ARACNE: Reverse-Engineering Transcriptional Networks

Ilya Nemenman

Joint Centers for Systems Biology

Columbia

Contributors

Implementation: Adam Margolin, Kai Wang

Wet validation: Riccardo Dalla Favera, Katia Basso, Ulf Klein

Bioinformatic validation: Nila Banerjee, Andrea Califano

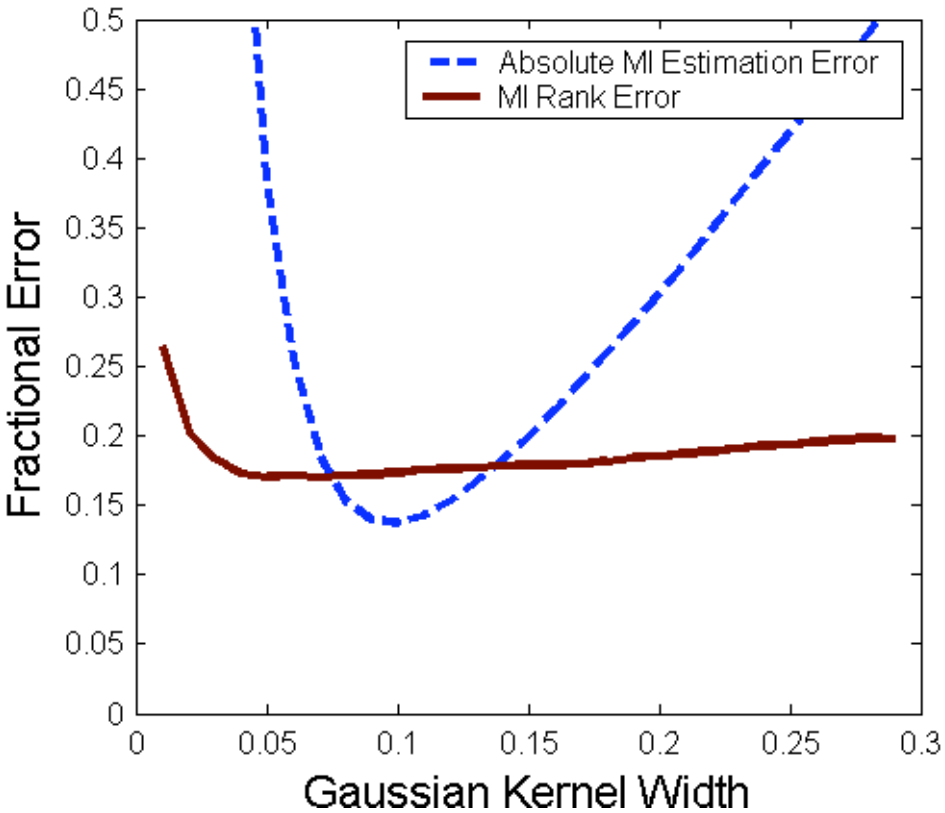
General discussions: Chris Wiggins, Gustavo Stolovitzky

Computations: Amdec group

Big boss: Andrea Califano

Estimating MI is not a problem

~100 samples



$$f(x,y) = \frac{1}{2\pi h^2 M} \sum_i \exp\left\{ -\frac{(x-x_i)^2 + (y-y_i)^2}{2h^2} \right\}$$

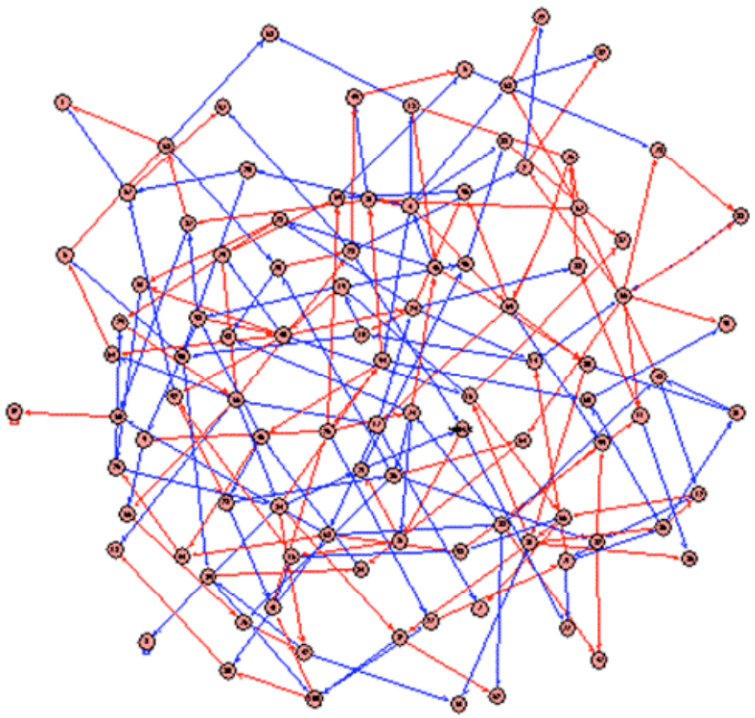
$$f(x) = \frac{1}{\sqrt{2\pi} hM} \sum_i \exp\left\{ -\frac{(x-x_i)^2}{2h^2} \right\}$$

