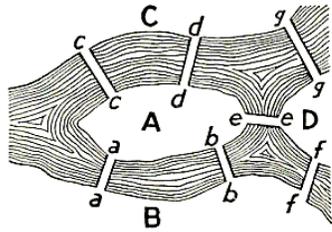


SYSTEMS BIOLOGY 1: NETWORKS



SYSTEMS BIOLOGY

- Starting around 2000 a number of biologists started adopting the term *systems biology* for an approach to biology that emphasized the systems-character of biology: how multiple parts are integrated together in biological organisms
 - In part, this reflected a growing frustration with reductionistic approaches to biology, especially molecular biology
 - While they generated lots of valuable information, reductionists couldn't answer fundamental questions about how organisms function
- Systems biologists adopted a more holistic perspective—how are the various components of living organisms organized into systems
 - Invoking network analyses to represent the integrated nature of biological organisms
 - Invoking computational modeling to understand the dynamical behavior of biological organisms

MAPPING GENOMES

- In October, 1990, a project was launched to map the human genome: to determine the sequence of the three billion base pairs making up the 23 chromosomes found in human beings
- Two groups, one university based, one industry based, competed but on February 12, 2001 together published the first draft of the total human genome
- Expectations were high that knowing the sequence of base pairs would lead to major breakthroughs in both basic science and medicine, but the results have been far less overwhelming
- One thing the project did accomplish was to reduce the estimate of number of genes in humans from >100,000 to 20,000 to 25,000
 - leading to a change in focus from genes coding for traits to understanding how their expression is coordinated in us and other organisms

THE OMICS

- Genome mapping gave rise to the field of *genomics*
- It involves more than gene mapping: focus on gene function, expression, etc. Goal of quantifying the class of molecules, specifying their structure and function, and characterizing their dynamic behavior
- It was soon complemented by other fields that focused on characterized the large numbers of entities that figure in living systems
 - Proteomics: focused on proteins
 - Lipidomics: focused on lipids
 - Metabolomics: focused on the small molecules that figure in metabolism

HIGH-THROUGHPUT DATA COLLECTION

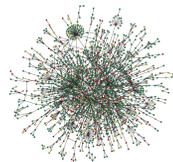
- In traditional biology
 - Individual researchers or labs conducted experiments, collected data, and interpreted their findings
 - Review articles brought multiple studies together to provide a view on the current state of inquiry
- Identifying large numbers of genes, proteins, small molecules, etc. involved in living organisms has required development of automated techniques that execute experiments, collect data, and perform analyses of it
 - Employing novel methods: microarrays, yeast two-hybrid screens, etc.
 - Pooling the data in large databases (sometimes curated by individual scientists)
 - Deploying automated techniques to interpret (or aid interpretation) of the data

NETWORK SCIENCE

- Similar networks found across a vast range of disciplines, raising the prospect of universal laws of organization
 - whole network level
 - small sub-graph level
 - meso-level (mechanism)



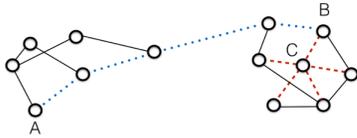
Eukaryotic metabolic network



Protein interaction network from yeast—nodes are proteins and edges represent binding

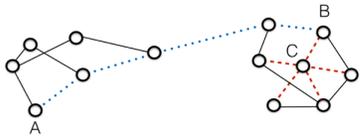
ABSTRACT ORGANIZATION OF NETWORKS

- In systems/mechanisms in the real world
 - Entities have distinctive properties
 - Relations between entities involve different types of relations
- Network representations abstract from these properties
 - Entities are represented as nodes
 - Relations as edges
 - undirected or directed



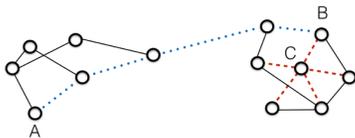
NETWORK MEASURES

- Path-length measures
 - Shortest path length between two nodes is the minimum number of links that need to be followed to traverse from one to the other
 - Efficiency: Inversely related to shortest path length
 - Diameter: longest shortest path length
 - Average or characteristic path length



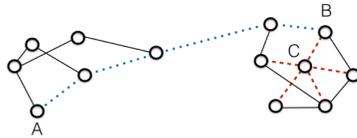
NETWORK MEASURES

- Clustering measures
 - Clustering coefficient of a node: proportion of the possible linkages between nearest neighbors that are realized
 - Clustering coefficient of a graph: average of clustering coefficients of nodes
 - Clusters or communities: subgraphs whose nodes are highly interconnected



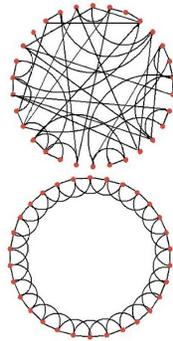
NETWORK MEASURES

- Connectivity Measures:
 - Degree or connectivity of a node k : the number of links between the node and other nodes
 - Degree distribution $P(k)$: probability that a randomly chosen node has degree k
 - Assortativity: correlation between degrees of connected nodes. Positive value indicates high-degree nodes tend to connect with each other
- Betweenness Measures
 - Centrality: Percentage of shortest paths going through a node (or link)



