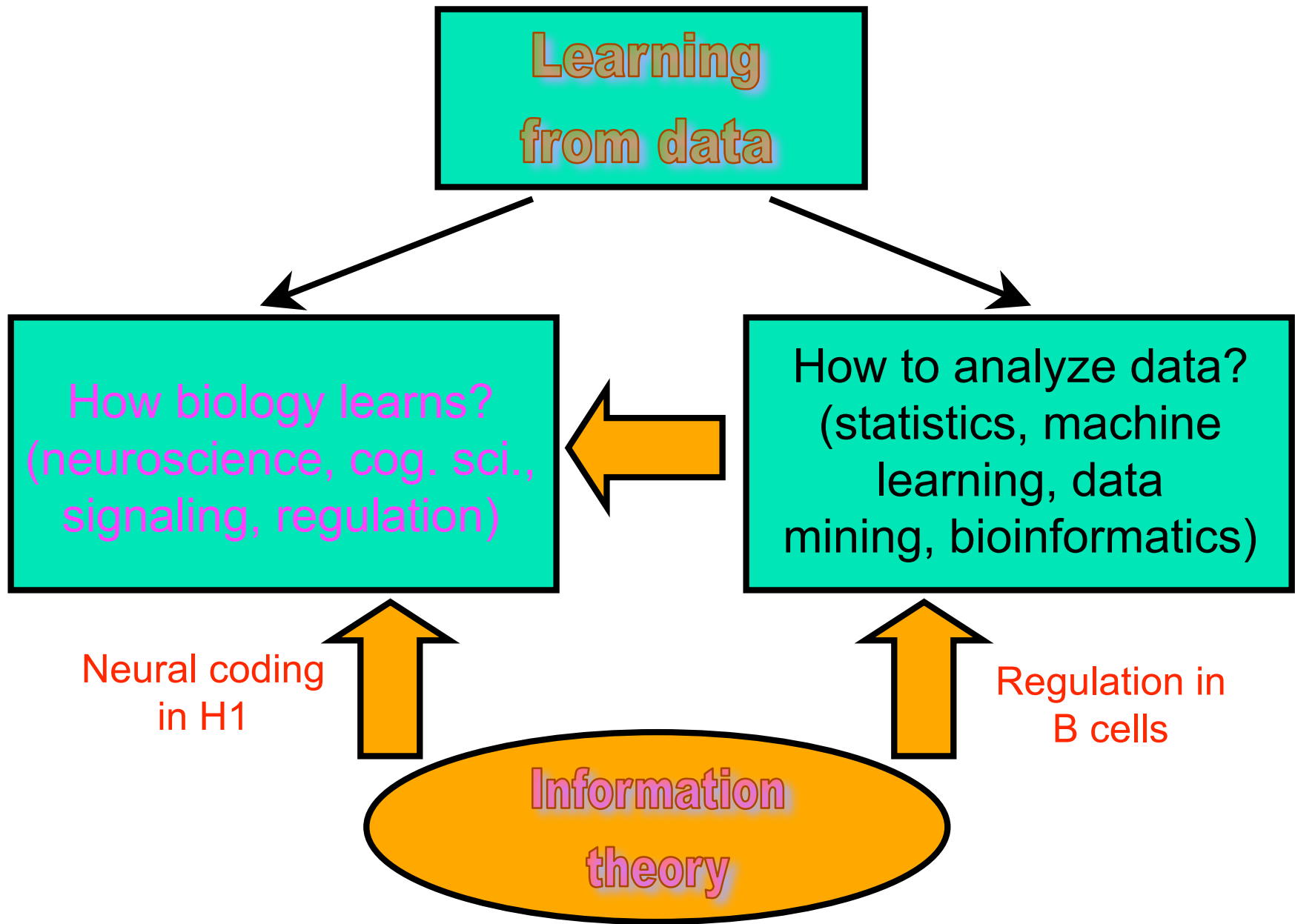


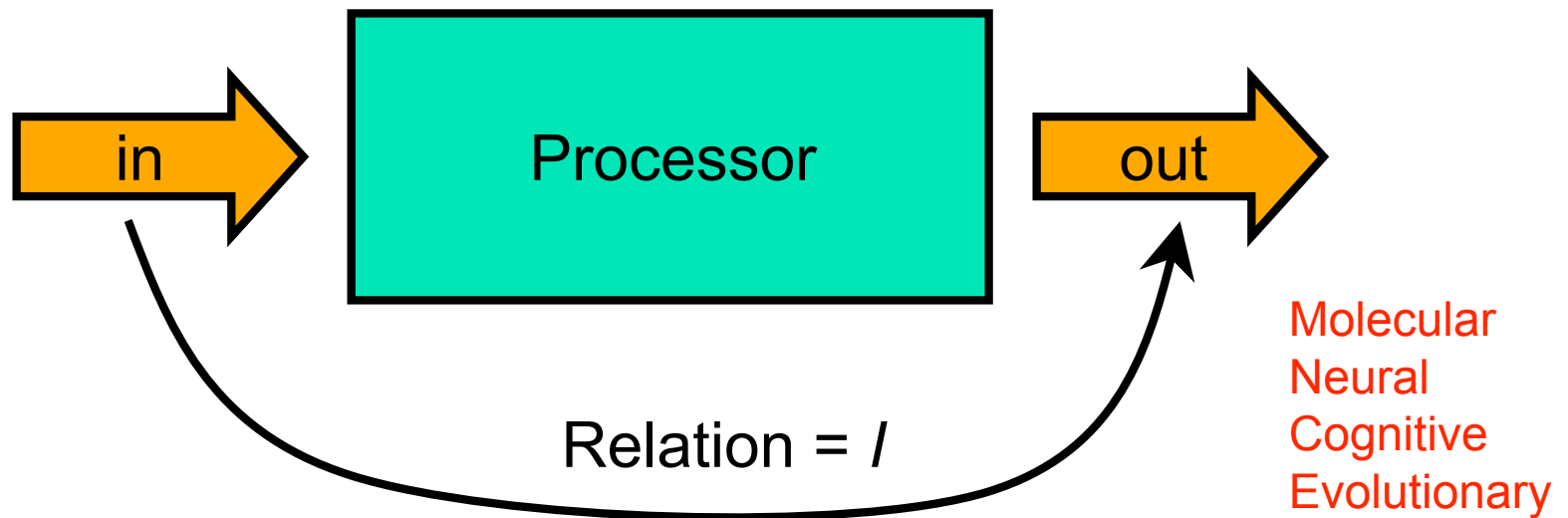
Information theory in systems biology



Ilya Nemenman
(JCSB, Columbia)

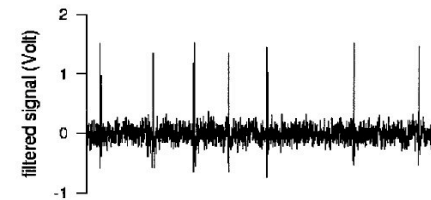
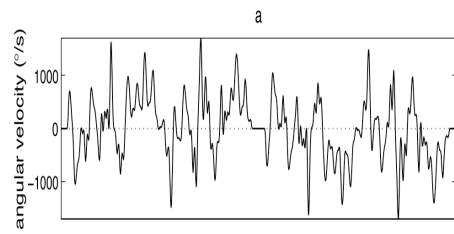


Studying signal transduction



What is the richness of ins/outs?
How faithful is the output to the input?
How does it coding input?

Studying signal transduction



Neural

What is the in/out *relation*?
Efficiency of estimation?
Efficiency of encoding?

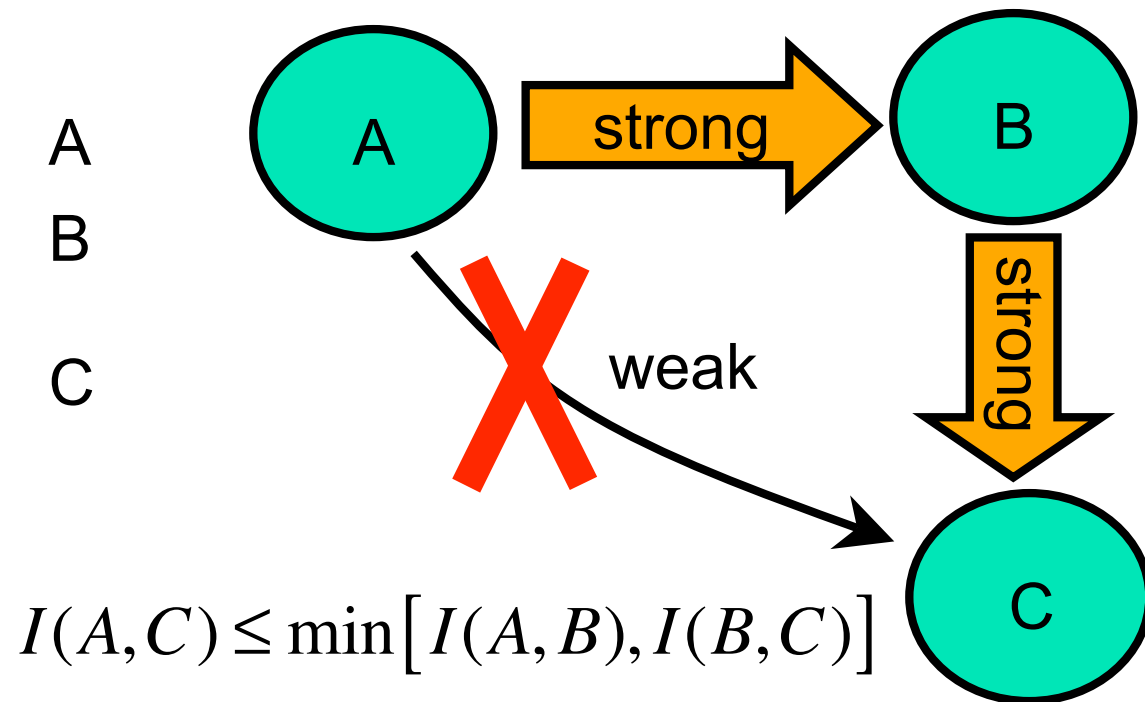
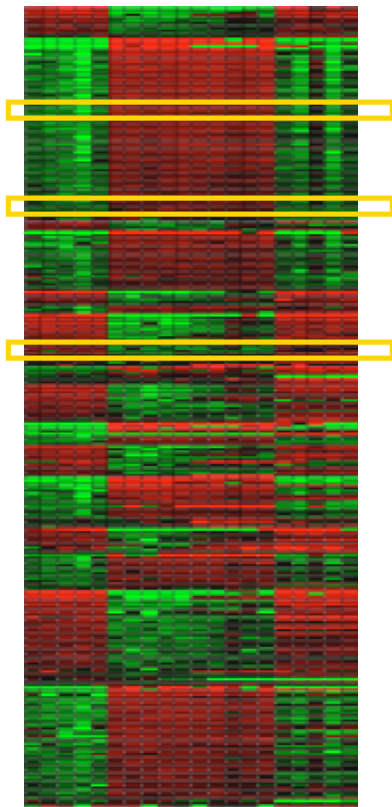
Studying signal transduction



Synergies for multiple ins/outs?

What is I?

Reconstructing interaction models



What is I?



Variances and Correlations

$$\sigma^2(x)$$

normal

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

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.

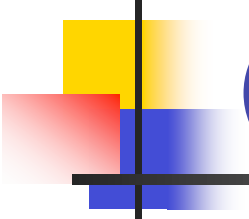


Entropy (unique measure of randomness, in bits)

$$S[X] = - \sum_{x=1}^K p_x \log p_x = - \langle \log p_x \rangle$$

$$0 \leq S[X] \leq \log K \quad (\text{number of "bins"})$$

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



Mutual Information

(interactions, shared data)

$$I[X;Y] = \left\langle \log \frac{p_{xy}}{p_x p_y} \right\rangle$$
$$= S[X] + S[Y] - S[X,Y]$$

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

$$N[(x_0, y_0), \Sigma] \Rightarrow I[X;Y] = -\frac{1}{2} \log(1 - \rho_{xy}^2)$$



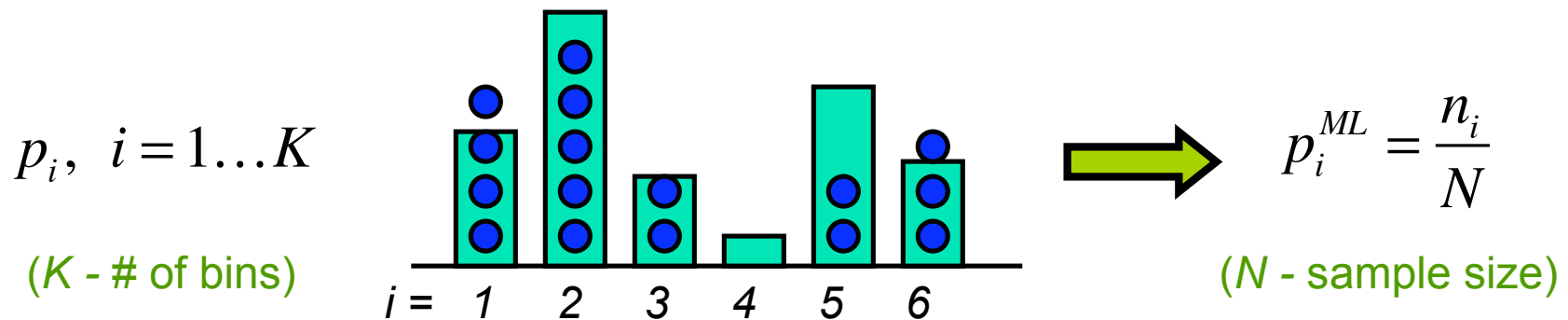
Why MI?

- Captures all dependencies (zero *iff* joint probabilities factorize)
- Reparameterization invariant
- Unique metric-independent measure of “how related”

(Nemenman and Tishby 2005)

Why is IT not common in statistics?

Maximum likelihood estimation:



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



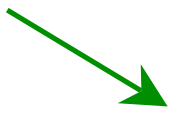
$$\langle S_{ML} \rangle \leq -\sum_i \frac{\langle n_i \rangle}{N} \log \frac{\langle n_i \rangle}{N} = S$$



Why is IT not common in statistics?

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

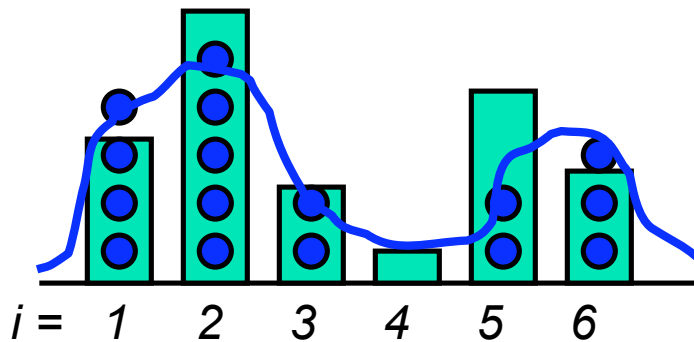
$\log K$


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

Fluctuations underestimate entropies and overestimate mutual informations.

(Need smoothing.)

Correct smoothing possible



$$S \leq \log N$$

(often not enough)

Incorrect smoothing = over- or underestimation.

Developed for problems ranging from mathematical finance to computational biology.

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



What if $S > \log N$?

But there is hope (Ma, 1981):

For uniform K -bin distribution the first coincidence occurs for

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

$$S \sim \textcolor{red}{2} \log N_c \leftarrow \text{Time of first coincidence}$$

Can make estimates for square-root-fewer samples!