$$h(M) \to 0 \qquad [h(M)]^2 M \to \infty$$

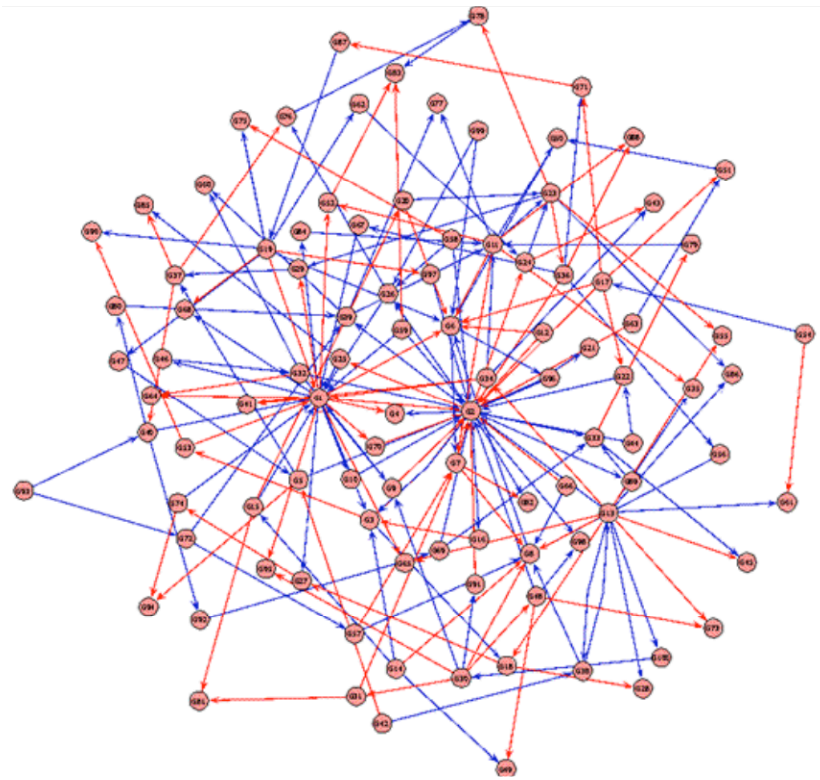
Use universal best *h*

Synthetic networks

Random



Scale-free



