VE444: Networks

Yifei Zhu, assistant professor
University of Michigan-Shanghai Jiao Tong University

Acknowledgment:
This lecture's slides are modified from the slides shared by Prof. Jure Leskovec

VE444：Administrivia

Logistics: Teaching Staff

- Instructor: Yifei Zhu
  - Office: JI-Rm440
  - Email: yifei.zhu@sjtu.edu.cn
  - Office hour: TBD
- TA:
  - Sheng Cen
  - Email:cens98@sjtu.edu.cn
  - Zibo Wang
  - Email: wangzibo@sjtu.edu.cn

Lecturing Schedule

- Weekly: Monday and Wednesday: regular lecturing
- Odd weeks: Friday: mainly discussion

- Recitation class: As requested, mixed with discussion class

Course Outline

Weeks	10:00-11:40	Contents
1	Sep.7	Overview
	Sep.9	Graph basics, Strong and weak ties
	Sep.11	Strong and weak ties, homophily
2	Sep.14	Postive and Negative relationships;
	Sep.16	Game basics
3	Sep.21	multiple Equilibria; Pareto Optimality
	Sep.23	Traffic as a game;
	Sep.25	Auction basics;
4	Sep.28	Matching markets
	Sep.30	Market-clearing price; price construction
5	Oct.5	Trading networks and its equilibria analysis
	Oct.7	Power in networks; two-person bargaining
	Oct.9	Two-person bargaining
6	Oct.12	Network exchange models
	Oct.14	Structure of the web; ranking and link analysis
7	Oct.19	PageRank
	Oct.21	Advertising; VCG
	Oct.23	VCG and Generalized second price auction
8	Oct.26	Herding; Bayes' Rule
	Oct.28	Cascade and its modelling
9	Nov.2	Network effects basics
	Nov.4	Makets with network effects
	Nov.6	Discussion
10	Nov.9	Power Laws
	Nov.11	Diffusion and its modelling
11	Nov.16	Cascasdes and clusters; Thresholds
	Nov.18	Six degree of seperation
	Nov.20	Decentralized search and its modelling
12	Nov.23	Epidemics
	Nov.25	Voting basics
13	Nov.30	Information aggregation
	Dec.2	Tragedy of commons
	Dec.4	Discussion
Exam week	Dec.7-Dec.11	

Logistics: Textbook

- **Textbook:**
  - Title: Networks, Crowds, and Markets: Reasoning About a Highly Connected World
  - Authors: David Easley and Jon Kleinberg
  - Online available: http://www.cs.cornell.edu/home/kleinber/networks-book/
- **Optional Readings:**
  - Mostly research papers and related chapters
- Other books:
- Title Network Science by Albert-Laszlo Barabasi
  - http://networksciencebook.com/

Logistics: Communication

- **Piazza Q&A website:**
  - Signup Link: piazza.com/sjtu.org/fall2020/ve444
  - Class Link: piazza.com/sjtu.org/fall2020/ve444/home
  - **Please participate and help each other!**
    - Don't post code, annotate your questions, search for answers before you ask
- **To reach course staff (prof/TAs), always use:**
  - **"VE444" in your subject**
  - **Grading: reach TA first**
- We will post all related course announcements in Canvas

Work for the Course & Grading

- **Final grade will be composed of:**
  - **Homework: 30% (3%+3*9%)**
    - Homeworks 0, 1, 2, 3
  - **Final exam: 30%**
  - **Course project: 40%**
    - Proposal: 20%
    - Project milestone: 20%
    - Final report: 50%
    - Poster presentation: 10%

I look forward to seeing it! ☺

Exam

- **Duration: 100 minutes**
- **Depending on the COVID-19**
  - **1-paper cheat note or open book**
- **We strictly enforce the JI Honor Code:**
  - https://www.ji.sjtu.edu.cn/academics/academic-integrity/honor-code/

Course Projects (TBD)

- **Course project:**
  - **Empirical analysis** of network data to develop a model of behavior
  - **Algorithms and models** to make predictions on a network dataset
  - **Scalable algorithms** for massive graphs
  - **Theoretical project** that considers a model/algorithm and derives a rigorous result about it
- **Performed in groups of up to 3 students**
  - Currently 82 students.
  - Fine to have groups of 1 or 2. The team size will be taken under consideration when evaluating the scope of the project in breadth and depth. But 3 person teams can be more efficient.
  - We will help with ideas, data and mentoring
  - Start thinking about this now!
- Do NOT COPY CODE ONLINE or FROM OTHERS

Course Schedule

Week	Assignment
1	Homework 0
4	Homework 1
6	Project proposal
8	Homework 2
9	Project milestone
12	Homework 3
13	Project report
	Poster session
14	Exam

Prerequisites

- **No single topic in the course is too hard by itself**
- **But we will cover and touch upon many topics and this is what makes the course hard**
  - **Good background in:**
    - Algorithms and graph theory
    - Probability and statistics
    - Linear algebra
  - **Programming:**
    - You should be able to write non-trivial programs (in C++ , Matlab or Python)

Questions so far?

Background Survey

- Which year are you in?
- Have you taken OS, Computer network course?
- Do you plan to take these courses?

What this subject is NOT

- Computer networking
- Detailed online social networking applications:
  - Facebooking, Wechatting, Twittering, TikTok

What this subject is

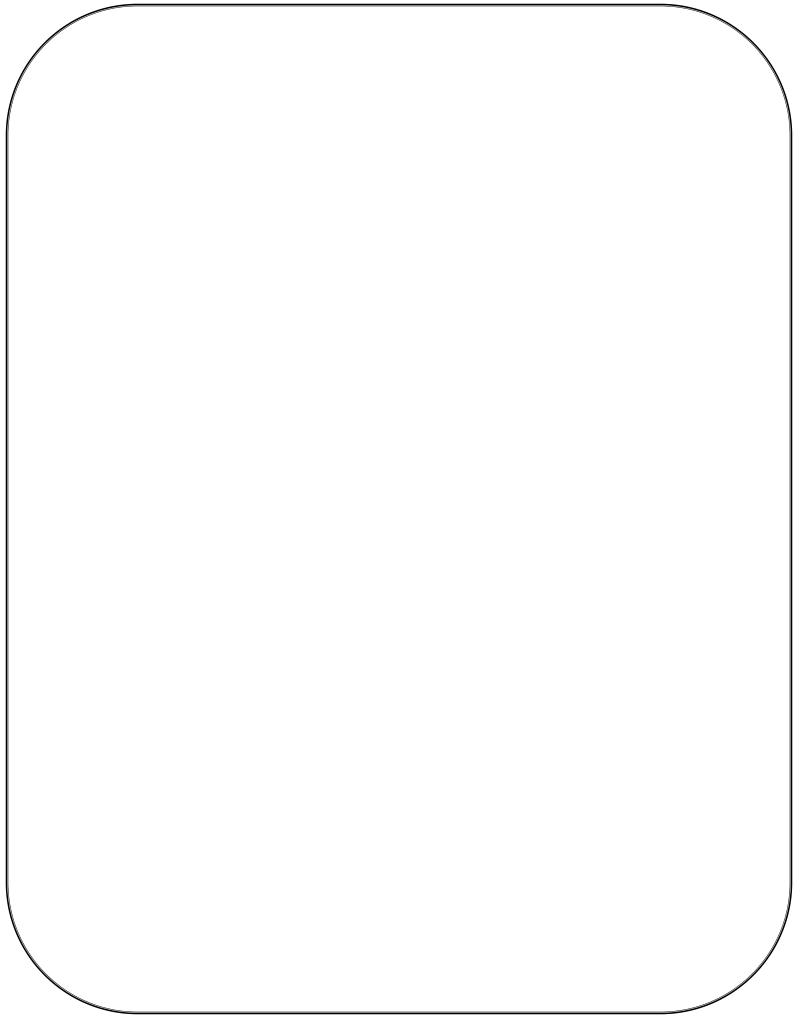
- From the discipline point of view:
  - Social sciences
  - Computer sciences
  - Economy
- Data science? Network Science
- Science
  - We study insights
  - More importantly, on the fundamentals
    - Rigorous quantification

What this subject is

- From the mathematical tool's point of view:
  - Graph Theory
  - Game Theory

Course Outline:Anatomy

Weeks	10:00-11:40	Contents
1	Sep.7	Overview
	Sep.9	Graph basics, Strong and weak ties
	Sep.11	Strong and weak ties, homophily
2	Sep.14	Postive and Negative relationships;
	Sep.16	Game basics
3	Sep.21	multiple Equilibria; Pareto Optimality
	Sep.23	Traffic as a game;
	Sep.25	Auction basics;
4	Sep.28	Matching markets
	Sep.30	Market-clearing price; price construction
5	Oct.5	Trading networks and its equilibria analysis
	Oct.7	Power in networks; two-person bargaining
	Oct.9	Two-person bargaining
6	Oct.12	Network exchange models
	Oct.14	Structure of the web; ranking and link analysis
7	Oct.19	PageRank
	Oct.21	Advertising; VCG
	Oct.23	VCG and Generalized second price auction
8	Oct.26	Herding; Bayes' Rule
	Oct.28	Cascade and its modelling
9	Nov.2	Network effects basics
	Nov.4	Makets with network effects
	Nov.6	Discussion
10	Nov.9	Power Laws
	Nov.11	Diffusion and its modelling
11	Nov.16	Cascasdes and clusters; Thresholds
	Nov.18	Six degree of seperation
	Nov.20	Decentralized search and its modelling
12	Nov.23	Epidemics
	Nov.25	Voting basics
13	Nov.30	Information aggregation
	Dec.2	Tragedy of commons
	Dec.4	Discussion
Exam week	Dec.7-Dec.11	



Similar Courses

University	Department	Name
UCB	CS	294-179. Network Structure and Epidemics
CMU	ECE	8-755: Networks in the real world
	Statistics & Data science/Economics	73449 Social, Economic and Information Networks
Stanford	CS	224W Analysis of Networks: Mining and learning with graphs
	Management Science	135: Networks
MIT	Economics	6.207/14.15: Networks
Caltech	CMS/CS/EE/IDS	144: Networks: Structures & Economics
Cornell	Economics/ Sociology/ Computer Science/ Information Science	2040/2090/2850/2040 Networks

