# Social and Information Networks

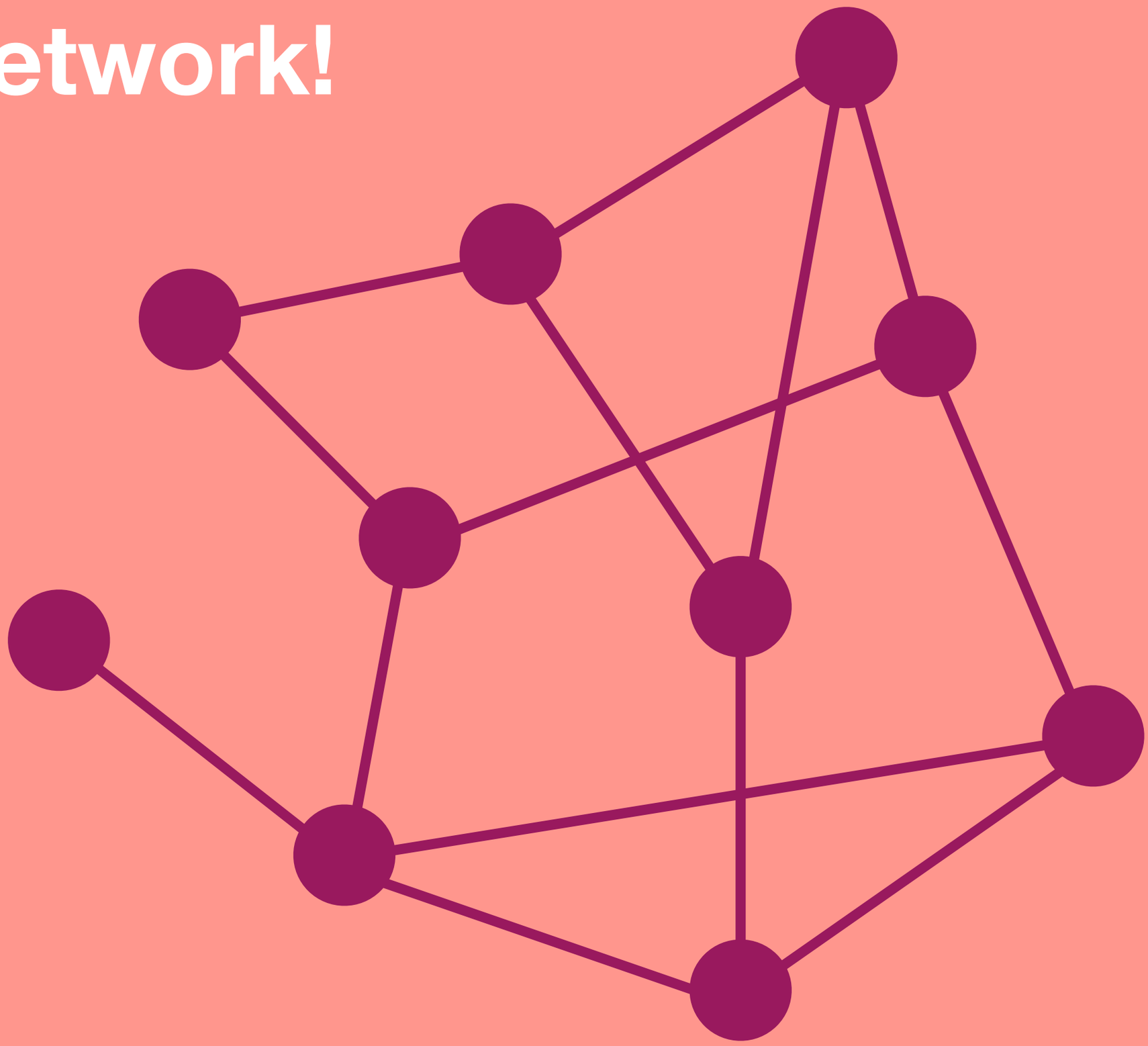
CSCC46H, Fall 2025
Lecture 1

Prof. Ashton Anderson
ashton@cs.toronto.edu

# Why Networks?

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

# A Network!



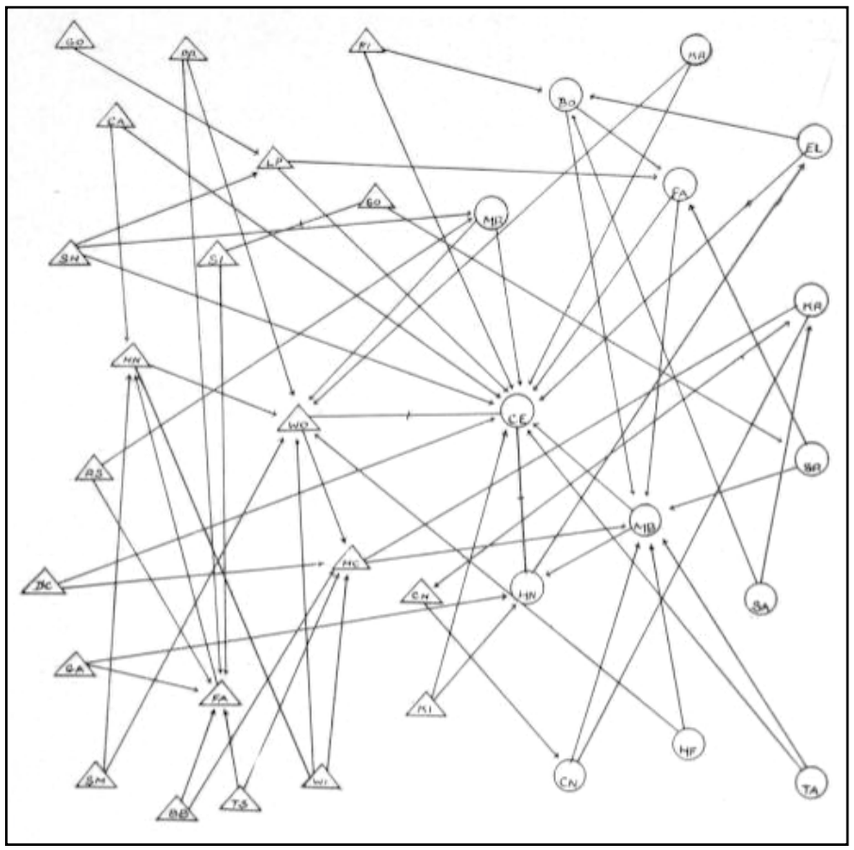
# How to understand the behaviours, decisions, beliefs, etc. of millions of people?


# Social network

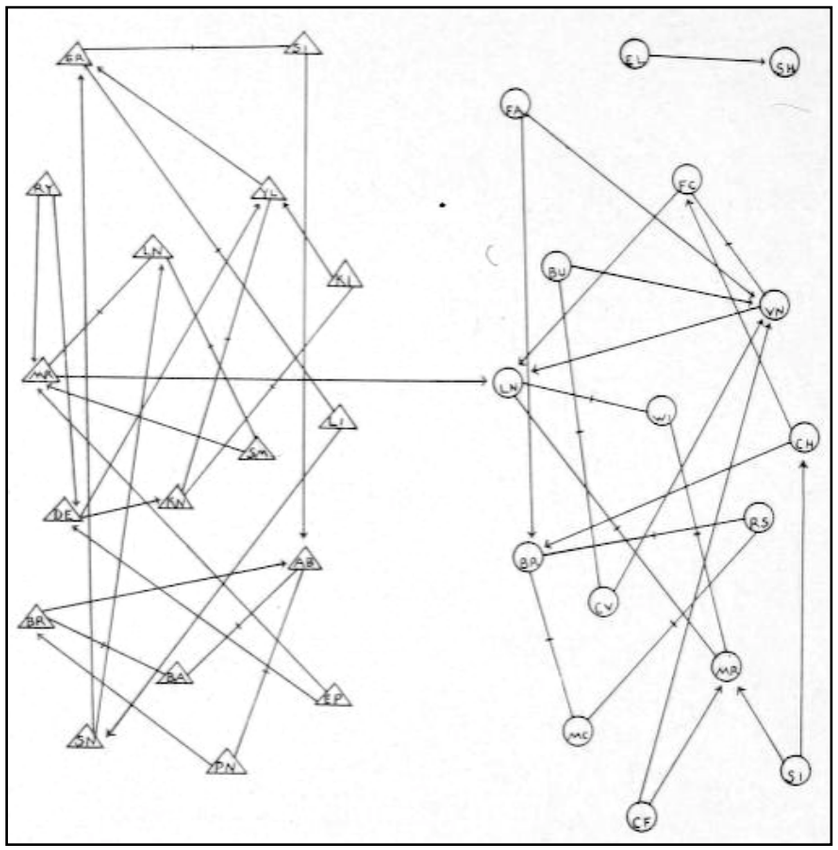
# Educational settings: people learn and interact with each other in complex ways

# EMOTIONS MAPPED BY NEW GEOGRAPHY

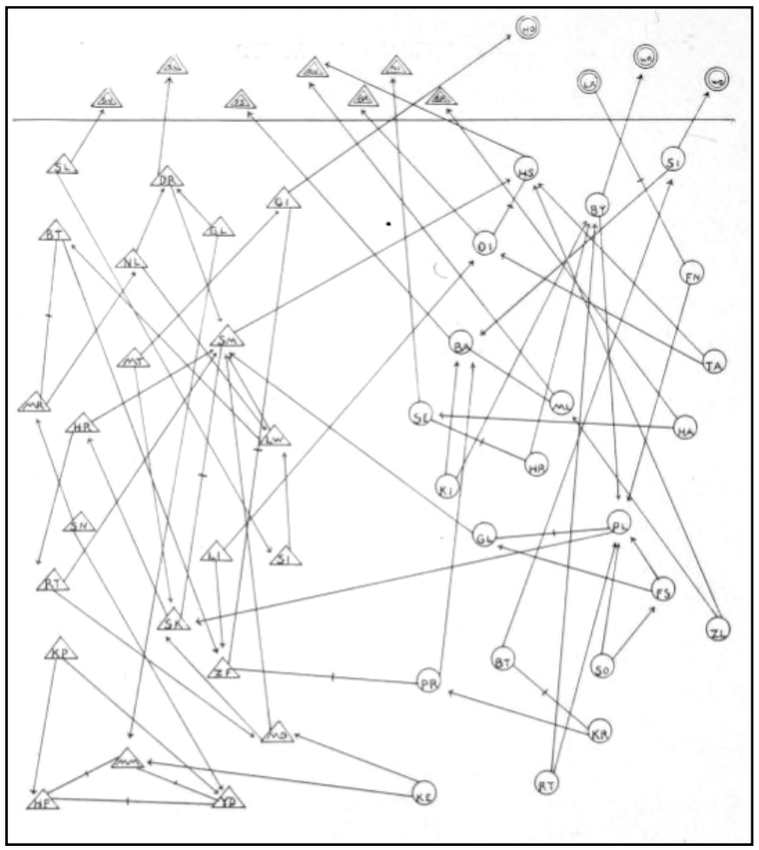
**Charts Seek to Portray the Psychological Currents of Human Relationships.**



