

Biological Networks Analysis

Network Motifs

Genome 373

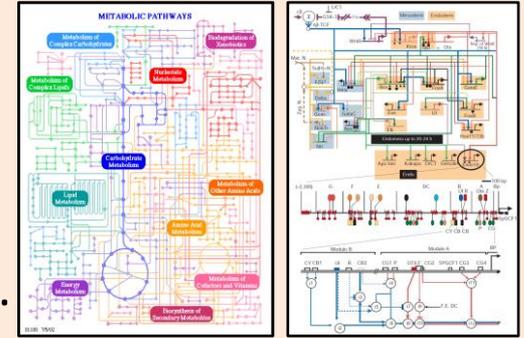
Genomic Informatics

Elhanan Borenstein

A quick review

- **Networks:**

- Networks vs. graphs
- A collection of **nodes** and **links**
- Directed/undirected; weighted/non-weighted, ...
- Networks as models vs. networks as tools

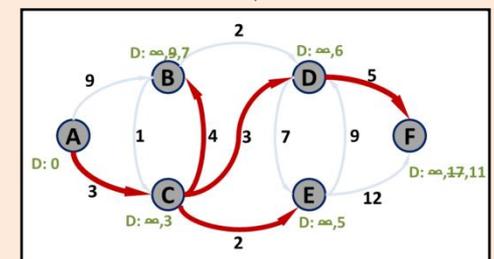
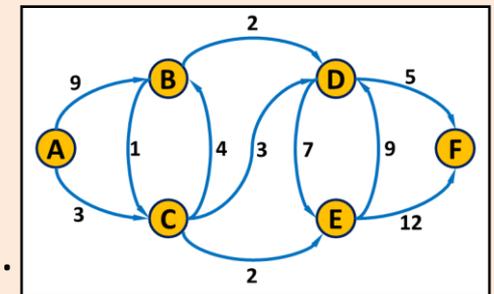


- Many types of biological networks

- The shortest path problem

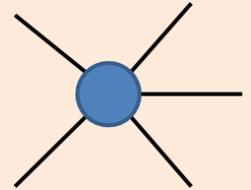
- Dijkstra's algorithm

1. **Initialize:** Assign a distance value, D , to each node. Set $D=0$ for *start* node and to infinity for all others.
2. **For each unvisited neighbor of the current node:** Calculate tentative distance, D^t , through current node and if $D^t < D$: $D \leftarrow D^t$. Mark node as visited.
3. **Continue with the unvisited node with the smallest distance**



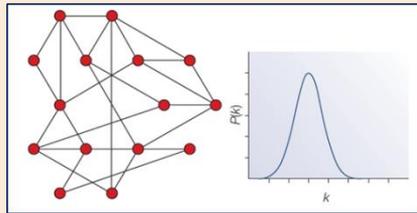
A quick review

- Degree = Number of neighbors
- Degree distribution



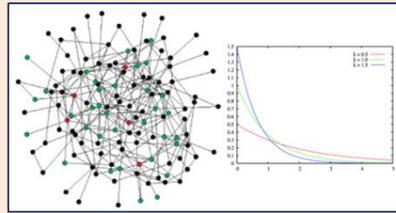
Poisson:

$$P(k) = \frac{e^{-d} d^k}{k!}$$



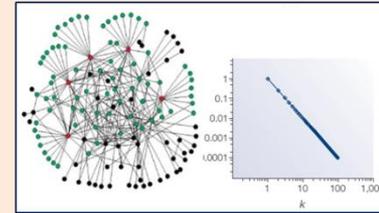
Exponential:

$$P(k) \propto e^{-k/d}$$



Power-law:

$$P(k) \propto k^{-c}, k \neq 0, c > 1$$



- Power-law degree distribution:

- Scale free networks
- Allows **hubs** in the network
- Affects **error** and **attack** tolerance



- Most (all) real-life networks seem to have a power-law degree distribution

Why do so many real-life networks exhibit a power-law degree distribution?

