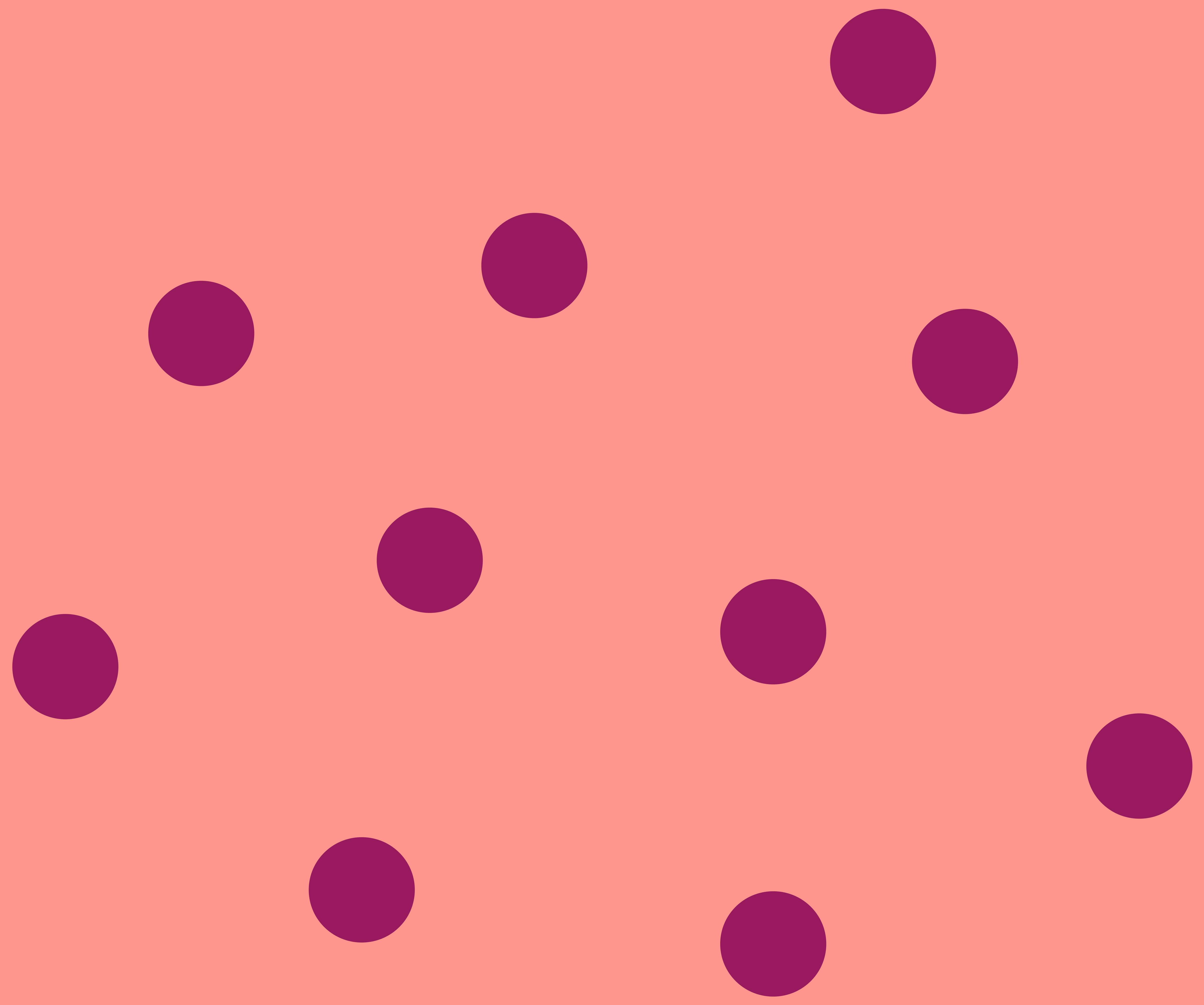
Social and Information Networks

CSCC46H, Fall 2022
Lecture 1

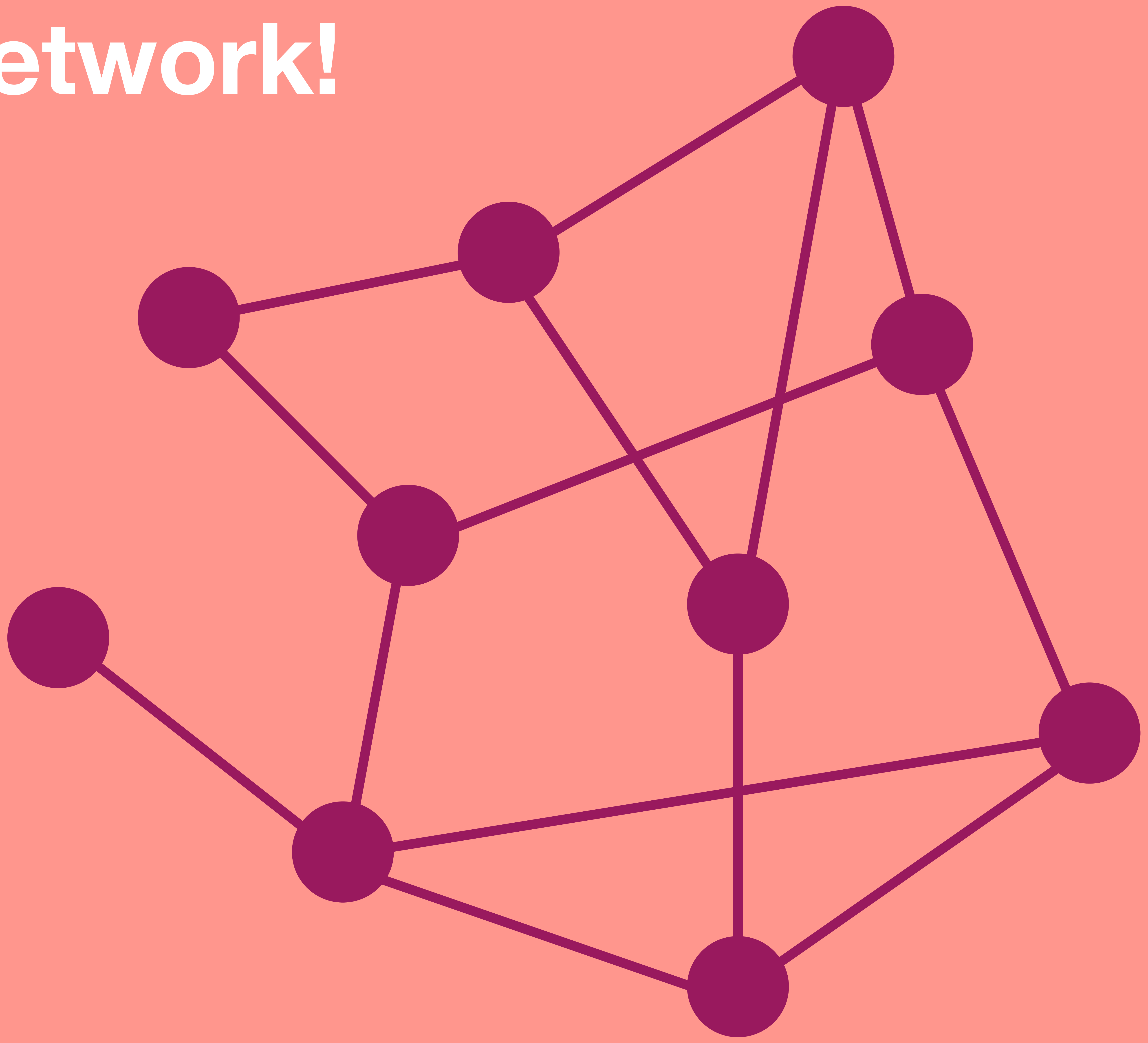
Prof. Ashton Anderson
ashton@cs.toronto.edu

What Are Networks?

Networks are a language for representing, describing, and understanding interconnected systems



A Network!



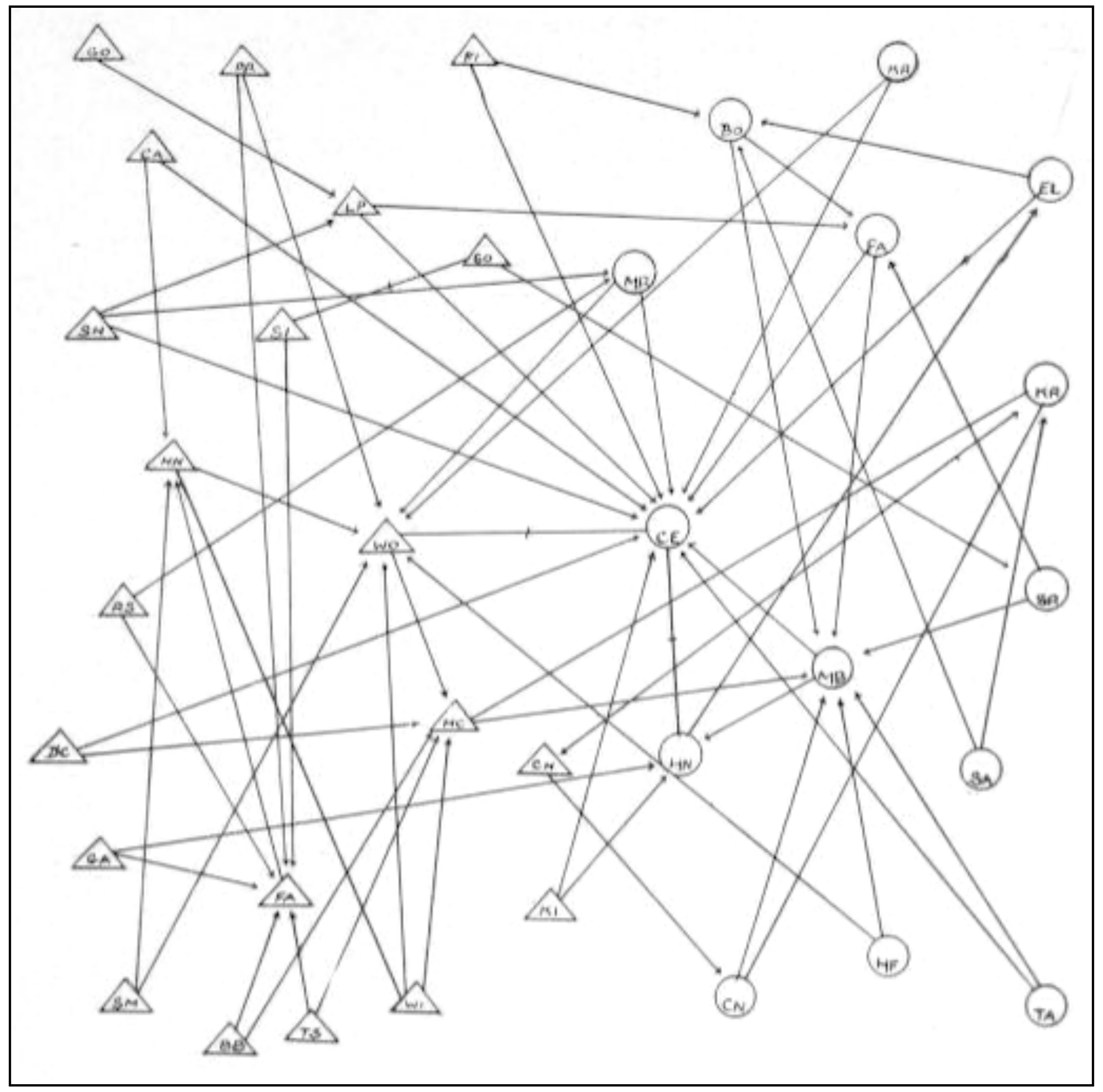
How to understand the behaviours, decisions, beliefs, etc. of millions of people?

Social network

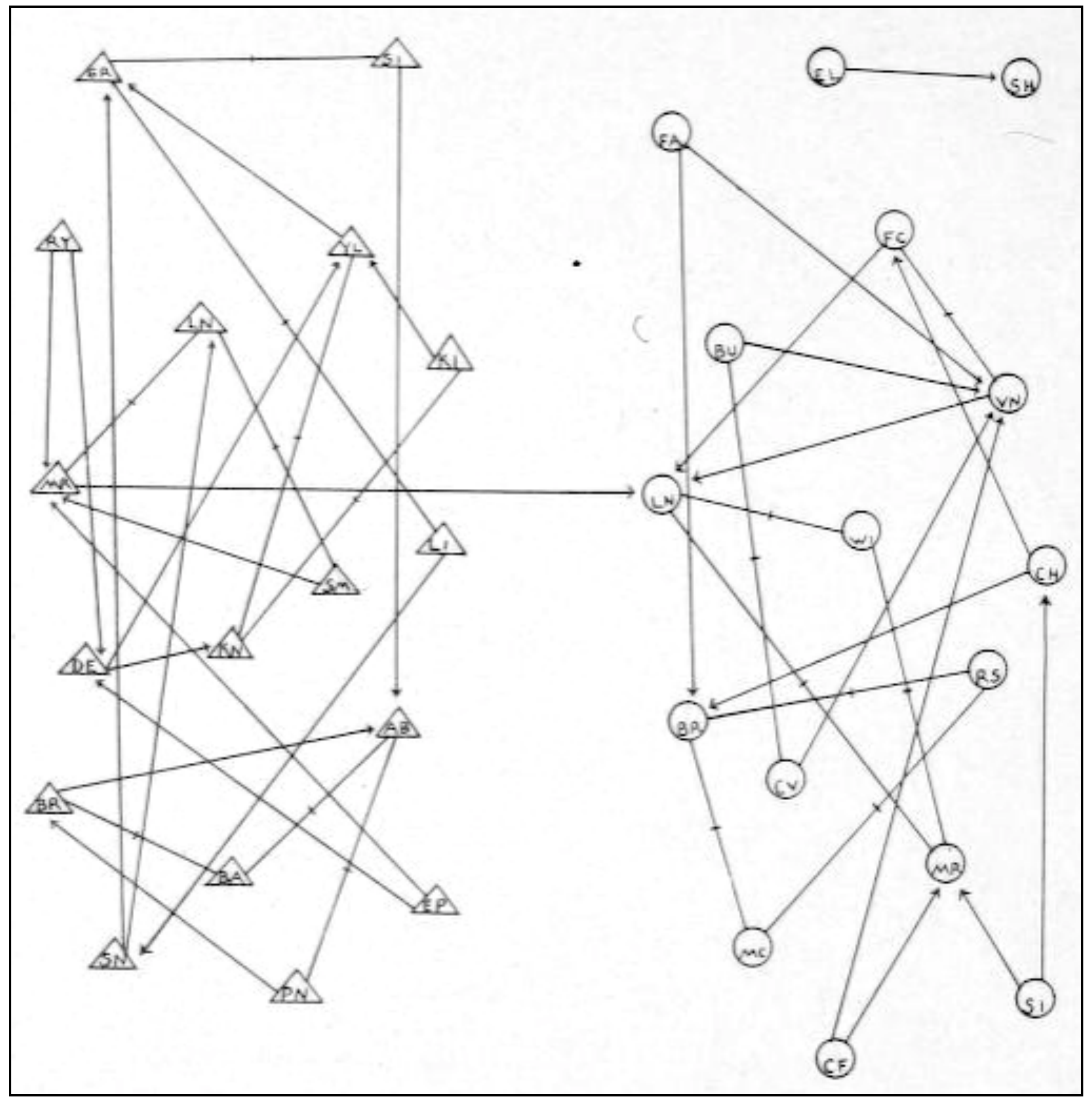
Educational settings: people learn and interact with each other in complex ways

EMOTIONS MAPPED BY NEW GEOGRAPHY

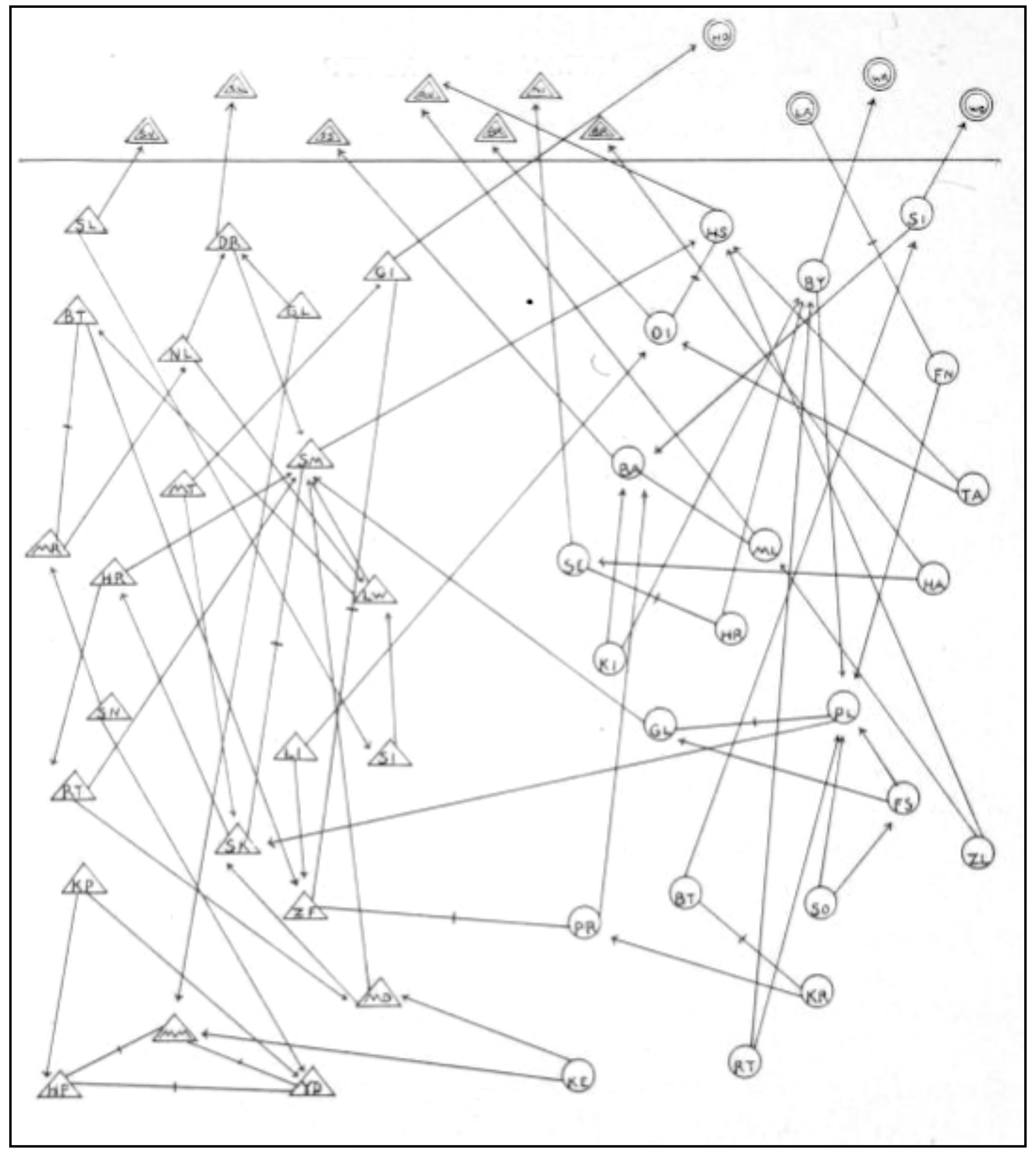
Charts Seek to Portray the Psychological Currents of Human Relationships.



1st grade



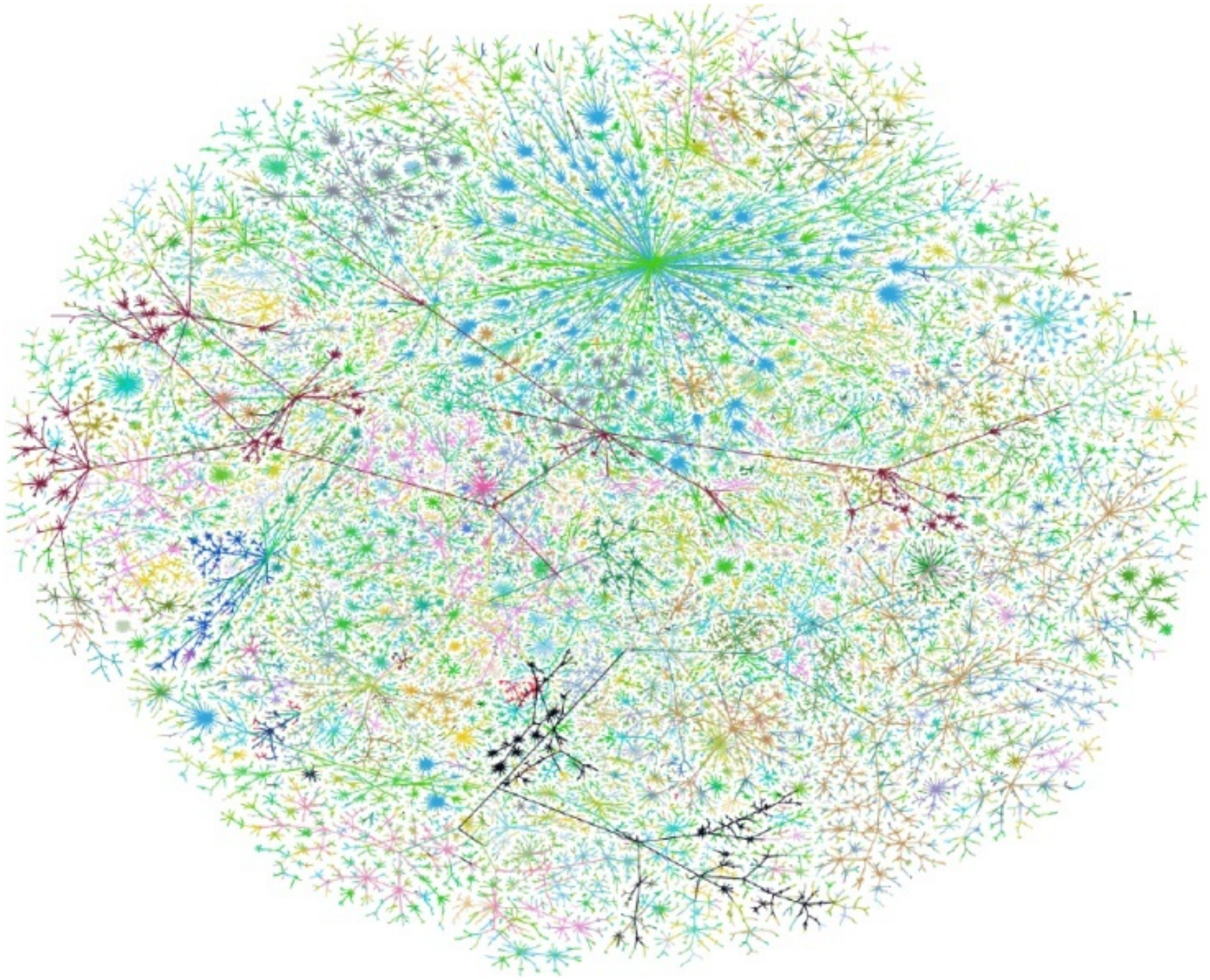
4th grade



8th grade

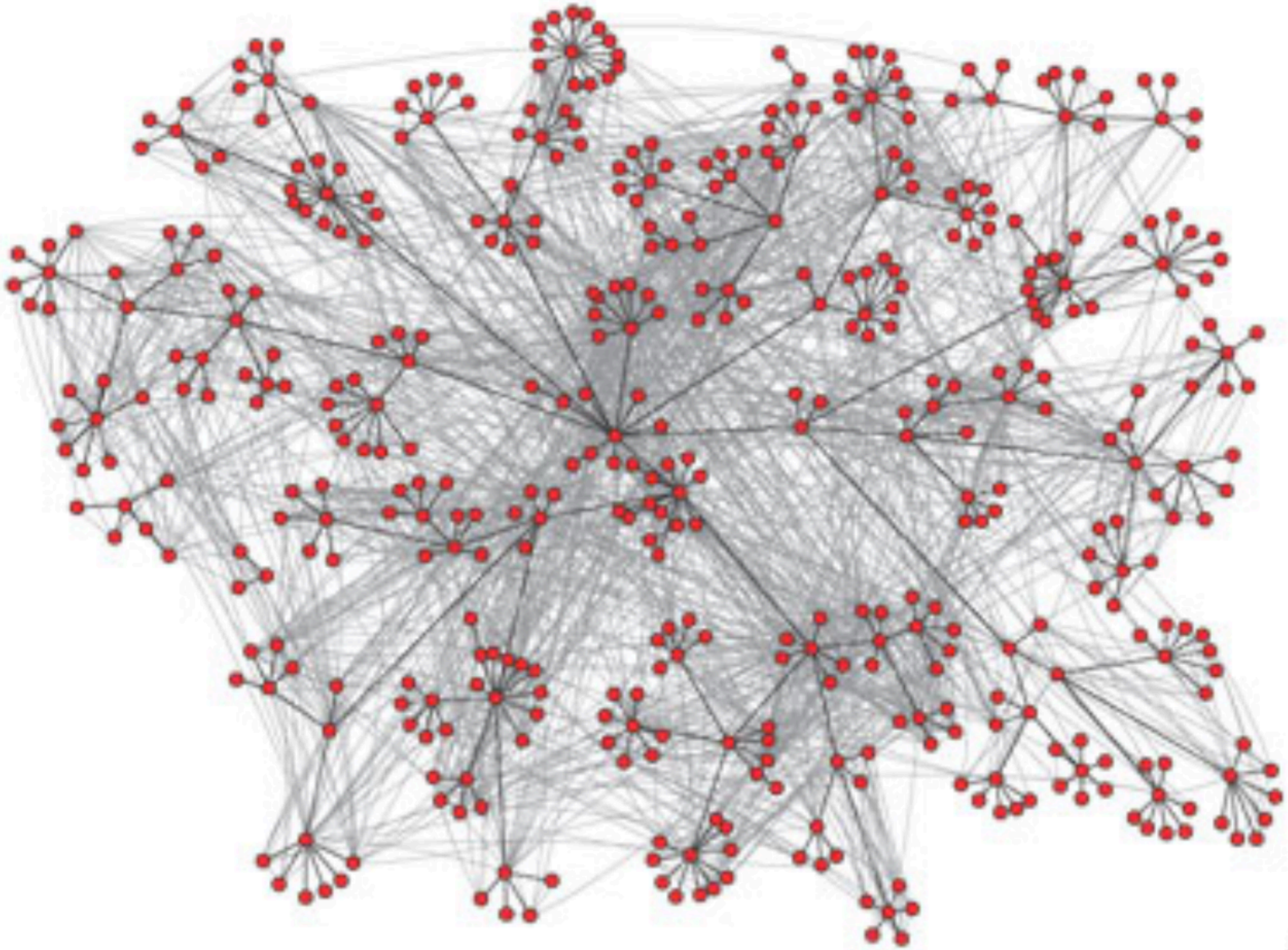
Moreno’s sociograms, 1934

Everything on the Internet is passed through autonomous systems (routers, etc.)

