

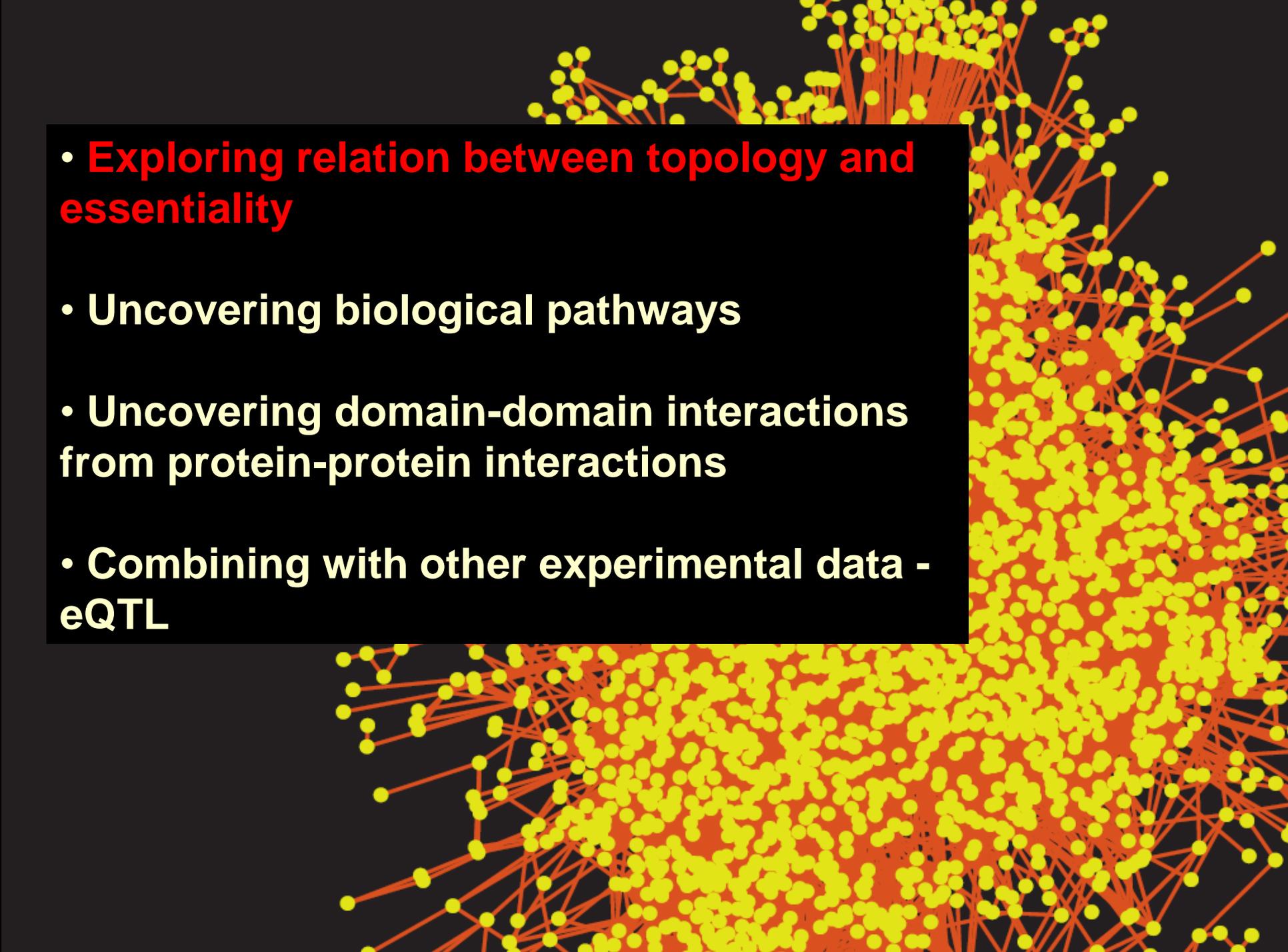
Reexamining connection between network topology and essentiality



