LECTURE 1:INTRODUCTION

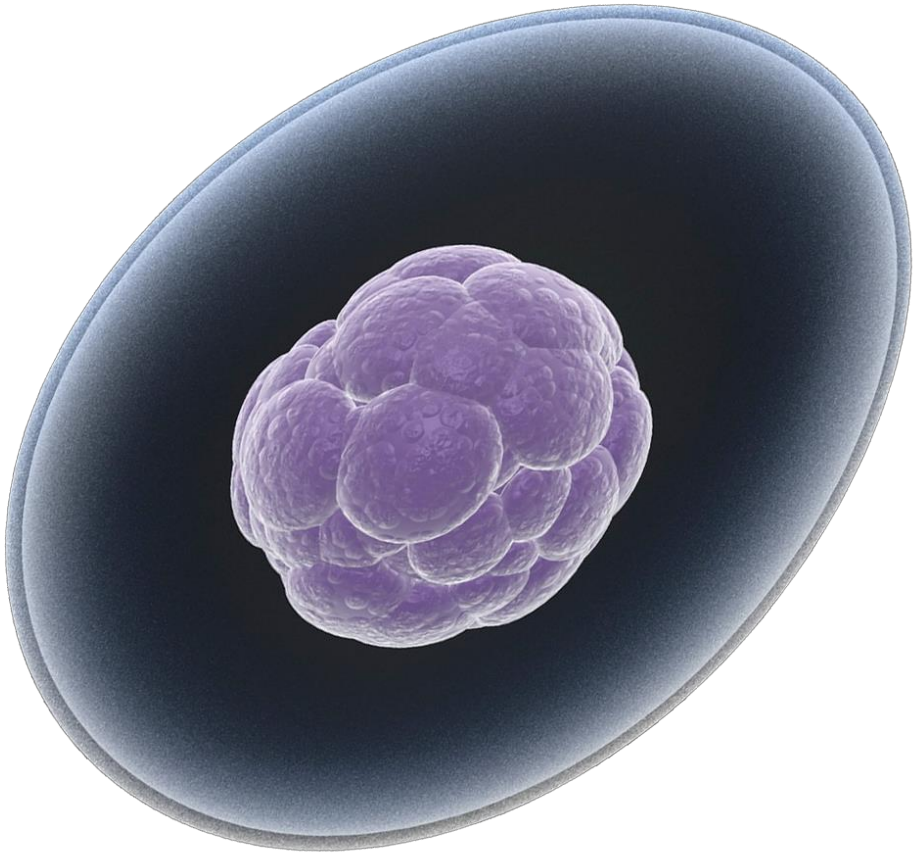
Prof. Pan Hui

Some course materials of this lecture are adopted from the Stanford University CS224W: Social and Information Network Analysis by Prof. Jure Leskovec, and the COMS W4995-1 Introduction to Social Networks by Prof. Augustin Chaintreau from Columbia University.

CSIT 6000K: Social Networks and Social Computing: A Data Science Perspective
Thursdays 07:30 PM - 10:20 PM

What do the following things have in common?

World economy

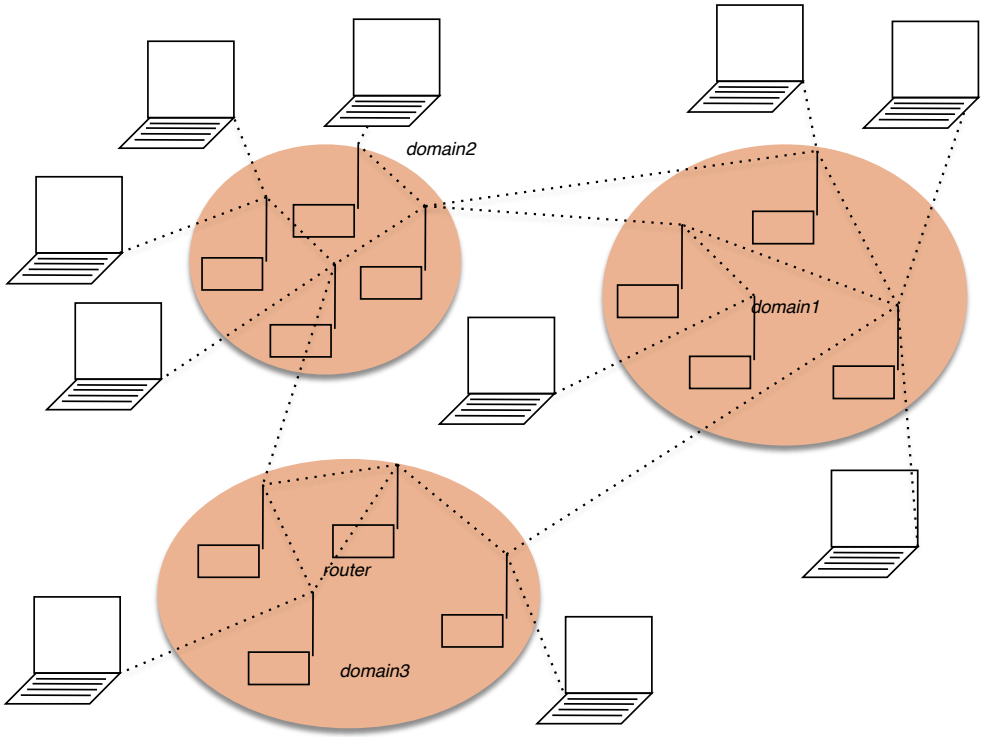


Human cell


Swiss Cottage
St. John's Wood
Mornington Crescent
King's Cross St. Pancras
Road & Barnsbury
Kilburn High Road
South Hampstead
Paddington
Edgware Road
Baker Street
Great Portland Street
Euston
Angel
Old Street
Warren Street
Euston Square
Farringdon
Marylebone
Regent's Park
Russell Square
Barbican
Bayswater
Holland Park
Notting Hill Gate
Lancaster Gate
Bond Street
Oxford Circus
Goodge Street
Holborn
Chancery Lane
Moorgate
Queensway
Marble Arch
Tottenham Court Road
St. Paul's
Bank
Covent Garden
Cannon Street
High Street Kensington
Hyde Park Corner
Green Park
Piccadilly Circus
Leicester Square
Mansion House
Monument
Knightsbridge
Charing Cross
Gloucester Road
Sloane Square
St. James's Park
Blackfriars
Temple
Earl's Court
South Kensington
Victoria
Westminster
Embankment
London Bridge
Fulham Broadway
Waterloo

