

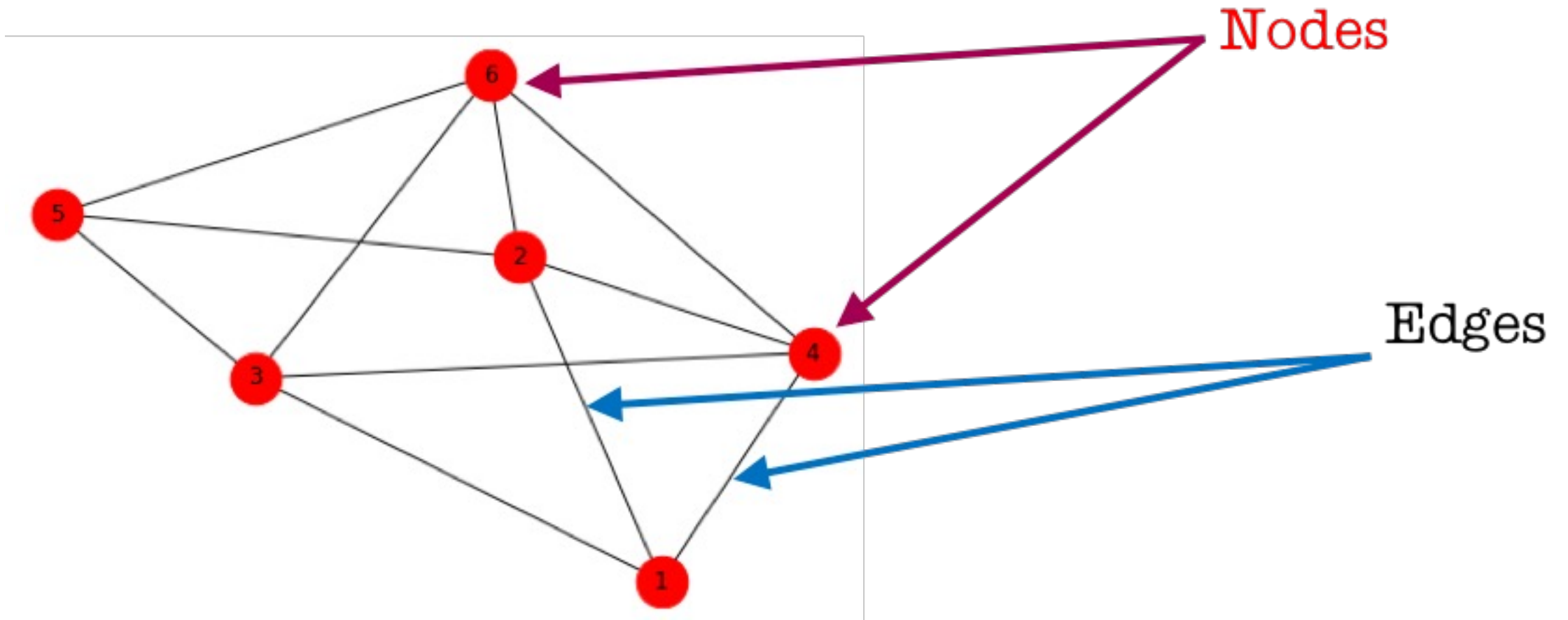
# Social Network Analysis

ICS 491

# Reminder about next few weeks

Tue Nov 21	Social Network Analysis (Class on Zoom)	
Thu Nov 23	Thanksgiving Holiday (No Class)	
Tue Nov 28	Watch final project videos (No Class)	<a href="#">Project Milestone #6: Final Presentation</a>
Thu Nov 30	Watch final project videos (No Class)	
Tue Dec 5	Multimedia Analytics	
Thu Dec 7	Course Overview	
Fri Dec 15		<a href="#">Final Project Infographic and Code</a>

