

CS 5984: COMPUTATIONAL SYSTEM BIOLOGY

Motifs, themes and thematic maps
of an integrated *Saccharomyces
cerevisiae* interaction network
Zhang et al. J.Biol 2005

Pallavi Sharma

INTRODUCTION

- Cellular system can be described as a web of relationship/interactions among different entities like genes, proteins etc.
- These give rise to various types of networks.
 - Protein – Protein interaction
 - DNA-Protein interaction etc..

- Such kinds of relationships often overlap each other.
- E.g. Protein which interact have similar expression profile.
- Challenge is to understand how are they organized in an integrated network.
- Also to find simple recurring patterns found in such integrated network.

DATASETS

- Five types of biological interactions
 - Protein – Protein
 - Stable interactions defined by shared membership in a protein complex.
 - Protein Complex was taken from MIPS
 - Total Interactions : 49,537
 - Transcriptional Regulation
 - Genome wide ChIP study of 106 transcription factors.
 - Total Interactions: 4357

- Cont..

- Sequence Homology

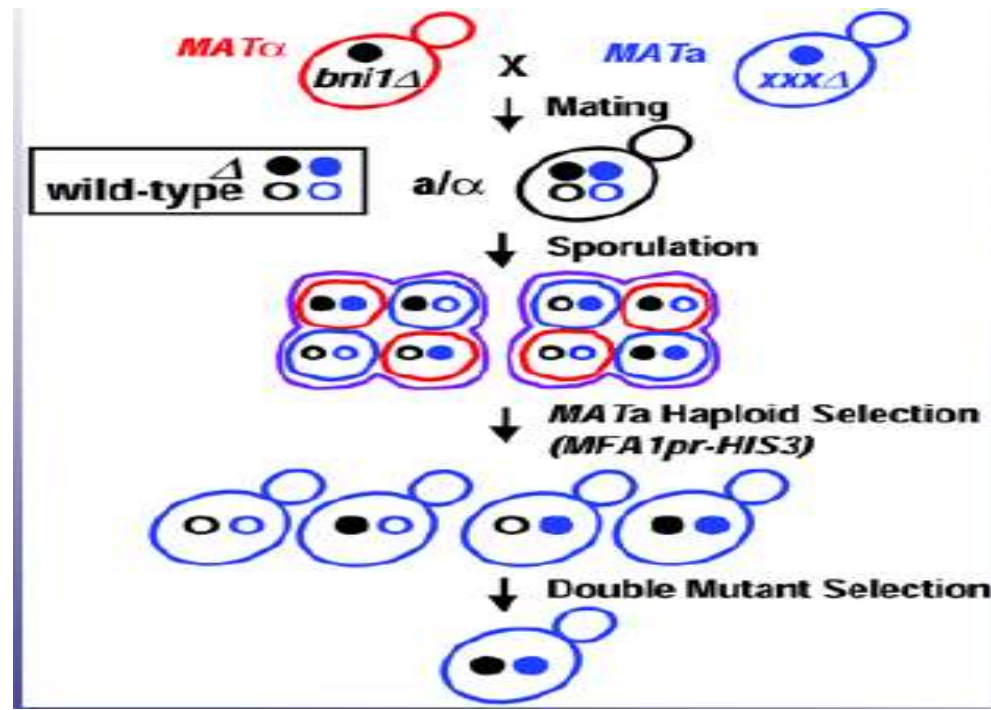
- Genome wide BLAST search.
 - Basic Local Alignment Search Tool.
 - Scoring Matrix, e.g. BLOSUM62
 - Pairs with $E < 10^{-3}$, homologous.
 - Total Interactions: 40,438

- Correlated Expression

- mRNA expression from Rosetta compendium microarray dataset.
 - Pearson correlation was calculated for all pairs of proteins
 - If it was > 0.6 , exists correlated expression.
 - Total Interactions: 57,367.