1st grade



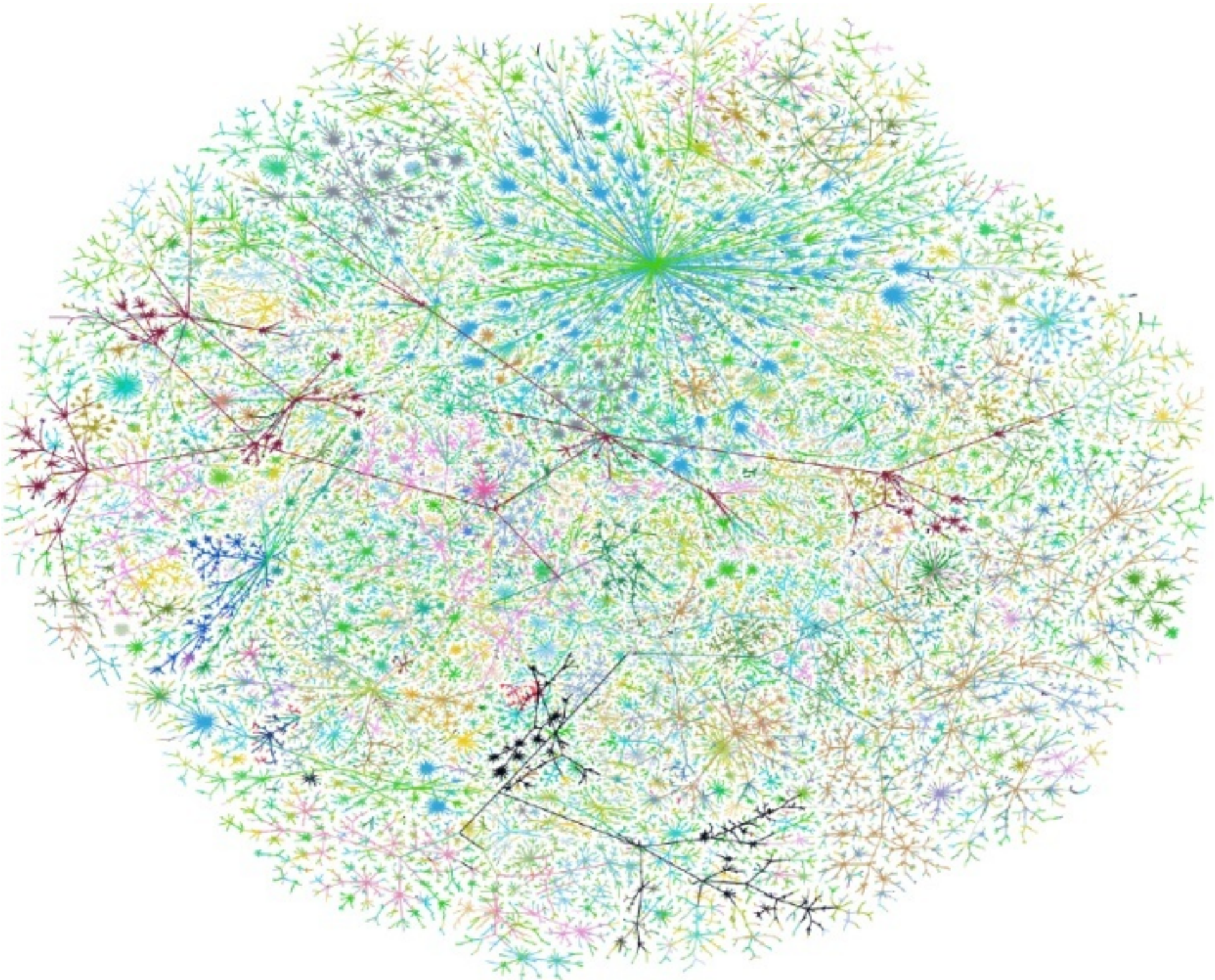
4th grade



8th grade

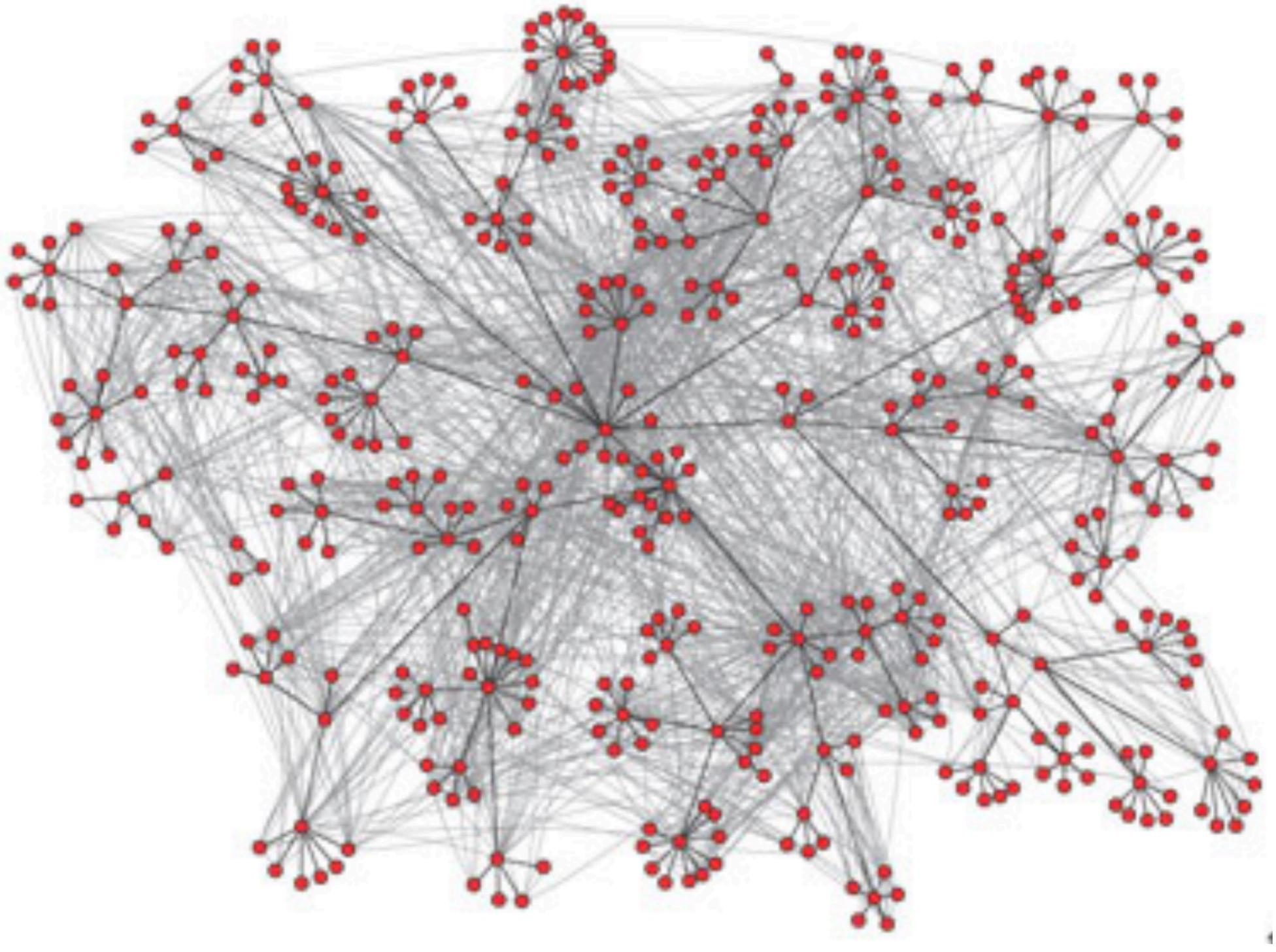
**Moreno’s sociograms, 1934**

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

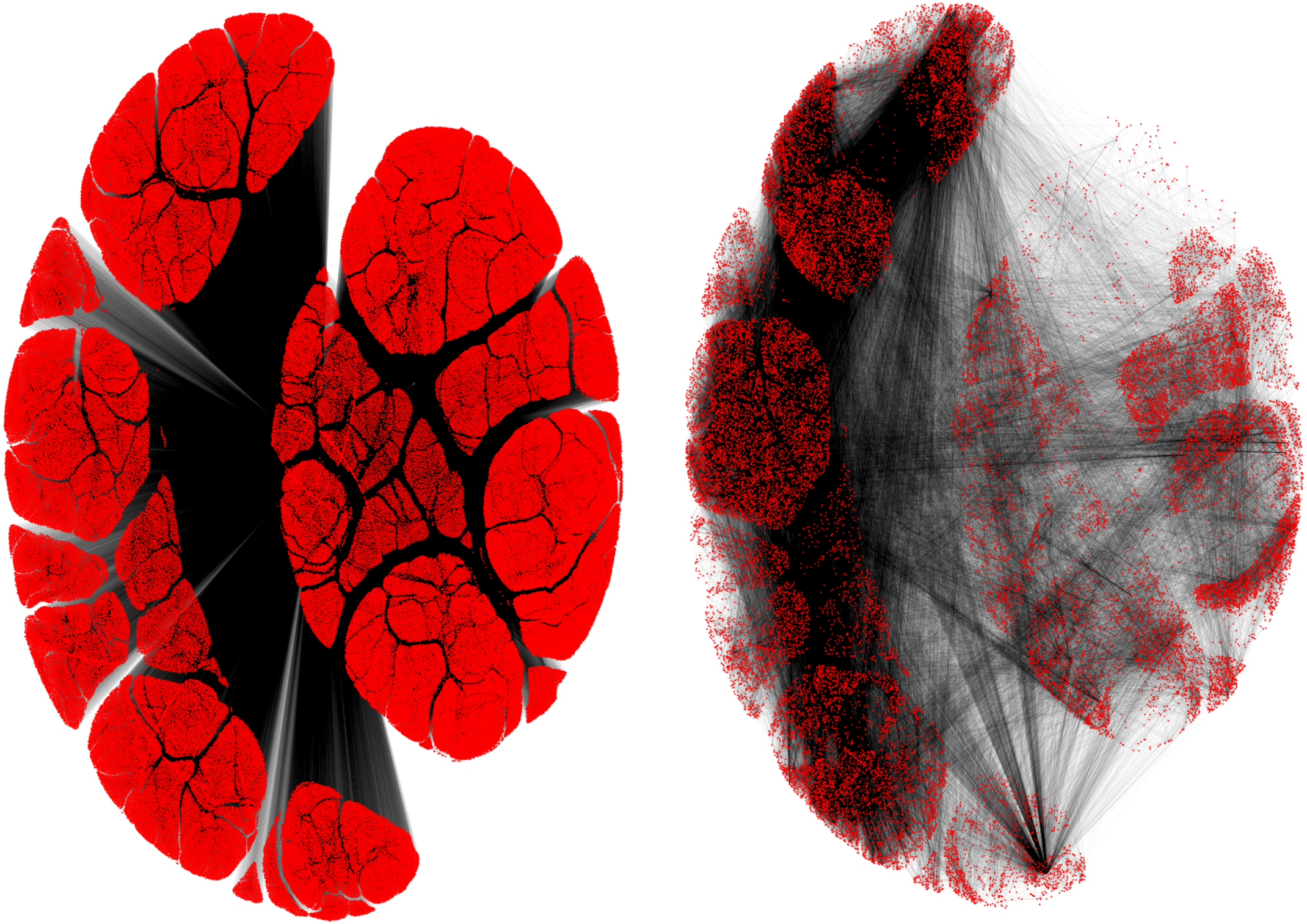


# Graph of the Internet (Autonomous Systems)

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

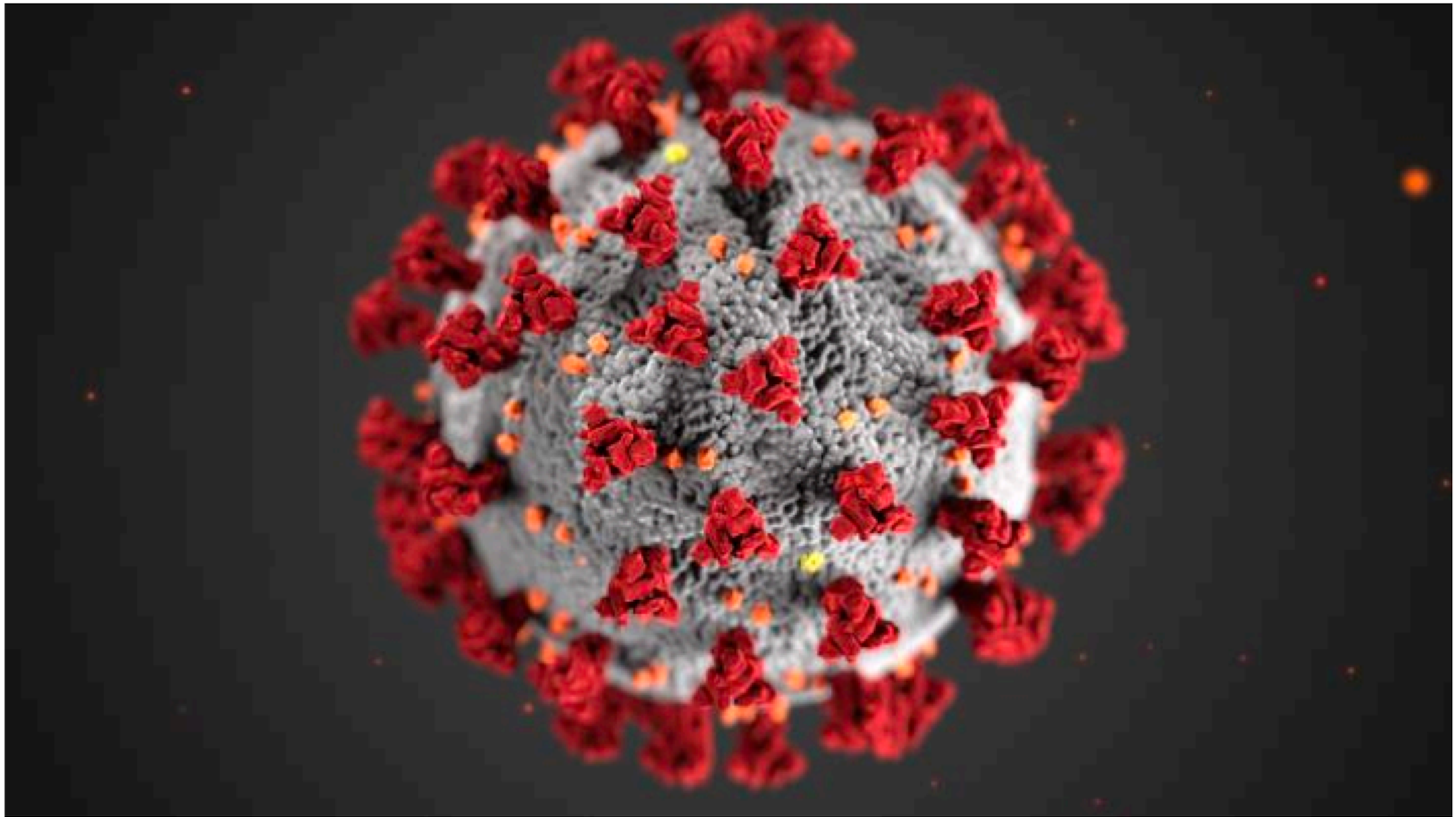


**Email communication network**
(HP Research, 436 employees)

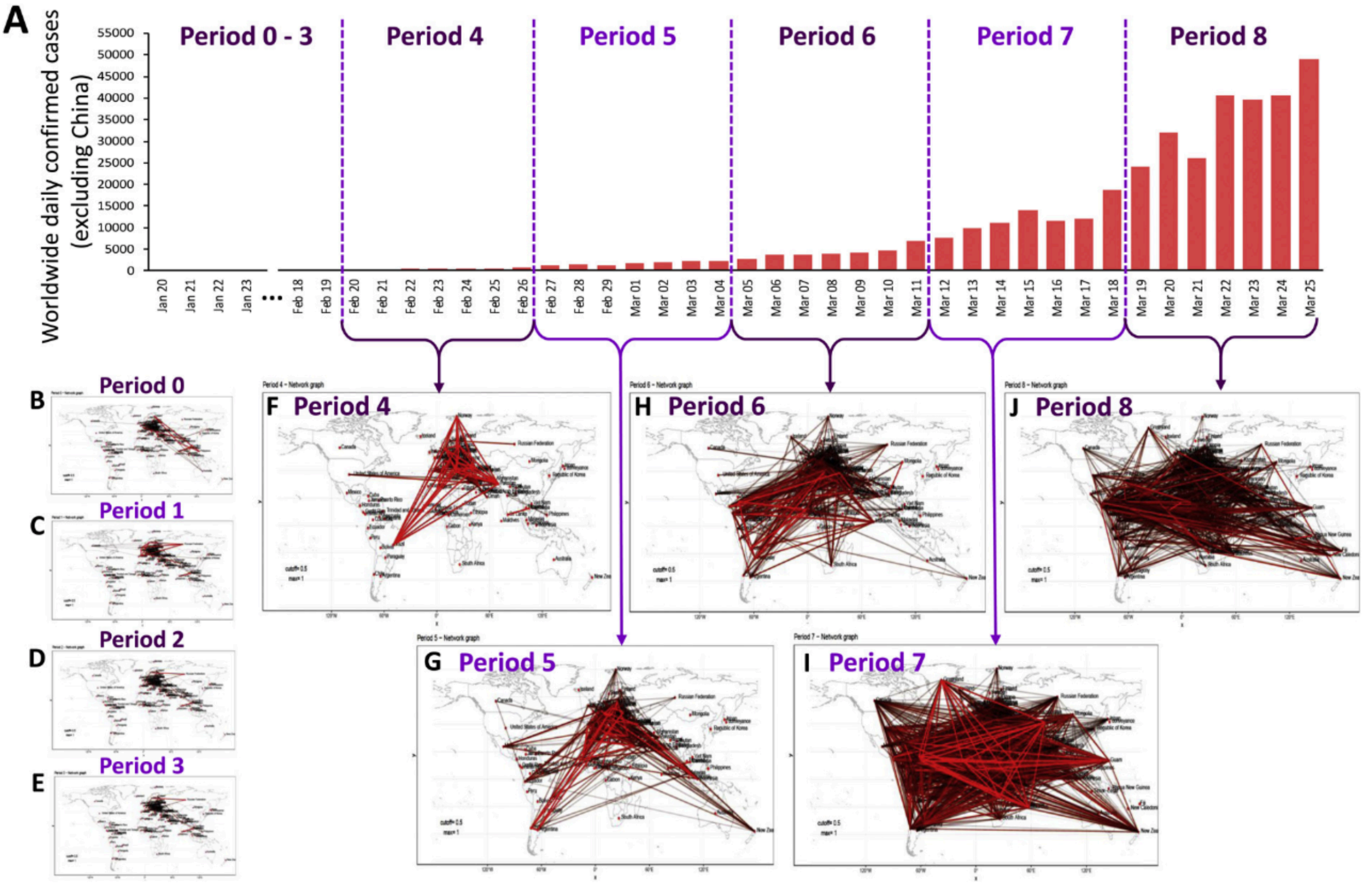


**Org chart (left) and Email communication network (right)**

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

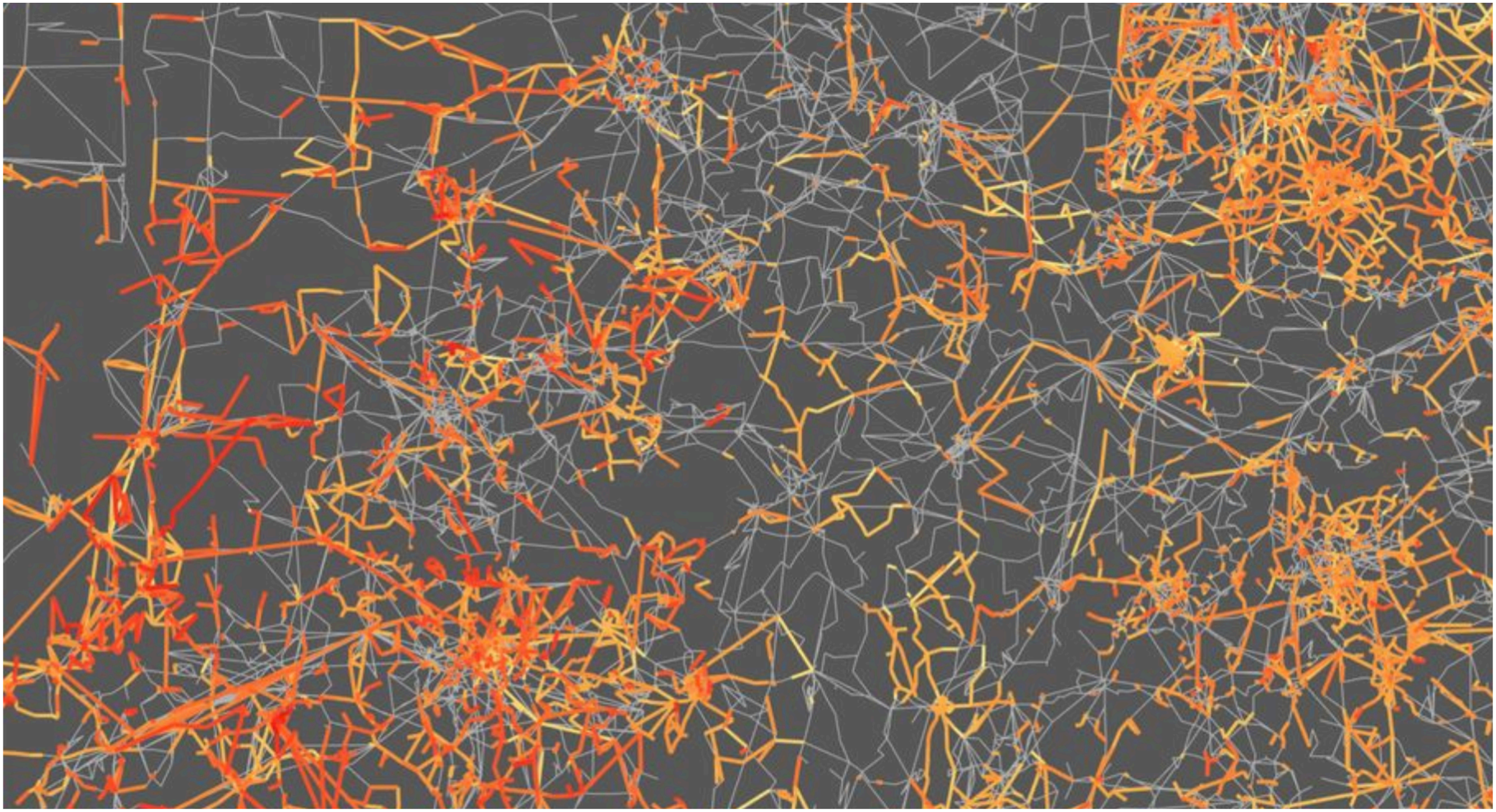


# Many diseases are transmitted socially (e.g. COVID-19)



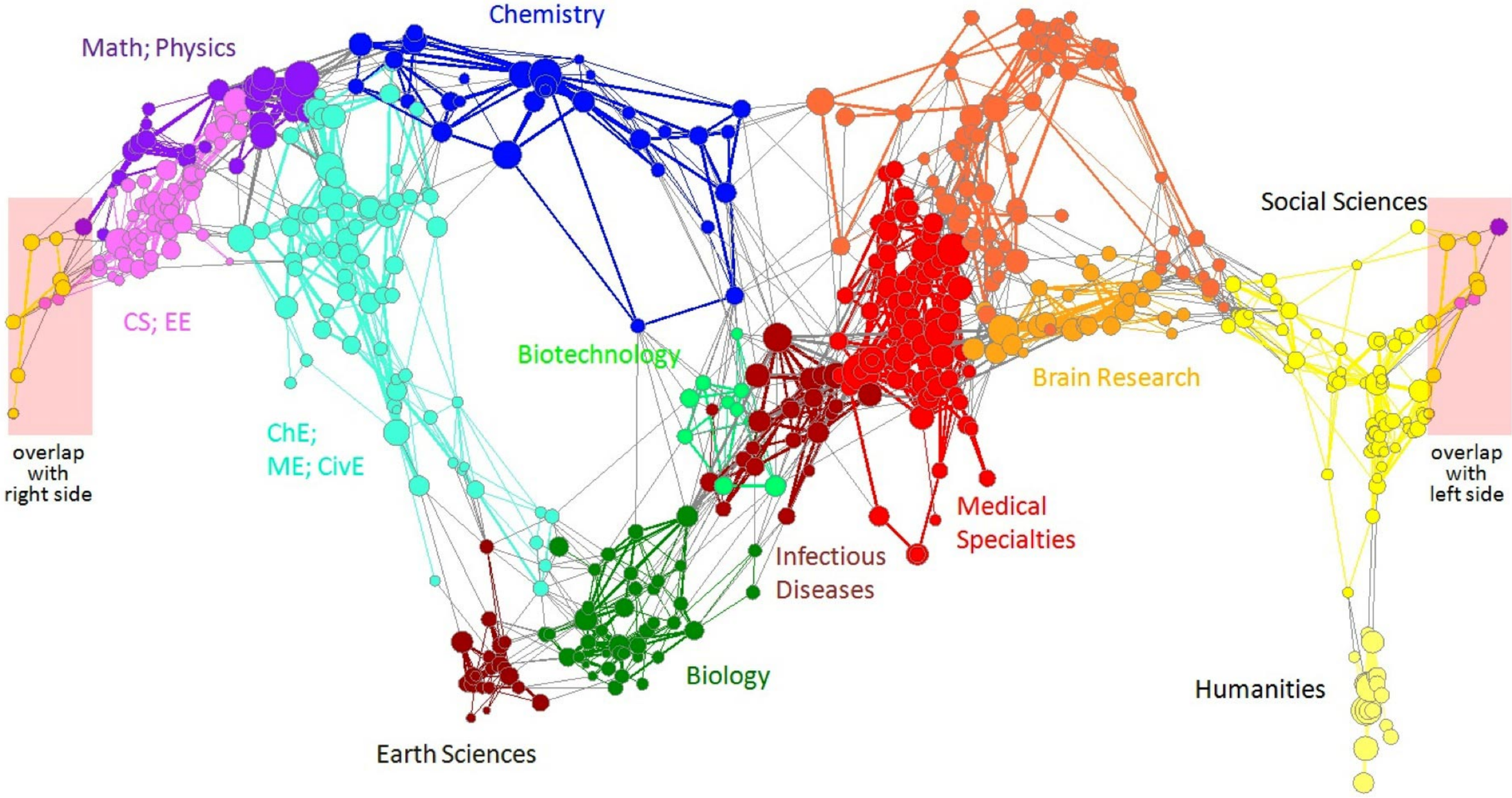

# Global spread of COVID-19

# Power is transported everywhere with interconnected stations and lines



# The power grid: a network

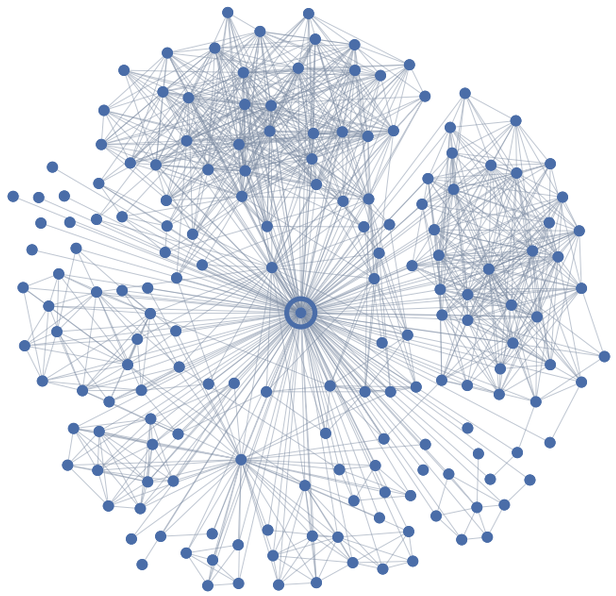
**Science is a complex system of academics working together and being influenced by each other**



