



MAPK Signaling Pathway Analysis

ESD.342 – Spring 2006

May 11th 2006

Presentation by:

Gergana Bounova, Michael Hanowsky and Nandan Sudarsanam

Faculty: Chris Magee



Agenda

- Background
- Network Statistics: the Usual Suspects
- Structural Analysis: Motifs and Communities
- Other Datasets, Benchmarking
- Future Directions and Scope



MAPK Signaling Pathway

- Cellular Level Biological Network
 - Signaling pathways are used to respond to external stimuli and regulate cellular activities
 - MAPK's (Mitogen Activated Protein Kinase) – Transfer information through chemical reactions and mechanistic physical interactions.
 - One pathway, three species – does that give us any information
- Literature and Data sources
 - Literature sources in Protein network analysis
 - The proteomics initiative
 - Nature of data – an incomplete map
 - KEGG database (Kyoto Encyclopedia of Genes and Genomes): signaling transduction pathways
 - DIP (Database of Interacting Proteins): all known interacting proteins



Network Statistics - Comparisons

	Drosophila	Yeast	Human
# nodes	19	56	148
# edges	19	56	187
Edge/node	1.000	1.000	1.264
Directed?	No	No	No
Connected?	Yes	No (5 comp)	No (16 comp)
Max,mean,min deg	6,2,1	8,2.154,1	15,2.831,1
Max degree node	9	6	98
Deg correlation	-0.630	-0.146	-0.306
Max,mean,min,betw	27,23.737,20	103,62.796,54	4639,1035.538,384
Max betw node	3,14,15,16,17	9	12
Clust coeff C1,C2	0	0.064	0.008
Max clust coeff node	-	1	1
# triangle loops	0	3	3
Mean path length	3.836 (20% n)	6.370 (11% n)	6.454 (4% n)
Network diameter	8.000	16.000	17

consistently close to 1

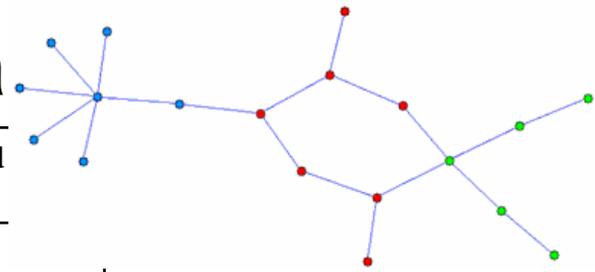
negative

consistently close to 0

Low reachability for small networks



Network statistics - Drosophila



Drosophila	Undirected	Directed	Random - Preserved Degree Sequence
# nodes	19	19	19
# edges	19	19	19
Edge/node			
Directed			
Connected			
Max,mean,min			
Max degree			
Deg correlation	-0.630		-0.203
Max,mean,min,betw	27,23.737,20	14,8,1	19, 19, 19
Max betw node			
Clust coeff CB			
Max clust coeff			
# triangle loops			
Mean path length	3.6	0.75	3.4
Diameter	10	10	7
	degree	betweenness	clustering coefficient
			path length

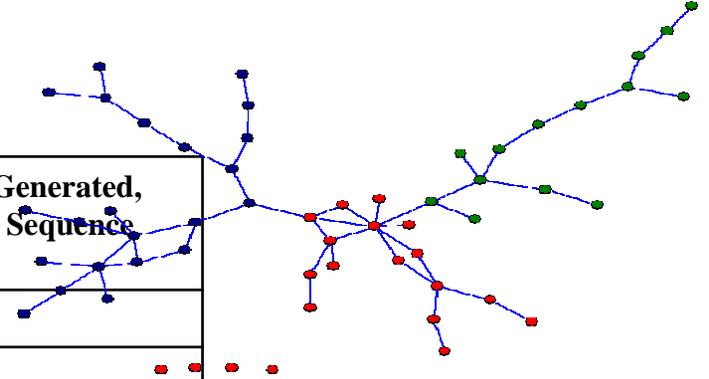
as expected

completely different!

consistent



Network statistics - Yeast



Yeast	Undirected	Directed	Yeast – Random Generated, Preserved Degree Sequence (GC)
# nodes	56	56	52 (GC)
# edges	56	56	56
Edge/node	1.000	1.000	1.077
Directed?	No	Yes	No
Connected?	Yes	Yes	Yes
Max,mean,min degree	8,2.1,0	8,2,0	2.154
Max degree	6	In: 6, Out: 6	9
Deg correlat	-0.14	0.092	0.092
Max,mean,min,betw	103,62.796,54	28,12.125,1	85, 56.7, 52
Max betw no	9	8,19	27
Clust coeff C	0.083	0.020	0.083
Max clust coeff	1	1	1
# triangle loop	3	0	3
Mean path length	6.376	4.779	10
Network diameter	16.000	11.000	10.000