Teresa Przytycka
NIH / NLM / NCBI



ISMB, 2009

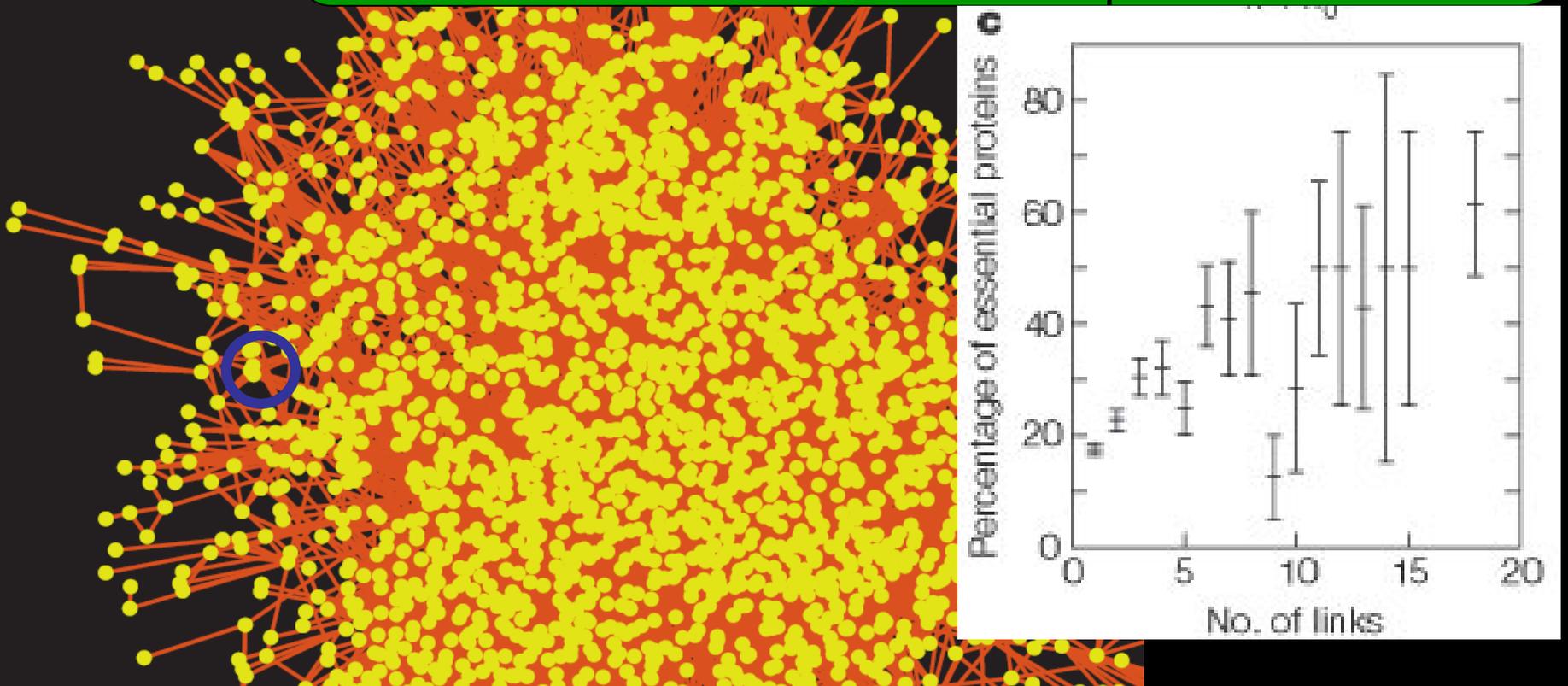
- 
- **Exploring relation between topology and essentiality**
 - **Uncovering biological pathways**
 - **Uncovering domain-domain interactions from protein-protein interactions**
 - **Combining with other experimental data - eQTL**

Is there a relation between graph-theoretical properties of a network and function?

The Centrality-Lethality Rule

(Jeong et al., Nature 2001)

High-degree nodes in a protein interaction network are enriched in essential proteins.



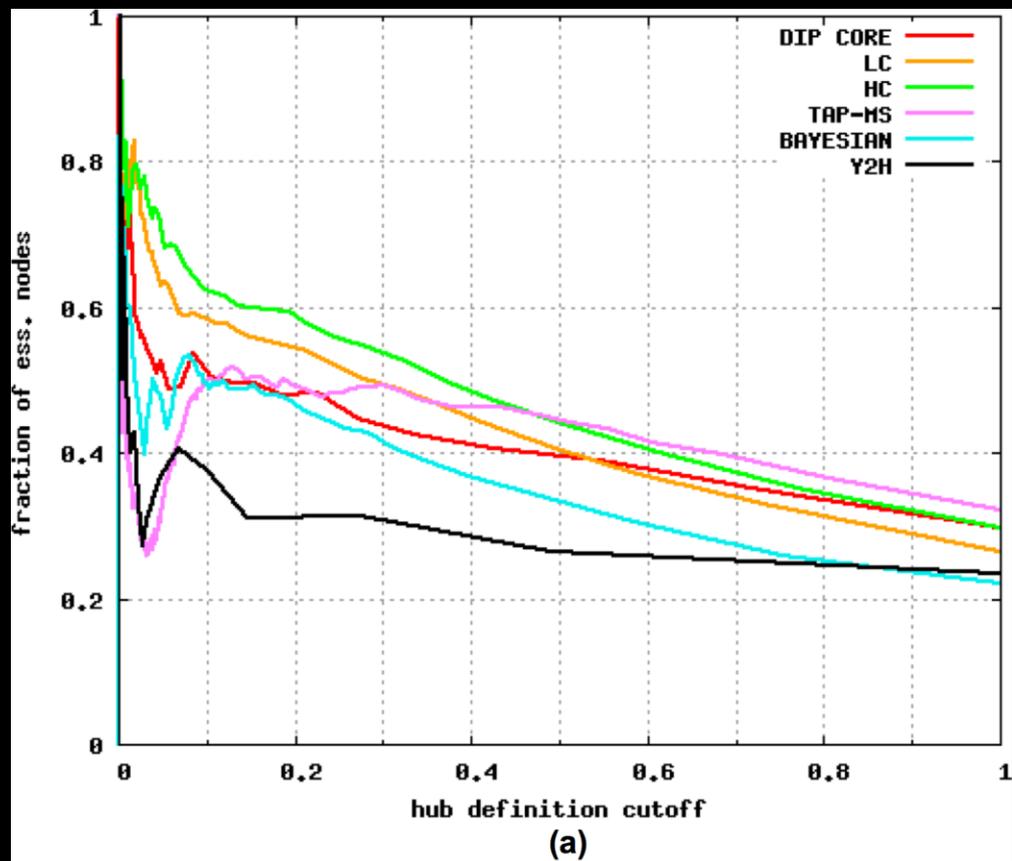
Are hubs indeed enriched in essential proteins?

Why are hubs enriched in essential proteins?

Protein Interaction Networks

- DIP CORE Deane et.al. 2002
 - high-confidence interactions from the DIP database
- LC (Literature Curated) Reguly et.al. 2006
 - interactions reported in small-scale experiments
- HC (High Confidence) Batada et.al. 2006
 - interactions reported by several independent studies
- TAP-MS Collins et.al. 2007
 - interactions derived from two high-throughput complex purification experiments
- BAYESIAN Jansen et.al. 2003
 - interactions derived in-silico (from experimental data) using Bayesian Networks formalism
- Y2H Ito et al. 2001

	Number of nodes	Number of edges	Avg. degree	Avg. clustering coeff.
DIP CORE	2,316	5,569	4.81	0.30
LC	3,224	11,291	7.00	0.36
HC	2,752	9,097	6.61	0.37
TAP-MS	1,994	15,819	15.87	0.60
BAYESIAN	4,135	20,984	10.15	0.26
Y2H	400	491	2.45	0.09



	Kendall's tau	Spearman's rho
DIP CORE	0.22 (1.1e-33)	0.25 (1.1e-34)
LC	0.32 (6.1e-99)	0.37 (3.3e-106)
HC	0.32 (1.1e-85)	0.37 (4.4e-92)
TAP-MS	0.24 (6.4e-37)	0.28 (3.6e-38)
BAYESIAN	0.27 (1.2e-91)	0.32 (2.4e-96)
Y2H	0.09 (2.6e-2)	0.10 (2.6e-2)

(b)

Why are hubs enriched in essential proteins?