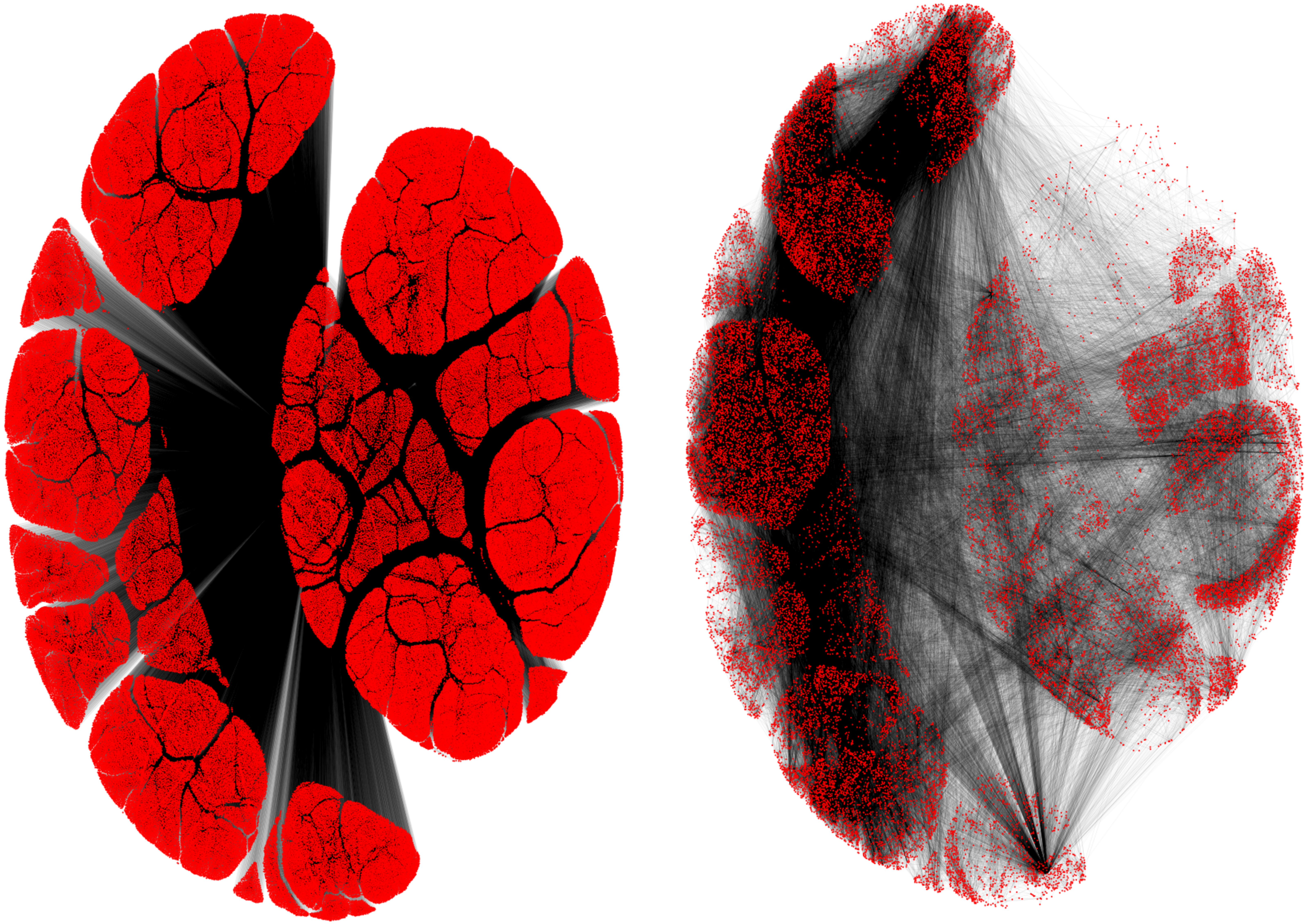


Graph of the Internet (Autonomous Systems)

In companies, people communicate and work together in large hierarchies and structures

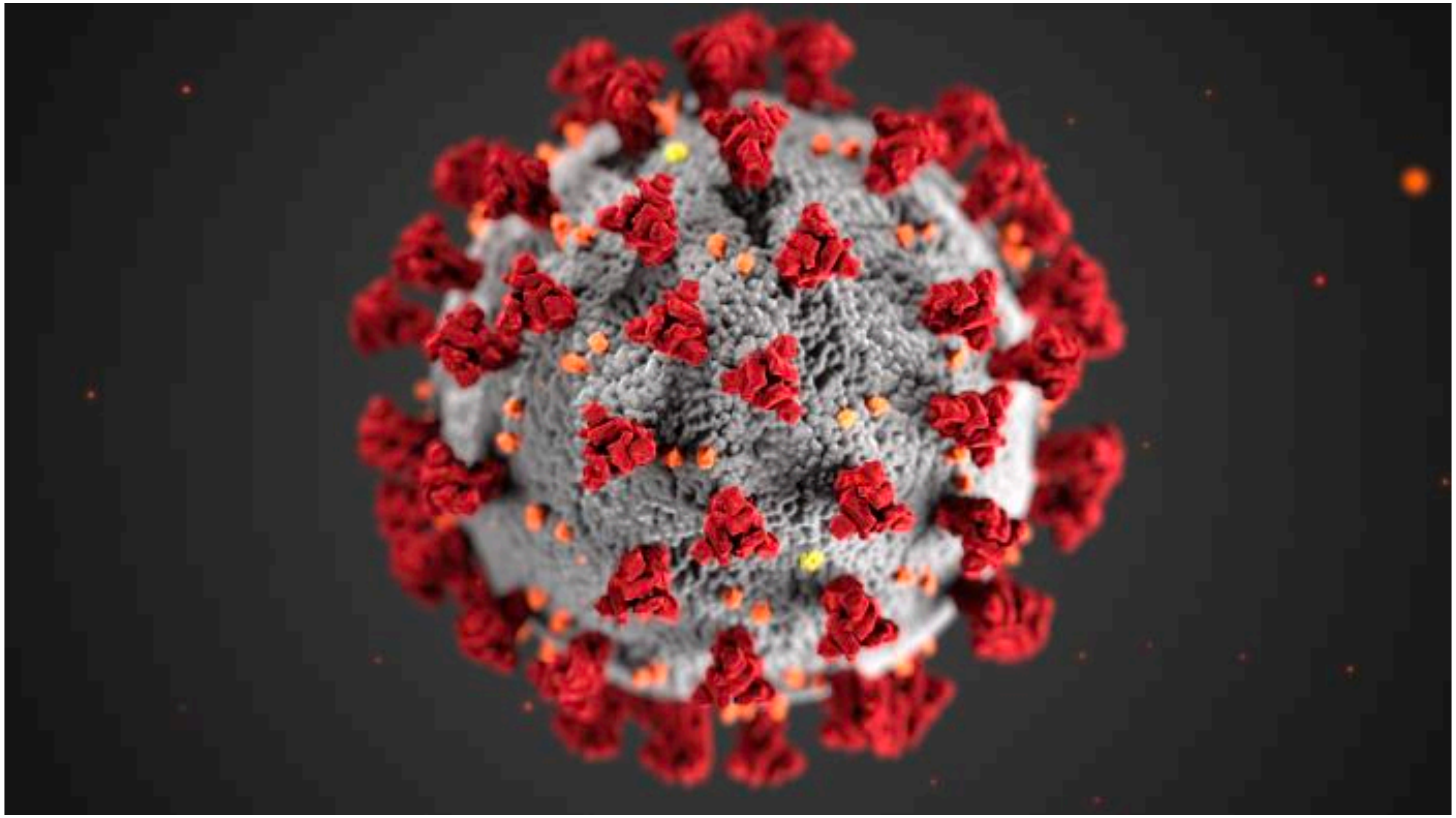


Email communication network
(HP Research, 436 employees)

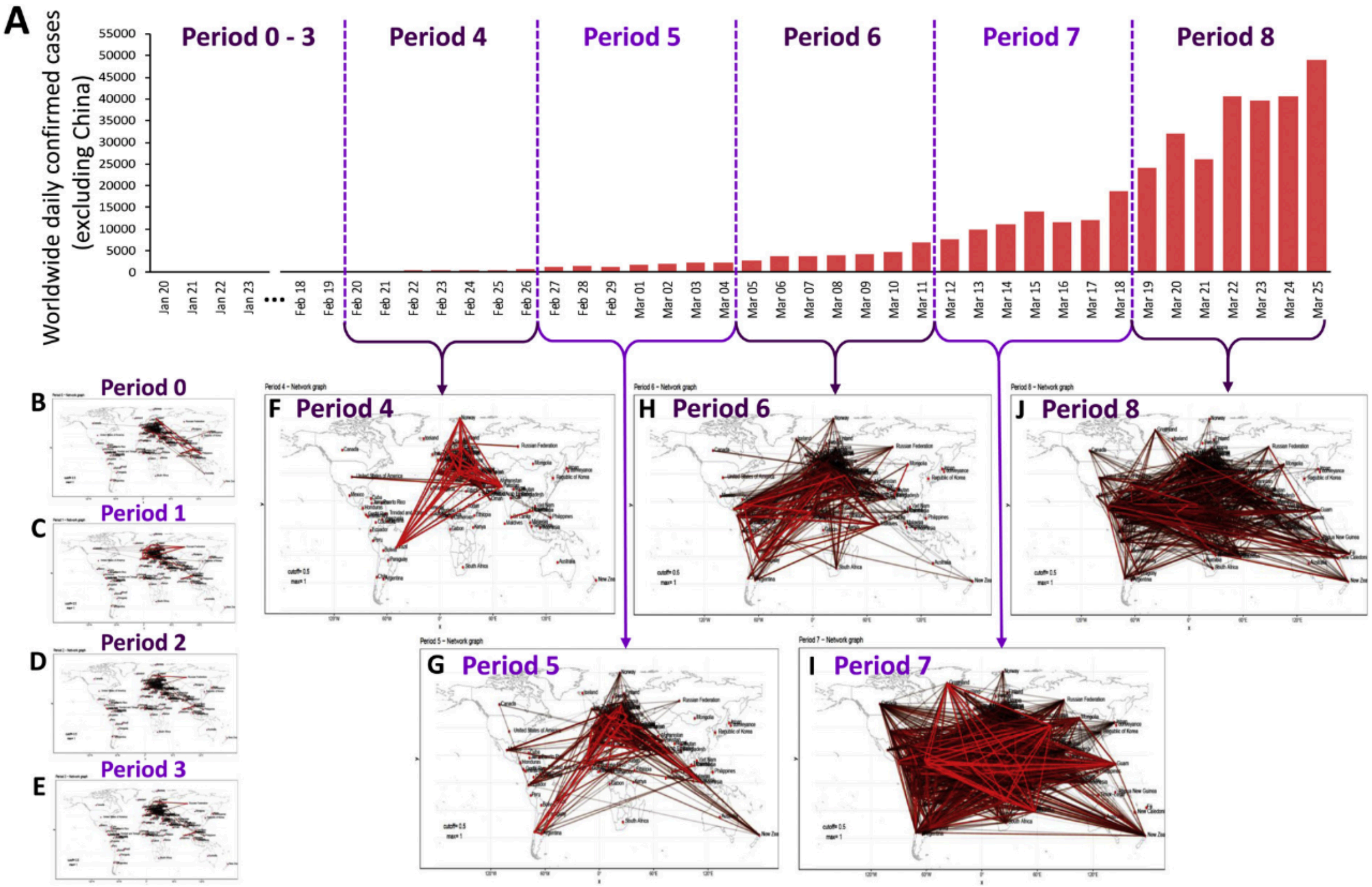


Org chart (left) and Email communication network (right)

Microsoft, 200,000 employees [Jake Hofman, 2018]

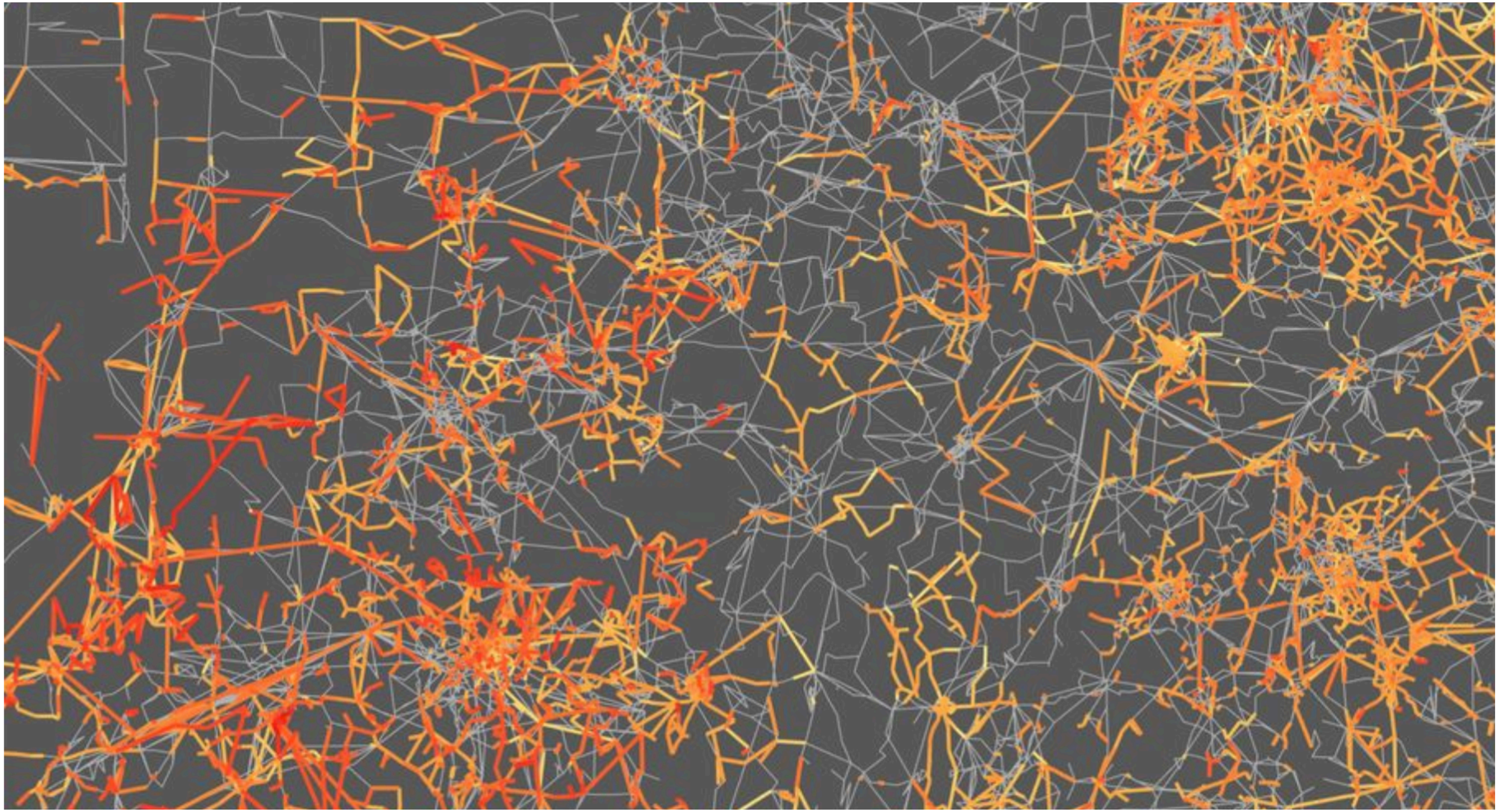


Many diseases are transmitted socially (e.g. COVID-19)



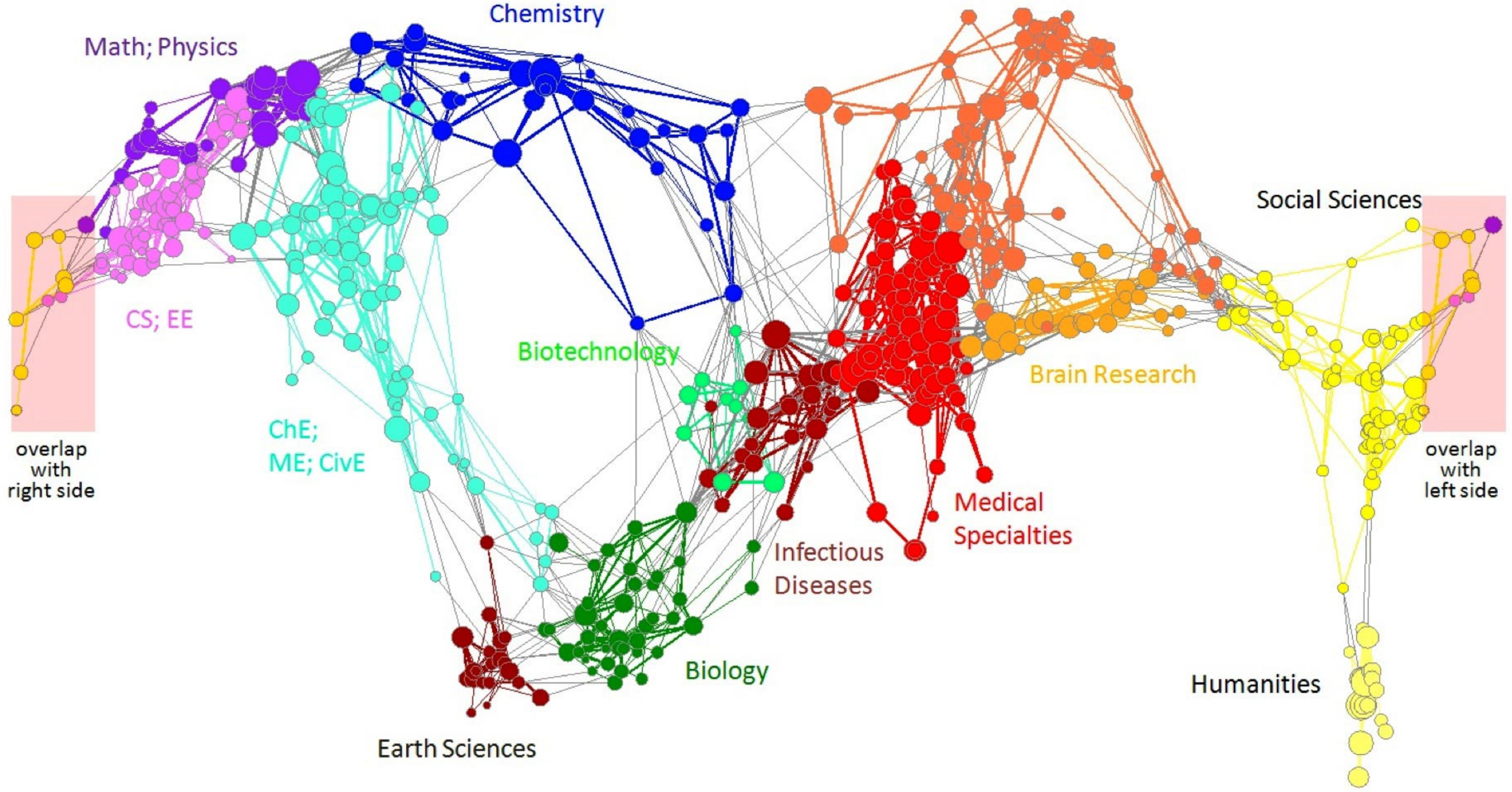
Global spread of COVID-19

Power is transported everywhere with interconnected stations and lines



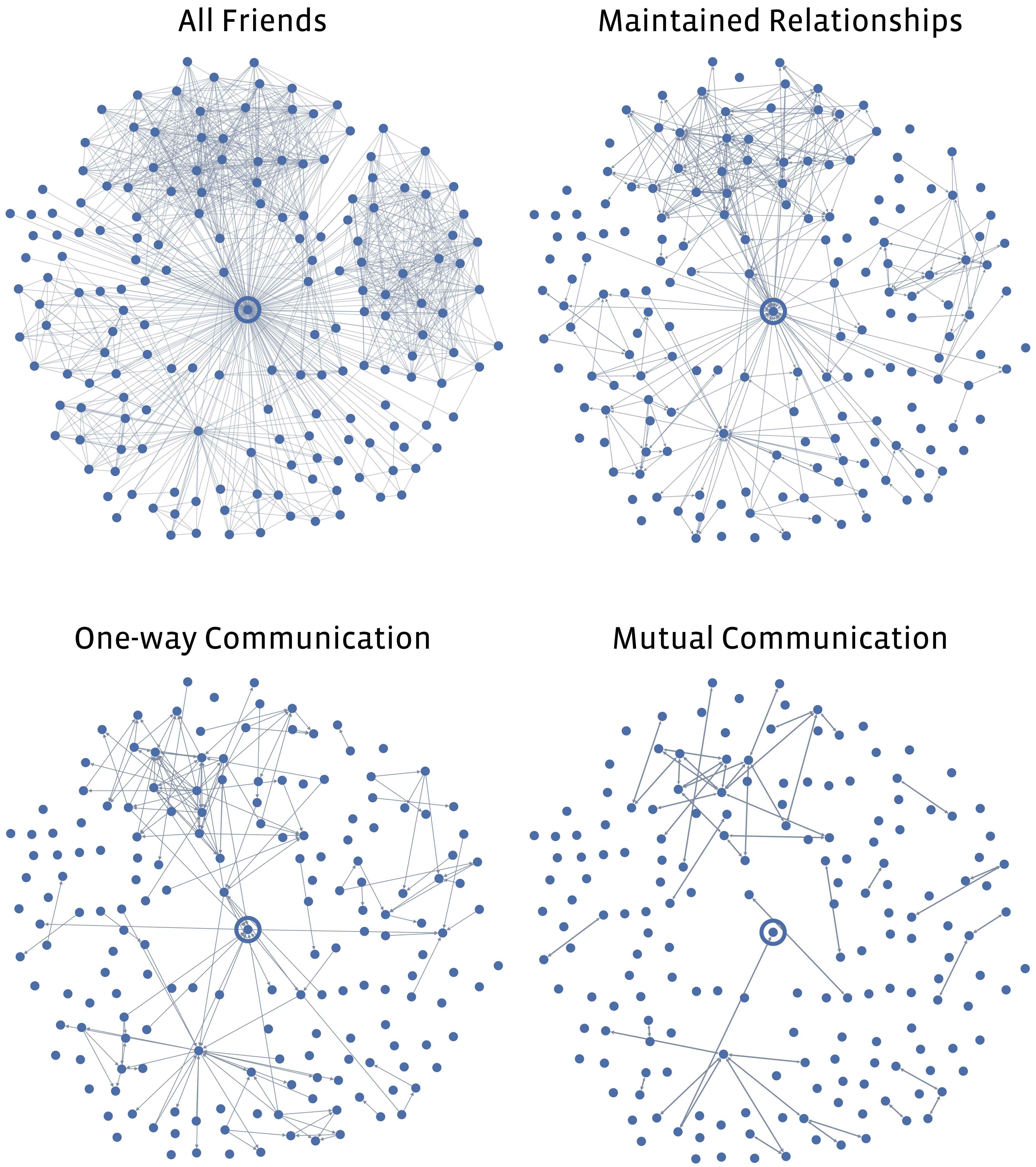
The power grid: a network

Science is a complex system of academics working together and being influenced by each other

