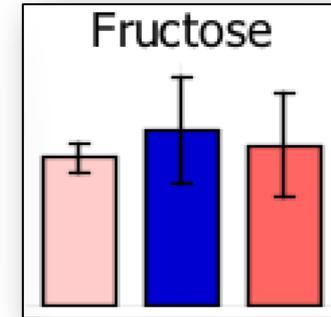


Graph Algorithms and Graph Measures for the Life Sciences

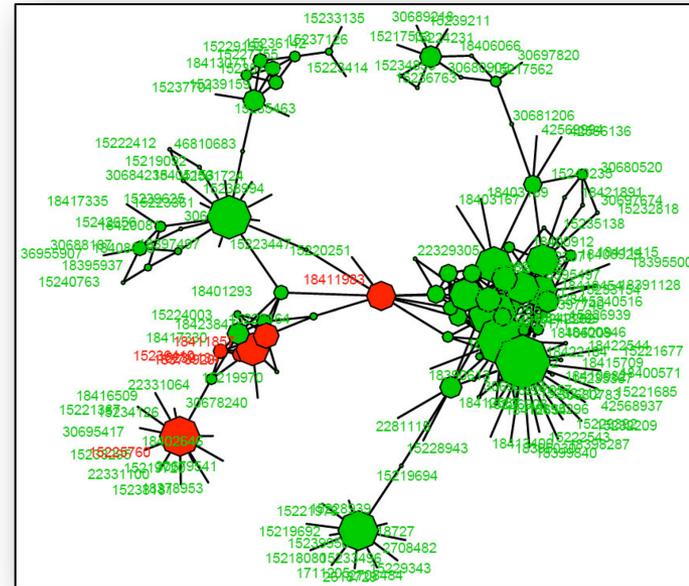
Falk Schreiber

Networks and Graphs in the Life Sciences

▶ Graph



▶ Network



Network Representation

- ▶ **Network** is an informal description for a set of elements with connections or interactions between them and data attached to them
- ▶ **Graph** is a formal description, it is a mathematical object consisting of vertices and edges representing elements and connections, respectively

Interactions → Networks → Pathways

- ▶ A collection of interactions and/or transformations defines a network
- ▶ Pathways are subsets of networks
 - ▶ All pathways are networks, however not all networks are pathways
 - ▶ Difference: level of annotation/understanding
 - ▶ We can define a pathway as a biological network that relates to a known physiological process or phenotype
- ▶ There is no precise biological definition of a pathway
 - ▶ Partitioning of networks into pathways is somewhat arbitrary

Networks a Decade Ago

SPECIAL SECTION

NETWORKS IN BIOLOGY

VIEWPOINT

Biological Networks: The Tinkerer as an Engineer

U. Alon

This viewpoint comments on recent advances in understanding the design principles of biological networks. It highlights the surprising discovery of "good-engineering" principles in biochemical circuitry that evolved by random tinkering.

François Jacob pictured evolution as a tinkerer, not an engineer (1). Engineers and tinkers arrive at their solutions by very different routes. Rather than planning structures in advance and drawing up blueprints (as an engi-

networks. Here are three of the most important shared principles, modularity, robustness to component tolerances, and use of recurring circuit elements.

The first principle, modularity (10–12), is

connection can be added that reduces modularity and increases the fitness of the network. This is the reason that NNs almost always display a nonmodular design. A clue to the reason that modules evolve in biology can be found in engineering (16). Modules in engineering convey an advantage in situations where the design specifications change from time to time. New devices or software can be

NETWORK BIOLOGY: UNDERSTANDING THE CELL'S FUNCTIONAL ORGANIZATION

Albert-László Barabási* & Zoltán N. Oltvai†

A key aim of postgenomic biomedical research is to systematically catalogue all molecules and how these molecules

NETWORKS IN BIOLOGY

SPECIAL SECTION

INTRODUCTION

Life and the Art of Networks

In the world at large, it is easy to see the impact of networks and the relevance of network theory. Steve Strogatz recently called the massive power failure in the United States the "grid's immune response" and called for the utilization of technology to have parts of the grid talk to each other and make decisions that would benefit the entire network rather than individual power plants (S. Strogatz, "How the blackout came to life," *New York Times*, 25 August 2003, p. A21). The impact of this way of thinking is percolating into fundamental biological research with increasing speed, and this issue of *Science* focuses on how we are making that transition. A Viewpoint by Bray (p. 1864) gives an overview of basic network properties. Alon's Viewpoint (p. 1866) notes the striking similarity of biological and human-made machines, which suggests that engineers and biologists may now find their efforts converging.

Biologists are striving to move beyond a "parts list" to more fully understand the ways in which network components interact with one another to influence complex processes. Thus attention has turned to the analysis of networks that operate at many levels. At the scale of networks of interacting proteins that govern cellular function, the flagellated bacterium *Caulobacter crescentus* has been a model system for cell cycle regulation for at least 25 years. McAdams and Shapiro (p. 1874) review the spatial and temporal



Network Motifs: Simple Building Blocks of Complex Networks

Milo,¹ S. Shen-Orr,¹ S. Itzkovitz,¹ N. Kashtan,¹ D. Chklovskii,² U. Alon^{1*}

REVIEWS

Complex networks are studied across many fields of science. To uncover their structural design principles, we defined "network motifs," patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks. We found such motifs in networks in biochemistry, neurobiology, ecology, and engineering. The motifs shared ecological food webs were distinct from the motifs shared by the genetic networks of *Escherichia coli* and *Saccharomyces cerevisiae* or from those found in the World Wide Web. Similar motifs were found in networks that perform information processing, even though they describe elements as different as molecules within a cell and synaptic connections between neurons in *Caenorhabditis elegans*. Motifs may thus define universal classes of networks. This approach may uncover the basic building blocks of most networks.

the complex networks that occur in nature have been shown to share global statistical properties (1–10). These include the "small-world" property (1–9), of short paths between nodes of networks, termed "scale-free networks" (4, 6), the fraction of nodes having k edges, $p(k)$, decays as a power law $p(k) \sim k^{-\gamma}$ (where γ is often between 2 and 3). To go beyond these

The large-scale organization of metabolic networks

H. Jeong*, B. Tombor†, R. Albert*, Z. N. Oltvai† & A.-L. Barabási*

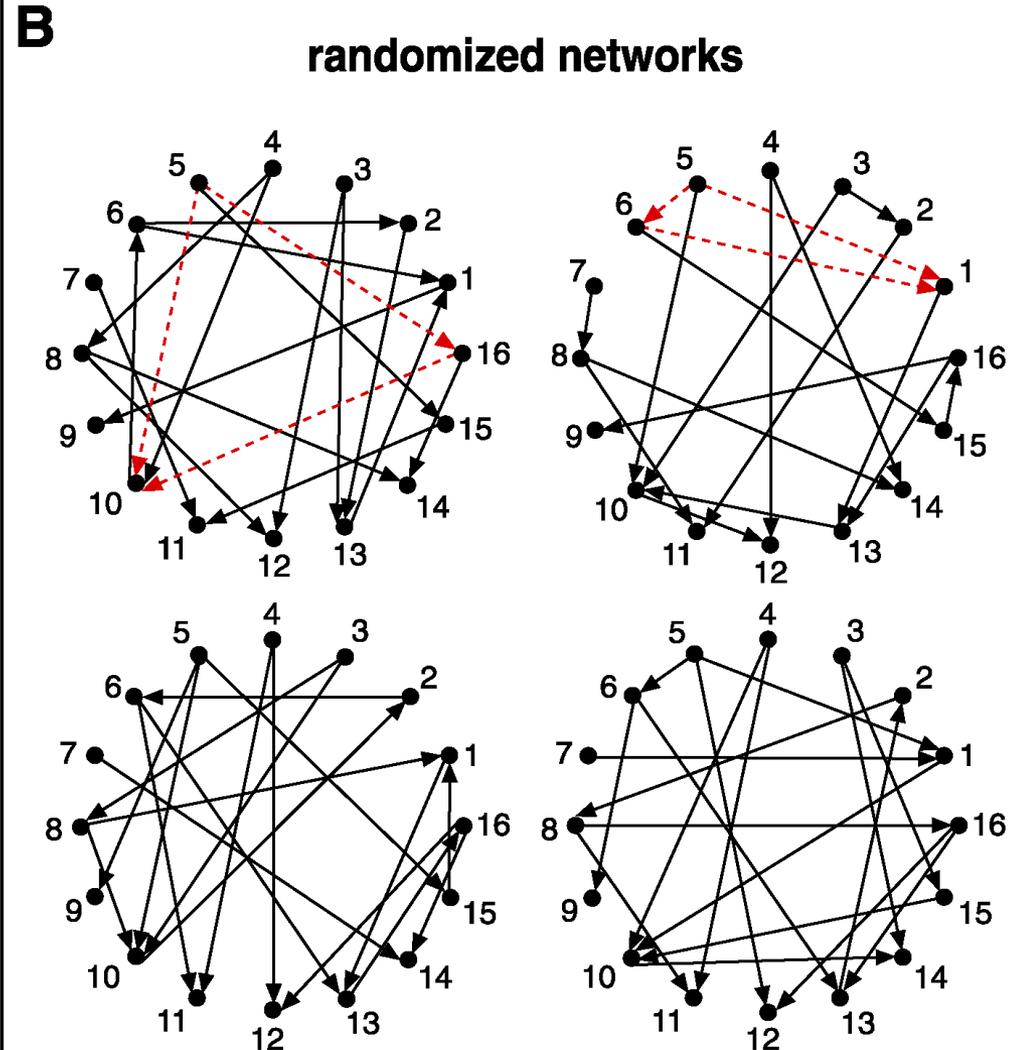
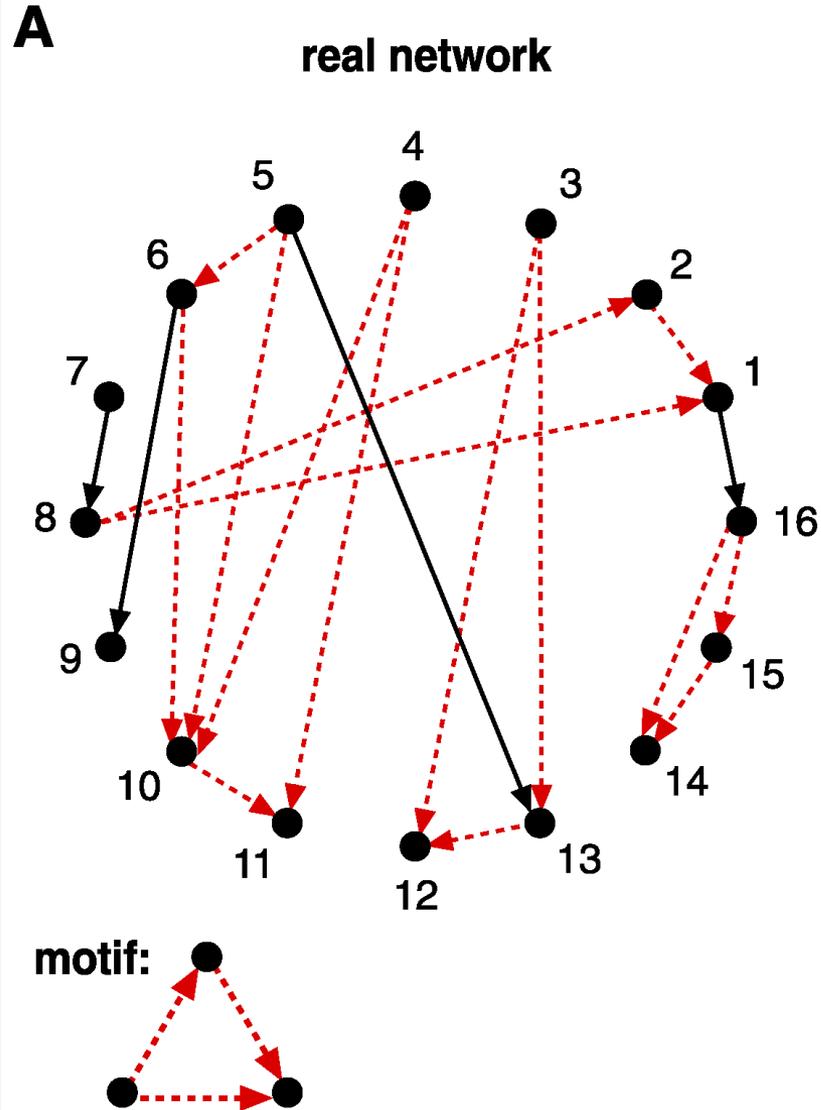
* Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA

† Department of Pathology, Northwestern University Medical School, Chicago, Illinois 60611, USA

In a cell or microorganism, the processes that generate mass, energy, information transfer and cell-fate specification are seamlessly integrated through a complex network of cellular constituents and reactions¹. However, despite the key role of these networks in sustaining cellular functions, their large-scale structure is essentially unknown. Here we present a systematic comparative mathematical analysis of the metabolic networks of

understanding particular to do this, we are constructing network models of interaction to a gene

Can you Spot the Error?



