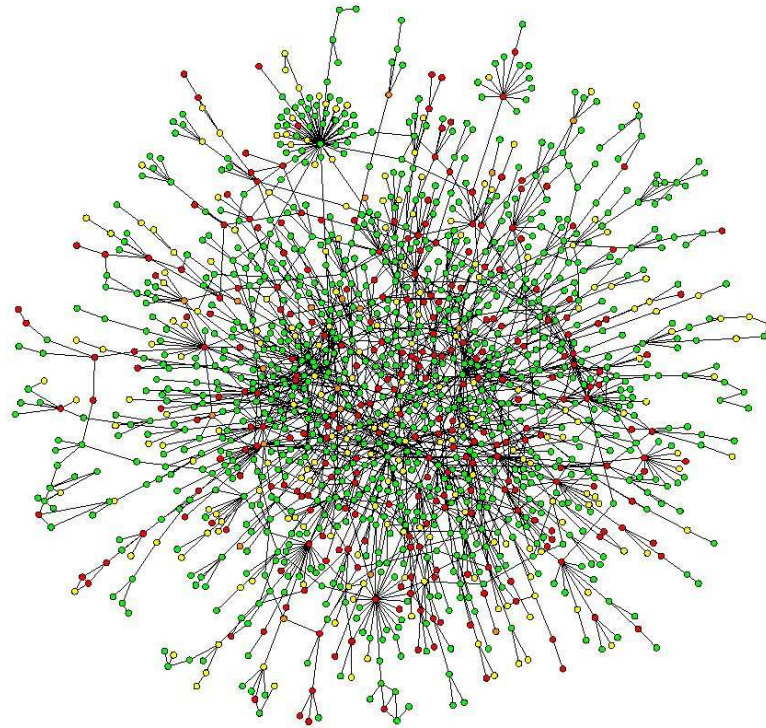


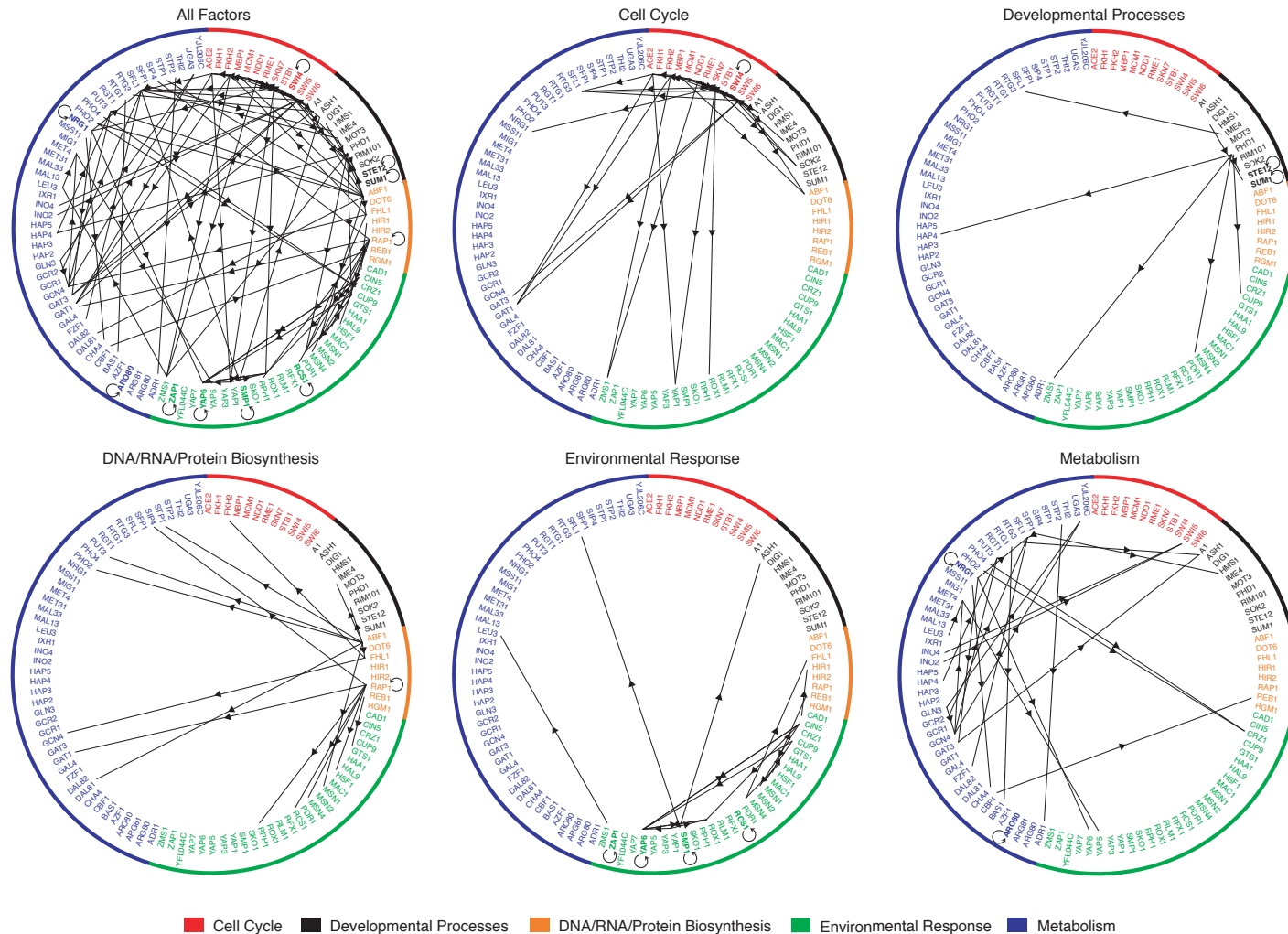
What do we do with genome-scale interaction networks?

