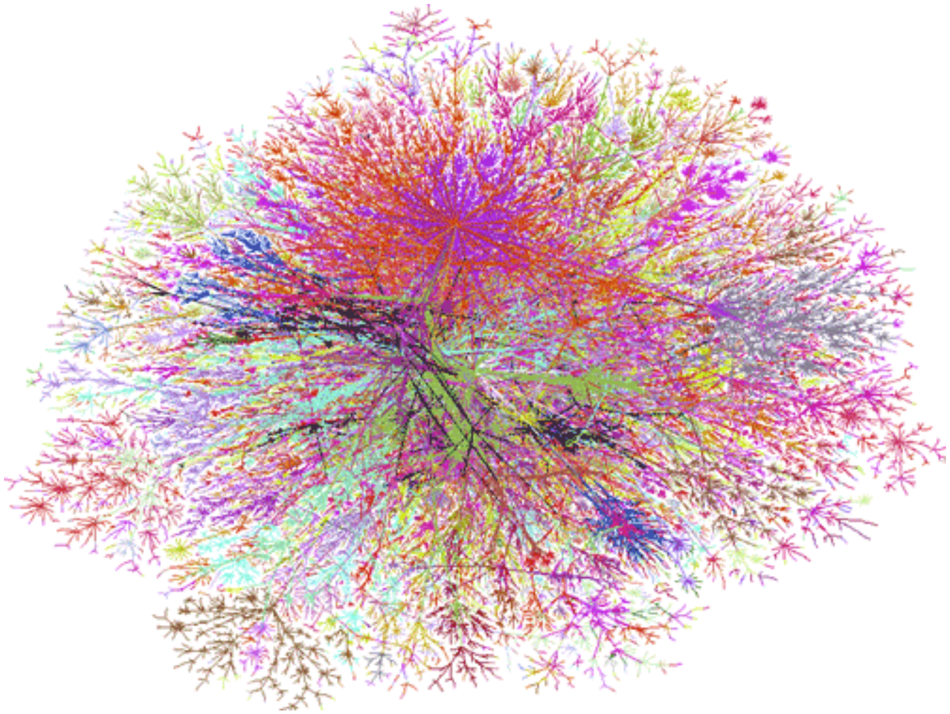
A Physics Approach to Understanding Complex Networks

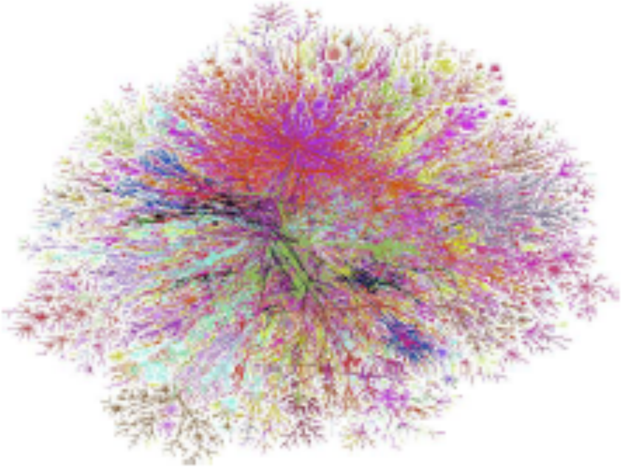


Michelle Girvan

Outline

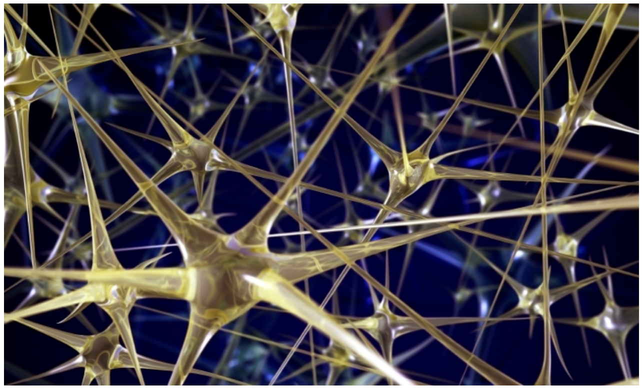
- Overview of the emerging field of “*network science*,” or “*complex networks*”
- Complex networks meets nonlinear science: Modeling the dynamics of gene networks
- Outlook for the field

Examples of Complex Networks



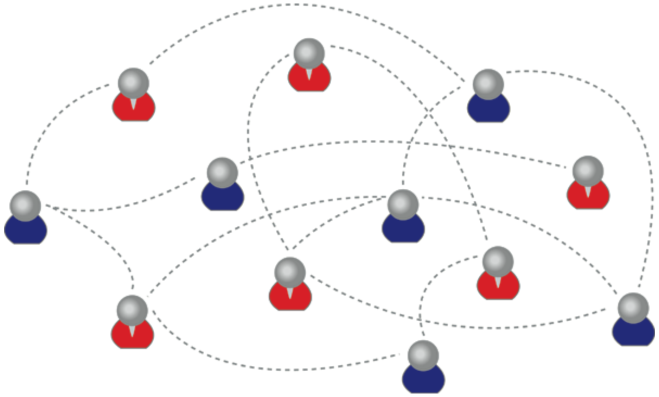
http://barabasilab.com/gallery

The Internet



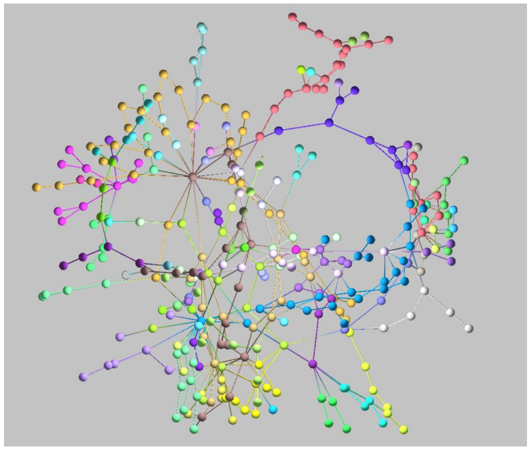
http://www.zmescience.com

A Neural Network



http://www.relenet.com/

A Social Network



http://www4.toulouse.inra.fr/toxalim_eng

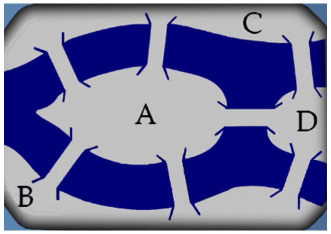
A Metabolic Network

Traditional vs. Complex Systems Approaches to Networks

Traditional Questions:

Social Networks:
Who is the most "important" person in the network?

Graph Theory:
Does there exist a cycle through the network that uses each edge exactly once?

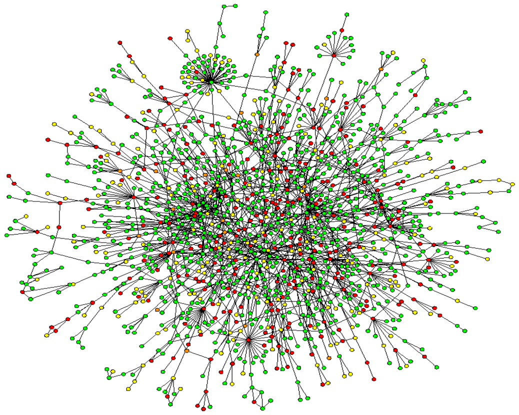


Complex Systems Questions:

What fraction of edges have to be removed to disconnect the graph?

What kinds of structures emerge from simple growth rules?

How does the network structure influence the system's dynamics?



Areas of Network Research

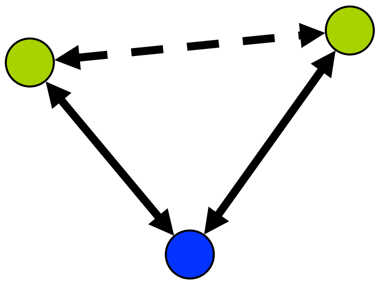
Structural Complexity

- The wiring diagram could be an intricate tangle, far from perfectly regular or perfectly random.
- The network could include different classes of nodes
- The edges could be heterogeneous with different weights, directions and signs.

Dynamical Complexity

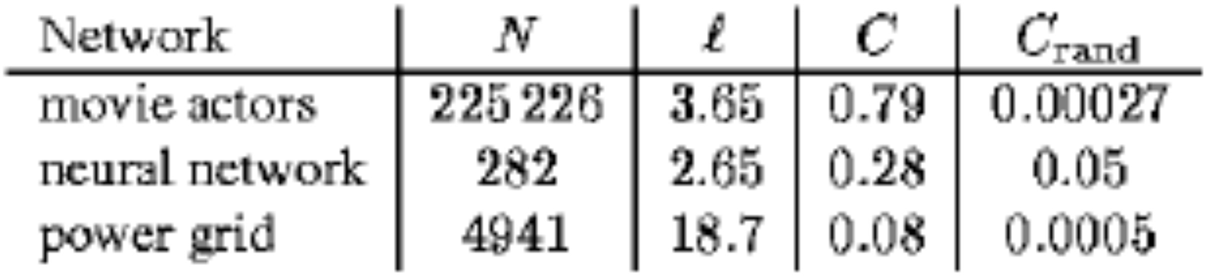
- Dynamics on the network: processes could be taking place on the fixed network. *Examples: disease spread, synchronization*
- Dynamics of the network: the network itself could be evolving in time.