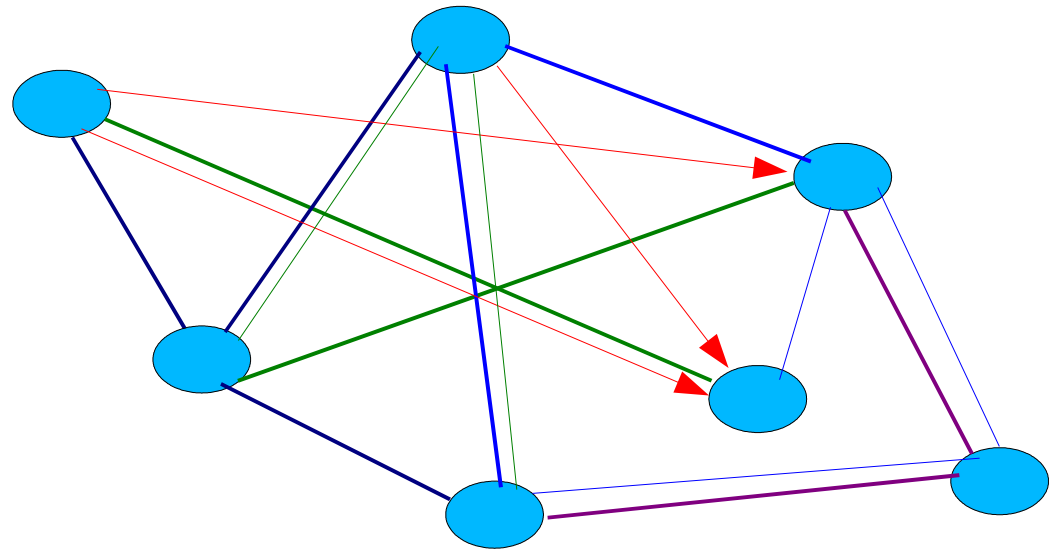
• Synthetic, Sick and Lethal

- Synergistic or antagonistic consequence of mutation in two or more genes.
- Reveals gene pairs where one gene compensates for the loss of other.
- 80% of yeast genome is non essential. Its a efficient way to determine functions of such genes.
- SGA Analysis, 132 query genes and 5000 array genes. Total Interactions: 3060



NETWORK

- Nodes: 5831
- Links: 154,759

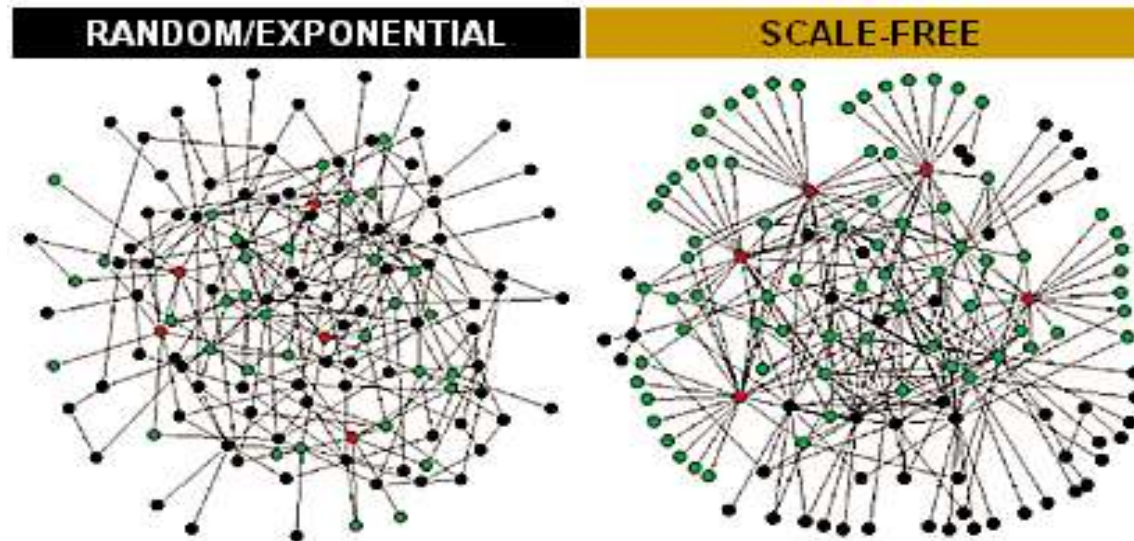


Key

- S: synthetic sickness or lethality
- H: sequence homology
- X: correlated expression
- P: stable physical interaction
- > R: transcriptional regulation