- Genome-wide “interaction networks” result from ChIP-chip, yeast two-hybrid, co-expression, sequence analysis and other studies.



- Hoped for “modules” or “hierarchy” not apparent

Medium-scale subnetworks not readily apparent in Lee et al.



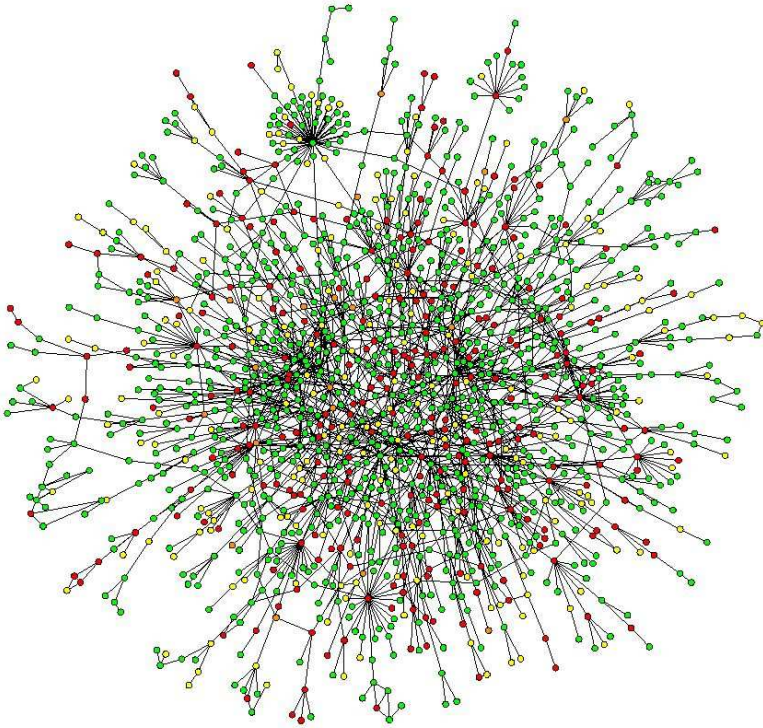
(Many links are missing, however.)

Why don't we see clear subnetworks?

- They're weak / hard to see? (Wang & Zhang, PLoS Comp Bio 2007)
 - Analyzed a hand-curated yeast protein-protein interaction network (with or without suspicious protein complex pulldowns) of ≈ 3700 – 3900 genes, average degree ≈ 3.5 – 3.75
 - Used simulated annealing to partition the network into “modules” so as to maximize a modularity criterion: $\sum_s \frac{l_s}{L} - \left(\frac{k_s}{2L}\right)^2$, where s ranges over modules, L is total edges, l_s is edges in module s , k_s is total degree in module s .
- ⇒ The network (with or without complexes) is significantly modular (statistically)
- ⇒ These modules correlate statistically significantly (but perhaps weakly) to function annotations of yeast gene function

Why don't we see clear subnetworks?

The graph omits important information? Interaction types? Activity?



