

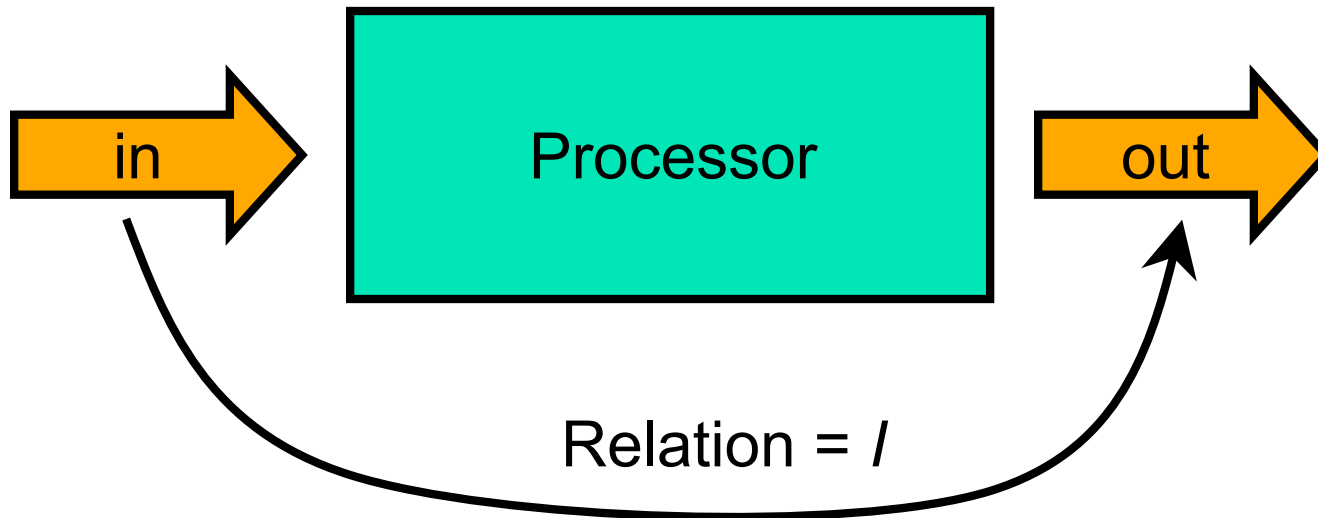
Information theory in systems biology



Ilya Nemenman
(JCSB, Columbia)

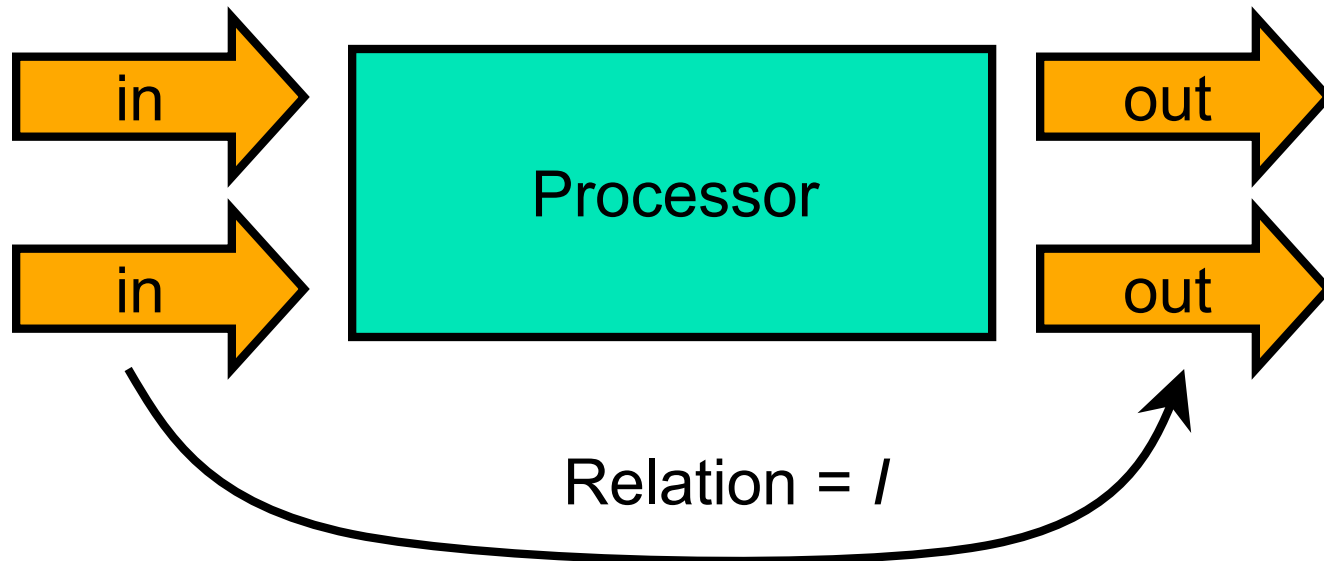


Studying signal transduction



How faithful is the output to the input?
How does it represent input?

Studying signal transduction

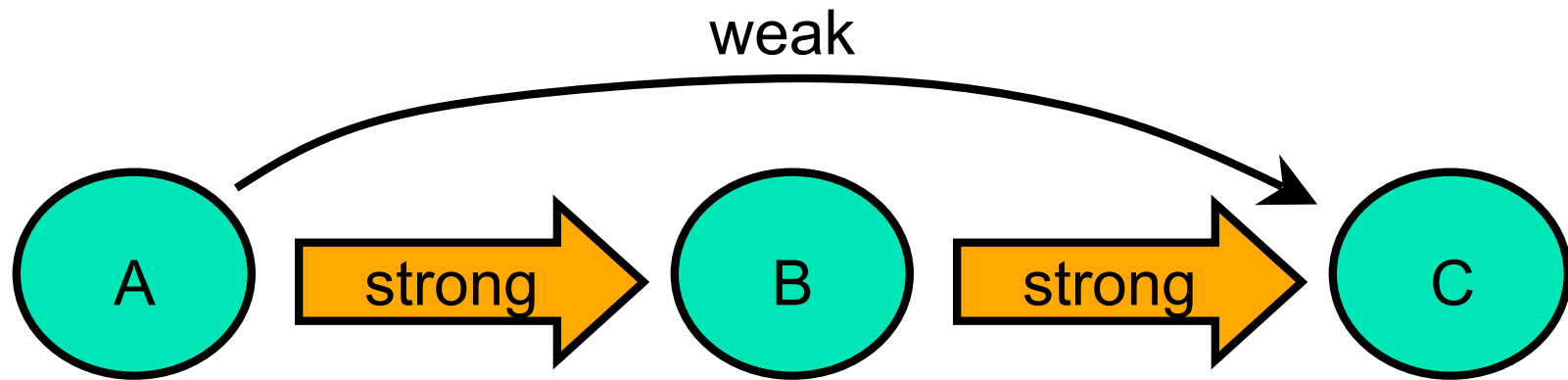


How faithful is the output to the input?
How does it represent input?
Synergies for multiple inputs/outputs?

But what is /?



Studying interaction models



$$I(A, C) \leq \min[I(A, B), I(B, C)]$$

But what is I ?



Correlation coefficients

$$\rho(x, x^2) = 0$$

linear

$$\rho(f(x), g(y)) \neq \rho(x, y)$$

not invariant



One-to-one transformations of microarray expression data completely destroys the ranking of correlations. Even sign of correlations may change.



Entropy (unique measure of randomness, in bits)

$$S[X] = - \sum_x p_x \log p_x = - \langle \log p_x \rangle$$

$$0 \leq S[X] \leq \log K$$

$$N(x_0, \sigma^2) \Rightarrow S[X] = \frac{1}{2} \log(2\pi e \sigma^2)$$

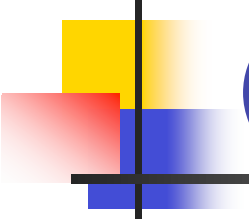


Kullback-Leibler divergence

$$D_{KL}[P \parallel Q] = \sum_x p_x \log \frac{p_x}{q_x}$$

$$0 \leq D_{KL}$$

How easy it is to mistake P for Q ?
Coding losses (in bits).



Mutual Information (interactions, shared data)

$$\begin{aligned} I[X; Y] &= \left\langle \log \frac{p_{xy}}{p_x p_y} \right\rangle = D_{KL}[p_{xy} \parallel p_x p_y] \\ &= S[X] + S[Y] - S[X, Y] \end{aligned}$$

$$0 \leq I[X; Y] \leq \min(S[X], S[Y])$$

$$N(\vec{x}_0, \Sigma) \Rightarrow I[X; Y] = -\frac{1}{2} \log(1 - \rho^2)$$



Why MI?

- All dependencies (zero *iff* joint probabilities factorize)
- Reparameterization invariant
- Metric-independent measure of “how related”



MI as MaxEnt

Find least constrained (highest entropy)
approximation q to p_{xy} , s.t.

$$p_x = q_x$$

$$p_y = q_y$$



$$q_{xy} = \frac{1}{Z} \exp[-\varphi_x - \varphi_y] = p_x p_y$$

$$I[X; Y] = D_{KL}[P \parallel Q]$$



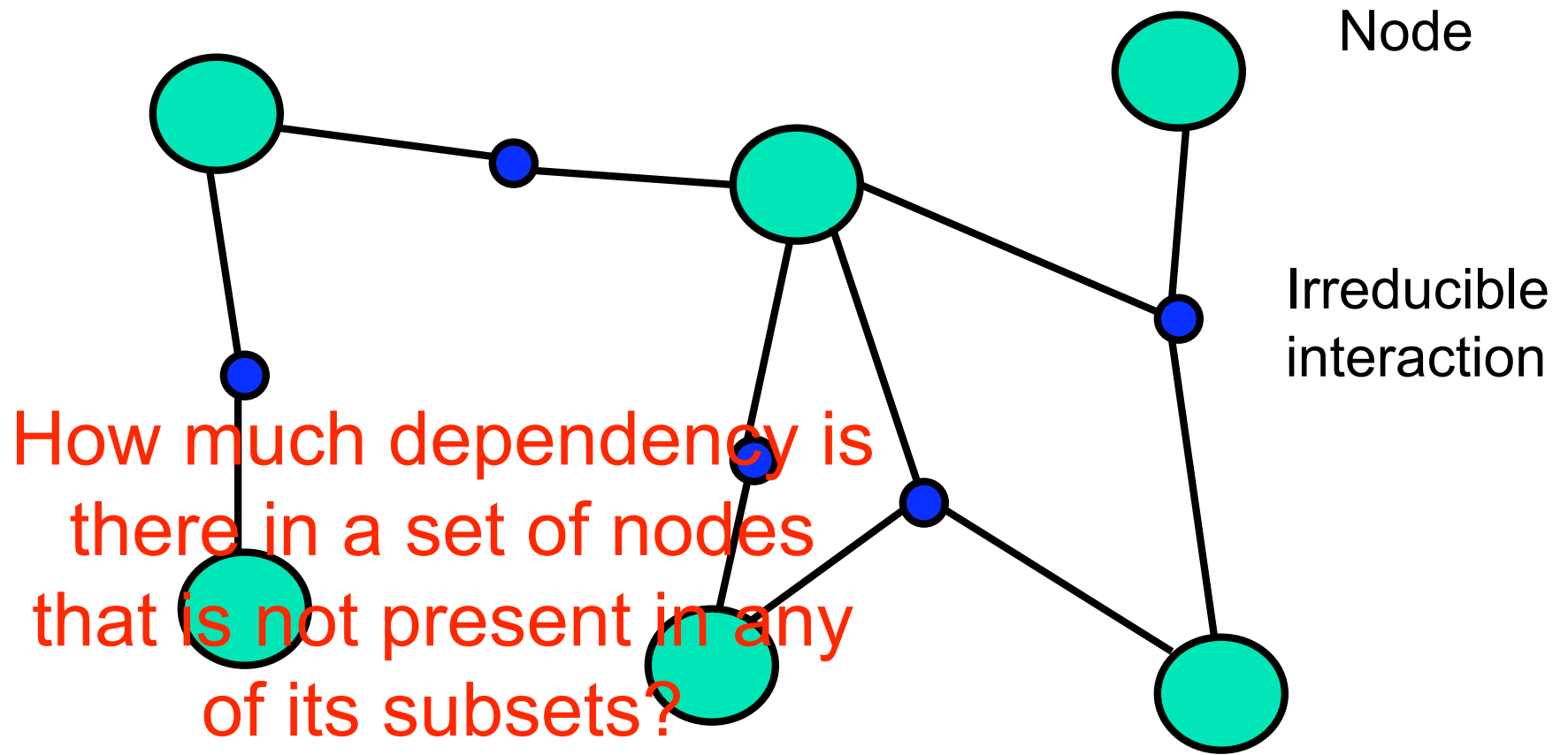
Higher order dependencies

$$I_{XYZ} = \left\langle \log \frac{p_{xyz}}{p_x p_y p_z} \right\rangle$$

(Axiomatically) Amount of *all* dependencies
(in bits) among variables.

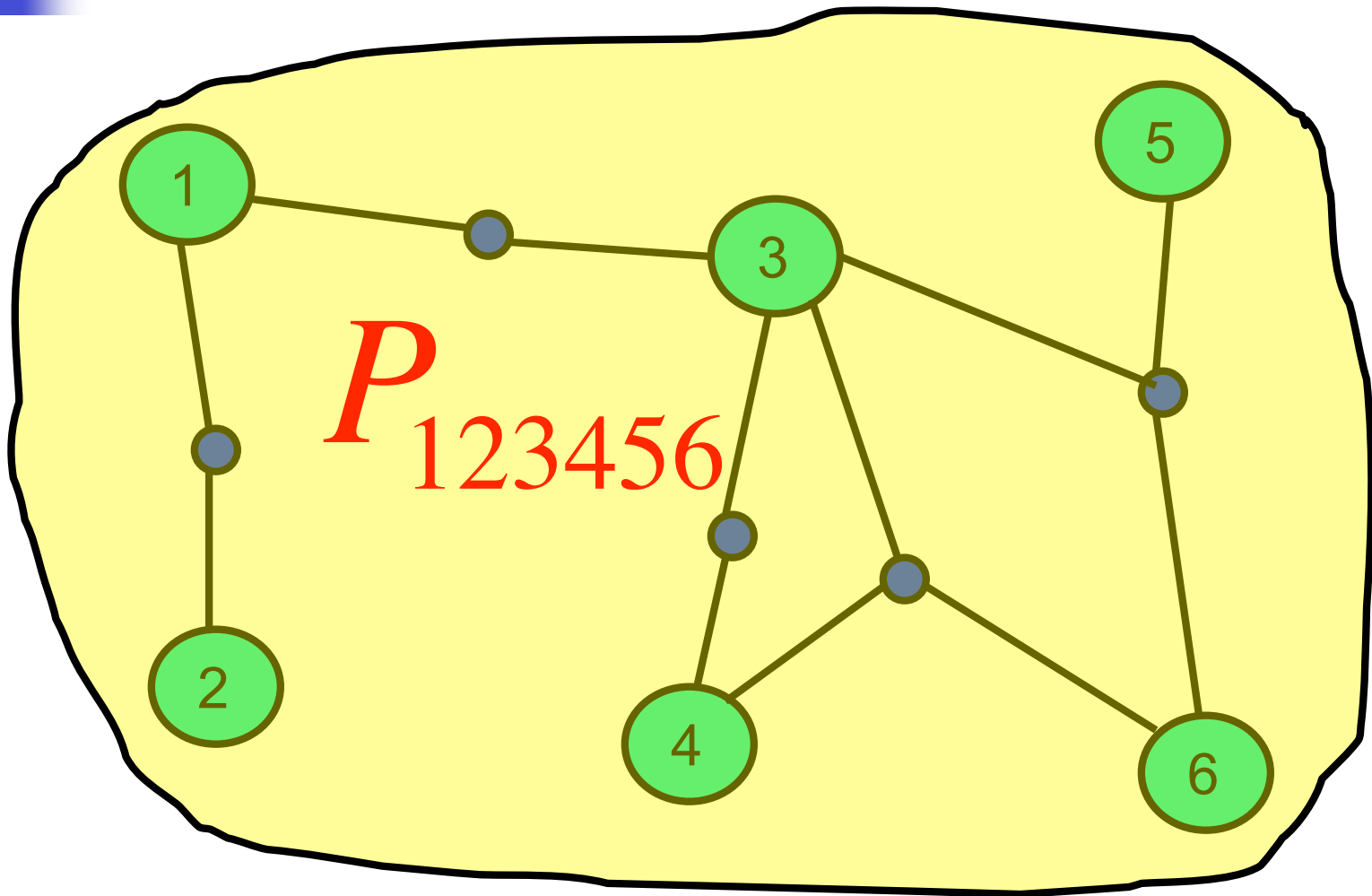
(Nemenman and Tishby 2005)