Retraction and Impact Factor

INFECTION AND IMMUNITY, Oct. 2011, p. 3855–3859

0019-9567/11/\$12.00 doi:10.1128/IAI.05661-11

Copyright © 2011, American Society for Microbiology. All Rights Reserved.

Vol. 79, No. 10

EDITORIAL

Retracted Science and the Retraction Index[▽]

Articles may be retracted when their findings are no longer considered valid, they represent misconduct or error, they plagiarize previously published work, or they are otherwise flawed. Using a novel measure that we call the “retraction index,” we found that the frequency of retracted articles among journals and shows a strong correlation with the journal impact factor. Although retracted articles are relatively rare, the retraction process is essential for correcting the literature and maintaining the scientific process.

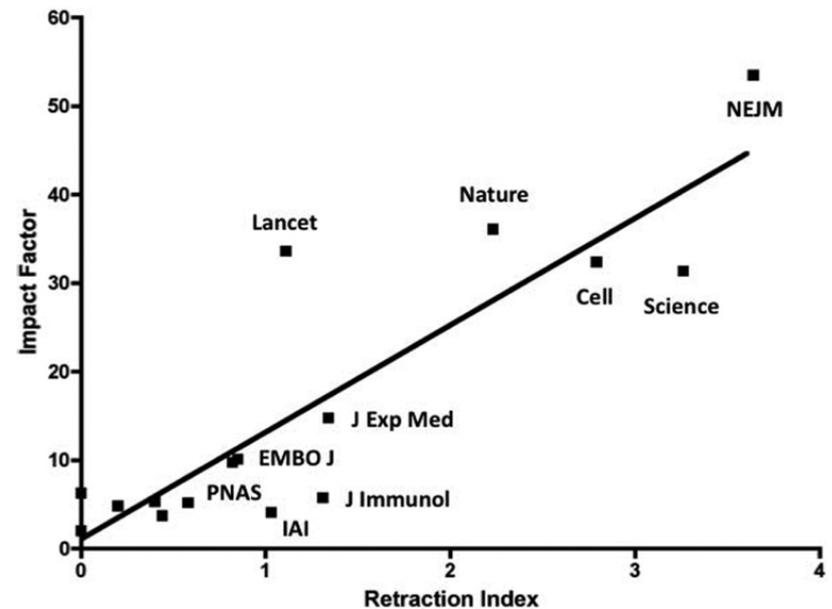


FIG. 1. Correlation between impact factor and retraction index. The 2010 journal impact factor (37) is plotted against the retraction index as a measure of the frequency of retracted articles from 2001 to 2010 (see text for details). Journals analyzed were *Cell*, *EMBO Journal*, *FEMS Microbiology Letters*, *Infection and Immunity*, *Journal of Bacteriology*, *Journal of Biological Chemistry*, *Journal of Experimental Medicine*, *Journal of Immunology*, *Journal of Infectious Diseases*, *Journal of Virology*, *Lancet*, *Microbial Pathogenesis*, *Molecular Microbiology*, *Nature*, *New England Journal of Medicine*, *PNAS*, and *Science*.

The web of human sexual contacts

Promiscuous individuals are the vulnerable nodes to target in safe-sex campaigns.

Unlike clearly defined 'real-world' networks¹, social networks tend to be subjective to some extent^{2,3} because the perception of what constitutes a social link may differ between individuals. One unambiguous type of connection, however, is sexual contact, and here we analyse the sexual behaviour of a random sample of individuals⁴ to reveal the mathematical features of a sexual-contact network. We find that the cumulative distribution of the number of different sexual partners in one year decays as a scale-free power law that has a similar exponent for males and females. The scale-free nature of the web of human sexual contacts indicates that strategic safe-sex campaigns are likely to be the most efficient way to prevent the spread of sexually transmitted diseases.

Many real-world networks¹ typify the 'small-world' phenomenon⁵, so called because of the surprisingly small average path lengths between nodes^{6,7} in the presence of a large degree of clustering^{3,6} (Fig. 1). Small-world networks are classed as single-scale, broad-scale or scale-free, depending on their connectivity distribution, $P(k)$, where k is the number of links

naïres. The response rate was 59%, which corresponds to 2,810 respondents. Two independent analyses of non-response error revealed that elderly people, particularly women, are under-represented in the sample; apart from this skew, the sample is representative in all demographic dimensions.

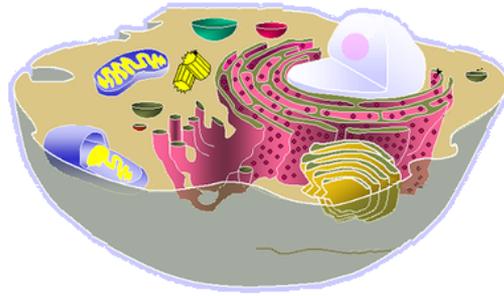
Connections in the network of sexual contacts appear and disappear as sexual relations are initiated and terminated. To investigate the connectivity of this dynamic network, in which links may be short-lived, we first analysed the number, k , of sex partners over a relatively short time period — the 12 months before the survey. Figure 2a shows the cumulative distribution, $P(k)$, for female and male respondents. The data closely follow a straight line in a double-logarithmic plot, which is consistent with a power-law dependence. Males report a larger number of sexual partners than females¹¹, but both show the same scaling properties.

These results contrast with the exponential or gaussian distributions — for which there is a well-defined scale — found for friendship networks⁸. Plausible explanations for the structure of the sexual-contact



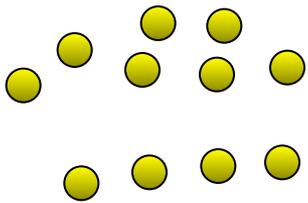
Figure 1 It's a small world: social networks have small average path lengths between connections and show a large degree of clustering. Painting by Idahlia Stanley.

From Biological Building Blocks to Complex Systems

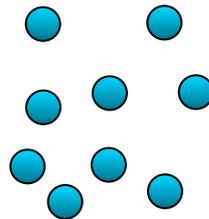


Genome

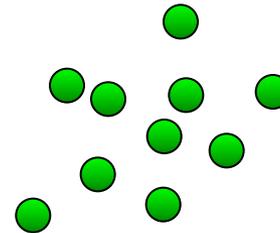
- ▶ Set of hereditary instructions needed to build, run and maintain a particular organism



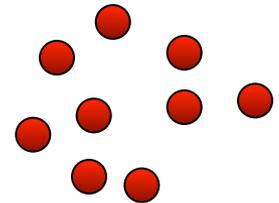
Genes



Transcripts

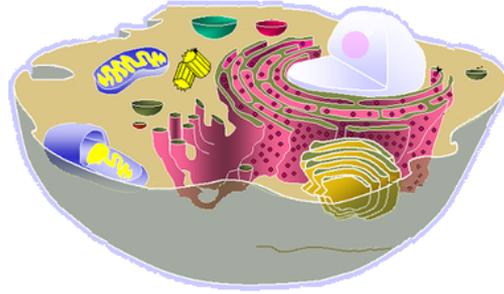


Proteins



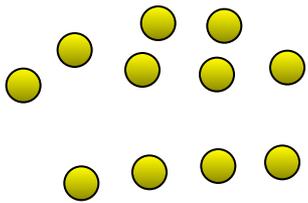
Metabolites

From Biological Building Blocks to Complex Systems

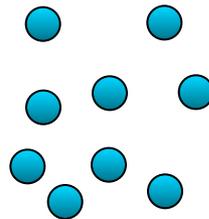


Transcriptome

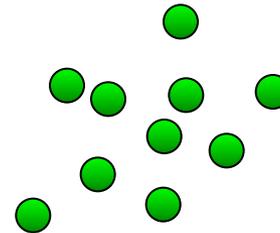
- ▶ Set of RNA transcribed from genes within the genome by a particular cell at a particular time
- ▶ Depends on the tissue, the developmental stage of the organism and the metabolic state of the cell



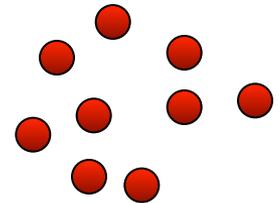
Genes



Transcripts

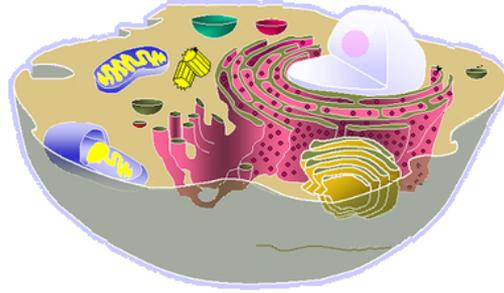


Proteins



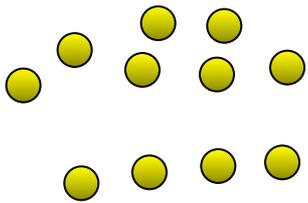
Metabolites

From Biological Building Blocks to Complex Systems

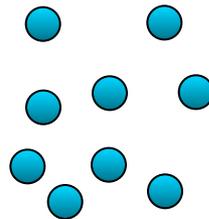


Proteome

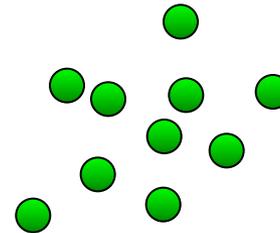
- ▶ Set of proteins translated from RNA within a transcriptome by a particular cell at a particular time
- ▶ *Complete proteome* of a cell: set of all potential proteins that could be synthesised by the cell