Clustering



C = Probability that two of a node's neighbors are themselves connected

In a random graph: $C_{\text{rand}} \sim 1/N$ (if the average degree is held constant)

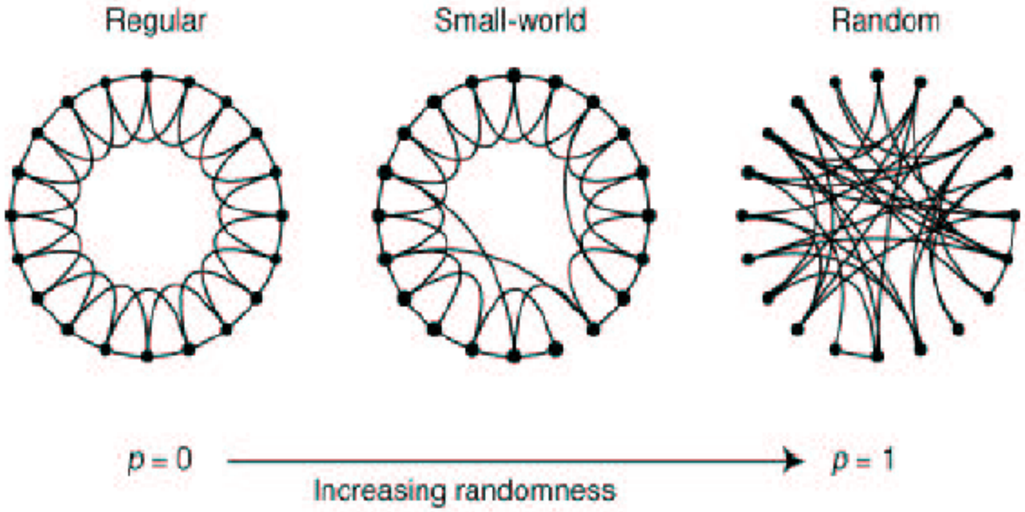


Network	N	ℓ	C	C_{rand}
movie actors	225 226	3.65	0.79	0.00027
neural network	282	2.65	0.28	0.05
power grid	4941	18.7	0.08	0.0005

Table from Watts & Strogatz, *Nature* (1998)

Watts-Strogatz 'Small World' Model

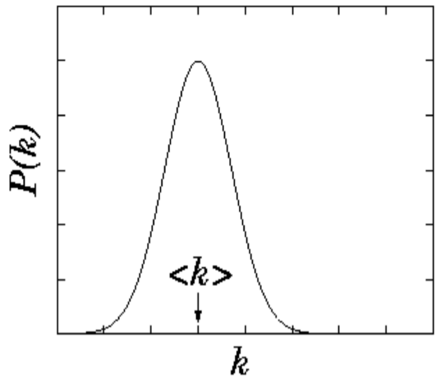
Watts and Strogatz introduced this simple model to show how networks can have both short path lengths and high clustering.



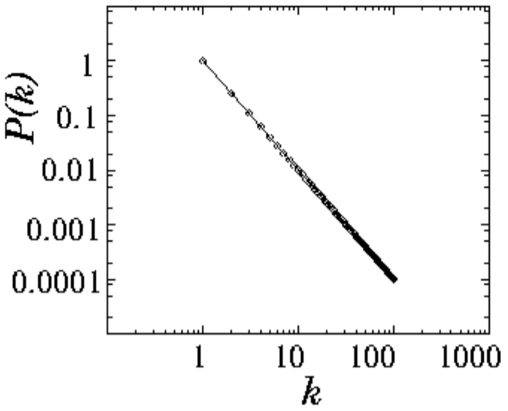
D. J. Watts and S. H. Strogatz, *Collective dynamics of "small-world" networks*, Nature, 393 (1998), pp. 440–442.

Degree Distributions

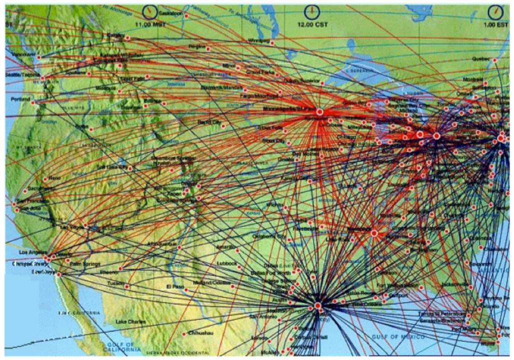
Poisson distribution



Power-law distribution



Exponential Network

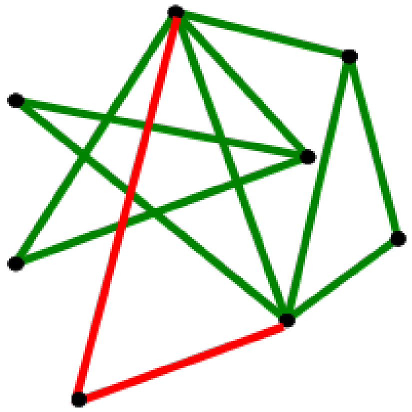


Scale-free Network

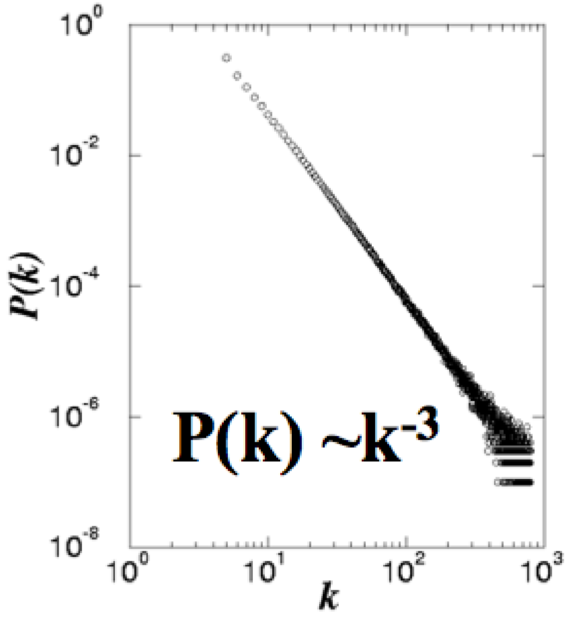
Images from the Barabasi Lab

How do power law degree distributions arise?

One possible answer: Barabasi-Albert model of preferential attachment



- Growth - At each time step, we add a node with *m* new edges (connecting to nodes already existent in the system)
- Preferential attachment - The probability that a new node connects to an existing node *i* depends on the connectivity, k_i of that node.



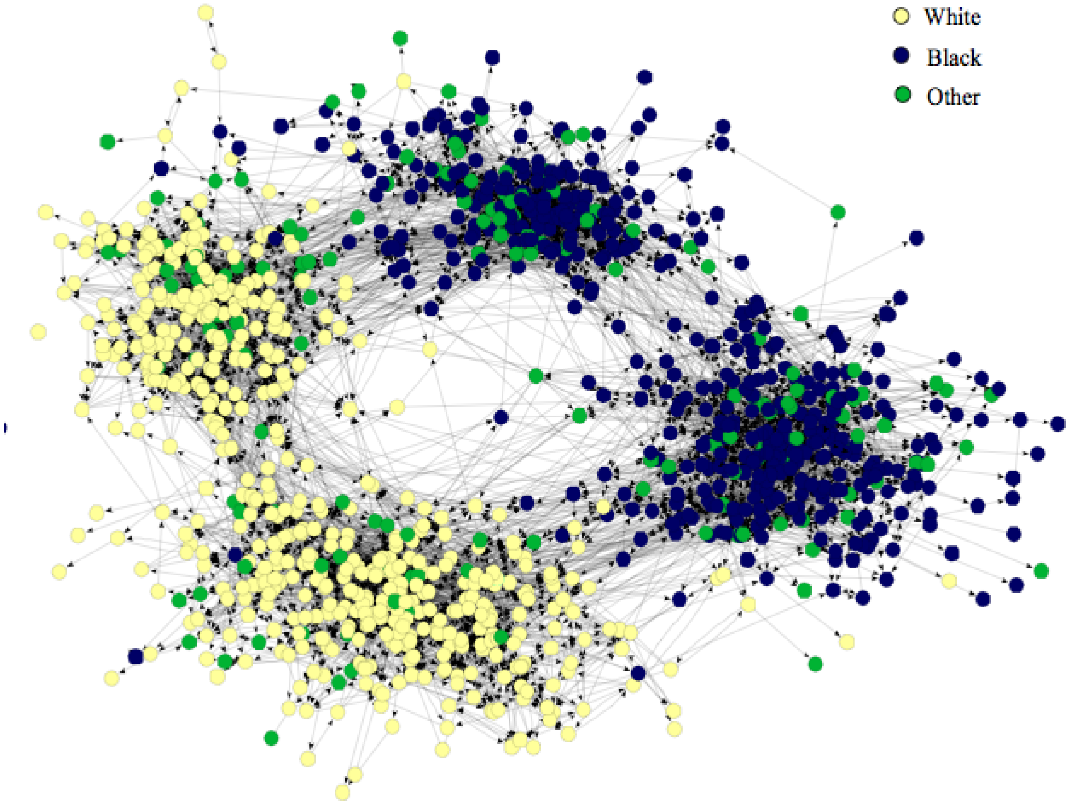
A.-L.Barabasi, R. Albert, Science 286,509 (1999)).

Implications of the Preferential Attachment Model

- Older nodes tend to have higher degrees. (This expected correlation does not appear to hold for WWW data)
- If the network is directed, cycles do not exist.
- Networks generated from the Barabasi-Albert model are assortatively neutral.

Assortative Mixing

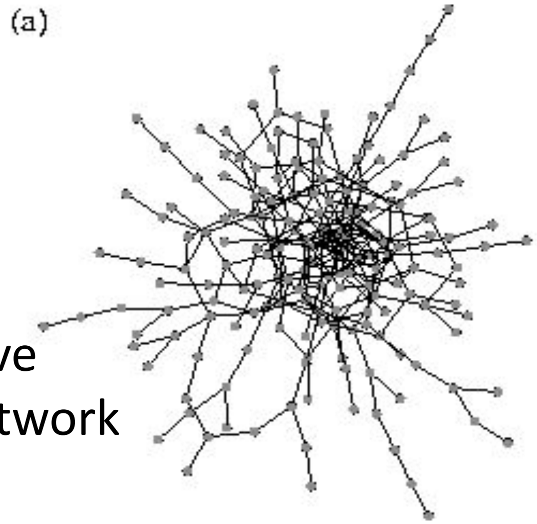
In assortatively mixed networks, like vertices tend to connect preferentially to one another.



