GENERAL ASSEMBLY

NETWORK ANALYSIS

Matt Brems

DSI+

LEARNING OBJECTIVES

- By the end of this lesson, students should be able to:
  - Describe the relevance of networks.
  - Identify undirected, directed, cyclic, acyclic, connected, disconnected networks, and multigraphs.
  - Identify the importance of degree and adjacency matrix.
  - Use the NetworkX package to do basic network analysis.

NETWORKS

“I think the next century will be the century of complexity.”
– Stephen Hawking

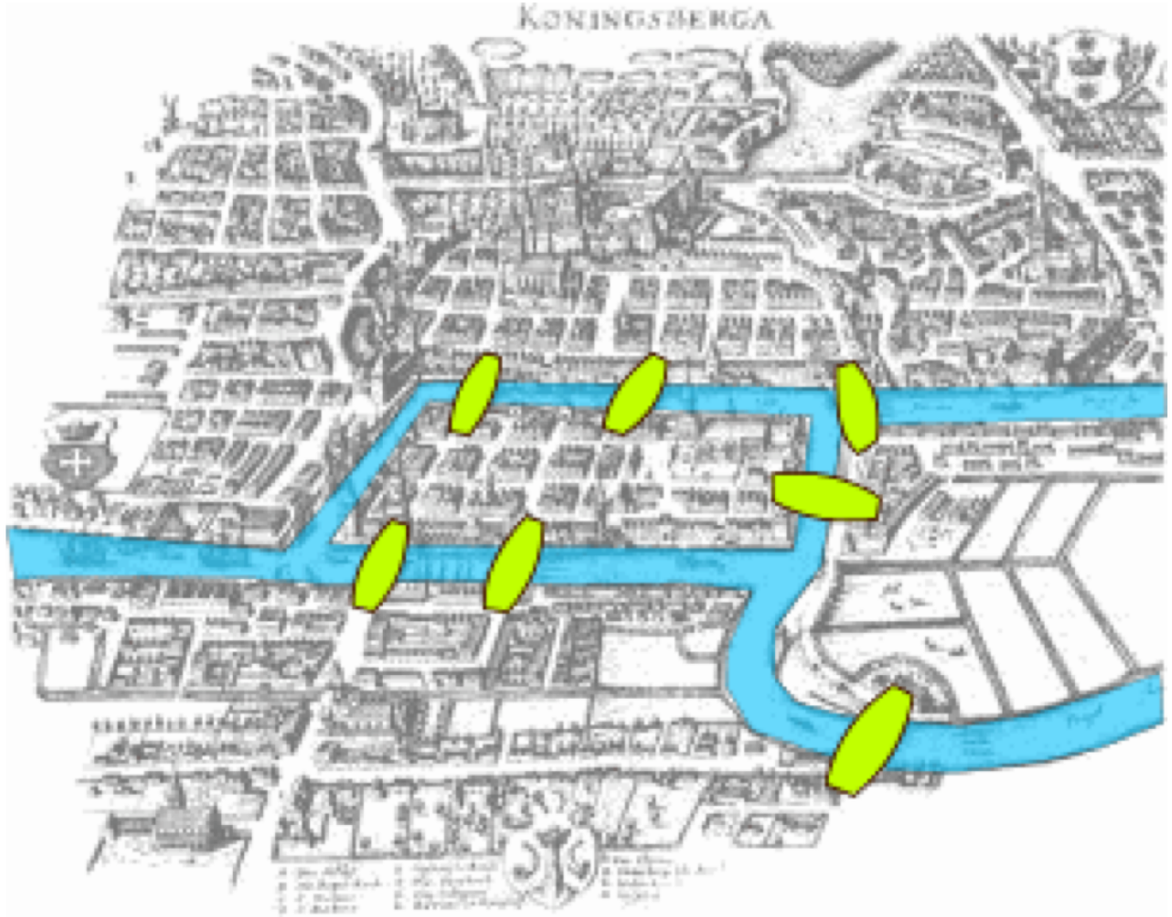
EXAMPLES OF NETWORKS

- Social network: professional, friendship, and family ties
- Utility network: power grid, generators, transmission lines
- Biological network: interactions among genes, proteins, and metabolites.
- Trade network: exchanging goods and services