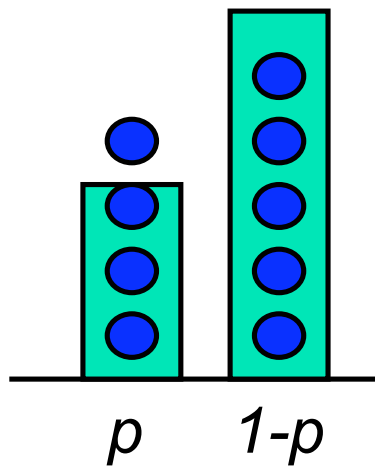
Can this be extended to nonuniform cases?

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

What is unknown?

Binomial distribution:

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



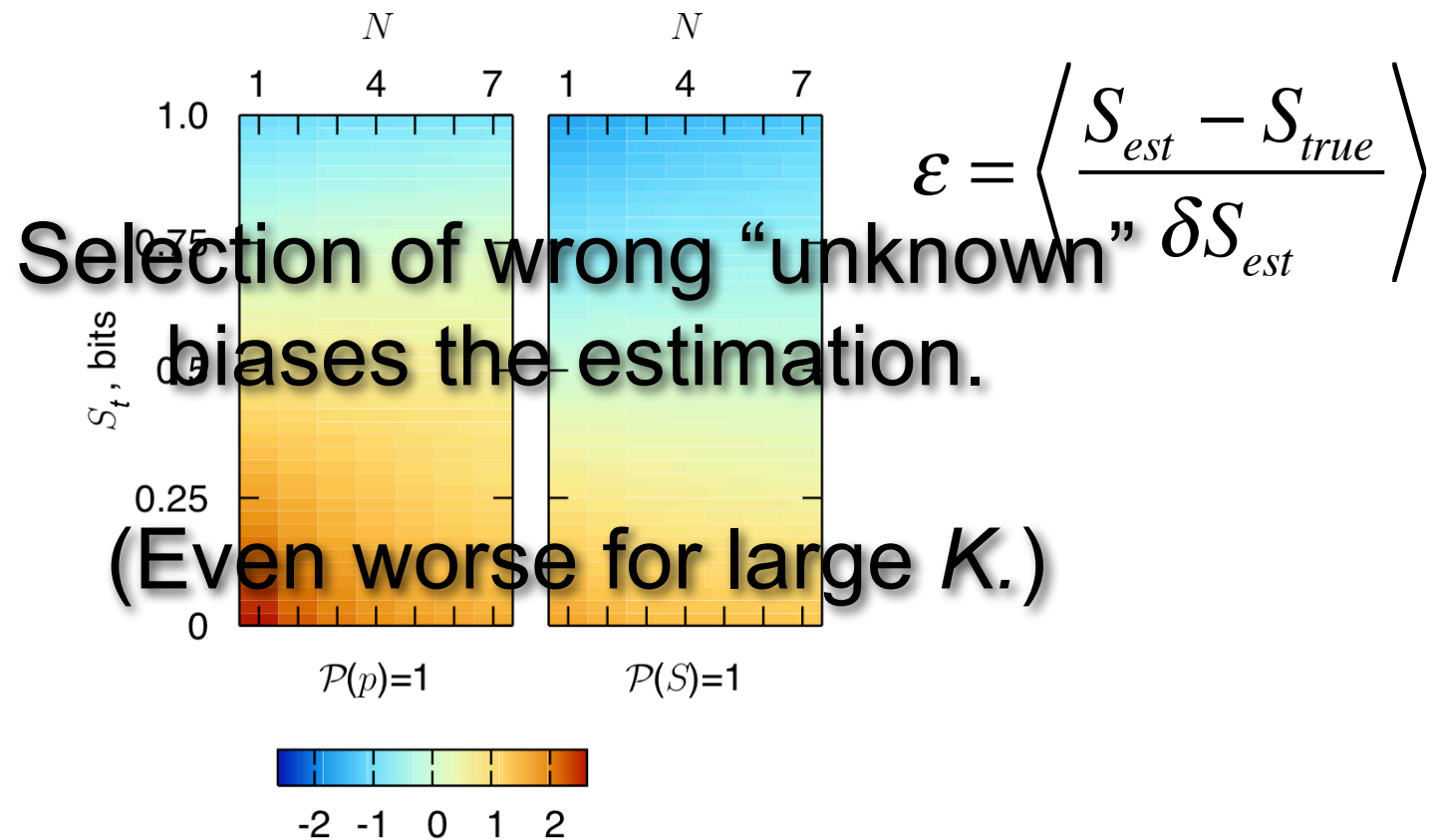
Assume (Bayes)

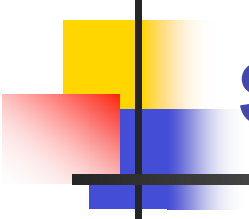
uniform (no assumptions)

p

S

What is unknown?



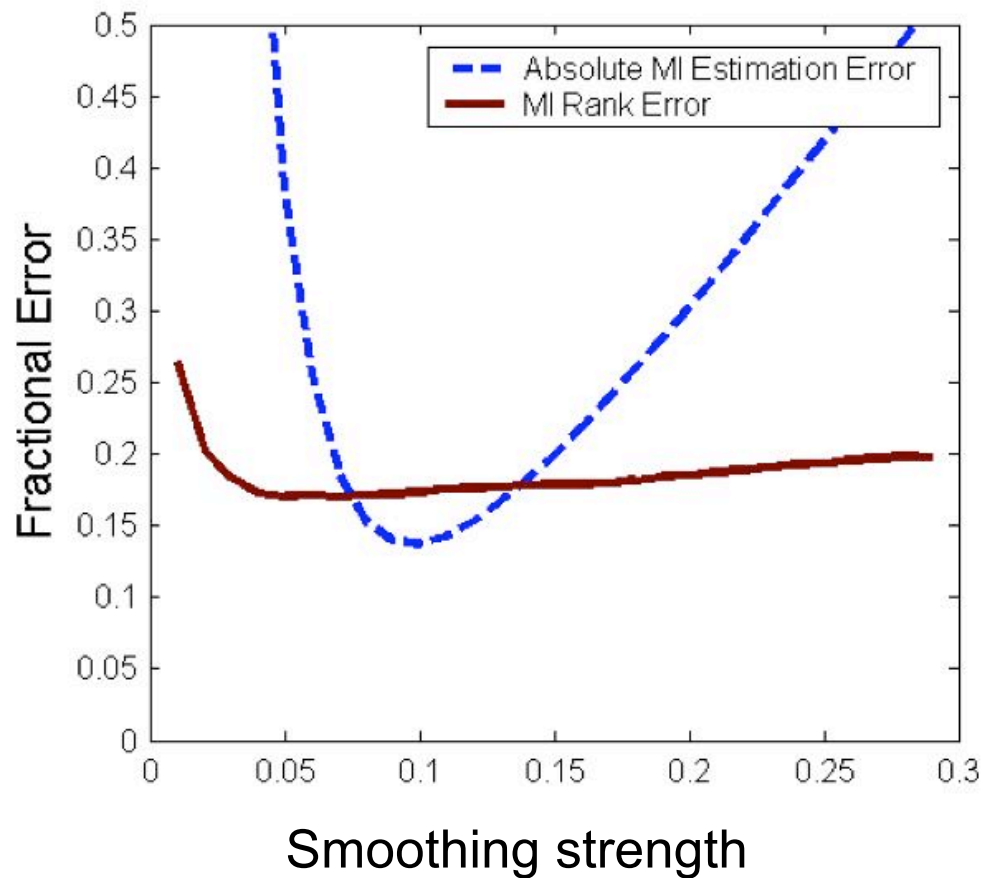


One possible uniformization strategy for S (NSB)

- Posterior variance scales as $1 / \sqrt{N}$
- Little bias, except in some known cases.
- Counts coincidences and works in Ma regime (if works).
- Is guaranteed correct for large N .
- Allows infinite # of bins.

(Nemenman et al. 2002, Nemenman 2003)

If fails: What if we need only S and $/$ ranks?



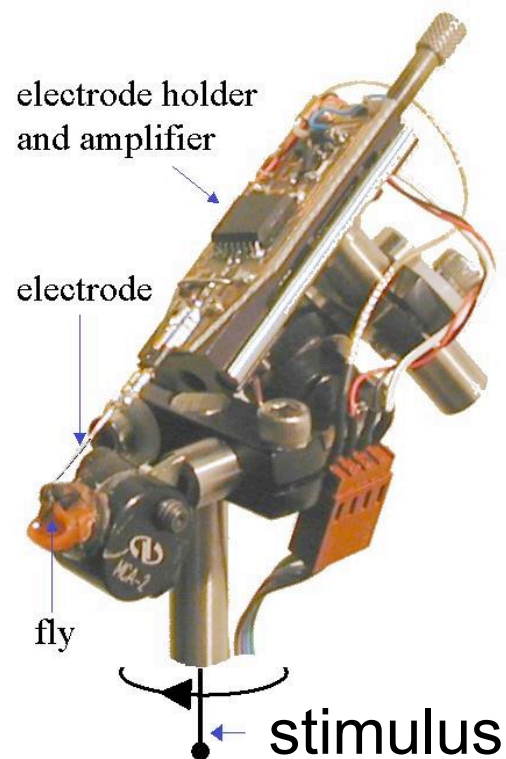


Now: apply all this to study neural coding

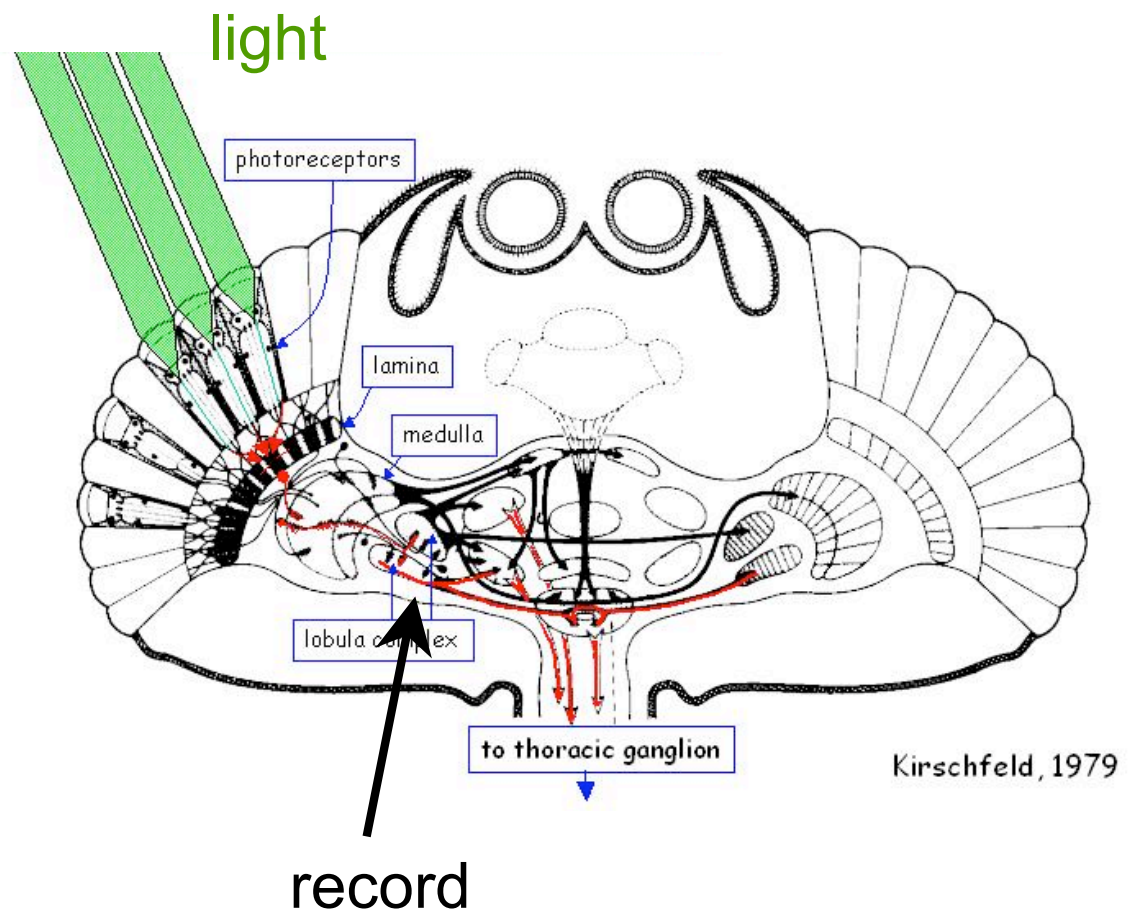
- Can we understand the code?
- Which features of it are important?
 - Is this a rate or a timing code?
- What/how much does the fly know?
- Is there an evidence for optimality?

Motion estimation is nontrivial and behaviorally important

Recording from fly's H1



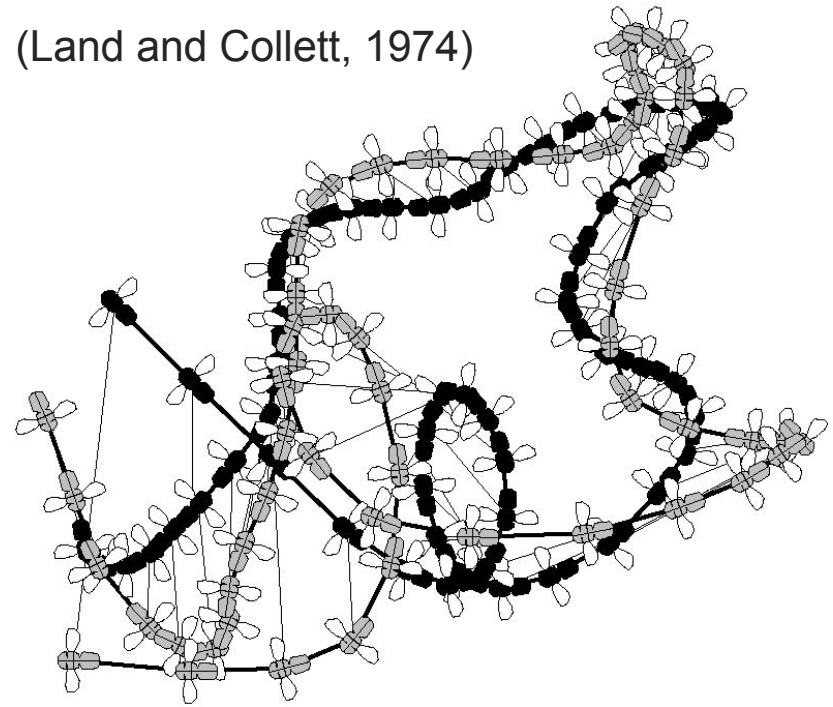
(Lewen et al, 2001)



Natural stimuli

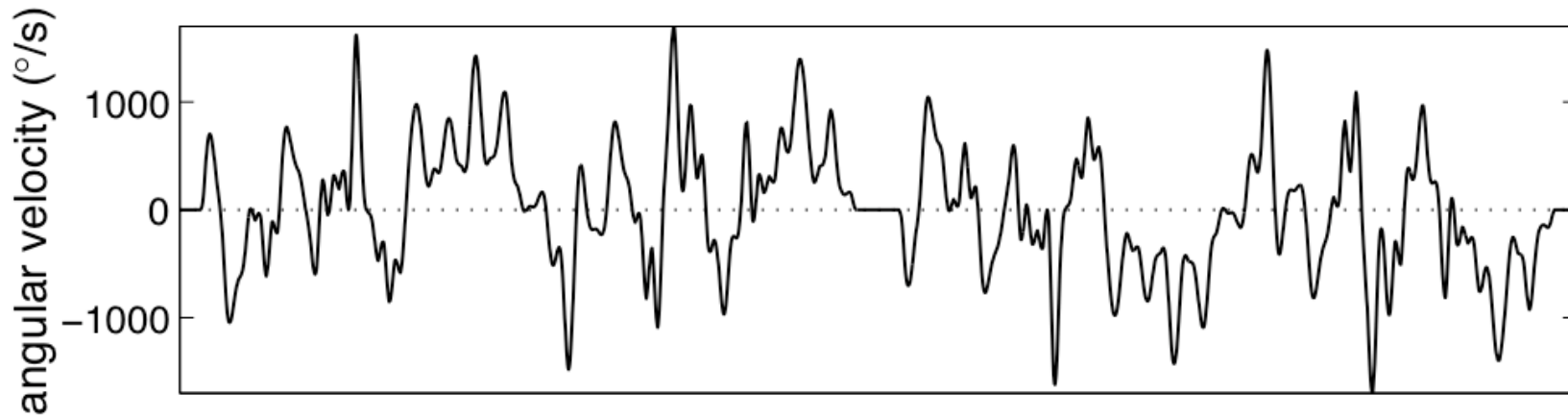
- ~2 ms resolution known to be important for white noise stimuli
- Could such “brisk” spikes be due to ~1 ms correlations in stimulus?
- What if stimulus has natural correlations?