- **The Centrality Hypothesis:** If removal of a node disrupts the “communication” between pairs of other nodes in the network, then the corresponding protein is likely to be essential (Jeong et al., Nature 2001)
- **The Essential PPIs Hypothesis:** All interactions are essential with uniform probability. High degree nodes are essential because they participate in many interactions and thus, with high probability, are adjacent to an essential interaction (He et al., PLoS Genetics 2006)
- **Our result:** Neither of the above is true. Alternative view is proposed.

Zotenko, Mestre, O’Leary, Przytycka. PLoS CB 2008
(highlighted in Nature Genetics Rev, Sept 2008)

Why are hubs enriched in essential proteins?

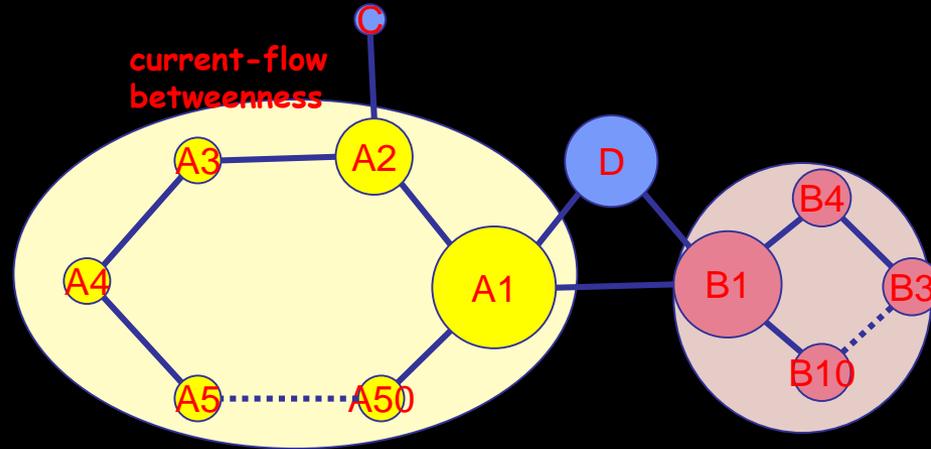
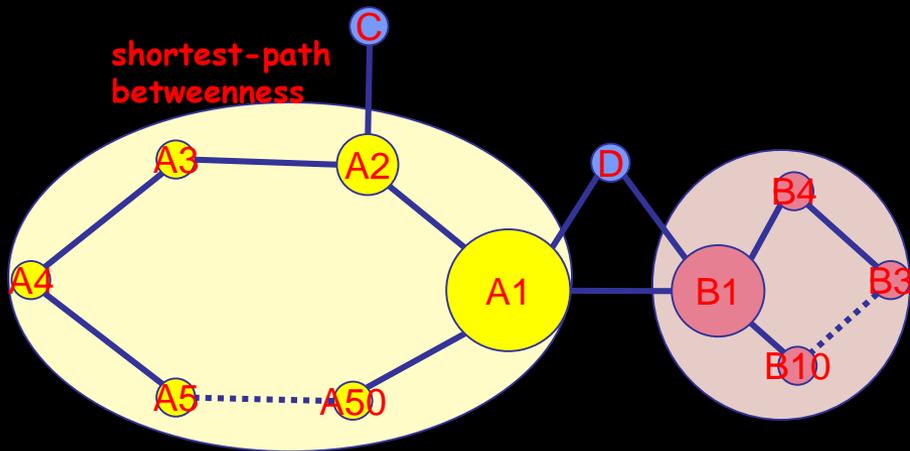
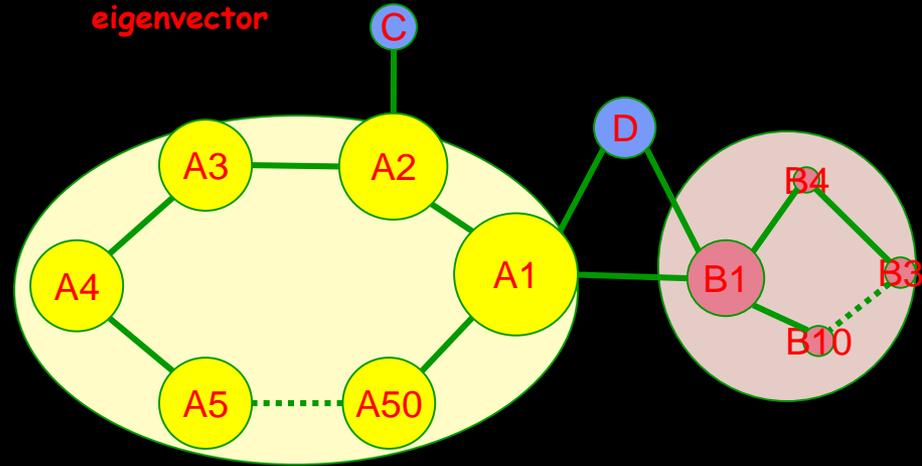
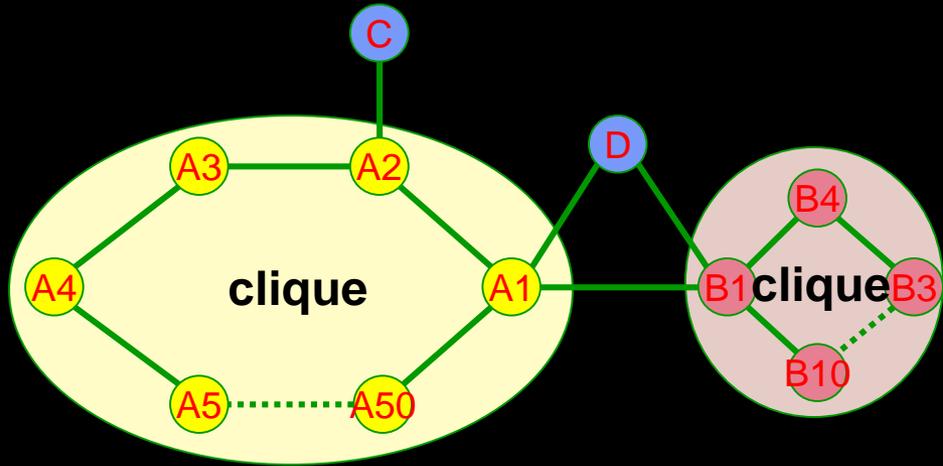
- **The Centrality Hypothesis:** If removal of a node disrupts the “communication” between pairs of other nodes in the network, then the corresponding protein is likely to be essential (Jeong et al., Nature 2001)
- **The Essential PPIs Hypothesis:** All interactions are essential with uniform probability. High degree nodes are essential because they participate in many interactions and thus, with high probability, are adjacent to an essential interaction (He et al., PLoS Genetics 2006)
- **Our result:** Neither of the above is true. Alternative view is proposed.