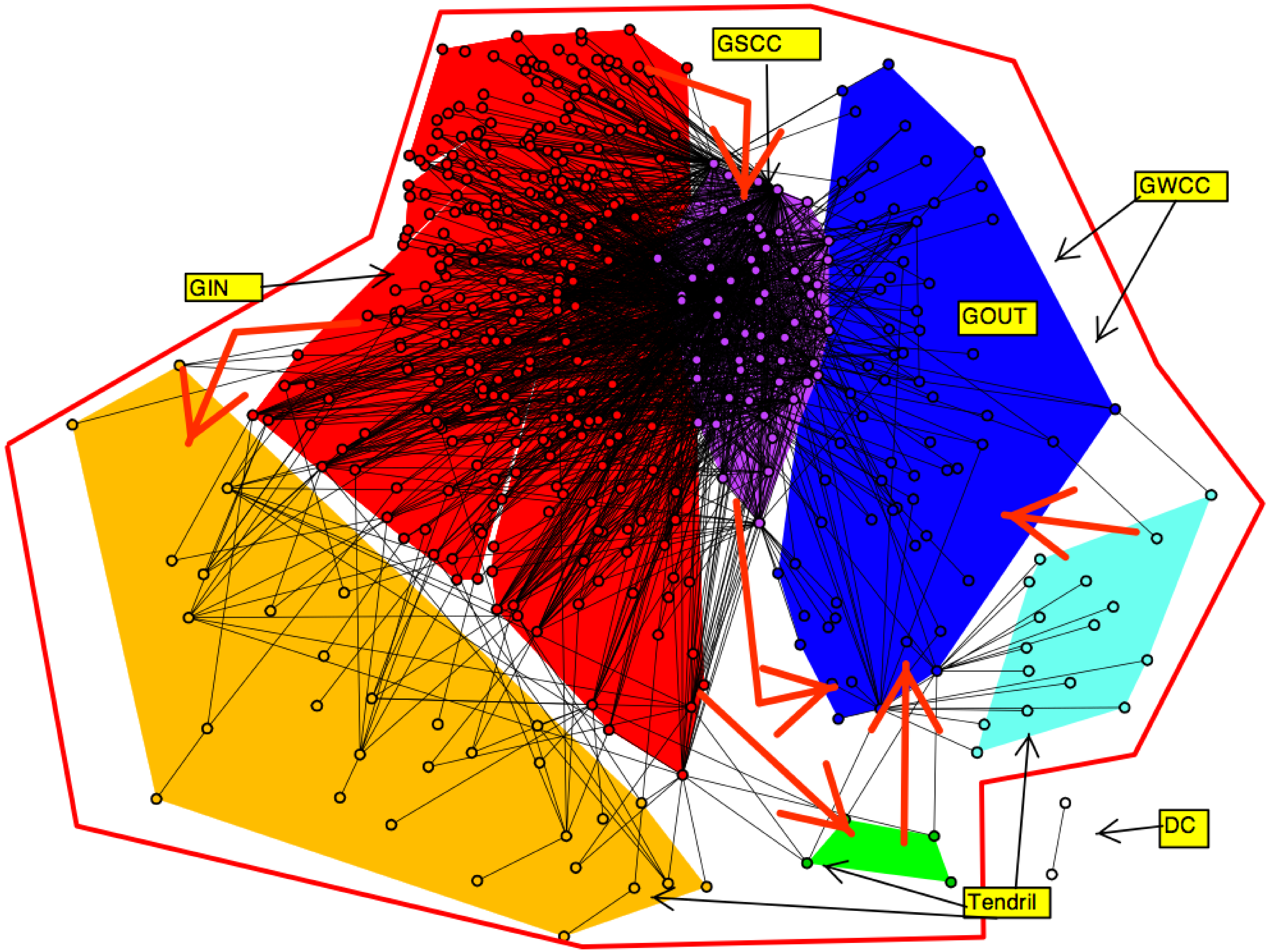


Citation networks and maps of science [Börner et al., 2012]

A single person’s interactions with friends and family are a huge part of their life

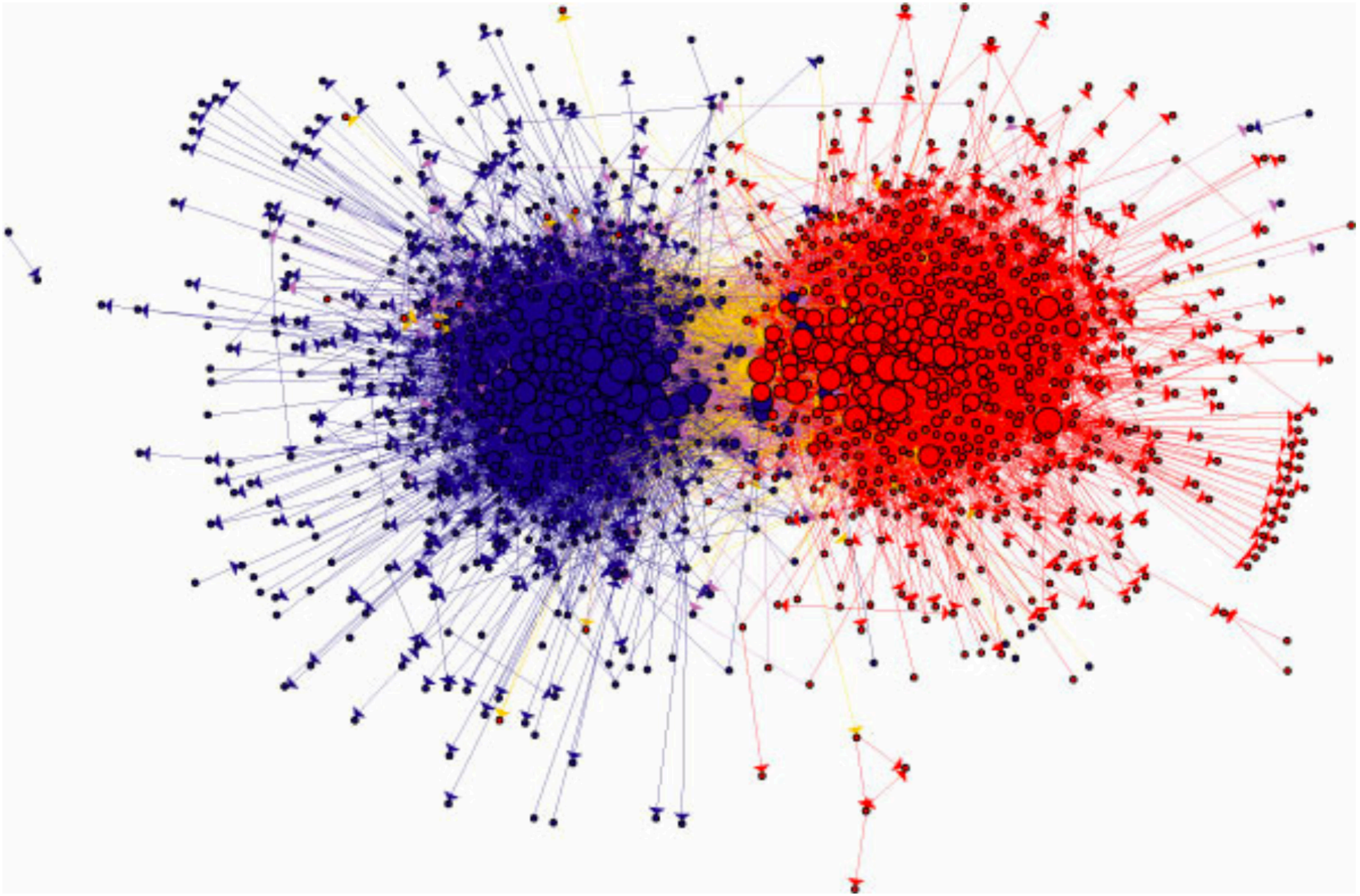


An "ego network": the neighbourhood around a single individual

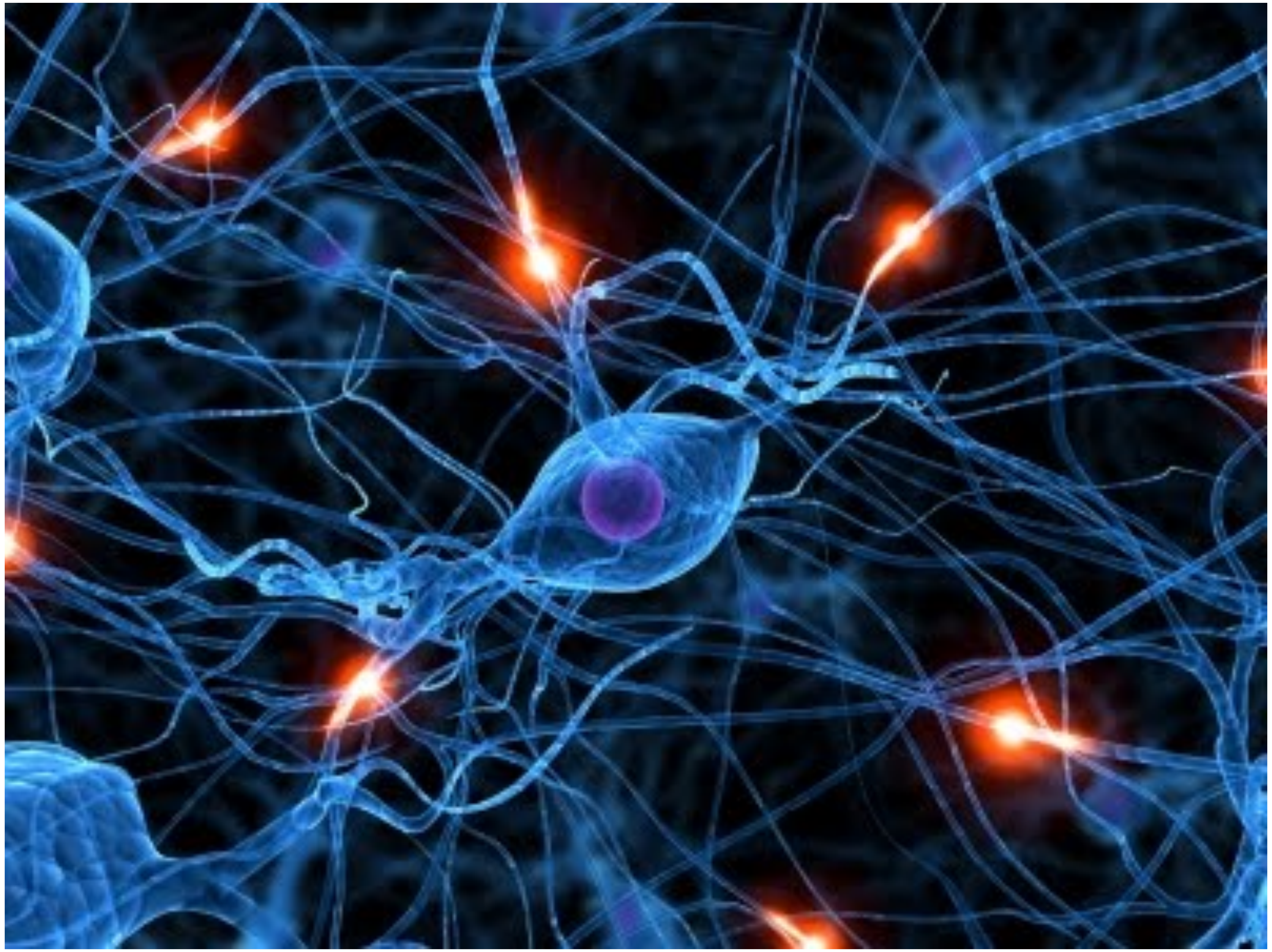


The Economy is a network:
e.g. Federal funds overnight lending

Transportation network (US only)

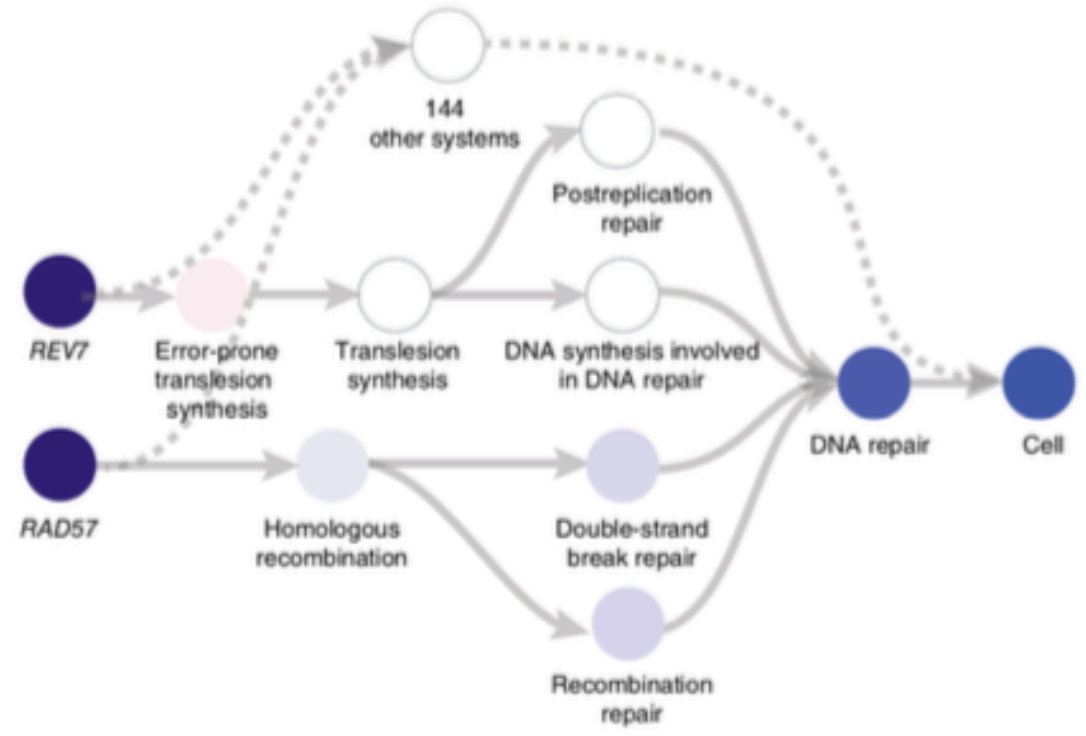


Political blogs prior to 2004 US Presidential election

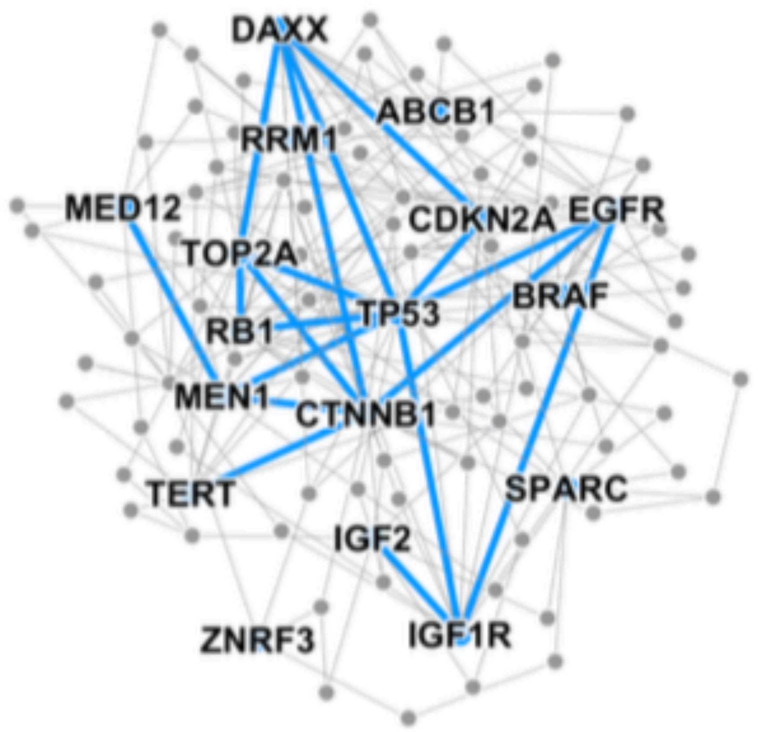


The human brain has between 10-100 billion neurons connected to each other in complex ways

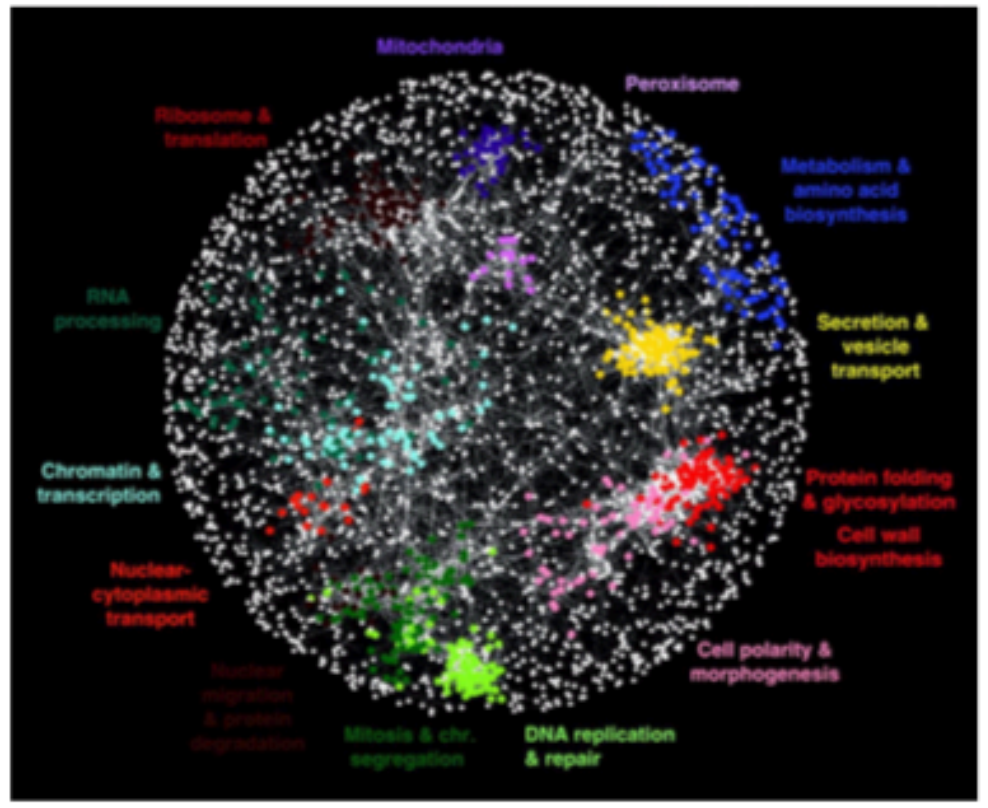
Patient networks



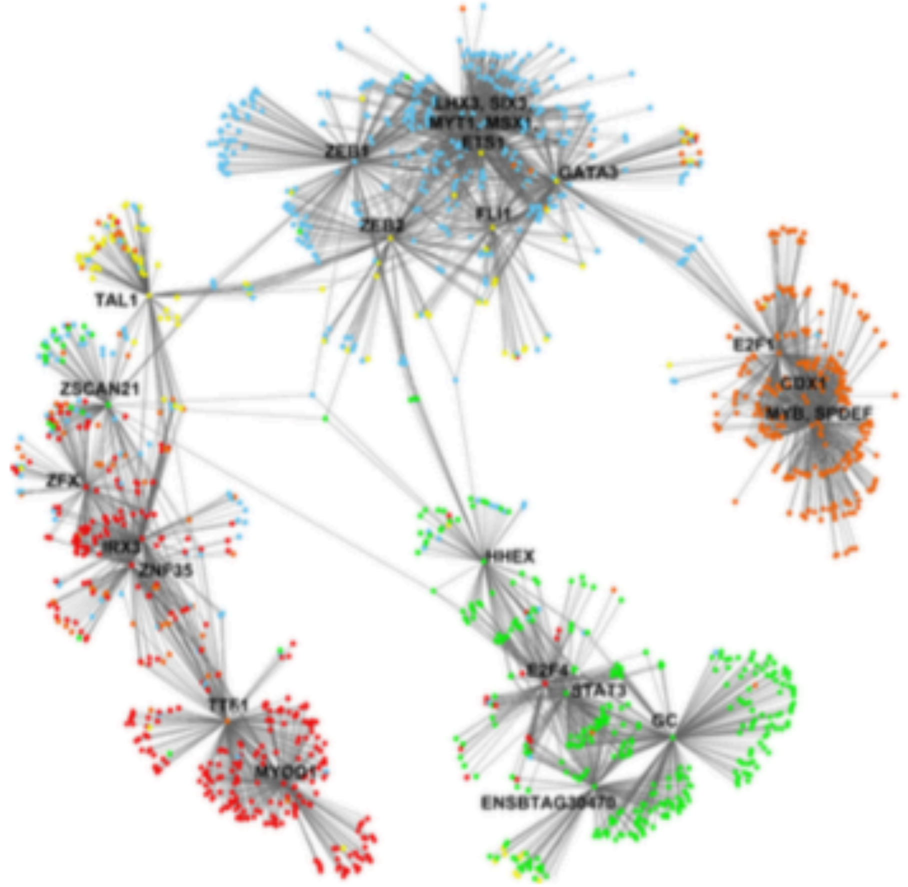
Hierarchies of cell systems



Disease pathways



Genetic interaction networks



Gene co-expression networks

Cell-cell similarity networks

Many, many more examples

But why should *you* care about networks?

Why study networks?

Networks are a **universal language for describing complex data**

- Networks from science, nature, and technology are more similar than you might expect

Shared vocabulary between fields

- CS, finance, tech, social sciences, physics, economics, statistics, biology

Data availability (and computational challenges)

- Web/mobile, bio, health, medical

Impact!

- Social networking, social media, drug design

Why study networks?

Complex systems are all around us

- **Society** is a collection of seven billion people
- **Communications systems** link electronic devices
- **Information and knowledge** is organized and linked
- Interactions between thousands of **genes** regulate life
- Our thoughts and selves are manifested in the connections between billions of neurons in the **brain**
- **Information and diseases** spread from person to person

Why now?