NETWORK ANALYSIS

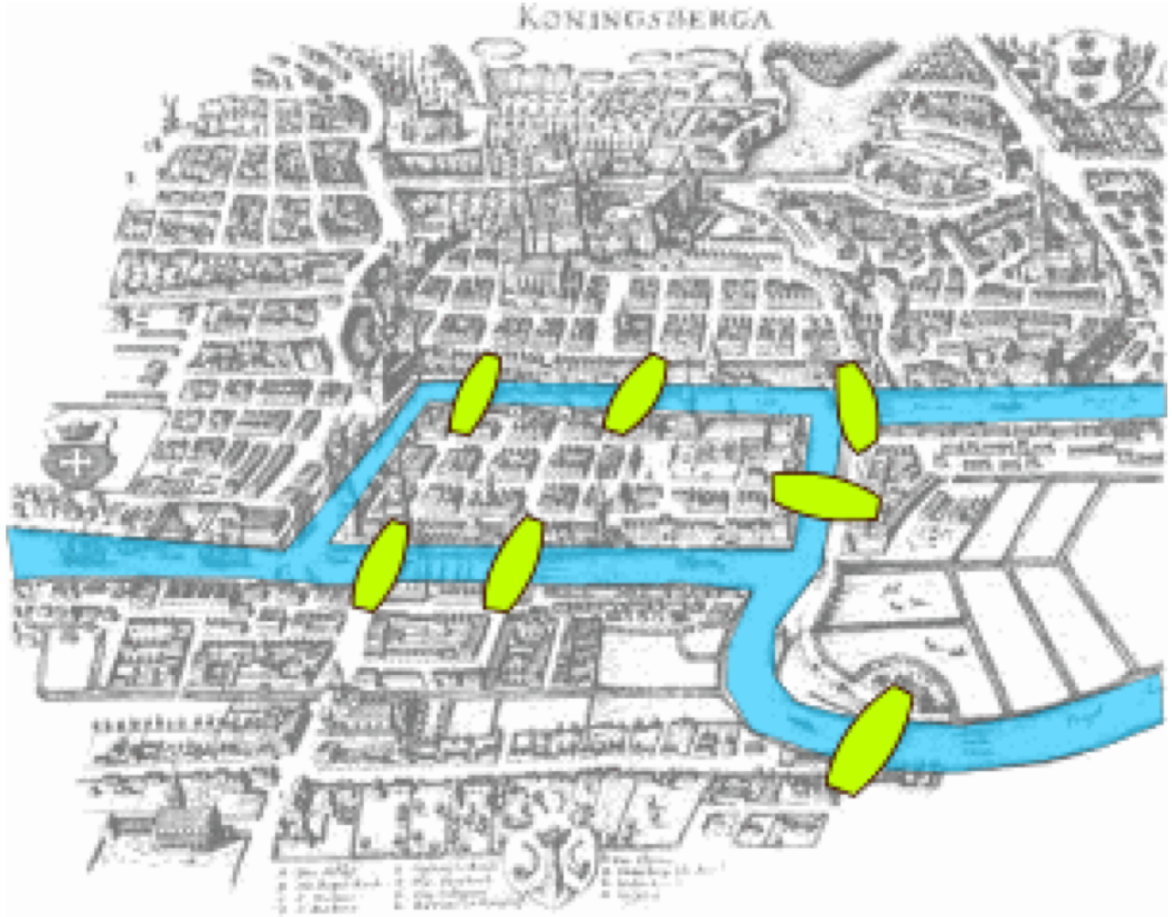
- Can we take these examples above, boil them down to very basic components, and come up with a set of rules that apply for all networks, regardless of application?
    - This is, roughly, the area of network science.
    - You may also hear the terms “graph theory” and “topology,” which are very closely related to “network science.”