Higher order irreducible dependencies

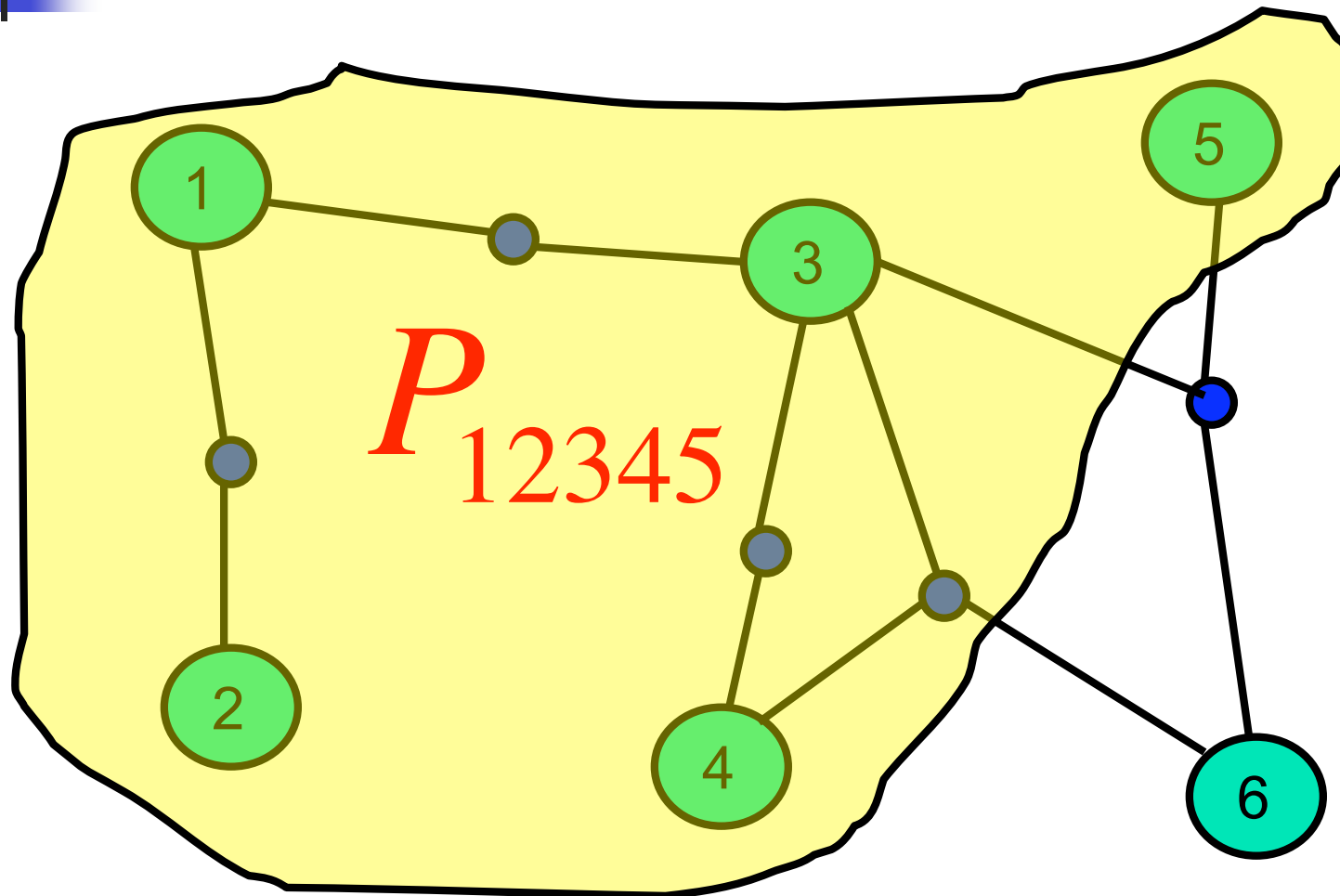


(Schneidman et al. 2003, Nemenman 2004)

