Network Analysis:

The Hidden Structures behind the Webs We Weave

17-338 / 17-668

Introduction & Course Overview

Tuesday, August 26, 2025

Patrick Park

Carnegie Mellon University
School of Computer Science

S3D
Software and Societal Systems Department

Networks are old news

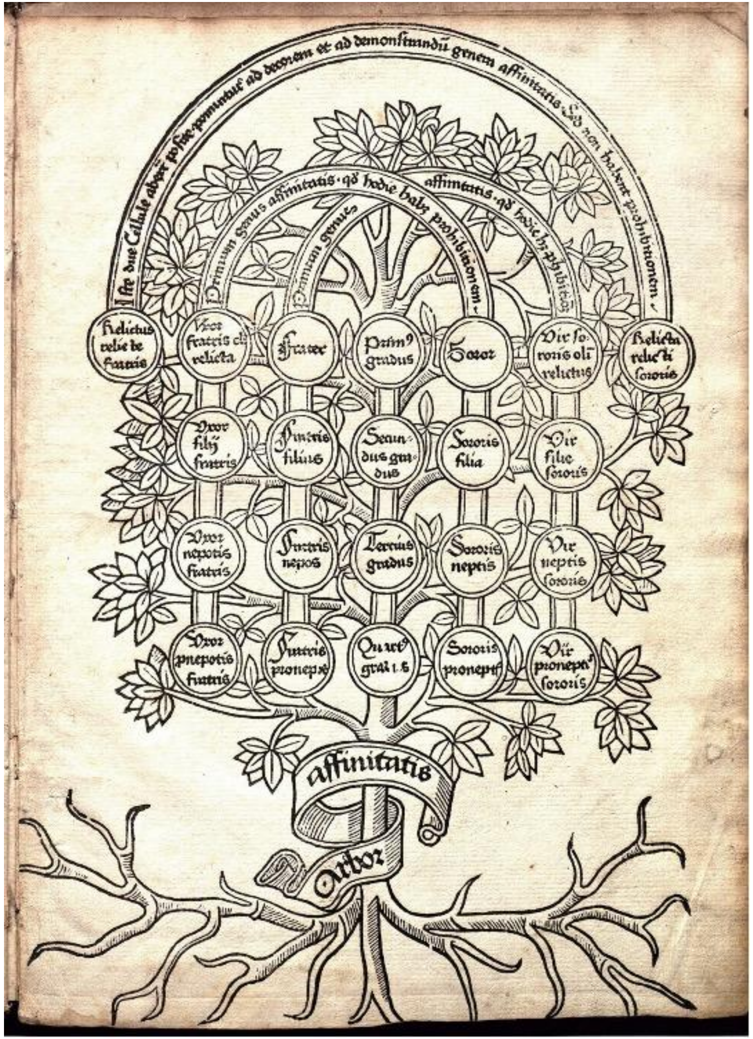
(networks: patterns of interconnections among a set of things)

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)

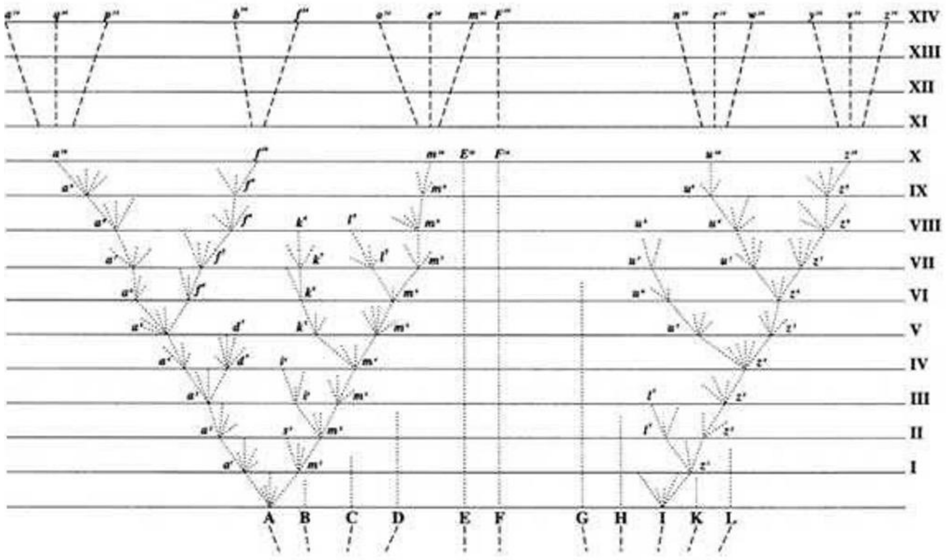
Maps laws and regulations on kinship and marriage decreed by the ecclesiastical authority of the Catholic Church.



Tree of Life by Charles Darwin in *On the Origin of Species*

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

(Lima 2013)