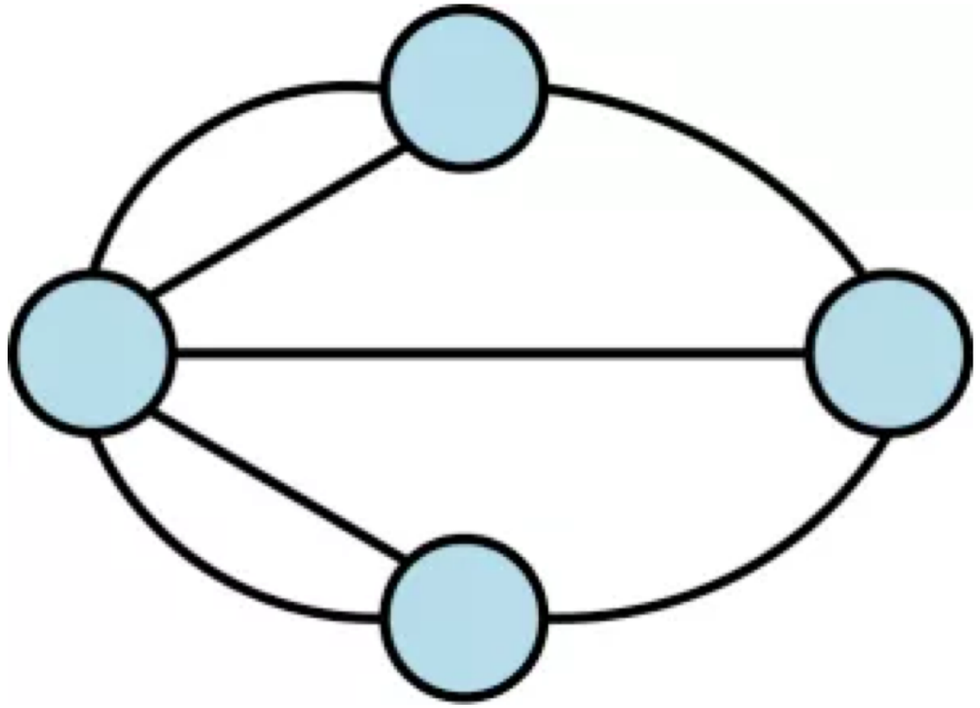
SEVEN BRIDGES OF KÖNIGSBERG



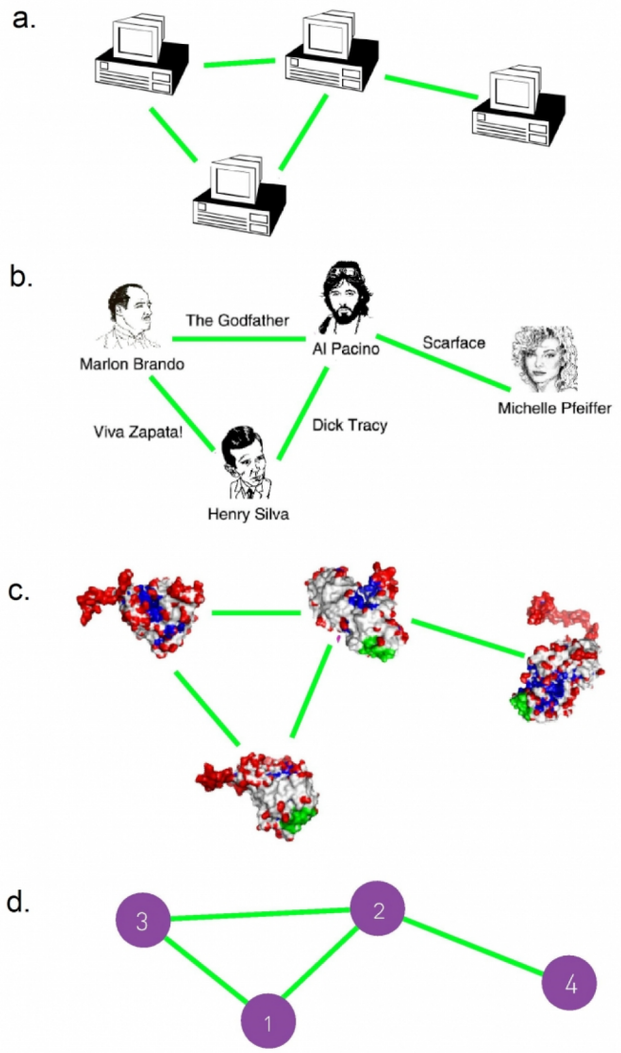
- Can we traverse all seven bridges exactly once?

SEVEN BRIDGES OF KÖNIGSBERG





NETWORKS



- By representing complex systems with **nodes** and **edges**, we can directly compare different systems.
    - If a rule applies to one network with a given structure, that rule should apply to all other networks with the same structure!

TYPES OF NETWORKS

- **Undirected:** An undirected network is one where the connection extends in both directions.

- **Directed:** A directed network is one where the connection may only flow in one direction.

TYPES OF NETWORKS

- **Cyclic:** A cyclic network is one that contains at least one cycle.
  - A cycle exists when a node can be connected to itself by traversing at least one edge.

- **Acyclic:** An acyclic network is one that contains no cycles.

TYPES OF NETWORKS

- **Multigraph:** A multigraph may have multiple links connecting the same pair of nodes.

OTHER NETWORK TERMINOLOGY

- **Degree:** The degree of a node in an undirected graph is given by the number of links of that node.
  - **In-degree:** The number of links heading **into** that node.
  - **Out-degree:** The number of links heading **out of** that node.