Network statistics - Human

Human	Undirected	Directed	Human undirected, randomly-generated (GC)
# nodes	148	148	130
# edges	187	187	184
Edge/node	1.264	1.264	1.415
Directed?	No	Yes	No
Connected?	No (16 comp)	No	Yes
Max,mean,min deg	15,2.831,1	15,2.527,0	15, 2.831, 1
Max degree node	98	In: 23, out: 98	82
Deg correlation	-0.306		-0.323
Max,mean,min,betw	4639,1035.538,384	383,37.669,1	362, 241.6, 160
Max betw node	12	145	25,33,41,70,72,73,82
Clust coeff C1,C2	0.0075, 0.0045	0, 0	0.122
Max clust coeff node	1	-	1
# triangle loops	3	0	9
Mean path length	6.454	3.931	4.286
Network diameter	17	11.000	9.000

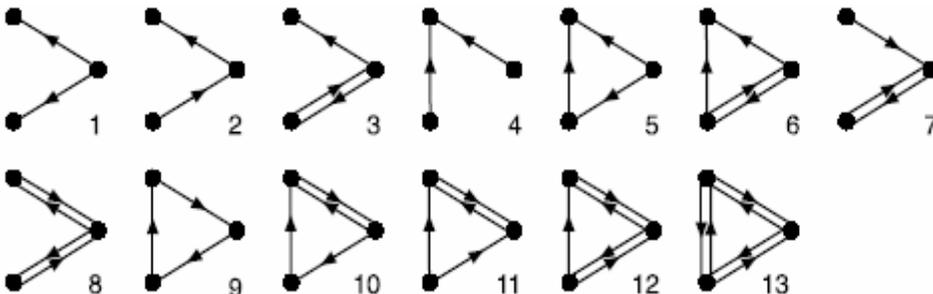
Becoming more evident: the real pathway has more built-in flexibility, even though reachability remains low



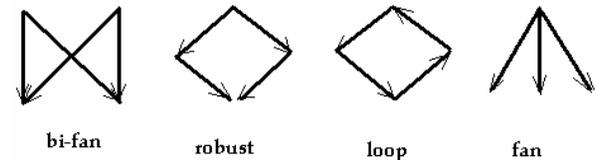
Motifs – Background

- **Coarse Graining** – an important bottom-up method of understanding network structure, by uncovering global patterns.
 - This helps us go beyond the global features and understand the relevance of certain structural elements.
 - Motifs are *statistically significant patterns* of connections that recur through out the network. They serve as the *basic building blocks* of the network.
 - Studies have shown that each network motif performs a key information processing function in biological networks.
- **Examples of Motifs studied in Biological networks:**

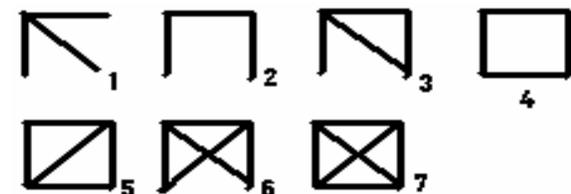
Directed 3-Node Motifs



Directed 4-Node Motifs

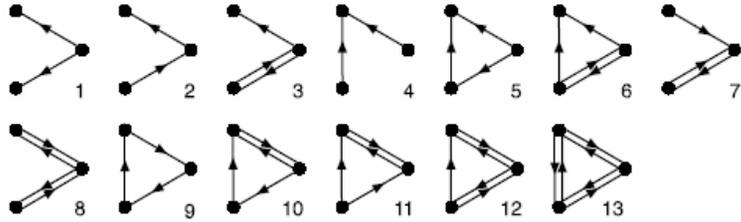


Undirected 4-Node Motifs



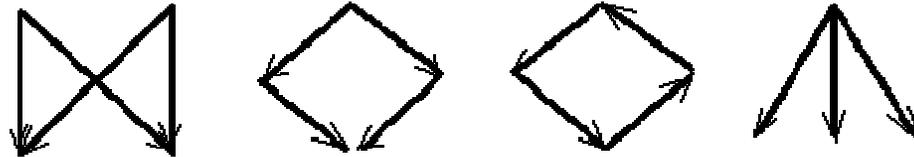
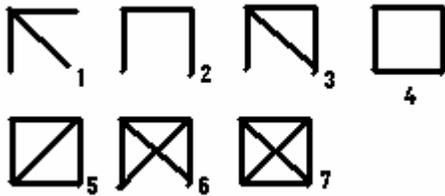