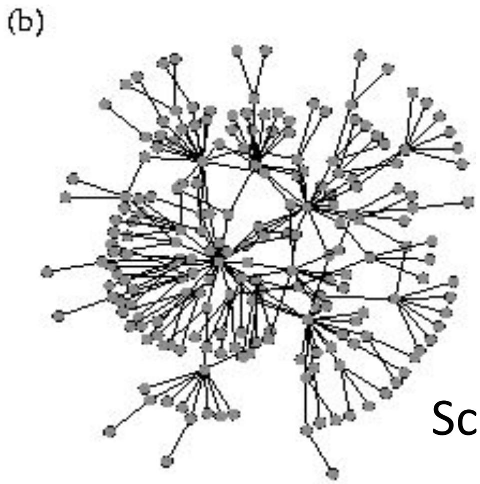
Friendship network of students in a U.S. school. Friendships are determined by asking the participants, and hence are directed, since A may say that B is their friend but not vice versa. Vertices are color coded according to race, as marked, and the split from left to right in the figure is clearly primarily along lines of race. The split from top to bottom reflects a division between middle school and high school students.

Assortative Mixing by Degree

- A network is said to be assortatively mixed by degree if high degree vertices tend to connect to other high degree vertices
- A network is disassortatively mixed by degree if high degree vertices tend to connect to low degree vertices.

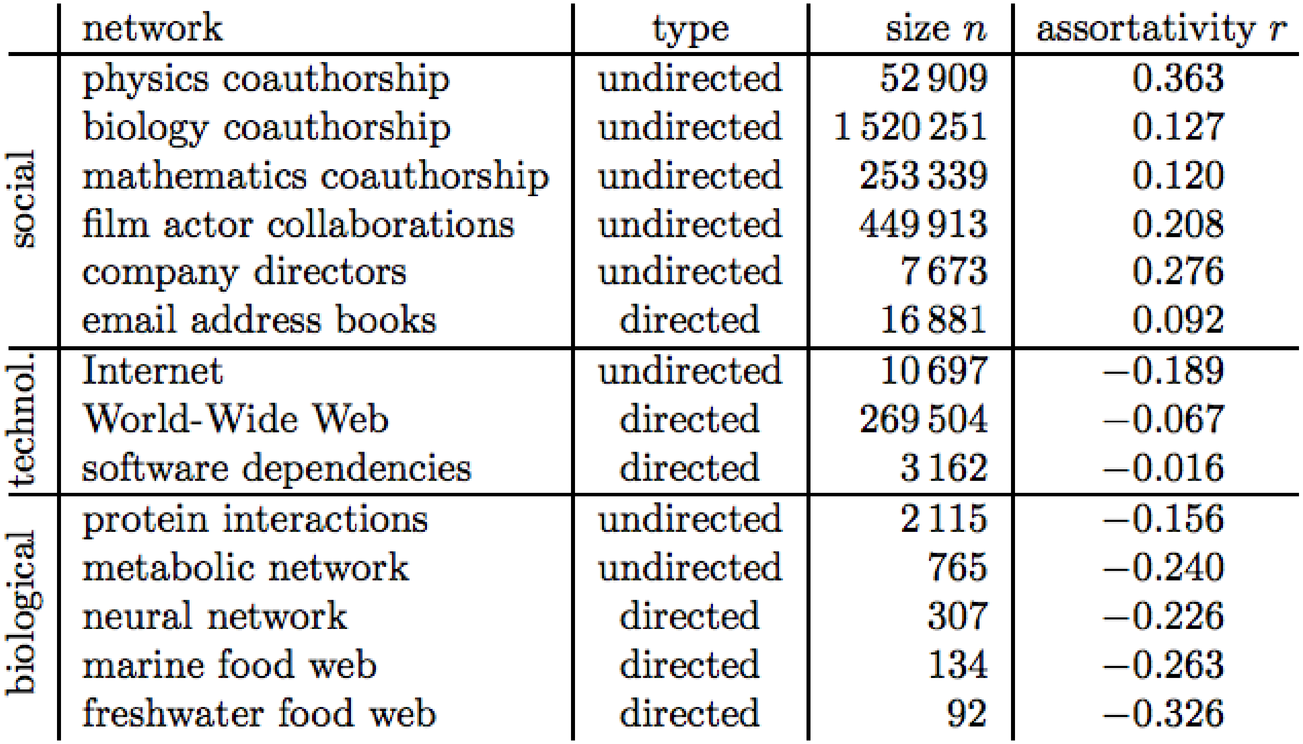


Assortative
Scale-free network



Disassortative
Scale-free network

Measured assortativity for various networks



	network	type	size n	assortativity r
social	physics coauthorship	undirected	52 909	0.363
	biology coauthorship	undirected	1 520 251	0.127
	mathematics coauthorship	undirected	253 339	0.120
	film actor collaborations	undirected	449 913	0.208
	company directors	undirected	7 673	0.276
	email address books	directed	16 881	0.092
technol.	Internet	undirected	10 697	−0.189
	World-Wide Web	directed	269 504	−0.067
	software dependencies	directed	3 162	−0.016
biological	protein interactions	undirected	2 115	−0.156
	metabolic network	undirected	765	−0.240
	neural network	directed	307	−0.226
	marine food web	directed	134	−0.263
	freshwater food web	directed	92	−0.326

M.E.J Newman and M. Girvan, *Mixing Patterns and Community Structure in Networks* (2002).

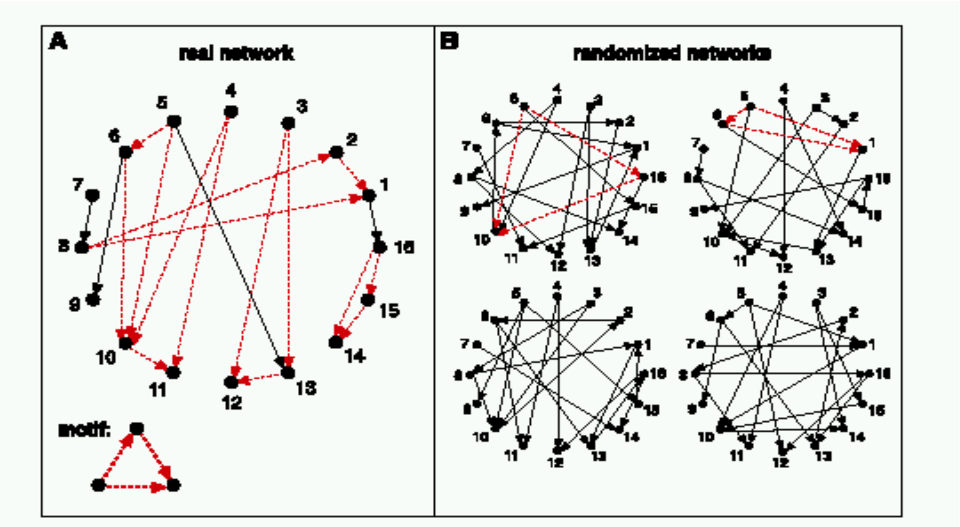
Network Motifs

Motifs

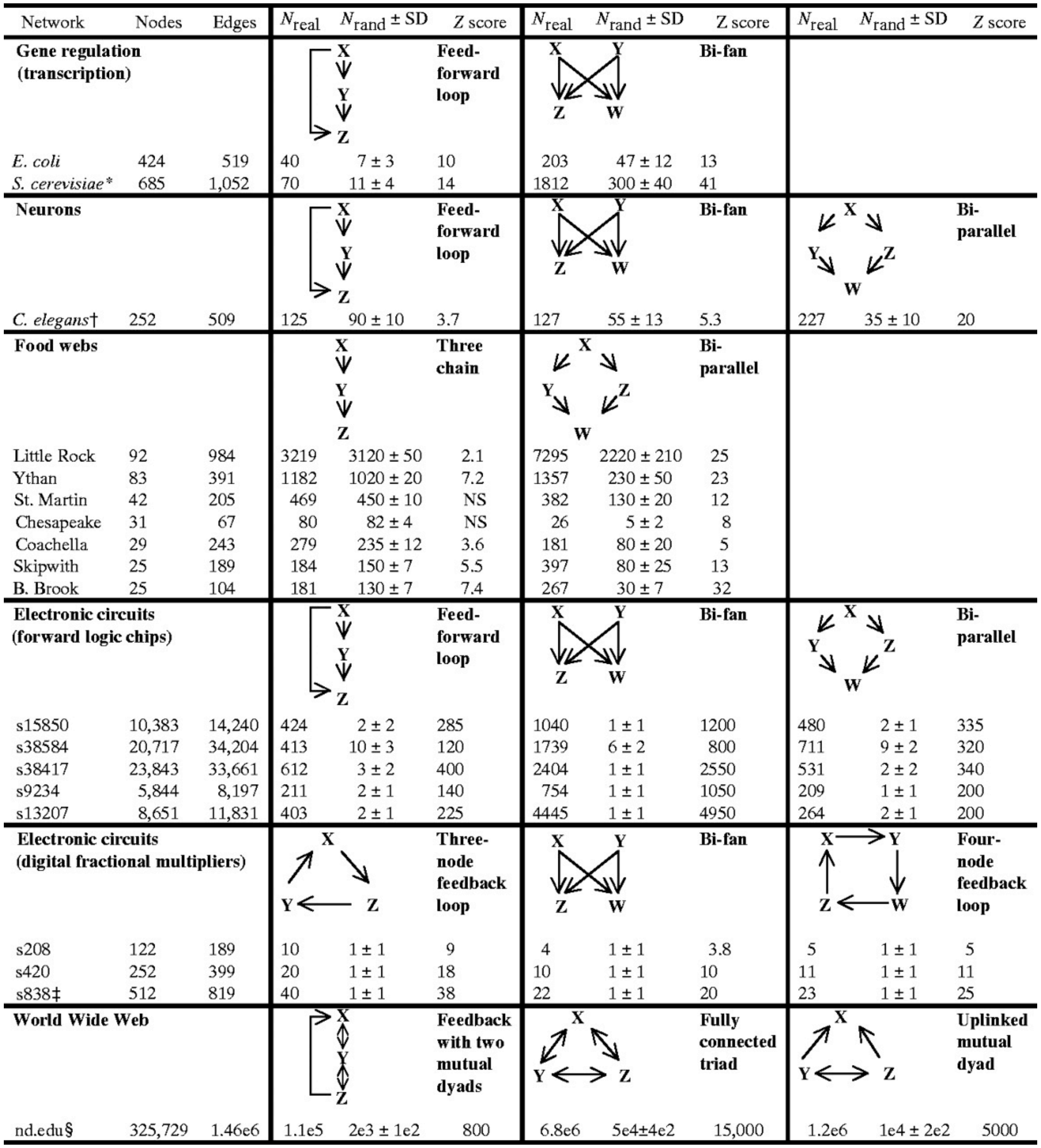
Subgraphs that have a significantly higher density in the observed network than in the randomizations of the same.

