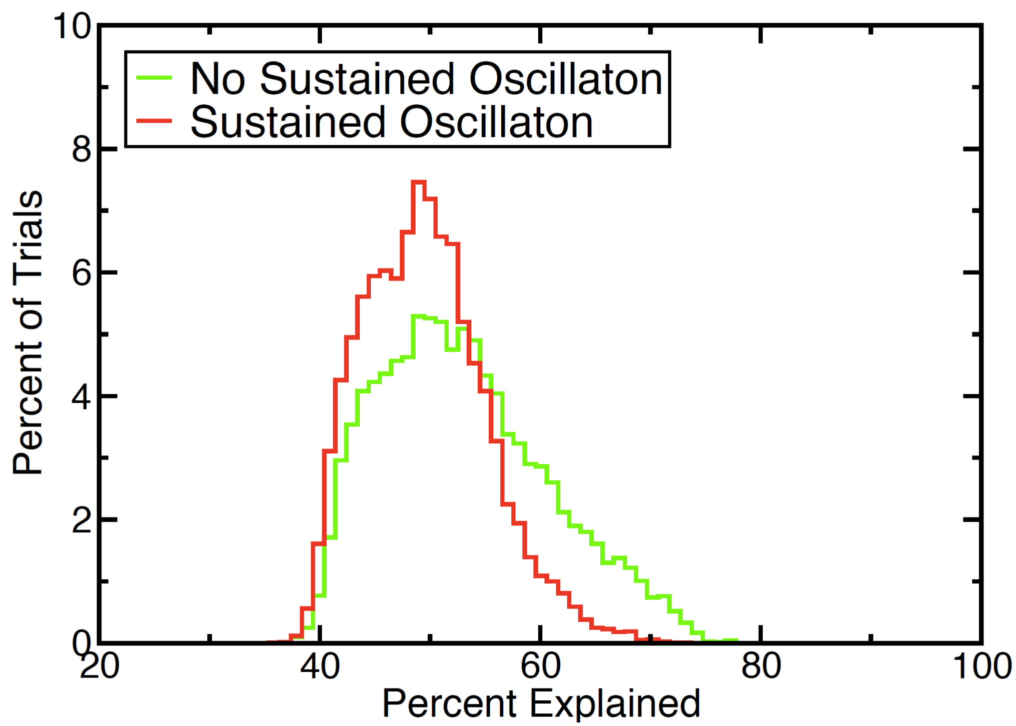
NF-κB Signaling Network

Eigenvalues show damped oscillations



NF-κB Signaling Network

Distribution of the minimum percent explained

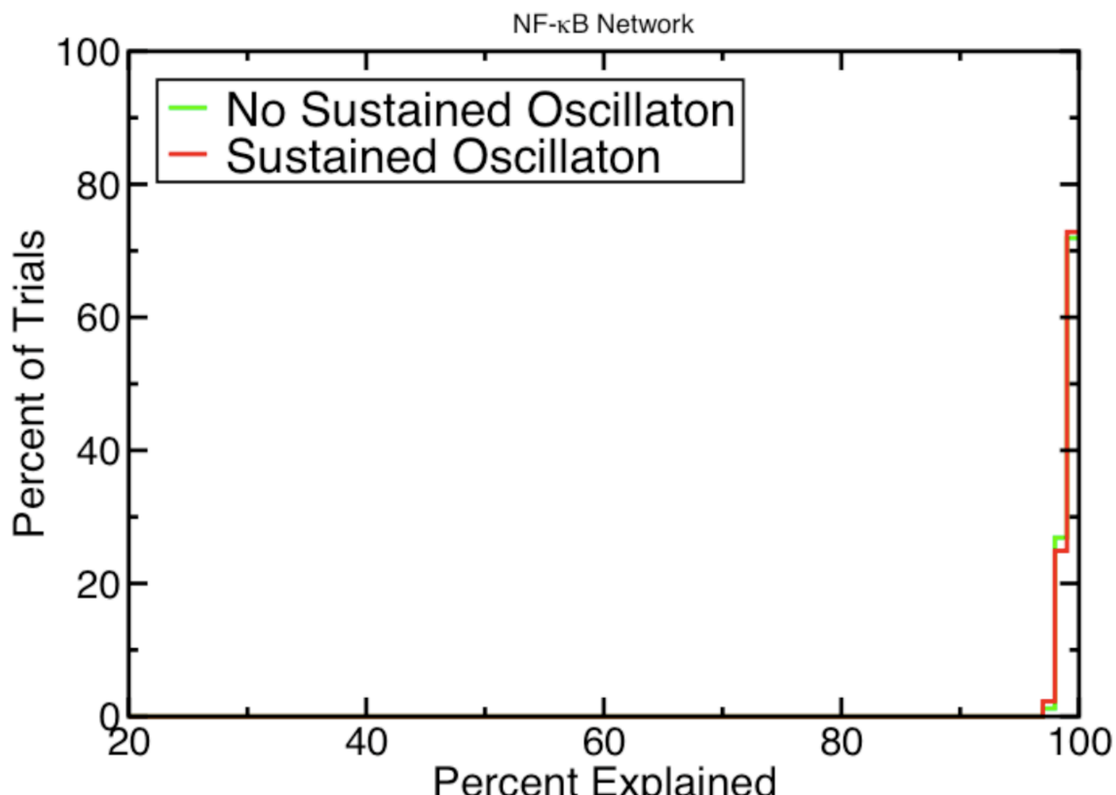


reduction
persists
in
nonlinear network
with oscillations