(Land and Collett, 1974)



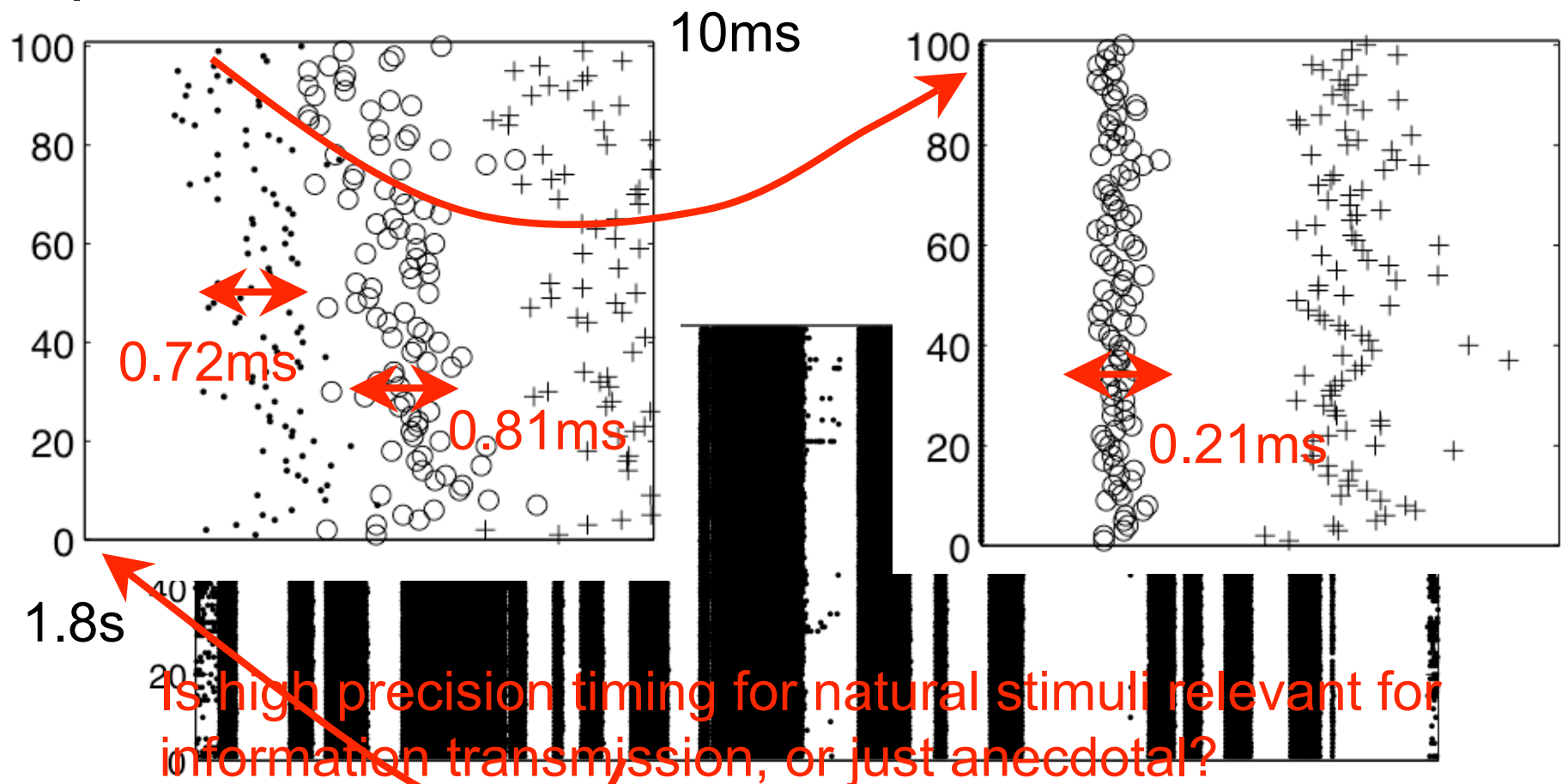
$$\tau = 60ms$$
$$\text{response} = 30ms$$

Natural stimulus and response



5s

Highly repeatable spikes (not rate coding)



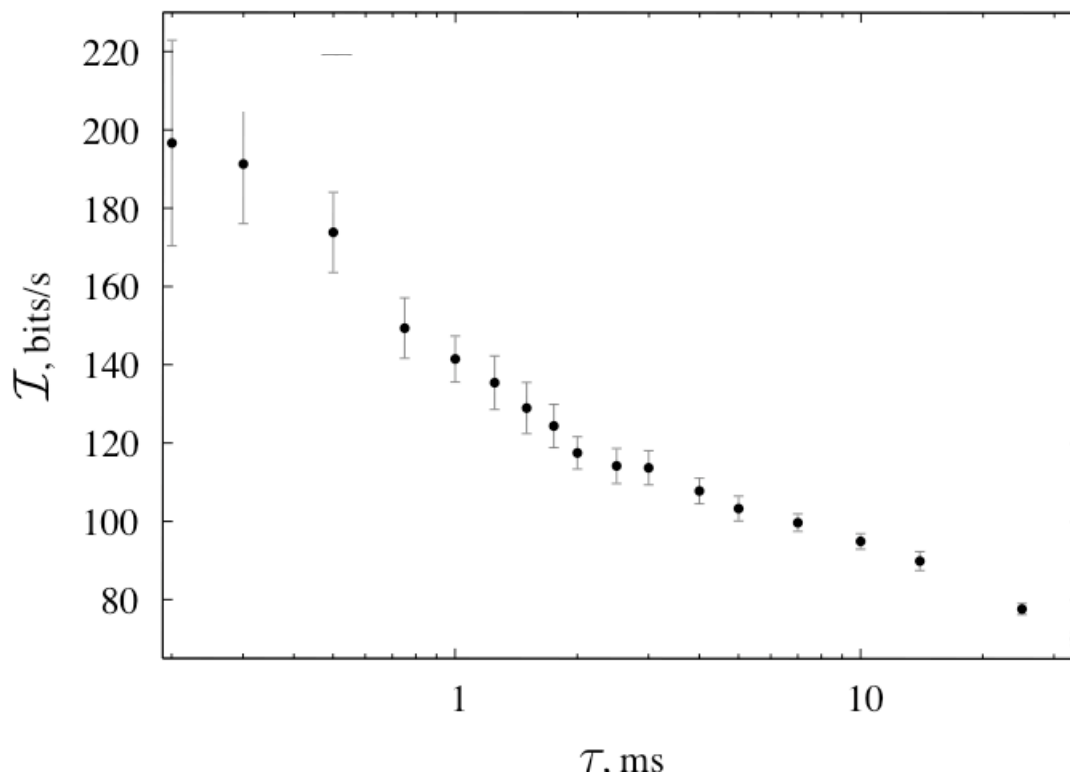


Analysis

- 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 -- words up to 150 symbols
- Severely undersampled (100 to 10000 samples). Couldn't be done before:

Use NSB!

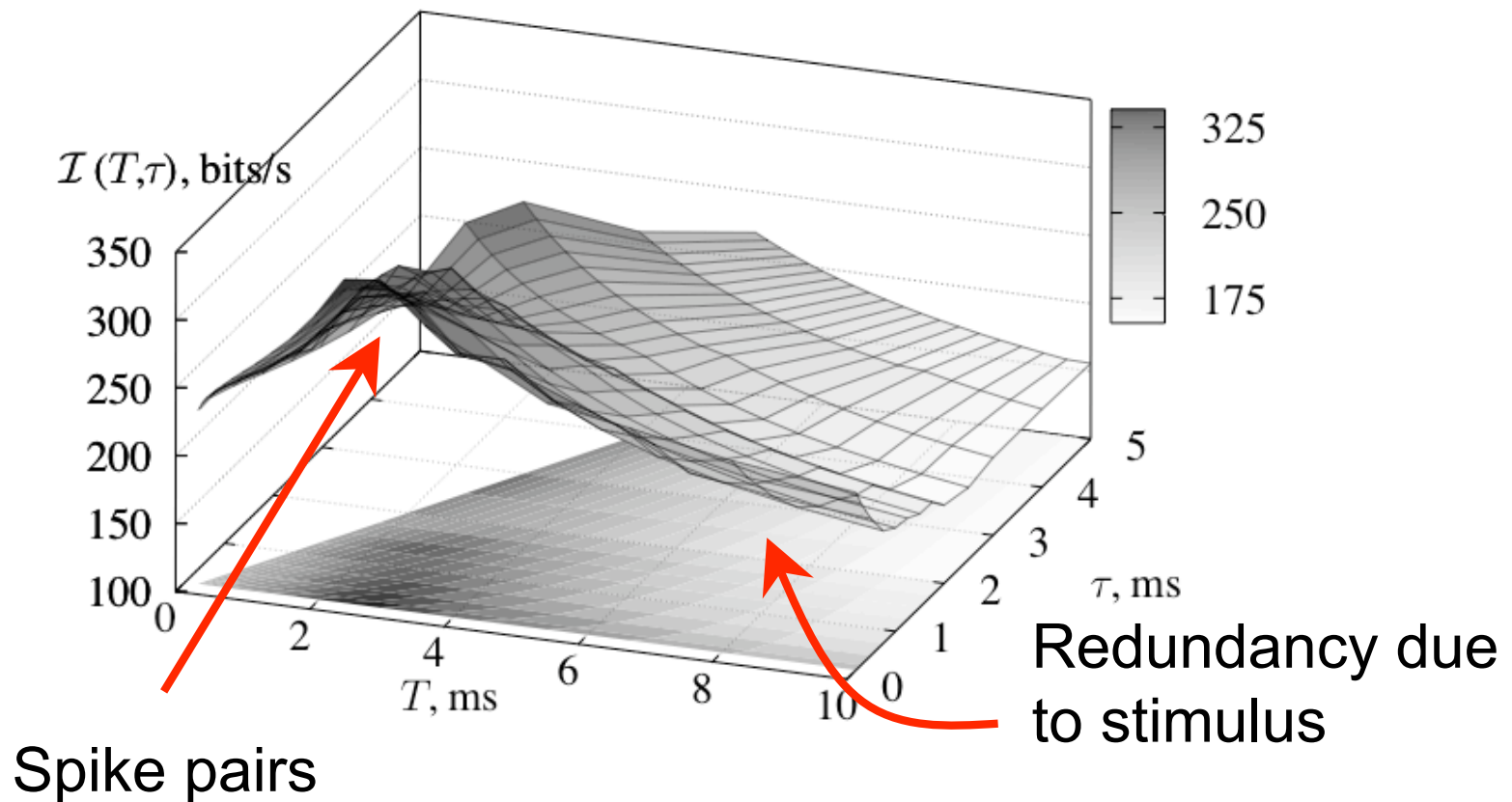
Information rate at $T=30\text{ms}$



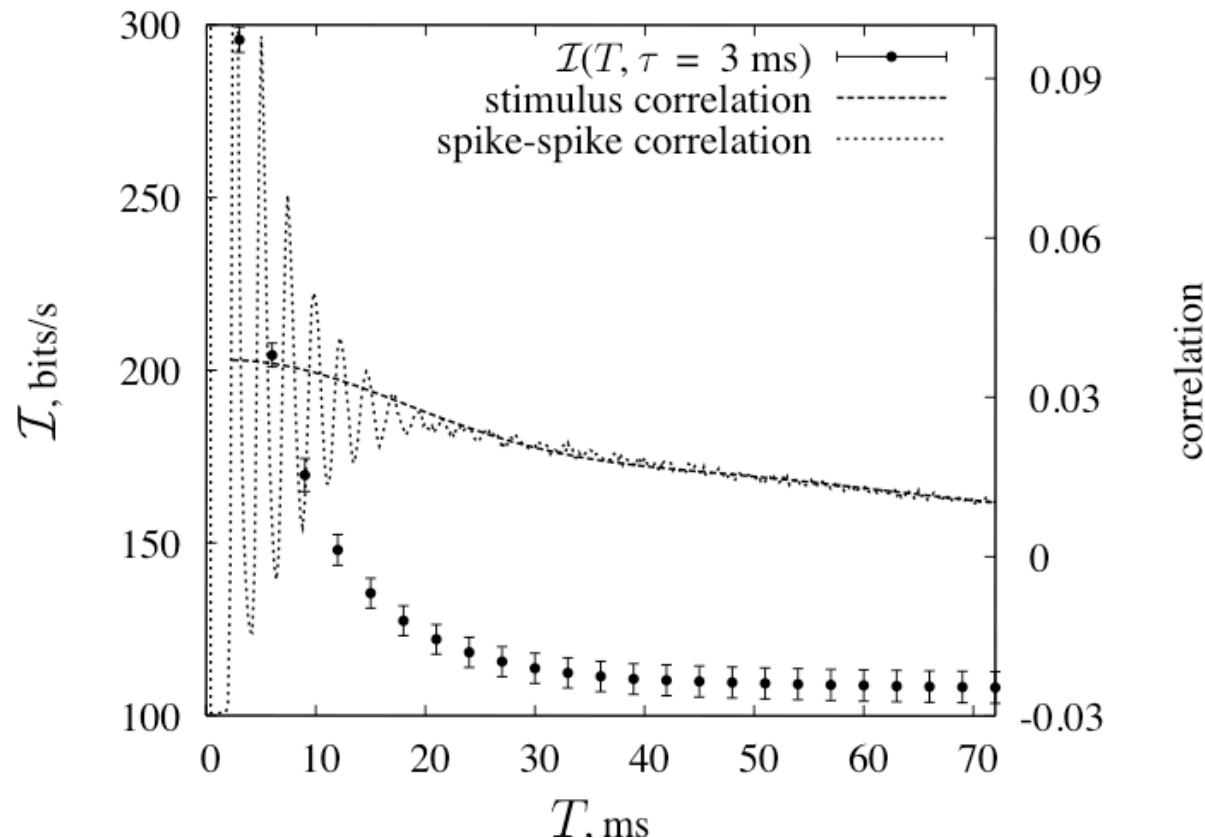
0.2 ms -- comparable to channel opening/
closing noise and experimental noise.

- Information present up to $\tau = 0.2\text{-}0.3$ 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 $\sim 75\%$ at 30ms. Optimized for natural statistics?