Randomized networks:

Ensemble of maximally random networks preserving the degree distribution (or some other feature(s)) of the original network.



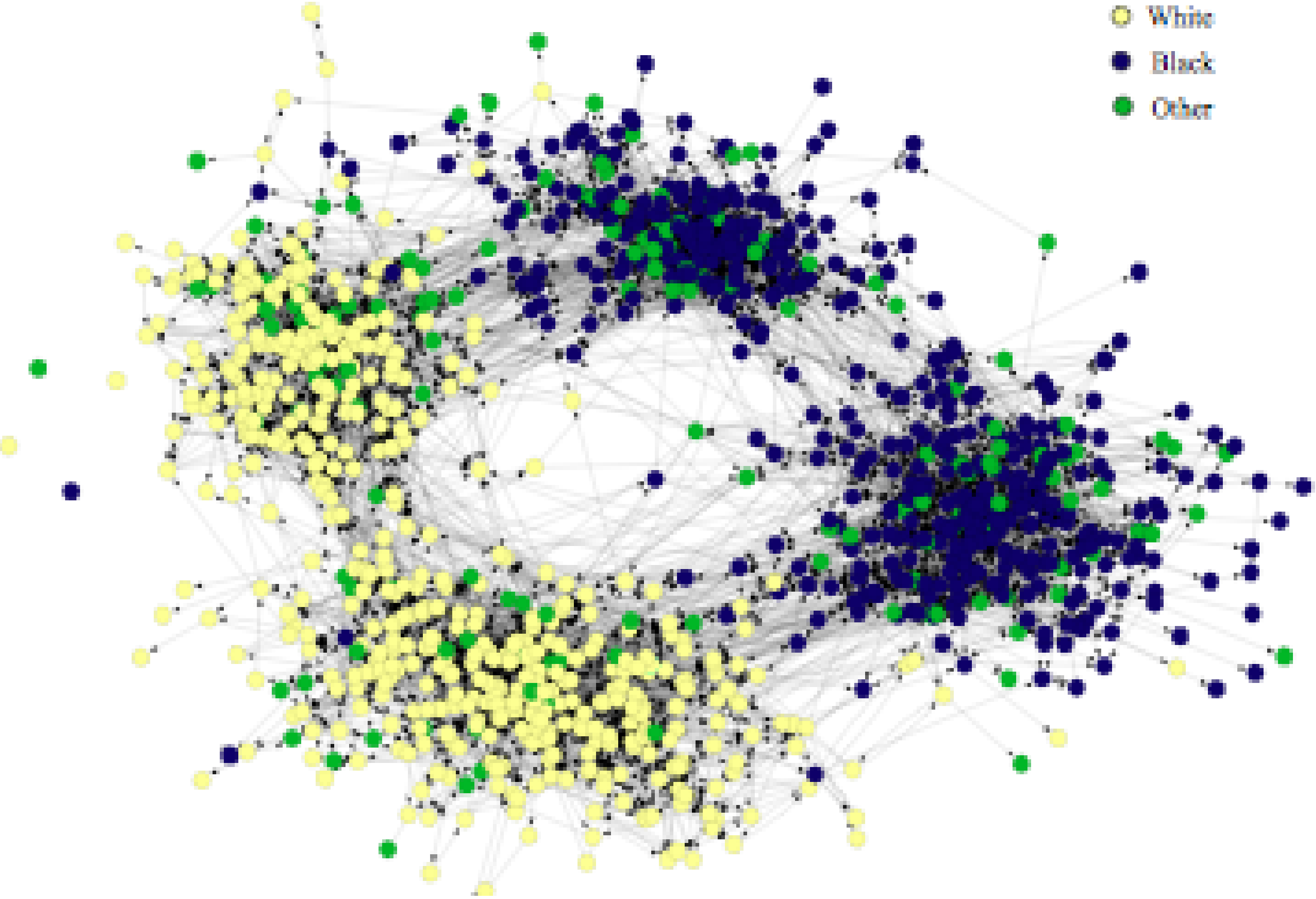
R. Milo et al., Science **298**, 824 (2002)



Network	Nodes	Edges	N_{real}	N_{rand} ± SD	Z score	N_{real}	N_{rand} ± SD	Z score	N_{real}	N_{rand} ± SD	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)					**Three-node feedback loop**			**Bi-fan**			**Four-node feedback loop**
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web					**Feedback with two mutual dyads**			**Fully connected triad**			**Uplinked mutual dyad**
nd.edu§	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4±4e2	15,000	1.2e6	1e4 ± 2e2	5000

R Milo et al., *Science* **298**, 824-827 (2002).

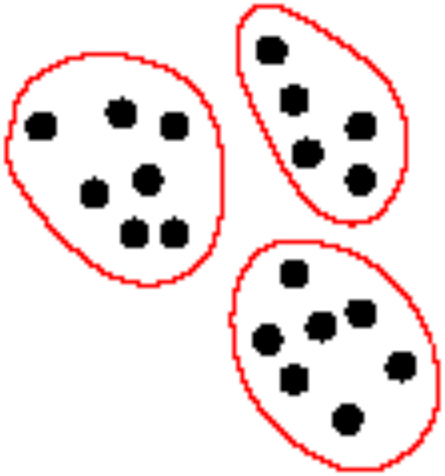
Community Structure in Social Networks



Friendship network of adolescents in a U.S. high school. Courtesy of James Moody

Detecting Communities

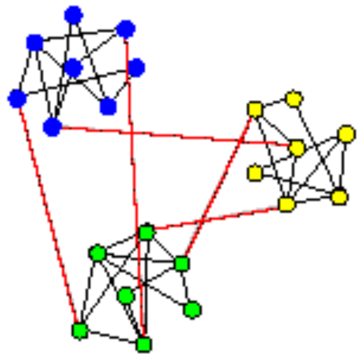
- We are interested in network clustering, which differs from ordinary data clustering.
- In network clustering, relationships between vertices are determined by flows through other vertices.
- In data clustering, relationships between vertices can be determined independently of other vertices
- Traditional methods for network clustering have involved transformation of the network into a data clustering problem.



Data Clustering

Network Clustering

Community Structure

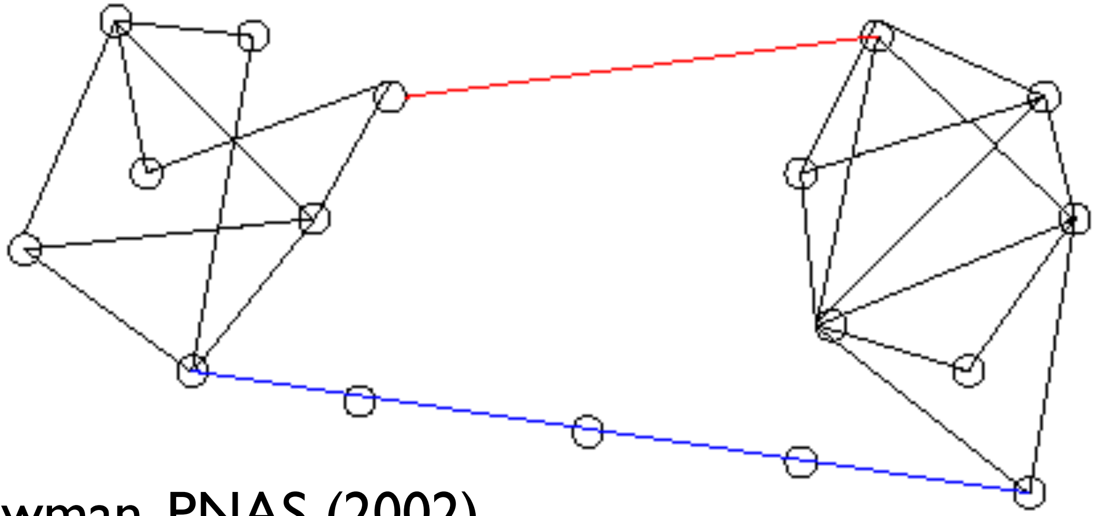


Consider a community detection scheme based on centrality indices:

- Node betweenness: The betweenness centrality of a vertex *i* is the number of shortest paths between pairs of other vertices which run through *i.*
- Edge betweenness: Similarly, the betweenness of an edge *j* is the number of shortest paths between pairs of nodes which run along *j.*

