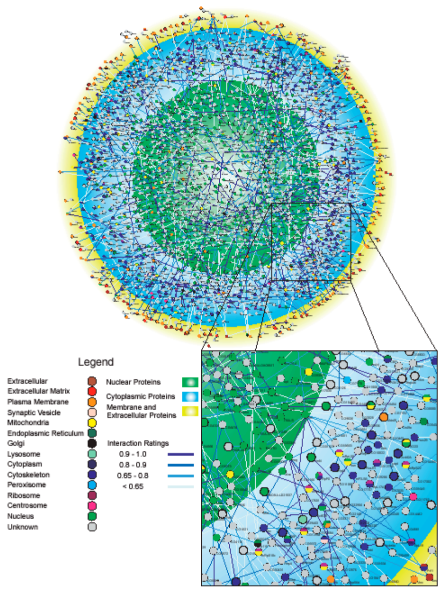
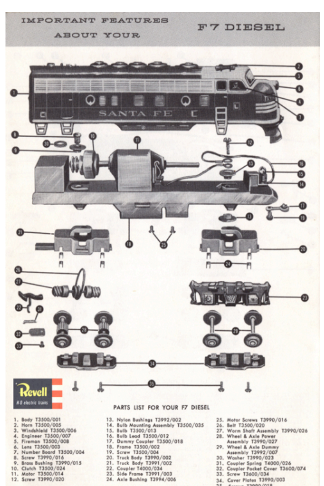
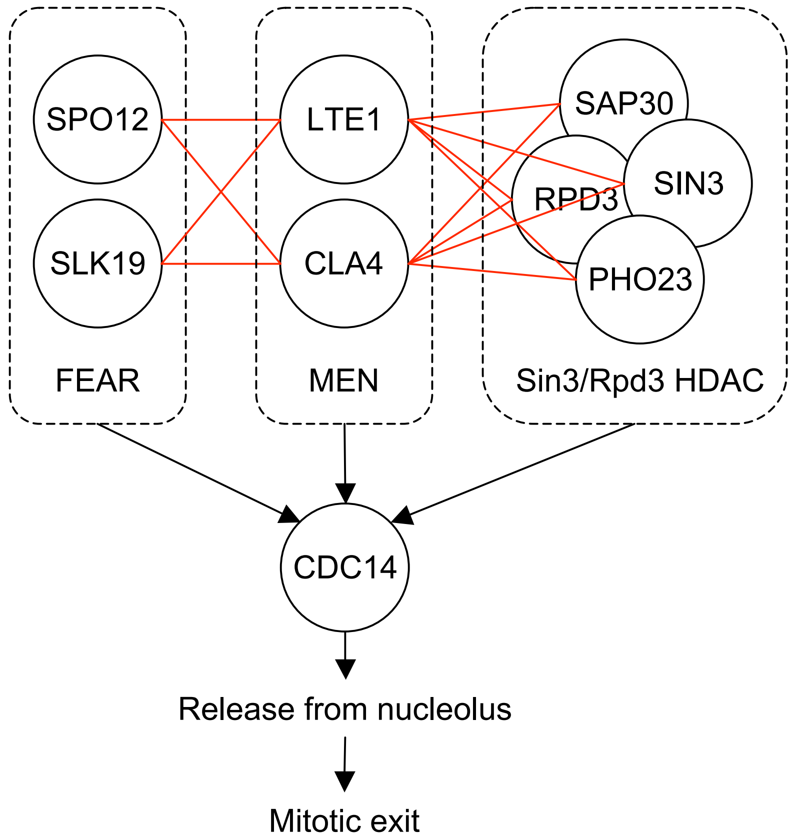
Finding meaning in biological networks

Physical interaction network





Genetic interaction network



Joel Bader
Department of Biomedical Engineering
High-Throughput Biology Center
Johns Hopkins University
joel.bader@jhu.edu

JHU BME / JHMI
HiT Center
Special K TCNP

joel.bader@jhu.edu

www.baderzone.org

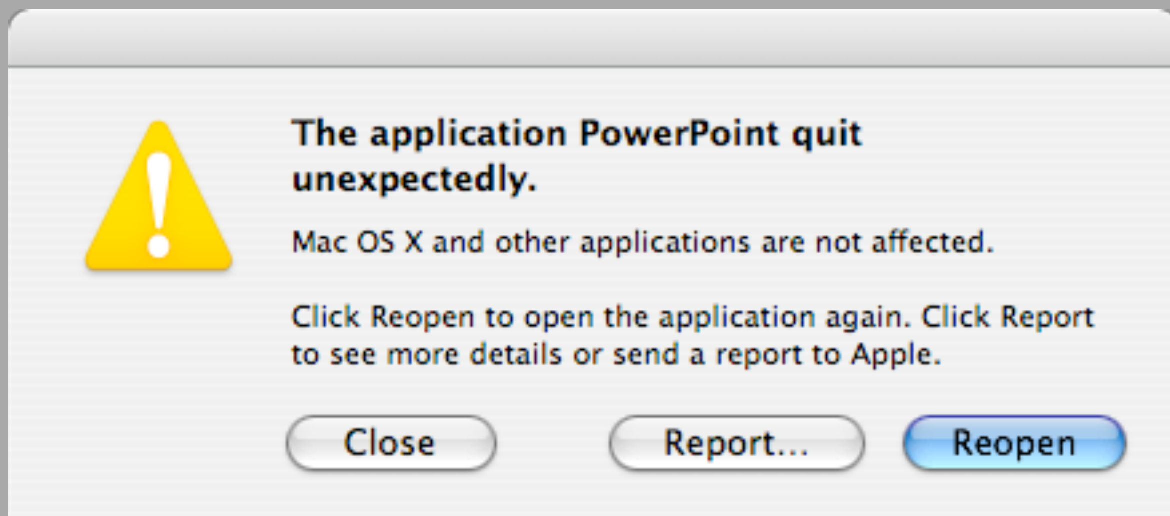
Santa Fe Institute

Statistical Inference for Complex Networks

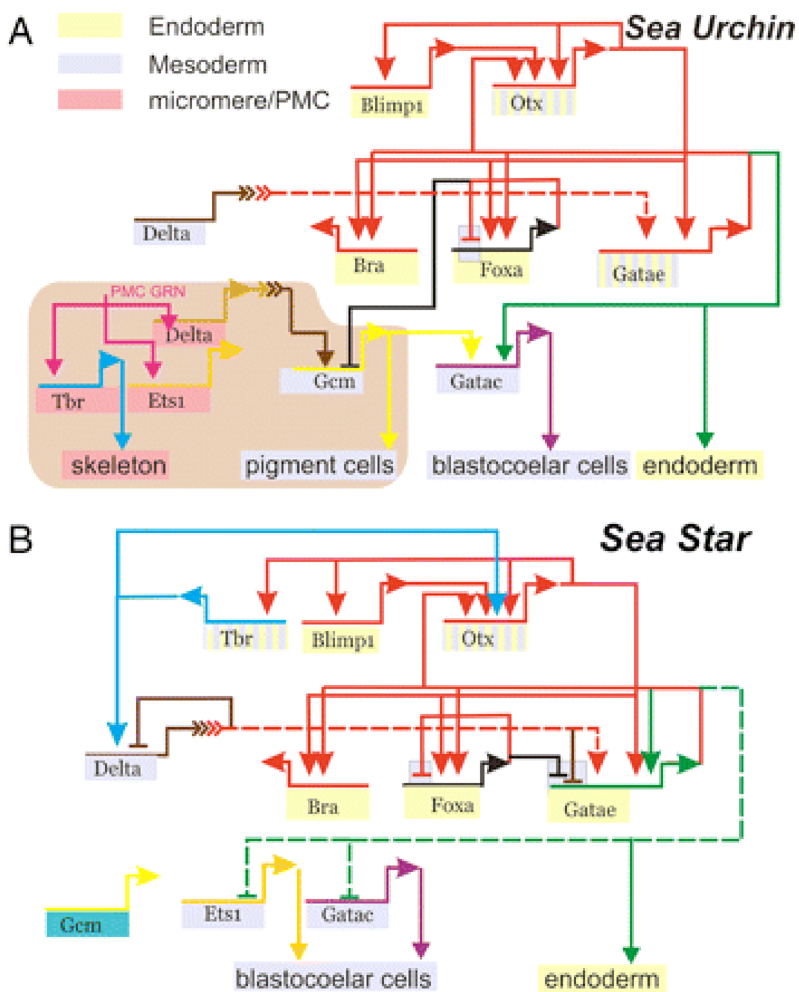
Dec 3-5, 2008

Main thanks

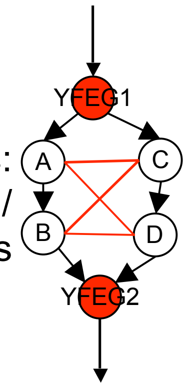
- Yan Qi (ABD): search on enemy networks
- Yongjin Park (CMU MS): extensions to CM, variational Bayes restaurant process
- Scott Patterson (stealth-mode biotech): degree-corrected stochastic block models


The application PowerPoint quit unexpectedly.
Mac OS X and other applications are not affected.
Click Reopen to open the application again. Click Report to see more details or send a report to Apple.
Close
Report...
Reopen

