Complex (Biological) Networks

**Today:** *Measuring Network Topology*

**Thursday:** *Analyzing Metabolic Networks*

Elhanan Borenstein

Spring 2011

Some slides are based on slides from courses given by Roded Sharan and Tomer Shlomi
Measuring Network Topology

- Introduction to network theory
- Global Measures of Network Topology
  - Degree Distribution
  - Clustering Coefficient
  - Average Distance
- Random Network Models
- Network Motifs
What is a Network?

- A collection of **nodes** and **links** (edges)
- A map of interactions or relationships
What is a Network?

- A collection of **nodes** and **links** (**edges**)
- A map of interactions or relationships
Networks vs. Graphs

- **Graph Theory**
- Definition of a graph: \( G=(V,E) \)
  - \( V \) is the set of nodes/vertices (elements)
  - \(|V|=N\)
  - \( E \) is the set of edges (relations)

- One of the most well studied objects in CS
  - Subgraph finding (e.g., **clique**, **spanning tree**) and alignment
  - Graph **coloring** and graph **covering**
  - Route finding (**Hamiltonian** path, traveling salesman, etc.)

- Many problems are proven to be **NP-complete**
## Networks vs. Graphs

<table>
<thead>
<tr>
<th>Network theory</th>
<th>Graph theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social sciences</td>
<td>Computer science</td>
</tr>
<tr>
<td>Biological sciences</td>
<td>Since 18\textsuperscript{th} century!!!</td>
</tr>
<tr>
<td>Mostly 20\textsuperscript{th} century</td>
<td>Modeling abstract systems</td>
</tr>
<tr>
<td>Modeling real-life systems</td>
<td>Solving “graph-related” questions</td>
</tr>
<tr>
<td>Measuring structure &amp; topology</td>
<td></td>
</tr>
</tbody>
</table>
Why Networks?

Networks as models

- Simple, visual representation of complex systems
- Focus on organization (rather than on components)
- Discovery (topology affects function)
- Predictive models
- Diffusion models (dynamics)

Networks as tools

- Problem representation (more common than you think)
- Algorithm development
The Seven Bridges of Königsberg

- Published by **Leonhard Euler, 1736**
- Considered the first paper in graph theory
Types of Graphs/Networks

- **Edges:**
  - Directed/undirected
  - Weighted/non-weighted
  - Simple-edges/Hyperedges

- **Special topologies:**
  - Directed Acyclic Graphs (DAG)
  - Trees
  - Bipartite networks
Computational Representation of Networks

- List/set of edges: (ordered) pairs of nodes
  \{ (A, C), (C, B), (D, B), (D, C) \}

- Connectivity Matrix

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

- Object Oriented

- Which is the most useful representation?
Networks in Biology

- Molecular networks:
  - Protein-Protein Interaction (PPI) networks
  - Metabolic Networks
  - Regulatory Networks
  - Synthetic lethality Networks
  - Gene Interaction Networks
  - Many more ...
Metabolic Networks

- Reflect the set of biochemical reactions in a cell
  - **Nodes**: metabolites
  - **Edges**: biochemical reactions
  - **Additional representations**!
- Derived through:
  - Knowledge of biochemistry
  - Metabolic flux measurements

*S. Cerevisiae*
1062 metabolites
1149 reactions
Protein-Protein Interaction (PPI) Networks

- Reflect the cell’s molecular interactions and signaling pathways (interactome)
  - **Nodes**: proteins
  - **Edges**: interactions (?)

