# LECTURE #1

JOHN GOUTSIAS
WHITAKER BIOMEDICAL ENGINEERING INSTITUTE
THE JOHNS HOPKINS UNIVERSITY
BALTIMORE, MD 21218

## INTRODUCTION & GRAPH THEORY

# Introduction



- What is a complex interaction network?
  - A system of relatively simple components that interact with each other to accomplish a desired objective.
- Complex interaction networks are at the core of many problems of scientific and engineering interest:
  - Chemical reaction networks.
  - Cellular (signaling, transcriptional, metabolic) networks.
  - Pharmacokinetic networks.
  - Epidemiological networks.
  - Ecological networks.
  - Social networks.
  - Neural networks.
  - Multi-agent networks.

# Introduction



- Modeling approaches:
  - Deterministic.
  - Stochastic.
- Stochastic more appropriate (in most situations) and more general !!
  - Based on a Master Equation.
  - Governs the time evolution of the joint probability mass function of the underlying population process (state of the system).
  - Leads to Markovian dynamics.

# Introduction

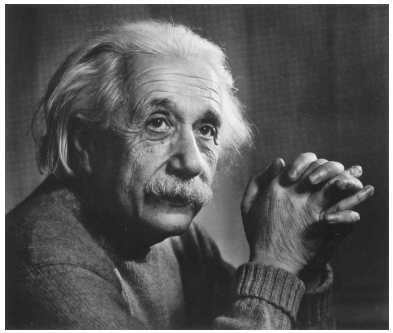


- Due to the nonlinear nature of most Markovian networks, computing the exact solution of the master equation is not possible in general.
- Analysis of nonlinear Markovian networks is difficult.
- Deterministic approaches are theoretically and computationally easier to handle than stochastic approaches.
- Deterministic approaches may fail to predict important system behavior observed in real networks.
  - e.g., cannot predict the emergence of noise-induced behavior, a fundamental property of nonlinear interaction networks with stochastic dynamics.

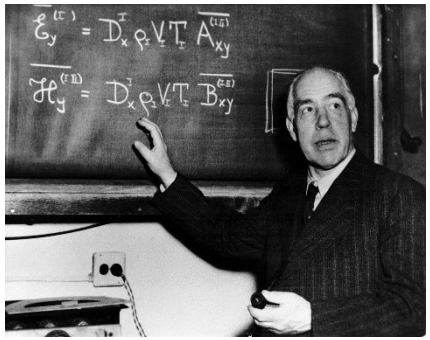
# Introduction



- What Albert Einstein got wrong?

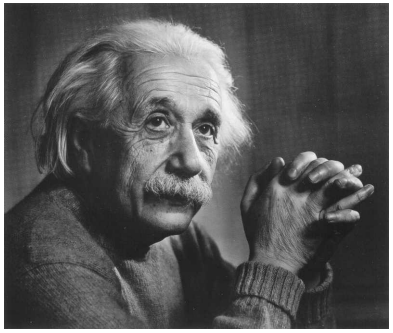


Albert Einstein

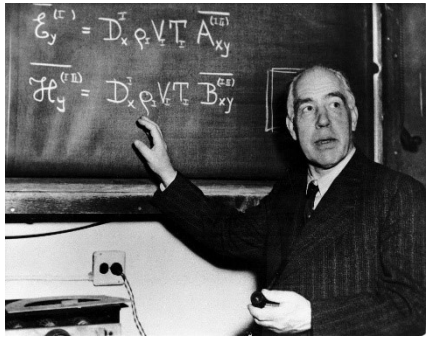


Niels Bohr

"God does not play dice"



Albert Einstein



Niels Bohr

"Einstein, stop telling God what to do"

# Introduction



- The main premise underlying the use of stochastic models for interaction networks is the realization that environmental, demographic, behavioral, and biological factors fluctuate randomly, and that the resulting stochasticity can cause dramatic deviation from what is predicted by deterministic approaches.

# Biochemical Systems and Reaction Networks



- Biochemical reaction networks can serve as archetypal systems when studying dynamics on complex networks.
- Comprised of molecular species and chemical reactions.
- Reactants ⇨ Products.