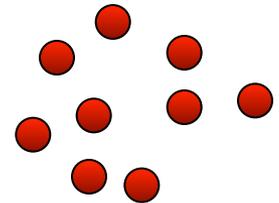
Genes



Transcripts

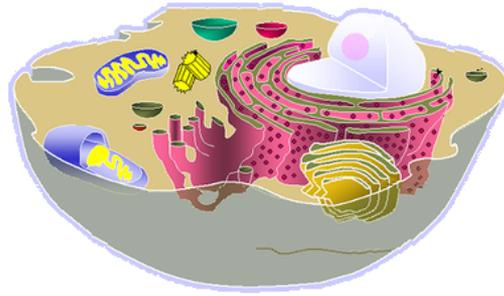


Proteins



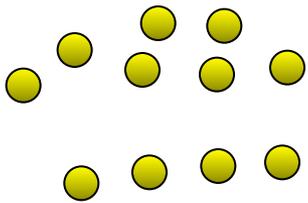
Metabolites

From Biological Building Blocks to Complex Systems

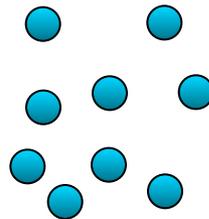


Metabolome

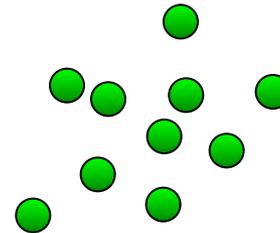
- ▶ Set of all the metabolites inside a particular cell at a particular time



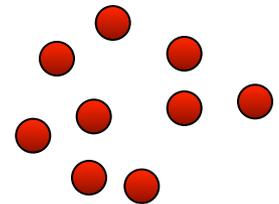
Genes



Transcripts

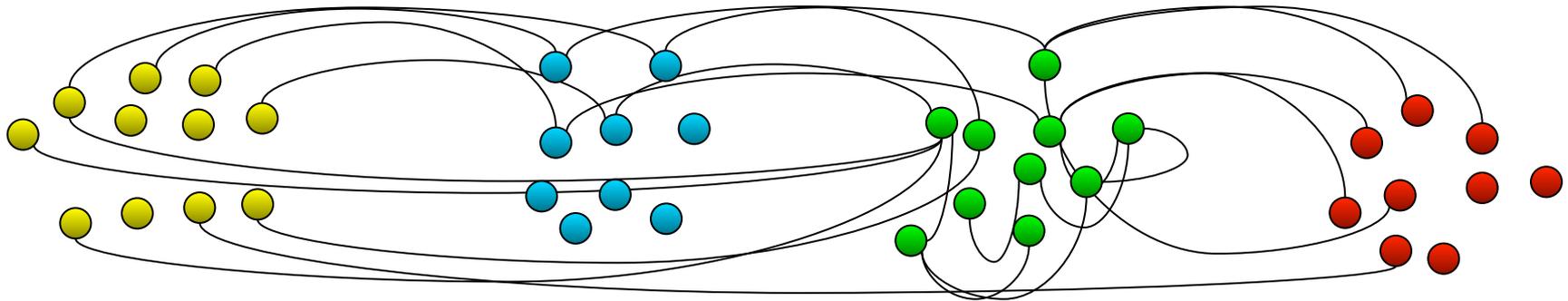
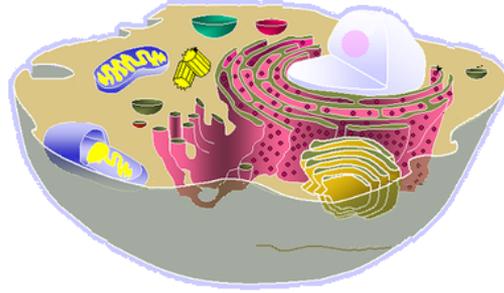


Proteins



Metabolites

From Biological Building Blocks to Complex Systems



Genes

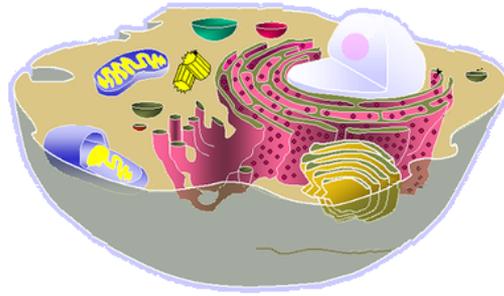
Transcripts

Proteins

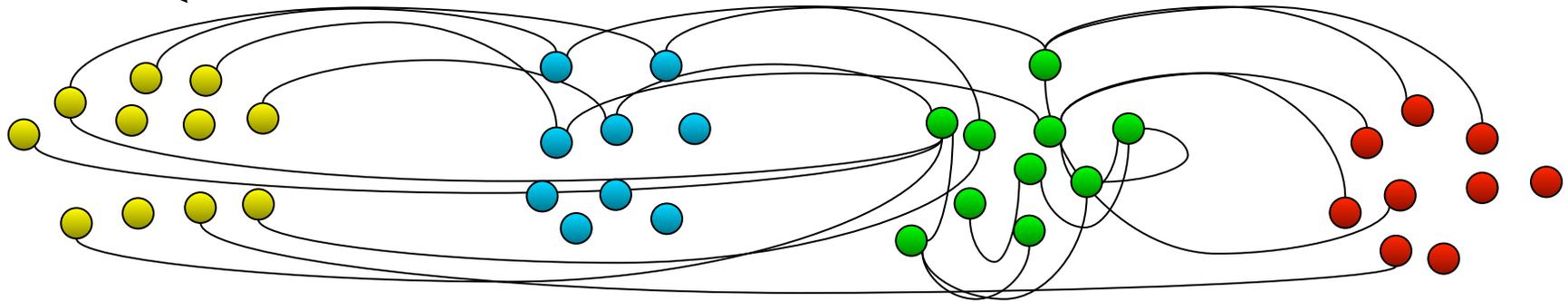
Metabolites

From Biological Building Blocks to Complex Systems

20th century biology
(reductionist approach)



Phenylketonuria
is caused by a mutated gene for the
enzyme phenylalanine hydroxylase (PAH)



Genes

Transcripts

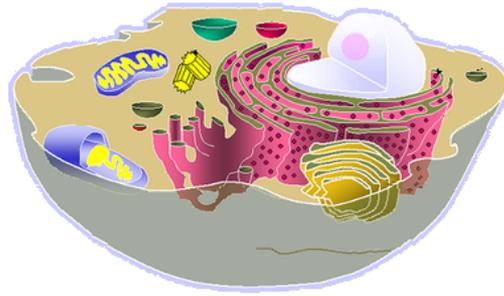
Proteins

Metabolites

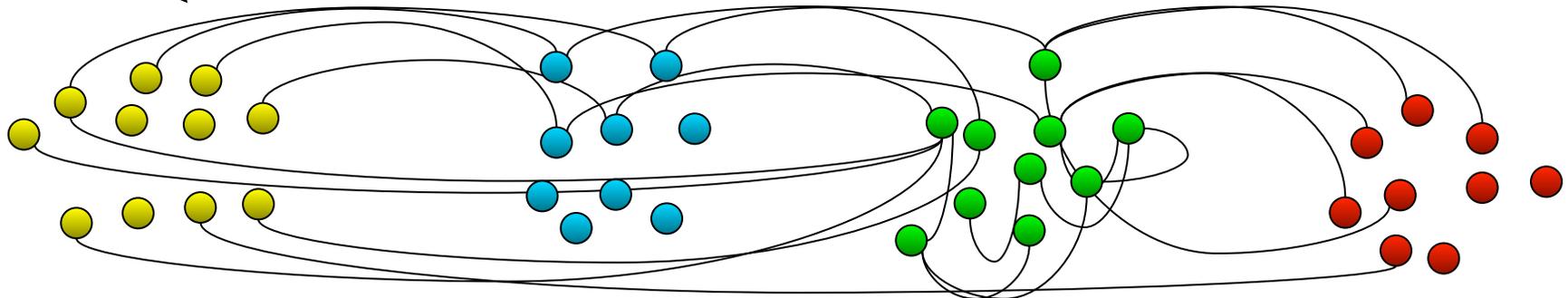
From Biological Building Blocks to Complex Systems

20th century biology
(reductionist approach)

21th century biology
(integrative approach)



