Network Analysis:

The Hidden Structures behind the Webs We Weave

17-213 / 17-668

Introduction & Course Overview

Tuesday, August 29, 2023

Patrick Park & Bogdan Vasilescu

Carnegie Mellon University
School of Computer Science

S3D
Software and Societal Systems Department

Networks are old news

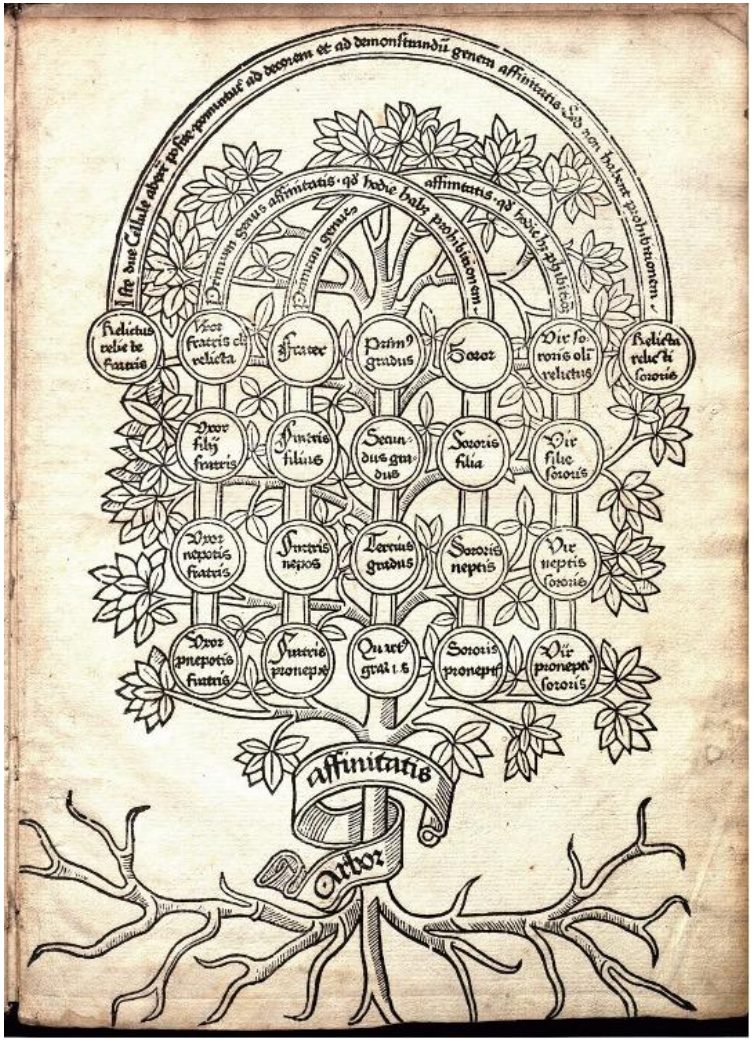
(networks: patterns of interconnections among a set of things)

Family tree of Marie de Medicis by Lucas Varsterman (1632)

Marie de Medici and her five children

Tree of Affinity by Johannes Andreae (1270-1348)

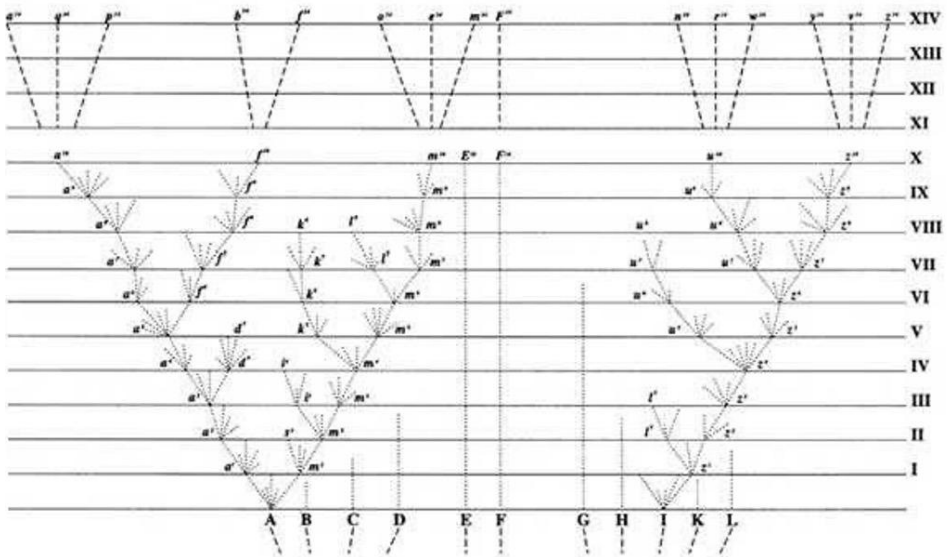
Maps laws and regulations on kinship and marriage decreed by the ecclesiastical authority of the Catholic Church.



Tree of Life by Charles Darwin

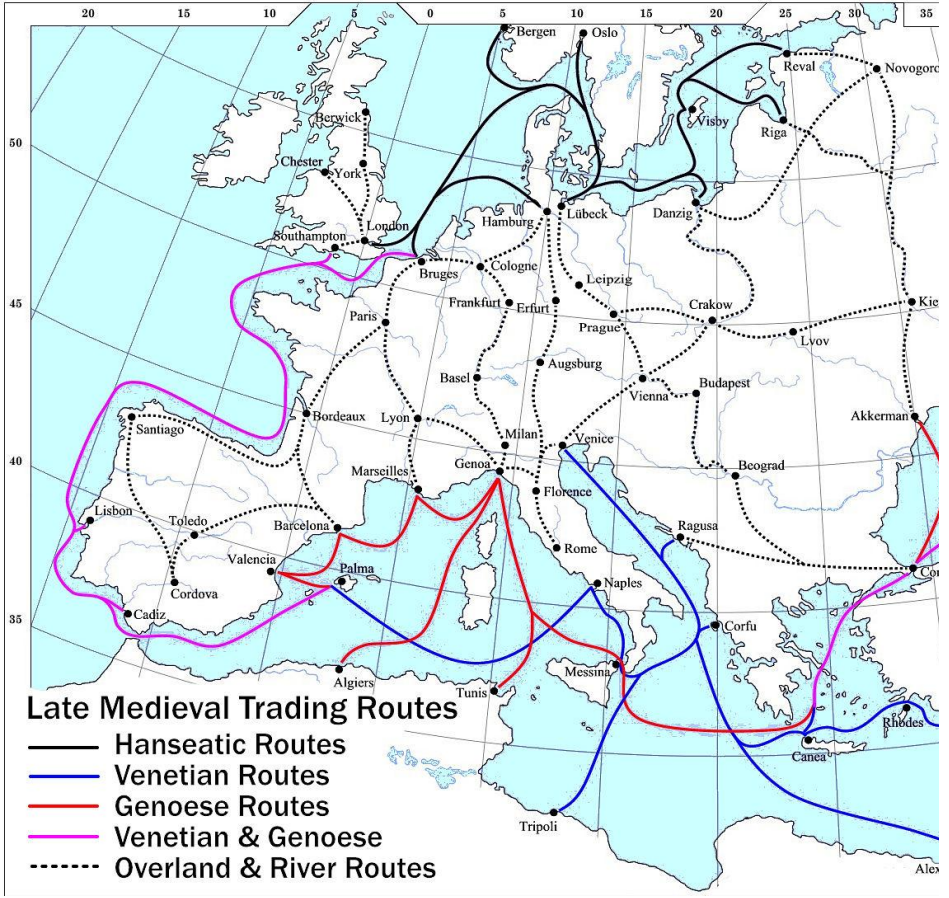
“A demonstration of Darwin’s evolutionary thinking and the theory of universal common descent.”

(Lima 2013)



Medieval trade routes in Europe.

