Analysis of signalling in large protein interaction networks

Ton Coolen

Dept of Mathematics and Randall Division
King's College London

=

Several research and teaching initiatives
at maths-computation-biomedicine interface
at King's College London

Dec 2011: Institute for Mathematical and Molecular Biomedicine

- biological networks:
  - graph theory for cellular signalling networks
  - network null models via MCMC processes
  - reaction dynamics in large protein interaction networks
- Bayesian analysis and biomedical statistics:
  - analysis of fluorescence lifetime data
  - clinical outcome prediction from biomarkers
  - survival analysis for heterogeneous cohorts with competing risks
- other topics:
  - theory of cell re-programming
  - immune networks

THIS TALK

- Problems with pathway analysis
- How to quantify protein network topology
- Analysis of signalling dynamics in large protein networks

Problems with proteomic pathway analysis

usual description:
reaction equations

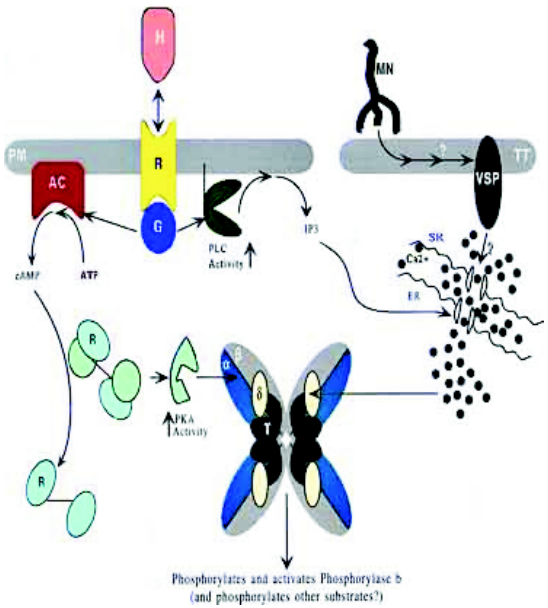


Table 2. Model Equations

$$
\begin{aligned}
d(RD)/dt &= k_{81}RDA - k_{18}RD \cdot A + k_{31}RDE - k_{13}RD \cdot E - k_{19}RD + k_{91}R \cdot D + k_{21}RT - k_{12}RD \cdot M \\
d(RT)/dt &= k_{52}RTE - k_{25}RT \cdot E + k_{92}R \cdot T - k_{29}RT - k_{21}RT + k_{62}RTA - k_{26}RT \cdot A - k_{2M}RT \cdot E + k_{M2}M + k_{12}RD \cdot M \\
d(RDE)/dt &= k_{13}RD \cdot E - k_{31}RDE + k_{43}RE \cdot D - k_{34}RDE + k_{53}RTE \\
d(RE)/dt &= k_{34}RDE - k_{43}RE \cdot D + k_{54}RTE - k_{45}RE \cdot T + k_{94}R \cdot E - k_{49}RE \\
d(RTE)/dt &= k_{45}RE \cdot T - k_{54}RTE + k_{25}RT \cdot E - k_{52}RTE - k_{53}RTE \\
d(RTA)/dt &= k_{26}RT \cdot A - k_{62}RTA - k_{68}RTA + k_{76}RA \cdot T - k_{67}RTA \\
d(RA)/dt &= k_{67}RTA - k_{76}RA \cdot T + k_{97}R \cdot A - k_{79}RA + k_{87}RDA - k_{78}RA \cdot D \\
d(RDA)/dt &= k_{68}RTA + k_{78}RA \cdot D - k_{87}RDA + k_{18}RD \cdot A - k_{81}RDA \\
d(R)/dt &= k_{29}RT - k_{92}R \cdot T + k_{49}RE - k_{94}R \cdot E + k_{19}RD - k_{91}R \cdot D + k_{79}RA - k_{97}R \cdot A \\
d(E)/dt &= k_{31}RDE - k_{13}RD \cdot E + k_{52}RTE - k_{25}RT \cdot E + k_{49}RE - k_{94}R \cdot E - k_{2M}RT \cdot E + k_{M2}M \\
d(A)/dt &= k_{81}RDA - k_{18}RD \cdot A + k_{62}RTA - k_{26}RT \cdot A + k_{79}RA - k_{97}R \cdot A \\
d(M)/dt &= k_{2M}RT \cdot E - k_{M2}M
\end{aligned}
$$

Model equations correspond to the reaction scheme shown in Figure 1. Numbering of the reaction rate constants follows the conventions introduced in Table 3.

- cannot solve equations analytically ...
- uncertainty regarding pathways & parameters ...
- too many components for numerical exploration ...

Problems with proteomic pathway analysis

usual description:
reaction equations

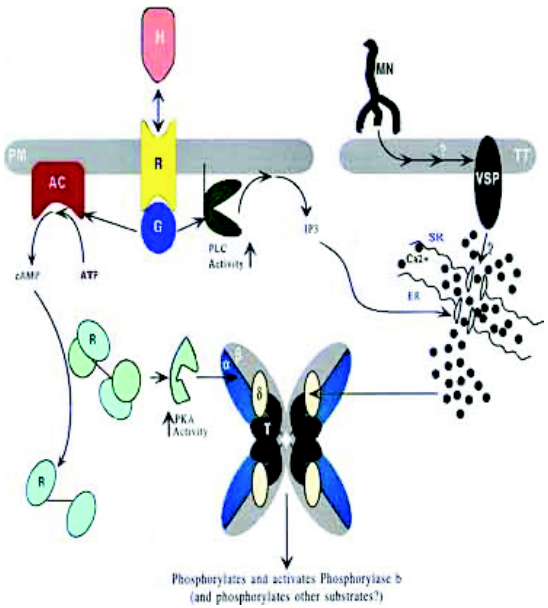


Table 2. Model Equations

$$
\begin{aligned}
d(RD)/dt &= k_{81}RDA - k_{18}RD \cdot A + k_{31}RDE - k_{13}RD \cdot E - k_{19}RD + k_{91}R \cdot D + k_{21}RT - k_{12}RD \cdot M \\
d(RT)/dt &= k_{52}RTE - k_{25}RT \cdot E + k_{92}R \cdot T - k_{29}RT - k_{21}RT + k_{62}RTA - k_{26}RT \cdot A - k_{2M}RT \cdot E + k_{M2}M + k_{12}RD \cdot M \\
d(RDE)/dt &= k_{13}RD \cdot E - k_{31}RDE + k_{43}RE \cdot D - k_{34}RDE + k_{53}RTE \\
d(RE)/dt &= k_{34}RDE - k_{43}RE \cdot D + k_{54}RTE - k_{45}RE \cdot T + k_{94}R \cdot E - k_{49}RE \\
d(RTE)/dt &= k_{45}RE \cdot T - k_{54}RTE + k_{25}RT \cdot E - k_{52}RTE - k_{53}RTE \\
d(RTA)/dt &= k_{26}RT \cdot A - k_{62}RTA - k_{68}RTA + k_{76}RA \cdot T - k_{67}RTA \\
d(RA)/dt &= k_{67}RTA - k_{76}RA \cdot T + k_{97}R \cdot A - k_{79}RA + k_{87}RDA - k_{78}RA \cdot D \\
d(RDA)/dt &= k_{68}RTA + k_{78}RA \cdot D - k_{87}RDA + k_{18}RD \cdot A - k_{81}RDA \\
d(R)/dt &= k_{29}RT - k_{92}R \cdot T + k_{49}RE - k_{94}R \cdot E + k_{19}RD - k_{91}R \cdot D + k_{79}RA - k_{97}R \cdot A \\
d(E)/dt &= k_{31}RDE - k_{13}RD \cdot E + k_{52}RTE - k_{25}RT \cdot E + k_{49}RE - k_{94}R \cdot E - k_{2M}RT \cdot E + k_{M2}M \\
d(A)/dt &= k_{81}RDA - k_{18}RD \cdot A + k_{62}RTA - k_{26}RT \cdot A + k_{79}RA - k_{97}R \cdot A \\
d(M)/dt &= k_{2M}RT \cdot E - k_{M2}M
\end{aligned}
$$

Model equations correspond to the reaction scheme shown in Figure 1. Numbering of the reaction rate constants follows the conventions introduced in Table 3.

- cannot solve equations analytically ...
- uncertainty regarding pathways & parameters ...
- too many components for numerical exploration ...

Problems with proteomic pathway analysis

usual description:
reaction equations

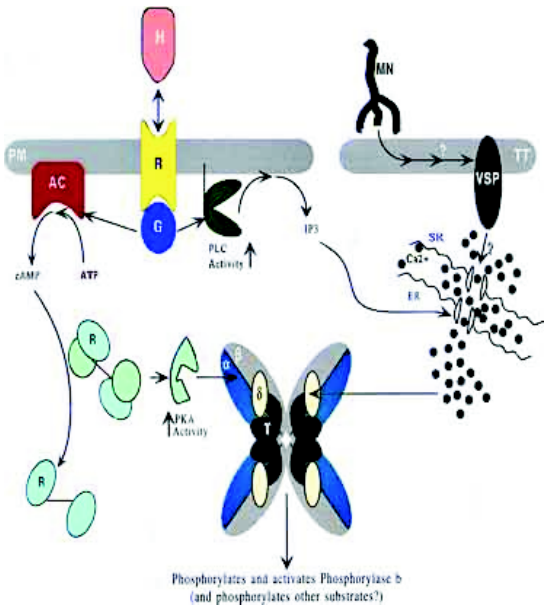


Table 2. Model Equations

$$
\begin{aligned}
d(RD)/dt &= k_{81}RDA - k_{18}RD \cdot A + k_{31}RDE - k_{13}RD \cdot E - k_{19}RD + k_{91}R \cdot D + k_{21}RT - k_{12}RD \cdot M \\
d(RT)/dt &= k_{52}RTE - k_{25}RT \cdot E + k_{92}R \cdot T - k_{29}RT - k_{21}RT + k_{62}RTA - k_{26}RT \cdot A - k_{2M}RT \cdot E + k_{M2}M + k_{12}RD \cdot M \\
d(RDE)/dt &= k_{13}RD \cdot E - k_{31}RDE + k_{43}RE \cdot D - k_{34}RDE + k_{53}RTE \\
d(RE)/dt &= k_{34}RDE - k_{43}RE \cdot D + k_{54}RTE - k_{45}RE \cdot T + k_{94}R \cdot E - k_{49}RE \\
d(RTE)/dt &= k_{45}RE \cdot T - k_{54}RTE + k_{25}RT \cdot E - k_{52}RTE - k_{53}RTE \\
d(RTA)/dt &= k_{26}RT \cdot A - k_{62}RTA - k_{68}RTA + k_{76}RA \cdot T - k_{67}RTA \\
d(RA)/dt &= k_{67}RTA - k_{76}RA \cdot T + k_{97}R \cdot A - k_{79}RA + k_{87}RDA - k_{78}RA \cdot D \\
d(RDA)/dt &= k_{68}RTA + k_{78}RA \cdot D - k_{87}RDA + k_{18}RD \cdot A - k_{81}RDA \\
d(R)/dt &= k_{29}RT - k_{92}R \cdot T + k_{49}RE - k_{94}R \cdot E + k_{19}RD - k_{91}R \cdot D + k_{79}RA - k_{97}R \cdot A \\
d(E)/dt &= k_{31}RDE - k_{13}RD \cdot E + k_{52}RTE - k_{25}RT \cdot E + k_{49}RE - k_{94}R \cdot E - k_{2M}RT \cdot E + k_{M2}M \\
d(A)/dt &= k_{81}RDA - k_{18}RD \cdot A + k_{62}RTA - k_{26}RT \cdot A + k_{79}RA - k_{97}R \cdot A \\
d(M)/dt &= k_{2M}RT \cdot E - k_{M2}M
\end{aligned}
$$

Model equations correspond to the reaction scheme shown in Figure 1. Numbering of the reaction rate constants follows the conventions introduced in Table 3.