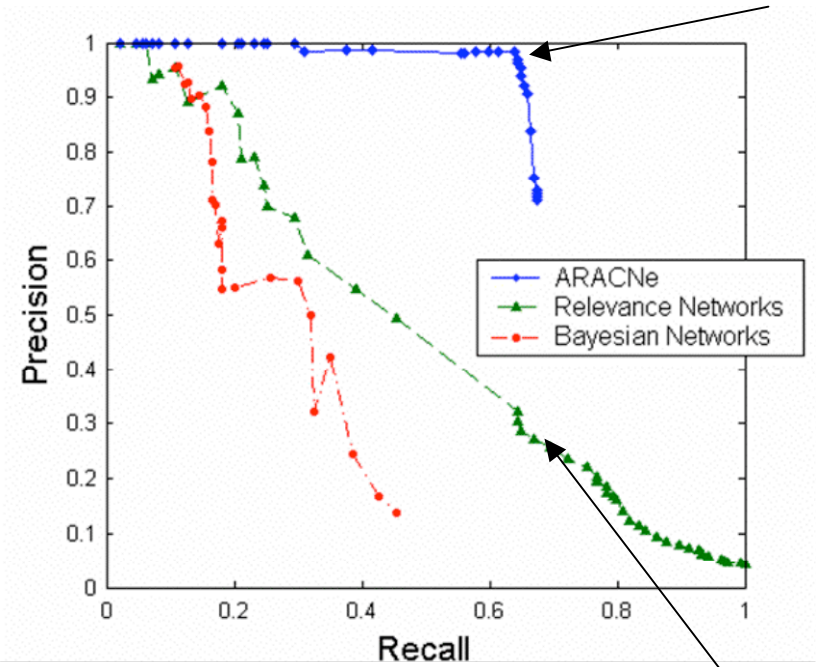
Hill dynamics

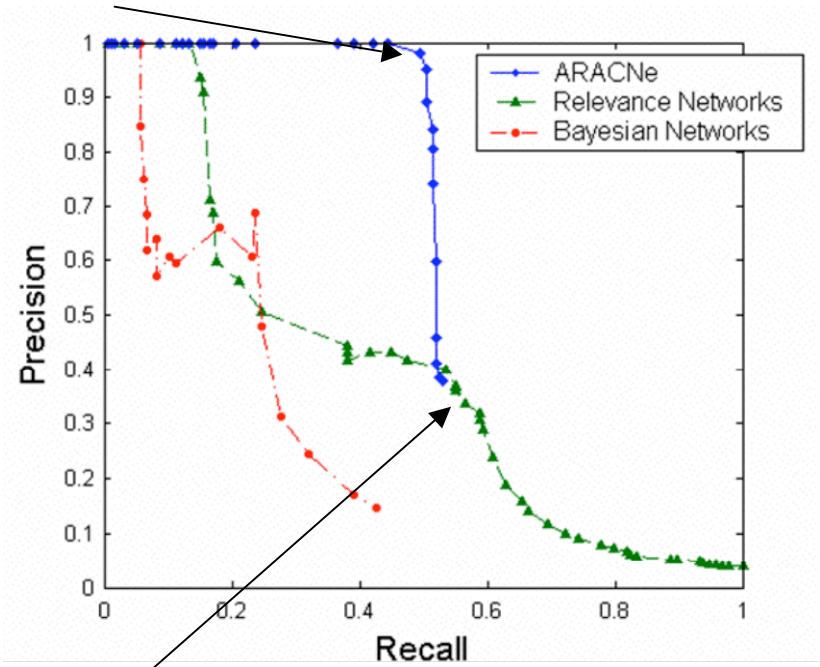
Performance vs BNs and RNs

Random

Scale-free

