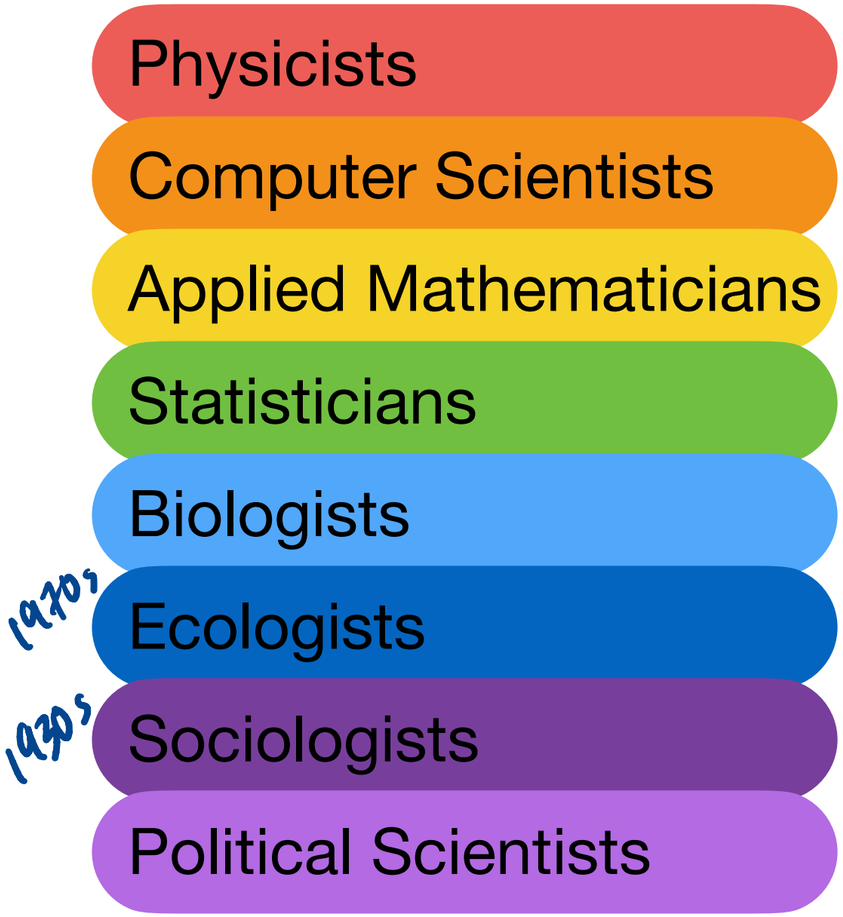
# Network Analysis & Modeling

lecture 0: what are networks and how do we talk about them?

- [x] slack (9) vs. Piazza (6)
- [x] Canvas
- [x] HW #1 By end of week "free throw"
- [x] no more waiting room ? ?

CU Computer Science
& BioFrontiers Institute
@danlarremore
daniel.larremore@colorado.edu

# who are network scientists?




recent

recent

recent older

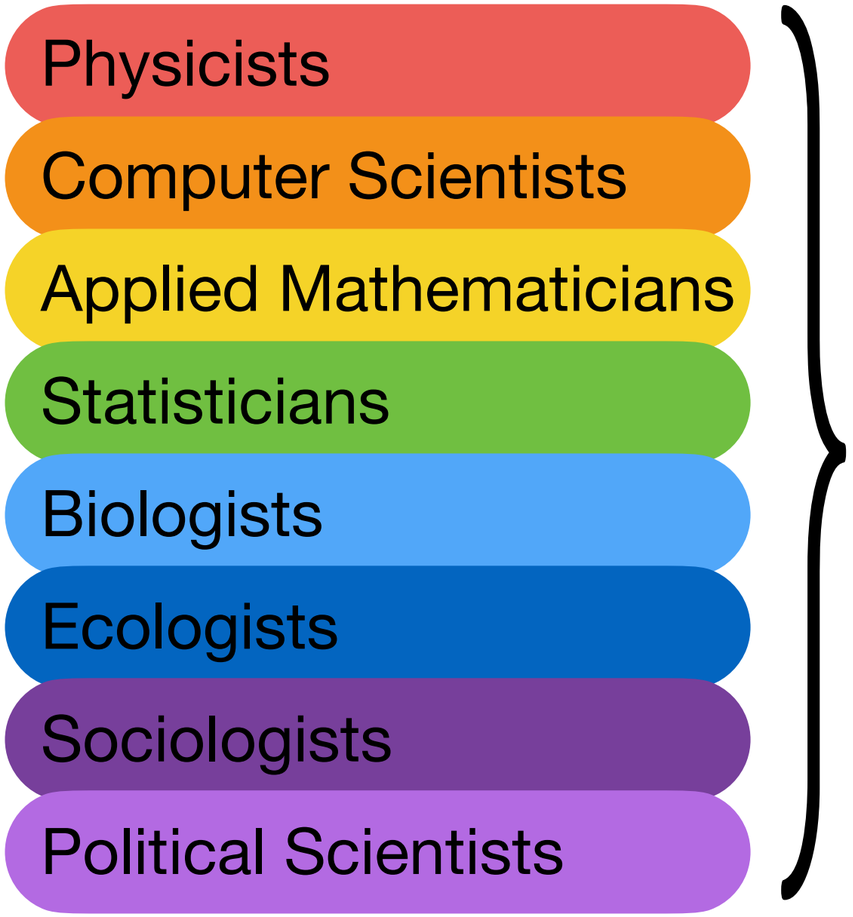
older

it's a big community!

