HEART: Statistics and Data Science With Networks

Joshua Agterberg

Johns Hopkins University

Fall 2021

Outline

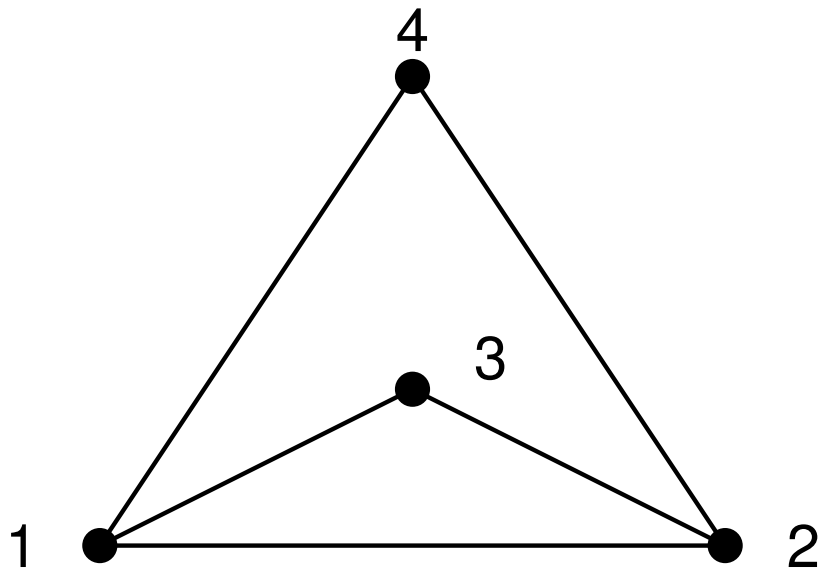
Outline

What Are Networks?

- A graph (or network) is a set of vertices V and edges E

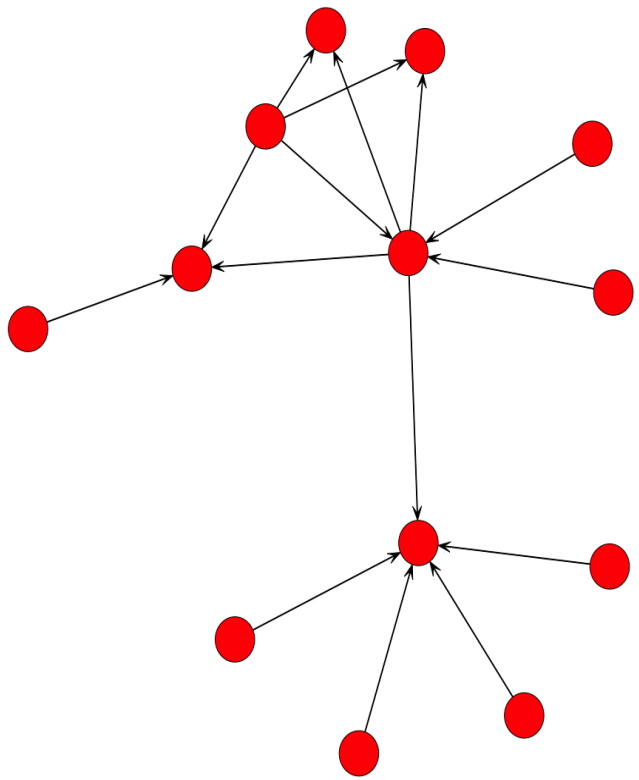
What Are Networks?

- A graph (or network) is a set of vertices V and edges E



What Are Networks?

- A graph (or network) is a set of vertices V and edges E



What Are Networks?

- A graph (or network) is a set of vertices V and edges E
- Social networks – Facebook, Twitter, Instagram
- Biological Networks – Protein-protein interaction, Gene co-expression networks, neuronal network
- Semantic networks – concepts and relationships (Google searches)
- ...