# Biochemical Systems and Reaction Networks



- Example (quadratic autocatalator with positive feedback):

$$S \rightarrow P$$
$$D + P \rightarrow D + 2P$$
$$2P \rightarrow P + Q$$
$$P + Q \rightarrow 2Q$$
$$P \rightarrow \varnothing$$
$$Q \rightarrow \varnothing$$

# Biochemical Systems and Reaction Networks



- If we set $X_1 = \mathrm{S}, X_2 = \mathrm{P}, X_3 = \mathrm{D}, X_4 = \mathrm{Q}$ and denote the reactions by $1, 2, \ldots, 6$ , then we have:

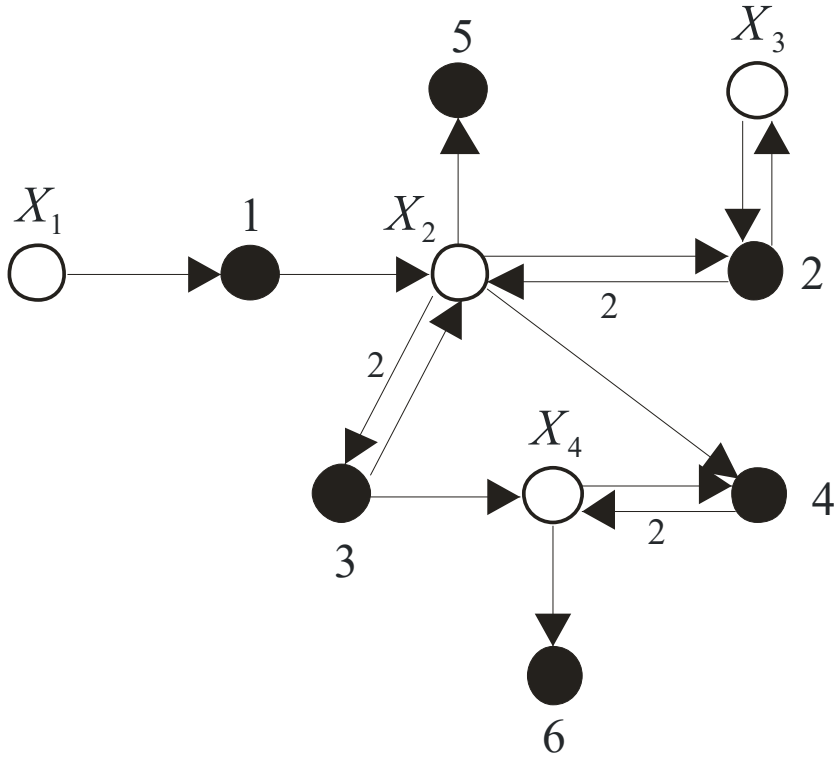
$$\text{reaction 1: } X_1 \rightarrow X_2$$
$$\text{reaction 2: } X_2 + X_3 \rightarrow 2\mathrm{X}_2 + X_3$$
$$\text{reaction 3: } 2X_2 \rightarrow \mathrm{X}_2 + X_4$$
$$\text{reaction 4: } X_2 + X_4 \rightarrow 2X_4$$
$$\text{reaction 5: } X_2 \rightarrow \varnothing$$
$$\text{reaction 6: } X_4 \rightarrow \varnothing$$

# Biochemical Systems and Reaction Networks



$$X_1 \rightarrow X_2$$
$$X_2 + X_3 \rightarrow 2X_2 + X_3$$
$$2X_2 \rightarrow X_2 + X_4$$
$$X_2 + X_4 \rightarrow 2X_4$$
$$X_2 \rightarrow \varnothing$$
$$X_4 \rightarrow \varnothing$$

Reaction Network




Directed, weighted, bipartite graph

# Biochemical Systems and Reaction Networks



- In general, we write:

$$\sum_n \nu_{nm} X_n \rightarrow \sum_n \nu'_{nm} X_n$$

  - $n = 1, 2, \ldots, N$ species
  - $m = 1, 2, \ldots, M$ reactions
  - $\nu_{nm}$ stoichiometric coefficients of reactants
  - $\nu'_{nm}$ stoichiometric coefficients of products
    - How many molecules of the $n$-th species are consumed or produced by the $m$-th reaction.
    - Occurrence of the $m$-th reaction changes the molecular count of the $n$-th species by $s_{nm} = \nu'_{nm} - \nu_{nm}$
    - $s_{nm}$ is the net stoichiometric coefficient.

# Biochemical Systems and Reaction Networks



- Stoichiometric matrices:

$$\mathbf{V} = [\nu_{nm}]$$
$$\mathbf{V}' = [\nu'_{nm}]$$
$$\mathbf{S} = [s_{nm}]$$

- Example

$$X_1 \rightarrow X_2$$
$$X_2 + X_3 \rightarrow 2X_2 + X_3$$
$$2X_2 \rightarrow X_2 + X_4$$
$$X_2 + X_4 \rightarrow 2X_4$$
$$X_2 \rightarrow \varnothing$$
$$X_4 \rightarrow \varnothing$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \quad \mathbf{V}' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & -1 \end{bmatrix}$$