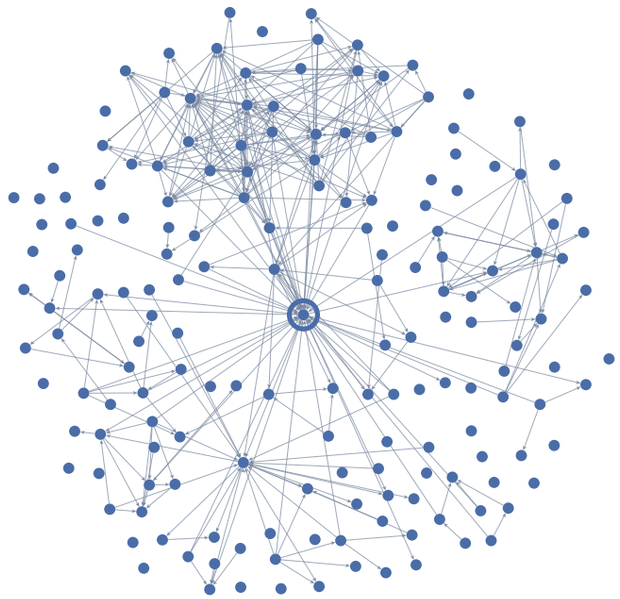

# Citation networks and maps of science [Börner et al., 2012]

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

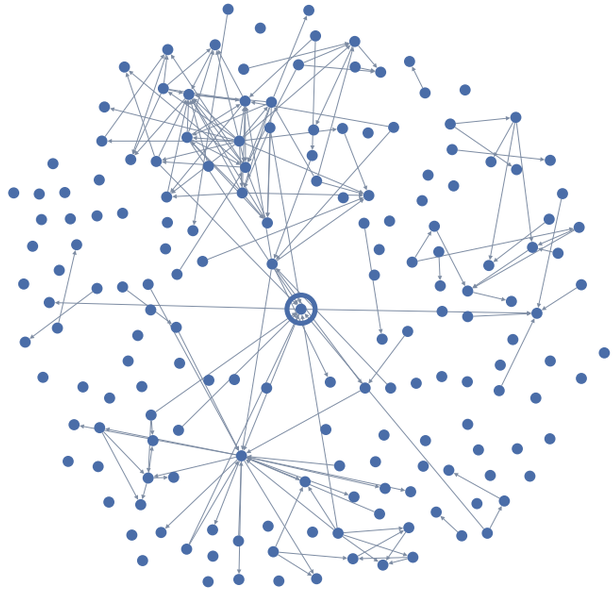
All Friends



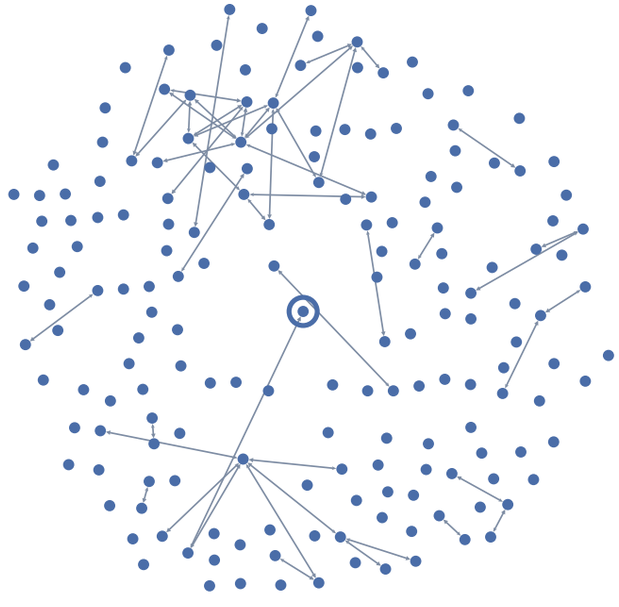
Maintained Relationships



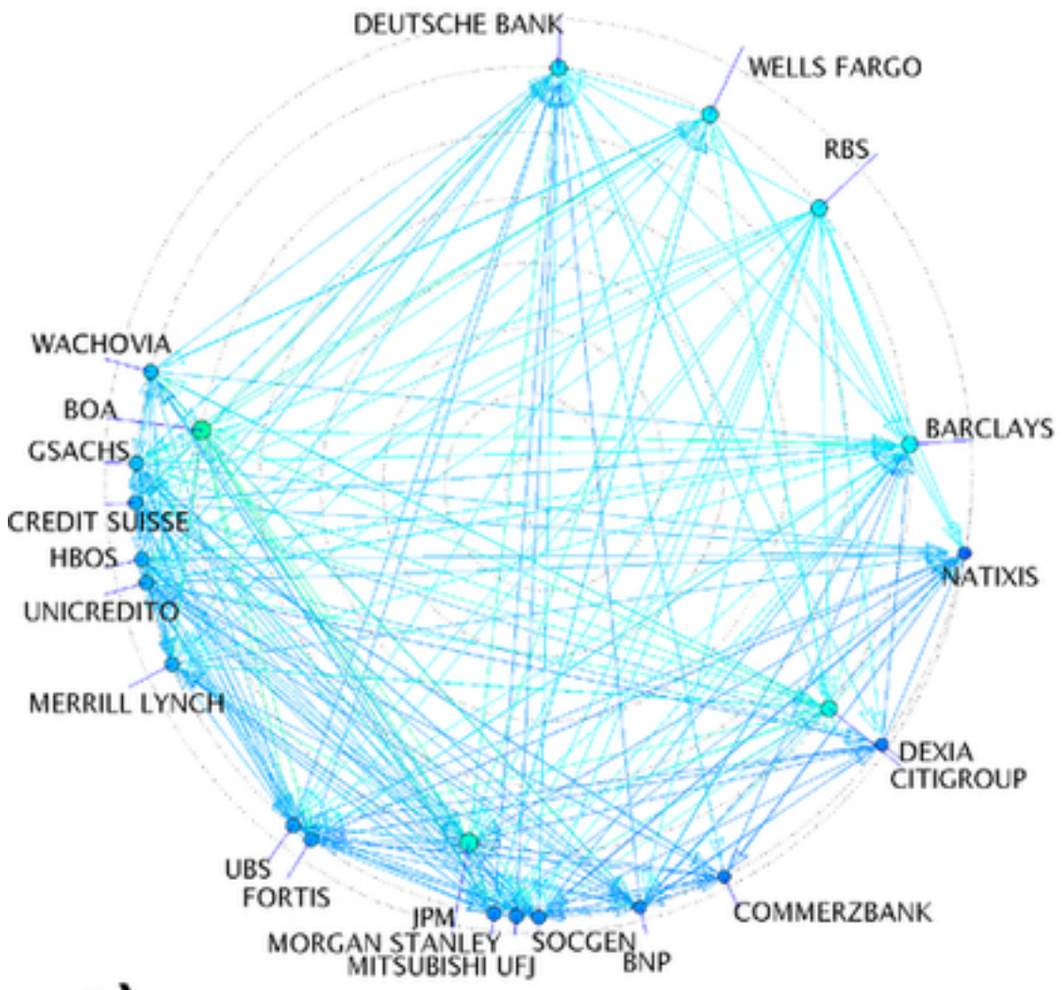
One-way Communication

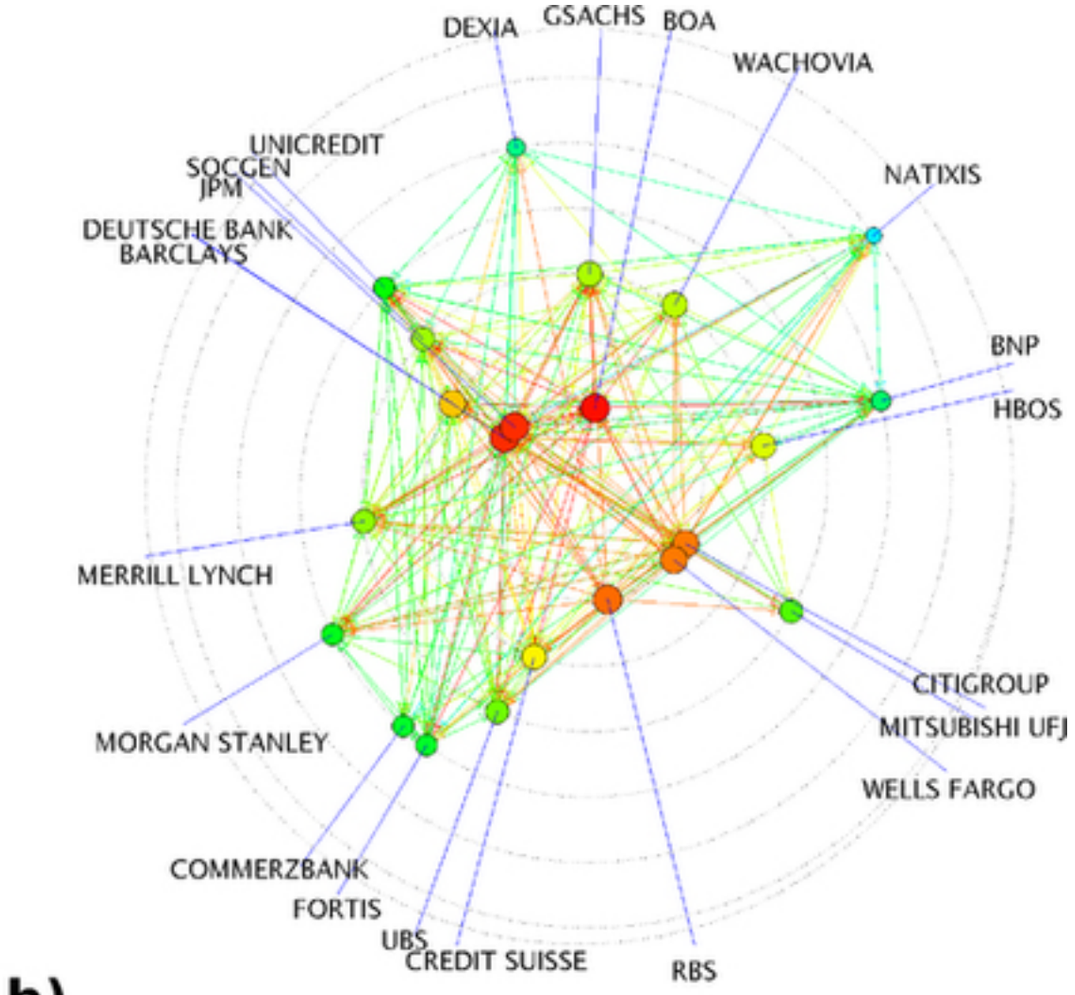


Mutual Communication

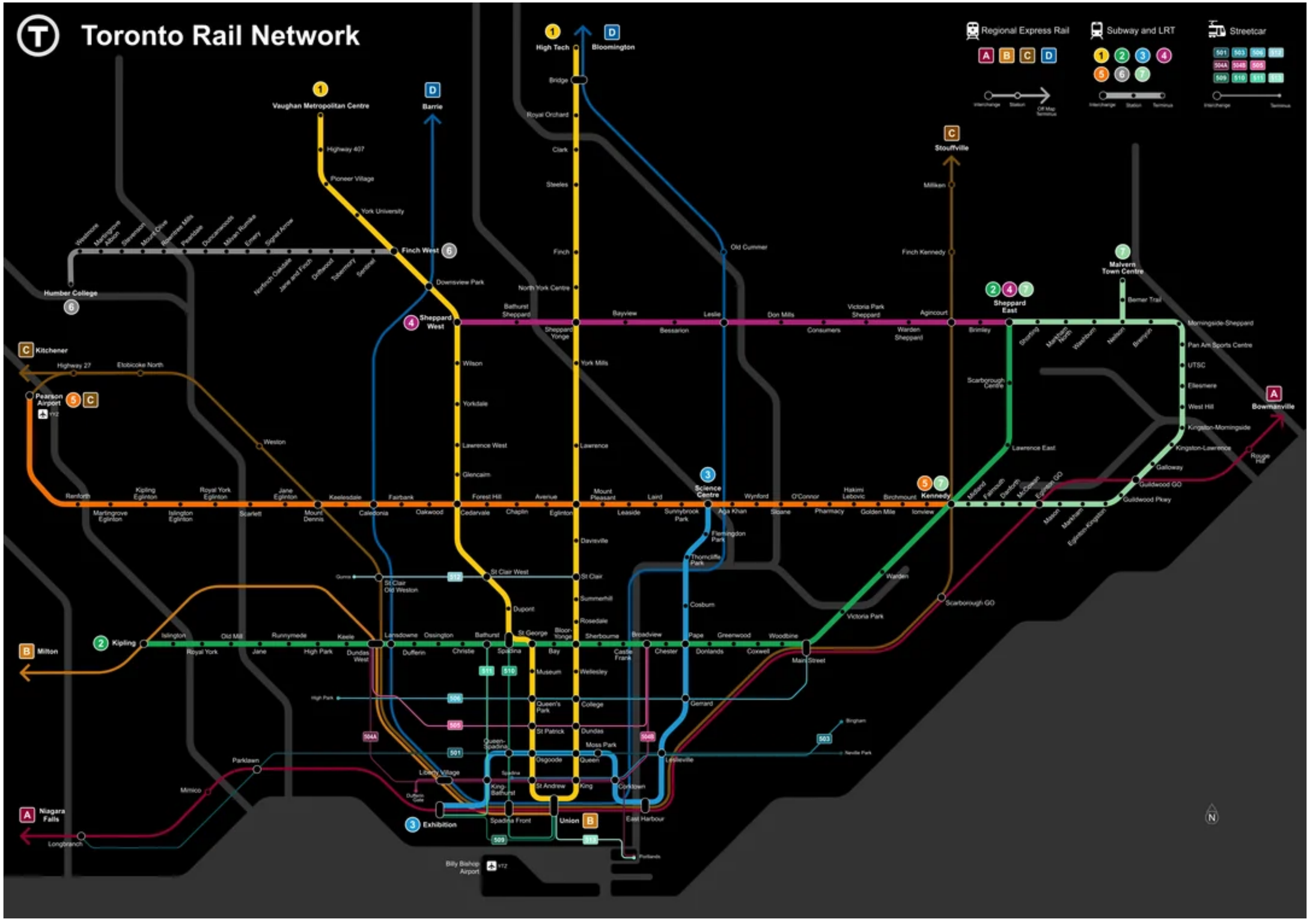


# An “ego network”: the neighbourhood around a single individual

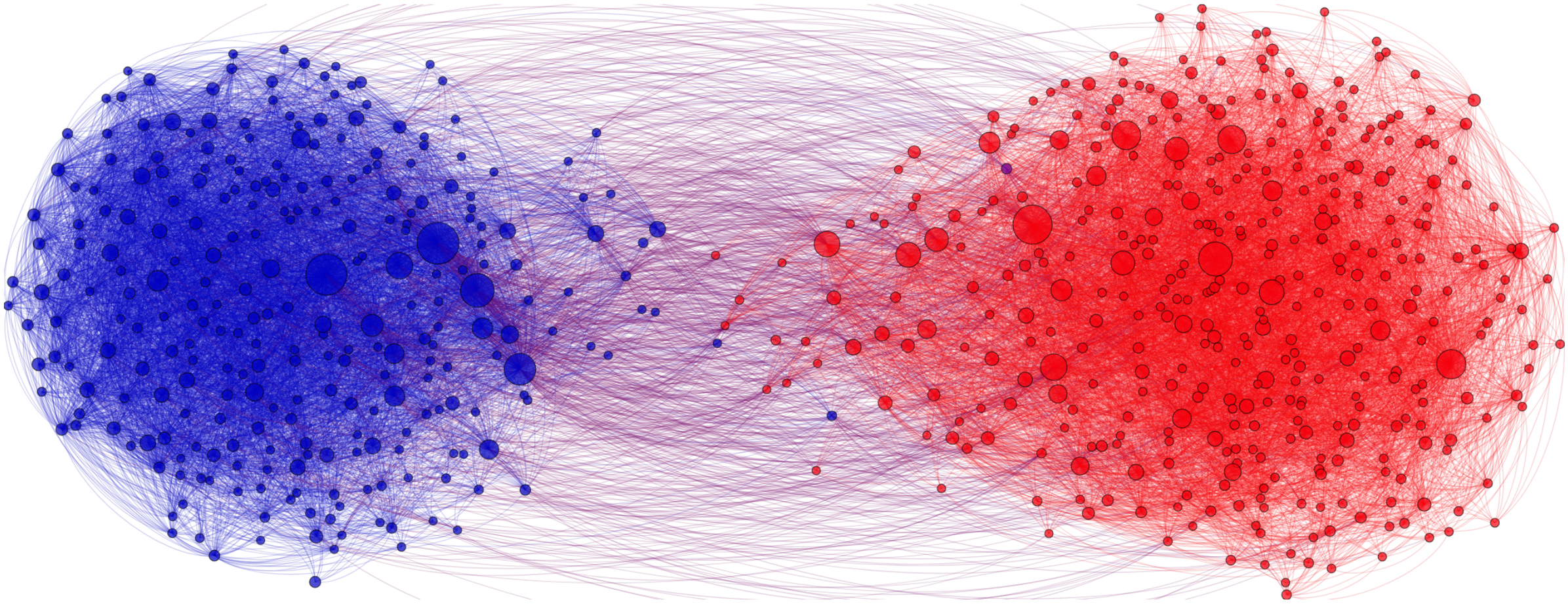



# The economy is a network: e.g. Bank lending



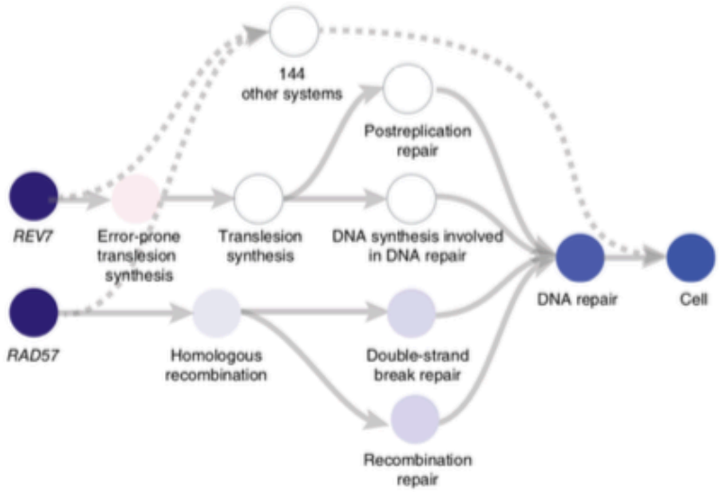

# Transportation network



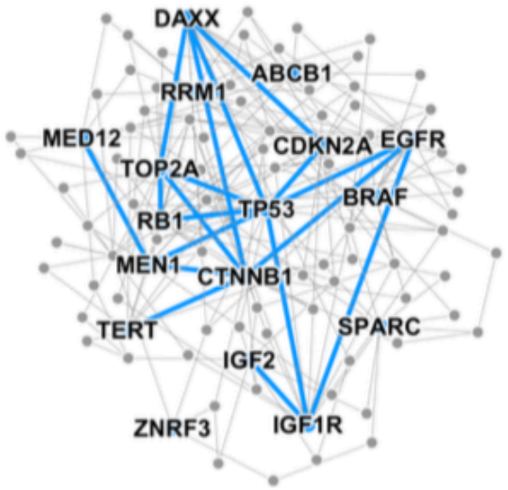
# Political blogs prior to 2004 US Presidential election

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

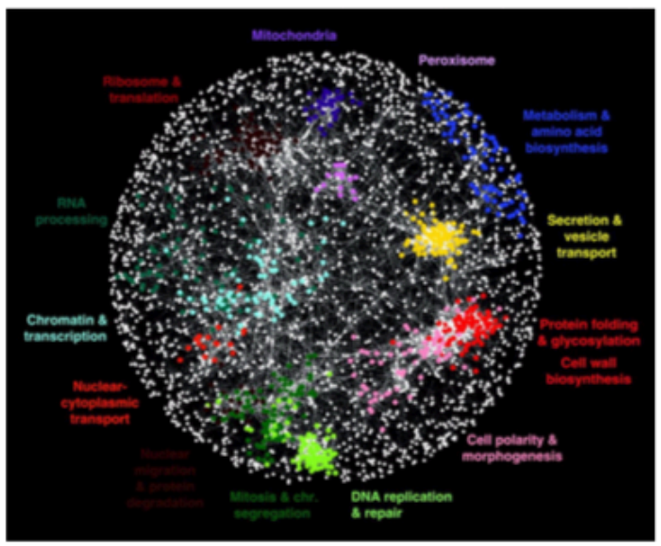
Patient networks



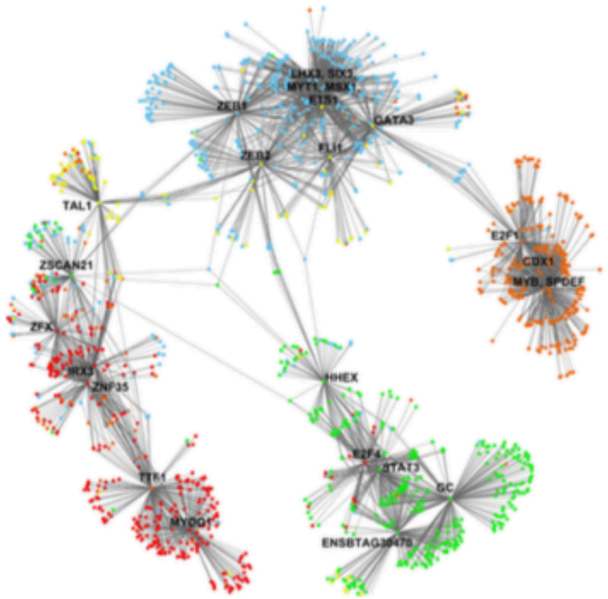

Hierarchies of cell systems




Disease pathways




Genetic interaction networks




Gene co-expression networks

Cell-cell similarity networks

**Many, many more examples**

# But why should *I* care about networks?

# Why study networks?

**Complex systems are all around us**

- **Society** is a collection of eight 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 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 now?


## Age and size of networks

# Networks: Impact


**Google/Alphabet**
Dominant tech giant

**Facebook/Meta**
Dominant social platforms

# Networks and Applications

# Ways to Analyze Networks

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

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

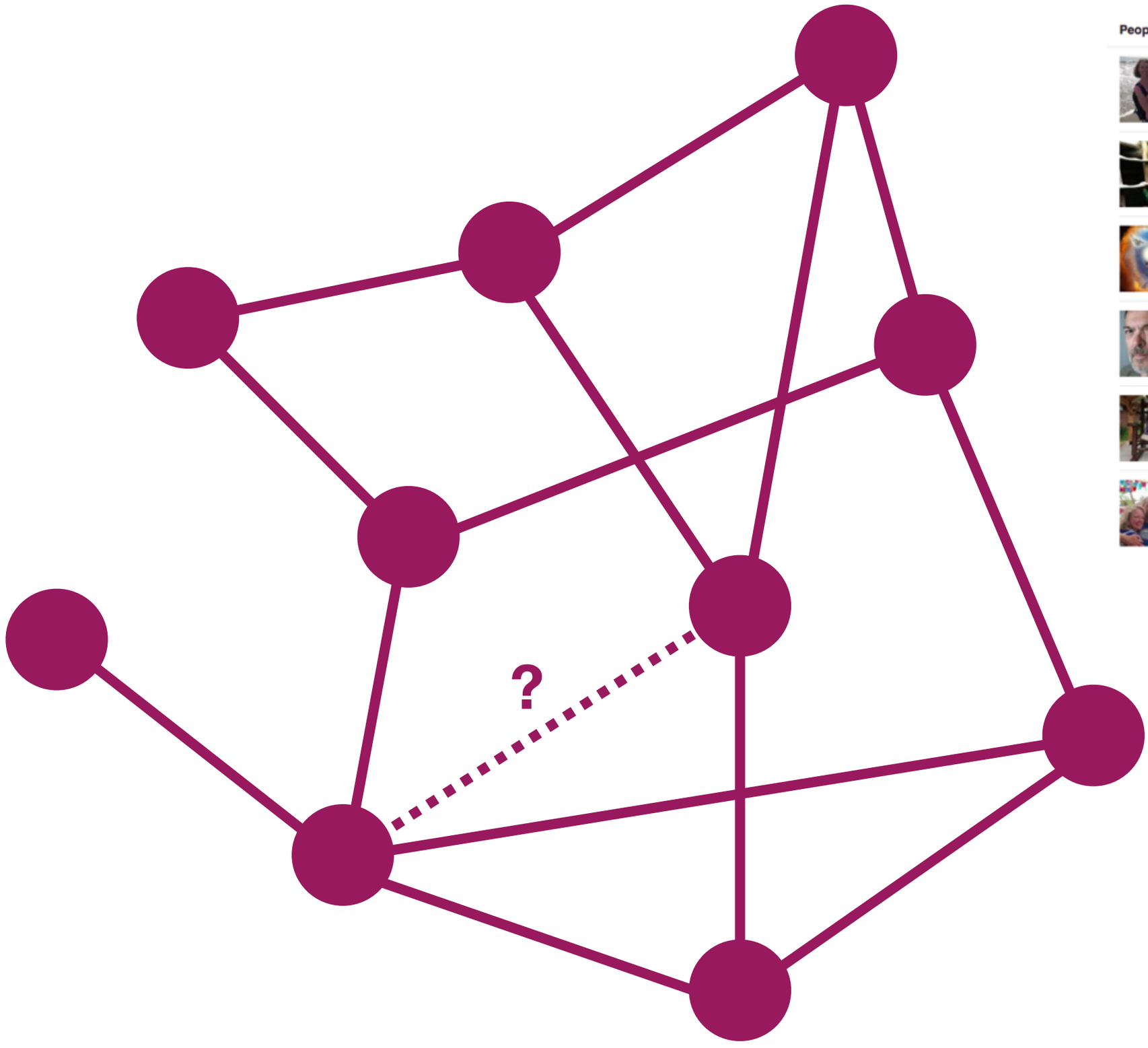
**Identify** densely linked clusters of nodes (*community detection*)

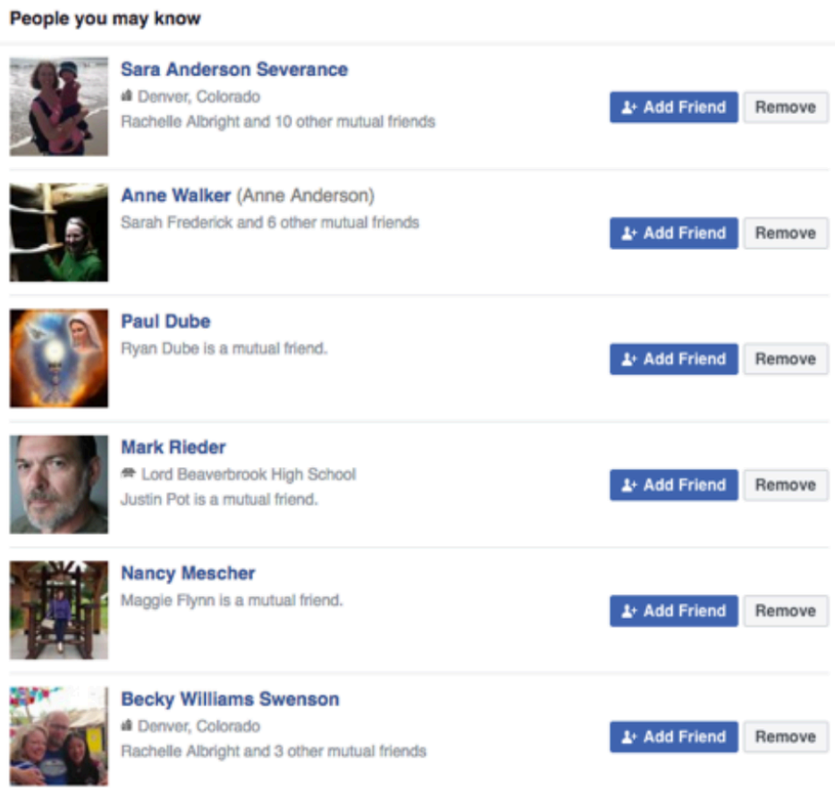
**Predict** common pathways (*social influence/propagation*)