(Easley & Kleinberg, 2010; image by Lampman, Public domain, via Wikimedia Commons)



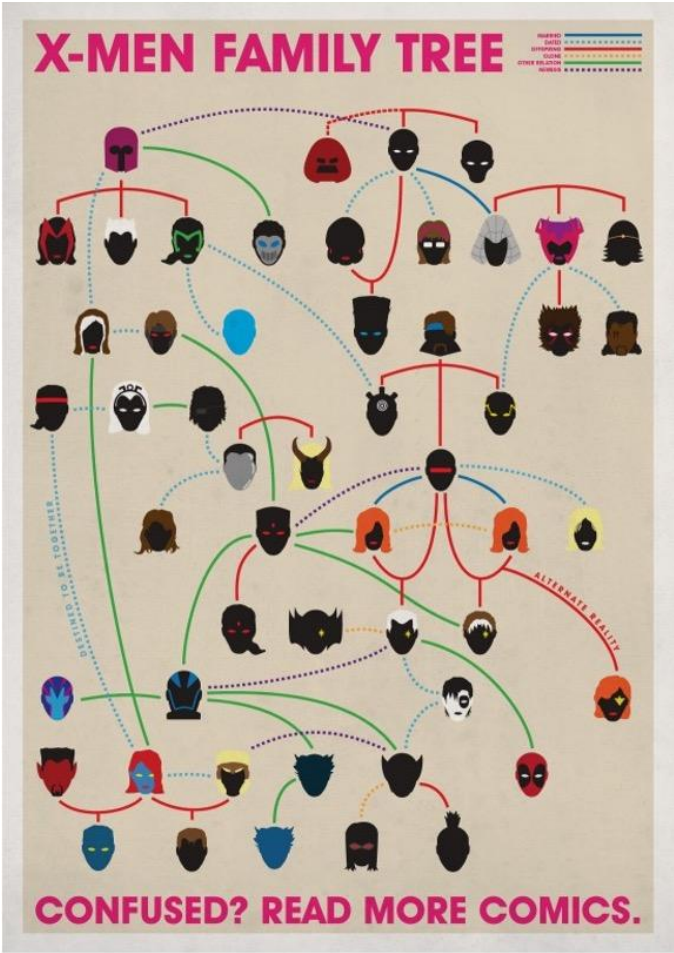
But networks are also recent news!

… sometimes for fun and art

X-Men Family Tree by Joe Stone

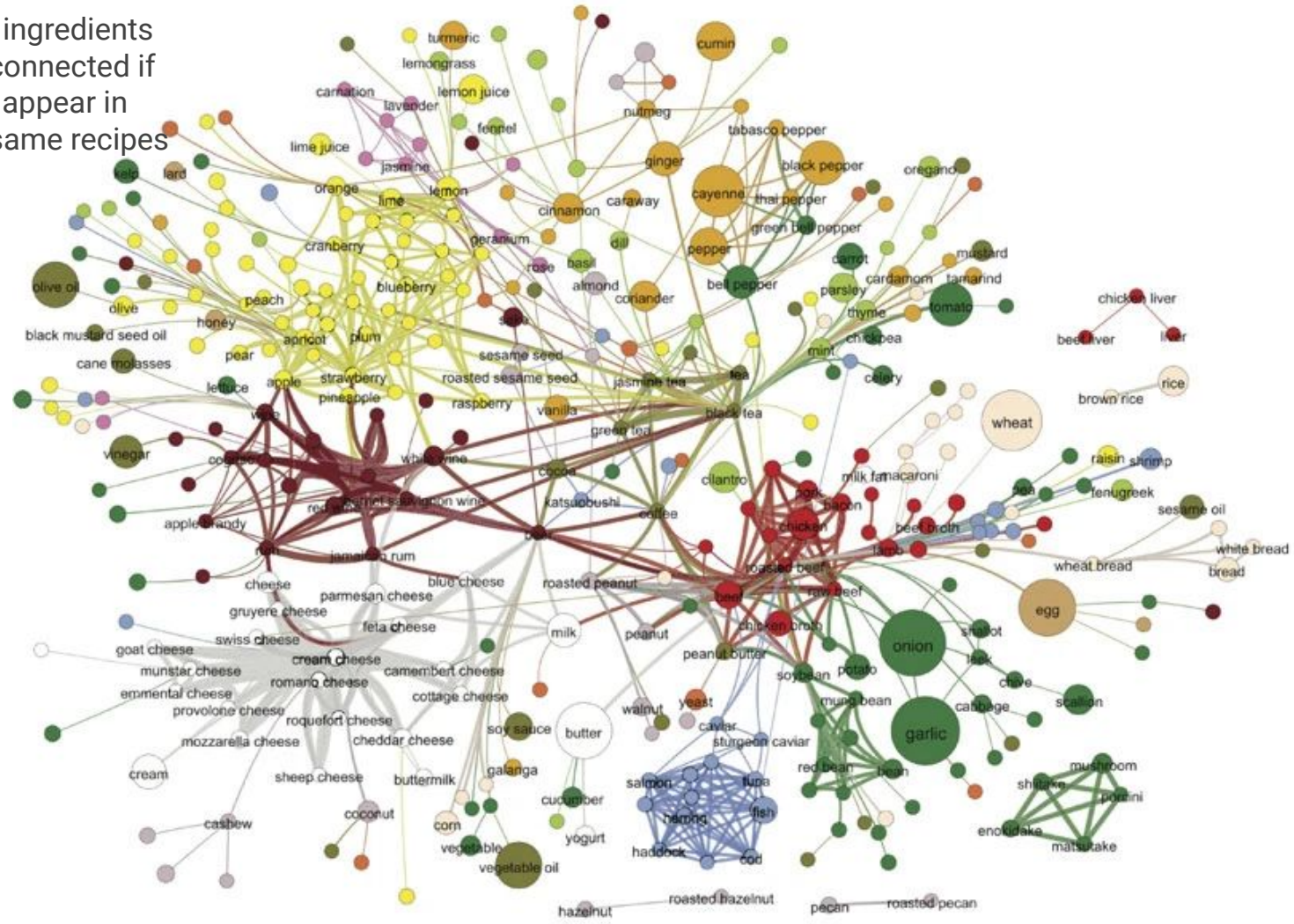
“An enticing and playful family tree charting the many convoluted relationships – romantic, genetic, or otherwise – of the X-Men characters from Marvel Comics.”

(Lima, 2013)




The international journal of science / 7 November 2019
nature
NATURE'S CO-CITATION NETWORK
Here, more than 88,000 papers published by *Nature* since 1900 are each represented by a dot, coloured by discipline. Papers are linked if another scientific paper cites both; the dot size reflects the number of these co-citation links. The complex network reveals the relationships between papers and captures the multidisciplinary scope of the journal.
150 YEARS OF NATURE
A web of multidisciplinary research and discovery
EXPLORE INTERACTIVE NETWORK
Arts
Biology
Biomedical research
Chemistry
Clinical medicine
Earth and space
Engineering and technology
Health
Humanities
Mathematics
Physics
Business and management
Psychology
Social sciences

Two ingredients are connected if they appear in the same recipes



Categories

- fruits
- dairy
- spices
- alcoholic beverages
- nuts and seeds
- seafoods
- meats
- herbs
- plant derivatives
- vegetables
- flowers
- animal products
- plants
- cereal