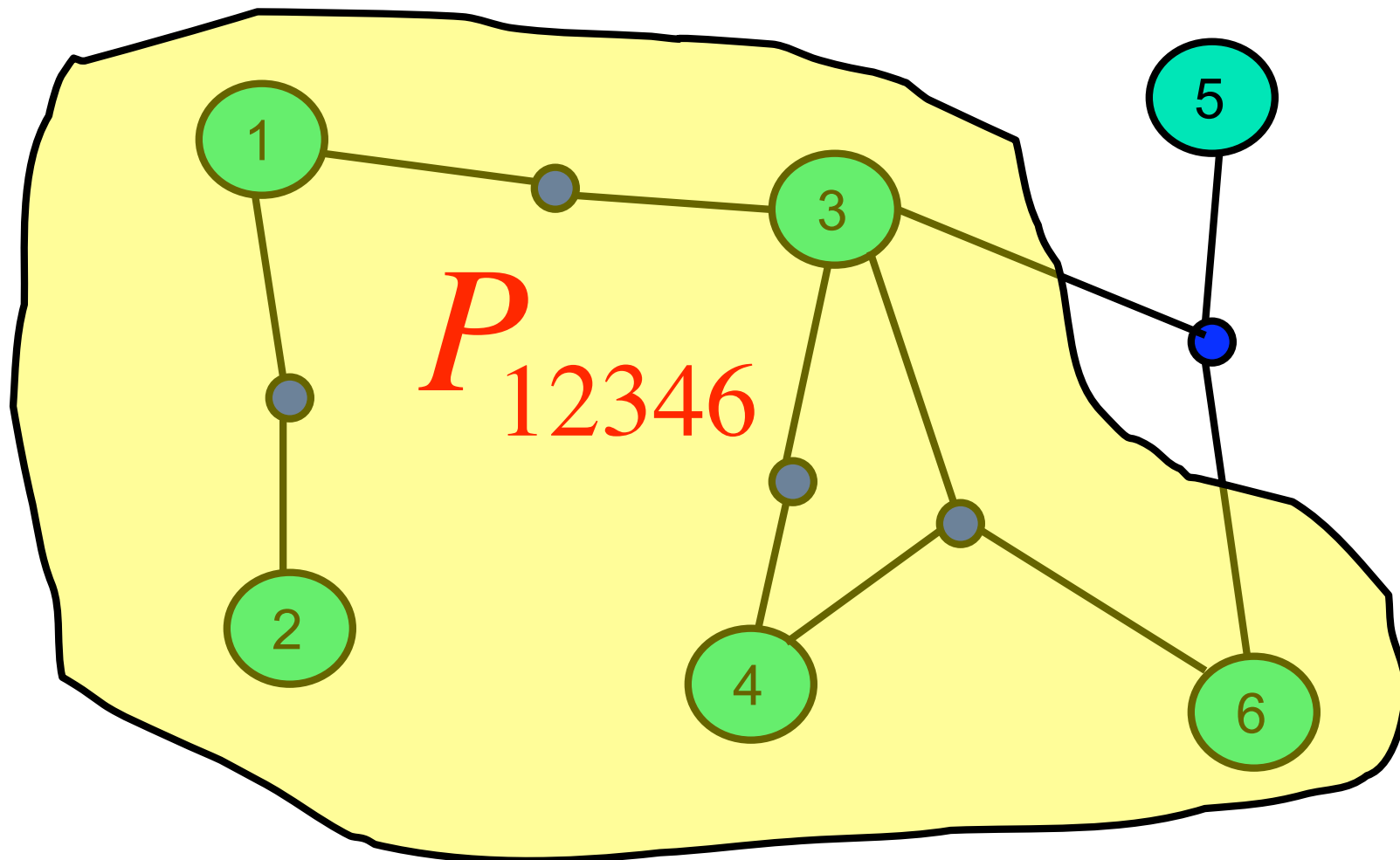
MaxEnt approximations



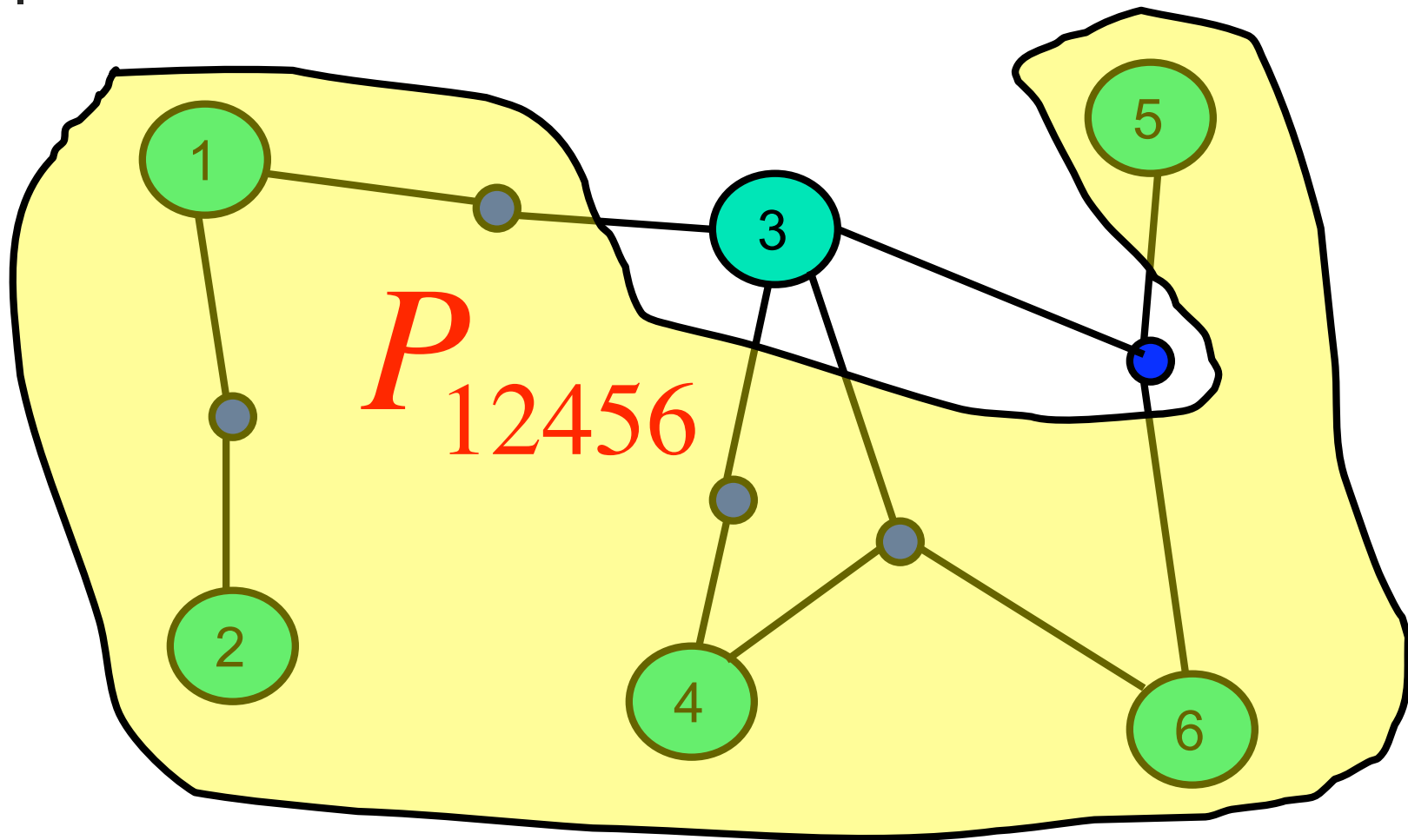
MaxEnt approximations



MaxEnt approximations



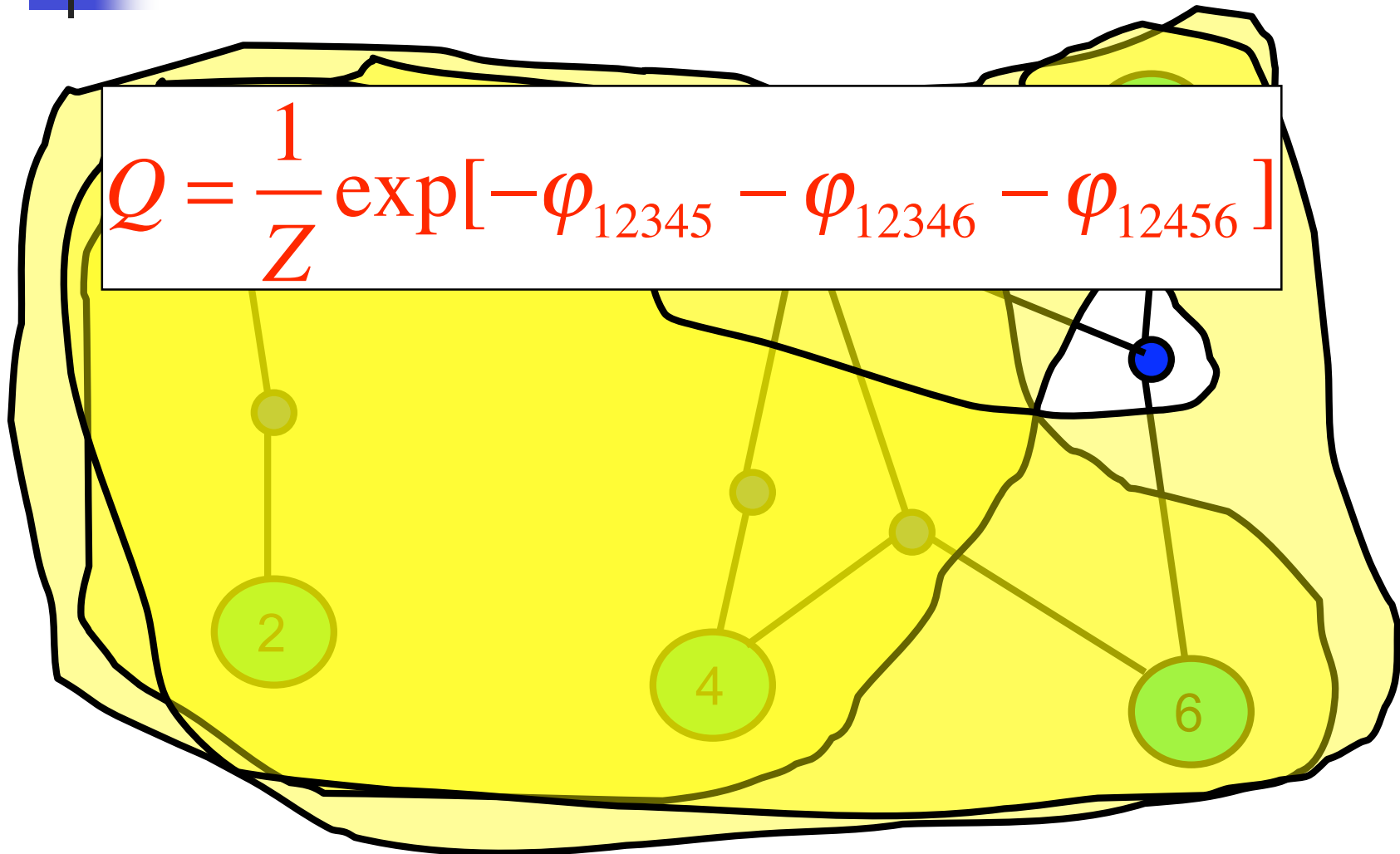
MaxEnt approximations





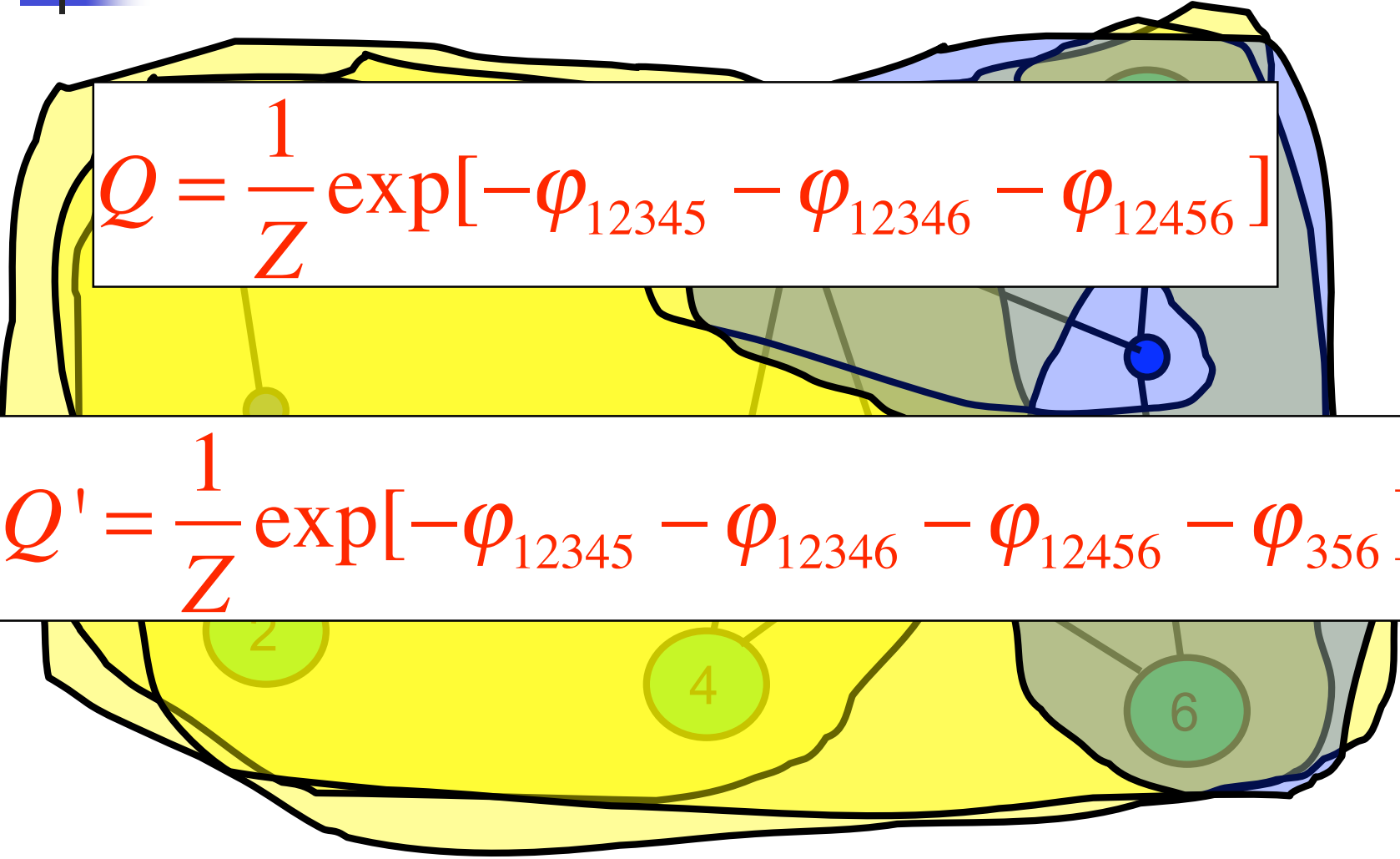
MaxEnt approximations

$$Q = \frac{1}{Z} \exp[-\varphi_{12345} - \varphi_{12346} - \varphi_{12456}]$$





MaxEnt approximations


$$Q = \frac{1}{Z} \exp[-\varphi_{12345} - \varphi_{12346} - \varphi_{12456}]$$

$$Q' = \frac{1}{Z} \exp[-\varphi_{12345} - \varphi_{12346} - \varphi_{12456} - \varphi_{356}]$$



MaxEnt approximations

$$I'_{356} = D_{KL}[Q' \parallel Q]$$

$$0 \leq I'_{356} \leq I_{356}$$

$I'_{356} > 0 \Rightarrow$ Irreducible interaction present



MaxEnt factorization of PDFs

$$P(x_1, \dots, x_M) =$$
$$= \exp \left[- \sum_i \varphi_i(x_i) - \sum_{ij} \varphi_{ij}(x_i, x_j) - \sum_{ijk} \varphi_{ijk}(x_i, x_j, x_k) - \dots \right]$$

- N -particle potentials
- Spin models (for discrete variables)
- Random lattices
- Message passing
- Markov Networks



Why is IT not common in statistics?

Maximum likelihood estimation:

$$S_{ML} = -\sum_i \frac{n_i}{N} \log \frac{n_i}{N}$$

$$\langle S_{ML} \rangle \leq -\sum_i \frac{\langle n_i \rangle}{N} \log \frac{\langle n_i \rangle}{N} = S$$

$$\text{bias} \propto -\frac{K_{eff}}{N} \propto -\frac{2^S}{N} \gg (\text{variance})^{1/2} \propto \frac{1}{\sqrt{N}}$$

Similarly, MI is often overestimated due to sampling irregularities.



Universally correct smoothing

$$S \propto \log K_{eff} \ll \log N \quad (\text{often not enough})$$

Incorrect smoothing = over- or underestimation.

- Bioinformatics, systems biology, ecology
- Computational linguistics, mathematical finances
- Dynamical systems

For estimation of entropy at $K / N \leq 1$ see:

Grassberger 1989, 2003, Antos and Kontoyiannins 2002, Wyner and Foster 2003, Batu et al. 2002, Paninski 2003, Panzeri and Treves 1996, Strong et al. 1998



No universal estimator for $S > \log N$

But there is hope (Ma, 1981):

For uniform K -bin distribution the first coincidence occurs for

$$N_c \sim \sqrt{K} = \sqrt{2^S}$$

$$S \sim 2 \log N_c$$

Can make estimates in the nonasymptotic regime!

Can this be extended to nonuniform cases?

- Assumptions needed (won't work always)
- Estimate entropies without estimating distributions.

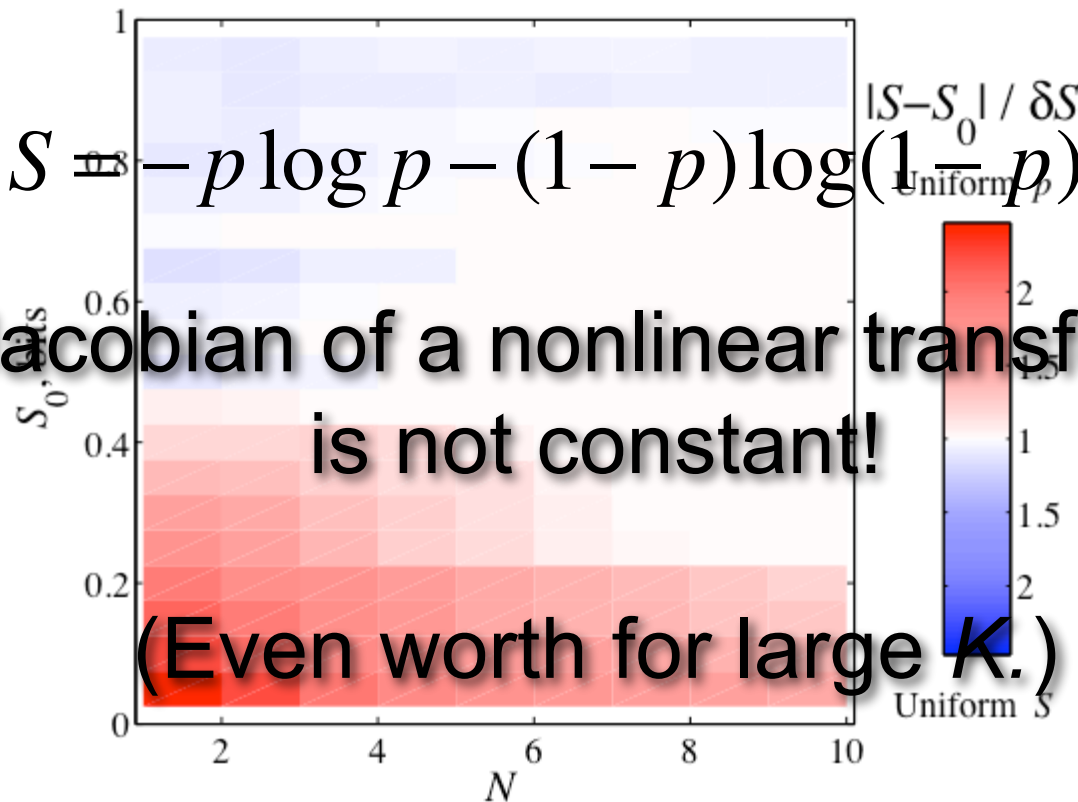
What is unknown?

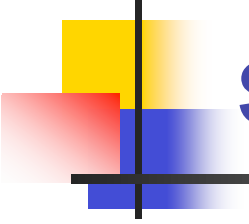
Binomial distribution with the prior uniform on p or S :

$$S = -p \log p - (1-p) \log(1-p)$$

The Jacobian of a nonlinear transformation
is not constant!

(Even worth for large K .)





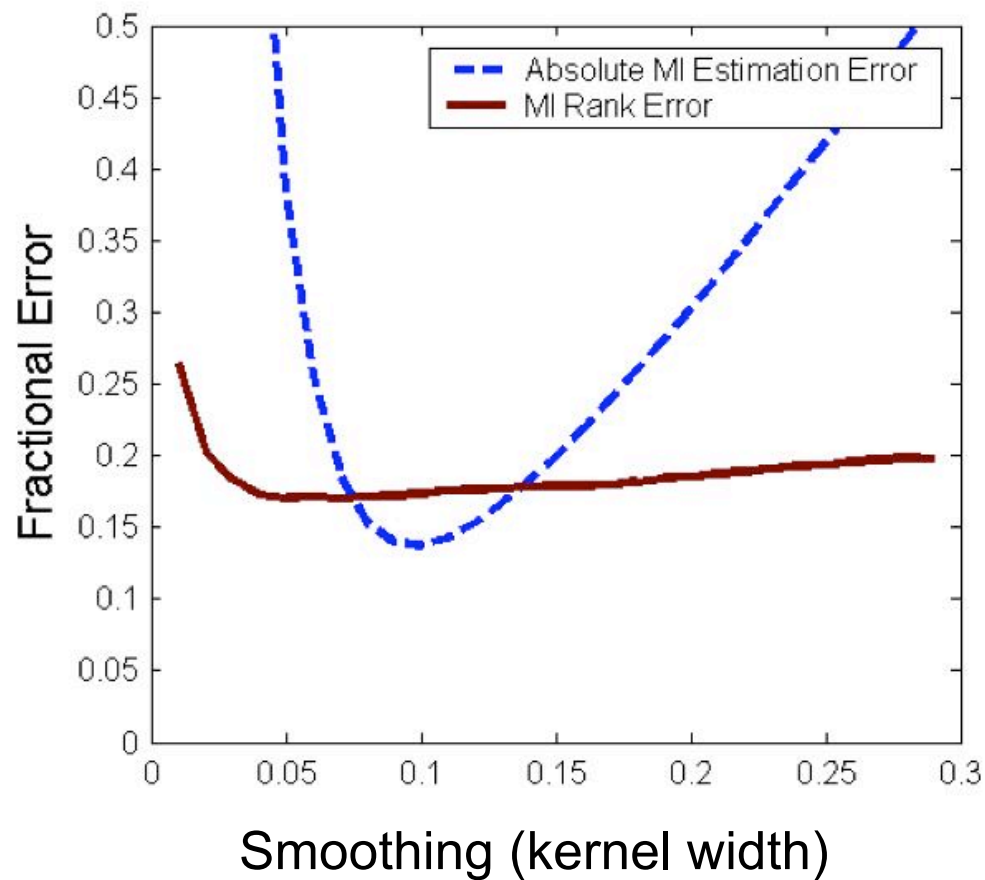
One possible uniformization strategy for S (NSB)

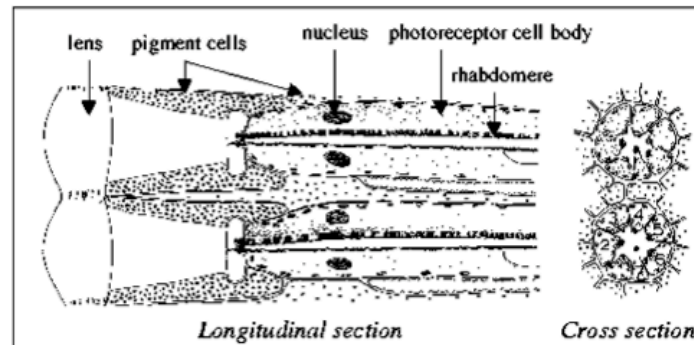
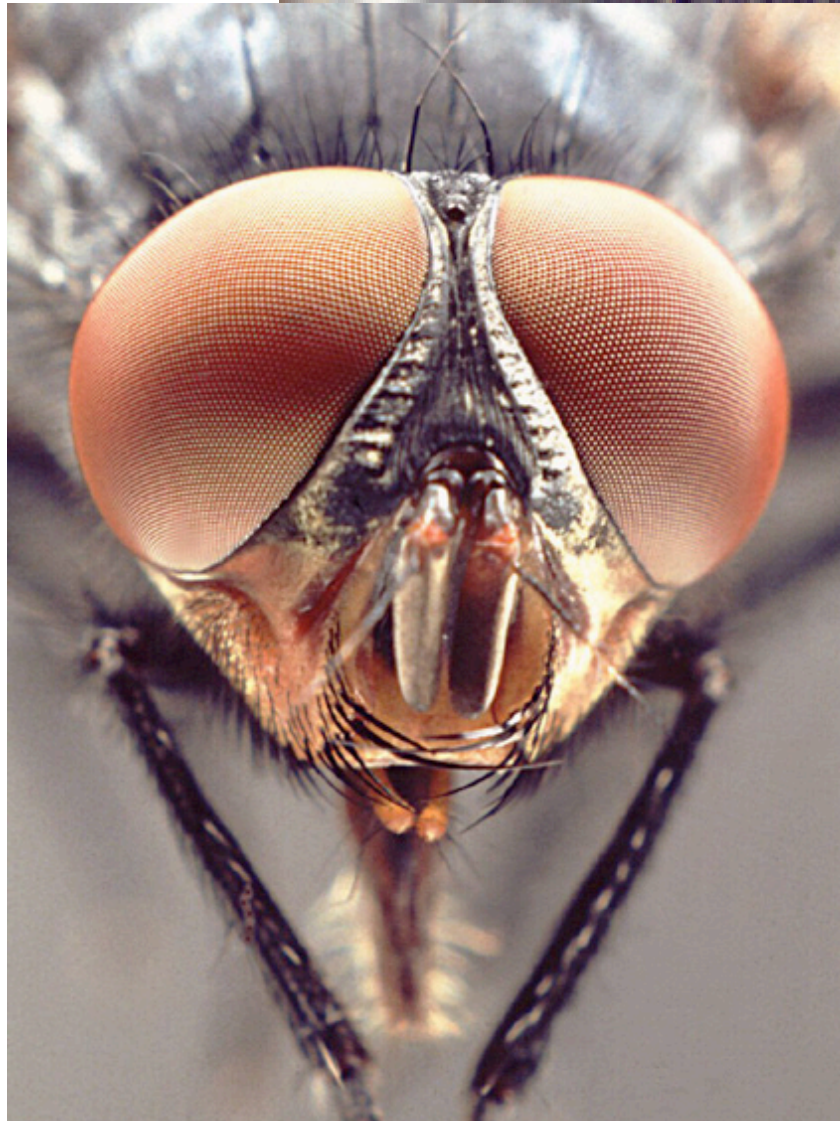
- Posterior variance scales as $1 / \sqrt{N}$
- Little bias, except for distribution with long rank-order tails.
- Counts coincidences and works in Ma regime (if works).
- Is consistent.
- Allows infinite K