# Biochemical Systems and Reaction Networks



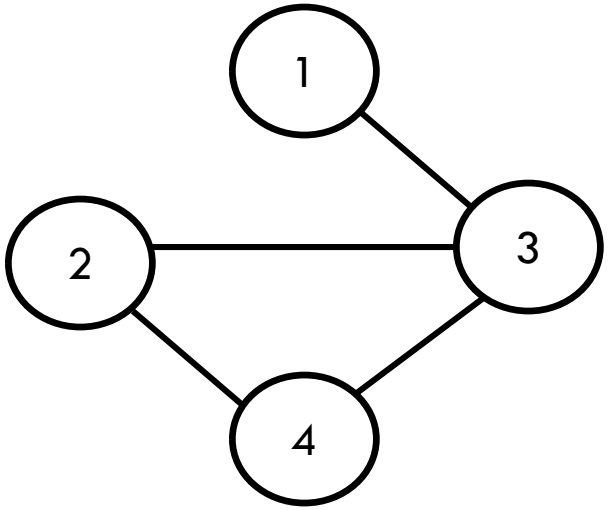
- Given the two stoichiometry matrices $\mathbf{V}$ and $\mathbf{V}'$, we can uniquely reconstruct the chemical reaction system and the network.
- Knowledge of the two stoichiometry matrices completely specify the network topology.
- What happens with the dynamics?
- We will deal with this question later.

# Graph Theory – Standard Graphs

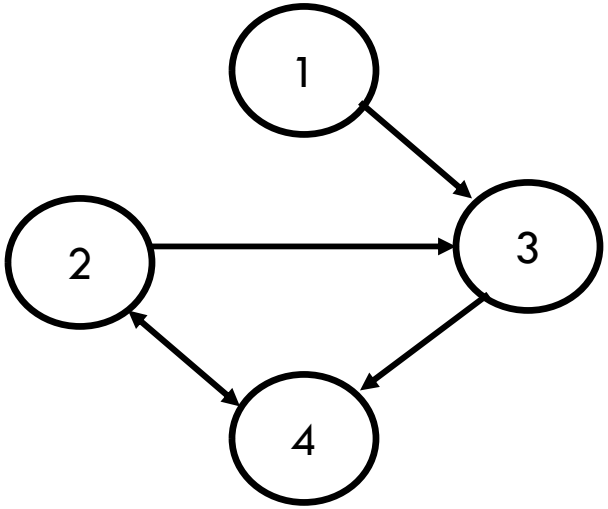


- Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
  - $\mathcal{V}$ set of nodes or vertices
  - $\mathcal{E}$ set of edges connecting nodes

Undirected Graph




Directed Graph



# Graph Theory – Adjacency Matrix



- Adjacency matrix

$$\mathbf{A} = [a_{ij}]$$

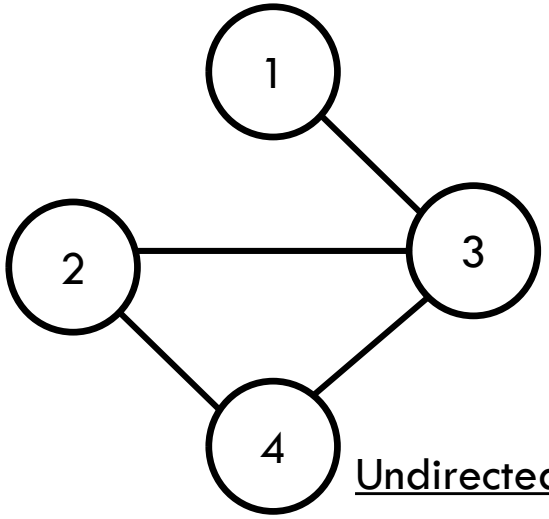
$$a_{ij} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}$$

- For an undirected graph

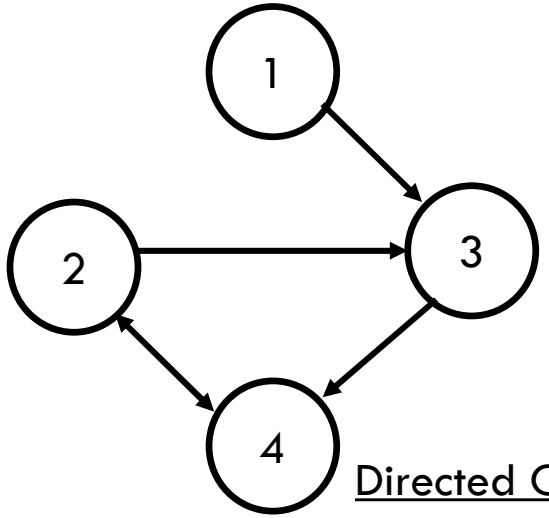
$$\mathbf{A}^T = \mathbf{A}$$

# Graph Theory – Adjacency Matrix






Undirected Graph




Directed Graph

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

# Graph Theory – Paths



- A path in a graph is a sequence of vertices such that from each of its vertices there is an edge to the next vertex in the sequence.
- The length of the path is given by the number of edges traversed in order to go from the initial node to the final node of the path.
- If a path exists between two nodes, then we say that the nodes are connected.