Biological vs Random Network

- Scale Free
- Small World
- Neighbourhood Clustering



Source: the journal Nature

•NETWORK MOTIFS

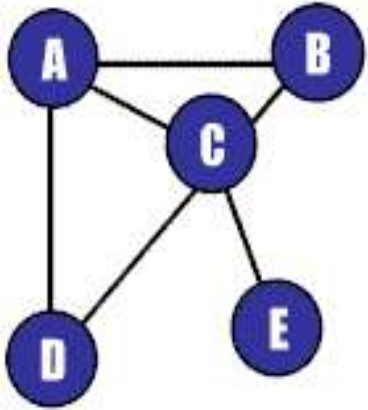
- Simple recurring patterns of interactions.
- Different networks exhibit different motif profiles. E.g. Feed forward motif of TRN.
- 3-node interconnection patterns = Triangles.
- E.g. Sequence homology, correlated expression.

•NETWORK THEMES

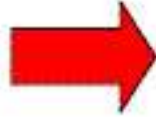
- Higher order recurring interconnection patterns that encompass multiple occurrence of network motifs.
- Biologically meaningful

ALGORITHMS

- Detect Network Motifs:
 - Find n node subgraphs in a connectivity matrix 'M', $n=3, 4$.
 - For each nonzero element (i, j) :
 - Loop through all connected elements $M_{ik} = 1, M_{ki} = 1, M_{jk} = 1$, and $M_{kj} = 1$. This is recursively repeated with elements $(i, k), (k, i), (j, k)$, and (k, j) until an n -node subgraph is obtained.
 - count the number of appearances of each type of subgraph in the network.

