Lecture 15
Graph Neural Networks

Foundations, GNNs, PPIs, Graphs

Manolis Kellis
Guest lecture: Neil Band (part 2)
Guest lecture: Maria Brbic / Jure Leskovec

Slides credit: Soheil Feizi

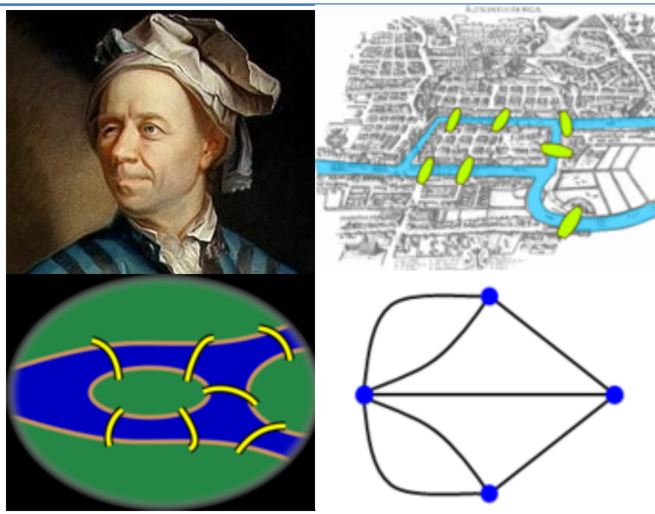
6.874/6.802/20.390/20.490/HST.506 Deep Learning in the Life Sciences Spring 2021 Lec:Tue/Thu 1-2:30pm Rec:Fri 3-4pm Proj:Fri 4-5pm

Psets	Date	Module	Week	Lec/R	Description
PS0: Set up Environment (Due Monday 2/22)	Tuesday, February 16, 2021	Module 1: ML models and interpretation	1	L01	Course Intro + Overview Foundations
	Thursday, February 18, 2021			L02	ML foundations
	Friday, February 19, 2021			R01	ML Review
	Friday, February 19, 2021			Proj1	Intro video + personal profile
PS1: Softmax warmup (MNIST) (out: Tue 2/23, due: Wed 3/10)	Tuesday, February 23, 2021		2	L03	Convolutional Neural Networks CNNs
	Thursday, February 25, 2021			L04	RNNs, GNNs
	Friday, February 26, 2021			R02	Neural Networks Review
	Friday, February 26, 2021			Proj2	Research Mentors Introductions and Breakouts
	Tuesday, March 2, 2021		3	L05	Interpretability, Dimensionality Reduction, tSNE
	Thursday, March 4, 2021			L06	Generative Models, GANs, VAEs
	Friday, March 5, 2021			R03	Interpreting ML Models
	Friday, March 5, 2021			Proj3	**Research Team Building Breakout Rooms**
PS2: CNN for TF binding prediction (out: Tue 3/16, Due: Mon 3/29)	Tuesday, March 9, 2021	Module 2: Gene Regulation	4		No Class (Monday Schedule)
	Thursday, March 11, 2021			L07	DNA accessibility, Promoters and Enhancers
	Friday, March 12, 2021			R04	Chromatin and gene regulation
	Friday, March 12, 2021			Proj4	**Initial Ideas 1-slide presentations (teams, or individual)**
	Tuesday, March 16, 2021		5	L08	Transcription factors, DNA methlyation
	Thursday, March 18, 2021			L09	Gene Expression, Splicing
	Friday, March 19, 2021			R05	RNA-seq, Splicing
	Friday, March 19, 2021			Proj5	**Meet with potential mentors (optional, asynchronous)**
	Tuesday, March 23, 2021		6		No Class (Student Holiday)
	Thursday, March 25, 2021			L10	Single-cell RNA-sequencing
	Friday, March 26, 2021			R06	scRNA-seq, dimensionality reduction
	Friday, March 26, 2021			Proj6	**Full Project Proposals Due (pdf, slides, team video)**
PS3: scRNA-seq tSNE analysis (out: Tue 3/30, due Mon 4/12)	Tuesday, March 30, 2021	Module 3: Genetic Variation / Disease	7	L11	Dimensionality reduction, PCA, t-SNE, NNMF
	Thursday, April 1, 2021			L12	GWAS, variant calling, variant interpretation
	Friday April 2, 2021			R07	Genetics
	Friday, April 2, 2021			Proj7	Meet with your mentors (optional, asynchronous)
	Tuesday, April 6, 2021		8	L13	eQTLs, intermediate molecular phenotypes
	Thursday, April 8, 2021			L14	Electronic health records and patient data
	Friday April 9, 2021			R08	ML for health data
	Friday, April 9, 2021			Proj8	End-to-End pipeline demo (team video)
PS4: Graph Neural Networks (Out: Tue 4/13, Due: Wed 4/28)	Tuesday, April 13, 2021	Module 4: Graphs and Proteins	9	L15	Graphs, GNNs, Protein-protein interactions
	Thursday, April 15, 2021			L16	GNNs for Protein Structure and Drug Design
	Friday April 16, 2021			R09	Graph Neural Networks
	Tuesday, April 20, 2021		10		No Class (Student Holiday)
	Thursday, April 22, 2021			L17	GNNs for Protein Structure and Drug Design
	Friday April 23, 2021			R10	Drug Development
	Friday, April 23, 2021			Proj9	Meet with your mentors (optional, asynchronous)
PS5: Image Analysis (Out: Wed 4/28, Due: Mon 5/10)	Tuesday, April 27, 2021	Quiz			In-class quiz
	Thursday, April 29, 2021	Module 5: Imaging	11	L19	Imaging, Morphology
	Friday, April 30, 2021			R11	Therapeutics, 3D structure, imaging
	Friday, April 30, 2021			Proj10	Midcourse report (google doc)
	Tuesday, May 4, 2021		12	L20	Imaging applications in healthcare
	Thursday, May 6, 2021			L21	Video processing, structure determination
	Friday May 7, 2021				No Class (Student Holiday)
Finalize Projects	Tuesday, May 11, 2021	Module 6: Frontiers	13	L22	Text applications in healthcare, clinical decision making
	Thursday, May 13, 2021			L23	Neuroscience
	Friday, May 14, 2021			R12	How to Present
	Friday, May 14, 2021			Proj11	How to Present
	Monday, May 17, 2021		14	Proj12	Final Reports due (Google doc + pdf)
	Tuesday, May 18, 2021			L24	Cancer and Infectious Disease
	Wednesday, May 19, 2021			Proj13	Final Presentations (slides, team video)
	Thursday, May 20, 2021			L25	Final Presentations
Course website:	http://compbio.mit.edu/6874/				

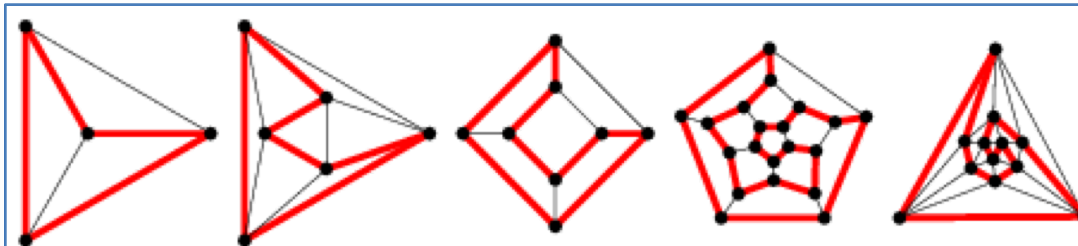
Goals for today: Network analysis

1. **Introduction to networks**
   – Network types: regulatory, metab., signal., interact., func.
   – Bayesian (probabilistic) and Algebraic views
2. **Network Centrality Measures**
   – Local centrality metrics (degree, betweenness, closeness, etc)
   – Global centrality metrics (eigenvector centrality, page-rank)
3. **Linear Algebra Review: eigenvalues, SVD, low-rank approximations**
   – Eigenvector and singular vector decomposition
   – Low rank approximations, Wigner semicircle law
4. **Sparse Principal Component Analysis**
   – Lasso and Elastic lasso
   – PCA and Sparse PCA
5. **Network Communities and Modules**
   – Guilt by association
   – Maximum cliques, density-based modules and spectral clustering
6. **Network Diffusion Kernels and Deconvolution**
   – Network diffusion kernels
   – Network deconvolution

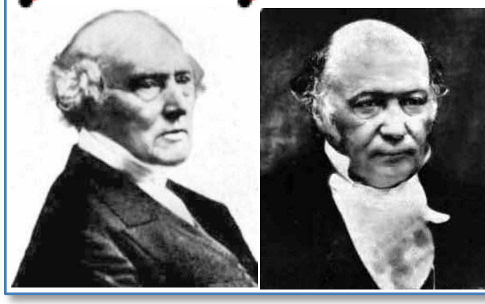
Graph Theory: Abstracting real-world into graphs



Leonhard Euler, *Bridges of Königsberg*, 1736.