Motif Occurrences



Motif index	D.Mel	D.Mel. rand	Yeast	Yeast rand	Human	Human rand
1	24	36	40	40	538	326
2	38	16	128	104	600	694
4	8	12	38	74	272	352
5	-	3	6	3	9	15
9	-	-	3	-	-	12

2
1
3
4
5

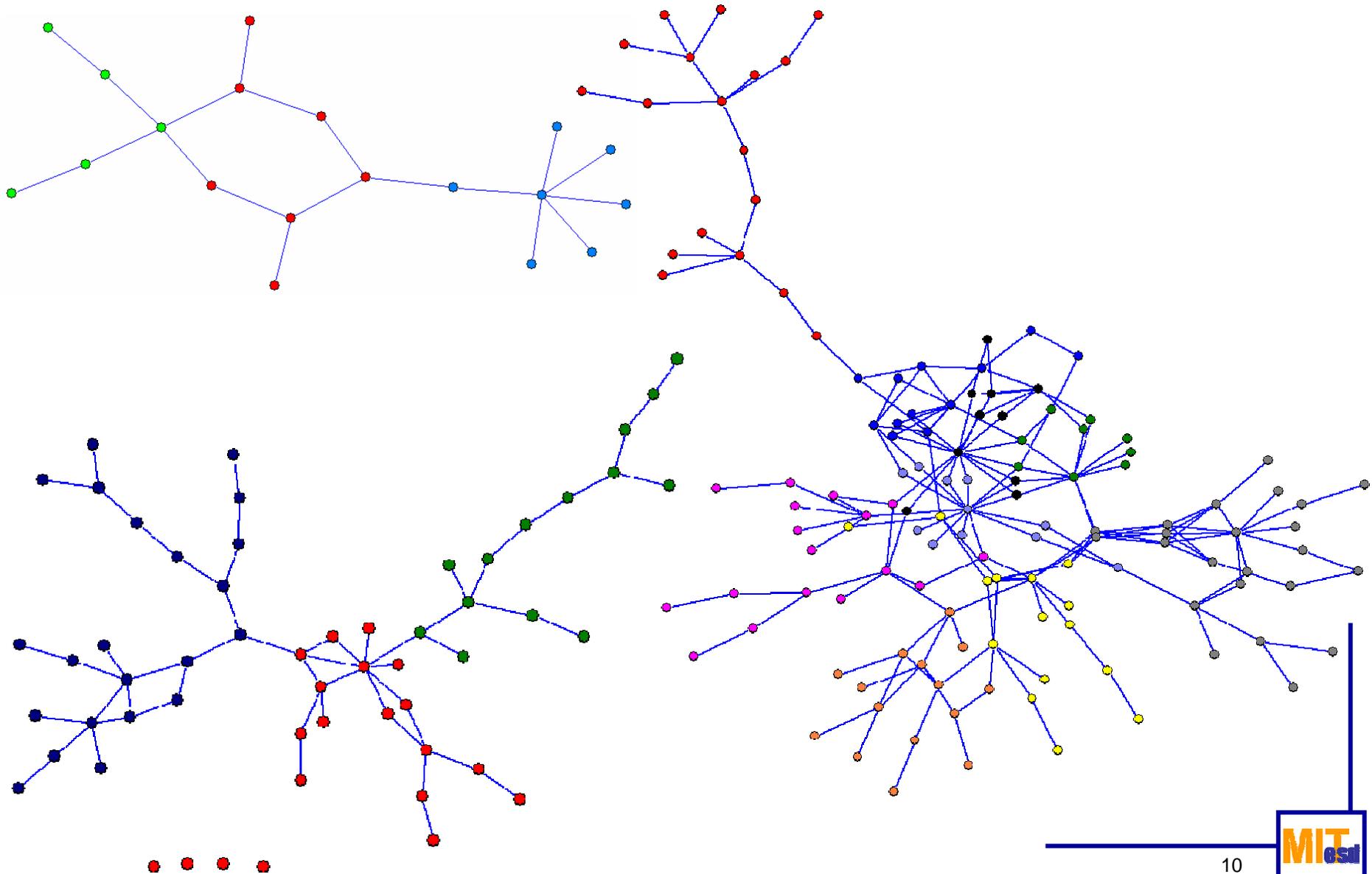


Motif index	D.Mel	D.Mel. rand	Yeast	Yeast rand	fan		loop		fan	
					Human	Human rand	Human	Human rand	Human	Human rand
1	81	66	228	297	4488	4269	-	-	-	-
2	64	82	264	406	2996	3334	4	-	204	4
3	-	25	105	-	150	545	4	-	164	16
4	-	-	8	4	400	28	-	-	-	12
5	-	-	8	-	24	30	15	66	1929	411