Examples of Networks: Social Network

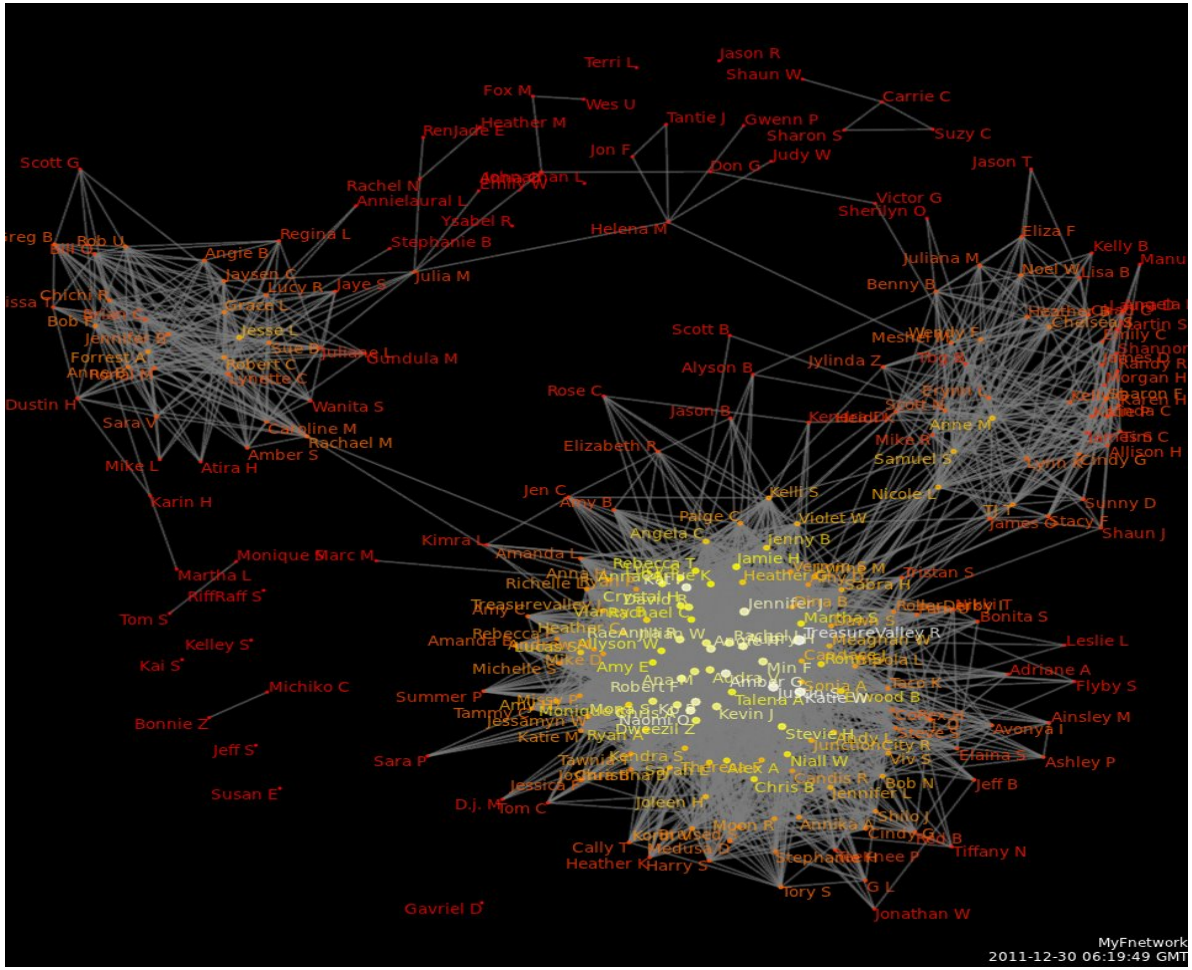


Figure: Source: By Kencf0618 - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=17857442