Cancer, heart diseases, ...
multiple, complex changes



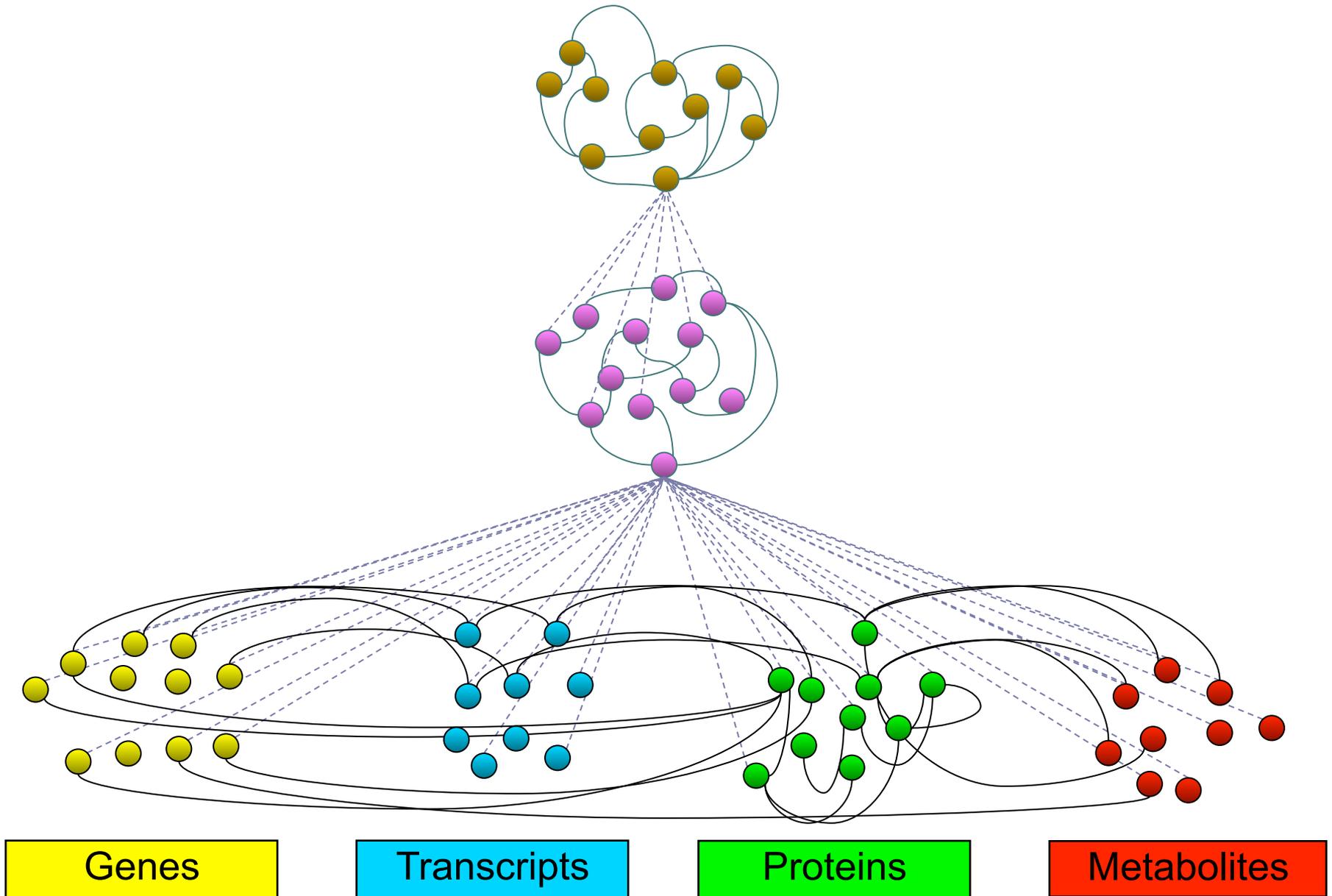
Genes

Transcripts

Proteins

Metabolites

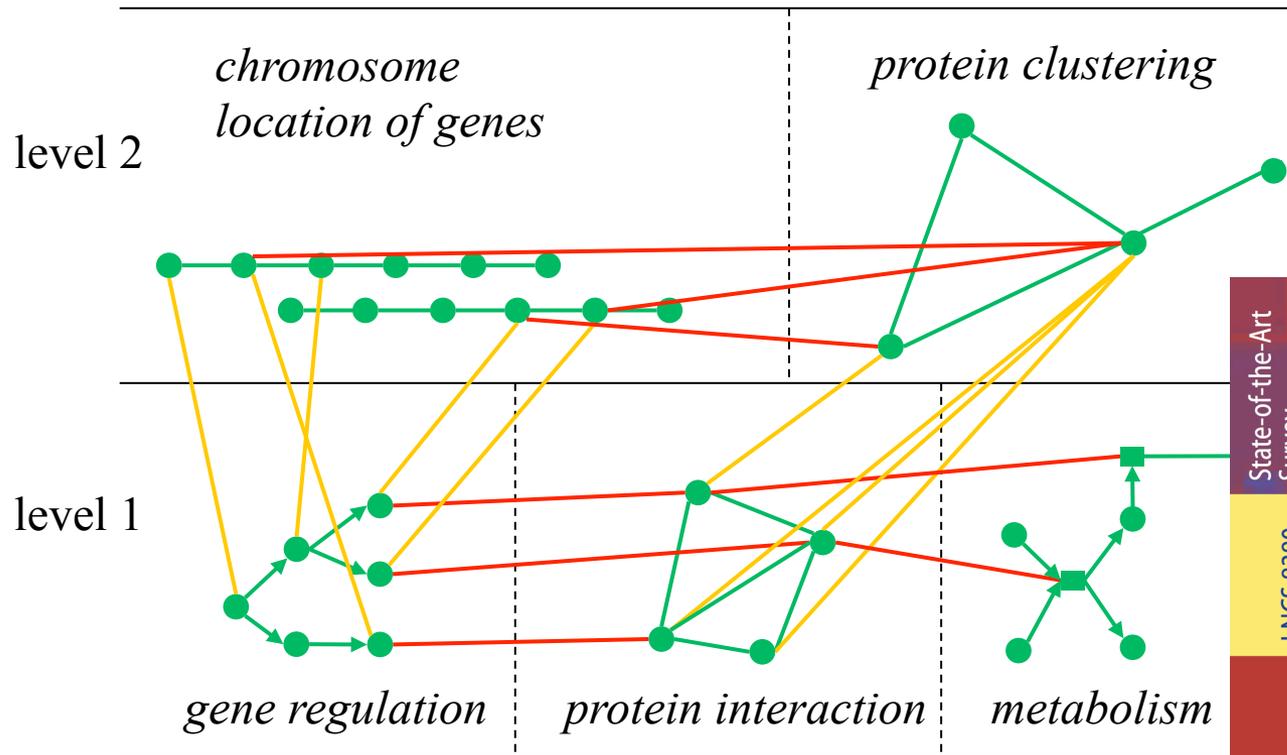
From Biological Building Blocks to Complex Systems



Biological Pathways and Networks - Examples

- ▶ Signal transduction pathway and networks
 - ▶ Cellular processes that recognize extra- or intra-cellular signals and induce appropriate cellular responses
- ▶ Gene regulatory networks
 - ▶ Pathways that regulate a cell's behaviors, including transcription and translation
- ▶ Metabolic pathway
 - ▶ A series of enzymatic reactions that produce a specific product
- ▶ Protein interaction networks
 - ▶ Interaction of proteins (e.g. activation, non-covalent binding)

Biological Pathways and Networks



State-of-the-Art
Survey

LNC8 8380

Andreas Kerren
Helen C. Purchase
Matthew O. Ward (Eds.)

Multivariate Network Visualization

Dagstuhl Seminar #13201
Dagstuhl Castle, Germany, May 12–17, 2013
Revised Discussions

Springer

Many Informatics Areas

- ▶ Evolutionary networks
- ▶ Infection networks
- ▶ Ecological networks / food webs
- ▶ Neuronal networks
- ▶ Hormonal networks
- ▶ Signalling networks
- ▶ Gene regulatory network
- ▶ Protein interaction networks
- ▶ Metabolic networks
- ▶ Chemical structure graphs

**Health informatics/
Environmental informatics**

Medical informatics

Bioinformatics

Chemoinformatics

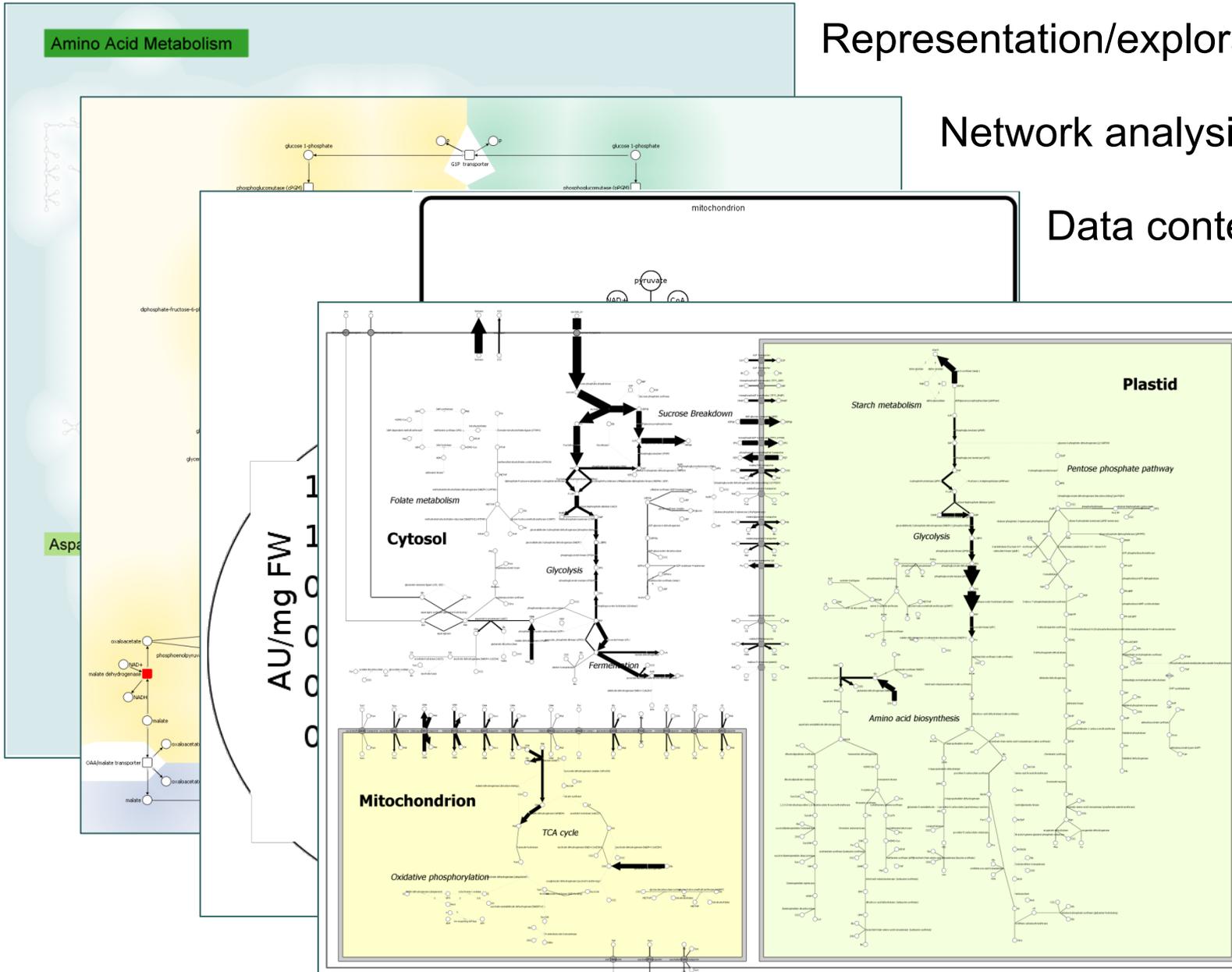
Network Usage - Examples

Representation/exploration

Network analysis

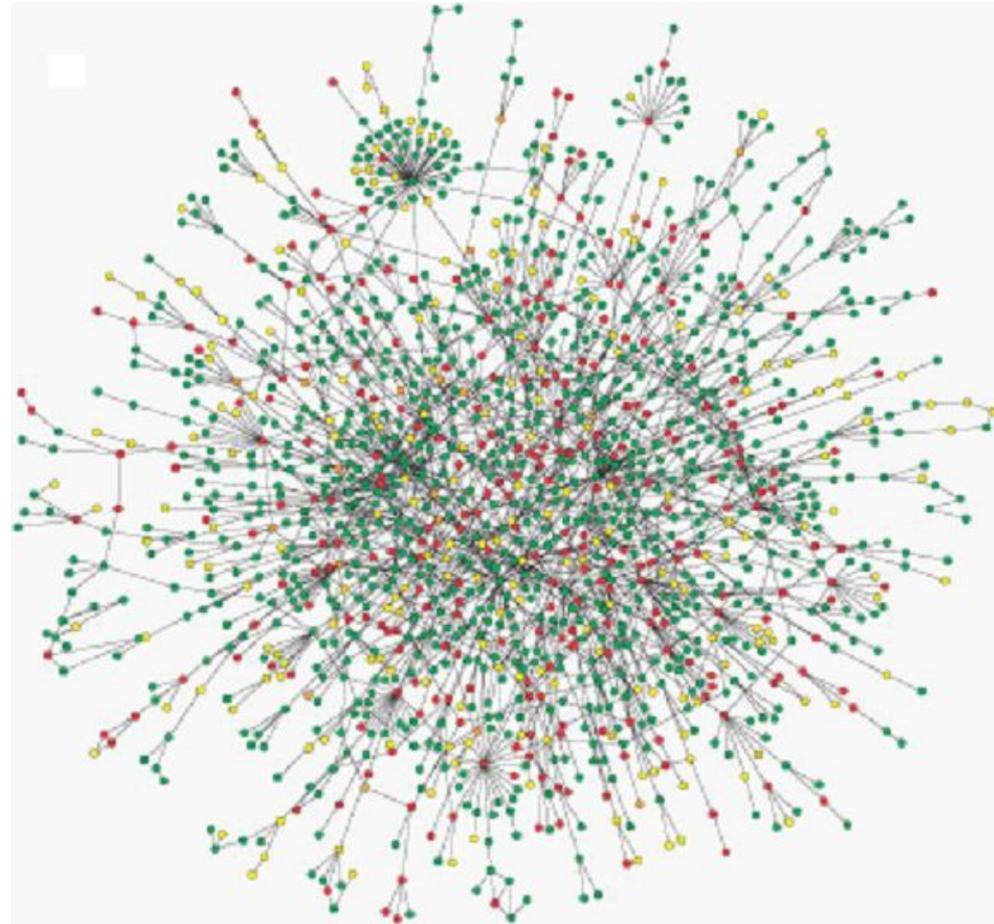
Data context/analysis

Simulation



Network Analysis - Network Centralities

- ▶ Centrality of graph $G=(V,E)$
 - ▶ Funktion $c:V\rightarrow\mathbf{R}$
 - ▶ With $c(u)>c(v)$, if $u\in V$ more important than $v\in V$
- ▶ Ranking of vertices
 - ▶ According to importance
 - ▶ Based on the network structure
- ▶ Application examples
 - ▶ Hypothesis generation for experiments
 - ▶ Which patients should be vaccinated first
- ▶ Problem
 - ▶ Works not well with existing algorithms