- cannot solve equations analytically ...
- uncertainty regarding pathways & parameters ...
- too many components for numerical exploration ...

Problems with proteomic pathway analysis

usual description:
reaction equations

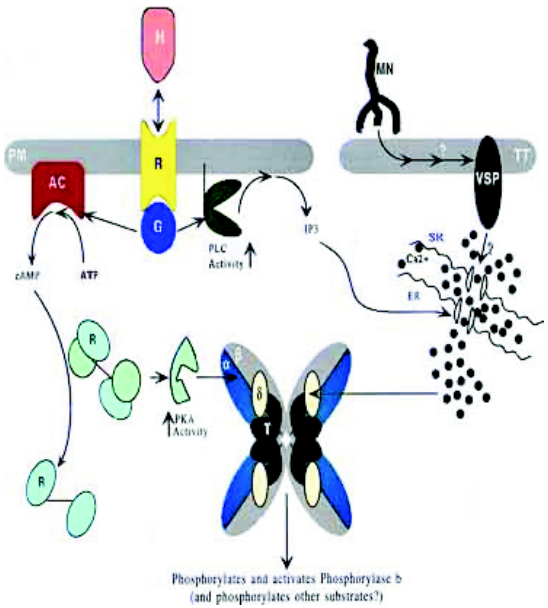


Table 2. Model Equations

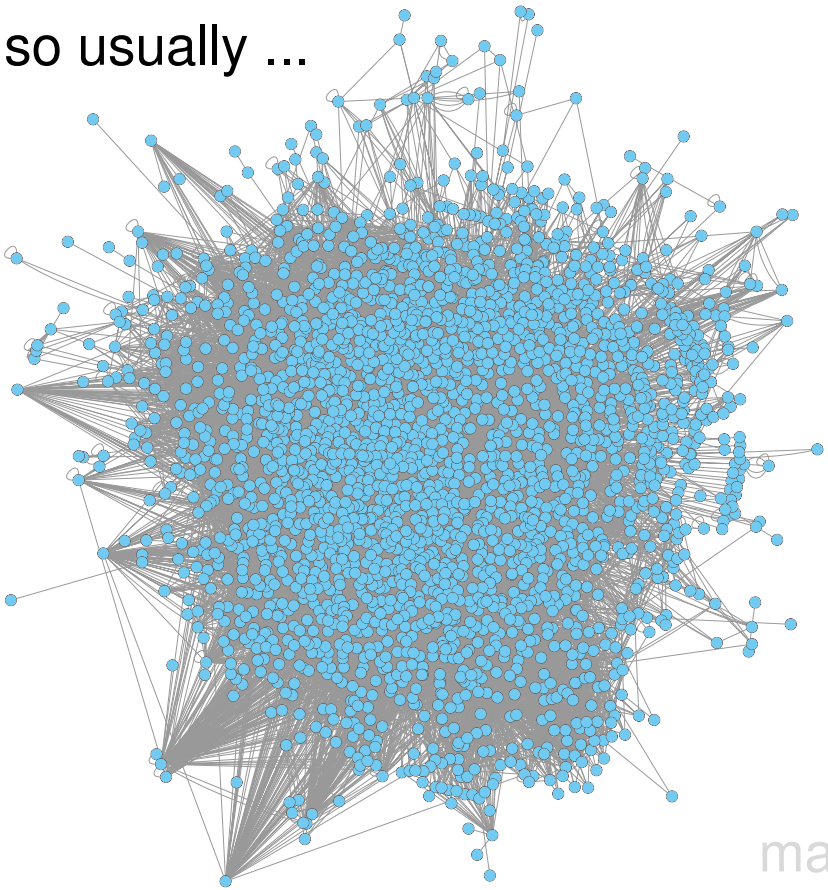
$$
\begin{aligned}
d(RD)/dt &= k_{81}RDA - k_{18}RD \cdot A + k_{31}RDE - k_{13}RD \cdot E - k_{19}RD + k_{91}R \cdot D + k_{21}RT - k_{12}RD \cdot M \\
d(RT)/dt &= k_{52}RTE - k_{25}RT \cdot E + k_{92}R \cdot T - k_{29}RT - k_{21}RT + k_{62}RTA - k_{26}RT \cdot A - k_{2M}RT \cdot E + k_{M2}M + k_{12}RD \cdot M \\
d(RDE)/dt &= k_{13}RD \cdot E - k_{31}RDE + k_{43}RE \cdot D - k_{34}RDE + k_{53}RTE \\
d(RE)/dt &= k_{34}RDE - k_{43}RE \cdot D + k_{54}RTE - k_{45}RE \cdot T + k_{94}R \cdot E - k_{49}RE \\
d(RTE)/dt &= k_{45}RE \cdot T - k_{54}RTE + k_{25}RT \cdot E - k_{52}RTE - k_{53}RTE \\
d(RTA)/dt &= k_{26}RT \cdot A - k_{62}RTA - k_{68}RTA + k_{76}RA \cdot T - k_{67}RTA \\
d(RA)/dt &= k_{67}RTA - k_{76}RA \cdot T + k_{97}R \cdot A - k_{79}RA + k_{87}RDA - k_{78}RA \cdot D \\
d(RDA)/dt &= k_{68}RTA + k_{78}RA \cdot D - k_{87}RDA + k_{18}RD \cdot A - k_{81}RDA \\
d(R)/dt &= k_{29}RT - k_{92}R \cdot T + k_{49}RE - k_{94}R \cdot E + k_{19}RD - k_{91}R \cdot D + k_{79}RA - k_{97}R \cdot A \\
d(E)/dt &= k_{31}RDE - k_{13}RD \cdot E + k_{52}RTE - k_{25}RT \cdot E + k_{49}RE - k_{94}R \cdot E - k_{2M}RT \cdot E + k_{M2}M \\
d(A)/dt &= k_{81}RDA - k_{18}RD \cdot A + k_{62}RTA - k_{26}RT \cdot A + k_{79}RA - k_{97}R \cdot A \\
d(M)/dt &= k_{2M}RT \cdot E - k_{M2}M
\end{aligned}
$$

Model equations correspond to the reaction scheme shown in Figure 1. Numbering of the reaction rate constants follows the conventions introduced in Table 3.

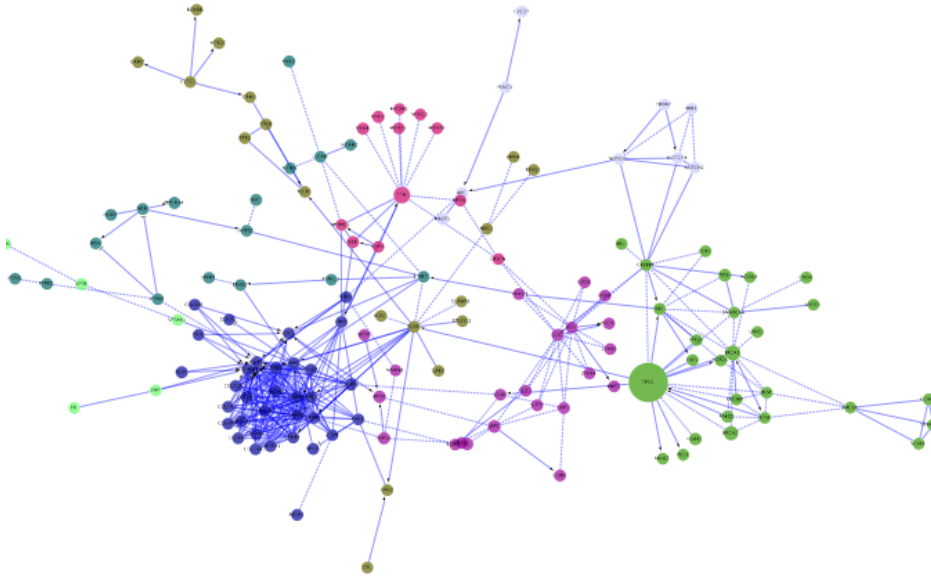
- cannot solve equations analytically ...
- uncertainty regarding pathways & parameters ...
- too many components for numerical exploration ...

‘The most significant challenges that face mechanistic modelling are the lack of quantitative kinetic data and the combinatorial increase in the number of distinct species ... of the protein network ...’ (Kholodenko 2006)

so usually ...