> 90% variance captured by the top three eigenvalues for all cases

NF-κB Signaling Network

Distribution of the minimum percent explained



reduction
persists
in
nonlinear network
with oscillations

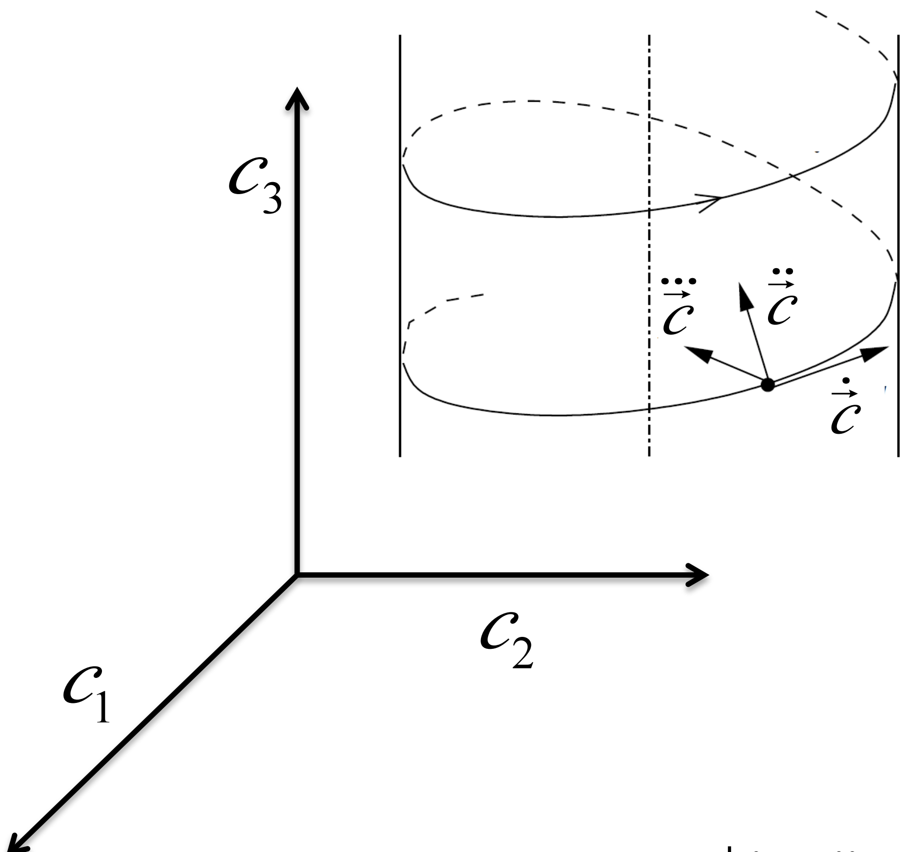
> 90% variance captured by the top three eigenvalues for all cases