- Is it “selected for”?
- Is it expected by chance?
- Does it have anything to do with the way networks evolve?
- Does it have functional implications?

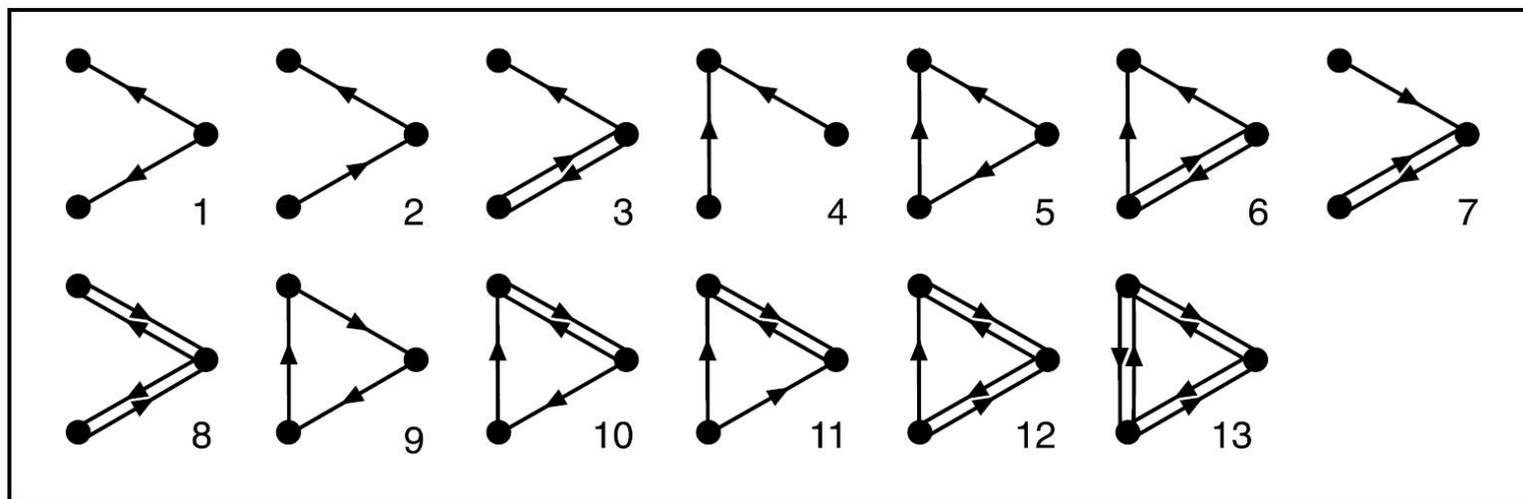


Network motifs

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

What are network motifs?

- Recurring patterns of interaction (*sub-graphs*) that are significantly **overrepresented** (w.r.t. a background model)