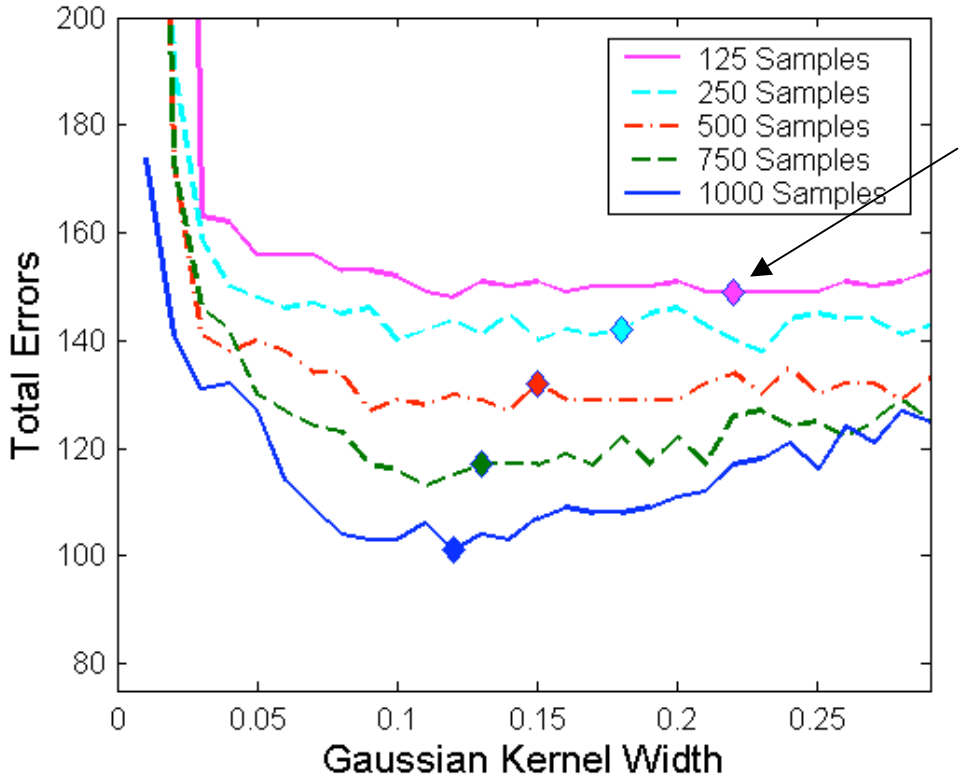
$p_0 = 10^{-5}; N_{TP} - N_{FP} = \max$





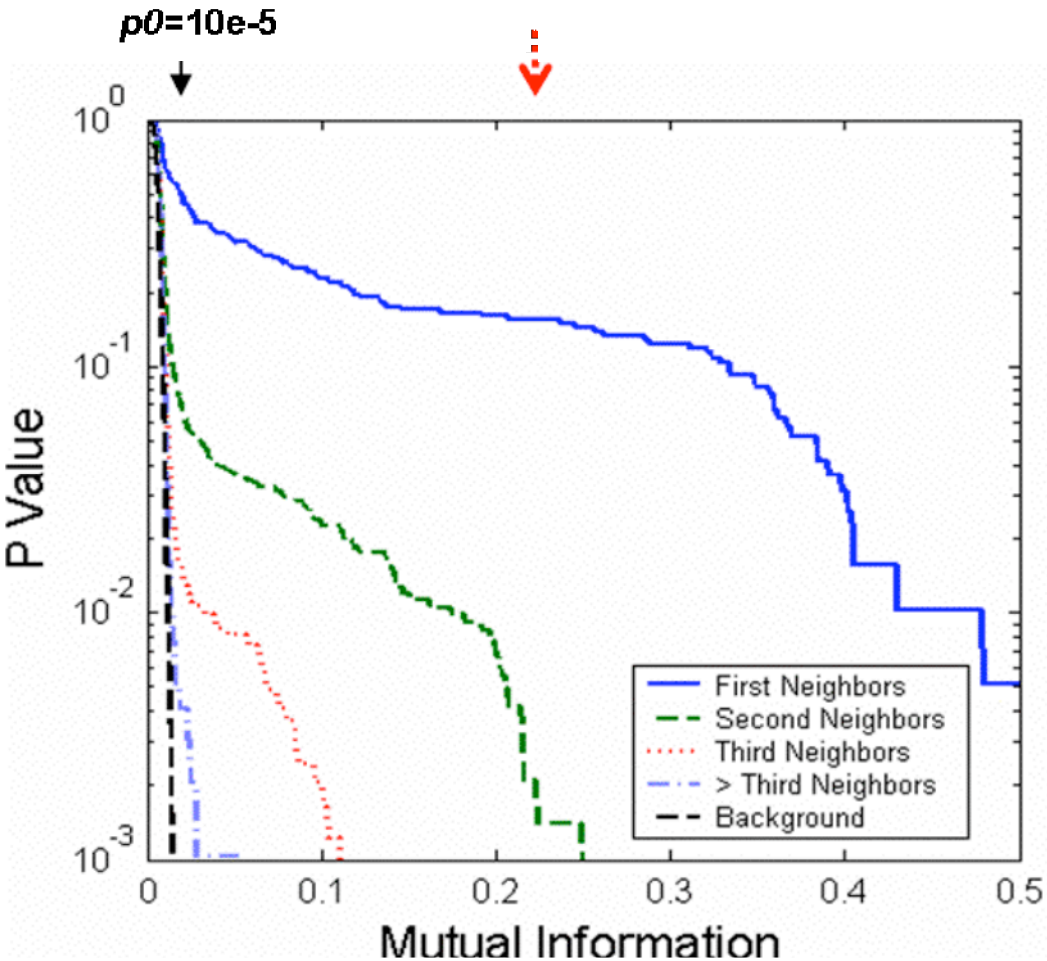
$p = 10^{-5}$

No sampling catastrophe!