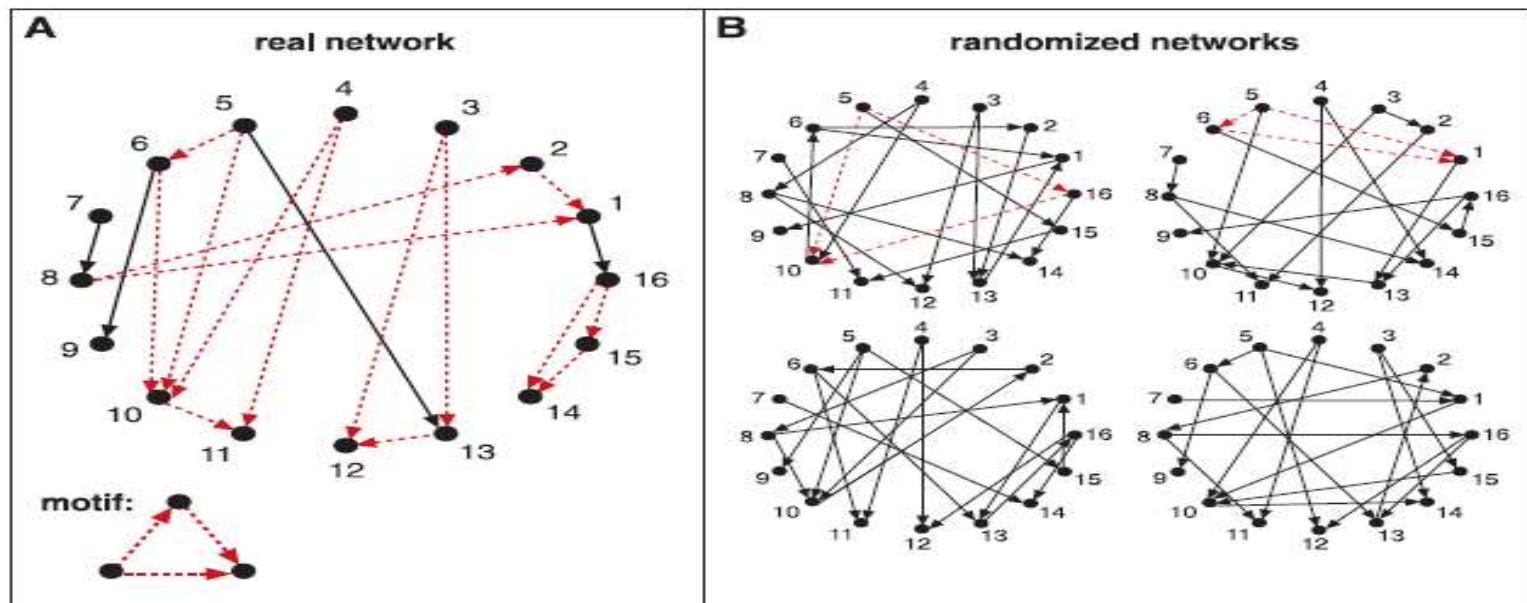
Network**Connectivity Matrix**

	A	B	C	D	E
A	0	1	1	1	0
B	1	0	1	0	0
C	1	1	0	1	1
D	1	0	1	0	0
E	0	0	1	0	0



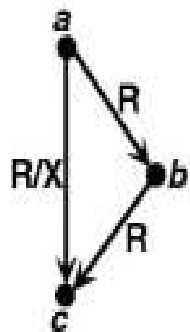
- For each interconnection pattern, find the number of subgraphs matching this pattern in real and random networks.
- A subgraph is a match to pattern if it can be transformed to pattern by any combination of node identity permutation or link removals.
- P value is the fraction of randomized networks having at least the same number of matching subgraphs as the real network.

- **Generating Randomized Networks**
 - Different interaction types in the integrated network were randomized independently and then overlaid.
 - Random networks are generated with property that expected degree of each node is same as in the real network. For details see Supplement of Network Motifs, Milo et al Science 2002.

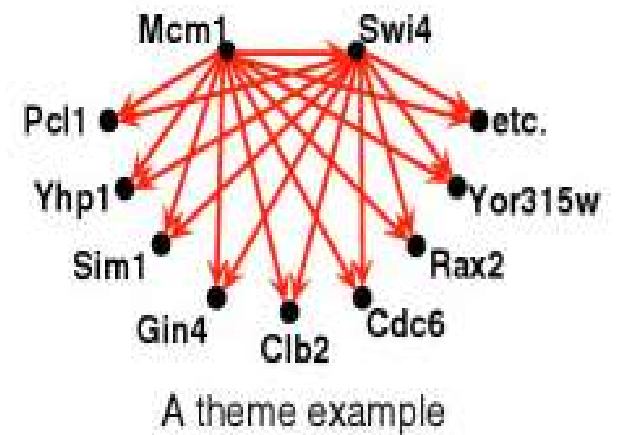
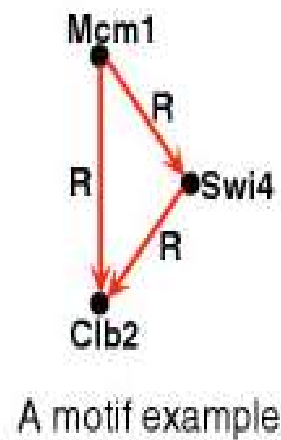


- **First motif set:**
 - Transcriptional regulation and correlated expression links.
 - **Feed Forward Theme:** Pairs of Tfs, one regulates other and both regulate a set of genes.
 - E.g. Swi4 and Mcm1