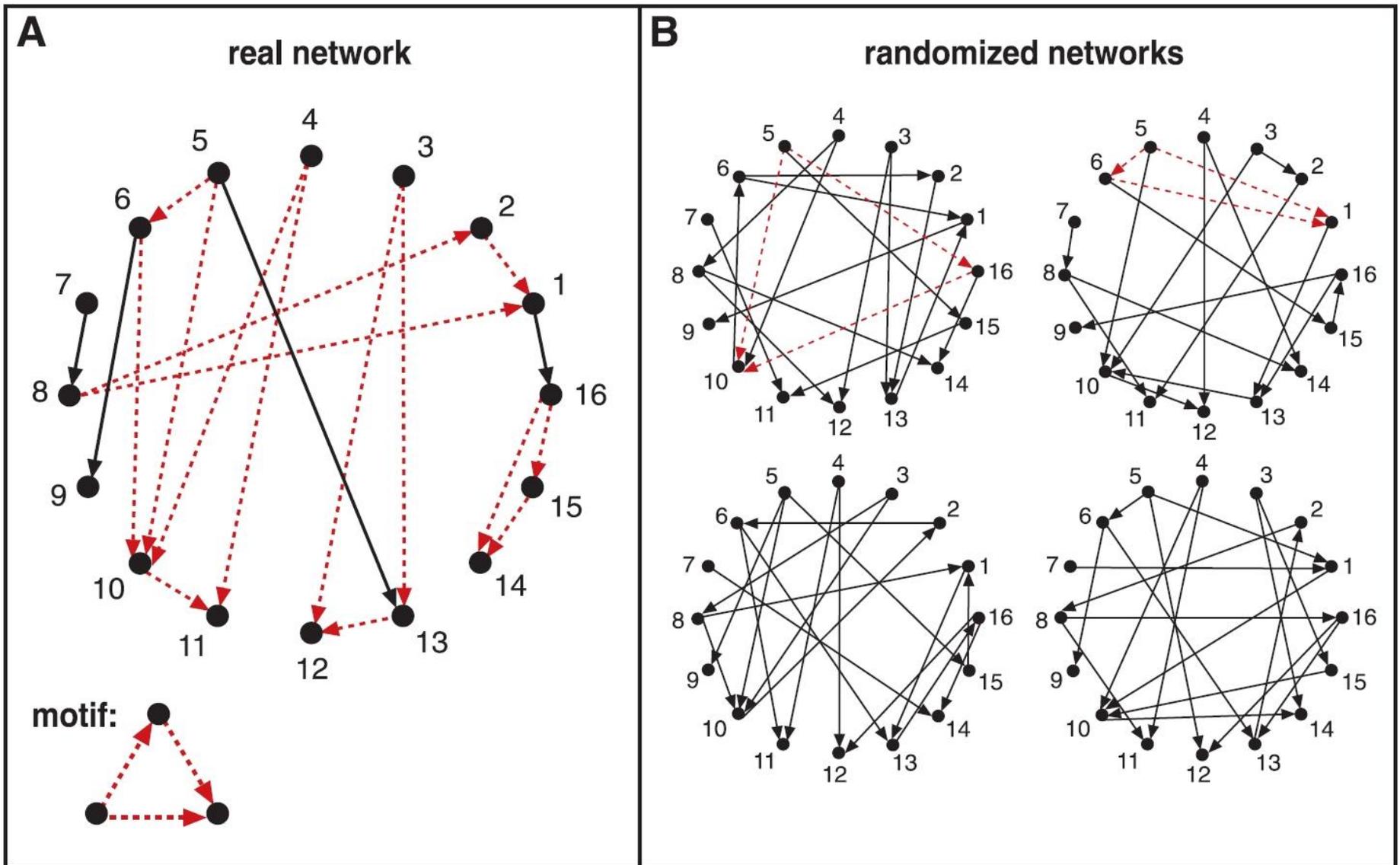


13 possible 3-nodes sub-graphs
(199 possible 4-node sub-graphs)

Finding motifs in the network

- 1a. Scan all n-node sub-graphs in the *real* network
- 1b. Record number of appearances of each sub-graph (*consider isomorphic architectures*)
2. Generate a large set of random networks
- 3a. Scan for all n-node sub-graphs in **random** networks
- 3b. Record number of appearances of each sub-graph
4. Compare each sub-graph's data and identify motifs

Finding motifs in the network

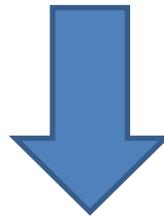


Network randomization

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

Network randomization

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

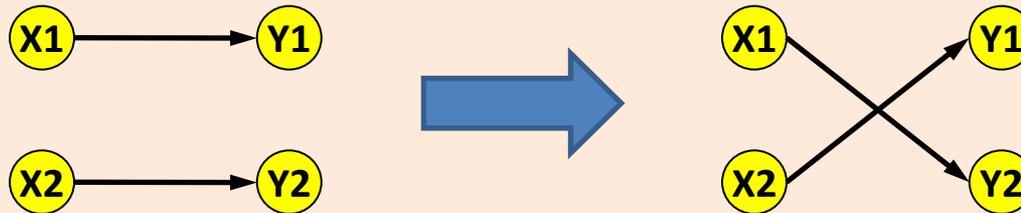


Preserve in- and out-degree

Generation of randomized networks

Network randomization algorithm :