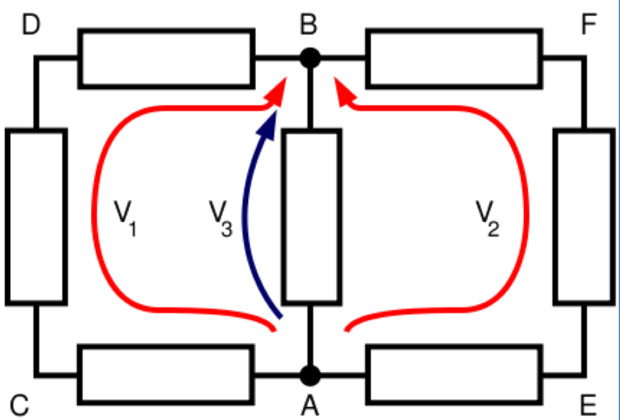
Cycles in Polyhedra



Thomas Kirkman
William Hamilton

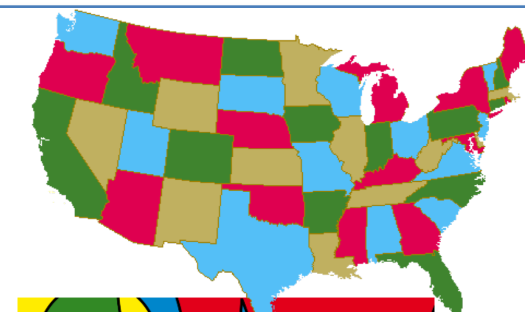
Hamiltonian cycles in Platonic graphs

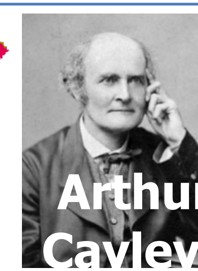
Gustav Kirchhoff



$$V_1 = V_2 = V_3$$

Trees in Electric Circuits

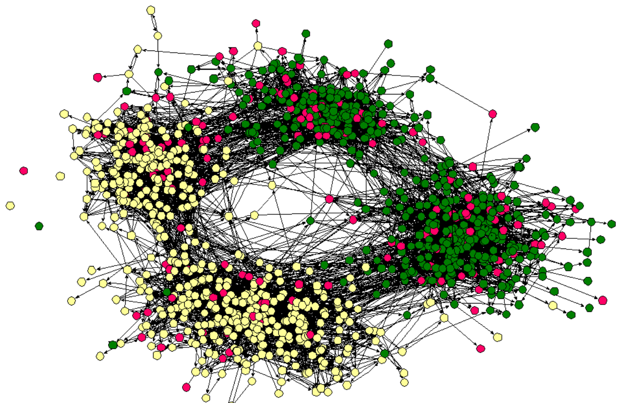




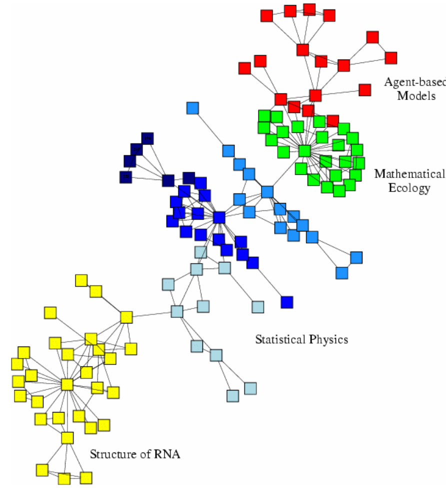
Auguste DeMorgan

Four Colors of Maps

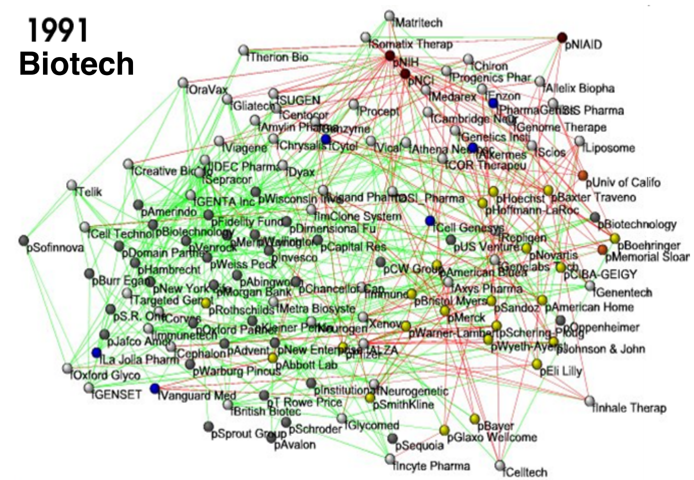
Networks are everywhere in the real world



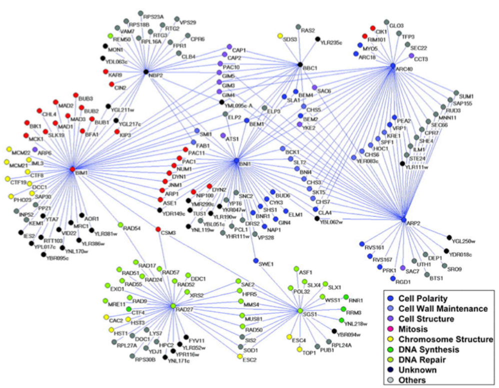
Social Network



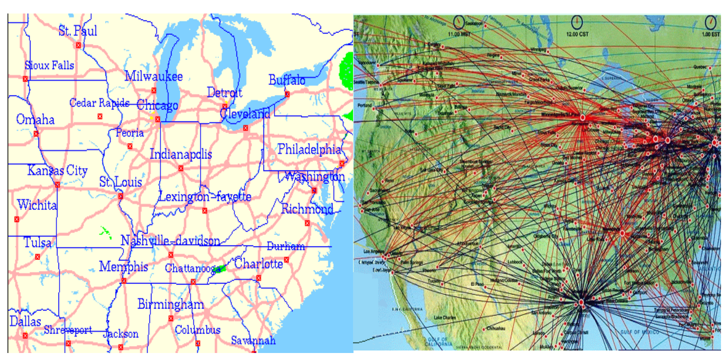
Collaboration Network



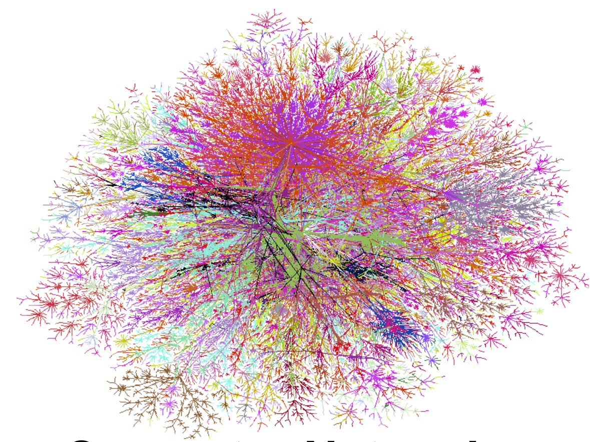
Commercial Networks



Biological networks



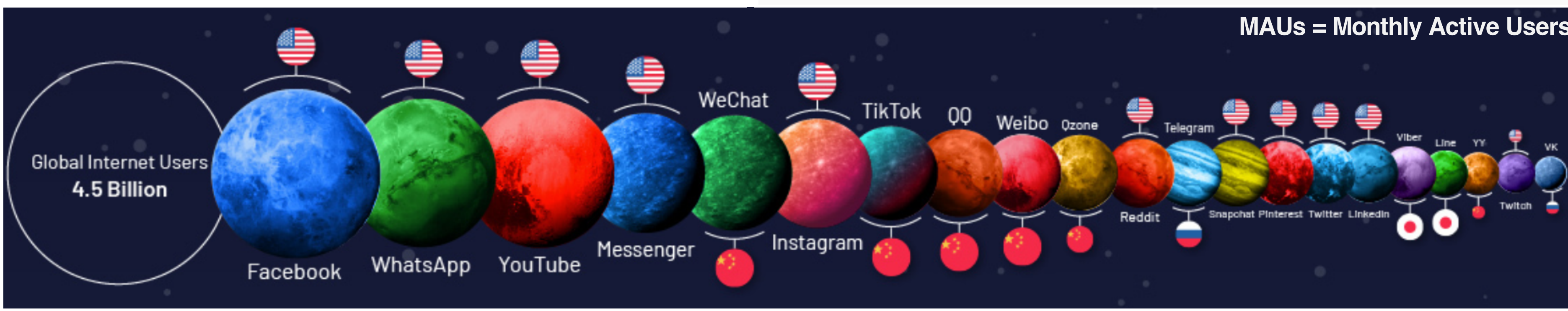
Transportation Networks



Computer Networks

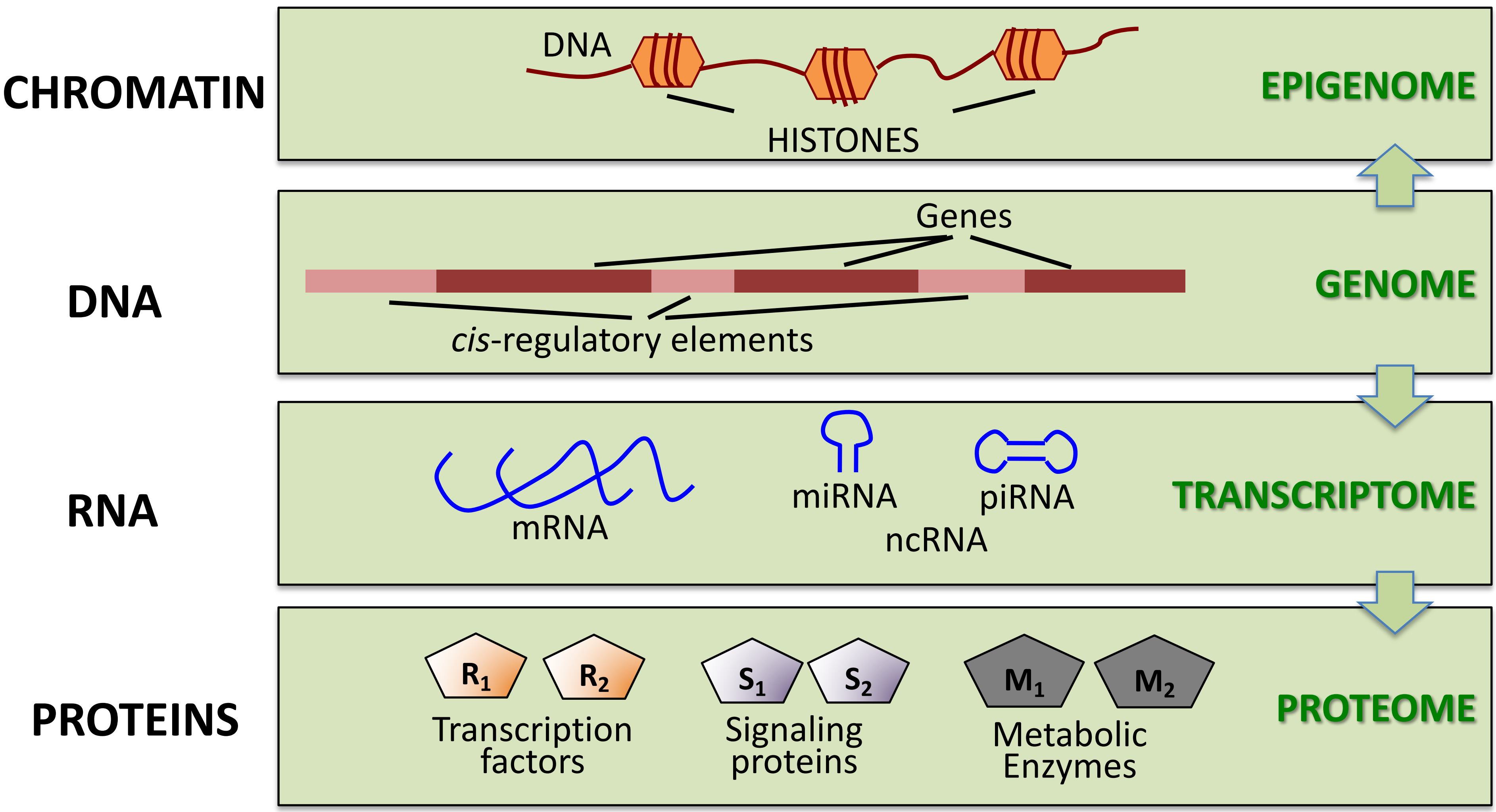
Social networks are most popular websites