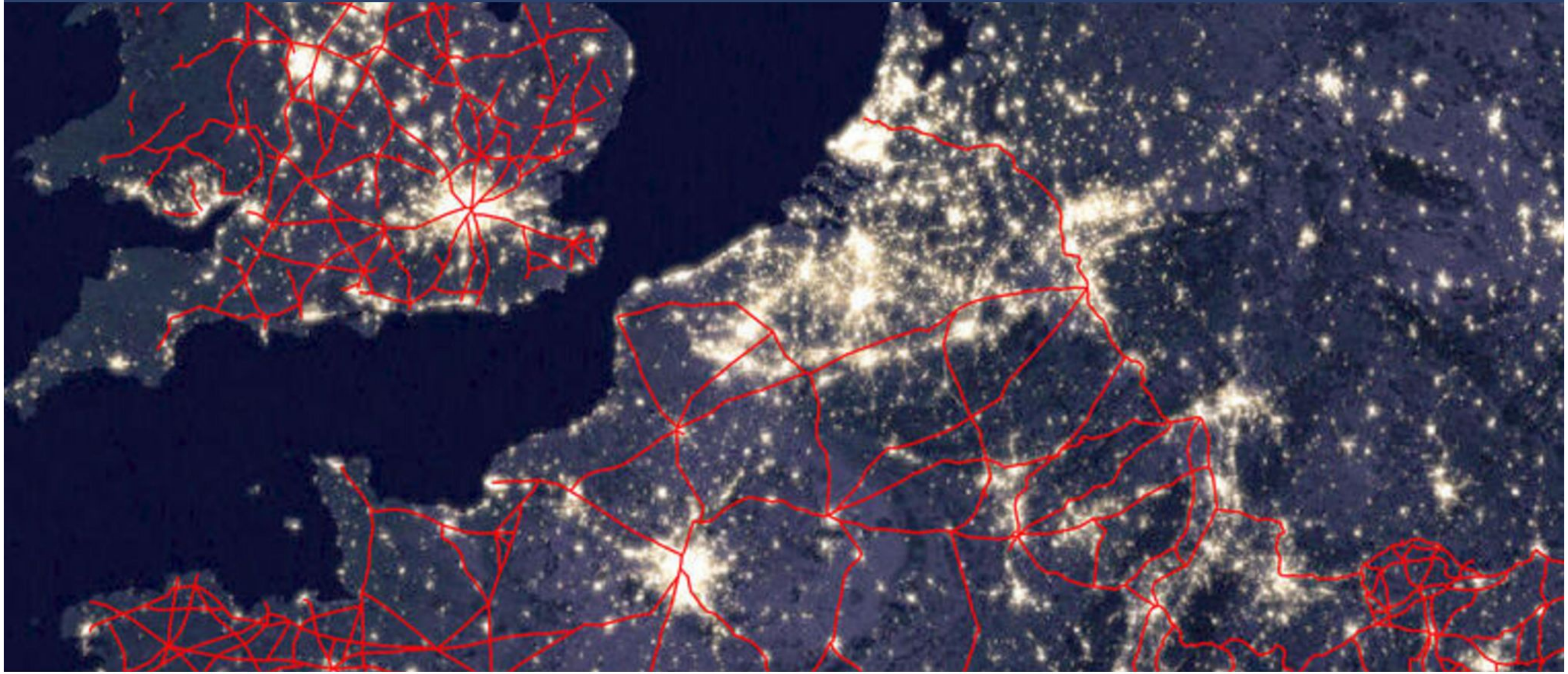
HARVARD UNIVERSITY

Mapping Past Societies

Formerly, the Digital Atlas of Roman and Medieval Civilizations

Search

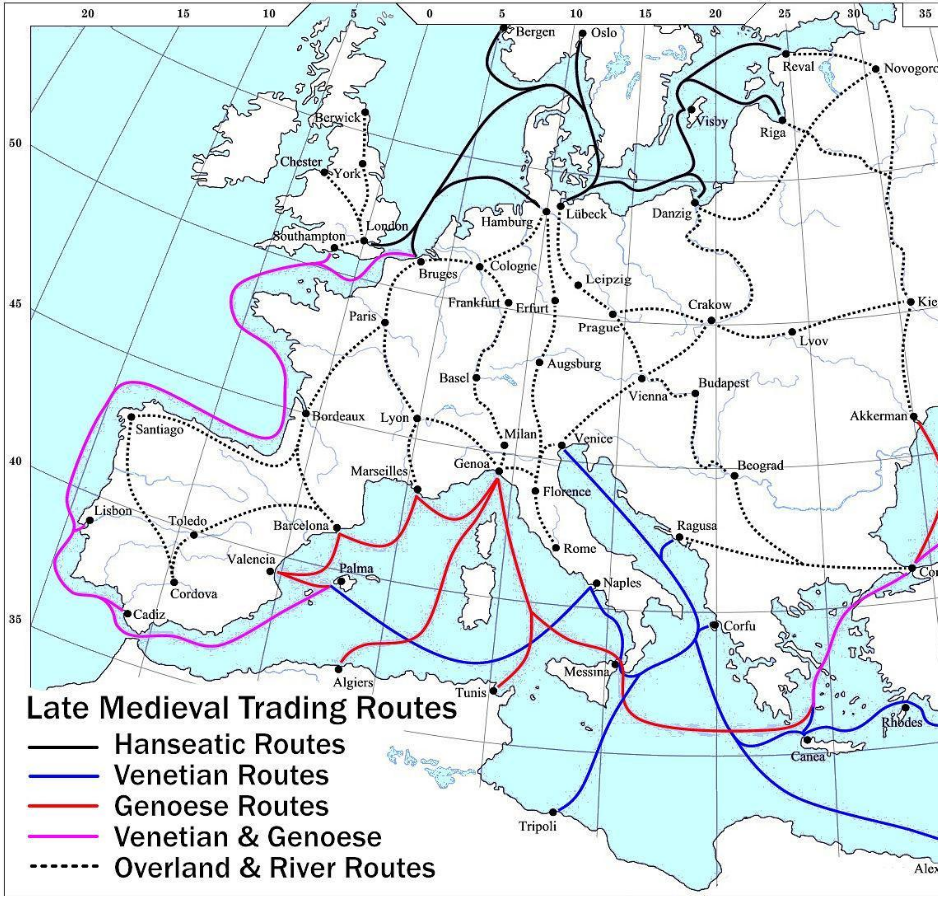
Home The Map Bibliography Team Map Sources Links Data Availability Contact



1 2 3

Medieval trade routes in Europe.

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



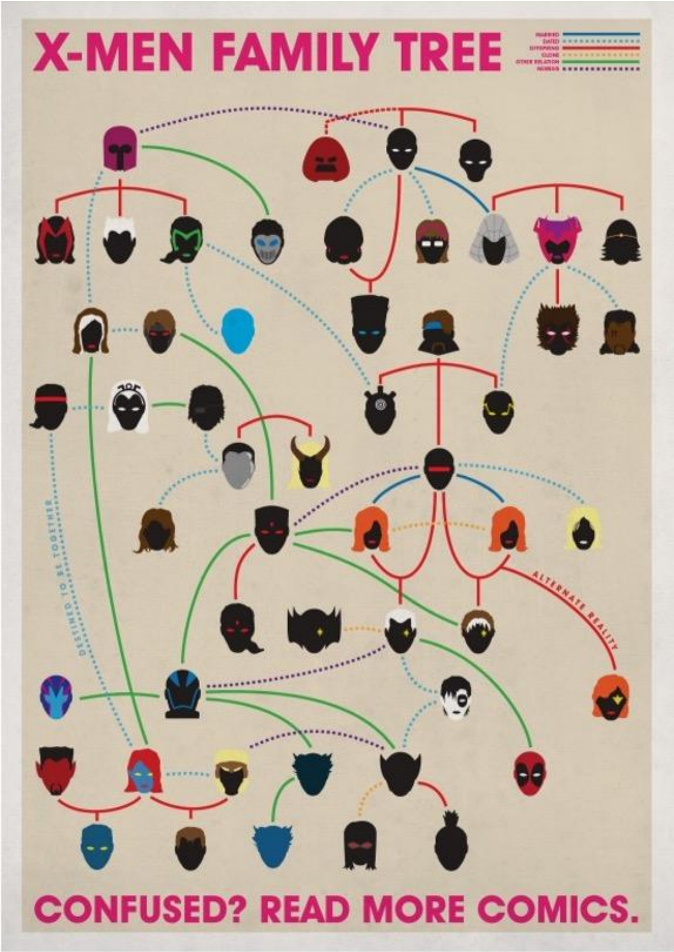
But networks are also recent news!

… sometimes for fun and art

X-Men Family Tree by Joe Stone

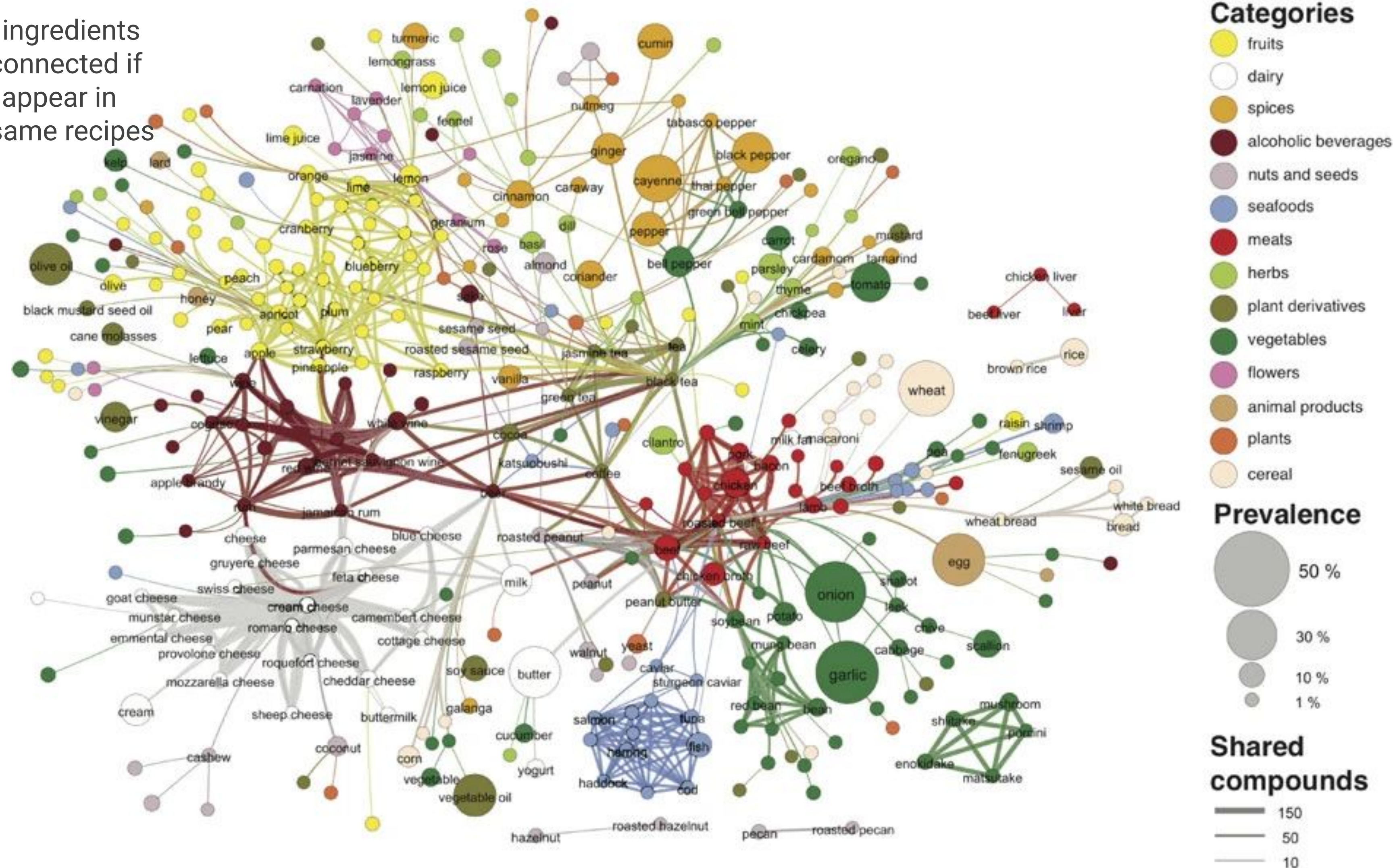
“An enticing and playful family tree charting the many convoluted relationships – romantic, genetic, or otherwise – of the X-Men characters from Marvel Comics.”