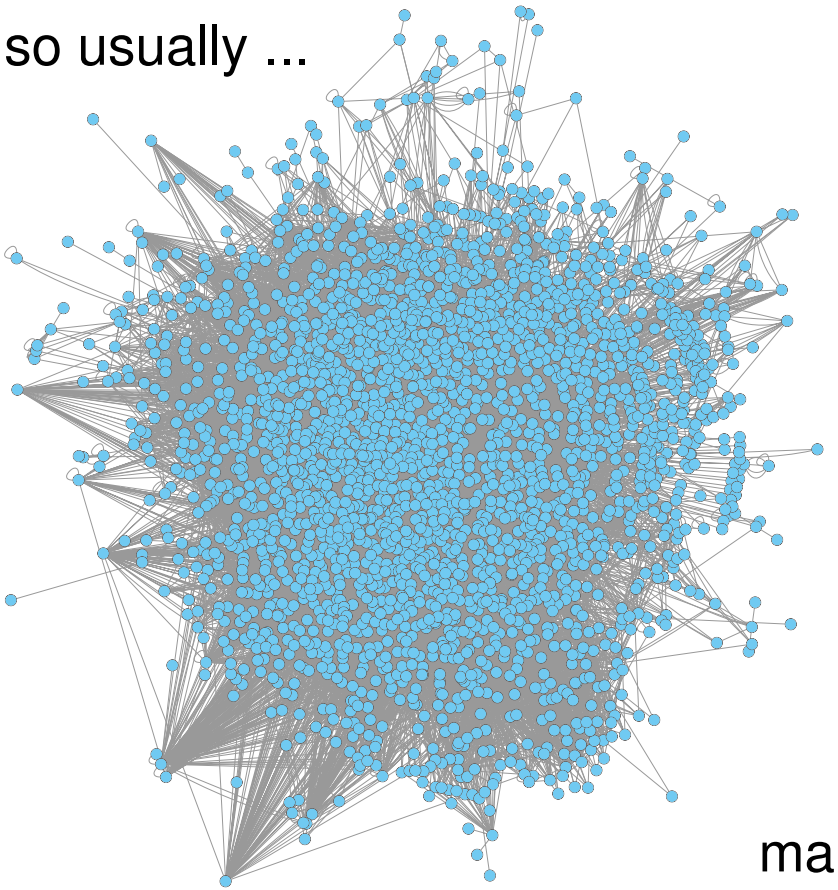
$\Rightarrow$



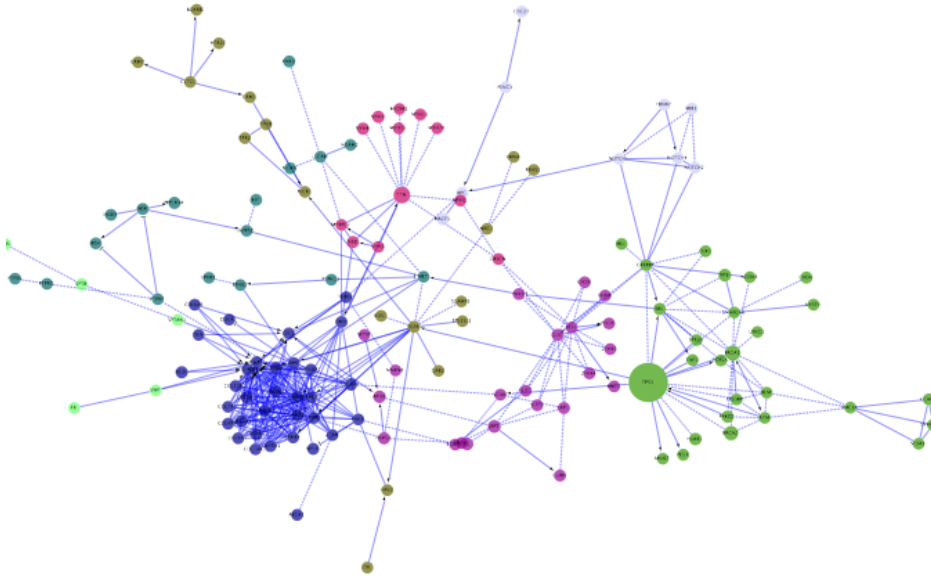
many-variable problem is not limited to biology
can we adapt methods from physics?

‘The most significant challenges that face mechanistic modelling are the lack of quantitative kinetic data and the combinatorial increase in the number of distinct species ... of the protein network ...’ (Kholodenko 2006)

so usually ...

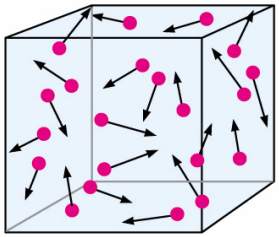


$\Rightarrow$



many-variable problem is not limited to biology
can we adapt methods from physics?

statistical physics



dynamical variables:

coordinates and velocities
$(\vec{x}_1, \vec{v}_1),\ (\vec{x}_2, \vec{v}_2),\ \ldots$

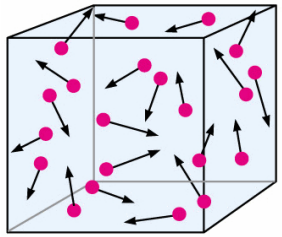
microscopic dynamics:

Newton's equations of motion
$\frac{d}{dt}(\vec{x}_1, \vec{v}_1) = \ldots,\ \frac{d}{dt}(\vec{x}_2, \vec{v}_2) = \ldots$ ← don't try to solve these!

macroscopic description:

densities, correlation functions,
perturbation response functions,
phase transitions ...

statistical physics



dynamical variables:

coordinates and velocities

$(\vec{x}_1, \vec{v}_1),\ (\vec{x}_2, \vec{v}_2),\ \ldots$

microscopic dynamics:

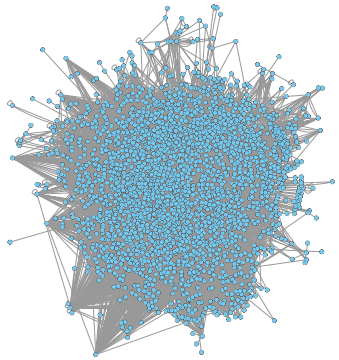
Newton's equations of motion

$\frac{d}{dt}(\vec{x}_1, \vec{v}_1) = \ldots,\ \frac{d}{dt}(\vec{x}_2, \vec{v}_2) = \ldots$

macroscopic description:

densities, correlation functions,
response functions (to perturbations),
phase transitions ...

statistical biology?



dynamical variables:

concentr of proteins & complexes

$\vec{x}_1,\ \vec{x}_2,\ \vec{x}_3,\ \ldots$

microscopic dynamics:

reaction equations

$\frac{d}{dt}\vec{x}_1 = \ldots,\ \frac{d}{dt}\vec{x}_2 = \ldots, \frac{d}{dt}\vec{x}_3 = \ldots$

macroscopic description:

???

Proteome:

heterogeneous many-particle system,
small number of partners per node

- math methods since $\sim$1980/2000 (statics/dynamics)
- what is many? $N = 1000$ or more ...
- biology is not physics:
  no evolution to equilibrium, conservation laws ...

What should we expect to get out?

- predictions for **macroscopic** quantities in **typical** proteomes
  (correlation functions, response functions, ...)
- collective phenomena (e.g. phase transitions)
- not dependent on *details* of network or reaction rates,
  only on network and rate **statistics** ('self-averaging')

potential to address both of Kholodenko's fundamental problems

Proteome:

heterogeneous many-particle system,
small number of partners per node

- math methods since $\sim$1980/2000 (statics/dynamics)
- what is many? $N = 1000$ or more ...
- biology is not physics:
  no evolution to equilibrium, conservation laws ...

What should we expect to get out?

- predictions for **macroscopic** quantities in **typical** proteomes
  (correlation functions, response functions, ...)
- collective phenomena (e.g. phase transitions)
- not dependent on *details* of network or reaction rates,
  only on network and rate **statistics** ('self-averaging')

potential to address both of Kholodenko's fundamental problems

Proteome:

heterogeneous many-particle system,
small number of partners per node

- math methods since $\sim$1980/2000 (statics/dynamics)
- what is many? $N = 1000$ or more ...
- biology is not physics:
  no evolution to equilibrium, conservation laws ...

What should we expect to get out?

- predictions for **macroscopic** quantities in **typical** proteomes
  (correlation functions, response functions, ...)
- collective phenomena (e.g. phase transitions)
- not dependent on *details* of network or reaction rates,
  only on network and rate **statistics** ('self-averaging')

potential to address both of Kholodenko's fundamental problems

Proteome:

heterogeneous many-particle system,
small number of partners per node

- math methods since $\sim$1980/2000 (statics/dynamics)
- what is many? $N = 1000$ or more ...
- biology is not physics:
  no evolution to equilibrium, conservation laws ...

What should we expect to get out?

- predictions for **macroscopic** quantities in **typical** proteomes
  (correlation functions, response functions, ...)
- collective phenomena (e.g. phase transitions)
- not dependent on *details* of network or reaction rates,
  only on network and rate **statistics** ('self-averaging')

potential to address both of Kholodenko's fundamental problems

Proteome:

heterogeneous many-particle system,
small number of partners per node

- math methods since $\sim$1980/2000 (statics/dynamics)
- what is many? $N = 1000$ or more ...
- biology is not physics:
  no evolution to equilibrium, conservation laws ...

What should we expect to get out?

- predictions for **macroscopic** quantities in **typical** proteomes
  (correlation functions, response functions, ...)
- collective phenomena (e.g. phase transitions)
- not dependent on *details* of network or reaction rates,
  only on network and rate **statistics** ('self-averaging')

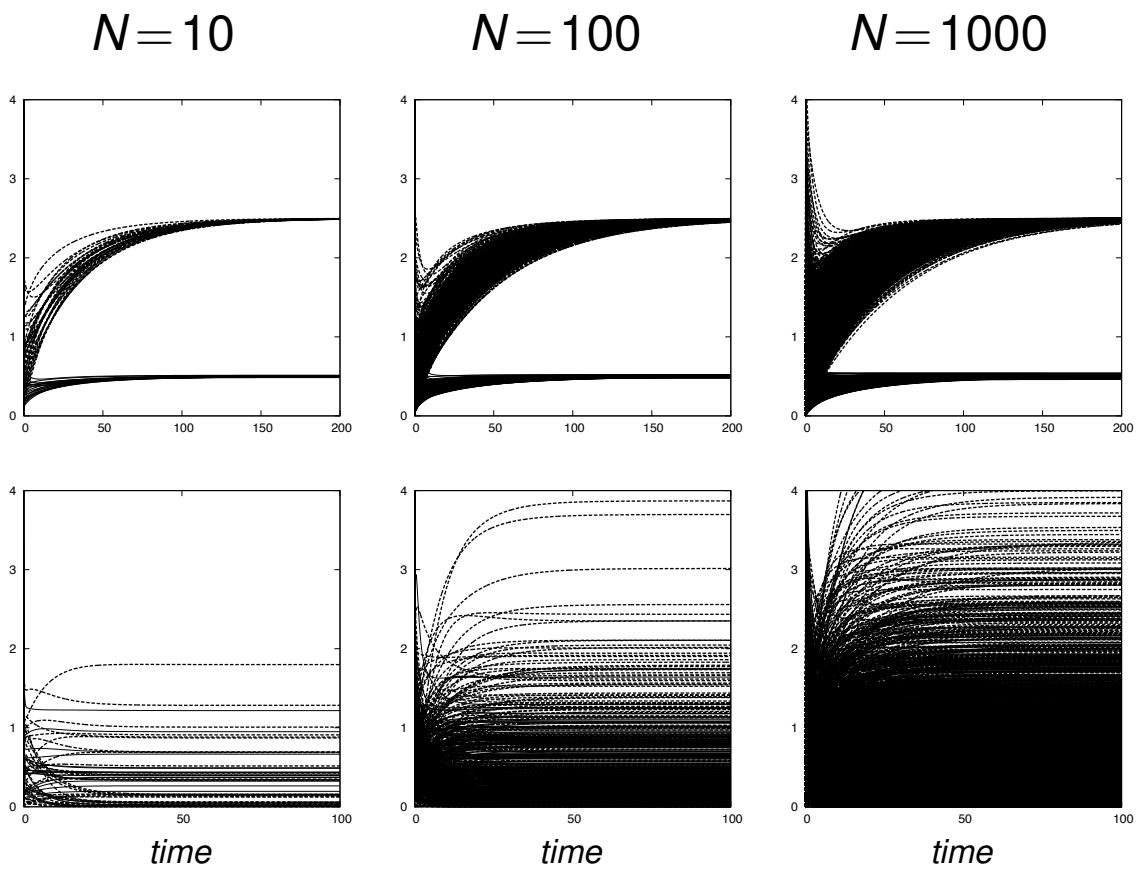
potential to address both of Kholodenko's fundamental problems

numerical illustration:

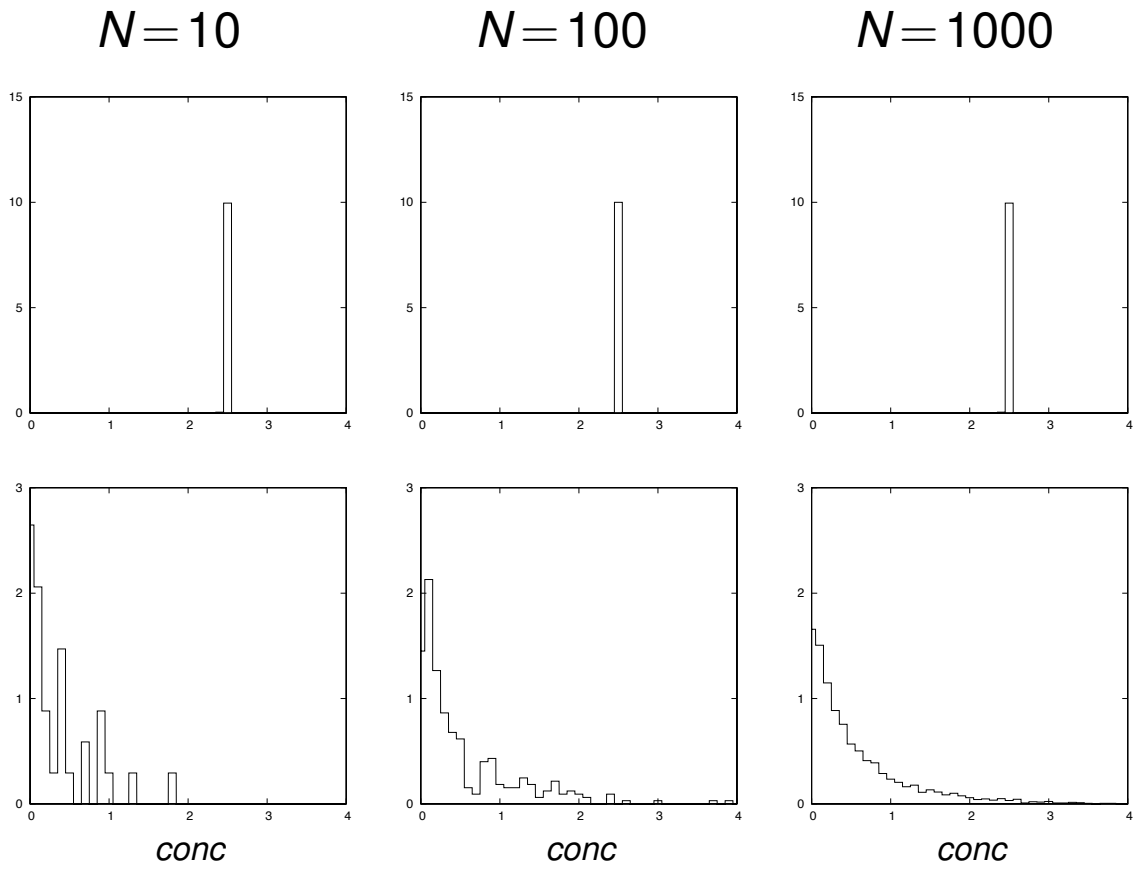
two states/protein,
binary complexes,

random topology,
average nr of partners: 7

dashed lines: conc of complexes
solid lines: conc of unbound proteins



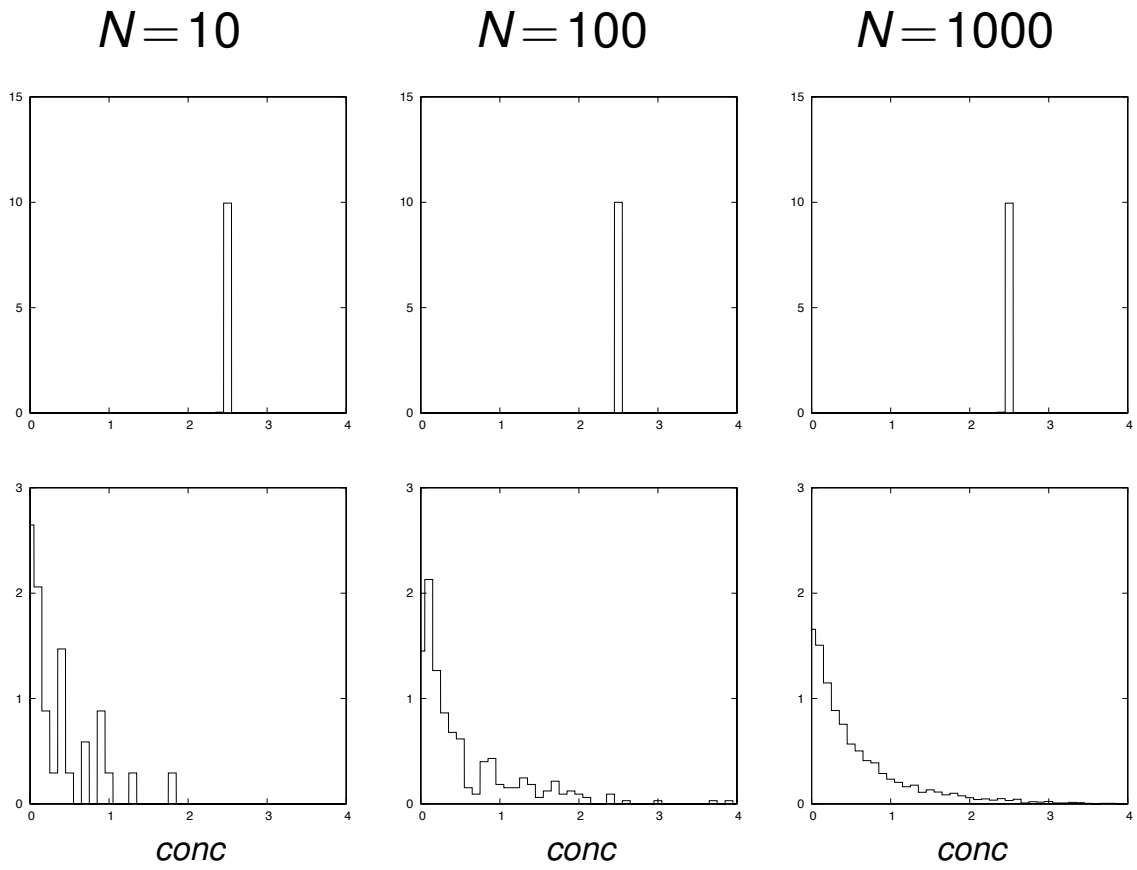
e.g. distribution of complex concentrations in stationary state



individual trajectories not predictable,

statistics of trajectories **predictable** as $N \to \infty$
(and dependent only on topology and rate distributions)

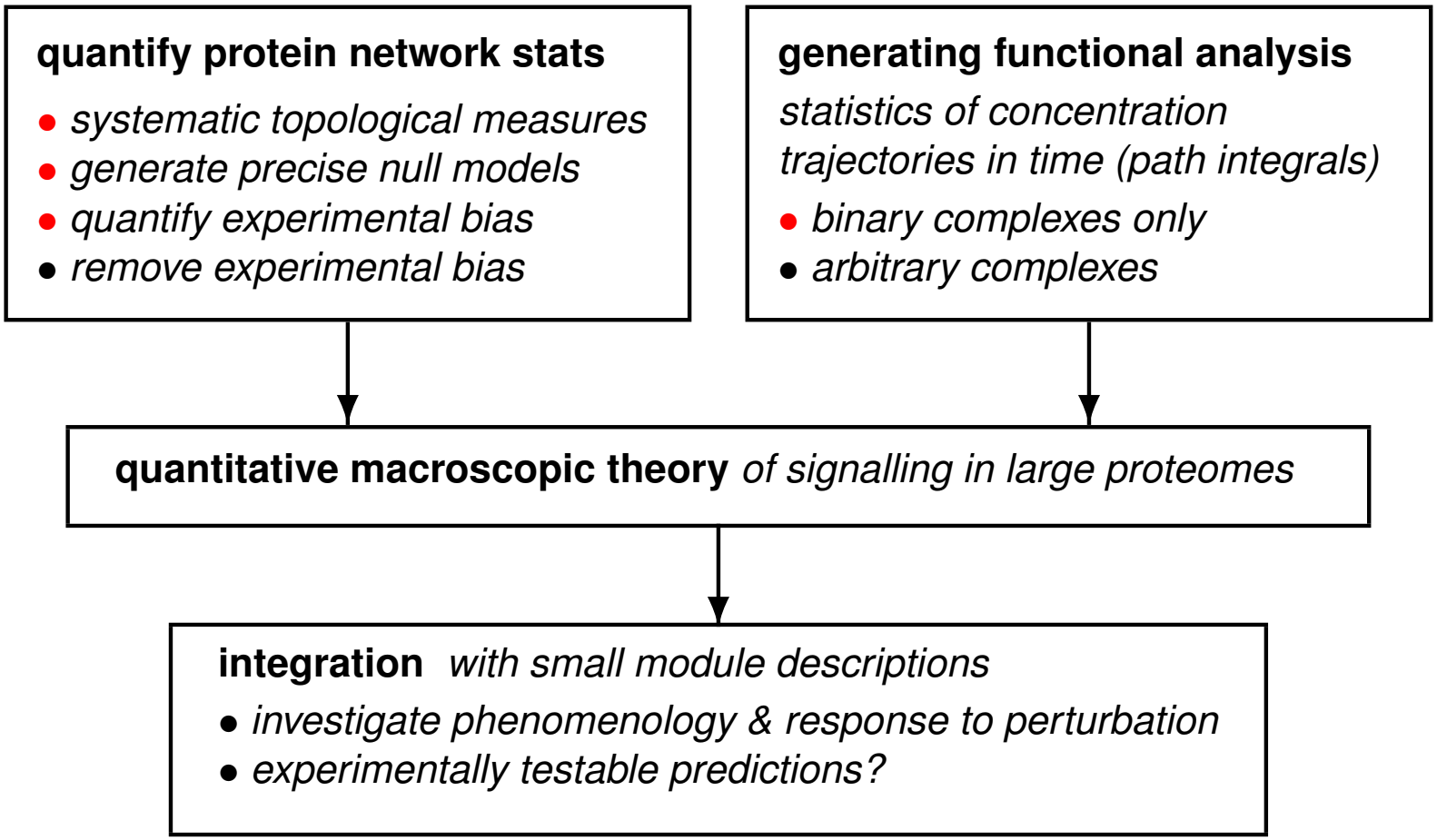
e.g. distribution of complex concentrations in stationary state