Age and size of networks

Networks: Impact

Google
Market cap: $1.4T

Facebook
Market cap: $500B

Cisco
Market cap: $220B

Networks and Applications

Ways to Analyze Networks

Predict the type of a given node (*node classification*)

Predict whether two nodes are linked (*link prediction*)

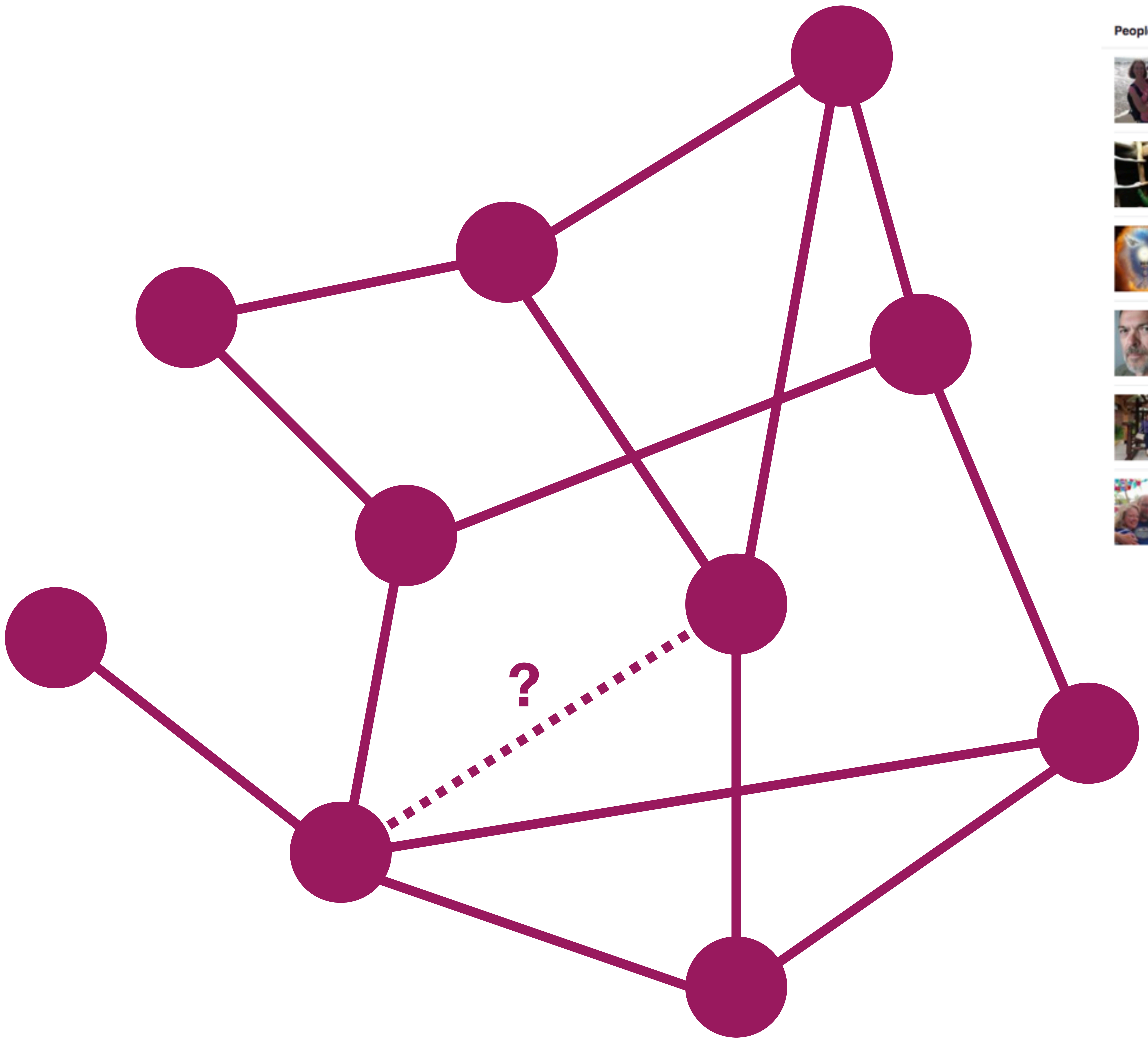
Identify densely linked clusters of nodes (*community detection*)

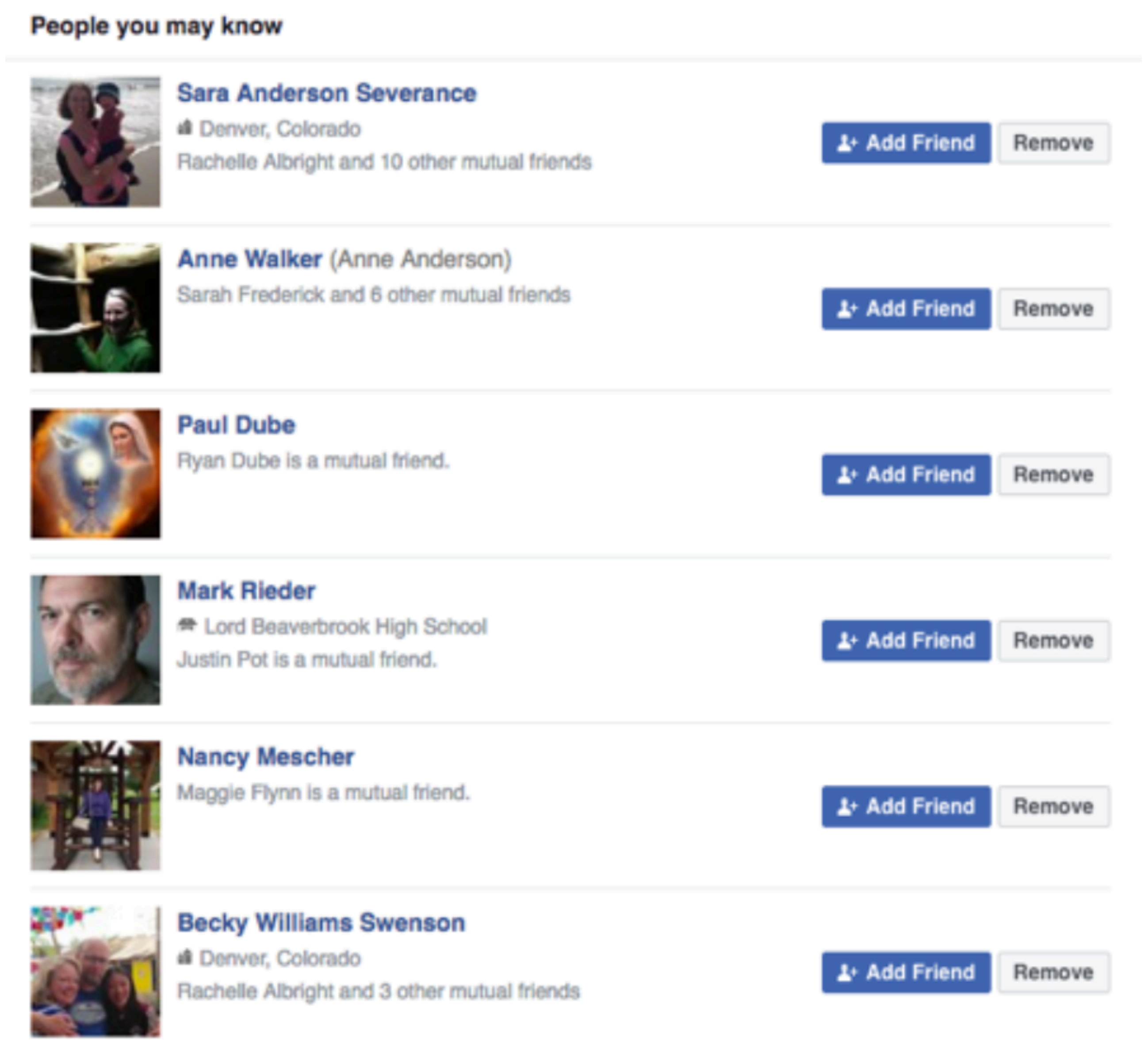
Predict common pathways (*social influence/propagation*)

Measure similarity between nodes/networks (*network similarity*)

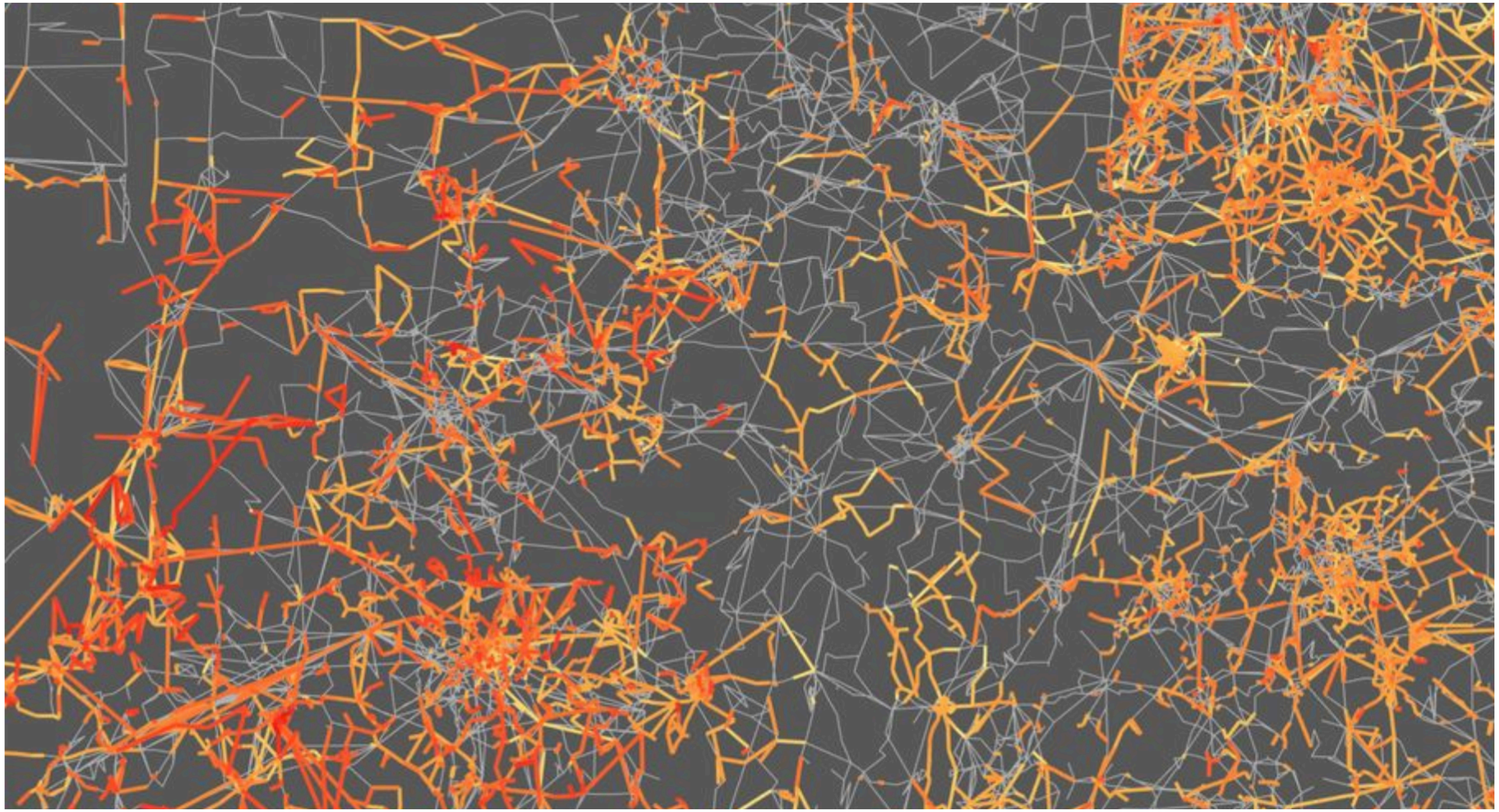
(1) Networks: Social

Application: Friend Prediction



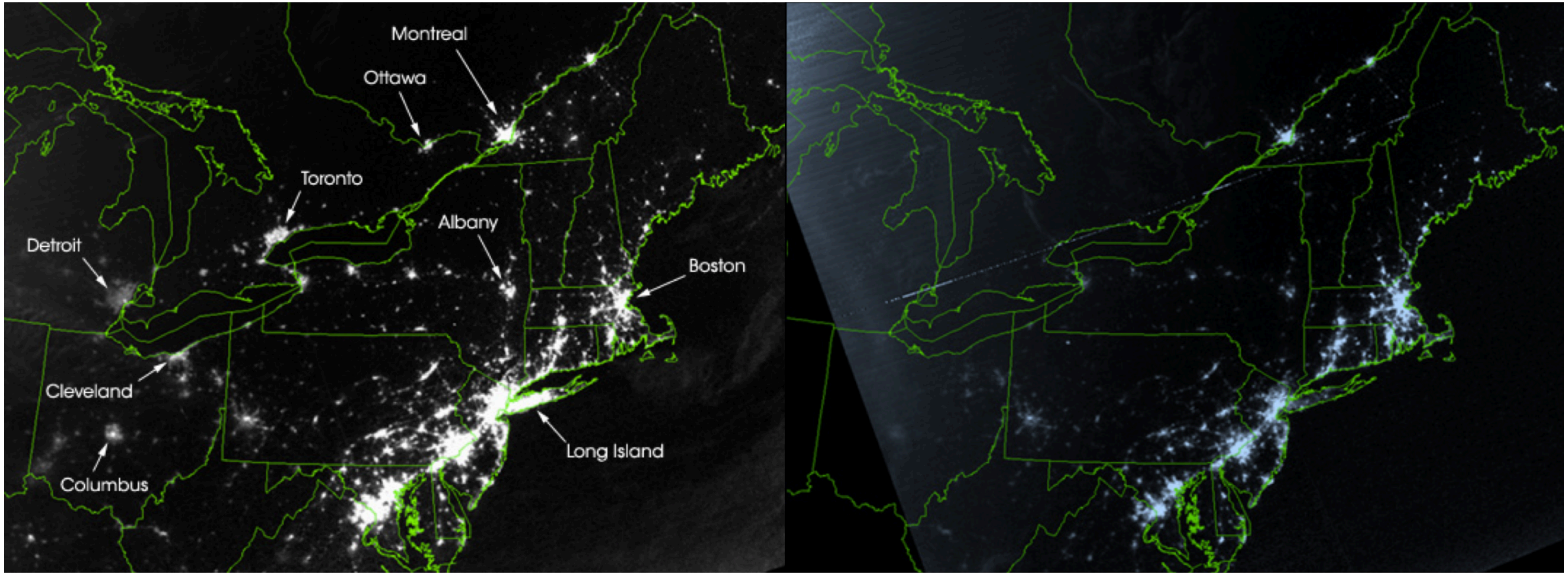


(2) Networks: Infrastructure



Power grids connect you to electricity

(2) Networks: Infrastructure



Aug 14, 2003, 9:29pm
20 hours before

Aug 15, 2003, 9:14pm
4 hours after

August 2003 blackout

(2) Networks: Infrastructure

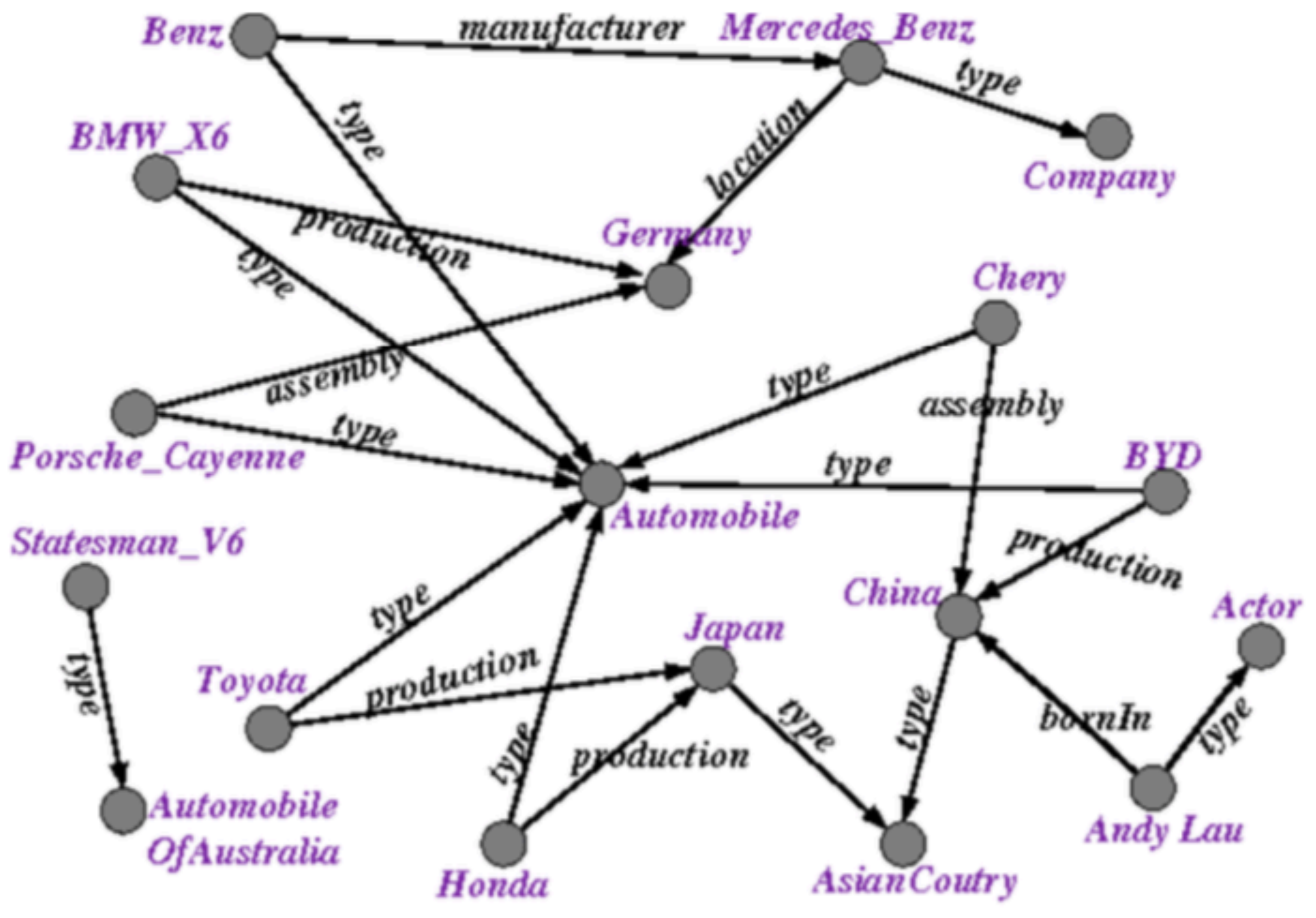
This illustrates important themes of this class:

- We must understand how network structure affects the system

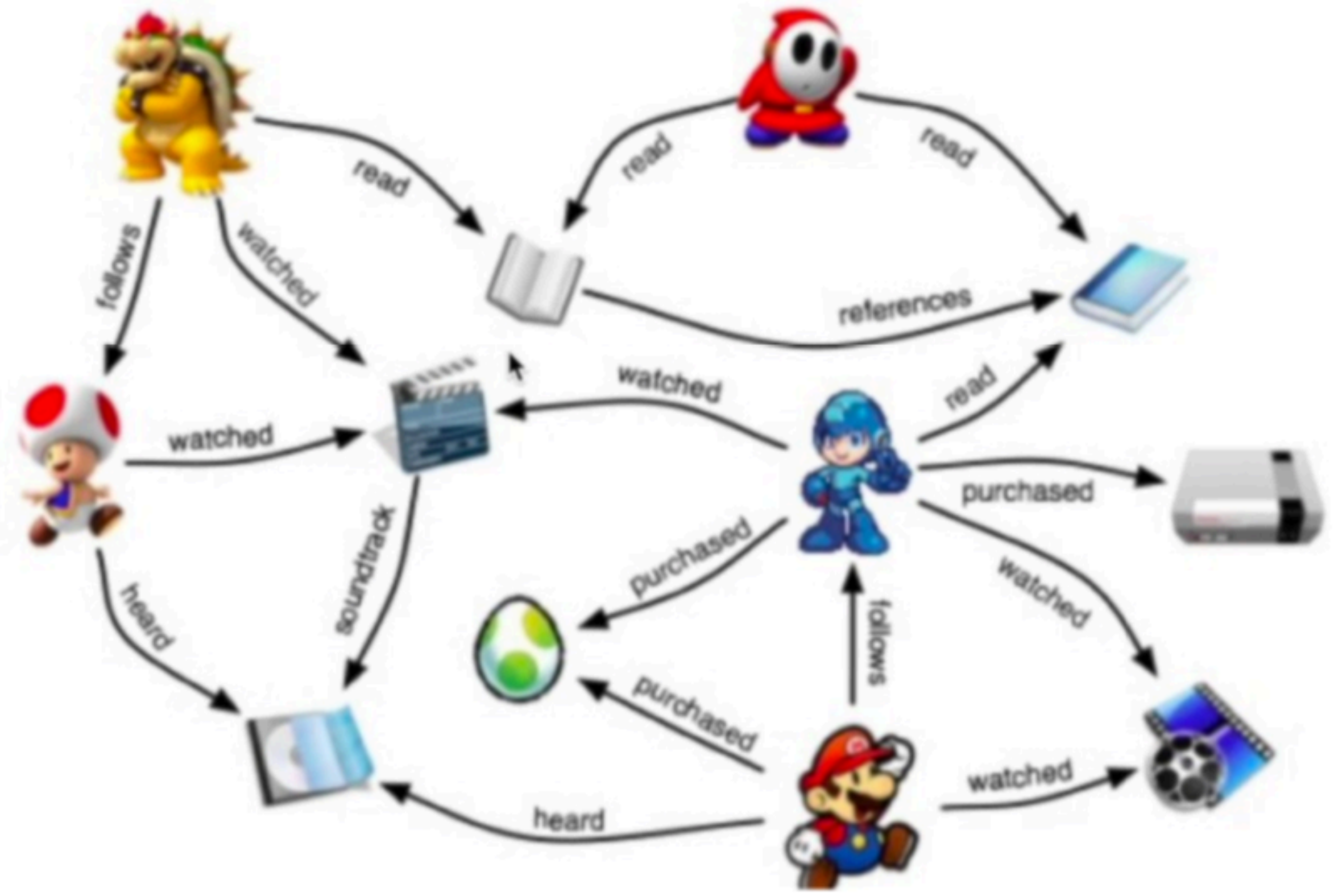
- We will develop quantitative tools to assess the interplay between network structure and the dynamic processes that happen on networks

- We will learn that in reality failures follow reproducible laws, and can be quantified, and to some extent predicted, using the language of network analysis