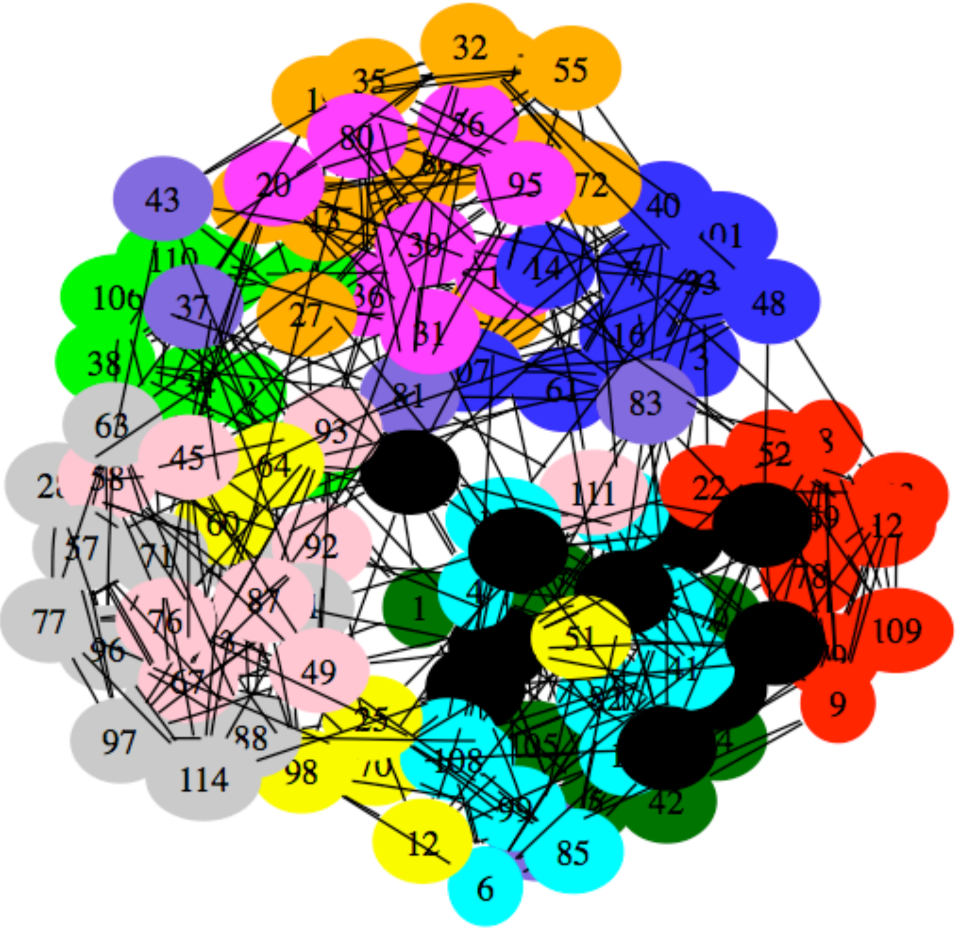
Algorithm for Detecting Communities

1. Calculate the betweenness for all edges in the network.
2. Remove the edge with the highest betweenness.
3. Recalculate betweennesses for all edges affected by the removal.
4. Repeat from step 2 until no edges remain.

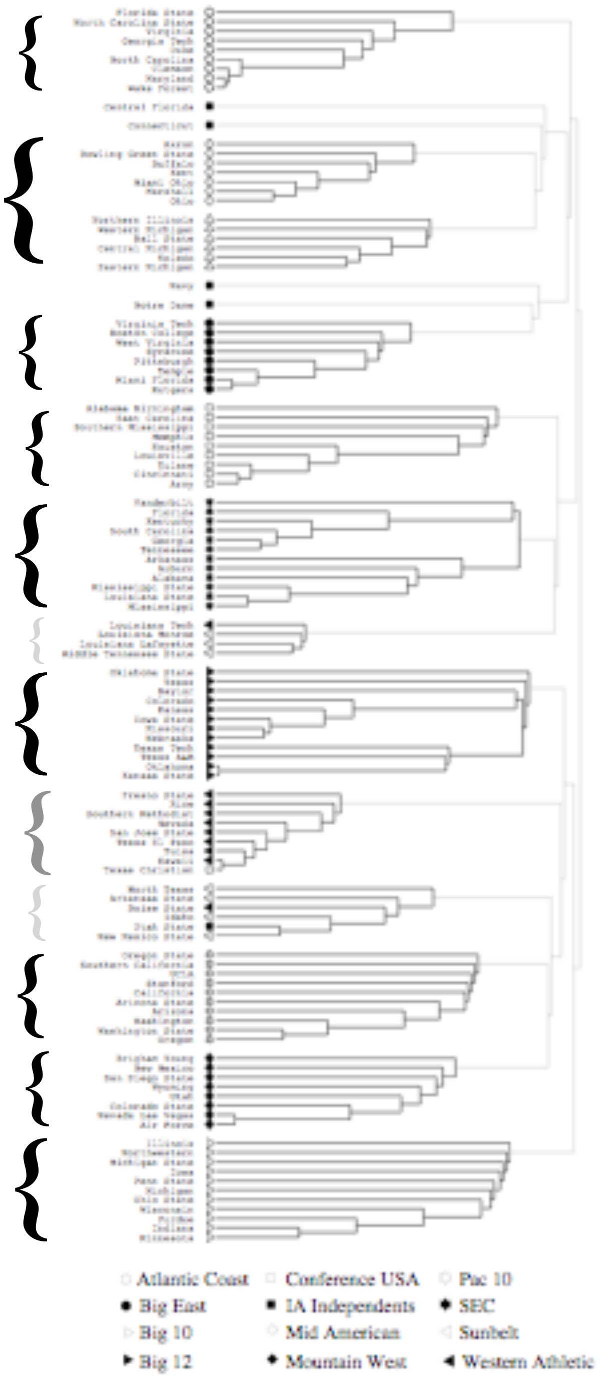


Ref: Girvan & Newman, PNAS (2002)

Illustration: Finding Community Structure in College Football Data



College Football



Atlantic Coast

Mid-American

Big East

Conference USA

SEC

Sunbelt

Big 12

Western Athletic

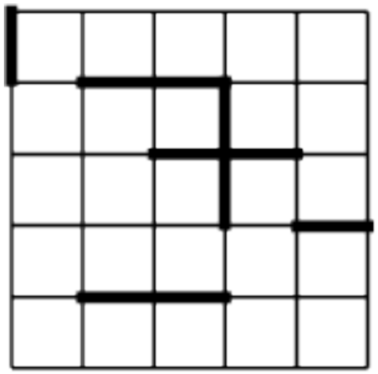
Sunbelt

Pac 10

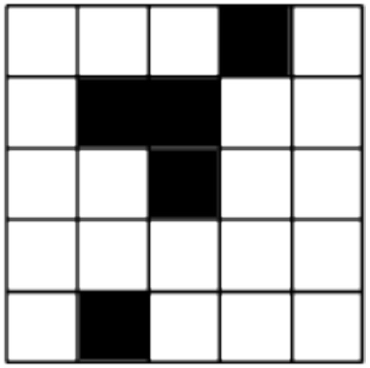
Mountain West

Big 10

Network Robustness and Resilience: Percolation as a Starting Point



bond percolation

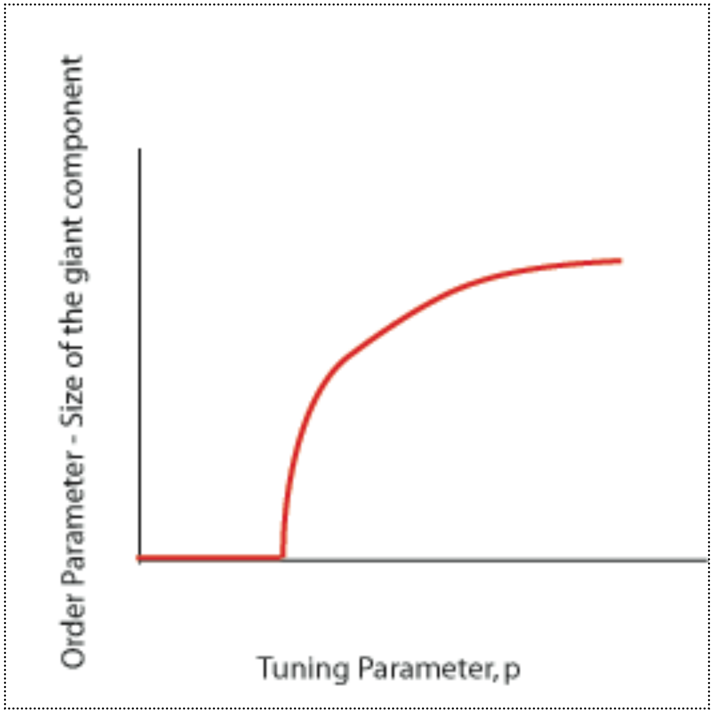


site percolation

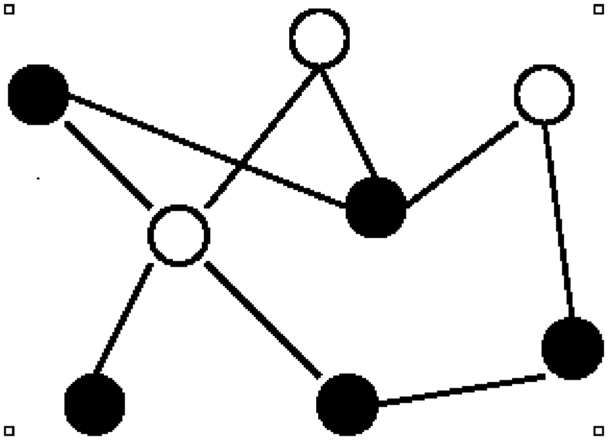
Ordinary Percolation on Lattices: Fill in each link (bond percolation) or site (site percolation) with probability p and ask questions about the sizes of connected components.

Q: What happens as we increase the probability, *p*, of filling in each site?

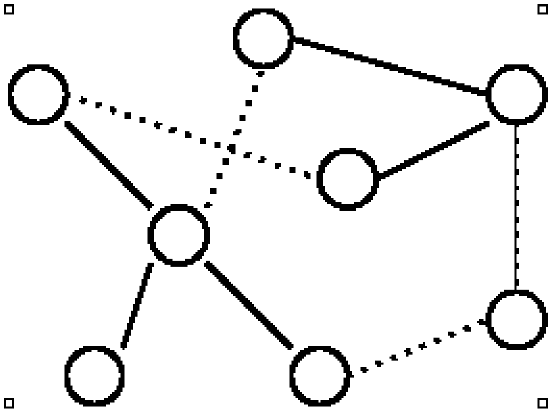
- For low values of p, we see small islands of connected components.
- At a critical value of p, a giant component forms. A giant component is a connected component that occupies a finite fraction of the system, in the limit of infinite system size. At the critical point, there is a power law distribution of the size of connected components.
- Above the critical value, the giant component occupies an increasingly large fraction of the system. If we look at the mean component size excluding the giant component, we observe a characteristic component size.