# Graph Theory – Paths



- The number of paths of length $n$ connecting the $i$-th node to the $j$-th node can be computed by $N_{ij}(n) = [\mathbf{A}^n]_{ij}$.

- The shortest path length between the $i$-th node and the $j$-th node is denoted by $l_{ij}$ .

- If $l_{ij} = \infty$ , then there is no path between nodes $i$ and $j$.

# Graph Theory – Node Degree



- The degree $k_i$ of the $i$-th node is the number of edges that connect to that node.
- In directed graphs we have in-degree and out-degree:

$$k_i = k_{in,i} + k_{out,i}$$

# Graph Theory – Degree Distribution



- Degree distribution:

$$p(k) = \frac{\#\text{ nodes with degree } k}{\text{total \# of nodes}}$$

$$\sim \text{ probability that a randomly selected node will have degree } k$$

- Directed graphs may have in- and out- degree distribution.

# Graph Theory

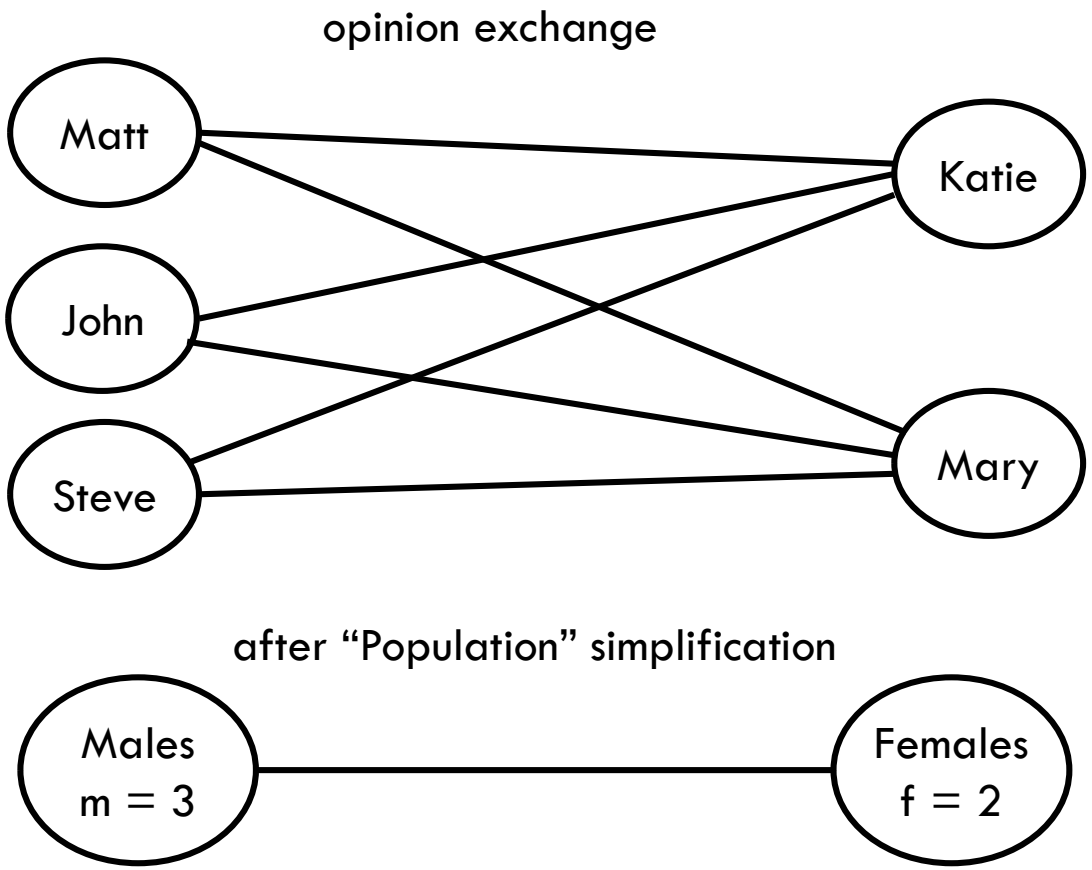


- Other quantities used to describe the topology of a graph:
  - Degree centrality.
  - Betweenness distribution.
  - Clustering coefficient.
  - Cliques.
  - Etc.