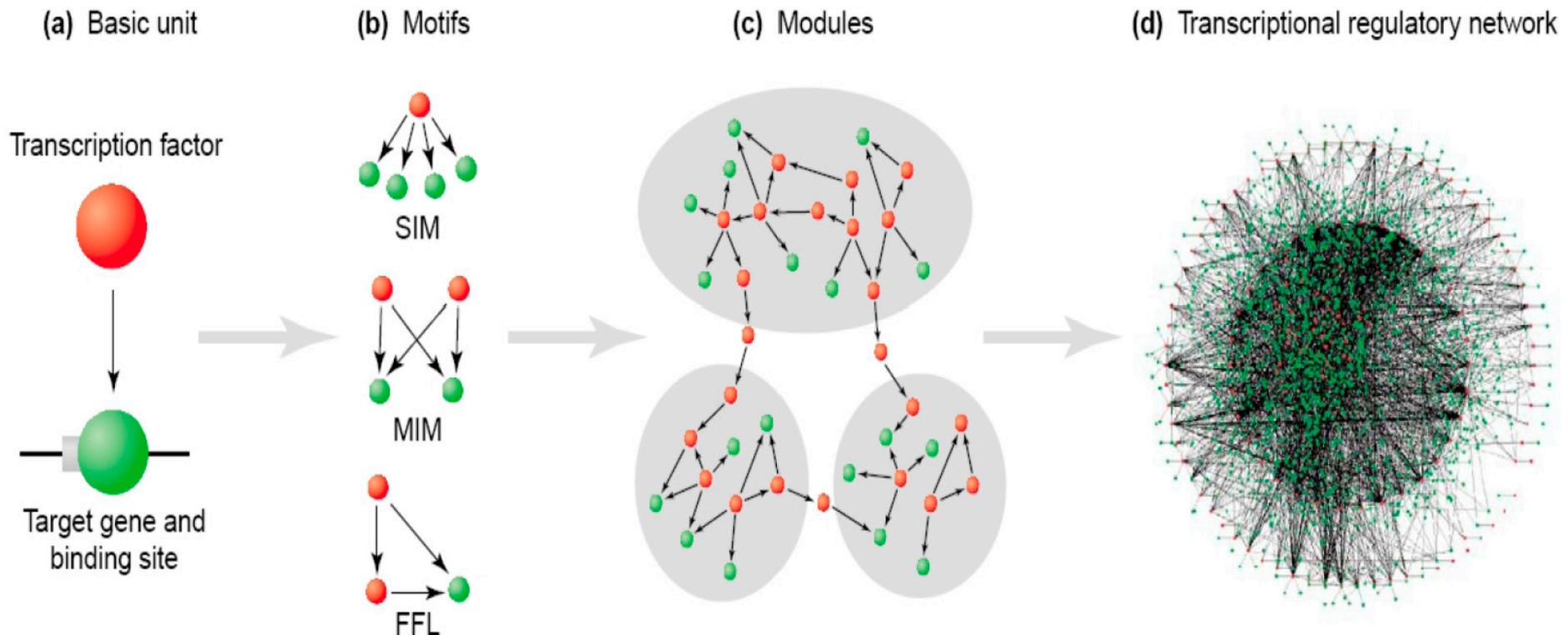


[from Jeong et al., Nature, 2001]

New Centrality Measure

Based on network motifs

- ▶ Sub-graphs representing patterns of local interconnections
- ▶ May represent basic building blocks and design patterns of functional modules

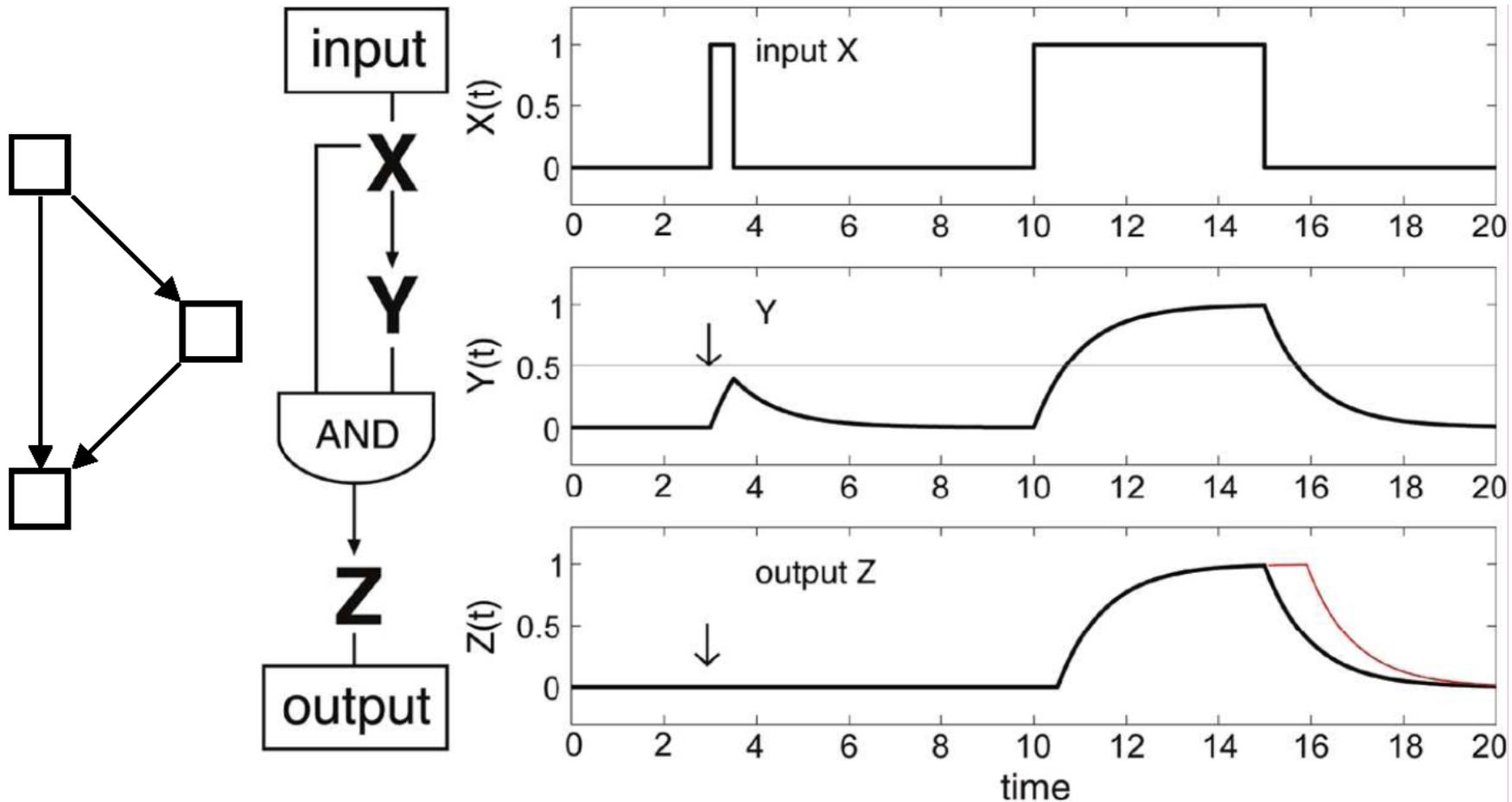


[from Babu et al., Current Opinion in Structural Biology, 2004]

Motifs in Gene Regulatory Networks: Feed-forward Loop

Example of functional properties

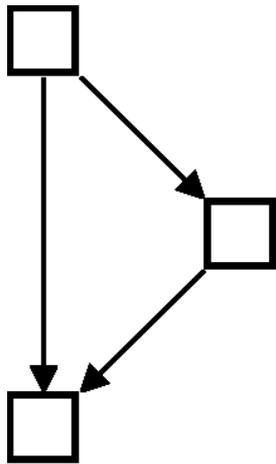
- ▶ Noise filtering: responds only to persistent activations



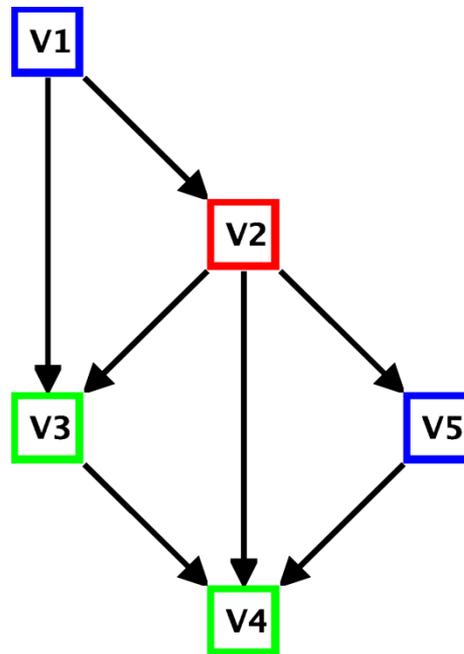
[from Shen-Orr et al., Nature Genetics, 2002]

Motif-based Centrality

- ▶ Combines centrality measures and network motifs
- ▶ Uses occurrences of a motif in the network
- ▶ Incorporates functional substructures into centrality analysis
- ▶ $\mathbf{G}_M = \{G_M | G_M \subseteq G \wedge G_M \approx M\}$
- ▶ $c(v) = |\{G_M | G_M \in \mathbf{G}_M \wedge v \in V(G_M)\}|$



Motif (Feed-forward loop)
M

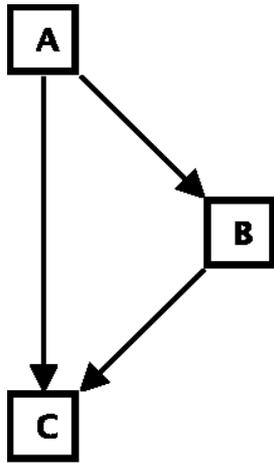


Target graph
G

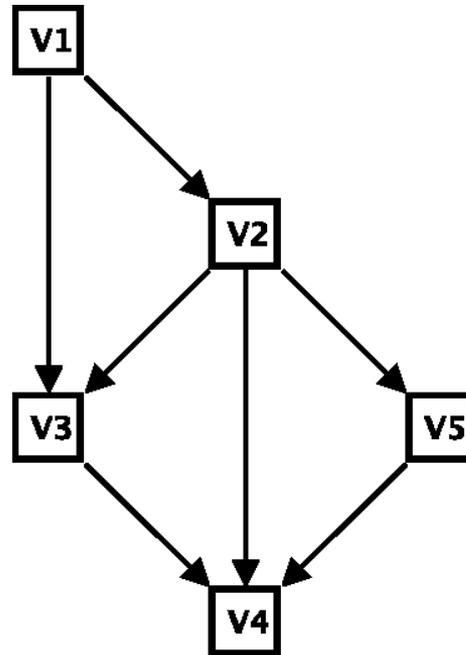
vertex	centrality
v2	3
v3	2
v4	2
v1	1
v5	1