Prevalence

50 %
30 %
10 %
1 %

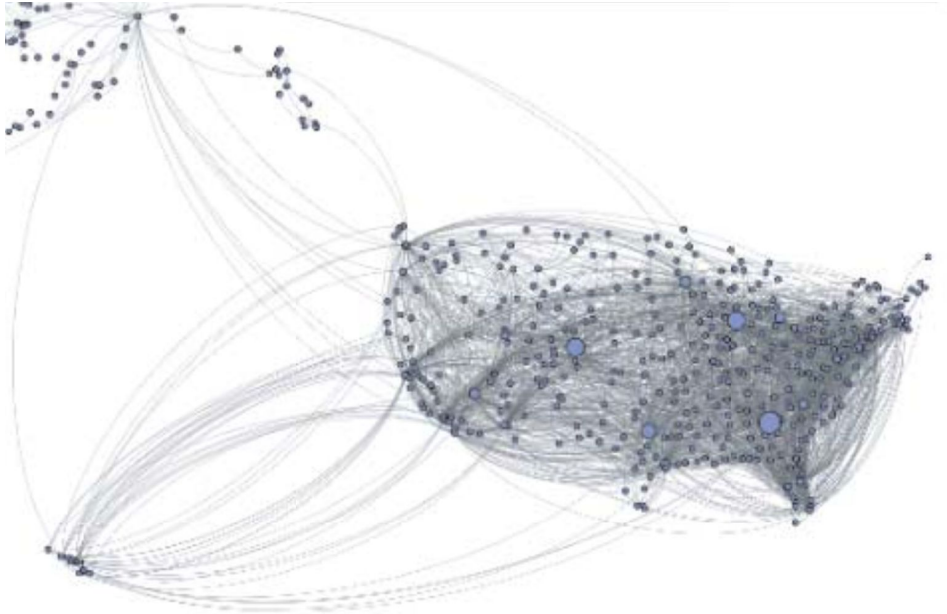
Shared compounds

150
50
10

… but mostly for serious applications

The US air transportation network. Note the “hub and spoke” structure: a few hubs have huge numbers of links, while the majority of nodes have few connections.

(Menczer, Fortunato, & Davis, 2020)

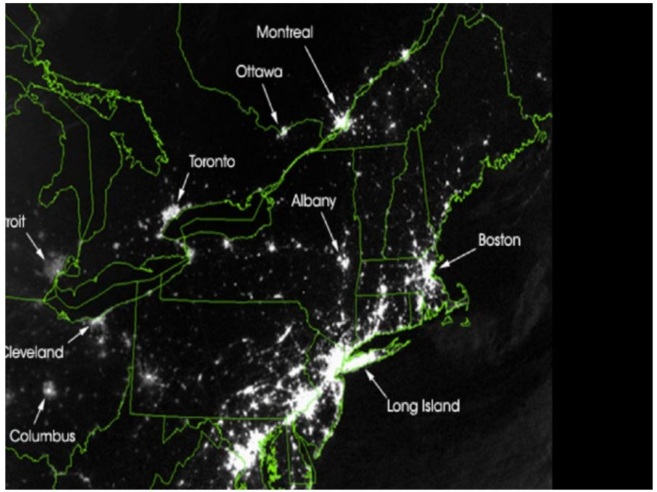


The 2003 Northeast blackout that left without power an estimated 45 million people in eight US states and another 10 million in Ontario.

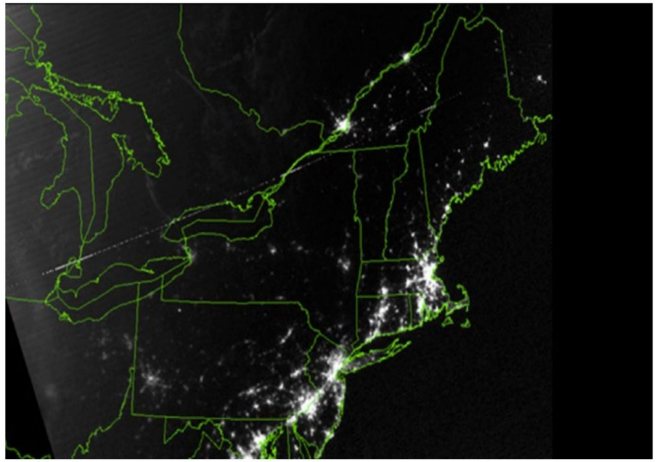
(a) Satellite image 20 hours before
(b) … and 5 hours after

(Barabasi, 2016)

(a)

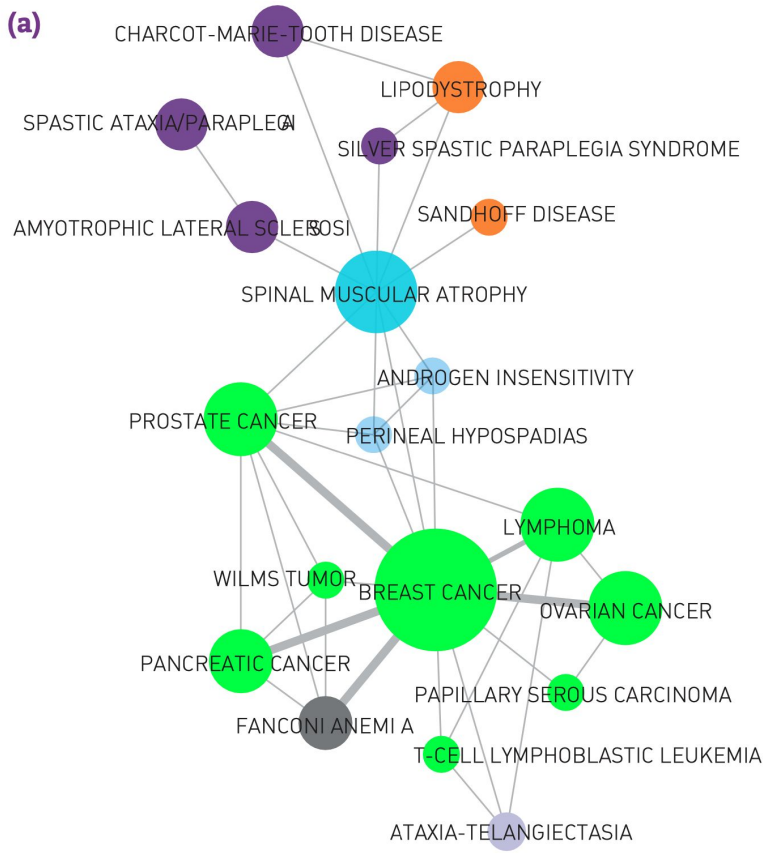


(b)



A subset of the “diseaseome,” i.e., a network where two genes are connected if they are associated with the same disease.

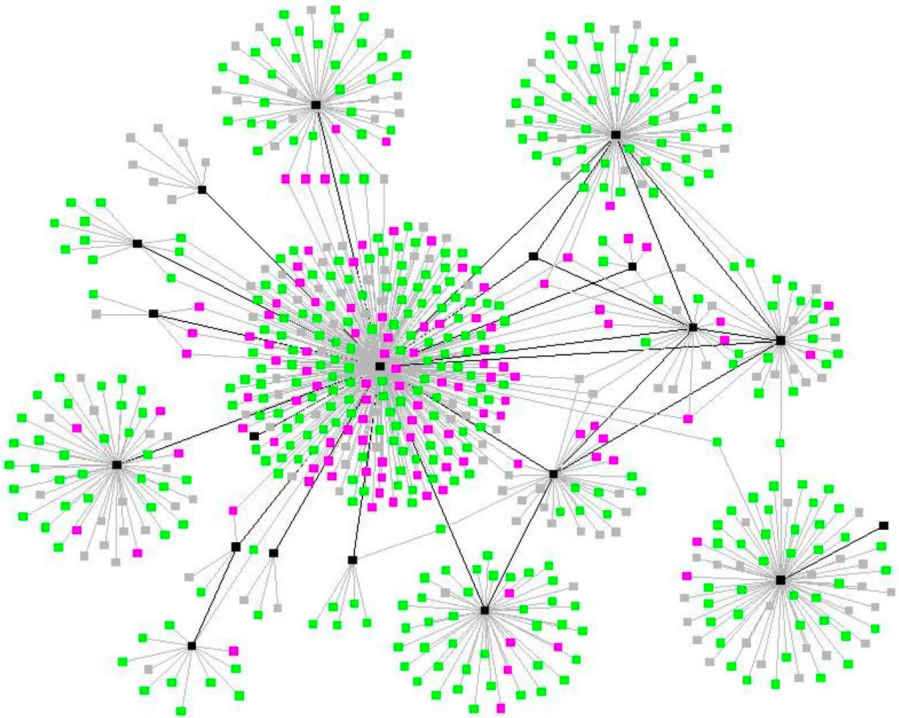
(Barabasi, 2016)



HUMAN DISEASE NETWORK

The spread of a tuberculosis as a form of *cascading behavior* in a network.