Gene regulatory vs. Genetic interaction



Physical reality: TF-DNA binding



Black edges:
Physical interactions /
metabolic pathways

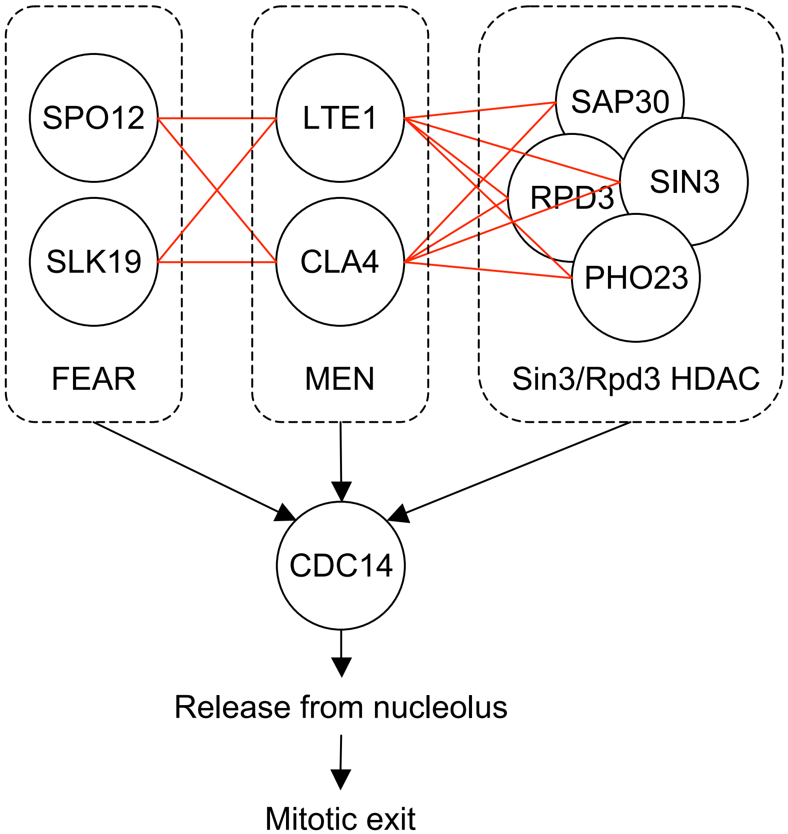
Red edges:
Genetic interactions /
synthetic lethals
orthogonal to pathway

Logical structure of network
Physical reality: phenotype of mutant
Relevant to network failure:
- Yeast viability
- Human disease
- Communication failure

Current use: logical structure of yeast biological pathways
- 6000 genes
- 1000 essential
- 5000 non-essential, test all 12.5 M pairs
- Higher-order combos: ask me

Leading groups at Toronto, UCSF, JHU

Dual prediction problems



1. Given red edges, predict genes/proteins in the same module.
   Share many red-edge neighbors
   Enriched for paths of length 2

2. Given incomplete / noisy set of tested red edges, predict untested red edges
   (Similar to predator/parasite-prey prediction)
   Enriched for paths of length 3
   No path of length 1

Haven't been using knowledge of tested/untested
Red edges are sparse (50/5000 = 1%)

Similar to spin-spin correlation in antiferromagnetic Ising lattice with network topology

+ correlation = same module
– correlation = probable red edge

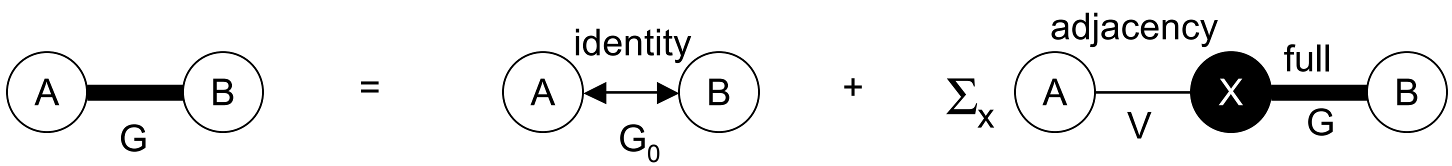
Linearized mean field theory = linear response theory = ...

Systematic path sums = Graph diffusion kernel

Recall Ornstein Zernicke / Poisson Boltzman: $h = c + c\,\rho\, h$, $c \sim -\beta u$

Recall path integrals: $U(t) = \exp(H\,t) = \exp[(H_0 + V)t] = [I + (H_0+V)t/n]^n$

kernel betw. A & B = (some constant if A = B) + (direct interactions betw. A & all of its neighbors) times (kernel betw. neighbor & B)

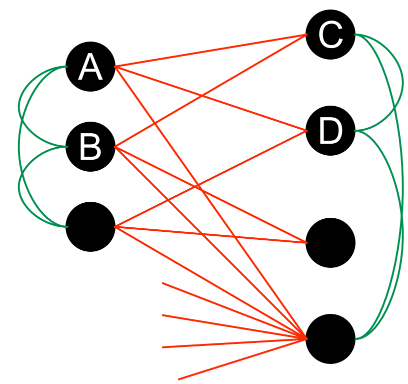


$$G = G_0 + G_0\, V\, G$$

$$= \boxed{G_0} + \boxed{G_0VG_0} + \boxed{G_0VG_0VG_0} + \boxed{G_0VG_0VG_0VG_0 +} \ldots$$

Even parity kernel: friends

Odd parity kernel: enemies



V = V(friend) + V(enemy)

Early integration: $G = [I - V(\text{friend}) - V(\text{enemy})]^{-1}$ Bad: scambles friends, enemies

Late integration: $G = [I - V(\text{friend})]^{-1} + [I - V(\text{enemy})^2]^{-1}$ Better, like naive Bayes

But what about predicting enemies?

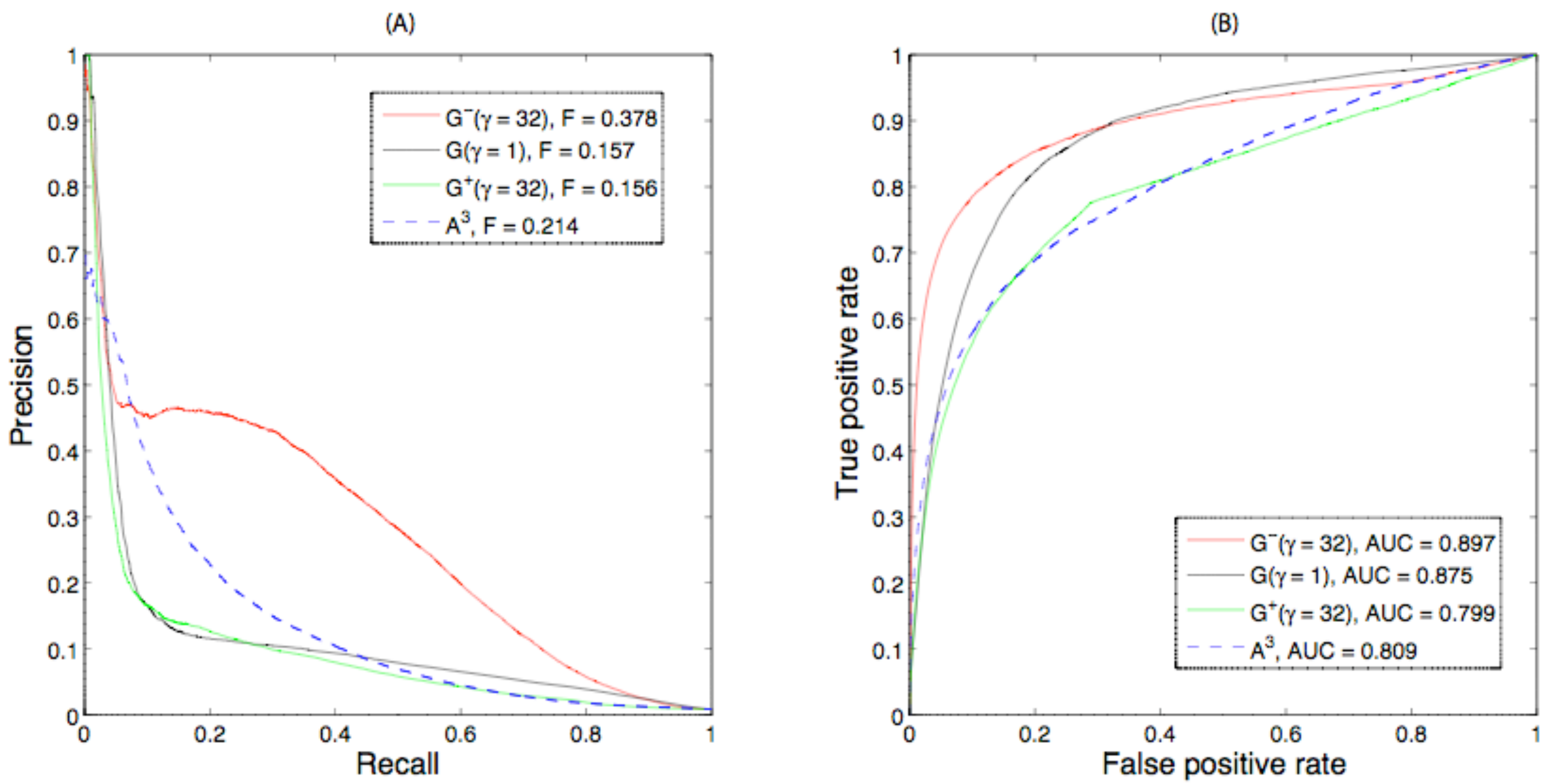
Renormalization:

$G_0 = [I - V(\text{friend})]^{-1}$

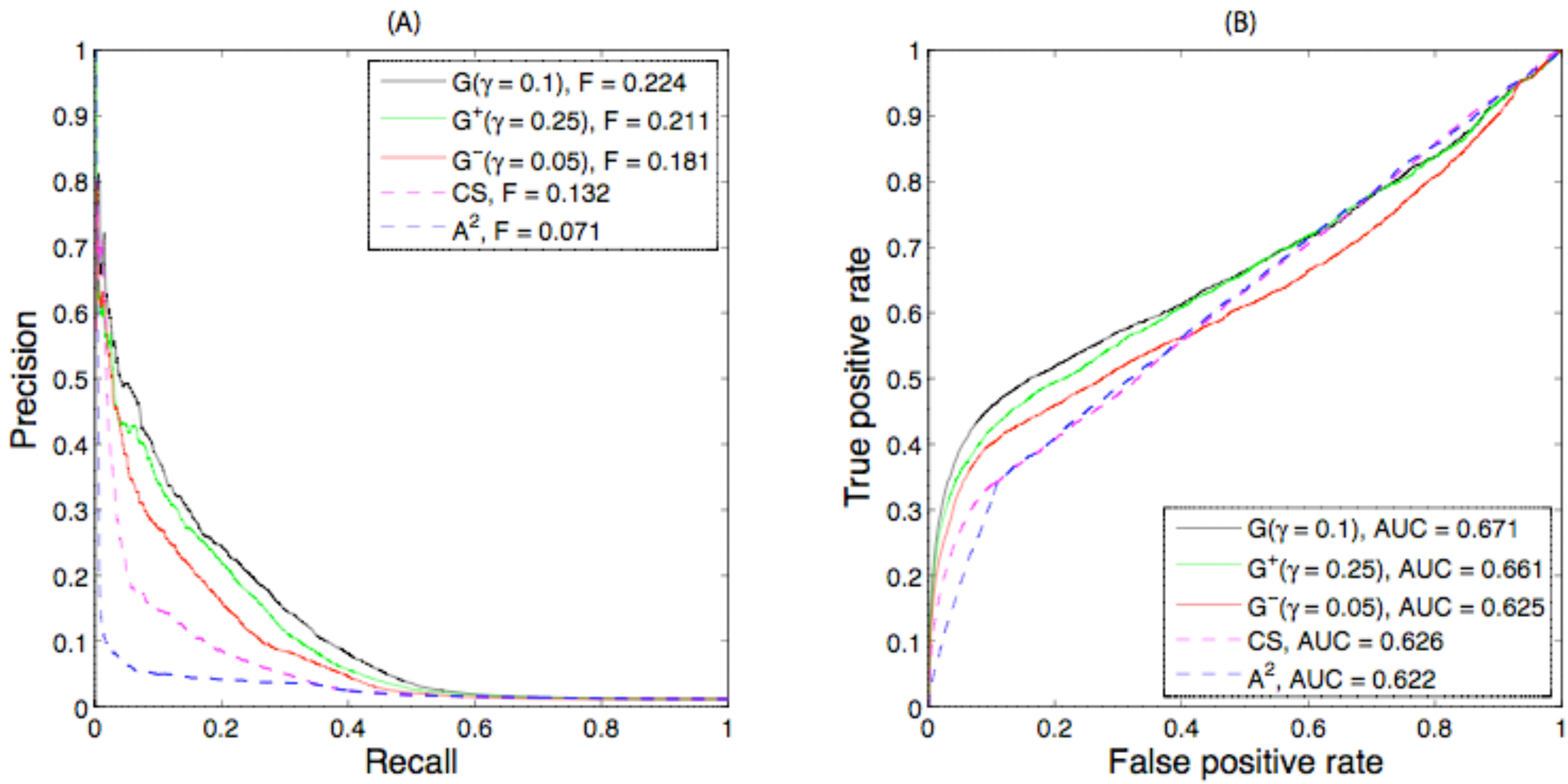
$G(\text{friend}) = [I - G_0\, V(\text{enemy})\, G_0\, V(\text{enemy})\,]^{-1} G_0$

$G(\text{enemy}) = [I - G_0\, V(\text{enemy})\, G_0\, V(\text{enemy})\,]^{-1} G_0\, V(\text{enemy})\, G_0$

Best method to date for predicting new edges

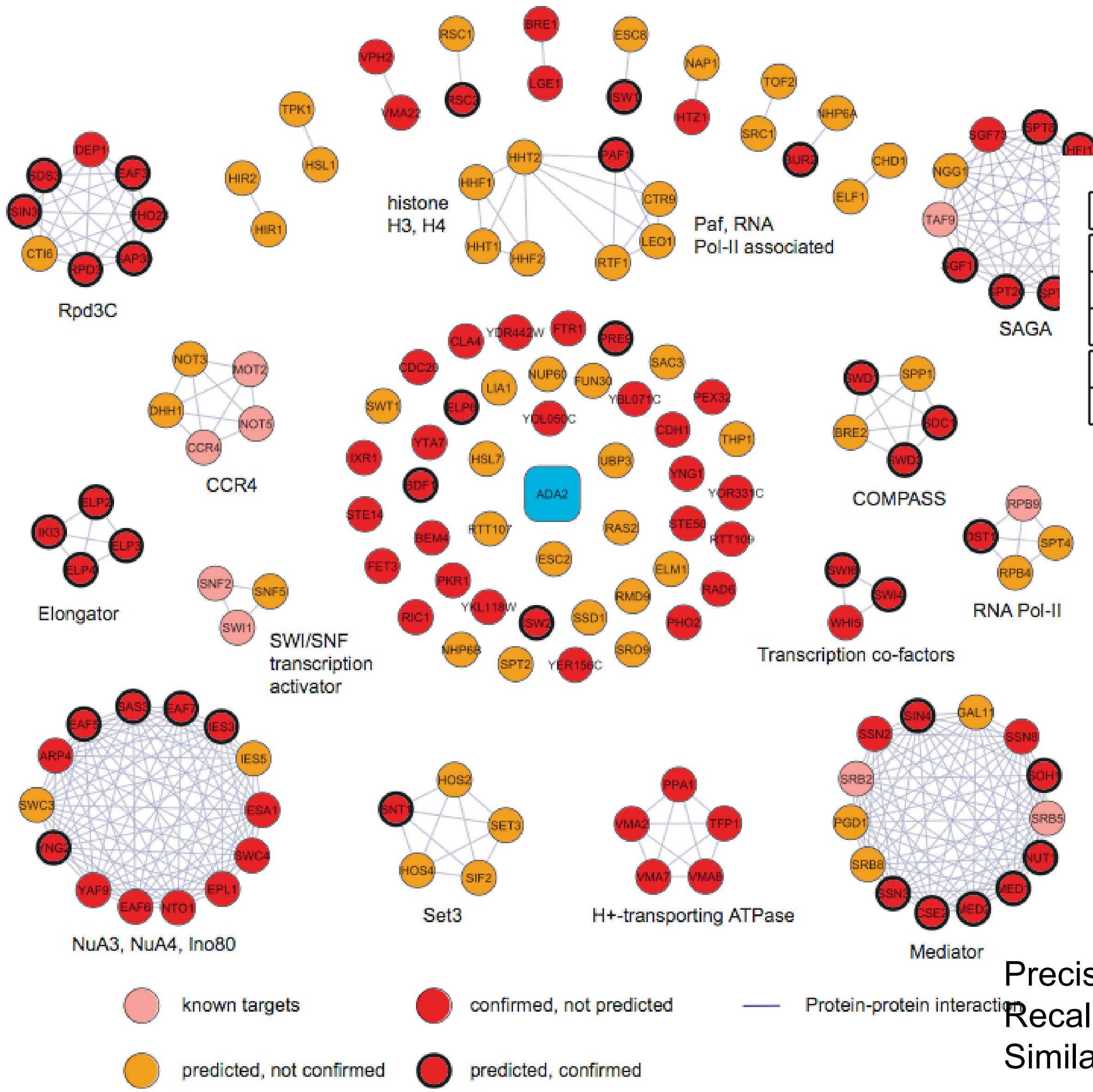


Also good for dual problem (communities)



Directed experimental search

Qi et al. 2008 Genome Research



ADA2	Pattern					Total
Top 100	Y	Y	Y	Y	N	100
HTS	Y	N	N	N	Y	75
Follow-Up	N	Y	N	NA	N	12
Count	29	12	38	21	46	
Category	TP1	TP2	FP	NA	FN	

Top 100 predictions
vs.
HTS 75 hits

29 TP

71 naive FP
(or FN from HTS)

Re-test 50 / 71

12 TP

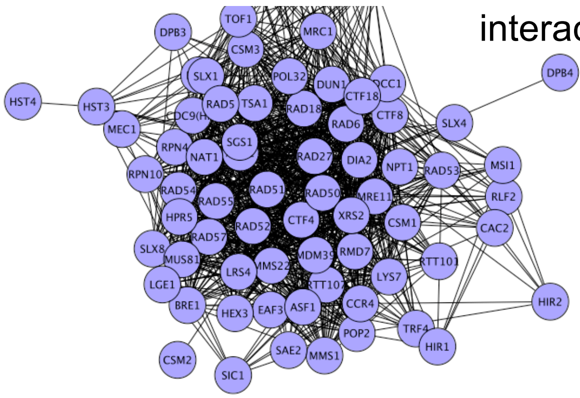
38 FP

Precision: at least 29+12/100 = 41%
Recall: at least 41/75+12 = 47%
Similar results for Esa1 (essential acetylase)

Adding to Clauset-Moore-Newman

... or imitation is the sincerest form of flattery

vertices = genes
edges = synthetic lethal genetic interactions (SLAM)



Expert curation

Our Karate club:
SL screen of ~100 genes involved in mainting DNA integrity (DNA damage sensing and repair)
(Pan et al. 2006 Cell)
Network analyzed by a biological expert, segmented.

Can a clustering algorithm reproduce the truth (or was the expert wrong?)

Multiple edge types:
SL (logical/emotional)
PPI (physical/location)

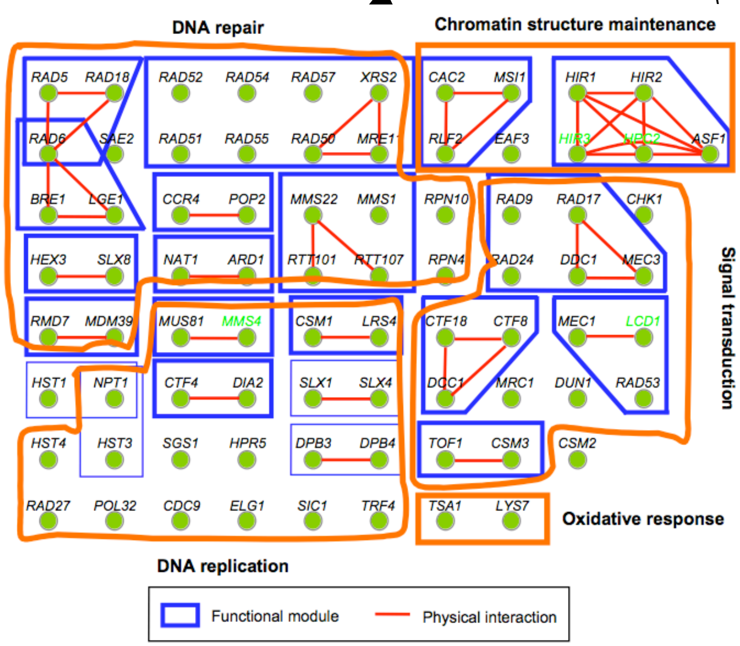


Table 1. Functionally Distinct Modules

Name of Module	Components of Module[a]	Congruency[b] of SL Profiles	SFL within Module[c]	Protein-Protei Interaction[d]
BRE1m	RAD6, BRE1, LGE1	22–109	No	Yes
CAF-I	CAC2, MSI1, RLF2	33–34	No	Yes
CCR4m	CCR4, POP2	133	No	Yes
CSM1m	CSM1, LRS4	65	No	Yes
CTF18m	CTF18, CTF8, DCC1	129–144	No	Yes
HEX3m	HEX3, SLX8	26	No	Yes
HIR	ASF1, HIR1, HIR2, **HIR3**, **HPC2**	13–40	No	Yes
HR	RAD50, MRE11, XRS2, RAD51, RAD52, RAD54, RAD55, RAD57	55–116	No	Yes (only among R Mre11p, and X
MEC1m	MEC1, **LCD1**, RAD53	10	No	Yes (only betw Mec1p and Lc
MMS22m	MMS22, MMS1, RTT101, RTT107	20–28	Yes[f] (only between RTT101 and RTT107)	Yes (only among M Rtt101p, and
MUS81m	**MMS4**, MUS81	6[e]	No	Yes
NAT1m	NAT1, ARD1	184	No	Yes
PRR	RAD6, RAD5, RAD18	19–50	No	Yes
RAD9m	RAD9, DDC1, RAD17, MEC3, RAD24	27–38	No	Yes (only among D Rad17p, and
RMD7m	RMD7, MDM39	187	No	Yes
TOF1m	TOF1, CSM3	78	No	Yes

