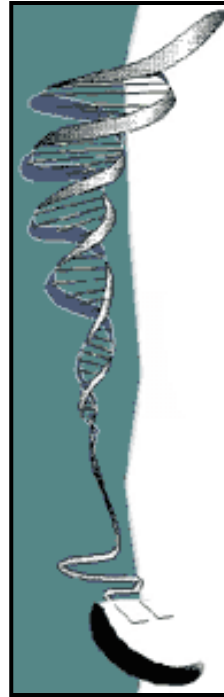
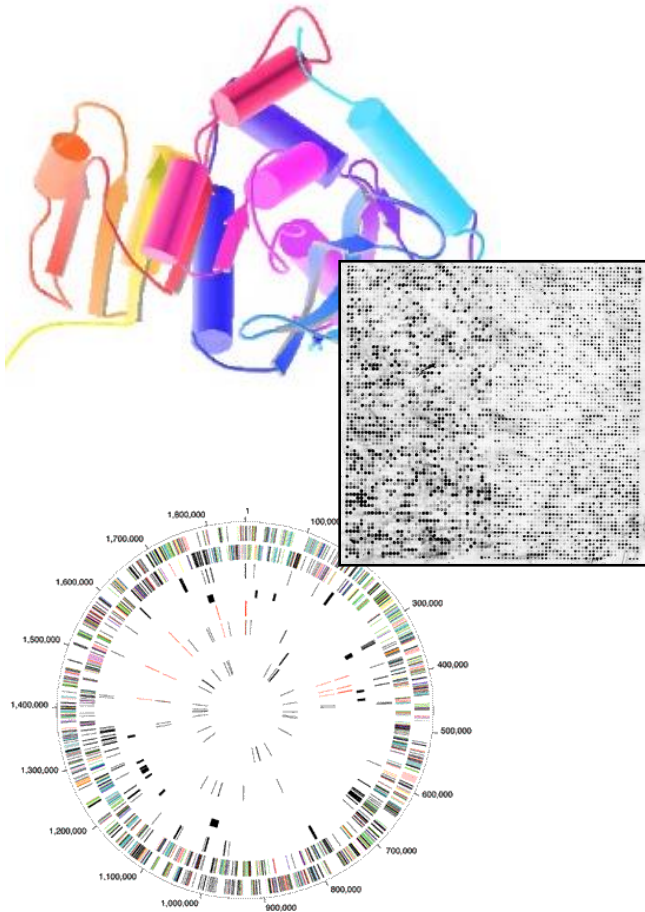


Biomedical Data Science: Analysis of Network Topology -- Network Quantities



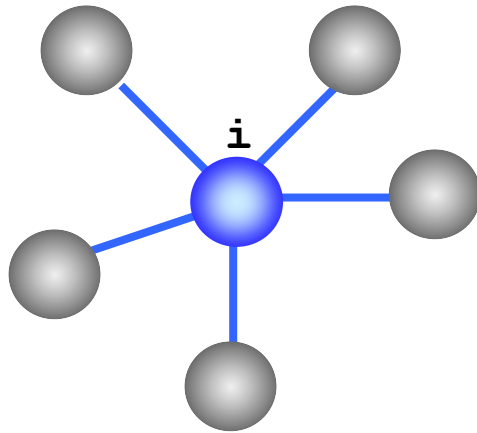
Mark Gerstein, Yale University
gersteinlab.org/courses/452

(Last edit in spring '22; pack 22m10b, very similar to M10b from '21.)

Network Topology

**What are the Main
Quantities that Can be
Calculated from Networks?**

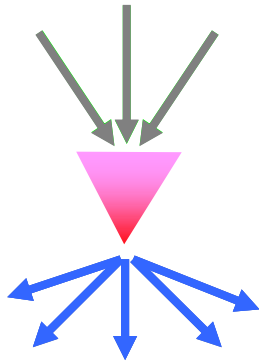
- **Degree of a node:** the number of edges incident on the node