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

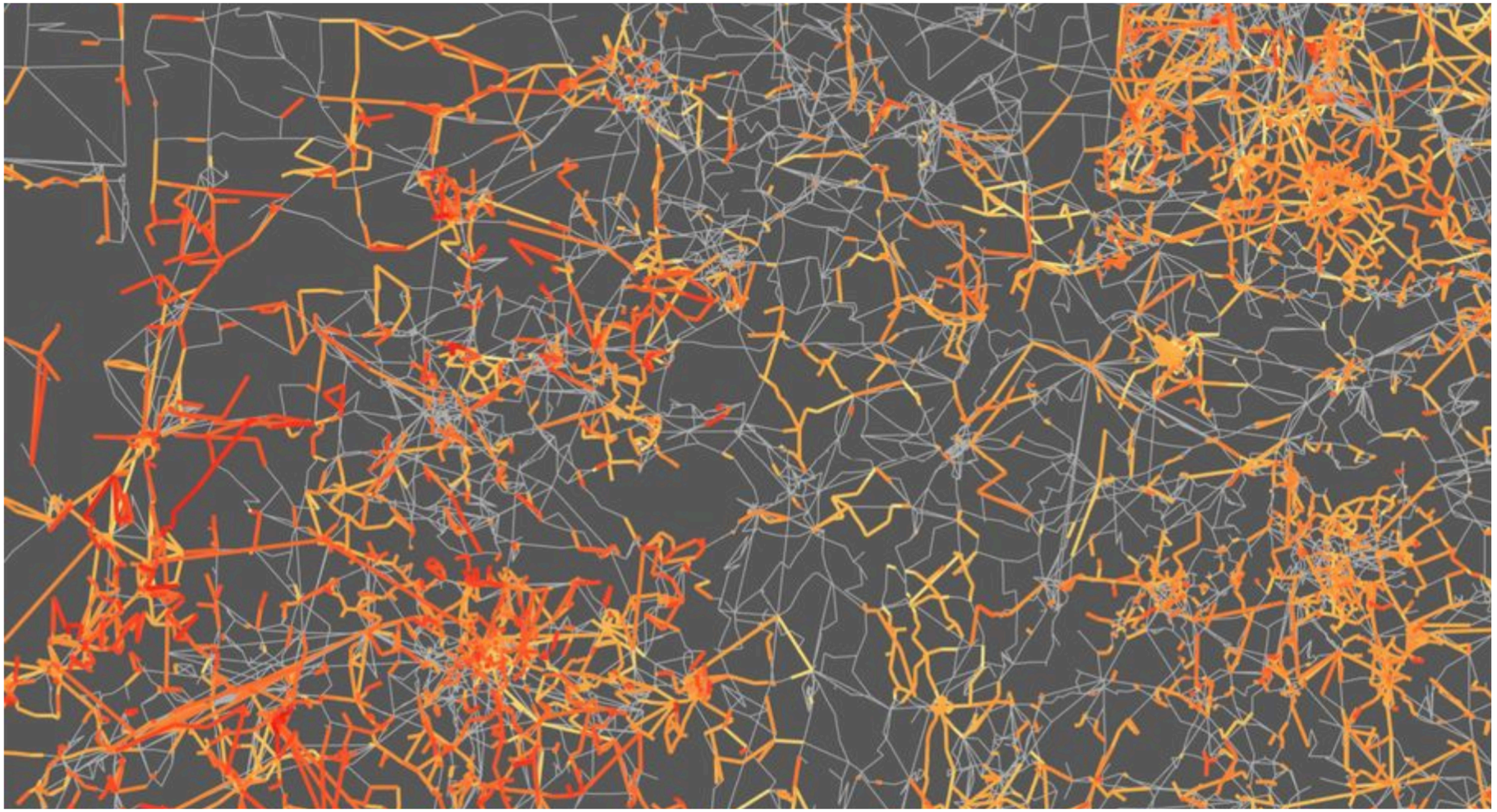
# (1) Networks: Social

# Application: Friend Prediction



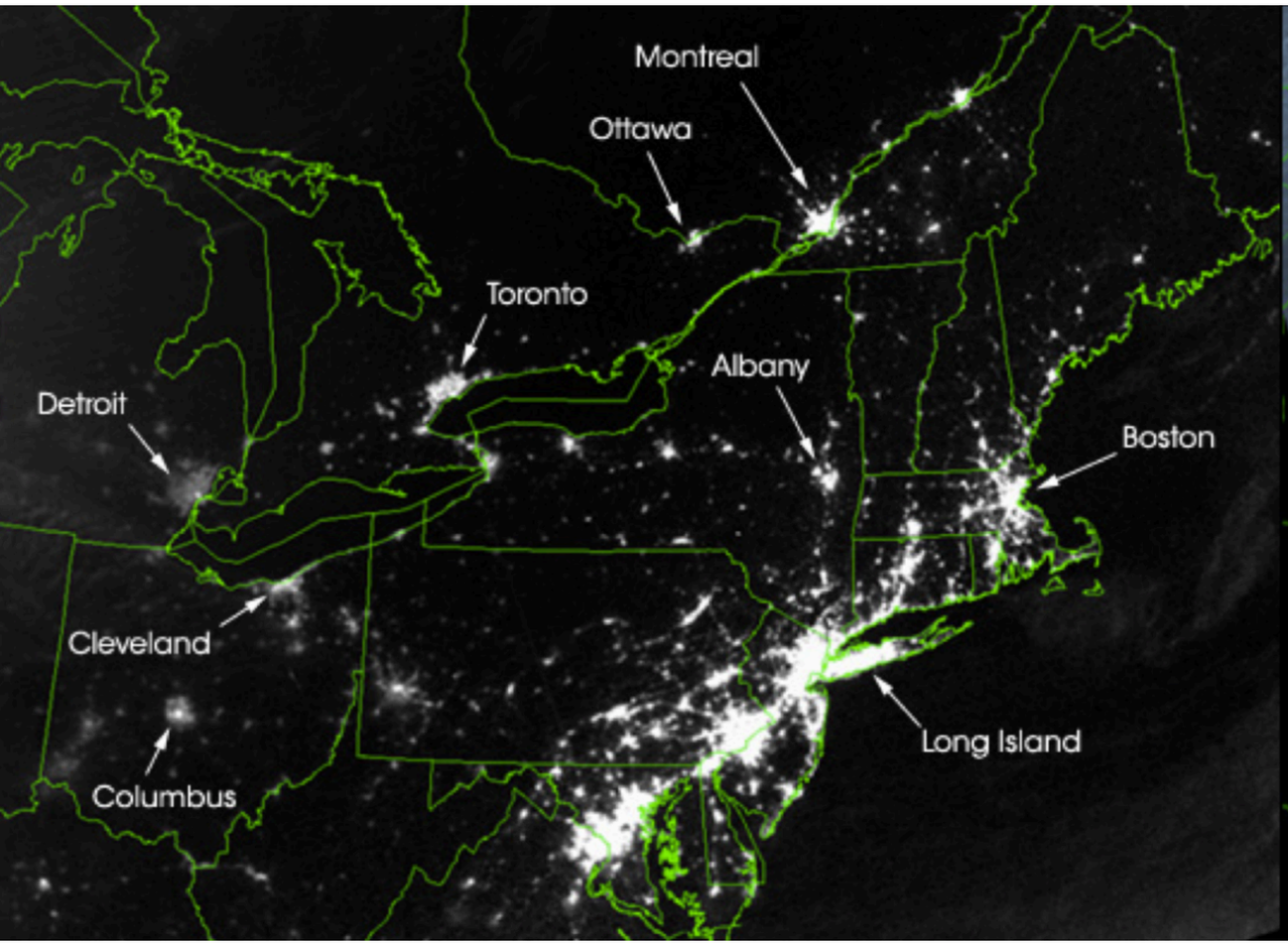



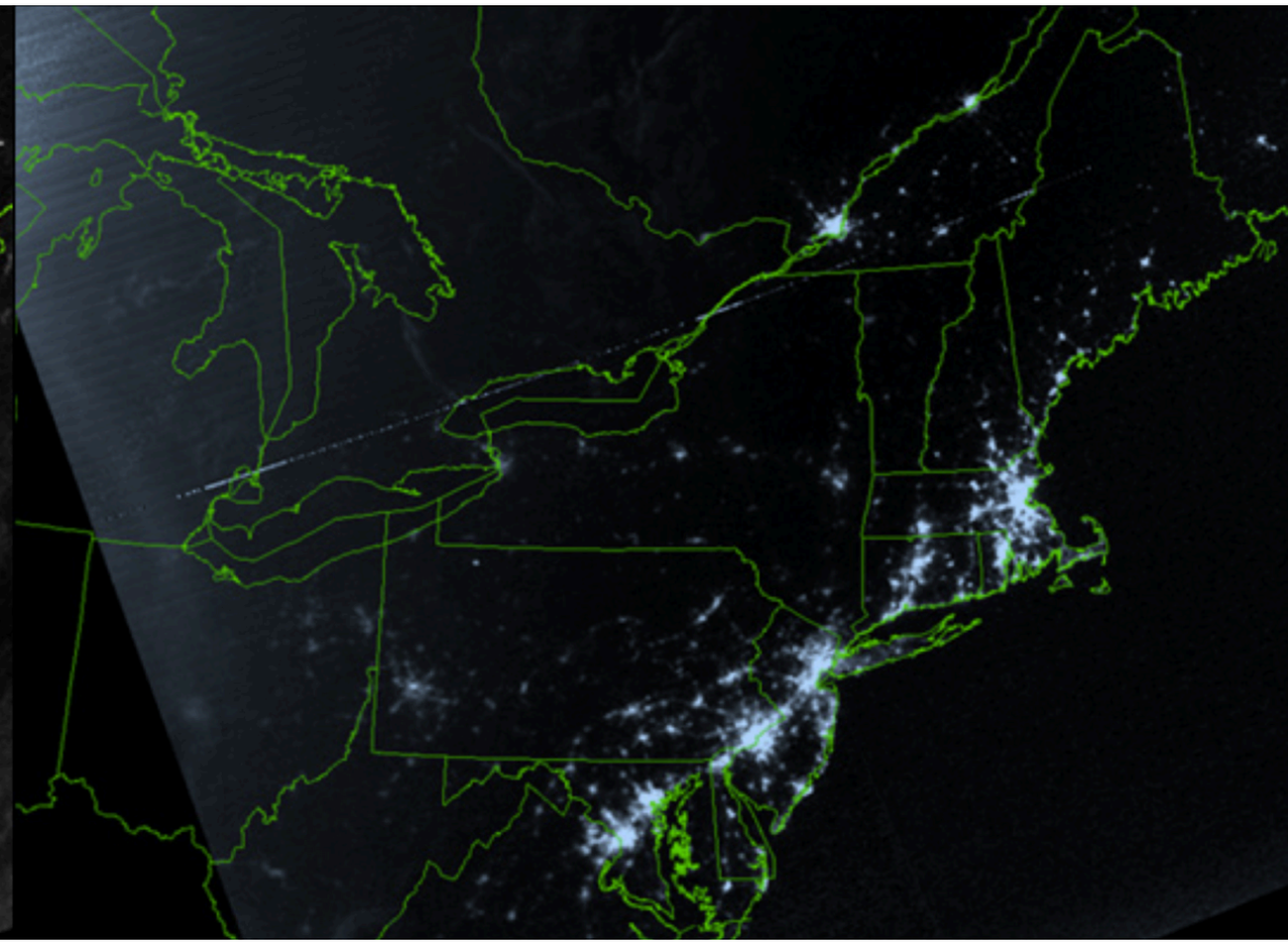
# (2) Networks: Infrastructure



**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

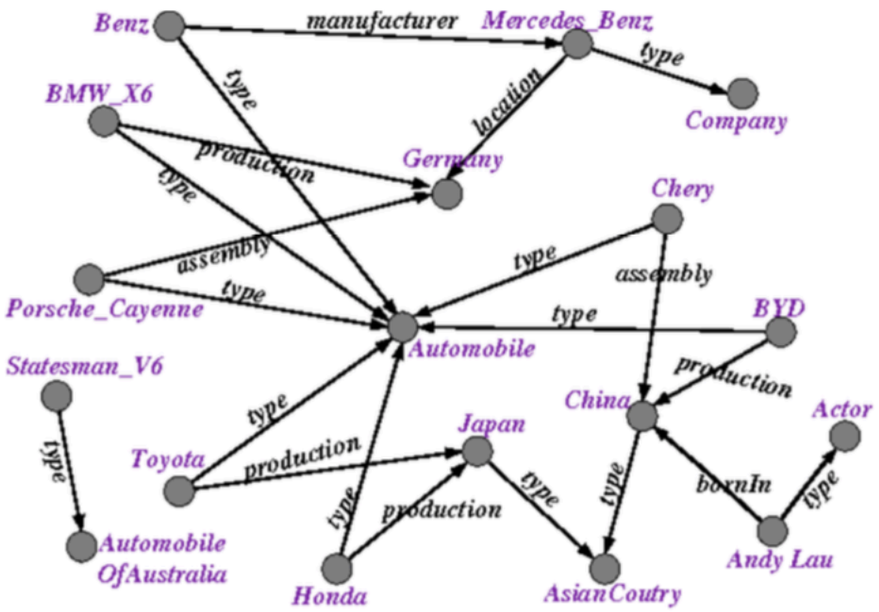
## August 2003 blackout

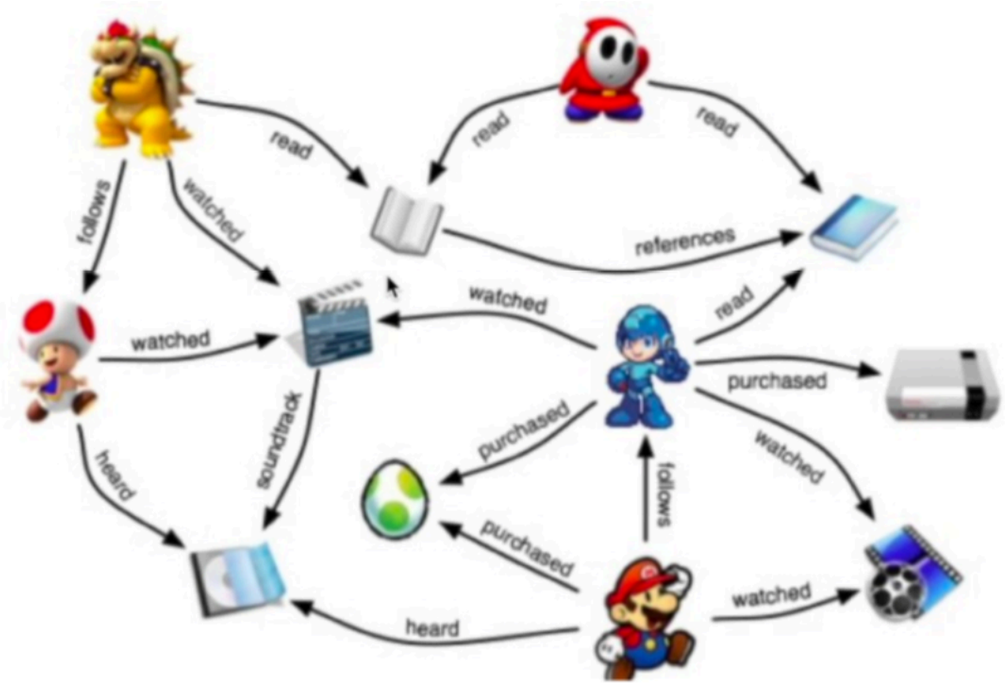
# (2) Networks: Infrastructure

**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

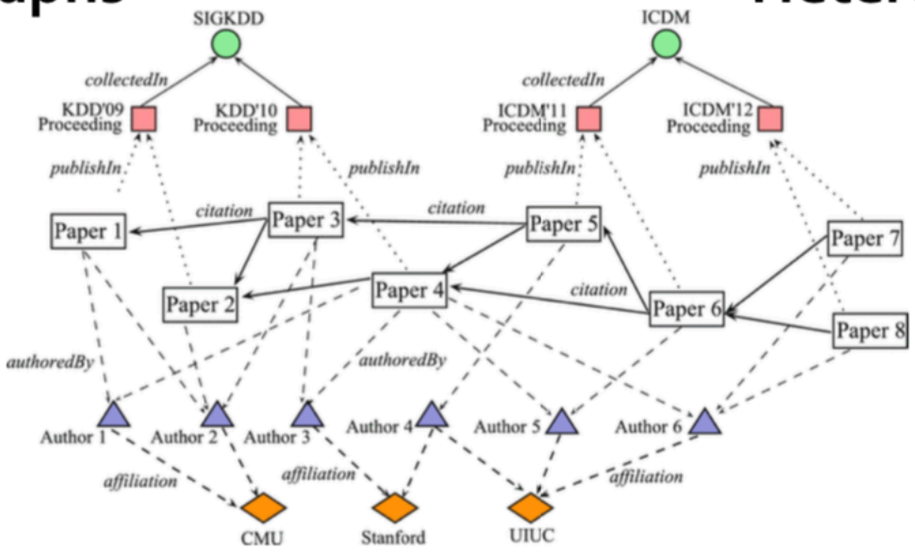
# (3) Networks: Information and Knowledge




## Knowledge Graphs

## Heterogeneous Graphs

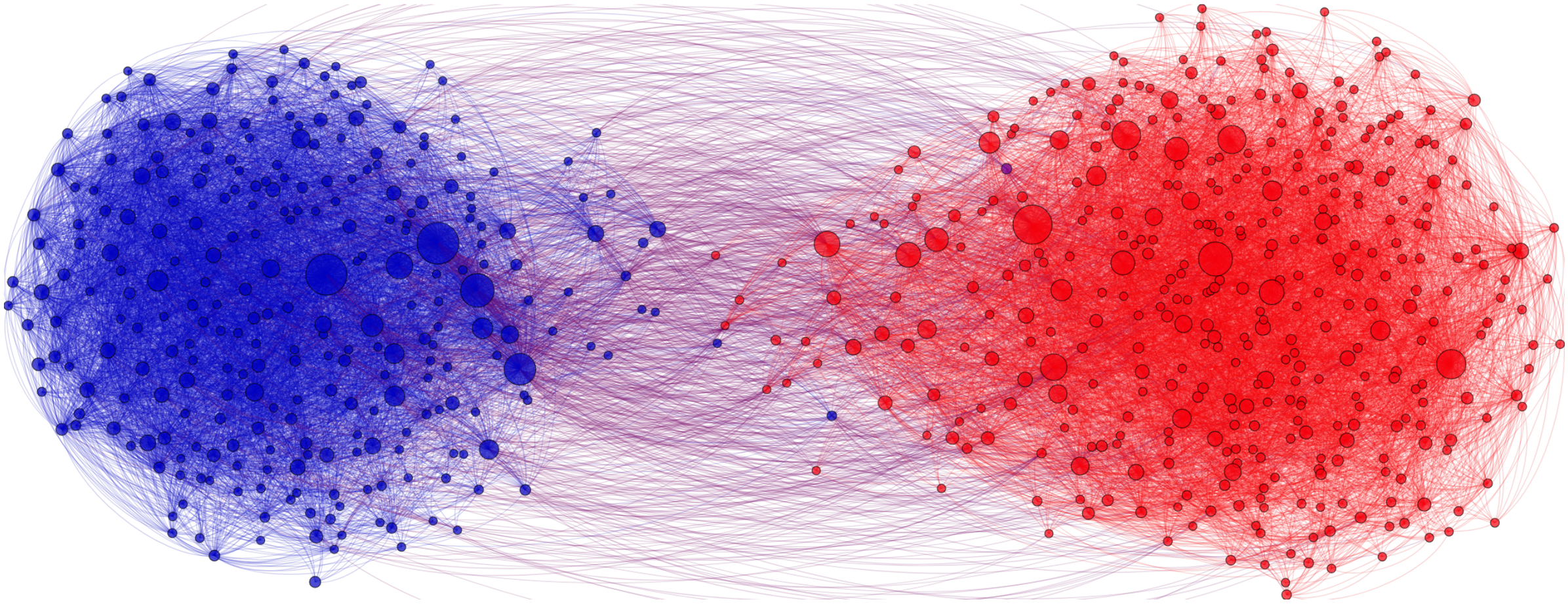



## Multimodal Graphs

# Application: Web Search

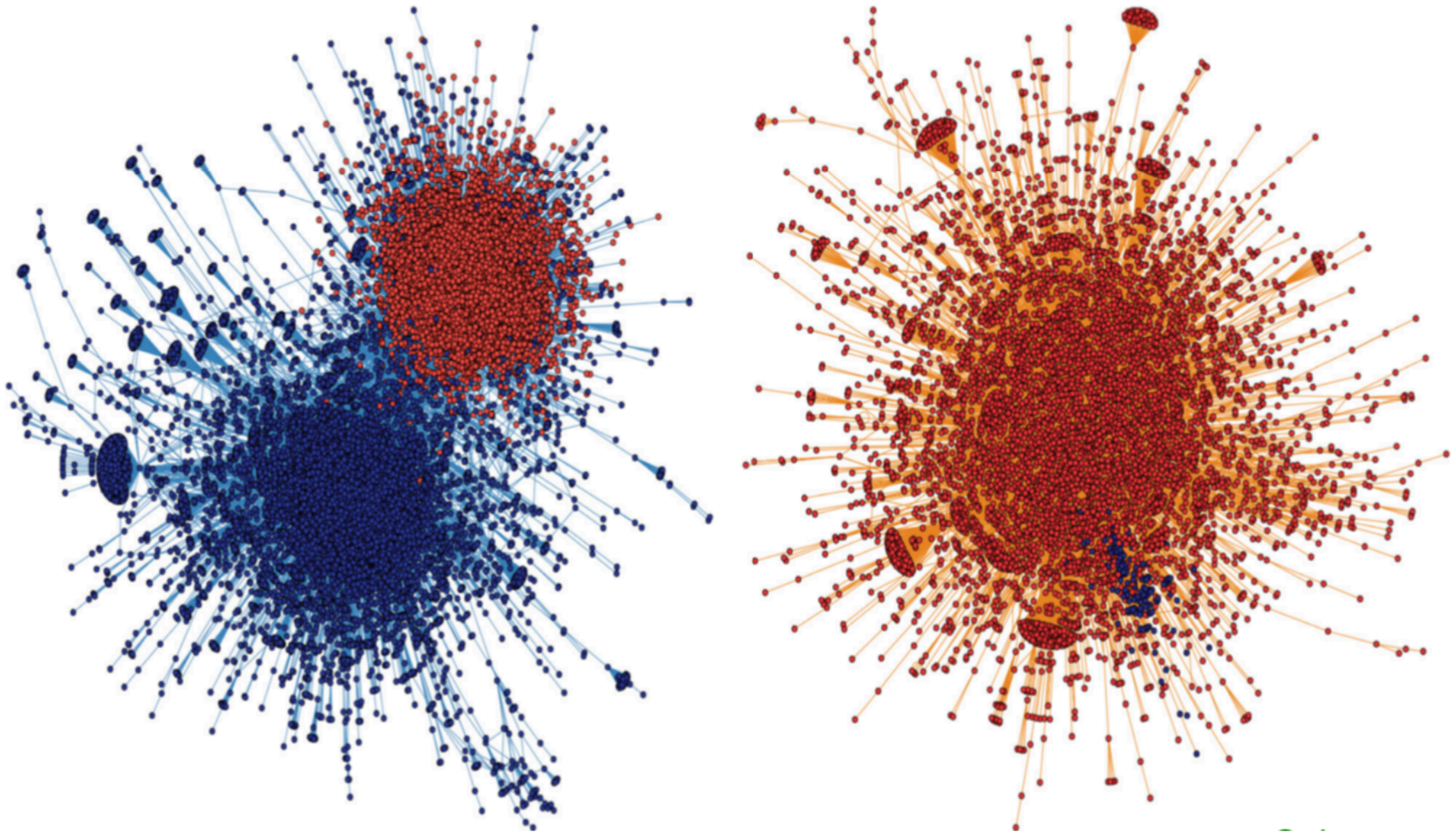

**How do you go from a tiny text string to the 10 most relevant sites out of billions of pages?**

# (4) Networks: Online Media



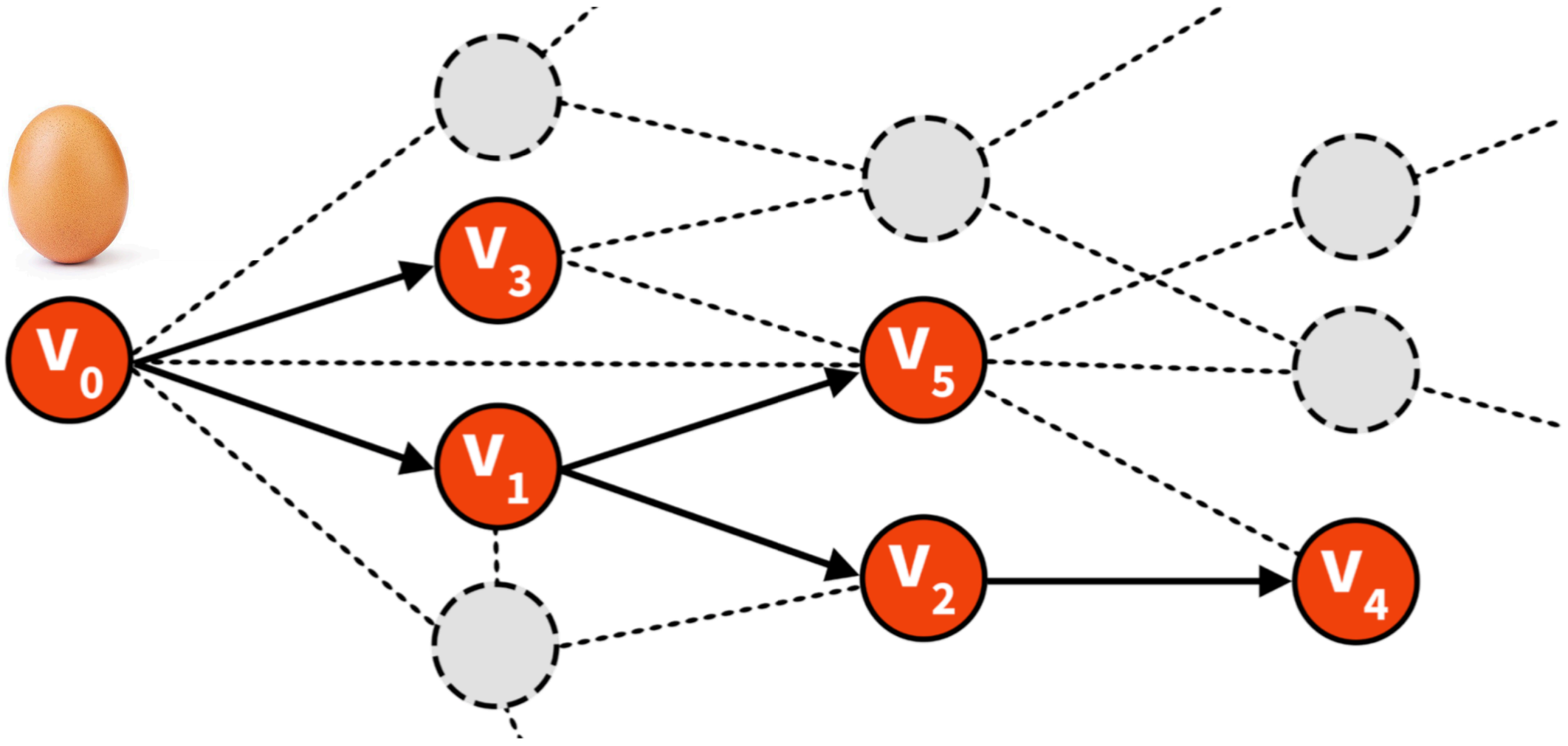
Connections between political blogs

# Application: Polarization on Twitter



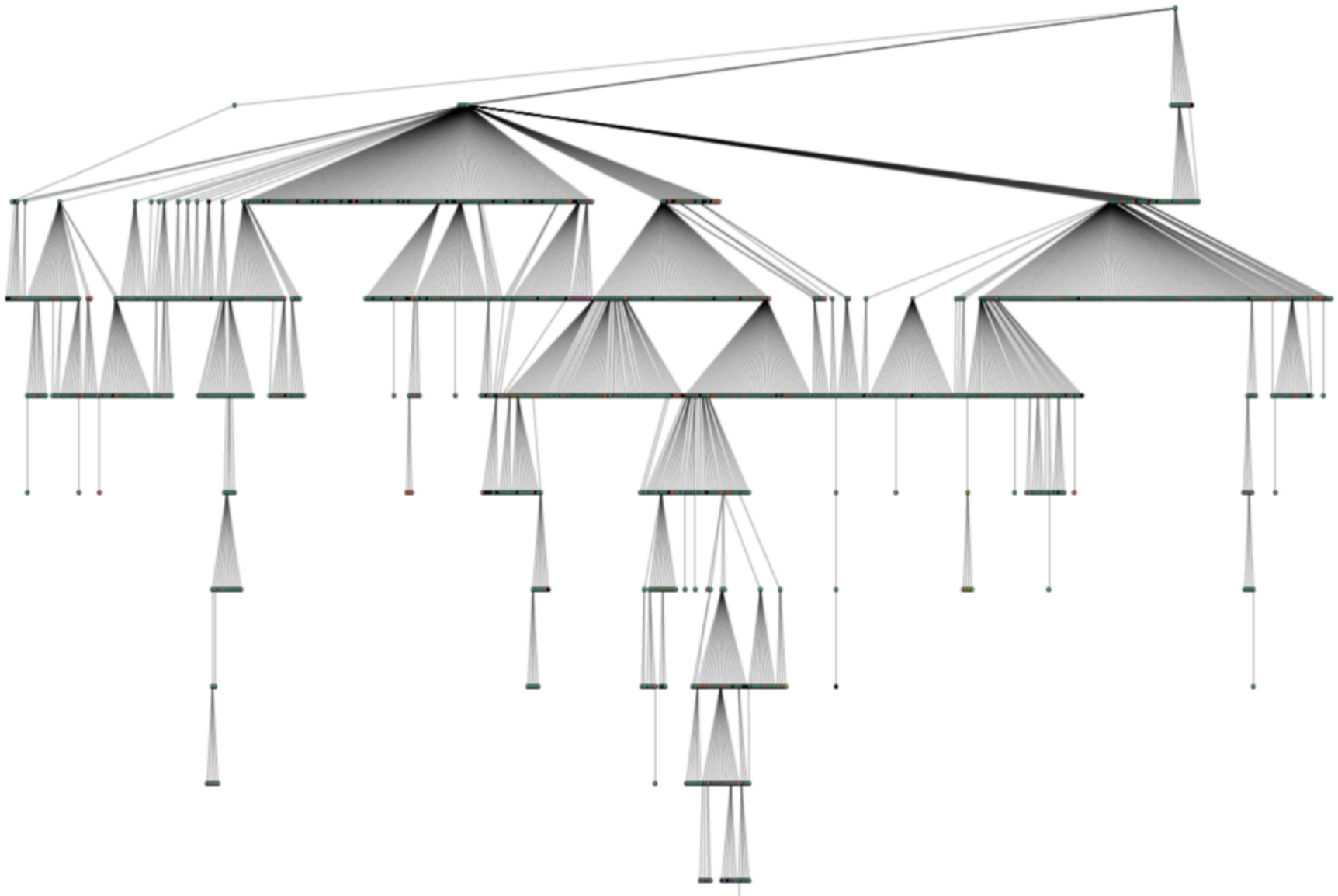
Retweet networks: polarized (left) and unpolarized (right)

# 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

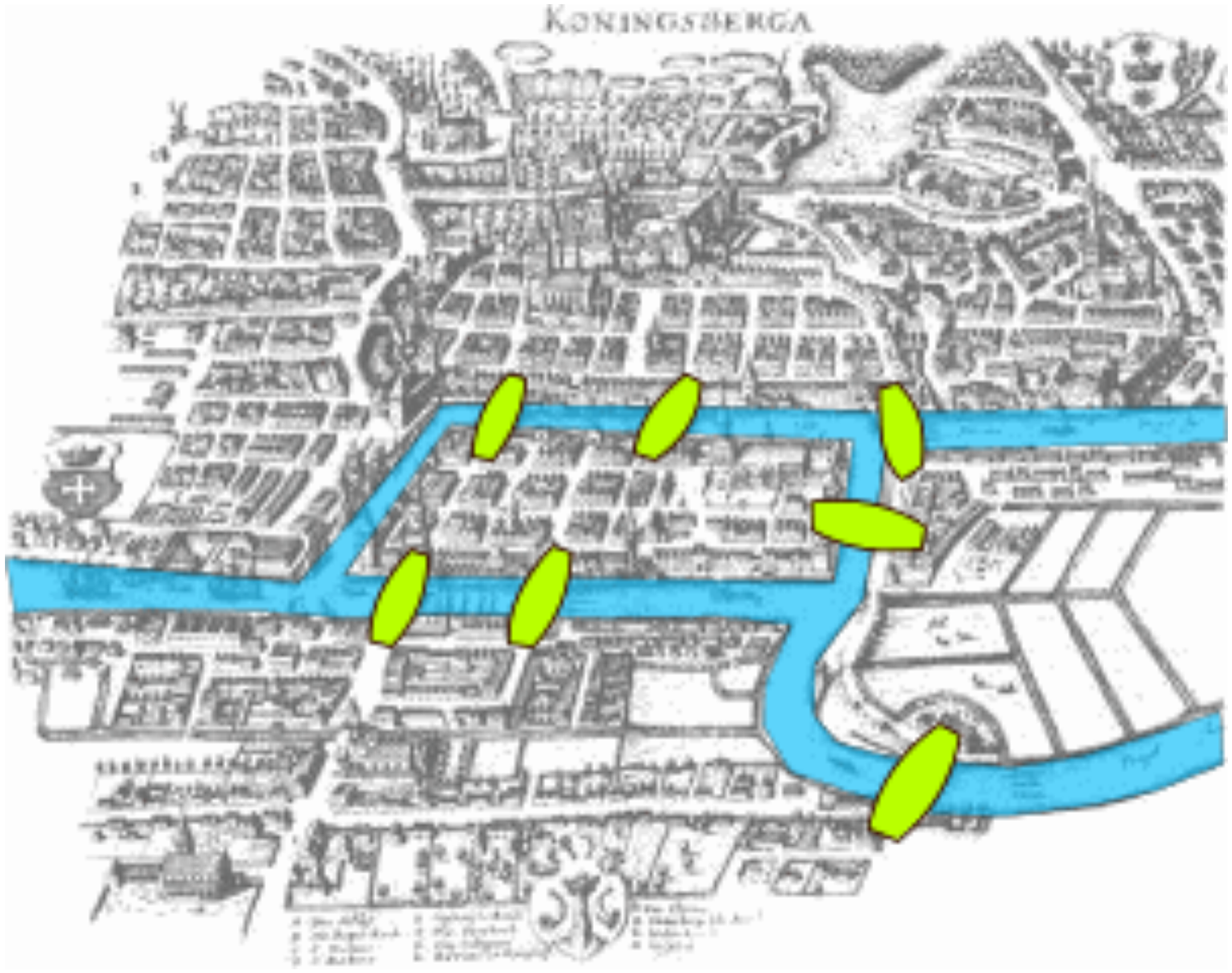
How do information and diseases spread?

### Incentives in networks

Behaviour is interconnected by depending on what others do

How do decisions and behaviours depend on network structure and dynamics?

# 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

**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)

# Course evaluation

25% 3 assignments

20% Midterm

50% Final exam

5% Participation/Collegiality

Assignments are due on Thursdays at 5pm

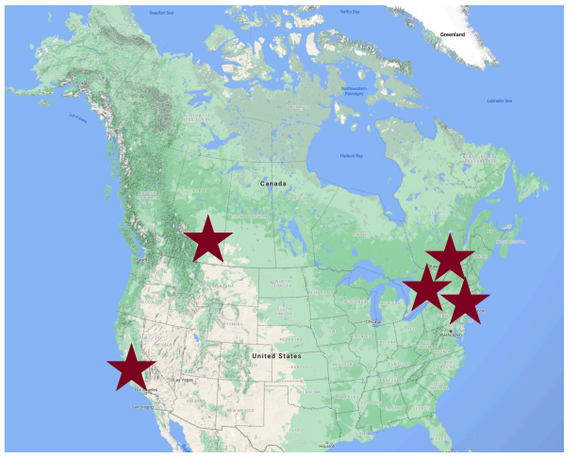
Because I understand sometimes stuff happens, you get 3 "flex days": 1 flex day is a 24-hour period that you can hand in assignments late with no penalty. After that, no late assignments will be accepted.