(Nemenman et al. 2002, Nemenman 2003)

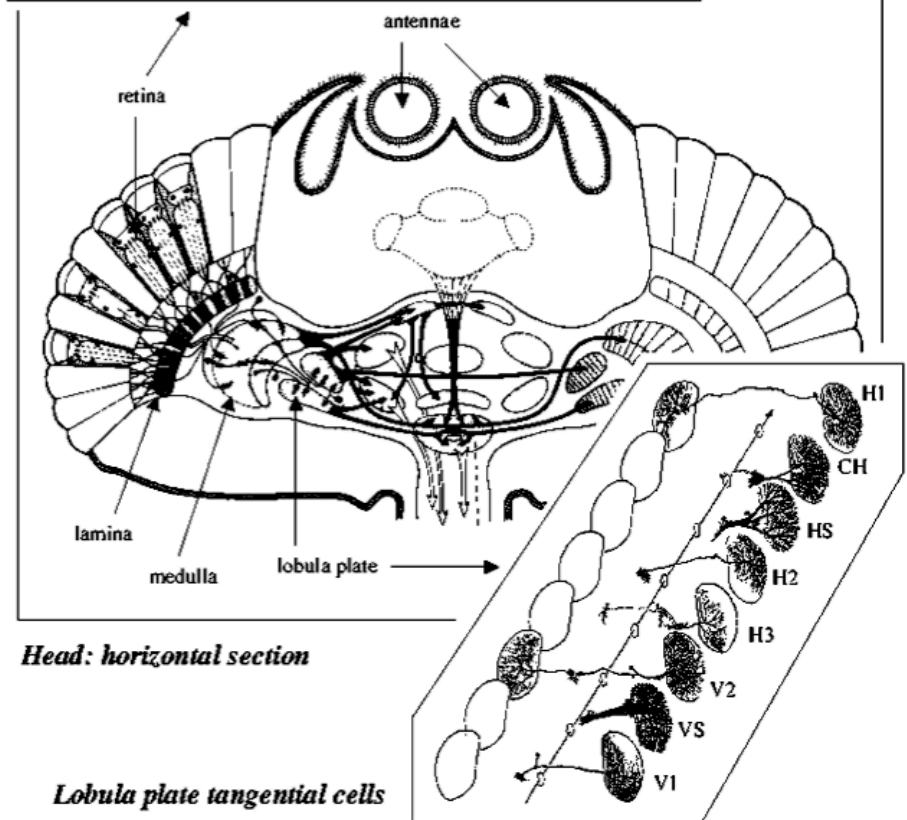
Another hope:

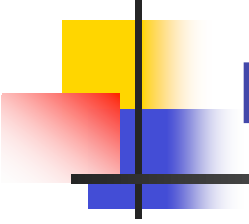
How stable are S and I ranks?





Structure of the retina





Why fly as a neurocomputing model system?

- Can record for long times
- Named neurons with known functions
- Nontrivial computation (motion estimation)
- Vision (specifically, motion estimation) is behaviorally important
- Possible to generate natural stimuli



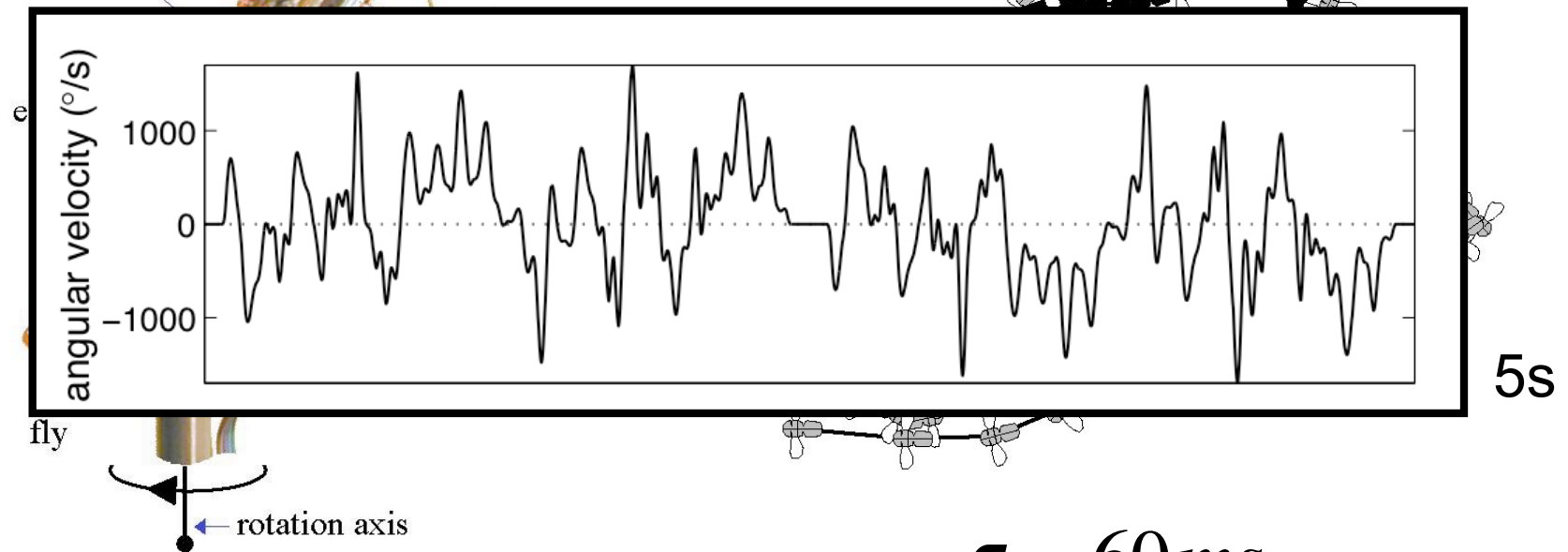
Questions

- Can we understand the code?
- Which features of it are important?
 - Is this a rate or a timing code?
 - Synergy between spikes?
- What does the fly code for?
- How much does it know?
- Is there an evidence for optimality?

Motion estimation in fly H1

(Land and Collett, 1974)

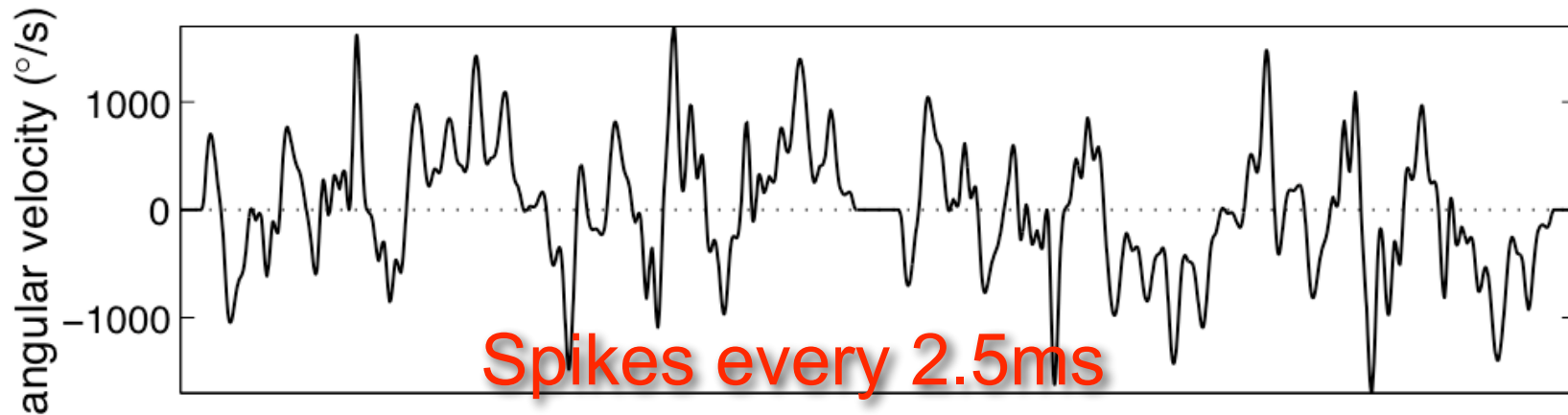
electrode holder
and amplifier



$$\tau = 60ms$$

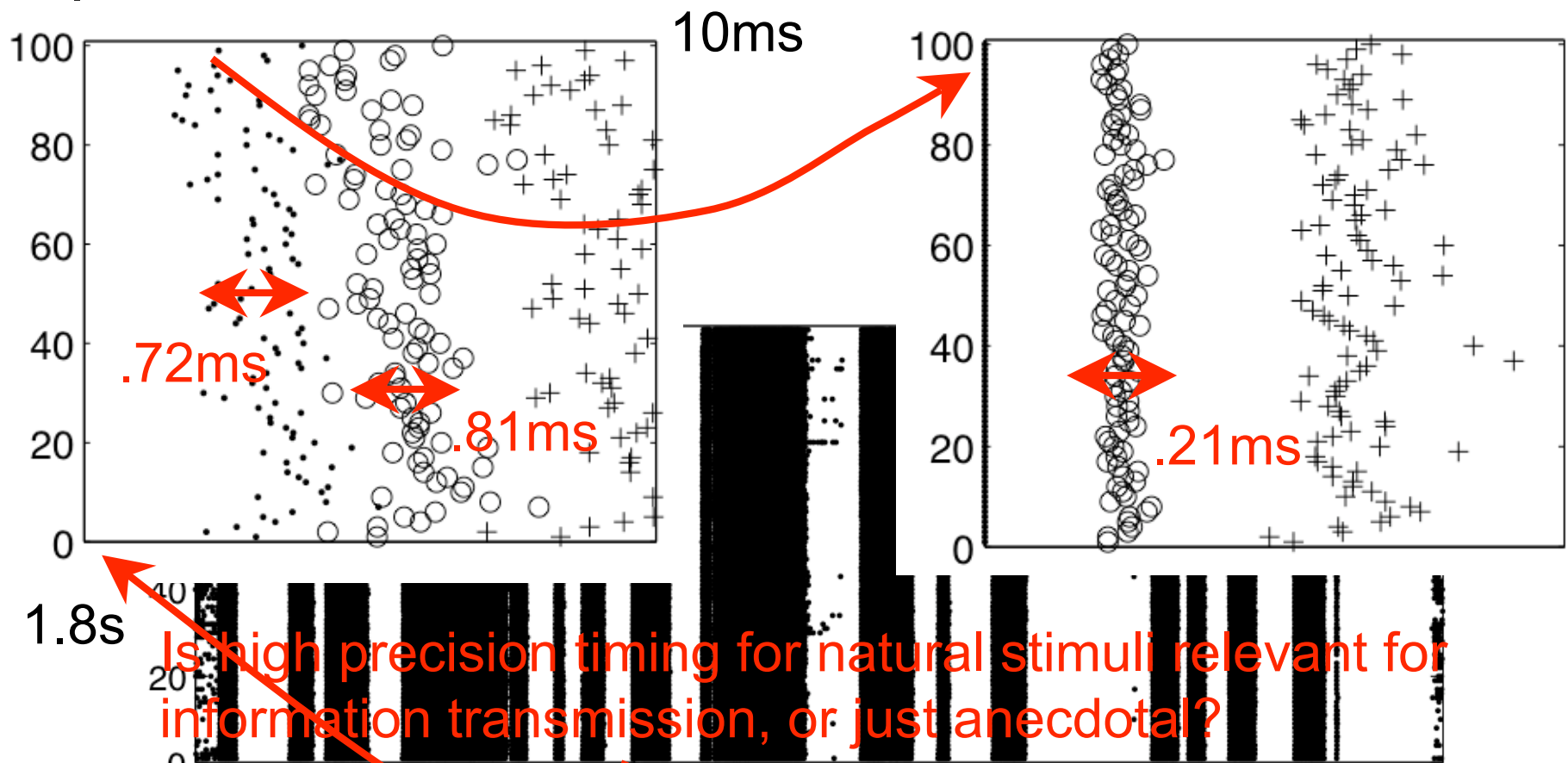
(Lewen et al, 2001)

Natural stimulus and response



5s

Highly repeatable spikes (not rate coding)



Is high precision timing for natural stimuli relevant for information transmission, or just anecdotal?

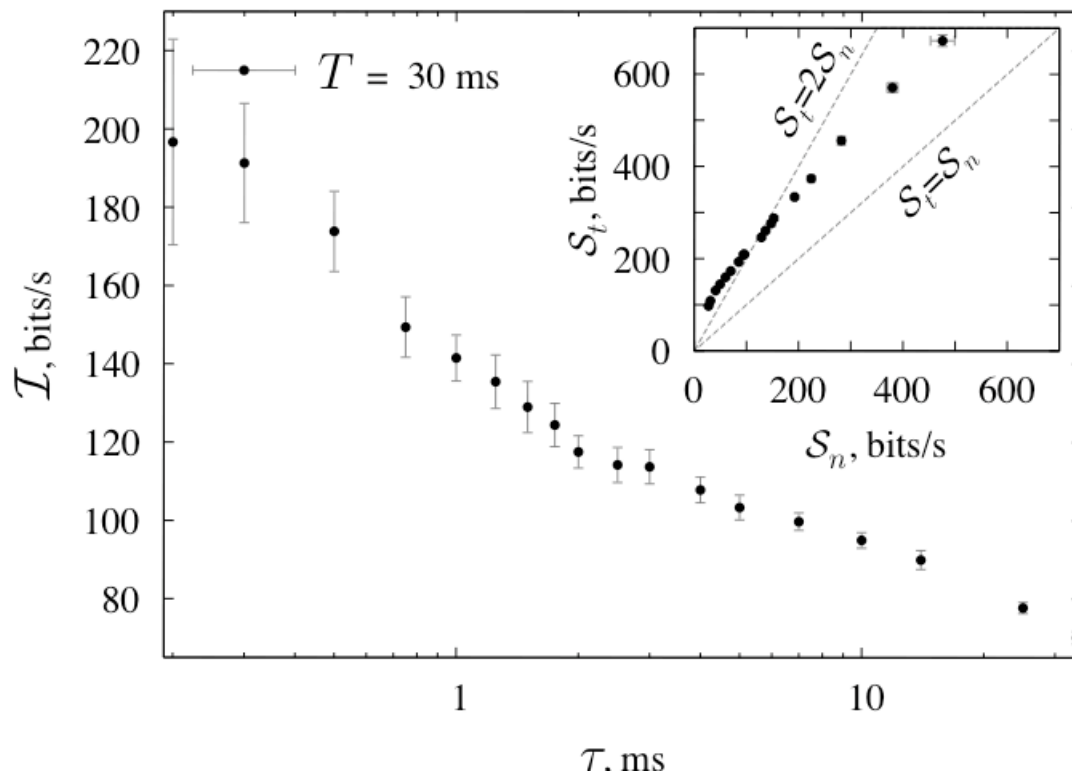
How to interpret the code?