Rank	Social Network	MAUs In Millions	Country of Origin
#1	Facebook	2,603 2.6B	U.S.
#2	WhatsApp	2,000 2B	U.S.
#3	YouTube	2,000	U.S.
#4	Messenger	1,300	U.S.
#5	WeChat	1,203	China
#6	Instagram	1,082	U.S.
#7	TikTok	800	China
#8	QQ	694	China
#9	Weibo	550	China
#10	Qzone	517	China
#11	Reddit	430	U.S.
#12	Telegram	400	Russia
#13	Snapchat	397	U.S.
#14	Pinterest	367	U.S.
#15	Twitter	326	U.S.
#16	LinkedIn	310	U.S.
#17	Viber	260	Japan
#18	Line	187	Japan
#19	YY	157	China
#20	Twitch	140	U.S.
#21	Vkontakte	100	Russia

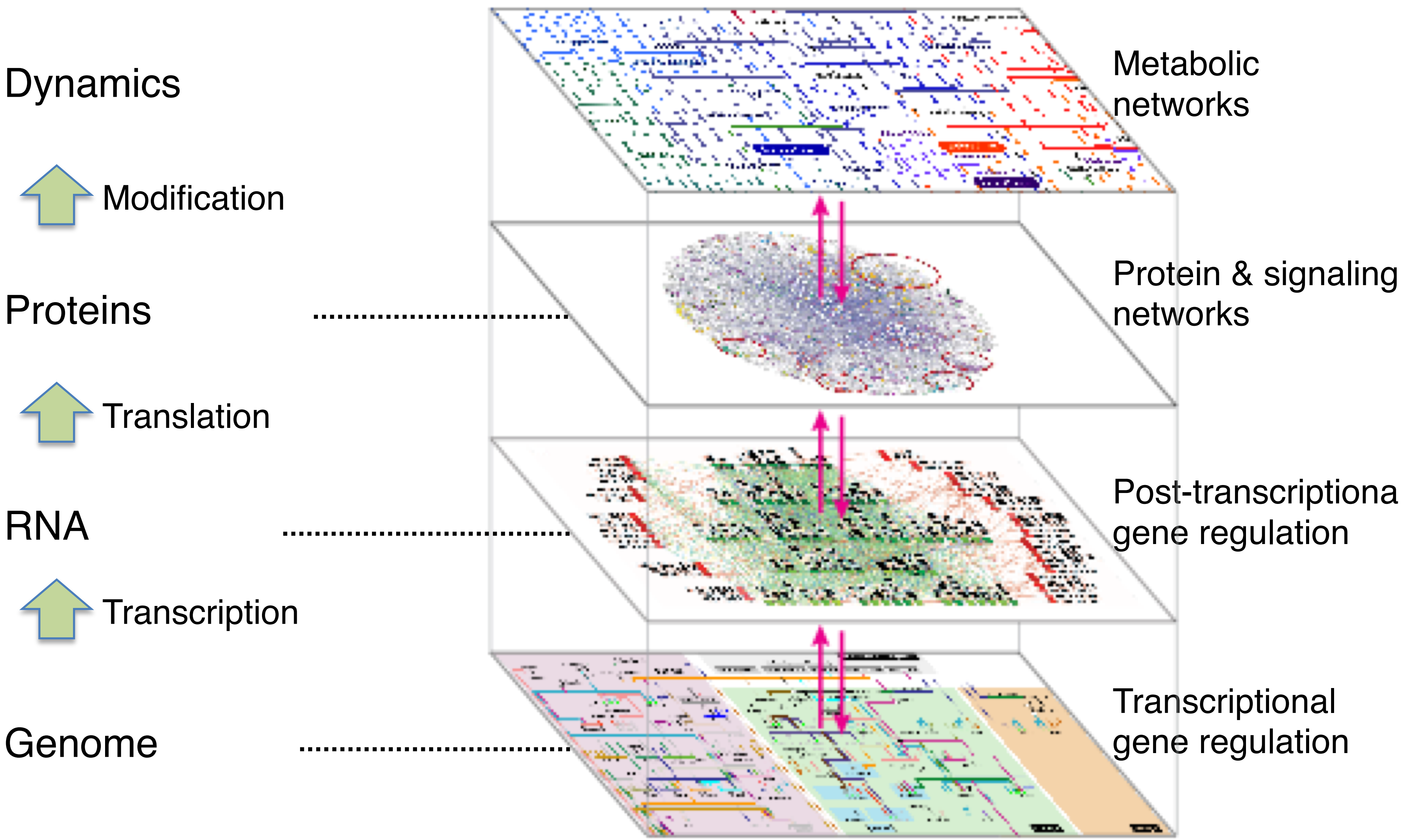


- **Social Network:** asocial structure of individuals/organizations (nodes) tied (connected) by interdependencies (eg. friendship, interests, etc)
- **Social Network Analysis (SNA):** can reveal patterns, properties, important nodes, subnetworks, classification of individuals, etc

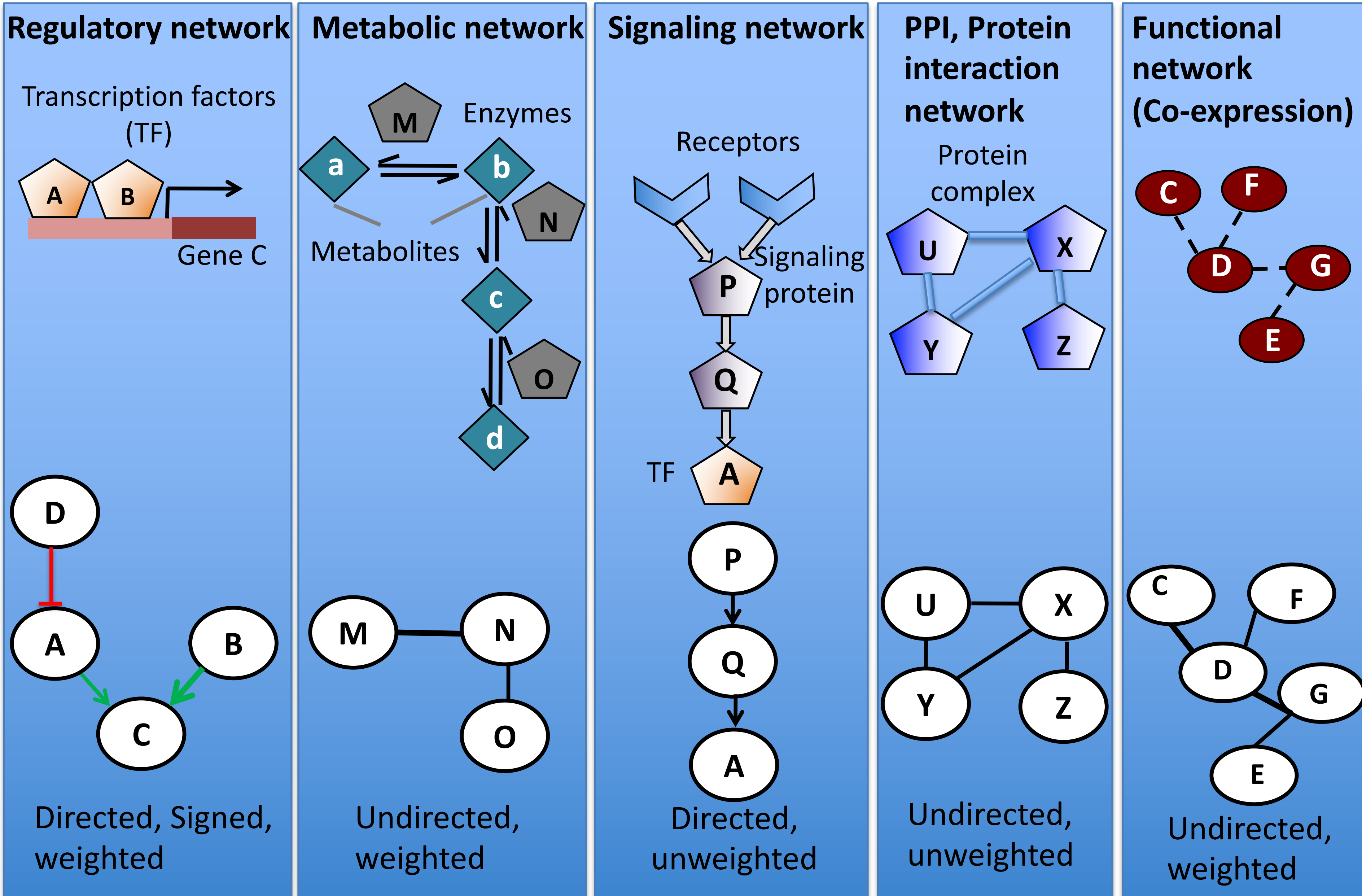
The multi-layered organization of information in living systems