Course

- **No single topic in the course is too hard by itself**
- **But we will cover and touch upon many topics and this is what makes the course hard**
  - **Good background in:**
    - Algorithms and graph theory
    - Probability and statistics
    - Linear algebra
  - **Programming:**
    - You should be able to write non-trivial programs (in C++ , Matlab or Python)

Learning strategy

- Take lectures
- Extra readings
- Leverage your project
- Online resources available: edX, Coursera, Youtube

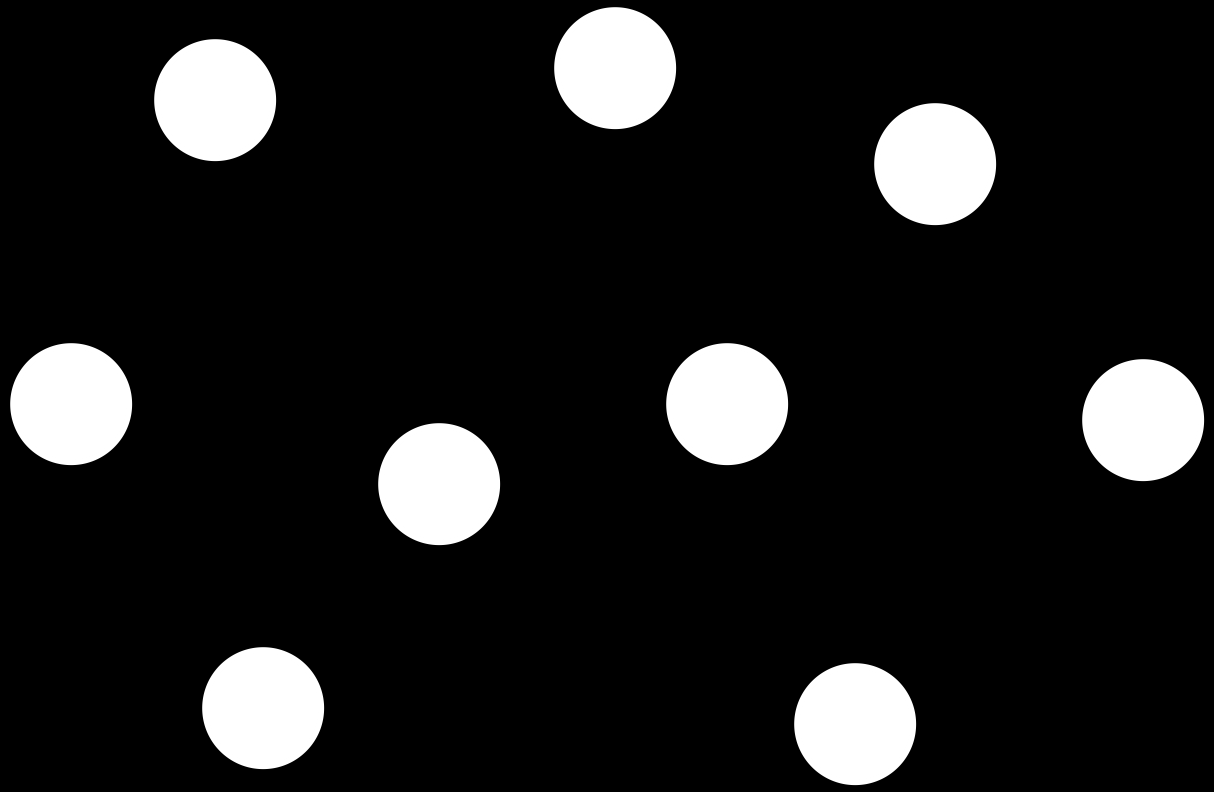
Questions so far?

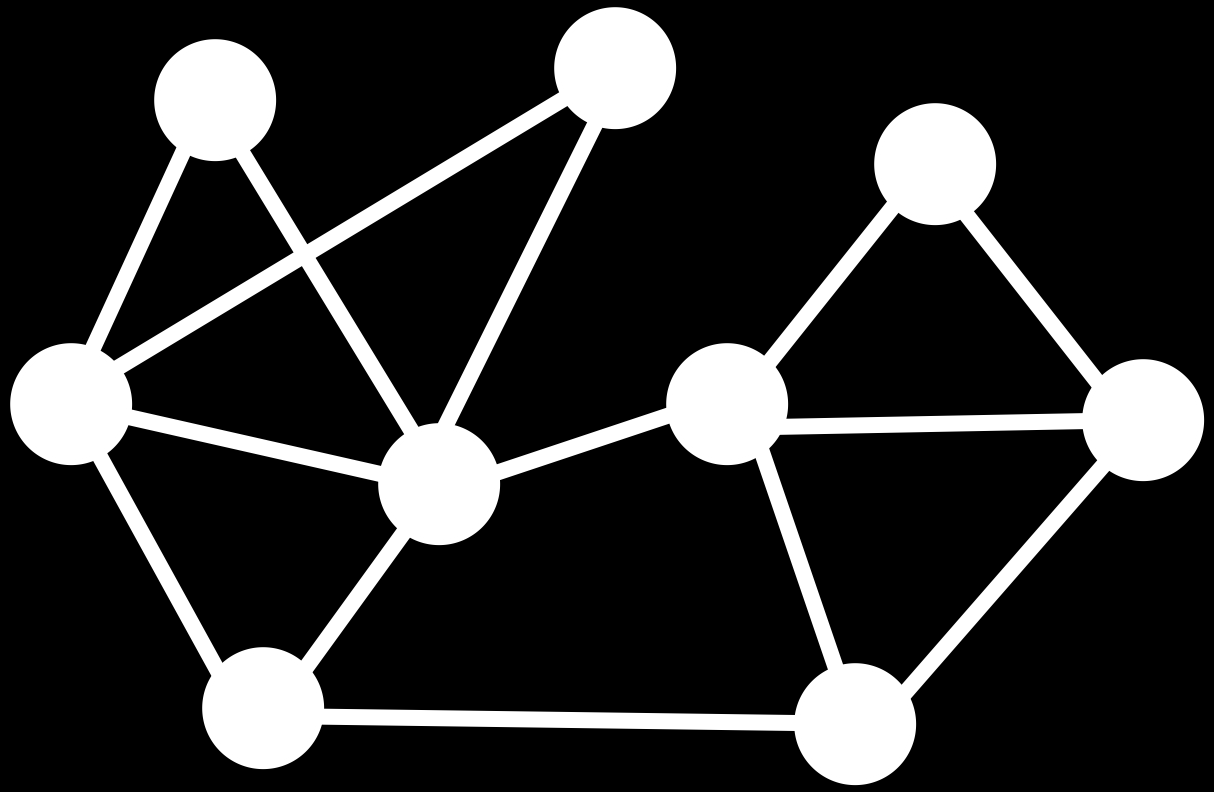
Let's start!

Why Networks?

Networks are a general language for describing complex systems of interacting entities

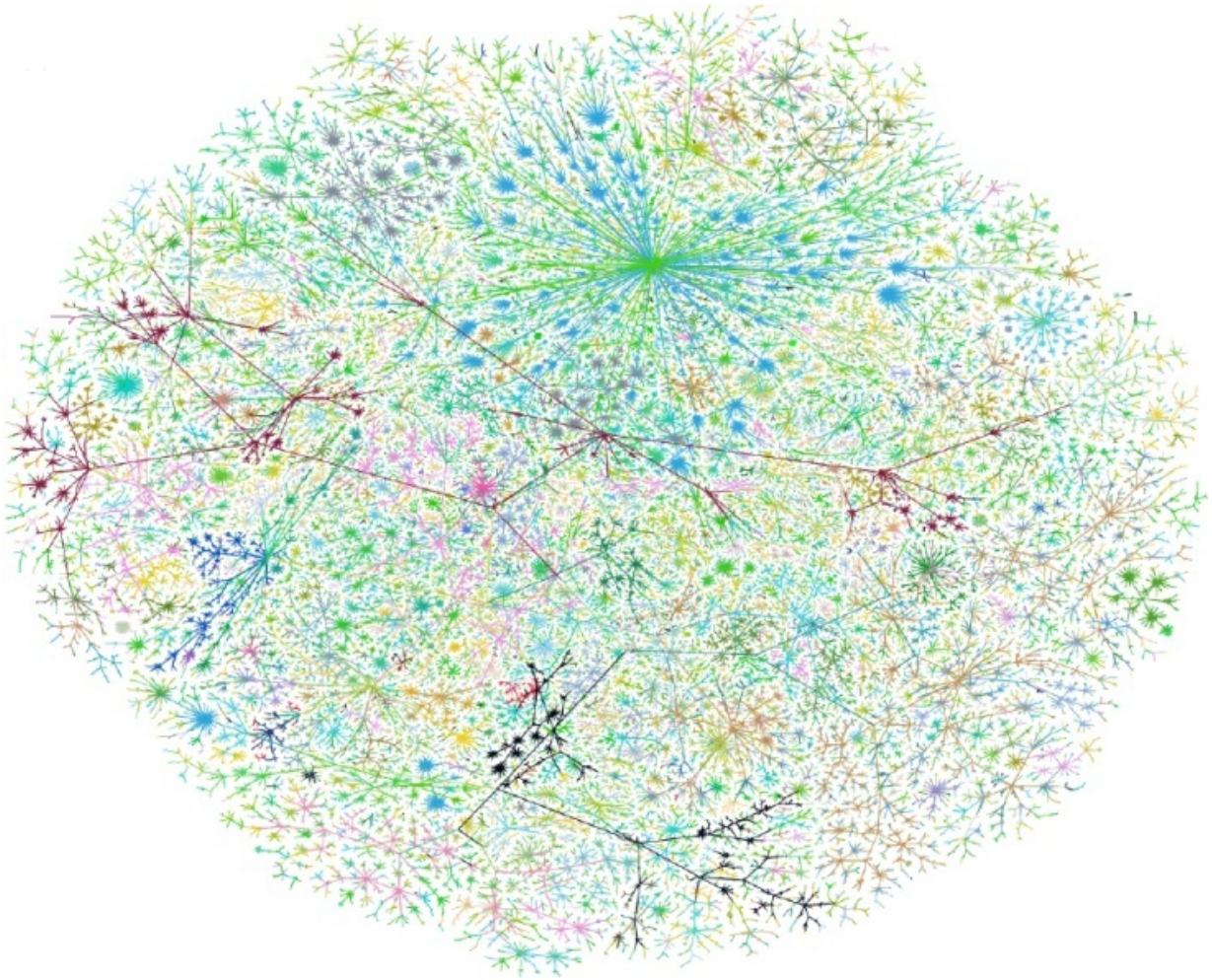
What is a Network?