Analysis

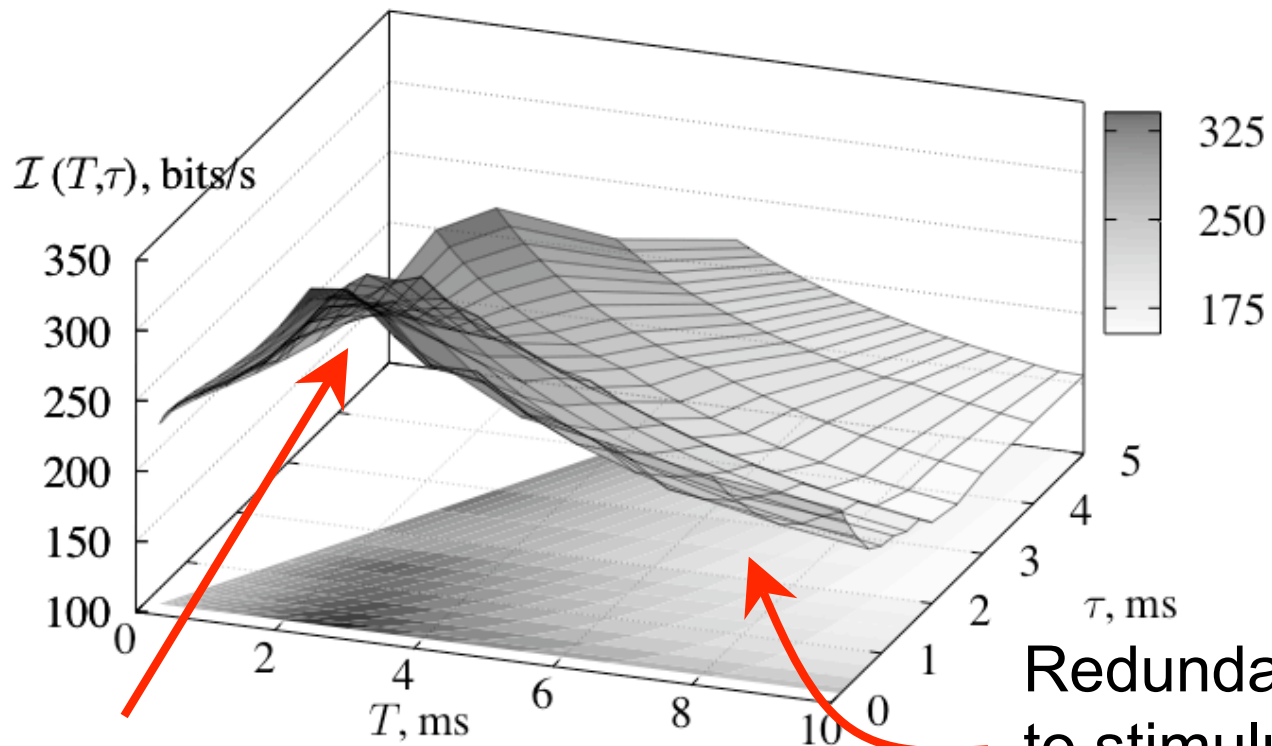
- Collect joint samples of stimuli and responses
- No useful linear features observed
- Analyze $I(s,r)$
- Analyze r up to 30-60 ms, at discretization up to 0.2 ms
- Severely undersampled

Information rate at $T=30\text{ms}$



- Information present up to $\tau = 0.2\text{-}0.3 \text{ ms}$
- 30% more information at $\tau < 1 \text{ ms}$. Encoding by refractoriness?
- ~ 1 bit/spike at 170 spikes/s and low-entropy correlated stimulus. Design principle?
- Efficiency $> 50\%$ for $\tau > 1 \text{ ms}$, and $\sim 80\%$ at 30ms. Optimized for natural statistics?

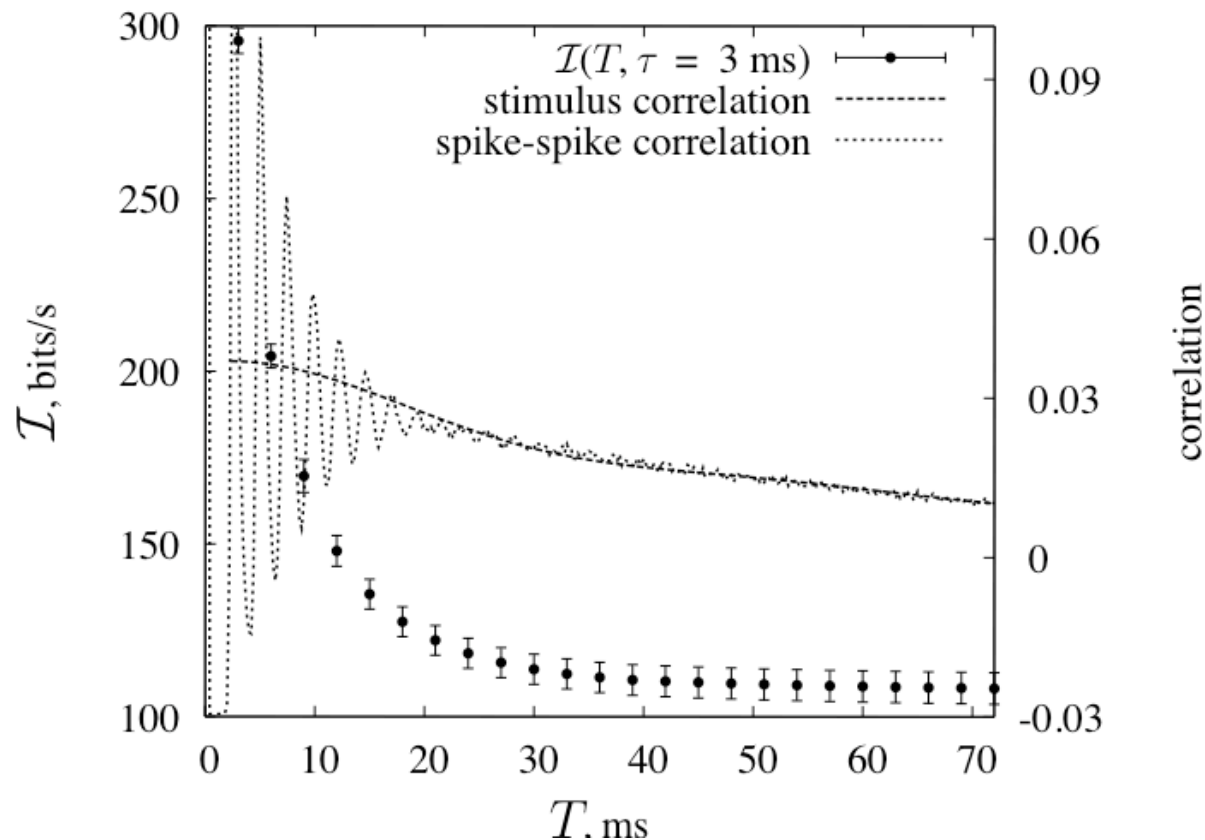
Synergy from spike combinations



Spike pairs

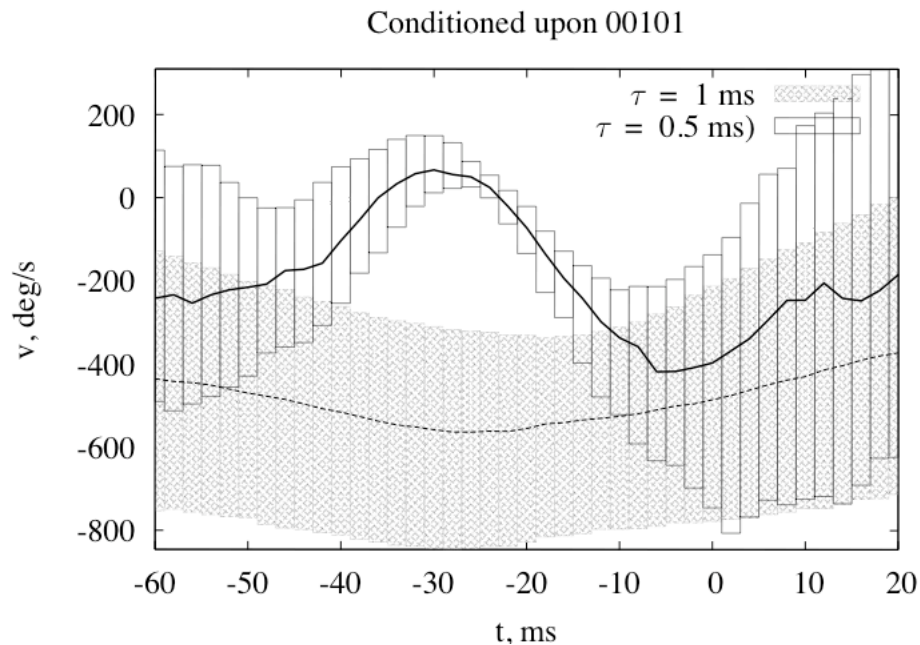
Redundancy due to stimulus

New bits (optimized code)

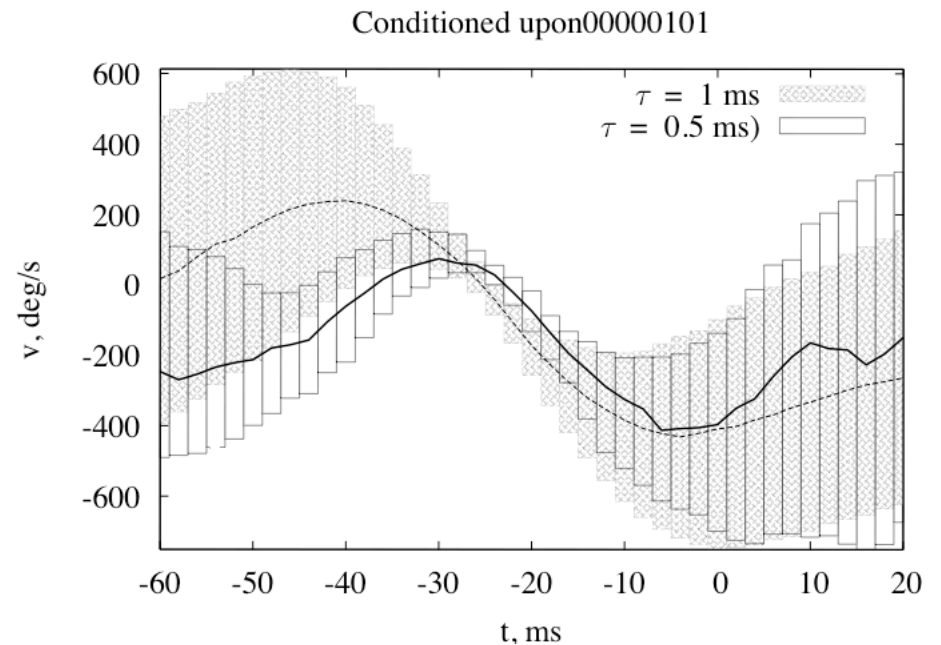


- Spikes are very regular (15 rings)
WKB or liquid decoder? Interspike potential?
- CF at half its value, but fly gets new bits every 30 ms
- Independent info (even though entropies are T dependent).

Information about...



Signal shape



Zero-crossings time

Best estimation at 25ms delay. Little time for reaction.

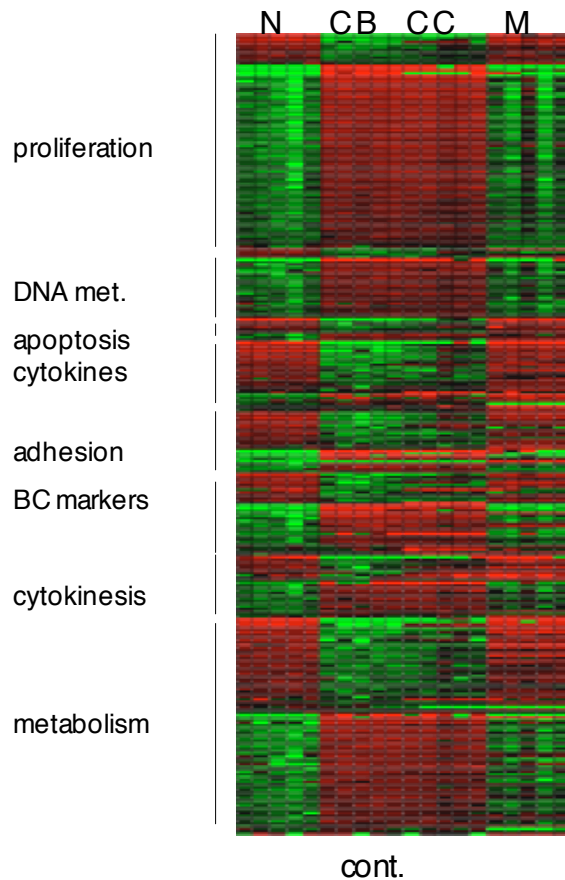


Other analysis

- Adaptation of the neural code to stimuli statistics (to maximize information transmission)
- Speed of adaptation
- Individuality of animals
- Effects of multiple neurons
- Effects of multiple spikes
- Predictive features selection by the fly

Example 2

Transcriptional networks



Biochemical interaction networks:
The Holy Grail

Search for irreducible steady
state statistical dependencies
(with biologically realistic
assumptions) and hope for the
best.