- different *traditions*
- different *tools*
- different *questions*

increasingly, not ONE community,
but MANY, only loosely interacting communities

# who are network scientists?




phase transitions, universality

data / algorithm oriented, predictions

dynamical systems, diff. eq.

inference, consistency, covariates

experiments, causality, molecules

observation, experiments, species

individuals, differences, causality

rationality, influence, conflict

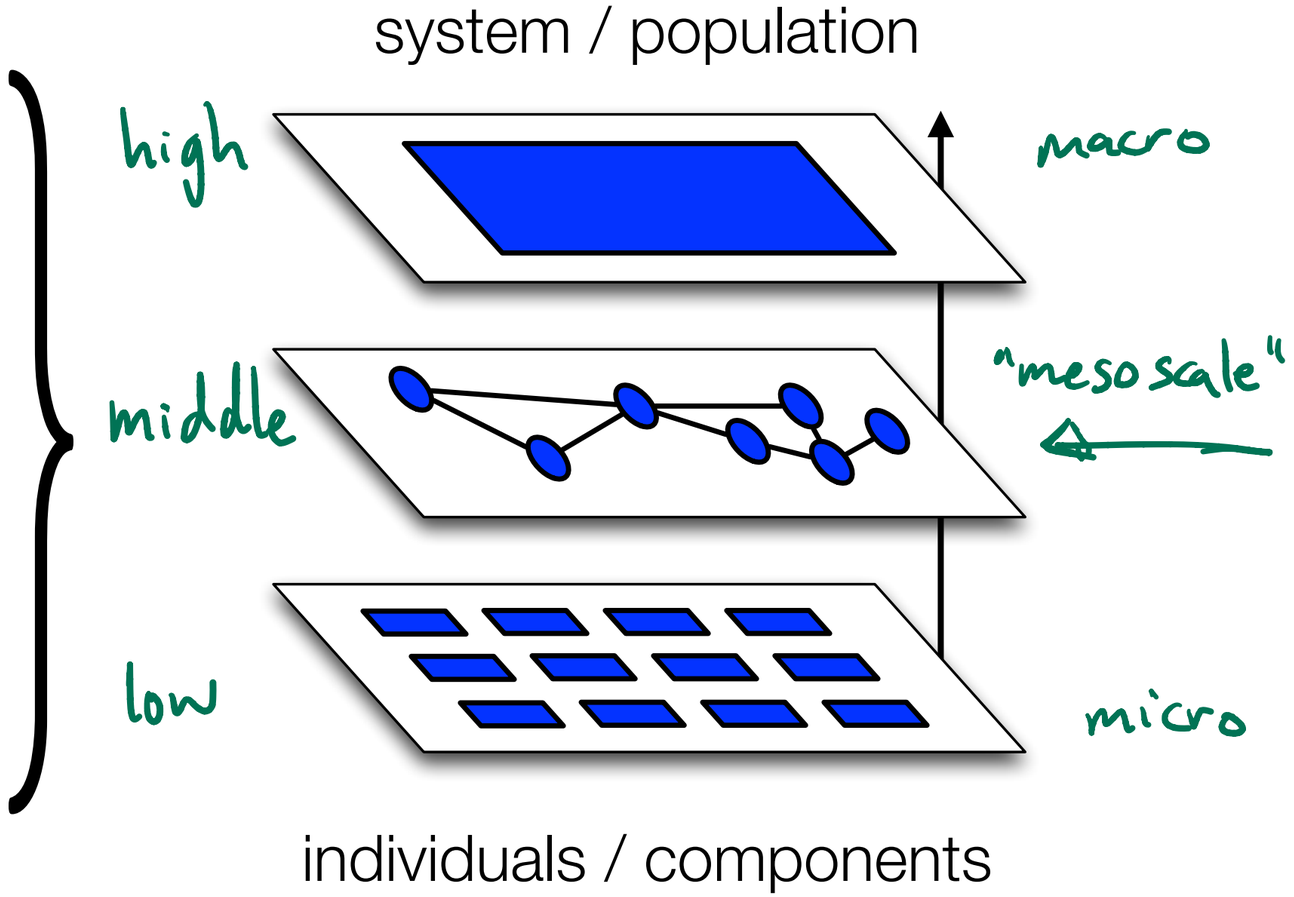
# what are networks?

networks aren't real!

An approach.
A mathematical representation
provide structure to complexity.

Structure that exists *above*
individuals / components

Or: structure that exists *below*
system / population



# csci 5352: learning goals

1. Develop a **network intuition** for reasoning about how structural patterns are related, and how they influence dynamics in / on networks.

2. Master **basic terminology** and **concepts**.

3. Master **practical tools** for analyzing / modeling structure of network data.

4. Build **familiarity** with key advanced techniques that will allow you to **explore and test hypotheses** about networks.

5. See how **the rubber meets the road for math**: calc, diff EQs, linear algebra, Lagrange multipliers, theoretical CS, Bayesian inference.