- High-throughput experiments:
  - Protein Complex-IP (Co-IP)
  - Yeast two-hybrid
  - Computationally

*S. Cerevisiae*
4389 proteins
14319 interactions
Transcriptional Regulatory Network

- Reflect the cell’s genetic regulatory circuitry
  - **Nodes**: transcription factors (TFs) and genes;
  - **Edges**: from TF to the genes it regulates; Directed; weighted?; “almost” bipartite

- Derived through:
  - Chromatin IP
  - Microarrays
  - Computationally
Other Networks in Biology/Medicine
Non-Biological Networks

- **Computer related networks:**
  - WWW; Internet backbone
  - Communication and IP

- **Social networks:**
  - Friendship (facebook; clubs)
  - Citations / information flow
  - Co-authorships (papers); Co-occurrence (movies; Jazz)

- **Transportation:**
  - Highway system; Airline routes

- **Electronic/Logic circuits**

- **Many more...**
Global Measures of Network Topology
Comparing networks

- We want to find a way to “compare” networks.
  - “Similar” (not identical) topology
  - Common design principles

- We seek measures of network topology that are:
  - Simple
  - Capture **global** organization
  - Potentially “important”
  (equivalent to, for example, GC content for genomes)
Node Degree / Rank

- Degree = Number of neighbors

- Node degree in PPI networks correlates with:
  - Gene essentiality
  - Conservation rate
  - Likelihood to cause human disease

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.
Degree Distribution

- Degree distribution $P(k)$: probability that a node has a degree of exactly $k$

- Common distributions:
  - **Poisson:**
    \[ P(k) = \frac{e^{-d} d^k}{k!} \]
  - **Exponential:**
    \[ P(k) \propto e^{-k/d} \]
  - **Power-law:**
    \[ P(k) \propto k^{-c}, k \neq 0, c > 1 \]
The Power-Law Distribution

- **Power-law distribution has a heavy ("fat") tail!**
  - Characterized by a small number of highly connected nodes, known as **hubs**
  - A.k.a. "**scale-free**" network

- **Hubs are crucial:**
  - ‘party’ hubs and ‘date’ hubs
The Internet

- **Nodes** – 150,000 routers
- **Edges** – physical links
- \( P(k) \sim k^{-2.3} \)

Govindan and Tangmunarunkit, 2000
Movie Actor Collaboration Network

- **Nodes** – 212,250 actors
- **Edges** – co-appearance in a movie
- \( <k> = 28.78 \)
- \( P(k) \sim k^{-2.3} \)

Barabasi and Albert, Science, 1999
Protein Interaction Networks

- **Nodes** – Proteins
- **Edges** – Interactions (yeast)

\[ P(k) \sim k^{-2.5} \]
Metabolic Networks

- **Nodes** – Metabolites
- **Edges** – Reactions
- \( P(k) \sim k^{-2.2 \pm 2} \)

Metabolic networks across all kingdoms of life are scale-free

A. Fulgidus (archae)

E. Coli (bacterium)

C. Elegans (eukaryote)

Averaged (43 organisms)

Jeong et al., Nature, 2000
Network Clustering

Costanzo et al., Nature, 2010
Clustering Coefficient (Watts & Strogatz)

- Characterizes tendency of nodes to cluster
  - “triangles density”
  - How often do my friends know each other (think “facebook”)

\[
C_i = \frac{\text{# of edges among neighbors}}{\text{Max. possible # of edges among neighbors}} = \frac{2E_i}{d_i(d_i - 1)}
\]

\[
C = \frac{1}{N} \sum_{v} C_i
\]

(if \(d_i = 0\) or \(1\) then \(C_i\) is defined to be \(0\))
Clustering Coefficient: Example

- Lies in \([0,1]\)
  - For cliques: \(C=1\)
  - For triangle-free graphs: \(C=0\)

\[
\begin{align*}
C_i &= \frac{10}{10} = 1 \\
C_i &= \frac{3}{10} = 0.3 \\
C_i &= \frac{0}{10} = 0
\end{align*}
\]
Average Distance

- **Distance:**
  
  Length of shortest (geodesic) path between two nodes

- **Average distance:**
  
  average over all connected pairs

\[ \ell = \frac{1}{\frac{1}{2}n(n + 1)} \sum_{i \geq j} d_{ij} \]
Small World Networks

- Despite their often large size, in most (real) networks there is a relatively short path between any two nodes

- “Six degrees of separation”
  (Stanley Milgram; 1967)

- Collaborative distance:
  - Erdös number
  - Bacon number

- Danica McKellar: 6
- Natalie Portman: 6
- Daniel Kleitman: 3
# Network Structure in Real Networks