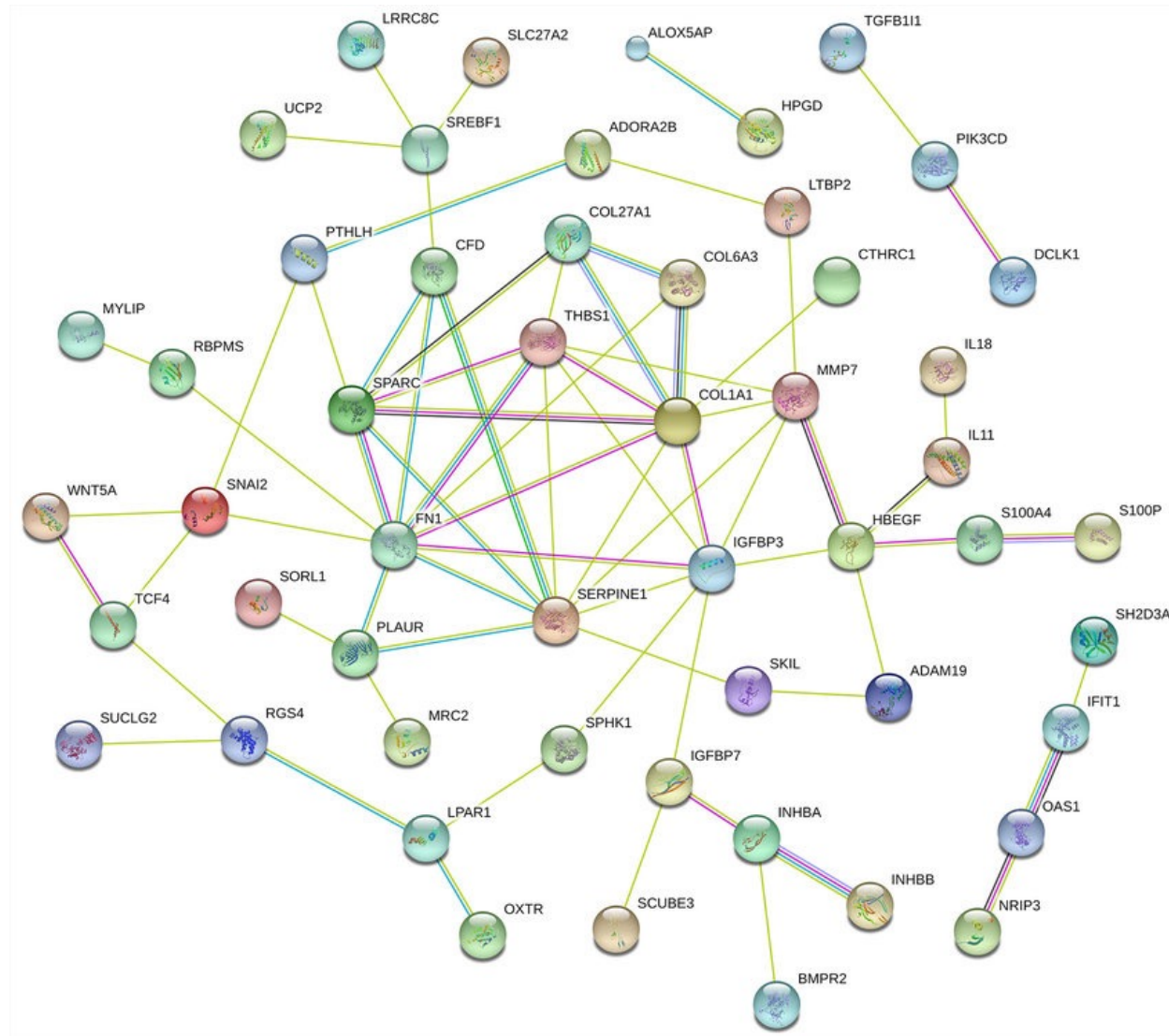
# Network Analysis



# (Human) Network Analysis



# (Protein) Network Analysis

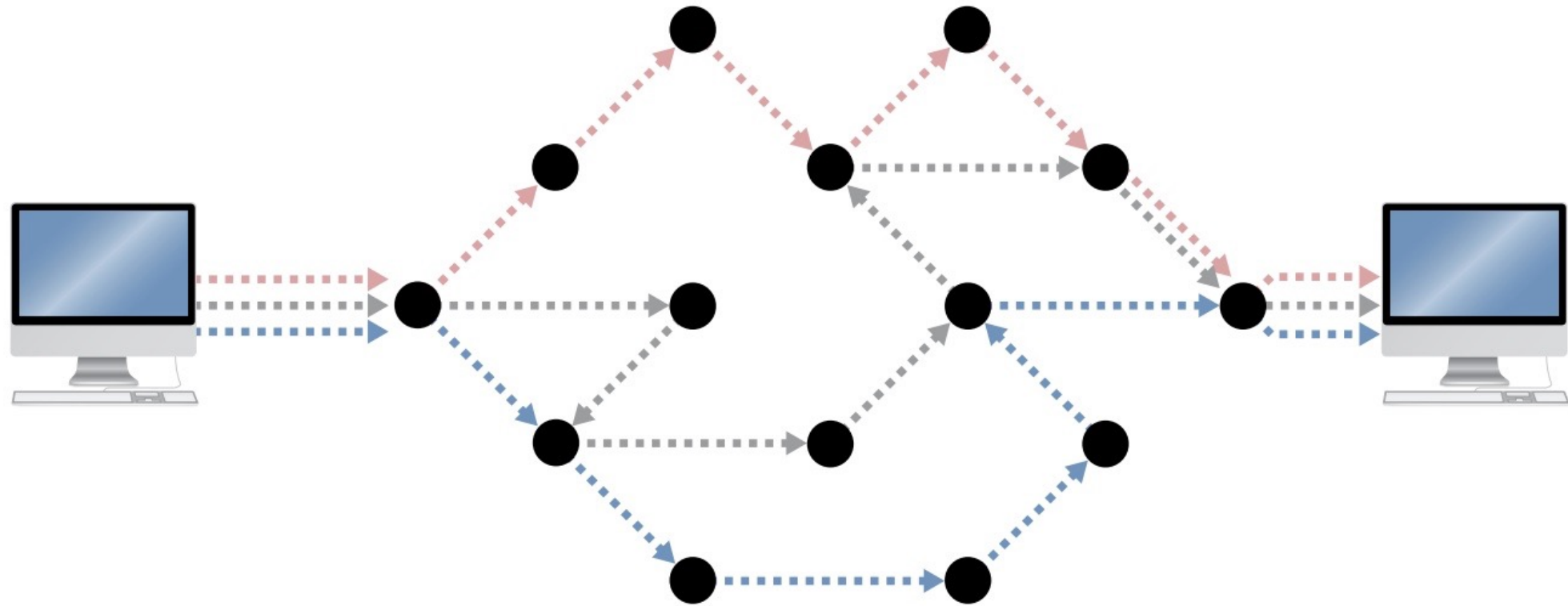




# (Neuroscience) Network Analysis

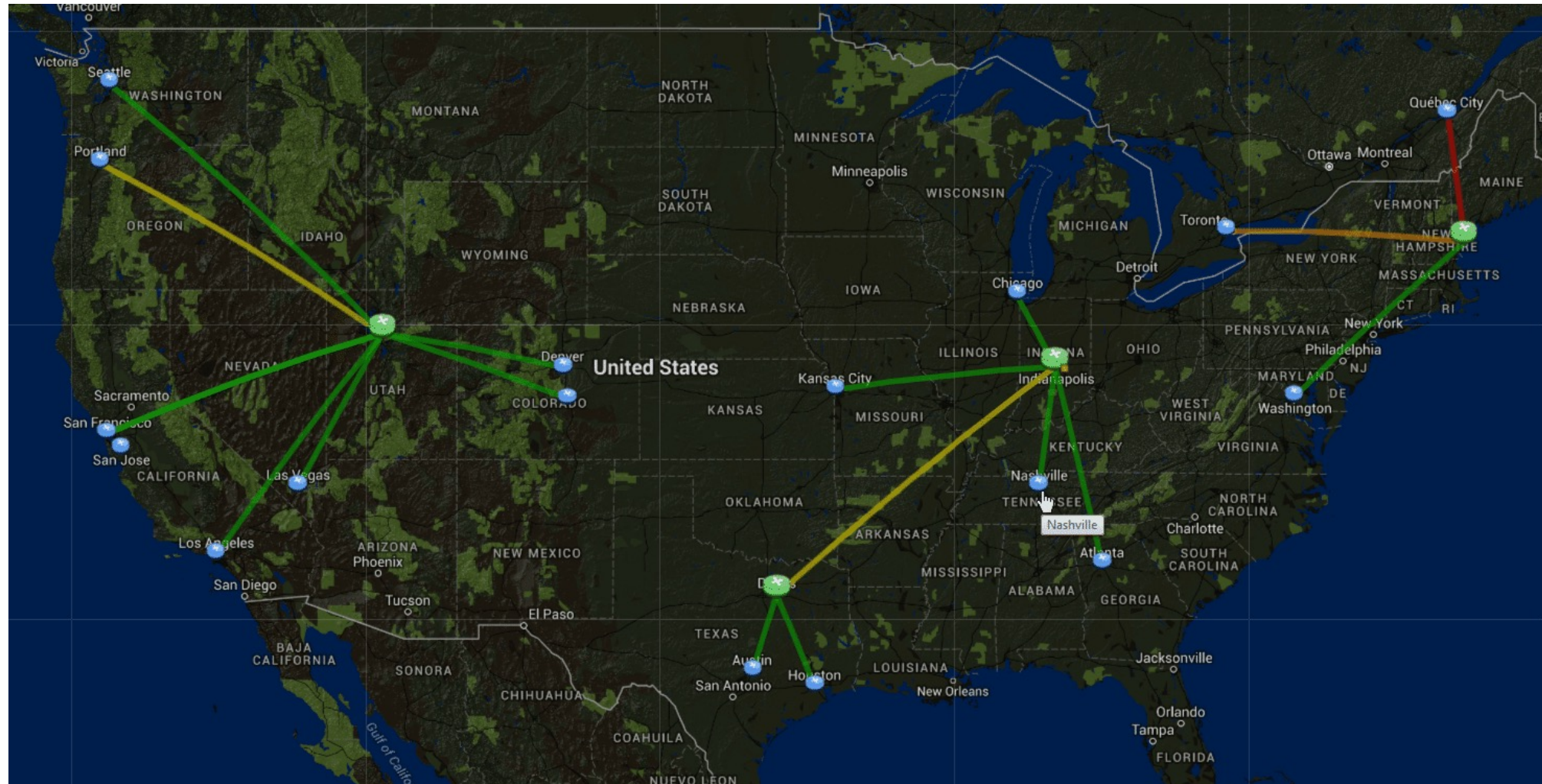


# (Internet) Network Analysis



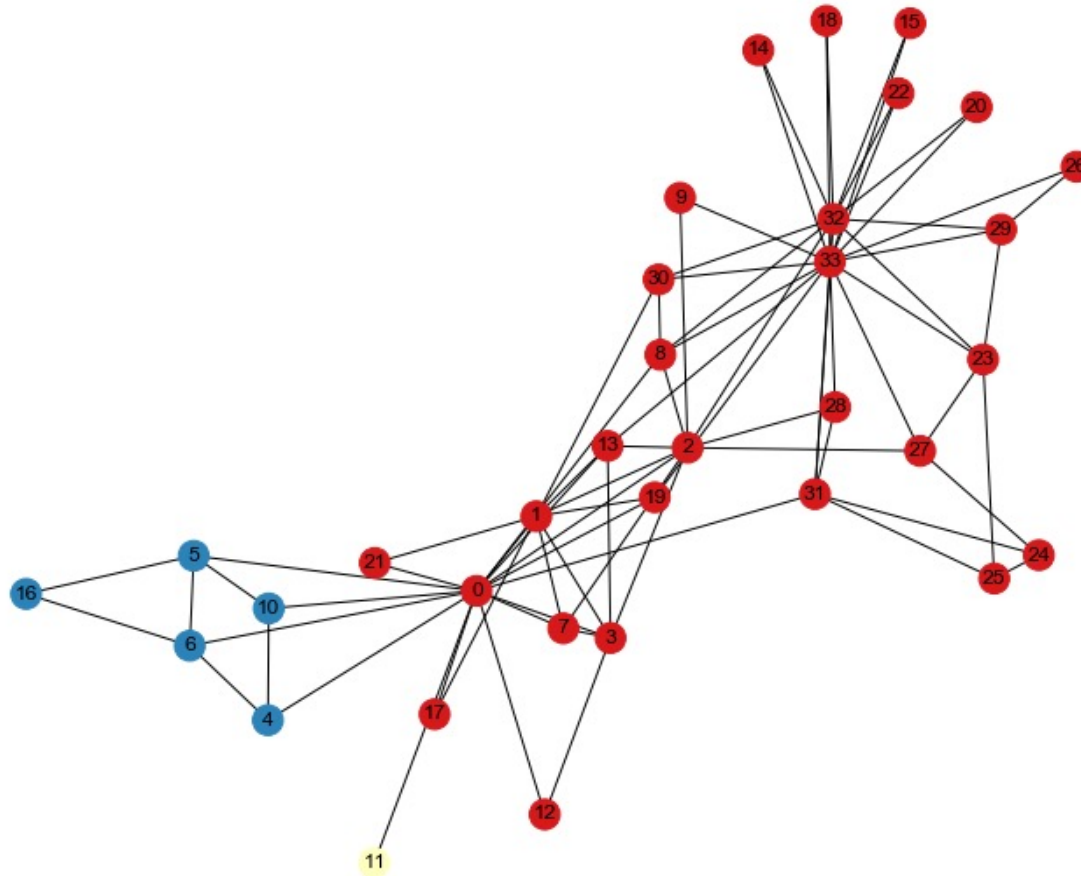


# (Transportation) Network Analysis



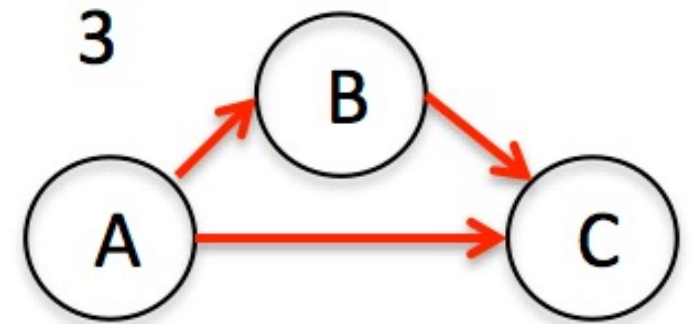
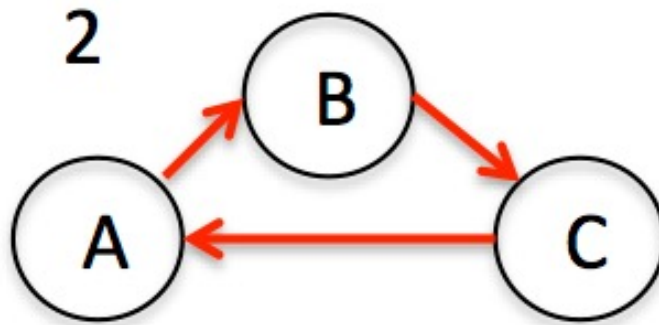
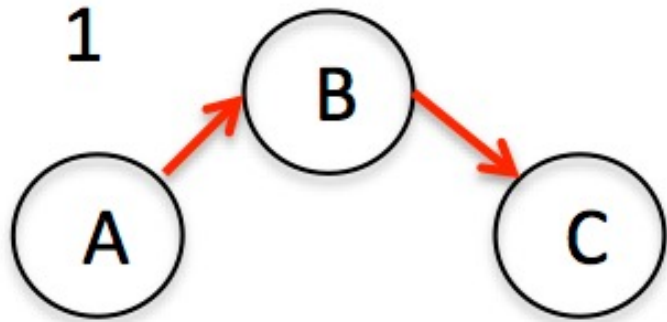


Many, many other applications...  
Can you name some?