Motif-based Centrality with Roles

- ▶ Different vertices have different roles
- ▶ Count number of matches according to roles
- ▶ $\mathbf{G}_M = \{G_M | G_M \subseteq G \wedge G_M \approx M\}$
- ▶ $c(v, r) = |\{G_M | G_M \in \mathbf{G}_M \wedge v \in V(G_M) \wedge \text{role}(v, G_M) = r\}|$



Motif (Feed-forward loop)
M

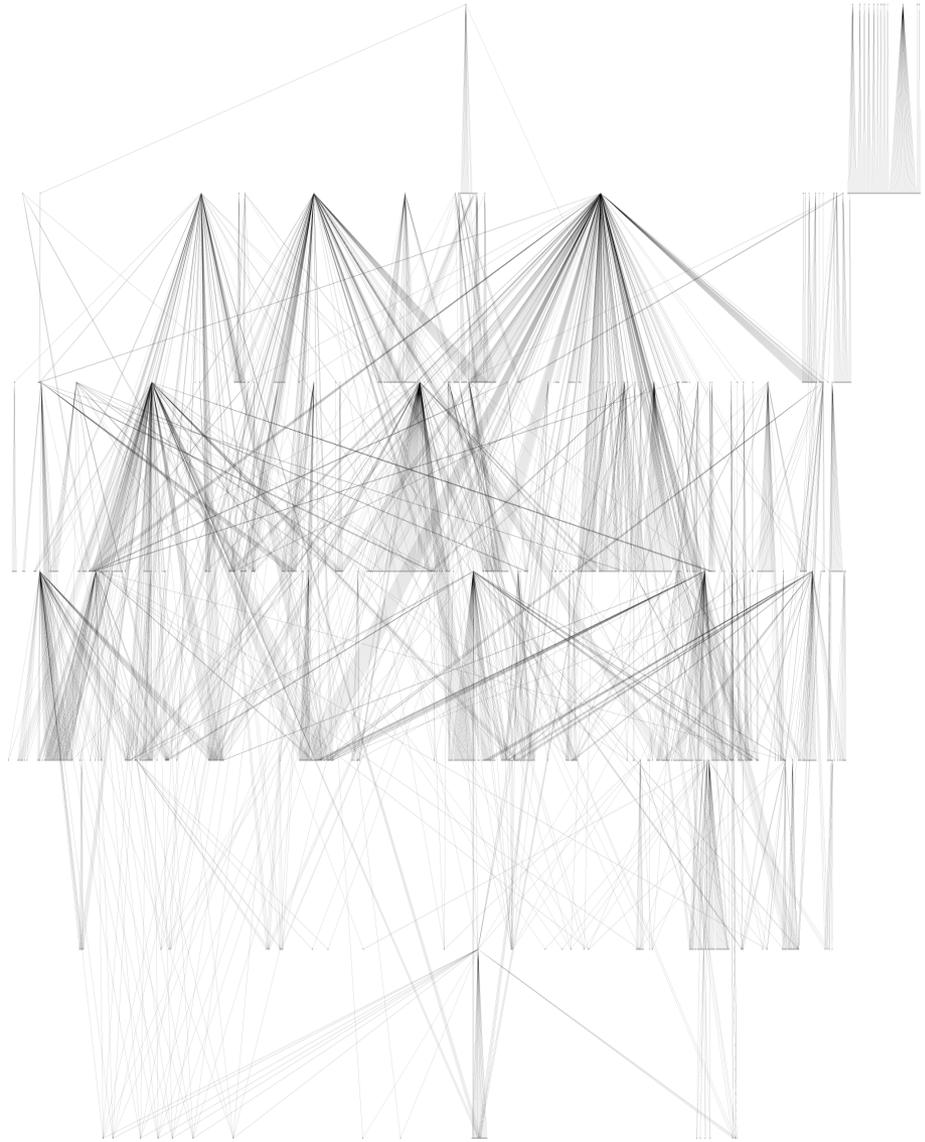


Target graph
G

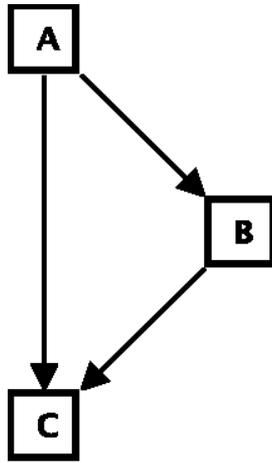
vertex	centrality		
	A	B	C
v1	1	0	0
v2	2	1	0
v3	0	1	1
v4	0	0	2
v5	0	1	0

Gene Regulatory Network of *E. coli*

- ▶ Based on data from RegulonDB (<http://regulondb.ccg.unam.mx/>)
- ▶ 1250 vertices and 2515 edges
- ▶ Global regulators?



Motif-based Centrality with Roles for *E. coli*



- ▶ Top 20 genes (of 1250)
 - ▶ 11 of 18 global regulators (Martínez-Antonio and Collado-Vides)
- ▶ Method works also for other networks
- ▶ Even better results with different motifs

gene	centrality		
	A	B	C
crp	254	0	0
fnr	150	53	0
ihfAB	61	0	0
arcA	58	53	0
fis	40	70	0
modE	18	0	0
soxS	18	1	0
hns	14	39	0
fhIA	11	0	0
gadE	11	0	0
cpxR	11	0	0
rob	10	0	0
galR	8	0	0
gadX	8	26	0
gntR	6	0	0
fur	6	36	1
oxyR	6	1	0
tdcR	6	0	0
srlR	5	11	1
narL	5	95	0

Two Vague Ideas

- ▶ Are scale-free and small-world networks relevant or more an artifact ?

Collective dynamics of 'small-world' networks

Duncan J. Watts* & Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Kimball Hall,
Cornell University, Ithaca, New York 14853, USA

Networks of coupled dynamical systems have been used to model biological oscillators¹⁻⁴, Josephson junction arrays^{5,6}, excitable

The metabolic world of *Escherichia coli* is not small

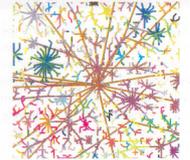
Masanori Arita*

Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency, and Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, 5-1-5 Kashiwanoha, Kashiwa 277-8561, Japan

Edited by Michael S. Waterman, University of Southern California, Los Angeles, CA, and approved December 9, 2003 (received for review October 7, 2003)

To elucidate the organizational and evolutionary principles of the metabolism of living organisms, recent studies have addressed the graph-theoretic analysis of large biochemical networks responsible for the synthesis and degradation of cellular building blocks [Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A. L. (2000) *Nature* 407, 651–654; Wagner, A. & Fell, D. A. (2001) *Proc. R. Soc. London Ser. B* 268, 1803–1810; and Ma, H.-W. & Zeng, A.-P. (2003) *Bioinformatics* 19, 270–277]. In such studies, the global properties of the network are computed by considering enzymatic reactions as links between metabolites. However, the pathways computed in this manner do not conserve their structural moieties

cannot be defined *per se* by compounds or reactions. The biochemical link between metabolites is context-sensitive; it depends on the conserved structural moieties in the adjacent reactions. To accurately compute the reaction connectivity as in the traditional metabolic map, we used digitally compiled atomic mappings, i.e., atomic position pairs between substrates and products corresponding to the substructural moieties conserved in each reaction (Figs. 1 and 2) (10). With this information, we reassessed the global properties of metabolic networks with special emphasis on the small-world hypothesis.



THE INTERNET, mapped on the opposite page, is a scale-free network in that some sites (starbursts and detail above) have a seemingly unlimited number of connections to other sites. This map, made on February 6, 2003, traces the shortest routes from a test Web site to about 100,000 others, using like colors for similar Web addresses.

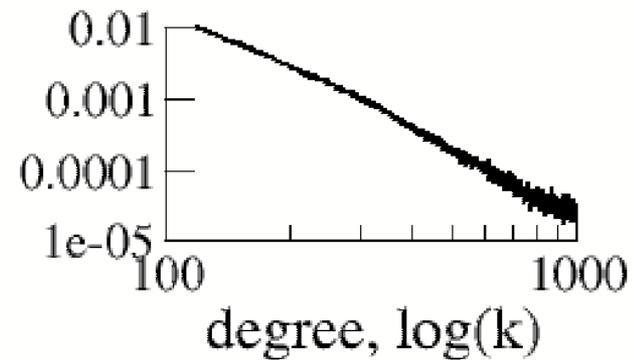
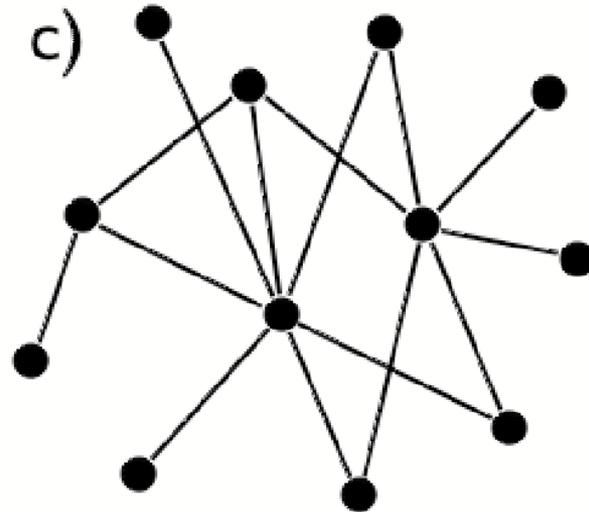
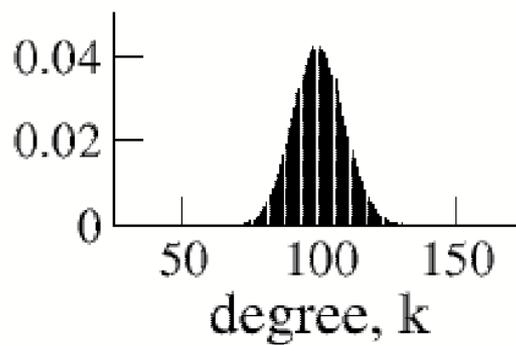
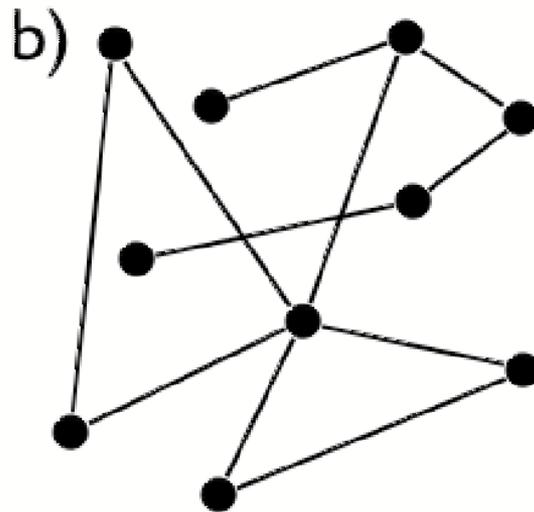
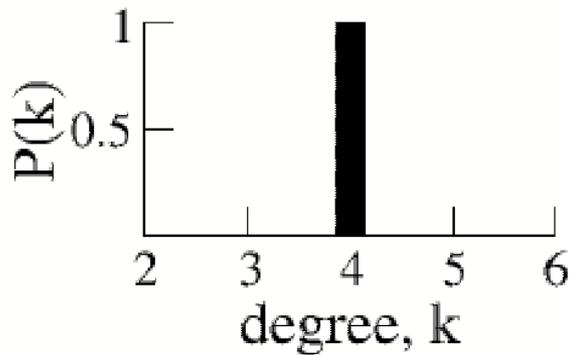
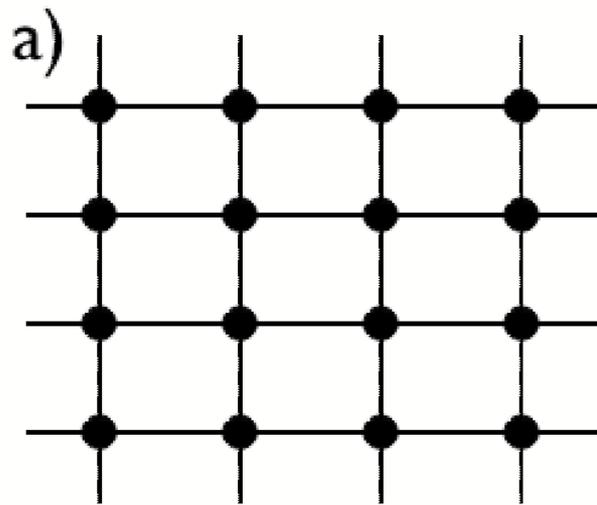
Scale-Free networks