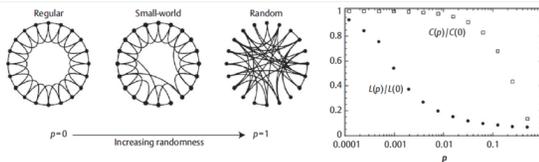
CLASSICAL FOCI: LATTICES AND RANDOM GRAPHS

- The 20th century pioneers of graph theory focused primarily on totally connected networks, random networks or regular lattices
- Characteristic path length (L): mean of the shortest paths (d) between all pairs of nodes
 - In a totally connected graph $L=1$
 - Regular lattice: L is very large
 - Random graph: L can be very small
- Clustering Coefficient (C)
 - For regular lattice: C is very high
 - For random graph: C is very low
- Apparent tradeoff between keeping path length short and maintaining high clustering



SMALL-WORLDS: BETWEEN LATTICES AND RANDOM GRAPHS

- Watts and Strogatz (1998) identified a class of networks referred to as small worlds
- They start with a regular lattice and gradually replace a local connection with a longer-distance one
 - Very quickly the characteristic path length drops to close to that for a random graph
 - The clustering coefficient remains high through many replacements
 - This identifies the range of small worlds



SMALL WORLDS

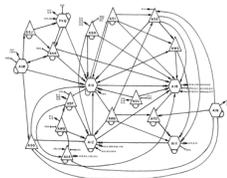
- Watts and Strogatz called these networks in the middle range *small-worlds*
- Showed that they occur widely in real world networks

Table 1 Empirical examples of small-world networks

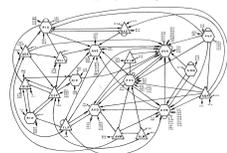
	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
<i>C. elegans</i>	2.65	2.25	0.28	0.05

E. ELEGANS NEURAL NETWORK

- Based on serial electron micrographs, White et al. (1986) reconstructed the entire neural network of the hermaphrodite nematode worm *C. elegans*
- 302 neurons
- approx. 5000 chemical synapses
- approx. 600 gap junctions
- approx. 2000 neuromuscular junctions
- Watts and Strogatz showed that the *C. elegans* network constitute a small-world



• Chemoreception circuit



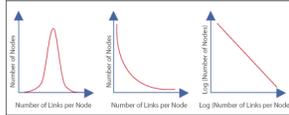
Other sensory neurons

THE FUNCTIONALITY OF SMALL-WORLDS

- “Models of dynamical systems with small-world coupling display enhanced **signal-propagation speed, computational power, and synchronizability**. In particular, infectious diseases spread more easily in small-world networks than in regular lattices.” (Watts and Strogatz)

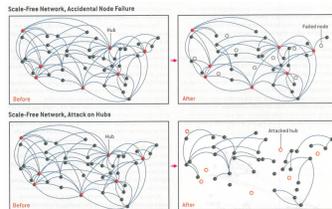
NODE DEGREE: FROM GAUSSIAN TO POWER LAW DISTRIBUTIONS

- Plausible assumption: node degree is distributed in a Gaussian fashion
- Exponential because "the probability that a node is connected to k other sites decreases exponentially for large k ."
- Surveying webpages at U. of Notre Dame, Barabási and his colleagues were surprised to find a few nodes that had far higher degree than expected
- Rather than exponential, they fit a power law: probability of the degree of a node was $1/k^n$
 - No peak to the distribution
- When plotted on log-log coordinates it forms a straight line



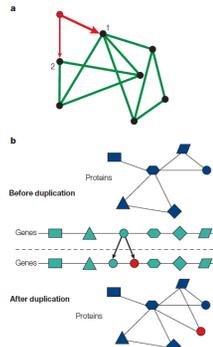
ROBUSTNESS AND VULNERABILITY

- Nodes with few connections can often be lost without compromising the integrity of the network
 - "as many as 80 percent of randomly selected Internet routers can fail and the remaining ones will still form a compact cluster in which there will still be a path between any two nodes. It is equally difficult to disrupt a cell's protein-interaction network: our measurements indicate that even after a high level of random mutations are introduced, the unaffected proteins will continue to work together."
- But loss of just a few highly-connected nodes (hubs) can be catastrophic
 - Networks fractionate



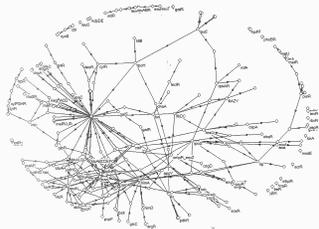
CREATING SCALE-FREE NETWORKS

- Barabási and his collaborators offered proposals for how scale-free networks might evolve
 - Preferential attachment
 - Duplication