Gram Determinant

$$\dot{\vec{c}} = \frac{d\vec{c}}{dt}$$

$$\ddot{\vec{c}} = \frac{d^2\vec{c}}{dt^2}$$

$$\dddot{\vec{c}} = \frac{d^3\vec{c}}{dt^3}$$

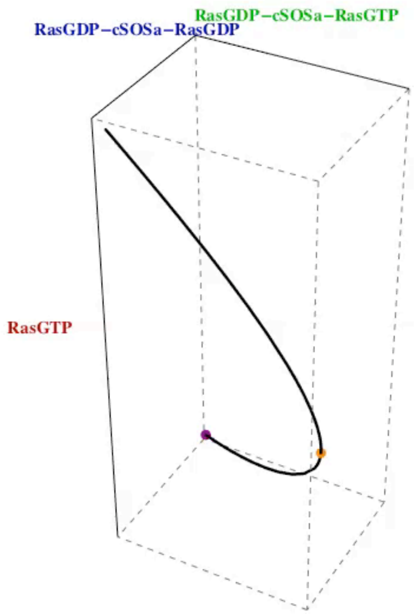


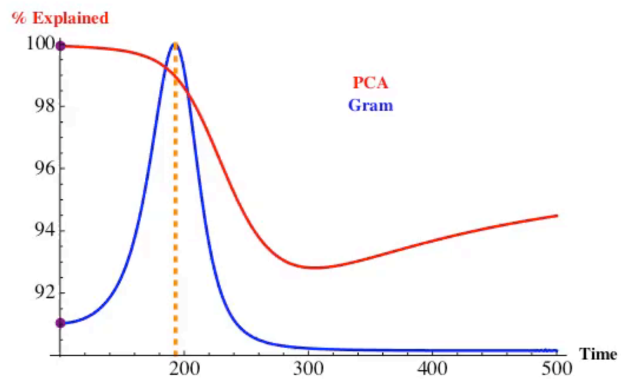
volume spanned by $(\dot{\vec{c}}, \ddot{\vec{c}}, \dddot{\vec{c}})$ contains information about local changes in direction.

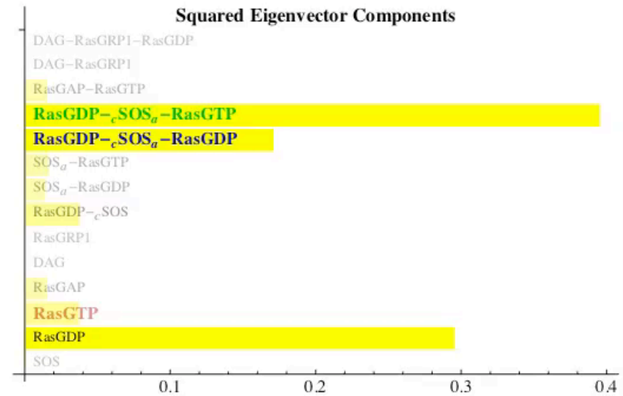
volume² spanned $(\dot{\vec{c}}, \ddot{\vec{c}}, \dddot{\vec{c}})$ by $= \left|\dot{\vec{c}} \cdot (\ddot{\vec{c}} \times \dddot{\vec{c}})\right|^2 = \det\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$