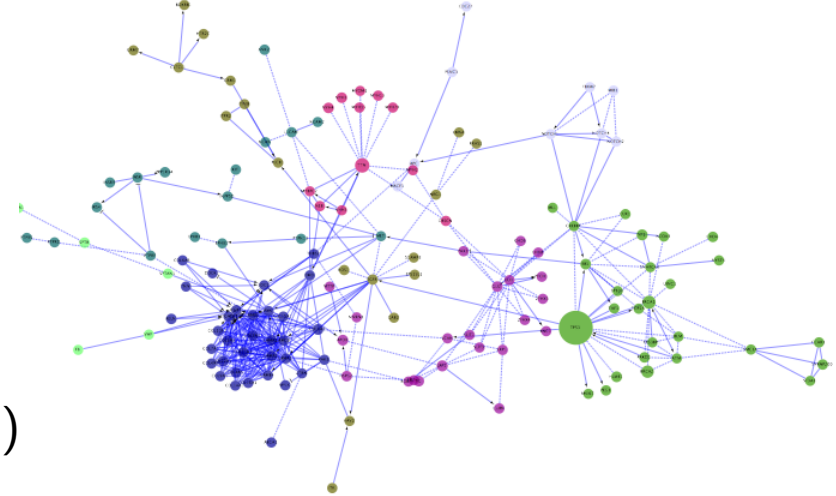
individual trajectories not predictable,

statistics of trajectories **predictable** as $N \to \infty$
(and dependent only on topology and rate distributions)

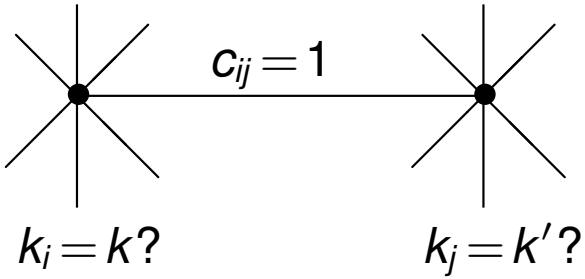
The research programme



Quantify protein network topology



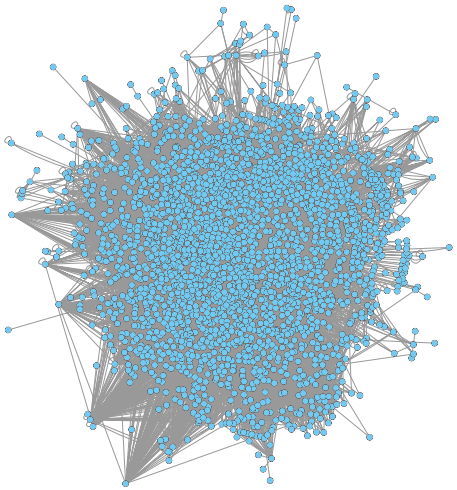
- Quantify topology beyond degrees:
  joint degree stats of connected nodes

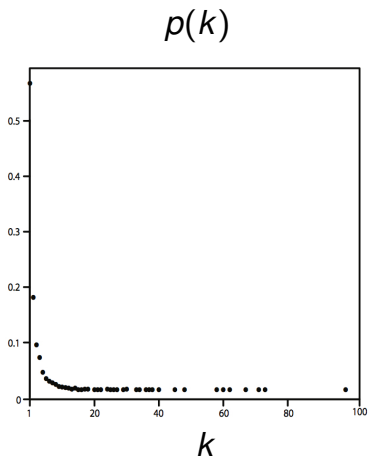


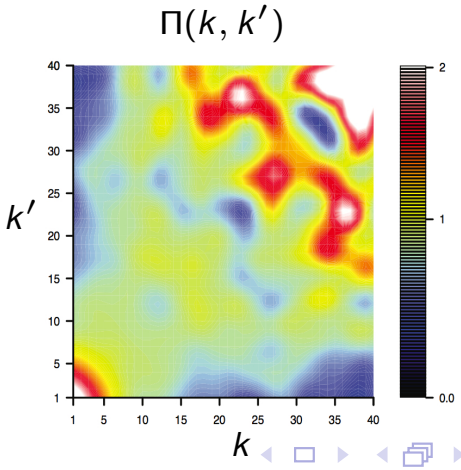
$W(k,k')$

- $W(k)=p(k)k/\langle k\rangle$: so use $\Pi(k,k')=W(k,k')/W(k)W(k')$

  $\Pi(k,k')\neq 1$:
  structural information in degree correlations

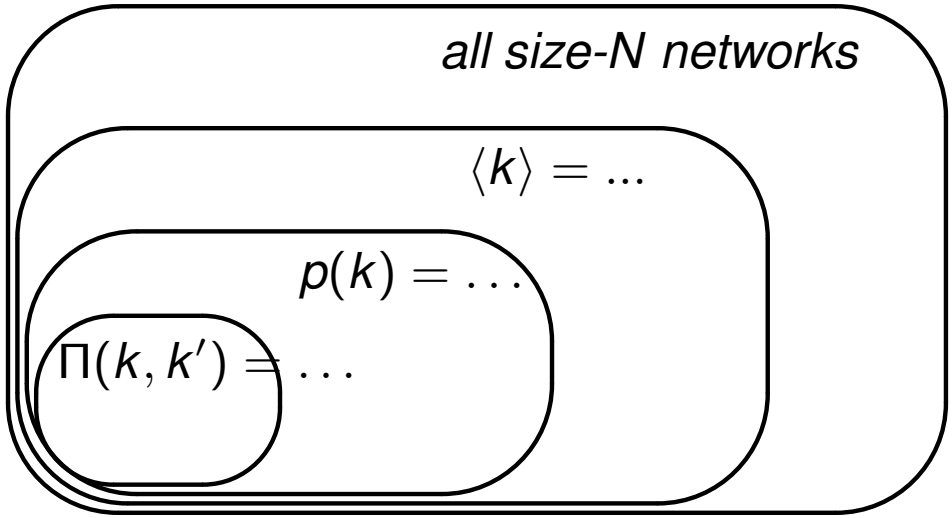






H sapiens
$N=9306$
$\langle k\rangle=7.53$

- network classification via increasingly detailed measurements



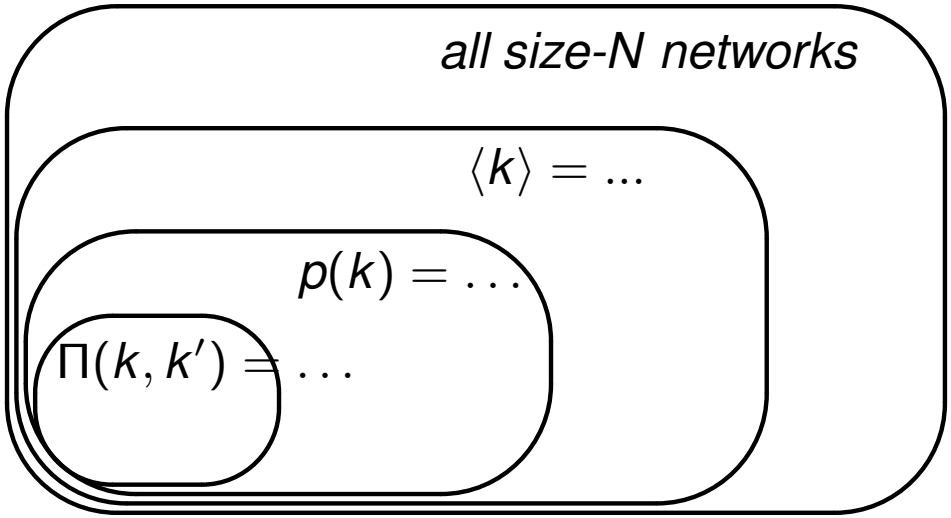
- Questions:
  - complexity: how many networks exist with given properties?
  - hypothesis testing: graphs with controlled features as null models (e.g. how 'special' are local modules?)
  - quantifying network dissimilarity

can al be done **analytically**
(information theory and statistical mechanics of complex graphs)

entropy & complexity in terms of $p(k)$ and $\Pi(k, k')$,
structural distances in terms of $p(k)$ and $\Pi(k, k')$

present focus: short loops

- network classification via increasingly detailed measurements



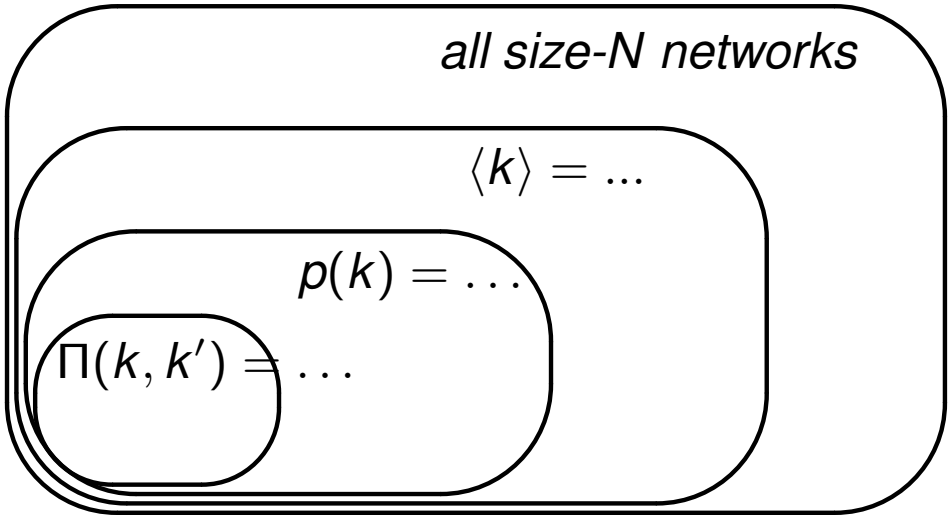
- Questions:
  - complexity: how many networks exist with given properties?
  - hypothesis testing: graphs with controlled features as null models (e.g. how 'special' are local modules?)
  - quantifying network dissimilarity