Network Centrality Indices

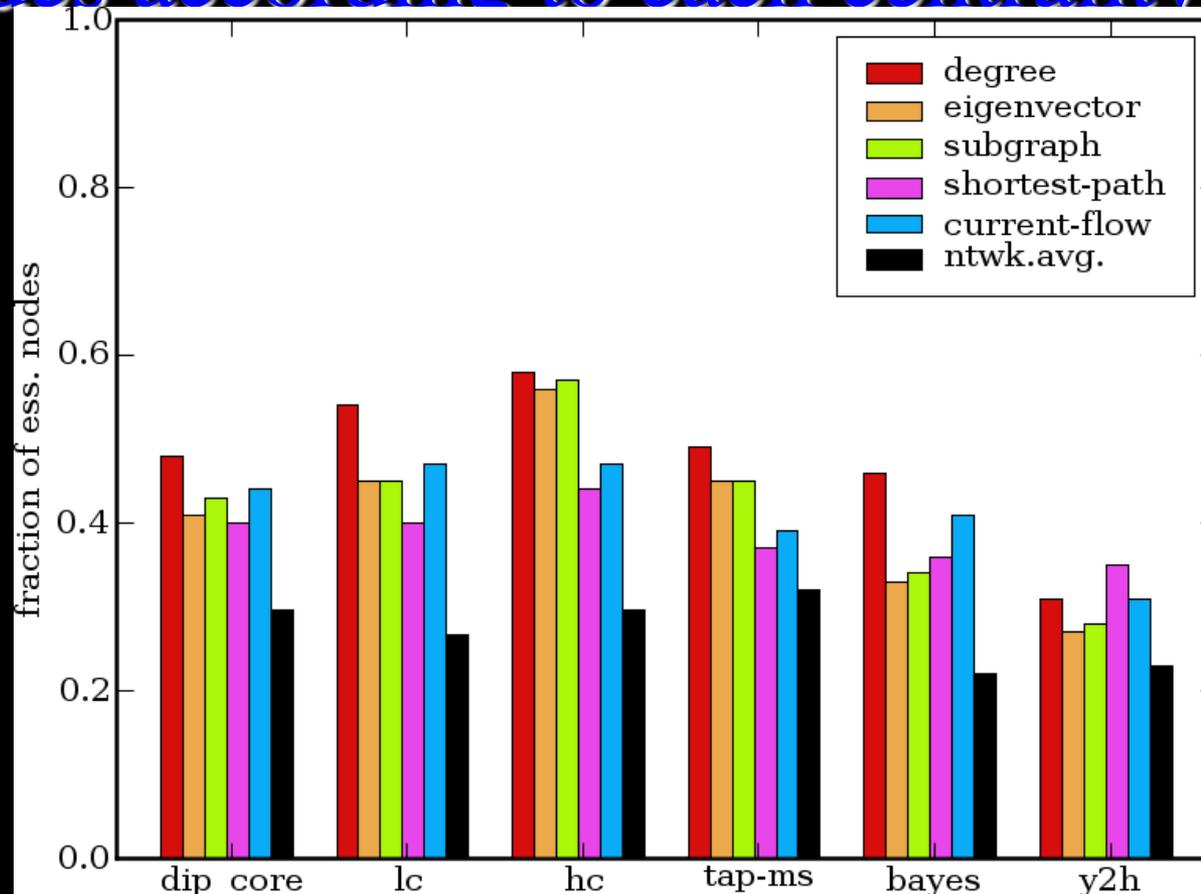
A centrality index assigns a centrality value to every node in the network which quantifies its topological prominence.