$a_{mn} = \vec{x}_m \cdot \vec{x}_n$

Gram Determinant

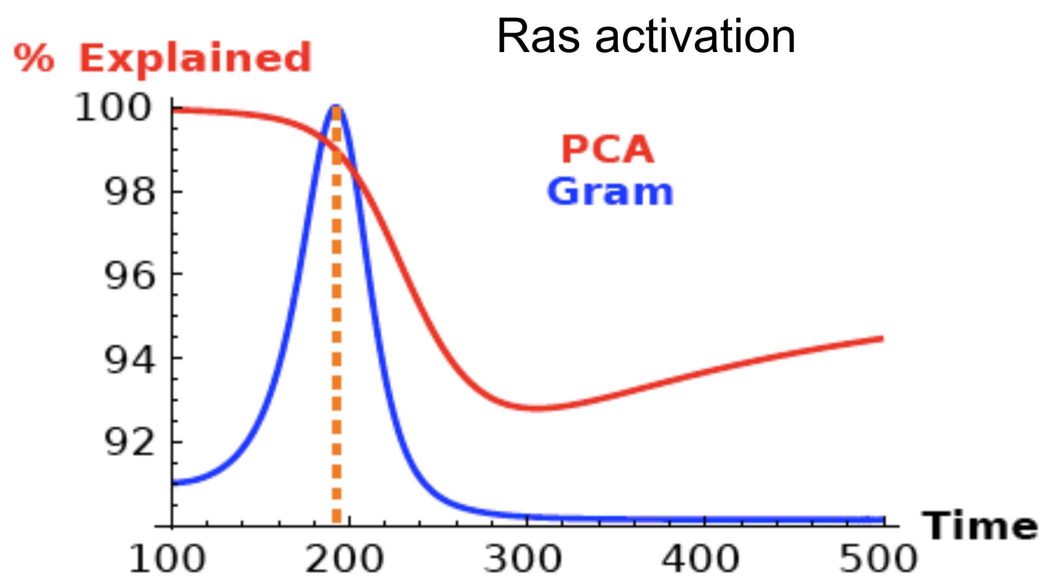


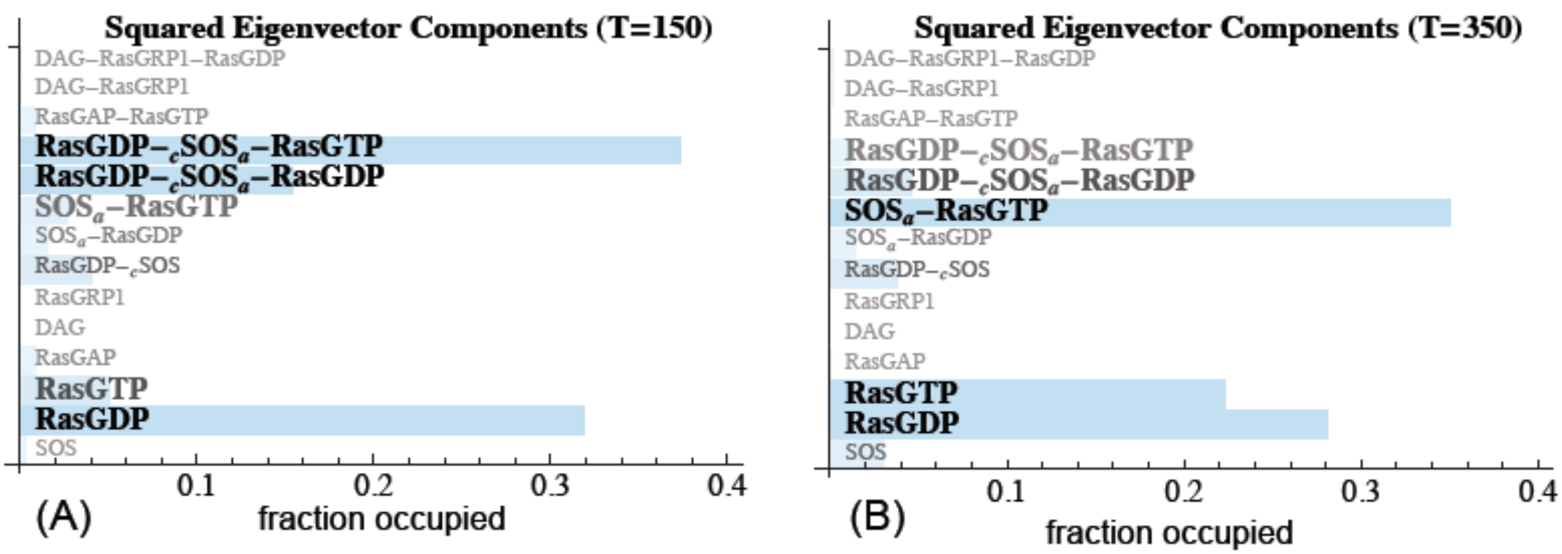




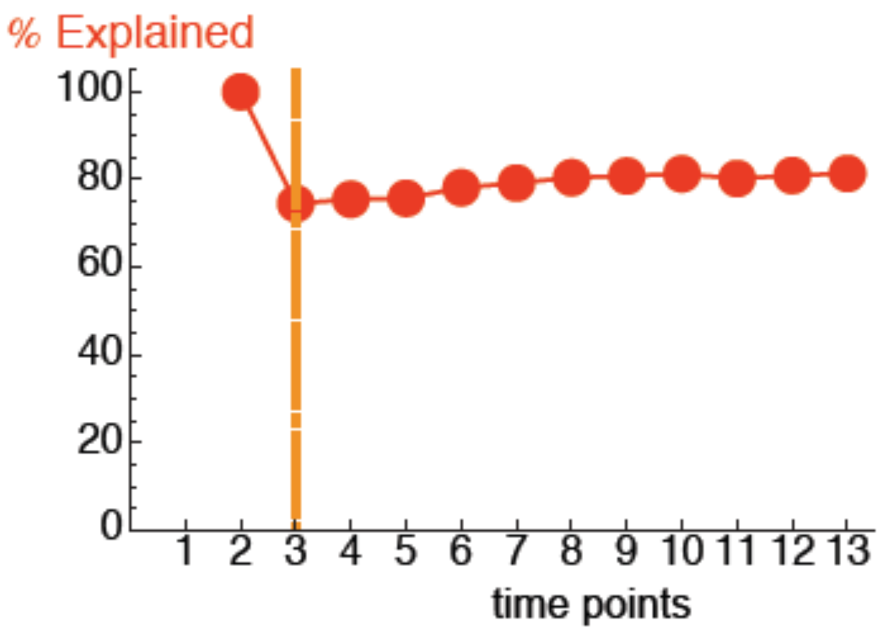
Largest eigenvalue kinetics displays time scale of Ras activation