# About Me

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

1M → 4M → 7M → 20M → 6M

Now: Associate Professor of Computer Science at U of T




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

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

# My path

| Stage | Interests |
| --- | --- |
| McGill<br>B.Soft.Eng ‘08 | Theoretical<br>Quantum algorithms and information<br>Anything practical was impure |
| Stanford | “Hmm…would be nice to feel more connected to the world”<br>Game theory: computational/economic lens on |
| Stanford Ph.D. | Discovered the joy and power of large-scale empirical analysis<br>Computational social science: social research in the digital age<br>Mostly empirical analysis supplemented with theoretical modeling, experimentation, and surveys |

# My research

## Artificial Intelligence

Study algorithms
Design algorithms
Algorithmic effects

## Data

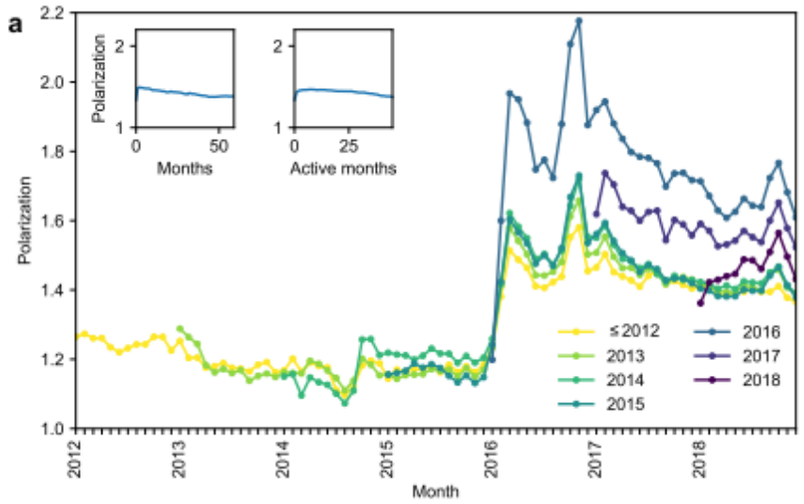
Large online data
Often behavioural

**CSS**

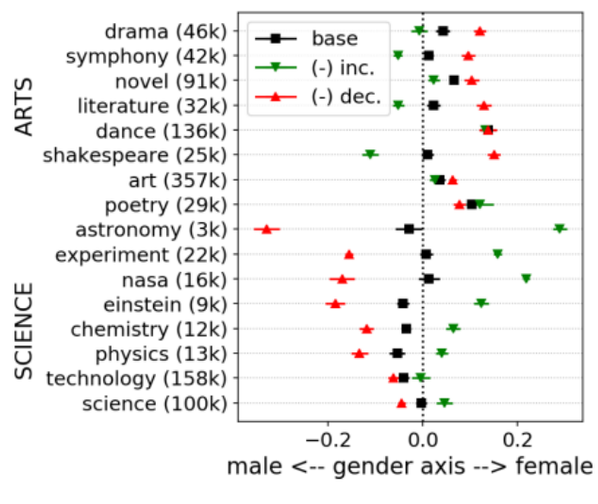
Important societal questions
Polarization, echo chambers, bias,
info diffusion, social media impact

## Society

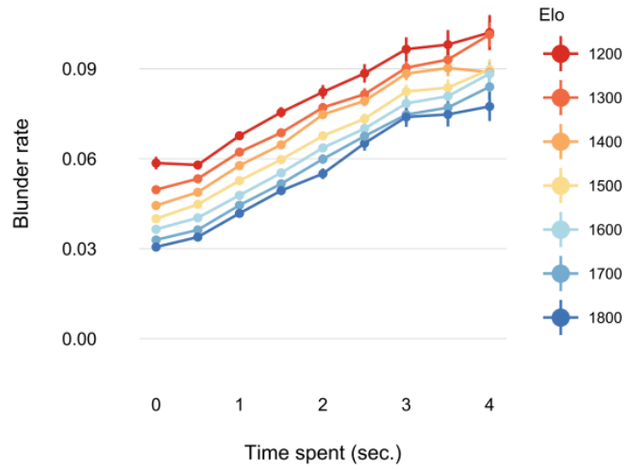
# My research



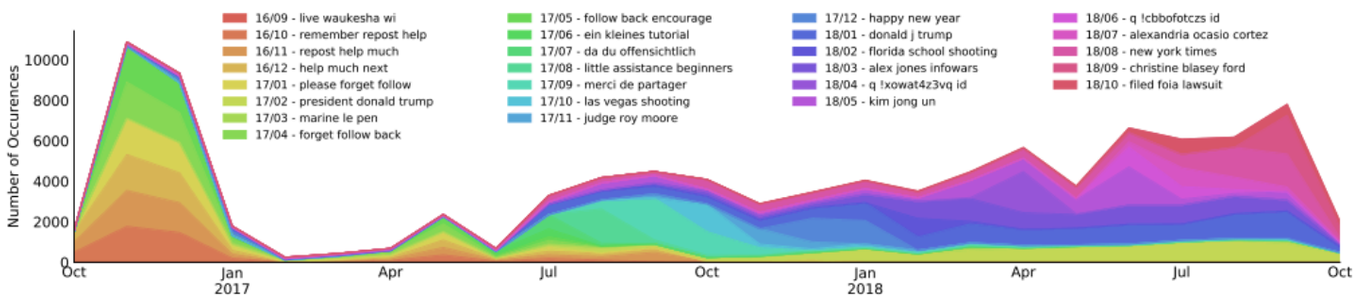

Political polarization on Reddit



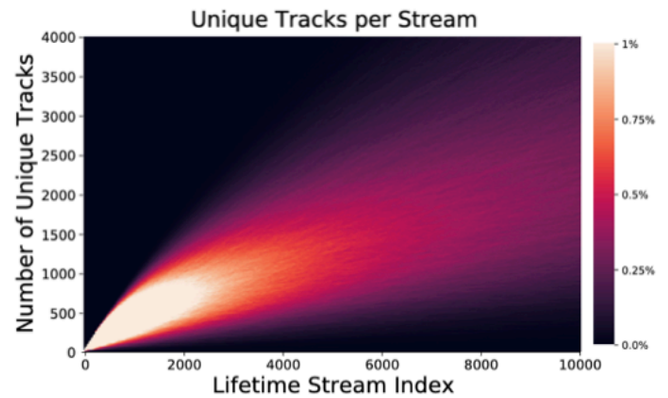

Gender bias in text algorithms




Nature of human error in chess




Discussion topics on Gab (alt-right platform)




Music exploration on Spotify

# Tutorials

- Time: Mondays 12–1 and 1–2
- Required
- Mostly working through concrete examples as a group, Q&A (group office hours), and assignment help

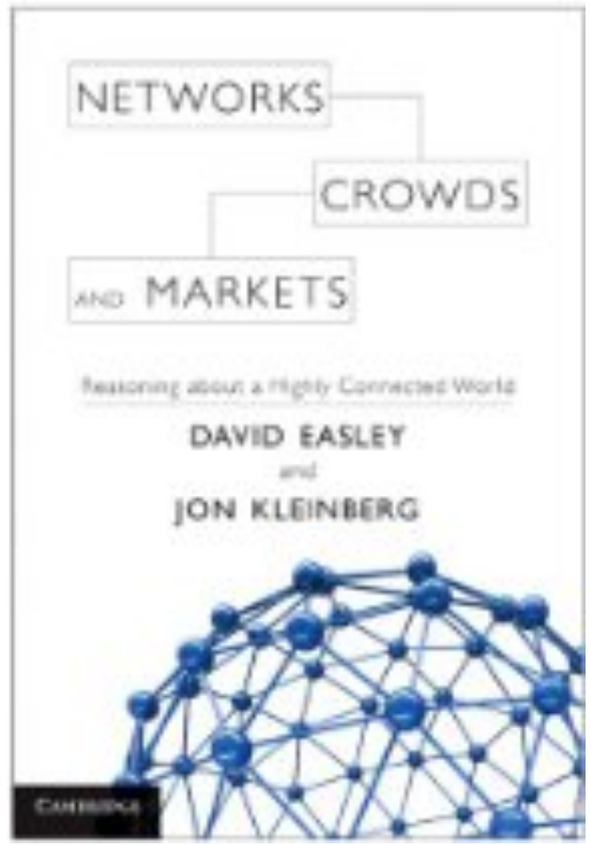
# Textbook

"Networks, Crowds & Markets" Easley & Kleinberg

Available free online / reasonably-priced hardcover

Very readable, engaging text

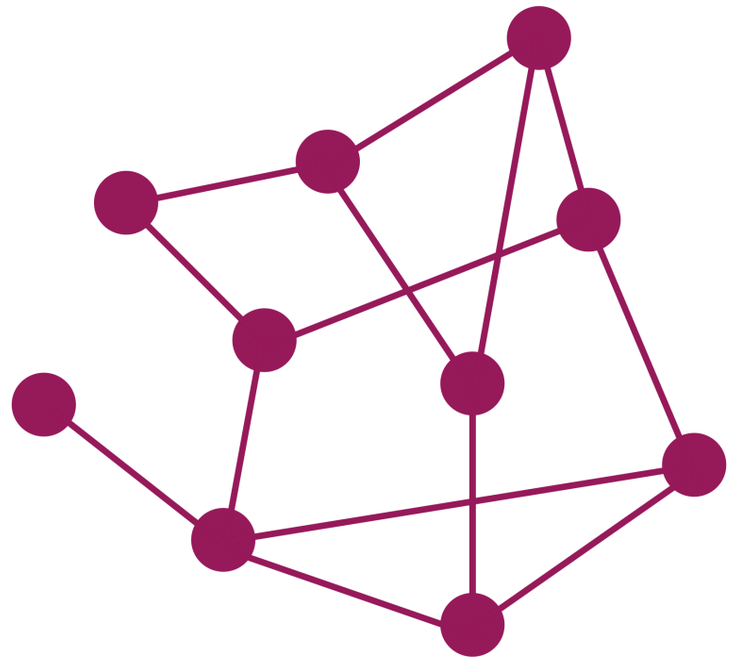
Some assignment exercises from the book



# Questions?

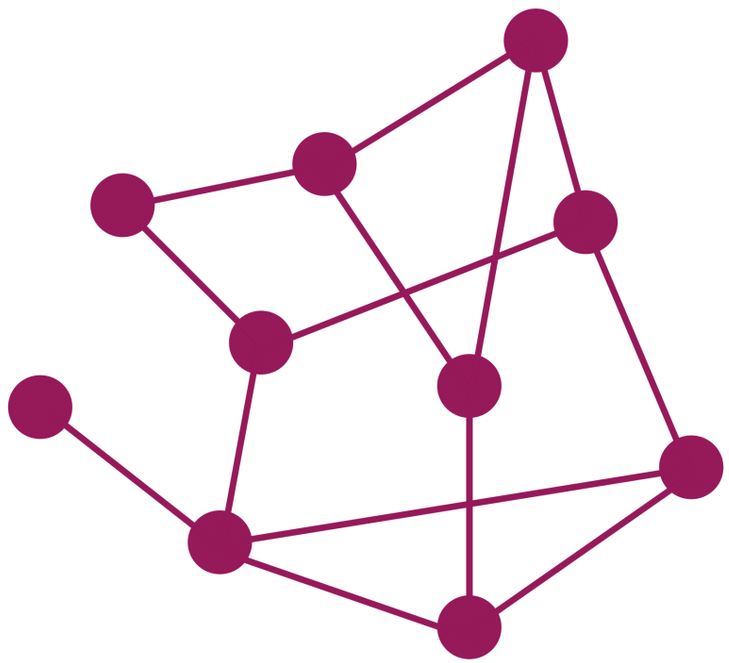
# First topic: Network Analysis Fundamentals and The Structure of the Web

# 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

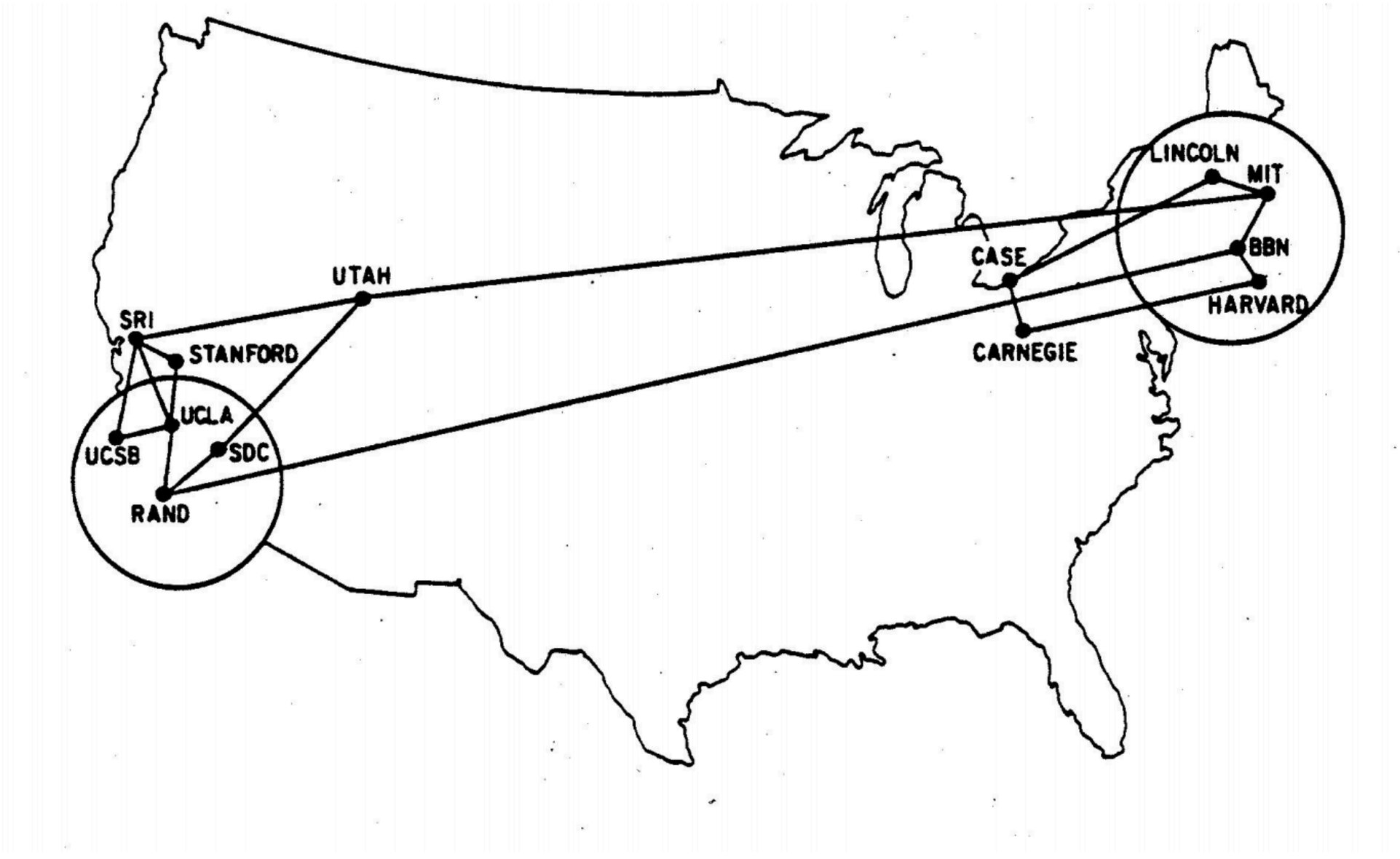
**Graph** is mathematical representation of a network

Web graph, Social graph (a Facebook term)

**Language:** Graph, vertex, edge

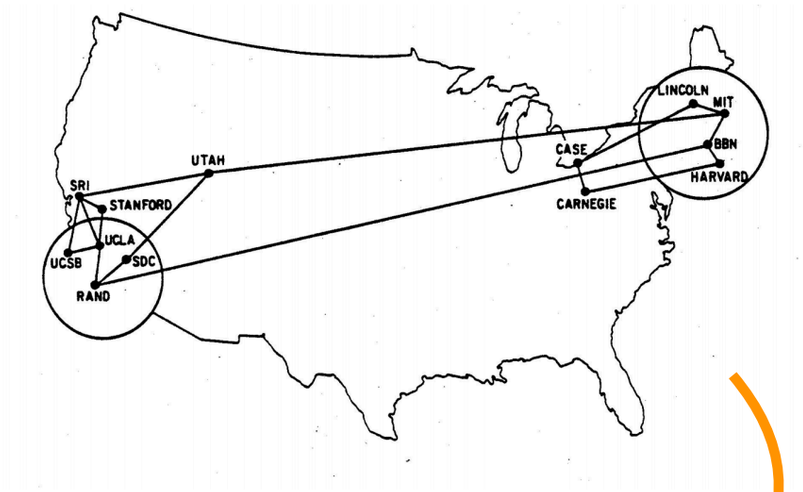
We will try to make this distinction whenever it is appropriate

# A first example

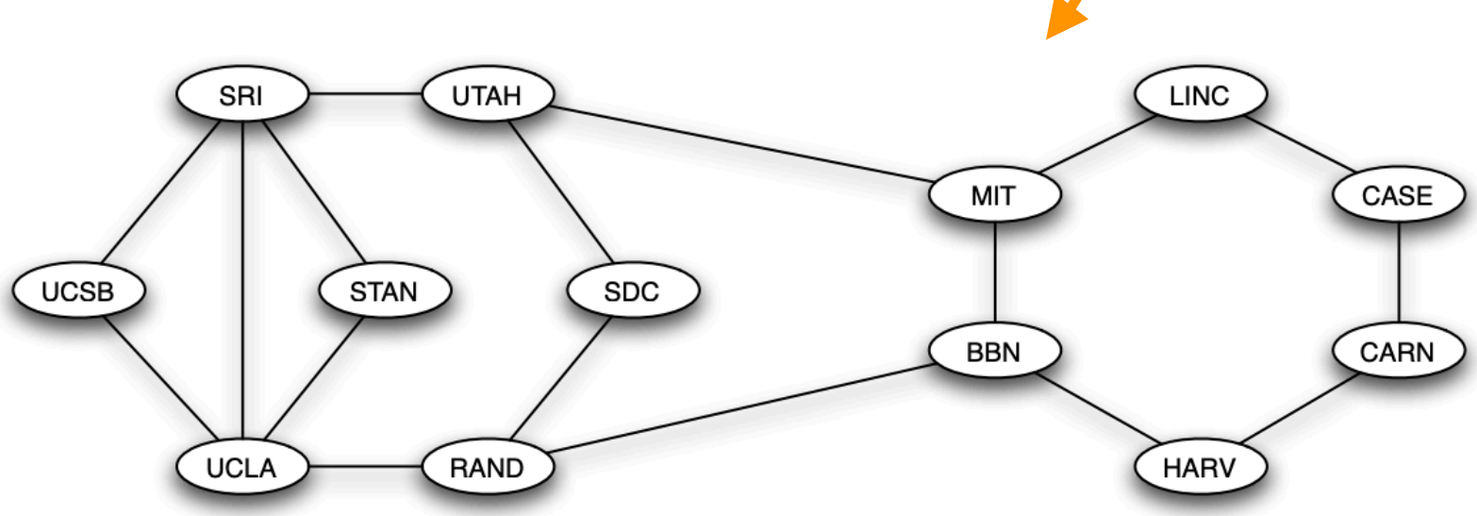



The Internet in 1970

# A first example

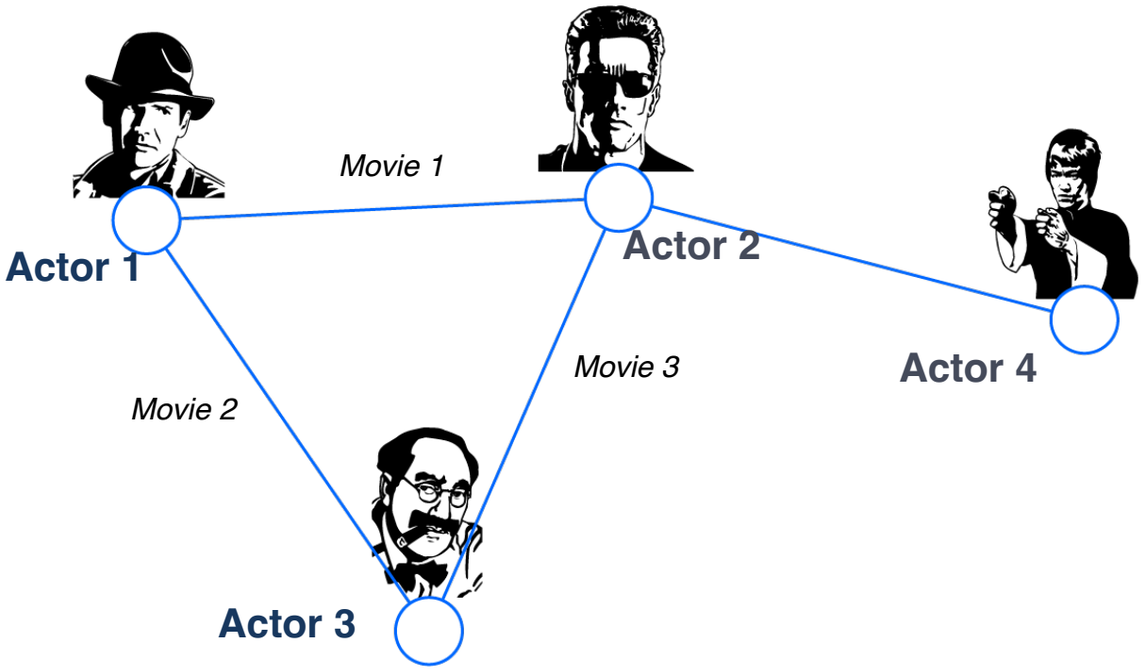


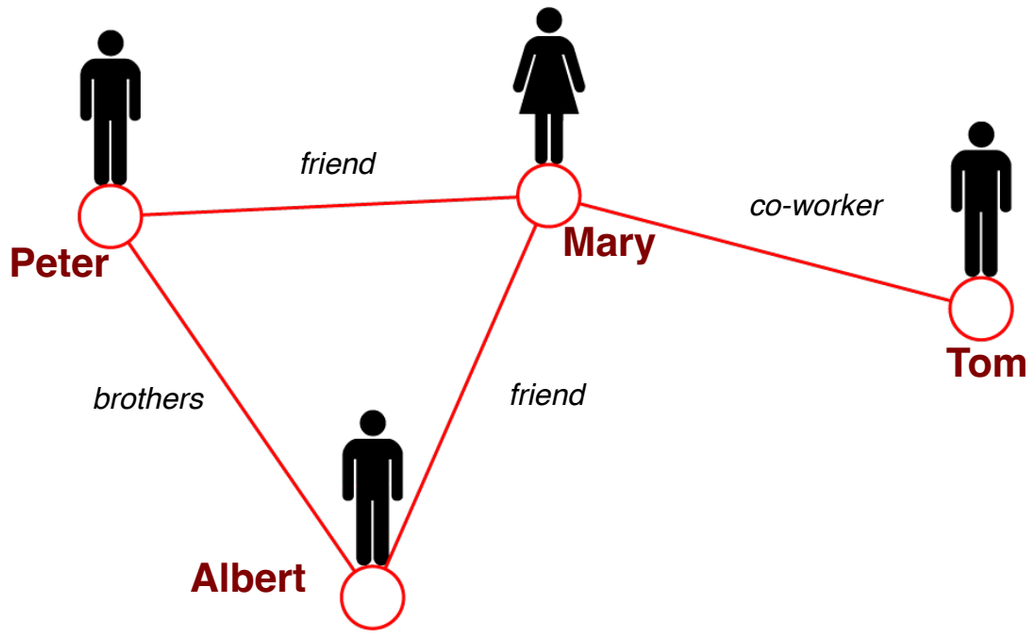

Translation

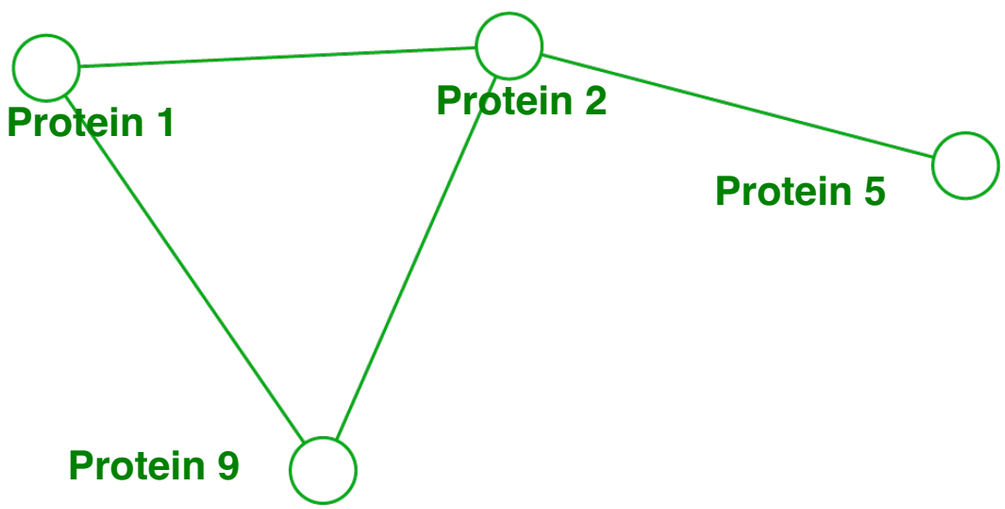


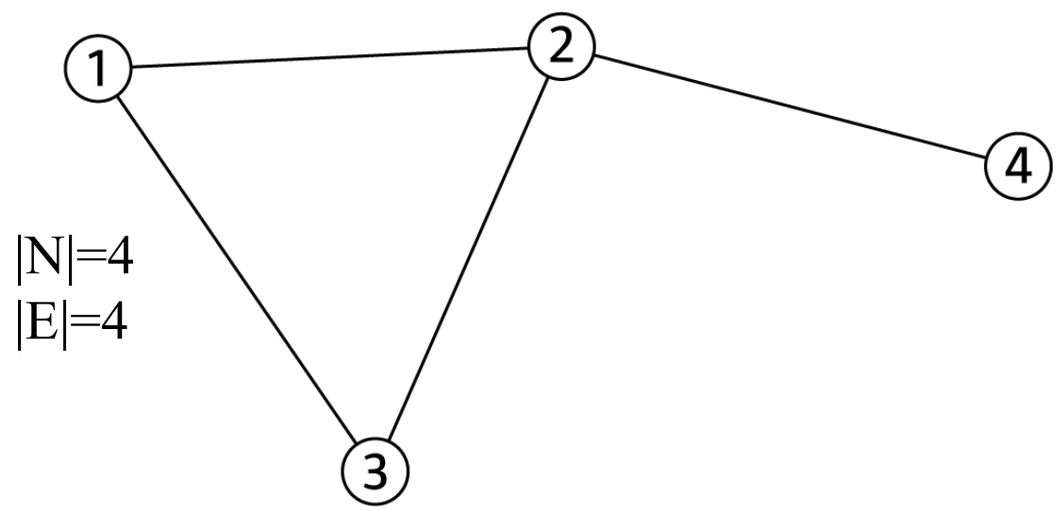

## The Internet in 1970

# 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**

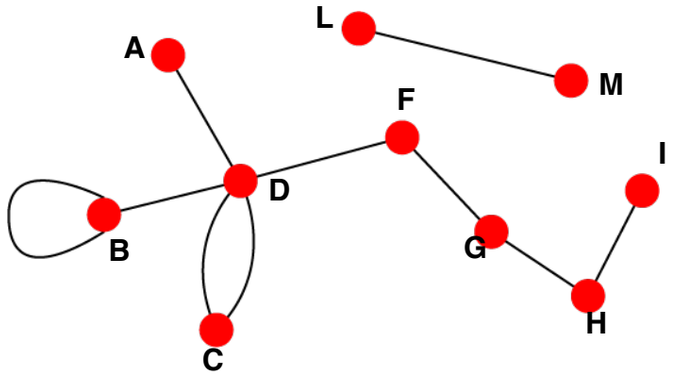
If you connect all people with first names that share the same first letter, what are you studying?

It is a network, but is it meaningful?