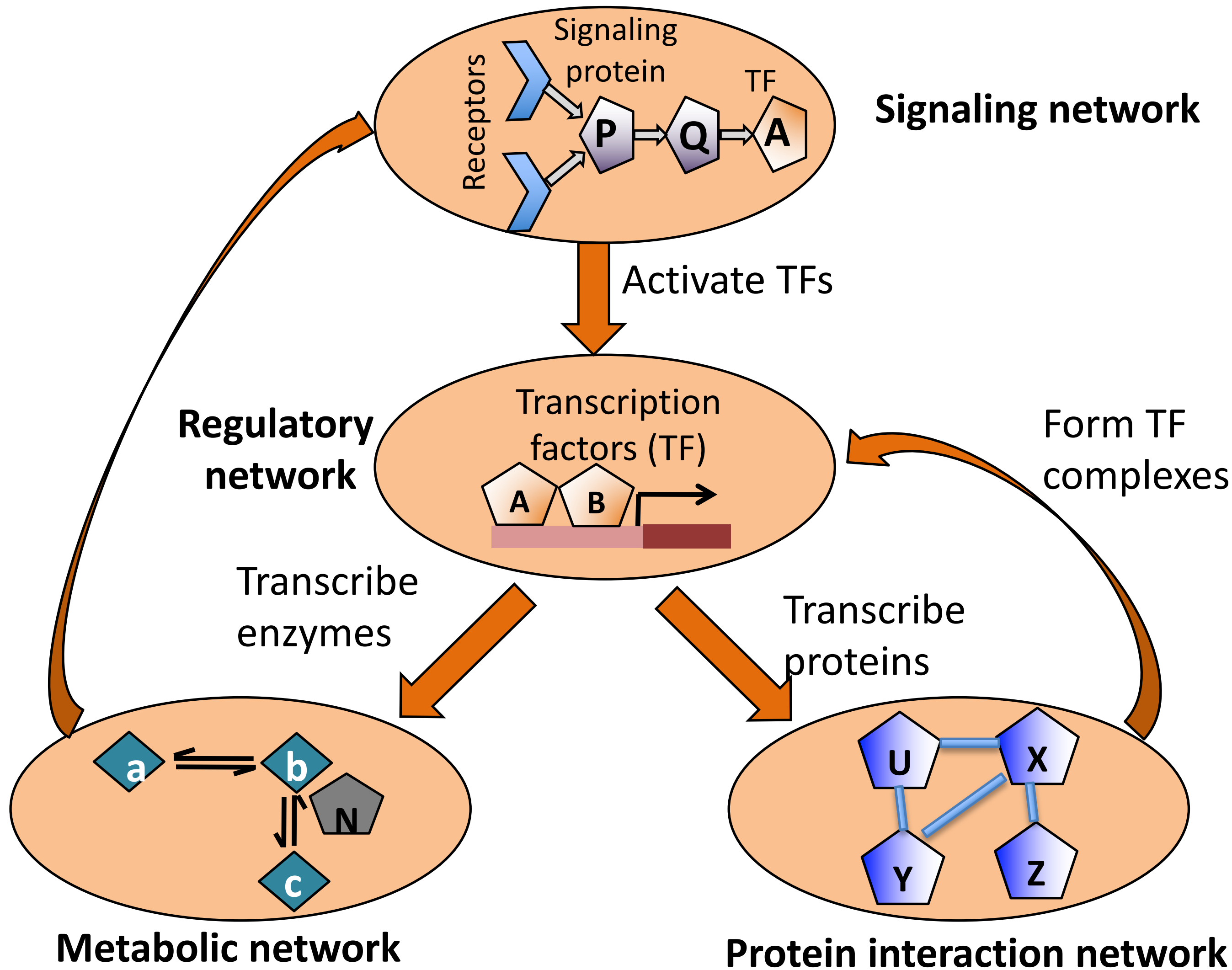
Biological networks at all cellular levels



Five major types of biological networks

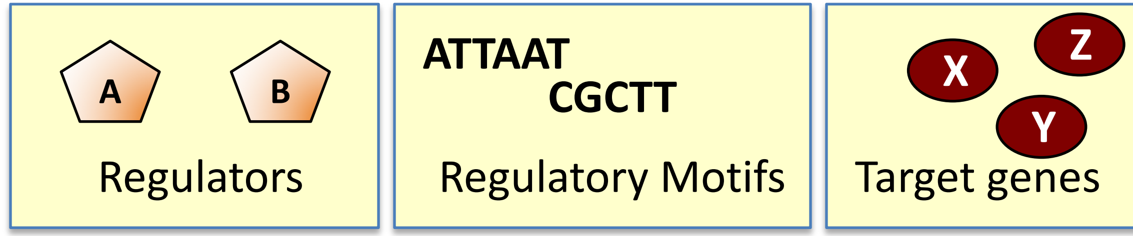


Information exchange across networks

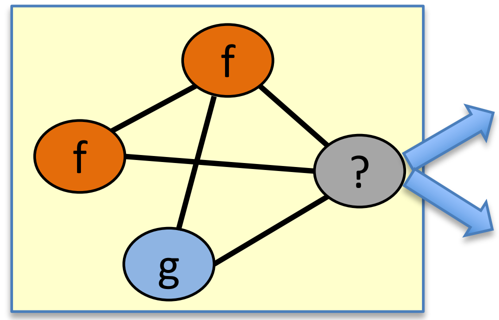


Network applications and challenges

1. **Element Identification (motif finding lecture)**



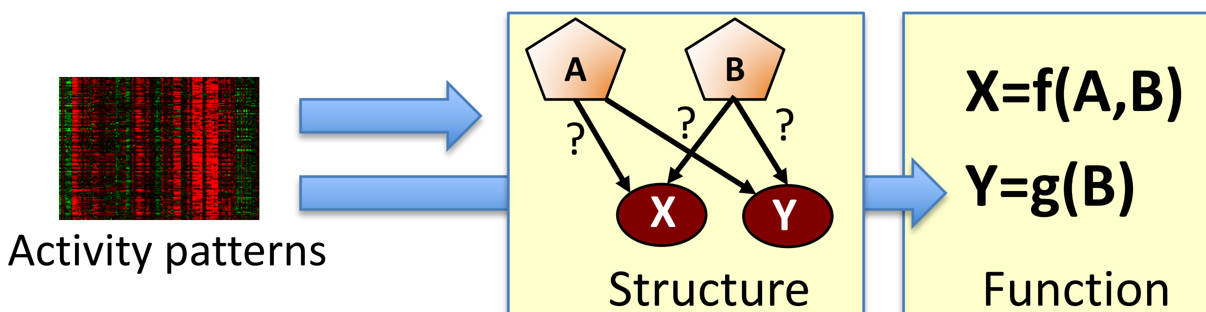
2. **Using networks to predict cellular activity**



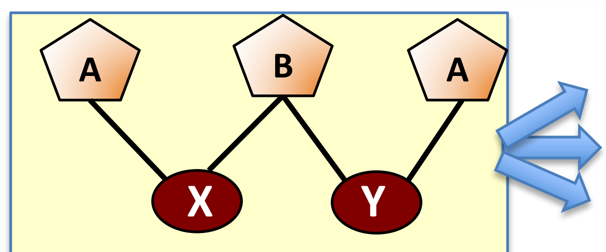
Predict expression levels

Predict gene ontology (GO) functional annotation terms

3. **Inferring networks from functional data**



4. **Network Structure Analysis**



- Hubs (degree-distribution)
- Network motifs
- Functional modules

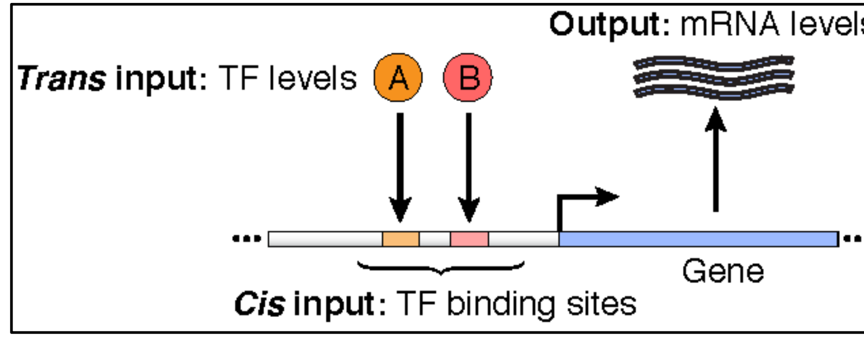
Beyond real-world networks

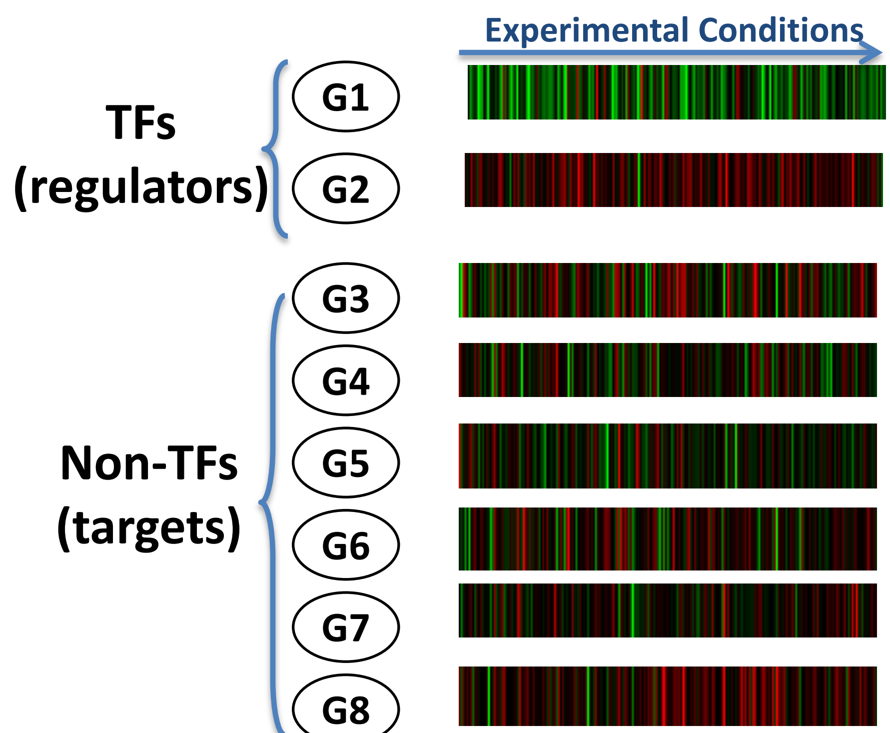
More abstractly, edges can represent relationships between data points