<table>
<thead>
<tr>
<th>Network</th>
<th>Size</th>
<th>$\langle k \rangle$</th>
<th>$\ell$</th>
<th>$\ell_{\text{rand}}$</th>
<th>$C$</th>
<th>$C_{\text{rand}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WWW, site level, undir.</td>
<td>153,127</td>
<td>35.21</td>
<td>3.1</td>
<td>3.35</td>
<td>0.1078</td>
<td>0.00023</td>
</tr>
<tr>
<td>Internet, domain level</td>
<td>3015 - 6209</td>
<td>3.52 - 4.11</td>
<td>3.7 - 3.76</td>
<td>6.36 - 6.18</td>
<td>0.18 - 0.3</td>
<td>0.001</td>
</tr>
<tr>
<td>Movie actors</td>
<td>225,226</td>
<td>61</td>
<td>3.65</td>
<td>2.99</td>
<td>0.79</td>
<td>0.00027</td>
</tr>
<tr>
<td>LANL coauthorship</td>
<td>52,909</td>
<td>9.7</td>
<td>5.9</td>
<td>4.79</td>
<td>0.43</td>
<td>$1.8 \times 10^{-4}$</td>
</tr>
<tr>
<td>MEDLINE coauthorship</td>
<td>1,520,251</td>
<td>18.1</td>
<td>4.6</td>
<td>4.91</td>
<td>0.066</td>
<td>$1.1 \times 10^{-5}$</td>
</tr>
<tr>
<td>SPIRES coauthorship</td>
<td>56,627</td>
<td>173</td>
<td>4.0</td>
<td>2.12</td>
<td>0.726</td>
<td>0.003</td>
</tr>
<tr>
<td>NCSTRL coauthorship</td>
<td>11,994</td>
<td>3.59</td>
<td>9.7</td>
<td>7.34</td>
<td>0.496</td>
<td>$3 \times 10^{-4}$</td>
</tr>
<tr>
<td>Math coauthorship</td>
<td>70,975</td>
<td>3.9</td>
<td>9.5</td>
<td>8.2</td>
<td>0.59</td>
<td>$5.4 \times 10^{-5}$</td>
</tr>
<tr>
<td>Neurosci. coauthorship</td>
<td>209,293</td>
<td>11.5</td>
<td>6</td>
<td>5.01</td>
<td>0.76</td>
<td>$5.5 \times 10^{-5}$</td>
</tr>
<tr>
<td><em>E. coli</em>, substrate graph</td>
<td>282</td>
<td>7.35</td>
<td>2.9</td>
<td>3.04</td>
<td>0.32</td>
<td>0.026</td>
</tr>
<tr>
<td><em>E. coli</em>, reaction graph</td>
<td>315</td>
<td>28.3</td>
<td>2.62</td>
<td>1.98</td>
<td>0.59</td>
<td>0.09</td>
</tr>
<tr>
<td>Ythan estuary food web</td>
<td>134</td>
<td>8.7</td>
<td>2.43</td>
<td>2.26</td>
<td>0.22</td>
<td>0.06</td>
</tr>
<tr>
<td>Silwood park food web</td>
<td>154</td>
<td>4.75</td>
<td>3.40</td>
<td>3.23</td>
<td>0.15</td>
<td>0.03</td>
</tr>
<tr>
<td>Words, cooccurrence</td>
<td>460,902</td>
<td>70.13</td>
<td>2.67</td>
<td>3.03</td>
<td>0.437</td>
<td>0.0001</td>
</tr>
<tr>
<td>Words, synonyms</td>
<td>22,311</td>
<td>13.48</td>
<td>4.5</td>
<td>3.84</td>
<td>0.7</td>
<td>0.0006</td>
</tr>
<tr>
<td>Power grid</td>
<td>4,941</td>
<td>2.67</td>
<td>18.7</td>
<td>12.4</td>
<td>0.08</td>
<td>0.005</td>
</tr>
<tr>
<td><em>C. Elegans</em></td>
<td>282</td>
<td>14</td>
<td>2.65</td>
<td>2.25</td>
<td>0.28</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Additional Measures