# Undirected and Directed Networks

## Undirected

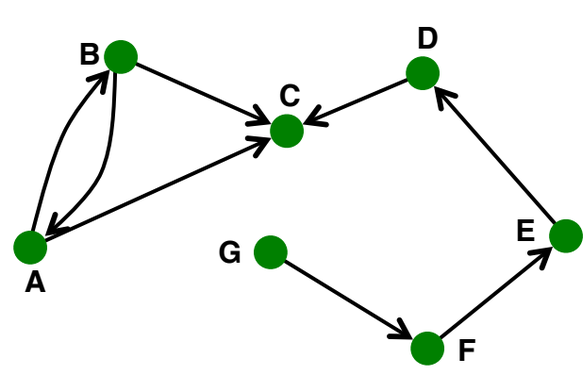
- **Links:** undirected (symmetrical, reciprocal)




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

## Directed

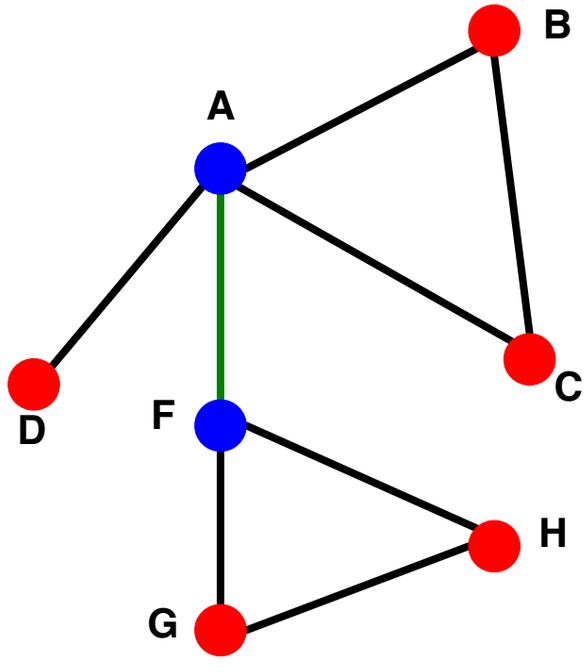
- **Links:** directed (arcs)

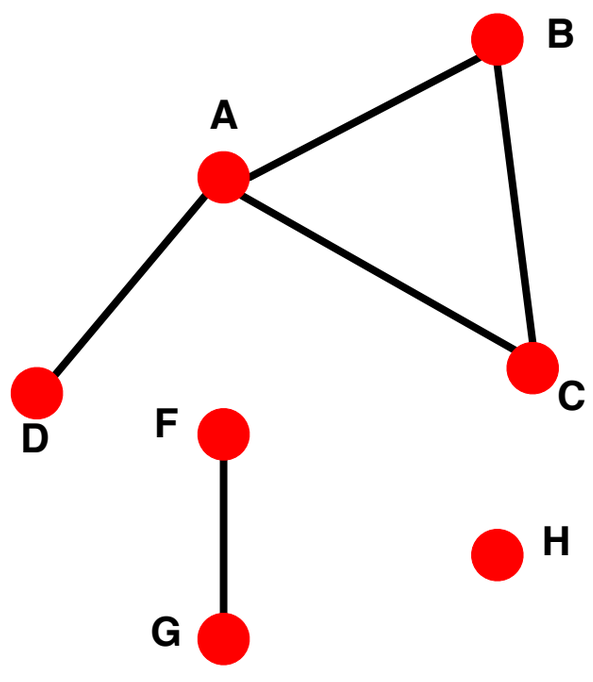



- **Examples:**
  - Phone calls
  - Following on Twitter

# 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



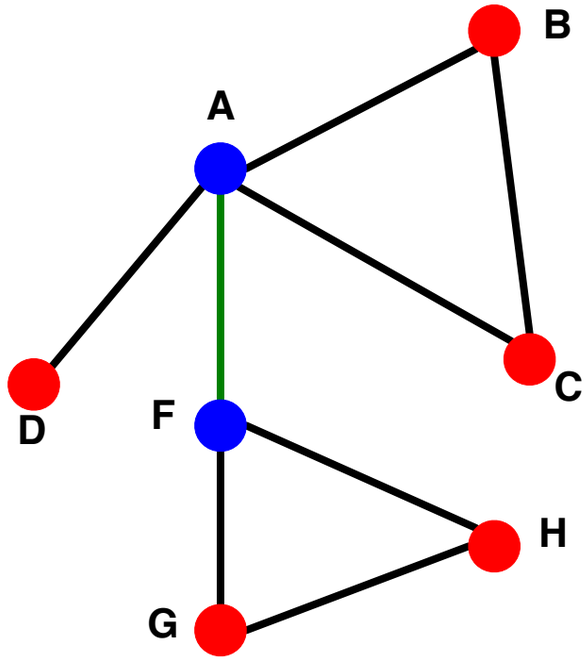



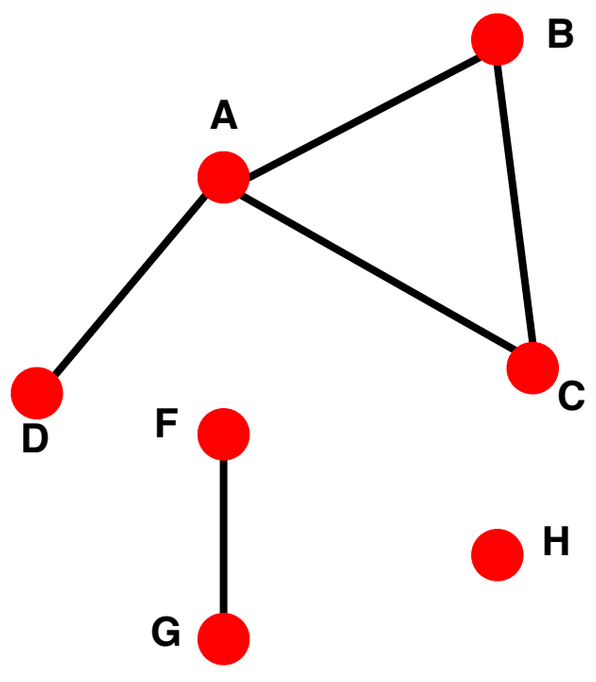

Largest Component:
**Giant Component**

**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



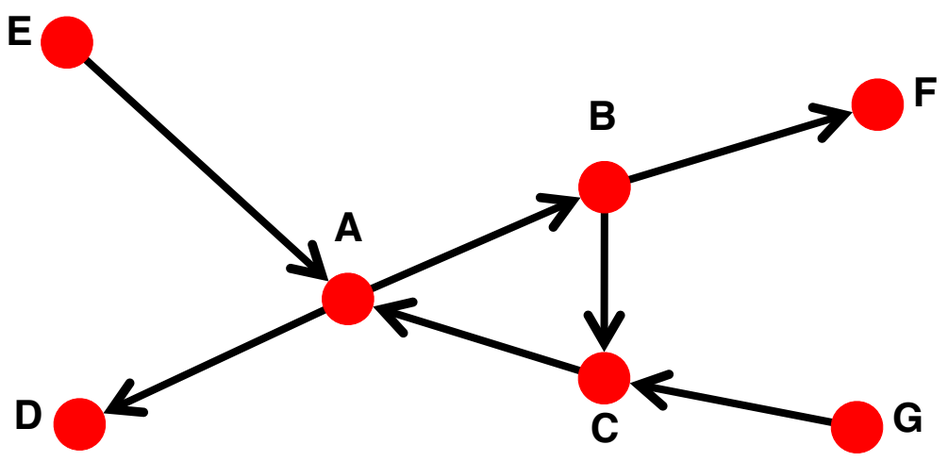

Largest Component:
**Giant Component**

**Isolated node** (node H)

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

# Connectivity of Directed Graphs

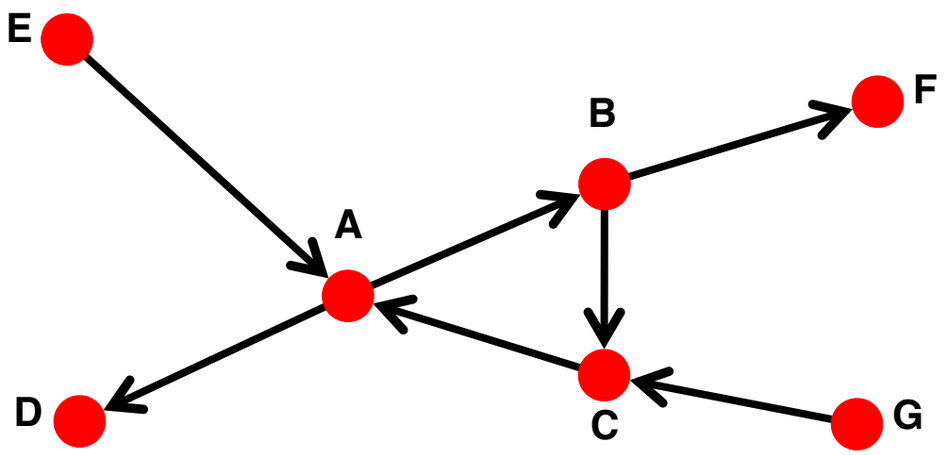
- **Strongly connected directed graph**
  - has a path from each node to every other node and vice versa (e.g., A-B path and B-A path)
- **Weakly connected directed graph**
  - is connected if we disregard the edge directions




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

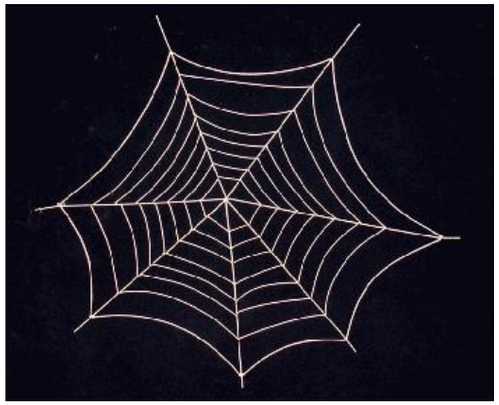



It is weakly connected but not strongly connected (e.g., there is no way to get from F to G by following the edge directions)

# What is the large-scale structure of the Web?

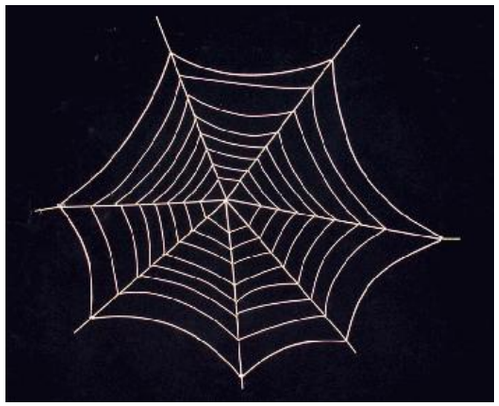
# The Structure of the Web

- Q: What does the Web “look like”?



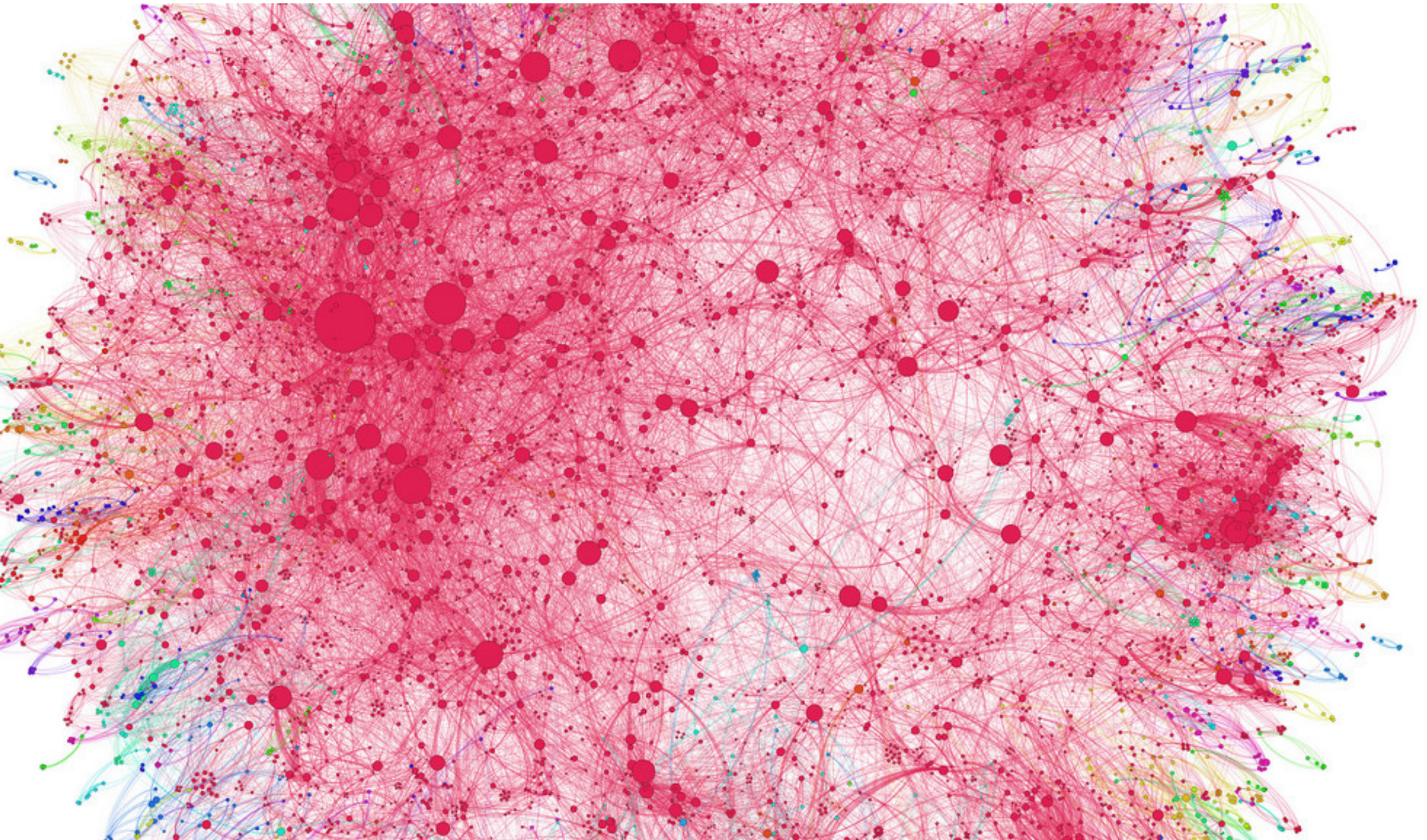
# The Structure of the Web

- Q: What does the Web “look like”?

# The Structure of the Web

## A network!



# Web as a Graph

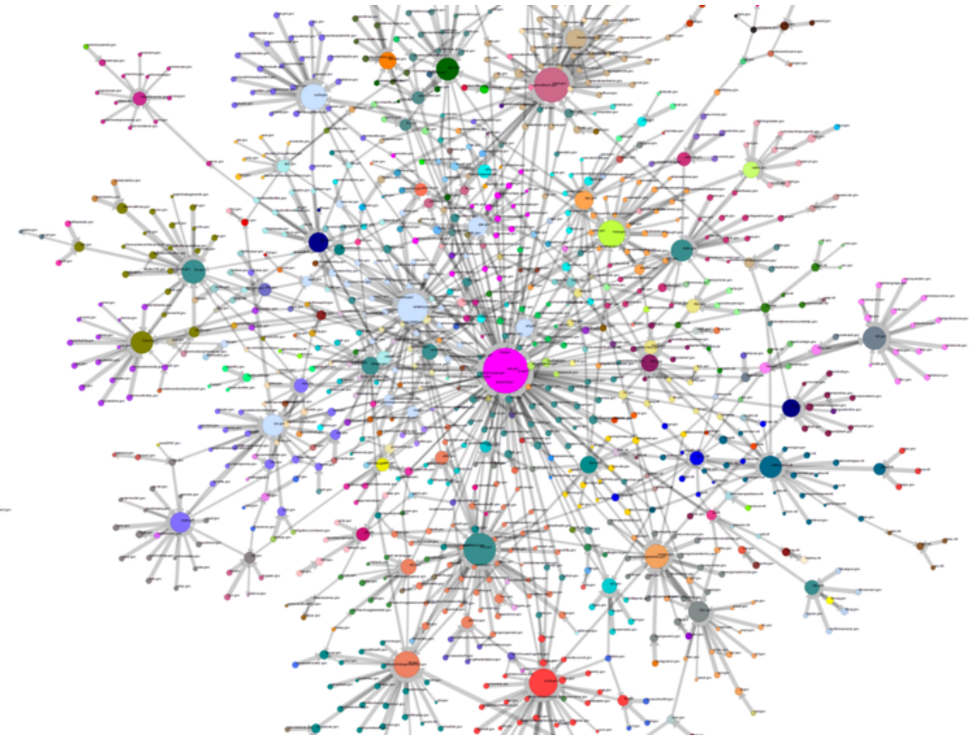
**Here is what we will do next:**

- We will take a real system (i.e., the Web)
- We will represent the Web as a graph
- We will use language of graph theory to reason about the structure of the graph
- Do a computational experiment on the Web graph
- **Learn something about the structure of the Web!**

# Web as a Graph

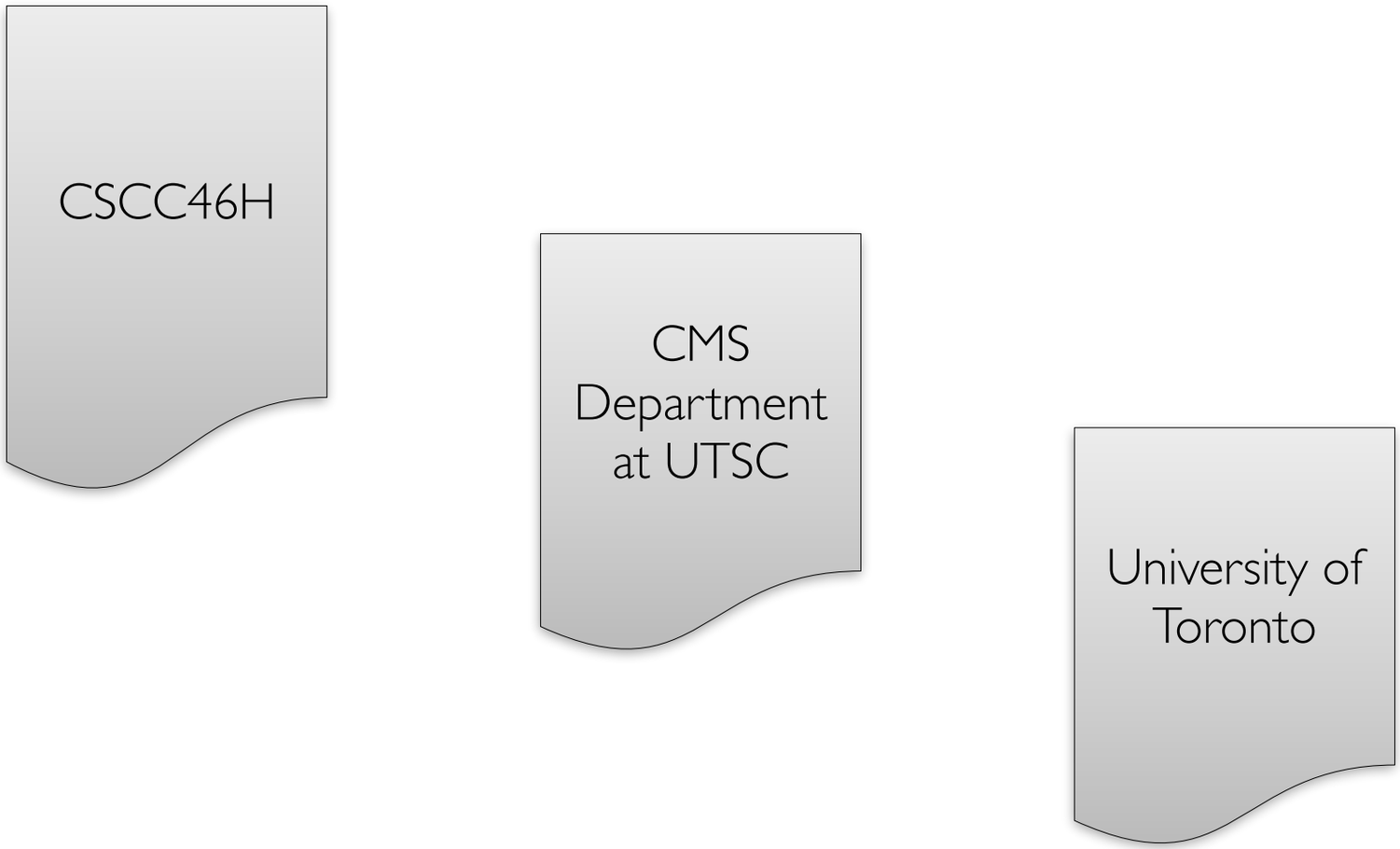
## Q: What does the Web "look like" at a global level?

- **Web as a graph:**
  - Nodes = web pages
  - Edges = hyperlinks

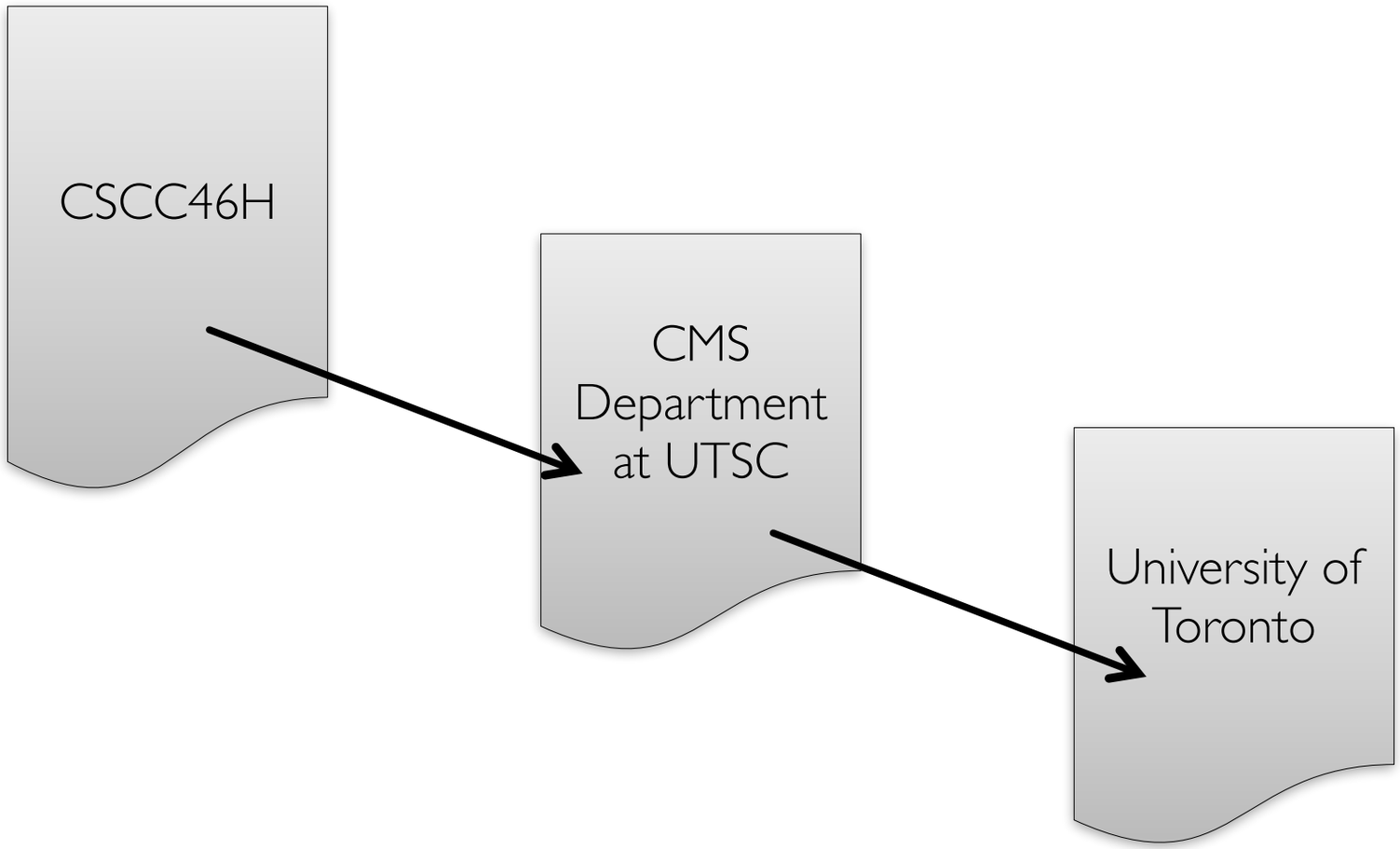


- **Side issue:** What is a node?
  - Dynamic pages created on the fly
  - "dark matter" – inaccessible database generated pages