Degree of node $i = 5$

Network parameters

Number of incoming and outgoing connections



Connectivity

Incoming connections = 2.2

→ each gene is regulated by ~ 2 TFs

In-degree

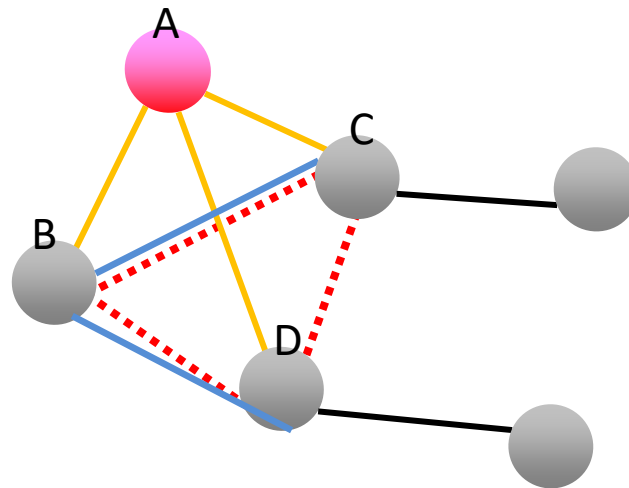
Outgoing connections = 20.2

→ each TF regulates ~ 20 genes

Out-degree

Clustering coefficient

- Clustering Coefficient:
 - Ratio of existing links to maximum number of links for neighbouring nodes
 - Example:
 - For A:
 - 3 neighbours
 - 2 existing link
 - 3 possible links
 - Clustering coefficient
 - $C_A = 2/3$



Example from: <http://www.learner.org/courses/mathilluminated/units/11/textbook/04.php>

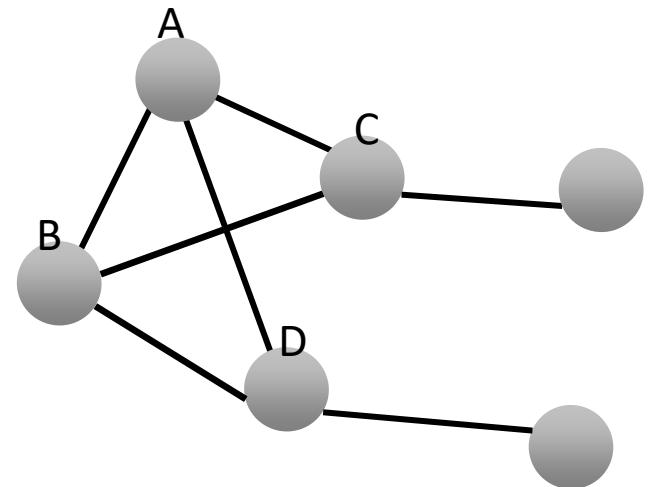
Clustering coefficient

- Average Coefficient:
 - Average of clustering coefficients of all nodes n

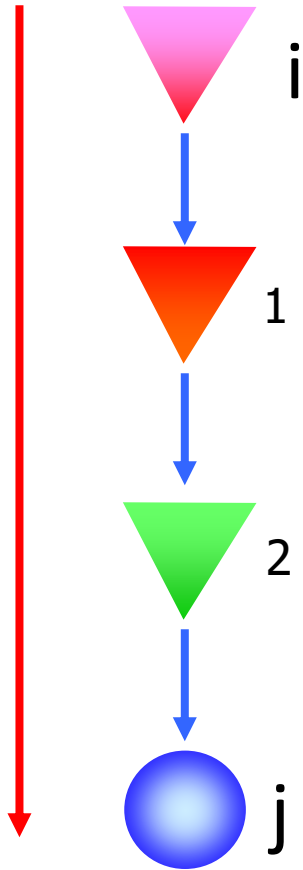
$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$$

- Measure of inter-connectedness of the network
 - Global property

- Example:
 - Clustering coefficient:
 $C_A=2/3$ $C_B=2/3$
 $C_C=1/3$ $C_D=1/3$
 - Average coefficient =
 $1/4(2/3+2/3+1/3+1/3)= 0.5$



Path length

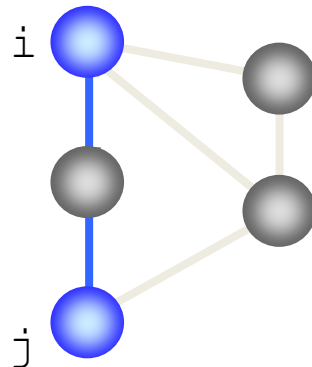


- Number of edges along a path
- Path length = 3
- Meaning:
 - Number of intermediate TFs to reach final target
 - Indication of how immediate a response is

Path length

- Shortest path length:

- $L_{(i,j)}$ is the minimum number of edges that must be traversed to travel from a vertex i to another vertex j of a graph G



$$L_{(i,j)} = 2$$

Graph Theory Terminology (Batten, pp. 92-105)

- Characteristic path length (Average path length)