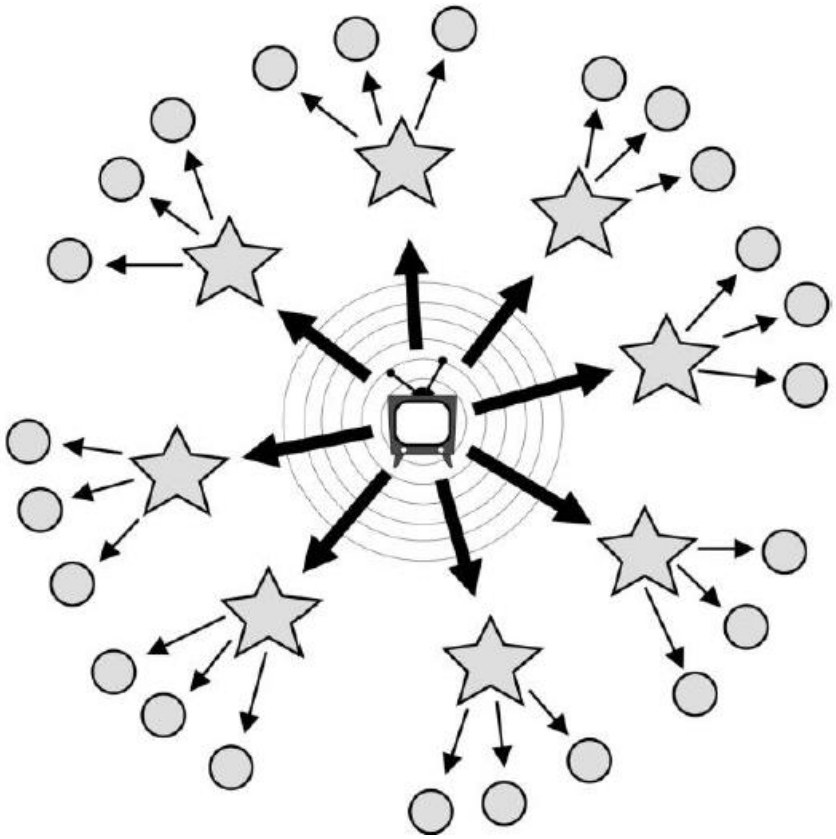
Roads

Brain

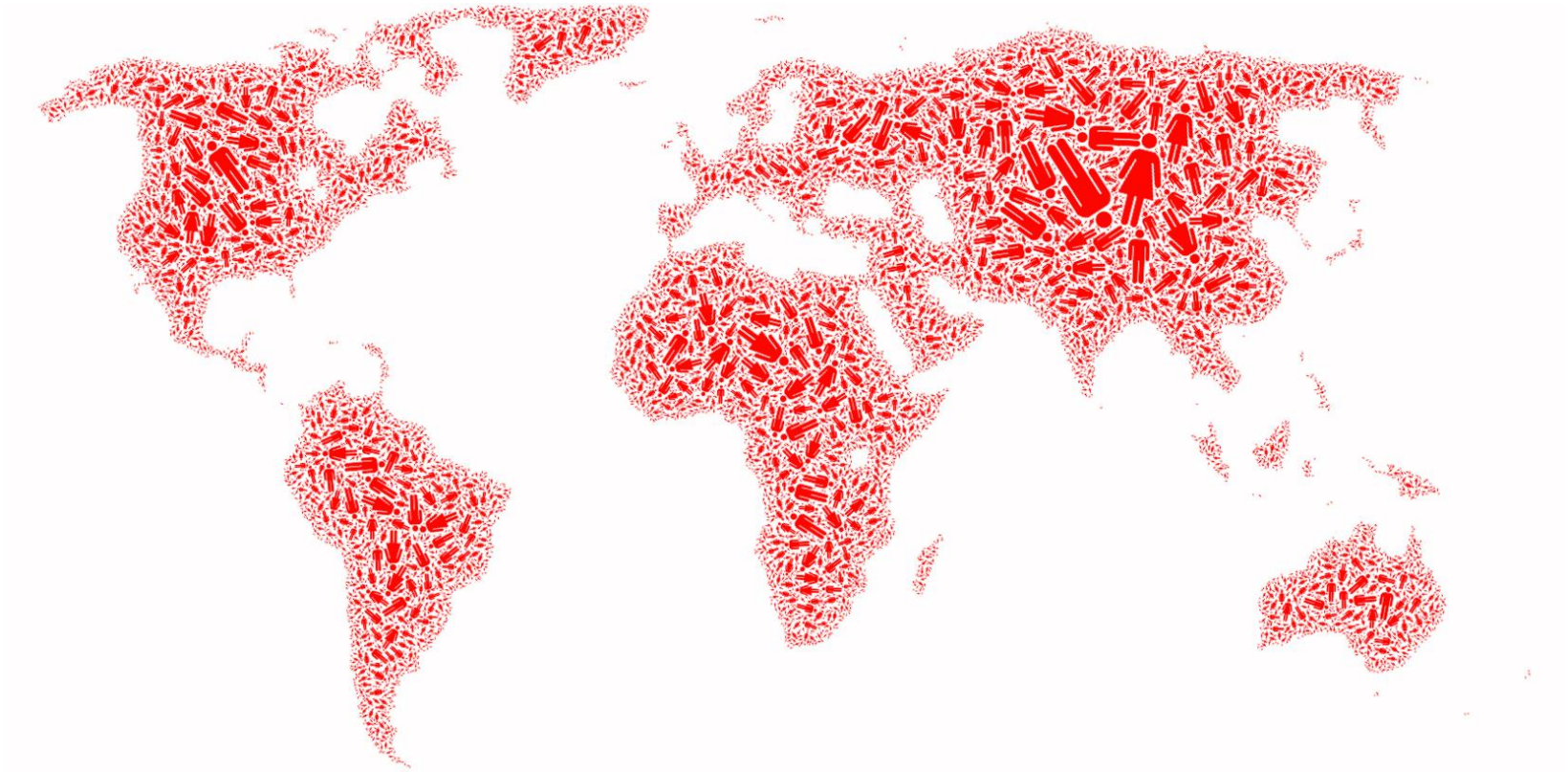


Internet

Friends & Family



Media & Information

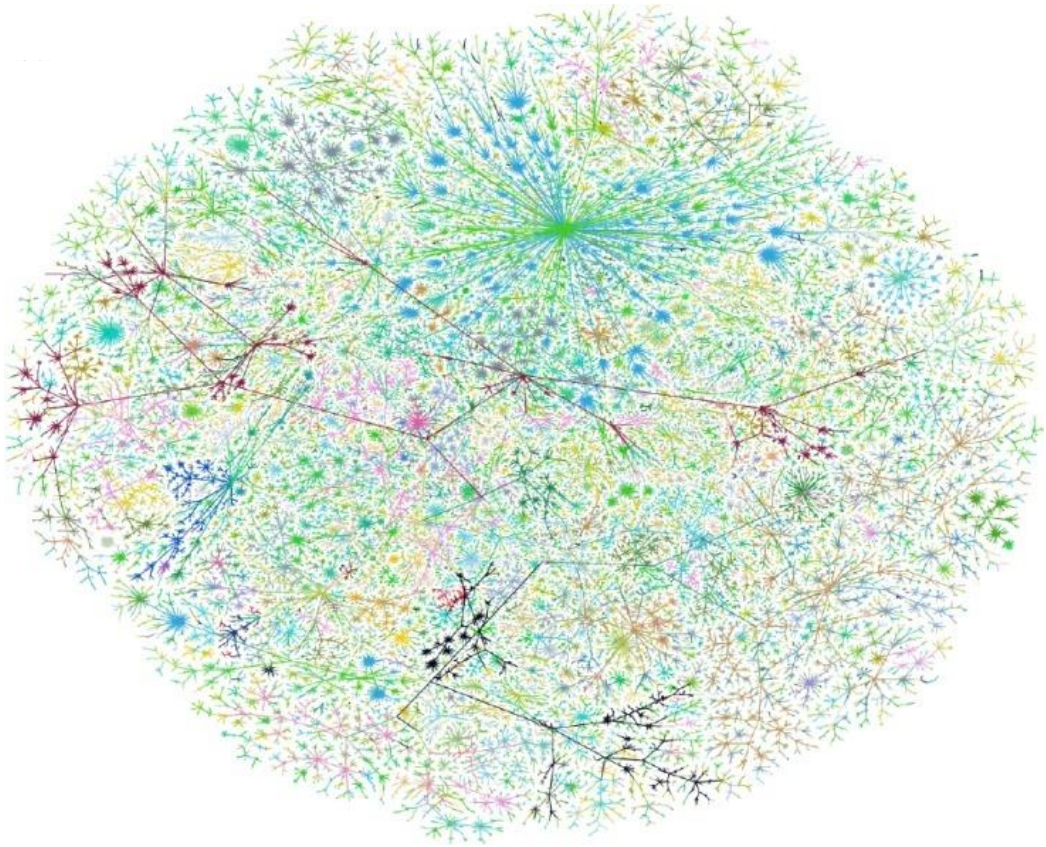


Society

Networks & Complex Systems

- **Hopelessly complex systems are around us:**
  - **Society** is a collection of six billion individuals
  - **Communication systems** link electronic devices
  - **Information** and **knowledge** is organized and linked