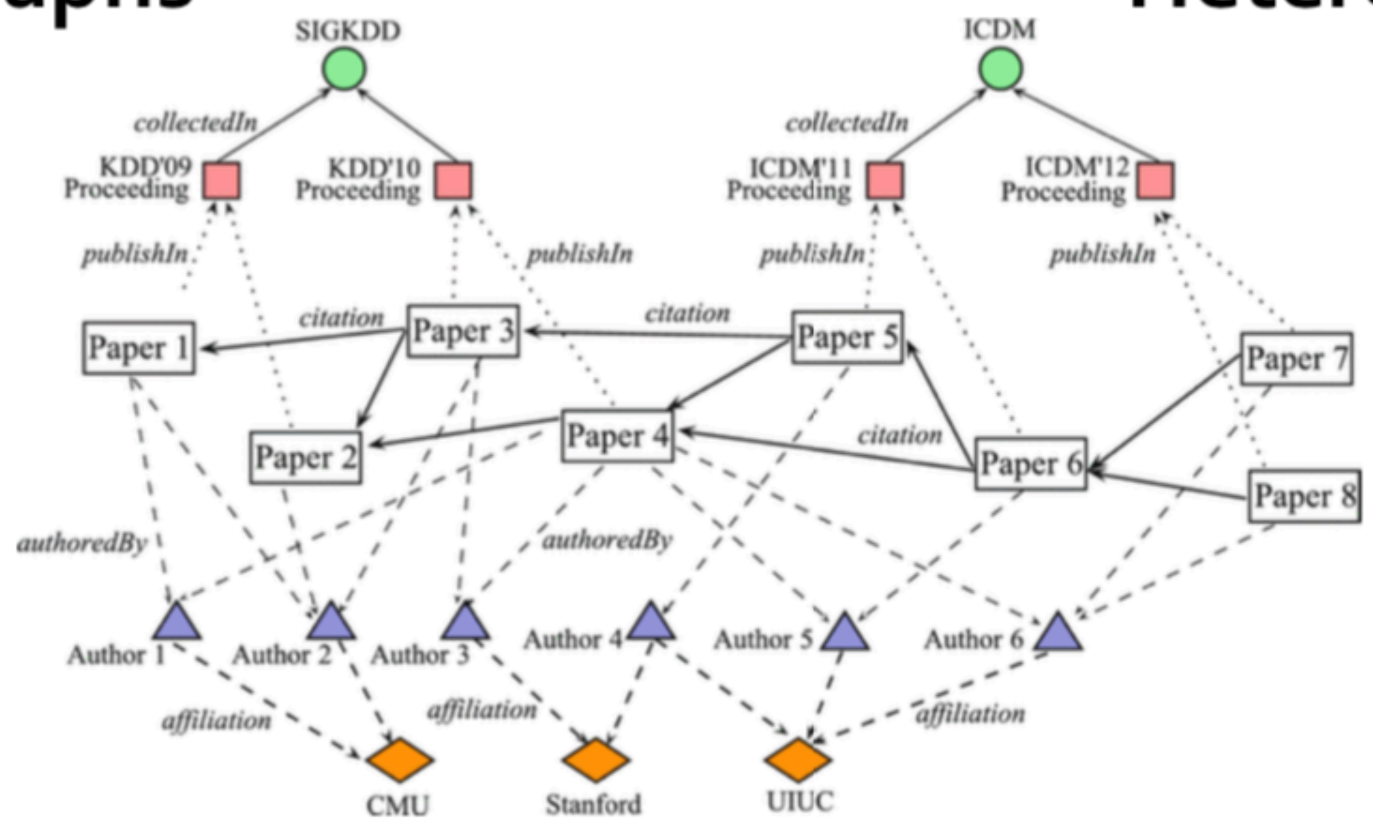
(3) Networks: Information and Knowledge



Knowledge Graphs

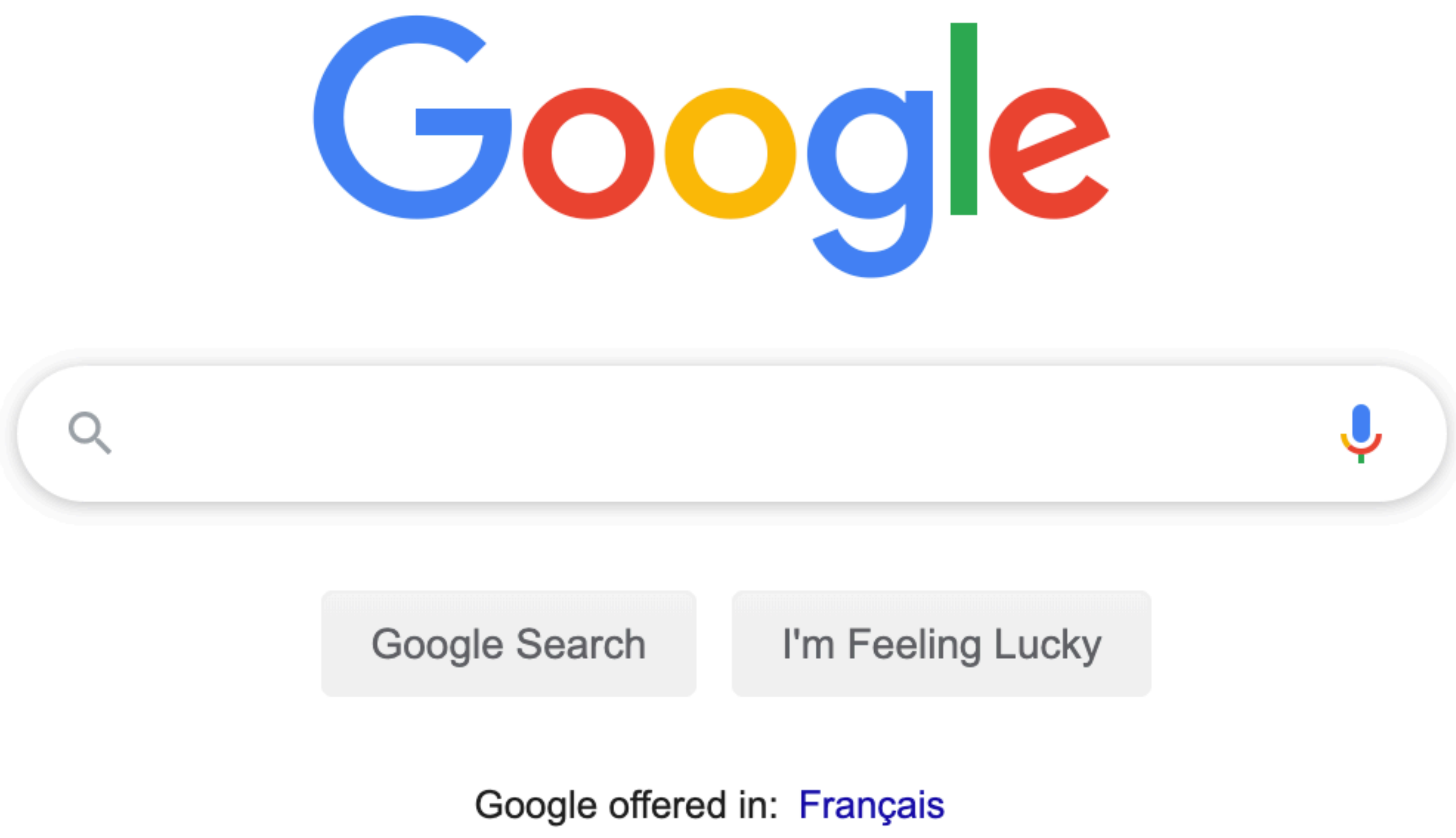


Heterogeneous Graphs



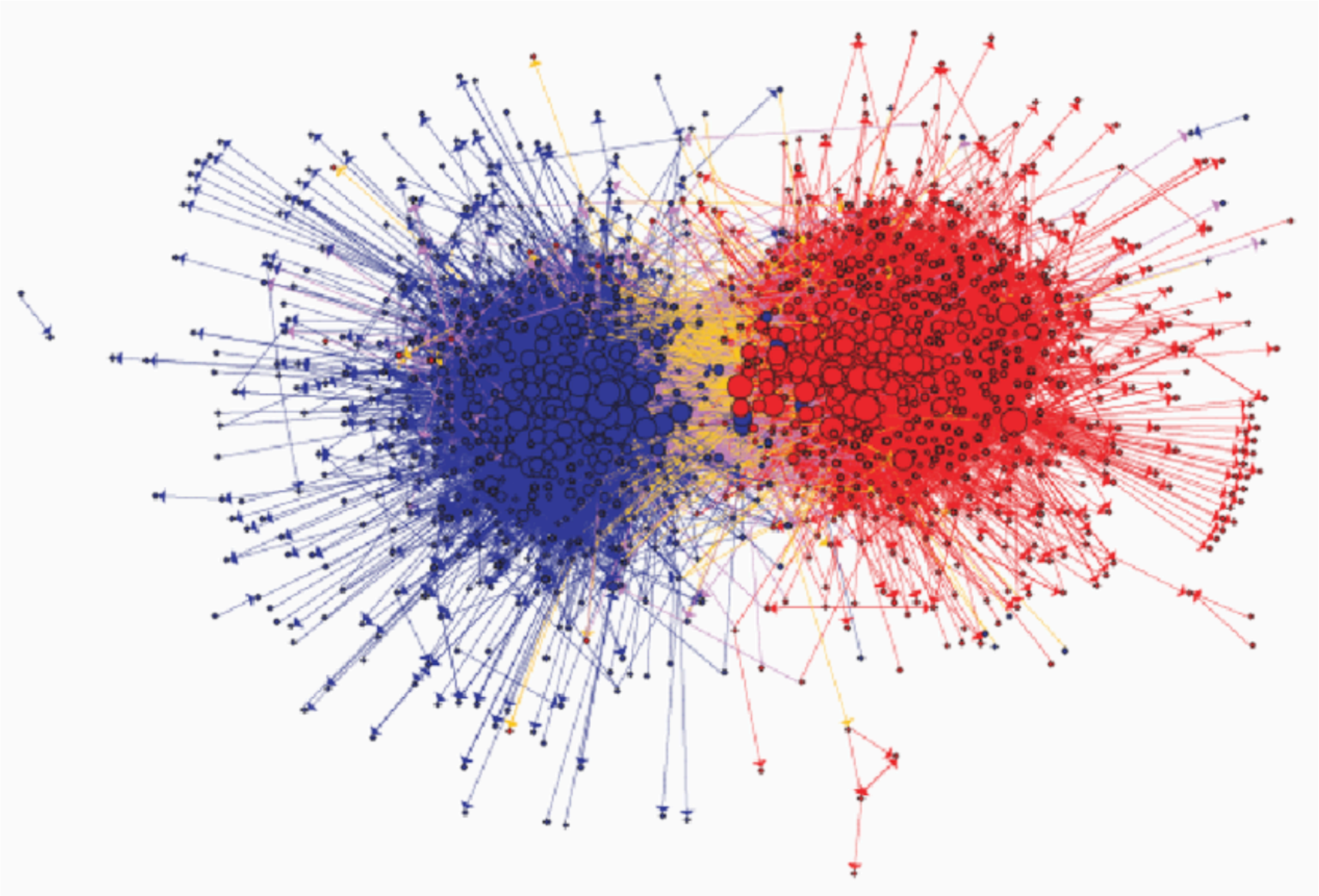
Multimodal Graphs

Application: Web Search



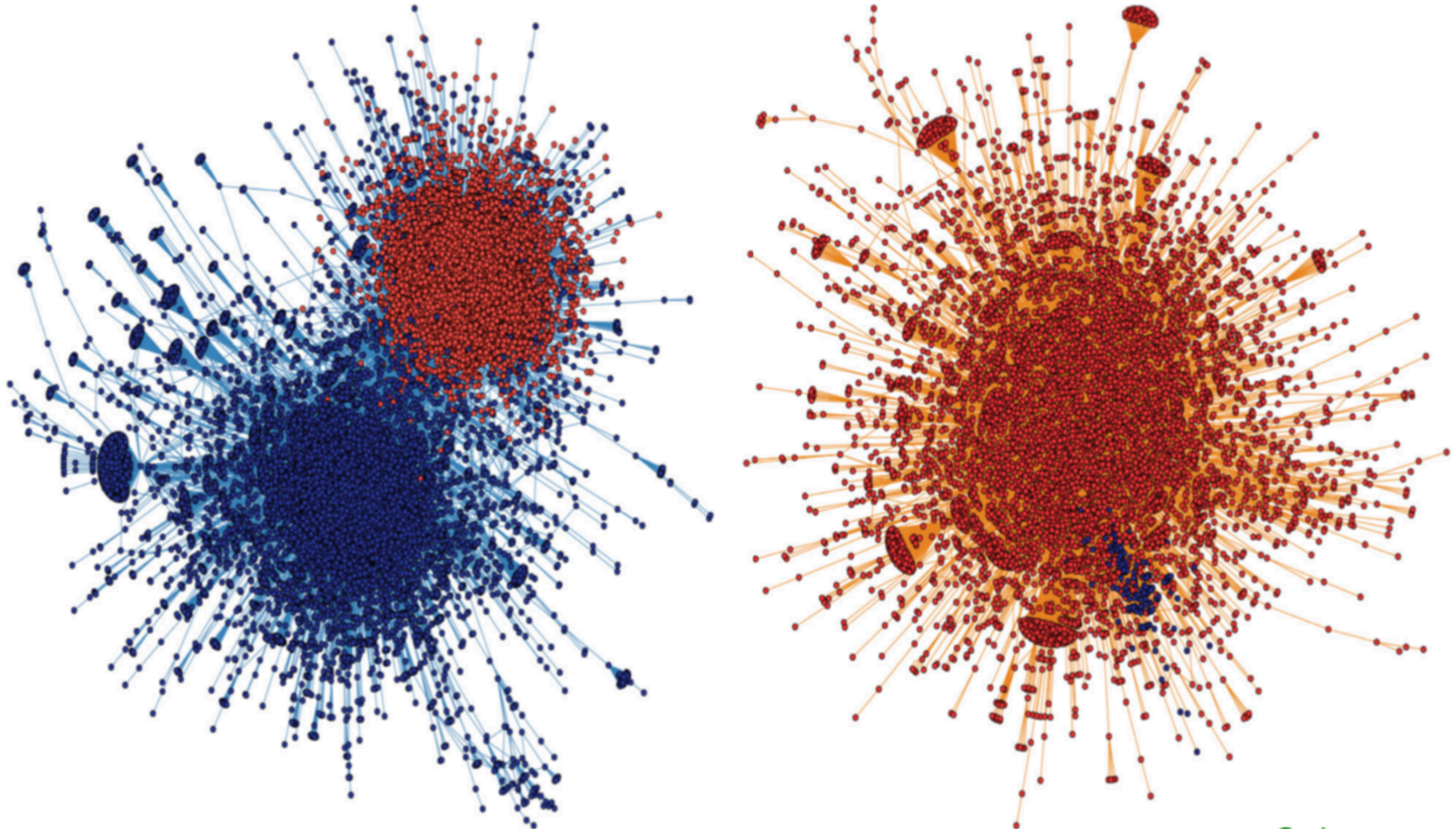
How do you go from a tiny text string to the 10 most relevant sites out of billions of pages?

(4) Networks: Online Media



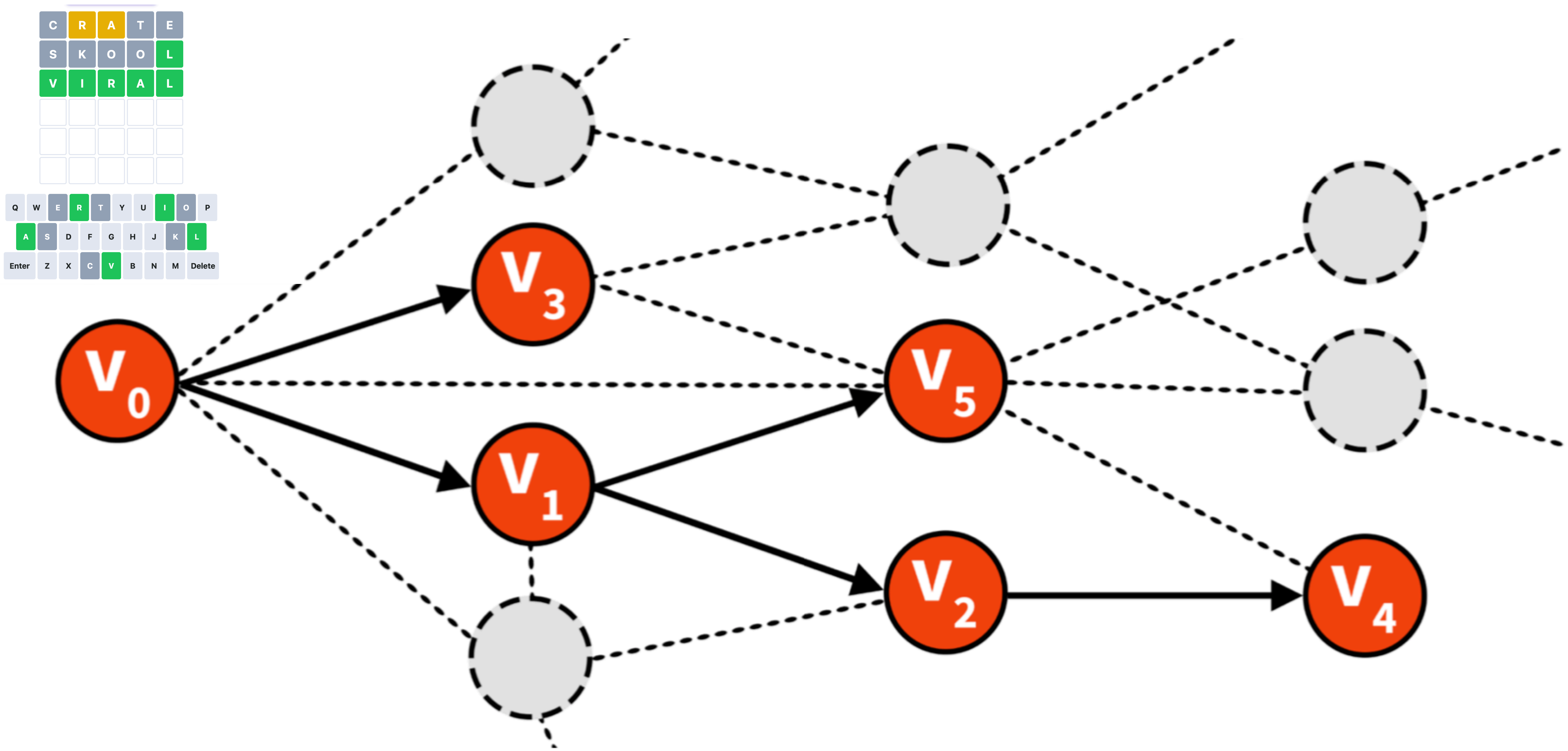
Connections between political blogs

Application: Polarization on Twitter



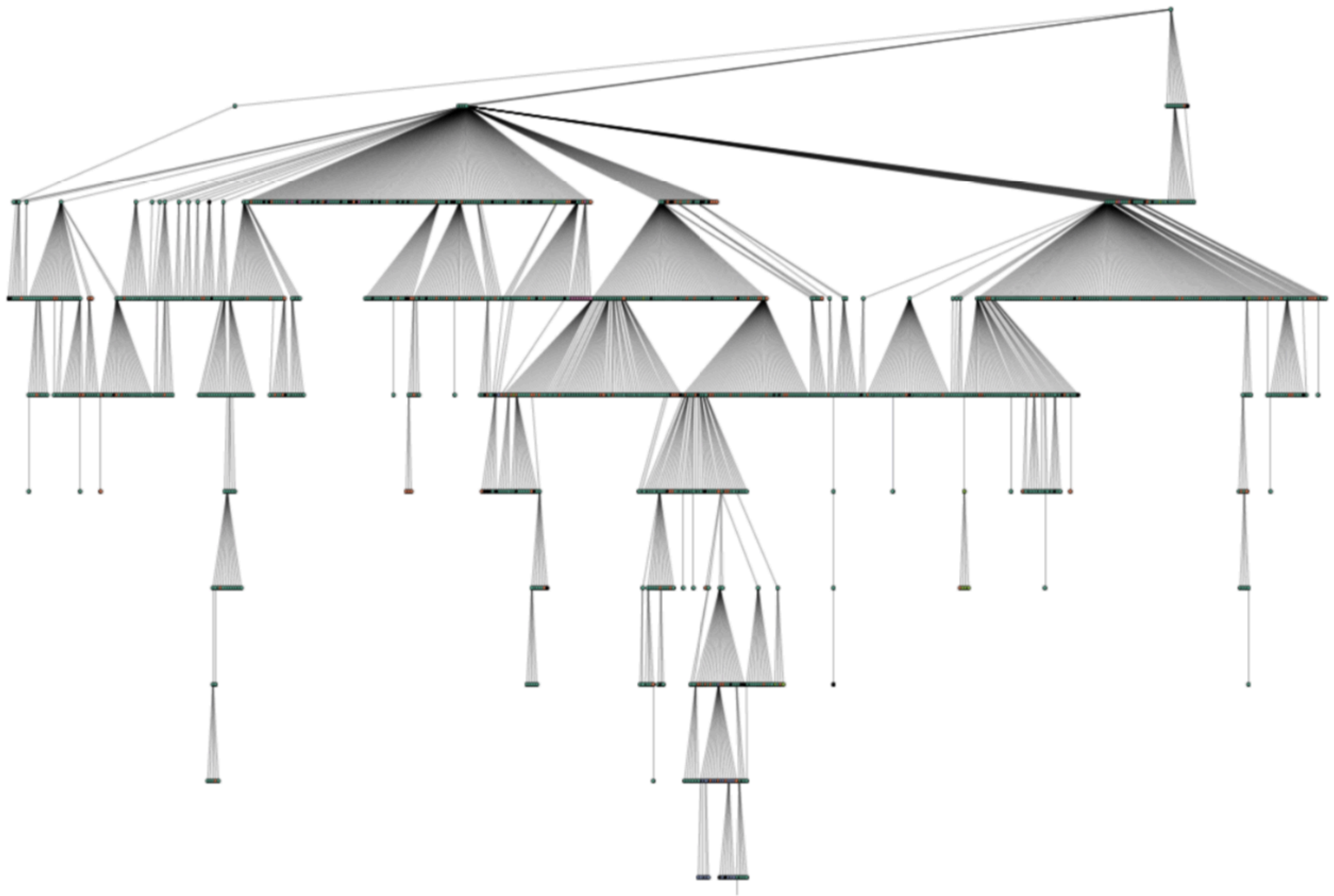
Retweet networks: polarized (left) and unpolarized (right)

Application: Understanding Virality



Information cascades in networks

Application: Product Adoption



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

[Anderson et al., WWW ’15]

Networks matter

If you want to **predict the spread of a disease**, you need to know who is in contact with whom

If you want to **understand the structure of the Web** (or serve relevant search results), you have to analyze the links

If you want to understand the **dissemination of news or the evolution of science**, you have to follow the flow

About CSCC46

Ways to Analyze Networks

What do we hope to achieve from studying networks?

Develop the language of interconnectedness

Learn the patterns and statistical properties of network data

Understand design principles and models of networks

Develop algorithmic understanding of processes in networked systems

Networks: Structure, Dynamics, Incentives

What do we study in networks?

Structure and evolution

What is the structure of networks?
Why and how do they come to have such structure?
How do we harness the structure to extract useful information?

Processes and dynamics

Networks are the "skeleton" on which information, behaviours, and diseases spread
How do information and diseases spread?

Incentives in networks

Behaviour is interconnected by depending on what others do
How do decisions and behaviours depend on network structure and dynamics?

Reasoning about Networks

How do we reason about networks and collective behaviour?

Empirical analyses: Study network data to find organizational principles

How do we measure and quantify networks?

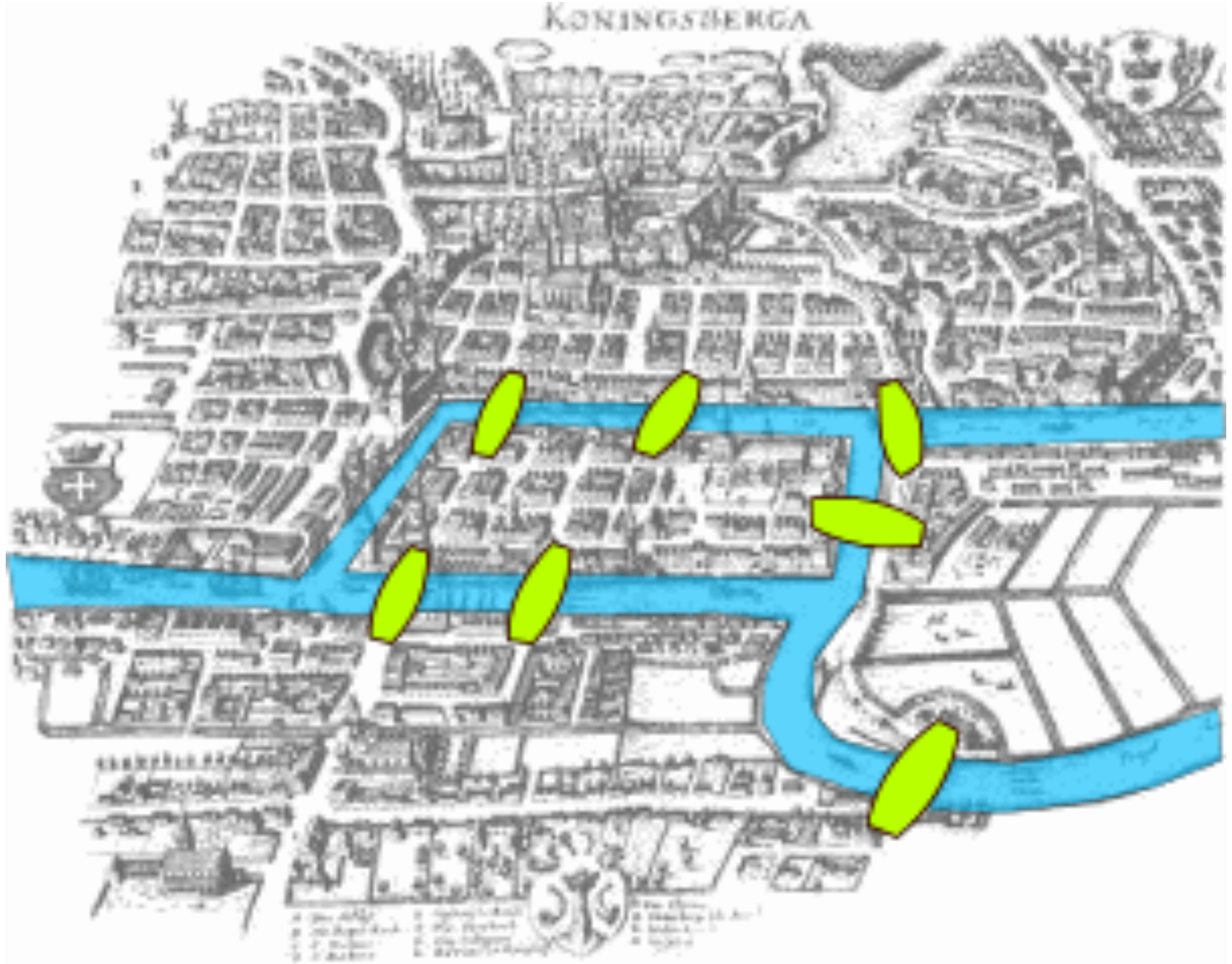
Mathematical models: Graph theory and statistical models

Models allow us to understand behaviors and distinguish surprising from expected phenomena

Algorithms for analyzing graphs: Computer science toolkit

Overcome hard computational challenges to solve important problems

Main Tool: Graph Theory



Seven Bridges of Königsberg [Euler, 1735]

We'll make extensive use of graph theory in this course

Main Tool: Game Theory

The mathematical theory of strategic interaction

We’ll also make extensive use of game theory in this course

(Tentative) Course Overview

Week 1: Course overview, Introduction to graph theory, The Web as a Network

Week 2: Network Representations, Affiliation, Homophily, Strong and Weak Ties, Structural Holes

Week 3: Strong and Weak Ties; Community Detection

Week 4: Signed Networks; Structural Balance; Homophily

Week 5: Six Degrees; Decentralized Search

Week 6: Power Laws and Rich-Get-Richer Phenomena

Week 7: Link Analysis; PageRank

Week 8: Game Theory

Week 9: Congestion; Decision Cascades; Information Cascades

Week 10: Contagion; Epidemics

Week 11: Voting

Week 12: Review

Course resources

Course webpage (http://www.cs.toronto.edu/~ashton/cscc46/)

Quercus (course announcements, assignments)

MarkUs (assignment submission)

Discord (chat)

Course evaluation

35% 4 assignments

10% 2 blog posts

5% Class citizenship

50% Final exam

Assignments are due on Wednesdays at 10am

Because I understand sometimes stuff happens, you get 4 "flex days": 1 flex day is a 24-hour period that you can hand in assignments late with no penalty. You can use up to 2 flex days per assignment. After that, no late assignments will be accepted.