Percolation on Complex Networks



Site Percolation



Bond Percolation

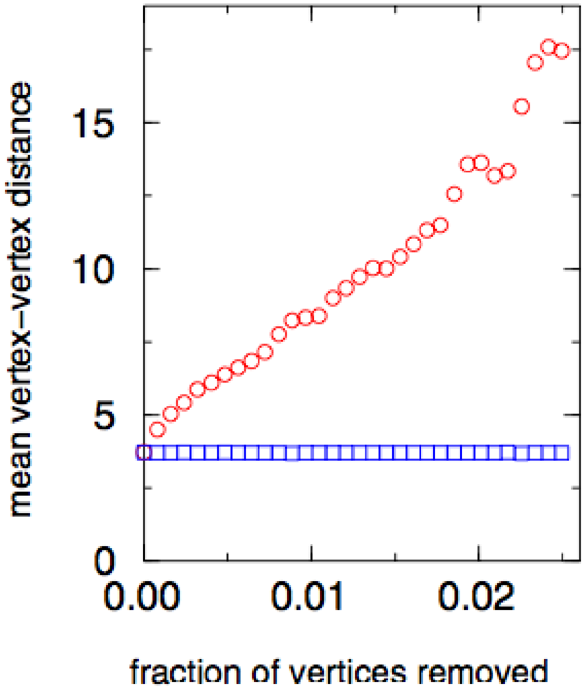
- Percolation can be extended to networks of arbitrary topology.
- We say the network percolates when a giant component forms.

How does percolation relate to network resilience?

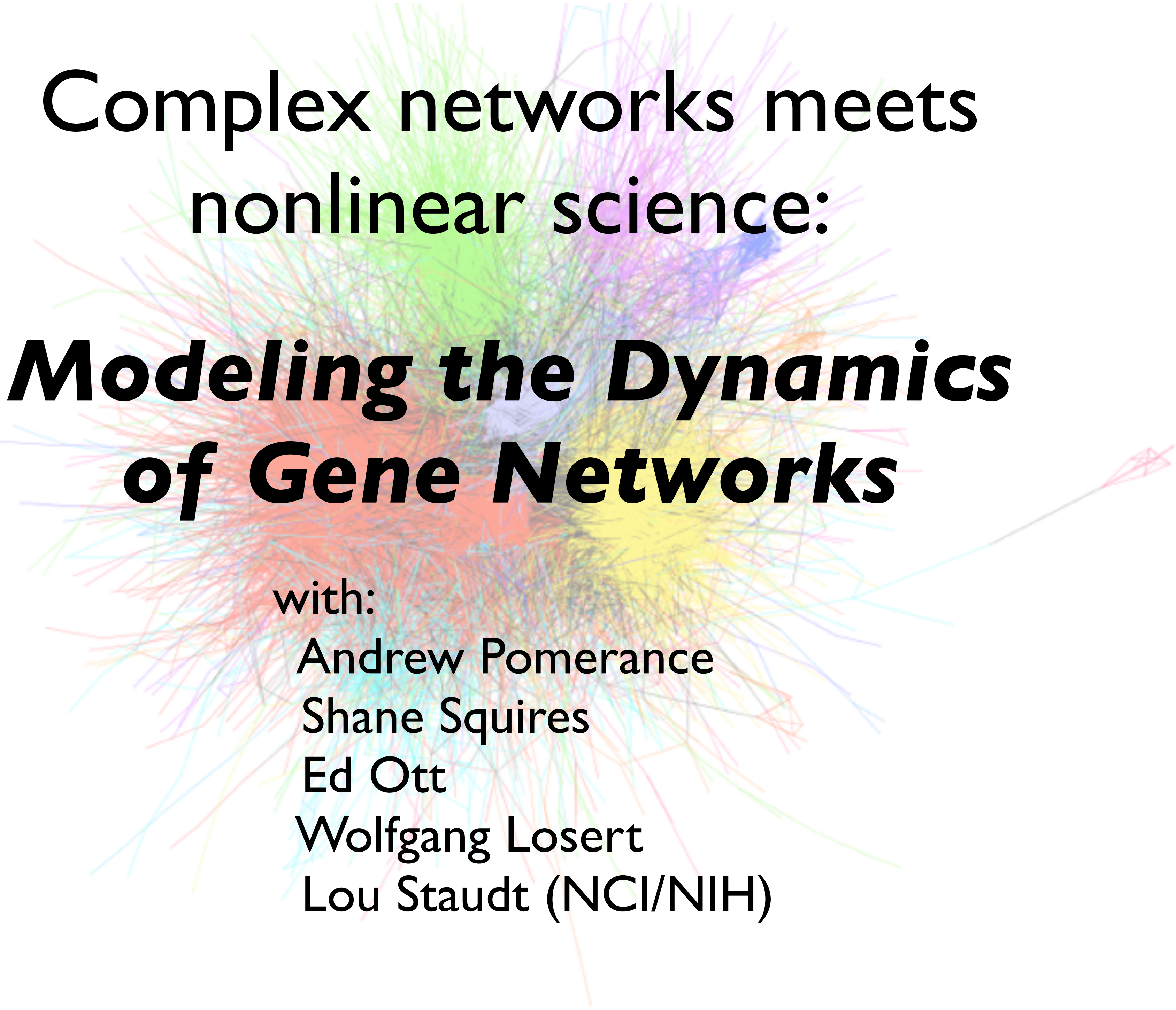
- We consider the resilience of the network to the removal of its vertices (site percolation) or edges (bond percolation).
- As vertices (or edges) are removed from the network, the average path length will increase.
- Ultimately, the giant component will disintegrate.
- Networks vary according to their level of resilience to vertex (or edge) removal.

Robustness and fragility of scale free networks

Mean vertex–vertex distance on a graph representation of the Internet at the autonomous system level, as vertices are removed one by one. If vertices are removed in random order (squares), distance increases only very slightly, but if they are removed in order of their degrees, starting with the highest degree vertices(circles), then distance increases sharply. We say the network is resilient to random removal of vertices, but sensitive to targeted removal.



R. Albert, H. Jeong, and A.-L. Barabasi, *Attack and error tolerance of complex networks*, Nature, 406 (2000), pp. 378–382.



Complex networks meets nonlinear science:

Modeling the Dynamics of Gene Networks

with:

Andrew Pomerance
Shane Squires
Ed Ott
Wolfgang Losert
Lou Staudt (NCI/NIH)

Overview

- **The goal:** To gain insights into the complex process of gene regulation.
- **The approach:** Considering a simple model of genetic control, we explore the effects of network topology.
- **The application:** We hypothesize that a dynamical instability in the gene network may be a causal mechanism contributing to the occurrence of some cancers.

A complex web of interactions in transcriptional regulation

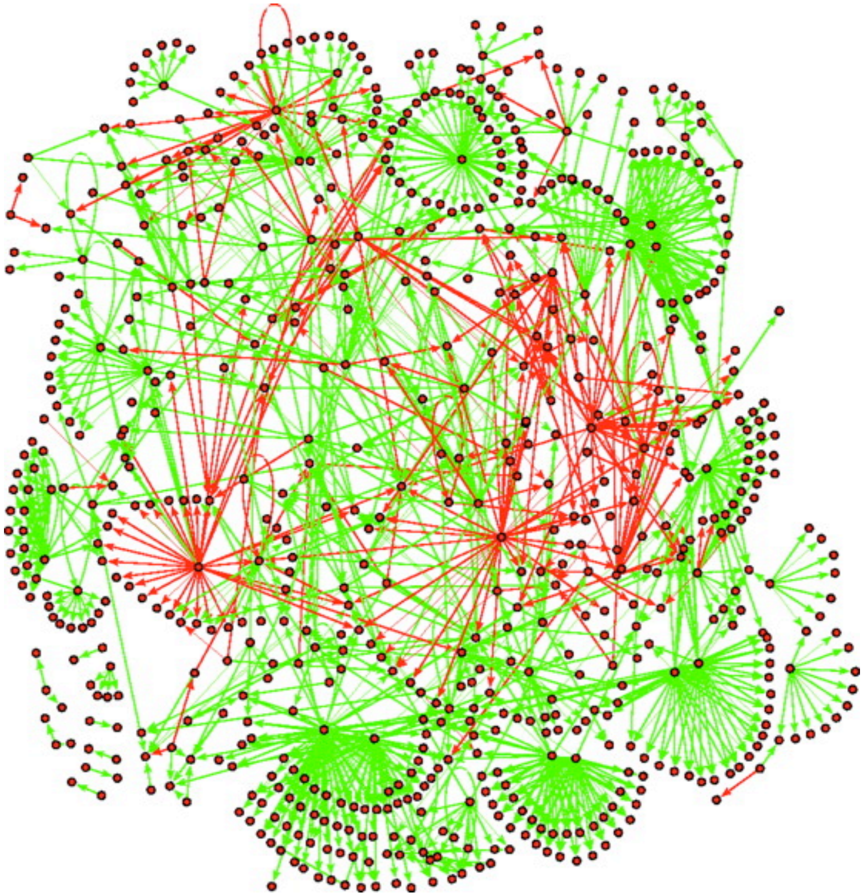


Figure taken from http://rsif.royalsocietypublishing.org/content/5/Suppl_1/S85.full

Modeling Gene Networks: The Boolean Approach

Kauffman's *N-K* model:

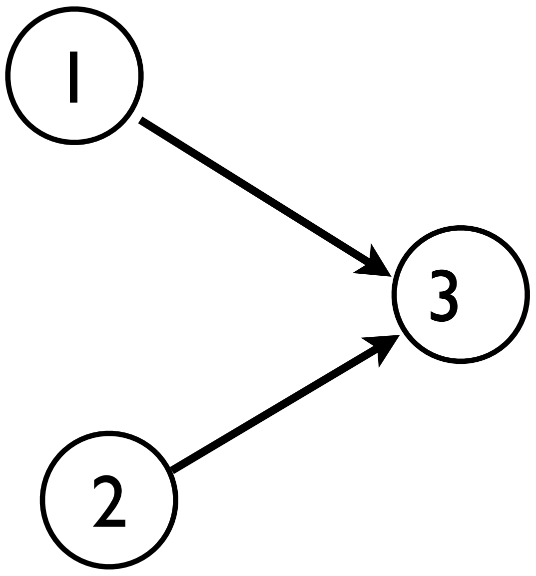
- *N* Genes on or off
- Each gene has exactly *K* inputs, which are randomly chosen
- Discrete updates
- Evolves by a random update function at each node