- Start with the real network and repeatedly swap randomly chosen pairs of connections
($X1 \rightarrow Y1, X2 \rightarrow Y2$ is replaced by $X1 \rightarrow Y2, X2 \rightarrow Y1$)

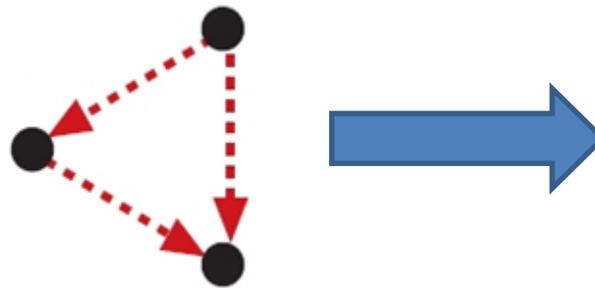


(Switching is prohibited if either of the $X1 \rightarrow Y2$ or $X2 \rightarrow Y1$ already exist)

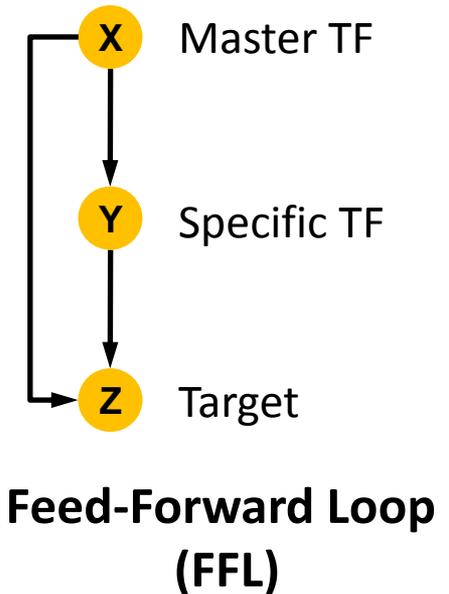
- Repeat until the network is “well randomized”