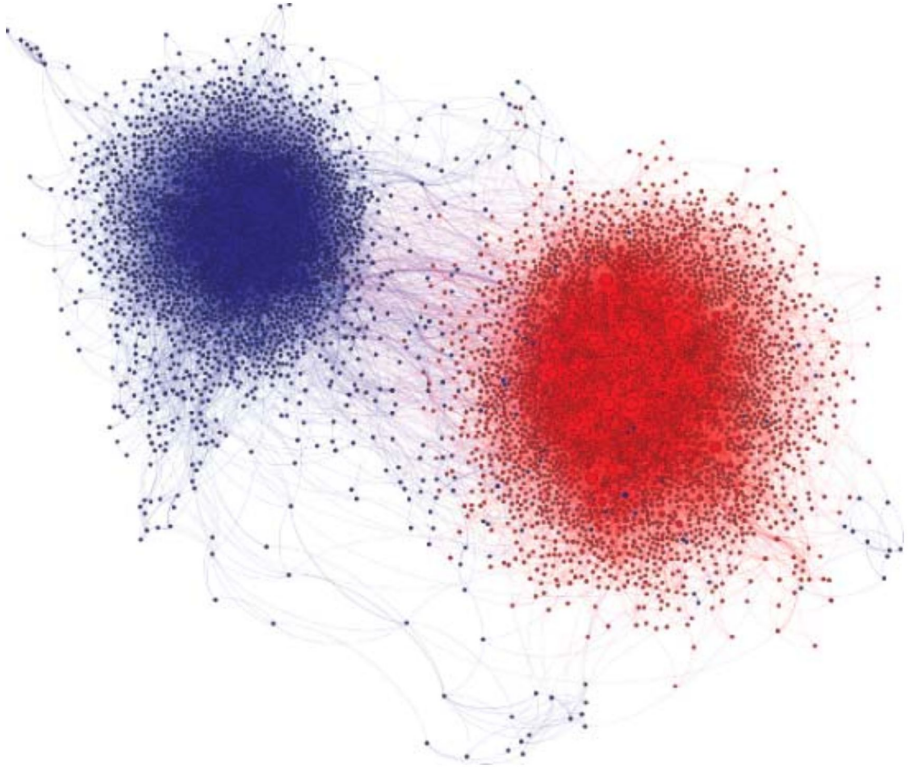
(Easley & Kleinberg, 2010)



Wood-wide web: Trees transmit danger messages and nutrition through fungi connections

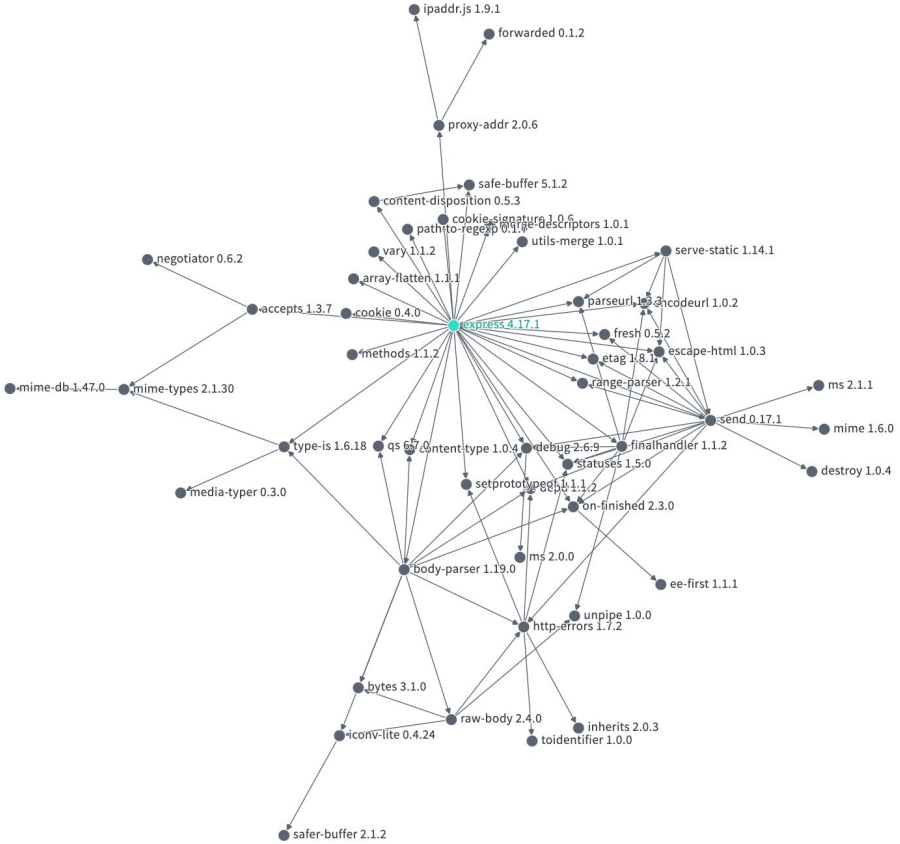
A retweet network on Twitter, among people sharing posts about US politics. Conservative users (red) mostly retweet messages from other conservatives, and vice versa.

(Menczer, Fortunato, & Davis, 2020)



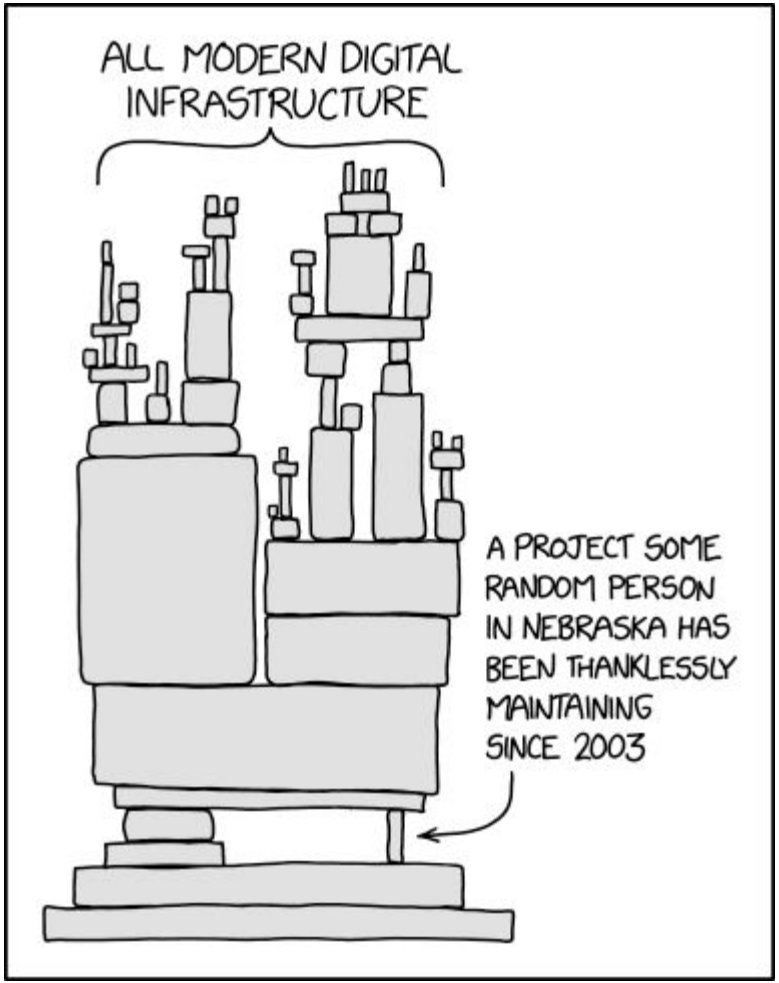
Dependency graph for the open-source framework express 4.17.1.