  - Thousands of **genes** in our cells work together in a seamless fashion
  - Our **thoughts** are hidden in the connections between billions of neurons in our brain

What do these systems have in common?
How can we represent them?



The Network!

Networks!!

Behind each such system there is an intricate wiring diagram, **a network**, that defines the **interactions** between the components

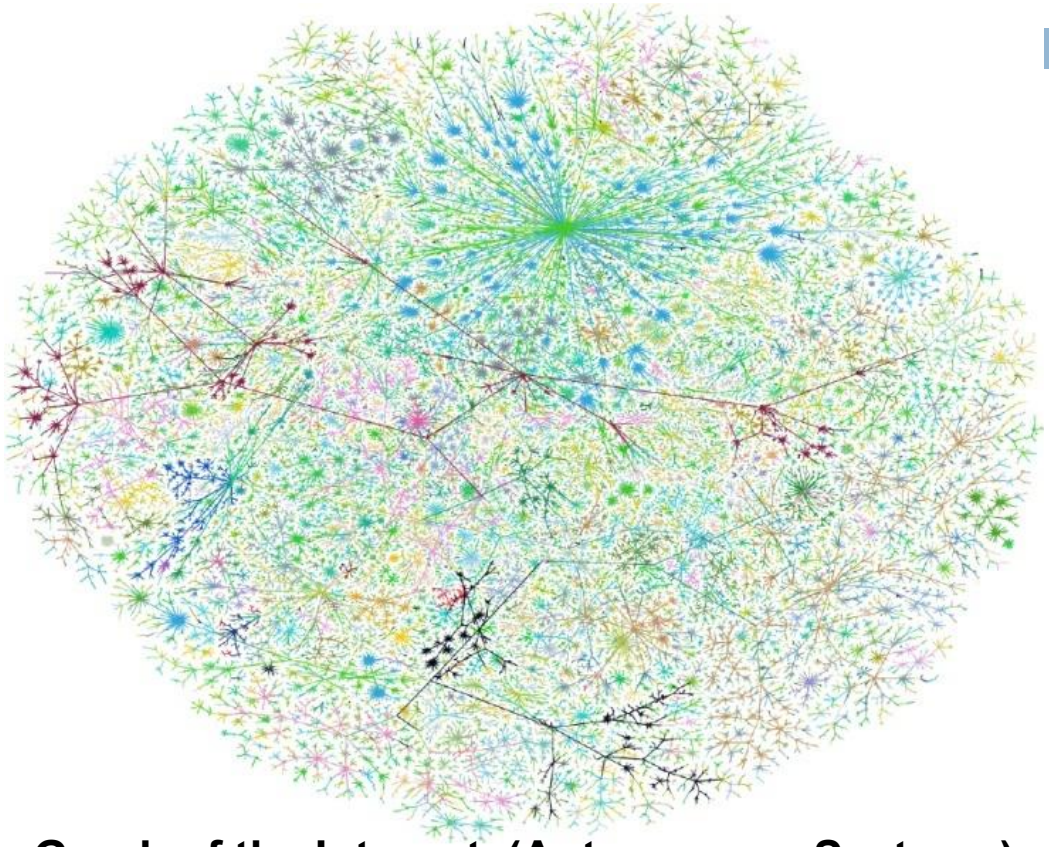
We will never understand these systems unless we understand the networks behind it