Motifs in transcriptional regulatory networks

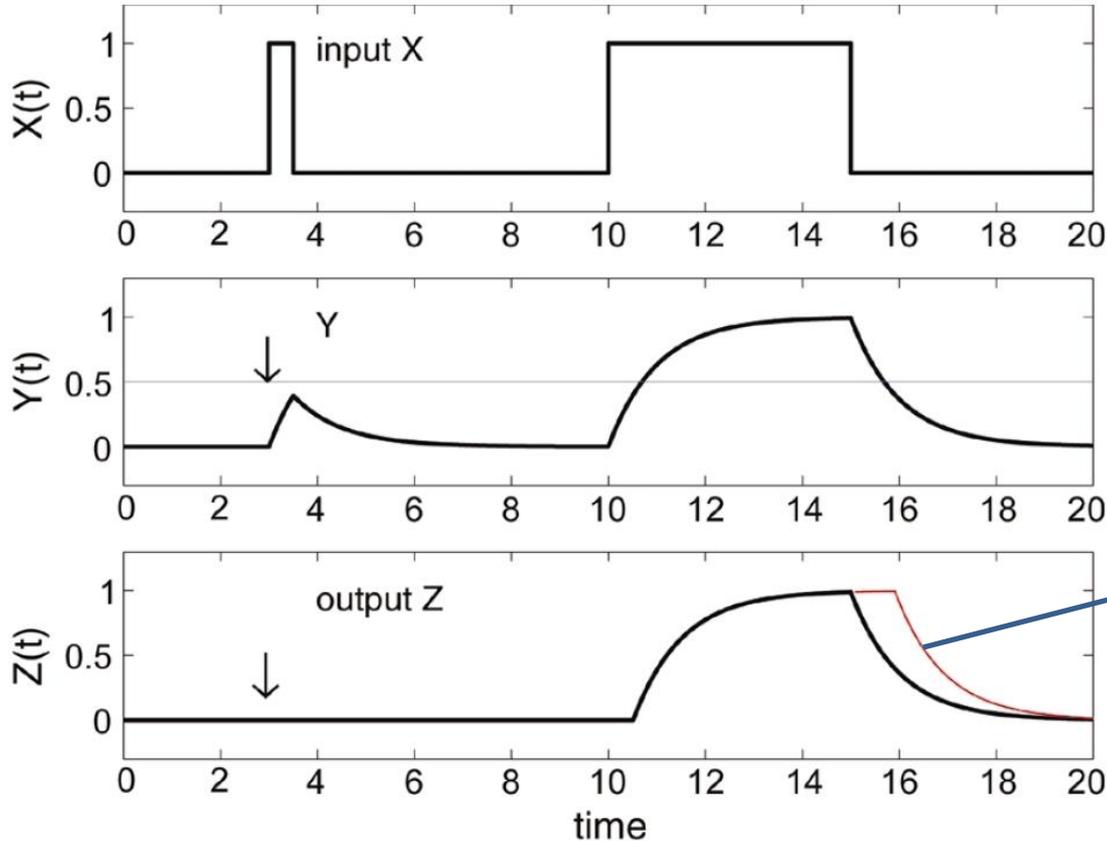
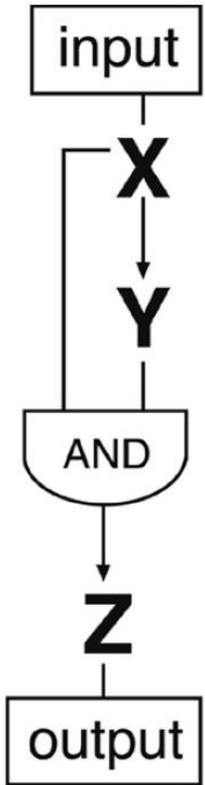
- E. Coli network
 - 116 TFs
 - 577 interactions
 - Significant enrichment of motif # 5



(40 instances vs. 7 ± 3)



What's so interesting about FFLs



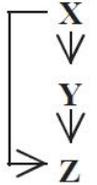
Boolean Kinetics

$$\frac{dY}{dt} = F(X, T_y) - aY$$
$$\frac{dZ}{dt} = F(X, T_y)F(Y, T_z) - aZ$$

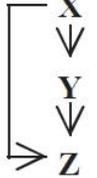
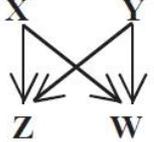
A simple cascade has slower shutdown

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 for a rapid system shutdown.

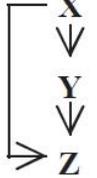
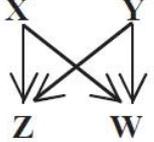
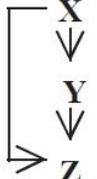
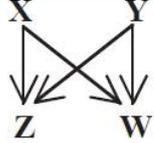
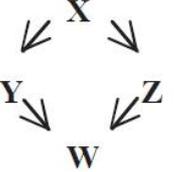
Network motifs in biological networks

Network	Nodes	Edges	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score
Gene regulation (transcription)					Feed- forward loop
<i>E. coli</i>	424	519	40	7 ± 3	10
<i>S. cerevisiae</i> *	685	1,052	70	11 ± 4	14