(Google Open Source Insights Project, 2021)



How resilient and sustainable is the overall open source ecosystem?

(xkcd 2347: Dependency)



Networks

Are everywhere.

Complex systems (social, economic, technological, biological, etc).

These systems are difficult to understand and their behavior increasingly difficult to reason about. Are risky to tinker with.

Network Science

Has a long history of scholarship.

Has become a mature interdisciplinary area that helps to explain many real-world phenomena and solve many real-world problems involving networks.

Computer Science

Mathematics

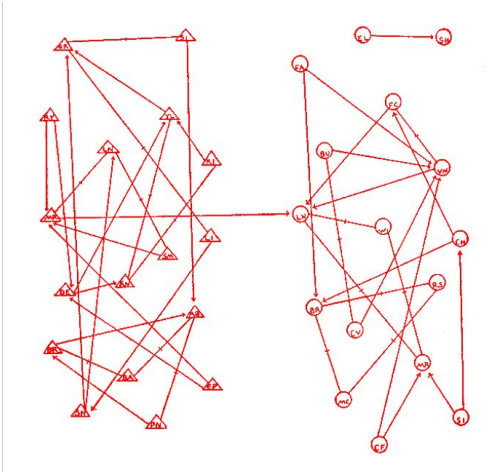
Physics

Social Science

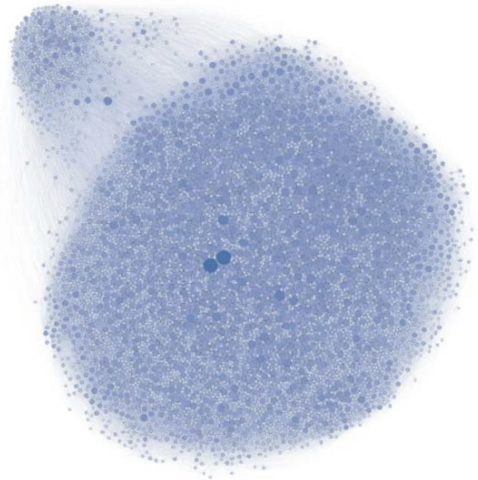
Statistics

Modern network science is propelled by computational power

Network maps



Hand-drawn friendship ties of a classroom (Moreno 1934)



Software-generated network visualization of Facebook friends at a university

Modern network science is propelled by computational power

Mathematical abstractions, algorithms, and software tools to compute on large networks of millions of nodes

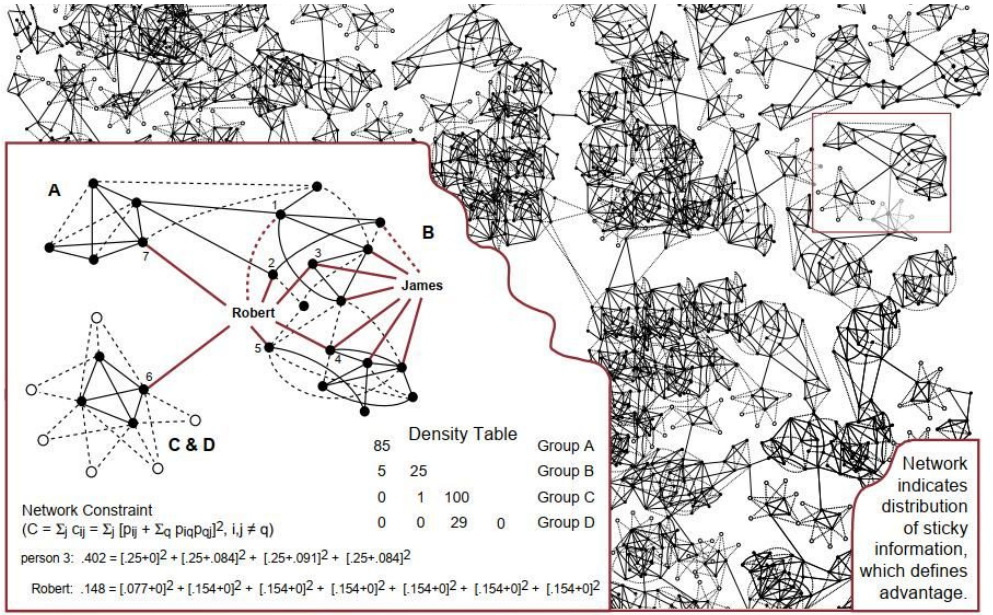


Figure 1. Network Bridge and Cluster Structure

Adapted from Burt (2005:14).

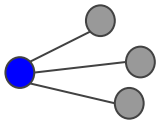
We are going to learn about, and practice using, many of them!

Main abstraction: Networks as graphs

Graph Basics: The Demography of a Graph

Node (or vertex, “N”): An entity (person, airport, neuron, etc.) 

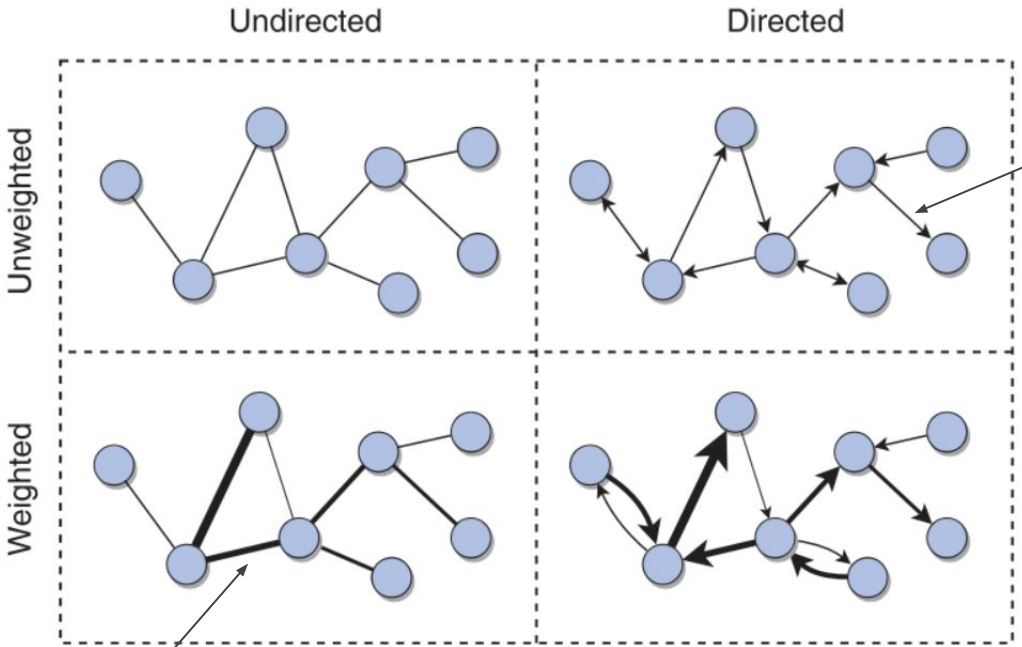
Edge (or link, “L”): Connection between two nodes 

Degree (“k”): Number of edges that a node has to other nodes 

Density: Number of edges (L), relative to the maximum possible edges

- Maximum possible edges: N(N-1)/2
- Density = L / (N(N-1)/2)

Graphs of interest: Four types



Arrow indicates directional relationship (e.g., advice giving)

Thickness represents edge weight (e.g., strength of a social relationship)

All sorts of relationships can be represented this way!