Network

Any collection of objects in which some pairs of these objects are connected



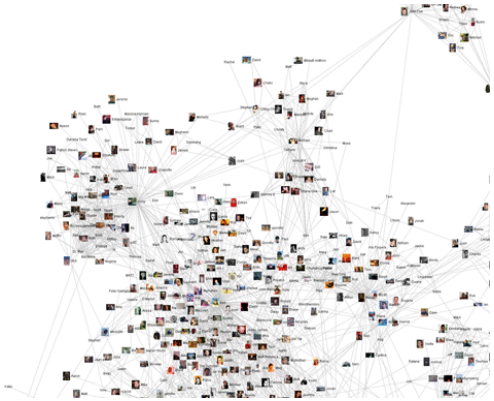
The Network!

Two Types of Networks/Graphs

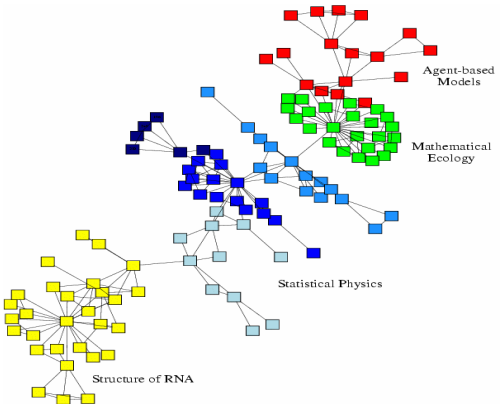
- **Networks (also known as Natural Graphs):**
  - **Society** is a collection of 7+ billion individuals
  - Interactions between **genes/proteins** regulate life
  - Our **thoughts** are hidden in the connections between billions of neurons in our brain
- **Information Graphs:**
  - **Information/knowledge** are organized and linked
  - **Scene graphs:** how objects in a scene relate
  - **Similarity networks:** take data, connect similar points
  - **Communication systems** link electronic devices

Sometimes the distinction is blurred

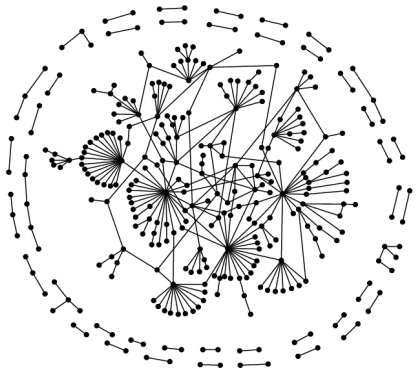
Many Types of Data are Networks



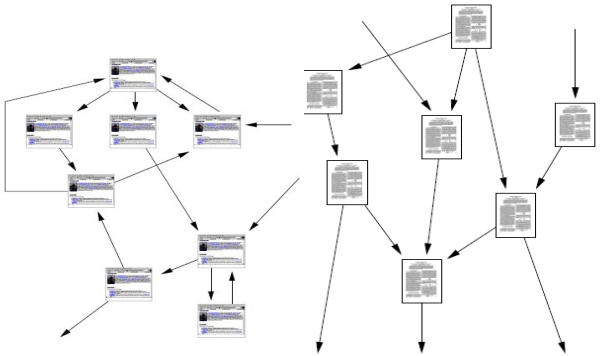
Social networks



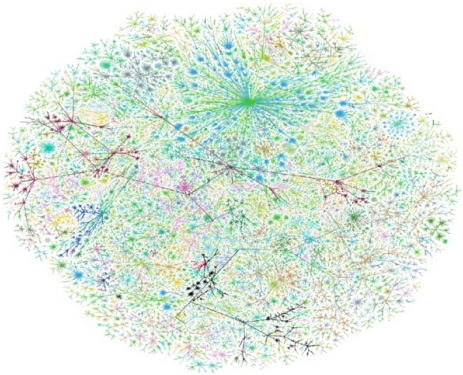
Economic networks



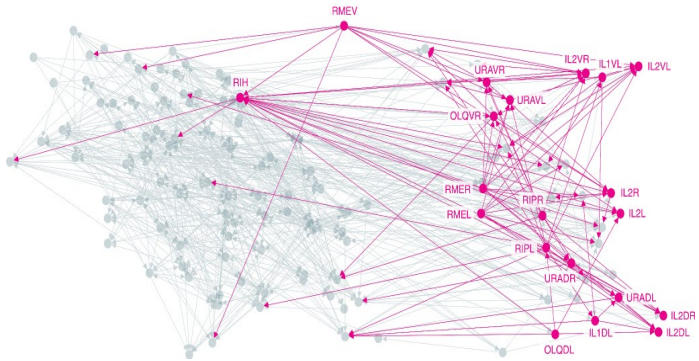
Communication networks



Information networks: Web & citations

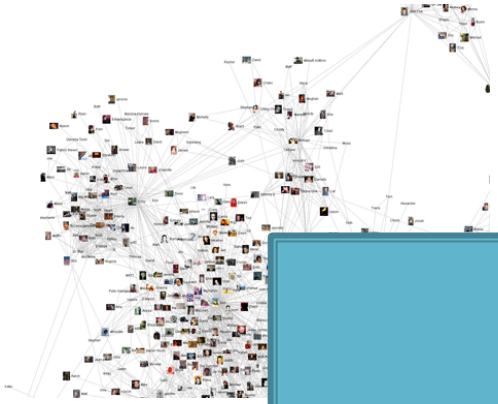


Internet

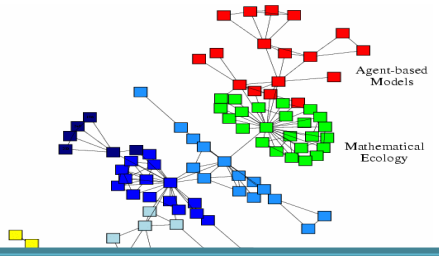


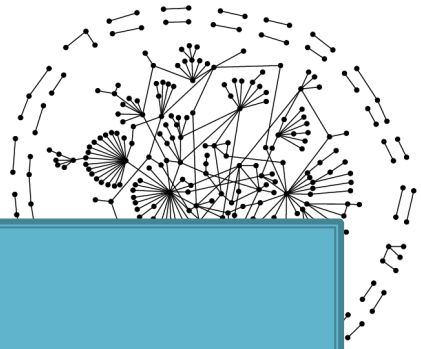
Networks of neurons

Many Types of Data are Networks



Social ne

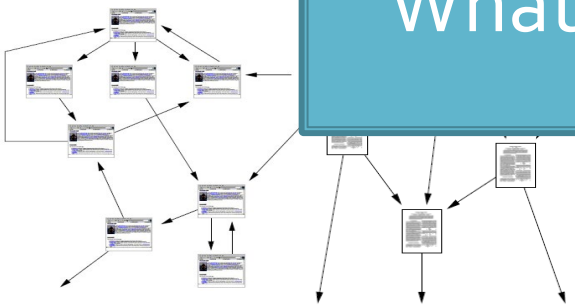




graphs

Main questions:

How are these systems organized?
What are their design properties?



Information networks:
Web & citations

Internet

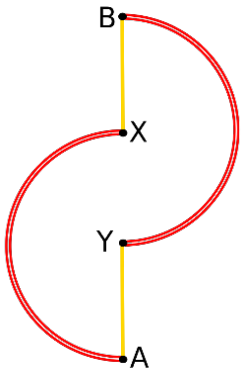
Networks of neurons

Two Types of Connectedness

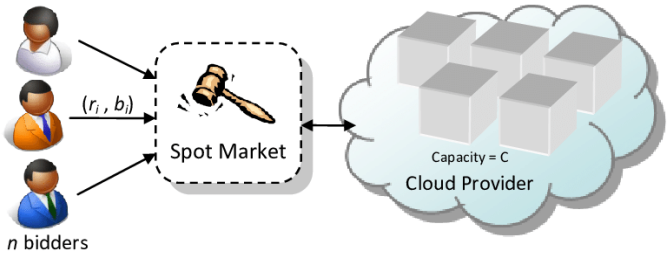
- **Structural level:**
  - **who is linked to whom**
- **Behavior level:**
  - **Individual's actions have implicit consequences for others**

Sometimes they could mix with each other

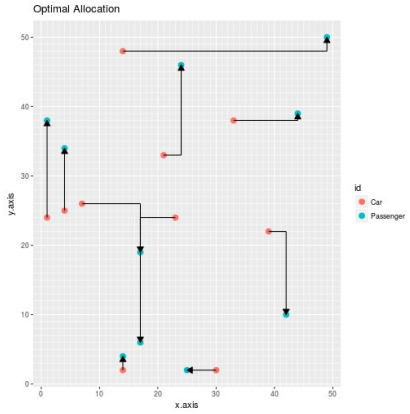
Many Types of Networked Systems



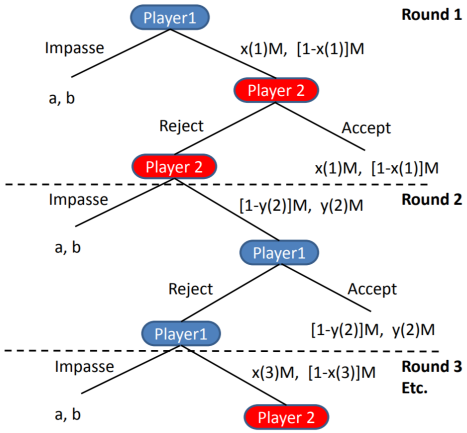
Traffic networks



Cloud networks

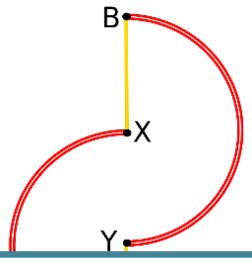


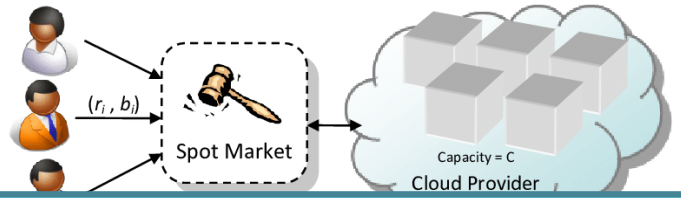
Car-sharing networks



Trading networks

Many Types of Networked Systems

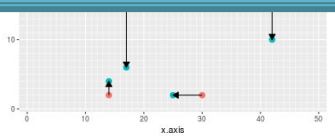




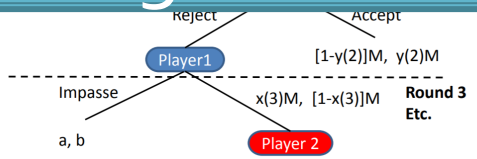
Main questions:

How to reason about the behavior and interaction?