Networks: Social

Facebook social graph

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

Networks: Communication

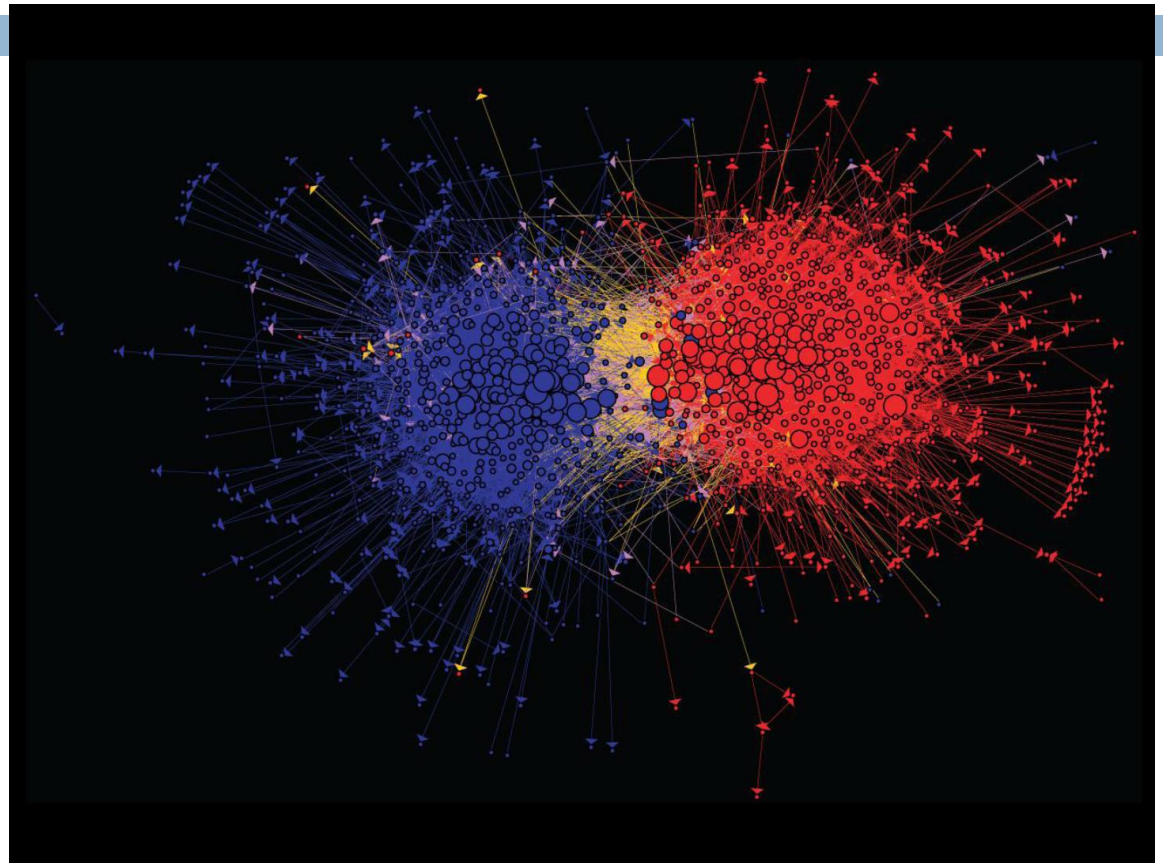


Graph of the Internet (Autonomous Systems)

Power-law degrees [Faloutsos-Faloutsos-Faloutsos, 1999]

Robustness [Doyle-Willinger, 2005]

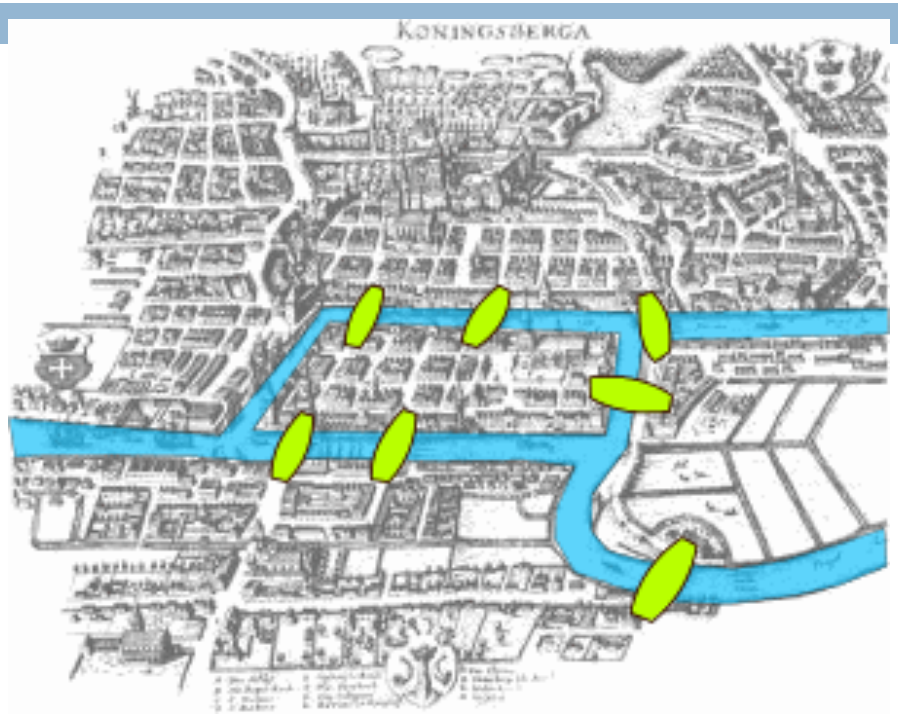
Networks: Media



Connections between political blogs

Polarization of the network [Adamic-Glance, 2005]

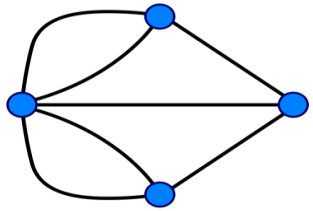
Networks: Technology



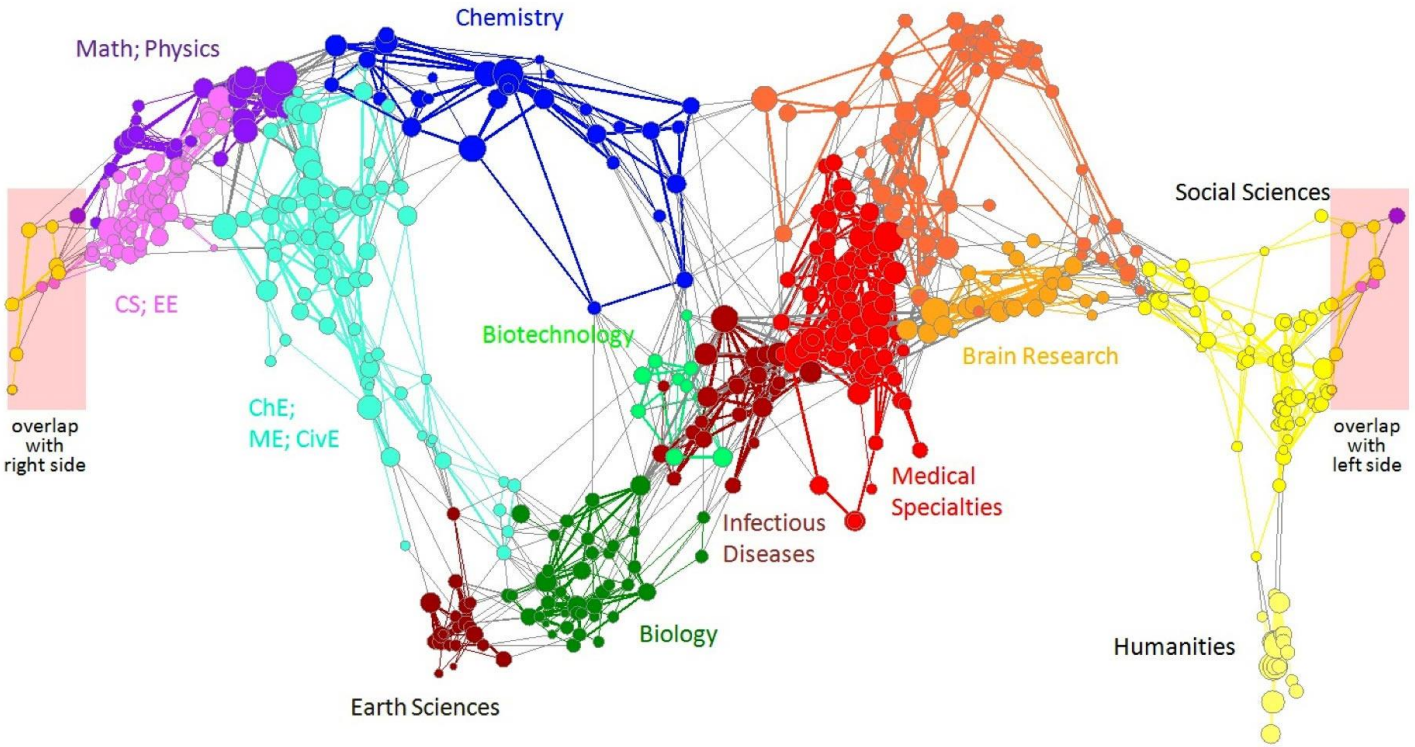
Seven Bridges of Königsberg

[Euler, 1735]

Return to the starting point by traveling each link of the graph once and only once.