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

(Barabasi, 2016)

Class Philosophy

Focus on networks involving humans:
"The Hidden Structures behind the Webs **We** Weave"

Pragmatism: Prioritize analysis and hands-on experience with theoretical guidance from different fields
"Network ~~Science~~Analysis"

Unapologetically at the confluence of two historically disconnected schools of thought:
Social science ~~vs.~~and Natural science

Two approaches to Network Science: universality vs. variation

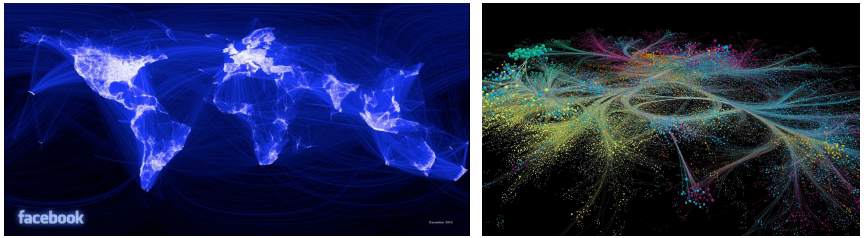
Natural sciences (e.g., physics, CS, mathematics) typically focused on

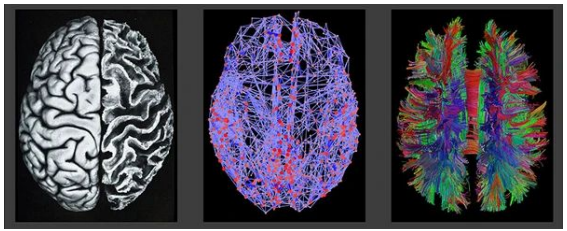
- Universality: search for common organizing principles across domains
- Level of analysis: entire network
- Study the networks as complex systems

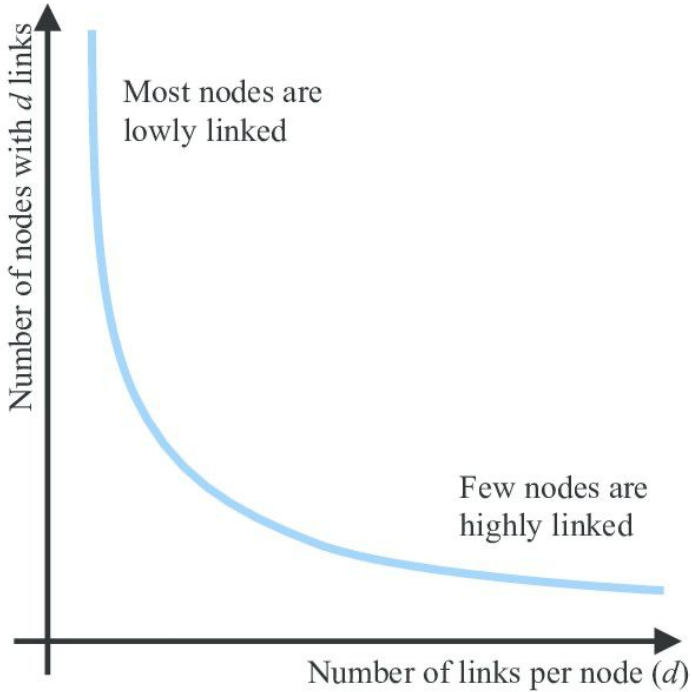
A key discovery of network science is that the architecture of networks emerging in various domains of science, nature, and technology are similar to each other, a consequence of being ***governed by the same organizing principles****. Consequently, we can use a common set of mathematical tools to explore these systems*. (Barabasi 2015)

Example of Universality

Understand **“universal” characteristics** common to networks in different domains







Two approaches to Network Science: universality vs. variation

Social sciences (e.g., sociology, political science, economics) typically focused on

- Connections between people through role relations (e.g., kinship, friendship)
- Variation: in network structure and social contexts that explain differences
- Level of analysis: Individual (node), social tie (edge), entire network

The social network approach is grounded in the intuitive notion that the patterning of social ties in which actors are embedded has important consequences for those actors. Network analysts, then, seek to ***uncover various kinds of patterns****. And they try to determine* ***the conditions under which those patterns arise*** *and to discover* ***their consequences****.* (Freeman 2004)

Example of embracing the variation

Distinguishing among different kinds of dyadic links both analytically and theoretically.

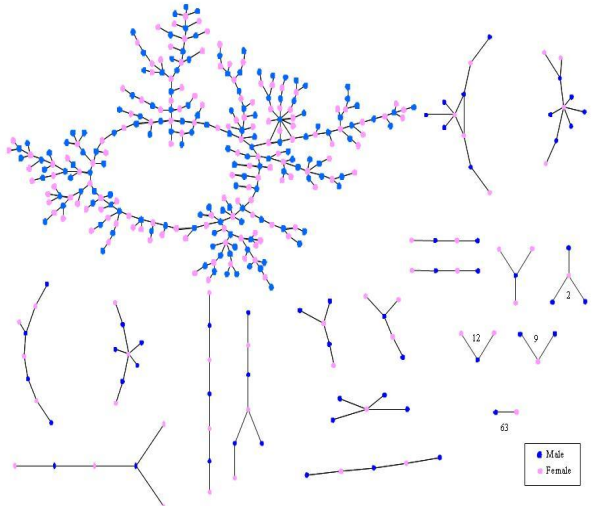
Similarities			Social Relations				Interactions	Flows
Location	**Membership**	**Attribute**	**Kinship**	**Other role**	**Affective**	**Cognitive**		
e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,
Same spatial and temporal space	Same clubs	Same gender	Mother of	Friend of	Likes	Knows	Sex with	Information
	Same events	Same attitude	Sibling of	Boss of	Hates	Knows about	Talked to	Beliefs
	etc.	etc.		Student of	etc.	Sees as happy	Advice to	Personnel
				Competitor of		etc.	Helped	Resources
							Harmed	etc.
							etc.	

(Borgatti, 2009)

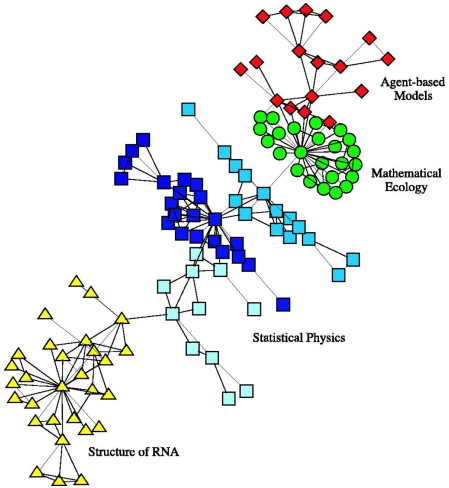
Course Objectives

In addition to universality, understand how **social context** is critical for constructing networks and interpreting results from network analysis: Know your nodes and edges!

Adolescent romantic relationships



Coauthorship relations among scientists



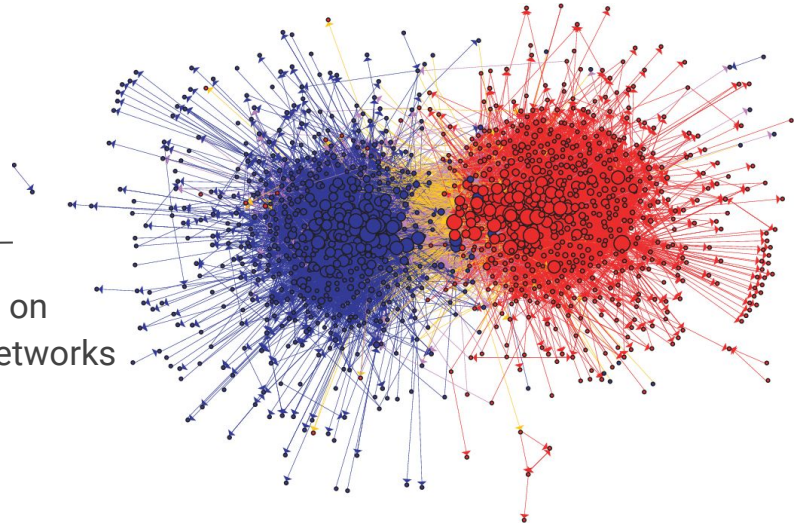
Course Objectives

Use network analysis to **describe** or **understand** puzzling social phenomena

Why does political polarization intensify?