- Network Modularity
- Giant component
- Betweenness centrality
- Current information flow
- Bridging centrality
- Spectral density
Random Network Models

1. Random Graphs (Erdös/Rényi)
2. Geometric Random Graphs
3. The Small World Model (WS)
4. Preferential Attachment
Random Graphs (Erdös/Rényi)

- $N$ nodes
- Every pair of nodes is connected with probability $p$
Random Graphs: Properties

- **Mean degree:** $d = (N-1)p \sim Np$

- **Degree distribution is binomial**
  - Asymptotically Poisson: $P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx \frac{d^k e^{-d}}{k!}$

- **Clustering Coefficient:**
  - The probability of connecting two nodes at random is $p$
  - $\rightarrow$ Clustering coefficient is $C = p$
  - In many large networks $p \sim 1/n \rightarrow C$ is lower than observed

- **Average distance:**
  - $l \sim \ln(N)/\ln(d)$ .... (think why?)
  - Small world! (and fast spread of information)
Geometric Random Graphs

- $G=(V, r)$
  - $V$ – set of points in a metric space (e.g. 2D)
  - $E$ – all pairs of points with distance $\leq r$
- Captures spatial relationships
The Small World Model (WS)

- Generate graphs with **high clustering coefficients** $C$ and **small distance** $l$
- Rooted in social systems

1. **Start with order** (every node is connected to its $K$ neighbors)
2. **Randomize** (rewire each edge with probability $p$)

- Degree distribution is similar to that of a random graph!

The Scale Free Model: Preferential Attachment

- A generative model (dynamics)
  - Growth: degree-m nodes are constantly added
  - *Preferential attachment*: the probability that a new node connects to an existing one is proportional to its degree

\[
P(u \text{ connects to } v) = \frac{d(v)}{\sum d(v)}
\]

- “The rich get richer” principle

\[
P(k) = \frac{2m(m+1)}{(k+2)(k+1)k} \sim k^{-3}
\]

Albert and Barabasi, 2002
Preferential Attachment: Clustering Coefficient

$C \sim N^{-0.1}$

$C \sim N^{-0.75}$

![Graph showing the clustering coefficient $C$ as a function of $N$. The graph includes data points for a scale-free model and a random graph, with the clustering coefficient decreasing as $N$ increases.](image)
Preferential Attachment: Empirical Evidence

- Highly connected proteins in a PPI network are more likely to evolve new interactions

Model Problems

- Degree distribution is fixed
  (although there are generalizations of this method that handle various distributions)
- Clustering coefficient approaches 0 with network size, unlike real networks
- Issues involving \textit{biological} network growth:
  - Ignores local events shaping real networks (e.g., insertions/deletions of edges)
  - Ignores growth constraints (e.g., max degree) and aging (a node is active in a limited period)
Conclusions

- No single best model!
  - Models differ in various network measures
  - Different models capture different attributes of real networks

- In literature, “random graphs” and “generalized random graphs” are most commonly used
Network Motifs
Network Motifs

- Going beyond degree distribution ...
- Generalization of sequence motifs
- Basic building blocks
- Evolutionary design principles

What are Network Motifs?

- Recurring patterns of interactions (*subgraphs*) that are significantly overrepresented (w.r.t. a background model)

13 possible 3-nodes subgraphs
(199 possible 4-node subgraphs)

Finding motifs in the Network

1. Generate randomized networks

2a. Scan for all n-node subgraphs in the real network

2b. Record number of appearances of each subgraph (consider isomorphic architectures)

3a. Scan for all n-node sub graphs in random networks

3b. Record number of appearances of each subgraph

4. Compare each subgraph’s data and choose motifs
Finding motifs in the Network

A. real network

B. randomized networks

motif:
Network Randomization

- How should the set of random networks be generated?
- Do we really want “completely random” networks?
- What constitutes a good null model?