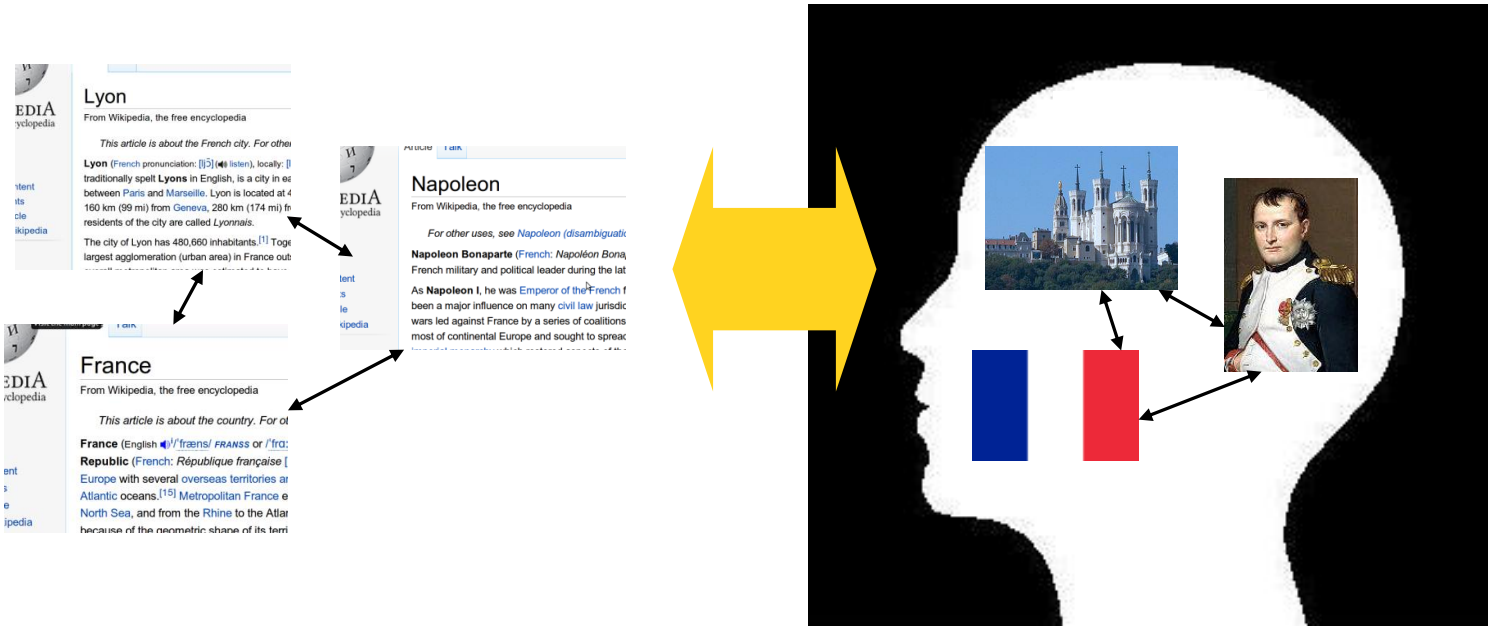
Networks: Information



Citation networks and Maps of science

[Börner et al., 2012]

Networks: Knowledge

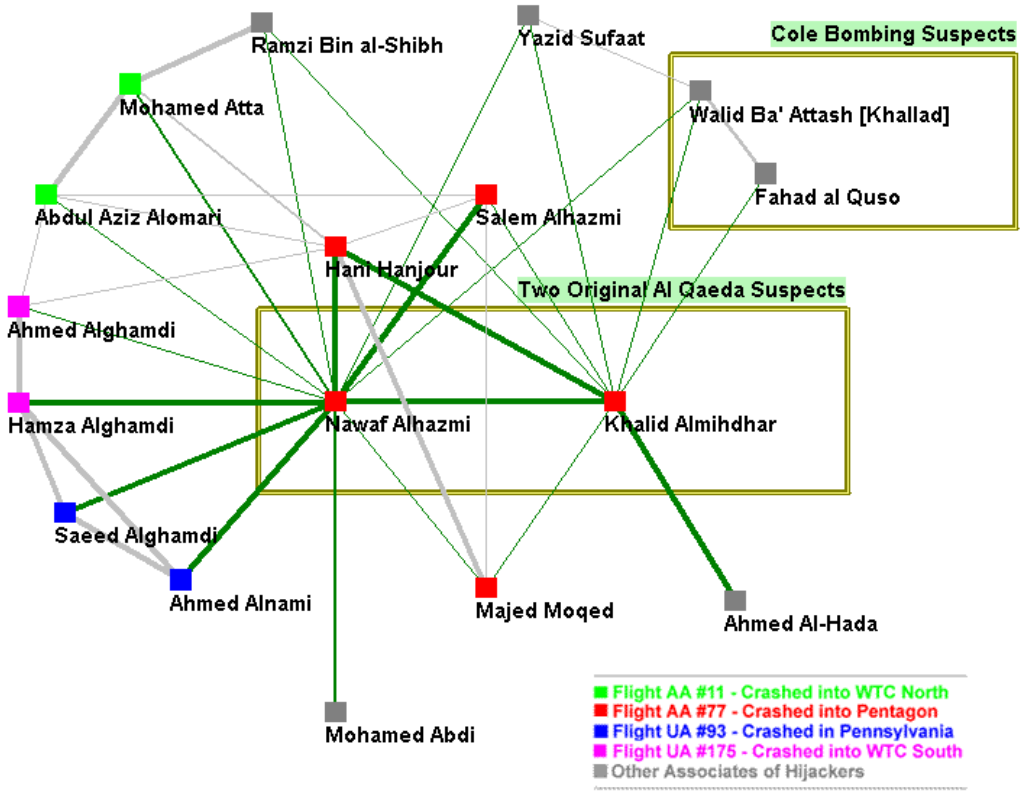


Understand how humans navigate Wikipedia

Get an idea of how people connect concepts

[West-Leskovec, 2012]