Even more abstractly, nodes themselves can simply be probabilistic variables

Physical and Relevance Networks

- **Physical Networks:**
  - edges represent "physical interaction" among nodes
  - Example: physical regulatory networks

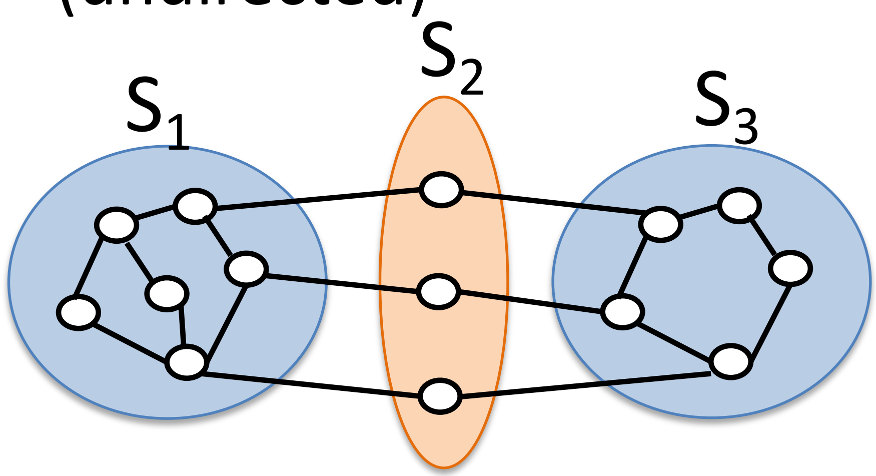




- **Relevance Networks:**
  - edge weights represent node similarities
  - Example: functional regulatory networks

Probabilistic networks and graphical model

- There are several types of networks, with different meanings, and different applications
- Networks as graphical models:
  - modeling joint probability distribution of variables using graphs
  - Bayesian networks (directed), Markov Random Fields (undirected)

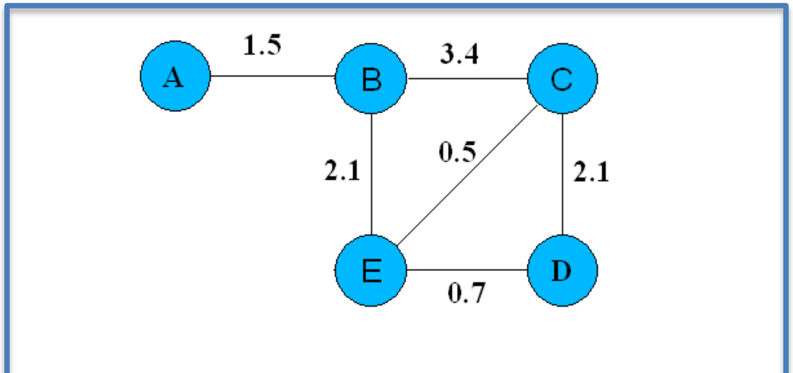


Next Lecture!

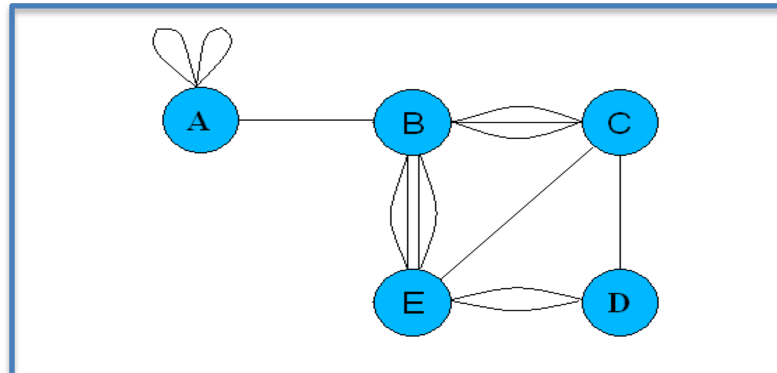
$$X_{S_1} \perp\!\!\!\perp X_{S_3} | X_{S_2}$$

Representing Networks as Graphs

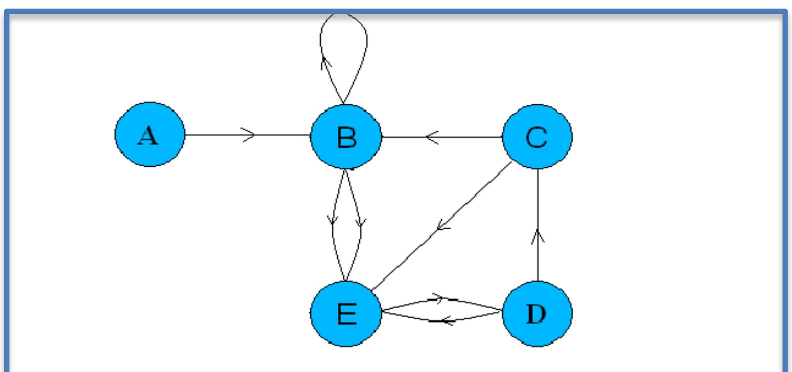
- **Weighted graphs:** weights associated to every edge, generally positive



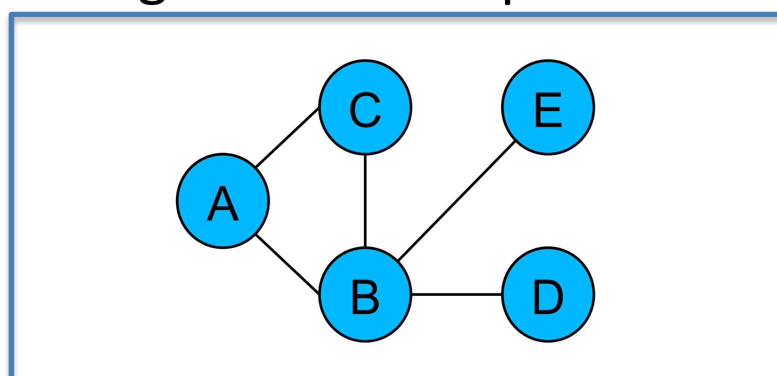
- **Multigraphs (Pseudographs):** multiple edges can exist among nodes



- **Digraphs:** edges have directions

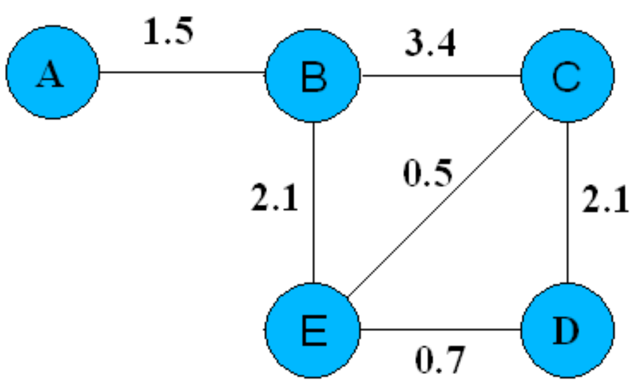


- **Simple graphs**: no multiple edges or self-loops



Matrix representation of networks

- A matrix representation of a network:
  - **Unweighted network:** binary adjacency matrix
  - **Weighted network:** real-valued matrix



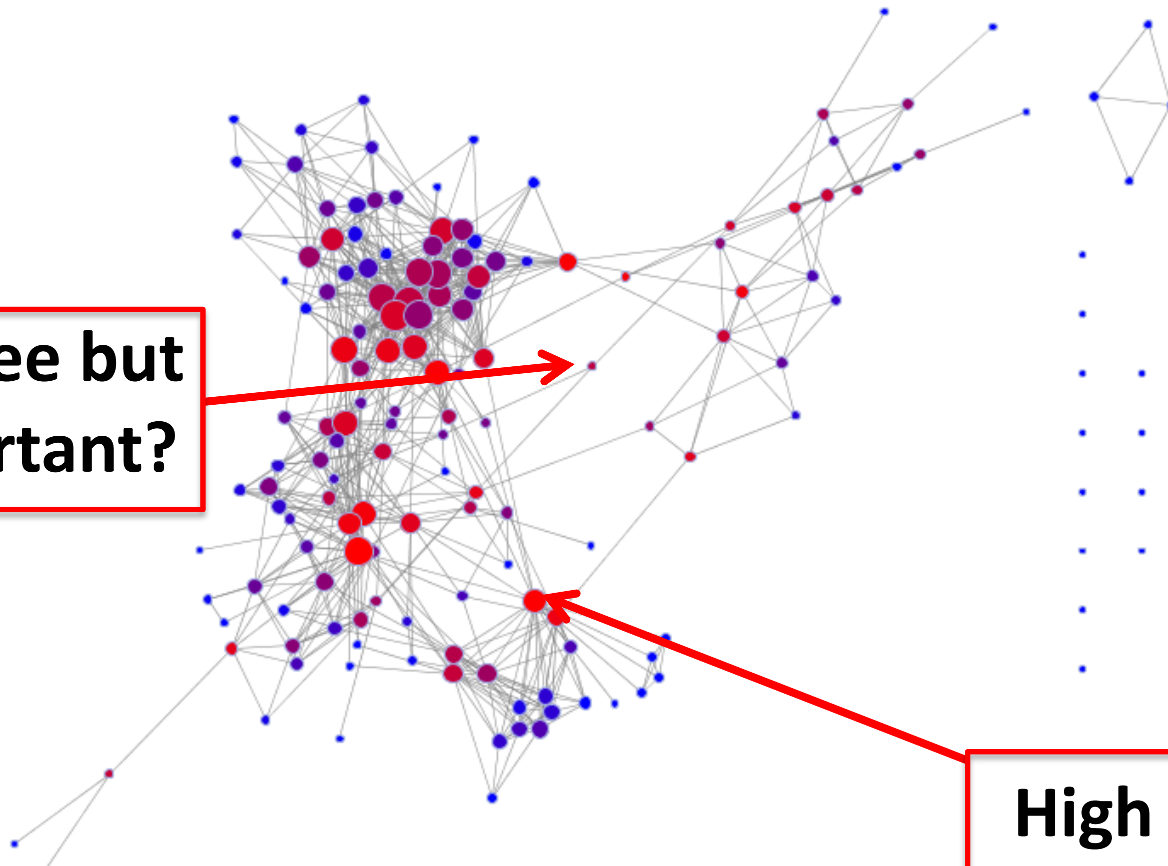
$$\begin{array}{ccc} & \begin{array}{ccccc} A & B & C & D & E \end{array} & \underline{\textbf{\textit{Degree}}} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{pmatrix} 0 & 1.5 & 0 & 0 & 0 \\ 1.5 & 0 & 3.4 & 0 & 0 \\ 0 & 3.4 & 0 & 2.1 & 0.5 \\ 0 & 0 & 2.1 & 0 & 0.7 \\ 0 & 2.1 & 0.5 & 0.7 & 0 \end{pmatrix} & \begin{array}{c} 1.5 \\ 4.9 \\ 6 \\ 2.8 \\ 3.3 \end{array} \end{array}$$