How to design mechanisms to better achieve certain goals?



Car-sharing networks



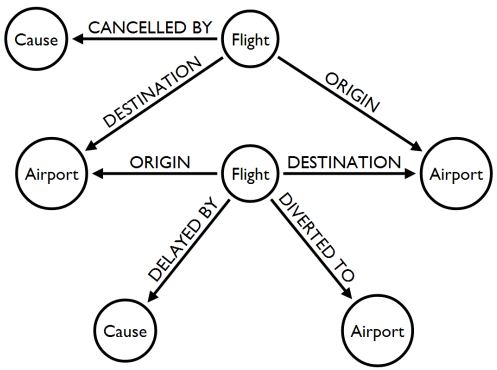
Trading networks

Networks: Knowledge Discovery

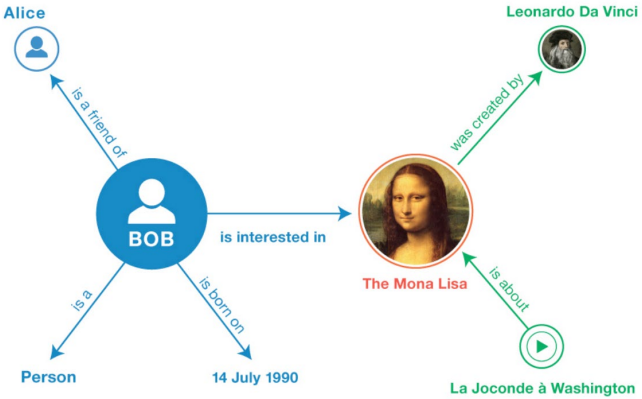
Behind many systems there is an intricate wiring diagram, **a network**, that defines the **interactions** between the components

We will never be able to model and predict these systems unless we understand the networks behind them

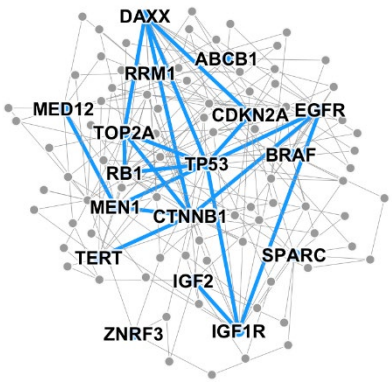
Many Types of Data are Graphs



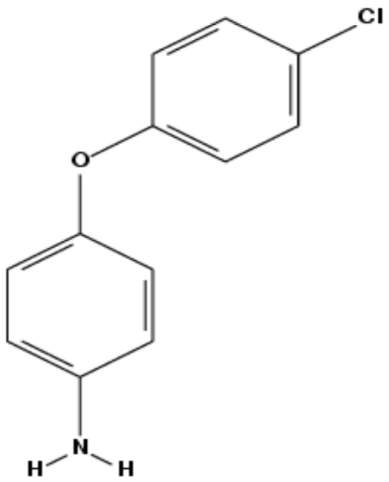
Event Graphs



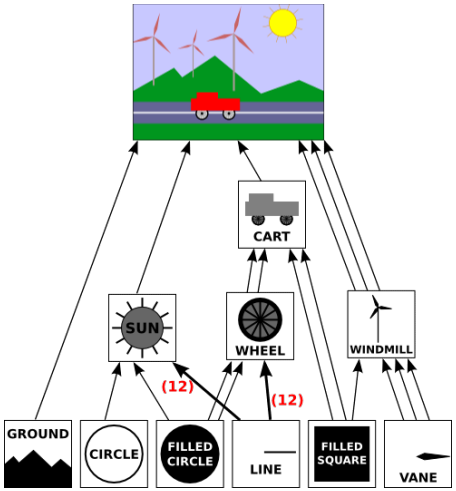
Knowledge Graphs



Disease pathways



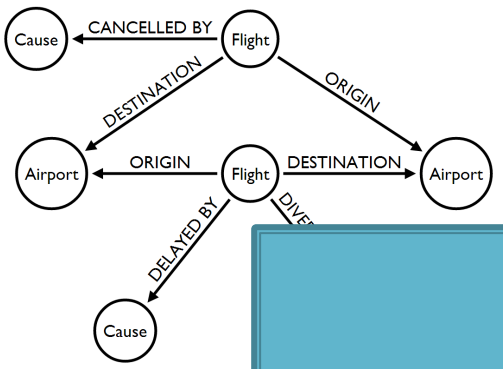
Molecules

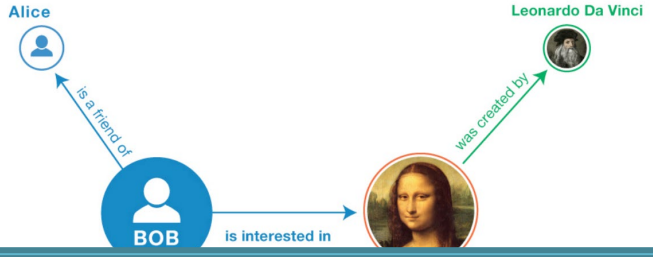


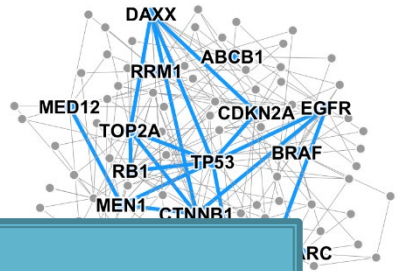
Scene Graphs

Cell-cell similarity networks

Many Types of Data are Graphs





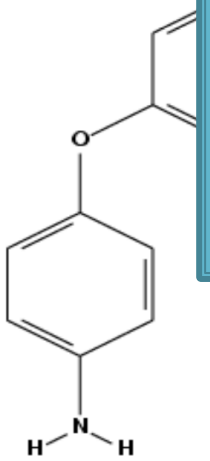


Event

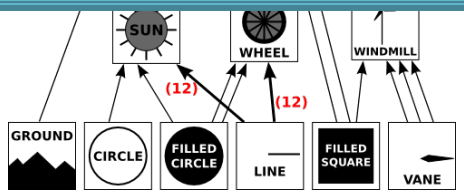
ways

Main questions:

How do we take advantage of relational structure for better prediction?



Molecules



Scene Graphs

Cell-cell similarity networks

Graphs: Machine Learning

Complex domains (knowledge, text, images, etc.) have rich relational structure, which can be represented as a **relational graph**

By explicitly modeling relationships we achieve better performance

But professor, why should I care about networks?

Why Networks? Why Now?

- **Universal language for describing complex data**
  - Networks from science, nature, and technology are more similar than one would expect
- **Shared vocabulary between fields**
  - Computer Science, Social Science, Physics, Economics, Statistics, Biology
- **Data availability & computational challenges**
  - Web/mobile, bio, health, and medical
- **Impact!**
  - Social networking, Drug design, AI reasoning

Networks: Impact

- Google
- Cisco
- Facebook
- Amazon
- Pinterest

Networks and Applications

Ways to Analyze Networks

- **Predict the type/color of a given node**
  - Node classification
- **Predict whether two nodes are linked**
  - Link prediction
- **Identify densely linked clusters of nodes**
  - Community detection
- **Measure similarity of two nodes/networks**
  - Network similarity

Ways to Analyze Networks

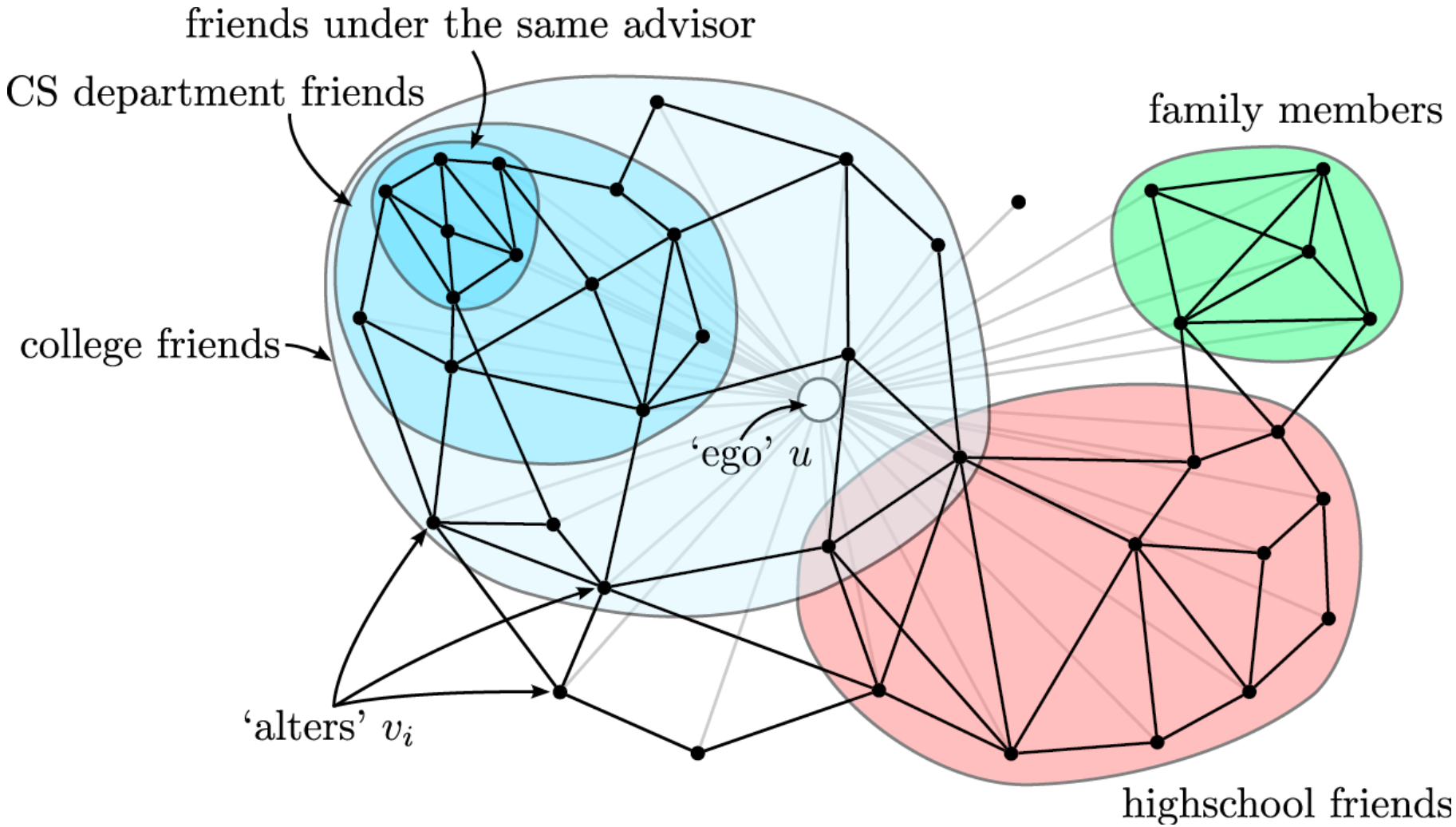
- **Predict the behavior of players**
  - Strategy analysis
- **Predict the stable state of the system**
  - Equilibria analysis
- **Design rules to achieve desired goals**
  - Mechanism design