# Graph Theory – Population Simplification



- Population simplification (coarse graining)



# Graph Theory – Weighted Graphs



- Weighted graphs.
  - An edge can be associated with a weight which quantifies the strength of connection between two nodes in the network or some other attribute.
  - The adjacency matrix $\mathbf{A}$ is now replaced by the weighting matrix

$$\mathbf{W} = [w_{ij}]$$

weight of the edge connecting the $i$-th node with the $j$-th node

  - Node strength:

$$s_i = \sum_{j \in \mathcal{V}} w_{ij}$$

# Graph Theory – Weighted Graphs



- For directed graphs we may also have in-strength and out-strength:

$$s_{in,i} = \sum_{j \in \mathcal{V}_{in}(i)} w_{ij} \ , \ \mathcal{V}_{in}(i) : \text{all nodes connecting to node } i \text{ via an in-edge}$$

$$s_{out,i} = \sum_{j \in \mathcal{V}_{out}(i)} w_{ij} \ , \ \mathcal{V}_{\text{out}}(i) : \text{all nodes connecting to node } i \text{ via an out-edge}$$

$$s_i = s_{in,i} + s_{out,i}$$

# Graph Theory – Bipartite Graphs



- Standard graphs cannot be sufficiently descriptive to epitomize a general complex network.
- Biochemical reaction networks require the use of two types of nodes, one for the species and the other for the reactions.
- In general, we have:
  - Interaction components (chemical compounds, animal species, individual people).
  - Processing units (reactions, predator-pray relationships, processing of social interactions).

# Graph Theory – Bipartite Graphs



- We need the notion of bipartite graphs:

$$\mathcal{B} = (\mathcal{C}, \mathcal{P}, \mathcal{E})$$

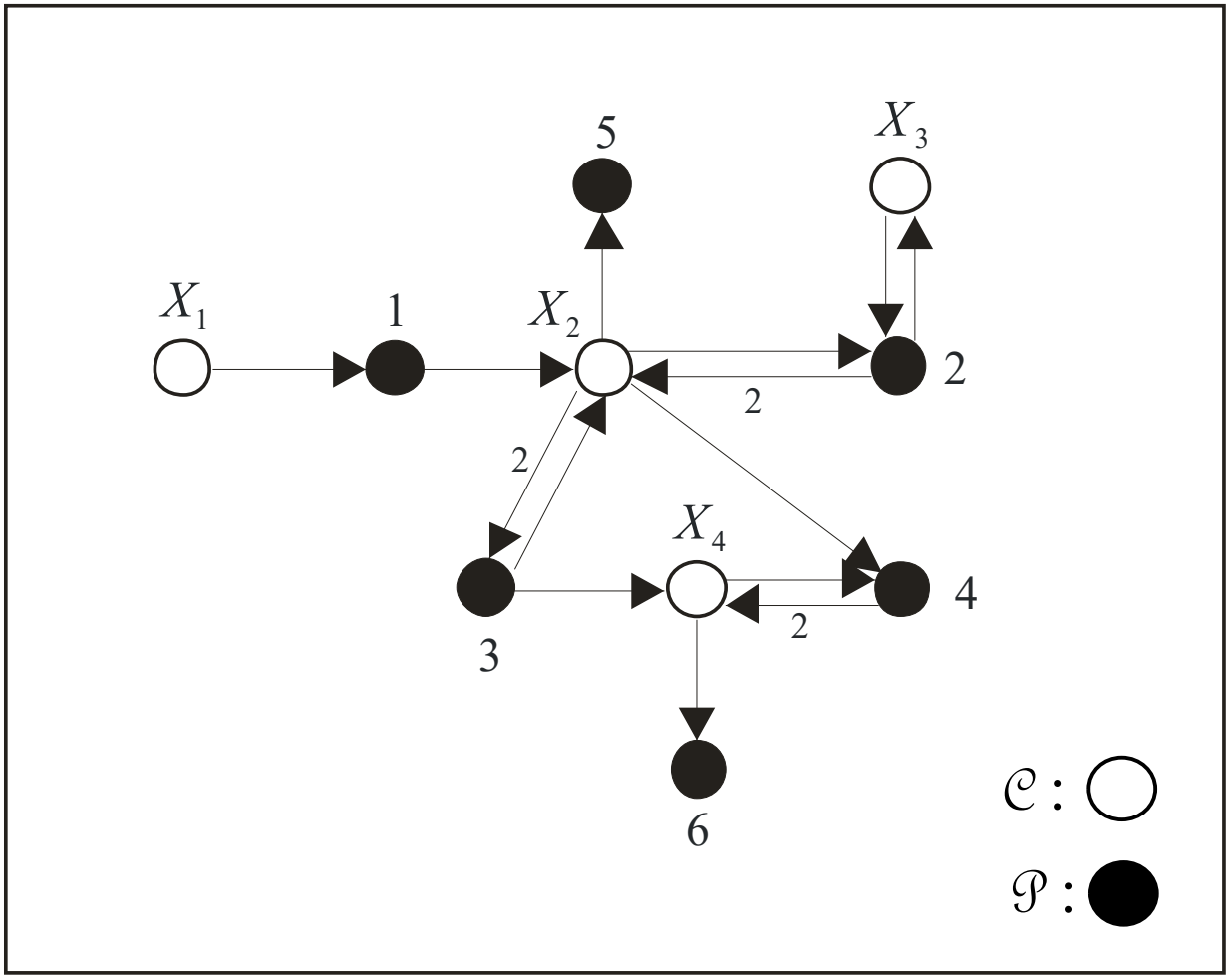
nodes of type 1 — nodes of type 2

- The edges may only connect nodes in $\mathcal{C}$ to nodes in $\mathcal{P}$ and vice-versa.
- Edges may be directed or undirected and weighted of weight-free.
- $\mathcal{C}$ represents interacting components.
- $\mathcal{P}$ represents processing units.

# Graph Theory – Bipartite Graphs



$$X_1 \rightarrow X_2$$
$$X_2 + X_3 \rightarrow 2X_2 + X_3$$
$$2X_2 \rightarrow X_2 + X_4$$
$$X_2 + X_4 \rightarrow 2X_4$$
$$X_2 \rightarrow \varnothing$$
$$X_4 \rightarrow \varnothing$$



# Network Topology



- Discussions on network topology are usually simplified by focusing on standard (i.e., not bipartite) graphical representations of networks.
- We can map any bipartite graph

$$\mathcal{B} = (\mathcal{C}, \mathcal{P}, \mathcal{E})$$

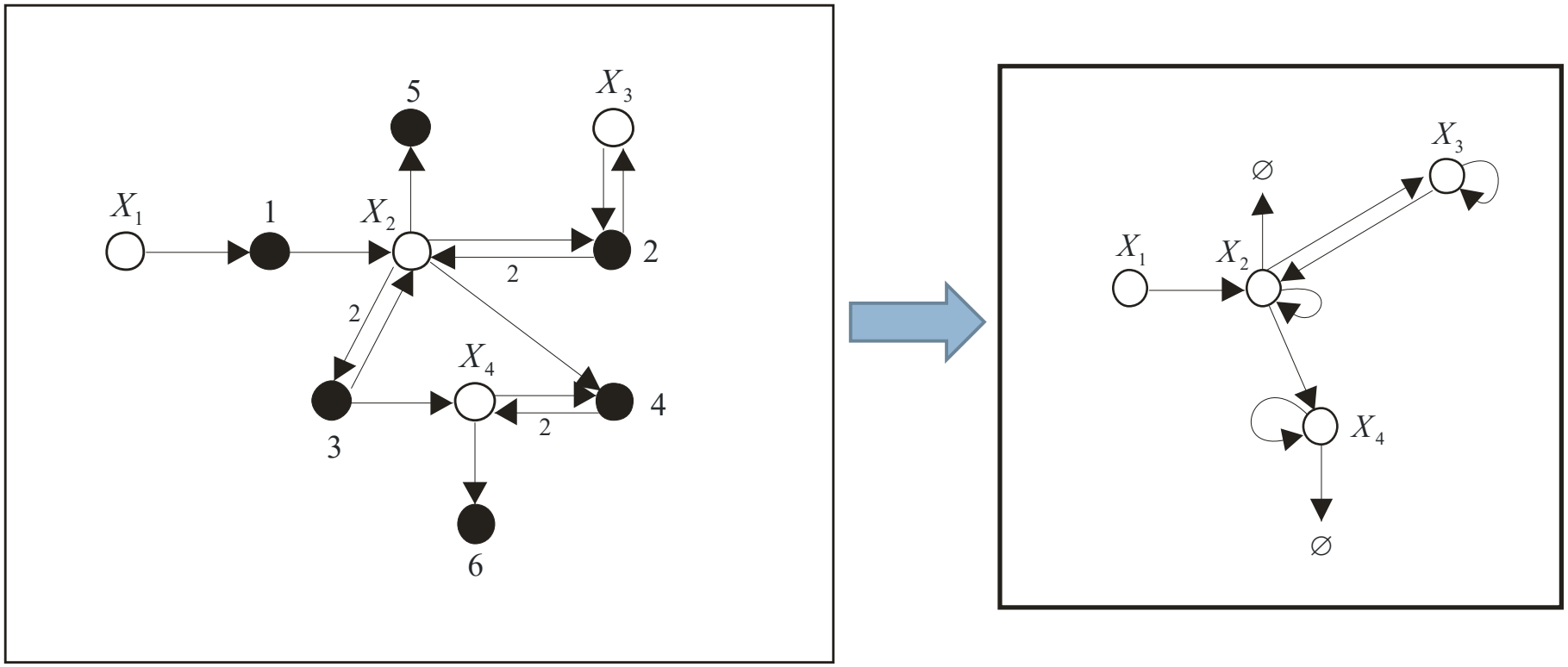
onto a standard graph

$$\mathcal{G} = (\mathcal{C}, \mathcal{E}_0)$$

by inducing an edge in $\mathcal{E}_0$ between nodes in $\mathcal{C}$ only when these nodes are connected to the same node in $\mathcal{P}$ .