Goals for today: Network analysis

1. **Introduction to networks**
   - Network types: regulatory, metab., signal., interact., func.
   - Bayesian (probabilistic) and Algebraic views
2. **Network Centrality Measures**
   - Local centrality metrics (degree, betweenness, closeness, etc)
   - Global centrality metrics (eigenvector centrality, page-rank)
3. **Linear Algebra Review: eigenvalues, SVD, low-rank approximations**
   - Eigenvector and singular vector decomposition
   - Low rank approximations, Wigner semicircle law
4. **Sparse Principal Component Analysis**
   - Lasso and Elastic lasso
   - PCA and Sparse PCA
5. **Network Communities and Modules**
   - Guilt by association
   - Maximum cliques, density-based modules and spectral clustering
6. **Network Diffusion Kernels and Deconvolution**
   - Network diffusion kernels
   - Network deconvolution

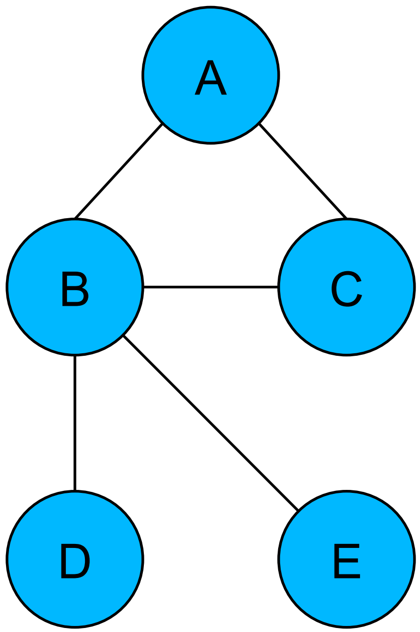
Centrality Measures in Networks

Question: how important is a node/edge to the structural characteristics of the system?



Degree Centrality

- **Example:**



$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}
\quad
\begin{array}{c} \text{degree} \\ 1 \\ 4 \\ 3 \\ 1 \\ 1 \end{array}$$

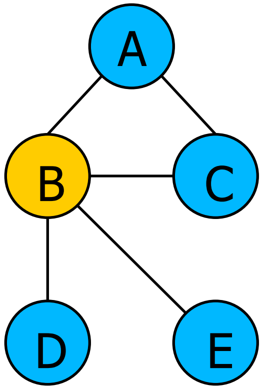
- Degree Centrality can be similarly defined for
  - Directed graphs, in- and out- degrees
  - Weighted graphs, weighted degrees

Betweenness centrality

- The number of shortest paths in the graph that pass through the node divided by the total number of shortest paths.

$$BC(k) = \sum_i \sum_j \frac{\rho(i,k,j)}{\rho(i,j)}, \quad i \neq j \neq k$$

- Nodes with a high betweenness centrality control information flow in a network.
- Edge betweenness is defined similarly.



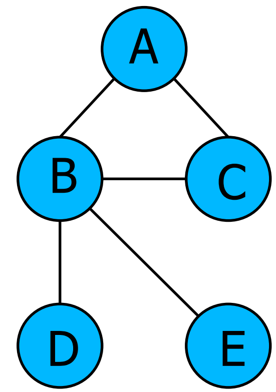
- Shortest paths are:
  - AB, AC, <u>ABD</u>, <u>ABE</u>, BC, BD, BE, <u>CBD</u>, <u>CBE</u>, <u>DBE</u>

$$\rho(A,B,D)=1; \quad \rho(A,D)=1$$
$$\rho(A,B,E)=1; \quad \rho(A,E)=1$$
$$\rho(C,B,D)=1; \quad \rho(B,D)=1$$
$$\rho(C,B,E)=1; \quad \rho(C,E)=1$$
$$\rho(D,B,E)=1; \quad \rho(D,E)=1$$

  - B has a BC of 5/10

Closeness Centrality

- The normalised inverse of the sum of topological distances in the graph.

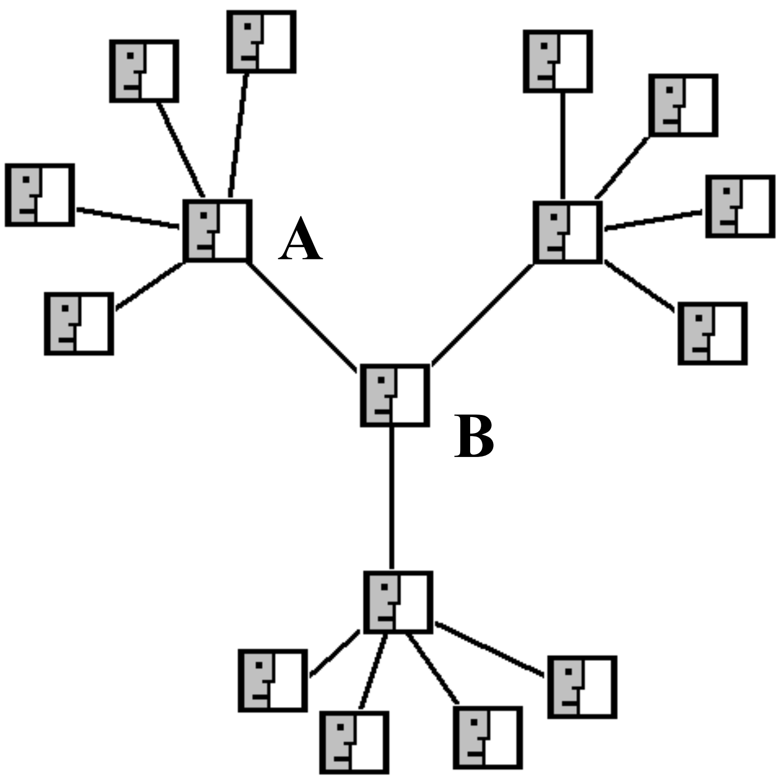


$$CC(i) = \frac{N-1}{\sum_j d(i,j)}$$

	A	B	C	D	E	$\sum_{j=1}^{n} d(i,j)$	CC
A	0	1	1	2	2	**6**	**0.67**
B	1	0	1	1	1	**4**	**1.00**
C	1	1	0	2	2	**6**	**0.67**
D	2	1	2	0	2	**7**	**0.57**
E	2	1	2	2	0	**7**	**0.57**

- Node B is the most central one in spreading information from it to the other nodes in the network.
- DC, BC and CC all agree

When closeness centrality and degree centrality are different



- A is the most central according to the degree
- B is the most central according to closeness and betweenness

Which is the most central node?

Eigenvector Centrality: Extending the Concept of Degree

- Make x_i proportional to the **average of the centralities** of its i's network neighbors

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j$$

where λ is a constant. In matrix-vector notation we can write

$$\mathbf{x} = \frac{1}{\lambda} \mathbf{A}\mathbf{x}$$

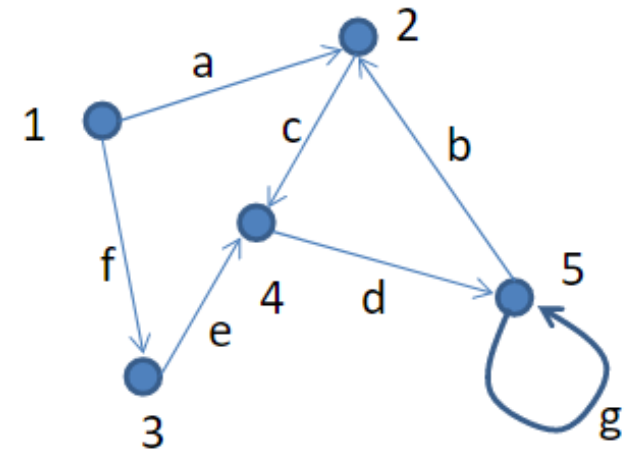
In order to make the centralities non-negative we select the ***eigenvector*** corresponding to the ***principal eigenvalue*** (Perron-Frobenius theorem).