Our work:

- Focuses on stability of these systems in response to small perturbations
- Explores the effect of network topology on stability
- Explores more realistic update functions

Local update rules: An example

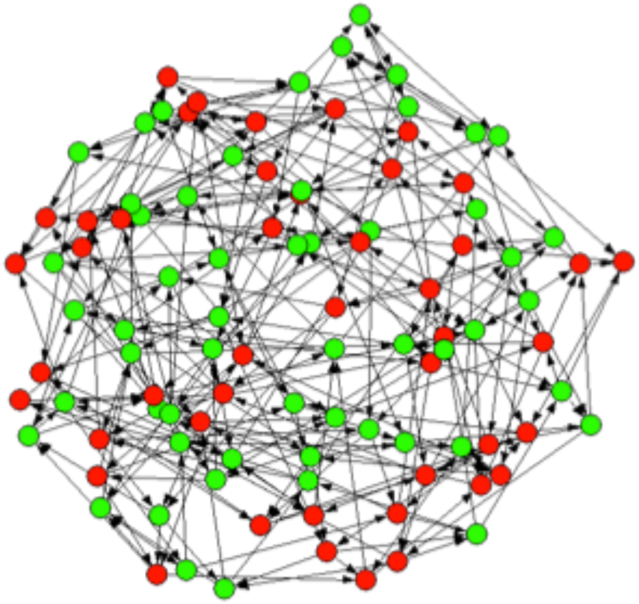
current state time t		State of gene 3 at t+1
Gene 1	Gene 2	
0	0	0
0	1	0
1	0	1
1	1	0

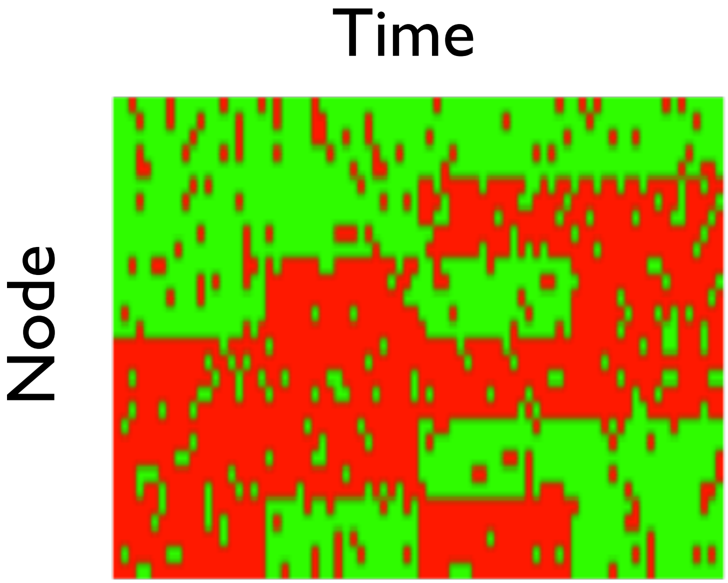


Node with 2 inputs

Output column filled in randomly with bias (probability of 0), p_i

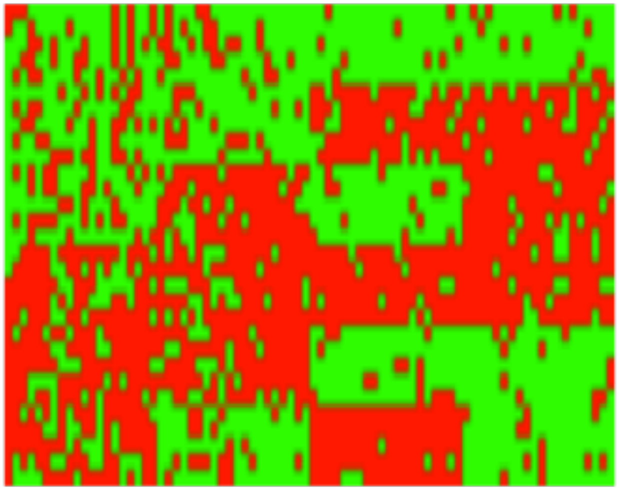
Local Rules Lead to Global Patterns



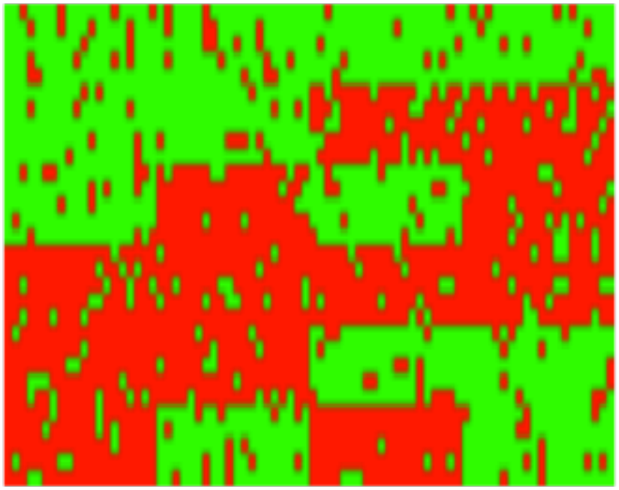


Is the network stable or chaotic?

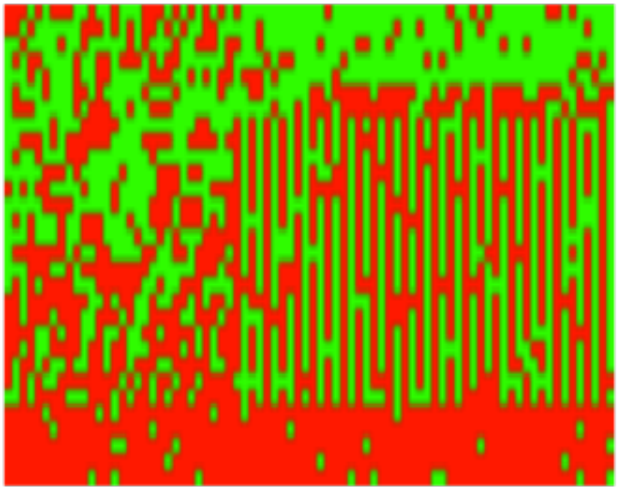
Flip the states of a few genes. Do we see the same pattern as before?