Joel Bader
Department of Biomedical Engineering
High-Throughput Biology Center
Johns Hopkins University
joel.bader@jhu.edu

JHU BME / JHMI
HiT Center
Special K TCNP

joel.bader@jhu.edu

www.baderzone.org

Santa Fe Institute

Statistical Inference for Complex Networks

Dec 3-5, 2008

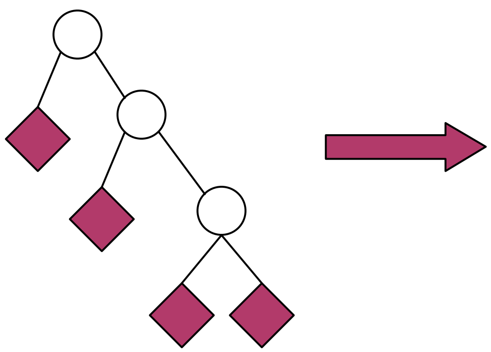
Modification #1: Fully Bayes

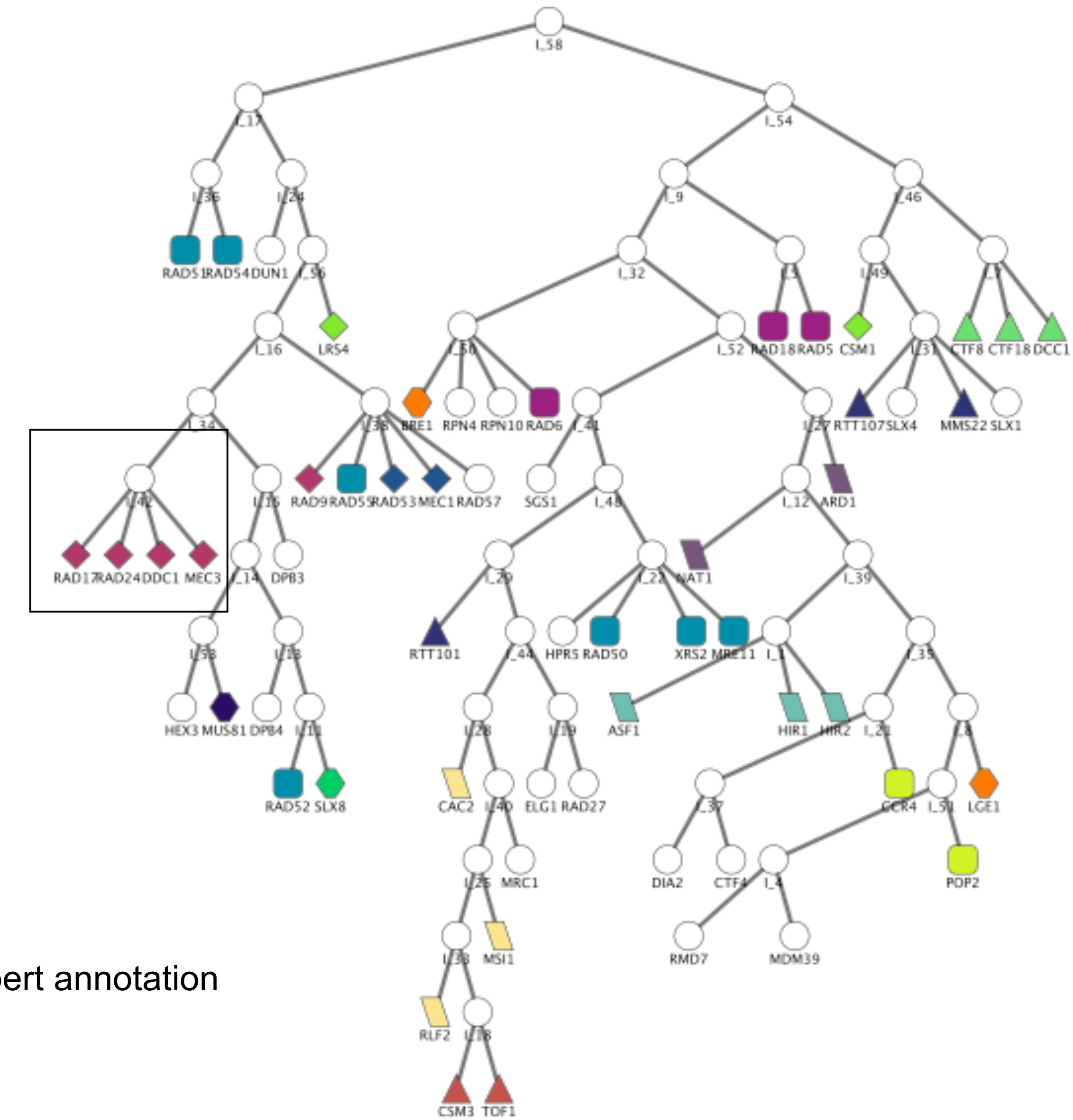
$$\Pr(D \mid \{\rho_r\}) = \prod_{r \in D} \rho_r^{E_r} (1-\rho_r)^{L_r R_r - E_r}$$

$$\Pr(\{\rho_r\} \mid a,b) = \prod_{r \in D} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \rho_r^{a-1} (1-\rho_r)^{b-1}$$

$$\Pr(D \mid a,b) = \left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right)^{|D|} \prod_{r \in D} \left[\frac{\Gamma(E_r + a)\Gamma(L_r R_r - E_r + b)}{\Gamma(L_r R_r + a + b)} \right]$$

Model Selection: Left/Right vs. Center





Terminal color = expert annotation

Modification #2: Left/Right vs. Center

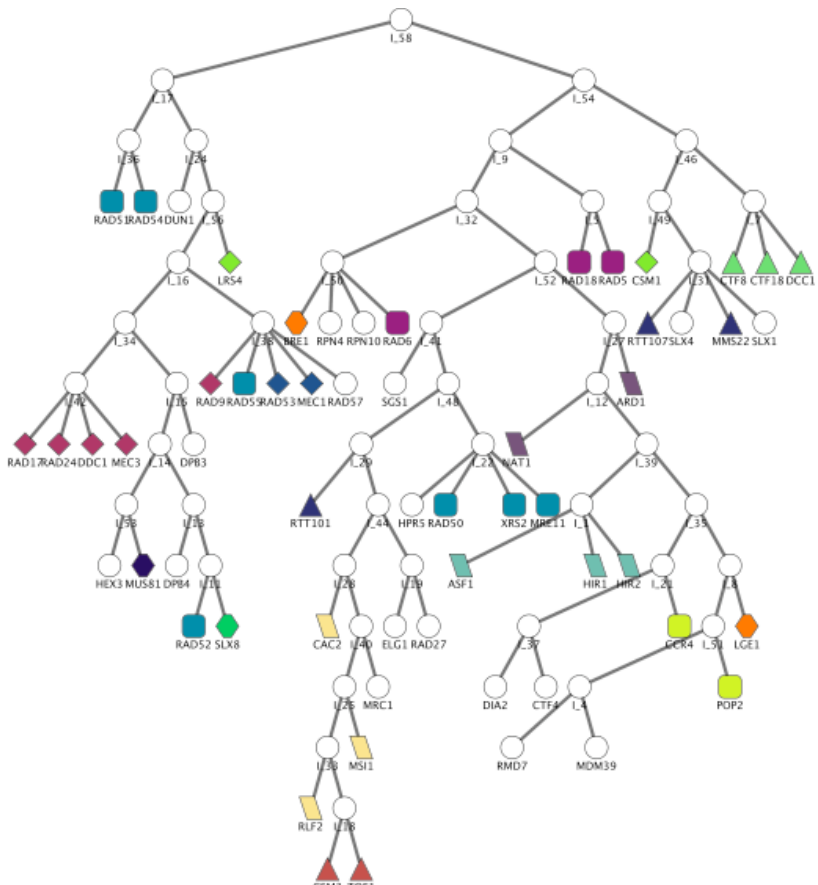
$$\Pr(D \mid \{\rho_r\}) = \prod_{r \in D} \rho_r^{E_r} (1-\rho_r)^{C(n,2)-E_r}$$

$$\Pr(\{\rho_r\} \mid a,b) = \prod_{r \in D} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \rho_r^{a-1} (1-\rho_r)^{b-1}$$