Goals for today: Network analysis

1. **Introduction to networks**
   - Network types: regulatory, metab., signal., interact., func.
   - Bayesian (probabilistic) and Algebraic views
2. **Network Centrality Measures**
   - Local centrality metrics (degree, betweenness, closeness, etc)
   - Global centrality metrics (eigenvector centrality, page-rank)
3. **Linear Algebra Review: eigenvalues, SVD, low-rank approximations**
   - Eigenvector and singular vector decomposition
   - Low rank approximations, Wigner semicircle law
4. **Sparse Principal Component Analysis**
   - Lasso and Elastic lasso
   - PCA and Sparse PCA
5. **Network Communities and Modules**
   - Guilt by association
   - Maximum cliques, density-based modules and spectral clustering
6. **Network Diffusion Kernels and Deconvolution**
   - Network diffusion kernels
   - Network deconvolution

Matrix interpretation of graphs

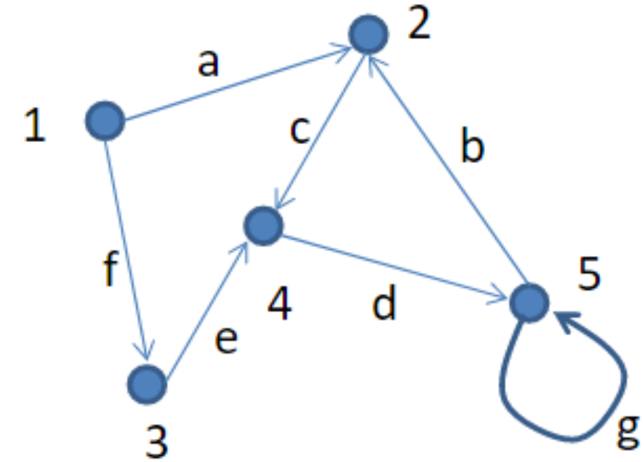
- Graph (V,E) as a matrix
  - Choose an ordering of vertices
  - Number them sequentially
  - Fill in |V|x|V| matrix
  - Called "incidence matrix" of graph
- Observations:
  - Diagonal entries: weights on self-loops
  - Symmetric matrix ←→ undirected graph
  - Lower triangular matrix ←→ no edges from lower numbered nodes to higher numbered nodes
  - Dense matrix ←→ clique (edge between every pair of nodes)



$$
\begin{array}{c c}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} &
\begin{bmatrix}
0 & a & f & 0 & 0 \\
0 & 0 & 0 & c & 0 \\
0 & 0 & 0 & e & 0 \\
0 & 0 & 0 & 0 & d \\
0 & b & 0 & 0 & g
\end{bmatrix}
\end{array}
$$

Matrix operations on graphs

- Matrix computation: y = Ax
- Graph interpretation:
  - Each node i has two values (labels) x(i) and y(i)
  - Each node i updates its label y using the x value from each of its neighbors j, scaled by the label on edge (i,j)
- Observation:
  - Graph perspective shows dense MVM is just a special case of sparse MVM

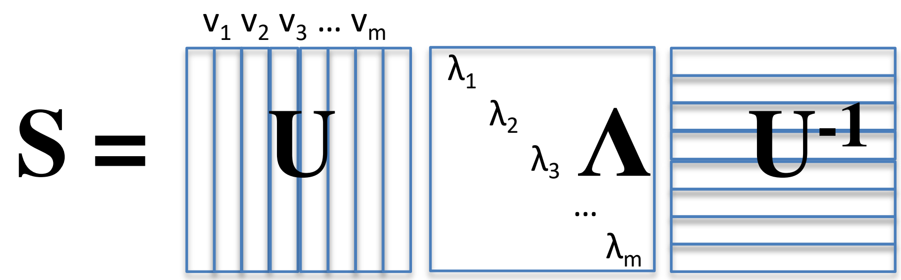


$$
\begin{array}{c}
\begin{array}{cccccc} & 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}
\begin{bmatrix}
0 & a & f & 0 & 0 \\
0 & 0 & 0 & c & 0 \\
0 & 0 & 0 & e & 0 \\
0 & 0 & 0 & 0 & d \\
0 & b & 0 & 0 & g
\end{bmatrix}
\end{array}
$$

A

Eigen/diagonal Decomposition

- Let $\mathbf{S} \in \mathbb{R}^{m \times m}$ be a **square** matrix with ***m*** **linearly independent eigenvectors** (a "non-defective" matrix)



- **Theorem**: Exists an **eigen decomposition**

diagonal

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{-1}$$

Unique for distinct eigen-values

  – (cf. matrix diagonalization theorem)

- Columns of ***U*** are **eigenvectors** of ***S***

- Diagonal elements of $\boldsymbol{\Lambda}$ are **eigenvalues** of $\mathbf{S}$

$$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_m), \quad \lambda_i \geq \lambda_{i+1}$$

Singular Value Decomposition

For an $m \times n$ matrix $\mathbf{A}$ of rank r there exists a factorization (Singular Value Decomposition = **SVD**) as follows:

$$A = \underbrace{U}_{m\times m}\underbrace{\Sigma}_{m\times n}\underbrace{V^T}_{V \text{ is } n\times n}$$

The columns of ***U*** are orthogonal eigenvectors of $\boldsymbol{AA^T}$.

The columns of ***V*** are orthogonal eigenvectors of $\boldsymbol{A^TA}$.

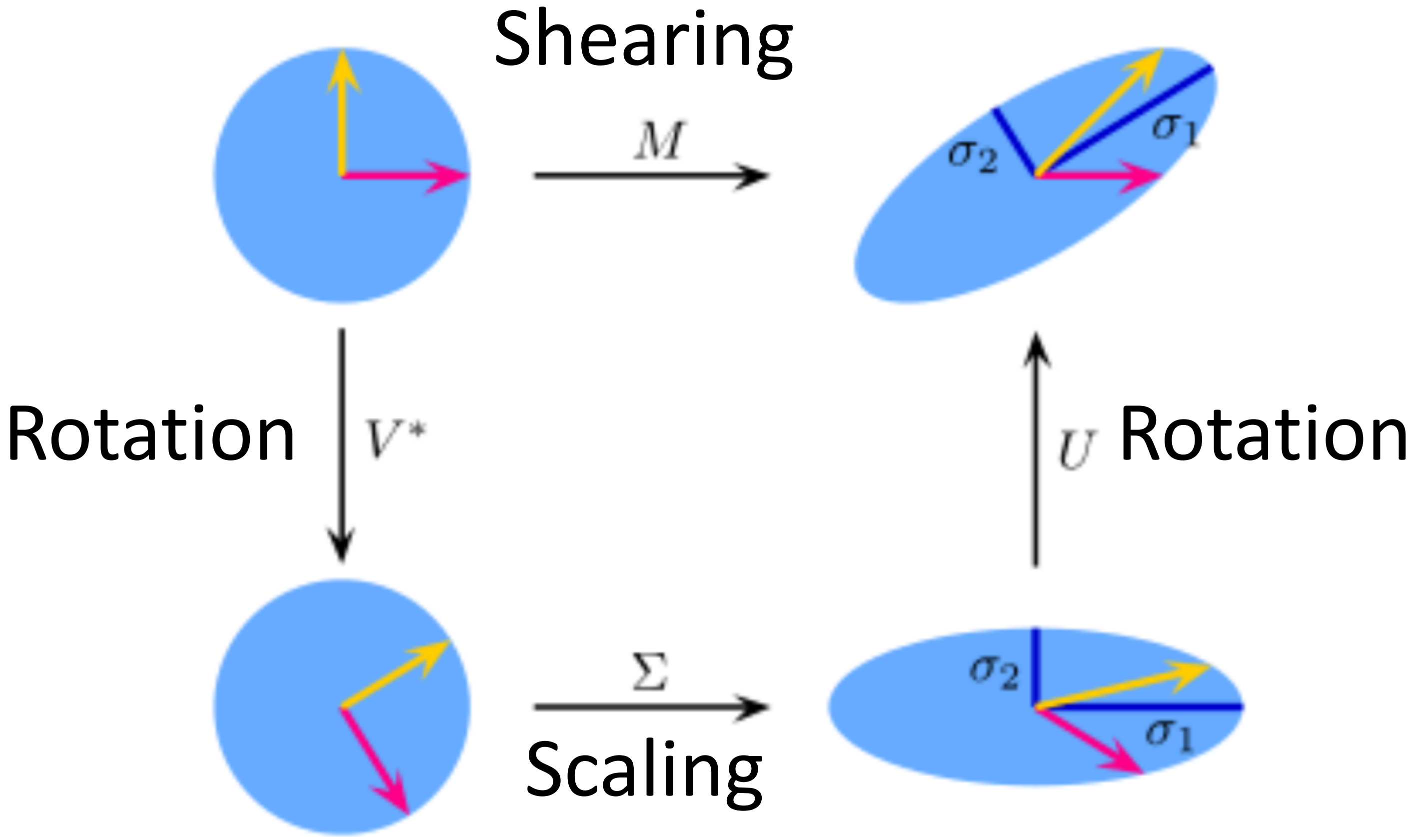
Eigenvalues $\lambda_1 \ldots \lambda_r$ of $\boldsymbol{AA^T}$ are the eigenvalues of $\boldsymbol{A^TA}$.

$$\sigma_i = \sqrt{\lambda_i}$$

$$\Sigma = diag(\sigma_1 \ldots \sigma_r)$$

Singular values.