(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

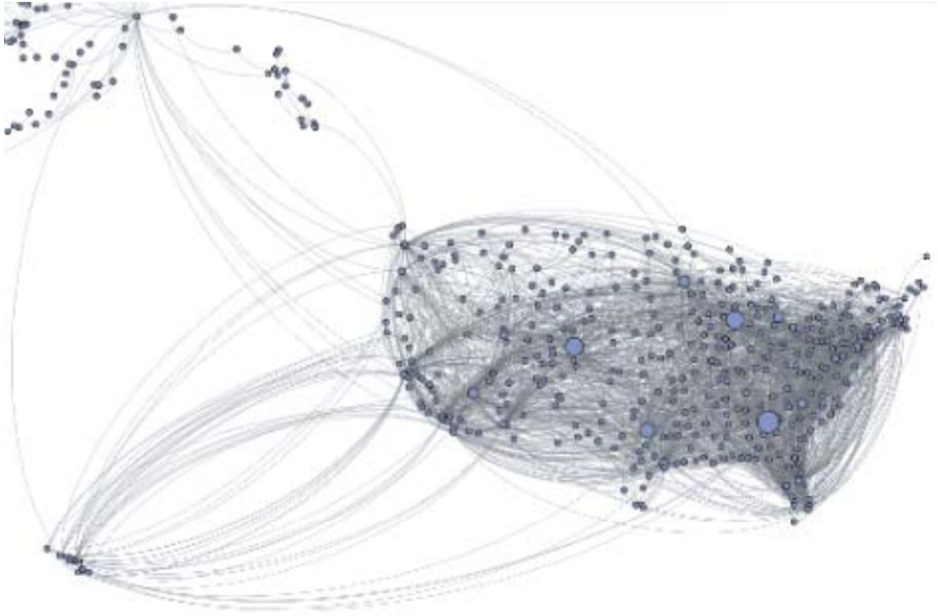
Two ingredients are connected if they appear in the same recipes



… 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)

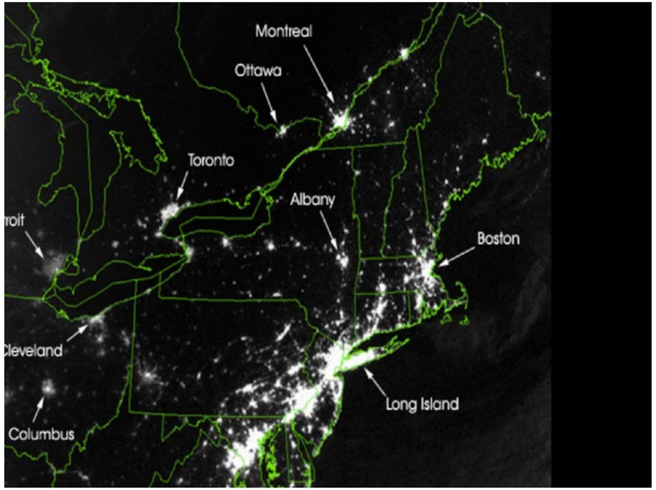


The 2003 Northeast blackout that left without power an estimated 45 million people in eight US states and another 10 million in Ontario.

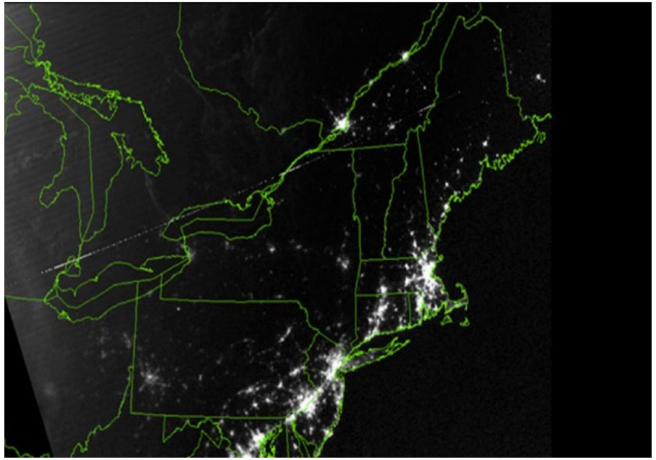
(a) Satellite image 20 hours before
(b) … and 5 hours after

(Barabasi, 2016)

(a)

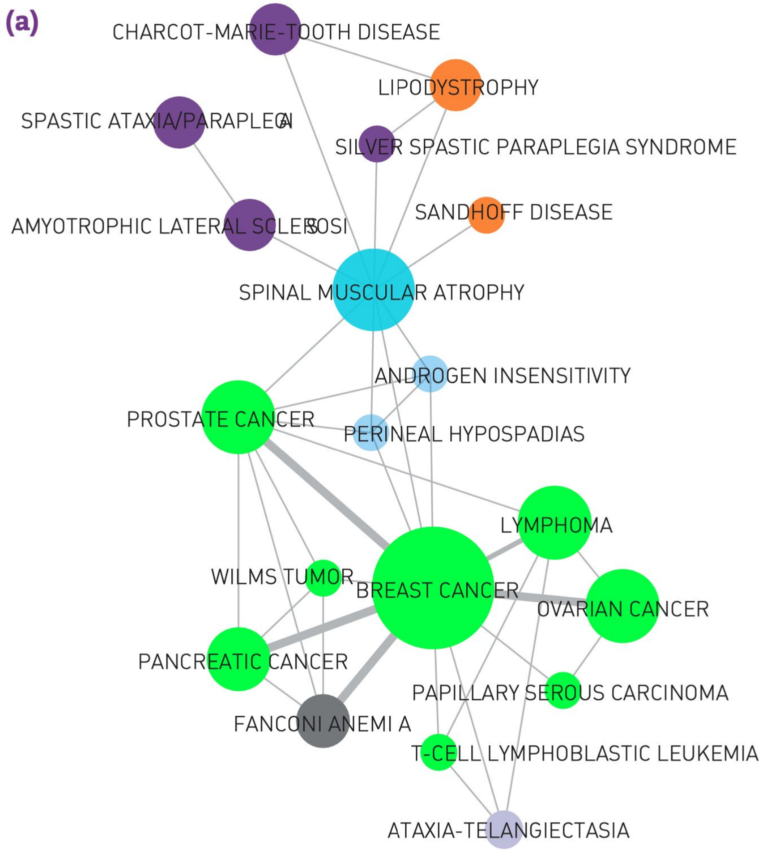


(b)