- Local indices
 - Degree Centrality (DC)
 - $c(v)$ is the number of neighbors
 - Eigenvector Centrality (EC)
 - $c(v)$ is the weighted number of neighbors
 - Sub-graph Centrality (SC)
 - $c(v)$ is the number of closed walks that start and terminate at v
- Betweenness indices
 - Shortest-Path Betweenness Centrality (SPBC)
 - $c(v)$ is the fraction of shortest paths that pass through v
 - Current-Flow Betweenness Centrality (CFBC)
 - $c(v)$ extends the shortest-path betweenness values by taking into account other paths

Illustration of differences in centrality measures

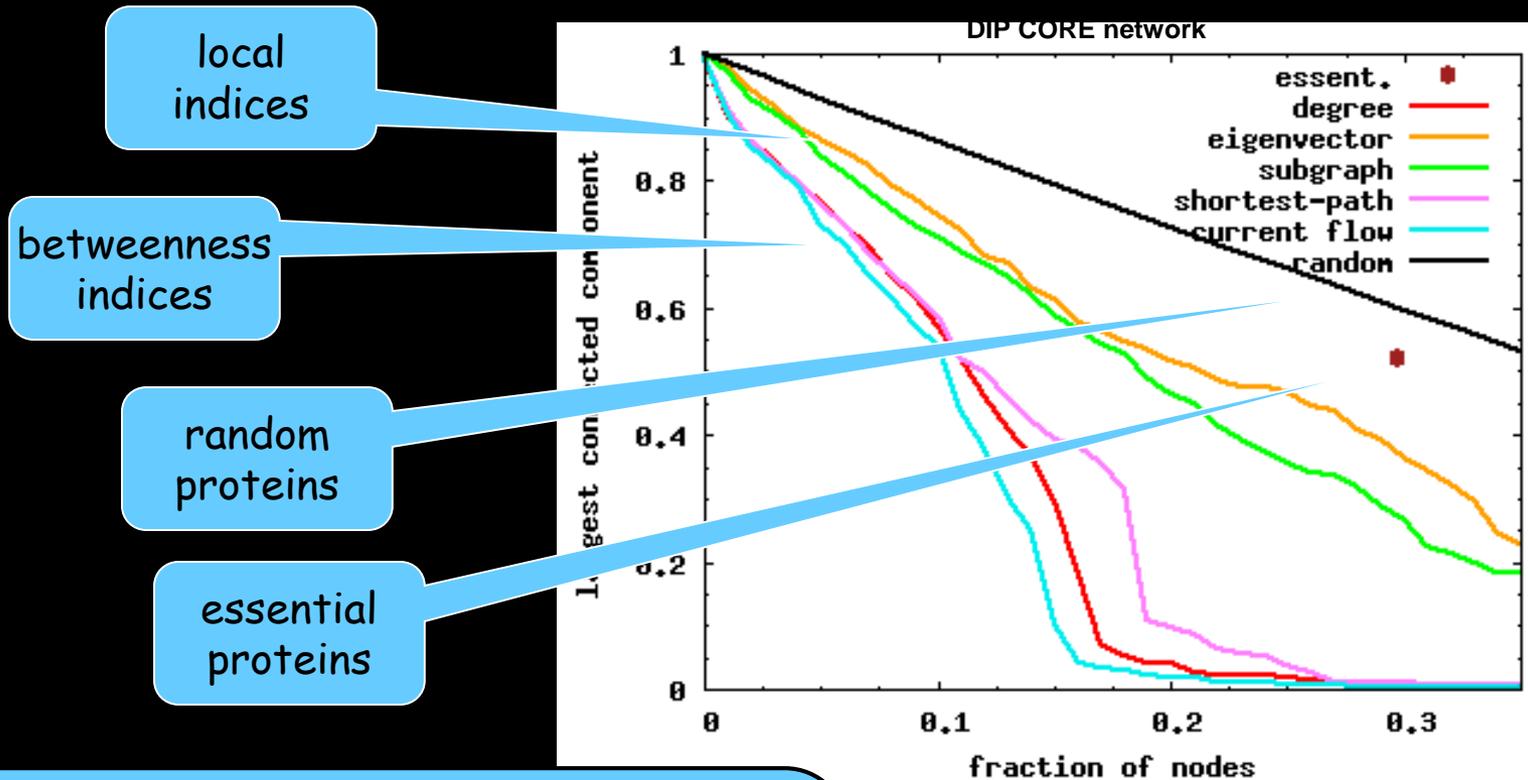


Essential proteins in top 20% of central nodes according to each centrality measure



- None of the centrality indexes is a better predictor of essentiality than degree

How destructive to network integrity is removal of central nodes



Network Integrity Measures

- fraction of nodes in the largest connected component
- increase in the average shortest path
- decrease in the number of edge-disjoint paths

Correlation of betweenness centrality with degree centrality, essentiality, and essentiality corrected for degree centrality

	eigenvector centrality		subgraph centrality	
	τ_{ess}	$\tau_{\text{ess.dc}}$	τ_{ess}	$\tau_{\text{ess.dc}}$
DIP CORE	0.15 (3.5e-19)	0.064 (8.6e-05)	0.17 (1.2e-24)	0.059 (2.5e-04)
LC	0.23 (7.9e-56)	0.094 (3.6e-11)	0.23 (1.2e-55)	0.093 (4.9e-11)
HC	0.24 (1.8e-54)	0.107 (2.9e-12)	0.24 (7.9e-55)	0.102 (3.4e-11)
TAP-MS	0.12 (8.42e-11)	-0.007 (6.5e-01)	0.12 (8.42e-11)	-0.007 (6.5e-01)
BAYESIAN	0.17 (5.7e-39)	0.046 (1.5e-04)	0.17 (5.1e-41)	0.051 (3.1e-05)
Y2H	0.05 (1.1e-01)	0.027 (2.5e-01)	0.03 (2.0e-01)	-0.024 (7.2e-01)
	shortest-path betweenness centrality		current flow betweenness	
	τ_{ess}	$\tau_{\text{ess.dc}}$	τ_{ess}	$\tau_{\text{ess.dc}}$
DIP CORE	0.15 (3.2e-18)	-0.002 (5.5e-01)	0.19 (2.7e-27)	0.012 (2.5e-01)
LC	0.21 (1.4e-46)	0.003 (4.25e-01)	0.26 (3.7e-70)	-0.007 (6.8e-01)
HC	0.20 (1.9e-36)	0.005 (3.7e-01)	0.24 (2.6e-53)	-0.005 (6.2e-01)
TAP-MS	0.12 (3.5e-11)	0.018 (1.8e-01)	0.16 (3.3e-18)	0.017 (1.8e-01)
BAYESIAN	0.18 (2.4e-41)	0.005 (3.43e-01)	0.23 (2.7e-69)	0.018 (8.1e-02)
Y2H	0.10 (1.2e-02)	0.048 (1.4e-01)	0.10 (1.4e-02)	0.041 (1.8e-01)