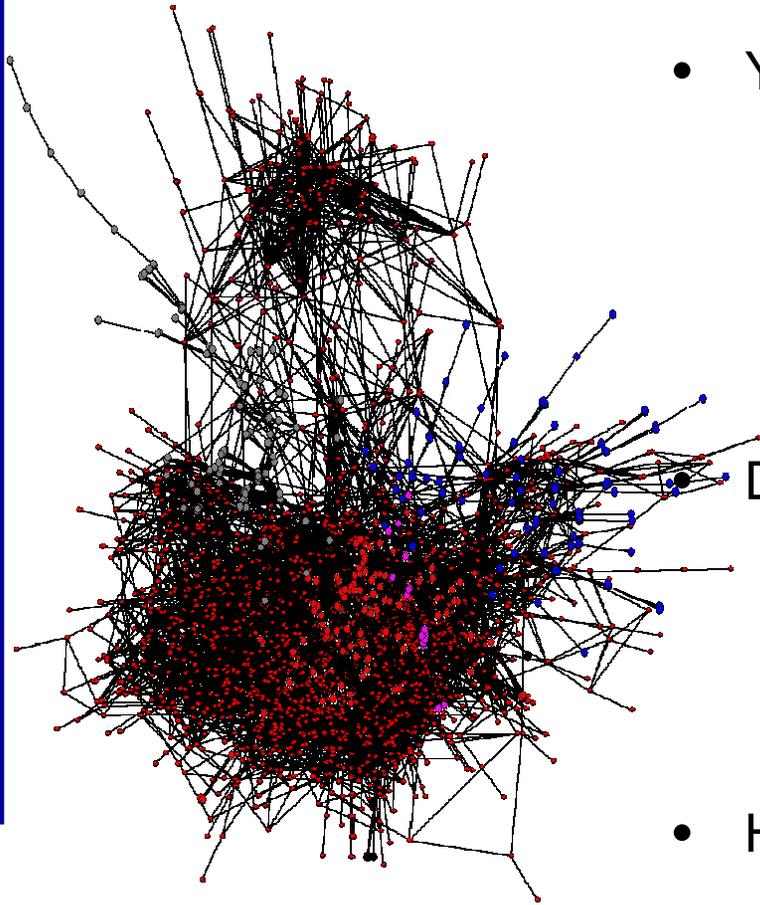


The Newman-Girvan View





All Protein Datasets



Yeast Core Proteins Main Component
Human Proteins Connected Component

- Yeast original: 8992 nodes, 5952 links
 - First refinement: 2554 nodes, 5728 edges
 - Second refinement: 2408 nodes, 5668 edges
 - *Edge/node: 2.4, clust=0.294, meanL=5.197, diam=14*
- Drosophila original: 28052 nodes, 22819 links
 - First refinement: 7451 nodes, 22636 edges
 - Second refinement: 7355 nodes, 22593 links
 - *Edge/node: 3.072, clust=0.016, meanL=8.009*
- Human original: 28155 nodes, 1397 links
 - First refinement: 1085 nodes, 1346 links
 - Second refinement: 939 nodes, 1276 links
 - *Edge/node: 1.359, clust=0.235, meanL=6.822*



Conclusions & Future Work

- Analyzed three pathway sets structurally and detected some similar patterns (modules, communities)
- Found betweenness as the best signature of pathways and a sign of flexibility in biological networks
- Building blocks are preserved across pathways, and certain motifs found by others do appear statistically significant (such as bi-fan, for example)

- Plan to perform coarse-graining on larger datasets to look for further structural matching
- Benchmarking of motifs and other statistical measures with whole known protein datasets



References

- Itzkovitz, S., Levitt, R., Kashtan, N., Milo, R., Itzkovitz, M., and Alon. U., Coarse-Graining and Self-Dissimilarity of Complex Networks. arXiv:q-bio.MN/0405011 v1, May 2004.
- Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N., *Lethality and centrality in protein networks*. Nature 411, 41–42 (2001).
- Maslov, S. & Sneppen, K. *Specificity and stability in topology of protein networks*. Science 296, 910–913 (2002).
- Milo, R., Shen-Orr, S. S., Itzkovitz, S., Kashtan, N. & Alon, U., *Network motifs: simple building blocks of complex networks*. Science **298**, 824–827 (2002).
- Sole, R. V., Pastor-Satorras, R., Smith, E., and Kepler, T. B., *A model of large-scale proteome evolution*, Advances in Complex Systems 5, 43{54 (2002).
- Wuchty, S., and Almaas, E., “Peeling the Yeast protein network,” Proteomics. 2005 Feb;5(2):444-9. 5(2), pp. 444–449, 2005.



Distribution Comparisons - Human