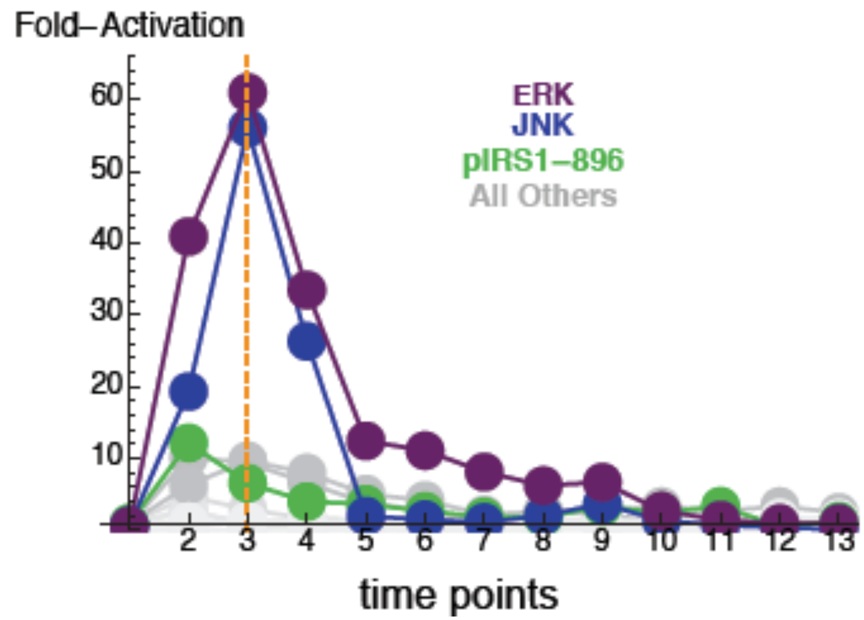
Mechanistic Insights





Mechanistic Insights





Data from Gaudet et al (2004)

Summary

strong correlations between species in a biochemical reaction networks produce dramatic reduction in dimensionality that is *insensitive* to

-rate constants and initial concentrations
-nonlinearities in the kinetics
-network topology

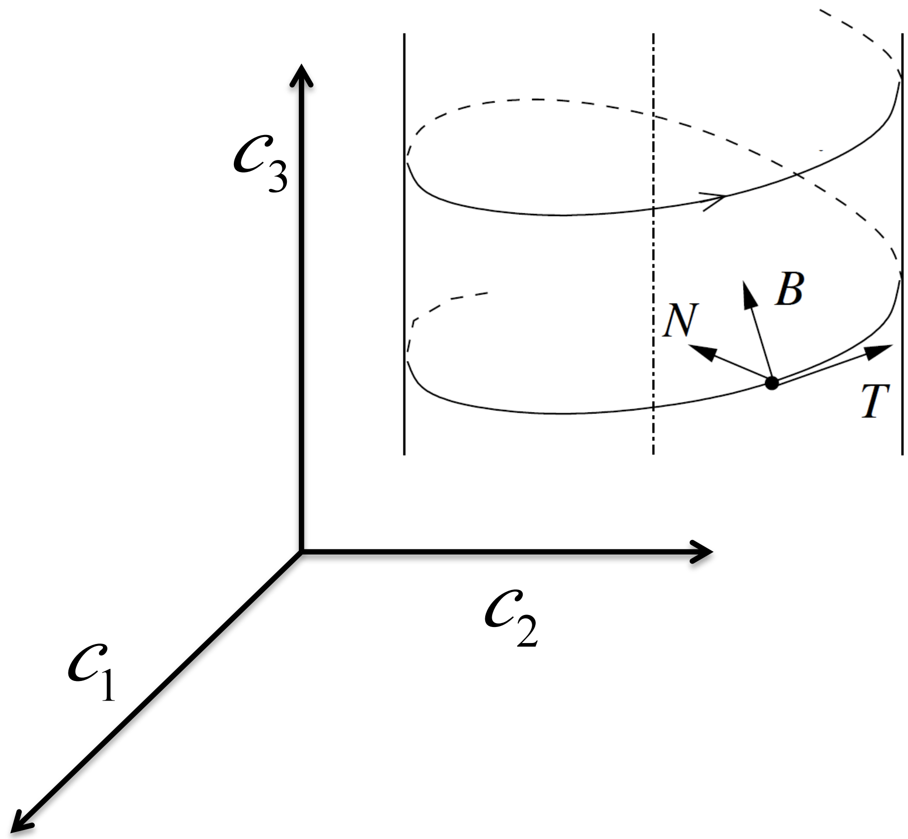
Time-scales associated with significant changes in the kinetics is reflected in the percent explained by the principal components