Motif set A

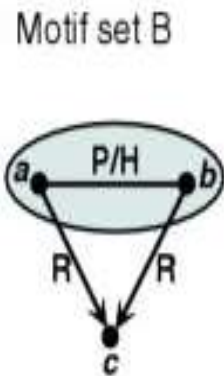


N_{real}	4.7×10^2	3.0×10^1
N_{rand}	$(2.6 \pm 0.5) \times 10^2$	5.4 ± 3.2

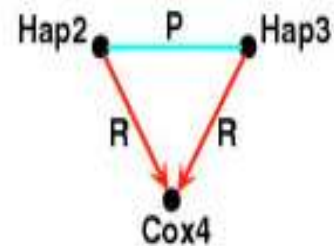


- Second Motif Set

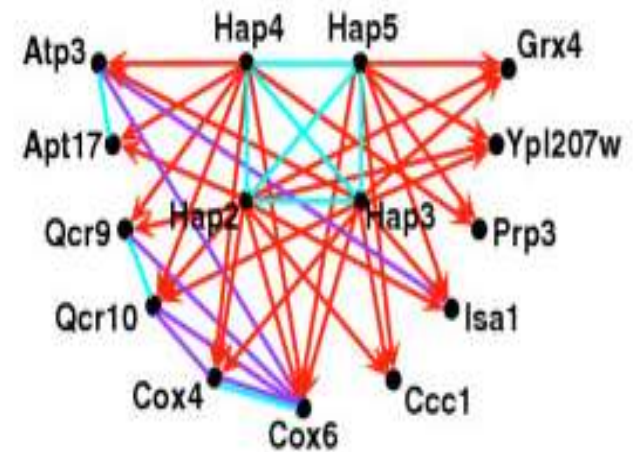
- Transcriptional regulation, protein interaction and sequence homology links.
- **Copointing Theme:** Multiple Tfs regulate a set of target genes and these are either derived from a same ancestral gene or function as a protein complex.



N_{real}	1.3×10^2	6.1×10^2
N_{rand}	3.3 ± 3.7	$(8.0 \pm 2.3) \times 10^1$



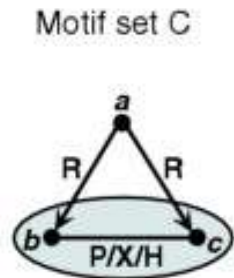
A motif example



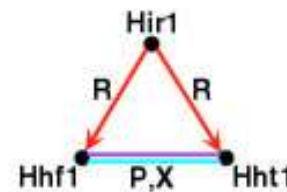
A theme example

- Third Motif Set

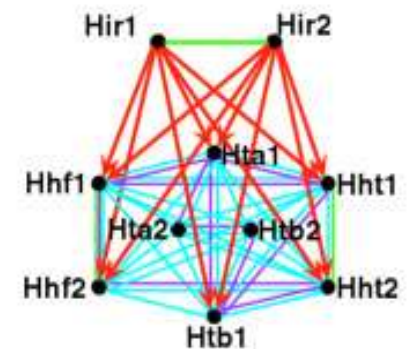
- Transcriptional regulation, correlated expression, sequence homology and protein-protein interactions.
- **Regulonic Complex Theme:** Co-regulated proteins, either components of a complex or related by gene duplication and divergence
- E.g. Members of histone octamer are regulated by Hir1, Hir2, histone



N_{real}	5.9×10^3	3.5×10^3	1.9×10^3
N_{rand}	$(5.4 \pm 0.5) \times 10^2$	$(2.7 \pm 0.3) \times 10^2$	$(5.3 \pm 0.5) \times 10^2$



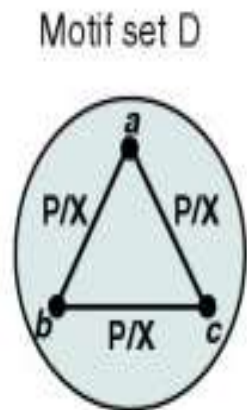
A motif example



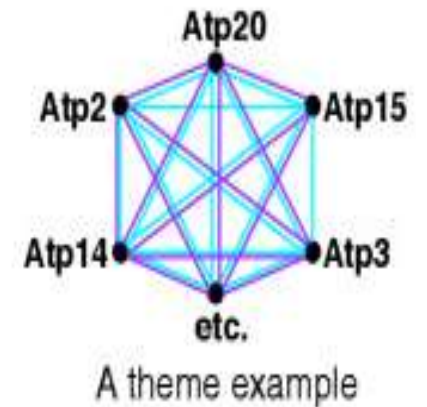
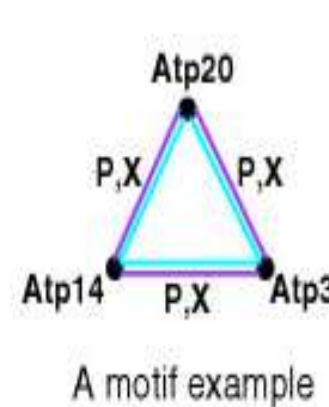
A theme example

- Fourth Motif Set

- Protein-protein interaction and correlated expression.
- **Protein Complex Theme:** Proteins corresponding to this motif are often members of same complex.
- E.g. ATP synthase complex.