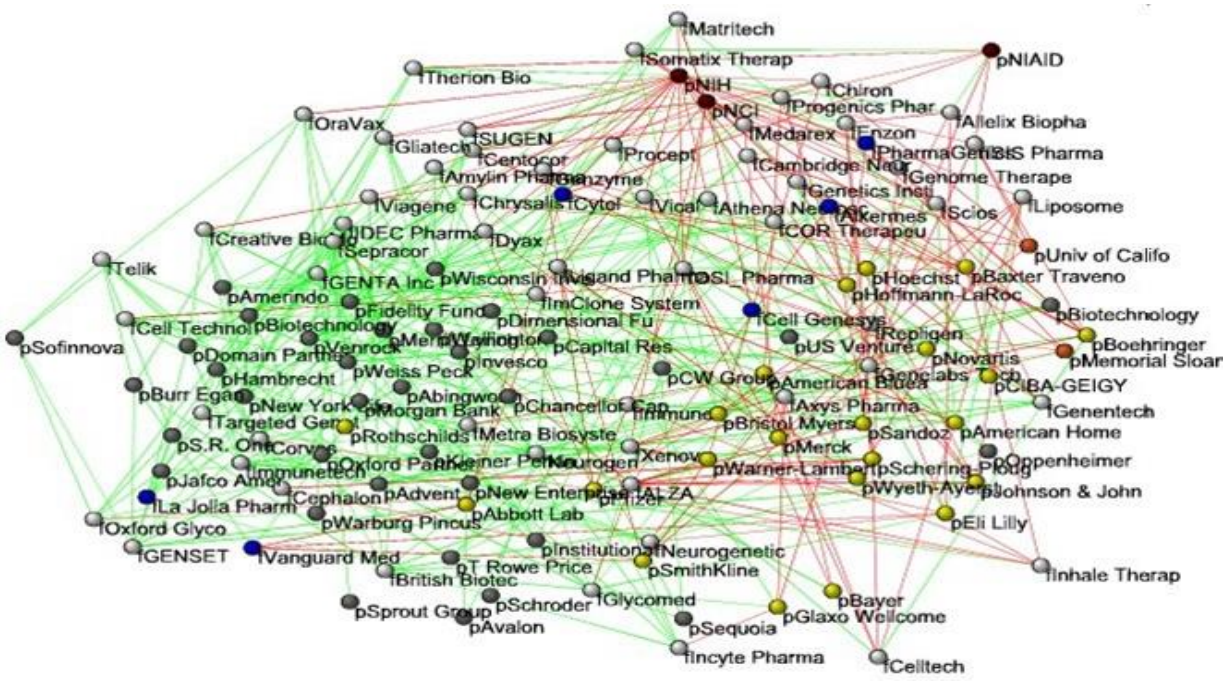
Networks: Organizations



9/11 terrorist network

[Krebs, 2002]

Networks: Economy



Nodes:

- Companies
- Investment
- Pharma
- Research Labs
- Public
- Biotechnology

Links:

- Collaborations
- Financial
- R&D

Bio-tech companies

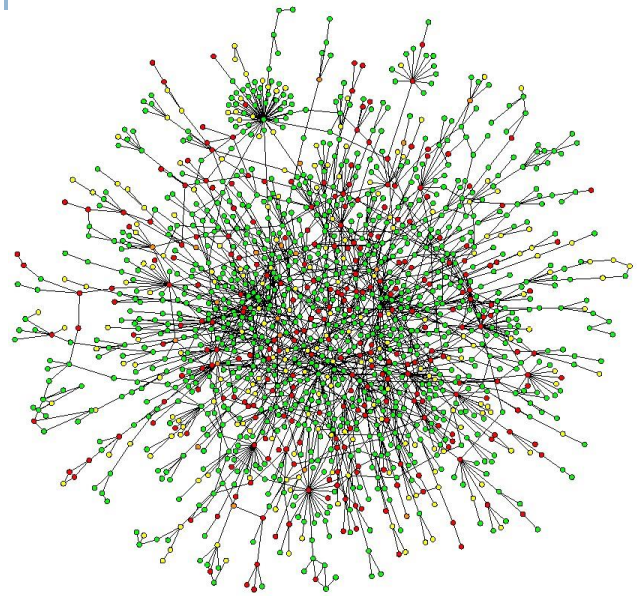
[Powell-White-Koput, 2002]

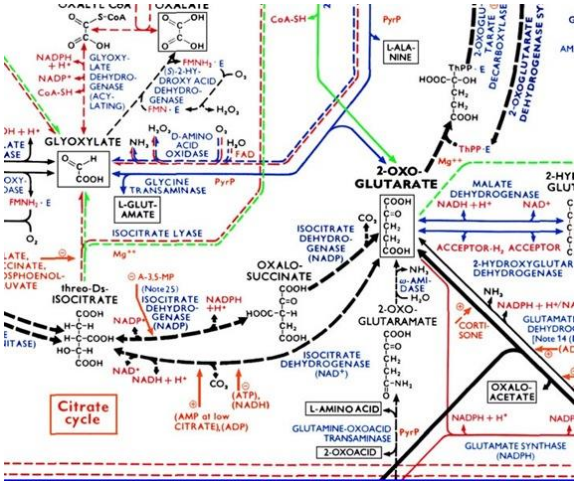
Networks: Brain

Human brain has between 10-100 billion neurons

[Sporns, 2011]