Synergy from spike combinations



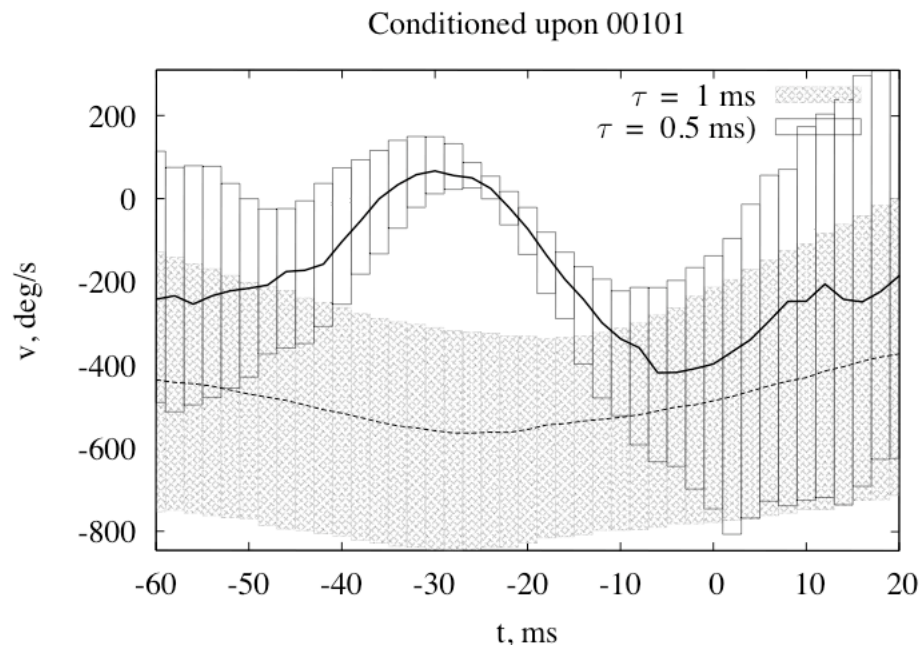
New bits



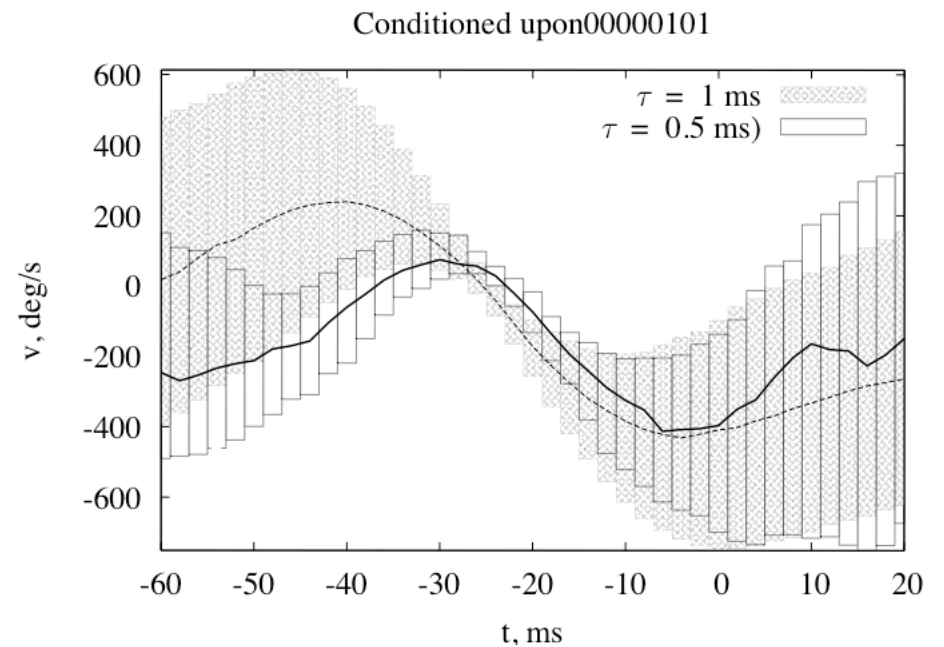
- Spikes are very regular (15 oscillat.); decoding?
- Corr. Func. at half its value, but fly gets new bits every 30 ms
- Independent info (even though entropies are T dependent).

Behaviorally optimized code!

Information about...



Signal shape



Zero-crossings time

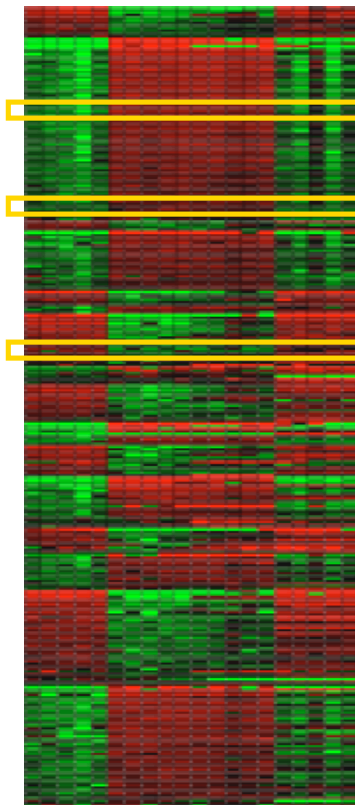
Best estimation at 25ms delay. Little time for reaction.



Same IT techniques needed (have been used) to study:

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

Reconstructing interaction networks



A

Small data requirements

Stat Co GM Biochem.

✗✓

✓

✗✓

✗

B

Robustness to fluct.

✓

✓

✗✓

✗

Computational complexity

✗

✓

✗

✗✓

C

Conditional interactions

✓

✗✓

✓

✗✓

Reparam inv., non-param.

✗✓

✗✓

✗✓

✓

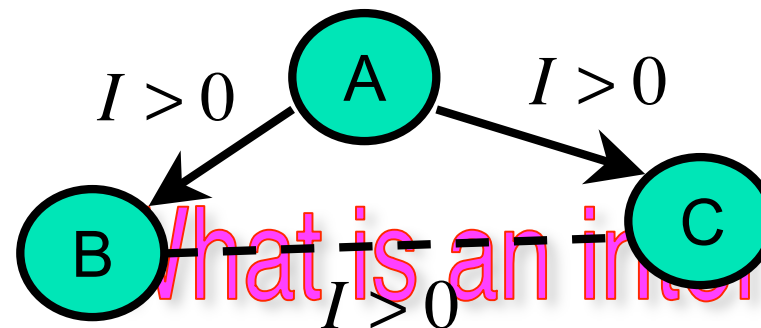
Irreducibility

✓

✗

✓

✗



What is an interaction in IT?



Two *separate* problems

- What is an interaction?
- Realistic algorithm to uncover them



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 ?
(KS test, etc.)



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

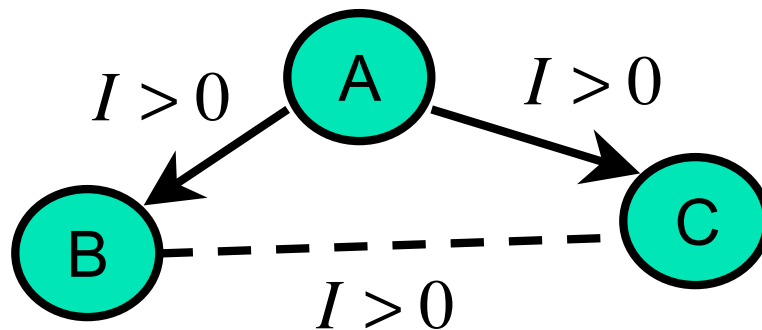
$$I_{XYZ} = \left\langle \log \frac{p_{xyz}}{p_x p_y p_z} \right\rangle$$

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

But these are not irreducible.

(Nemenman and Tishby 2005)

By analogy: Example of irreducibility



$$P_{ABC} = \frac{P_{AB}P_{AC}}{P_A} = \frac{1}{Z} f_{AB} f_{BC}$$

MaxEnt approximation without BC:

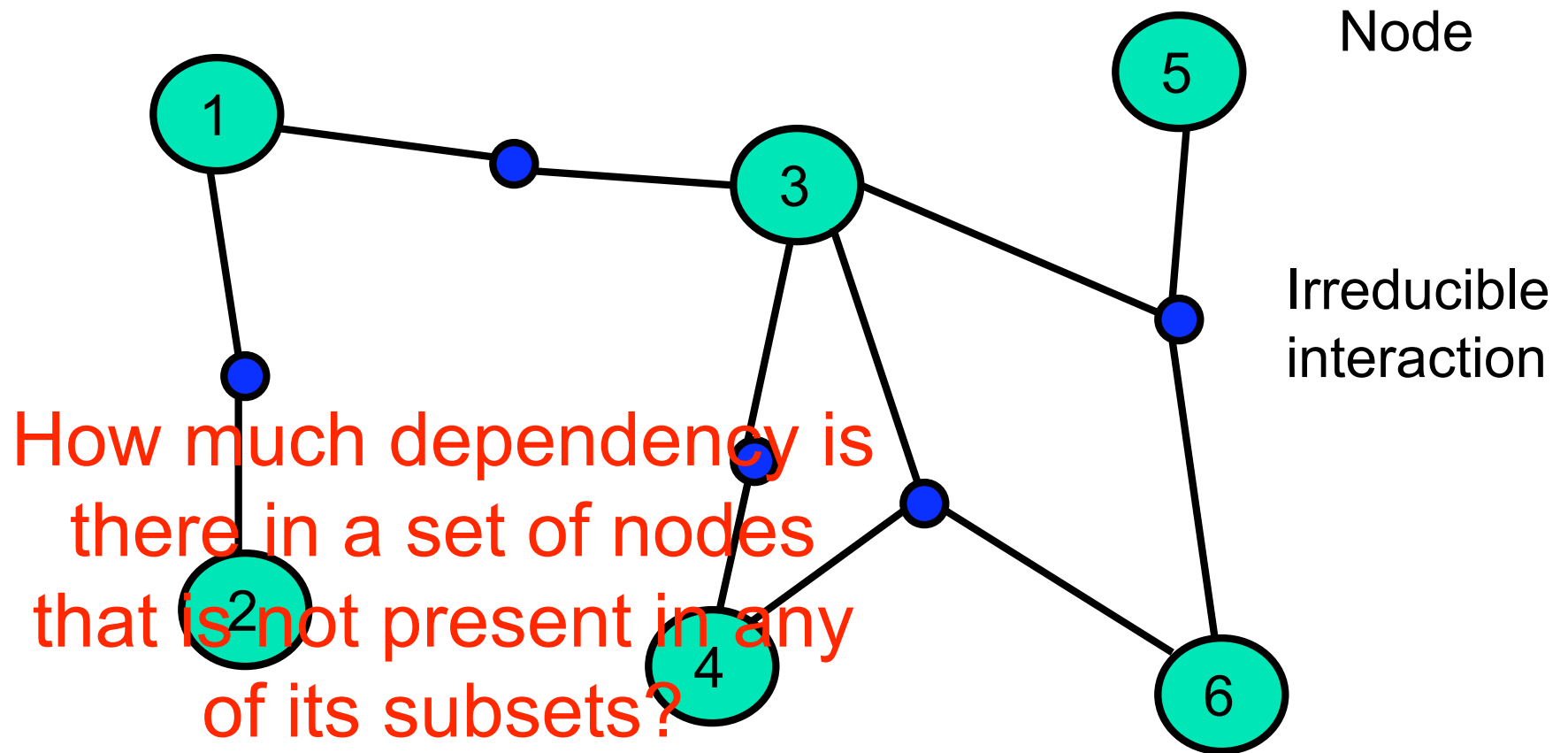
$$Q_{ABC} = \frac{1}{Z} \exp(-\varphi_{AB} - \varphi_{AC}) \Rightarrow D_{KL}[P_{ABC} \parallel Q_{ABC}] = 0$$

No irreducible interaction!

For other links, e.g., AB: $D_{KL}[P_{ABC} \parallel Q_{ABC}] > 0$

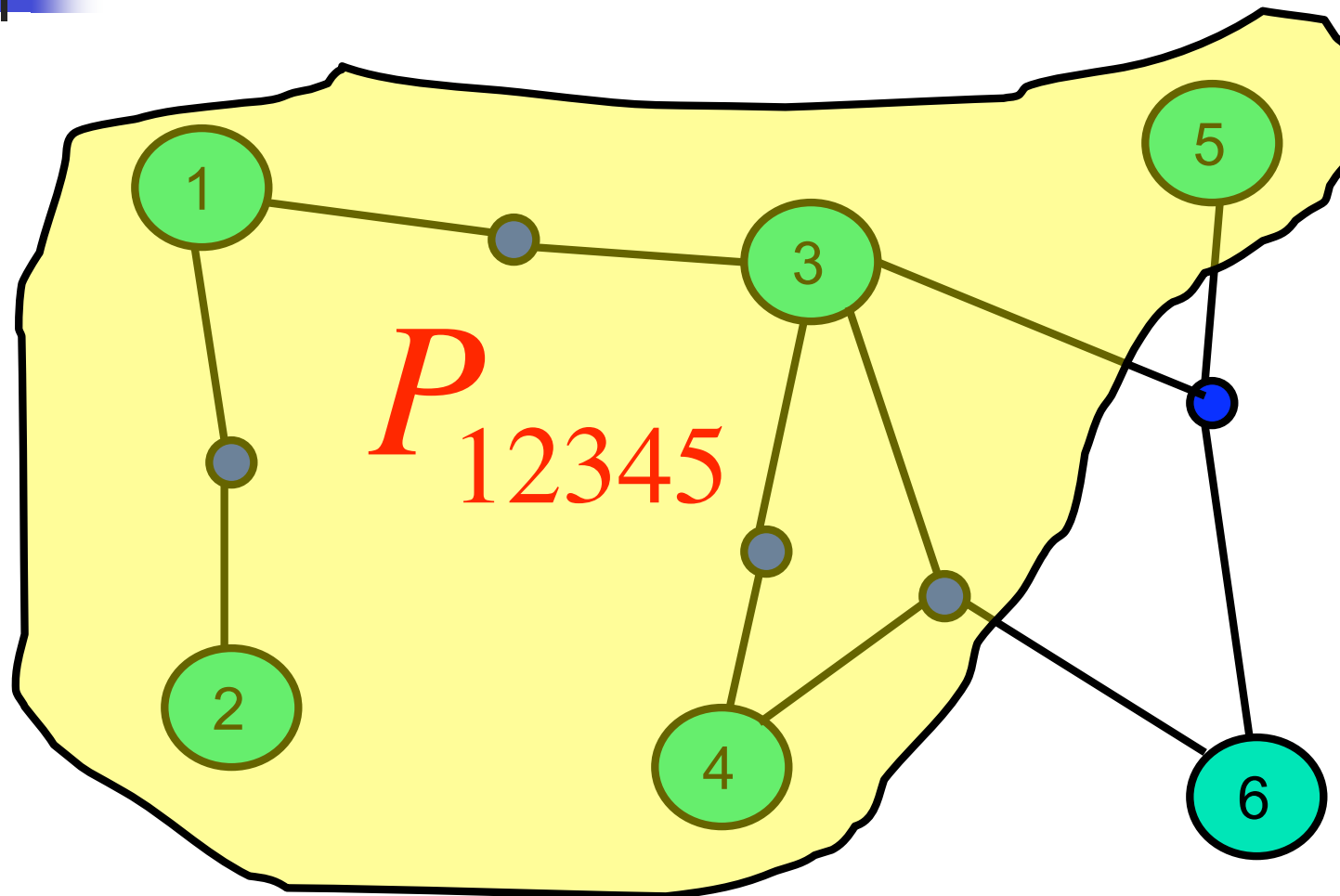
Irreducible interaction.