# The Web as a Graph

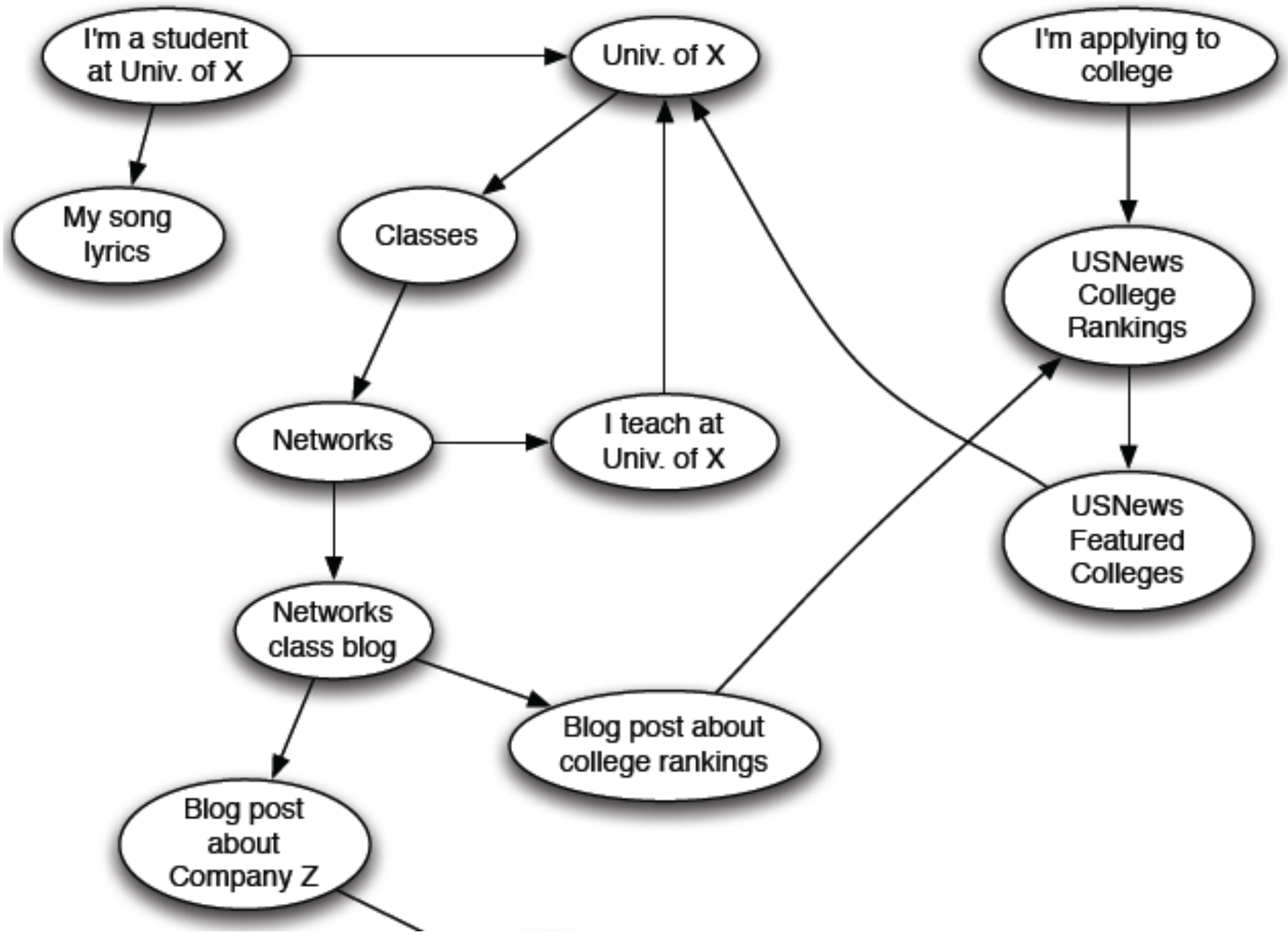


# The Web as a Graph

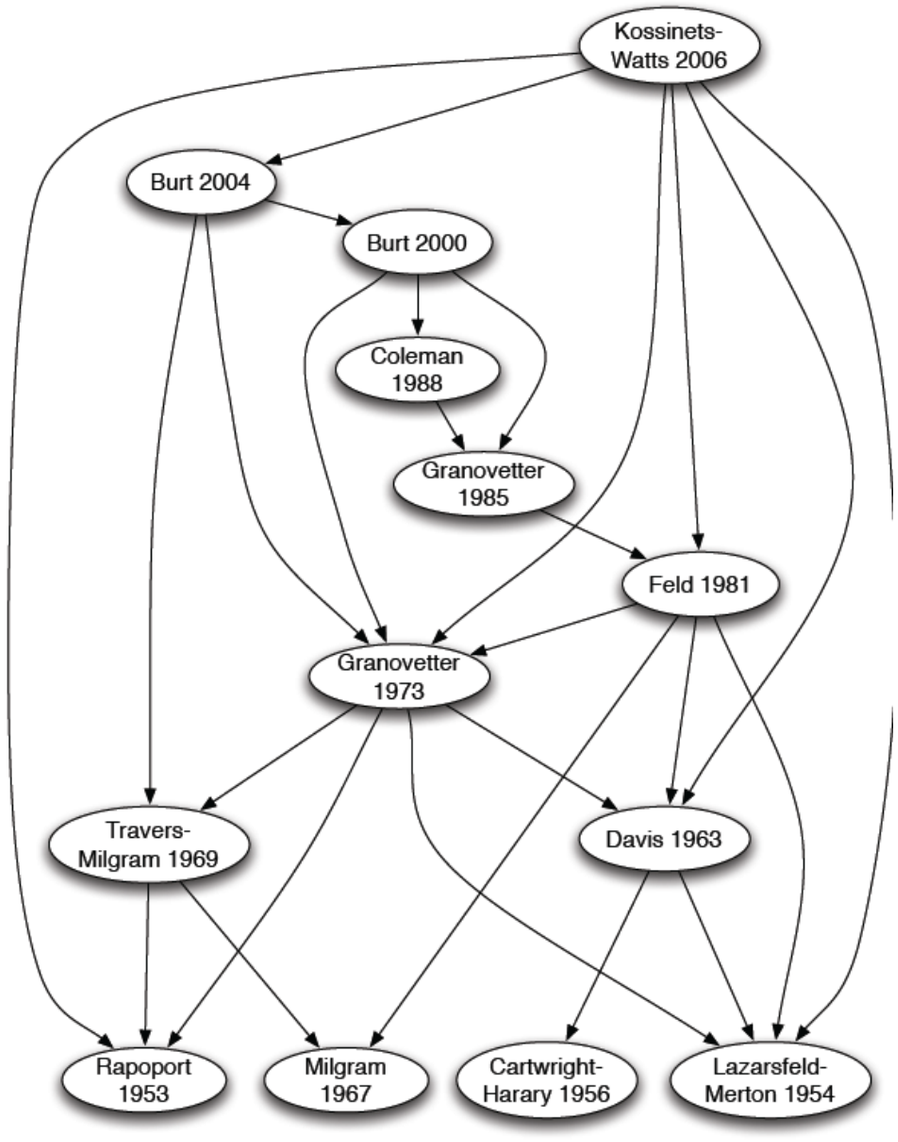



- In early days of the Web links were **navigational**
- Today many links are **transactional**

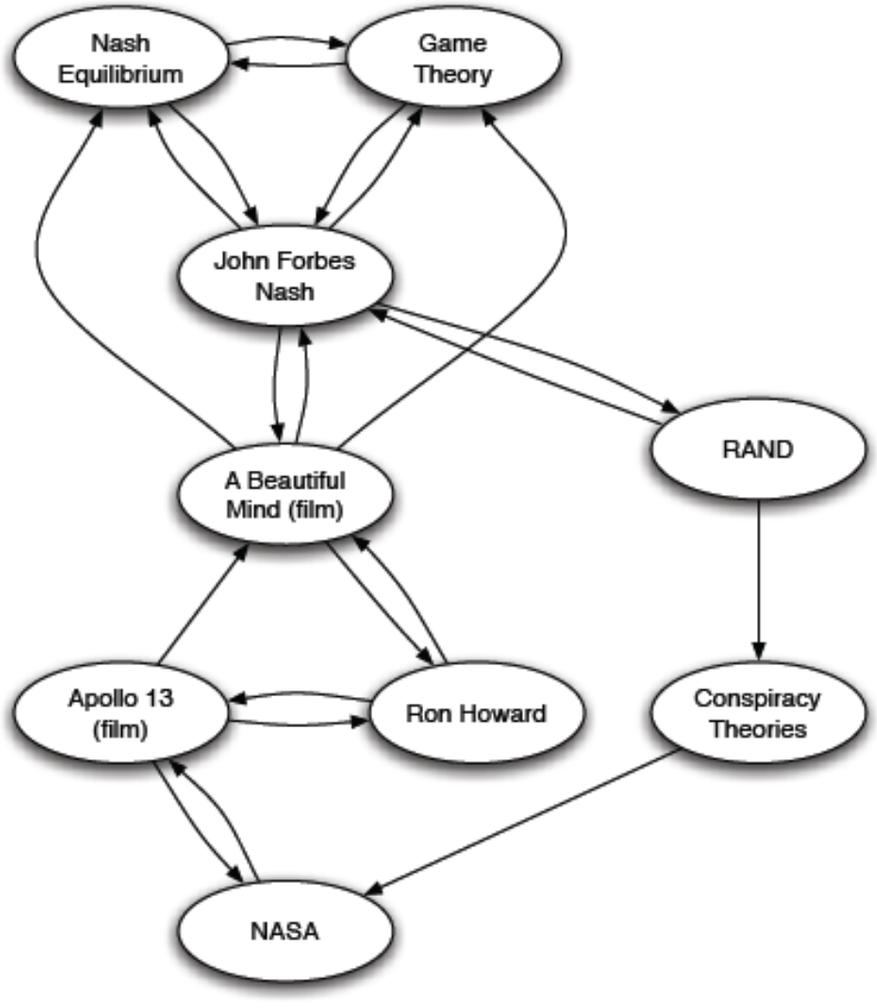
# The Web as a Directed Graph



# Other Information Networks

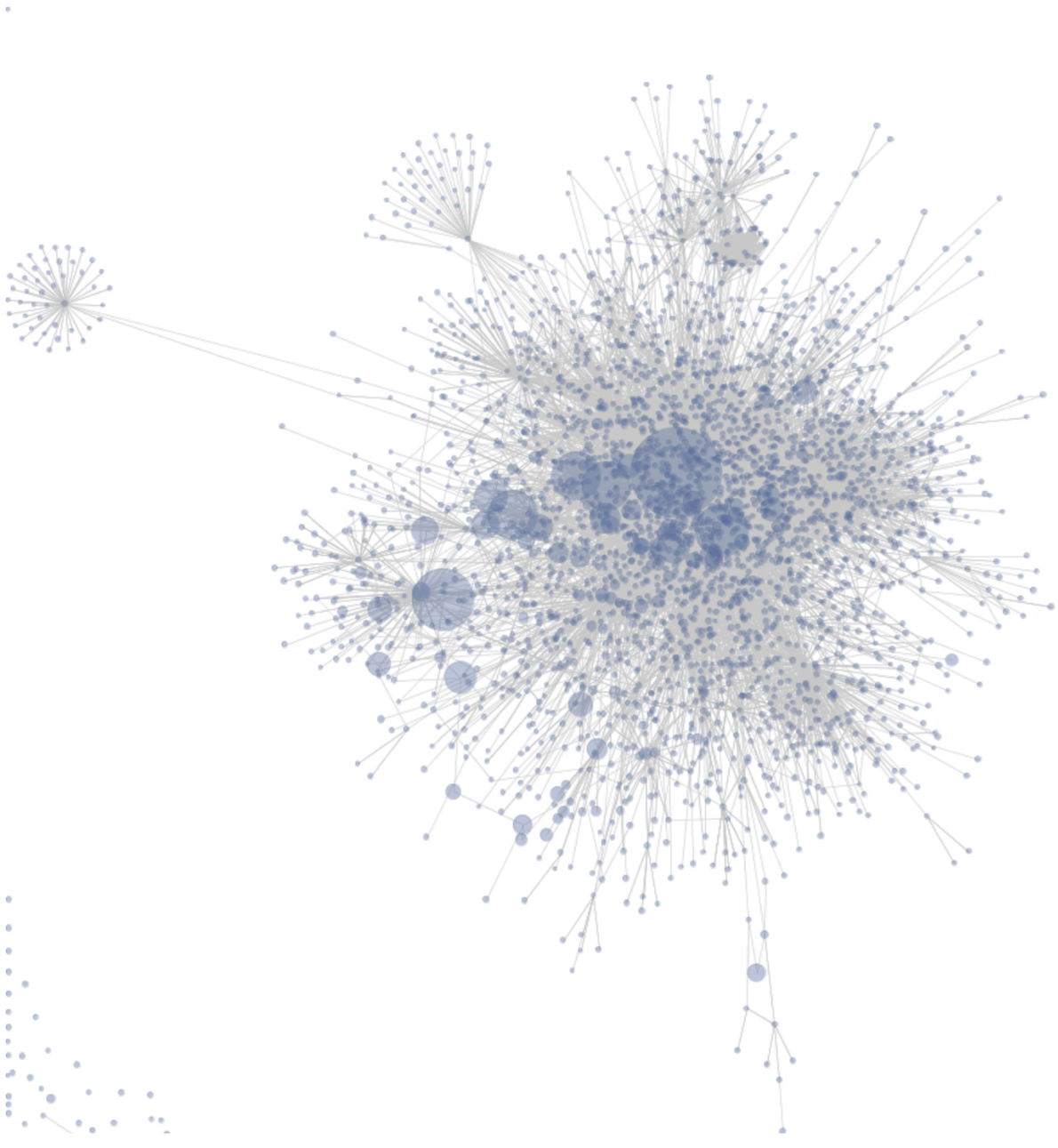



Citations




References in an encyclopedia

# Other Information Networks



References between pages in a part of Wikipedia

# What Does the Web Look Like?

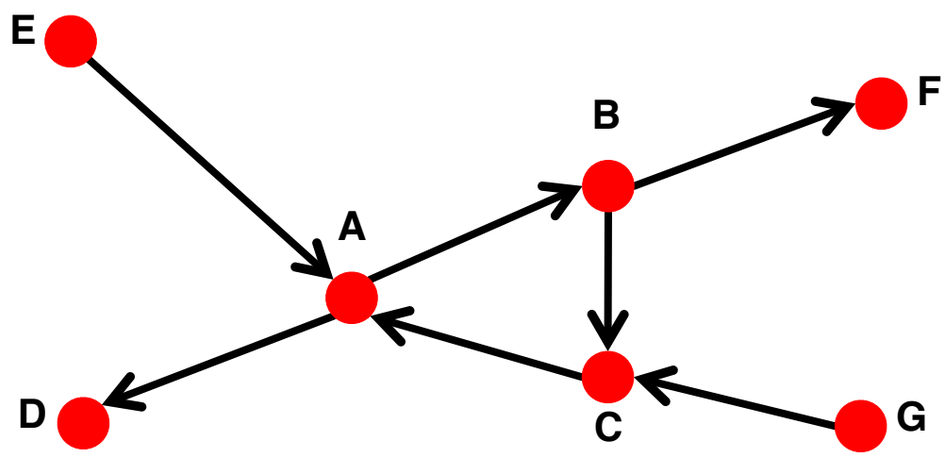
- How is the Web linked?
- What is the “map” of the Web?

# What Does the Web Look Like?

- **How is the Web linked?**
- **What is the "map" of the Web?**

**Web as a directed graph** [Broder et al. 2000]:

- Given node ***v***, what can ***v*** reach?
- What other nodes can reach ***v***?




$In(v) = \{w \mid w \text{ can reach } v\}$

$Out(v) = \{w \mid v \text{ can reach } w\}$

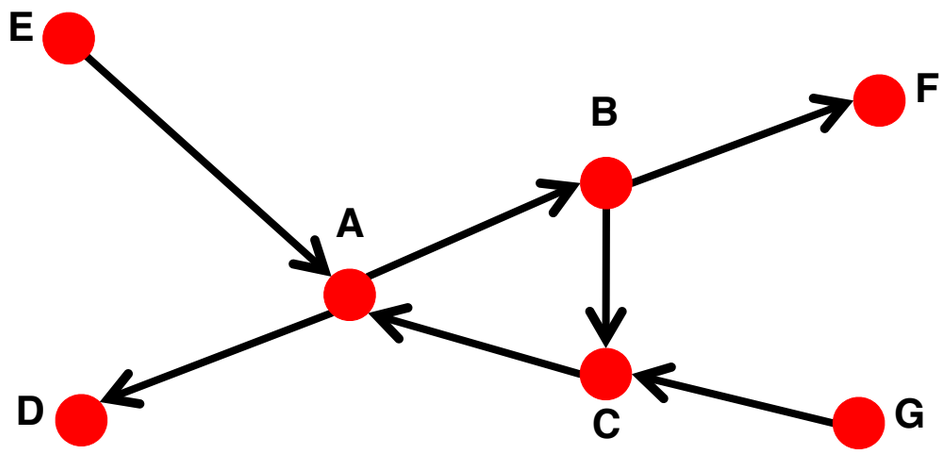
**For example:**
In(A) = {?}
Out(A)={?}

# What Does the Web Look Like?

- **How is the Web linked?**
- **What is the "map" of the Web?**

**Web as a directed graph** [Broder et al. 2000]:

- Given node ***v***, what can ***v*** reach?
- What other nodes can reach ***v***?



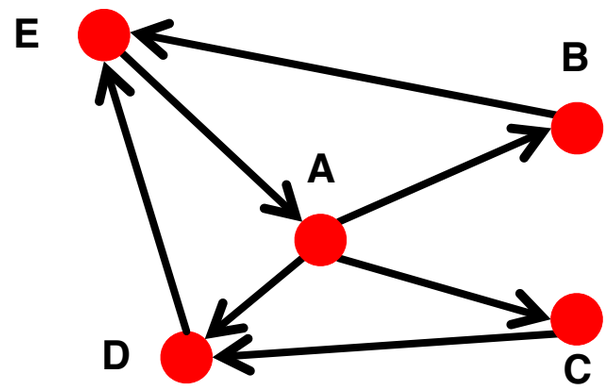

$$In(v) = \{w \mid w \text{ can reach } v\}$$
$$Out(v) = \{w \mid v \text{ can reach } w\}$$

**For example:**
In(A) = {A,B,C,E,G}
Out(A)={A,B,C,D,F}

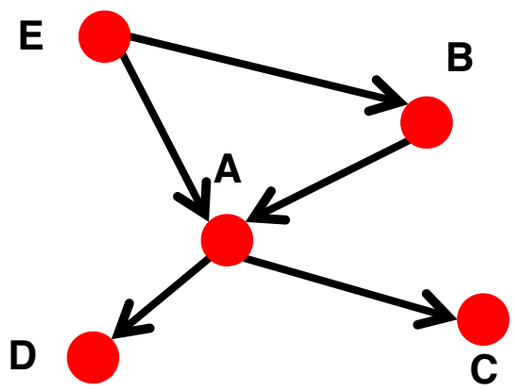
# Directed Graphs

- **Two types of directed graphs:**
  - **Strongly connected graph:**
    - Any node can reach any node via a directed path

      *In(A)=Out(A)={A,B,C,D,E}*



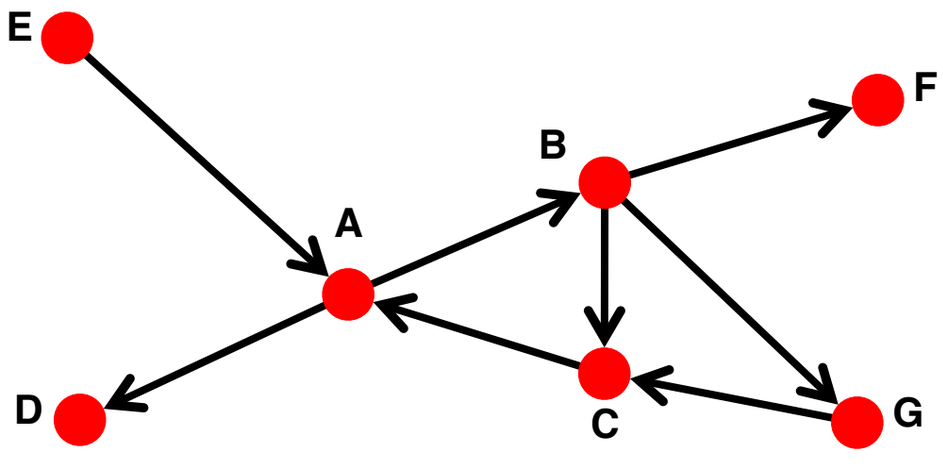

  - **DAG – Directed Acyclic Graph:**
    - Has no cycles: if ***u*** can reach ***v***, then ***v*** can not reach ***u***




**Any directed graph can be expressed in terms of these two types!**

# Strongly Connected Component

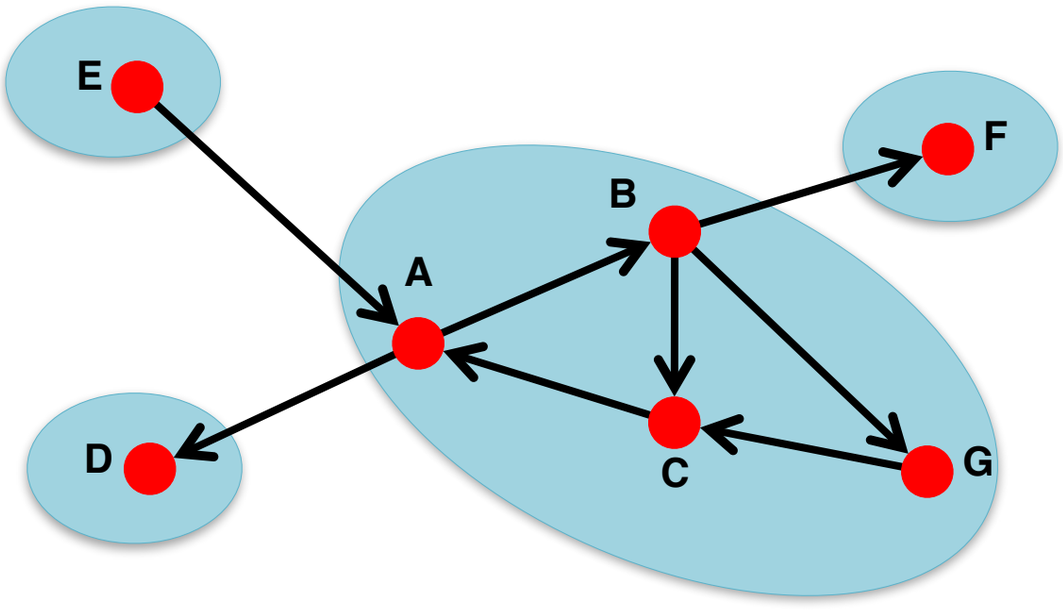
- **Strongly connected component (SCC)** is a set of nodes ***S*** so that:
  - Every pair of nodes in ***S*** can reach each other
  - There is no larger set containing ***S*** with this property




What are the strongly connected components of this graph?

# Strongly Connected Component

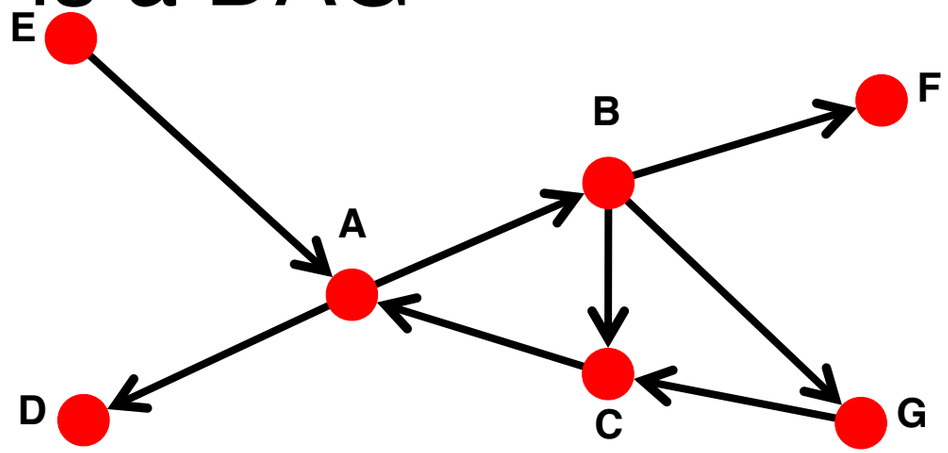
- **Strongly connected component (SCC)** is a set of nodes ***S*** so that:
  - Every pair of nodes in ***S*** can reach each other
  - There is no larger set containing ***S*** with this property




Strongly connected components of the graph:
{A,B,C,G}, {D}, {E}, {F}

# Strongly Connected Component

- **Fact: Every directed graph is a DAG on its SCCs**
  - **(1)** SCCs partitions the nodes of ***G***
    - That is, each node is in exactly one SCC
  - **(2)** If we build a graph ***G'*** whose nodes are SCCs, and with an edge between nodes of ***G'*** if there is an edge between corresponding SCCs in ***G***, then ***G'*** is a DAG

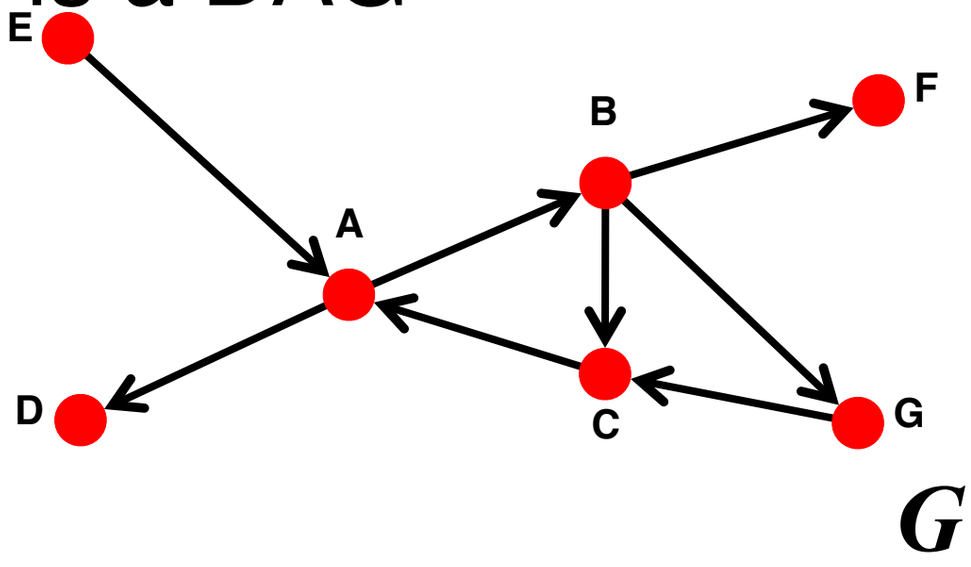



***G***

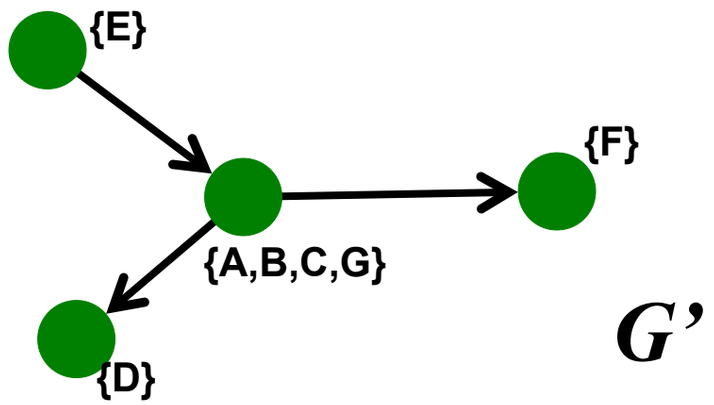
***G'?***

# Strongly Connected Component

- **Fact: Every directed graph is a DAG on its SCCs**
  - **(1)** SCCs partitions the nodes of ***G***
    - That is, each node is in exactly one SCC
  - **(2)** If we build a graph ***G'*** whose nodes are SCCs, and with an edge between nodes of ***G'*** if there is an edge between corresponding SCCs in ***G***, then ***G'*** is a DAG

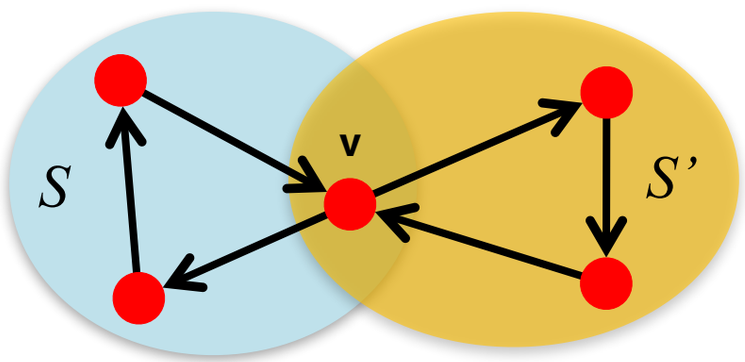



(1) Strongly connected components of graph G: {A,B,C,G}, {D}, {E}, {F}
(2) G' is a DAG:



# Proof of (1)

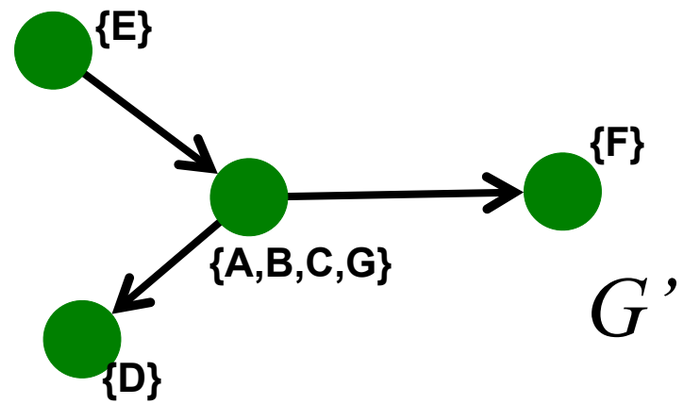
- **Claim: SCCs partitions nodes of G.**
  - This means: Each node is member of exactly 1 SCC
- Proof by contradiction:
  - Suppose there exists a node ***v*** which is a member of two SCCs ***S*** and ***S'***

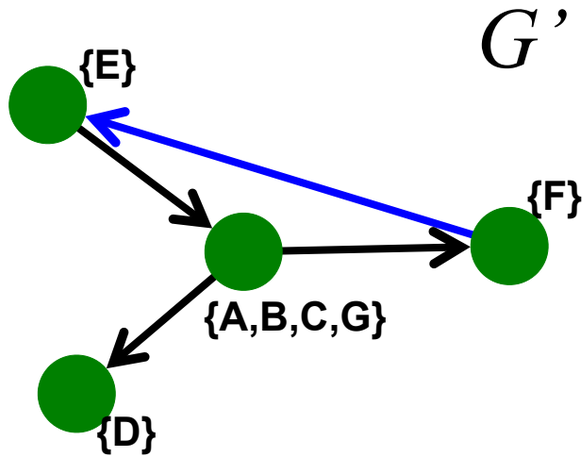



  - But then ***S*** ∪ ***S'*** is one large SCC!
    - Contradiction!