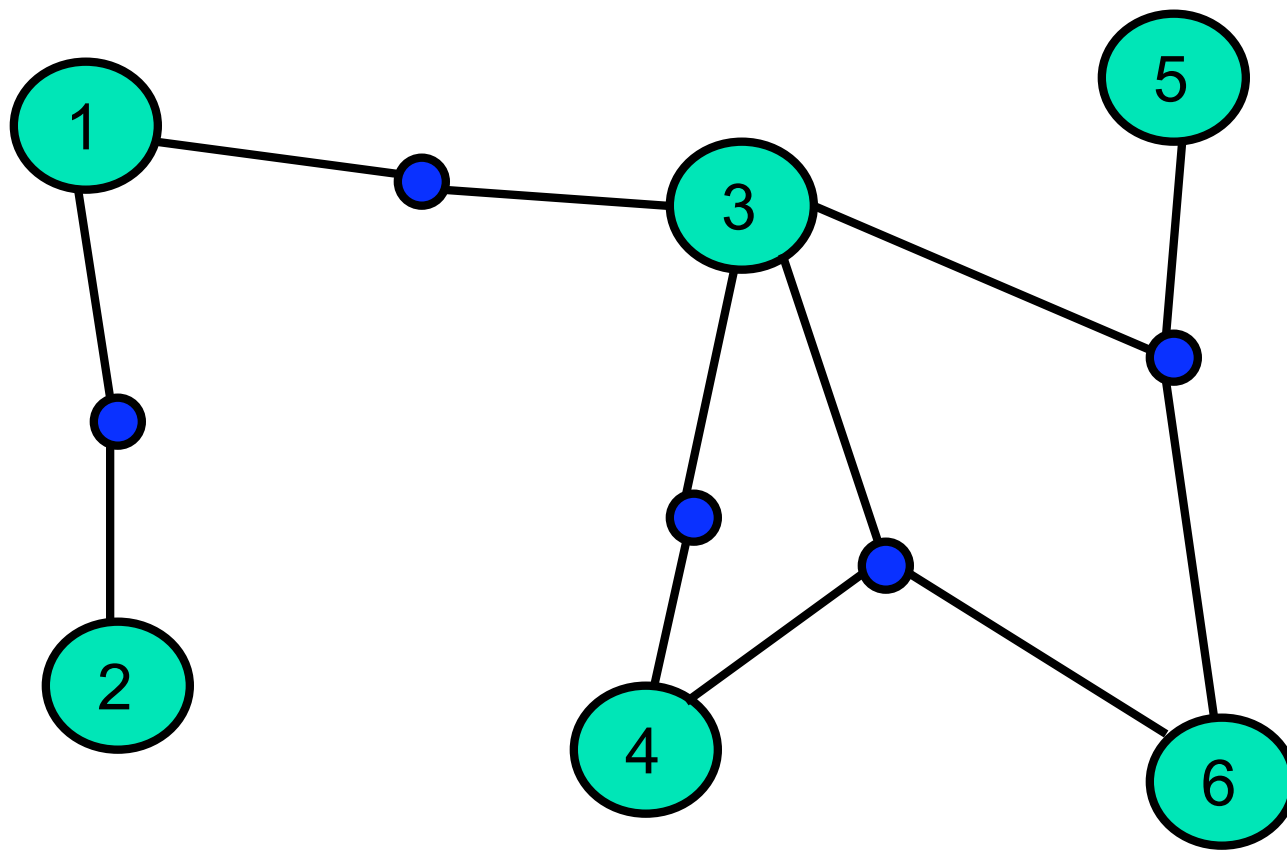


Problems

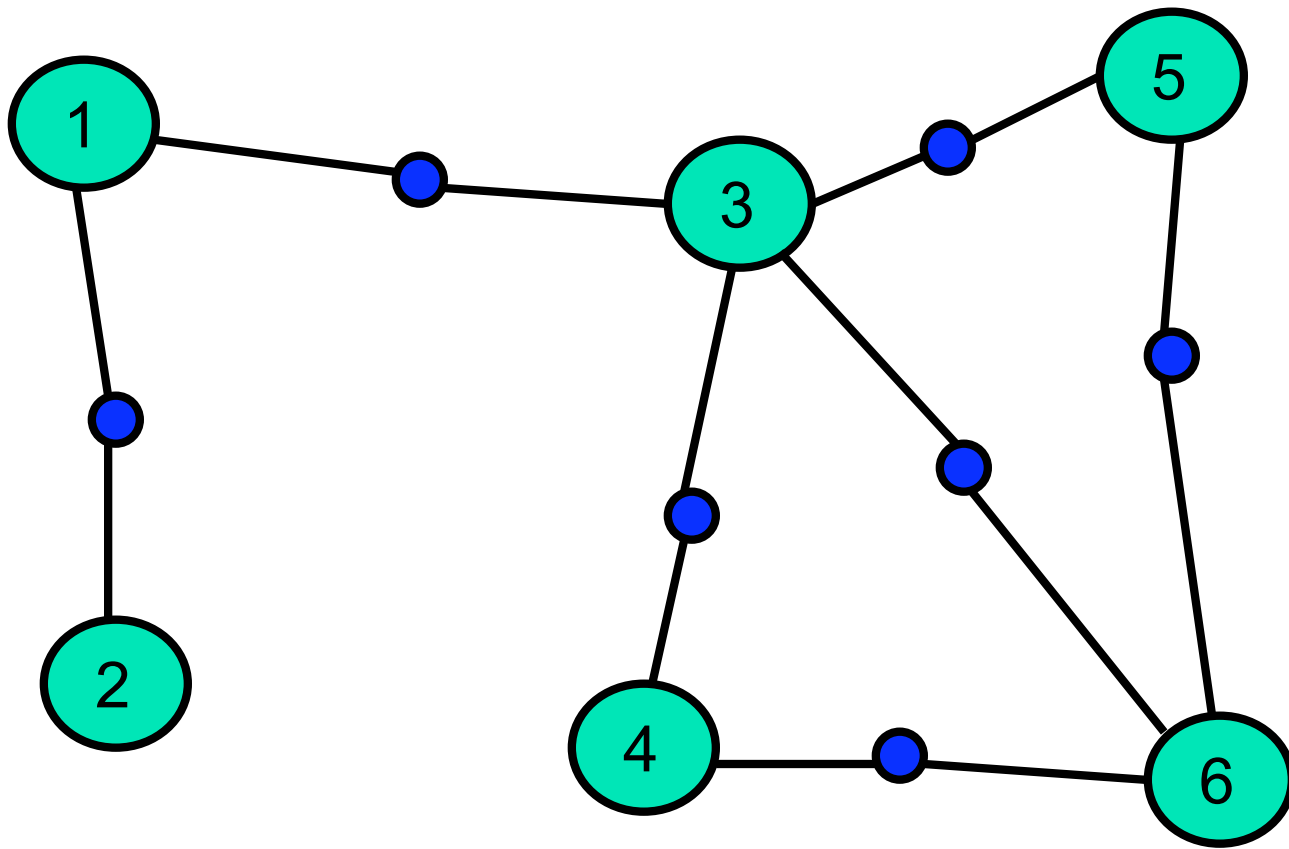
	Co	BN	Reg	ARACNE
Small data requirements	✓	✗✓	✗	✓
Robustness of reconstruction	✓	✗✓	✗	✗✓
Computational complexity	✓	✗	✗✓	✓
Scalability	✓	✗	✗✓	✓
Conditional interactions	✗✓	✓	✗✓	✓
Overfitting	✓	✗✓	✗✓	✓
Confounding	✗	✓	✗	✓



Interaction network

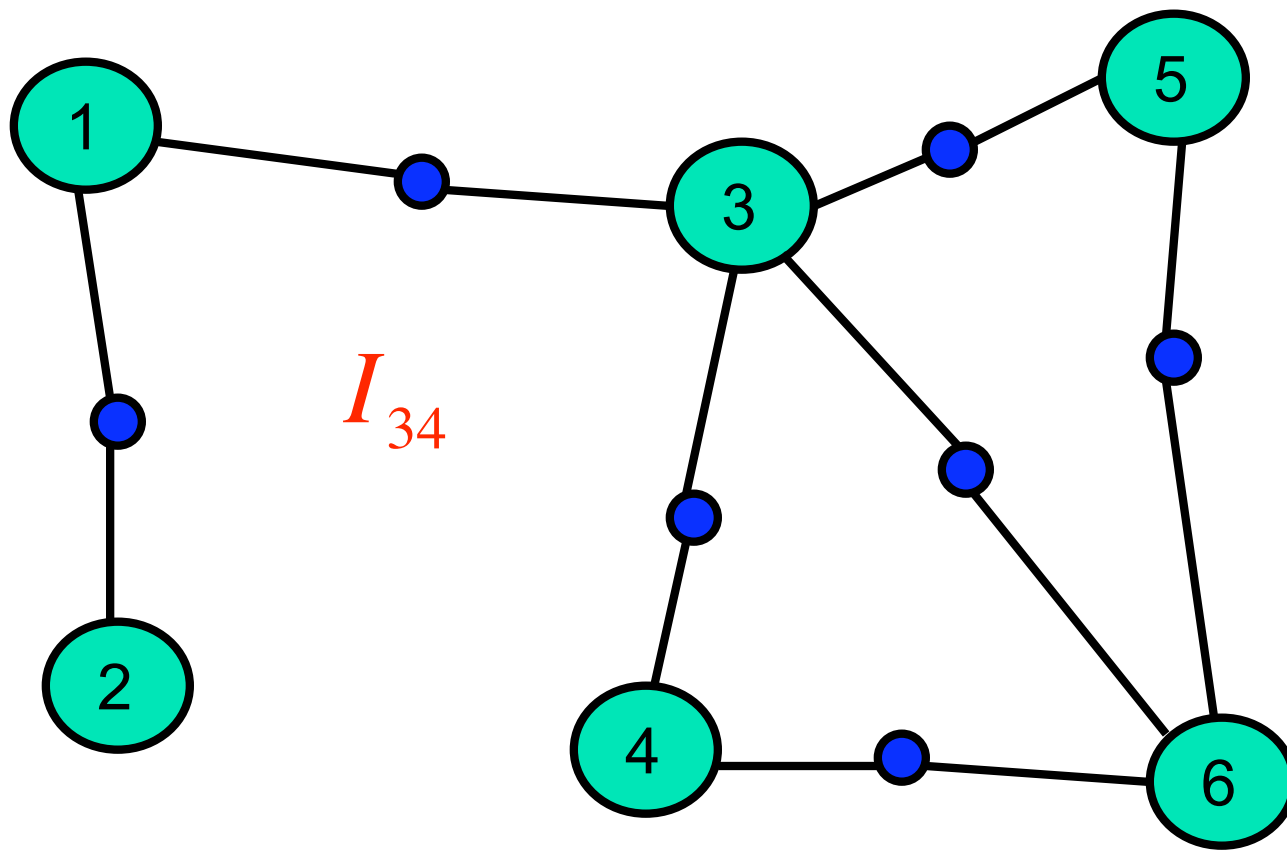


Approximate by 2-way
network (few data)