Preserve in- and out-degree
(For motifs with n>3 also preserve distribution of smaller sub-motifs)
Generation of Randomized Networks

- Algorithm A (Markov-chain algorithm):
  - Start with the real network and repeatedly swap randomly chosen pairs of connections (X1→Y1, X2→Y2 is replaced by X1→Y2, X2→Y1)
  - Repeat until the network is well randomized
  - Switching is prohibited if the either of the connections X1→Y2 or X2→Y1 already exist
Generation of Randomized Networks

- Algorithm B (Generative):
  - Record marginal weights of original network
  - Start with an empty connectivity matrix $M$
  - Choose a row $n$ & a column $m$ according to marginal weights
  - If $M_{nm} = 0$, set $M_{nm} = 1$; Update marginal weights
  - Repeat until all marginal weights are 0
  - If no solution is found, start from scratch
Exact Criteria for Network Motifs

- Subgraphs that meet the following criteria:

1. The probability that it appears in a randomized network an equal or greater number of times than in the real network is smaller than $P = 0.01$

2. The number of times it appears in the real network with distinct sets of nodes is at least 4

3. The number of appearances in the real network is significantly larger than in the randomized networks: ($N_{\text{real}} - N_{\text{rand}} > 0.1N_{\text{rand}}$)
Feed-Forward Loops in Transcriptional Regulatory Networks

- E. Coli network
  - 424 operons (116 TFs)
  - 577 interactions
  - Significant enrichment of motif # 5
    (40 instances vs. 7±3)

- Coherent FFLs:
  - The direct effect of x on z has the same sign as the net indirect effect through y
  - 85% of FFLs are coherent

S. Shen-Orr et al. Nature Genetics 2002
A coherent feed-forward loop can act as a circuit that rejects transient activation signals from the general transcription factor and responds only to persistent signals, while allowing a rapid system shutdown.

Boolean Kinetics

\[
\frac{dY}{dt} = F(X, T_Y) - aY
\]

\[
\frac{dZ}{dt} = F(X, T_Y)F(Y, T_Z) - aZ
\]
## Network Motifs in Biological Networks

<table>
<thead>
<tr>
<th>Network</th>
<th>Nodes</th>
<th>Edges</th>
<th>(N_{\text{real}})</th>
<th>(N_{\text{rand}} \pm \text{SD})</th>
<th>Z score</th>
<th>(N_{\text{real}})</th>
<th>(N_{\text{rand}} \pm \text{SD})</th>
<th>Z score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene regulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(transcription)</td>
<td>424</td>
<td>519</td>
<td>40</td>
<td>7 ± 3</td>
<td>10</td>
<td>205</td>
<td>47 ± 3</td>
<td>13</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td></td>
<td></td>
<td>70</td>
<td>11 ± 4</td>
<td>14</td>
<td>1812</td>
<td>300 ± 40</td>
<td>41</td>
</tr>
<tr>
<td><em>S. cerevisiae</em></td>
<td>685</td>
<td>1,052</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neurons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>125</td>
<td>96 ± 10</td>
<td>3.7</td>
<td>127</td>
<td>55 ± 15</td>
<td>5.3</td>
</tr>
<tr>
<td><em>C. elegans</em>†</td>
<td>252</td>
<td>509</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Food webs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little Rock</td>
<td>92</td>
<td>984</td>
<td>3219</td>
<td>3120 ± 50</td>
<td>2.1</td>
<td>7235</td>
<td>2220 ± 210</td>
<td>25</td>
</tr>
<tr>
<td>Ythan</td>
<td>83</td>
<td>391</td>
<td>1182</td>
<td>1020 ± 20</td>
<td>7.2</td>
<td>1357</td>
<td>230 ± 50</td>
<td>23</td>
</tr>
<tr>
<td>St. Martin</td>
<td>42</td>
<td>205</td>
<td>469</td>
<td>450 ± 10</td>
<td>NS</td>
<td>382</td>
<td>130 ± 20</td>
<td>12</td>
</tr>
<tr>
<td>Chesapeake</td>
<td>31</td>
<td>67</td>
<td>80</td>
<td>82 ± 4</td>
<td>NS</td>
<td>26</td>
<td>5 ± 2</td>
<td>8</td>
</tr>
<tr>
<td>Coachella</td>
<td>29</td>
<td>243</td>
<td>279</td>
<td>235 ± 12</td>
<td>3.6</td>
<td>181</td>
<td>80 ± 20</td>
<td>5</td>
</tr>
<tr>
<td>Skipwith</td>
<td>25</td>
<td>189</td>
<td>184</td>
<td>150 ± 7</td>
<td>5.5</td>
<td>397</td>
<td>80 ± 25</td>
<td>13</td>
</tr>
<tr>
<td>B. Brook</td>
<td>25</td>
<td>104</td>
<td>181</td>
<td>130 ± 7</td>
<td>7.4</td>
<td>267</td>
<td>30 ± 7</td>
<td>32</td>
</tr>
</tbody>
</table>