Examples of Networks: Protein-Protein Interaction

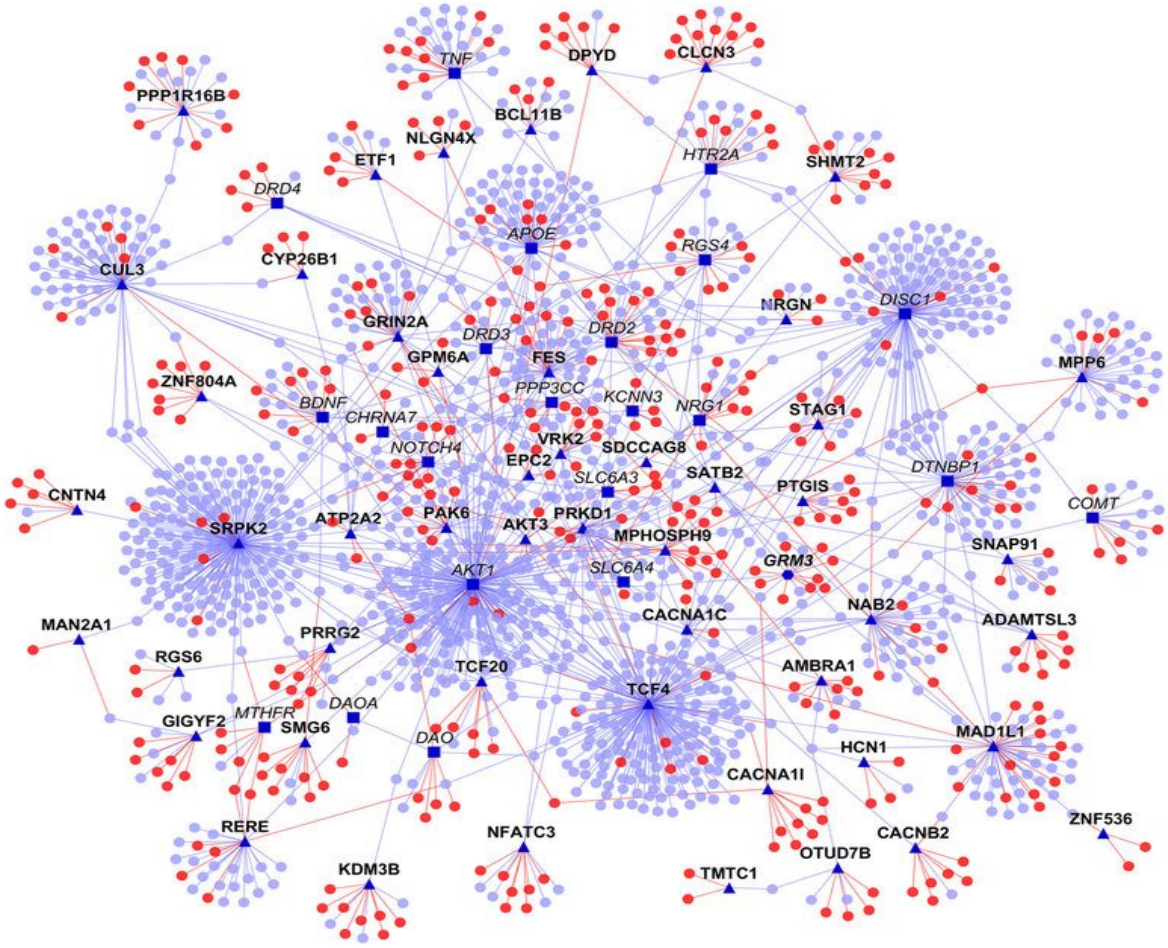


Figure: Source: By Madhavicmu - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=48447204