- The characteristic path length L of a graph is the average of the $L_{(i,j)}$ for every possible pair (i,j)

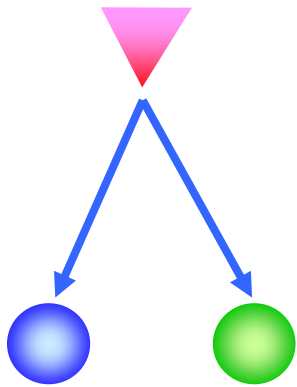
$$L = \frac{1}{n(n-1)} \sum_{i,j} L_{(i,j)}$$

- Networks with small values of L are said to have the “small world property”

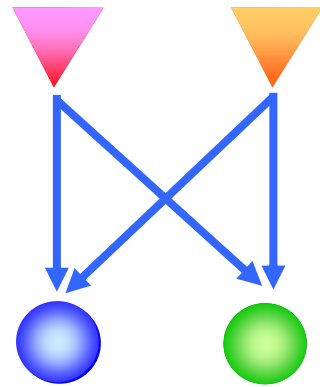
http://en.wikipedia.org/wiki/Average_path_length

Network motifs

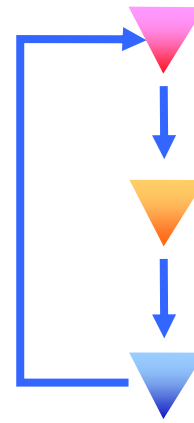
Regulatory modules within the network



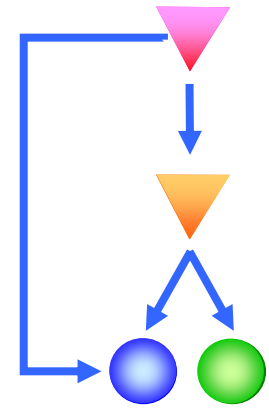
SIM



MIM



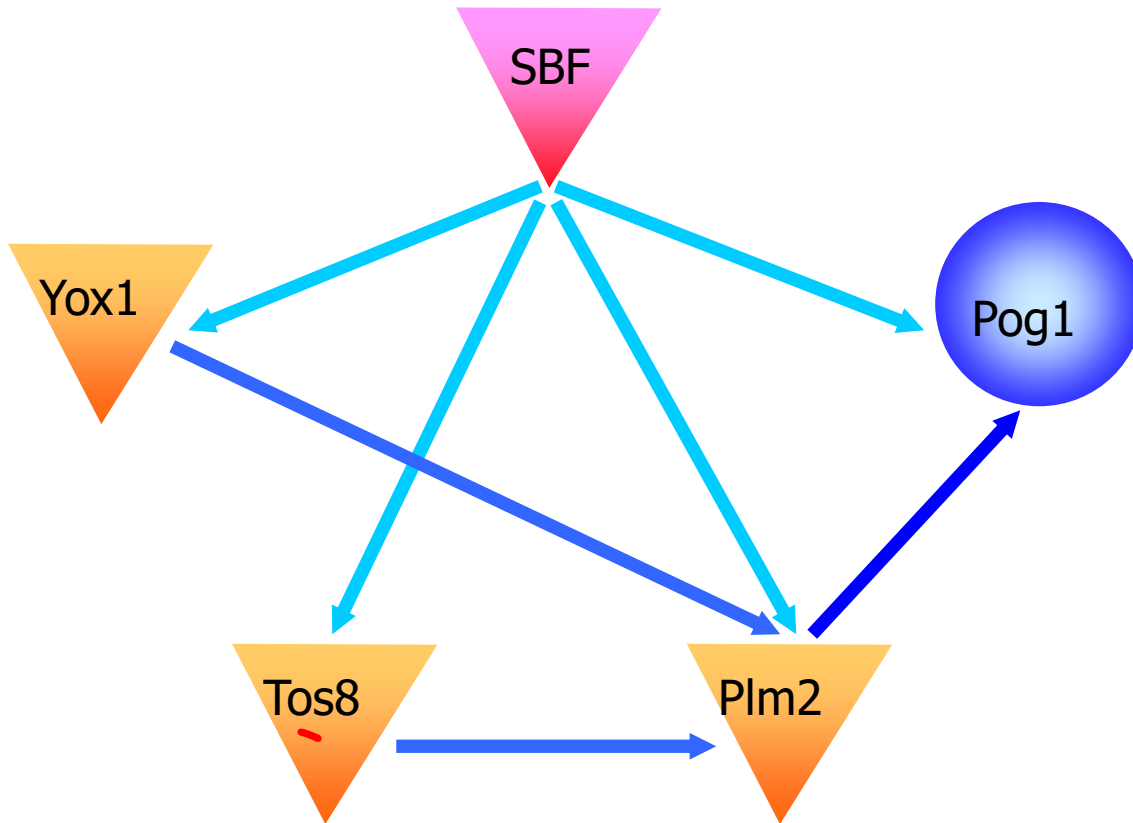
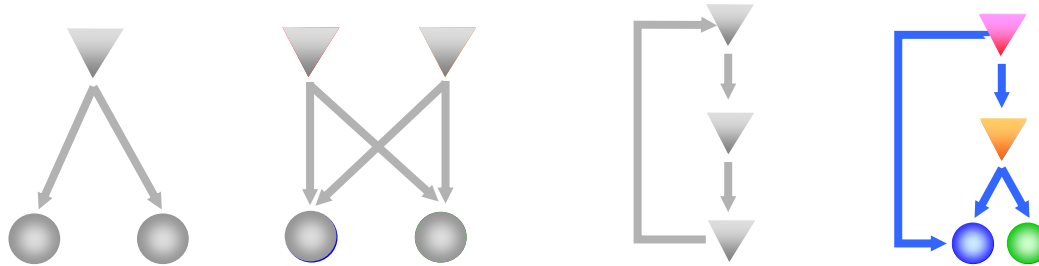
FBL