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

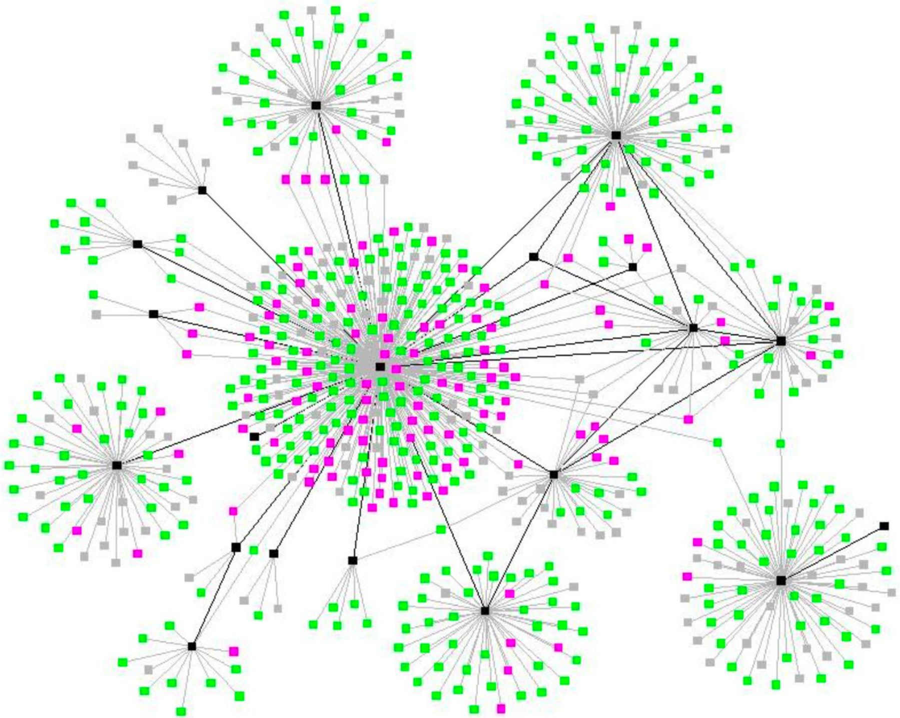
(Barabasi, 2016)



HUMAN DISEASE NETWORK

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

(Easley & Kleinberg, 2010)

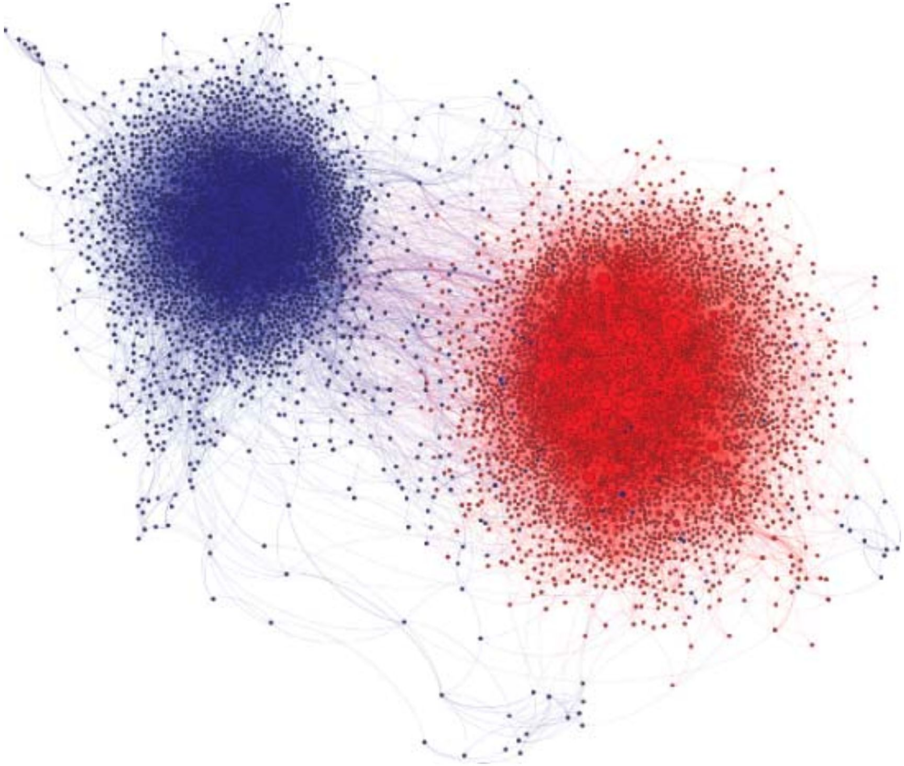


“Wood-wide web:” Trees transmit danger messages and nutrition through fungi connections.

(Steidinger, Crowther, et al, 2019)

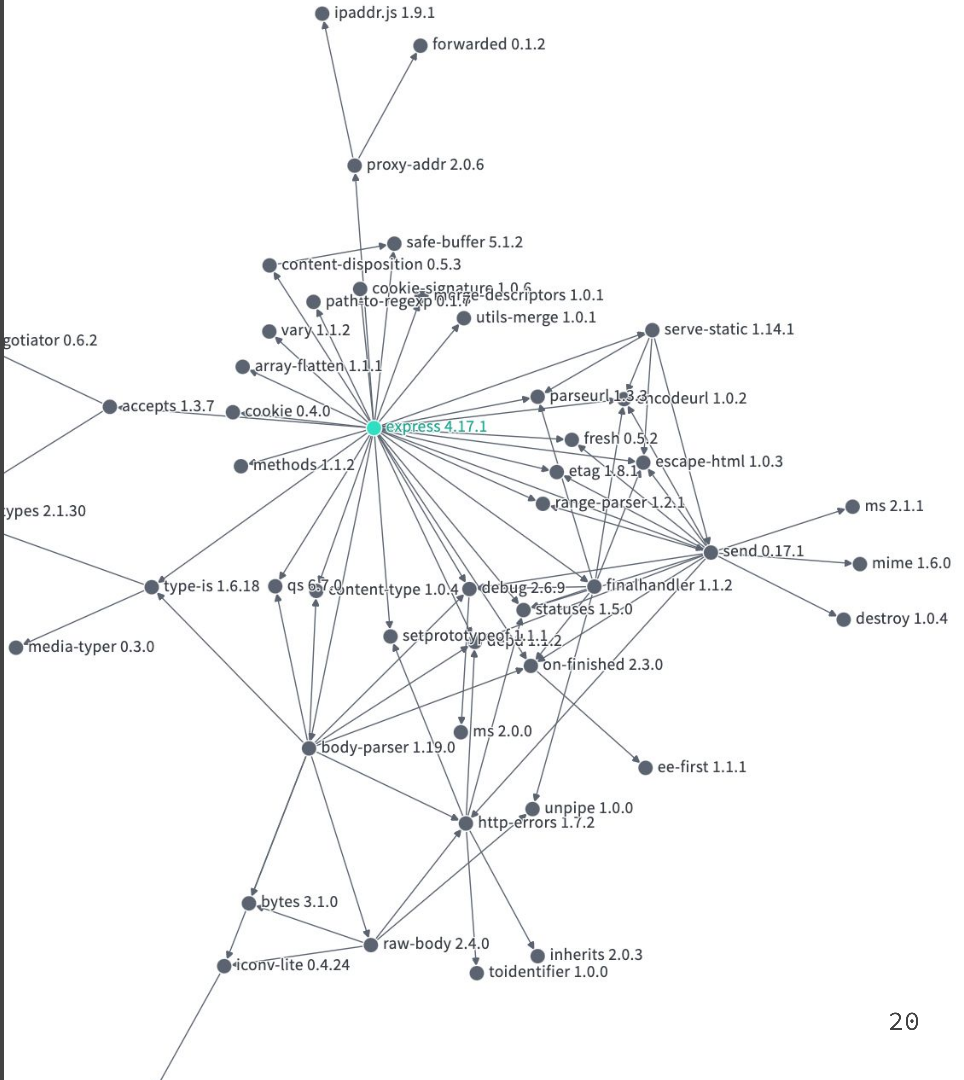
A retweet network on Twitter, among people sharing posts about US politics. Conservative users (red) mostly retweet messages from other conservatives, and vice versa.