Examples of Networks: Semantic Network

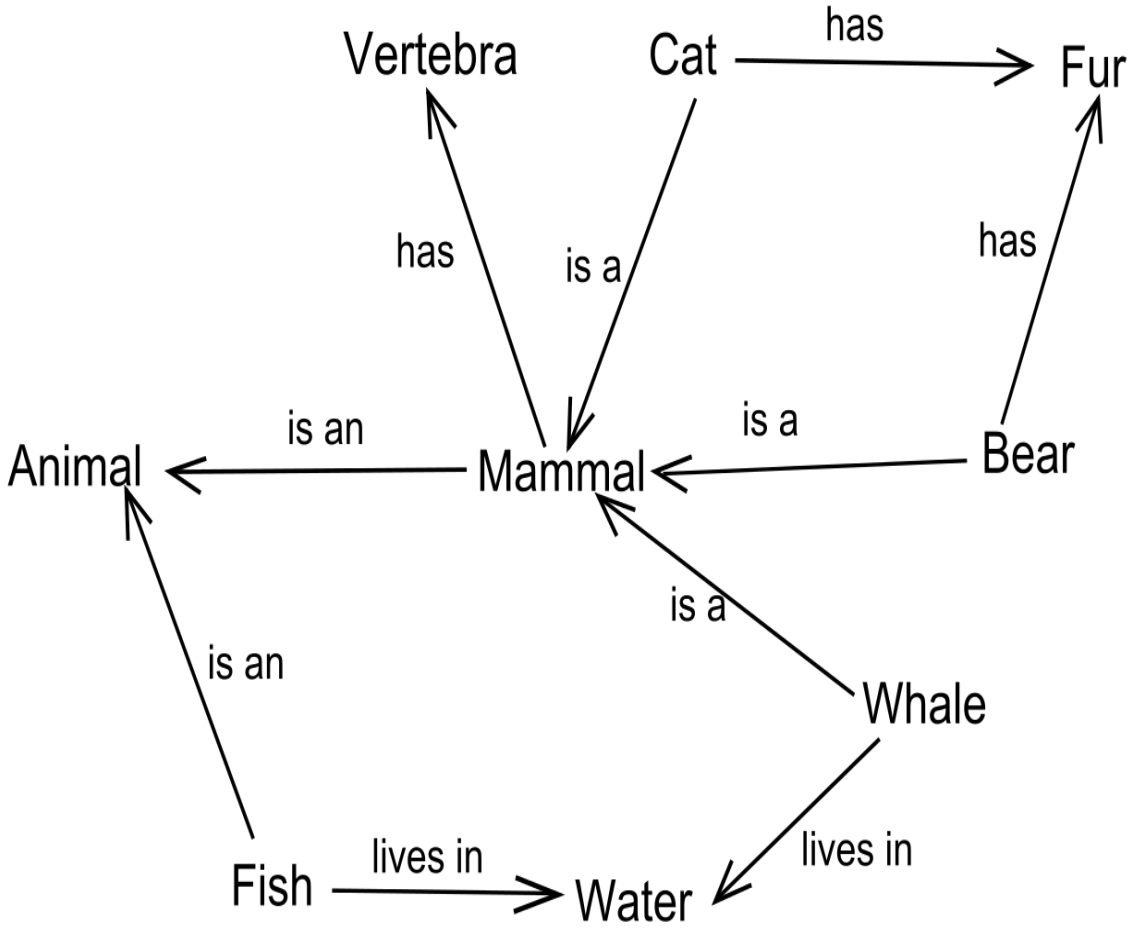


Figure: Source: https://commons.wikimedia.org/w/index.php?curid=1353062

Outline

What is Data Science?