FFL

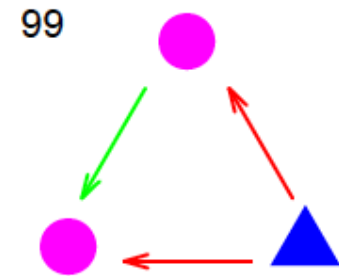
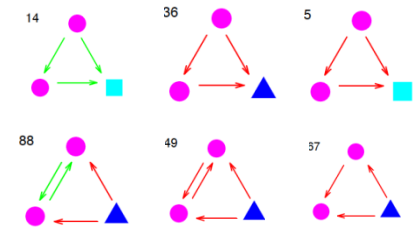
[Alon U, 2007, Nature Reviews Genetics, Network motifs: theory and experimental approaches]

FFL = Feed-forward loops

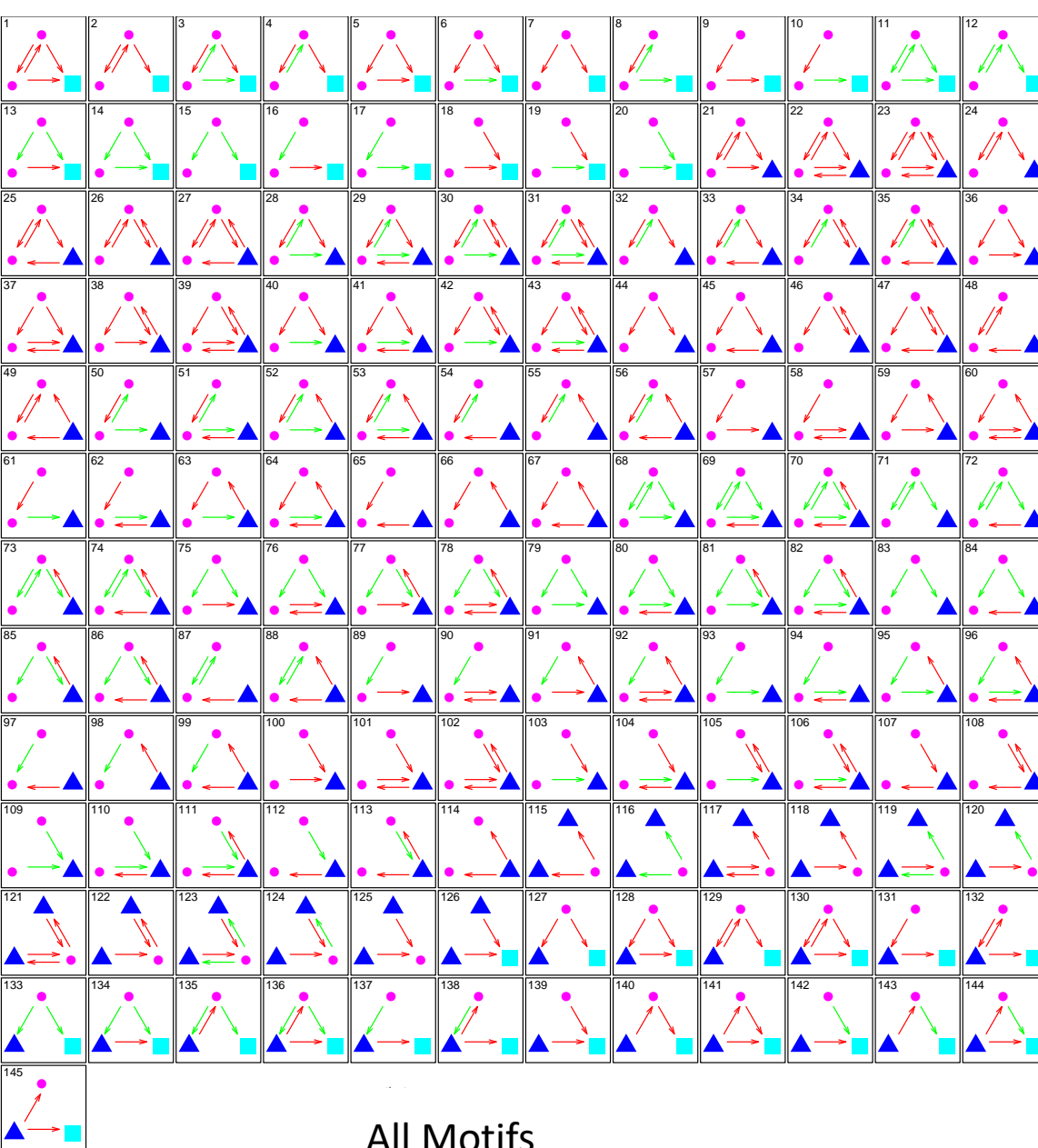


Network Motifs

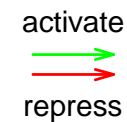
7 Motifs
Over-represented



FFL involving
miRNA & 2 TFs



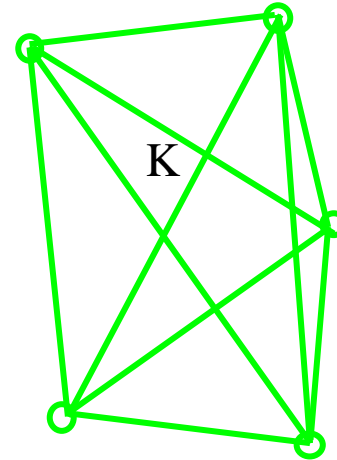
All Motifs



Cliques

- Fully connected sub-components
- Related measures