Networks: Biology

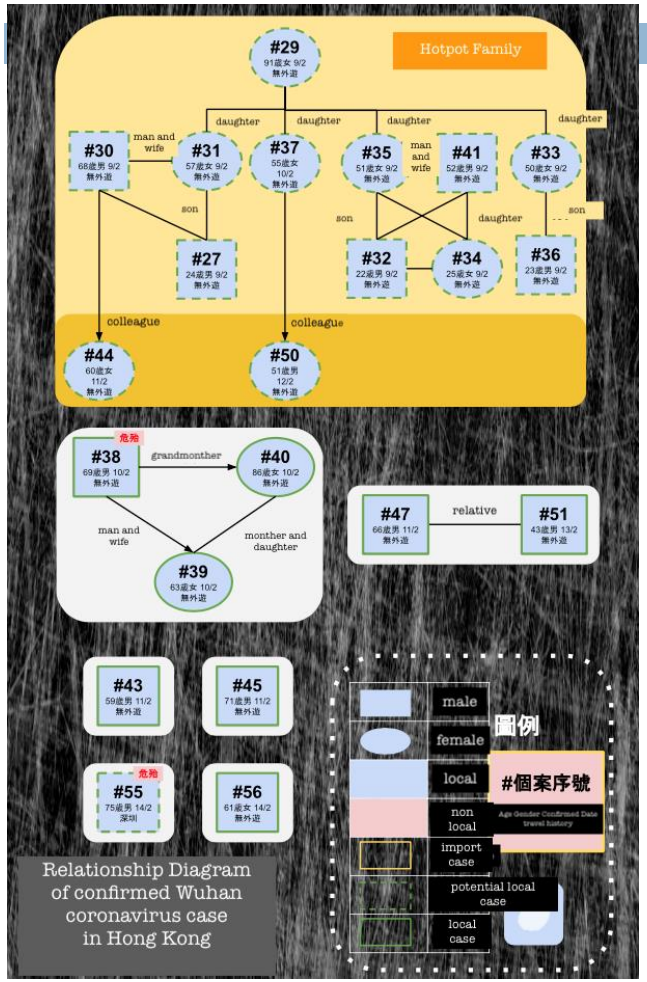


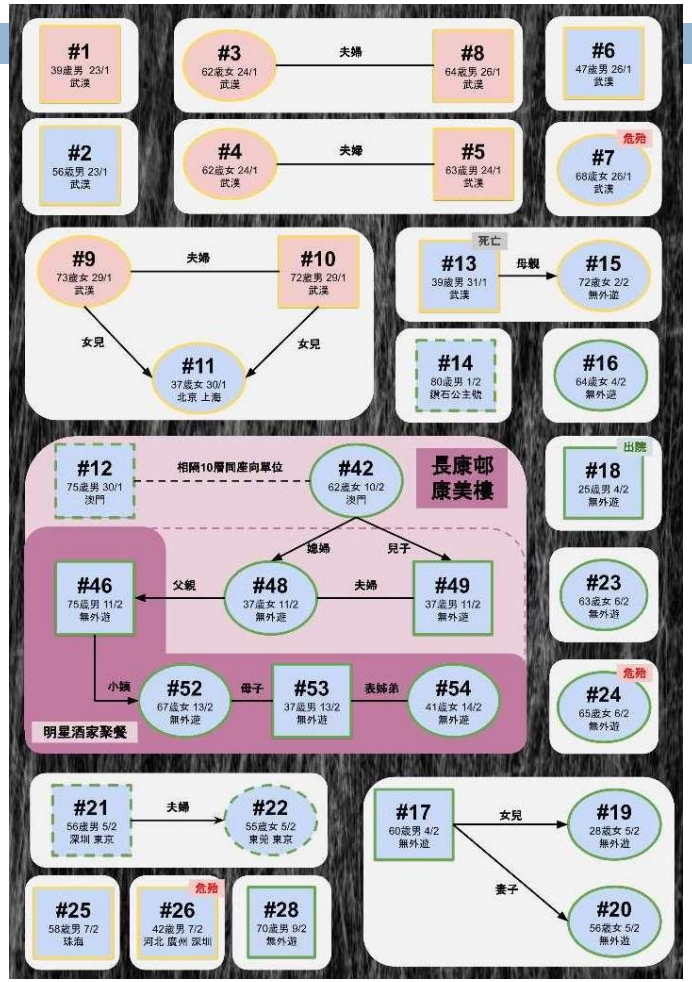


Protein-Protein Interaction Networks:
Nodes: Proteins
Edges: ‘physical’ interactions

Metabolic networks:
Nodes: Metabolites and enzymes
Edges: Chemical reactions

Networks: Epidemic (COVID-19)





Reasoning about Networks

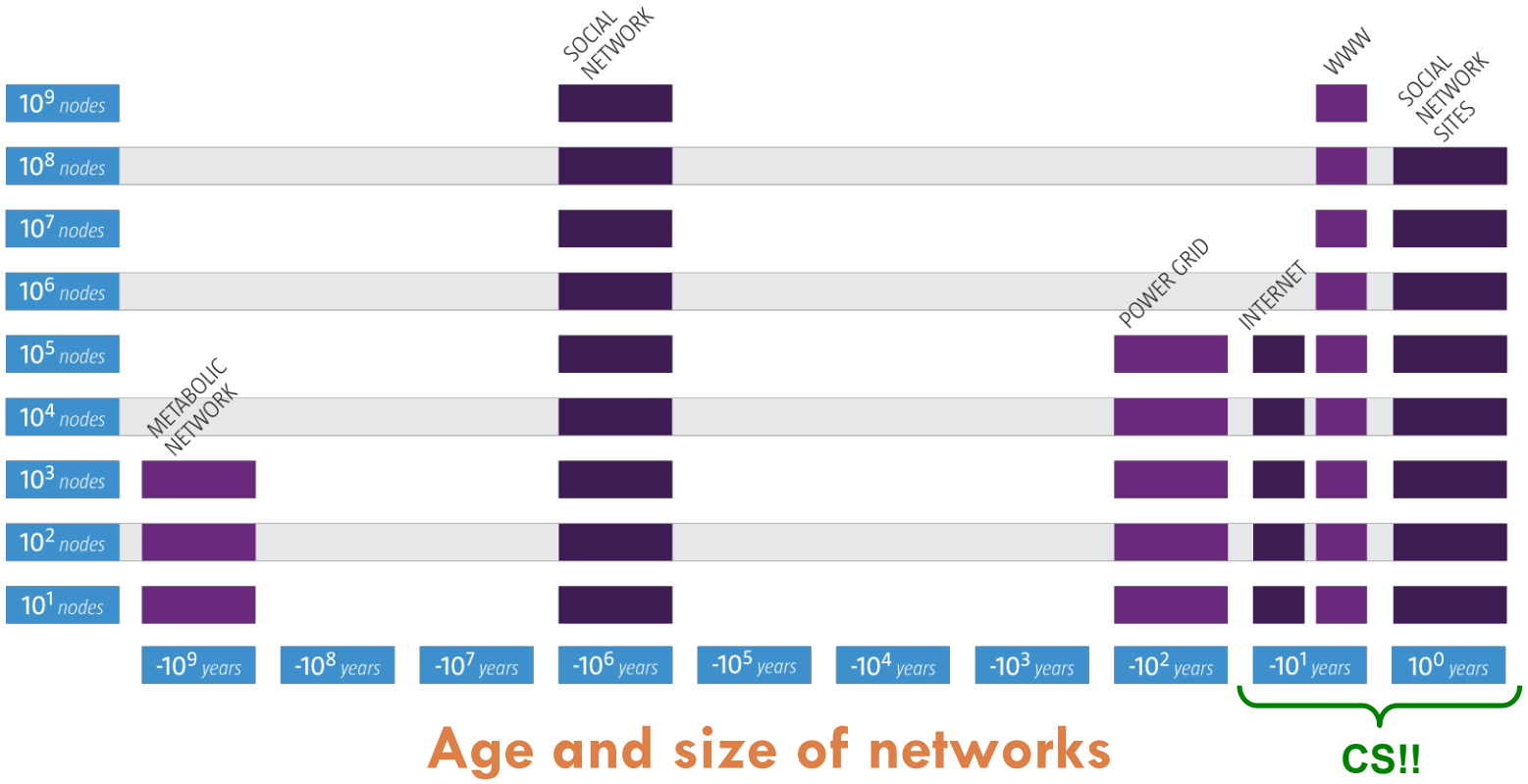
- **How do we reason about networks?**
  - **Empirical:** Study network data to find organizational principles
  - **Mathematical models:** Probabilistic, graph theory
  - **Algorithms** for analyzing graphs
- **What do we hope to achieve from studying networks?**
  - Patterns and statistical **properties** of network data
  - **Design principles** and **models**
  - **Understand** why networks are organized the way they are (Predict behavior of networked systems)

Why Networks? Why Now?