(Menczer, Fortunato, & Davis, 2020)



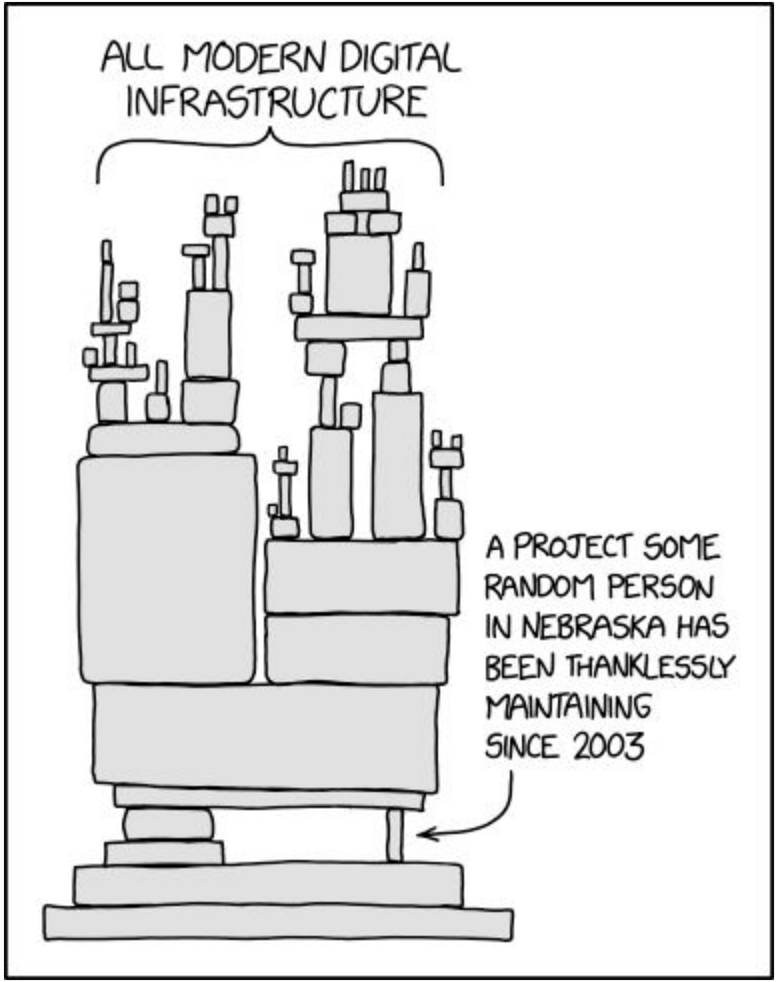
Dependency graph for the open source framework express 4.17.1.

(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

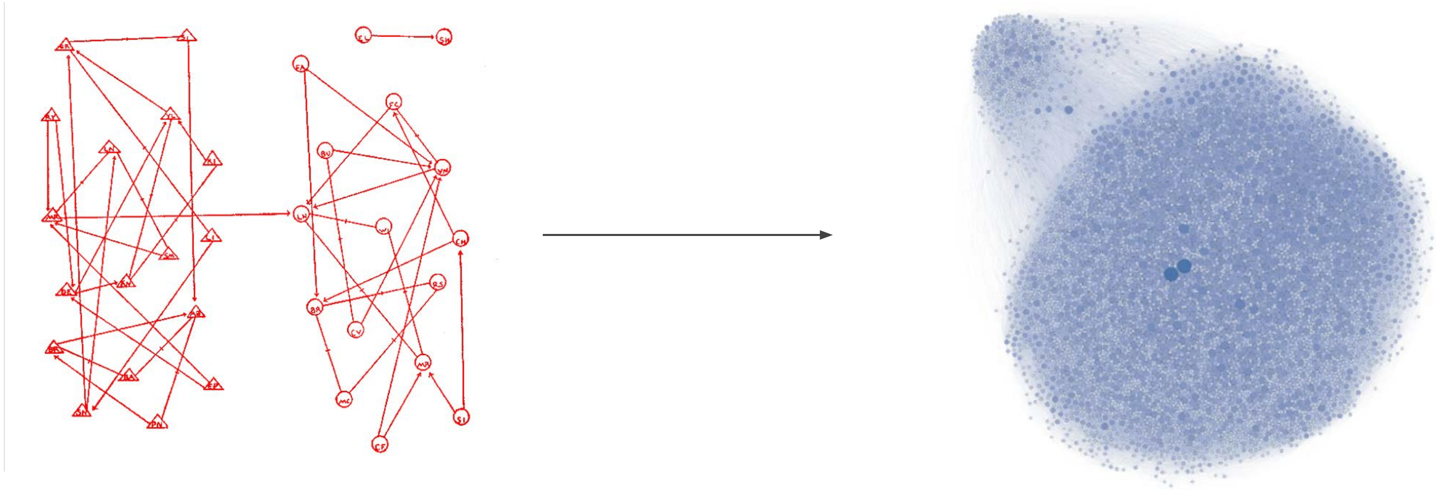
Physics

Social Science

Statistics

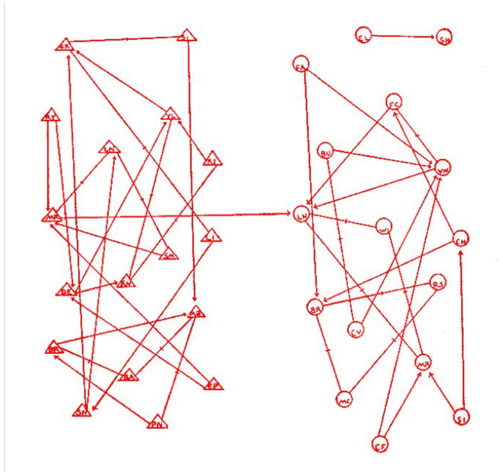
Modern network science is propelled by computational power

Network maps

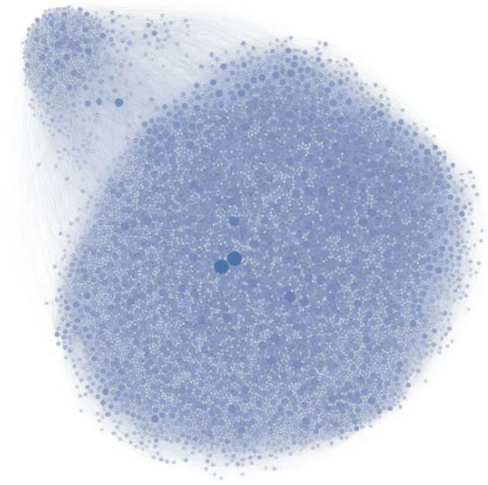


Modern network science is propelled by computational power

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

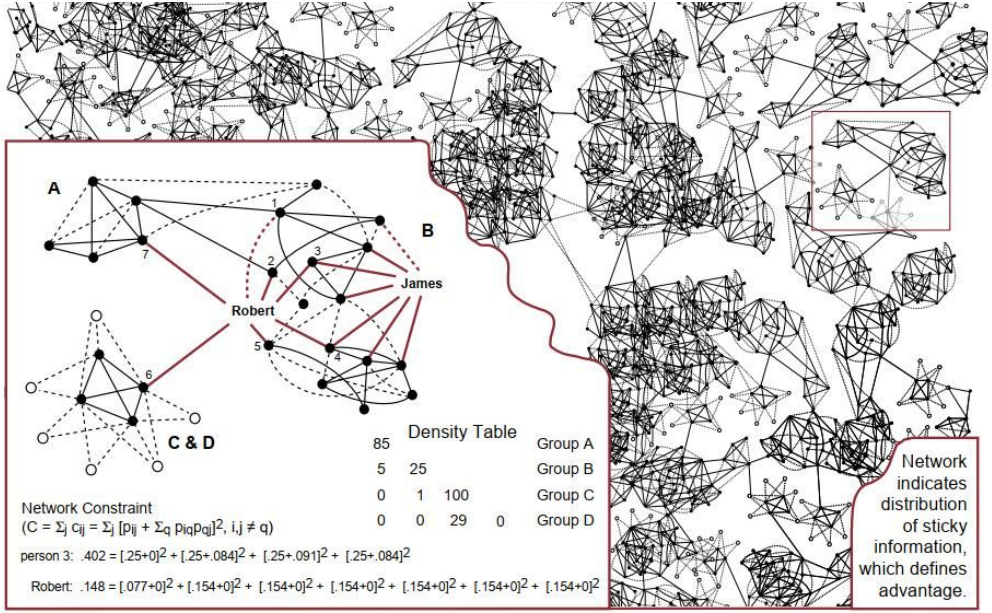


Figure 1. Network Bridge and Cluster Structure

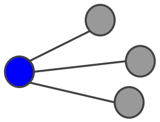
Adapted from Burt (2005:14).