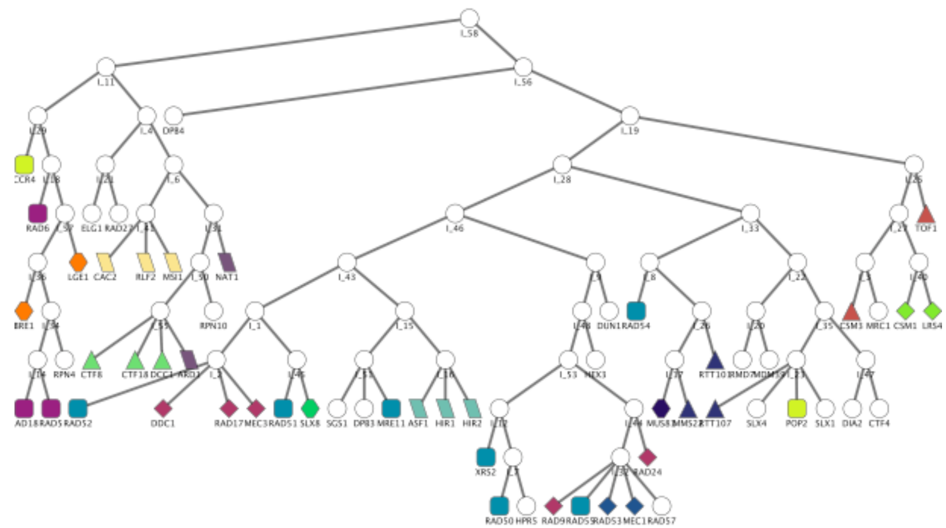
$$\Pr(D \mid a,b) = \left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right)^{|D|} \prod_{r \in D} \left[\frac{\Gamma(E_r + a)\Gamma(C(n,2) - E_r + b)}{\Gamma(C(n,2)+a+b)} \right]$$

Modification #3: Multiple edge types

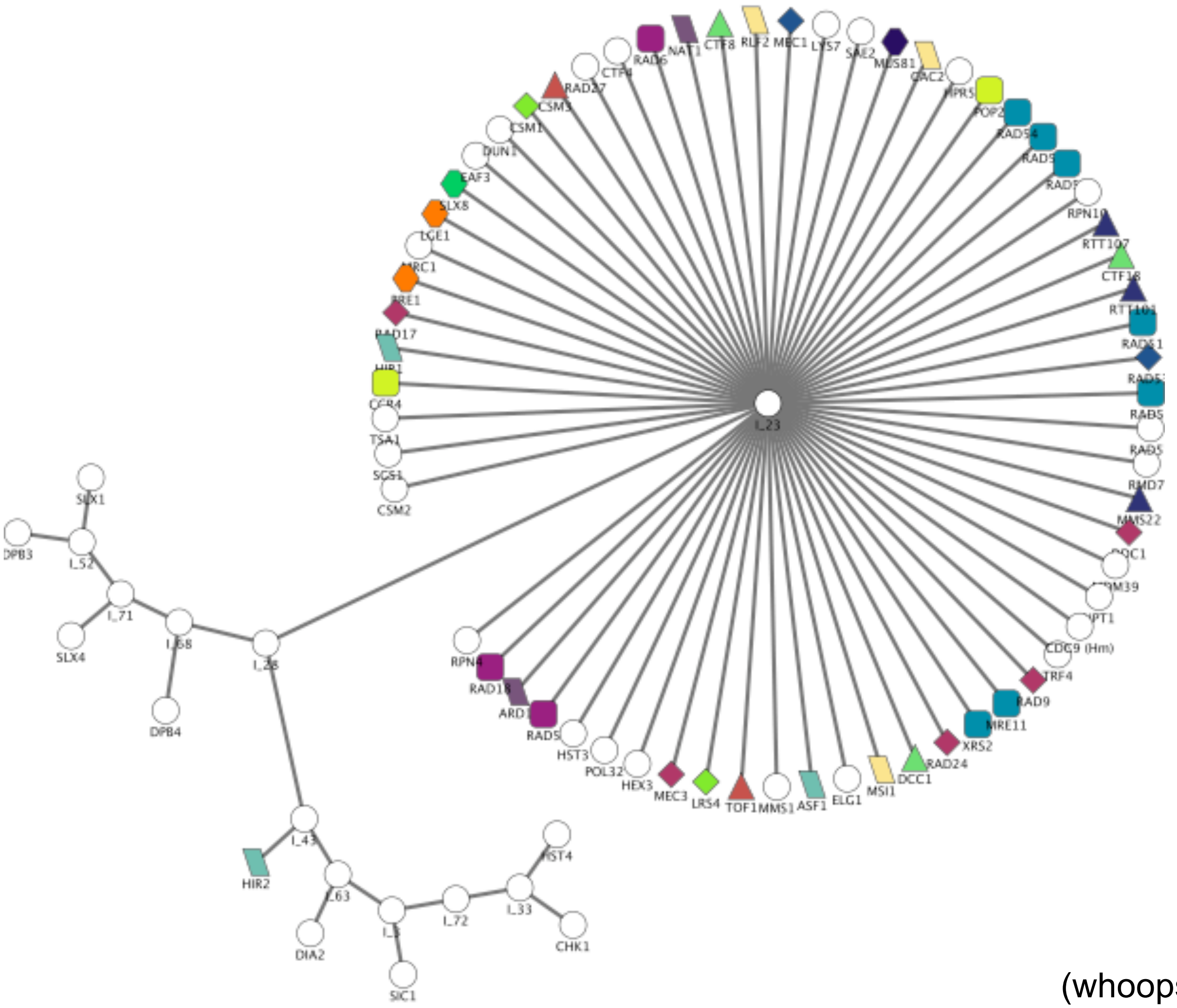
Just PPI/Physical/Location



Just SL/logical/social preference

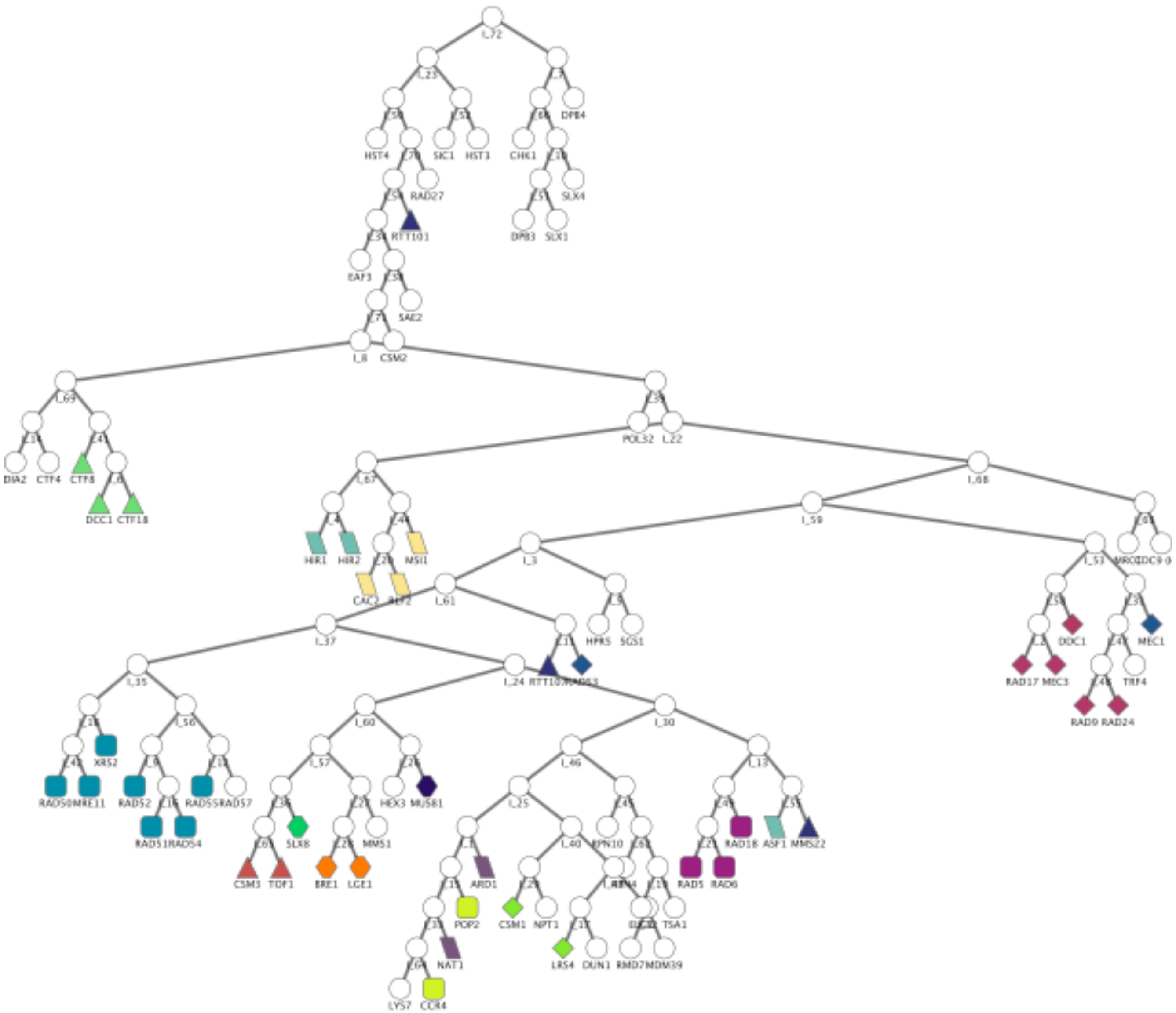


Both



(whoops: early integration)

Both

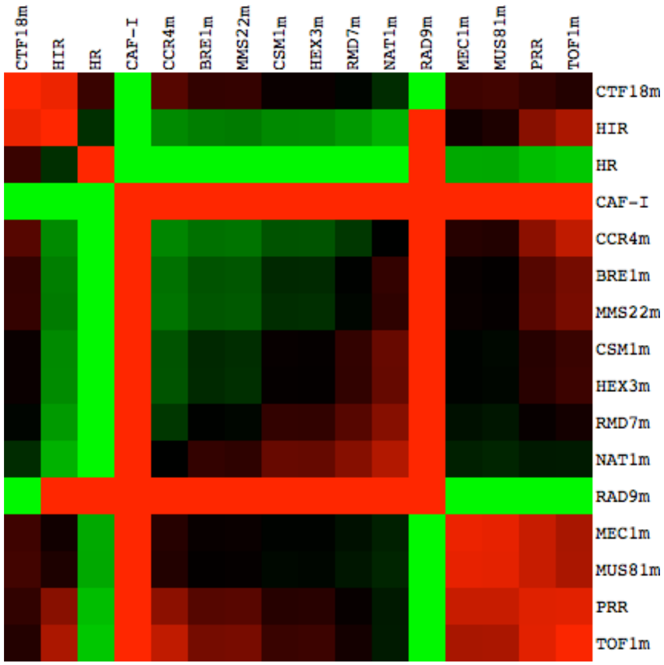


White terminal nodes: unannotated by expert

Problem: long run time

Fix with LR/C collapsing?
Have to implement detailed balance

Degree-corrected block models



Problems to address:

Long branch attraction
High-degree vertices grouped together

Multiple edge types

Model selection
How many clusters?
Form of prob. distribution?