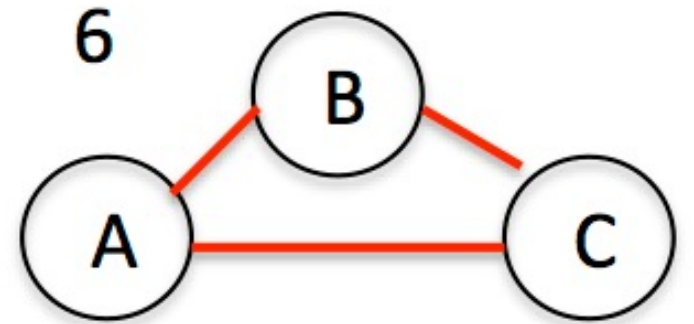
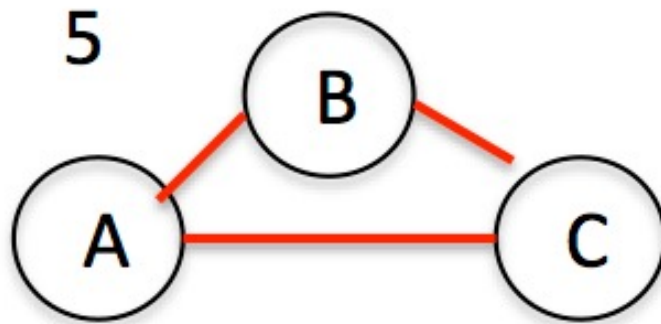
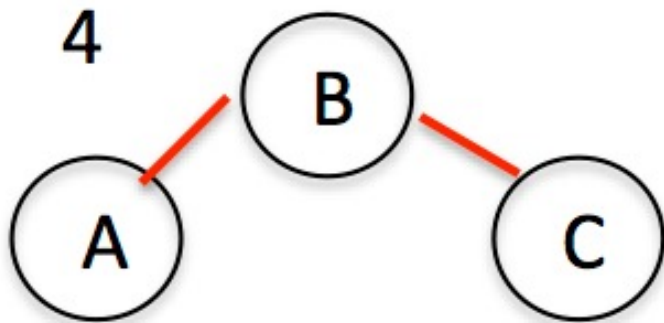