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

Main abstraction: Networks as Graphs

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 

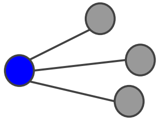
Density: Number of edges (L), relative to the maximum possible edges

- Maximum possible edges: ?

Main abstraction: Networks as Graphs

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

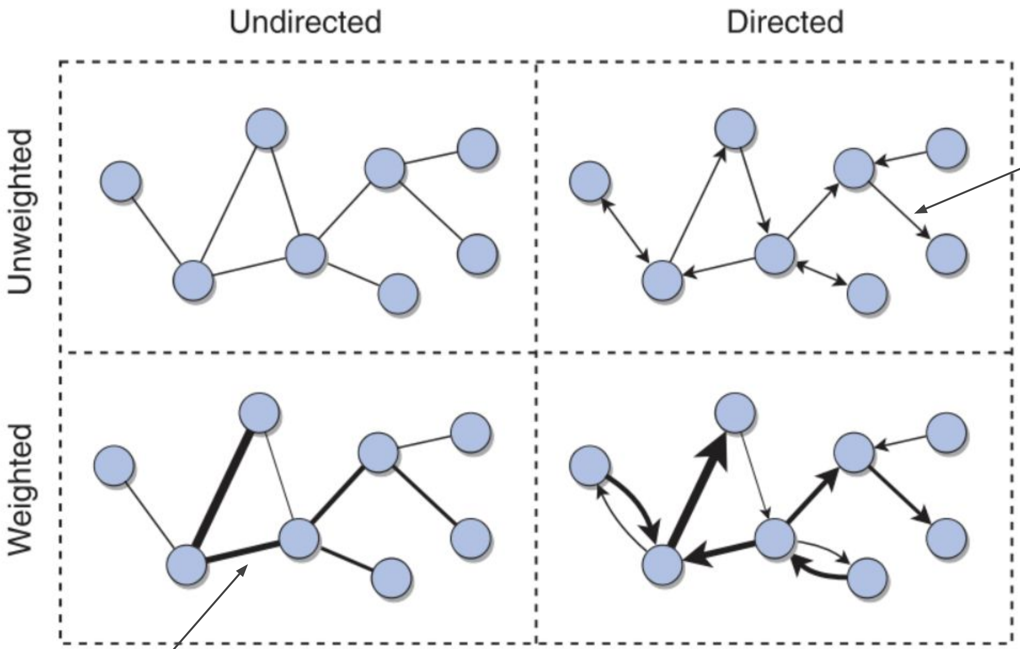
Edge (or link, “L”): Connection between two nodes 

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

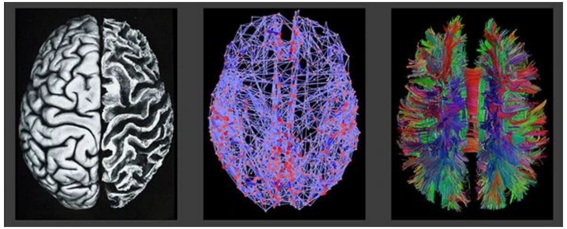
Natural sciences (e.g., physics, CS, mathematics) typically focused on

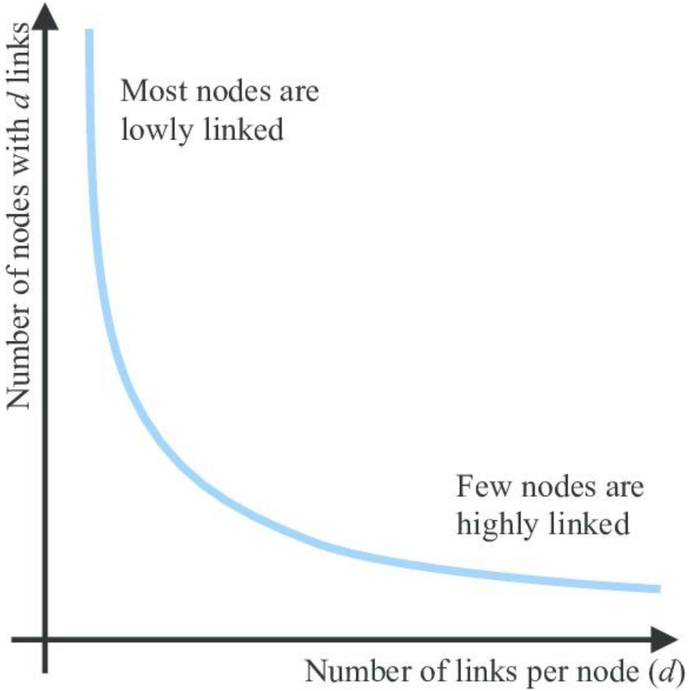
- Universality: search for common organizing principles across domains
- Level of analysis: entire network
- Study the networks as complex systems

*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: 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 [individuals] *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

Different kinds of dyadic links, both analytical and theoretical.

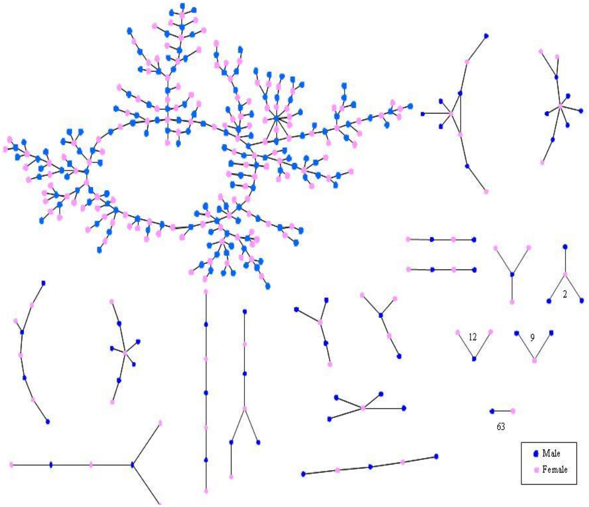
Similarities			Social Relations				Interactions	Flows
Location	**Membership**	**Attribute**	**Kinship**	**Other role**	**Affective**	**Cognitive**	e.g.,	e.g.,
e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	Sex with	Information
Same spatial and temporal space	Same clubs	Same gender	Mother of	Friend of	Likes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

(Borgatti, 2009)