Geometric interpretation of SVD



$$M = U \cdot \Sigma \cdot V^*$$

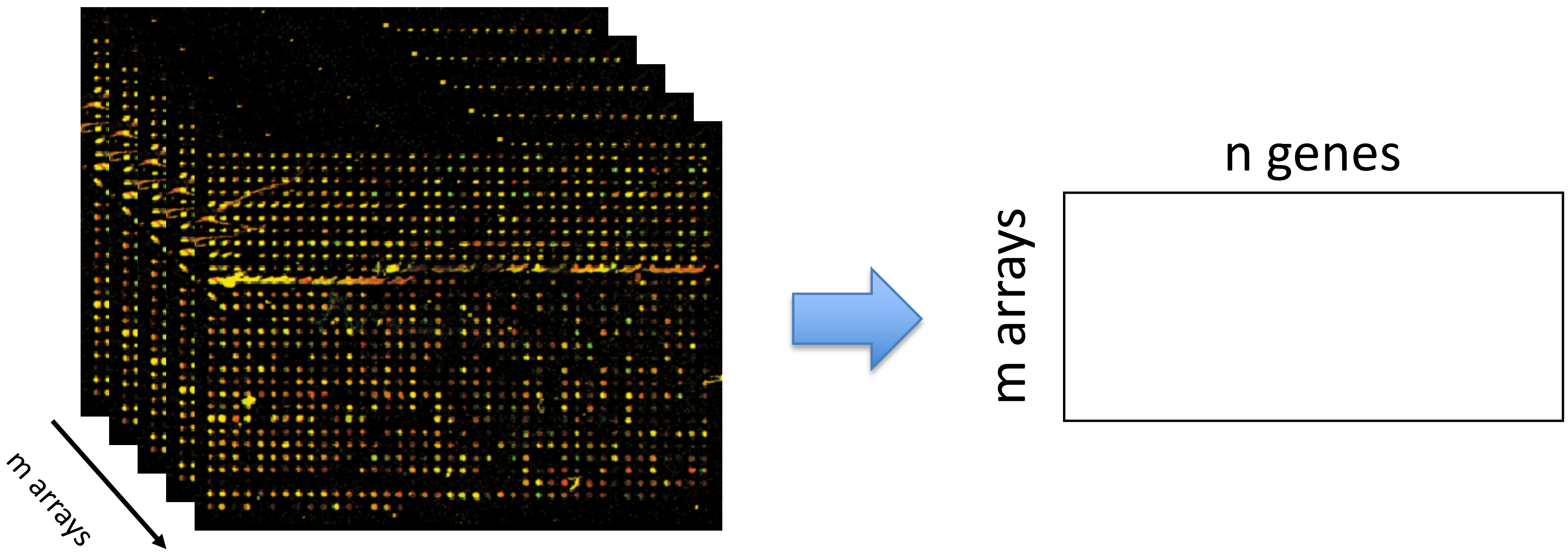
$$Mx = M(x) = U(\ S(\ V^*(x)\ )\ )$$

Goals for today: Network analysis

1. **Introduction to networks**
   - Network types: regulatory, metab., signal., interact., func.
   - Bayesian (probabilistic) and Algebraic views
2. **Network Centrality Measures**
   - Local centrality metrics (degree, betweenness, closeness, etc)
   - Global centrality metrics (eigenvector centrality, page-rank)
3. **Linear Algebra Review: eigenvalues, SVD, low-rank approximations**
   - Eigenvector and singular vector decomposition
   - Low rank approximations, Wigner semicircle law
4. **Sparse Principal Component Analysis**
   - Lasso and Elastic lasso
   - PCA and Sparse PCA
5. **Network Communities and Modules**
   - Guilt by association
   - Maximum cliques, density-based modules and spectral clustering
6. **Network Diffusion Kernels and Deconvolution**
   - Network diffusion kernels
   - Network deconvolution

Sparse Principal Component Analysis

Gene Expression Data
(RNA-Seq, Microarray, …)



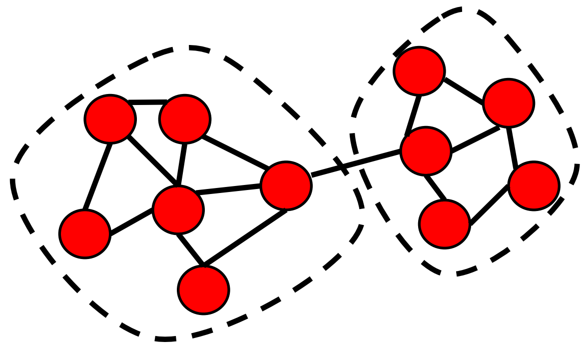
- n = 20k genes, m = 100 arrays
- $n >> m$

Goals for today: Network analysis

1. **Introduction to networks**
   - Network types: regulatory, metab., signal., interact., func.
   - Bayesian (probabilistic) and Algebraic views
2. **Network Centrality Measures**
   - Local centrality metrics (degree, betweenness, closeness, etc)
   - Global centrality metrics (eigenvector centrality, page-rank)
3. **Linear Algebra Review: eigenvalues, SVD, low-rank approximations**
   - Eigenvector and singular vector decomposition
   - Low rank approximations, Wigner semicircle law
4. **Sparse Principal Component Analysis**
   - Lasso and Elastic lasso
   - PCA and Sparse PCA
5. **Network Communities and Modules**
   - Guilt by association
   - Maximum cliques, density-based modules and spectral clustering
6. **Network Diffusion Kernels and Deconvolution**
   - Network diffusion kernels
   - Network deconvolution

Modularity of regulatory networks

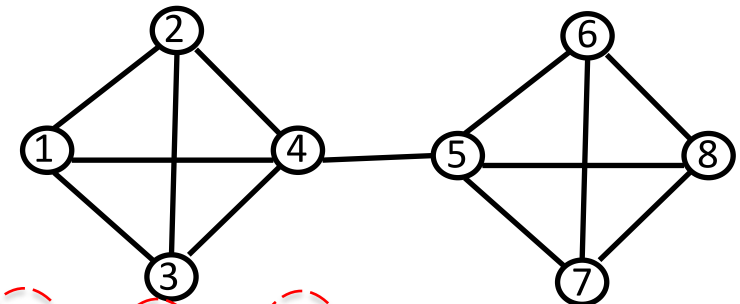
- Modular: Graph with densely connected subgraphs



- Genes in modules involved in similar functions and co-regulated
- Modules can be identified using graph partitioning algorithms
  – Markov Clustering Algorithm (random walks on graph)
  – Girvan-Newman Algorithm (hierarchical communities)
  – Spectral partitioning (eigenvalue of Laplacian matrix)

Newman PNAS 2007

Eigen decomposition-example



$$L = U \Sigma U^{-1}$$

$$U = \begin{matrix}
0.3536 & -0.3825 & 0.2714 & -0.1628 & -0.7783 & 0.0495 & -0.0064 & -0.1426 \\
0.3536 & -0.3825 & 0.5580 & -0.1628 & 0.6066 & 0.0495 & -0.0064 & -0.1426 \\
0.3536 & -0.3825 & -0.4495 & 0.6251 & 0.0930 & 0.0495 & -0.3231 & -0.1426 \\
0.3536 & -0.2470 & -0.3799 & -0.2995 & 0.0786 & -0.1485 & 0.3358 & 0.6626 \\
0.3536 & 0.2470 & -0.3799 & -0.2995 & 0.0786 & -0.1485 & 0.3358 & -0.6626 \\
0.3536 & 0.3825 & 0.3514 & 0.5572 & -0.0727 & -0.3466 & 0.3860 & 0.1426 \\
0.3536 & 0.3825 & 0.0284 & -0.2577 & -0.0059 & -0.3466 & -0.7218 & 0.1426 \\
0.3536 & 0.3825 & 0.0000 & 0.0000 & 0.0000 & 0.8416 & -0.0000 & 0.1426
\end{matrix}$$

$$\Sigma = \begin{matrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.3542 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 4.0000 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 4.0000 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 4.0000 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 4.0000 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 4.0000 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 5.6458
\end{matrix}$$

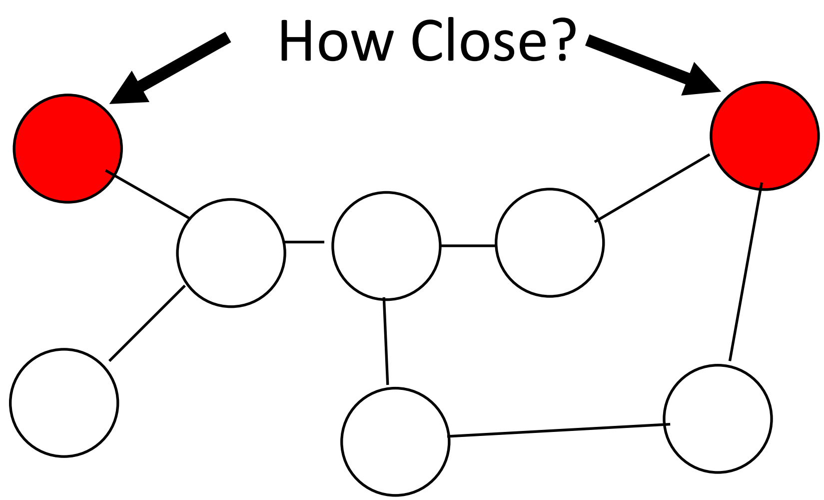
- First smallest eigenvalue of Laplacian matrix is always zero.
- Second smallest eigenvector of Laplacian matrix characterizes a network partition.

Goals for today: Network analysis

1. **Introduction to networks**
   - Network types: regulatory, metab., signal., interact., func.
   - Bayesian (probabilistic) and Algebraic views
2. **Network Centrality Measures**
   - Local centrality metrics (degree, betweenness, closeness, etc)
   - Global centrality metrics (eigenvector centrality, page-rank)
3. **Linear Algebra Review: eigenvalues, SVD, low-rank approximations**
   - Eigenvector and singular vector decomposition
   - Low rank approximations, Wigner semicircle law
4. **Sparse Principal Component Analysis**
   - Lasso and Elastic lasso
   - PCA and Sparse PCA
5. **Network Communities and Modules**
   - Guilt by association
   - Maximum cliques, density-based modules and spectral clustering
6. **Network Diffusion Kernels and Deconvolution**
   - Network diffusion kernels
   - Network deconvolution

Network Diffusion Kernels

- Define closeness of two nodes in the network



- One way: use weighted shortest path
- **Invariant** to the position of edges over a path

Conclusion: Network analysis

1. **Introduction to networks**
   - Network types: regulatory, metab., signal., interact., func.
   - Bayesian (probabilistic) and Algebraic views
2. **Network Centrality Measures**
   - Local centrality metrics (degree, betweenness, closeness, etc)
   - Global centrality metrics (eigenvector centrality, page-rank)
3. **Linear Algebra Review: eigenvalues, SVD, low-rank approximations**
   - Eigenvector and singular vector decomposition
   - Low rank approximations, Wigner semicircle law
4. **Sparse Principal Component Analysis**
   - Lasso and Elastic lasso
   - PCA and Sparse PCA
5. **Network Communities and Modules**
   - Guilt by association
   - Maximum cliques, density-based modules and spectral clustering
6. **Network Diffusion Kernels and Deconvolution**
   - Network diffusion kernels
   - Network deconvolution