(1) Networks: Social Networks

Facebook social graph

4-degrees of separation [Backstrom-Boldi-Rosa-Ugander-Vigna, 2011]

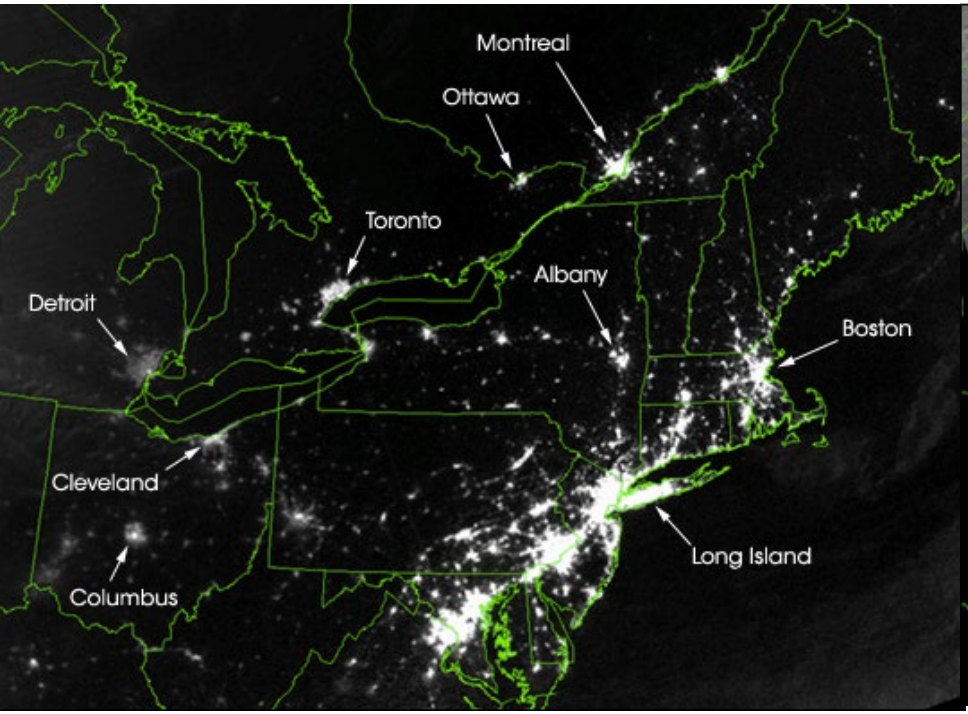
Application: Social Circle Detection



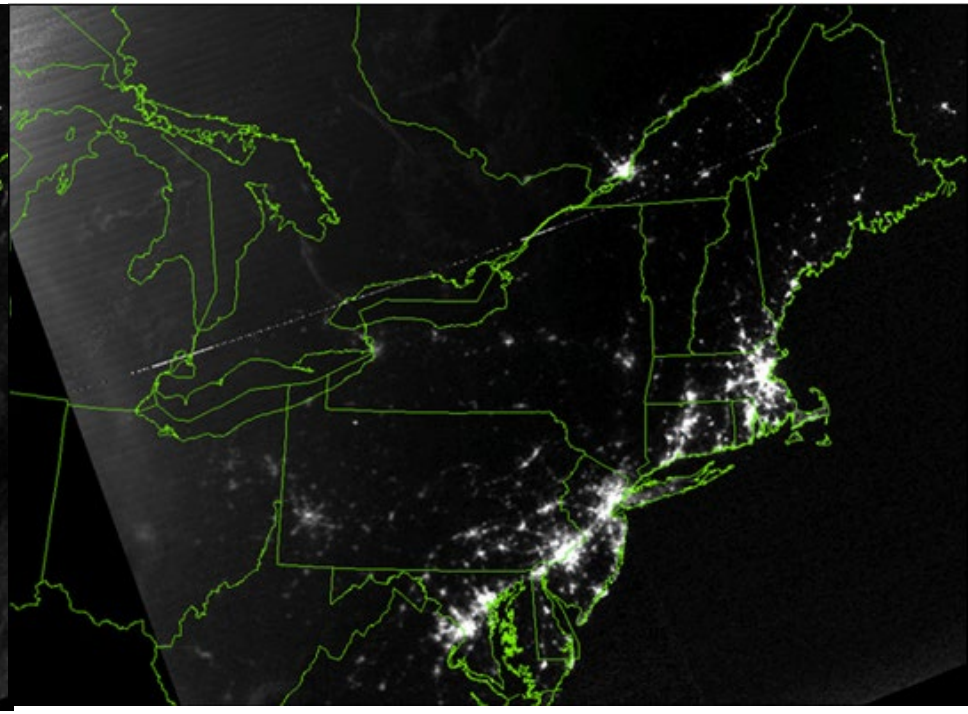
Discover circles and why they exist

(2) Networks: Infrastructure

- **August 15, 2003 blackout**



August 14, 2003: 9:29pm EDT
20 hours before



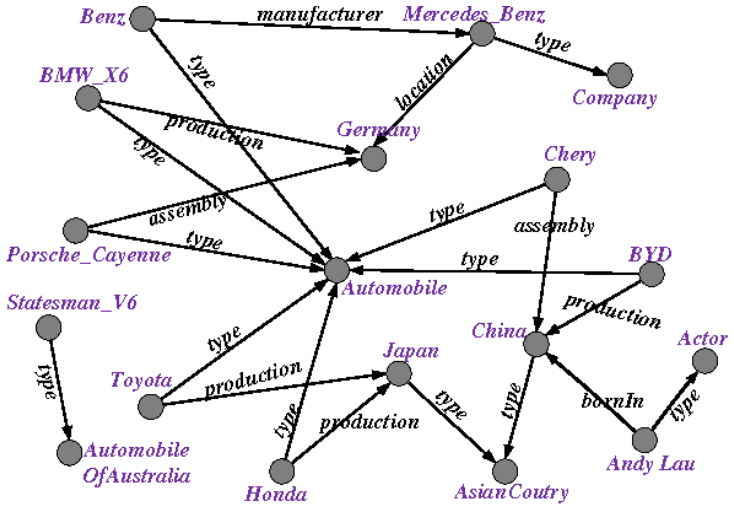
August 15, 2003: 9:14pm EDT
7 hours after

Application: Aug 15, 2003 blackout

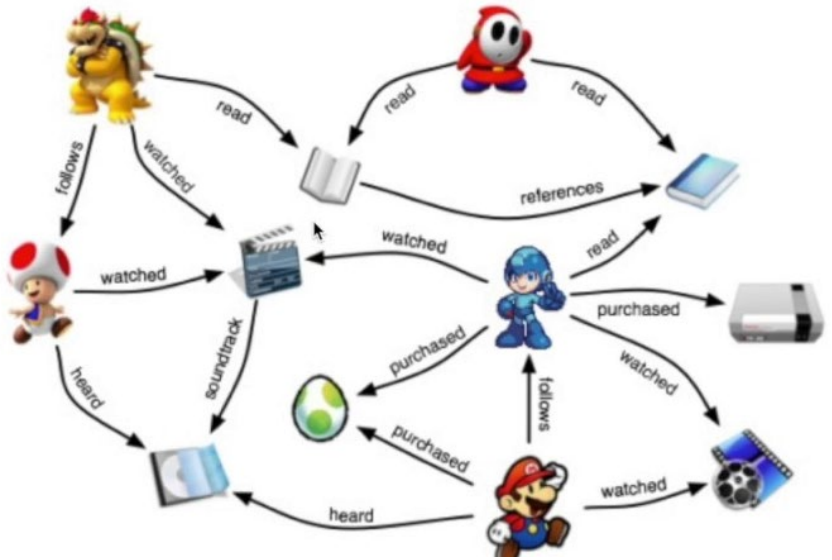
This reveals two important themes of this class:

- We must understand how network structure affects the robustness of a system
- Develop quantitative tools to assess the interplay between network structure and the dynamical processes on the networks, and their impact on failures
- We will learn that in reality failures follow reproducible laws, that can be quantified and even predicted using the tools of networks