# Directed vs Undirected Graphs

## Directed Graphs

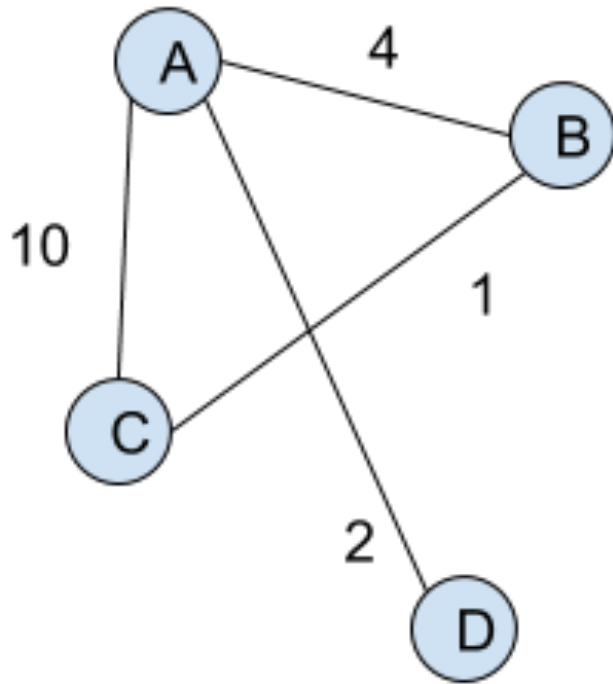


## Undirected Graphs

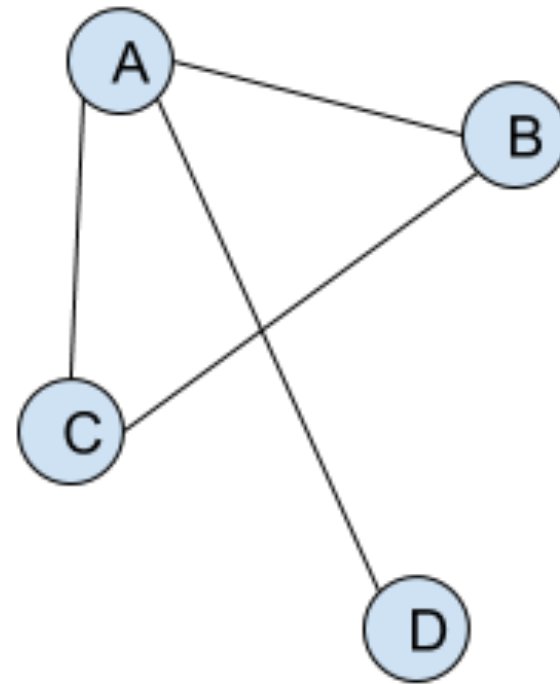


# Weighted vs Unweighted Graphs

Weighted Graph

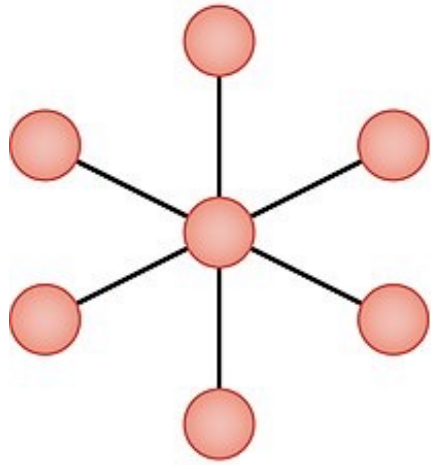


Unweighted Graph

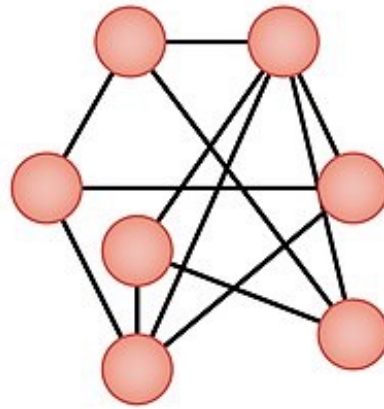


# Network Characteristics

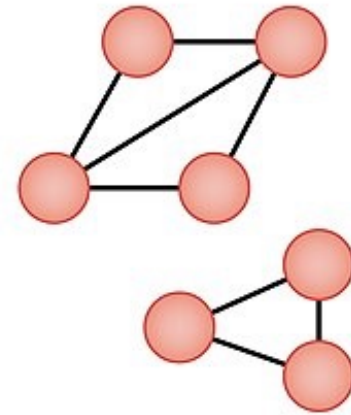
**a** Centralized



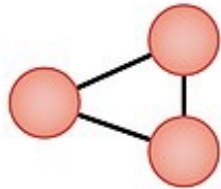
**b** Dense, not centralized



**c** Fragmented



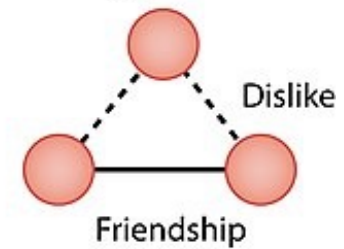
**d** Closure



**e** Ties between actors with different attributes

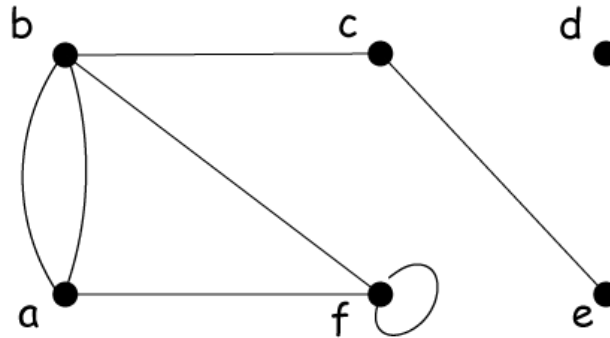


**f** Two types of ties



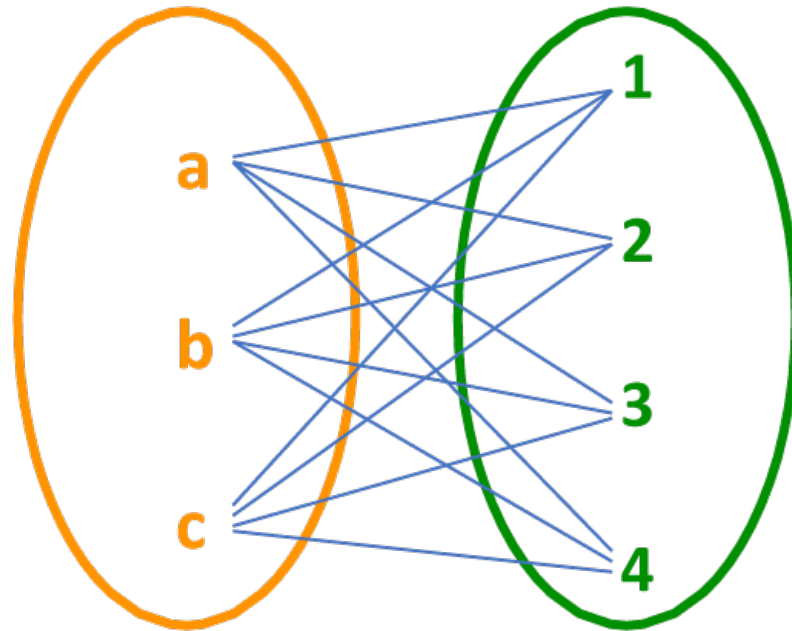


# Degree and Neighborhood

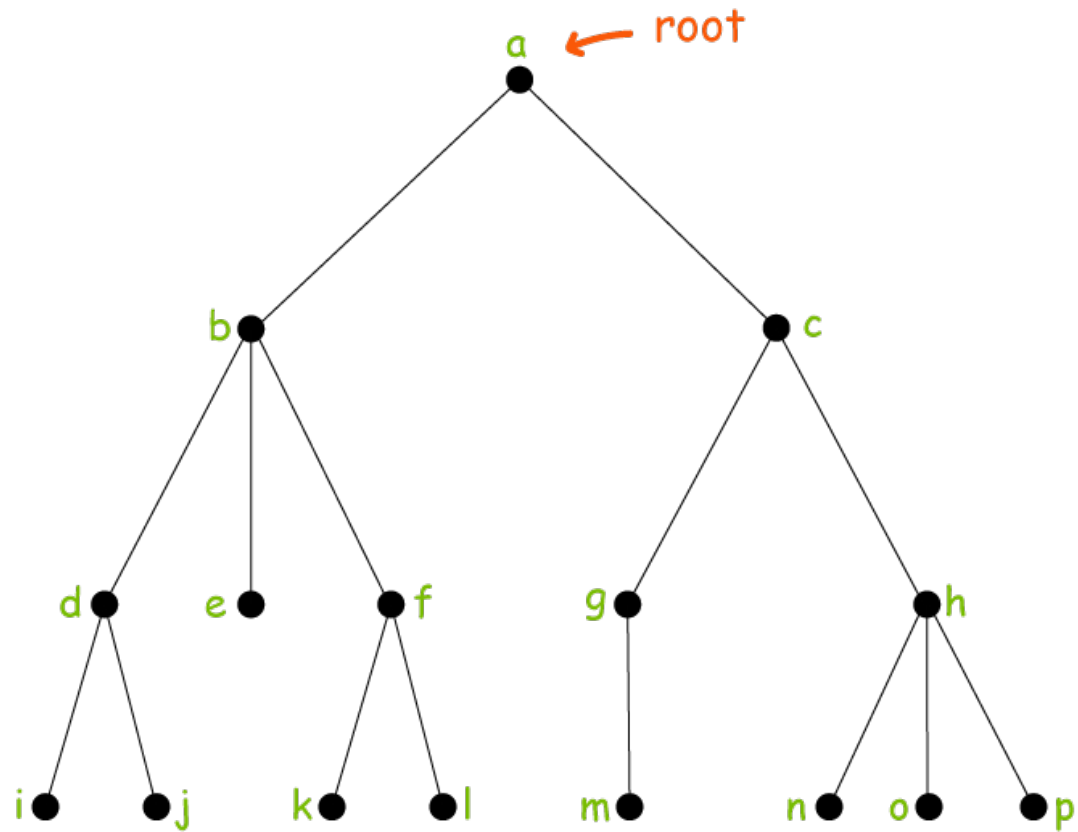


Degree	Neighborhood
$\deg(a) = 3$	$N(a) = \{b, f\}$
$\deg(b) = 4$	$N(b) = \{a, c, f\}$
$\deg(c) = 2$	$N(c) = \{b, e\}$
$\deg(d) = 0$ ( <i>isolated</i> )	$N(d) = \emptyset$
$\deg(e) = 1$ ( <i>pendant</i> )	$N(e) = \{c\}$
$\deg(f) = 4$	$N(f) = \{a, b, f\}$

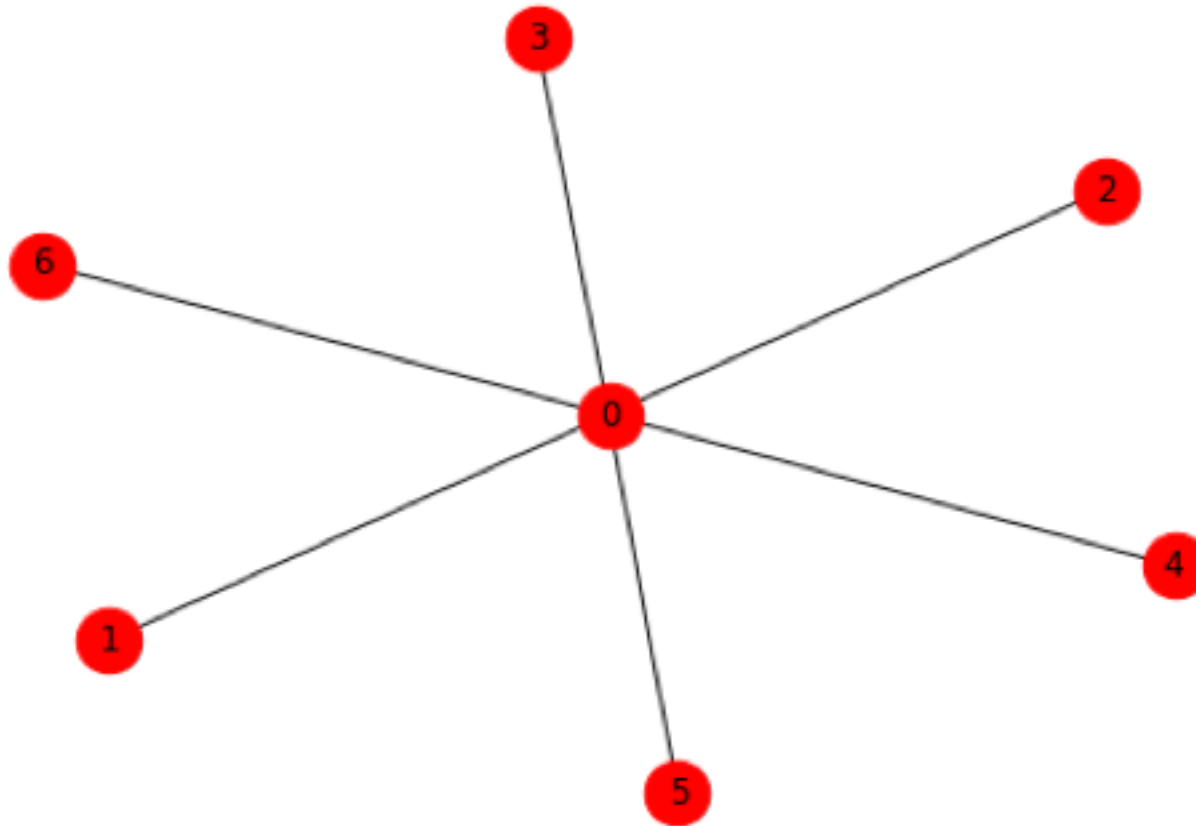
# Bipartite Graphs



# Trees are a Special Case of Graphs



# Important Task: Identifying Important Nodes (“Central” Nodes)





# Centrality Measures

## Comparing Common Centralities

Name	Formula	Description	Translates To
Degree	$C_D(p_i) = \sum_{k=1}^N a(p_i, p_k)$	# of nodes connected	most connected node
Closeness	$C_c(p_i) = \frac{N-1}{\sum_{k=1}^N d(p_i, p_k)}$	shortest avg. geodesic distance	fastest communicating node
Betweenness	$C_B(n_i) = \sum_{j < k} g_{jk}(n_i) / g_{jk}$	times found in geodesic distances	most influential node

Network Science is a huge field, and there are many more complex centrality measures, each with their own pros and cons

# Centrality measures in words

- **Degree centrality:** Number of links that a node has.

$$C_D(v) = \text{deg}(v)$$

- **Closeness centrality:** Average length of the shortest path between the node and all other nodes in the graph. Thus, the more central a node is, the closer it is to all other nodes.

$$C(v) = \frac{N - 1}{\sum_u d(u, v)}$$

- **Betweenness centrality:** Number of times a node acts as a bridge along the shortest path between two other nodes.

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$\sigma_{ST}$  = # shortest paths  
between s and t

1. For each pair of vertices (s,t), compute the shortest paths between them.
2. For each pair of vertices (s,t), determine the fraction of shortest paths that pass through the vertex in question (here, vertex v).
3. Sum this fraction over all pairs of vertices (s,t).

# Pagerank Centrality

(Simplified; used in paper and GOOGLE SEARCH)



Example: 
$$PR(A) = \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)}$$

Formula: 
$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$

- Iterative algorithm
- Initial PageRank of all pages is  $1 / N$
- $PR(x)$  is the PageRank of page  $x$
- $L(x)$  is the number of outbound links of page  $x$

# Pagerank Centrality (Simplified)

Damping factor  $d$  ( $0 < d < 1$ ) is used to account for people who will randomly click on links.  
Common choice is 0.85.

$$PR(A) = \frac{1-d}{N} + d \left( \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \dots \right).$$

Therefore, final PageRank equation is:

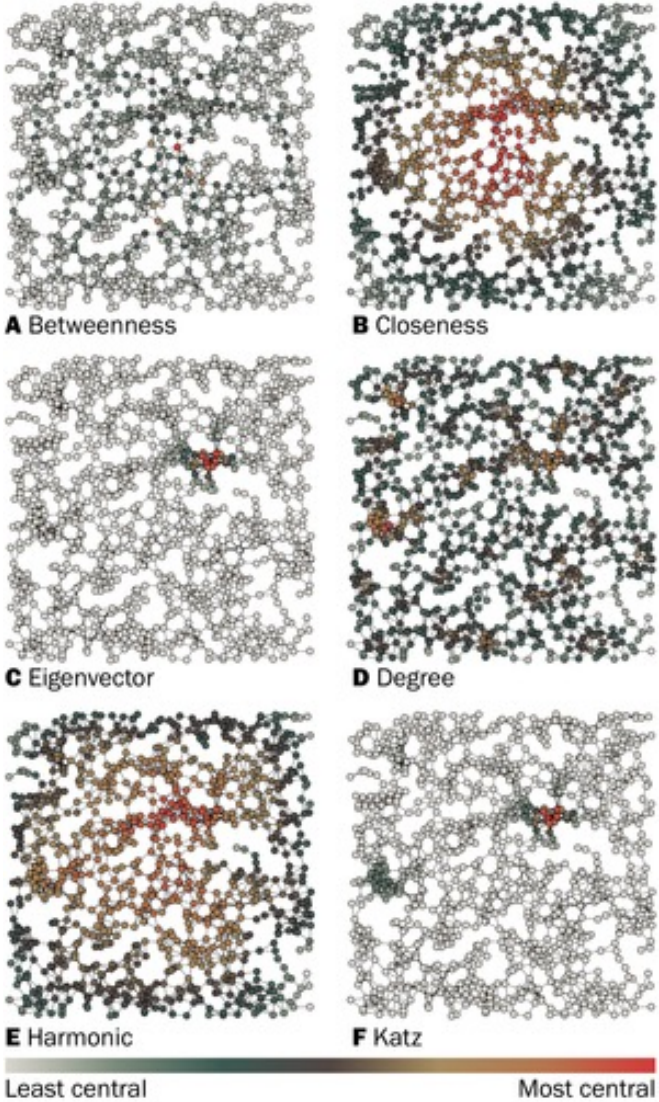
$$PR(p_i; t+1) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j; t)}{L(p_j)}$$