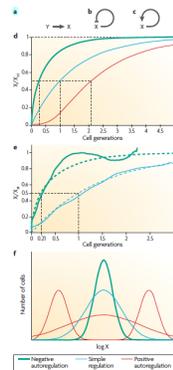
FROM WHOLE NETWORKS TO SUBGRAPHS: DISCOVERY OF MOTIFS

- Uri Alon and collaborators identified subgraphs (consisting of a small number of nodes) that occur repeatedly within networks such as the transcription regulation networks in *E. coli*
- They dubbed those subgraphs that occur far more frequently than would be expected by chance *motifs*
 - Raises important question about what counts as chance
- Major claims for motifs:
 - They are building blocks of larger networks
 - They carry out specific information processing functions



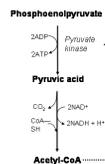
SIMPLE REGULATION

- Alon begins his analysis with single directed edges either between two nodes or from one node back onto itself
- In a transcription network, a single directed edge between two nodes will allow the first to increase the transcription of the second until it reaches steady-state
- Negative feedback allow for a faster rise in the concentration of the transcript but then leveling off due to the feedback. It can also reduce variability
- Positive feedback can yield slower rise followed by an inflection (s-shaped curve) and greater variability



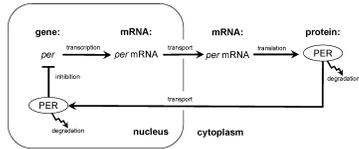
NEGATIVE FEEDBACK AS A DESIGN PRINCIPLE

- A gene regulating its own transcription is a pretty simple idea that might not be thought to have much significance
 - but it came to be recognized as a primary tool for regulating systems
 - and applied in a wide variety of contexts from biology to social systems to engineering



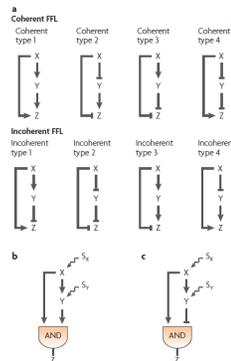
NEGATIVE FEEDBACK AS A DESIGN PRINCIPLE

- Engineers recognized that negative feedback systems typically result in oscillation—under and overshooting the target rather than simply reaching it and stopping
- As biologists recognized the prevalence of oscillatory phenomena in biological systems, they appealed to negative feedback mechanisms to explain it
 - Example: circadian rhythms



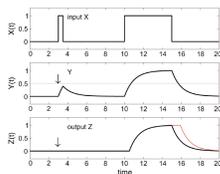
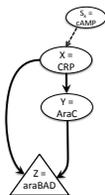
FEED-FORWARD LOOPS

- Three node subgraphs containing a direct route from X to Z and indirect route through Y
- Either route can produce activation or repression
 - Coherent if both routes have the same sign
 - Incoherent if opposite sign
- Treats Z as executing a Boolean function on its inputs
 - Either an AND or an OR gate
- In *E. coli* and yeast transcription networks two of these appear more frequently than by chance



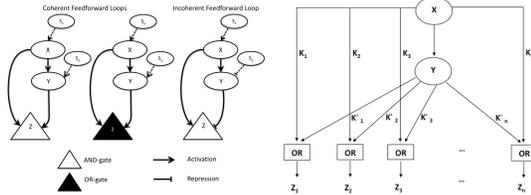
PERSISTENCE DETECTOR

- Sign-sensitive delay element: input X must persist long enough for Y to be expressed in a sufficient amount before transcription of Z begins
- Prevents transcribing Z in response to random perturbations
- Also stops transcribing Z promptly when X is inactivated



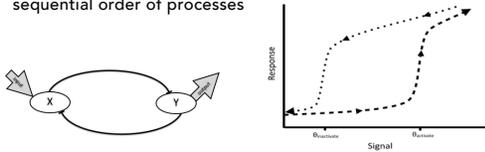
OTHER FEED-FORWARD LOOPS

- When the AND-gate is replaced by an OR-gate in a coherent feedforward loop, transcription of Z starts immediately when X is activated, but continues in the face of short disruptions due to Y
 - Multiple outputs with time delays enables sequential synthesis (e.g., of a flagellum in bacteria)
- With an incoherent loop, X will initially cause transcription of Z, but as Y builds up, transcription will stop, resulting in a pulse
 - enables rapid generation of Z without the risk of generating too much



BI-STABLE SWITCH

- With appropriate parameters, a double negative feedback loop generates an especially useful property
 - With increased input to one unit, it will switch off the other unit
 - But the input to that unit must drop to a much lower value for the that unit to become active again turn it off
 - The switch will remain stable in either position, not easily reverted to the other
 - Useful in context in which it is important to maintain the sequential order of processes



COMBINING MOTIFS

- Network that governs spore development in *Bacillus subtilis*
 - A process that bacteria undertake only in desperate circumstances
 - The outputs Z_1 , Z_2 , and Z_3 each represent large numbers of genes
 - The network combines two incoherent FFLs that each produce pulses and two coherent FFLs that generates a steady output