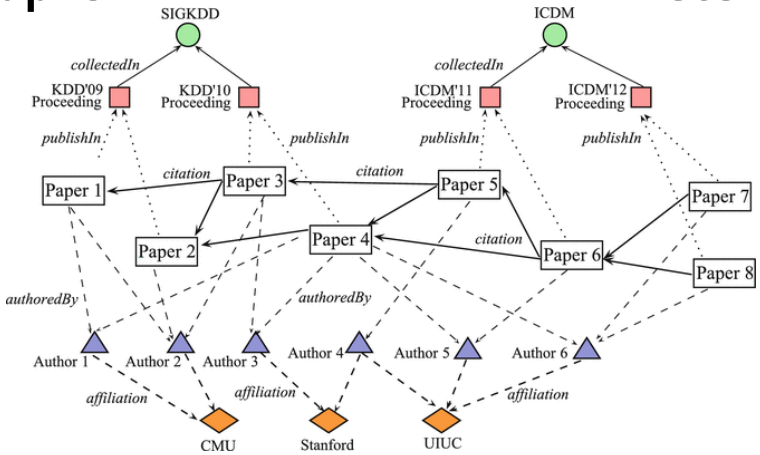
(3) Networks: Knowledge



Knowledge Graphs

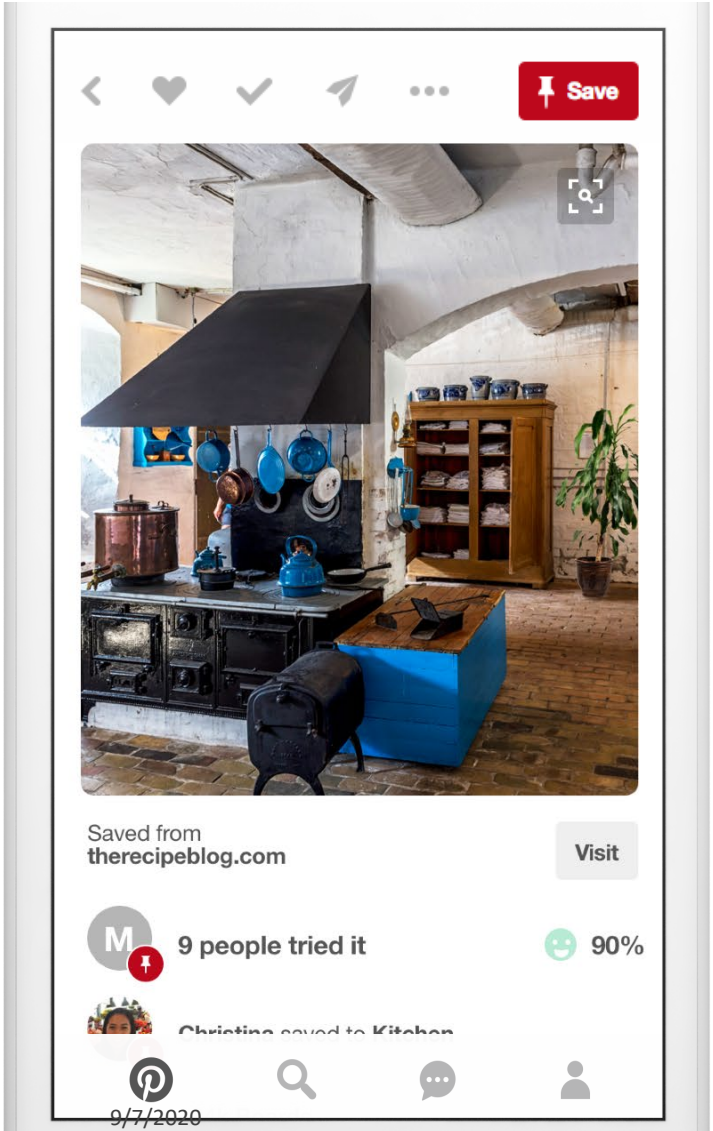


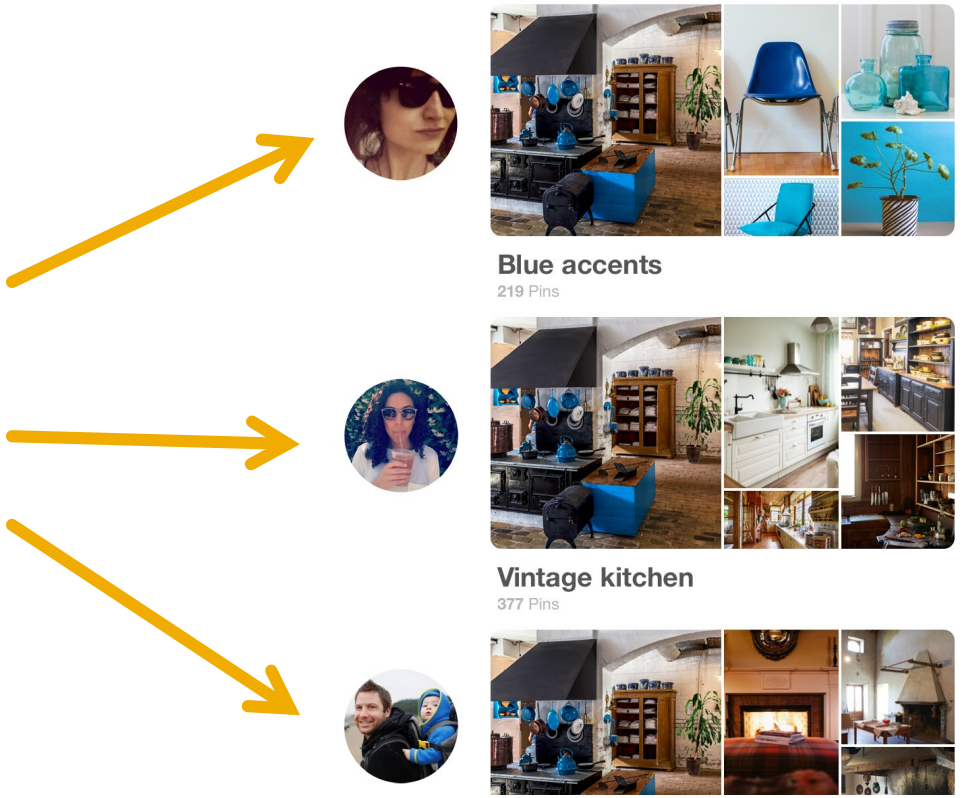
Heterogeneous Graphs



Multimodal Graphs

Application: Knowledge Graphs

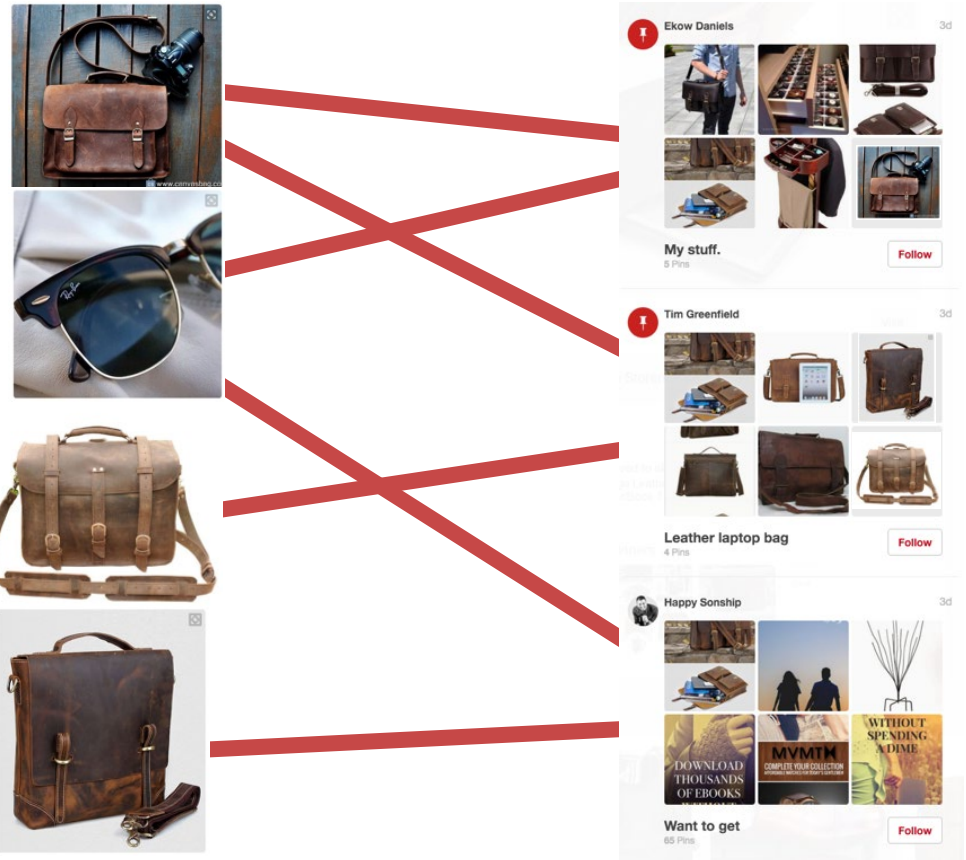




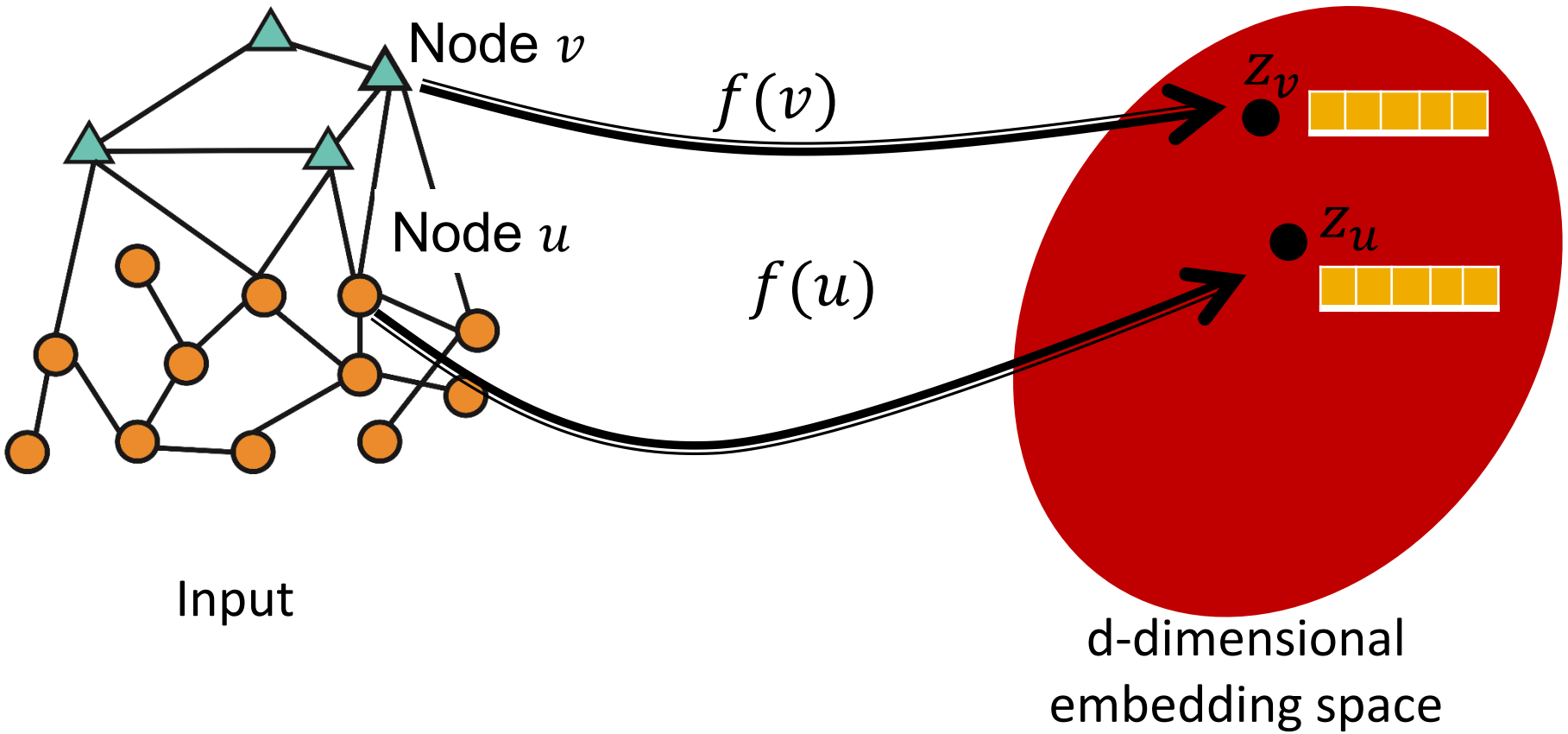
- 300M users
- 4+B pins, 2+B boards

Application: Link Prediction

Content recommendation is link prediction