Color = edge probability within/between blocks
Red = depleted
Green = enriched

Parameters
# of groups:
fixed K
sum over all K (restaurant process, Hofman/Wiggins, Jordan)
block-block probabilities:
2 parameters (within/between)
K(K+1)/2 parameters (each unique block pair)

Joel Bader
Department of Biomedical Engineering
High-Throughput Biology Center
Johns Hopkins University
joel.bader@jhu.edu

JHU BME / JHMI
HiT Center
Special K TCNP

joel.bader@jhu.edu

www.baderzone.org

Santa Fe Institute

Statistical Inference for Complex Networks

Dec 3-5, 2008

Trial #1: Newman's asymmetric block model

- Parameters are groups x vertices
- 1. For each v in V, sample v's membership according to prior distribution of mixture (e.g., DPM)
- 2. Given v's membership, sample adjacency profile (a column vector) from Multinomial distribution

$$z_i^k = \begin{cases} 1 & v_i \in k \\ 0 & otherwise \end{cases}$$

Asymmetric mixture model

$$\Pr(\overrightarrow{A_i} \mid \{\rho_{kj}\}, z_i^k) = \prod_{k'=1}^{K} \left[\prod_{j=1}^{n} \rho_{k'j}^{A_{ij}} \right]^{z_i^{k'}}$$

$$\Pr(\{\rho_{kj}\} \mid \lambda) = \mathrm{Dir}(\{\rho_{kj}\} \mid \lambda / n, \ldots, \lambda / n) = \frac{\Gamma(\lambda)}{\Gamma(\lambda / n)^n} \prod_{j=1}^{n} \rho_{kj}^{(\lambda/n)-1}$$

If "K" is fixed and with a uniform prior, we can derive the marginal as:

$$\Pr(A \mid \{z_i^k\}, \lambda) = \frac{\Gamma(\lambda)}{\Gamma(\lambda / n)^n} \prod_{k=1}^{K} \left[\frac{\prod_{j=1}^{n} \Gamma(\lambda / n + \sum_{i=1}^{n} A_{ij} \cdot z_i^k)}{\Gamma(\lambda + \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \cdot z_i^k)} \right]$$

CRP for the asymmetric mixture

CRP enables to sample latent membership
from Dirichlet Process Mixture by Gibbs sampling

More efficiently, we can simulate the mixture
using the collapsed Gibbs sampling (Neal 2000),
where parameters are integrated out beforehand.

CRP: prior distribution

If "k" is a new cluster,

$$\Pr(z_i^{k*} \mid z_{\neg i}) = \frac{\alpha}{n-1+\alpha}$$

where $z_{\neg i}$ denotes all the fixed membership except for the i-th one.

If "k" is one of clusters that already exist,

$$\Pr(z_i^{k} \mid z_{\neg i}) = \frac{|C_k|}{n-1+\alpha} = \frac{\sum_{j=1}^{n} z_j^k}{n-1+\alpha}$$

CRP: predictive distributions for a collapsed Gibbs sampling

If "k" is a new cluster,

$$\Pr(A_i \mid z_i^k) = \frac{\Gamma(N\lambda)}{\Gamma(\lambda)^N} \frac{\prod_{j=1}^{n} \Gamma(\lambda + A_{ij})}{\Gamma(\sum_{j=1}^{n} \lambda + A_{ij})}$$

If "k" is one of clusters that already exist,

$$\Pr(A_i \mid z_i^k, z_{\neg i}, A_{\neg i}) = \frac{\Gamma(N\lambda + \sum_{j=1}^{n} \sum_{l:l \neq i} A_{lj} z_l^k)}{\Gamma(N\lambda + \sum_{j=1}^{n} \sum_{l:l \neq i} A_{lj} z_l^k + \sum_{j=1}^{n} A_{ij})} \prod_{j=1}^{n} \left[\lambda + \sum_{l:l \neq i} A_{lj} z_l^k \right]^{A_{ij}}$$

CRP: collapsed Gibbs sampling

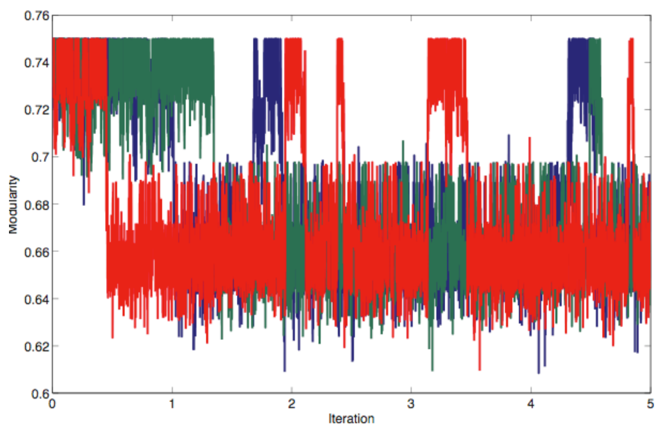
From previous distributions,
we can either assign to a new cluster with a probability of

$$\Pr(z_i^{k*} \mid A_i, \alpha) \propto \Pr(z_i^{k*} \mid \alpha) \Pr(A_i \mid z_i^{k*})$$

Or assign to one of existing cluster with a probability of

$$\Pr(z_i^{k} \mid A_i, z_{\neg i}, A_{\neg i}) \propto \Pr(z_i^{k} \mid z_{\neg i}, A_{\neg i}) \Pr(A_i \mid z_i^{k}, z_{\neg i}, A_{\neg i})$$

Non-ergodic?



DPM: Stick-breaking process

Another realization of DPM is the stick-breaking process (Sethuraman 1994), which provides a more explicit framework for a variational calculation (Blei and Jordan 2006).

DPM: Variational inference

$$\Pr(V \mid \alpha) = \prod_{k=1}^{\infty} \mathrm{Beta}(v_k \mid 1, \alpha)$$

v_k is the amount of mass in group k

$$\Pr(z_i \mid V) = \prod_{k=1}^{\infty} v_k^{z_i^k} (1 - v_k)^{\sum_{l>k} z_i^l}$$

$$\Pr(\overrightarrow{A_i} \mid \{\rho_k\}, z_i^k) = \prod_{k=1}^{\infty} \left[\prod_{j=1}^{n} \rho_k^{A_{ij}} \right]^{z_i^k}$$

$$\Pr(\rho_k \mid \lambda) = \mathrm{Dir}(\rho_k \mid (\lambda / n, ..., \lambda / n))$$

$$q(V, Z, \rho) = \prod_{k=1}^{K_{Tr}-1} \mathrm{Beta}(v_k \mid \gamma_{k1}, \gamma_{k2}) \prod_{k=1}^{K_{Tr}} \mathrm{Dir}(\rho_k \mid \tau_k) \prod_{i=1}^{n} \mathrm{Mult}(z_i \mid 1, \phi_i)$$

DPM: Variational updates

1.

$$\langle \ln(v_i)\rangle = \psi(\gamma_{k1}) - \psi(\gamma_{k1}+\gamma_{k2}) \qquad \gamma_{k1} = 1 + \sum_{i=1}^{n} \langle z_i^k \rangle$$

$$\langle \ln(1-v_i)\rangle = \psi(\gamma_{k2}) - \psi(\gamma_{k1}+\gamma_{k2}) \qquad \gamma_{k2} = \alpha + \sum_{i=1}^{n} \sum_{l>k}^{K_{Tr}} \langle z_i^l \rangle$$

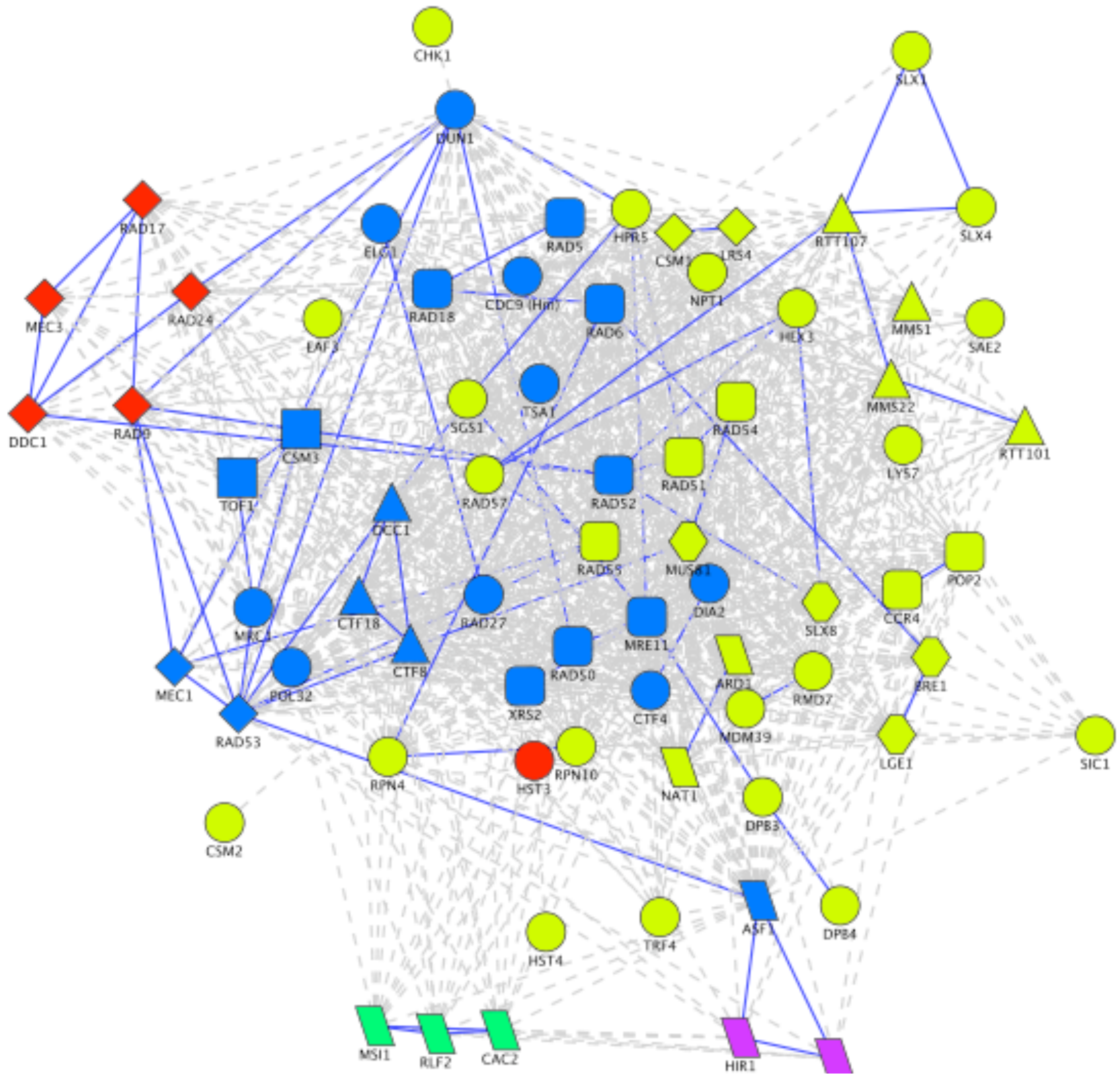
2.