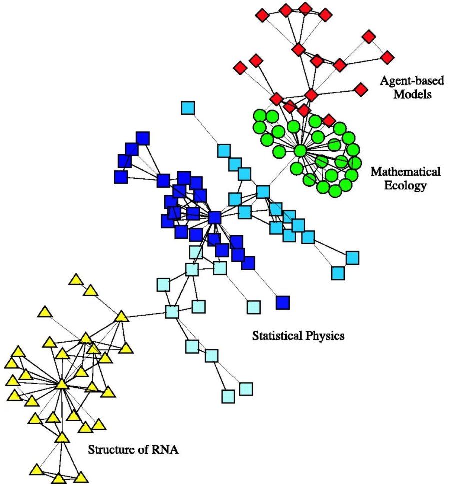
Course Objectives

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



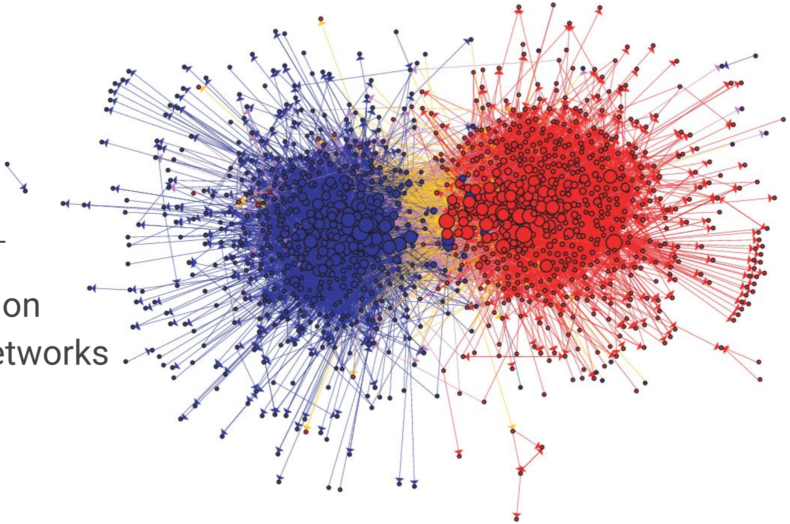
Course Objectives

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

Why does political polarization intensify?

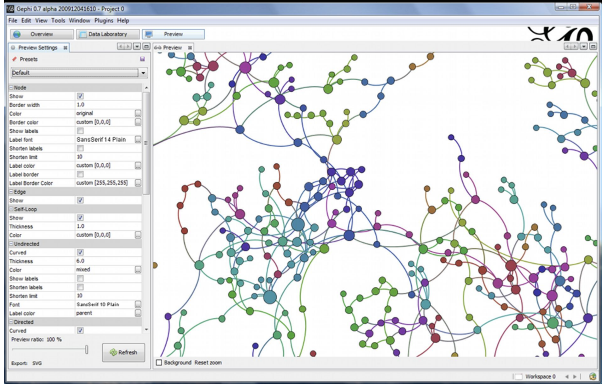
Opinion dynamics on communication networks

Network of political blogs

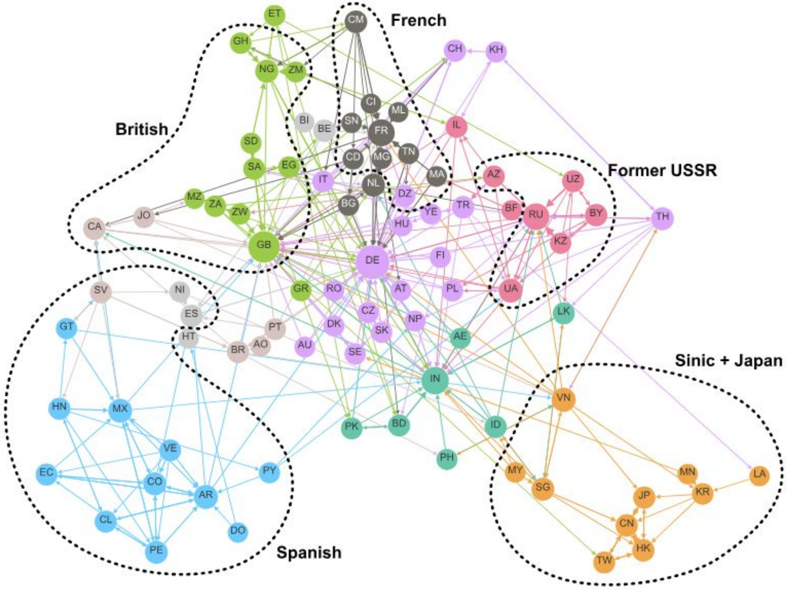


Course Objectives

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



Have fun learning together and from one another!



Class collaboration project in 2023

[CIKM'25](https://doi.org/10.1145/3746252.3761237)

The Structure of Cross-National Collaboration in Open-Source Software Development

Henry Xu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA
henryxu@mit.edu

Katy Yu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
tieliny@andrew.cmu.edu

Hao He
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
haohe@andrew.cmu.edu

Hongbo Fang
University of Chicago
Chicago, Illinois, USA
fanghongdoublebo@gmail.com

Bogdan Vasilescu
Carnegie Mellon University
Pittsburgh, USA
vasilescu@cmu.edu

Patrick S. Park
Carnegie Mellon University
Pittsburgh, USA
patpark@cmu.edu

Abstract

Open-source software (OSS) development platforms, such as GitHub, expand the potential for cross-national collaboration among developers by lowering the geographic, temporal, and coordination barriers that limited software innovation in the past. However, as shown in previous research, these technological affordances that lower barriers for cross-national collaboration do not uniformly benefit all countries. Using the GitHub Innovation Graph dataset, which aggregates the complete cross-country collaborations among the entire population of GitHub developers, we show the overwhelming dominance of the U.S. in global OSS collaboration and, in stark contrast, the paucity of collaborations among the majority of non-Western country pairs in the data. However, by removing the U.S. and its ties from the collaboration network, we find quantitative evidence of deep-seated religious and cultural affinities, shared colonial histories, and geopolitical factors structuring the collaborations between non-U.S. country pairs. This study highlights the opportunities to develop decentralizing strategies to facilitate new collaborations between developers in non-U.S. countries, thereby fostering the development of novel, innovative solutions. More generally, this study also underscores the importance of contextualizing user behavior and knowledge management in information systems with long-term, macro-social conditions in which these systems are inextricably embedded. [1]

Keywords

open-source software, international collaboration, homophily, network analysis, GitHub, exponential random graph models, block models, world systems theory, colonialism, geopolitics

ACM Reference Format:
Henry Xu, Katy Yu, Hao He, Hongbo Fang, Bogdan Vasilescu, and Patrick S. Park. 2025. The Structure of Cross-National Collaboration in Open-Source Software Development. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25), November 10–14, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3746252.3761237

[1]The replication data and code are available at: https://github.com/hehao98/github-innovation-graph

CIKM '25, Seoul, Republic of Korea

ACM ISBN 979-8-4007-2040-6/2025/11
https://doi.org/10.1145/3746252.3761237

1 Introduction

How do countries collaborate in open-source software development? Throughout history, dense economic trade relations in commodities, goods, services, and labor tended to promote international peace whereas sparse connections and fragmentation heightened the probability of war [13]. These international connections among countries have been subject to a myriad of social, economic, and political forces with shifting influences over time. For example, Western capitalist versus Eastern communist countries during the Cold War era formed connections along ideological fault lines while the post-Cold War divisions have been structured in part by deep-seated culture and religion [12]. Their significance has been shown to persist in online social interactions among millions of email and social media users, despite the friction-less connectivity that these communication technologies enabled [24].

While studies show the persistence of such ideological, cultural, and geopolitical forces on cross-national online communication, it is an open question whether cross-national collaborations with a clearer instrumental focus, such as those observed in open-source software development, also form under the influence of such macro-social forces. Presumably, these bottom-up collaborations between millions of software developers across countries should appear oblivious to these social forces insofar as the online collaboration platform helps overcome spatio-temporal barriers of collaboration in the creation of technological and economic value. Indeed, previous studies of open-source software development have shown that cross-country collaborations largely exhibit such instrumental rationality, reflecting the economic dominance of the U.S. and, to a lesser extent, of a handful of technologically advanced European countries. Although this heavy reliance on the U.S. is taken for granted, it is puzzling from the vantage point of the potential for frictionless and untethered collaborations that OSS platforms afford.