Higher order irreducible dependencies

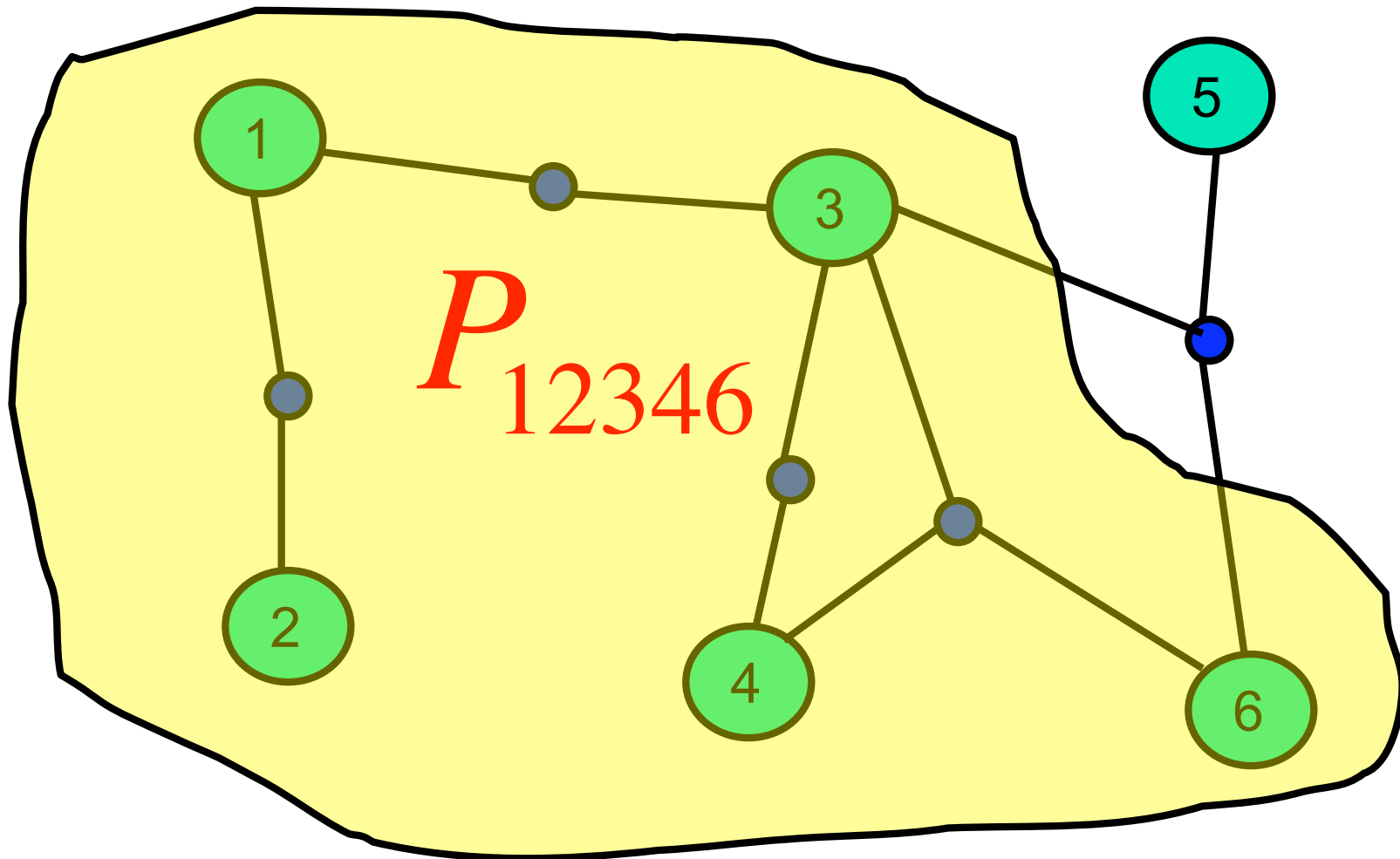


(Schneidman et al. 2003, Nemenman 2004)

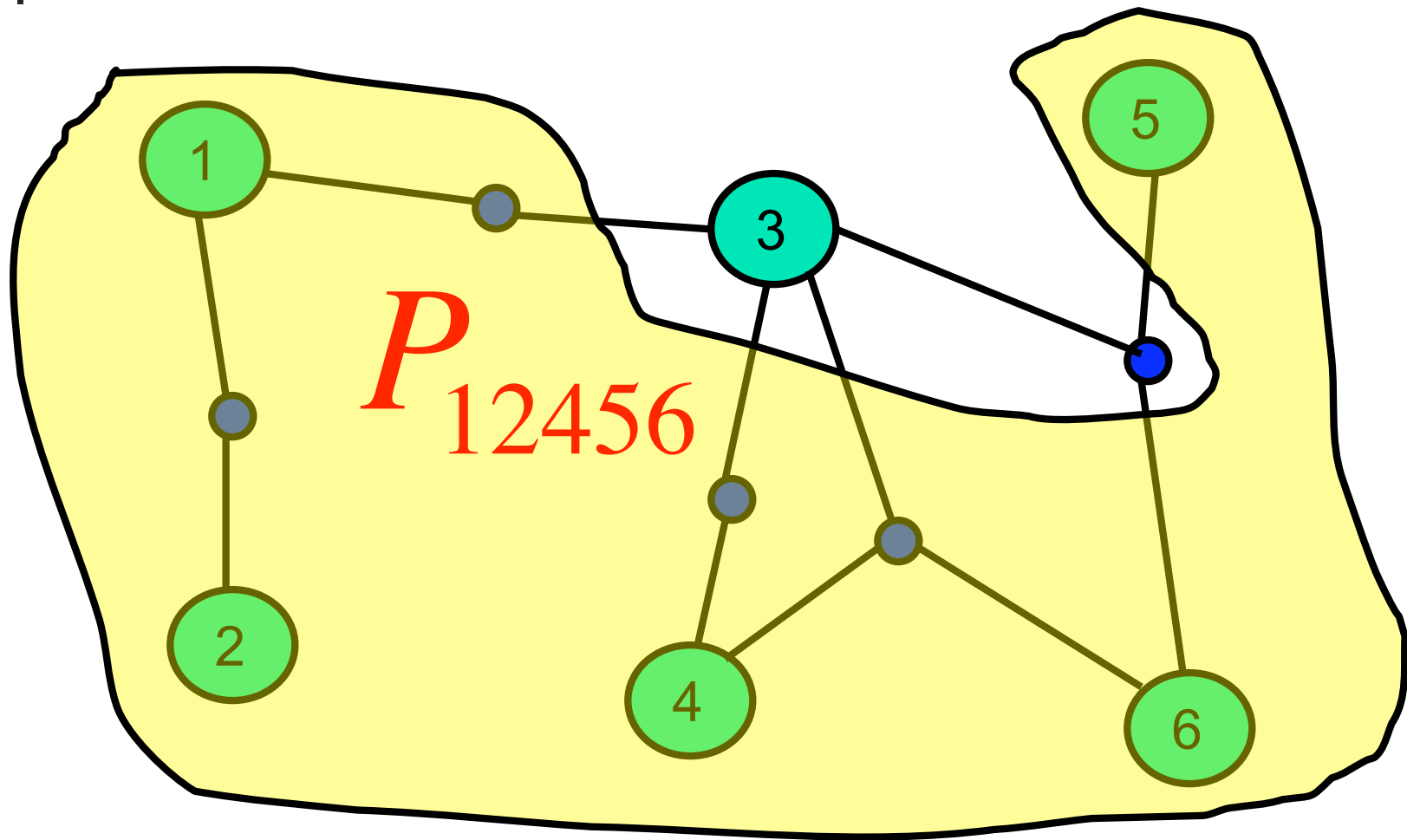
MaxEnt approximations



MaxEnt approximations



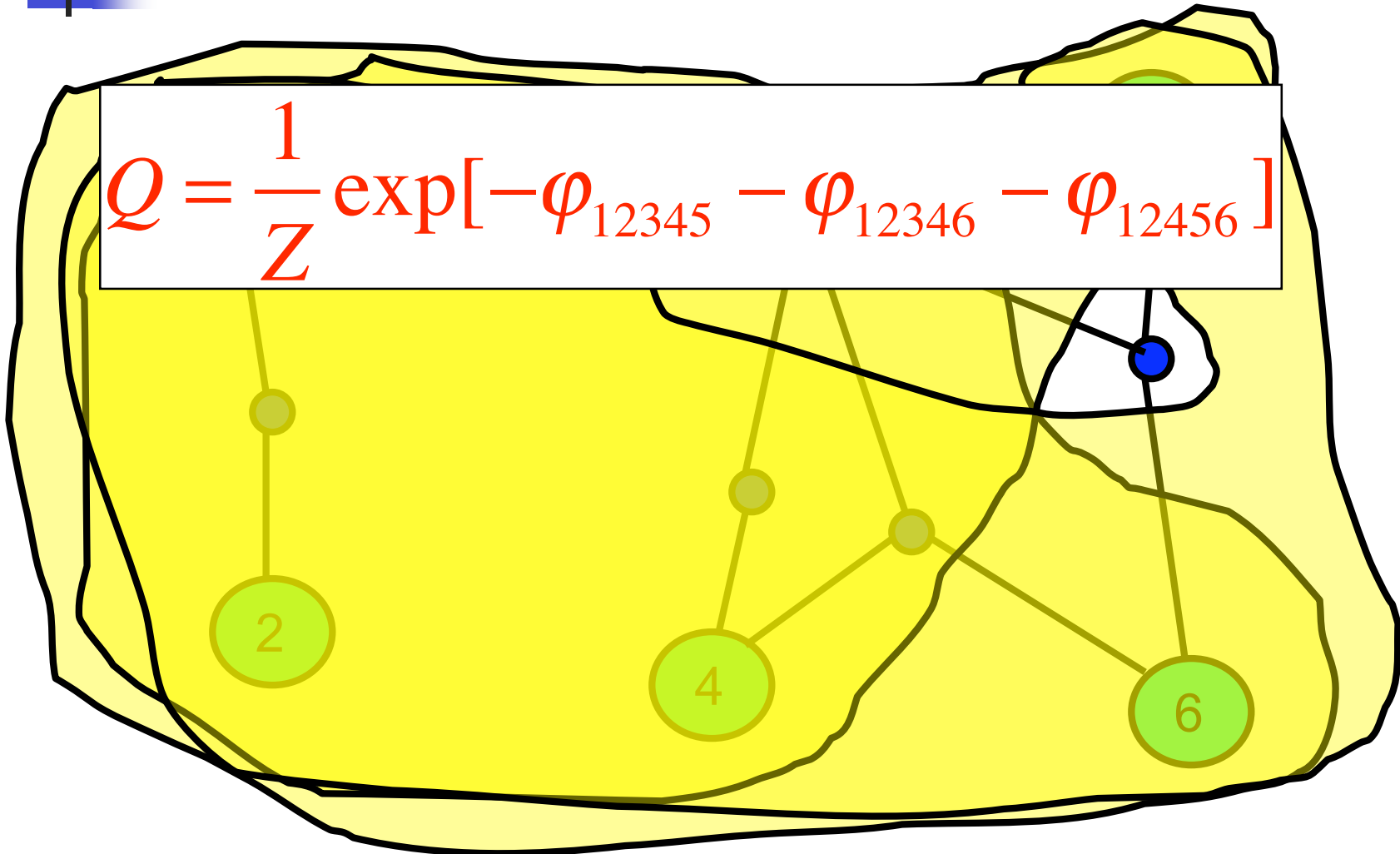
MaxEnt approximations





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

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

$I'_{356} > 0 \Rightarrow$ 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



Two *separate* problems

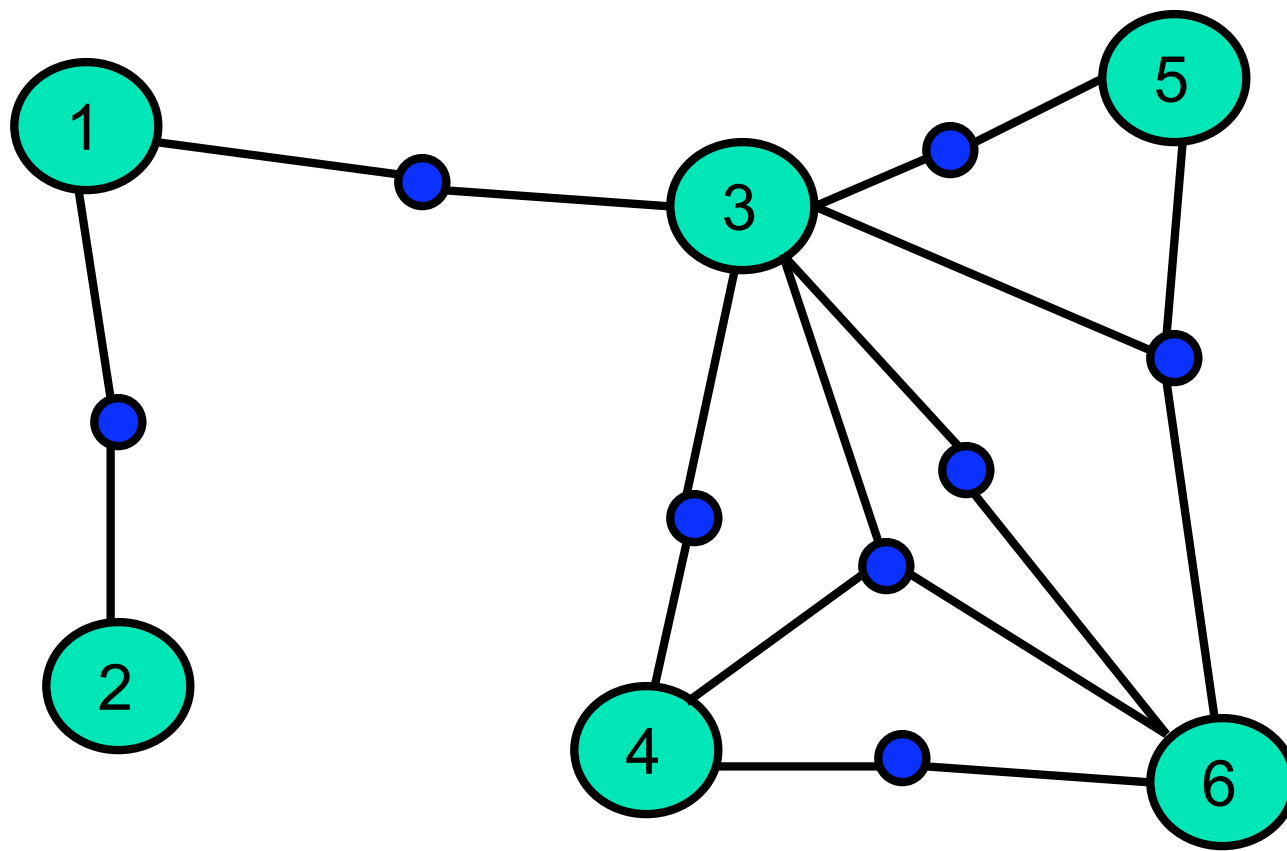
- What is an interaction?

An irreducible statistical dependency.