# Network Topology





# Degree Distribution



- Degree distribution:

$$p(k) = \frac{\#\text{ nodes with degree } k}{\text{total \# of nodes}}$$

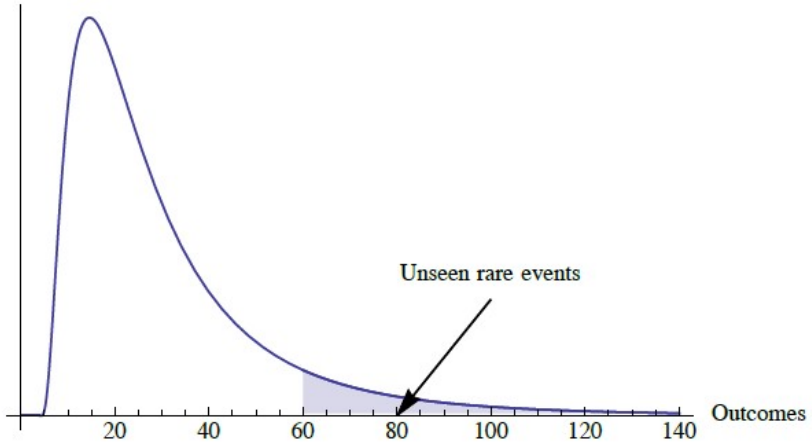
$$\sim \text{ probability that a randomly selected node will have degree } k$$

- Directed graphs may have in- and out- degree distribution.

# Scale-free degree distribution



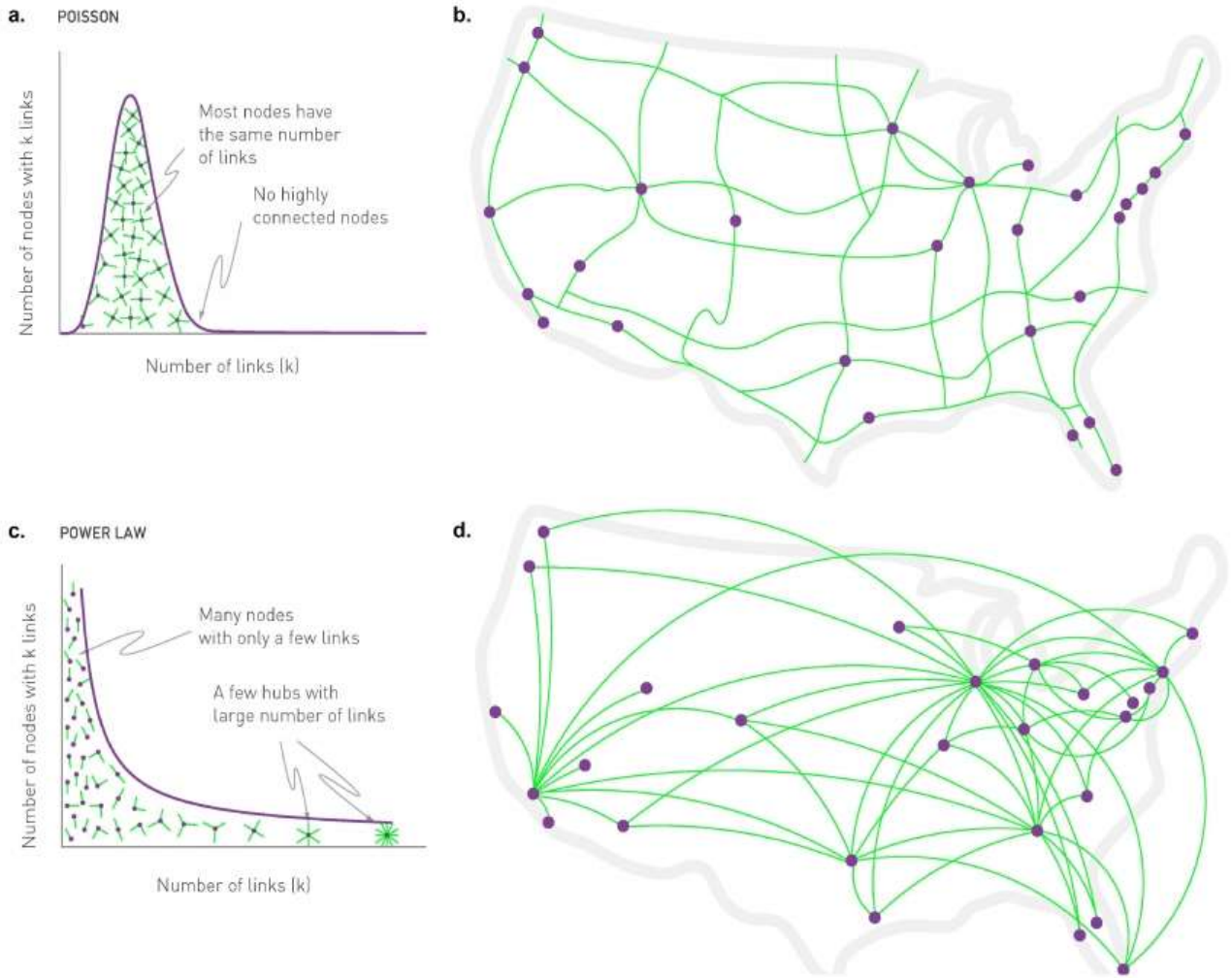
- The degree distribution of many complex networks is characterized by "heavy tails."