NETWORK SUMMARY STATISTICS

- What does the average degree of a network tell us?

- What does the degree distribution of a network tell us?

REPRESENTING A NETWORK

- Visually representing a network can be difficult. (We'll get into that later.)

- We often represent a network with an **adjacency matrix**.

USES OF ADJACENCY MATRICES

- Our computers and modeling techniques do well with matrices.

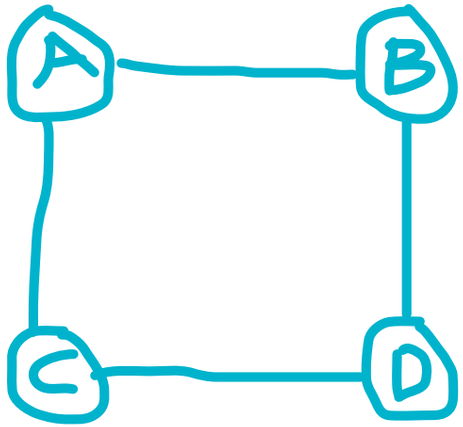
- If we take our adjacency matrix A and raise it to the n^{th} power, we can see how many paths of length n exist from node to node.

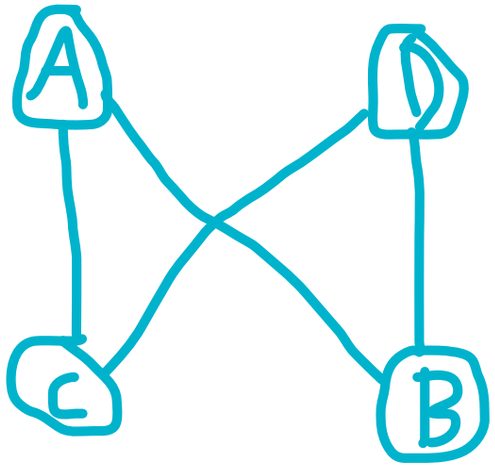
- It's useful for detecting communities in networks.
  - Leading eigenvectors can be used to detect communities.

FINAL TERMS

- **Minimum Path Length:** The minimum number of edges required to traverse to connect node *i* and node *j*.
- **Diameter:** The maximum shortest path length of a network.
- **Connected:** Two nodes *i* and *j* are connected if there exists a path from *i* to *j*.
- **Isomorphic:** Informally, we say two graphs are isomorphic if we can twist one graph to look exactly like the other without cutting or gluing.

EXAMPLE OF ISOMORPHIC GRAPH





UNSUPERVISED LEARNING

- **Unsupervised learning** is where we have, as part of our training data, no observed Y values.

- **Supervised learning** is where we have observed Y values as part of our training data.

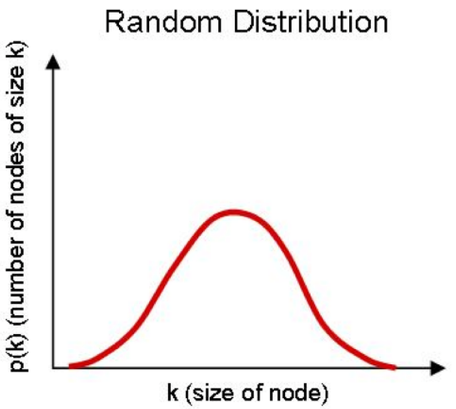
UNSUPERVISED LEARNING APPLICATIONS ON NETWORKS

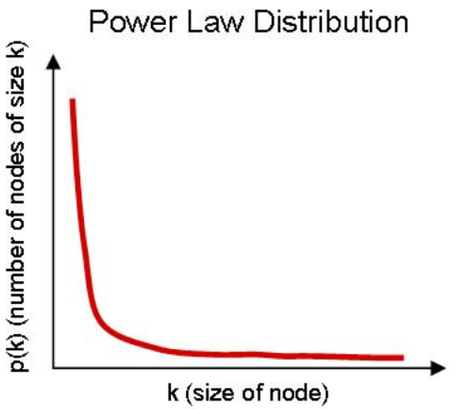
DIFFICULTIES WITH NETWORK ANALYSIS

- Storage
- Inference
- Visualization

DIFFICULTIES WITH NETWORK ANALYSIS

- Storage





DIFFICULTIES WITH NETWORK ANALYSIS

- Inference

DIFFICULTIES WITH NETWORK ANALYSIS

- Visualization