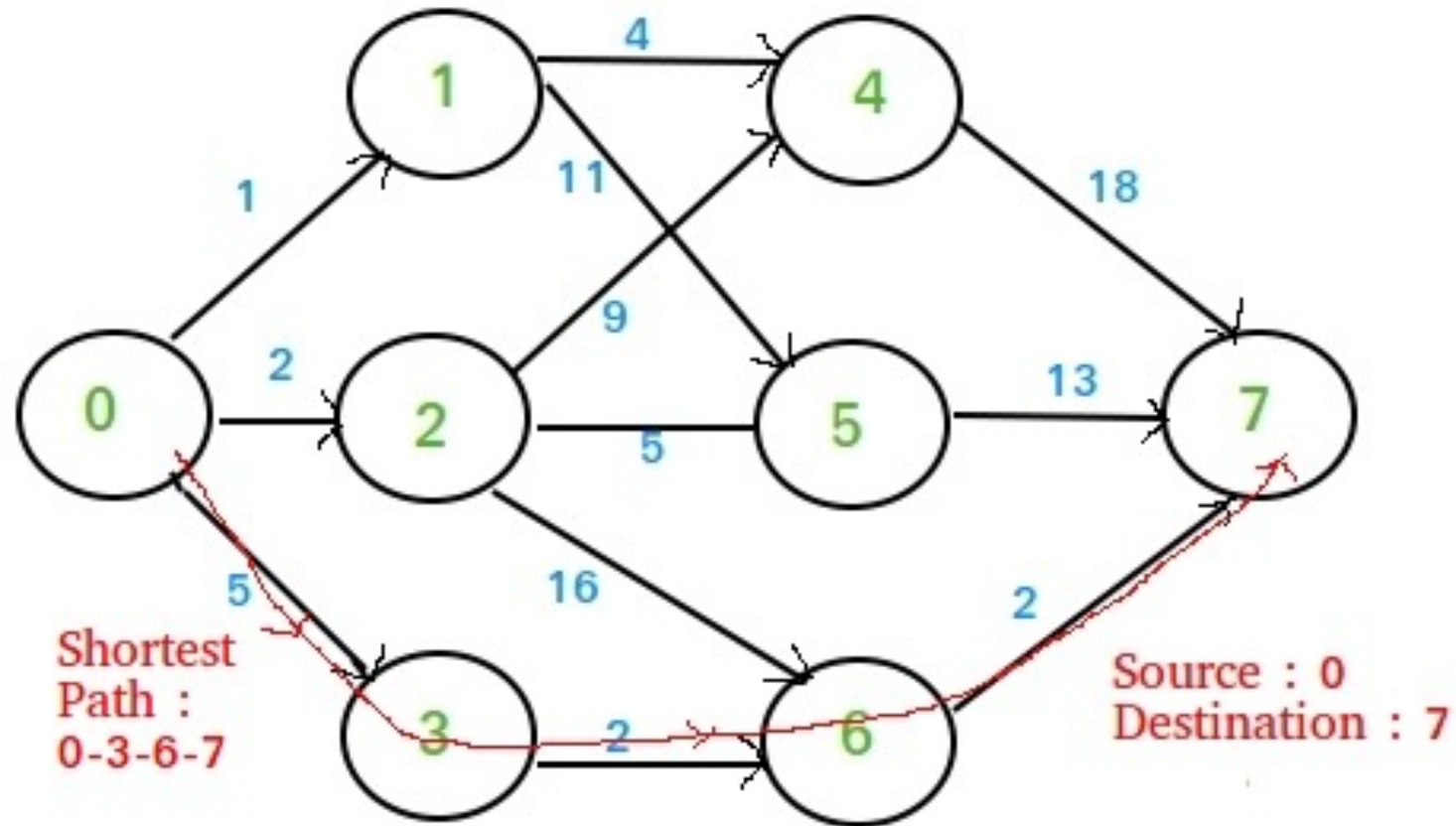
# PageRank in Python: the trillion dollar algorithm

```
import numpy as np
def pagerank(M, num_iterations: int = 100, d: float = 0.85):
    N = M.shape[1]
    v = np.ones(N) / N
    M_hat = (d * M + (1 - d) / N)
    for i in range(num_iterations):
        v = M_hat @ v ### @ operator is matrix multiplication in
Python
    return v
```

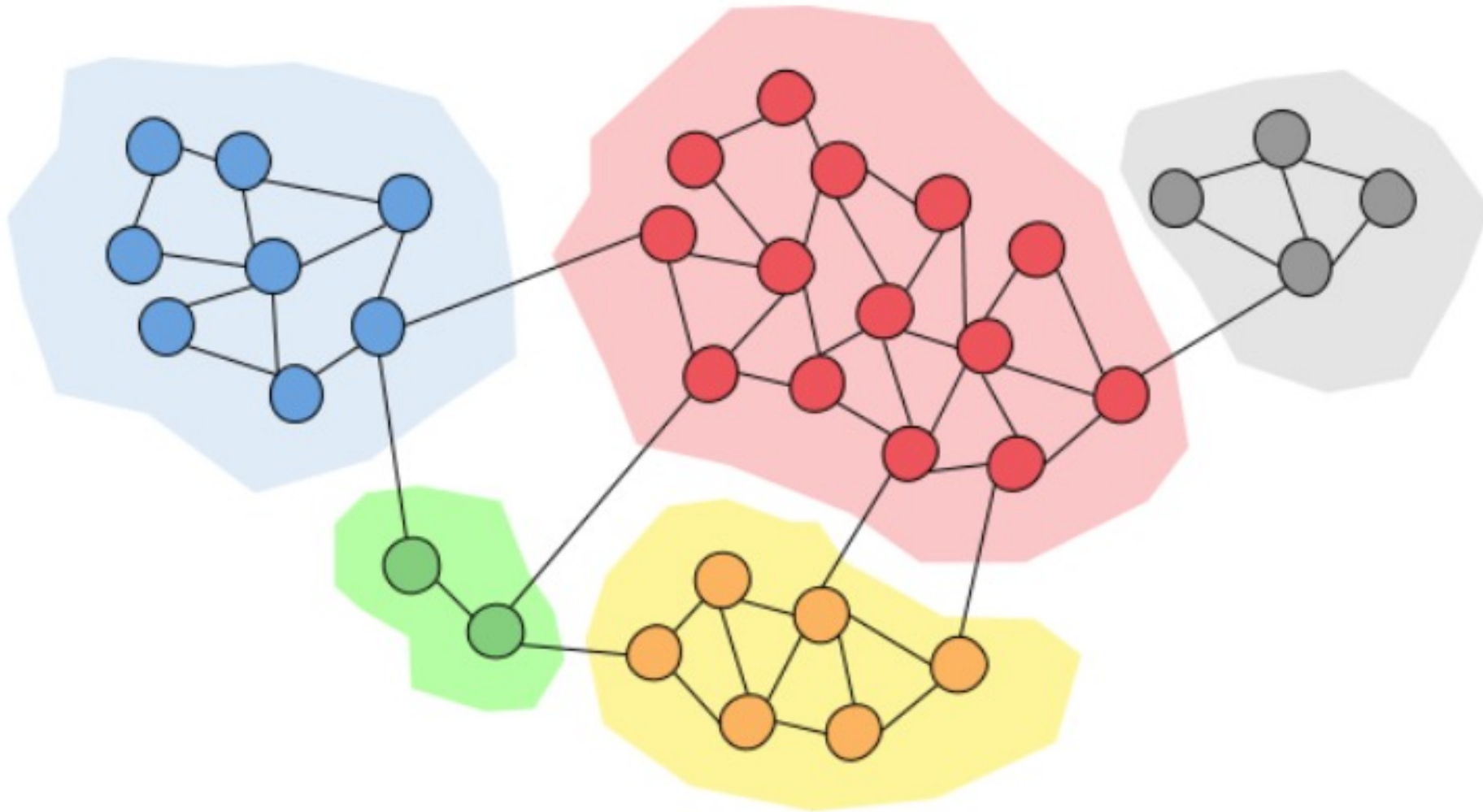
# Qualitative Comparison of Centrality Measures






# Important Task: Finding Shortest Path



# Important Task: Community Detection



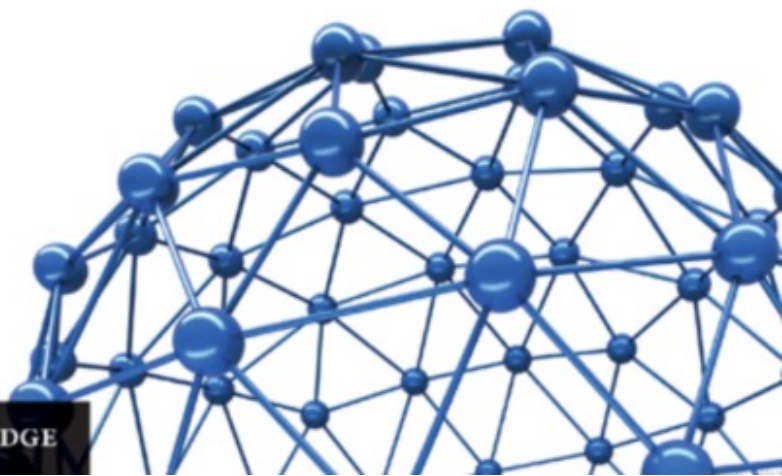
# Summary of Graph Algorithms

Algorithm Type	Graph Problem	Examples
 <b>Pathfinding &amp; Search</b>	Find the optimal path or evaluate route availability and quality	<ul style="list-style-type: none"><li>• Find the quickest route to travel from A to B</li><li>• Telephone call routing</li></ul>
 <b>Centrality</b>	Determine the importance of distinct nodes in the networks	<ul style="list-style-type: none"><li>• Determine social media influencers</li><li>• Find likely attack targets in communication and transportation networks</li></ul>
 <b>Community Detection</b>	Evaluate how a group is clustered or partitioned	<ul style="list-style-type: none"><li>• Segment customers</li><li>• Find potential members of a fraud ring</li></ul>