Results are published in Dworkin et al. *J. R. Soc. Interface* (2012)

Contact: das.70@osu.edu and http://www.mathmed.org/#Jayajit_Das

Summary

Mechanistic Models?



Frenet-Serret Formula

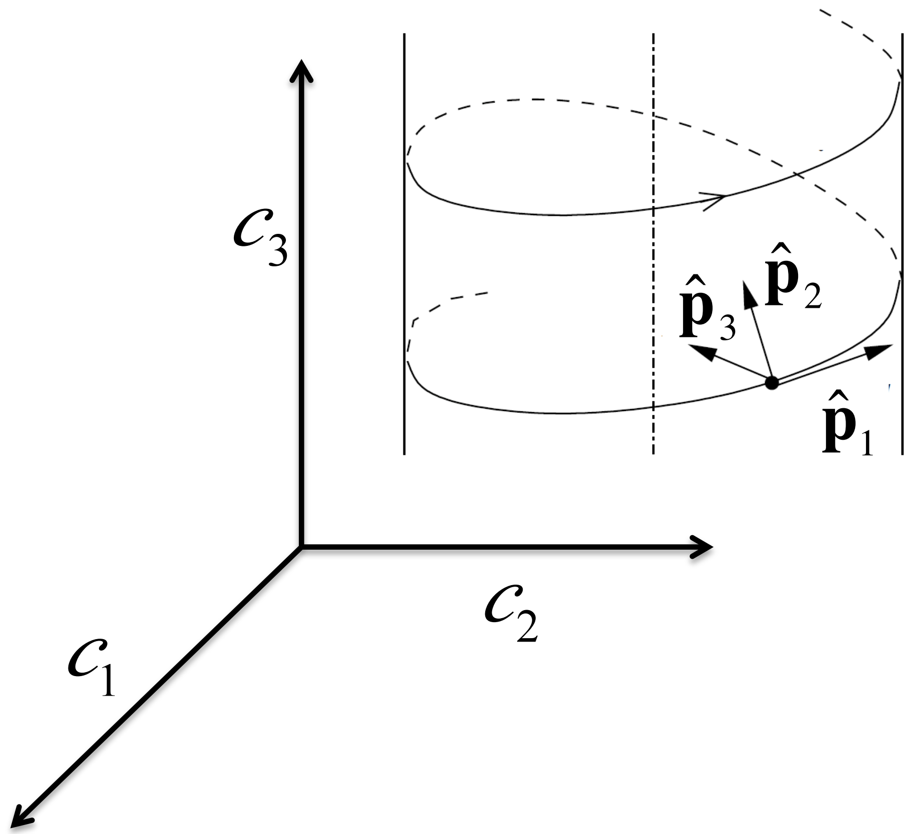
$$\frac{d\mathbf{T}}{ds} = \kappa(s)\mathbf{N}$$

$$\frac{d\mathbf{N}}{ds} = -\kappa(s)\mathbf{T} + \tau(s)\mathbf{B}$$

$$\frac{d\mathbf{B}}{ds} = -\tau(s)\mathbf{N}$$

Summary

Mechanistic Models?



Effective kinetics ?

$$\frac{d\hat{\mathbf{p}}_1}{ds} = \kappa(s)\hat{\mathbf{p}}_2$$

$$\frac{d\hat{\mathbf{p}}_2}{ds} = -\kappa(s)\hat{\mathbf{p}}_1 + \tau(s)\hat{\mathbf{p}}_3$$

$$\frac{d\hat{\mathbf{p}}_3}{ds} = -\tau(s)\hat{\mathbf{p}}_2$$

Acknowledgements

Michael Dworkin

Sayak Mukherjee

Ciriyam Jayaprakash
Physics, OSU

Funding

Nationwide Children's Hospital

Ohio Supercomputer Center
Empower. Partner. Lead.