So far we observed the following:

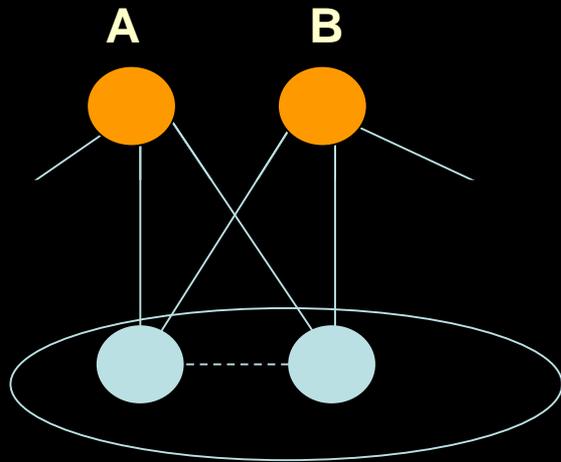
- Local centrality measures matter (degree, sub-graph centrality)
- Correlation of global centrality measures with essentiality is not statistically significant when correcting for correlation with vertex degree

Why are hubs enriched in essential proteins?

- **The Centrality Hypothesis:** If removal of a node disrupts the “communication” between pairs of other nodes in the network, then the corresponding protein is likely to be essential (Jeong et al., Nature 2001)
- **The Essential PPIs Hypothesis:** All interactions are essential with uniform probability. High degree nodes are essential because they participate in many interactions and thus, with high probability, are adjacent to an essential interaction (He et al., PLoS Genetics 2006)
- **Our result:** Neither of the above is true. Alternative view is proposed.

Rejecting Essential PPIs

Hypothesis:



Common neighbors

According to the essential interaction hypothesis, essentiality of A should be independent of essentiality of B.

The independence of such pairs was rejected with high probability

	total number of pairs	number of pairs of the same type	expected number of pairs of the same type		
			simulation	line fitting	weighted line fitting
DIP CORE	1,849	1,135	945 (3.6e-10)	928 (8.6e-12)	938 (8.0e-11)
LC	10,777	6,143	5,691 (6.6e-10)	5,556 (1.1e-15)	5,589 (3.9e-14)
HC	5,907	3,516	3,213 (2.0e-08)	2,997 (2.2e-16)	2,994 (2.2e-16)
Y2H	3,254	2,167	1,976 (9.6e-07)	2,025 (2.6e-04)	2,052 (3.3e-03)

Adding new observation:

- Local centrality measures matter (degree, sub-graph centrality)
- Correlation of global centrality measures with essentiality is not statistically significant when correcting for correlation with vertex degree
- Clustering effect of essential and non-essential nodes

Why are hubs enriched in essential proteins?

- **The Centrality Hypothesis:** If removal of a node disrupts the “communication” between pairs of other nodes in the network, then the corresponding protein is likely to be essential (Jeong et al., Nature 2001)
- **The Essential PPIs Hypothesis:** All interactions are essential with uniform probability. High degree nodes are essential because they participate in many interactions and thus, with high probability, are adjacent to an essential interaction (He et al., PLoS Genetics 2006)
- **Our result:** Neither of the above is true. Alternative view is proposed.

Zotenko, Mestre, O’Leary, Przytycka. PLoS CB 2008
(highlighted in Nature Genetics Rev, Sept 2008)

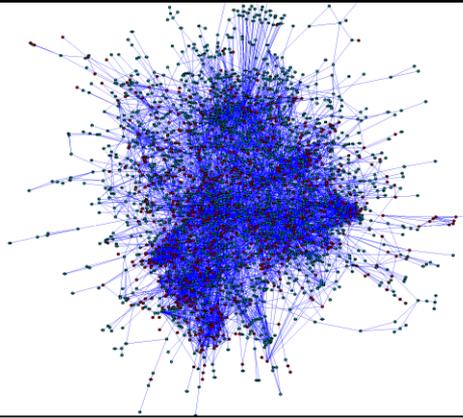
Essentiality of hubs is explained by
membership in
Essential COmplex Biological Modules
(ECOBIMs)

Complex Biological Module (COBIM) is a group of proteins that:

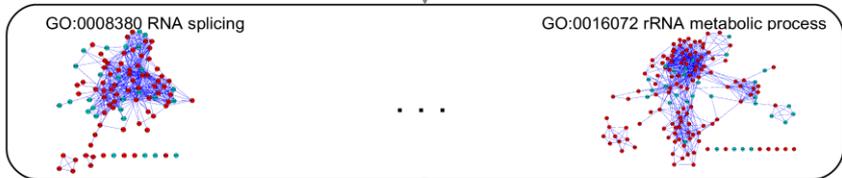
- share a biological function (Biological Module)
- interact extensively with each other (Complex)

COBIMs are clearly partitioned into two classes:

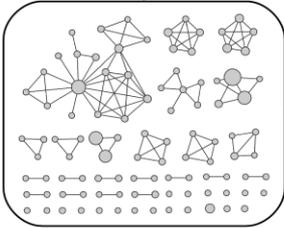
- enriched in essential proteins (ECOBIMs)
- depleted of essential proteins



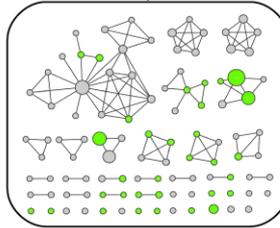
Delineate Biological Processes
Form subnetworks induced by proteins annotated with the same biological process GO term.