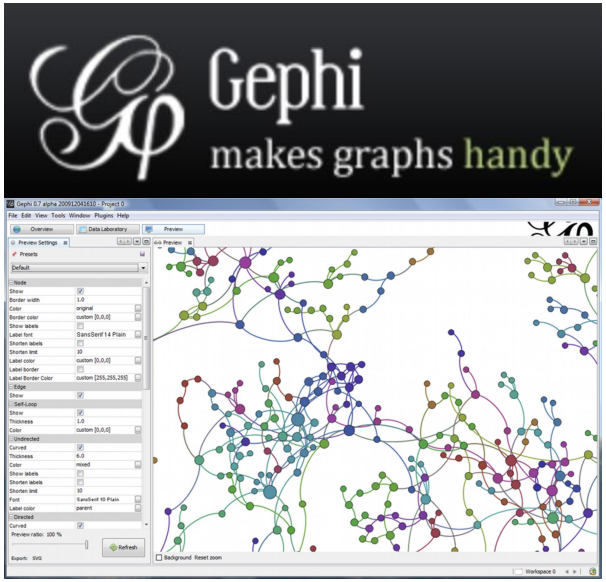
Opinion dynamics on communication networks

Network of political blogs



Course Objectives

Proficiently **analyze empirical network data** using network analysis and visualization software



Logistics

Overview of topics (tentative)

Graph theory basics

Edges vs. social ties

Degree correlation and homophily

Random networks: setting the baseline

Scale-free networks and preferential attachment

Power and centrality

Building blocks of social groups: Triads and clustering

Detecting communities

Small-world networks: clustering and short paths

Structural basis of advantage: diversity and closure

Ethical Issues in network research

Information diffusion

Social contagion

Structural equivalence

Network of affiliations (2-mode network, bipartite graphs)

Structure of online communities

Network mechanisms of political polarization

Guest lectures

Class format & grading

Lectures Tue / Thu 10%

Slides on the class website: http://bvasiles.github.io/networks

Assignments on Canvas https://canvas.cmu.edu/courses/37456 20%:

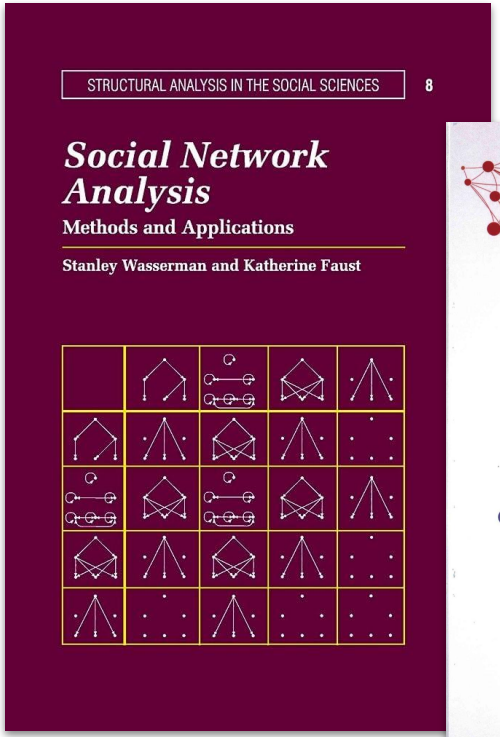
- Occasional readings
- Homework problems: coding, analysis, visualization of real-world datasets
- Quizzes on lecture material and readings

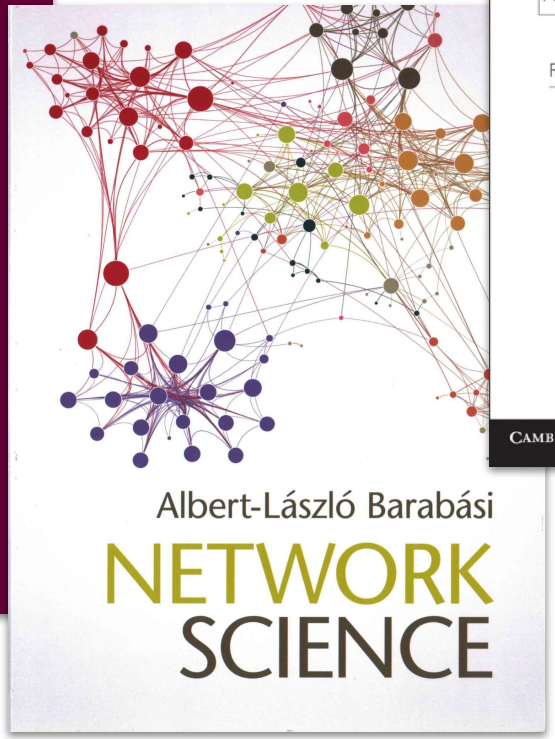
Midterm exam (open book) 30%

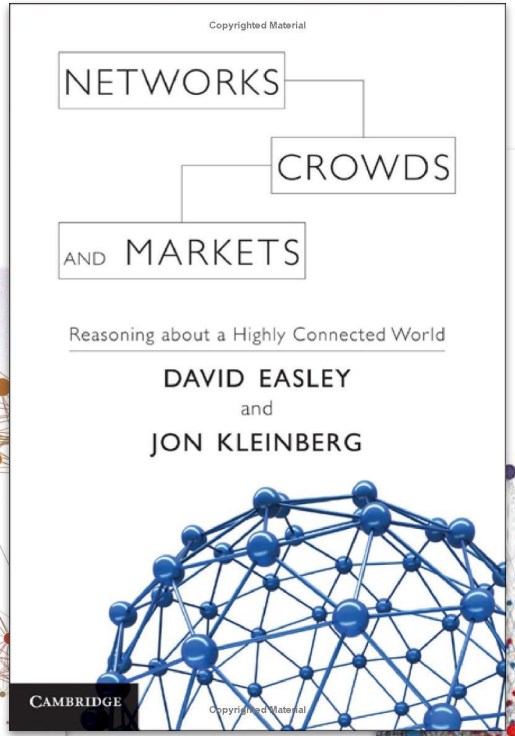
Final project: researchy paper 40%

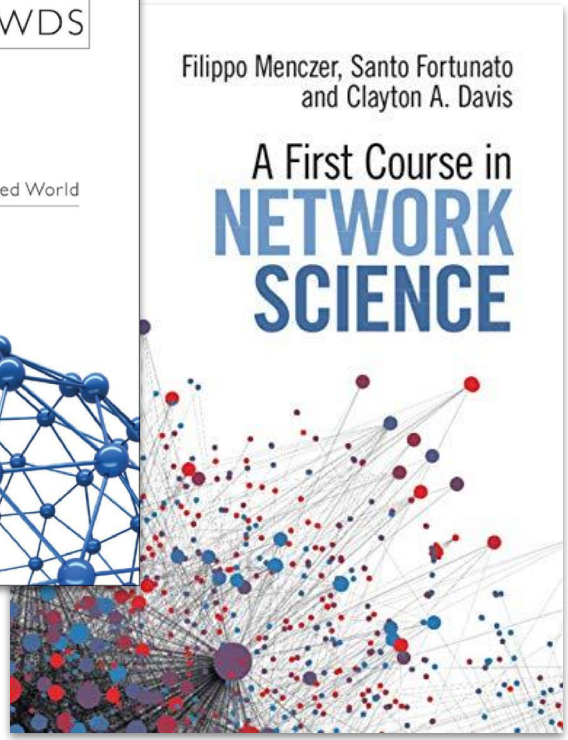
- Replication study (default)
- New study (possible, talk to us first)