k-cores : For all vertices in a graph G have degree at least k

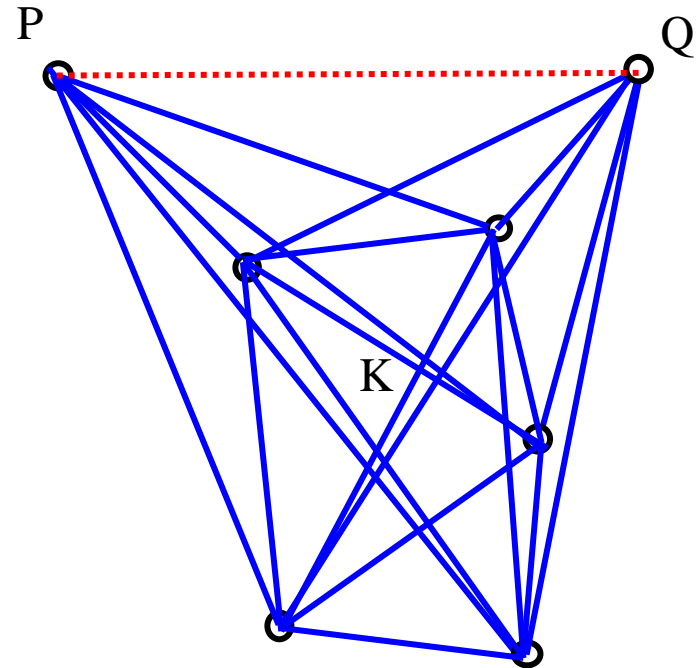
[Hogue et al, BMC BIOINFORMATICS, 2003]



Problem: High-throughput experiments are prone to missing interactions

One solution—defective cliques

- If proteins P and Q interact with a clique K of proteins which all interact with each other, then P and Q are more likely to interact with each other
- P, Q, and K form a **defective clique**



Predicting protein interactions by completing defective cliques