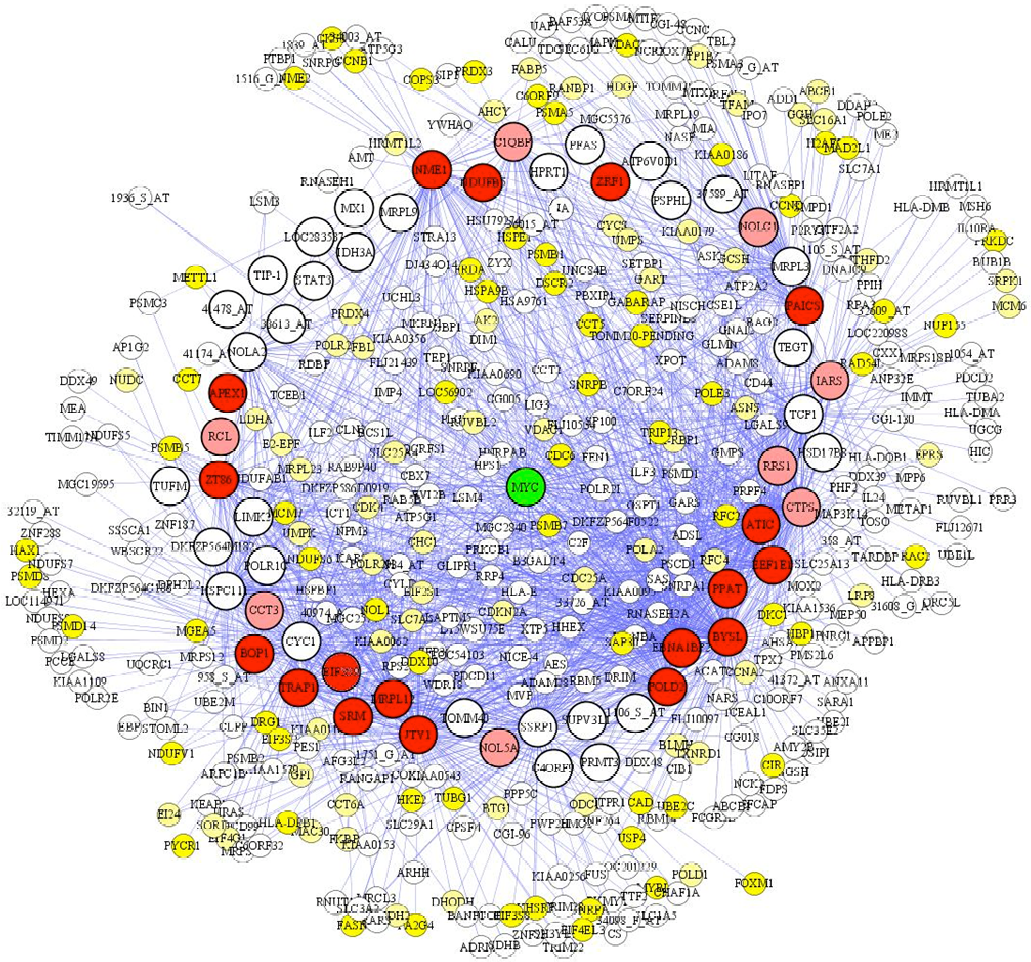


Ensemble best *h*
= best reconstruction

Why RNs fail


p0=10e-5
P Value
Mutual Information
First Neighbors
Second Neighbors
Third Neighbors
> Third Neighbors
Background