- Realistic algorithm to uncover them

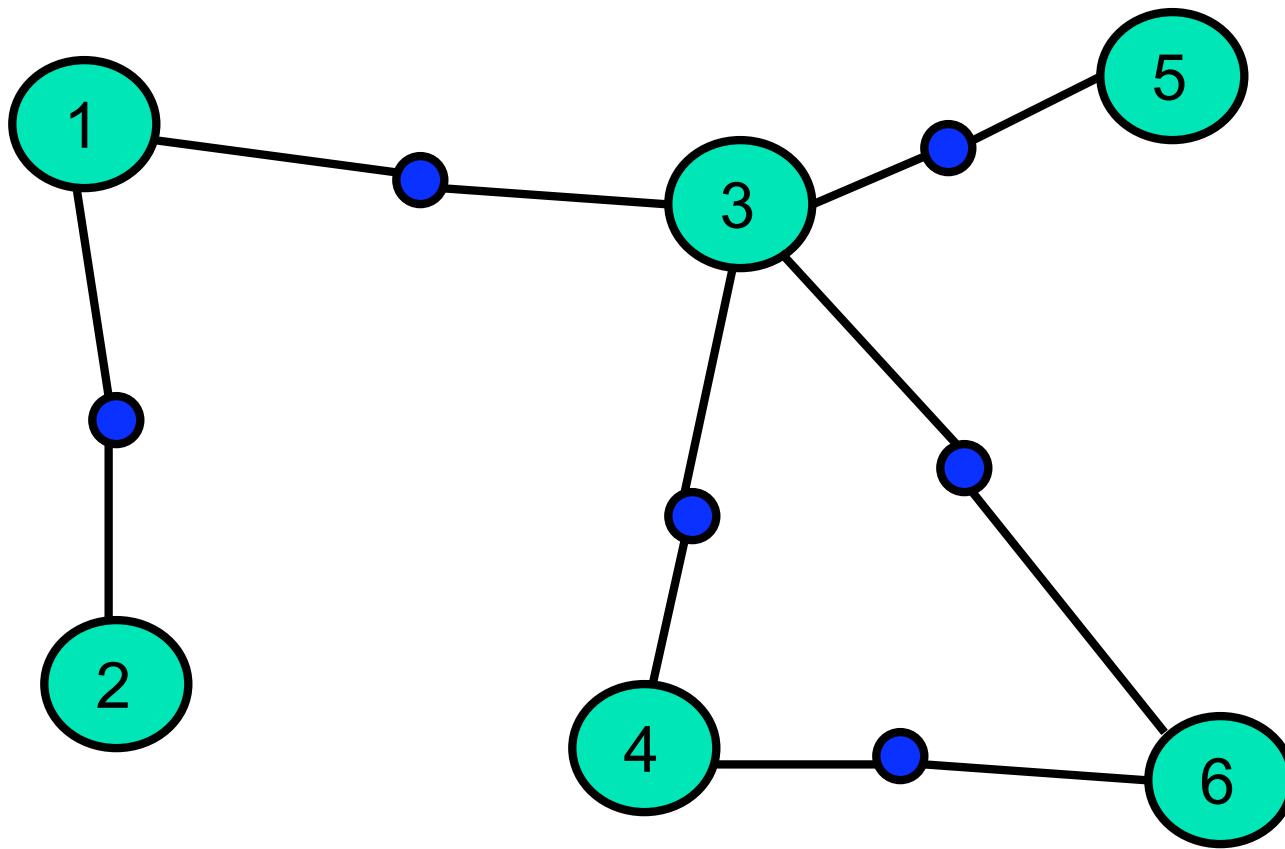
- Biologically sound assumptions (new knowledge from verifying assumptions).
- Know the order.
- Focus on high precision (irreducibility, no false positives), not so much on high recall (no false negatives).

Interaction network



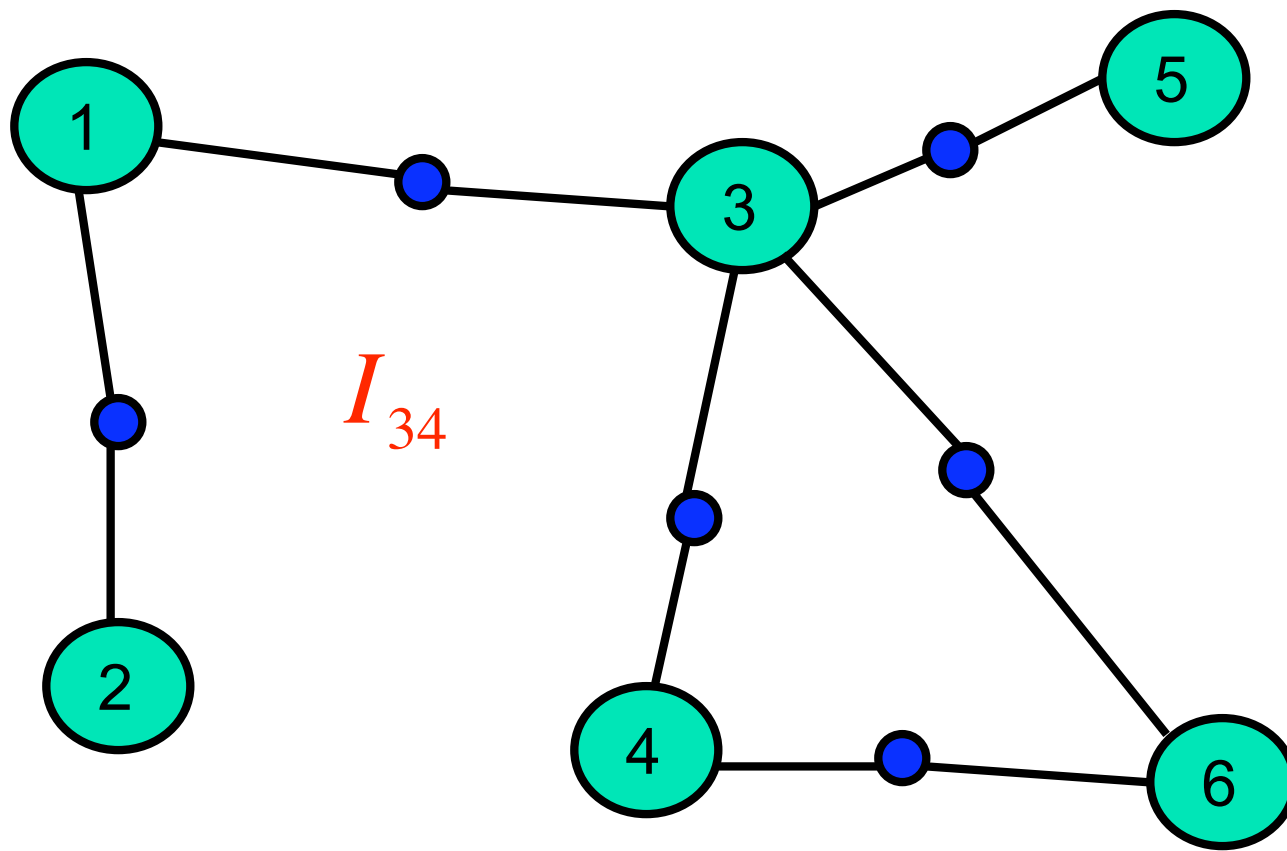
(Basso et al. 2005, Margolin et al. 2005)

Disregard
high orders (few data)