Blog posts

- During the term, write two blog posts on the course blog about topics related to the course
- Should be fun! Pick your favourite topic and explore it in more detail
- Short essay-like posts aimed at your peers
- Commenting on others' posts counts as participation

TAs

Very talented students! Feel free to go to them for help

Tutorials

Mostly working through concrete examples as a group,
Q&A (group office hours), and assignment help

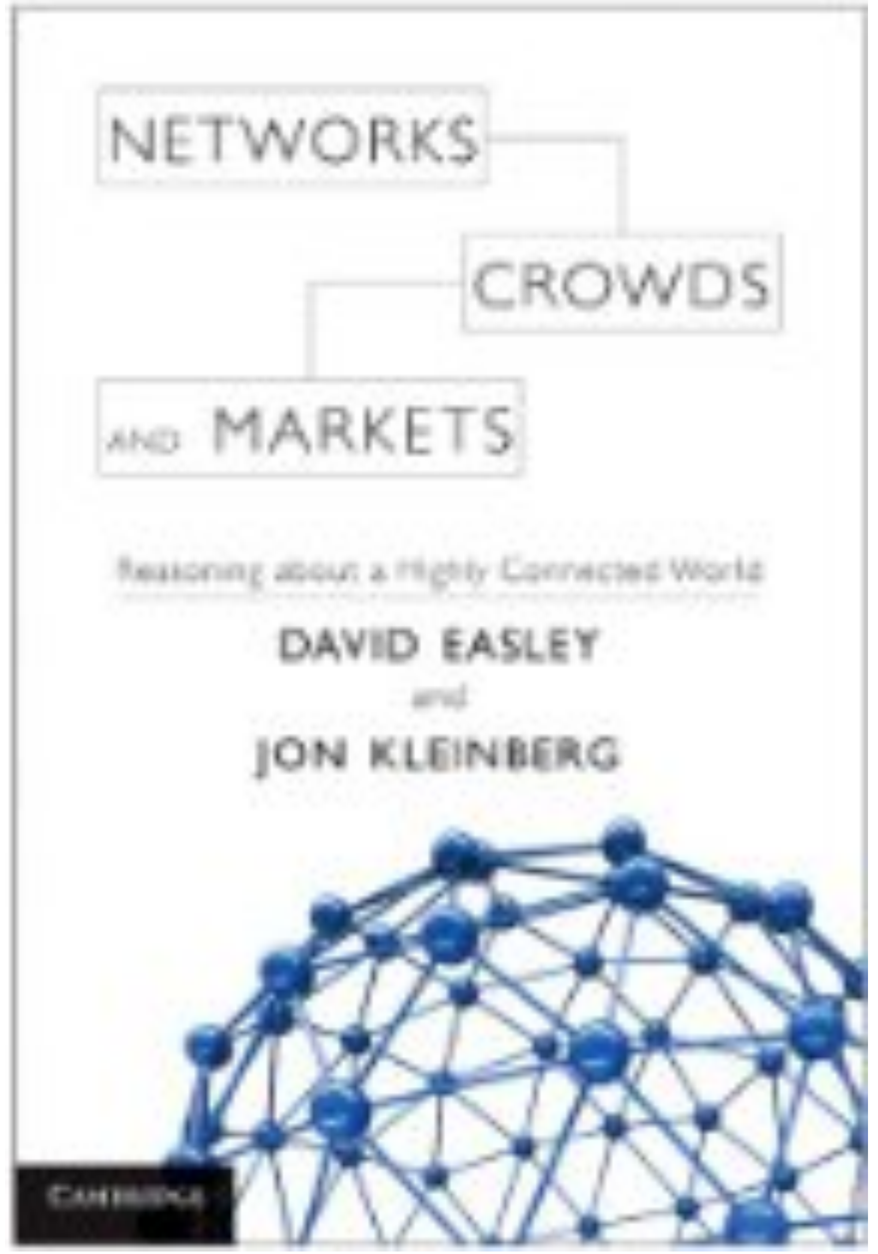
Textbook

“Networks, Crowds & Markets” Easley & Kleinberg

Available free online / reasonably-priced hardcover

Very readable, engaging text

Some assignment exercises from the book



Questions?

About Me

0–18 → 18–22 → 22–29 → 29–31 → 31+

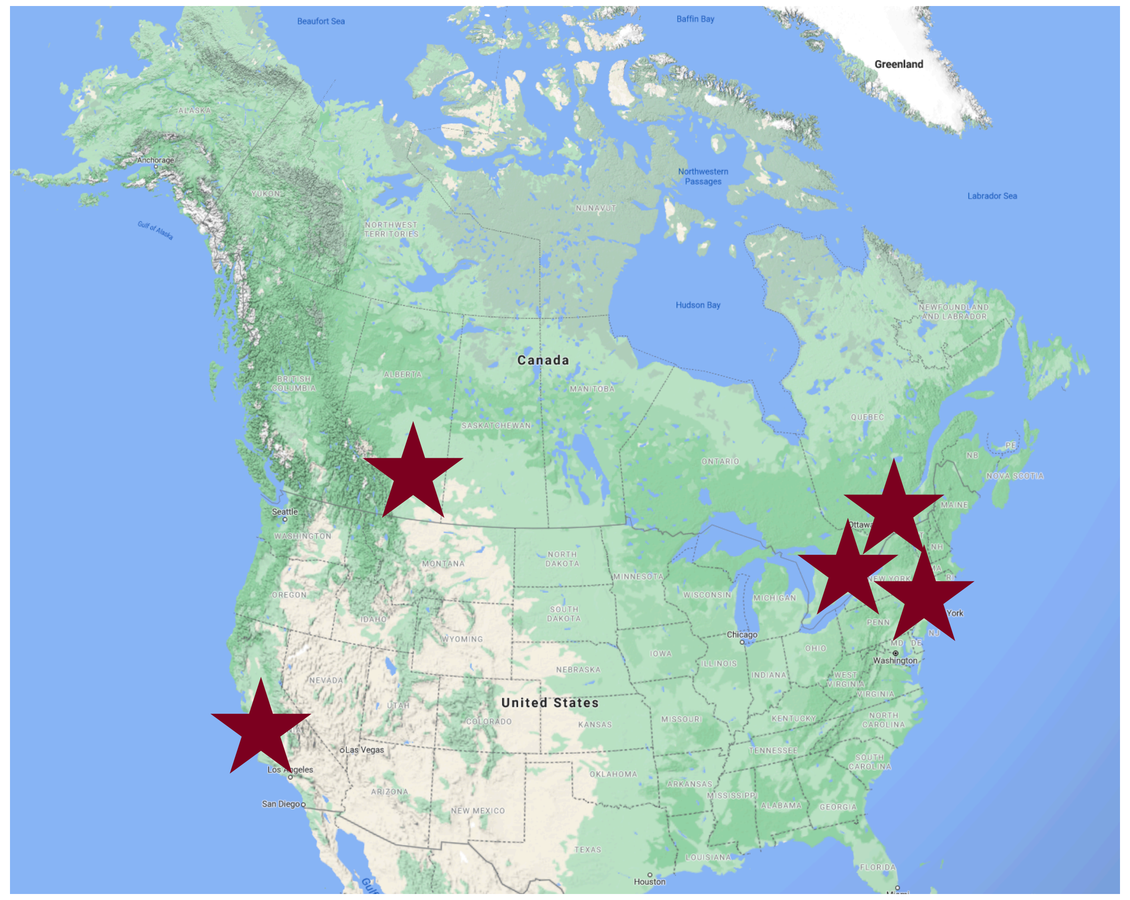
Calgary → Montreal → San Francisco → New York City → Toronto

1M → 4M → 7M → 20M → 6M

Now: Assistant Professor of Computer Science at U of T

Head of the Computational Social Science Lab (researching questions in AI, data, and society) 🤓

Computational Social Science Lab | Computer Science UNIVERSITY OF TORONTO

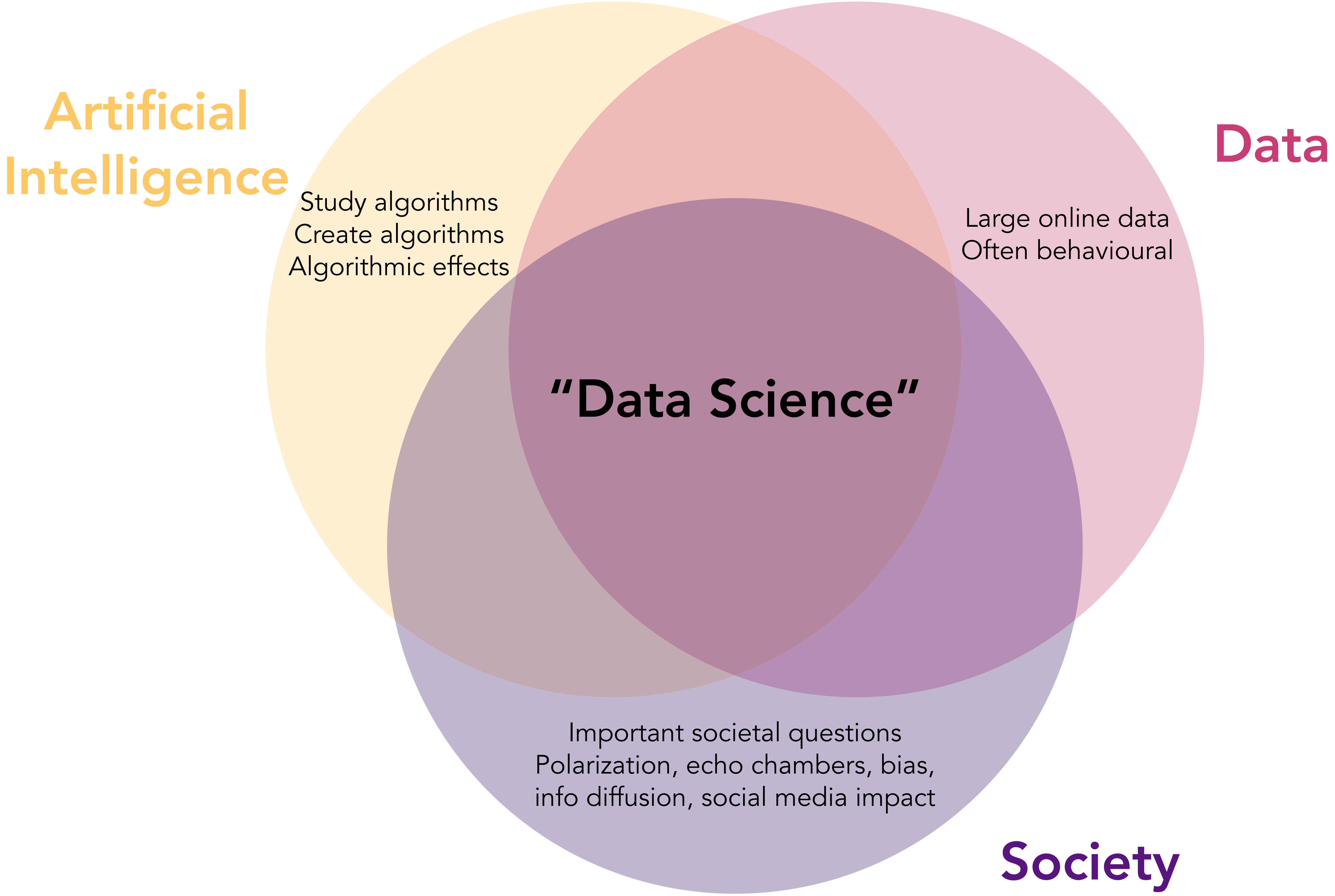


(Want to get involved? Email me after the course!)

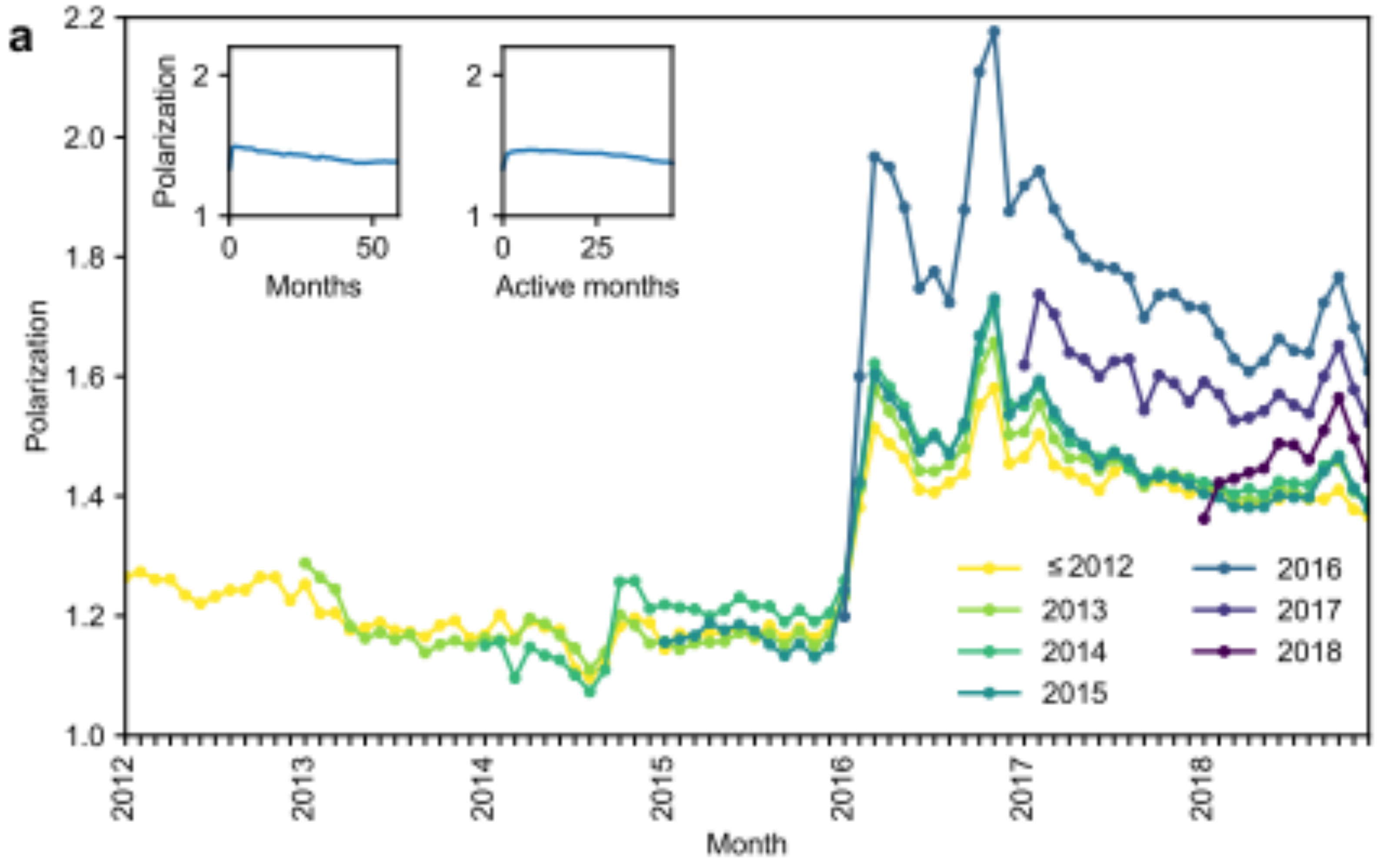
My path

Stage	Interests
McGill B.Soft.Eng '08	Theoretical Quantum algorithms and information Anything practical was impure
Stanford Master's '10	"Hmm…would be nice to feel more connected to the world" Game theory: computational/economic lens on strategic interaction Mix of theoretical and applied
Stanford Ph.D. '15	Discovered the joy and power of large-scale empirical analysis Computational social science: social research in the digital age Mostly empirical analysis supplemented with theoretical modeling, experimentation, and surveys

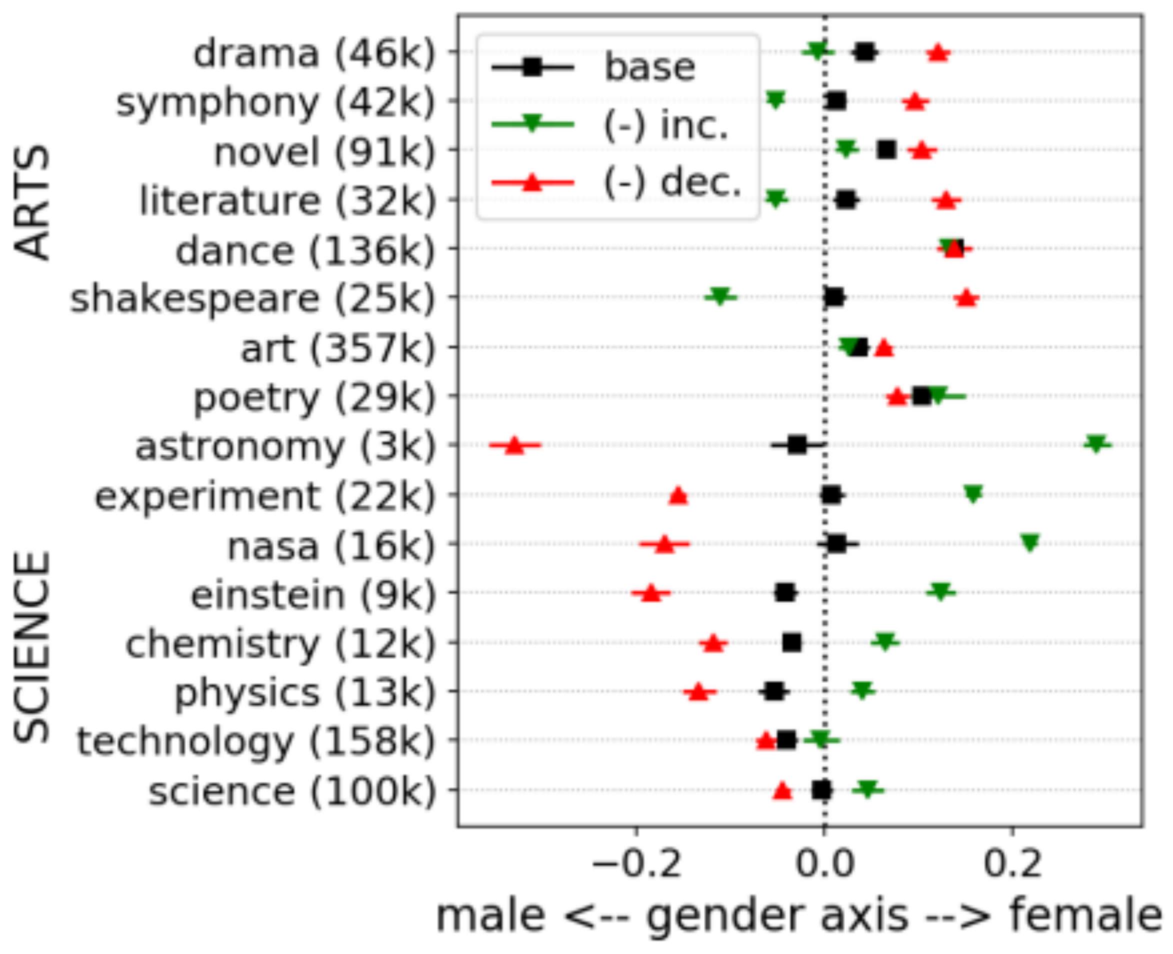
My research


Artificial Intelligence
Study algorithms
Create algorithms
Algorithmic effects
Data
Large online data
Often behavioural
"Data Science"
Important societal questions
Polarization, echo chambers, bias,
info diffusion, social media impact
Society