Embedding Nodes



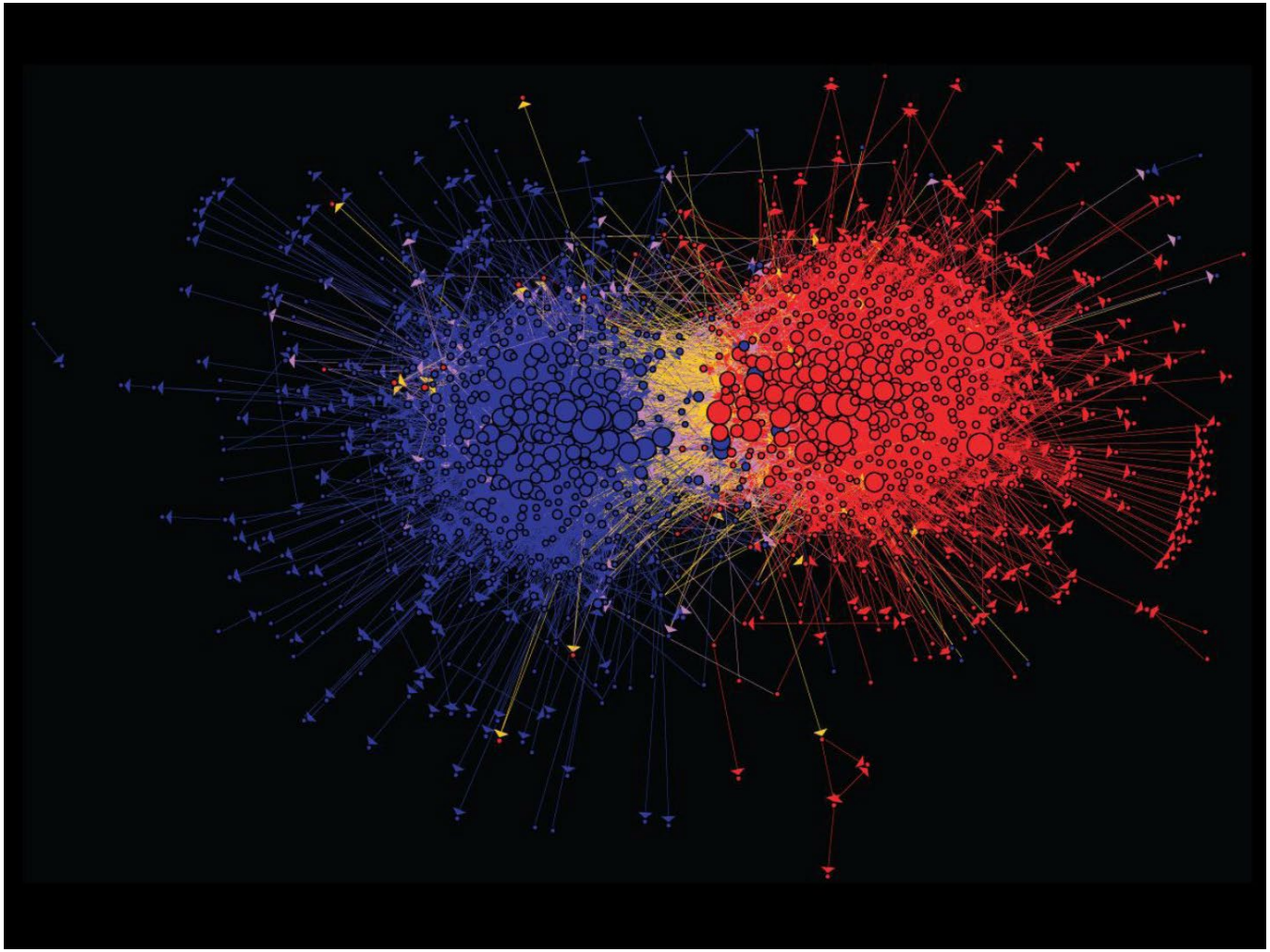
Goal: Map nodes to d-dimensional embeddings such that nodes with similar network neighbourhoods are embedded close together

Example Recommendations

Query

Graph-based algorithm

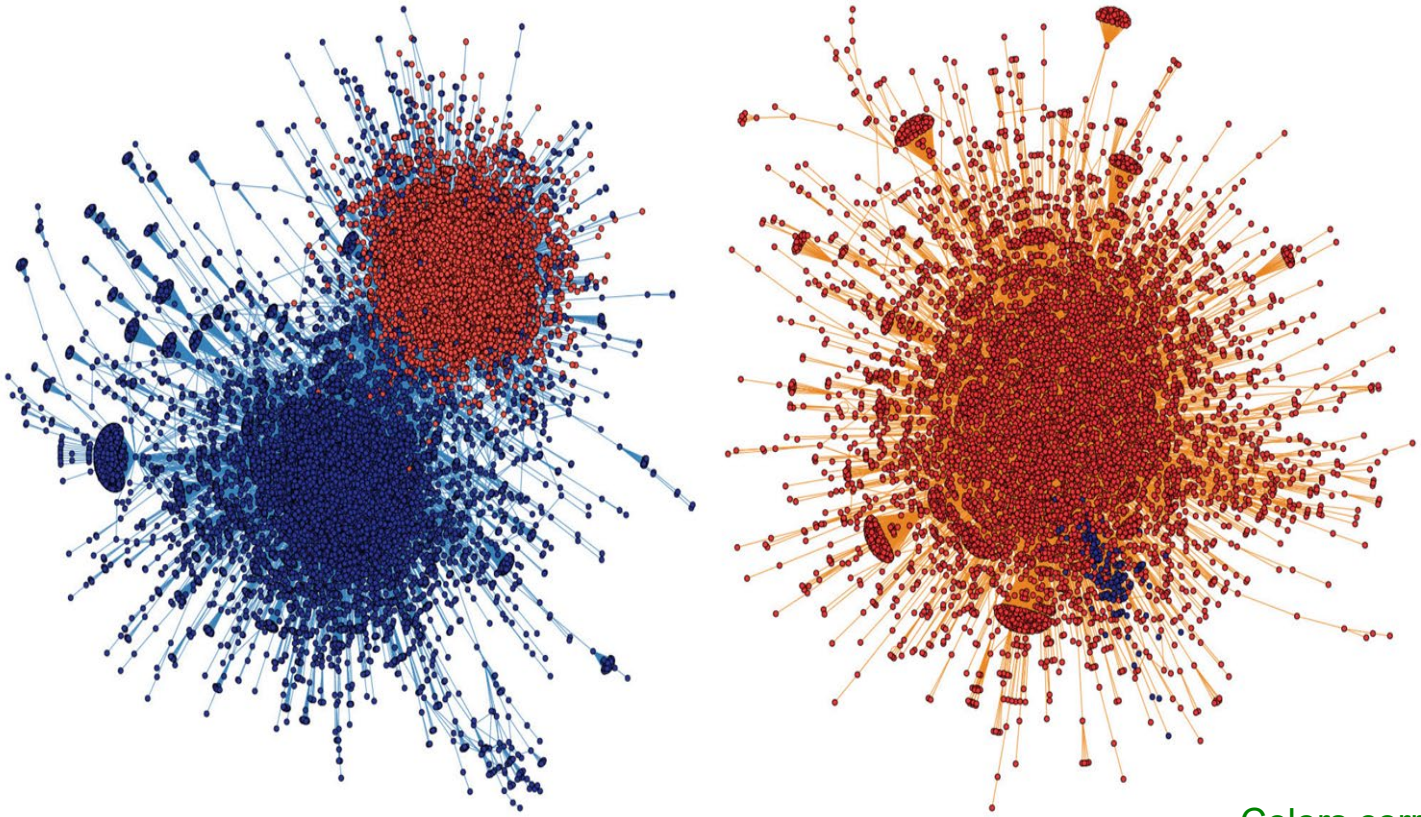
(4) Networks: Online Media



Connections between political blogs

Polarization of the network [Adamic-Glance, 2005]