... discovered that various complex systems have architecture governed by shared organizing principles. Important implications for a host of ... from drug development to Internet security

BY ALBERT-LÁSZLÓ BARABÁSI AND ERIC BONABEAU

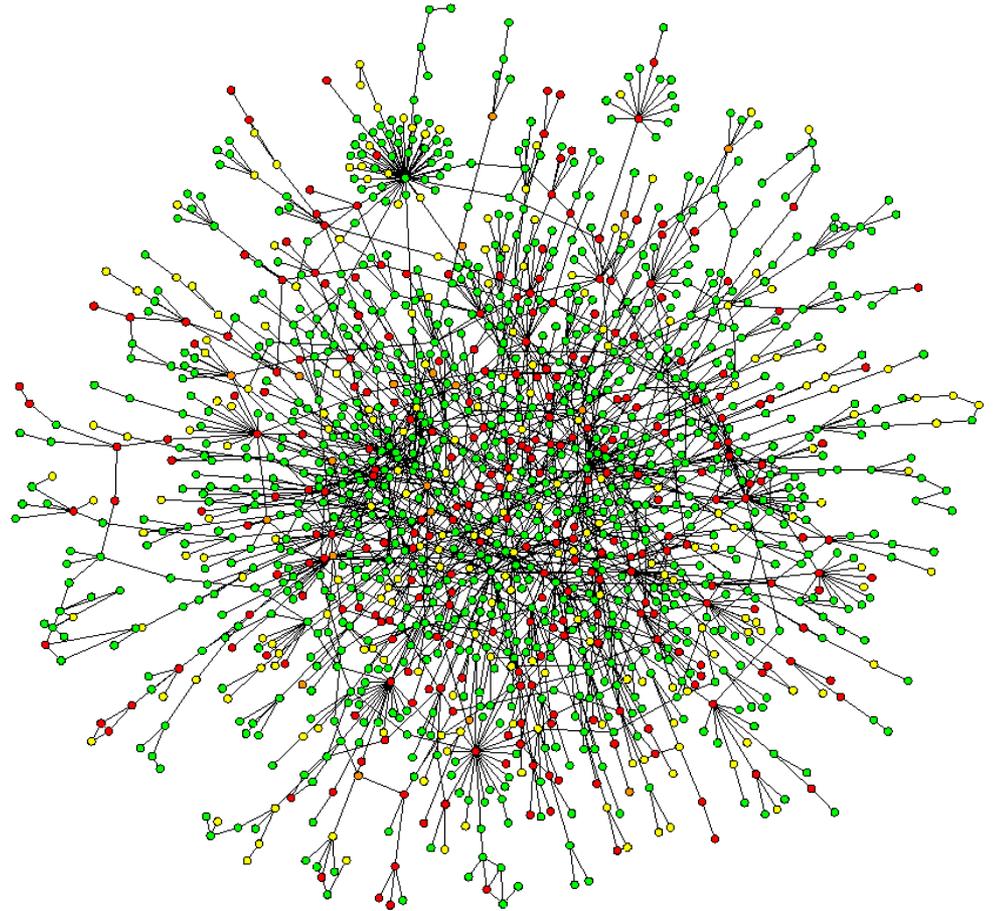
MAY 2003

Degree Distribution - Examples



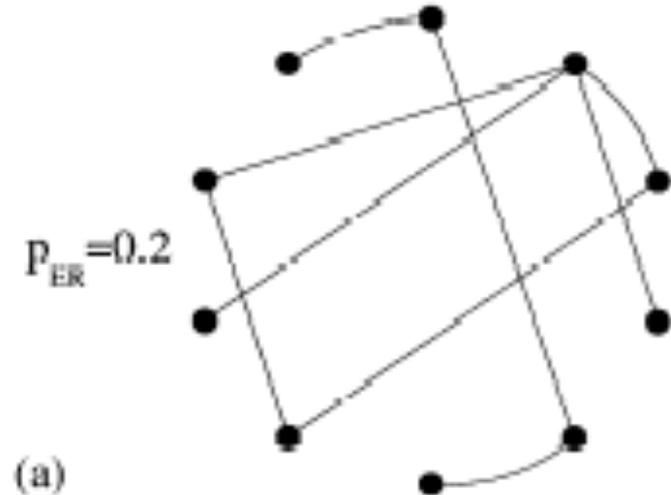
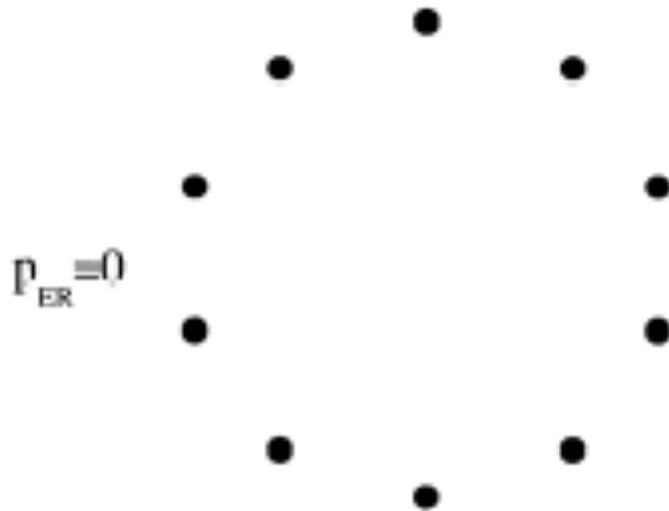
Models for Networks of Complex Topology

- ▶ Erdős-Rényi (1960)
- ▶ Watts-Strogatz (1998)
- ▶ Barabási-Albert (1999)



The Erdős-Rényi [ER] Model (1960)

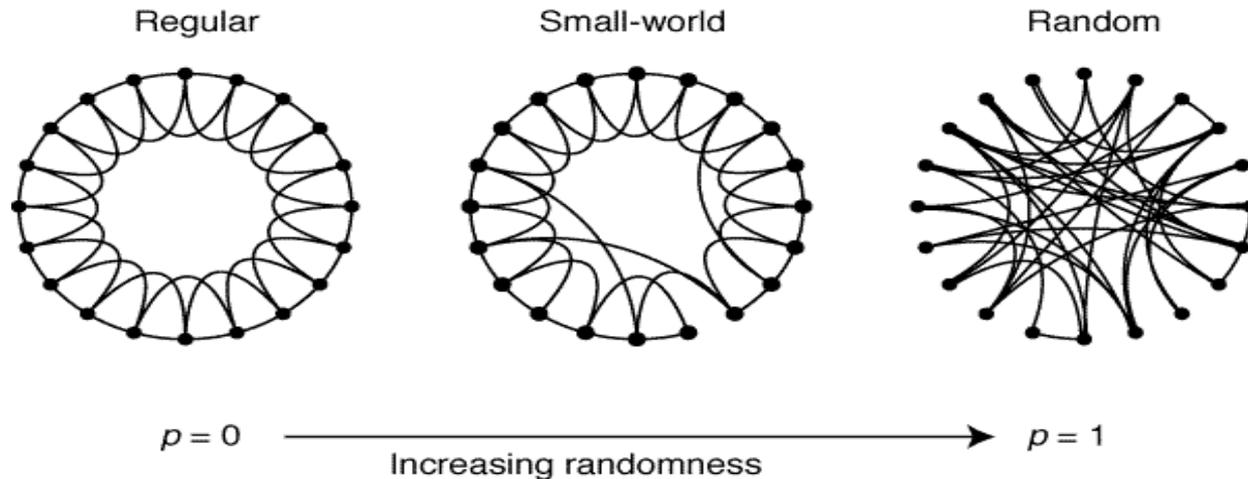
- ▶ Start with n nodes and 0 edges
- ▶ Connect each pair of vertices with probability p_{ER}



- ▶ Many properties in these graphs appear quite suddenly, at a threshold value of p_{ER}
 - ▶ If $p_{ER} \sim c/n$ with $c < 1$, then almost all nodes belong to isolated trees

The Watts-Strogatz [WS] Model (1998)

- ▶ Start with a regular network with n nodes
- ▶ Rewire each edge with probability p



- ▶ For $p=0$ (regular networks)
 - ▶ High clustering coefficient C , high characteristic path length L
- ▶ For $p=1$ (random networks)
 - ▶ Low clustering coefficient C , low characteristic path length L

The Watts-Strogatz [WS] Model (1998)

- ▶ There is a broad interval of p for which characteristic path length L is small but clustering coefficient C remains large
- ▶ Small world networks are common