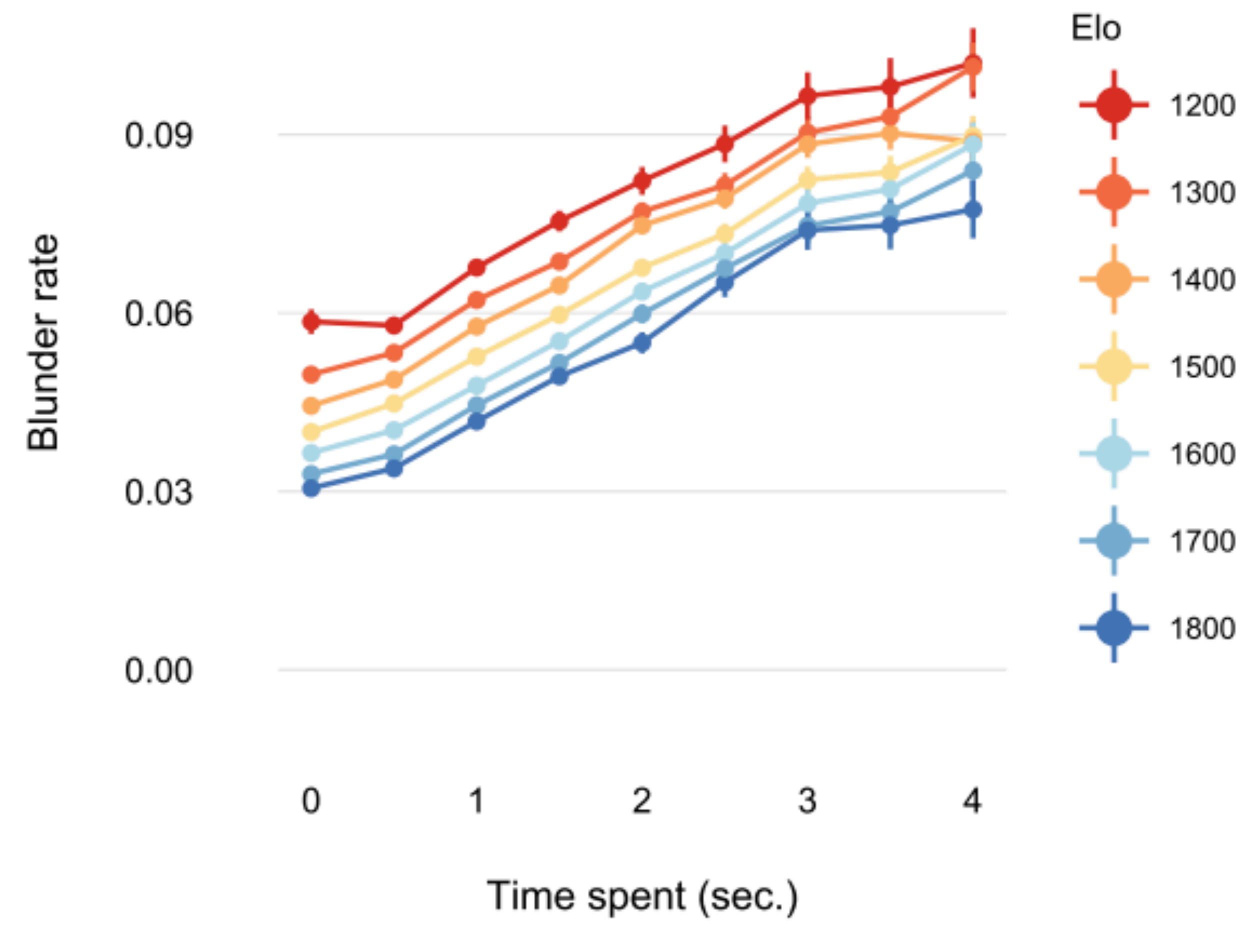
My research



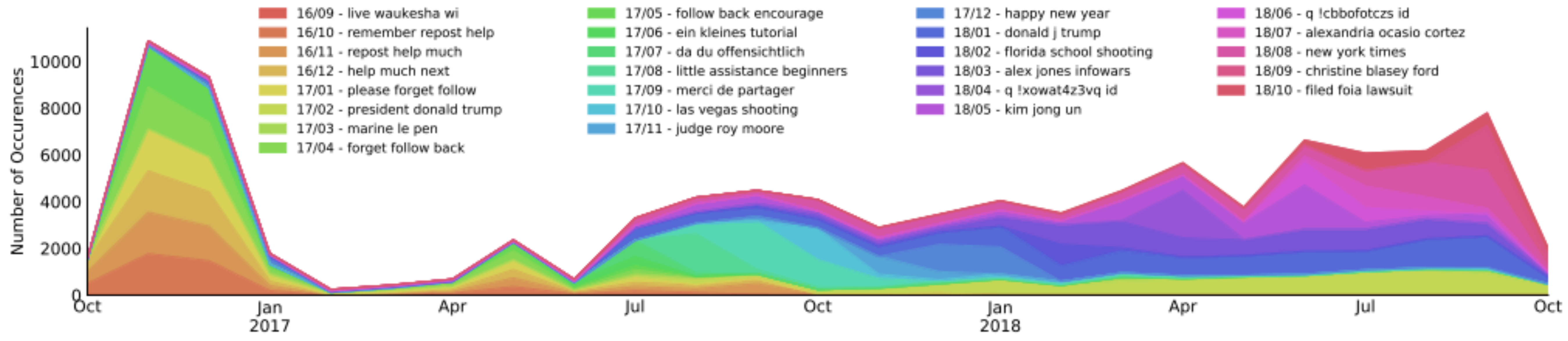
Political polarization on Reddit



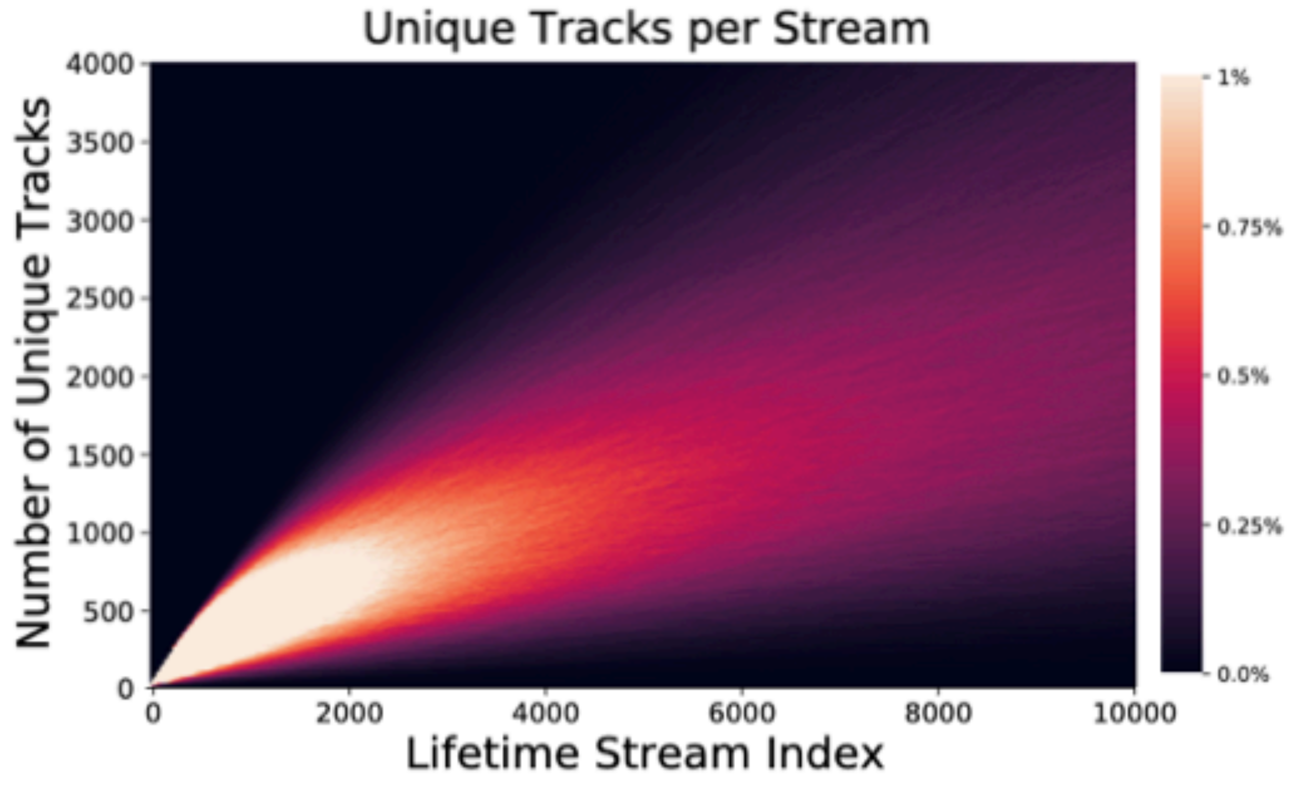
Gender bias in text algorithms



Nature of human error in chess



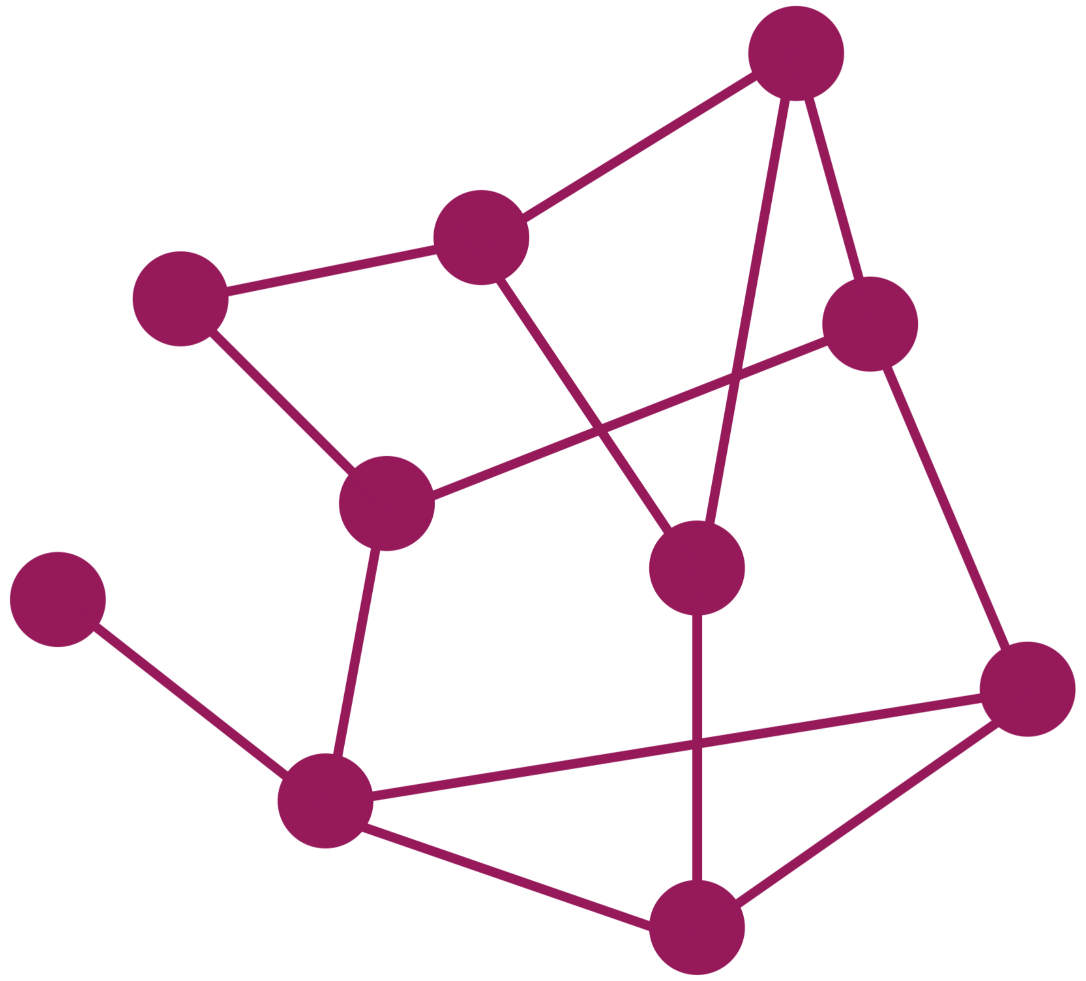
Discussion topics on Gab (alt-right platform)



Music exploration on Spotify

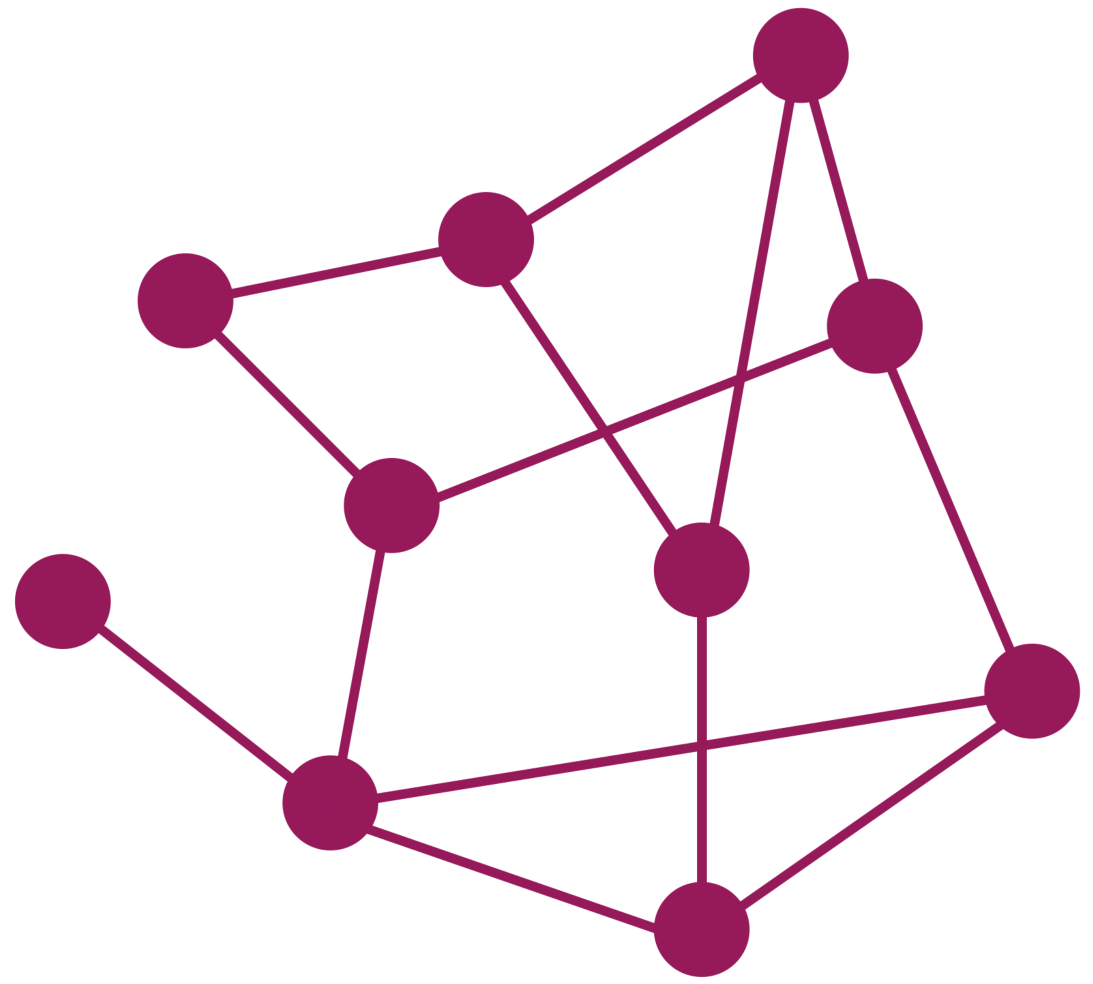
First topic: Network Analysis Fundamentals and The Structure of the Web

A Network



A network is a collection of objects where some pairs of objects are connected by links

Components of a Network



Objects: nodes, vertices — N
Interactions: links, edges — E
System: network, graph — $G(N,E)$

Networks or Graphs?

Network often refers to real systems

Web, Social network, Metabolic network

Language: Network, node, link

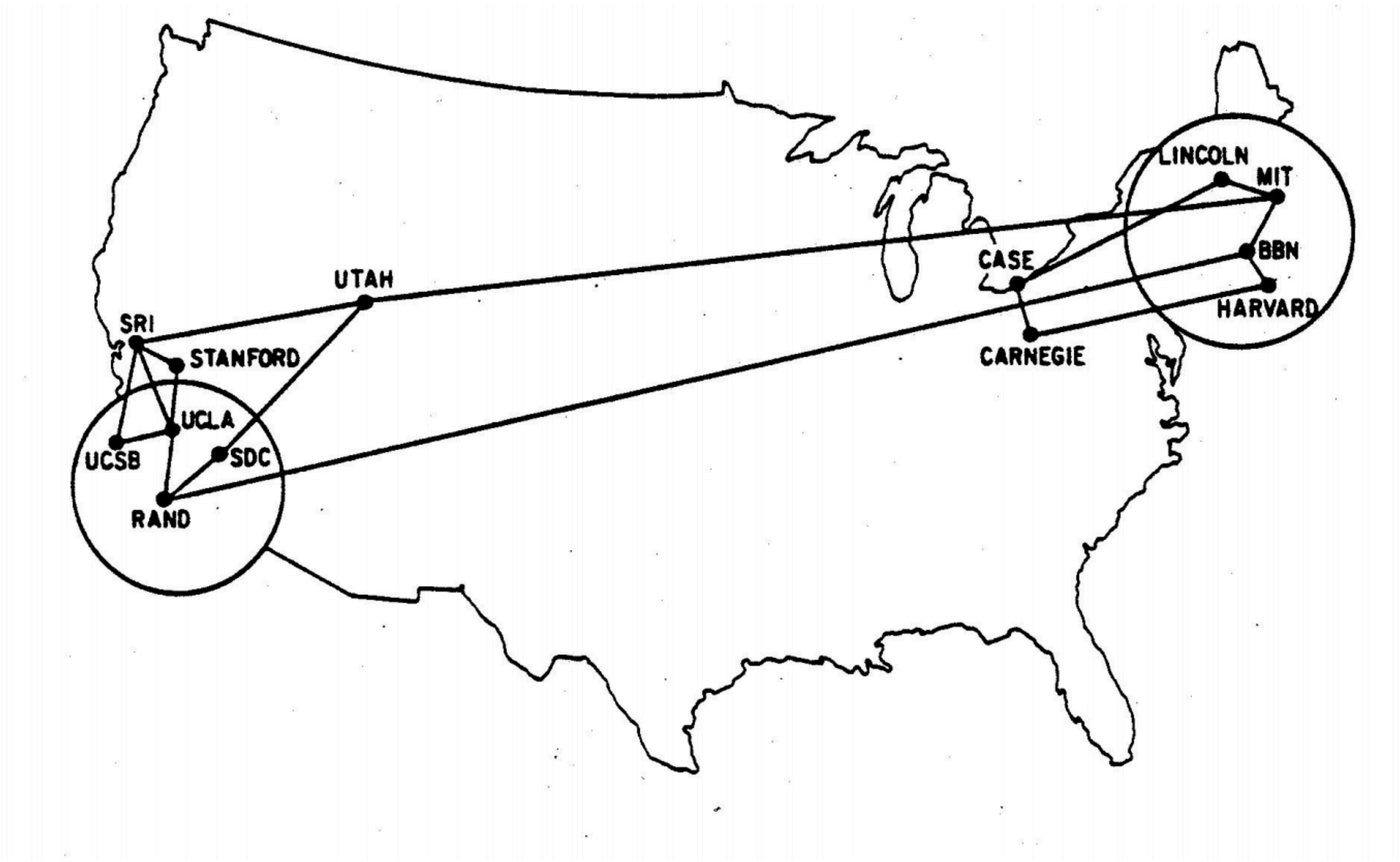
Graph is mathematical representation of a network

Web graph, Social graph (a Facebook term)

Language: Graph, vertex, edge

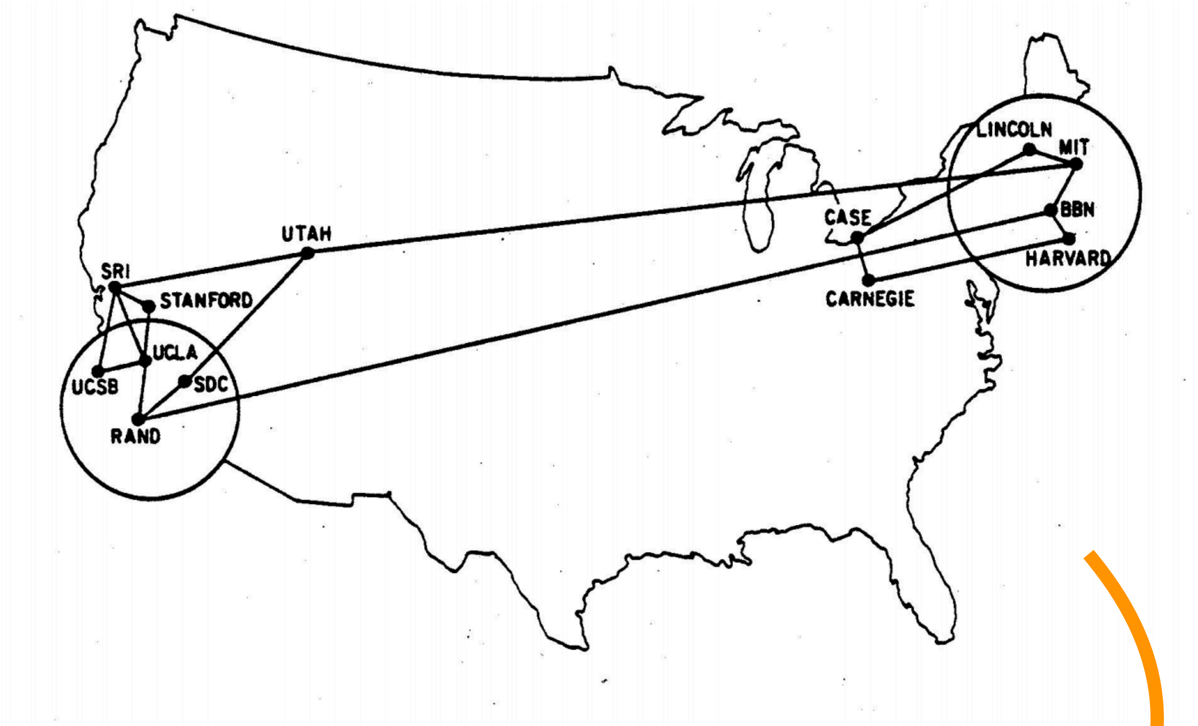
We will try to make this distinction whenever it is appropriate

A first example

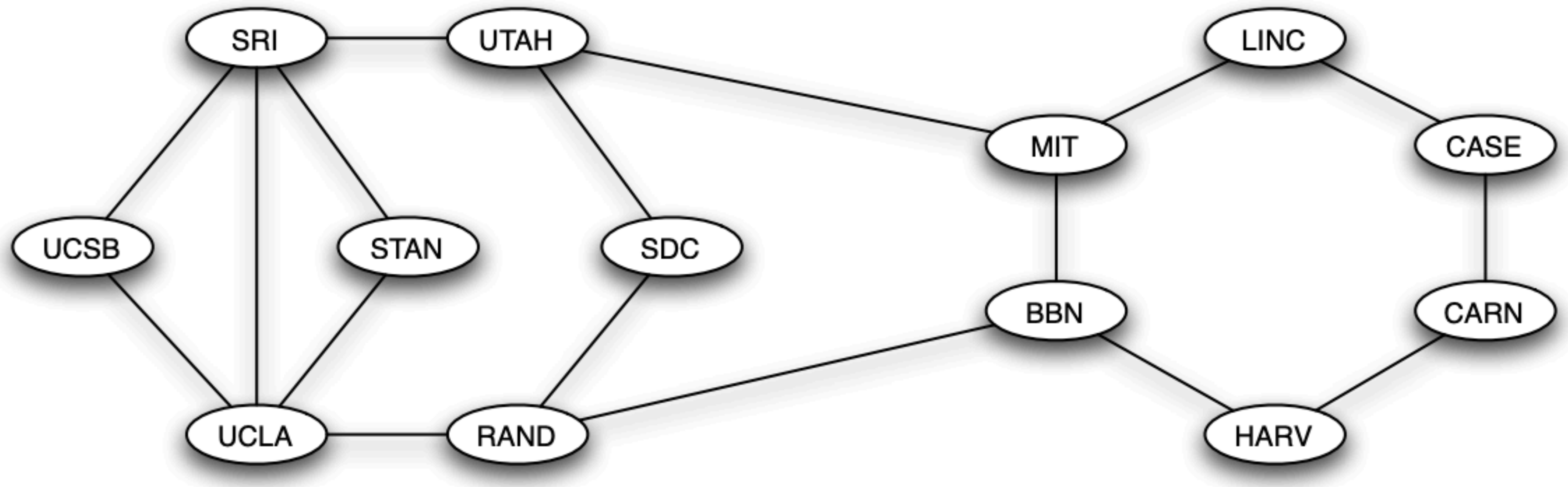


The Internet in 1970

A first example

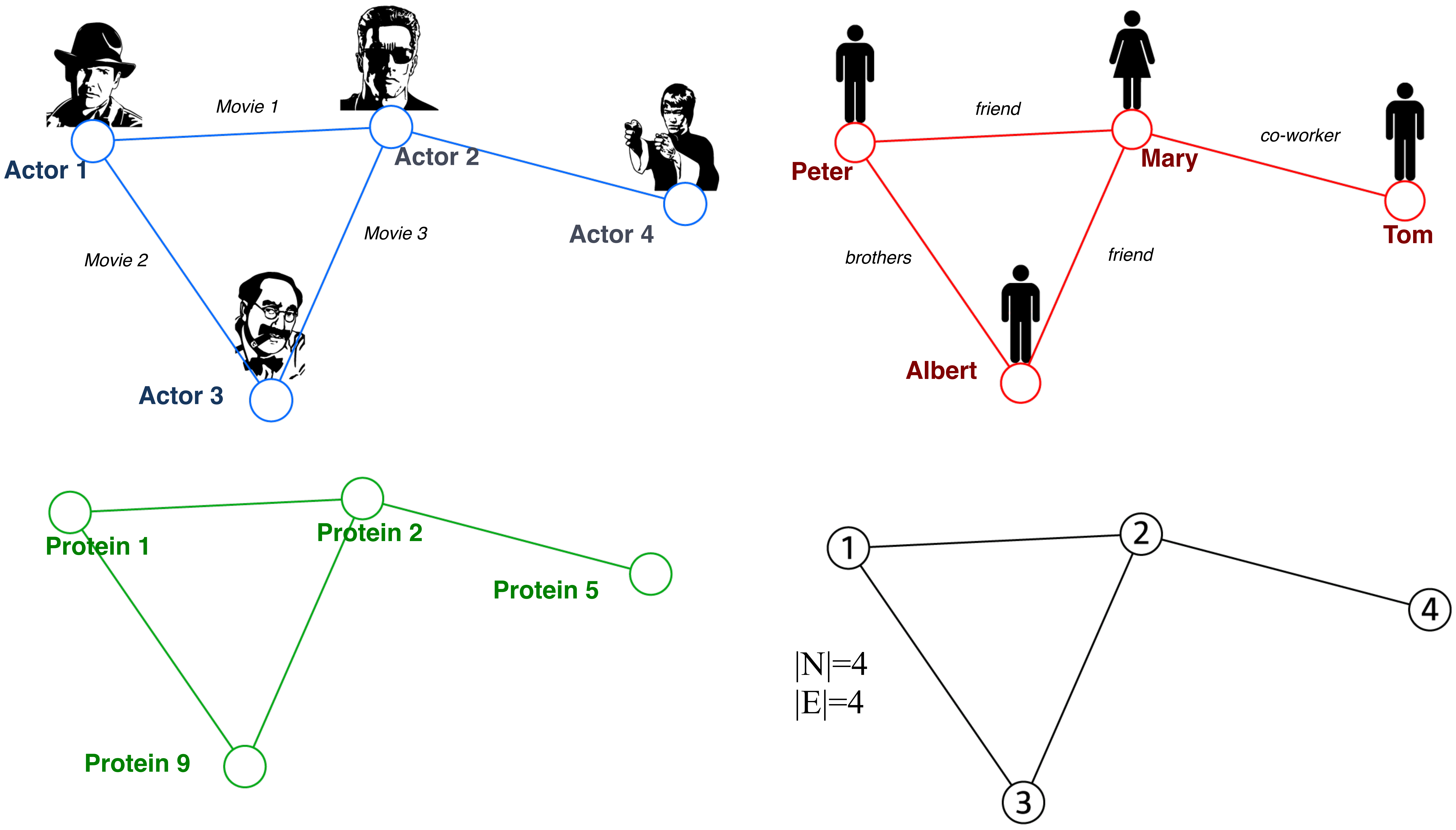


Translation



The Internet in 1970

Networks: a shared language



Choosing a Proper Representation

- **How to build a graph:**
  - What are nodes?
  - What are edges?