can al be done **analytically**
(information theory and statistical mechanics of complex graphs)

entropy & complexity in terms of $p(k)$ and $\Pi(k, k')$,
structural distances in terms of $p(k)$ and $\Pi(k, k')$

present focus: short loops

- network classification via increasingly detailed measurements



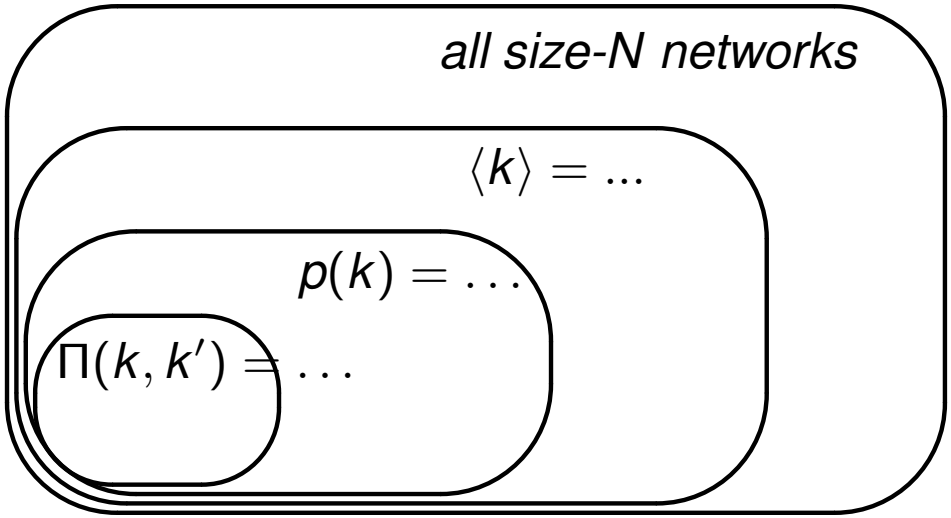
- Questions:
  - complexity: how many networks exist with given properties?
  - hypothesis testing: graphs with controlled features as null models (e.g. how 'special' are local modules?)
  - quantifying network dissimilarity

can al be done **analytically**
(information theory and statistical mechanics of complex graphs)

entropy & complexity in terms of $p(k)$ and $\Pi(k, k')$,
structural distances in terms of $p(k)$ and $\Pi(k, k')$

present focus: short loops

- network classification via increasingly detailed measurements



- Questions:
  - complexity: how many networks exist with given properties?
  - hypothesis testing: graphs with controlled features as null models (e.g. how 'special' are local modules?)
  - quantifying network dissimilarity

can al be done **analytically**
(information theory and statistical mechanics of complex graphs)

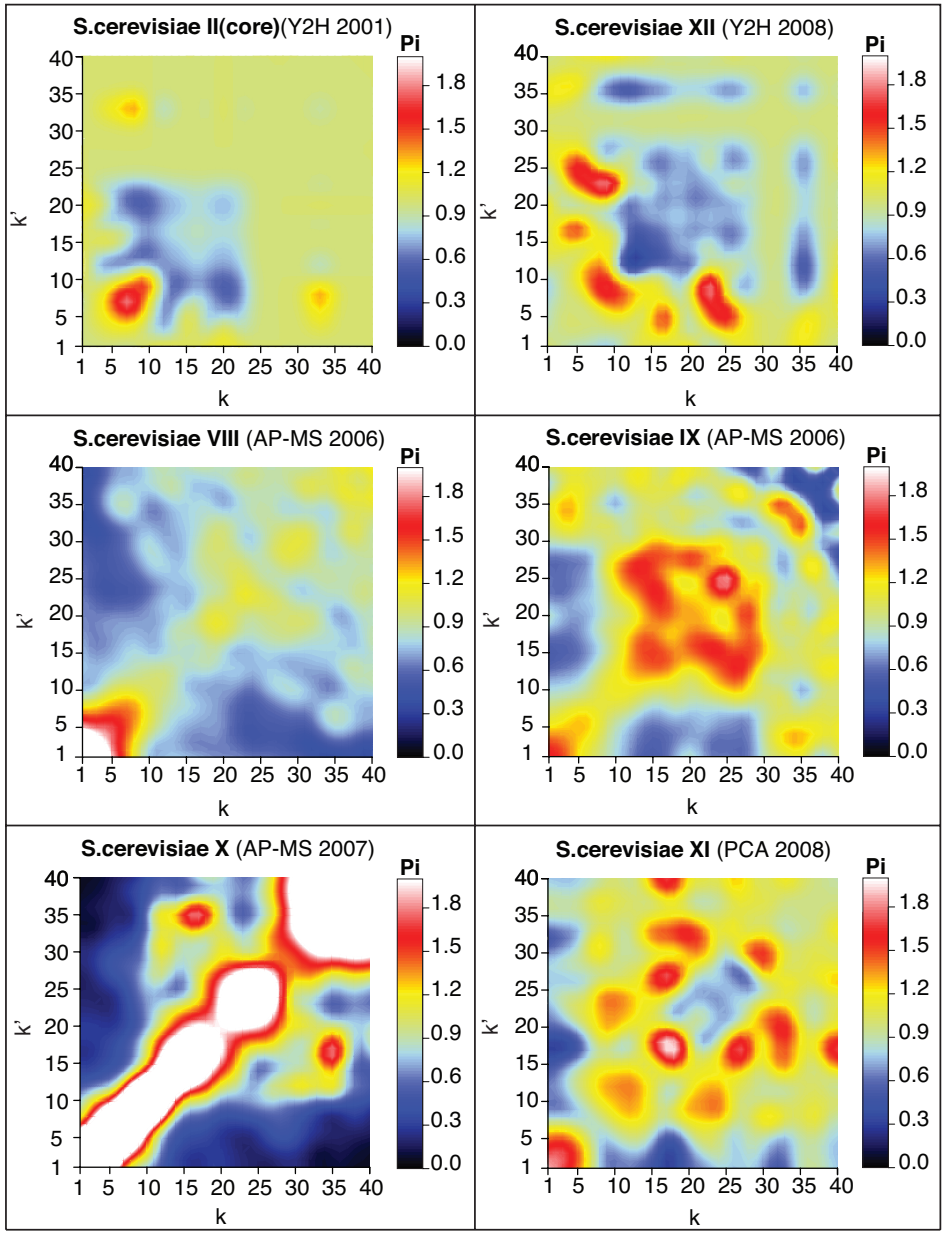
entropy & complexity in terms of $p(k)$ and $\Pi(k, k')$,
structural distances in terms of $p(k)$ and $\Pi(k, k')$

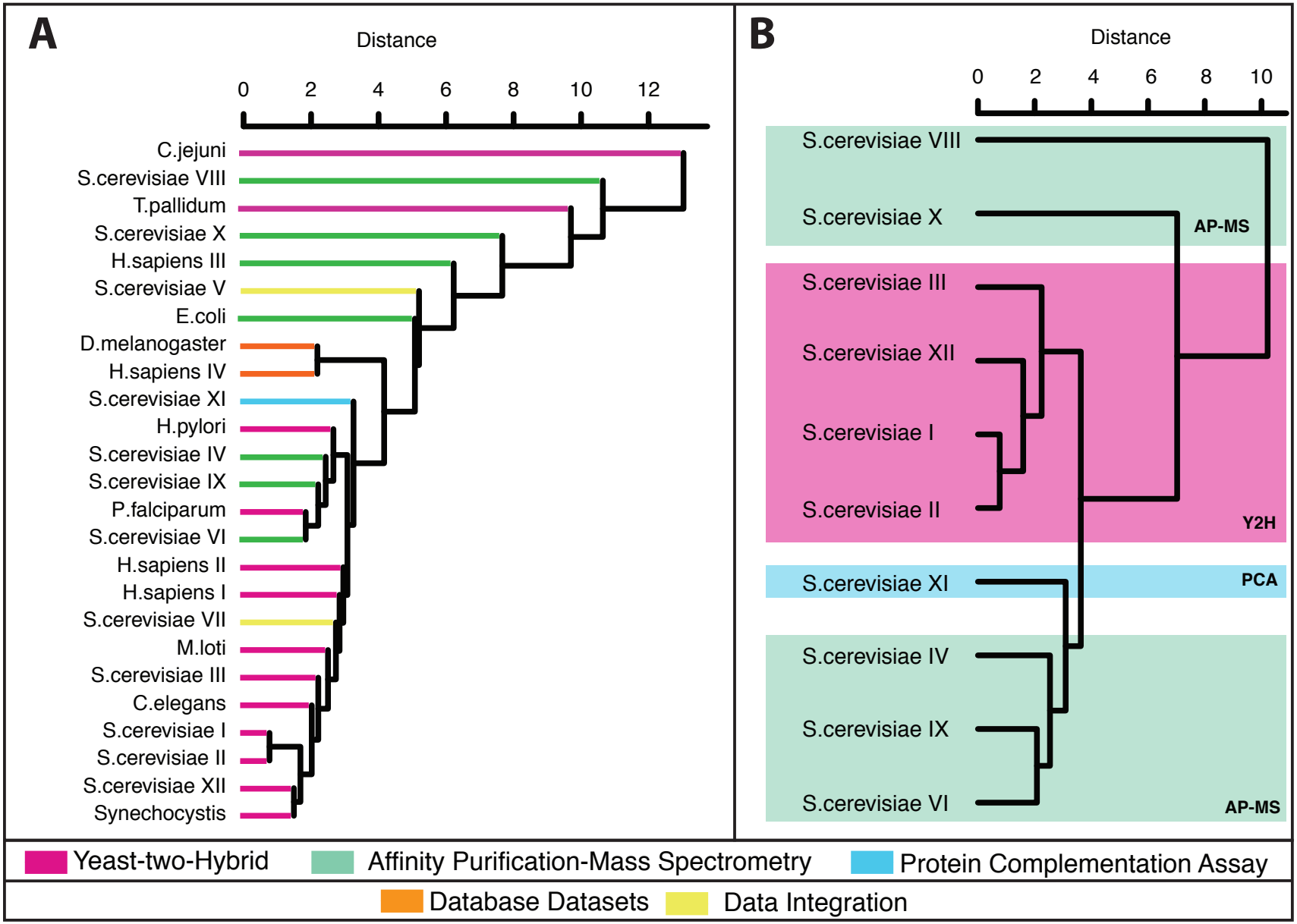
present focus: short loops

Can we trust protein interaction data?

e.g. yeast

$\Pi(k, k') \neq 1$:
degree-degree correlations





quantify sampling bias: see poster of Alessia Annibale

Signalling dynamics in the proteome

adapt techniques from many-particle physics
to do *many-particle biology*

- notation:

  $i = 1 \ldots N$ labels proteins
  x_i^α: concentr of protein i in state α
  x_{ij}: concentration of dimer $i \asymp j$