# Proof of (2)

- **Claim: *G'* (graph of SCCs) is a DAG.**
  - This means: ***G'*** has no cycles
- Proof by contradiction:
  - Assume ***G'*** is not a DAG
  - Then ***G'*** has a directed cycle
  - Now all nodes on the cycle are mutually reachable, and all are part of the same SCC
  - But then ***G'*** is not a graph of connections between SCCs (SCCs are defined as maximal sets)
    - Contradiction!



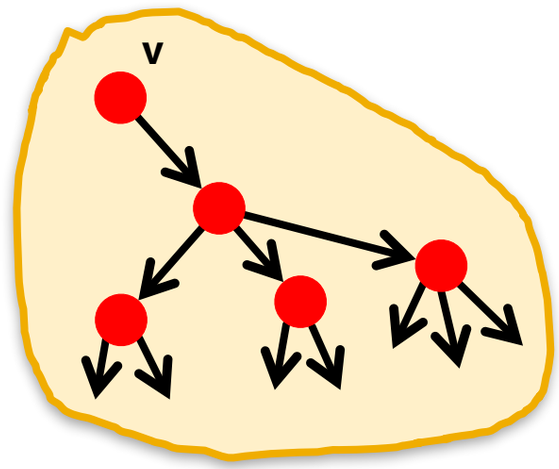




Now {A,B,C,G,E,F} is a SCC!

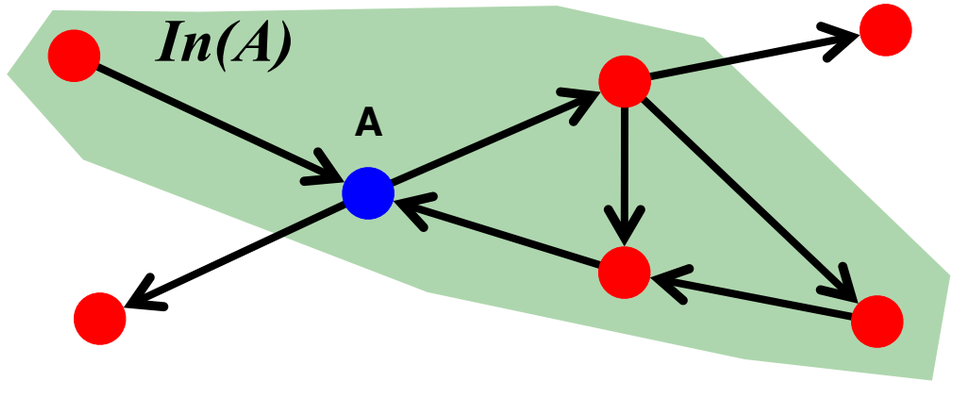
# Graph Structure of the Web

- **Goal:** **Take a large snapshot of the Web and try to understand how its SCCs "fit together" as a DAG**
- **Computational issue:**
  - Want to find a SCC containing node ***v***?
  - **Observation:**
    - ***Out(v)*** … nodes that can be reached from *v*
    - **SCC containing *v* is:** $Out(v) \cap In(v)$

      $= Out(v,G) \cap Out(v,\overline{G})$, where $\overline{G}$ is $G$ with all edge directions flipped

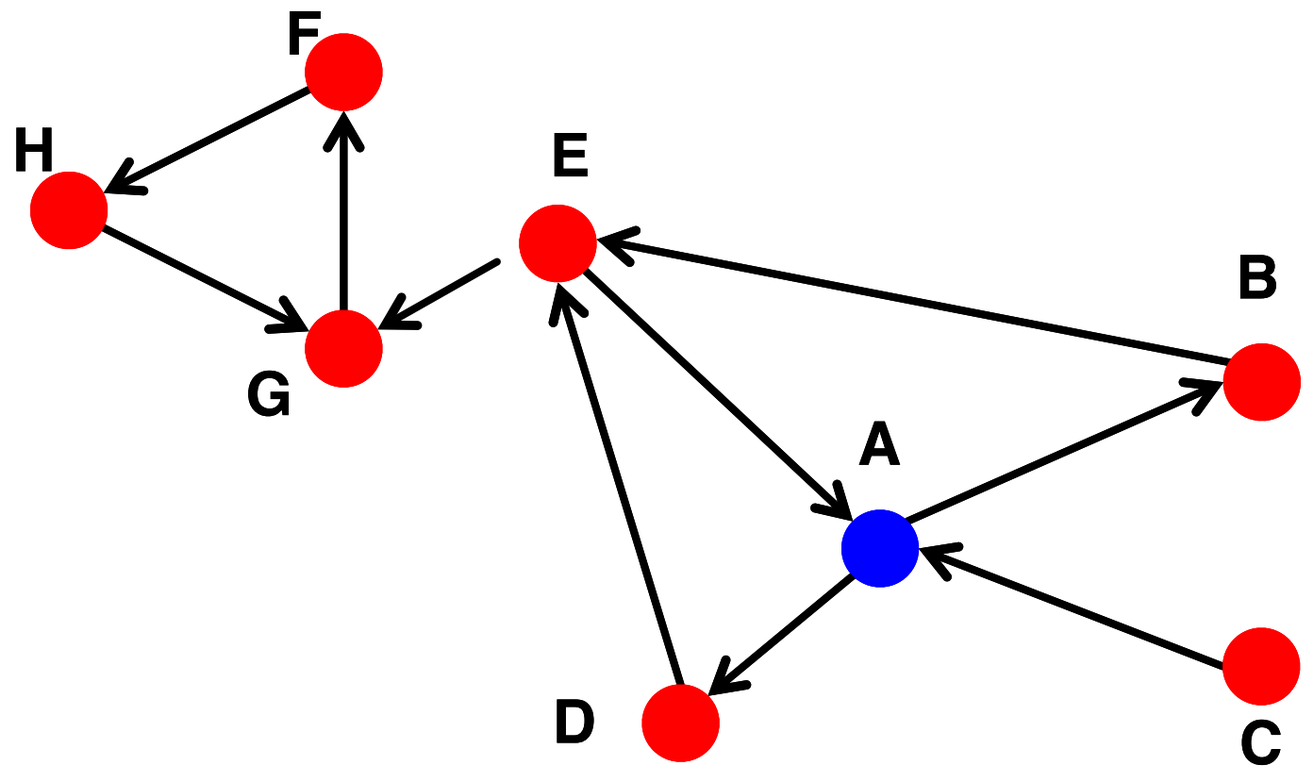


*Out(v)*



# Out(A) ∩ In(A) = SCC

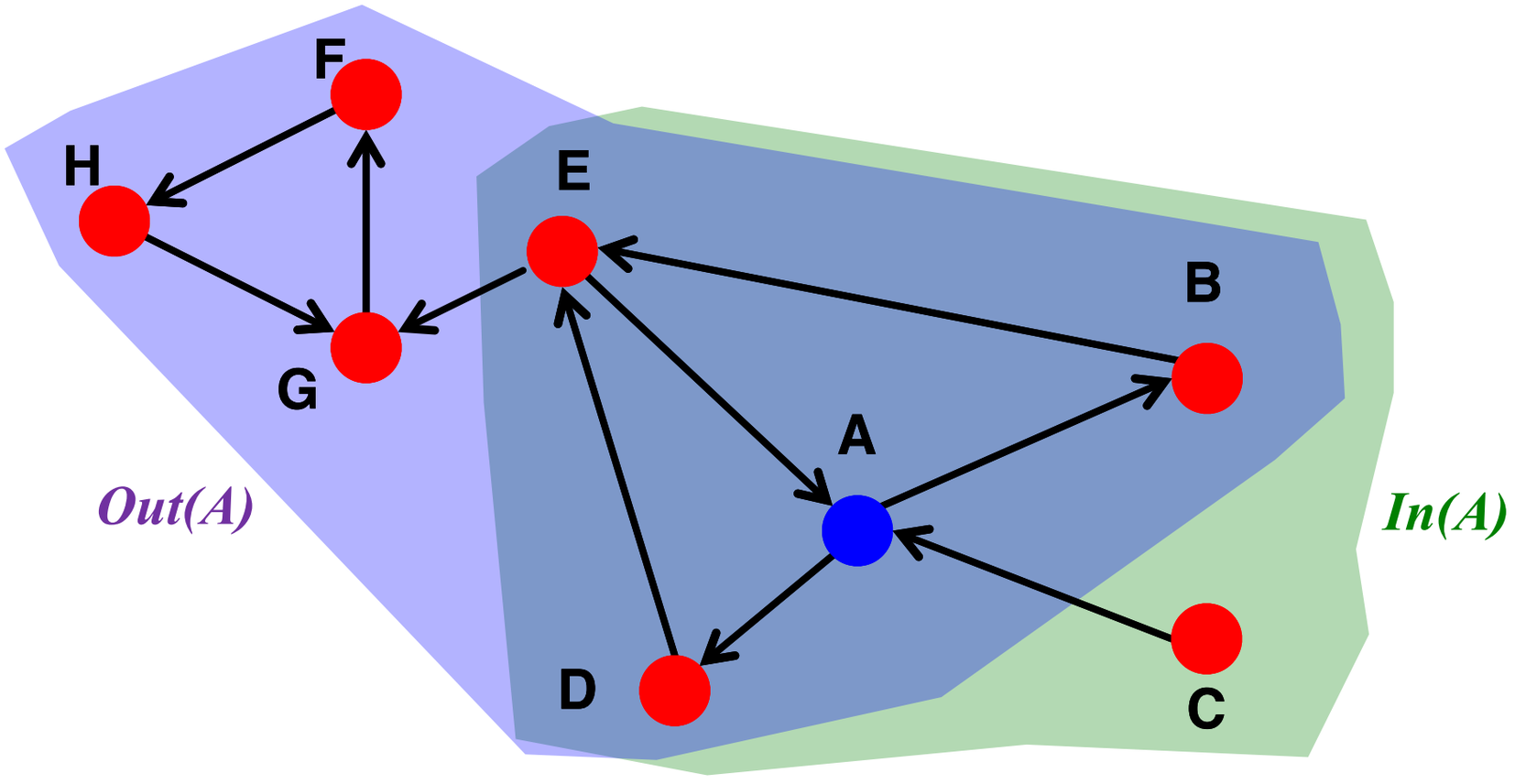
## Example:




- Out(A) = {?}
- In(A) = {?}

# Out(A) ∩ In(A) = SCC

## Example:



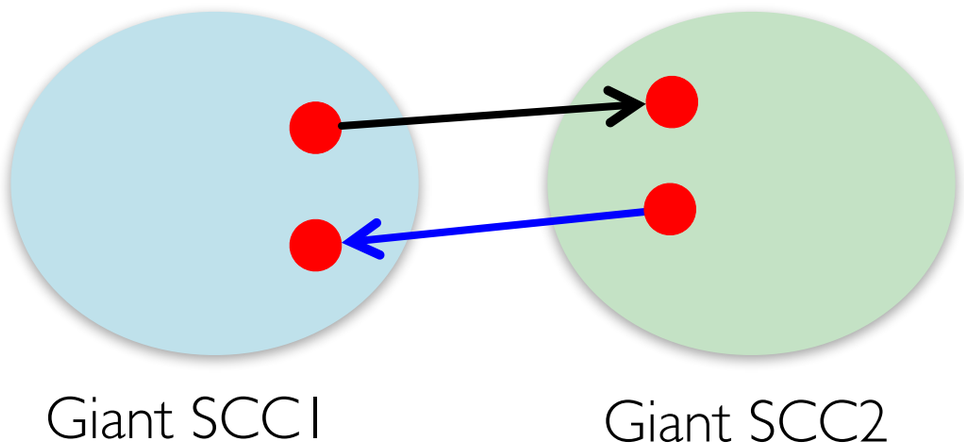

- Out(A) = {A,B,D,E,F,G,H}
- In(A) = {A,B,C,D,E}
- Therefore, SCC(A) = {A,B,D,E}

# Graph Structure of the Web

- How many “big” SCCs?

# Graph Structure of the Web

- **There is a single giant SCC**
  - That is, there won't be two SCCs
- **Heuristic argument:**
  - It just takes 1 page from one SCC to link to the other SCC
  - If the 2 SCCs have millions of pages the likelihood of this not happening is very very small



# Structure of the Web

- **Broder et al., 2000:**
  - Historical Altavista crawl from October 1999
    - 203 million URLS
    - 1.5 billion links
    - Computer: Server with 12GB of memory
- **Undirected version of the Web graph:**
  - 91% nodes in the largest weakly connected component
  - Are hubs making the web graph connected?
    - Even if they deleted links to pages with in-degree >10 WCC was still ≈50% of the graph

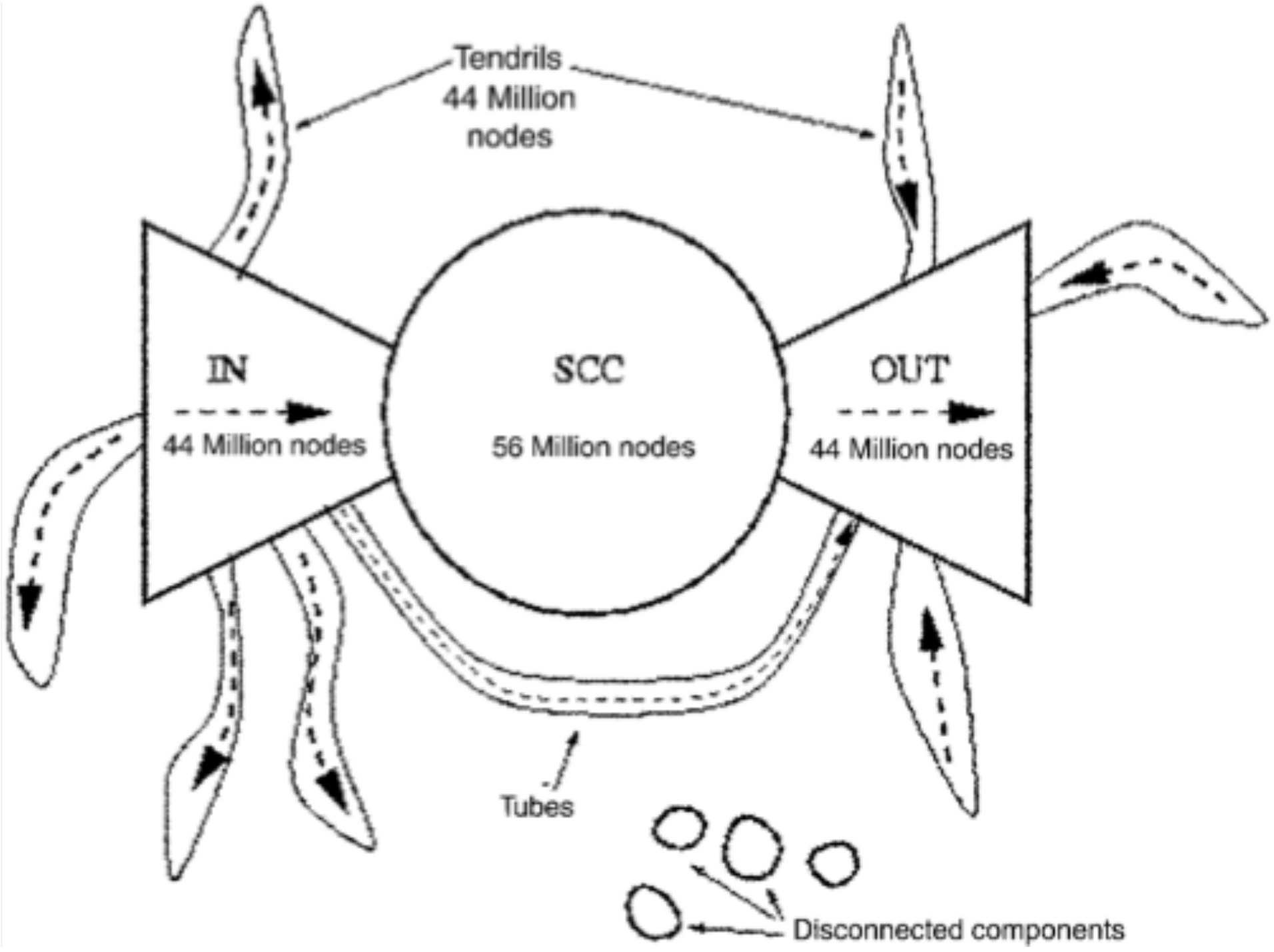
# Structure of the Web

- **Directed version of the Web graph:**
  - **Largest SCC:** 28% of the nodes (56 million)
  - Taking a random node $v$
    - **Out($v$) ≈ 50%** (100 million)
    - **In($v$) ≈ 50%** (100 million)

- **What does this tell us about the conceptual picture of the Web graph?**

# Bow-tie Structure of the Web

203 million pages, 1.5 billion links [Broder et al. 2000]



# What did we do?

- **Here is what we've already done**
  - We took a real system (the Web)
  - We represented the Web as a graph
  - We used the language of graph theory to reason about the structure of the graph
  - We did a computational experiment on the Web graph
  - **Learned something about the structure of the Web!**

# What did We Learn/Not Learn ?

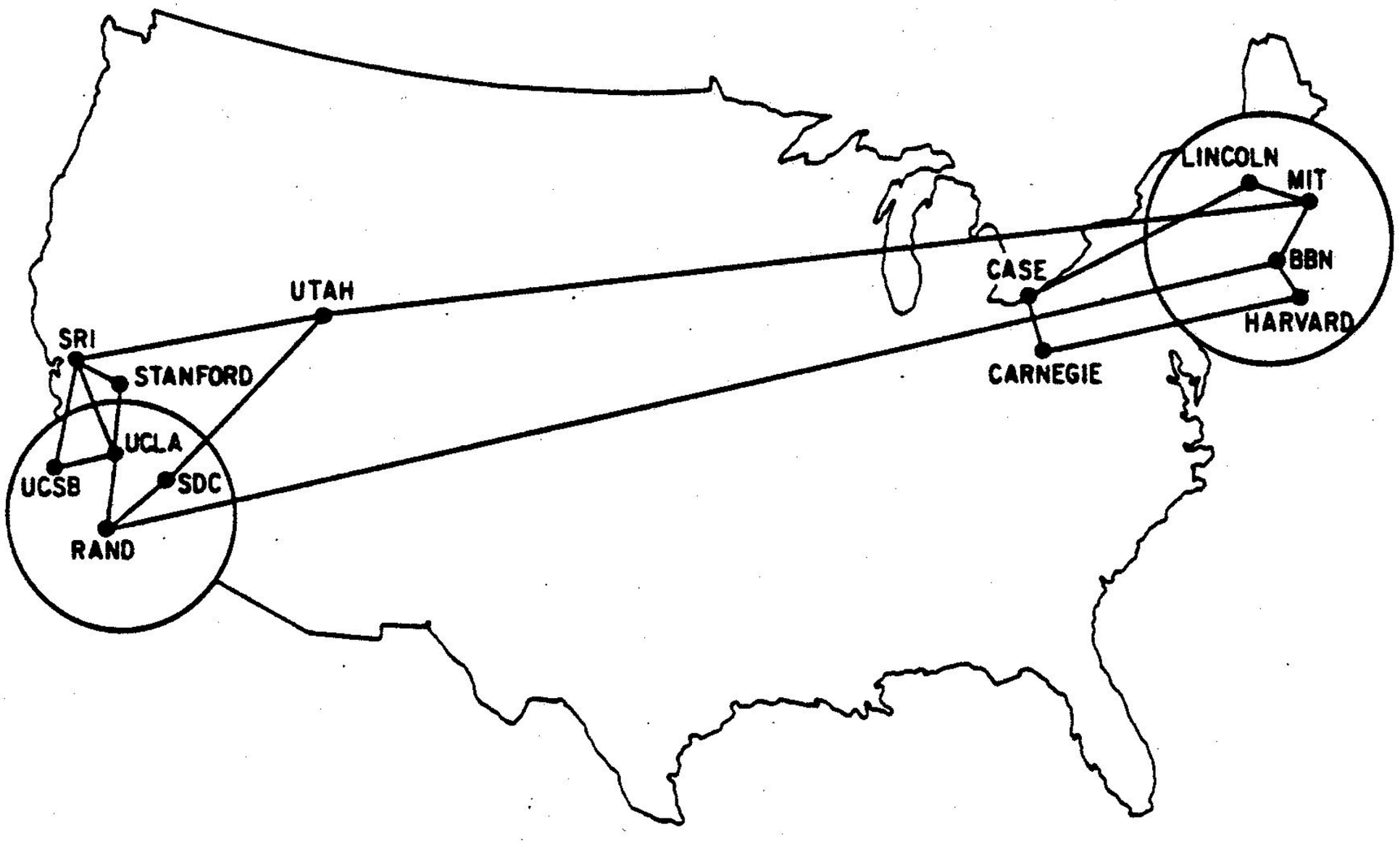
- **What did we learn:**
  - Some conceptual organization of the Web (i.e., the bowtie)
- **What did we not learn:**
  - **Treats all pages as equal**
    - Google's homepage == my homepage
  - **What are the most important pages**
    - How many pages have *k* in-links as a function of *k*?

      The degree distribution: $\sim k^{-2}$
    - Link analysis ranking -- as done by search engines (PageRank)
  - **Internal structure inside giant SCC**
    - Clusters, implicit communities?
  - **How far apart are nodes in the giant SCC:**
    - Distance = # of edges in shortest path
    - Avg = 16 [Broder et al.]

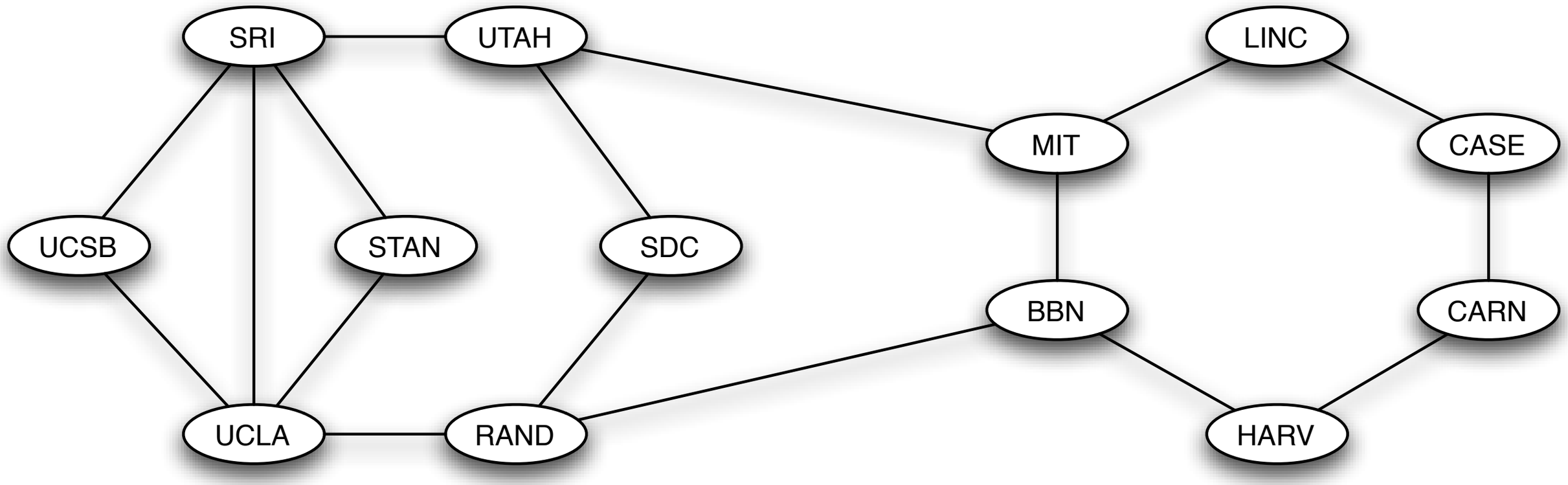
# Recap

- **Network analysis is the language of connectedness**
  - Represent real-world networks from many different domains as graphs, use graph theory and algorithms to reason about them
  - Social networks, information networks, knowledge networks, biological networks, etc.
- **Network analysis fundamentals**
  - Nodes, edges, paths, cycles, un/directed, connected components (weak and strong)
  - Choices of representation
  - Every directed graph is a DAG on its SCCs
- **Structure of the Web**
  - Looks like a bow-tie: big giant component, IN & OUT components, tendrils, disconnected components

# Arpanet, December 1970



# Arpanet, December 1970



# End!

**This week:**

- **Sign up for MarkUs**
- **Log in to Quercus**
- **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**