- This means that real world networks have an unusually high probability of containing nodes with many connecting edges (representing "hubs").

# Scale-free degree distribution






From: A.-L. Barabási. *Linked: The New Science of Networks*. Plume, New York, 2002.

# Scale-free degree distribution



- It has been suggested that real-world networks (metabolic, protein, social, collaborative, trophic, etc.) are scale-free:

$$p(k) \sim \frac{1}{k^{\gamma}}, \text{ where } \gamma \text{ is a constant that is usually estimated to take values between 2 and 3}$$

- Such degree distribution is called scale-free because

$$p(ck) \sim \frac{1}{(ck)^{\gamma}} = \frac{1}{c^{\gamma}} \frac{1}{k^{\gamma}} \sim \frac{1}{k^{\gamma}} \sim p(k), \text{ for any constant } c > 0$$

- Therefore,

$$p(k) \sim p(ck)$$

- However, the universality of scale-free networks is currently controversial.

# Scale-free Networks



- The most notable characteristic in a scale-free network is the relative commonness of vertices with a degree that greatly exceeds the average.
- The highest-degree nodes are "hubs," which are thought to serve specific purposes in a network.
- The scale-free property strongly correlates with the network's robustness to failure.
- It turns out that the major hubs are closely followed by smaller ones.
- In turn, these smaller hubs are followed by other nodes with an even smaller degree and so on.

# Scale-free Networks



- The previous hierarchy allows for a fault tolerant behavior.
- If failures occur at random and most nodes are those with small degree, the likelihood that a hub would be affected is almost negligible.
- Even if a hub-failure occurs, the network will generally not lose its connectedness, due to the remaining hubs.
- On the other hand, if a few major hubs are taken out of the network, the network is turned into a set of rather isolated graphs.
- Thus, hubs are both a strength and a weakness of scale-free networks.

# Preferential Attachment

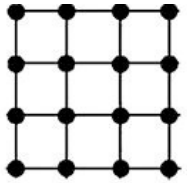


- Preferential attachment is a self-organizing property that can be used to explain the power-law nature of the degree distribution.
- A newly added node to the system may be more likely to form an edge with a node that already has many attached nodes.
- It has been shown that, under appropriate conditions, this behavior produces scale-free results.

# Small-World Property



- A $d$-dimensional rectangular lattice has an average shortest path length that scales as $N^{1/d}$, where $N$ is the number of nodes.



- Many complex networks however have an average shortest path length between two nodes that scales as $\ln N$.
- This is known as small-world property.
- The presence of hubs that connect to many nodes (and thus serve as shortcuts) ensure that most nodes maintain a short path between each other.

# Petri Nets



- Have extensively been used to describe discrete-event distributed systems (of interest to computer science).
- A Petri net is a weighted, directed, bipartite graph in which the nodes represent places and transitions.
- Places model passive system components.
- Transitions correspond to events that interconvert places.

https://en.wikipedia.org/wiki/Petri_net

# Petri Nets



- Directed arcs join places to transitions (connect places that can be converted during a transition) and transitions to places (connect a transition with the corresponding products).
- Weights associated with arcs indicate the multiplicity of the arc.
- Each place is associated with tokens, indicating the number of existing places.

# Petri Nets



- Whether or not a transition takes place is described by a rule, which may be deterministic or stochastic.
- This rule depends on the number of tokens available in the places connecting to the transition by incoming arcs.
- The occurrence of a transition results in removing a token from the input places and adding a token to the output places of the transition.

# Petri Nets



- Petri nets are identical to the reaction networks discussed before.
  - Places ⇒ species
  - Transitions ⇒ reactions
  - Tokens ⇒ species population.