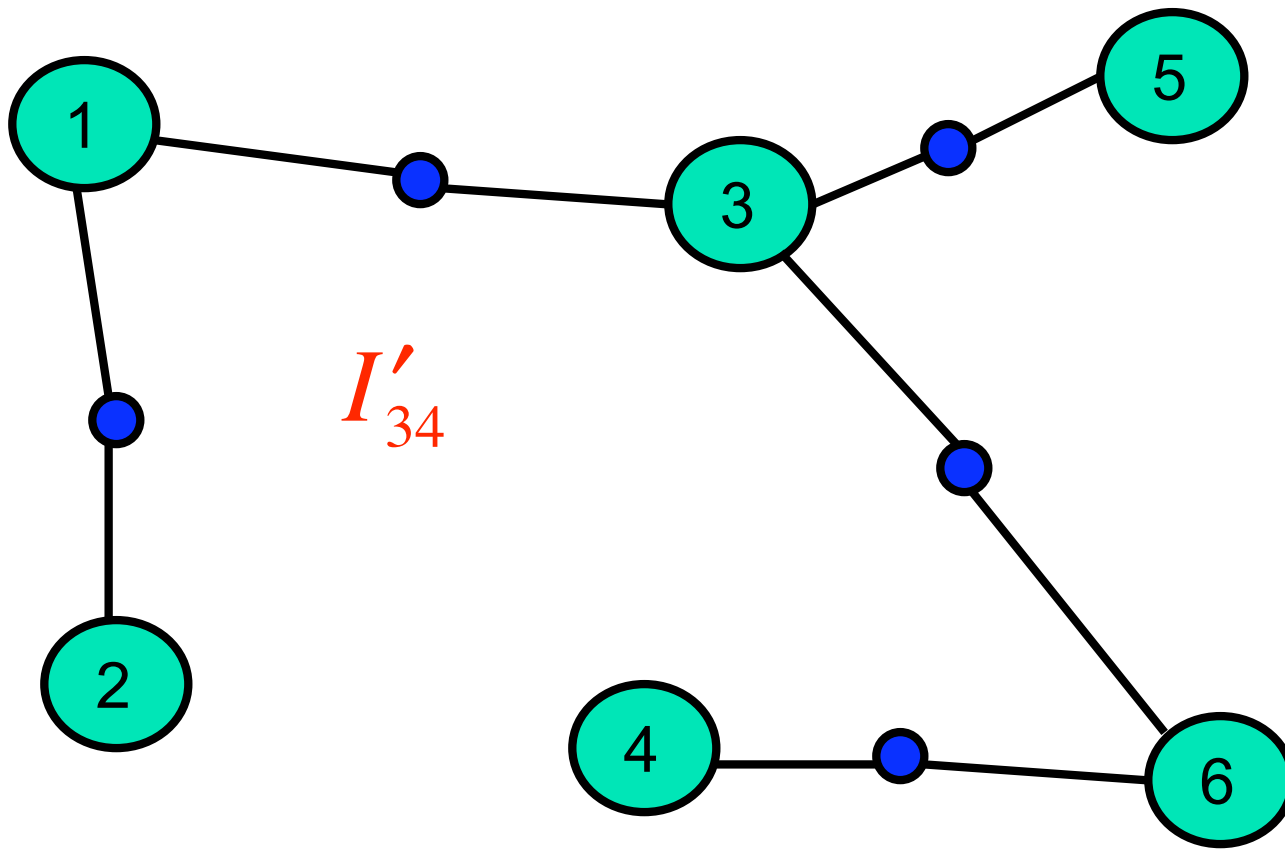




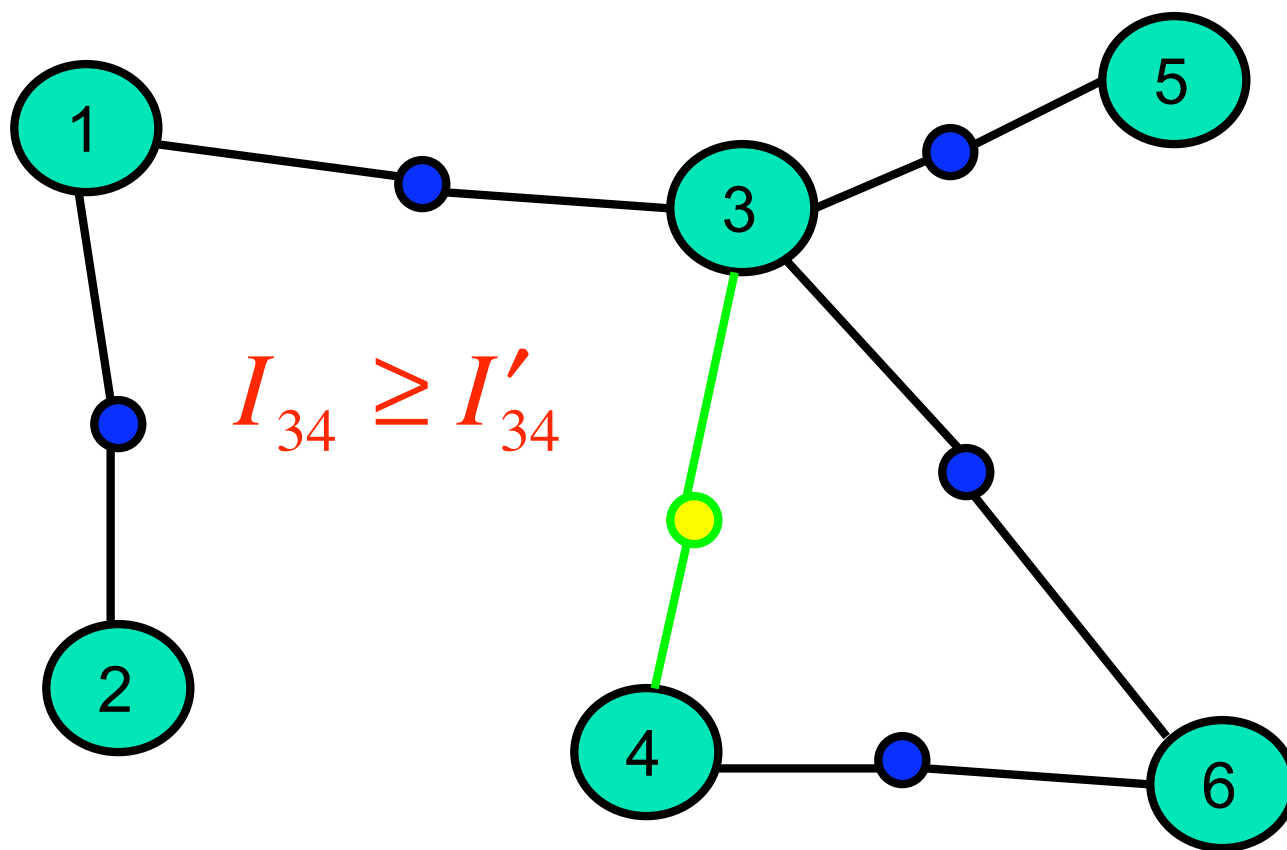
Locally tree-like approximation



Locally tree-like approximation

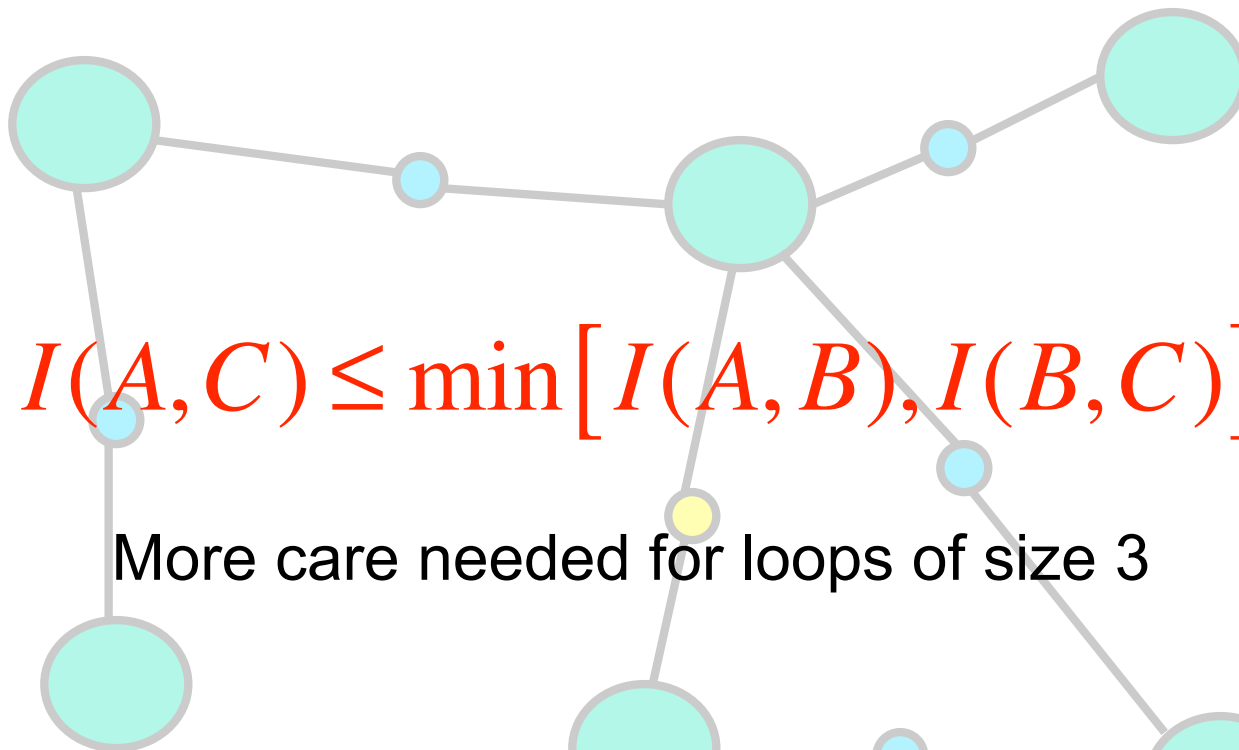


Locally tree-like:
signals decorrelate fast



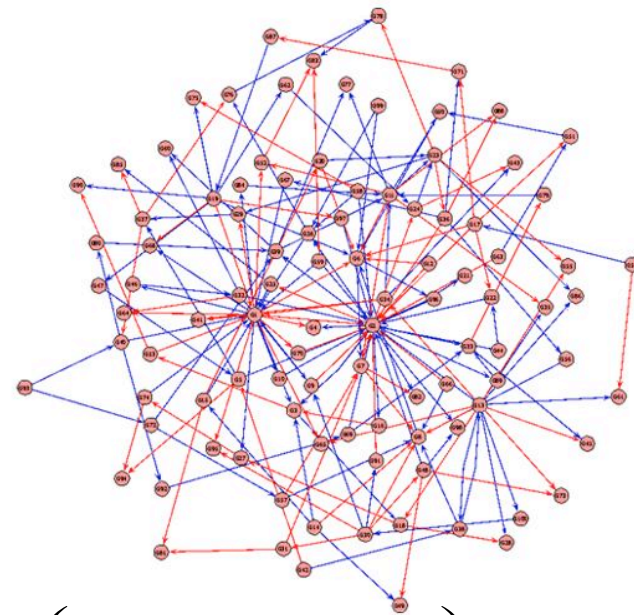
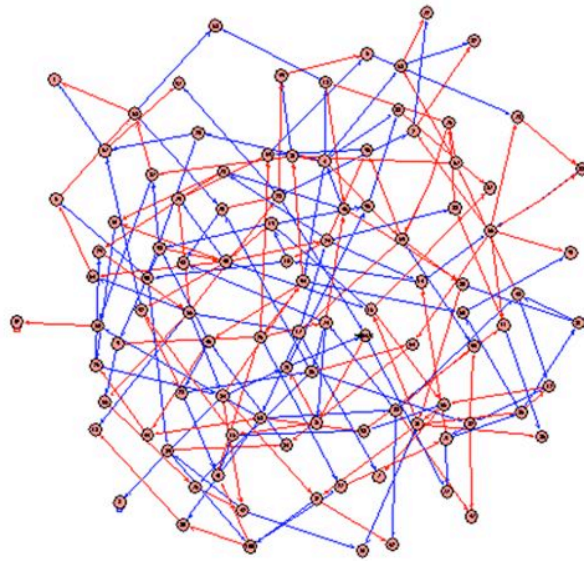
ARACNE: No false positives

Where 2-way -- it's 2-way



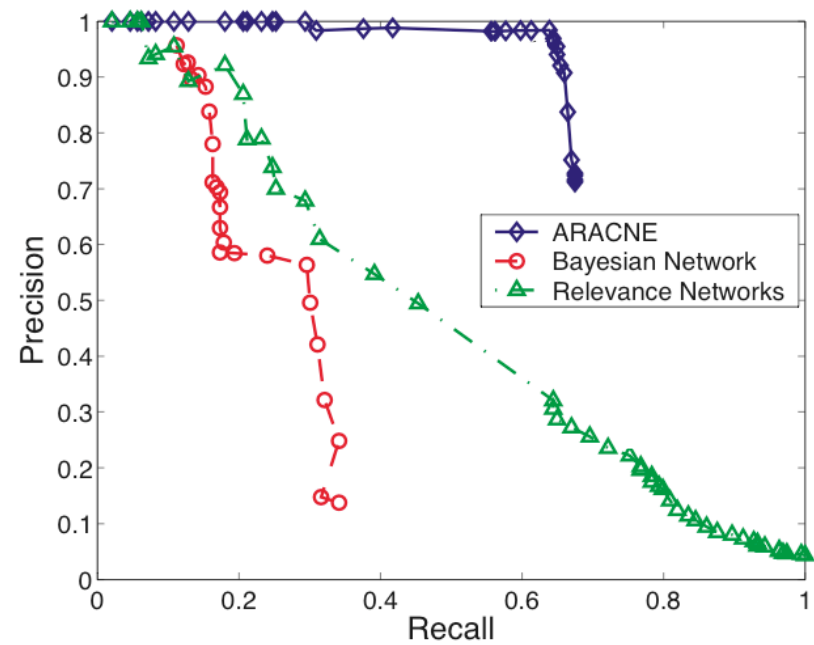
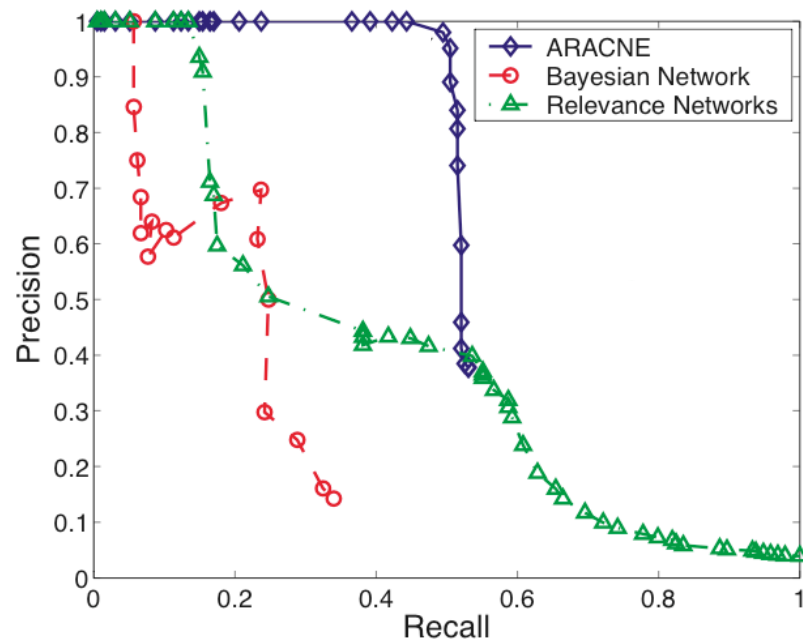
Techniques for MI estimation needed again!

Synthetic networks



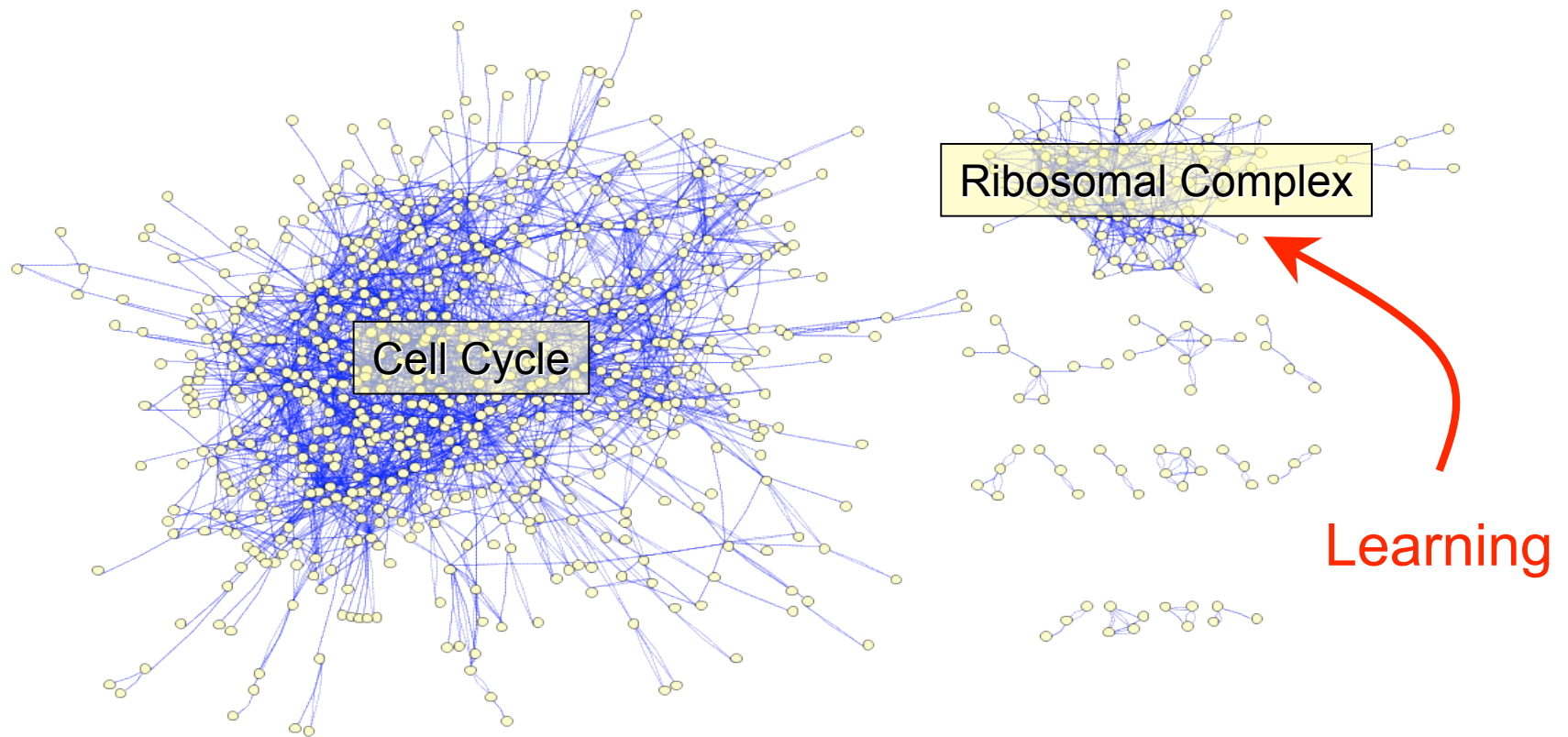
$$\frac{dx_i}{dt} = a_i \prod_j \frac{I_{0,j}^{v_j}}{I_j^{v_j} + I_{0,j}^{v_j}} \prod_j \left(1 + \frac{A_{0,j}^{v_j}}{A_j^{v_j} + A_{0,j}^{v_j}} \right) - b_i x_i$$

Synthetic networks benchmarks ($N=1000$)



Graceful decay for smaller N

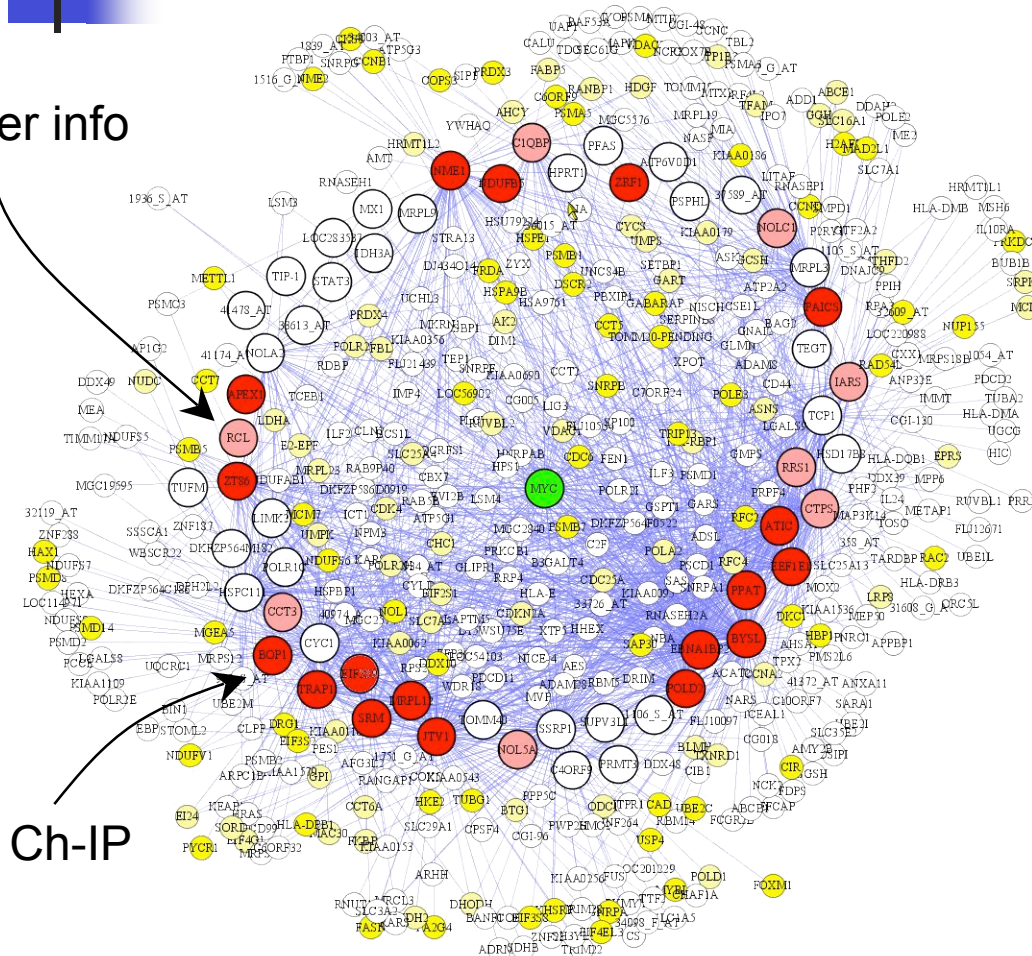
Complete B-cell network (400 arrays)



~129000 interactions

c-MYC subnetwork

other info

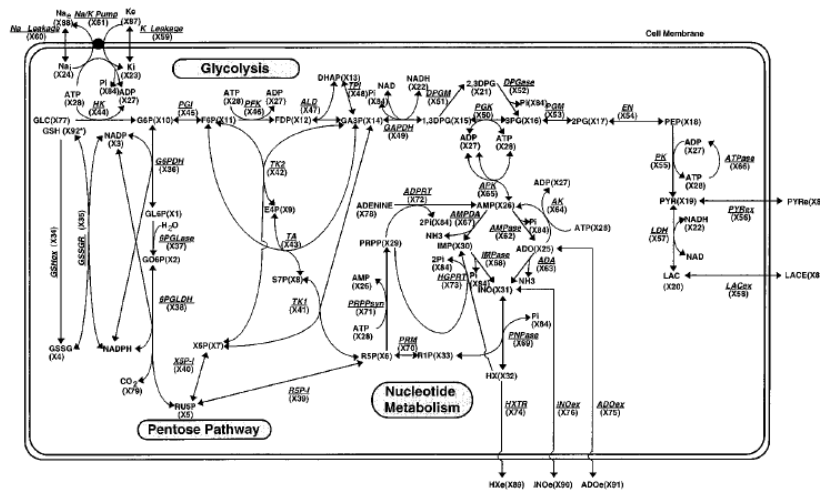


- Protooncogene,
- 12% background binding,
- one of top 5% hubs
- significant MI with 2000 genes