- events:

		rate:
complex formation:	$(i,\alpha) + (j,\beta) \to (i \asymp j)$	$k_{ij}^{\alpha\beta+} x_i^\alpha x_j^\beta$
complex dissociation:	$(i \asymp j) \to (i,\alpha) + (j,\beta)$	$k_{ij}^{\alpha\beta-} x_{ij}$
conformation change:	$(i,\alpha) \to (i,\beta)$	$\lambda_i^{\alpha\beta} x_i^\alpha$
protein degradation:	$(i,\alpha) \to \emptyset$	$\gamma_i^\alpha x_i^\alpha$
protein synthesis:	$\emptyset \to (i,\alpha)$	θ_i^α

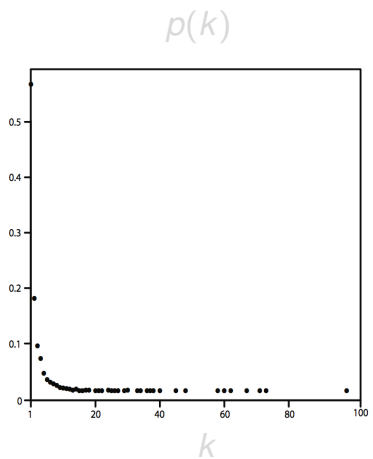
- reaction eqns:

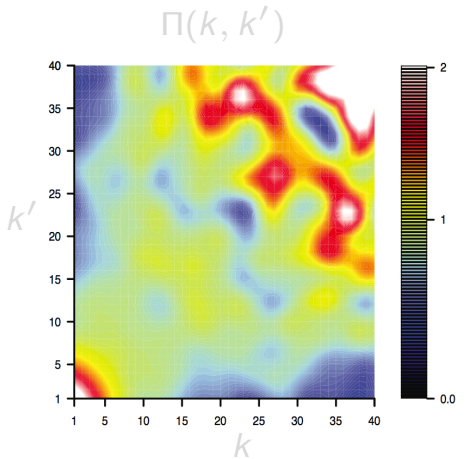
$$\frac{\mathrm{d}}{\mathrm{d}t}x_i^\alpha = \sum_j c_{ij} \overbrace{\sum_\beta [k_{ij}^{\alpha\beta-} x_{ij} - k_{ij}^{\alpha\beta+} x_i^\alpha x_j^\beta]}^{\textit{complex formation \& dissociation}} + \overbrace{\sum_\beta [\lambda_i^{\beta\alpha} x_i^\beta - \lambda_i^{\alpha\beta} x_i^\alpha]}^{\textit{post-transl modification}} + \overbrace{\theta_i^\alpha}^{\textit{synthesis}} - \overbrace{\gamma_i^\alpha x_i^\alpha}^{\textit{decay}}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}x_{ij} = c_{ij} \sum_{\alpha\beta} [k_{ij}^{\alpha\beta+} x_i^\alpha x_j^\beta - k_{ij}^{\alpha\beta-} x_{ij}]$$

- tailored random interaction network,

  prescribed degrees $p(k)$,
  and degree correlations $W(k, k')$





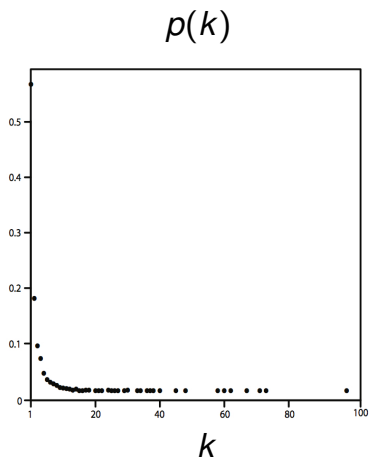
- reaction eqns:

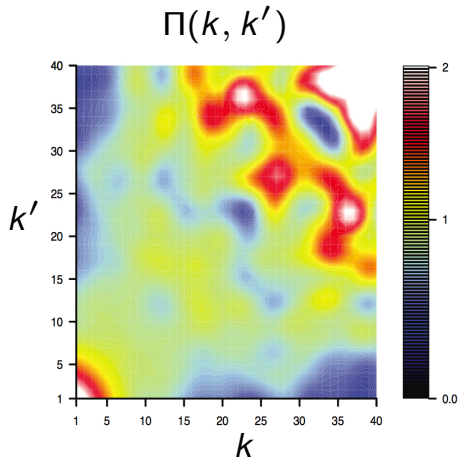
$$\frac{\mathrm{d}}{\mathrm{d}t}x_i^\alpha = \sum_j c_{ij} \overbrace{\sum_\beta [k_{ij}^{\alpha\beta-} x_{ij} - k_{ij}^{\alpha\beta+} x_i^\alpha x_j^\beta]}^{\textit{complex formation \& dissociation}} + \overbrace{\sum_\beta [\lambda_i^{\beta\alpha} x_i^\beta - \lambda_i^{\alpha\beta} x_i^\alpha]}^{\textit{post-transl modification}} + \overbrace{\theta_i^\alpha}^{\textit{synthesis}} - \overbrace{\gamma_i^\alpha x_i^\alpha}^{\textit{decay}}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}x_{ij} = c_{ij} \sum_{\alpha\beta} [k_{ij}^{\alpha\beta+} x_i^\alpha x_j^\beta - k_{ij}^{\alpha\beta-} x_{ij}]$$

- tailored random interaction network,

  prescribed degrees $p(k)$,
  and degree correlations $W(k, k')$





preparation:

- solve equations for $\{x_{ij}\}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}x_i^\alpha(t) = F_i^\alpha[t,\{x\}]$$

$$F_i^\alpha[t,\{x\}] = \theta_i^\alpha - \gamma_i^\alpha x_i^\alpha + \sum_\beta [\lambda_i^{\beta\alpha} x_i^\beta - \lambda_i^{\alpha\beta} x_i^\alpha] + \sum_j c_{ij} \int \mathrm{d}s \sum_{\rho\lambda} \underbrace{W_{\alpha;\rho\lambda}(t-s|\mathbf{k}_{ij})\, x_i^\rho(t-s) x_j^\lambda(t-s)}$$

effective delayed free-protein interaction

$$W_{\alpha;\rho\lambda}(\tau|\mathbf{k}) = k^{\rho\lambda+}\Big[\sum_\beta k^{\alpha\beta-}\theta[\tau]\mathrm{e}^{-k^-\tau} - \delta_{\alpha\rho}\delta(\tau)\Big]$$

closed equations for unbound protein concentrations,
price paid: equations are nonlocal in time

generating functional analysis:

calculate correlations, response functions etc ...
without solving reaction equations!

- generating functional:

$$Z[\boldsymbol{\psi}] = \int \Big[\prod_{i\alpha t} \mathrm{d}x_i^\alpha(t)\Big]\; \mathrm{e}^{\mathrm{i}\sum_{i\alpha}\int \mathrm{d}t\;\psi_i^\alpha(t)x_i^\alpha(t)} \prod_{i\alpha t}\delta\Big[x_i^\alpha(t+\mathrm{d}t)-x_i^\alpha(t)-F_i^\alpha[t,\{x\}]\mathrm{d}t\Big]$$

path integral over all possible concentration trajectories in time

$$x_i^\alpha(t) = -\mathrm{i}\lim_{\boldsymbol{\psi}\to 0}\frac{\delta Z[\boldsymbol{\psi}]}{\delta\psi_i^\alpha(t)} \qquad x_i^\alpha(t)x_j^\beta(t') = -\lim_{\boldsymbol{\psi}\to 0}\frac{\delta^2 Z[\boldsymbol{\psi}]}{\delta\psi_i^\alpha(t)\delta\psi_j^\beta(t')}$$

- for $N\to\infty$ (large systems),

  $Z[\psi]$ will no longer depend on network details, just on statistics
  so calculate instead its average over all tailored networks

generating functional analysis:

calculate correlations, response functions etc ...
without solving reaction equations!

- generating functional:

$$Z[\psi] = \int \Big[\prod_{i\alpha t} \mathrm{d}x_i^\alpha(t)\Big]\; \mathrm{e}^{\mathrm{i}\sum_{i\alpha}\int \mathrm{d}t\;\psi_i^\alpha(t)x_i^\alpha(t)} \prod_{i\alpha t}\delta\Big[x_i^\alpha(t+\mathrm{d}t)-x_i^\alpha(t)-F_i^\alpha[t,\{x\}]\mathrm{d}t\Big]$$

path integral over all possible concentration trajectories in time

$$x_i^\alpha(t) = -\mathrm{i}\lim_{\psi\to 0}\frac{\delta Z[\psi]}{\delta\psi_i^\alpha(t)} \qquad x_i^\alpha(t)x_j^\beta(t') = -\lim_{\psi\to 0}\frac{\delta^2 Z[\psi]}{\delta\psi_i^\alpha(t)\delta\psi_j^\beta(t')}$$

- for $N\to\infty$ (large systems),

  $Z[\psi]$ will no longer depend on network details, just on statistics
  so calculate instead its average over all tailored networks

after further calculations ...
(path integral techniques, saddle-point integration, etc)

- key macroscopic quantities:

$$D[\{x\}|\{y\}] = \frac{1}{N}\sum_j \left\langle \delta[\{x\} - \{x_j\}] \right\rangle \Big|_{\theta_j^\alpha(t)\to\theta_j^\alpha(t)+y_\alpha(t)\ \forall\alpha}$$