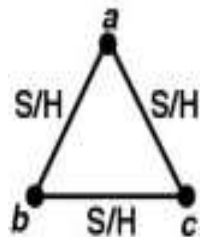
	D1	D2	D3	D4
N_{real}	5.7×10^5	9.9×10^4	6.7×10^4	1.2×10^6
N_{rand}	$(1.1 \pm 0.0) \times 10^5$	$(8.2 \pm 0.3) \times 10^3$	$(5.2 \pm 0.2) \times 10^3$	$(2.7 \pm 0.1) \times 10^4$



• Fifth Motif Set

- SSL interactions and sequence homology.
- SSL: Neighbours of same gene interact forming triangles.
- Homology: Relationship is Transitive.
- Homologous genes have a greater tendency to show SSL interactions.

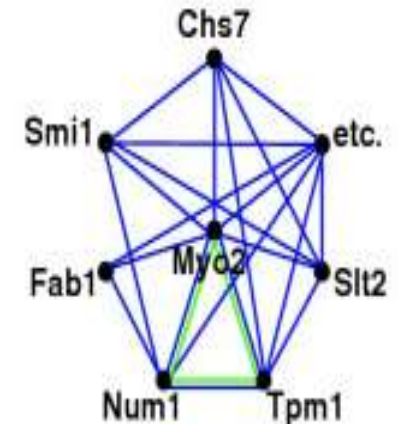
Motif set E



	E1	E2	E3	E4
N_{real}	2.7×10^3	9.8×10^2	3.2×10^3	5.6×10^5
N_{rand}	$(1.7 \pm 0.1) \times 10^3$	$(3.8 \pm 0.4) \times 10^2$	$(1.3 \pm 0.1) \times 10^3$	$(1.0 \pm 0.2) \times 10^5$



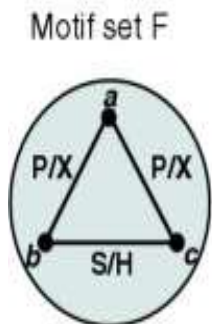
A motif example



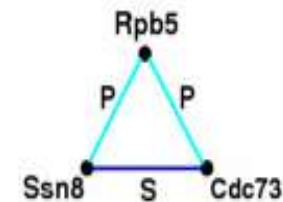
A theme example

• Sixth Motif Set

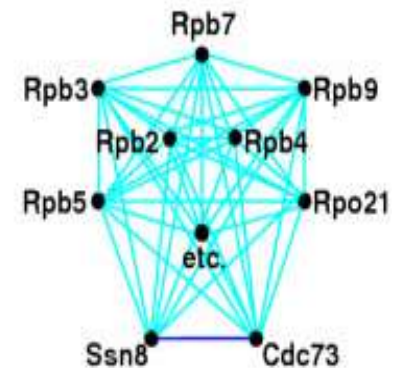
- SSL interactions, sequence homology, protein -protein interaction and correlated expression.
- **Protein complex with partial compensatory theme:** In this, only 2 out of 3 proteins is sufficient for the function. Proteins may coexist or be mutually exclusive.
- E.g. Ssn8 and Cdc73 associate with RNA plovmerase II



	F1	F2	F3	F4	F5	F6
N_{real}	1.2×10^3	2.7×10^2	2.8×10^2	4.1×10^4	1.1×10^4	4.4×10^4
N_{rand}	$(7.6 \pm 0.7) \times 10^2$	$(1.5 \pm 0.3) \times 10^2$	$(1.3 \pm 0.2) \times 10^2$	$(1.1 \pm 0.0) \times 10^4$	$(2.0 \pm 0.1) \times 10^3$	$(2.4 \pm 0.1) \times 10^3$