# course format

This course will meet via Zoom.

I would prefer to meet in the classroom! But that's not the world we live in.
[We will study infectious disease dynamics in this class!]

I'll ask you for your feedback at multiple points in this semester. I **need** to know what is working and what isn't, for *you*.

I would love to know about effective online teaching tricks other profs have used.

Homework will come in 12 chunks this year instead of 6 megachunks. There will also be new "free throw" type mini-problems.

# course schedule (roughly)

**Course schedule (roughly) :**

1. network basics
2. centrality measures
3. random graphs ($G_{nm}$, $G_{np}$)
4. random graphs (configuration model)
5. large-scale structure (assortativity and modularity)
6. large-scale structure (stochastic block models)
7. spreading processes (social, biological, SI, SIR, SIS)
8. ranking, pairwise comparison, choice modeling
9. data wrangling + data sampling (artifacts)
10. data wrangling + stats and tests
11. spatial networks
12. growing networks
13. dynamic networks
14. student project presentations

- building intuition
- basic concepts, tools
- practical tools
- advanced tools

# course webpage(s)

LarremoreLab.github.io/5352

Slides,
Readings,
Links,
Notebooks

## Network Analysis & Modeling

## CSCI 5352 - Network Analysis & Modeling

**Fall 2020**

- Time: Monday, Wednesday, Friday, 1:50pm - 2:40pm
- Place: **all streamed & recorded**; ECCS 1B28
- Lecturer: Dan Larremore
- Office: –Online–
- Office hours: –TBD–
- Email: daniel.larremore
- Teaching Assistant: None
- Syllabus: PDF