c-MYK centered network

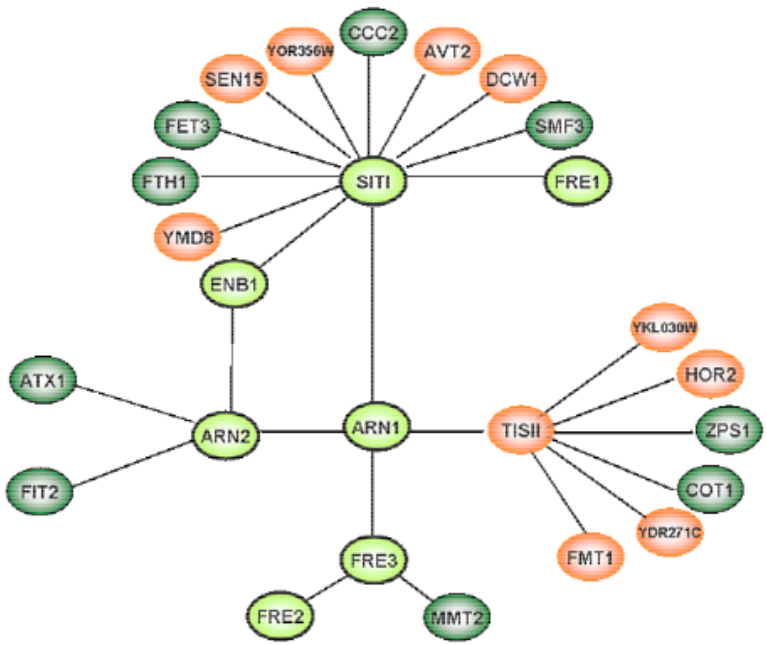


1st neighbors:
56
Of these putative targets:
Pre-known
22
ChIp-confirmed
11/12

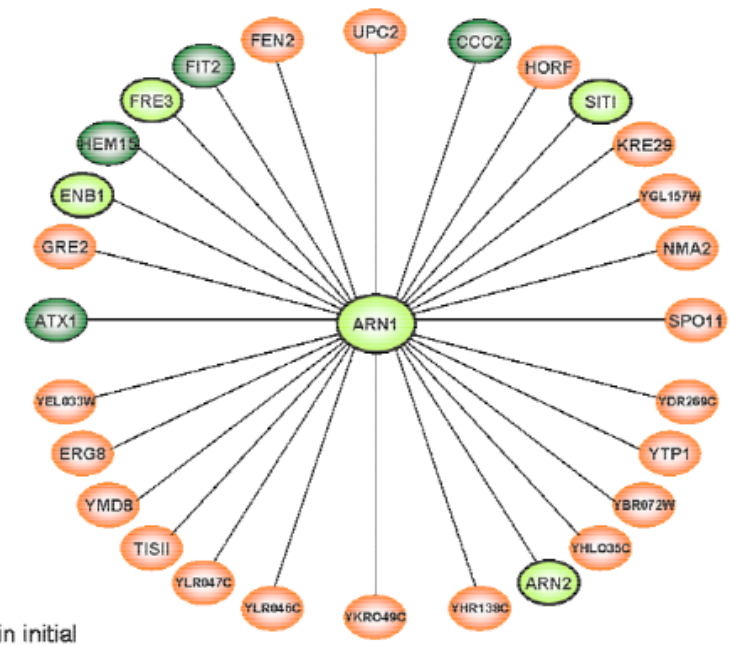
2nd neighbors:
weaker enrichment

ARN1 iron uptake in Yeast

0% tolerance



15% tolerance



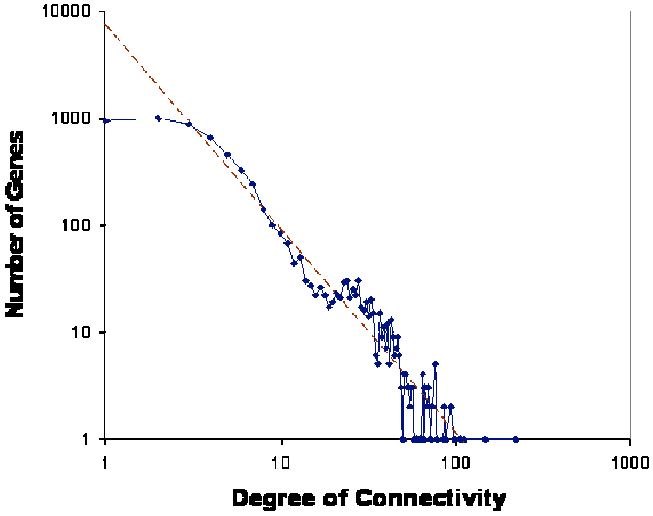
Directly involved in initial steps of iron-uptake pathway