Network motifs in biological networks

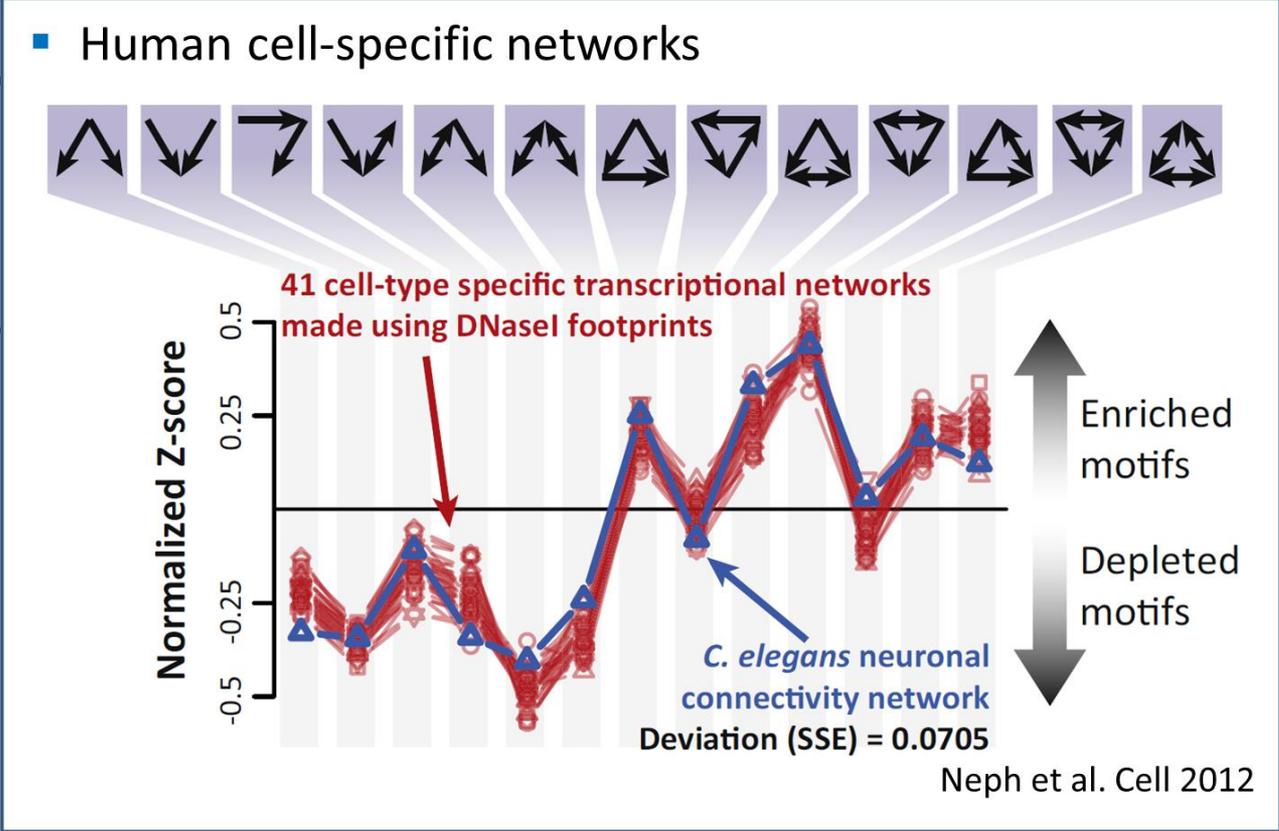
Network	Nodes	Edges	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score
Gene regulation (transcription)											
					Feed-forward loop			Bi-fan			
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> *	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			

Network motifs in biological networks

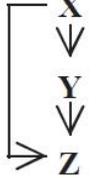
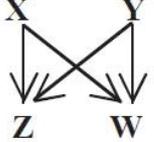
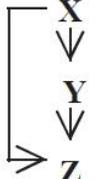
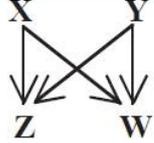
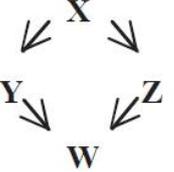
Network	Nodes	Edges	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score
Gene regulation (transcription)				Feed-forward loop			Bi-fan				
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> *	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons				Feed-forward loop			Bi-fan			Bi-parallel	
<i>C. elegans</i> †	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20