**On this page**



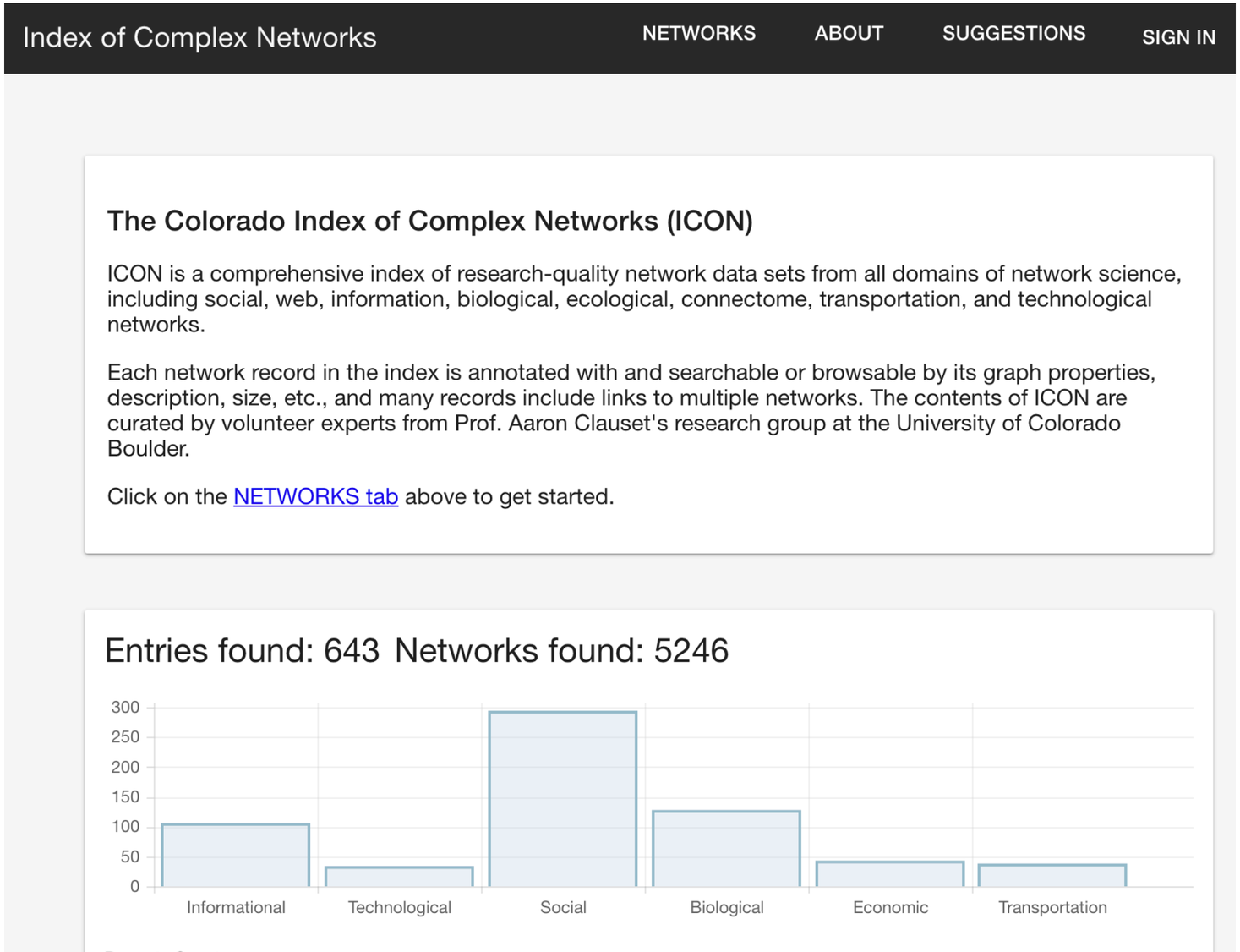
### Description

Network science is a thriving and increasingly important cross-disciplinary domain that focuses on the representation, analysis, and modeling of complex social, biological and technological systems as networks or graphs. Modern data sets often include some kind of network. Nodes can have locations, directions, memory, demographic characteristics, content, and preferences. Edges can have lengths, directions, capacities, costs, durations, and types. And, these variables and the network structure itself can vary, with edges and nodes appearing, disappearing and changing their characteristics over time. Capturing, modeling and understanding networks and rich data requires understanding both the mathematics of networks and the computational tools for identifying and explaining the patterns they contain.

This graduate-level course will examine modern techniques for analyzing and modeling the structure and dynamics of complex networks. The focus will be on statistical

# network data for assignments

icon.colorado.edu



# lessons learned

**what's difficult:**

1. students need to know many different things:

| | |
|---|---|
| some probability | Erdos-Renyi, configuration, calculations |
| some mathematics | physics-style calculations, phase transitions |
| some statistics | basic data analysis, correlations, distributions |
| some machine learning | prediction, likelihoods, features, estimation algorithms |
| some programming | data wrangling, coding up measures and algorithms |

2. can't teach all of these things to all types of students!

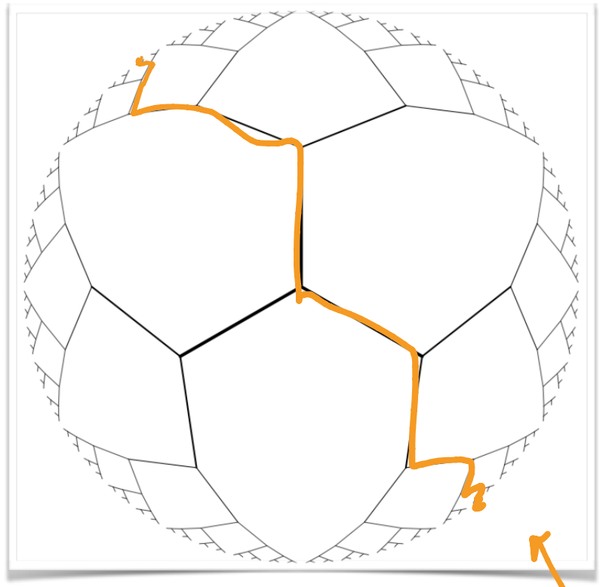
- vast amounts of advanced material in each of these directions
- students have little experience / intuition of what makes *good* science ← muscle to build

projects

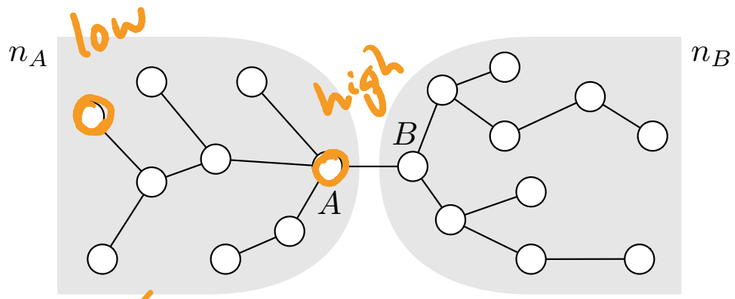
# lessons learned

**what works well**

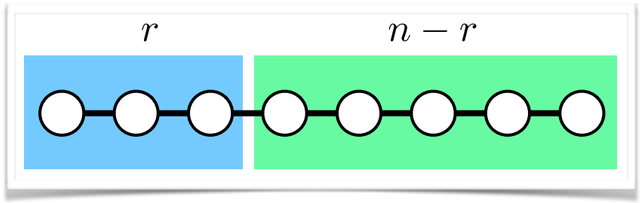
1. simple mathematical problems—build intuition & practice with concepts