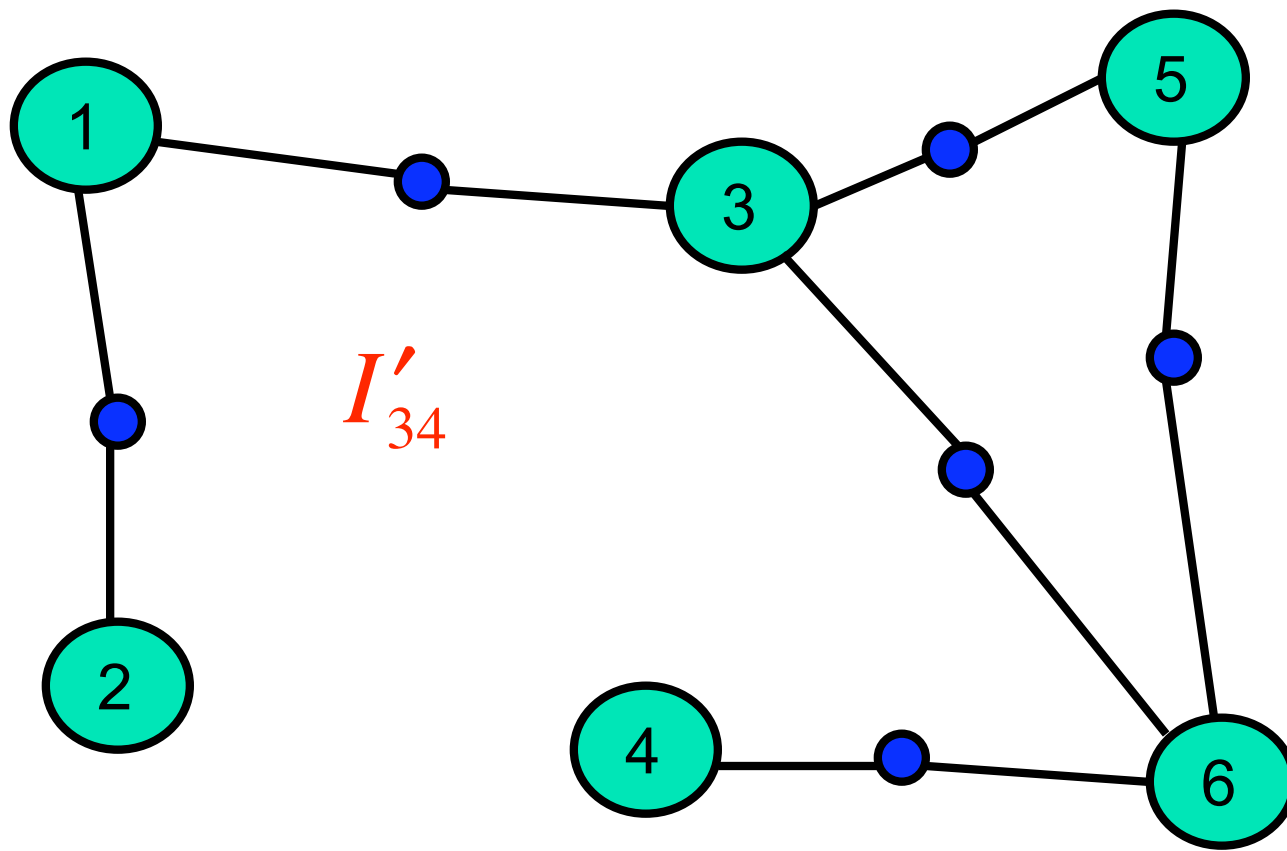


Locally tree-like approximation

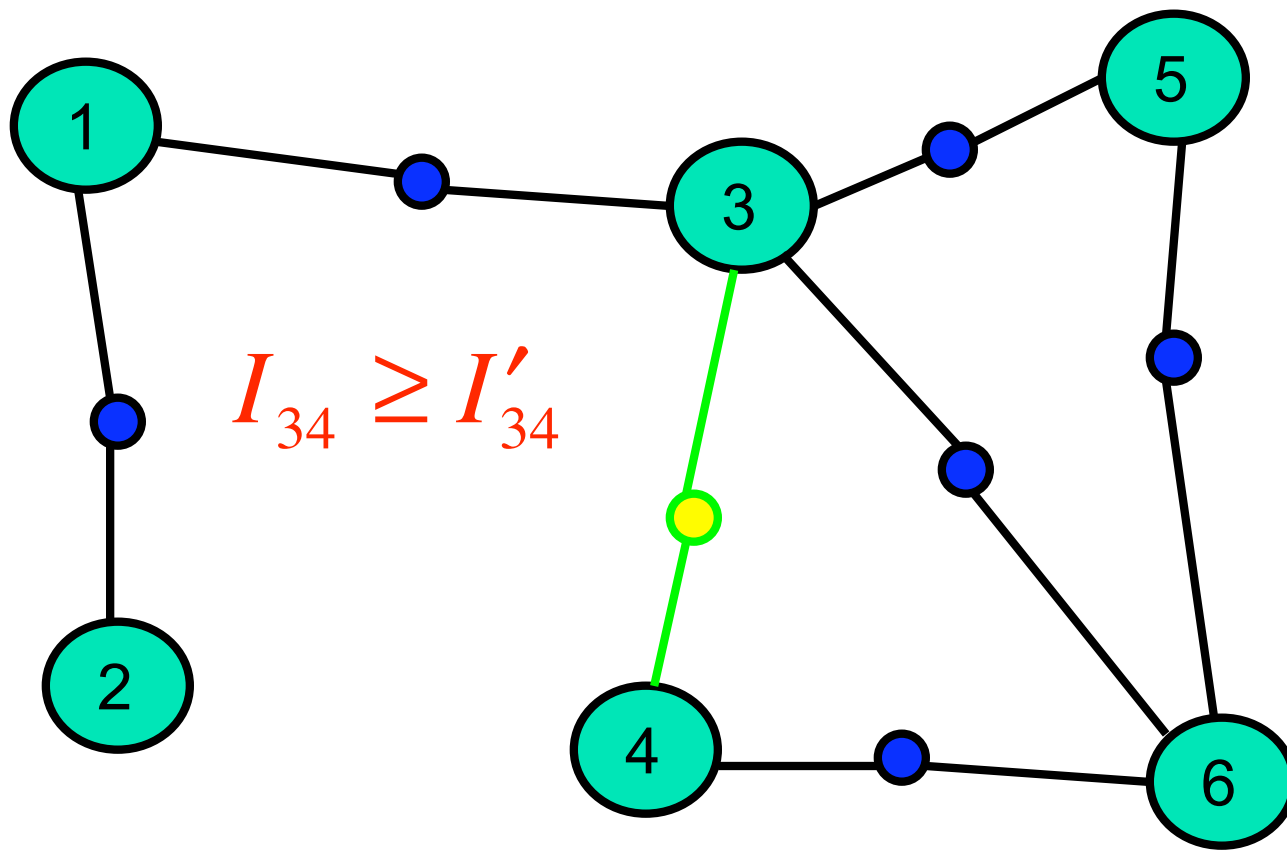




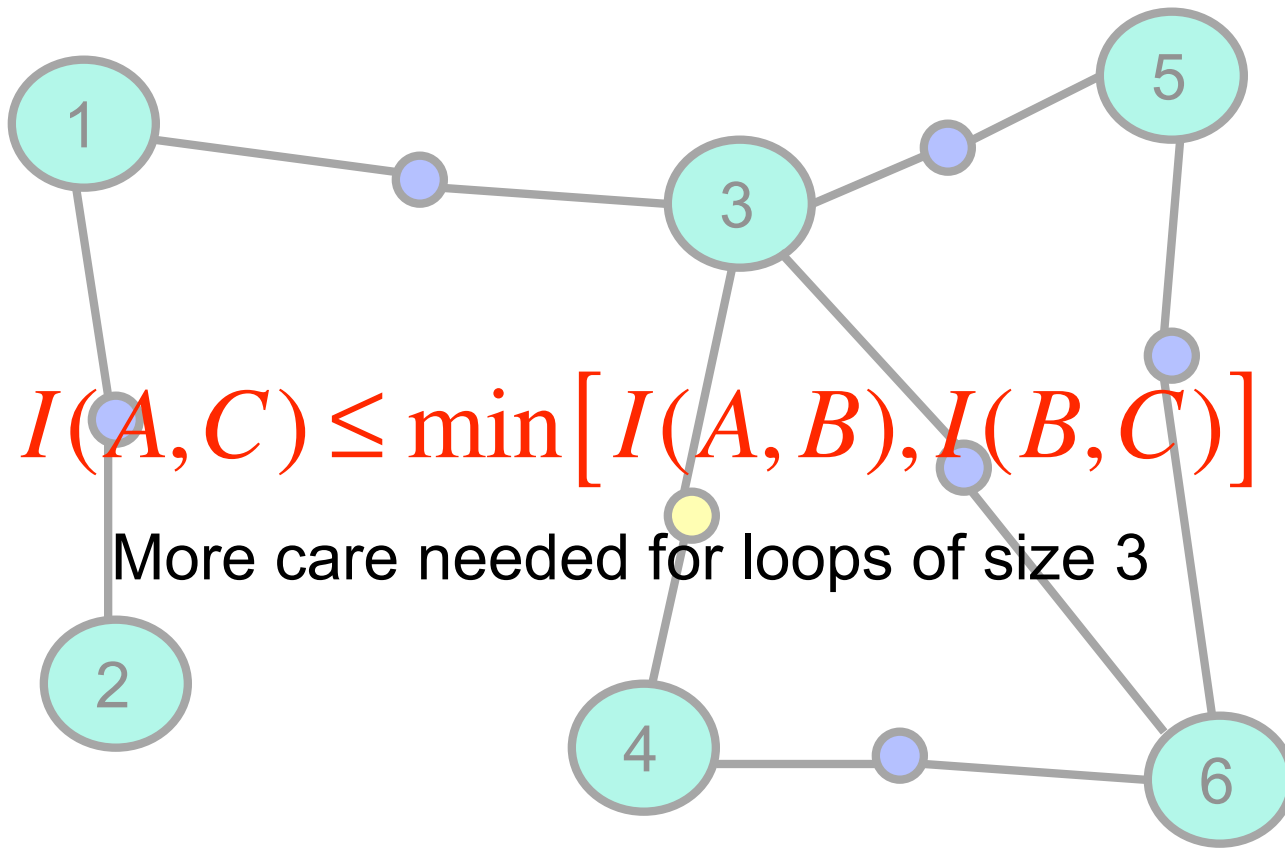
Locally tree-like approximation



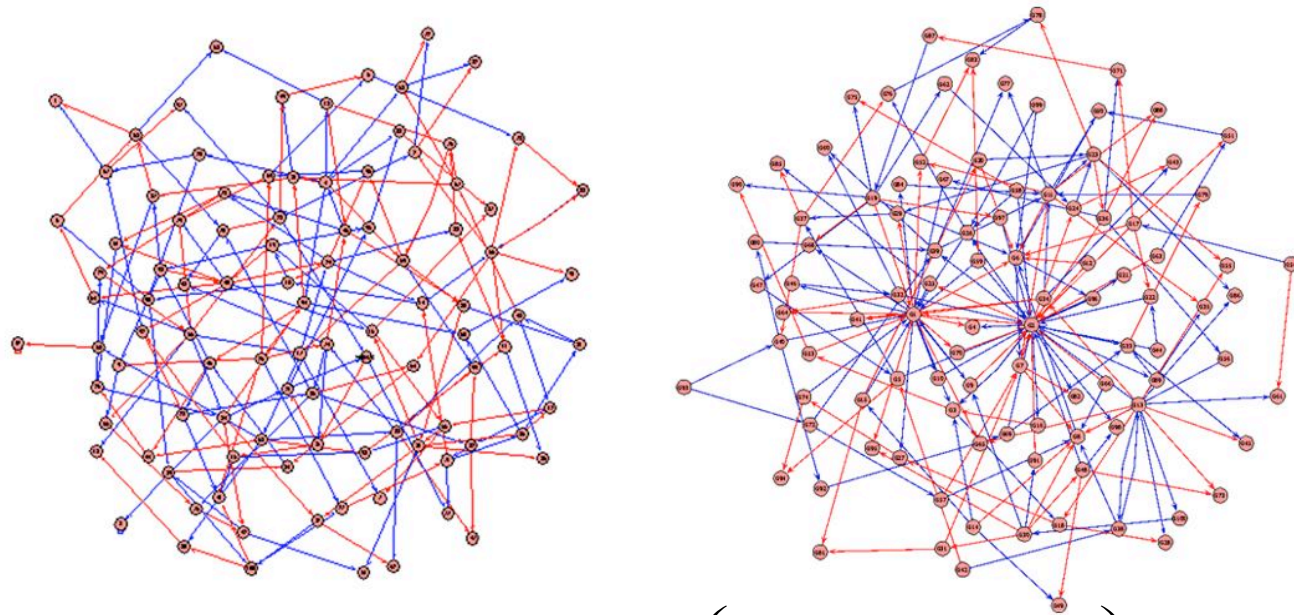
Locally tree-like:
signals decorrelate fast



No false positives and
(almost) no false negatives



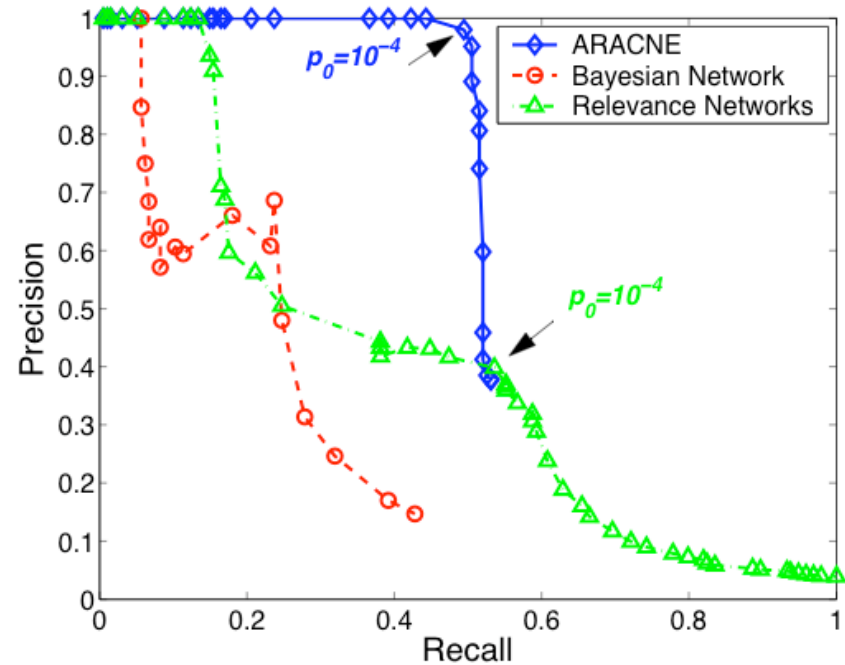
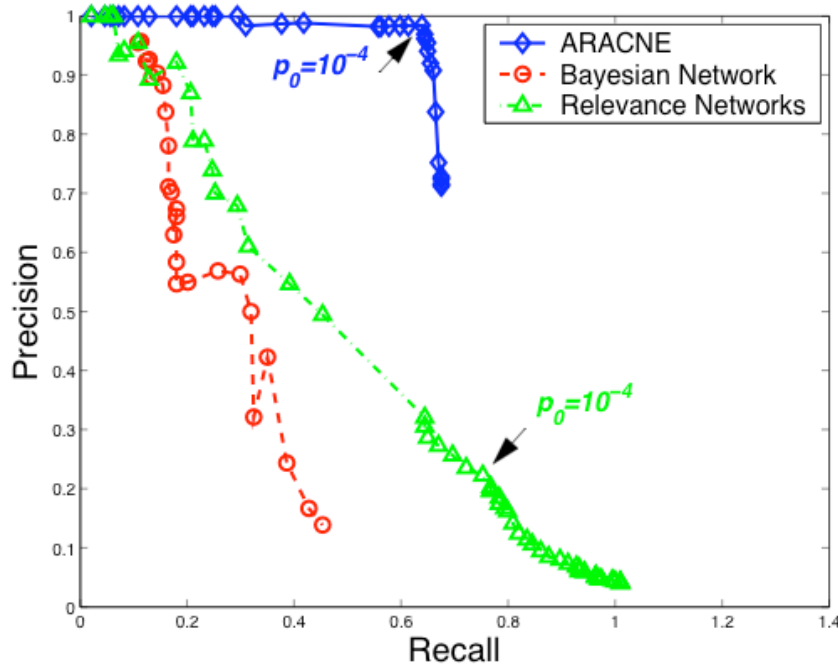
Synthetic networks



$$\frac{dx_i}{dt} = a_i \prod_j \frac{I_{0,j}^{v_j}}{I_j^{v_j} + I_{0,j}^{v_j}} \prod_j \left(1 + \frac{A_{0,j}^{v_j}}{A_j^{v_j} + A_{0,j}^{v_j}} \right) - b_i x_i$$

Randomize rates to sample different steady states

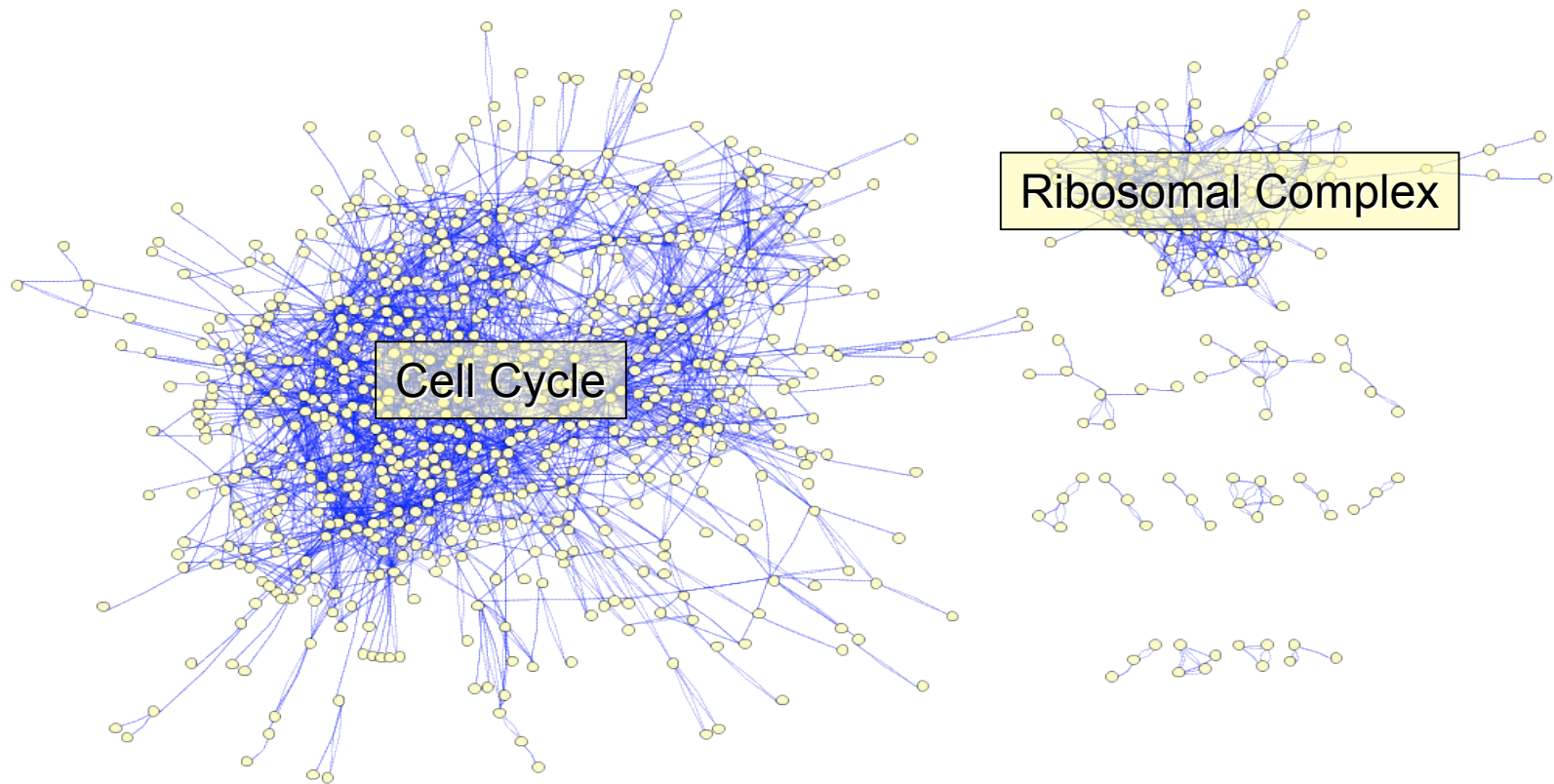
Synthetic networks benchmarks ($N=1000$)