A motif example

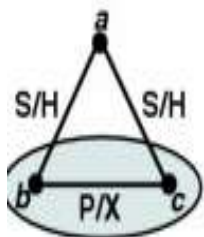


A theme example

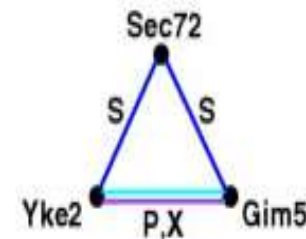
• Seventh Motif Set

- Protein interaction, Correlated expression, sequence homology, SSL interactions.
- **Compensatory protein and complex/process theme:** protein and a complex/process has compensatory function that causes SSL between protein and any member of complex. Also that protein can be a part of another complex/process.
- E.g. Gim complex and Sec72

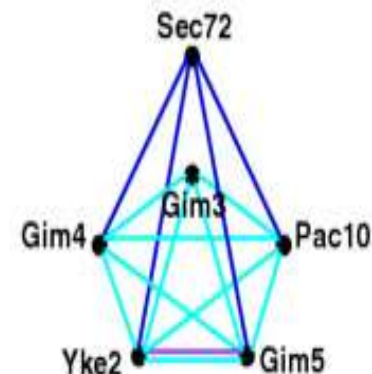
Motif set G



	G1	G2	G3	G4	G5	G6
N_{real}	2.0×10^3	7.2×10^2	3.0×10^4	2.8×10^2	2.5×10^2	4.3×10^4
N_{rand}	$(2.4 \pm 0.3) \times 10^2$	$(3.5 \pm 0.3) \times 10^2$	$(1.2 \pm 0.1) \times 10^4$	$(7.0 \pm 1.5) \times 10^1$	$(1.2 \pm 0.2) \times 10^2$	$(4.0 \pm 0.2) \times 10^3$



A motif example



A theme example

Thematic Maps

- A map of network rendered in terms of a theme.
- Two thematic maps generated
 - Compensatory Complexes
 - Regulonic Complexes.

Compensatory Complexes

- Only protein-protein interactions (MIPS) and SSL links (SGA) were considered.
- Nodes=Protein Complex and Links=SSL interaction. Multiple nodes belonging to same protein complex was merged into a single node.
- For each complex pair, no. of links was recorded and probability of getting same or more links was calculated if SSL partner was chosen randomly

- Each complex had 3 different degree types:
 - Query gene in complex-query gene outside complex
 - Query gene within complex-non query gene outside complex
 - Non query gene within complex-query gene outside complex.