$$\langle \ln(\rho_{kj})\rangle = \psi(\tau_{kj}) - \psi\left(\sum\nolimits_{j=1}^{n} \tau_{kj}\right) \qquad \tau_{kj} = \frac{\lambda}{n} + \sum_{i=1}^{n} \langle z_i^k \rangle A_{ij}$$

3.

$$\langle z_i^k \rangle \propto \begin{matrix} \exp\{\langle \ln(v_k)\rangle + \sum\nolimits_{k'=1}^{k-1} \langle \ln(1-v_{k'})\rangle + \\ \sum\nolimits_{j=1}^{n} A_{ij} \langle \ln(\rho_{kj})\rangle + \sum\nolimits_{j=1}^{n} (\lambda/n - 1)\langle \ln(\rho_{kj})\rangle\} \end{matrix}$$

Too much lumping



Trial #2: Boltzmann Machine

Construction of the probability distribution by symmetric (vertex-vertex) metric:

$$\Pr(\{w_{ij}\} \mid \{z_i^k\}) \propto \prod_k^K \exp\left\{ \sum_{i<j} w_{ij} z_i^k z_j^k \right\}$$

Extension using DPM prior

$$\Pr(z_i \mid V) = \prod_{k=1}^{\infty} v_k^{z_i^k} (1 - v_k)^{\sum_{l>k} z_i^l}$$

$$\Pr(V \mid \alpha) = \prod_{i=1}^{\infty} \mathrm{Beta}(v_i \mid 1, \alpha)$$

DPM-BM: Newman-Girvan Modularity

Negative sign for disassortative (SL) edges

$$\sum_{i,j} w_{ij} z_i^k z_j^k = \sum_{ij} (A_{ij} - E_{H_0}[A_{ij}]) z_i^k z_j^k$$

where

$$E_{H_0}[A_{ij}] = \frac{\deg(i)}{|E|} \times \frac{\deg(j)}{|E|} \cdot |E|$$

Newman and Girvan, Phys.Rev.E (2004)
Clauset *et al.* Phys.Rev.E (2004)

DPM-BM: variational inference

$$z \approx \underset{\{z\}}{\arg\min} \frac{1}{T} \langle -\ln \Pr(\{w\} \mid \{z\}) \rangle + \langle \ln \Pr(\{z\}, \{v\} \mid \alpha) \rangle$$
$$+ \langle \ln q(\{z\} \mid \{\mu\}) \rangle + \langle \ln q(\{v\} \mid \{\gamma\}) \rangle$$

using the following mean-field distribution

$$q(V, Z) = \prod_{k=1}^{K_{Tr}-1} \mathrm{Beta}(v_k \mid \gamma_{k1}, \gamma_{k2}) \prod_{i=1}^{n} \prod_{k=1}^{K_{Tr}} \mu_{ik}^{z_i^k}$$

DPM-BM: variational updates

1.

$$\langle \ln(v_i) \rangle = \psi(\gamma_{k1}) - \psi(\gamma_{k1} + \gamma_{k2}) \qquad \gamma_{k1} = 1 + \sum_{i=1}^{n} \langle z_i^k \rangle$$

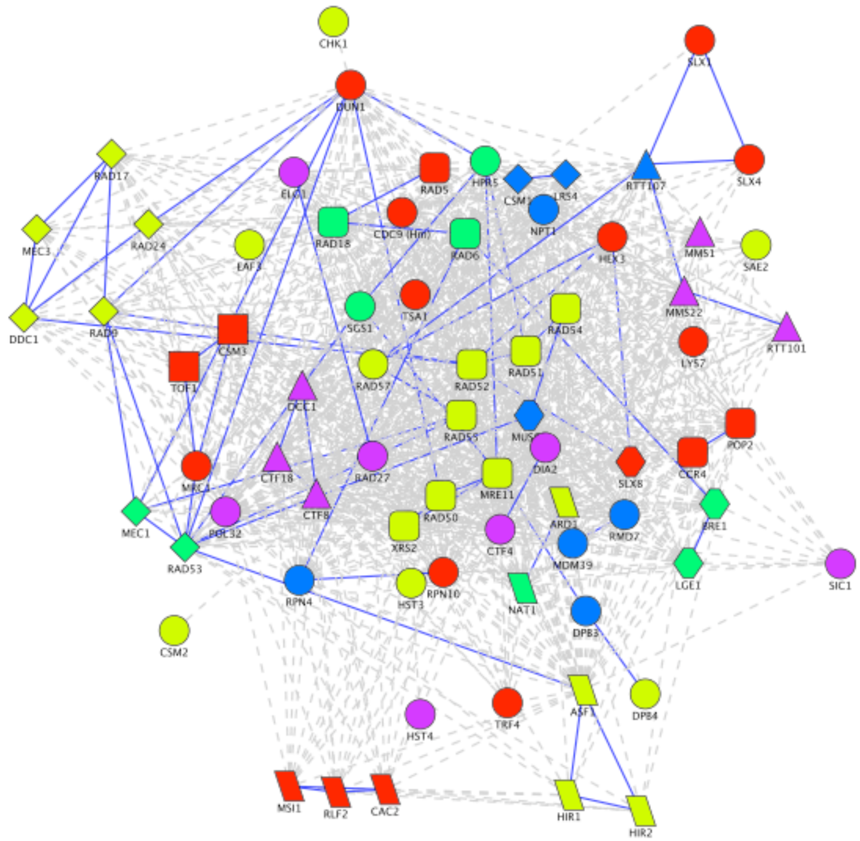
$$\langle \ln(1 - v_i) \rangle = \psi(\gamma_{k2}) - \psi(\gamma_{k1} + \gamma_{k2}) \qquad \gamma_{k2} = \alpha + \sum_{i=1}^{n} \sum_{l>k}^{K_{Tr}} \langle z_i^l \rangle$$

2.

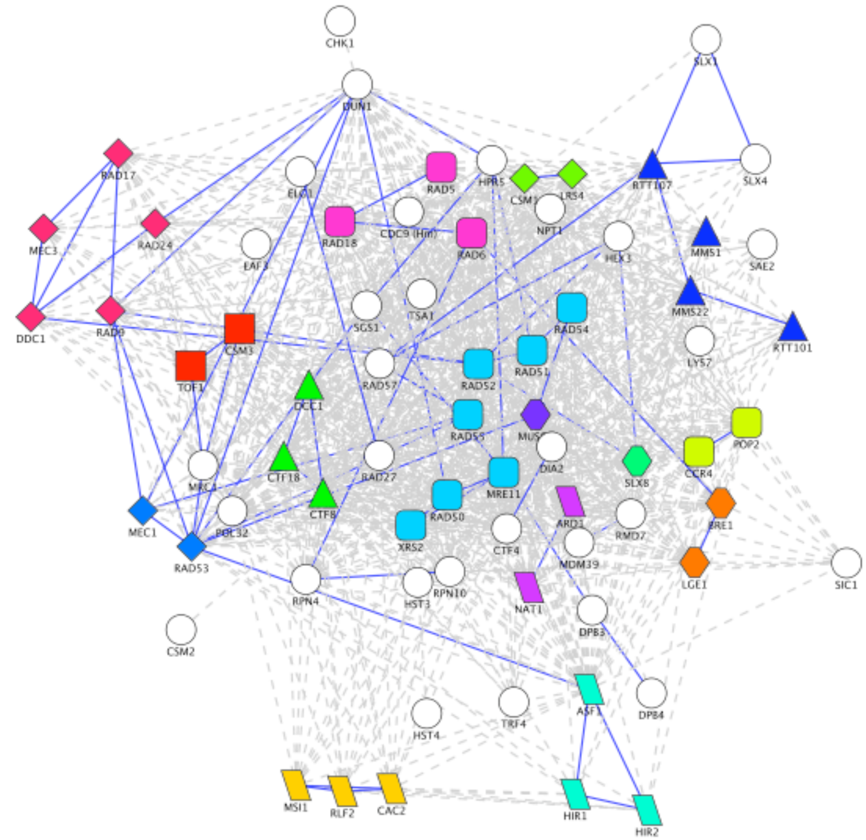
$$\langle z_i^k \rangle = \mu_i^k \propto \exp\left\{ \langle \ln(v_k) \rangle + \sum_{k'=1}^{k-1} \langle \ln(1 - v_{k'}) \rangle + \frac{1}{T} \sum_{j:\, j \neq i}^{n} w_{ij} \mu_j^k \right\}$$