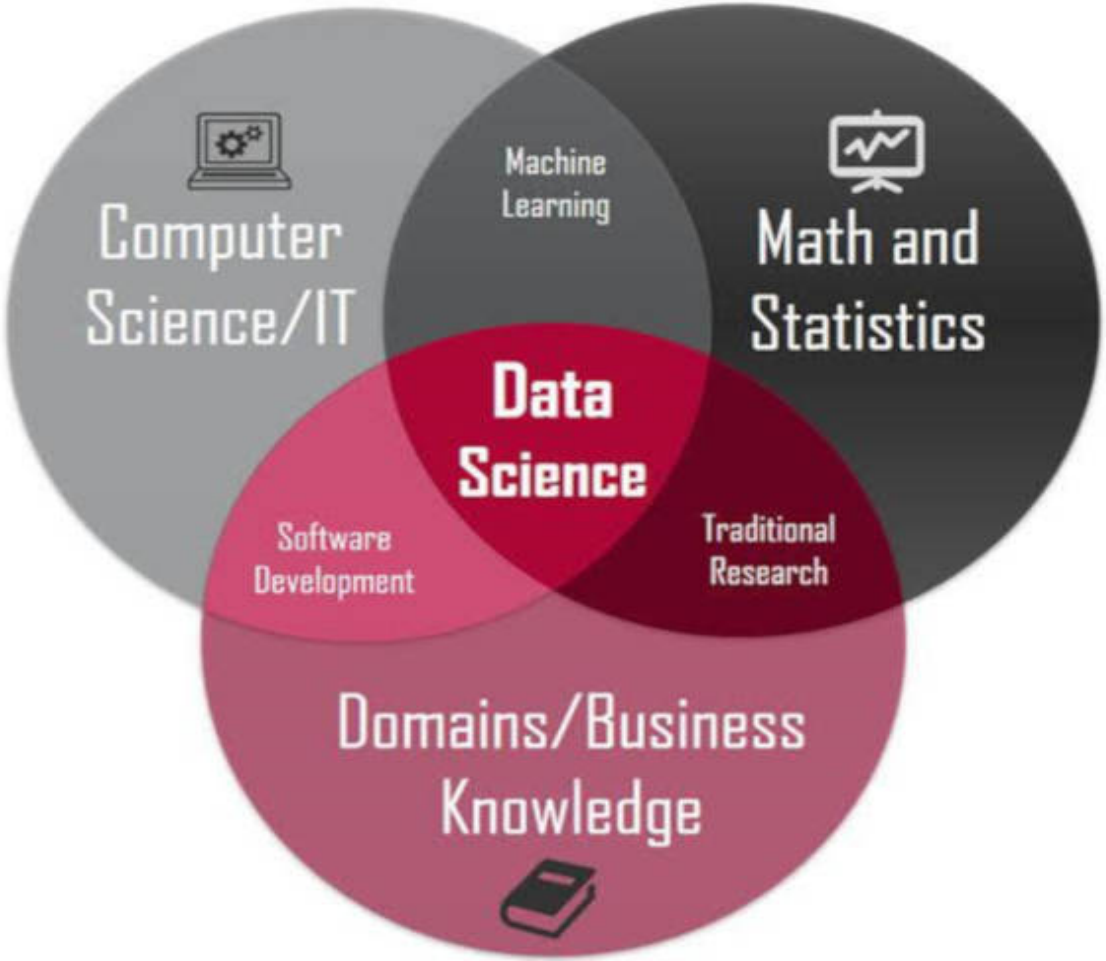


Figure: Source: `https%3A%2F%2Fwww.kdnuggets.com%2F2020%2F08%2Ftop-10-lists-data-science.html&psig=AOvVaw0WtwSe5c_5uk-leat1JfEN&ust=1630343927744000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCPCljMze1vICFQAAAAAdAAAAABAD`

What is Data Science?

- The art and science of extracting information from data

What is Data Science?

- The art and science of extracting information from data
  - Art: needs domain expertise and sometimes “hacks”

What is Data Science?

- The art and science of extracting information from data
  - Art: needs domain expertise and sometimes “hacks”
  - Science: Should be *principled*

What is Data Science?

- The art and science of extracting information from data
  - Art: needs domain expertise and sometimes “hacks”
  - Science: Should be *principled*
- The application of statistics and mathematics to design principled algorithms to extract information from data

What is Data Science?

- The art and science of extracting information from data
  - Art: needs domain expertise and sometimes “hacks”
  - Science: Should be *principled*
- The application of statistics and mathematics to design principled algorithms to extract information from data
- The application of theoretical computer science to design efficient algorithms to extract information from data

Outline

What are Network Data Science Problems?

- Graph Clustering/Community Detection
- Hypothesis Testing
- Multiple Network Analysis

What are Network Data Science Problems?

- Graph Clustering/Community Detection
- Hypothesis Testing
- Multiple Network Analysis

These require first assuming a *statistical network model* and using the assumptions to study properties of the network

Graph Clustering/Community Detection

- Assume each vertex belongs to one of a small number of communities
- Goal is to group vertices according to each community