Why is the role of networks expanding?

- **Data availability**
  - Rise of Mobile, Web 2.0 and Social media
- **Universality**
  - 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
- **Impact!**
  - Social networking, Social media, Drug design

Networks: Why Now?



Networks: Size Matters

- **Network data: Orders of magnitude**
  - 436-node network of email exchange at a corporate research lab [Adamic-Adar, SocNets '03]
  - 43,553-node network of email exchange at an university [Kossinets-Watts, Science '06]
  - 4.4-million-node network of declared friendships on a blogging community [Liben-Nowell et al., PNAS '05]
  - 240-million-node network of communication on Microsoft Messenger [Leskovec-Horvitz, WWW '08]
  - 800-million-node Facebook network [Backstrom et al. '11]

Web – The Lab for Humanity

Networks: Impact

- **Alphabet**
  Market cap: $1.961 trillion
- **Cisco**
  Market cap: $232 billion
- **Facebook**
  Market cap: $677 billion

Networks Really Matter

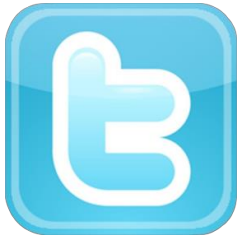
- If you were to understand the spread of diseases, **can you do it without social networks?**
- If you were to understand the WWW structure and information, **hopeless without invoking the Web's topology.**
- If you want to understand dissemination of news or evolution of science, **it is hopeless without considering the information networks**

Social Computing

- The next generation could be the one with access to an unprecedented amount of behavioral data
- This can solve real problems

  … not just finding a movie or a restaurant

  - ensuring energy efficiency
  - monitoring our environment
  - reduce inequality
  - informing social decision



Only convinced by numbers?

+40%

- How much data production grows / year
  - Enough to double every 24months

(>500hours of videos upload on YouTube in 1 min).

€260b

- How much data can save on health care
  - In Europe [McKinsey] (U.S. save $300b)

+300-1000%

- How much lifts improve when ads are using behavioral targeting

What are Social Networks?

- Large set of ***personal information*** about users
  - History of Browsing, Purchasing, Rating
  - Sociological profile (age, gender, location, income)
  - Community of interests
- Large set of ***relational information*** about users
  - Connections (friendship, collaboration, schoolmate)
  - Contacts (email IM phone calls etc., meeting)

A key principle

- What *primarily* matters is your social environment!
  - For Business: how to best advertise a product?
  - For Media: how to find most relevant information?
  - For Engineers-CS: how to best design an application?
  - For Science and Society at large: how to understand human behavior? Take advantage of it?

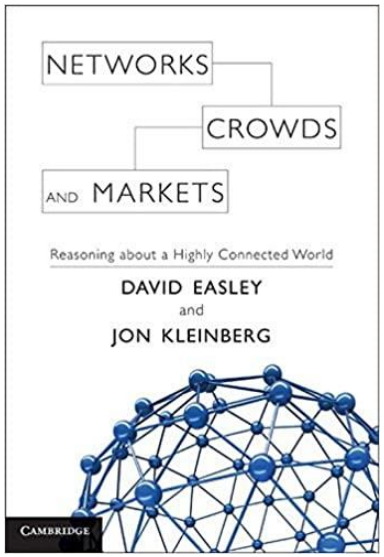
    …

… 4 (classical) questions, being reinvented *today*

COURSE LOGISTICS

Course Logistics

- We will post course announcements to **Canvas** (Please check regularly for updates)
- Slides posted at least 30 min before the class
- **Readings:**
  - Many chapters from Easley and Kleinberg
  - Papers
- **Optional readings:**
  - Papers and pointers to additional literature
  - **This will be very useful for project proposals**
- **Teacher Assistant:**
  - Reza HADI MOGAVI (rhadimogavi@connect.ust.hk)



Logistics: Communication

- **Instructor: Prof. Pan Hui**
  - Email: panhui@ust.hk
  - Office: Room 4338
  - Office Hours: By appointment
  - http://www.cse.ust.hk/~panhui

- **For Q&A**
  - Send emails to Reza: rhadimogavi@connect.ust.hk

Work for the Course & Grading

- **Final grade will (tentatively) be composed of:**
  - **Homeworks: 30%**
    - Homeworks 1,2,3: 10% each
  - **Substantial class project: 70%**
    - Proposal: 10%
    - Project milestone: 10%
    - Presentation: 10%
    - Final report: 40%
  - Extra credit for camera turned-on (during online-classes) and active class participation: 10%

Course Schedule (tentative)

Week	Assignment	Due on
6	Homework 1	March 9
7	**Project proposal**	**March 21**
10	Homework 2	April 1
	Work on the project	
11	**Project milestone**	**April 18**
13	Homework 3	May 2
14	**Final report** (no late days!)	May 27
	Project presentation	*Option 1: Last 2 or 3 lectures* *Option 2: After the exams*

Homeworks, Write-ups

- **Assignments take time. Start early!**
- **How to submit?**
  - **Canvas:** Assignments, and project write-ups (proposal, milestone, final report) have to be submitted electronically
  - Max 1 late day per assignment (will receive no marks after that)

Course Projects

- **Substantial course project:**
  - **Experimental evaluation** of algorithms and models on an interesting network dataset
  - A **system project** that involves software implementation of social information services or applications
  - A **theoretical project** that considers a model, an algorithm and derives a rigorous result about it
  - Develop **scalable algorithms** for massive graphs or large-scale social information systems
- **Performed in groups of 3 students**
- Project is the **main work** for the class

Prerequisites

- **Basic background in:**
  - Algorithms
  - Graph theory
  - Probability and Statistics
  - Linear algebra
- **Programming:**
  - You should be able to write non-trivial programs
- **Lab and tutorial sessions:**
  - Review programming tools (SNAP, NetworkX)
  - Review basic mathematical concepts
  - Review social networking services

Course Syllabus

Introduce properties, models and tools for

- Large real-world networks
- Processes taking place on networks

through real applications and case studies

- **Goal:** find **patterns, rules, clusters, outliers, …**
  - … in large static and evolving graphs
  - … in processes spreading over the networks
  - … in scalable computer networking system design

Course Syllabus

- Covers a wide range of **network analysis techniques** – from basic to state-of-the-art
- **You will learn about things you heard about:**

  Six degrees of separation, small-world, page rank, network effects, P2P networks, network evolution, virus propagation, link prediction, power-laws, scale free networks, core-periphery, network communities, hubs and authorities, bipartite cores, information cascades, influence maximization, tipping points, social engineering, altruism, malicious behaviors, mobile social networks, graph neural network …

- **Covers algorithms, theory, system and applications**
- **It's going to be fun**