Network motifs in biological networks

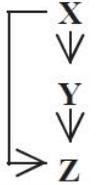
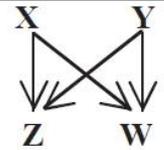
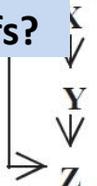
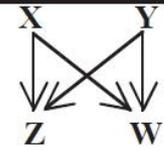
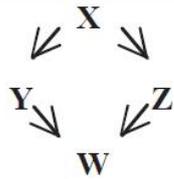
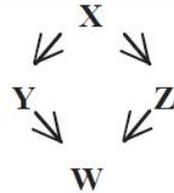
Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)					Feed-forward loop			Bi-fan			
<i>E. coli</i>	424										
<i>S. cerevisiae</i> *	685										
Neurons											Bi-parallel
<i>C. elegans</i> †	252										± 10 20



Network motifs in biological networks

Network	Nodes	Edges	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score
Gene regulation (transcription)				Feed-forward loop			Bi-fan				
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> *	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons				Feed-forward loop			Bi-fan			Bi-parallel	
<i>C. elegans</i> †	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20

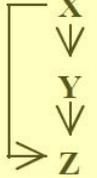
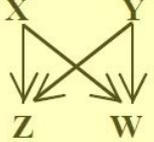
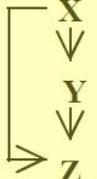
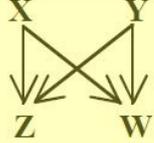
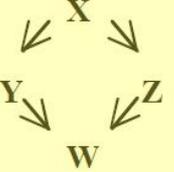
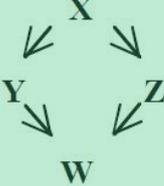
Network motifs in biological networks

Network	Nodes	Edges	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score
Gene regulation (transcription)					Feed-forward loop			Bi-fan			
<i>E. coli</i>	424			7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> *	685			11 ± 4	14	1812	300 ± 40	41			
Neurons					Feed-forward loop			Bi-fan			Bi-parallel
<i>C. elegans</i> †	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					Three chain			Bi-parallel			FFL motif is under-represented!
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			

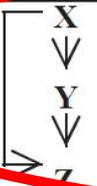
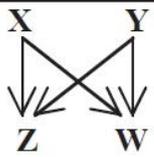
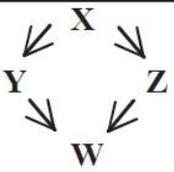
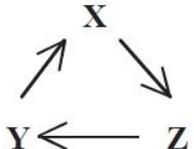
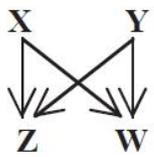
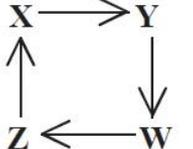
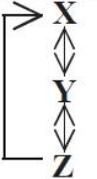
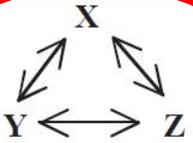
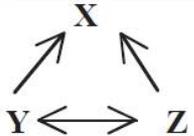
Why do these networks have similar motifs?

Why is this network so different?

Information Flow vs. Energy Flow

Network	Nodes	Edges	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score
Gene regulation (transcription)											
					Feed-forward loop			Bi-fan			
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> *	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons											
					Feed-forward loop			Bi-fan			Bi-parallel
<i>C. elegans</i> †	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs											
					Three chain			Bi-parallel	FFL motif is under-represented!		
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			

Network Motifs in Technological Networks

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	$2e3 \pm 1e2$	800	6.8e6	$5e4 \pm 4e2$	15,000	1.2e6	$1e4 \pm 2e2$	5000

Motif-based network super-families