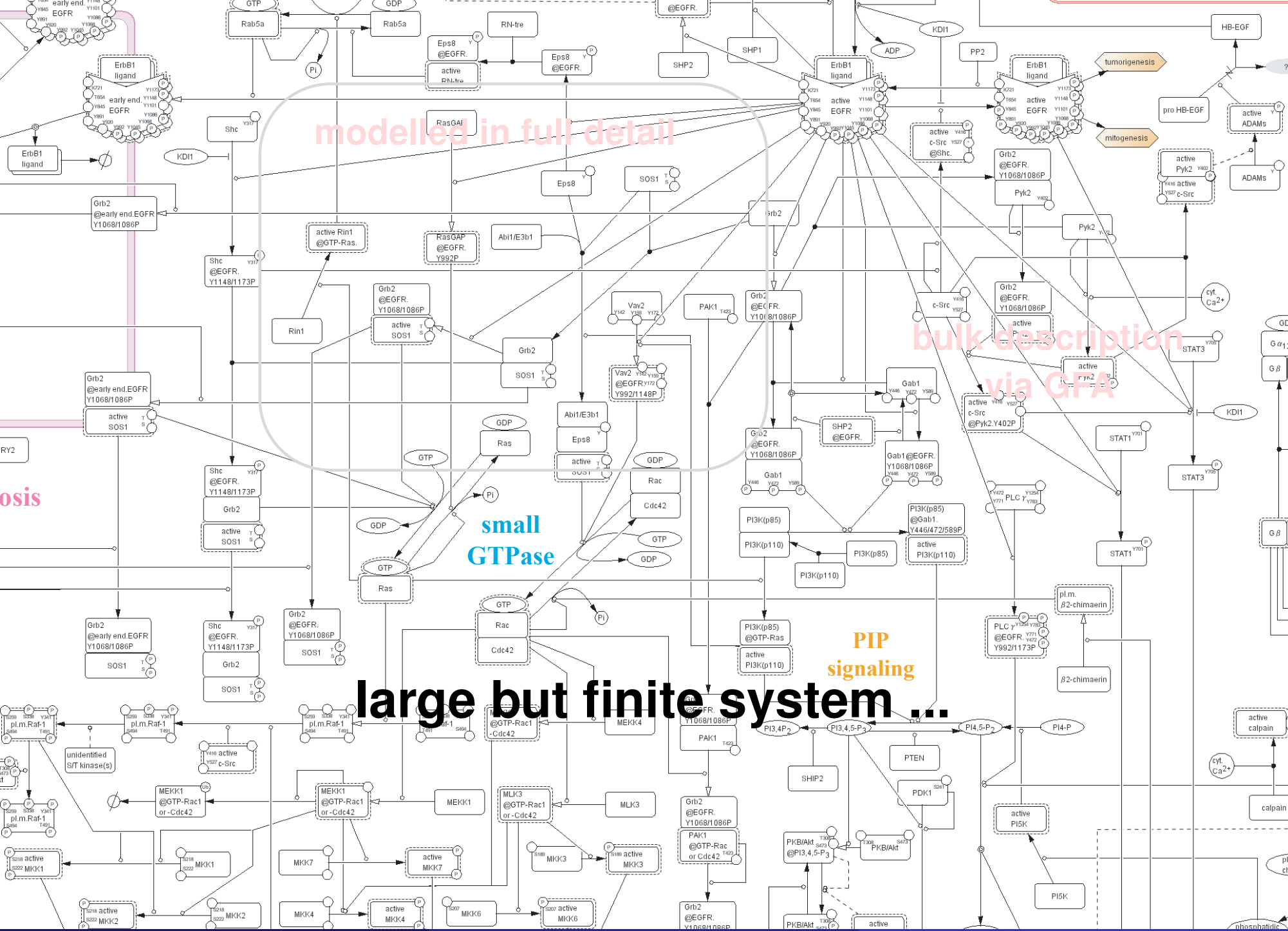
$$W[\{x\}|\{y\}] = \frac{1}{N}\sum_j \left\langle \delta[\{x\} - \{x_j\}] \right\rangle \Big|_{k_j\to k_j-1,\ \ \theta_j^\alpha(t)\to\theta_j^\alpha(t)+y_\alpha(t)\ \forall\alpha}$$

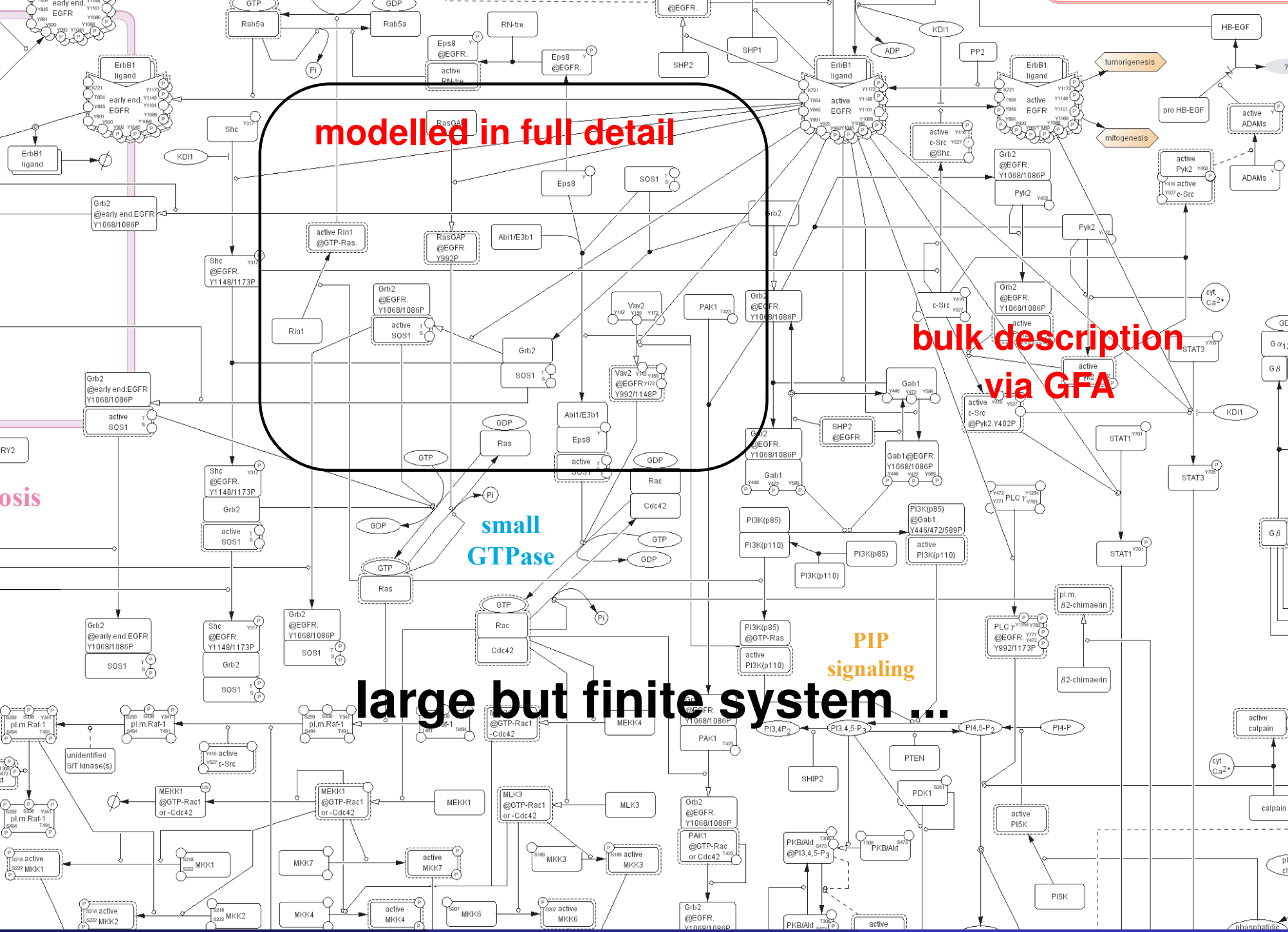
$\{x\}$ = trajectories $x_\alpha(t)$ for all α
$\{y\}$ = time dep production rate perturbations $y_\alpha(t)$ for all α

- macroscopic equations:

$$W = \mathcal{G}_1[W], \quad D = \mathcal{G}_2[W], \qquad \mathcal{G}_{1,2}: \ \textit{complicated but exact formulas}$$

equations interpreted in terms of
response to single-node perturbations


modelled in full detail
bulk description via GFA
small GTPase
PIP signaling
large but finite system ...


modelled in full detail
bulk description
via GFA
small
GTPase
PIP
signaling
large but finite system ...

Summary

- signalling in large protein interaction networks
  can be studied by adapting methods from many-particle physics
- requires systematic characterisation of network topologies
  (many spin-offs)
- macroscopic theory in terms of W and D
  (concentration trajectory response to time-dep perturbation)

- Ongoing:
  - solving equations for W and D
  - phase diagrams, analysis of instabilities
  - remove bias from protein network data
  - include short loops in network characterisation

- Next:
  - inclusion of higher order complexes
  - integration with 'small module' reaction equations
  - connections with experiment, verifiable predictions

Summary

- signalling in large protein interaction networks
  can be studied by adapting methods from many-particle physics
- requires systematic characterisation of network topologies
  (many spin-offs)
- macroscopic theory in terms of W and D
  (concentration trajectory response to time-dep perturbation)

- Ongoing:
  - solving equations for W and D
  - phase diagrams, analysis of instabilities
  - remove bias from protein network data
  - include short loops in network characterisation

- Next:
  - inclusion of higher order complexes
  - integration with 'small module' reaction equations
  - connections with experiment, verifiable predictions

Summary

- signalling in large protein interaction networks
  can be studied by adapting methods from many-particle physics
- requires systematic characterisation of network topologies
  (many spin-offs)
- macroscopic theory in terms of W and D
  (concentration trajectory response to time-dep perturbation)
- Ongoing:
  - solving equations for W and D
  - phase diagrams, analysis of instabilities
  - remove bias from protein network data
  - include short loops in network characterisation
- Next:
  - inclusion of higher order complexes
  - integration with 'small module' reaction equations
  - connections with experiment, verifiable predictions

Summary

- signalling in large protein interaction networks
  can be studied by adapting methods from many-particle physics
- requires systematic characterisation of network topologies
  (many spin-offs)
- macroscopic theory in terms of W and D
  (concentration trajectory response to time-dep perturbation)

- Ongoing:
  - solving equations for W and D
  - phase diagrams, analysis of instabilities
  - remove bias from protein network data
  - include short loops in network characterisation

- Next:
  - inclusion of higher order complexes
  - integration with 'small module' reaction equations
  - connections with experiment, verifiable predictions

Summary

- signalling in large protein interaction networks
  can be studied by adapting methods from many-particle physics
- requires systematic characterisation of network topologies
  (many spin-offs)
- macroscopic theory in terms of W and D
  (concentration trajectory response to time-dep perturbation)

- Ongoing:
  - solving equations for W and D
  - phase diagrams, analysis of instabilities
  - remove bias from protein network data
  - include short loops in network characterisation

- Next:
  - inclusion of higher order complexes
  - integration with 'small module' reaction equations
  - connections with experiment, verifiable predictions

Summary

- signalling in large protein interaction networks
  can be studied by adapting methods from many-particle physics
- requires systematic characterisation of network topologies
  (many spin-offs)
- macroscopic theory in terms of W and D
  (concentration trajectory response to time-dep perturbation)

- Ongoing:
  - solving equations for W and D
  - phase diagrams, analysis of instabilities
  - remove bias from protein network data
  - include short loops in network characterisation

- Next:
  - inclusion of higher order complexes
  - integration with 'small module' reaction equations
  - connections with experiment, verifiable predictions

thanks to:

Alessia Annibale, Nima Shayeghi, Ekaterina Roberts, James Barett
Luis Fernandes, Franca Fraternali, Jens Kleinjung
Tony Ng

EPSRC
Engineering and Physical Sciences Research Council

BBSRC
bioscience for the future