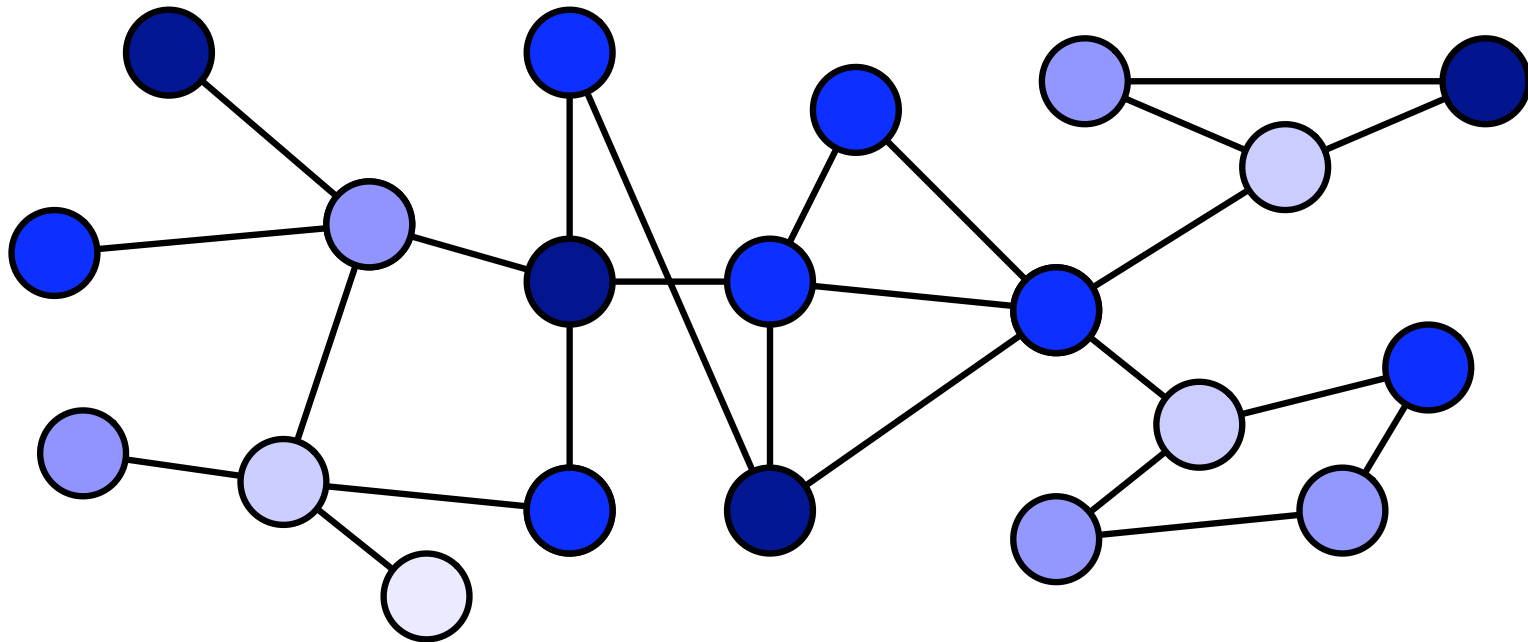
Graph shows interactions that may not occur simultaneously! May occur at different times, under different conditions, in presence of certain cofactors, ... or never, under in vitro conditions.

⇒ Can condition-specific information help us identify meaningful subnetworks?

Using expression information

(Ideker et al., Bioinformatics, 2002)

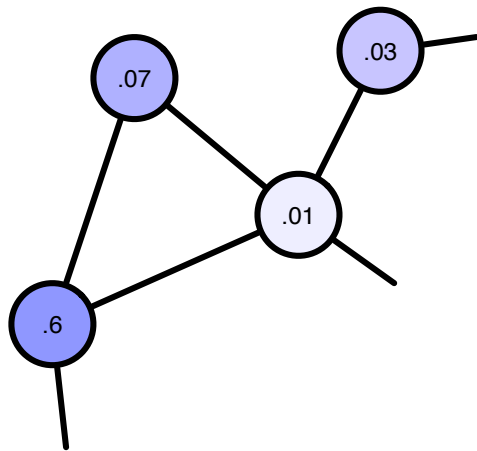
Idea: Look for subnetworks with high degree of differential expression (under some condition(s)).



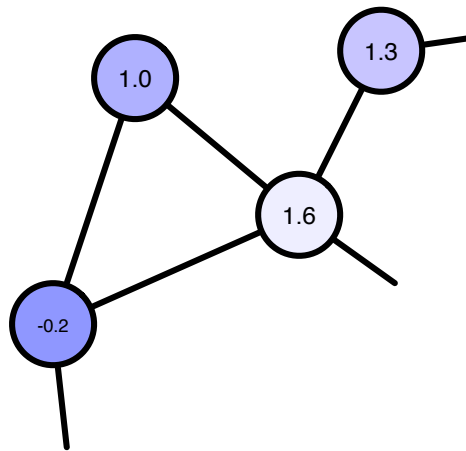
Ideker details

1. μ array(s) test for differential expression (\Rightarrow p -values)
2. Convert p -values to z -scores via inverse CDF of Gaussian
3. Look for subnetworks with most significant z -score sum, compared to random subnetworks of same size.