NETWORKS  
CROWDS  
AND MARKETS

Reasoning about a Highly Connected World

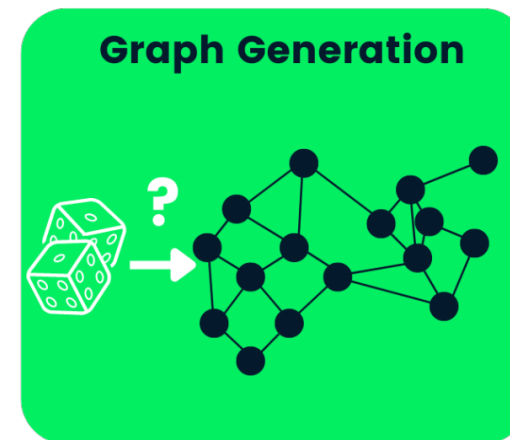
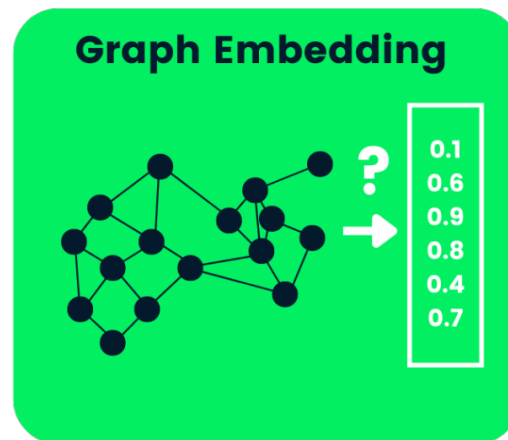
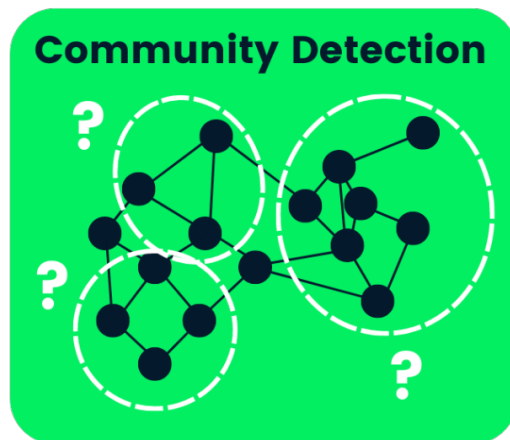
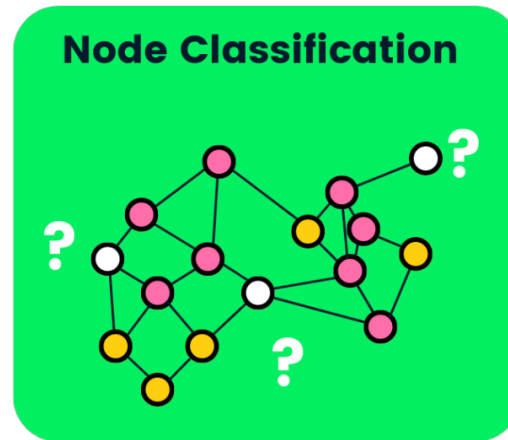
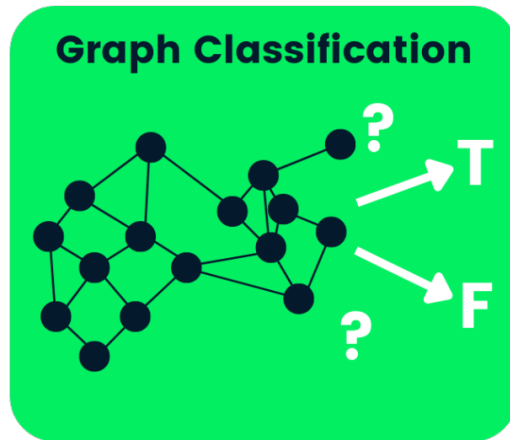
DAVID EASLEY  
and  
JON KLEINBERG



CAMBRIDGE



# Graph Neural Networks (GNNs): Machine Learning with Graphs



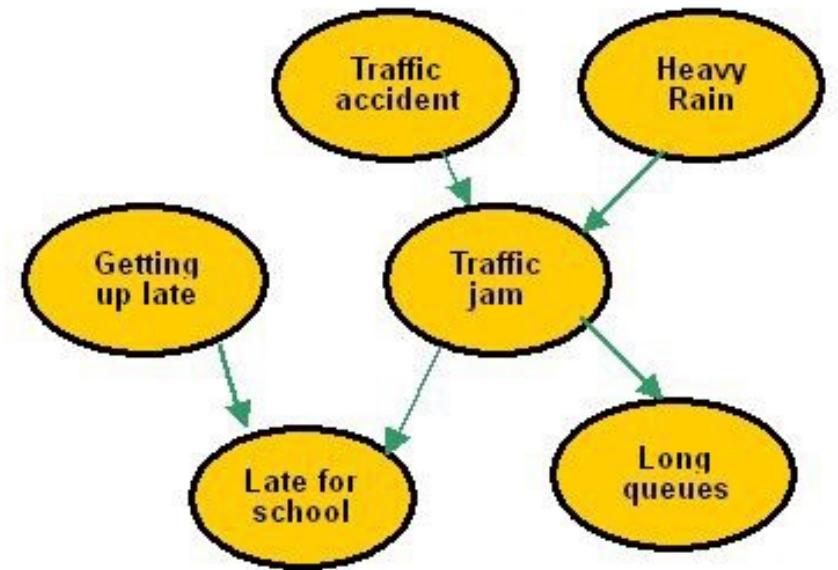


# Causal Inference

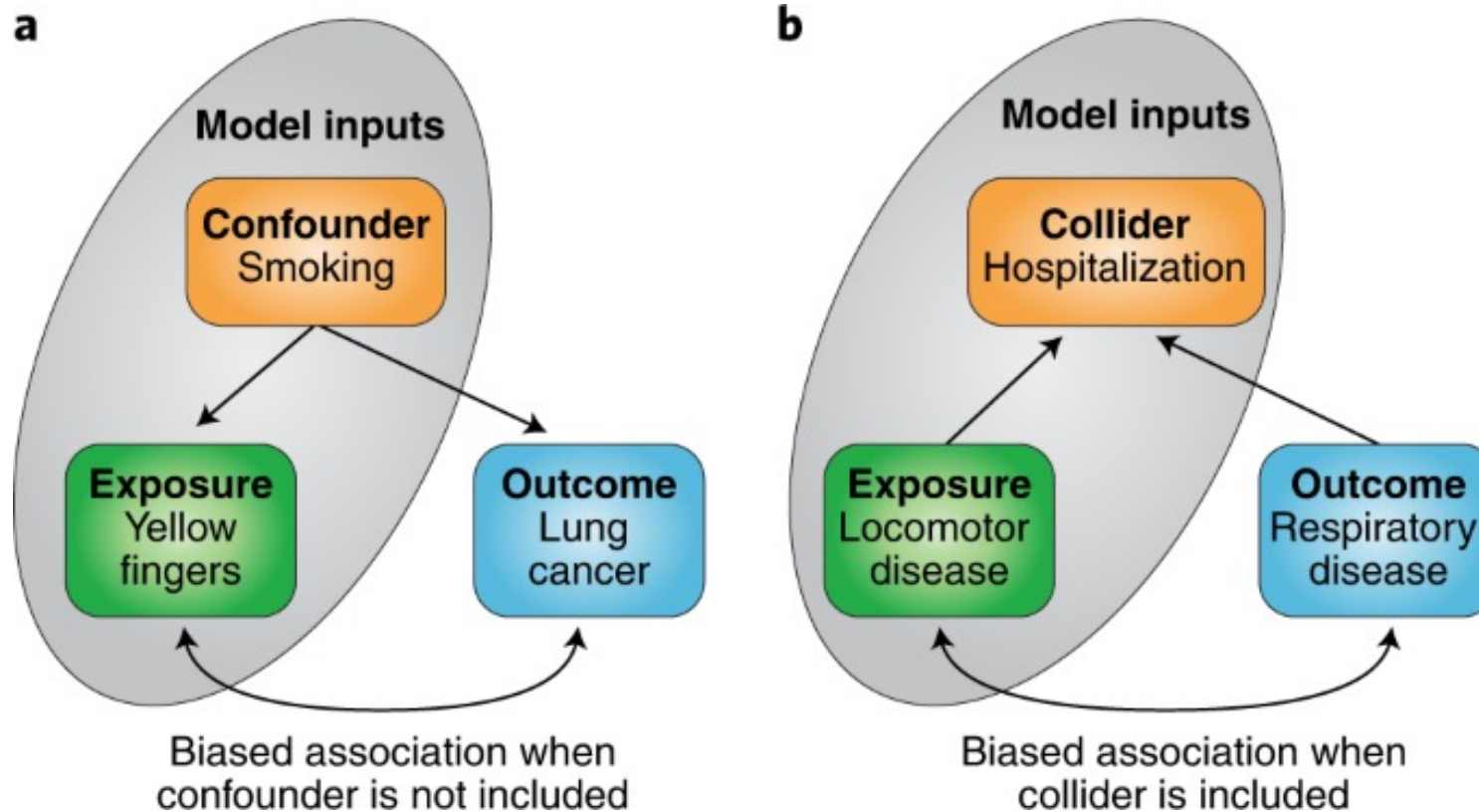
Correlation  $\neq$  Causation

See:

- PH 664 Principles of Epidemiology II
- SOC 176: Introduction to Data Analysis
- Prof. Susanne Still



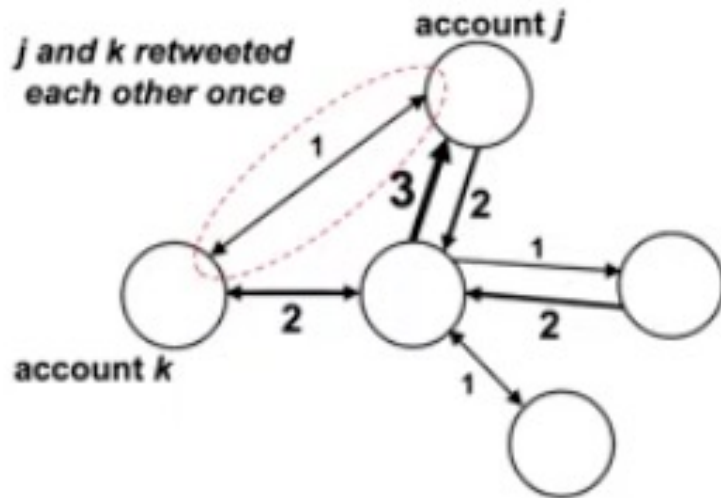
# Causal Inference



**Confounding** occurs when there exists a variable that causes the outcome and the effect  
**Collider:** variable that is caused by both the exposure (or causes of the exposure) and the outcome (or causes of the outcome)

# Network Causal Inference

Example retweet network



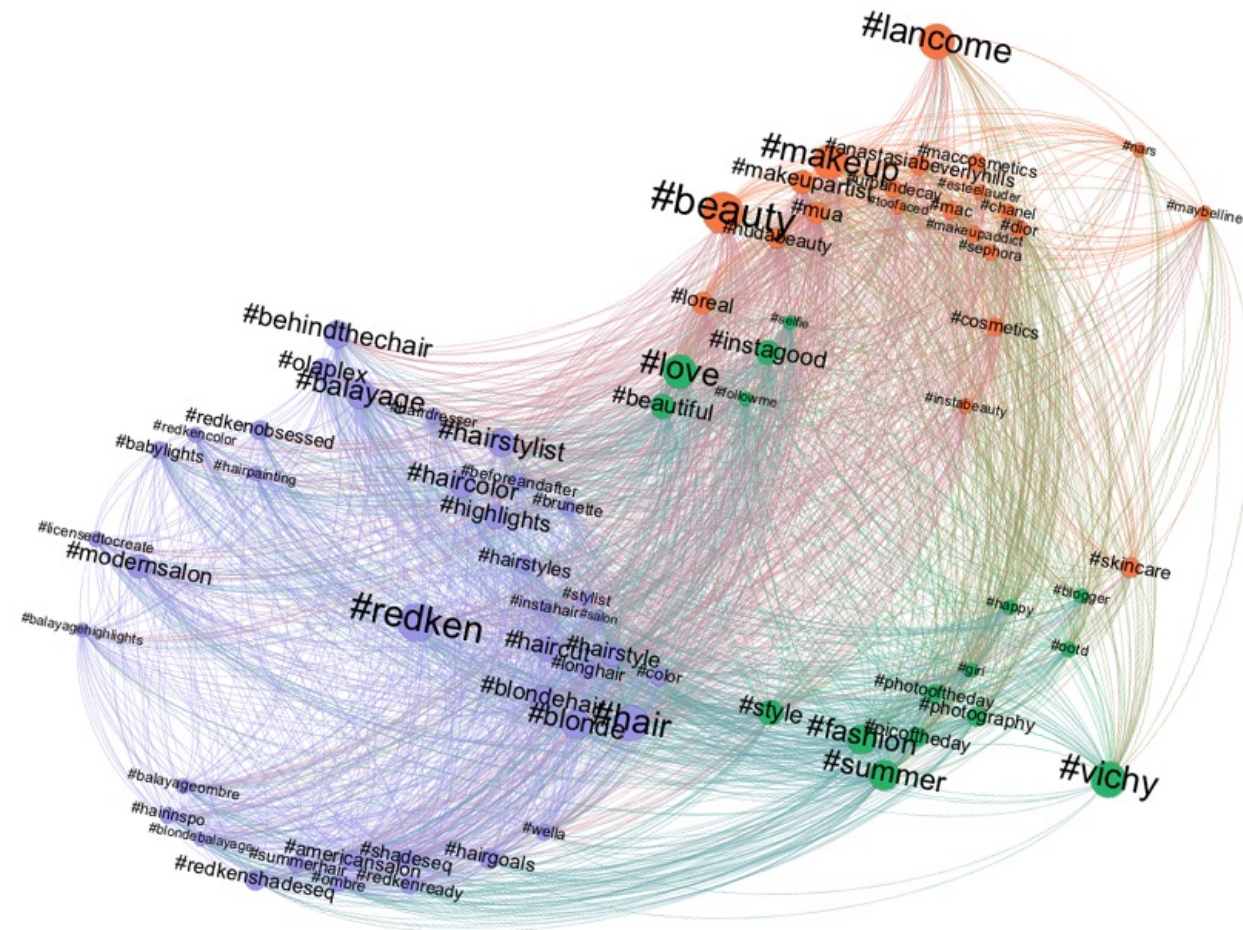
## Challenge:

How to map out the IO narrative network of participating accounts and their interactions?

## Approach:

- Represent network as directed graph where nodes are accounts and edges are interactions
- Construct edges using past interactions (e.g. retweets) and relationships (e.g. followers) between accounts

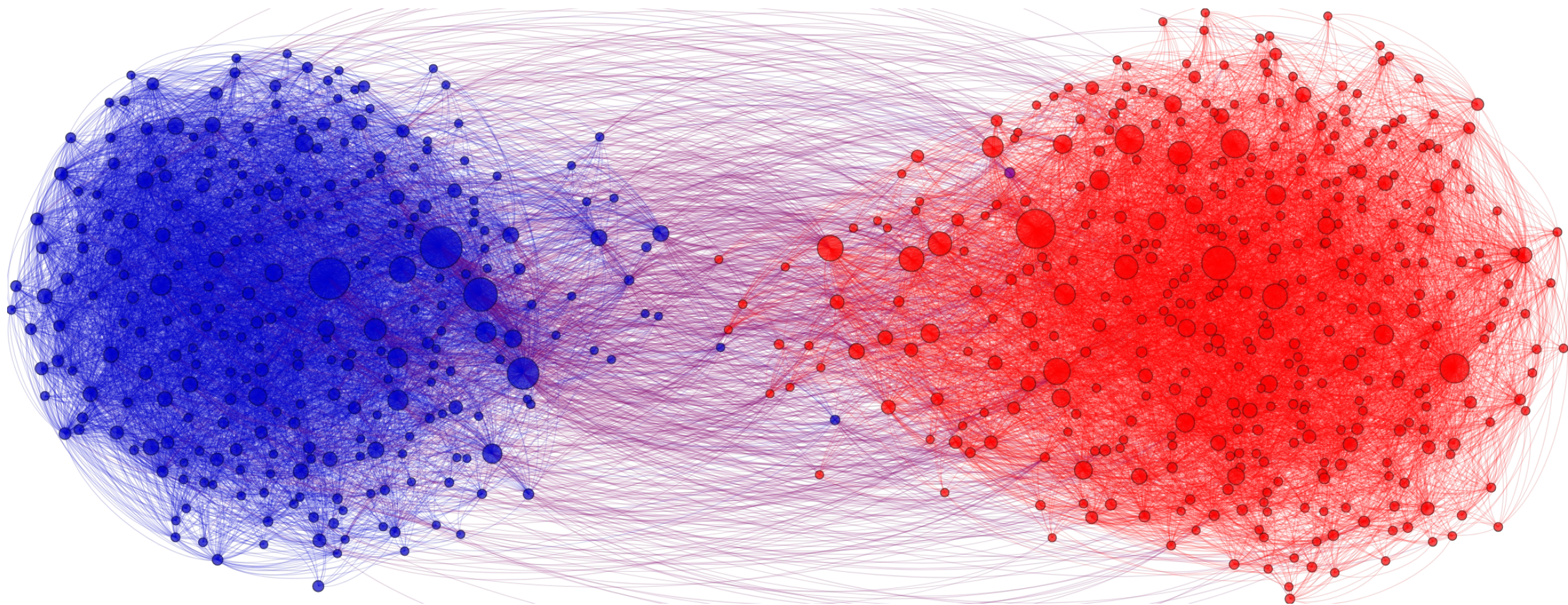
# Applications of Network Analysis: Social Network Analysis