Based on a recent cross-country collaboration dataset released by GitHub, we construct a network of OSS collaborations between countries and study the patterns hidden under the global dominance of the United States. We reveal the structural markers of

Logistics

Overview of topics

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: https://patpark.org/networks_17338

Assignments on Canvas https://canvas.cmu.edu/courses/50123 (20%)

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

Midterm exam (open book) and project proposal (20%)

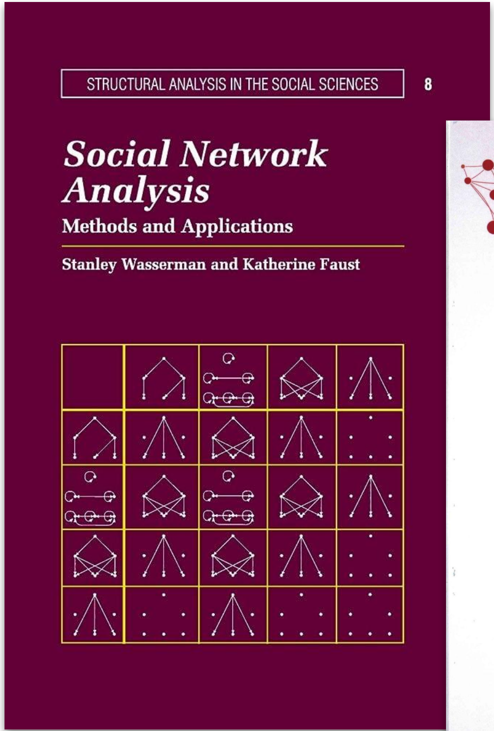
Project proposal (10%)

- Due soon after fall break

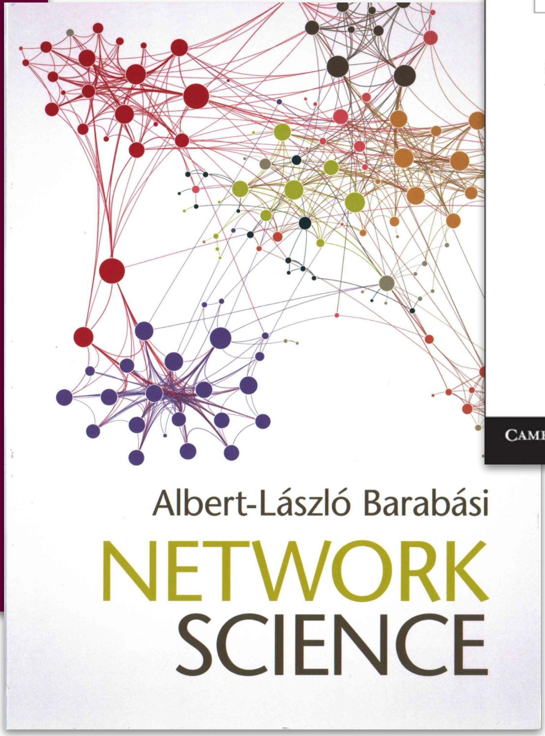
Final project (30%)

- Research report (Replication study or new study)
- Open source network software code contribution

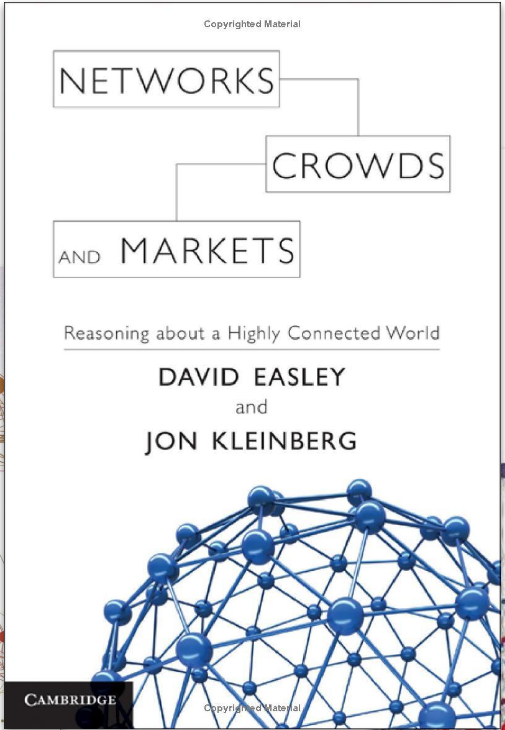
Textbooks



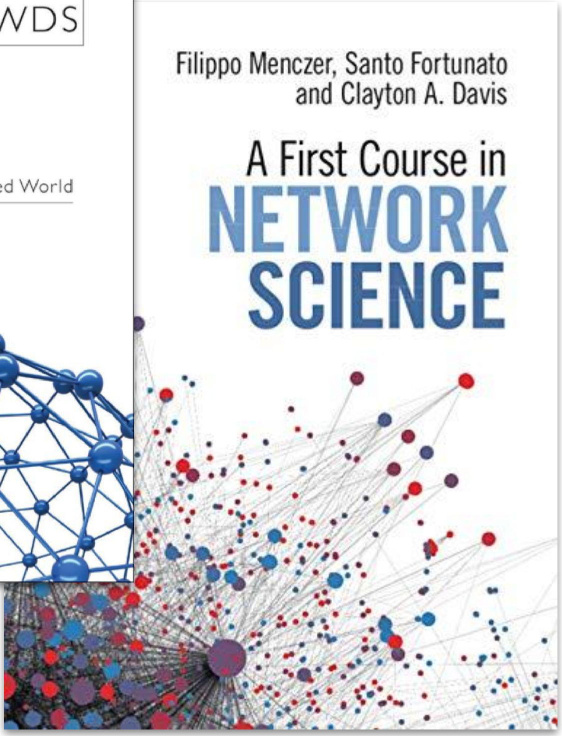
STRUCTURAL ANALYSIS IN THE SOCIAL SCIENCES
8
Social Network Analysis
Methods and Applications
Stanley Wasserman and Katherine Faust



Albert-László Barabási
NETWORK
SCIENCE



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

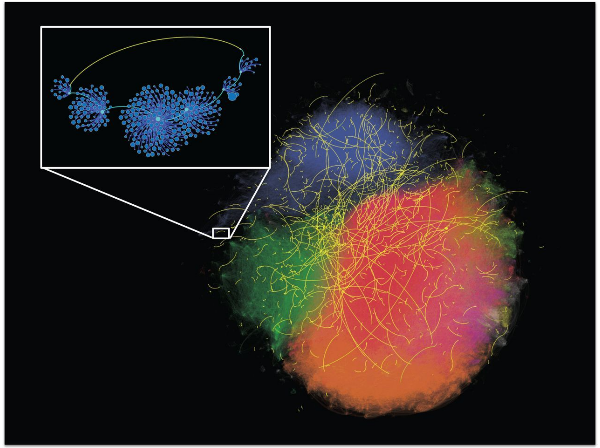


Filippo Menczer, Santo Fortunato
and Clayton A. Davis
A First Course in
NETWORK
SCIENCE

Instructor Patrick Park's background

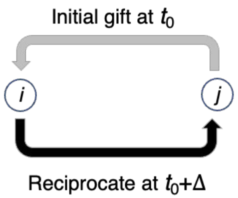
Assistant Prof. Software & Societal Systems Dept.

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



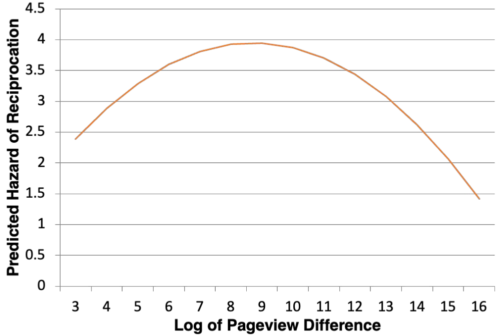
Why socially distant ties can be relationally strong

Gift exchange records from Cyworld



Dependent variable:
Time to reciprocation

Status competition through online gift exchange



Who are you?
What drives your interest in networks?