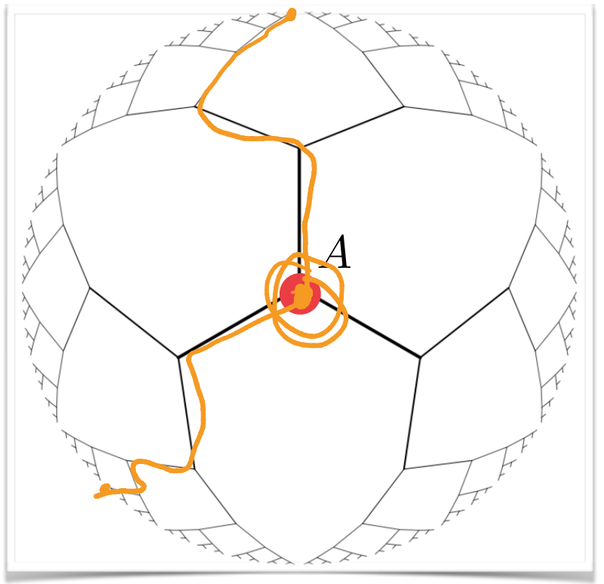
**calculate the diameter**




**closeness centrality**




**modularity of a line graph** $Q(r)$



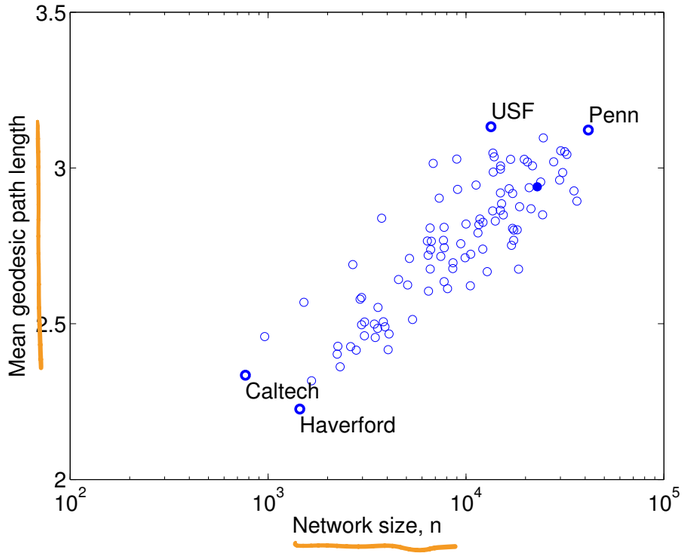

**betweenness of** $A$

centrality.

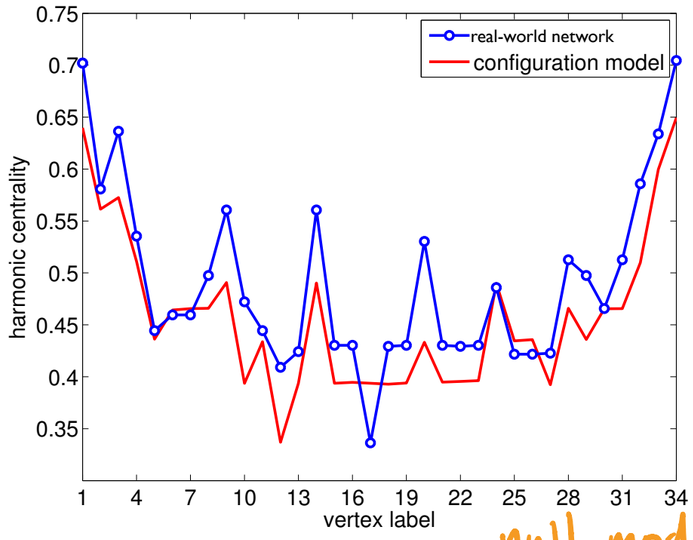
# lessons learned

**what works well**

2. analyze *real* networks—test understanding & practice with *implementing* methods



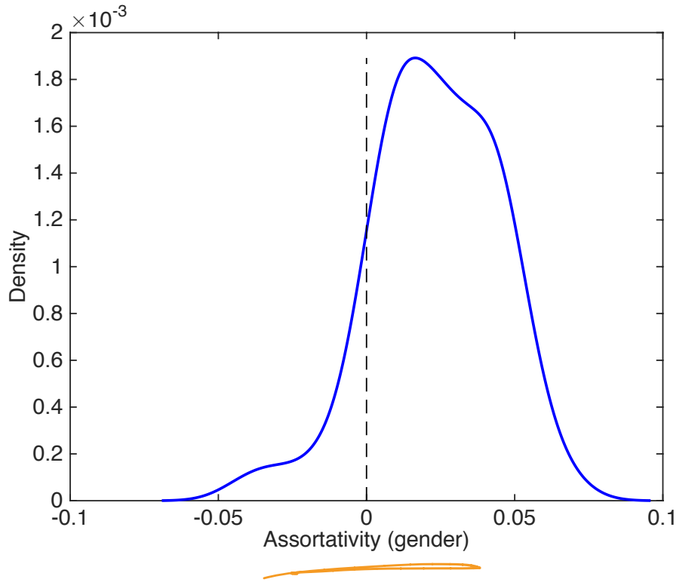

**mean geodesics and** $O(\log n)$




null models

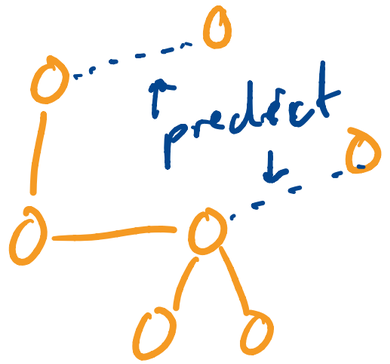
**node centrality vs configuration model**

(when is a pattern *interesting*?)