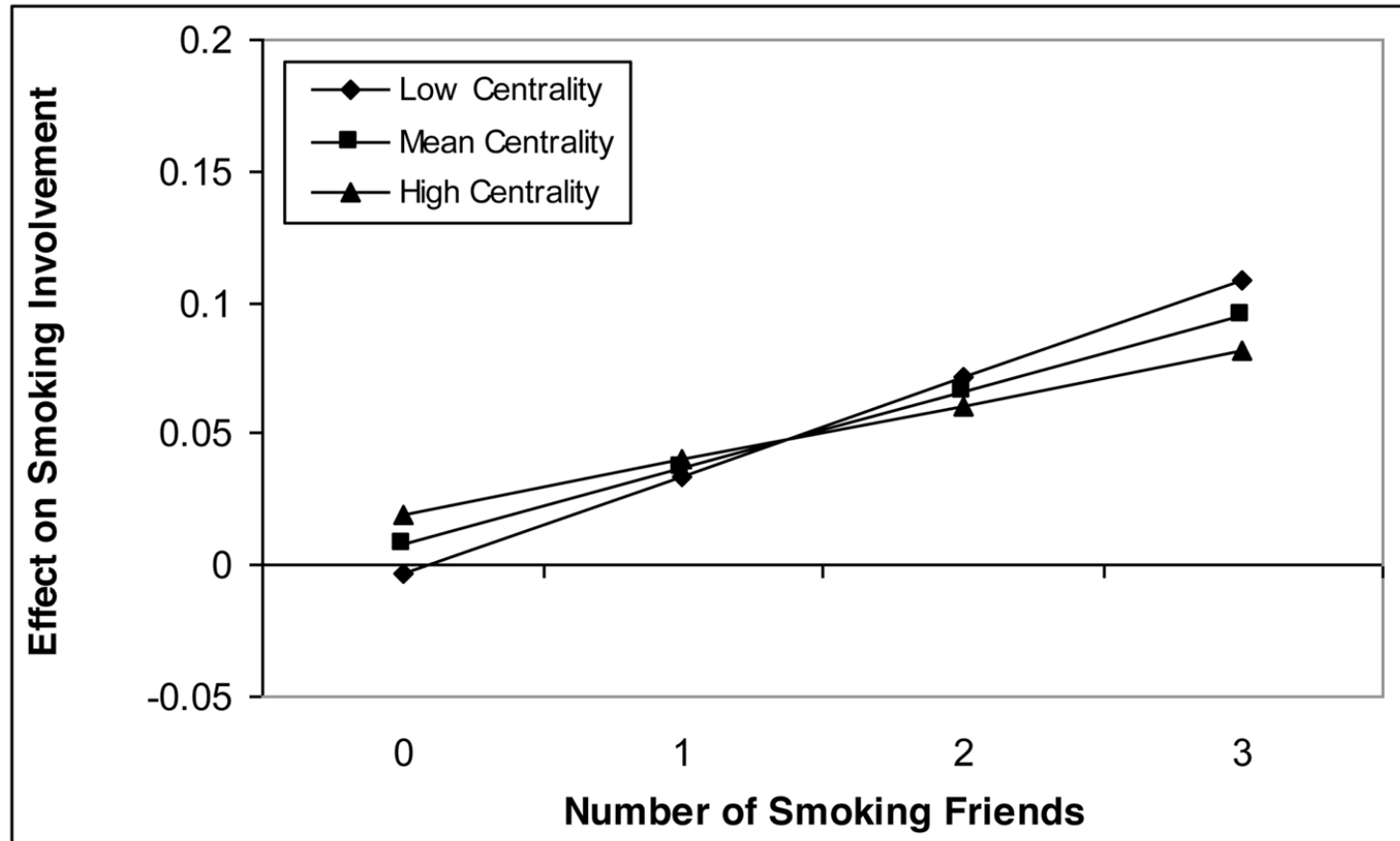
InstaCrawlR



# Political Polarization



# Smoking and Peer Pressure




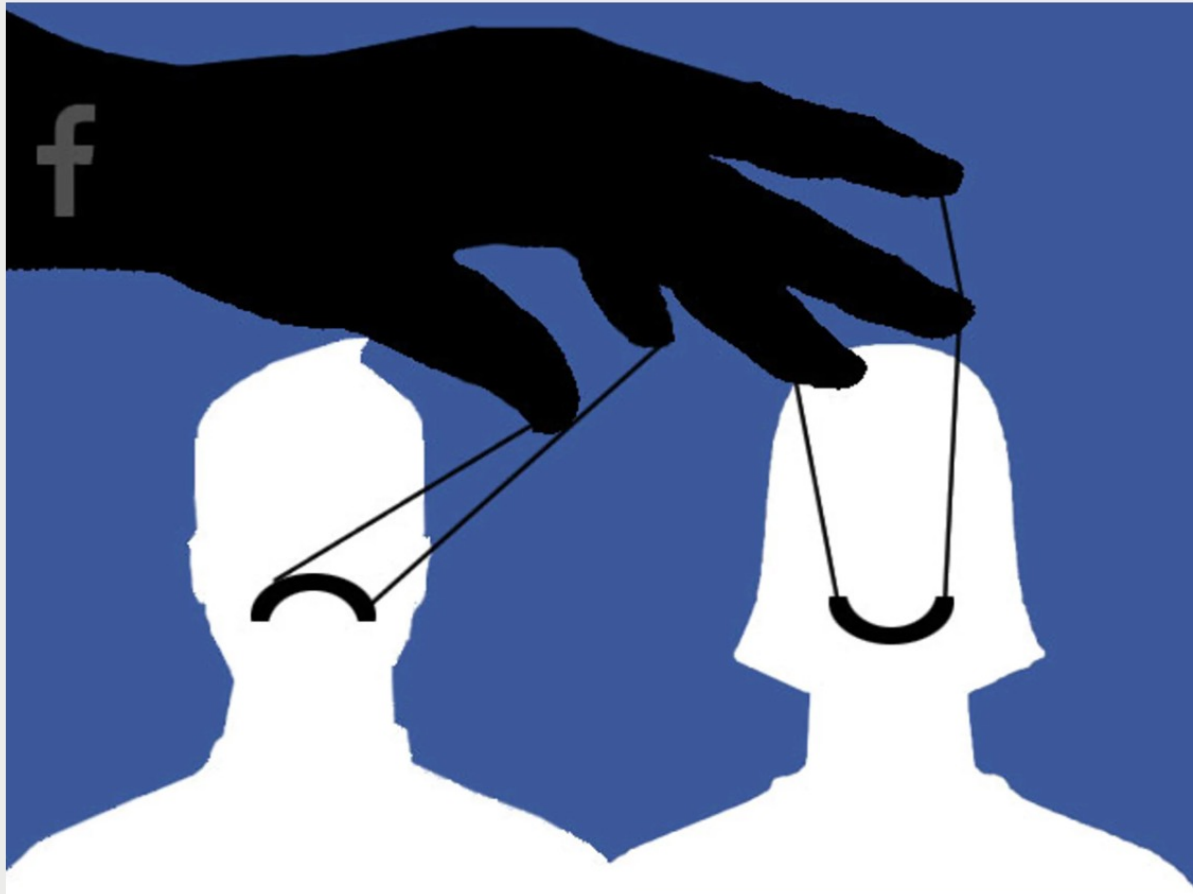
**Figure 1.**  
Effect on smoking involvement of the number of smoking friends by betweenness centrality  
(N=6,579)



# Social Contagion

## Facebook's Secret Experiment: The Era of Manipulation Has Begun > Worrying about privacy is so last week—start worrying about manipulation

BY TEKLA S. PERRY | 30 JUN 2014 | 3 MIN READ | 



Last week, the news began spreading that [Facebook](#), in the name of research, had manipulated users news feeds for a week back in 2012, skewing the distribution of posts to determine the impact on the users' moods. Over the weekend, the rumble turned into an outcry, with privacy activist [Lauren Weinstein](#) tweeting "I wonder if Facebook killed anyone with their emotion manipulation stunt."

The revelation came as a result of the publication, earlier this month, of a [2012 study](#) by researchers at Facebook, Cornell University, and the University of California, San Francisco. The study demonstrated that seeing a lot of positive news on Facebook is likely to lead you to produce more positive posts yourself. The reverse is also true: Seeing lots of negative news pushes people to post more negative material. That example of an emotional contagion was fairly interesting: It countered earlier theories that seeing all your friends having fun on Facebook might make someone's own life seem bleak by comparison. I would much prefer to see happiness as a contagious.

It turned out that the study was conducted by newsfeed manipulation, not just monitoring positive and negative posts.

Without explicit consent (barring a user agreement that says Facebook can pretty much do anything to your page that it wants), and certainly without the knowledge of its users, Facebook turned 700,000 users into research subjects, skewing their news feeds to be more positive, by removing negative posts, more negative, by removing positive posts, or neutral, by removing random posts. Or, in the shorthand of headlines, "[Facebook made users depressed.](#)"



# Terrorist Networks

## How The NSA Uses Social Network Analysis To Map Terrorist Networks

2013 JUNE 12

by Greg Satell

tags: Duncan Watts, Network Theory, Privacy, Social Network Analysis, Social Networks



Ever since [The Guardian reported](#) that the National Security Agency (NSA) has been collecting the phone record metadata of millions of Americans, the cable talk circuit has been ablaze with pundits demanding answers to what should be obvious questions.

Who knew about the program to collect data? (Apparently, [all three branches of government](#)). Who else has been supplying data? (Just about everybody, [according to the Washington Post](#)). What is [metadata](#)? (It's data about data).

# Surveillance

Authorities can collect and analyze details about personal relationships, spiritual beliefs, and sexual preferences, and share them with third parties.

Human and algorithmic bias perpetuates harmful and incorrect stereotypes, disproportionately impacting marginalized communities.

Immigration officials can deny individuals entry based on their political, social, or religious views expressed on social media, or that of their friends and family.

People refrain from speaking out on political, social, and religious issues when they fear their speech could be recorded and potentially used against them.

Individuals become less likely to join organizations and groups if authorities can monitor their memberships and activities.

Authorities can disrupt nonviolent demonstrations before they begin, and track the names of individuals in attendance.

Monitoring eschews democratic legal standards of “reasonable suspicion” and “probable cause,” and instead treats everyone as a suspect of wrongdoing.

**VIOLATES  
PRIVACY**

**ENABLES  
DISCRIMINATION**

**THREATENS MIGRANTS’ RIGHTS**

**RESTRICTS FREE EXPRESSION**

**DISCOURAGES FREEDOM OF ASSOCIATION**

**DISRUPTS FREEDOM OF ASSEMBLY**

**UNDERMINES DUE PROCESS**

## Social Media Surveillance Erodes Rights

Democracy requires vibrant public spaces free from constant surveillance.

# Policy Decision Making

## **Social network analysis for health policy**

---

### **Community detection, roles and information flows in social networks for the analysis of public opinion and patient engagement in public health policy**

*Professor Sophia Yaliraki (Chemistry), Professor Ara Darzi (IGHI, Surgery/Cancer)*

*Others: Dr Mariano Beguerisse (Mathematical Institute, University of Oxford), Sabine Vuik (Surgery & Cancer) Guillermo Garduno (Sinnia, Mexico)*

Other applications you can think of?