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

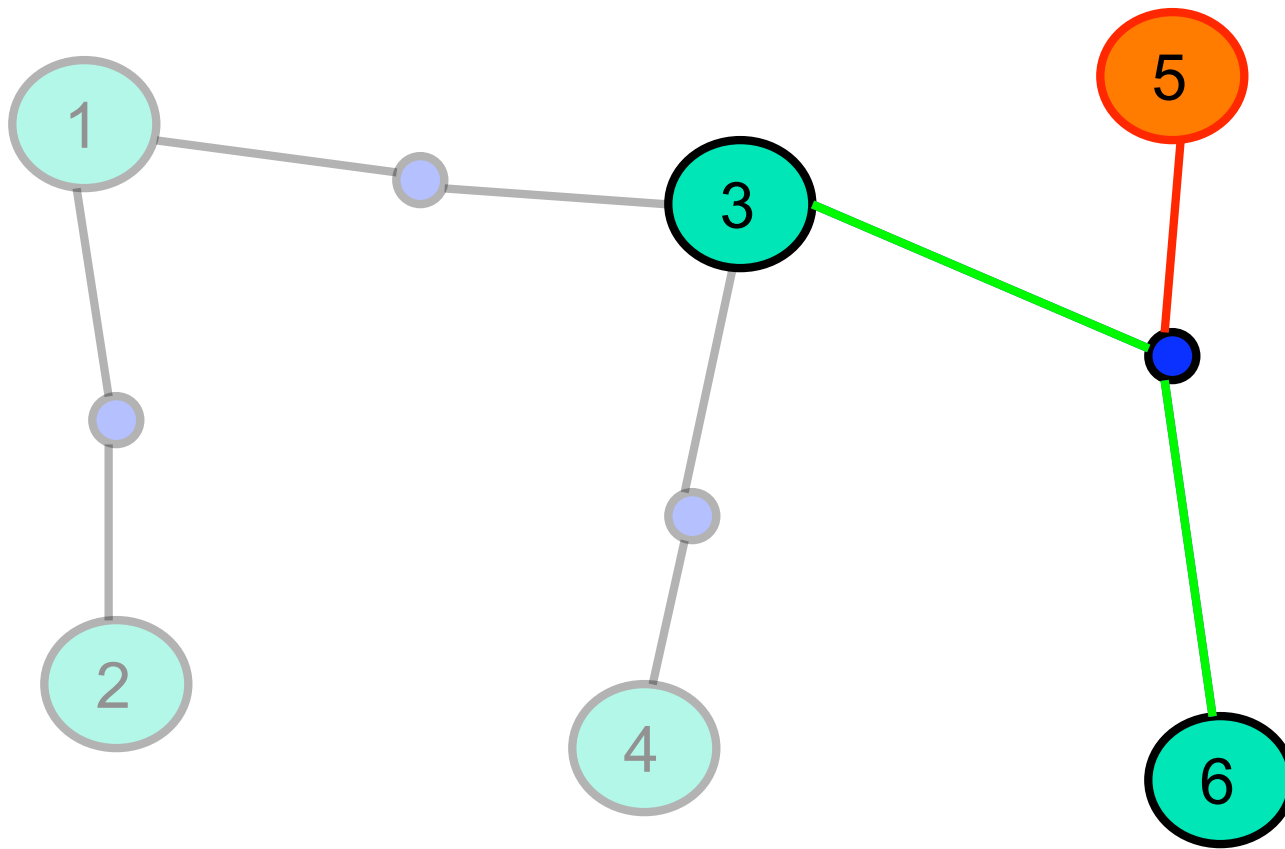
Also validated in...

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



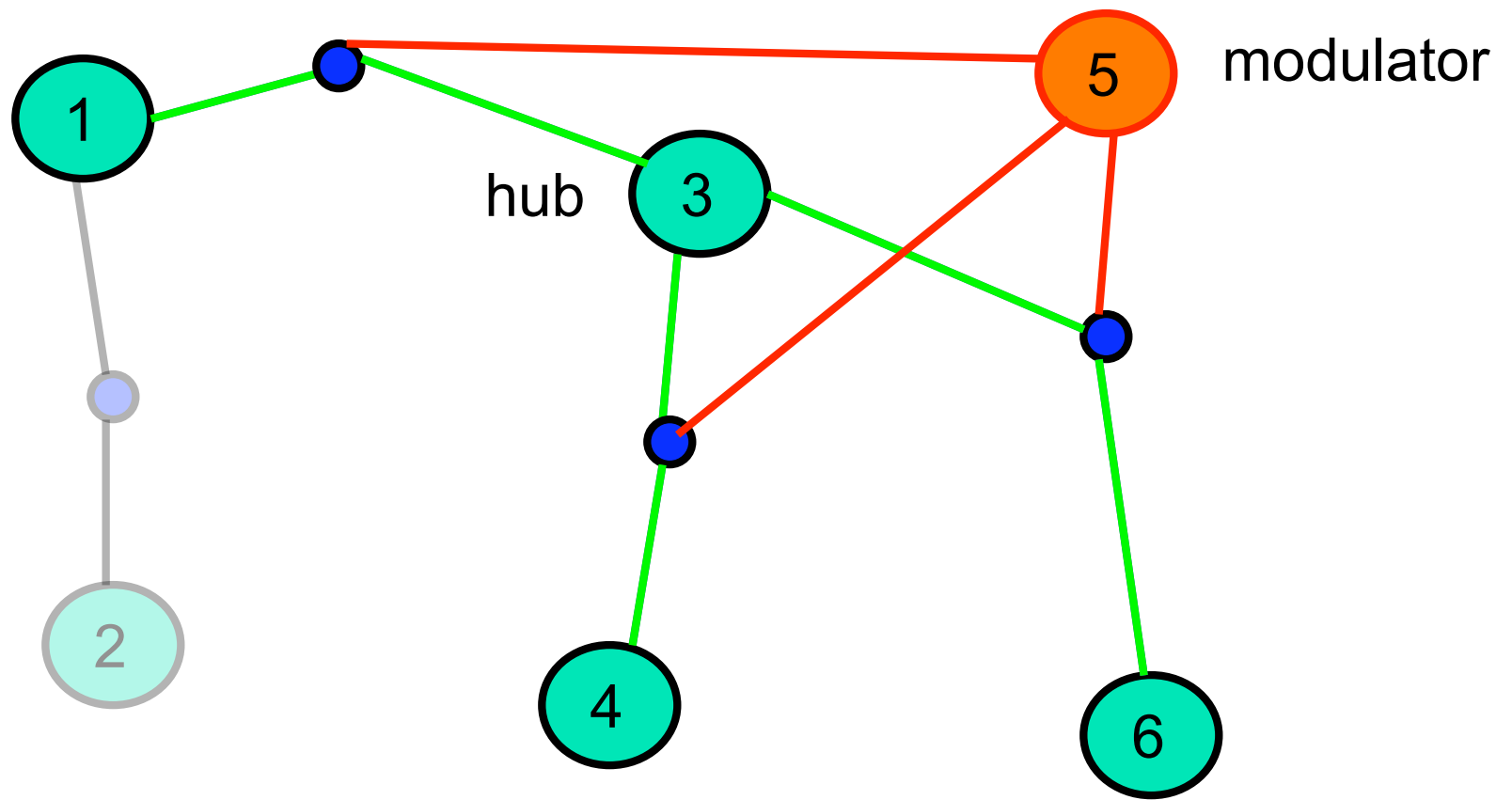
~80% precision
20-80% recall (depending on N)

3rd order interactions (modulated, conditional)



Nontranscriptional modulators!

Computational constraint: large modulators/hubs only





3rd order interactions

- Focus on important hubs (c-MYC)
- Pre-filter candidate modulators by dynamic range and other conditions.
- Find modulators whose expression inflicts large changes on hubs' interactions
- No guarantee of irreducibility
- Validate in GO w.r.t. to transcription factors and kinases among modulators

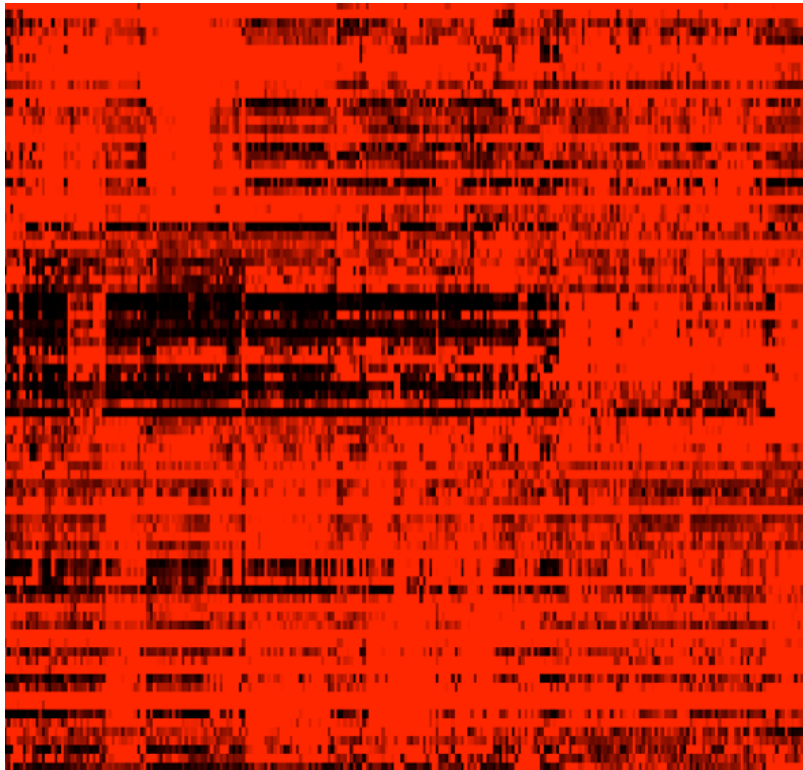


c-MYC modulators

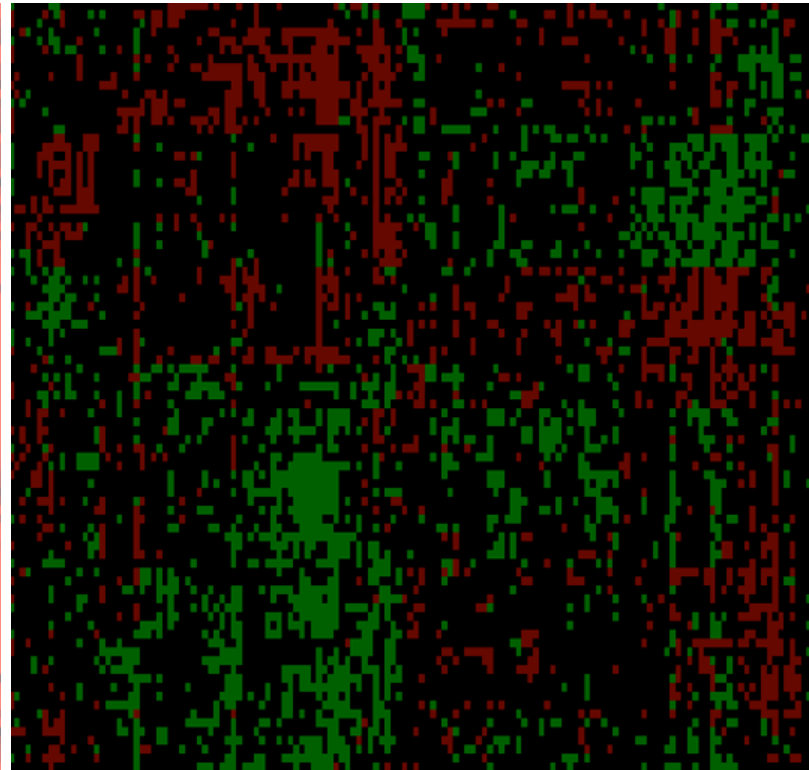
- 1117 candidate modulators (825 with known molecular function in GO)
- 82 (69) candidate modulators identified
- Kinases: 10/69 (backgr. 42/825), $p=1e-3$
- TFs: 15/69 (backgr. 56/825), $p=1e-6$; binding signature for co-TFs (E2F5, MEF2B) found.
- Total: 25/69 (backgr. 98/825), $p=3e-8$
- Other modulators: ubiquitin conjugating enzyme, mRNA stability, DNA/chromatin modification, known protein-protein target.

Many correlated modulators

|expression|

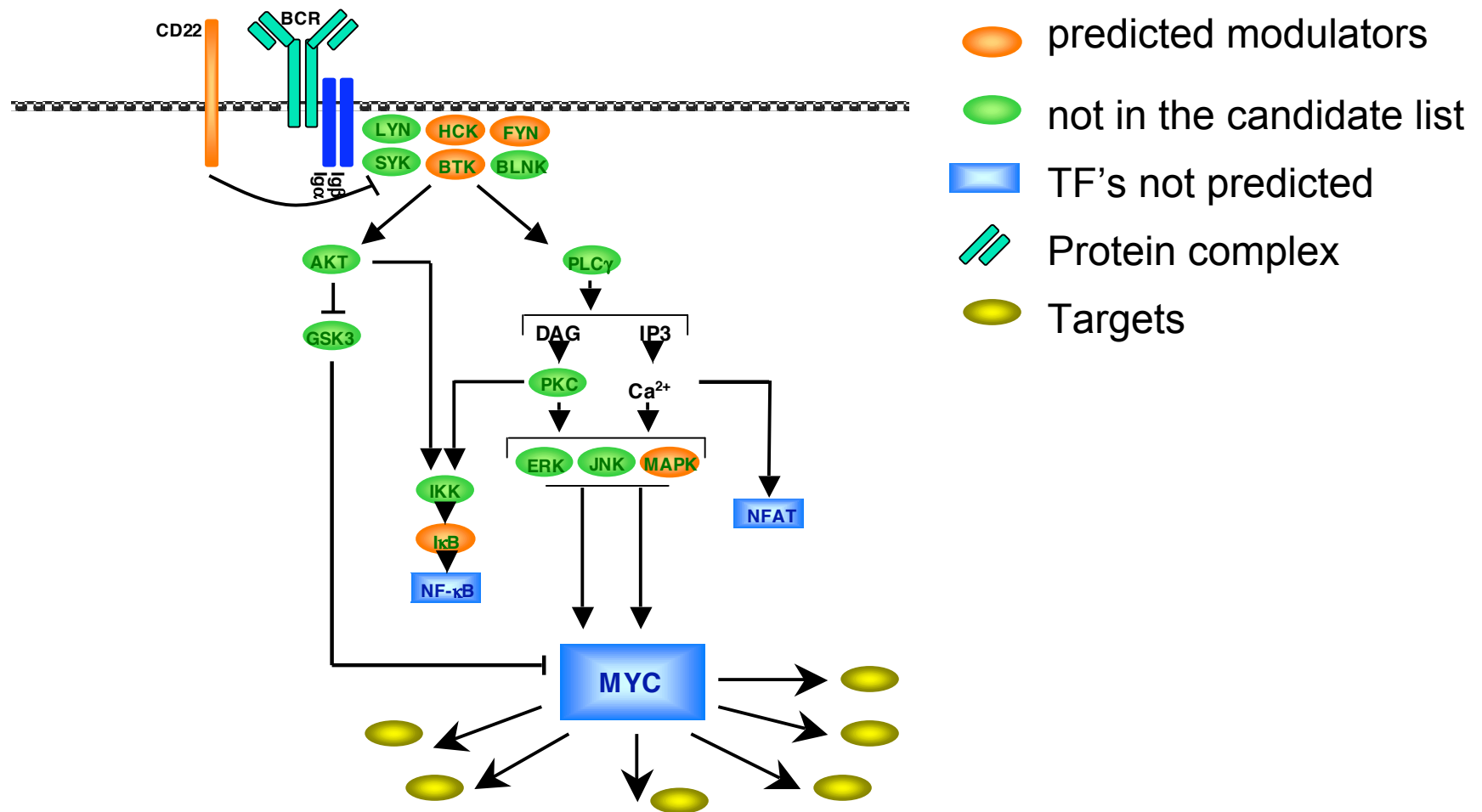


change in interactions



Over 70% cluster overlap

Reducibility: modulating pathways





Currently

- Biochemical validation
- Search for irreducible modulators



Summary

- IT quantities better measures of dependency
- **Problem:** estimation. **Solutions:** NSB (“don’t know” about entropies), stability of ranks
- **Application:** analysis of neural code at high resolution. **Found:** timing code, synergy, redundancy removal, photon counting -- optimality?
- **Problems:** what is an irreducible interaction? Algorithms with controlled approximations? **Solutions:** MaxEnt approximations, ARACNE, conditional ARACNE
- **Application:** B-cells microarrays analysis. **Found:** great performance on synthetic data, c-MYC targets (high precision validation), c-MYC modulators (to be validated, many confirmed by literature)



Thanks

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