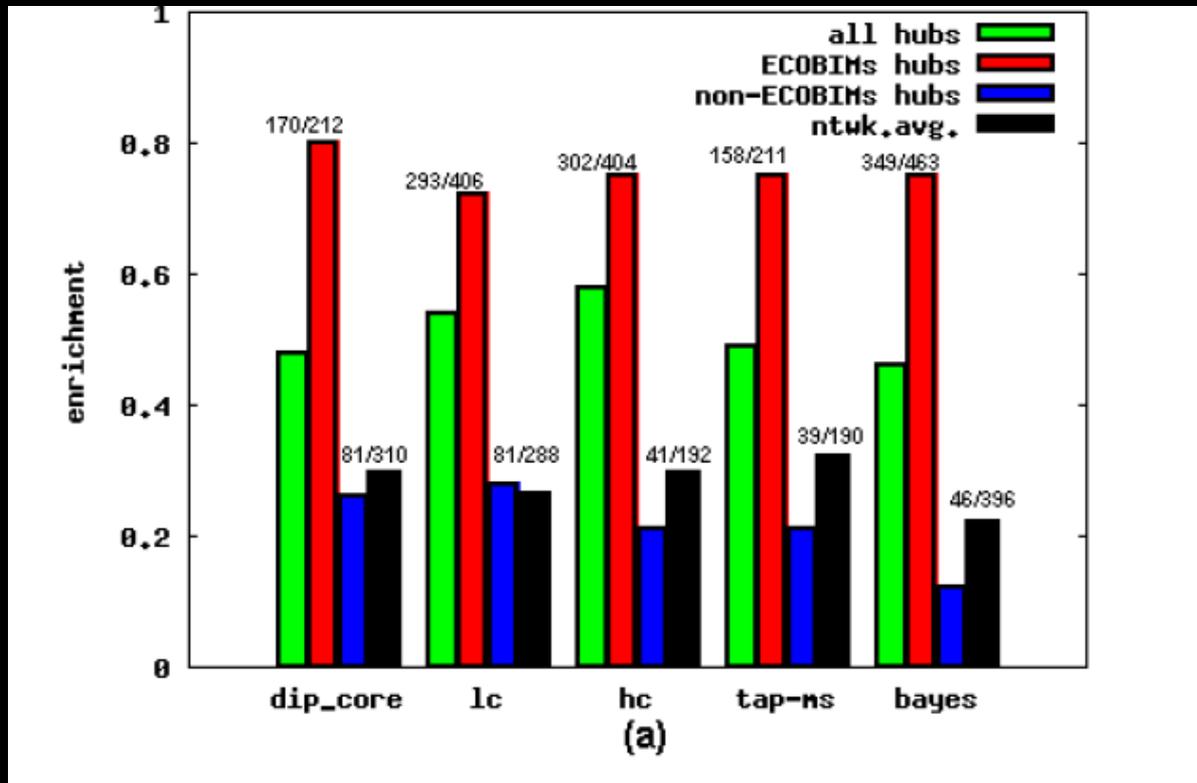
Delineate Complex Biological Modules (COBIMs)
Select groups of proteins that induce densely connected subnetworks: (i) start from seeds of densely connected proteins and (ii) extend seeds through iterative addition of proteins that maintain high density of connections.



Delineate ECOBIMs
Select COBIMs that are enriched in essential proteins, i.e., the fraction of essential proteins in the COBIM is larger than expected.



Non-ECOBIM hubs are depleted in essential proteins.



Enrichment in ECOBIMs and non-ECOBIM COBIM in essential proteins

	enrich. ECOBIM proteins			enrich. non-ECOBIM COBIM proteins		
	obs.	rand.	p-value	obs.	rand.	p-value
DIP CORE	0.77	0.65	<1.0e-05	0.06	0.21	<1.0e-05
LC	0.77	0.65	1.00e-05	0.10	0.17	1.56e-03
HC	0.81	0.68	<1.00e-05	0.12	0.18	2.31e-02
TAP-MS	0.74	0.64	<1.00e-05	0.09	0.17	1.87e-03
BAYESIAN	0.76	0.65	<1.00e-05	0.08	0.18	<1.00e-05
Y2H	0.79	0.63	9.93e-03	0.06	0.17	3.00e-05

Table 9 – Enrichment of ECOBIM and non-ECOBIM COBIM nodes for GO subnetworks in the DIP CORE network

For each GO subnetwork that contributed at least one ECOBIM, the fraction of essential proteins among the subnetwork nodes, subnetwork ECOBIM nodes and subnetwork non-ECOBIM COBIM nodes is shown.

GO term			
	sub-network nodes	ECOBIM nodes	non-ECOBIM COBIM nodes
GO:0016072 rRNA metabolic process	0.83	0.91	n/a
GO:0006352 transcription initiation	0.82	1.00	n/a
GO:0006383 transcription from RNA polymerase III pro	0.77	1.00	0.00
GO:0042254 ribosome biogenesis and assembly	0.72	0.87	n/a
GO:0008380 RNA splicing	0.71	0.79	0.50
GO:0006839 mitochondrial transport	0.64	0.80	n/a
GO:0006360 transcription from RNA polymerase I pro	0.64	0.80	0.00
GO:0016071 mRNA metabolic process	0.63	0.75	0.40
GO:0006260 DNA replication	0.61	0.93	n/a
GO:0031123 RNA 3'-end processing	0.59	0.93	0.29
GO:0006399 tRNA metabolic process	0.50	1.00	0.00
GO:0007059 chromosome segregation	0.49	0.76	n/a
GO:0006944 membrane fusion	0.48	0.75	0.22
GO:0006508 proteolysis	0.46	0.77	n/a
GO:0051169 nuclear transport	0.44	0.80	0.47
GO:0006997 nuclear organization and biogenesis	0.43	1.00	0.33
GO:0000278 mitotic cell cycle	0.43	0.81	0.19
GO:0015931 nucleobase, nucleoside, nucleotide and n	0.42	0.63	n/a
GO:0006913 nucleocytoplasmic transport	0.42	0.80	0.41
GO:0051236 establishment of RNA localization	0.42	0.63	n/a
GO:0006366 transcription from RNA polymerase II pro	0.40	0.75	0.29
GO:0007010 cytoskeleton organization and biogenesis	0.40	0.78	0.00
GO:0048308 organelle inheritance	0.39	0.86	n/a
GO:0006401 RNA catabolic process	0.38	0.83	0.41
GO:0006461 protein complex assembly	0.38	1.00	n/a
GO:0045184 establishment of protein localization	0.37	0.89	0.38
GO:0009100 glycoprotein metabolic process	0.37	0.63	n/a
GO:0006412 translation	0.36	0.85	0.00
GO:0007005 mitochondrion organization and biogenes	0.35	0.91	n/a
GO:0006512 ubiquitin cycle	0.34	0.82	n/a