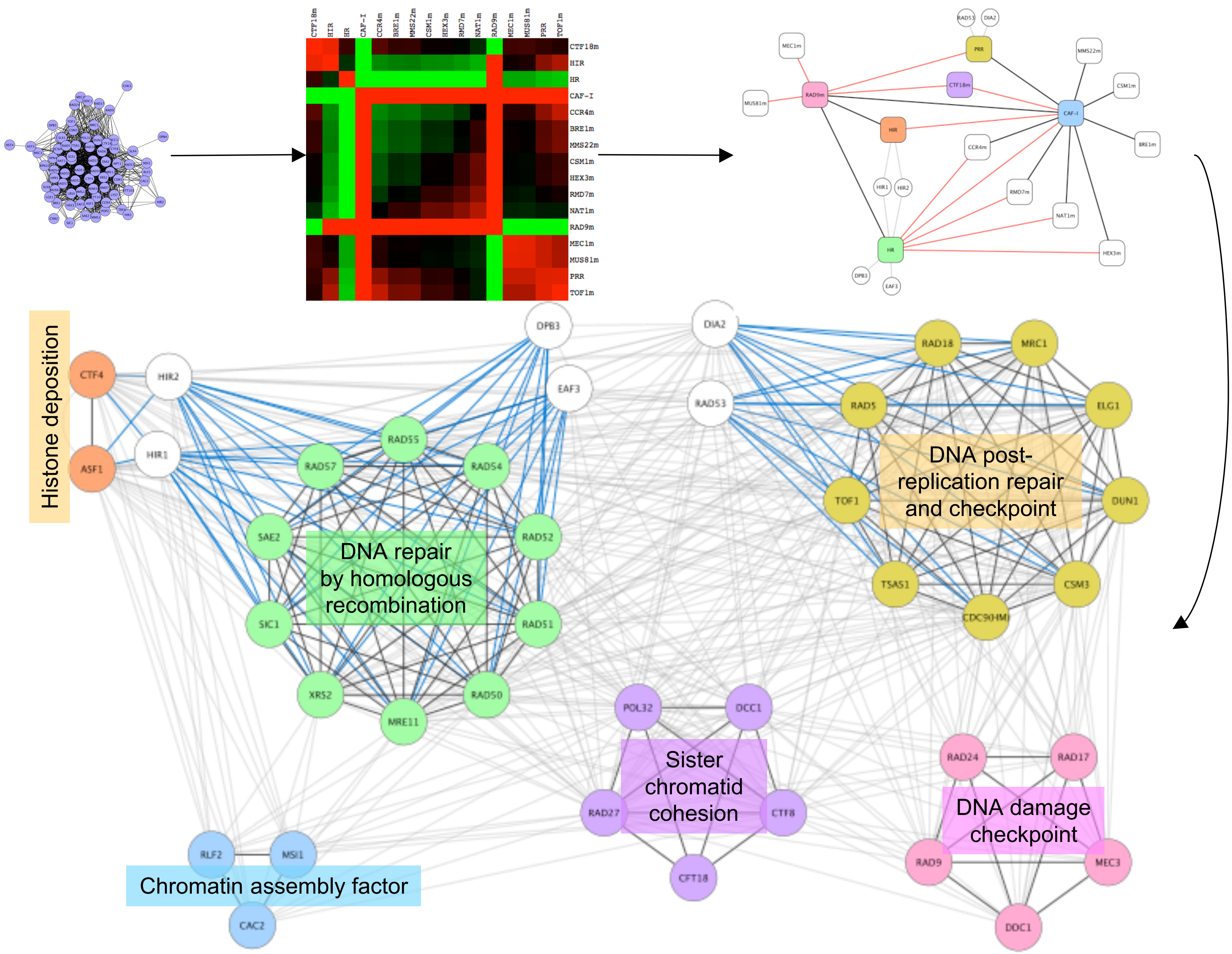
Boltzmann machine results

Boltzmann machine



Expert




CTF18m
HIR
HR
CAF-I
CCR4m
BRE1m
MMS22m
CSM1m
HEX3m
RMD7m
NAT1m
RAD9m
MEC1m
MUS81m
PRR
TOF1m
Histone deposition
CTF4
ASF1
HIR2
HIR1
DPB3
EAF3
DIA2
RAD53
RAD18
MRC1
RAD5
ELG1
DNA post-replication repair and checkpoint
TOF1
DUN1
TSA51
CSM3
CDC9OHM
RAD55
RAD57
RAD54
SAE2
RAD52
DNA repair by homologous recombination
SIC1
RAD51
XRS2
MRE11
RAD50
POL32
DCC1
Sister chromatid cohesion
RAD27
CTF8
CFT18
RAD24
RAD17
DNA damage checkpoint
RAD9
MEC3
DOC1
RLF2
MSI1
Chromatin assembly factor
CAC2

Missing nodes and edges

The Unknown

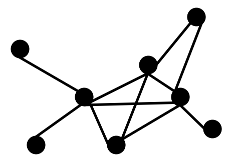
As we know,
There are known knowns.
There are things we know we know.
We also know
There are known unknowns.
That is to say
We know there are some things
We do not know.
But there are also unknown unknowns,
The ones we don't know
We don't know.

Donald Rumsfeld, Feb. 12, 2002, Department of Defense news briefing

Capture-Recapture for Networks Edges

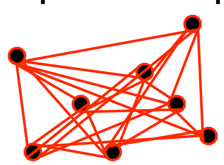
Sample edges with replacement

True Network (relevant connections)



Draw from true network
(true positive)
Prob = $1 - \alpha$

Complete Graph (red herrings)



Draw spurious edge
(false-positive)
Prob = α

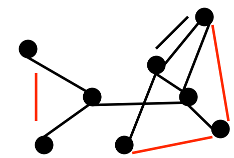
(Could improve,
have to start
somewhere)

Can we estimate k and f from {n, w, s}?

Chicken-and-egg problem:
If we knew parameters, we could estimate hiddens.
If we knew hiddens, we could fit parameters.
Solution: Expectation Maximization

The Problem:
Given a sampled network of strong and weak edges for a graph with an unknown strong-edge degree distribution

(1) Estimate the number of strong edges we've missed
(2) Estimate the probability that an edge observed once is strong



<u>Observed variables</u>
n = total number of draws (12)
w = number unique (11)
s = number of singletons (10)

<u>Hidden variables</u>
k = number of true edges (11)
f = number of FP's (3)

<u>Parameters</u>
α = FP rate (expect f ~ αn)
Pr(k) (or uniform)

Expectation: Bayes Rule

Hidden variables:
interaction partners
false positives in the catch

Observed variables:
singletons, # unique partners, catch size

Parameters:
Error model
Degree distribution

$$
\begin{aligned}
\Pr(k_j, f_j \quad | \quad & s_j, w_j, n_j, \alpha_j, \Phi) = \Pr(k_j|\Phi) \\
\times \quad & [\alpha_j^{f_j}(1-\alpha_j)^{n_j-f_j}/f_j!(s_j-f_j)!] \\
\times \quad & \boxed{[k_j!/(k_j-w_j+f_j)!k_j^{n_j-f_j}]} \\
\times \quad & \delta(0 \leq f_j \leq s_j)\delta(k_j \geq w_j - f_j) \\
/ \quad & \sum_{f=0}^{s_j} \sum_{k=w_j-f_j}^{\infty} \{\Pr(k|\Phi) \\
\times \quad & [\alpha_j^{f}(1-\alpha_j)^{n_j-f}/f!(s_j-f)!] \\
\times \quad & [k!/(k-w_j+f_j)!k^{n_j-f}]\},
\end{aligned}
$$

Ewens sampling formula
for equal frequencies

Huang, Jedynak, Bader, PLoS Comp Bio 2007
Improved method: Beta distribution for strong/weak edge mixing parameter (in review)

	Yeast	Worm	Fly
Screen properties			
Total # proteins	6,697	20,069	14,086
Total # baits	1,532	729	3,639
Total # preys	2,520	2,116	5,479
Total # used as bait and as prey	772	212	2,109
Fraction screened per bait	0.376	0.105	0.389
Fraction screened overall	0.086	0.004	0.100
False-pos. rates			
Per prey ($\bar{\alpha}$)	0.093	0.122	0.157
Per unique interaction	0.24	0.44	0.41
Per singleton interaction	0.36	0.66	0.65
True-pos. rates			
Systematic (p_{syst})	0.31(2)	0.45(4)	0.15(1)
Sampling (p_{samp})	0.47	0.53	0.67
Total	0.15	0.24	0.10
Mean # partners			
Unique preys per bait, full	3.0	5.6	5.7
Unique preys per bait, core	1.8	4.3	1.8
Corrected for false positives	2.3	3.1	3.4
… and sampling loss	4.8	5.9	5.0
… and systematic loss	15.4	13.1	33.9
… and fraction screened	40.8	124.4	87.0
Median # partners			
Corrected for FP's and sampling loss	1.0	2.9	2.7
… and systematic loss	3.3	6.4	18
… and fraction screened	8.8	61	46
Total # protein interactions			
… based on mean	137,000	1,250,000	613,000
… based on median	30,000	610,000	325,000

Only 5 to 10% of pairs tested ...

Of these,
only 10 to 20% of TPs captured

Huang, Jedynak, Bader
2007 PLoS Comp Bio

JHU CS Faculty Search

Faculty Applications

The Department of Computer Science at Johns Hopkins University is seeking applications for a tenure-track faculty position. Our primary interest is hiring at the Assistant Professor level, but candidates of all ranks will be considered. All areas will be considered, but candidates with research agendas in **security**, applied algorithms, computer systems, or **bioinformatics** will receive special attention. All applicants must have a Ph.D. in computer science or a related field and are expected to show evidence of ability to establish a strong, independent, multidisciplinary, internationally recognized research program.

Commitment to quality teaching at the undergraduate and graduate levels will be required of all candidates considered. The department webpage at **http://www.cs.jhu.edu** provides information about the department, including links to research laboratories and centers.

For full consideration, applicants should apply online before January 5 2009. Questions should be directed to fsearch@cs.jhu.edu. The Department is committed to building a diverse educational environment: women and minorities are strongly encouraged to apply. The Johns Hopkins University is an EEO/AA employer.

Faculty Search
Department of Computer Science
Room 224 New, Engineering Building
Johns Hopkins University
Baltimore, MD 21218-2694
Fax: 410-516-6134
Phone: 410-516-8775
fsearch@cs.jhu.edu