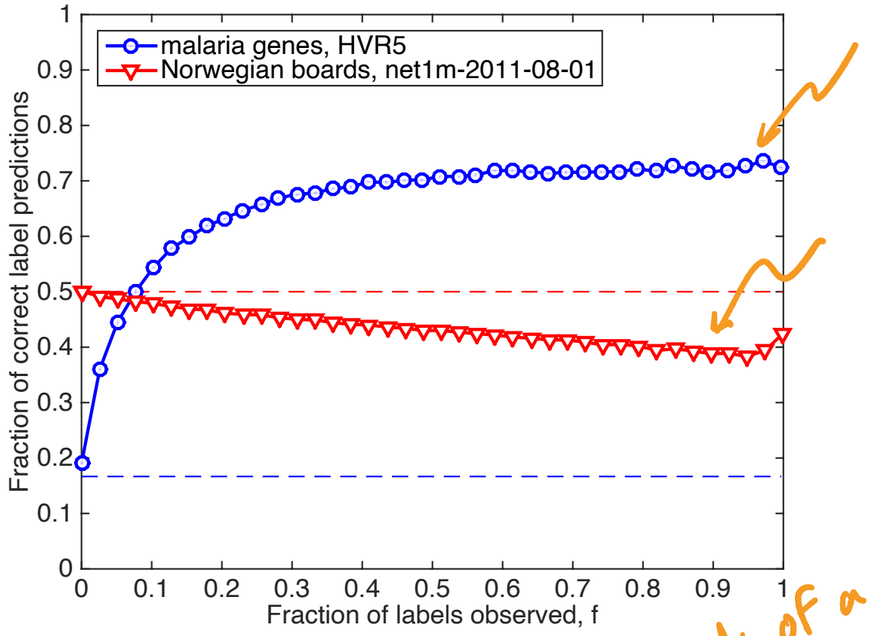

**attribute assortativity**

# lessons learned



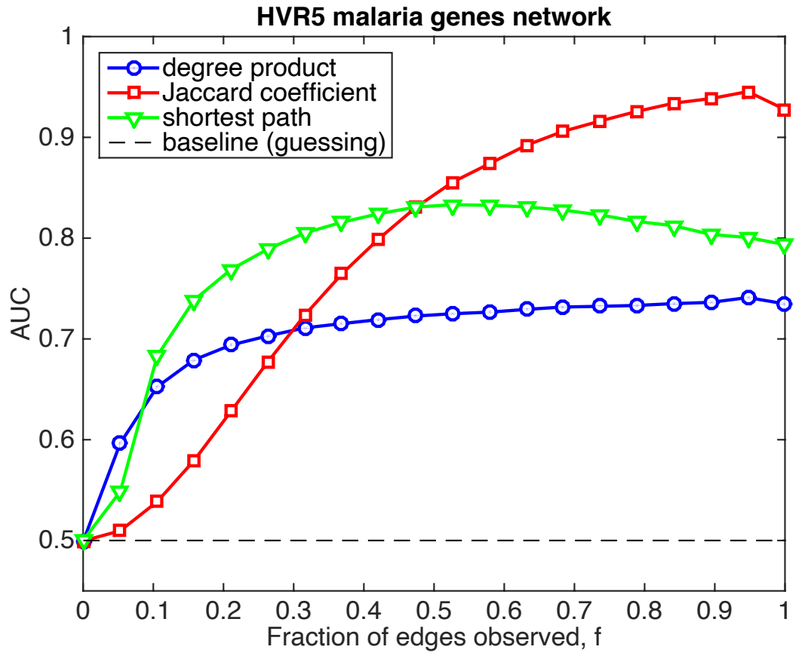

**what works well**

3. simple prediction tasks—test intuition & run numerical experiments




birds of a feather flock together.