✂ Author's Choice

A Complex-based Reconstruction of the *Saccharomyces cerevisiae* Interactome*[§]

Haidong Wang‡, Boyko Kakaradov‡§, Sean R. Collins§¶**[¶], Lena Karotki‡‡, Dorothea Fiedler¶¶**[¶], Michael Shales¶¶, Kevan M. Shokat¶¶**[¶], Tobias C. Walther‡‡, Nevan J. Krogan¶¶§§, and Daphne Koller‡¶¶

Molecular and cellular proteomics, 2009

- large complexes are enriched in essential proteins
- the enrichment is increasing with complex size

High-Quality Binary Protein Interaction Map of the Yeast Interactome Network

Haiyuan Yu,^{1,2*} Pascal Braun,^{1,2*} Muhammed A. Yıldırım,^{1,2,3*} Irma Lemmens,⁴
Kavitha Venkatesan,^{1,2} Julie Sahalie,^{1,2} Tomoko Hirozane-Kishikawa,^{1,2} Fana Gebreab,^{1,2}
Na Li,^{1,2} Nicolas Simonis,^{1,2} Tong Hao,^{1,2} Jean-François Rual,^{1,2} Amélie Dricot,^{1,2}
Alexei Vazquez,⁵ Ryan R. Murray,^{1,2} Christophe Simon,^{1,2} Leah Tardivo,^{1,2} Stanley Tam,^{1,2}
Nenad Svrzikapa,^{1,2} Changyu Fan,^{1,2} Anne-Sophie de Smet,⁴ Adriana Motyl,⁶
Michael E. Hudson,⁶ Juyong Park,^{1,7} Xiaofeng Xin,⁸ Michael E. Cusick,^{1,2} Troy Moore,⁹
Charlie Boone,⁸ Michael Snyder,⁶ Frederick P. Roth,^{1,10} Albert-László Barabási,^{1,7}
Jan Tavernier,⁴ David E. Hill,^{1,2} Marc Vidal^{1,2†}

Science 2008

- **No centrality-lethality relation in this larger set**

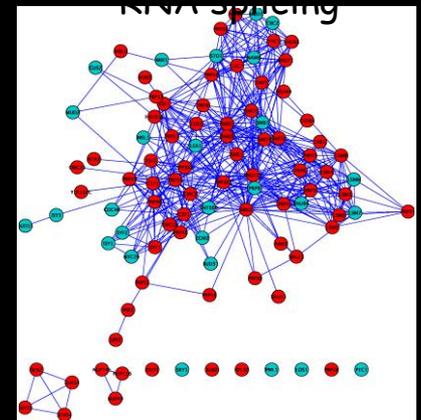
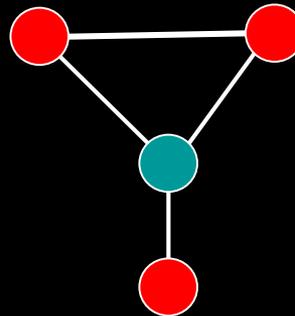
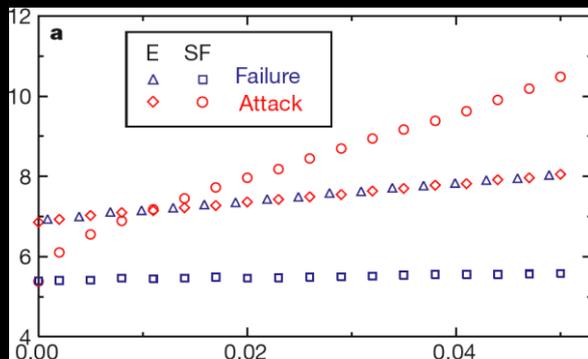
SUMMARY

Why proteins associated with high degree nodes are essential

Jeong et al., Nature 2001
The Betweenness Hypothesis

He et al., PLoS Genetics 2006
The Essentiality Hypothesis

Our explanation
The ECOBIMs Hypothesis



Acknowledgments

Elena Zotenko (currently Max Planck) – poster L11)
Dianne O’Leary, UMD
Julian Mestre (currently Max Planck)

Eigen vector centrality

The computation of the eigenvector centrality values can be cast as an iterative process:

(i) start with an initial vector of centrality scores $\vec{x}_0 = (x_0 \dots x_n)$ (ii) in iteration $k + 1$

update the centrality score of a node i using the scores of its neighbors from the

previous iteration: $x_i^{k+1} = \sum_j x_j^k$, and then normalize the scores $\vec{x}_{k+1} = \vec{x}_{k+1} / |\vec{x}_{k+1}|$. It can

be shown that this process converges to the eigenvector that corresponds to the largest

eigenvalue of the adjacency matrix of the network.



Subgraph centrality

The subgraph centrality value of a node is equal to the number of closed walks that start and terminate at the node. As there is an infinite number of such walks, to obtain finite index values the number of closed walks of length k is weighted by $1/k!$. Therefore, short walks dominate the subgraph centrality values.