- The choice of the proper network representation of a given domain/problem **determines our ability to use networks successfully**:
  - In some cases there is a unique, unambiguous representation
  - In other cases, the representation is by no means unique
  - The way you assign links will determine the nature of the question you can study

Choosing a Proper Representation

- If you connect individuals that work with each other, you will be exploring a **professional network**
- If you connect those that have a friendship relationship, you will be exploring a **friendship network**
- If you connect scientific papers that cite each other, you will be studying a **citation network**

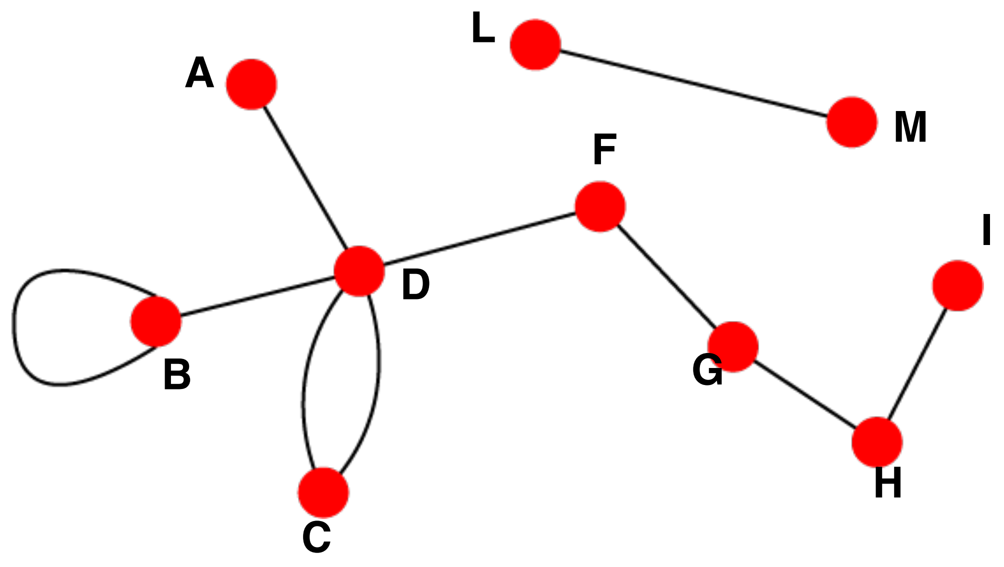
If you connect all people with first names that share the same first letter, what are you studying?

It is a network, but is it meaningful?

Undirected and Directed Networks

Undirected

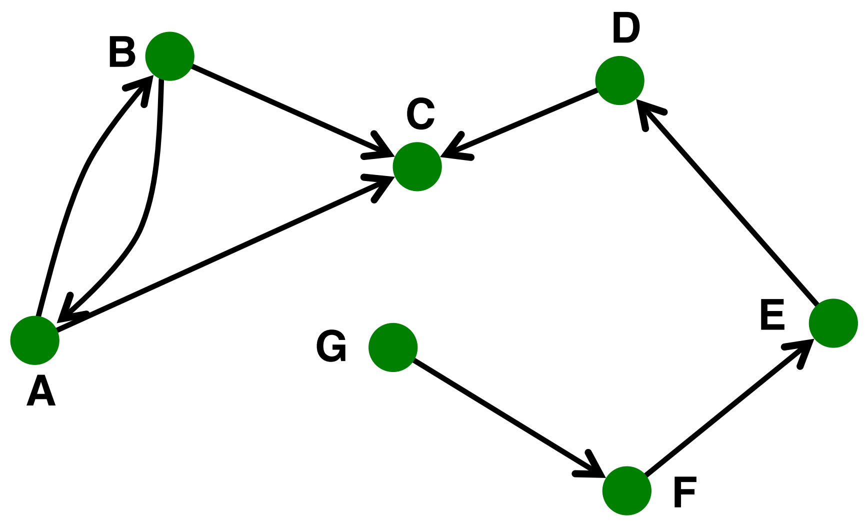
- **Links:** undirected (symmetrical, reciprocal)



- **Examples:**
  - Collaborations
  - Friendship on Facebook

Directed

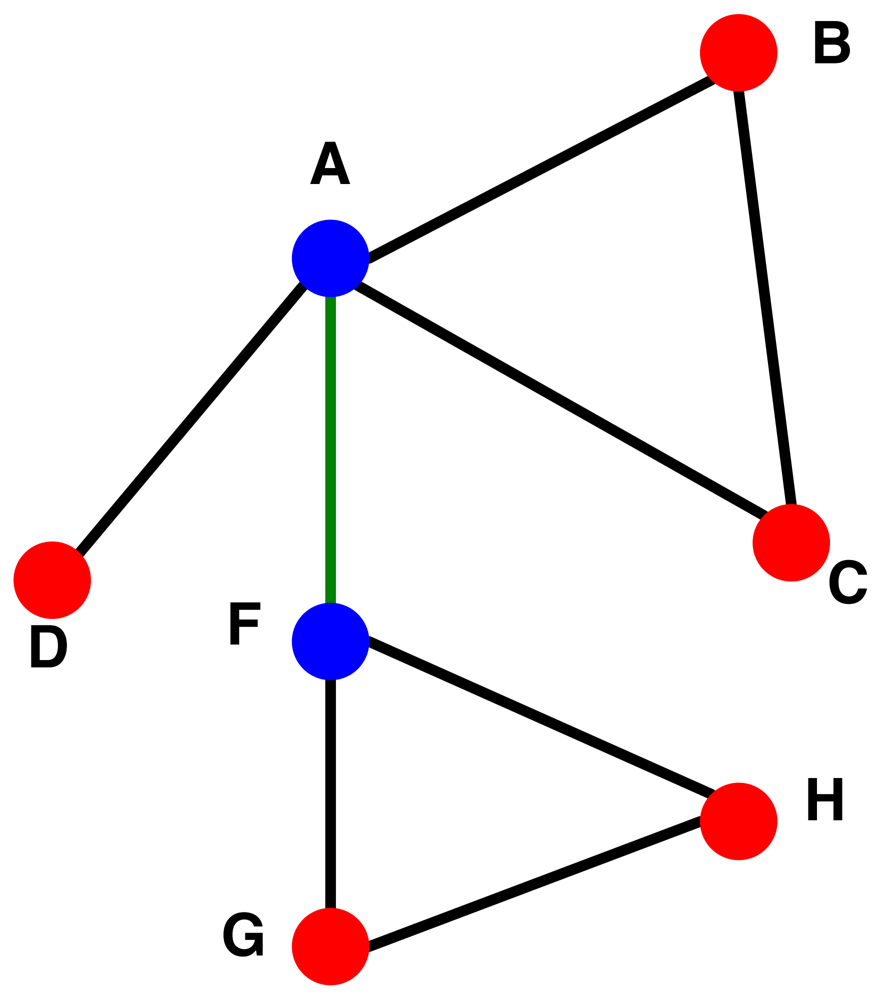
- **Links:** directed (arcs)

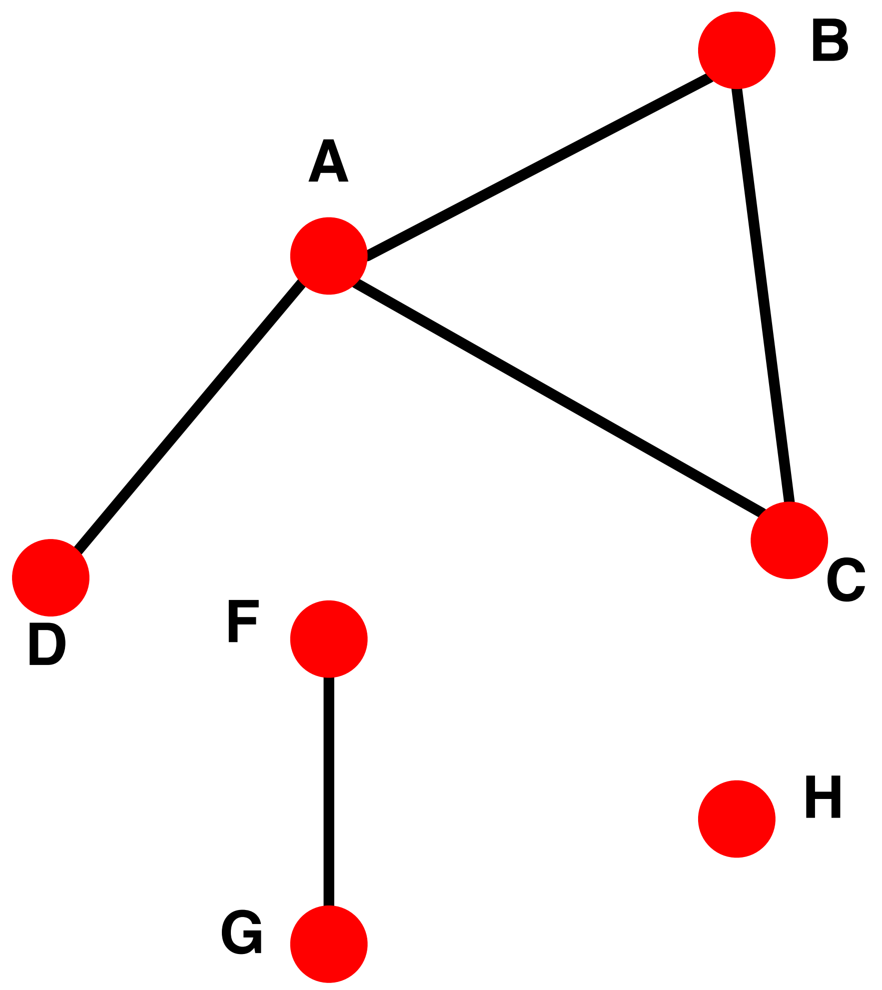


- **Examples:**
  - Phone calls
  - Following on Twitter and TikTok

Connectivity of Graphs

- **Connected component (undirected):**
  - Any two vertices can be joined by a path
  - No superset with the same property
- A disconnected graph is made up of two or more connected components



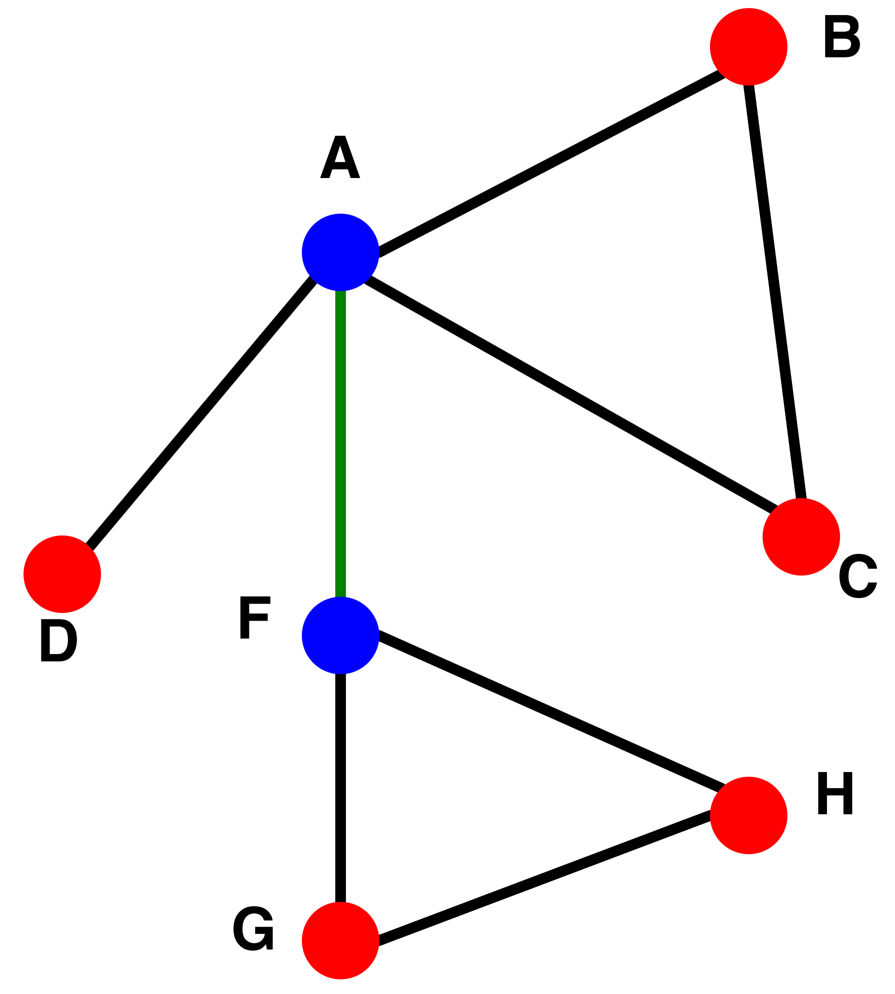


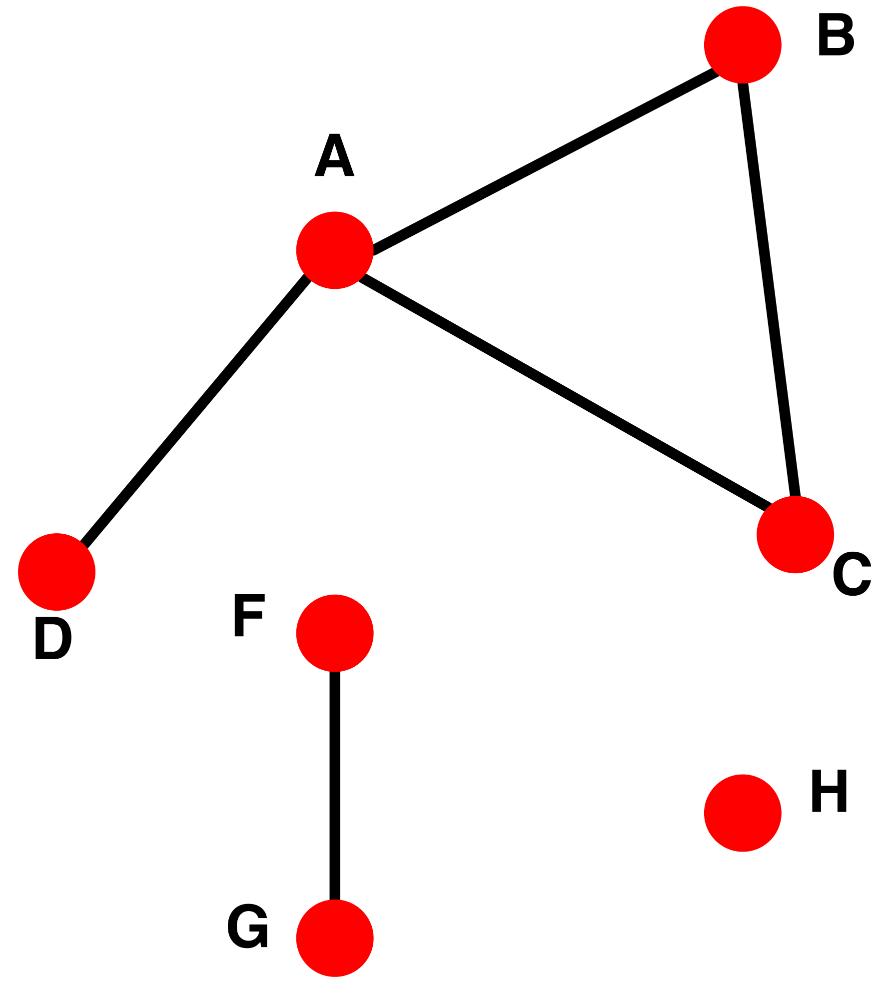
Largest Component:
Giant Component

Isolated node (node H)

Connectivity of Graphs

- **Connected component (undirected):**
  - Any two vertices can be joined by a path
  - No superset with the same property
- A disconnected graph is made up of two or more connected components





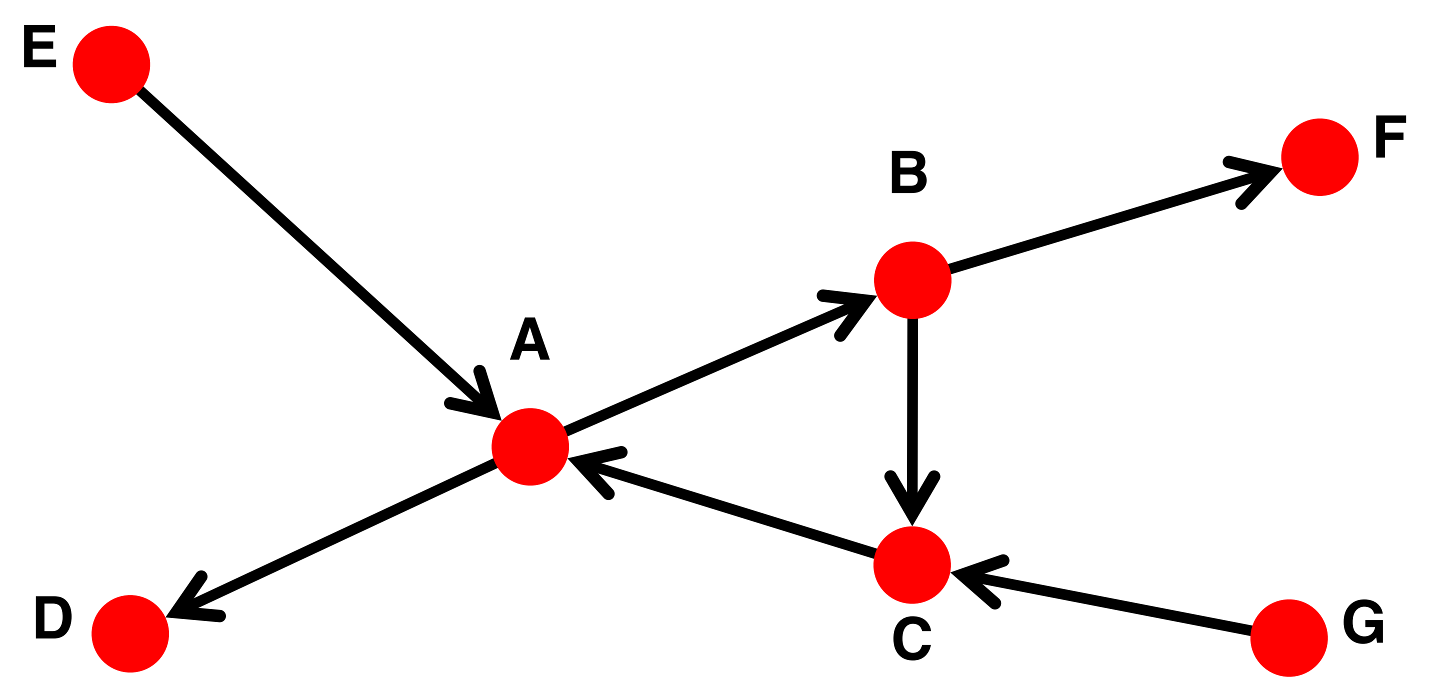
Largest Component:
Giant Component

Isolated node (node H)

Bridge edge: If we erase it, the graph becomes disconnected.

Connectivity of Directed Graphs

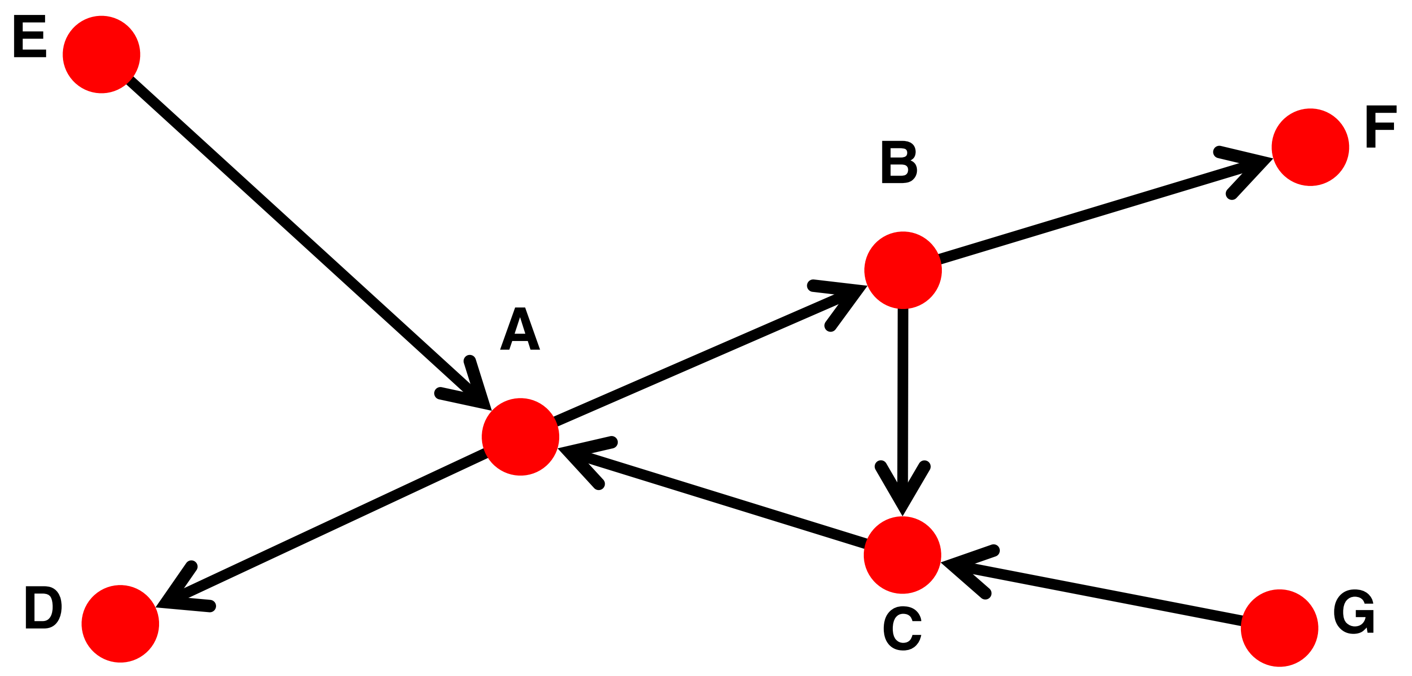
- **Strongly connected directed graph**
  - has a path from each node to every other node and vice versa (e.g., A-B path and B-A path)
- **Weakly connected directed graph**
  - is connected if we disregard the edge directions



Is this graph weakly connected?
Strongly connected?

Connectivity of Directed Graphs

- **Strongly connected directed graph**
  - has a path from each node to every other node and vice versa (e.g., A-B path and B-A path)
- **Weakly connected directed graph**
  - is connected if we disregard the edge directions



It is weakly connected but not strongly connected (e.g., there is no way to get from F to G by following the edge directions)

End!

This week:

- Sign up for MarkUs (to be released on website)
- Log in to Quercus, MarkUs, and Discord
- Read Ch. 1, 2.1-2.4, 13.1-13.4

Next week:

- Network Representations, Affiliation, Homophily, Strong and Weak Ties, Structural Holes
- Read Ch. 3.1-3.3, 4.1-4.3

Contact me: ashton@cs.toronto.edu

Please introduce yourselves!

On Discord, please write a quick blurb about yourself in #general! Tell us who you are, where you are, and a fun fact (and anything else you'd like to share)