Application: Polarization on Twitter



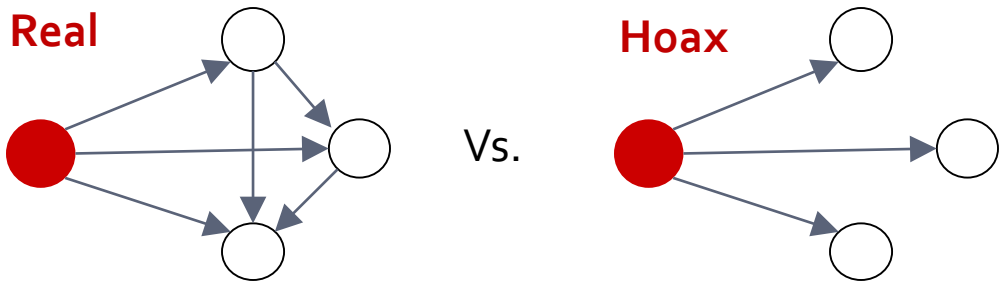
Colors correspond to clusters in the network

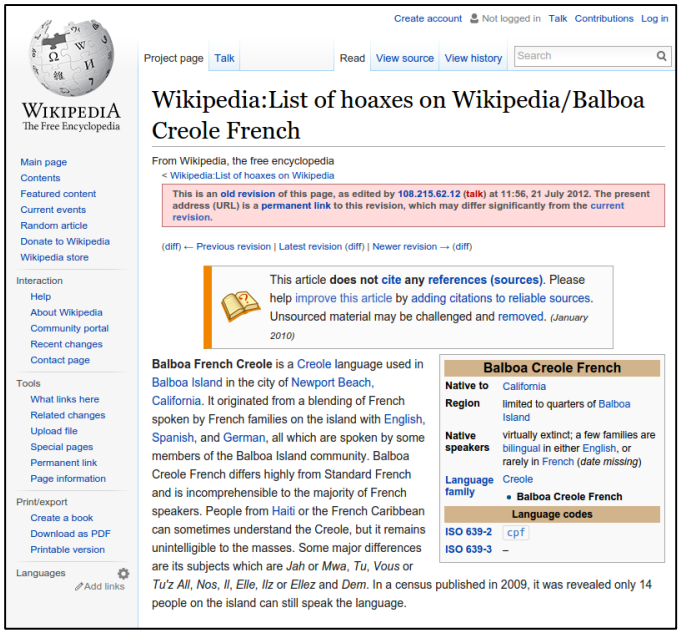
- **Retweet networks:**
  Polarized (left), Unpolarized (right)

Conover, M., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., & Flammini, A. "Political Polarization on Twitter." (2011)

Application: Misinformation

- **Q: Is a given Wikipedia article a hoax?**
  - Real articles link more coherently:





Hoax article detection performance:

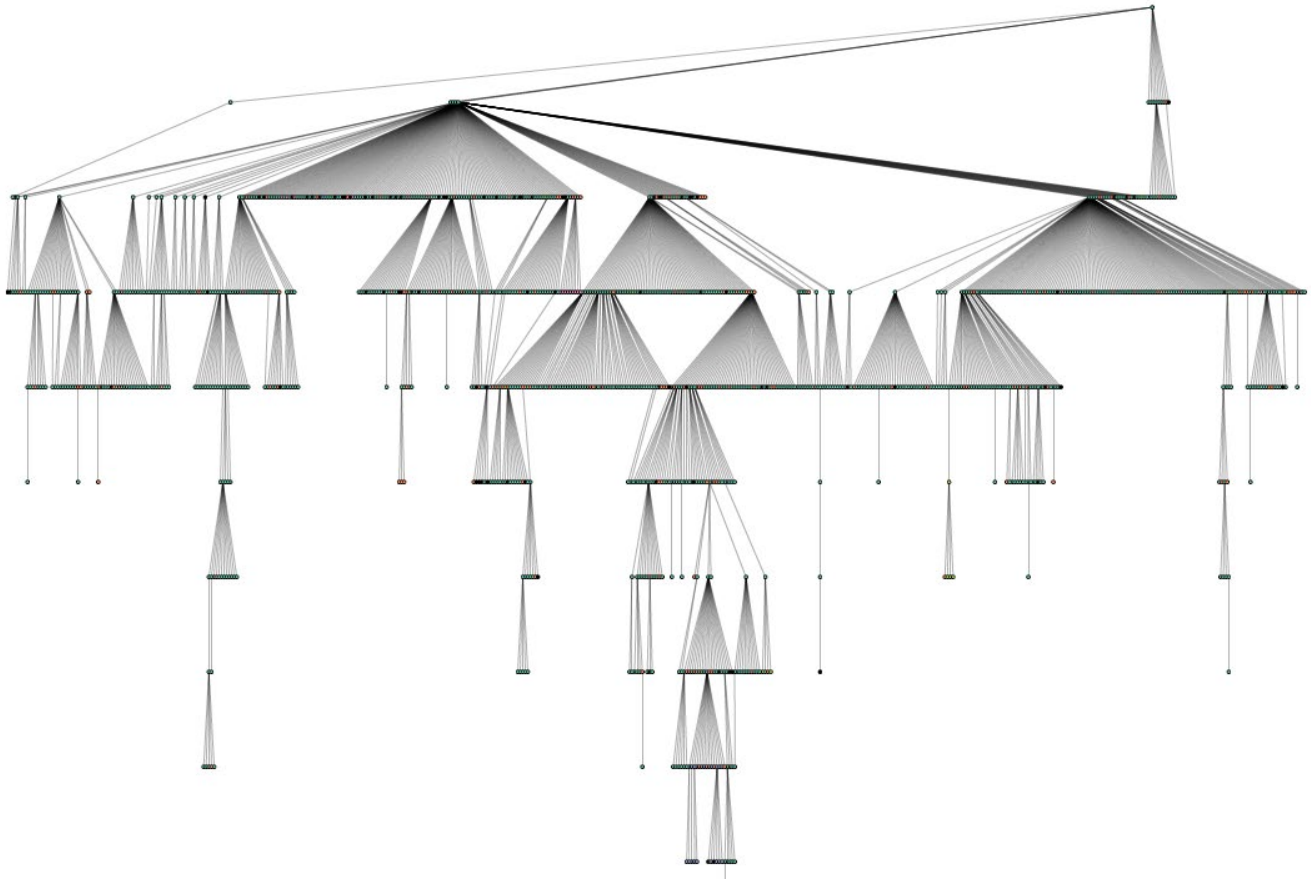
50%	66%	86%
Random	Human	Network

Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes. Kumar et al. WWW '16.

Application: Predicting Virality

Information cascade in social networks

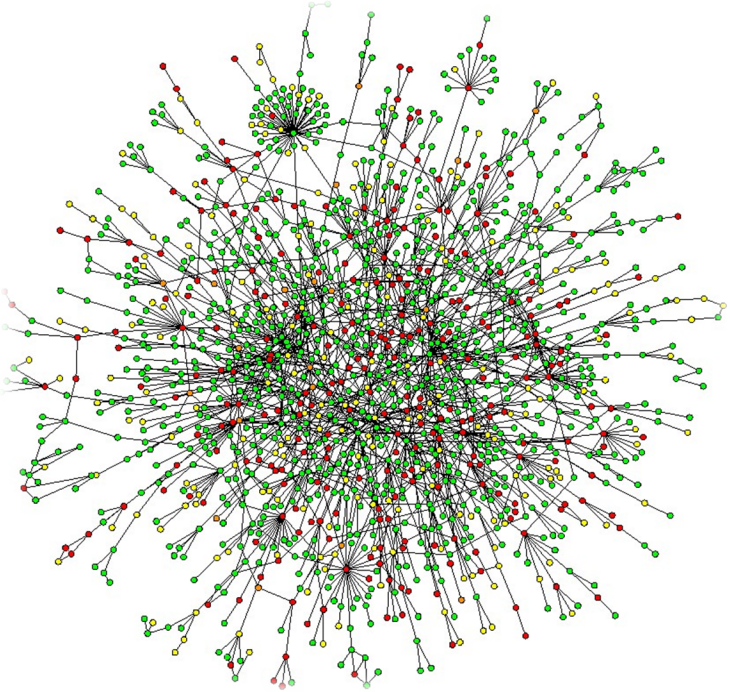
Application: Product Adoption

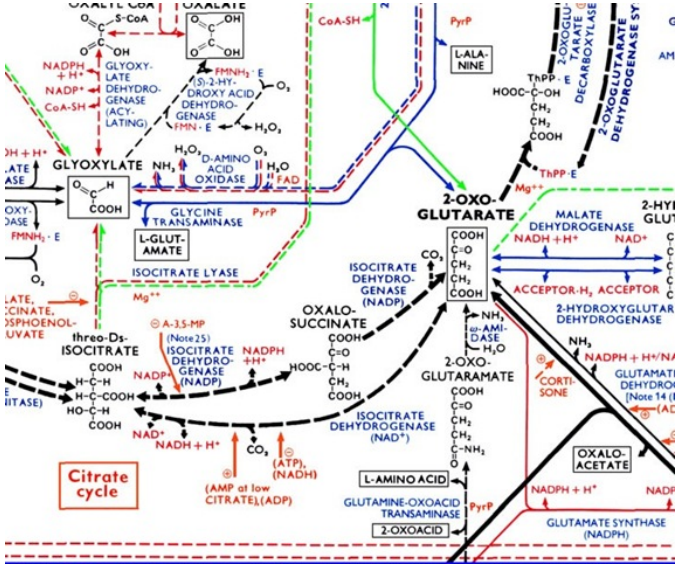


Invitation cascades: 60-90% of LinkedIn users signed up due to an invitation from another user.

Global Diffusion via Cascading Invitations: Structure, Growth, and Homophily. Anderson et al., WWW '15.

(5) Networks: Biomedicine





Protein-protein interaction (PPI) networks:

Nodes: Proteins

Edges: 'Physical' interactions

Metabolic networks:

Nodes: Metabolites and enzymes

Edges: Chemical reactions