Enhances uptake of iron or other heavy metals

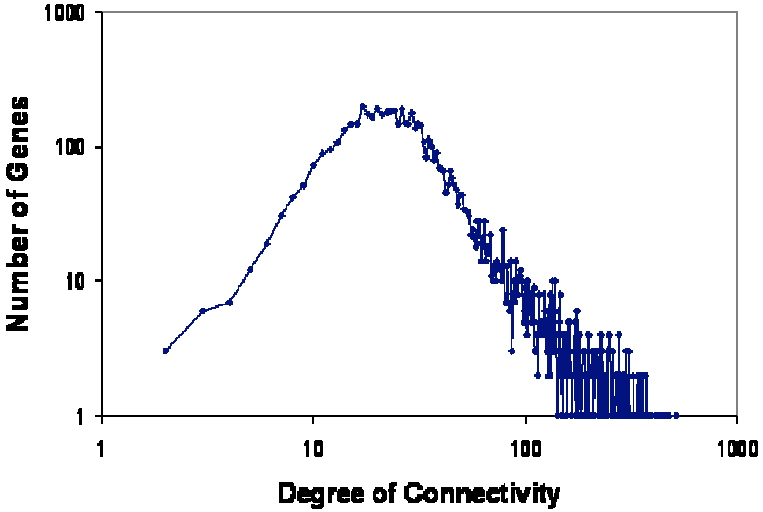
Not known to be related to uptake of iron or other heavy metals.

Yeast global network properties

Rosetta



Stress response



Hub three-way interactions (Conditional analysis)

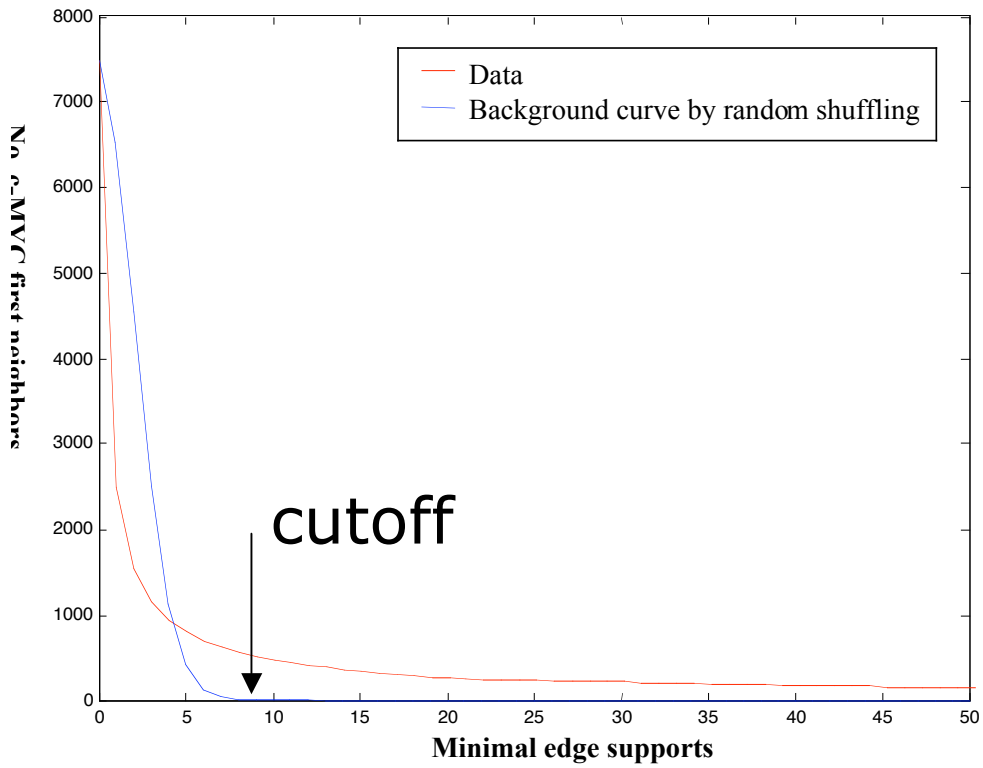
$$I(g_i, g_j \mid G_\mu^*)$$

G_μ^* – coarse conditions (+/-) of correlated gene clusters;
1) Independent of hub
2) Large dynamic range

(does not distinguish 3-way from 2-way loops)

	G_1^+	G_1^-	G_2^+	G_2^-	…	…	G_M^+	G_M^-
Edge 1	1	0	1	0	0	0	1	0
Edge 2	0	1	0	0	0	1	0	0
:	0	0	1	0	1	0	0	0
Edge N	1	0	0	1	0	0	0	1