Textbooks


STRUCTURAL ANALYSIS IN THE SOCIAL SCIENCES
8
Social Network Analysis
Methods and Applications
Stanley Wasserman and Katherine Faust


Albert-László Barabási
NETWORK SCIENCE

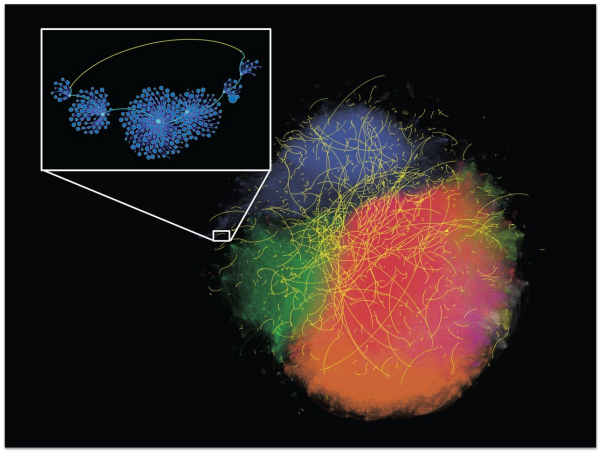

Copyrighted Material
NETWORKS
CROWDS
AND MARKETS
Reasoning about a Highly Connected World
DAVID EASLEY
and
JON KLEINBERG
CAMBRIDGE
Copyrighted Material


Filippo Menczer, Santo Fortunato and Clayton A. Davis
A First Course in
NETWORK SCIENCE

Instructor Patrick Park's background

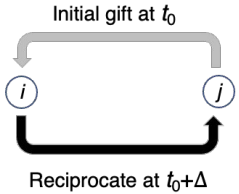
Assistant Prof. Software & Societal Systems Dept.

Research areas: Network science, computational social science, online communities



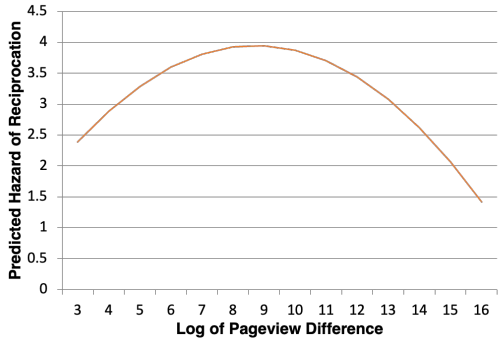
Why socially distant ties can be relationally strong

Gift exchange records from Cyworld



Dependent variable:
Time to reciprocation

Status competition through online gift exchange

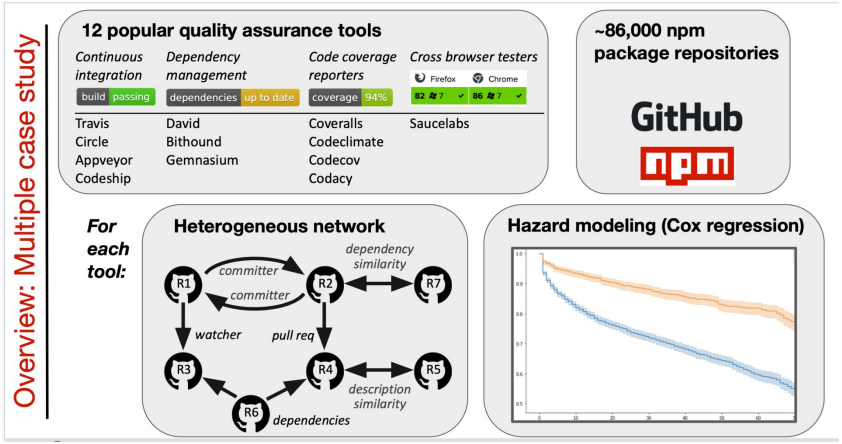


Co-instructor Bogdan Vasilescu's background

Associate Prof., Software & Societal Systems Dept.

Quantitative & mixed-methods empirical research in software engineering and online communities.

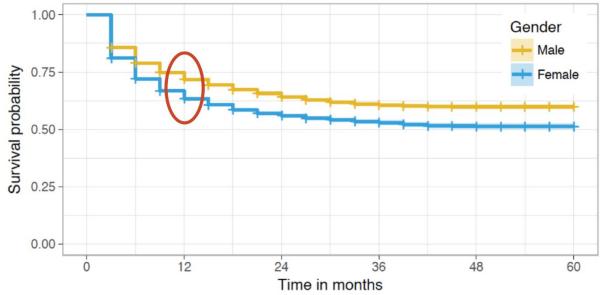
How do tools and practices spread through GitHub? (Lamba et al, 2020)



How does social capital explain OSS contributor participation? (Qiu et al, 2019)

On GitHub, women disengage earlier than men

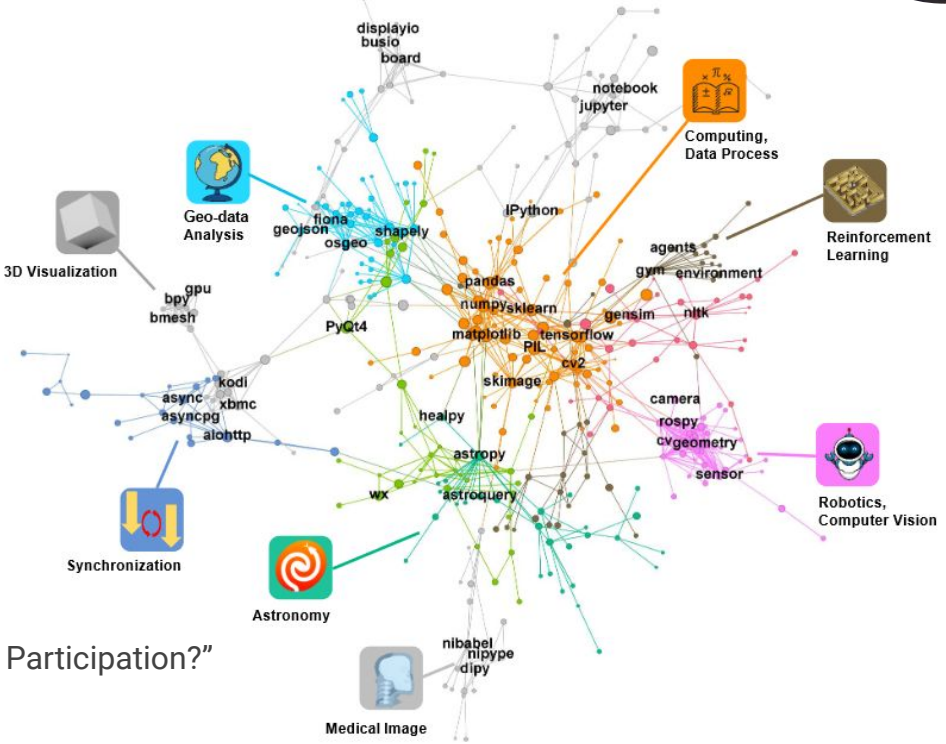
After one year ca. 70% of men are still active but only ca. 60% of women



TA Hongbo Fang's background

5th year Phd student in Societal Computing program at S3D.

Network analysis on the sustainability & novelty of open-source software projects



"Novelty Begets Long-Term Popularity, But Curbs Participation?"
(Fang et al, 2024, ICSE)

Who are you?
What drives your interest in networks?