**label prediction via homophily**

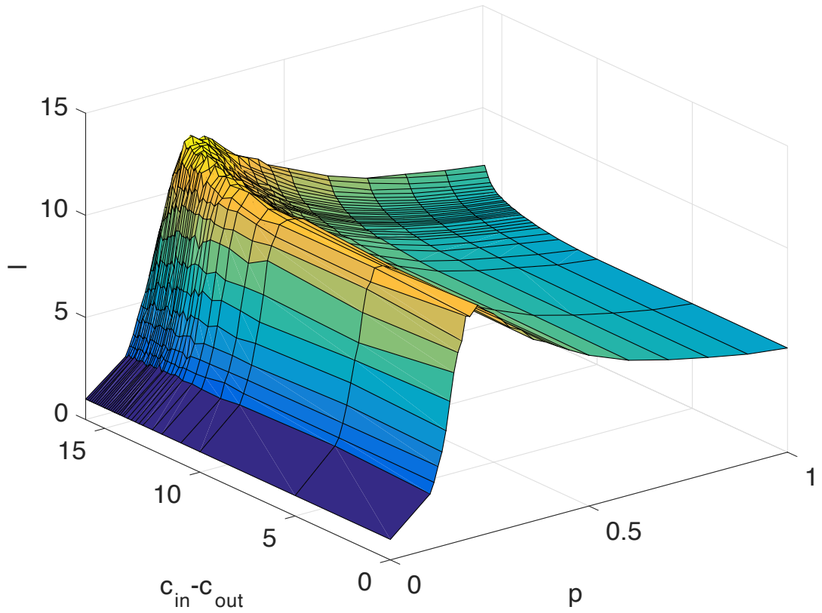



**link prediction via heuristic**

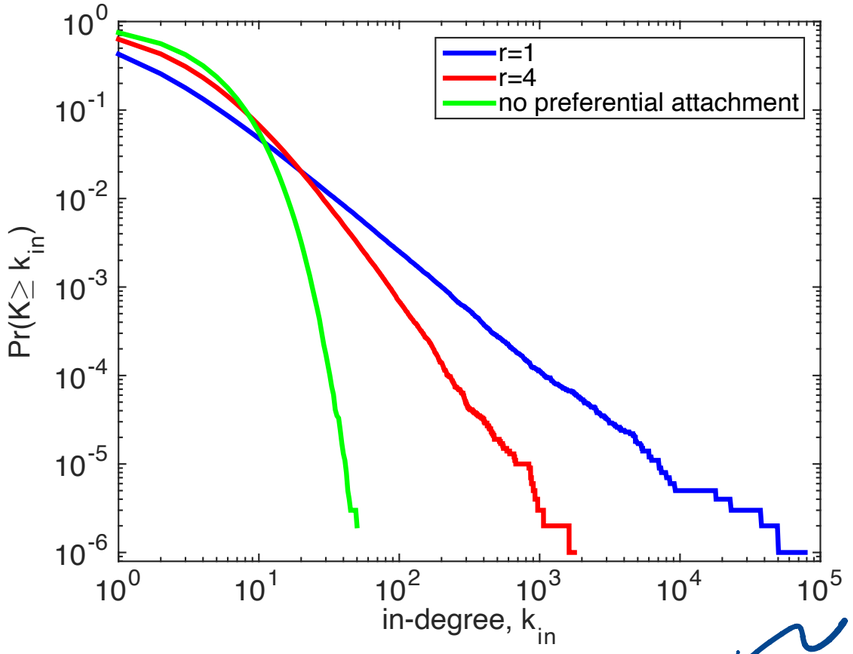
# lessons learned

**what works well**

4. simple simulations—explore dynamics vs structure & numerical experiments




**simulate epidemics (SIR) on planted partitions**




citations accumulate

**simulate Price's model**

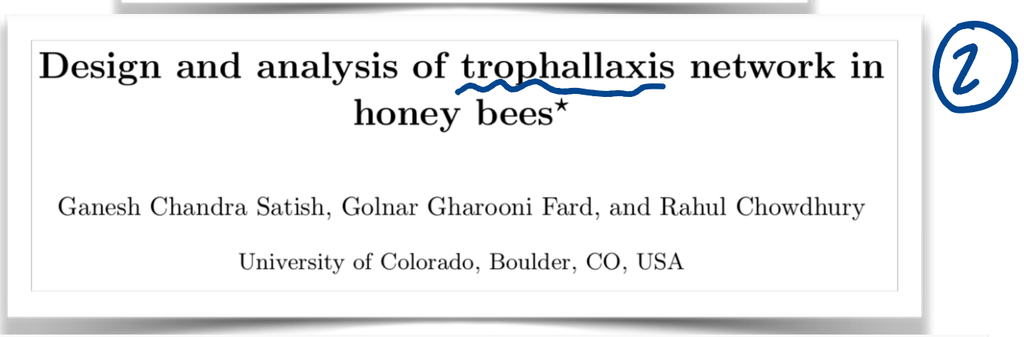
# lessons learned

## what works well

4. team projects—teamwork & exploring their (your!) own ideas

**Infinite Jest: An Elegant Hairball**

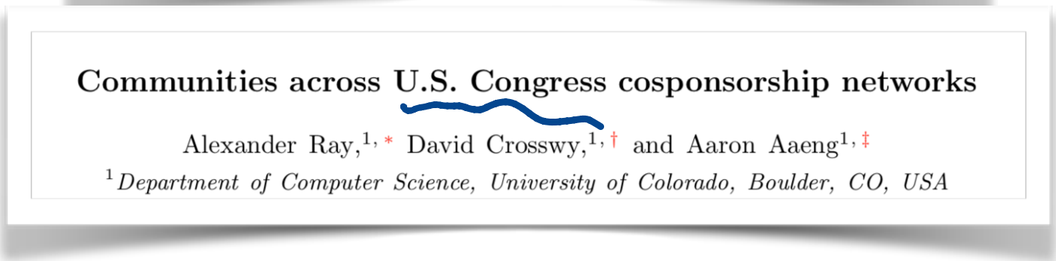
K. Hunter Wapman
Brian Lubars
Carl Mueller