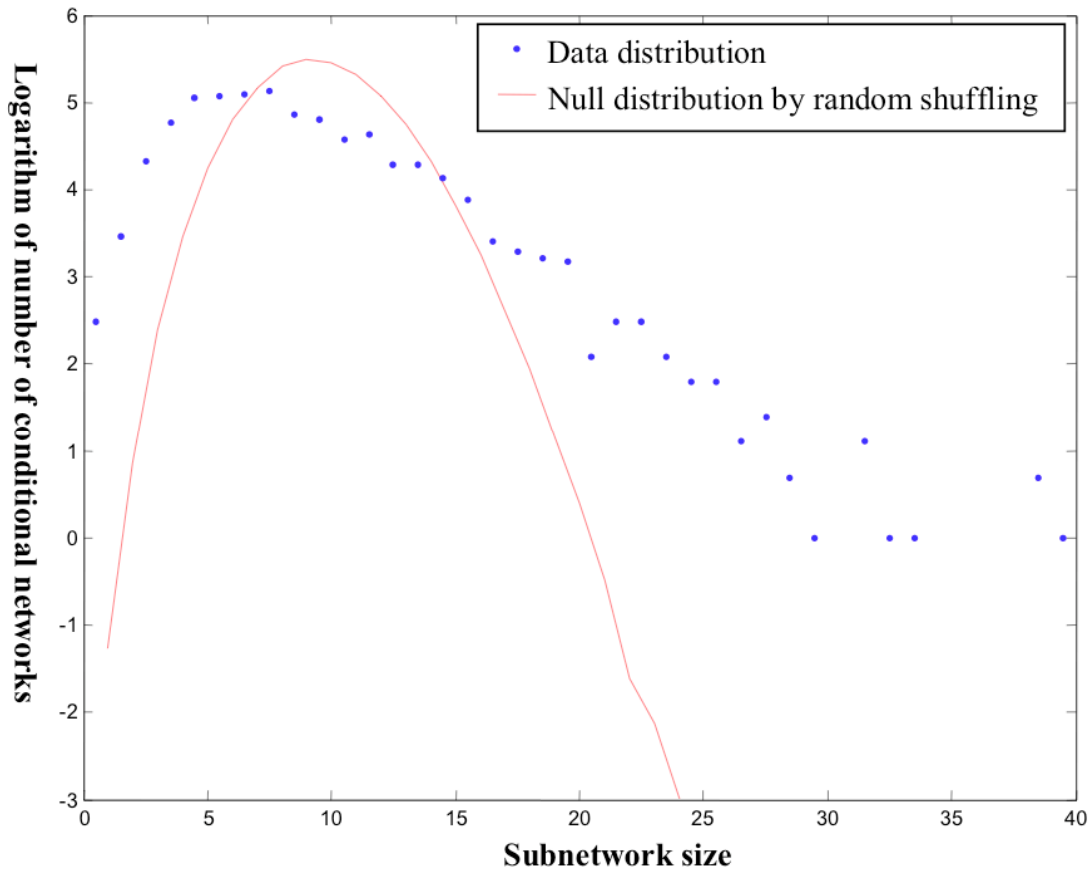
Edge support conditions set size



#	N_P	N_V	E.	N_{FP}	P
1	2422	437	0.18	6520.1	1
2	1458	278	0.19	4541.5	1
3	1066	224	0.21	2514.1	1
4	847	182	0.21	1131.6	1
5	710	157	0.22	423.13	0.60
6	591	132	0.22	136.04	0.23
7	511	119	0.23	37.03	0.072
8	459	110	0.24	9.18	0.02
9	406	104	0.26	1.9	< 0.01
10	367	96	0.26	0.37	< 0.01

Probably better reconstruction than original ARACNE

Conditional network sizes



Regulators, indeed

Of 168 regulators:

GO Category	N_C	N_{Tot}	GO N_C	GO N_{Tot}	P
Transcription Regulator Activity (MF)	21	116	2089	21014	0.0049
Protein Kinase (MF)	13	116	1004	21014	0.003
IKBK/NFKB cascade (BP)	5	117	222	24373	0.004
Immune Response (BP)	19	117	1664	24373	0.0003
Humoral IR (BP)	9	117	378	24373	0.0001
Reg. of Transcription, DNA-dep. (BP)	26	117	1697	24373	1.2×10^{-7}