Graph Clustering

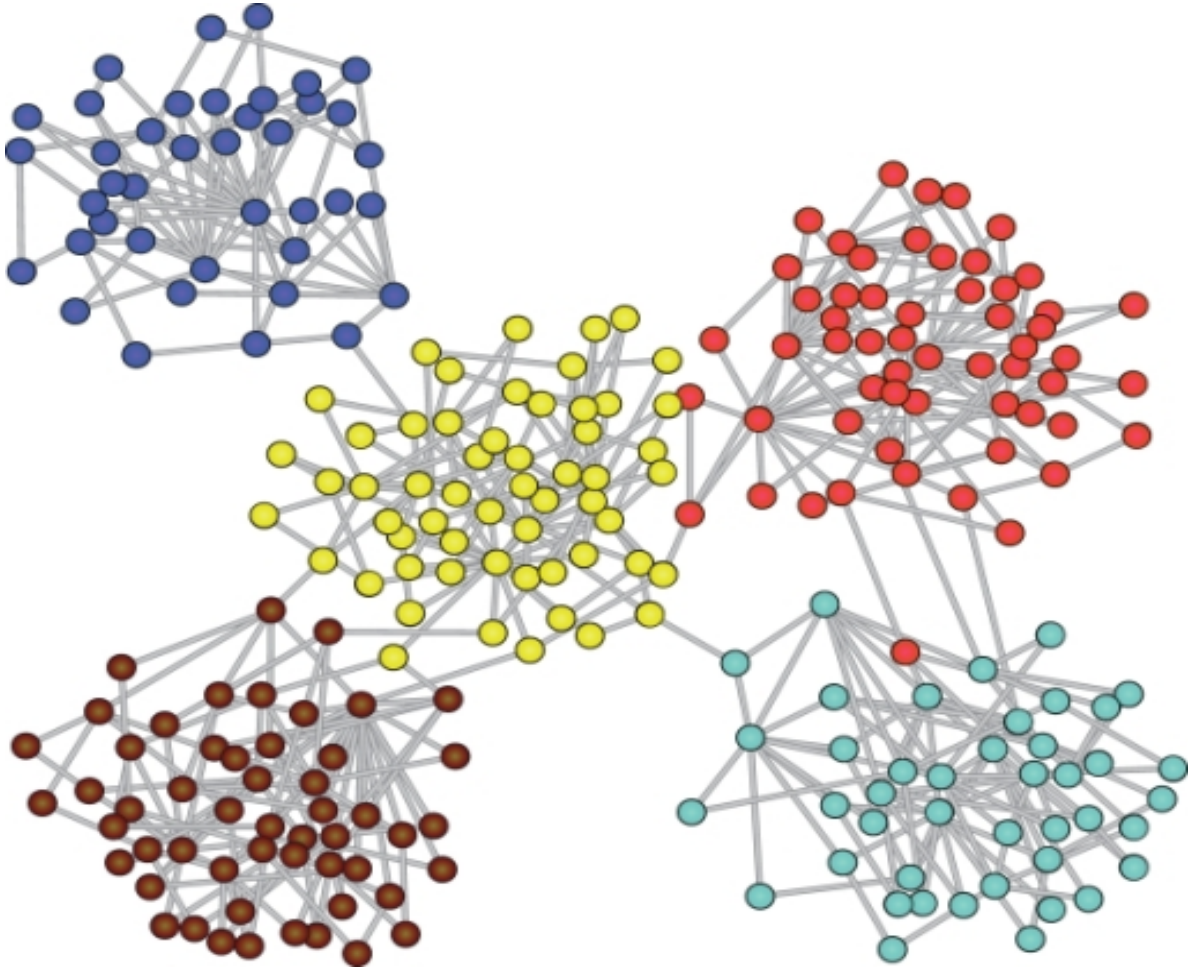


Figure: Source: https://github.com/benedekrozemberczki/awesome-community-detection/blob/master/coms.png

Graph Clustering/Community Detection

Examples:

- Priebe et al.
- Sussman et al
- Ji and Jin
- Jin et al

Hypothesis Testing

- Want to test if a particular statistical network model is a good fit
- Assume each vertex belongs to a community. Want to test whether two specific vertices belong to the same community
- Want to test whether two different networks have the same statistical distribution

Hypothesis Testing

Examples:

- Du and Tang
- Fan et al
- Tang et al
- Chung et al
- Zhang and Amini

Multiple Network Analysis

- Have many networks with some shared structure
- Want to extract information about same structure while respecting individual network properties
- Example: community memberships are the same, but each network forms edges differently

Multiple Network Analysis

Examples:

- Jones and Rubin-Delanchy
- Levin et al.
- Lei and Lin
- Jing et al.
- Arroyo et al

This Class

- Statistical Network Analysis:

This Class

- Statistical Network Analysis:
  - Learn some standard statistical models for network analysis

This Class

- Statistical Network Analysis:
  - Learn some standard statistical models for network analysis
  - Learn the basics of statistical network analysis

This Class

- Statistical Network Analysis:
  - Learn some standard statistical models for network analysis
  - Learn the basics of statistical network analysis
  - Learn how to apply principled data science techniques to networks

This Class

- Statistical Network Analysis:
  - Learn some standard statistical models for network analysis
  - Learn the basics of statistical network analysis
  - Learn how to apply principled data science techniques to networks
- Data Science:

This Class

- Statistical Network Analysis:
  - Learn some standard statistical models for network analysis
  - Learn the basics of statistical network analysis
  - Learn how to apply principled data science techniques to networks
- Data Science:
  - Learn standard algorithms for common data science problems

This Class

- Statistical Network Analysis:
  - Learn some standard statistical models for network analysis
  - Learn the basics of statistical network analysis
  - Learn how to apply principled data science techniques to networks
- Data Science:
  - Learn standard algorithms for common data science problems
  - Learn basic ideas behind dimensionality reduction

This Class

- Statistical Network Analysis:
  - Learn some standard statistical models for network analysis
  - Learn the basics of statistical network analysis
  - Learn how to apply principled data science techniques to networks
- Data Science:
  - Learn standard algorithms for common data science problems
  - Learn basic ideas behind dimensionality reduction
  - Learn how to apply principled data science techniques to networks

Common and Freely Available Network Datasets

- Political Blogs
- Karate Club Data
- Network Repository

Graph Clustering Example: What you can learn to do