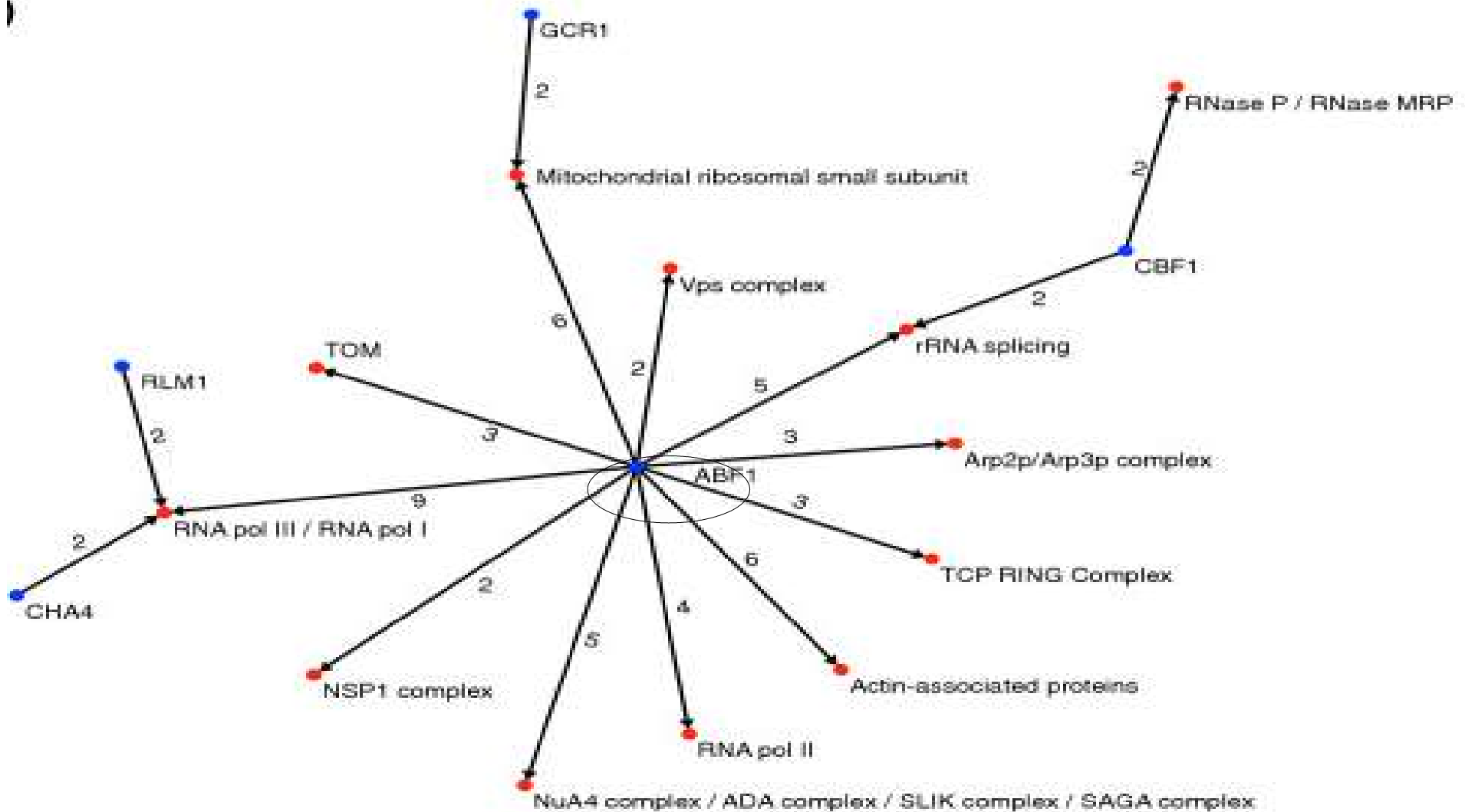
P value was calculated on the sum of three hypergeometric distributions.

Two complex are connected if $p < 0.05$ and ≥ 2 intercomplex SSL links.

Serve as guide to redundant systems in *S.cerevisiae*. Each complex act as a fail safe mechanism for other.

Regulon Complex

- protein protein interactions and transcriptional regulations.
- Nodes= Tfs or complexes. Links =Transcriptional Regulation.
- For each TF-complex pair, no.of links are recorded and probability of getting same or more links is found if chosen randomly.
- A TF pair is connected if $p < 0.05$ and have ≥ 2 regulatory links between TF and complex.



Blue= Transcription Factor

Red=Protein Complex

Link= If promoters of significantly large number of complex members are bound by TF.

<http://jbiol.com/content/supplementary/jbiol23->

<http://jbiol.com/content/supplementary/jbiol23->

Results and Discussions

- Network themes are more fundamental level of abstraction unlike network motifs which have artificial restrictions and often do not exist in isolation.
- Make predictions based on topological relationships involving multiple types of links.
- E.g. Predict a certain type of a link b/w nodes if it completes the structure matching an enriched motif or a theme.

- Novel predictions for SSL interactions using compensatory thematic maps.
- E.g. Sec62 and Sec63 has SSL interaction with Gim complex.
- Predict functions for genes involved in specific themes
- e.g. YOR315 gene involved in cell cycle. specific RNA polymerase II transcription factor activity.
Fujita et al Dec 2005, PMID: 16289536.

References

- Motifs, themes and thematic maps of an integrated *S.cerevisiae* interaction network. Zhang et.al *J.Biol*, 2005.
- Network Motifs: Simple building blocks of complex networks. R.Milo et.al *Science*, 2002.
- Transcriptional Regulatory networks in *S.cerevisiae*. Lee et.al *Science*, 2002.
- Global Mapping of the yeast genetic interaction network. Tong et.al. *Science*, 2004.
- The origin of degree correlations in the internet and other networks. Park et.al *Phys Rev*, 2003.
- <http://jbiol.com/content/supplementary/jbiol23-s2.pdf>