- **FFL motif is under-represented!**
### Information Flow vs. Energy Flow

<table>
<thead>
<tr>
<th>Network</th>
<th>Nodes</th>
<th>Edges</th>
<th>$N_{\text{real}}$</th>
<th>$N_{\text{rand}} \pm \text{SD}$</th>
<th>Z score</th>
<th>$N_{\text{real}}$</th>
<th>$N_{\text{rand}} \pm \text{SD}$</th>
<th>Z score</th>
<th>$N_{\text{real}}$</th>
<th>$N_{\text{rand}} \pm \text{SD}$</th>
<th>Z score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene regulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(transcription)</td>
<td>424</td>
<td>519</td>
<td>40</td>
<td>7 ± 3</td>
<td>10</td>
<td>203</td>
<td>47 ± 12</td>
<td>13</td>
<td>1812</td>
<td>300 ± 40</td>
<td>41</td>
</tr>
<tr>
<td>E. coli</td>
<td>685</td>
<td>1,052</td>
<td>70</td>
<td>11 ± 4</td>
<td>14</td>
<td>1812</td>
<td>300 ± 40</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. cerevisiae*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neurons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>252</td>
<td>509</td>
<td>125</td>
<td>90 ± 10</td>
<td>3.7</td>
<td>127</td>
<td>55 ± 13</td>
<td>5.3</td>
<td>227</td>
<td>35 ± 10</td>
<td>20</td>
</tr>
<tr>
<td>C. elegans†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Food webs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little Rock</td>
<td>92</td>
<td>984</td>
<td>3219</td>
<td>3120 ± 50</td>
<td>2.1</td>
<td>7295</td>
<td>2220 ± 210</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ythan</td>
<td>83</td>
<td>391</td>
<td>1182</td>
<td>1020 ± 20</td>
<td>7.2</td>
<td>1357</td>
<td>230 ± 50</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>St. Martin</td>
<td>42</td>
<td>205</td>
<td>469</td>
<td>450 ± 10</td>
<td>NS</td>
<td>382</td>
<td>130 ± 20</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chesapeake</td>
<td>31</td>
<td>67</td>
<td>80</td>
<td>82 ± 4</td>
<td>NS</td>
<td>26</td>
<td>5 ± 2</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coachella</td>
<td>29</td>
<td>243</td>
<td>279</td>
<td>235 ± 12</td>
<td>3.6</td>
<td>181</td>
<td>80 ± 20</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skipwith</td>
<td>25</td>
<td>189</td>
<td>184</td>
<td>150 ± 7</td>
<td>5.5</td>
<td>397</td>
<td>80 ± 25</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Brook</td>
<td>25</td>
<td>104</td>
<td>181</td>
<td>130 ± 7</td>
<td>7.4</td>
<td>267</td>
<td>30 ± 7</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FFL motif is under-represented!**
## Network Motifs in Technological Networks