Graceful decay for smaller N

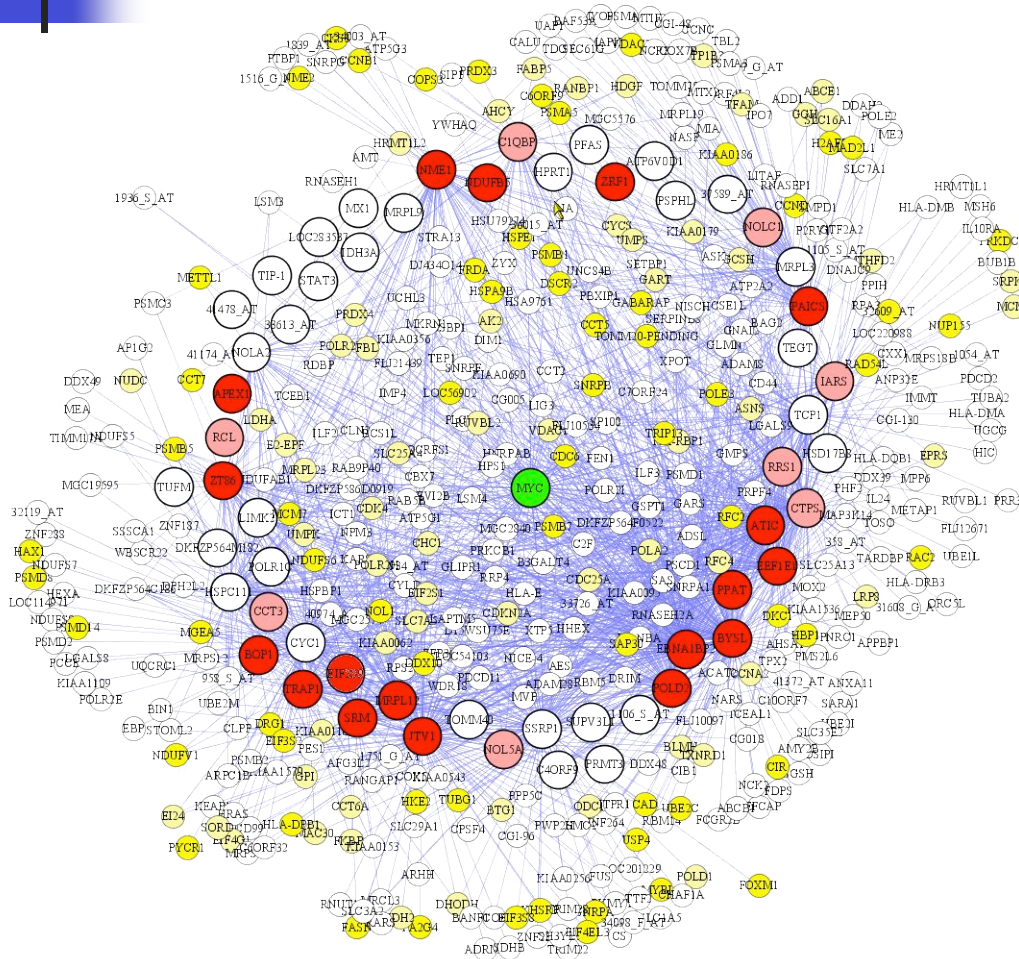


Complete B-cell network



~129000 interactions

c-MYC subnetwork



- Protooncogene,
- 12% background binding,
- one of top 5% hubs
- significant MI with 2000 genes

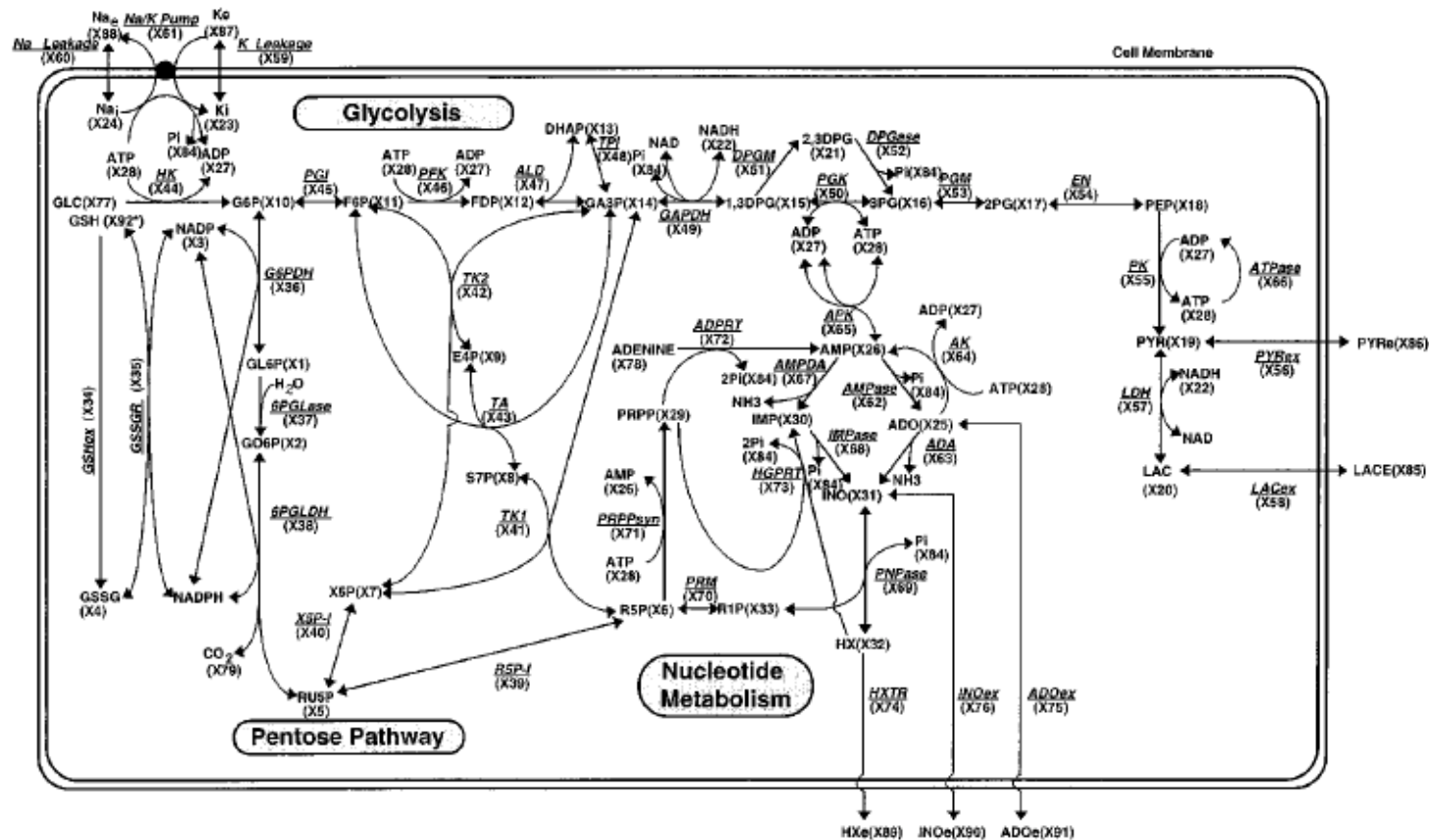
Total interactions: 56
Pre-known: 22
Ch-IP validated: 11/12



Also validated in...

- Other hubs
- Various yeast data sets
- RBC metabolic network

RBC metabolic network

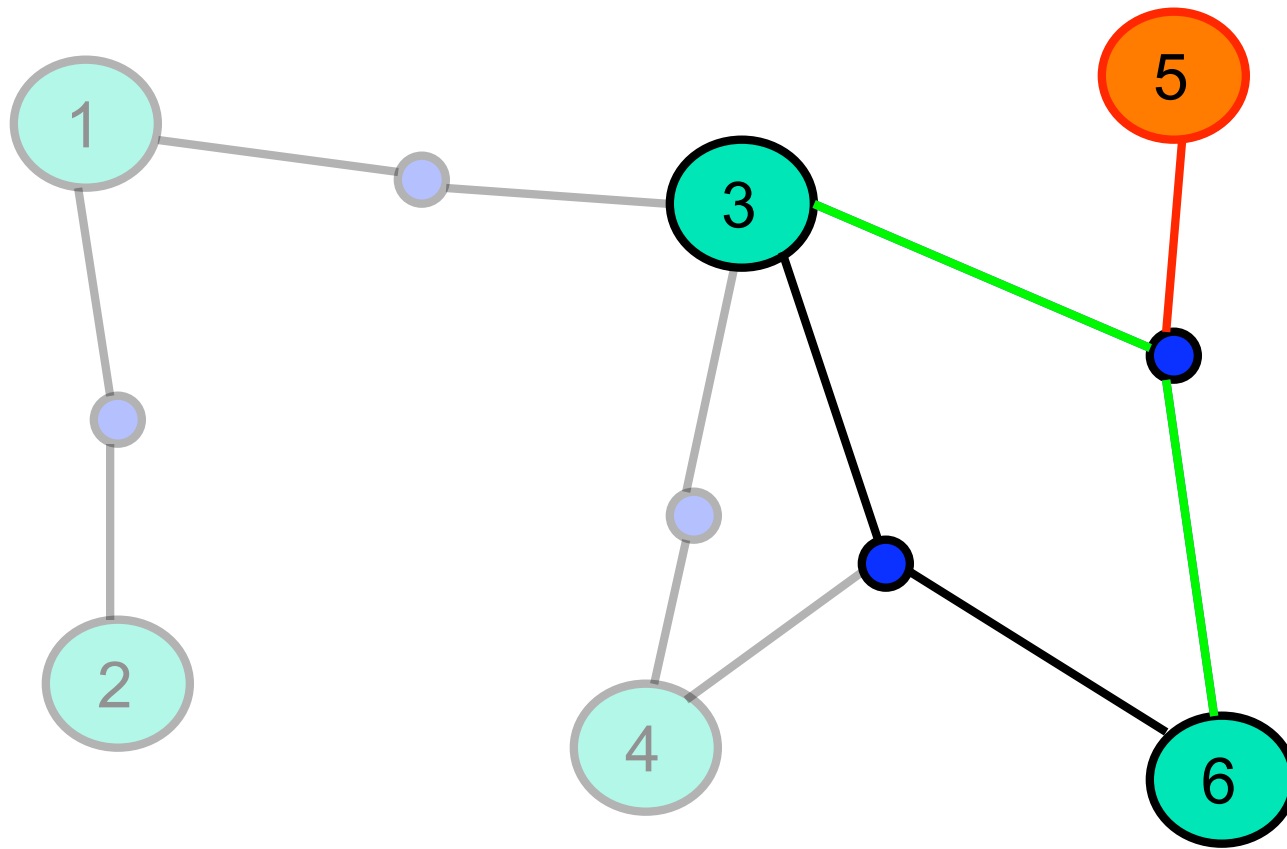




ARACNE deconvolution of synthetic data

- 39 metabolites, 19 samples
- 14 predicted, 11 (possibly 12) confirmed
- Done analysis on 828 samples, 40 predicted, still confirming
- Unrealistically small noise
- Questionable network stability for small data sets

3rd order (regulated) interactions





3rd order interactions

- Many triplets -- focus on important hubs (c-MYC, BCL-6)
- Search for modulators among genes with large dynamic range and small MI with the hub
- Find modulators whose under- or over-expression changes interaction properties of a hub
- No guarantee of irreducibility
- Validate in GO w.r.t. to transcription factors and kinases among modulators

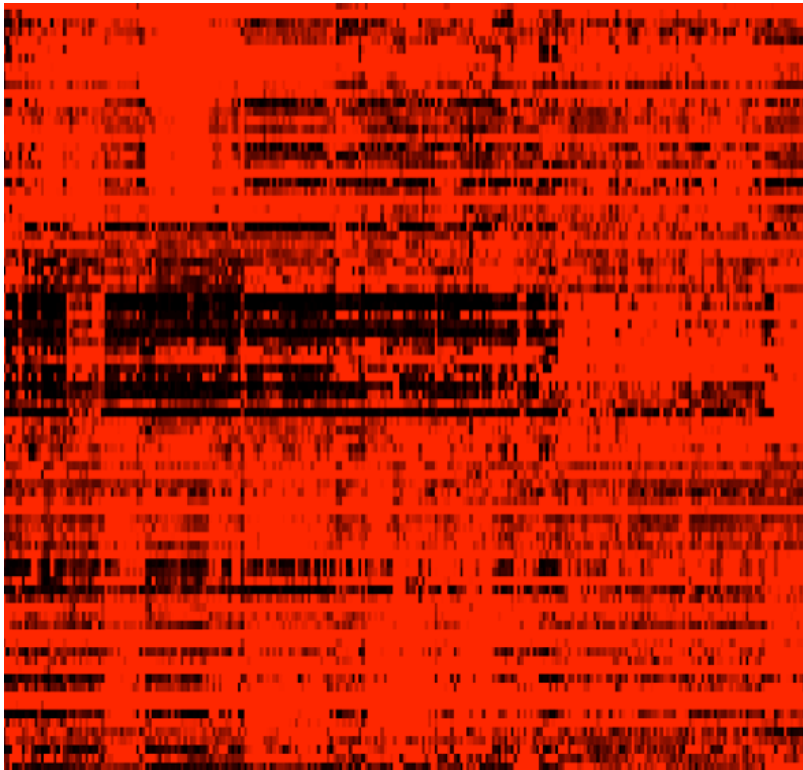


c-MYC modulators

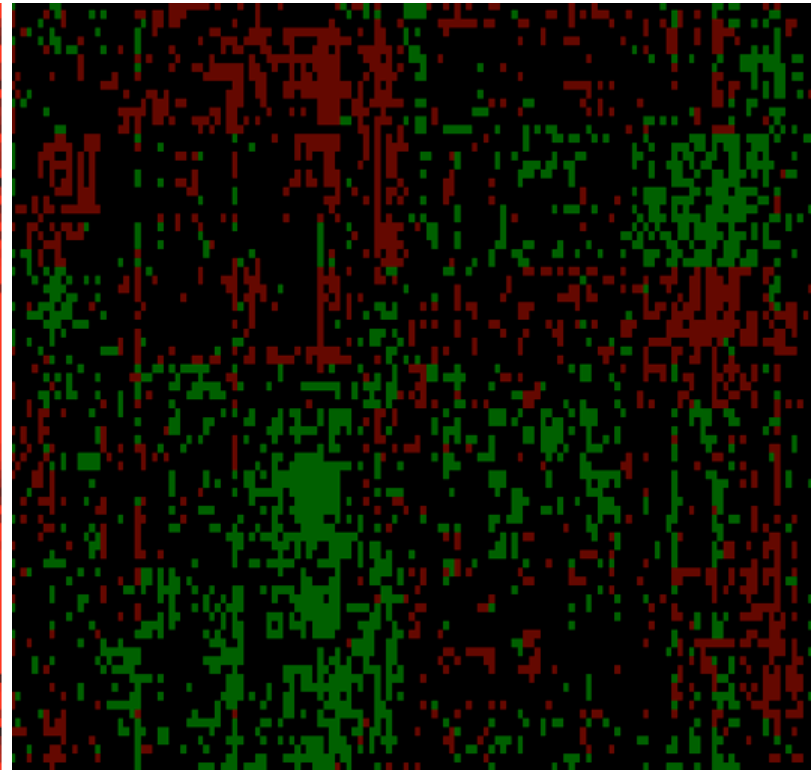
- 1117 candidate modulators (825 with known function in GO)
- 69 candidate modulators identified
- Kinases: 10/69 (background 42/825), p-value $1e-3$
- TFs: 15/69 (background 56/825), p-value $1e-6$
- Total: 25/69 (background 98/825), p-value $3e-8$

Many correlated modulators

|expression|



change in interactions



Over 70% cluster overlap



Currently

- Biochemical validation
- Search for irreducible modulators



Thanks

- Columbia: Andrea Califano, Chris Wiggins, Riccardo dalla Favera, Katia Basso, Adam Margolin, Kai Wang, Nila Banerjee, Omar Antar
- IBM: Gustavo Stolovitzky
- Princeton: William Bialek, Fariel Shafee
- Indiana: Rob de Ruyter van Steveninck
- Jerusalem: Naftali Tishby
- OSDN/SourceForge