pattern in stable network

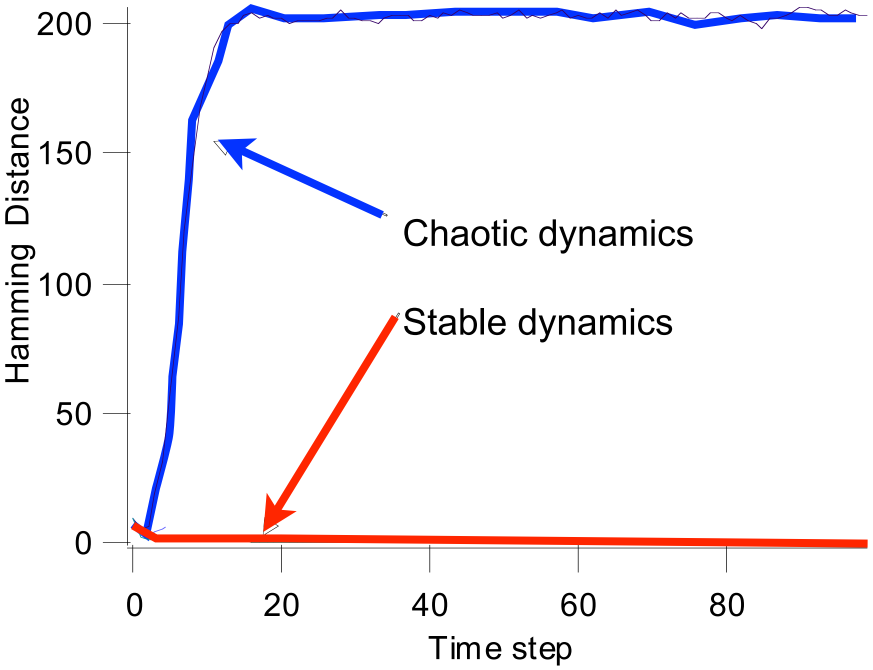


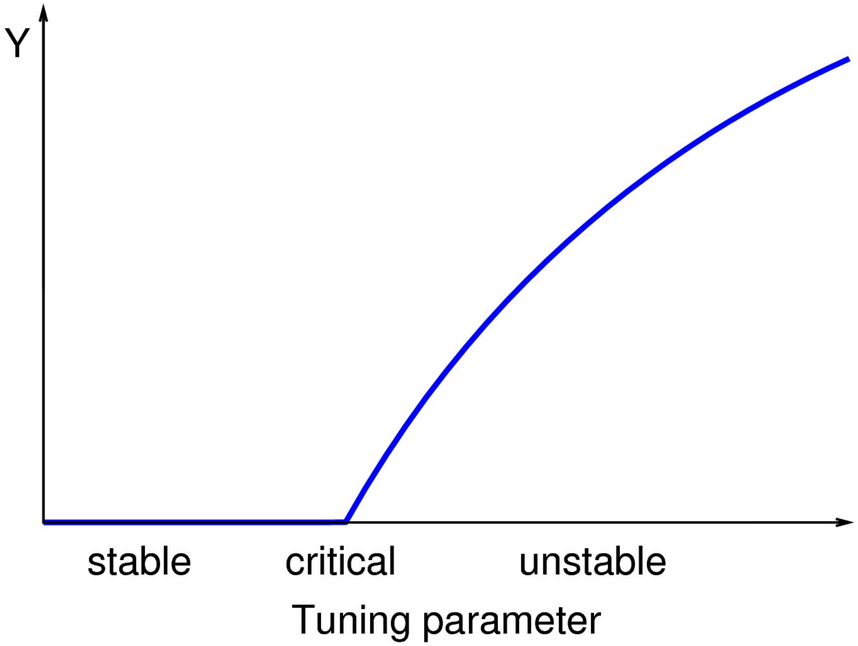
original pattern



pattern in chaotic network

Chaotic and stable dynamics for different networks



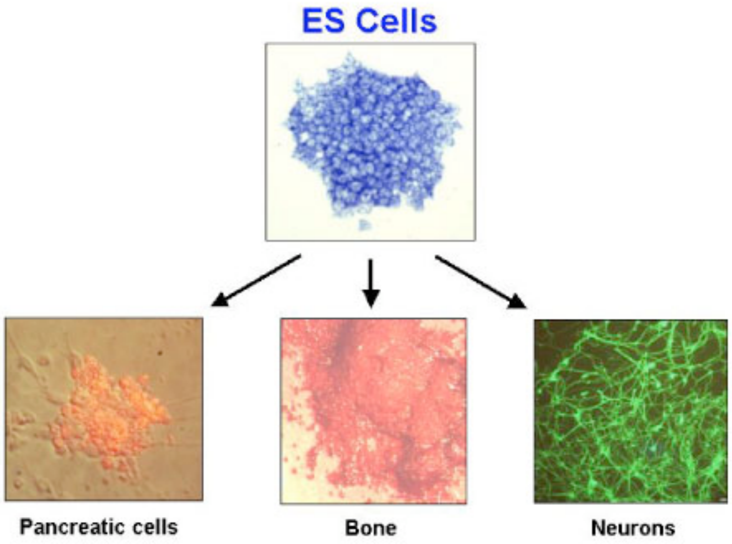


Significance of the patterns

- The patterns of activity may define a cell's character

- In single celled organisms this could correspond to different cell states: growing, dividing, starving, etc.

- In multicellular organisms these could correspond to different cell types.



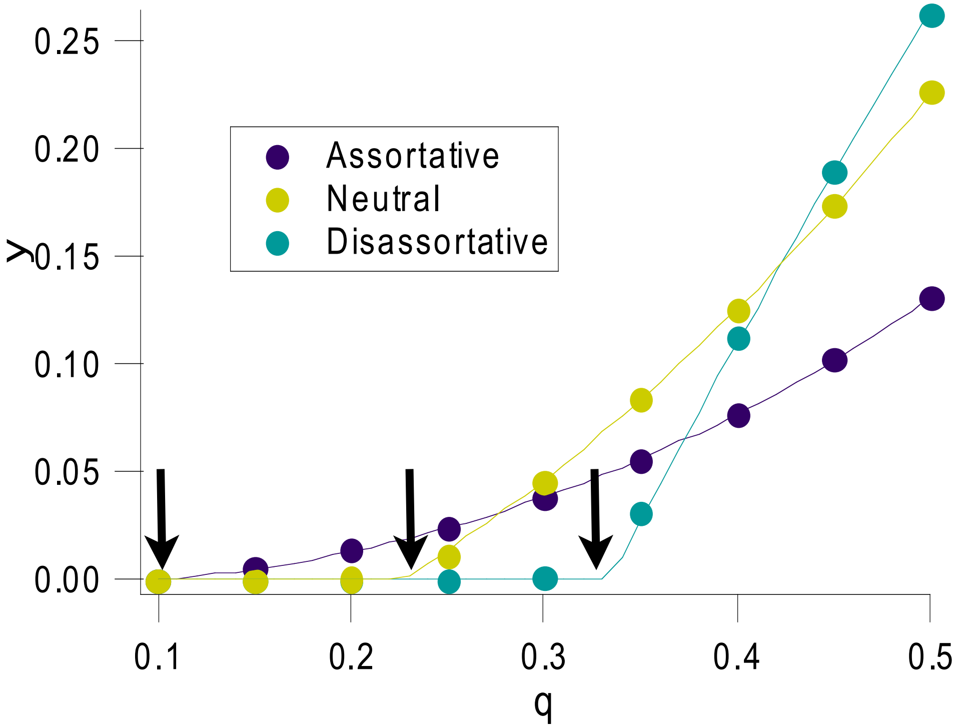
Motivation for our work

- Real networks are far from the idealized models studied previously
- We would like to be able to analyze any fixed network, and we are interested in the effects of:
    - Assortativity
    - Community structure
    - Network motifs
    - Heterogeneous gene sensitivities

What we can calculate with our model

Given an arbitrary network of connectivity, and a distribution for the sensitivities of the genes (and otherwise random update rules), we can predict whether we will see chaotic or stable dynamics.

We have also extended our approach to handle more realistic update rules.



Here we vary the average sensitivity for three different networks. We plot the average distance between initially close states.

Semi-annealed analysis

- Consider two state vectors, $\underline{\sigma}(t)$ and $\tilde{\underline{\sigma}}(t)$, that have evolved from slightly different initial conditions

- Let $y_i(t) =$ the probability that $\sigma_i(t)$ and $\tilde{\sigma}_i(t)$ differ

- Let $q_i =$ the probability that $\sigma_i(t)$ and $\tilde{\sigma}_i(t)$ differ, given a difference in the states of the inputs to i at time $t - 1$

$$q_i = 1 - [p_i^2 + (1 - p_i)^2] = 2p_i(1 - p_i)$$

Update equation for $y_i(t)$

$$y_i(t) = q_i \left\{ \overbrace{1 - \prod_{j\,,\,A_{ij}=1} \underbrace{\left[1 - y_j(t-1)\right]}_{\text{Probability that the input from node } j \text{ is the same}}}^{\text{Probability that the inputs at } t\text{-}1 \text{ to } i \text{ are not all the same}} \right\}$$

Perturb around $\underline{\sigma} = \underline{\tilde{\sigma}}$ $(y_i \ll 1)$, linearization gives:

$$y_i(t) \cong q_i \sum_{j=1}^{N} A_{ij} y_j(t-1) = \sum_{j=1}^{N} Q_{ij} y_j(t-1)$$

where the $Q_{ij} = q_i A_{ij}$ are the elements of a modified adjacency matrix

Stability Criterion

$$\underline{y}(t) = \underline{\underline{Q}}\,\underline{y}(t-1)$$

λ_Q is the largest eigenvalue of Q, which, according to the Perron-Frobenius theorem is real and positive ($Q_{ij} \geq 0$).

Stability Conditions:

If $\lambda_Q < 1$: stable
If $\lambda_Q > 1$: unstable
If $\lambda_Q = 1$: "edge of chaos"

Numerical tests

We numerically test the predictions of

- λ_Q stability criterion
- Saturated normalized Hamming distance between $\underline{\sigma}$ and $\underline{\tilde{\sigma}}$:

$$\overline{y} = \lim_{t \to \infty} \frac{1}{N} \sum_i y_i(t)$$

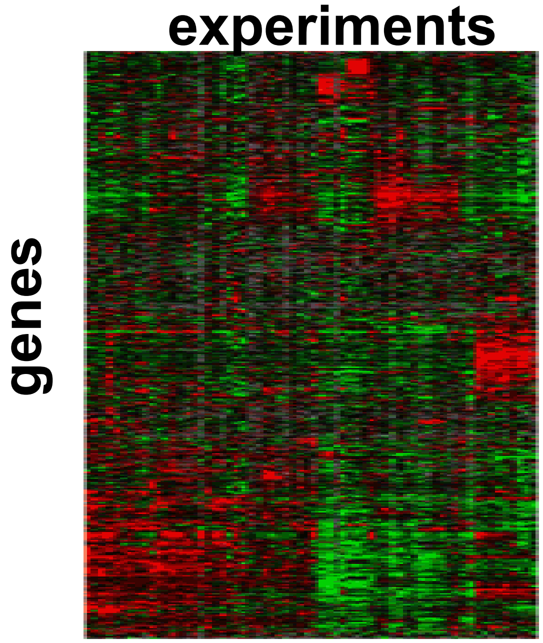
Stability and Cancer

- Gene expression profiles from tumor dissections show that nearby cells have vastly different gene expression profiles.
- Could these fluctuations imply a breakdown of genetic control due to dynamical instability?
- What kind of data do we need to answer these questions?

Elucidating the network and the sensitivities from data

- **Network:** Undirected links can be inferred from data by looking at co-expression patterns across a range of perturbation experiments

- **Sensitivities** can be determined from clinical expression data



Summary and future directions

- Simple Boolean models of genetic control, starting with random Boolean models and progressing to the more realistic Boolean update rules, can be used to gain insights into the effects of network structure in the process of gene regulation.
- A major challenge in this kind of research is to test the model predictions with real data.
- Future directions: This kind of modeling approach may also be useful for studying the evolution of gene regulatory networks. For example, we might study what kinds of networks and truth tables optimize an organism's fitness (which we might think of as some kind of tradeoff between diversity of behaviors and stability) and how might these structures arise through evolutionary processes.

References

A. Pomerance†, E. Ott*, M. Girvan*, and W. Losert*, "The effect of network topology on the stability of discrete state models of genetic control," Proc. Natl. Acad. Sci. USA 106, 8209-8214 (2009).

A. Pomerance*, M. Girvan*, and E. Ott*, "Stability of Boolean networks with generalized canalizing rules," Phys. Rev. E 85, 046106 (2012).

S. Squires, E. Ott, and M. Girvan, "Dynamical instability in Boolean networks as a percolation problem," Phys. Rev. Lett. 109, 085701 (2012)

Concluding Remarks

- The emerging field of network science demonstrates how characterizing complex connectivity patterns can be key to understanding many systems.
- Foundational work in this area gives us insight into the role of network topology in numerous applications.
- Many open questions remain. Areas of active research include:
    - Temporal networks
    - Multiplex networks
    - Uncertainty in networks
- Words of caution when taking a physics approach to understanding complex networks:
    - Stay up-to-date on network science advances coming from other disciplines
    - Simple models are appealing and can help us gain insights into complex systems, but we need to be careful that our assumptions are reasonable and our conclusions are not overstated.