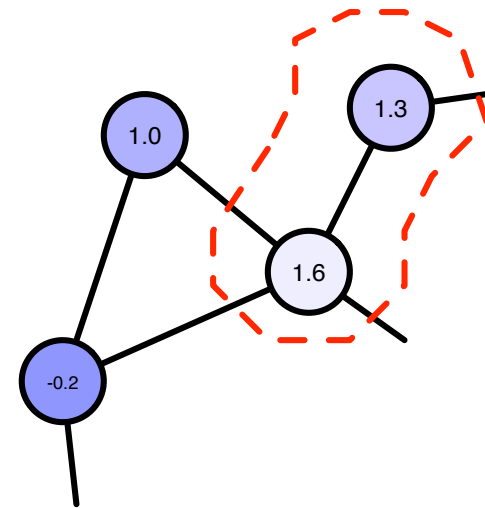
(1)



(2)



(3)

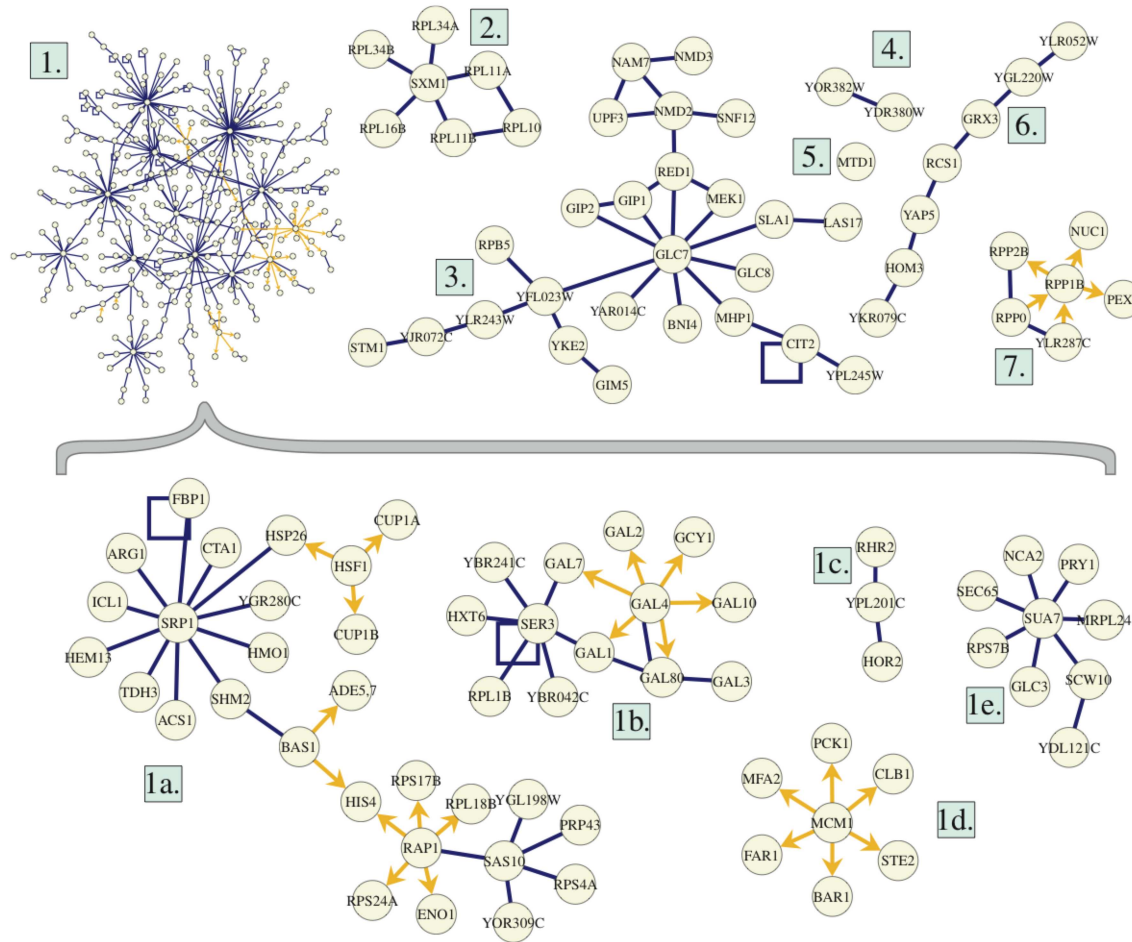


Exact optimization intractable; Ideker suggests simulated annealing.

One of Ideker's studies

- Constructed interaction network using 7145 protein-protein interactions from BIND¹ and 317 protein-DNA interactions from Transfac
 - Differential expression based on 20 yeast microarray measurements under Gal gene knockouts
-
1. BIND = Biochemical Interaction Network Database — an online database of interaction culled from the literature — privatized several years ago, now defunct?
 2. Transfac = Online database of transcription factors and association information

Results



1 Highest-scoring sub-network (340 genes)

1b Gal subnetwork

1d Mating response & cell cycle arrest