<table>
<thead>
<tr>
<th>Electronic circuits (forward logic chips)</th>
<th>Feed-forward loop</th>
<th>Bi-fan</th>
<th>Bi-parallel</th>
</tr>
</thead>
<tbody>
<tr>
<td>s15850</td>
<td>10,383</td>
<td>14,240</td>
<td>285</td>
</tr>
<tr>
<td>s38584</td>
<td>20,717</td>
<td>34,204</td>
<td>120</td>
</tr>
<tr>
<td>s38417</td>
<td>23,843</td>
<td>33,661</td>
<td>400</td>
</tr>
<tr>
<td>s9234</td>
<td>5,844</td>
<td>8,197</td>
<td>140</td>
</tr>
<tr>
<td>s13207</td>
<td>8,651</td>
<td>11,831</td>
<td>225</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Electronic circuits (digital fractional multipliers)</th>
<th>Three-node feedback loop</th>
<th>Bi-fan</th>
<th>Four-node feedback loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>s208</td>
<td>122</td>
<td>189</td>
<td>9</td>
</tr>
<tr>
<td>s420</td>
<td>252</td>
<td>399</td>
<td>18</td>
</tr>
<tr>
<td>s838‡</td>
<td>512</td>
<td>819</td>
<td>38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>World Wide Web</th>
<th>Feedback with two mutual dyads</th>
<th>Fully connected triad</th>
<th>Uplinked mutual dyad</th>
</tr>
</thead>
<tbody>
<tr>
<td>nd.edu§</td>
<td>325,729</td>
<td>1.46e6</td>
<td>1.1e5</td>
</tr>
</tbody>
</table>
Criticism of the Randomization Approach

- An incomplete null model?
- Local clustering:
  - Neighboring neurons have a greater chance of forming a connection than distant neurons
- Similar motifs are obtained in random graphs devoid of any selection rule
  - Gaussian toy network
  - Preferential-attachment rule

Gaussian “toy network”

Network Comparison:
Motif-Based Network Superfamilies

R. Milo et al. Superfamilies of evolved and designed networks. Science, 2004
### Evolutionary Conservation of Motif Elements

<table>
<thead>
<tr>
<th>#</th>
<th>Motifs</th>
<th>Number of yeast motifs</th>
<th>Natural conservation rate</th>
<th>Random conservation rate</th>
<th>Conservation ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Motif 1" /></td>
<td>9,266</td>
<td>13.67%</td>
<td>4.63%</td>
<td>2.94</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2" alt="Motif 2" /></td>
<td>167,304</td>
<td>4.99%</td>
<td>0.81%</td>
<td>6.15</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3" alt="Motif 3" /></td>
<td>3,846</td>
<td>20.51%</td>
<td>1.01%</td>
<td>20.28</td>
</tr>
<tr>
<td>4</td>
<td><img src="image4" alt="Motif 4" /></td>
<td>3,649,591</td>
<td>0.73%</td>
<td>0.12%</td>
<td>5.87</td>
</tr>
<tr>
<td>5</td>
<td><img src="image5" alt="Motif 5" /></td>
<td>1,763,891</td>
<td>2.64%</td>
<td>0.18%</td>
<td>14.67</td>
</tr>
<tr>
<td>6</td>
<td><img src="image6" alt="Motif 6" /></td>
<td>9,646</td>
<td>6.71%</td>
<td>0.17%</td>
<td>40.44</td>
</tr>
<tr>
<td>7</td>
<td><img src="image7" alt="Motif 7" /></td>
<td>164,075</td>
<td>7.67%</td>
<td>0.17%</td>
<td>45.56</td>
</tr>
<tr>
<td>8</td>
<td><img src="image8" alt="Motif 8" /></td>
<td>12,423</td>
<td>18.68%</td>
<td>0.12%</td>
<td>157.89</td>
</tr>
<tr>
<td>9</td>
<td><img src="image9" alt="Motif 9" /></td>
<td>2,339</td>
<td>32.53%</td>
<td>0.08%</td>
<td>422.78</td>
</tr>
<tr>
<td>10</td>
<td><img src="image10" alt="Motif 10" /></td>
<td>25,749</td>
<td>14.77%</td>
<td>0.05%</td>
<td>279.71</td>
</tr>
<tr>
<td>11</td>
<td><img src="image11" alt="Motif 11" /></td>
<td>1,433</td>
<td>47.24%</td>
<td>0.02%</td>
<td>2,256.67</td>
</tr>
</tbody>
</table>

*Wuchty et al. Nature Genetics, 2003*