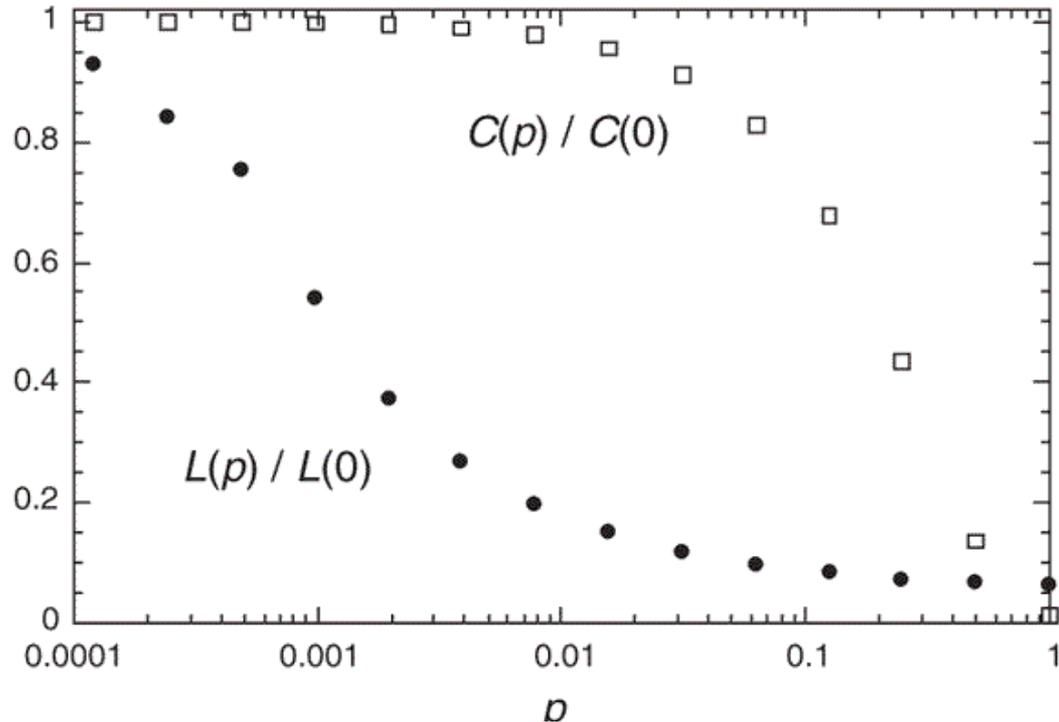
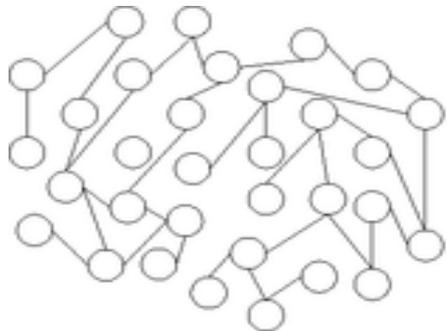


Table 1 Empirical examples of small-world networks

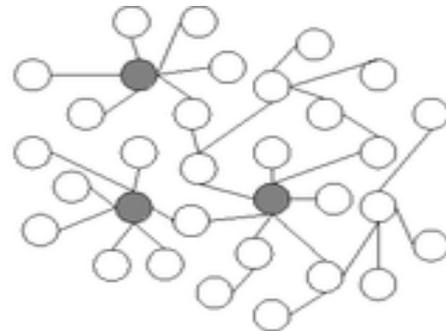
	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
<i>C. elegans</i>	2.65	2.25	0.28	0.05

The Barabási-Albert [BA] Model (1999)

- ▶ Look at the distribution of degrees k
- ▶ A scale-free network is a network where small proportion of the nodes have high degree of connection ("highly connected hubs")
 - ▶ The probability of finding a highly connected node decreases exponentially with k
 - ▶ $p(k) \sim k^{-\gamma}$, a given node has k connections to other nodes with probability as the power law distribution with $\gamma = [2, 3]$



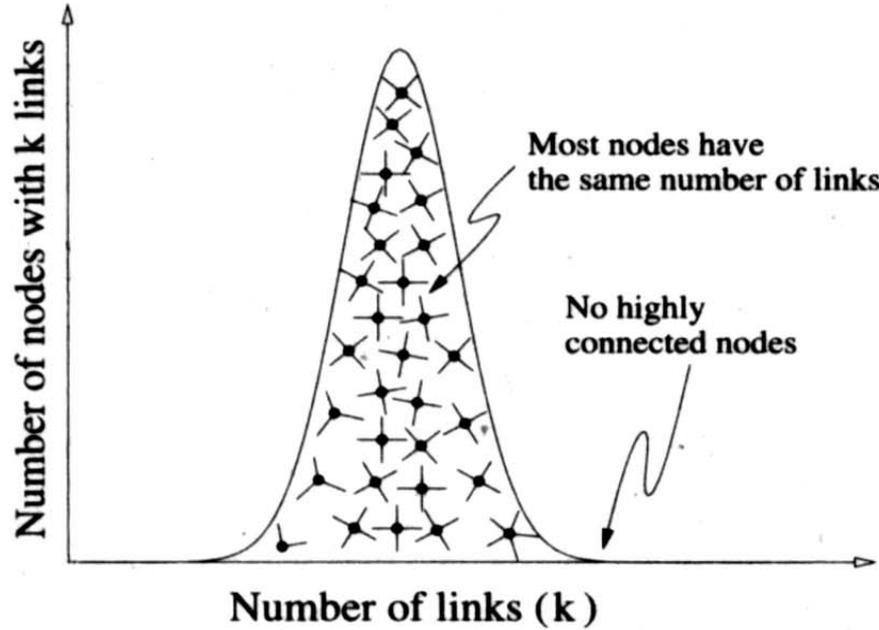
(a) Random network



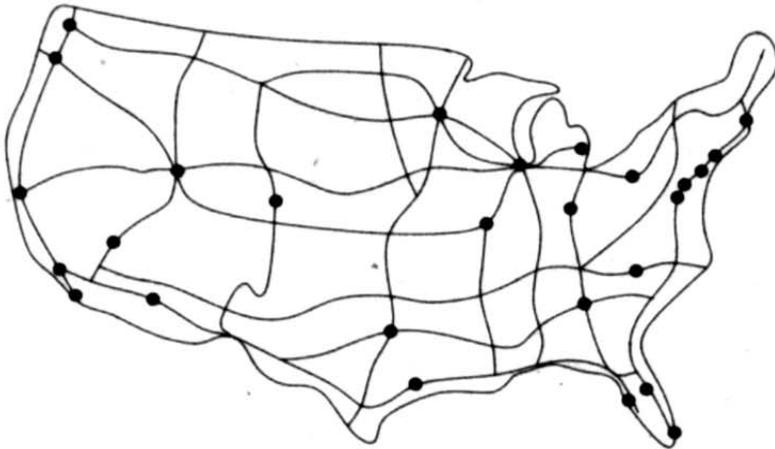
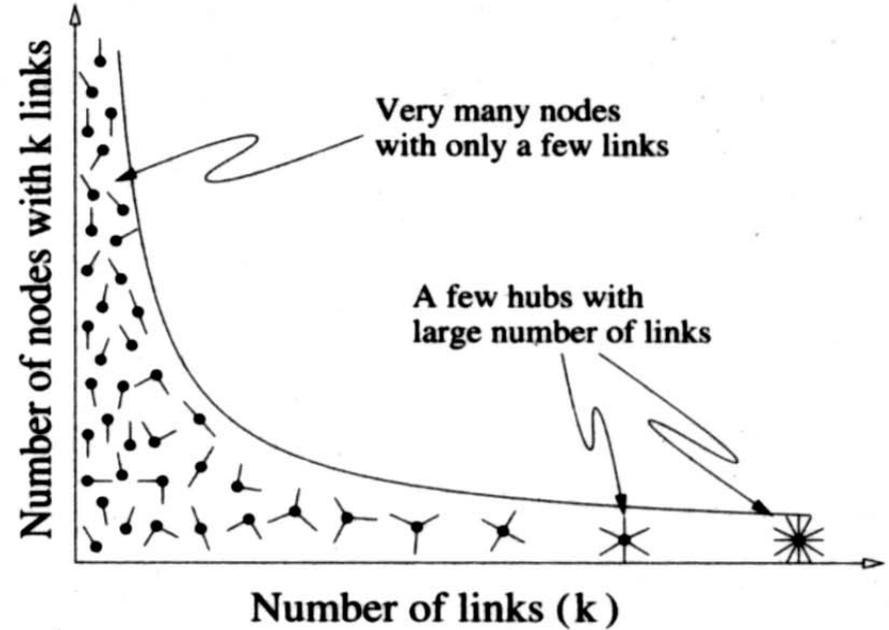
(b) Scale-free network

The Barabási-Albert [BA] Model (1999)

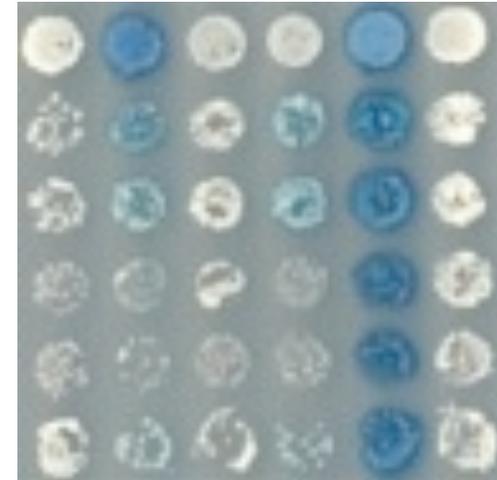
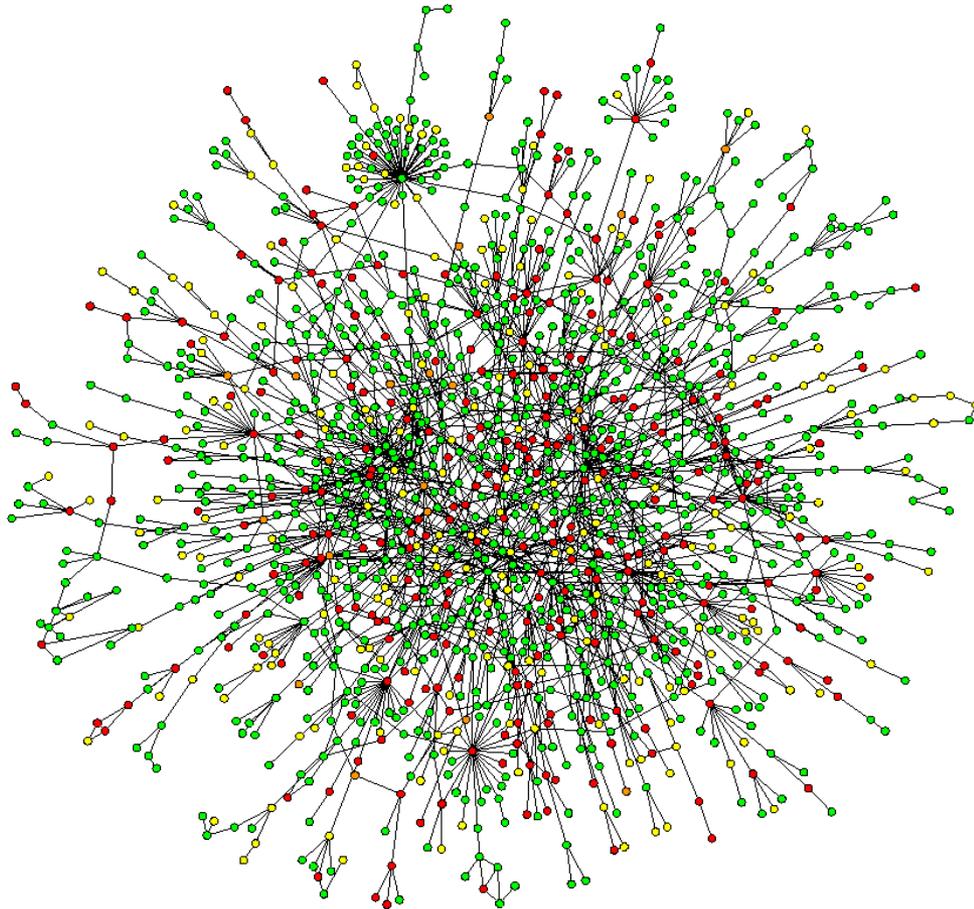
Bell Curve



Power Law Distribution



Protein Interaction Networks



- ▶ Also other networks, e.g. transcript correlation networks

Two Vague Ideas

- ▶ Are scale-free and small-world networks relevant or more an artifact ?
- ▶ Taxonomy for centrality measures

Taxonomy for Centrality Measures