**Design and analysis of trophallaxis network in honey bees***

Ganesh Chandra Satish, Golnar Gharooni Fard, and Rahul Chowdhury

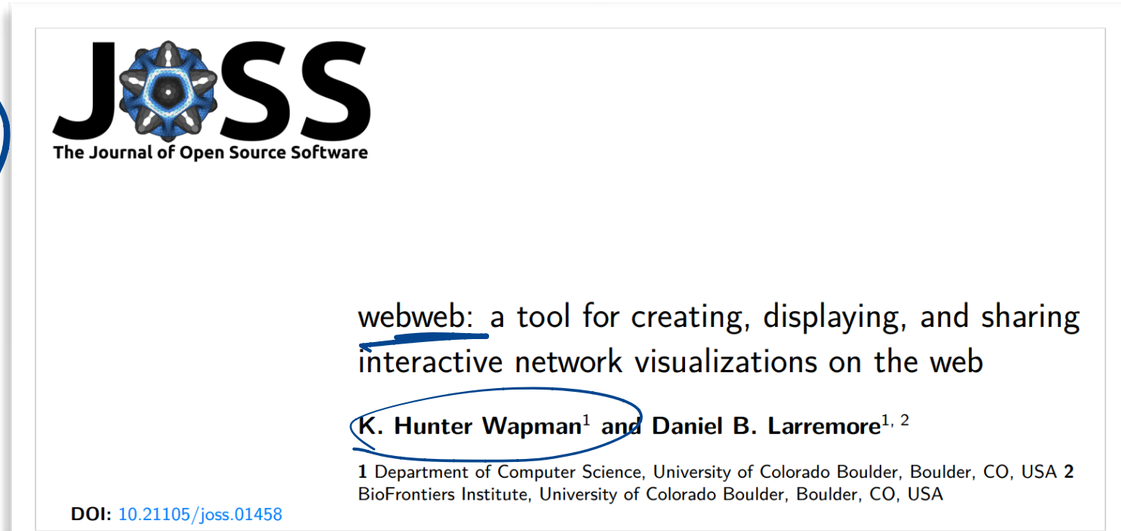
University of Colorado, Boulder, CO, USA



**Communities across U.S. Congress cosponsorship networks**

Alexander Ray,[1,*] David Crosswy,[1,†] and Aaron Aaeng[1,‡]

[1] *Department of Computer Science, University of Colorado, Boulder, CO, USA*



JOSS
The Journal of Open Source Software

webweb: a tool for creating, displaying, and sharing interactive network visualizations on the web

**K. Hunter Wapman**[1] **and Daniel B. Larremore**[1, 2]

**1** Department of Computer Science, University of Colorado Boulder, Boulder, CO, USA **2** BioFrontiers Institute, University of Colorado Boulder, Boulder, CO, USA

**DOI:** 10.21105/joss.01458

Morgan et al. *EPJ Data Science* (2018) 7:40
https://doi.org/10.1140/epjds/s13688-018-0166-4

EPJ Data Science
a SpringerOpen Journal

EPJ.org

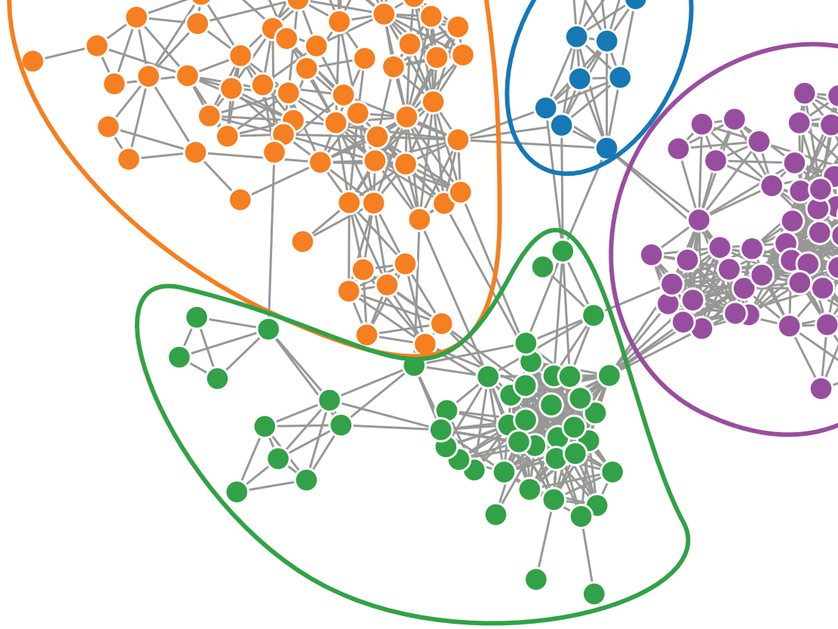
REGULAR ARTICLE Open Access

CrossMark

**Prestige drives epistemic inequality in the diffusion of scientific ideas**

Allison C. Morgan[1*], Dimitrios J. Economou[1], Samuel F. Way[1] and Aaron Clauset[1,2,3]

# key takeaways



- **network intuition is hard to develop!**
  good intuition draws on many skills
  (probability, statistics, computation, causal dynamics, etc.)

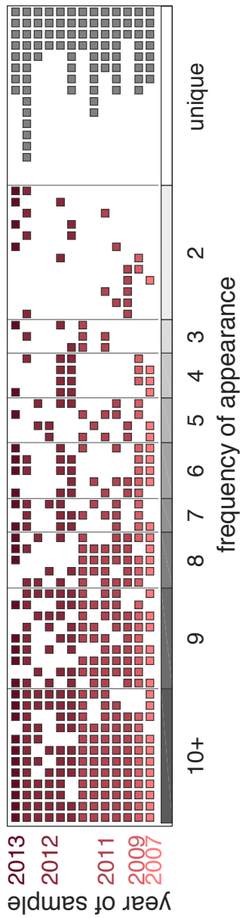
- best results come from
  1. exercises to get practice with calculations
  2. practice analyzing diverse real-world networks
  3. conducting out numerical experiments & simulations

- practical tasks are a pedagogical tool (e.g., link and label prediction)
- interpreting the results requires a good intuition
- null models are key conceptual idea: is a pattern interesting?
- networks are fun!

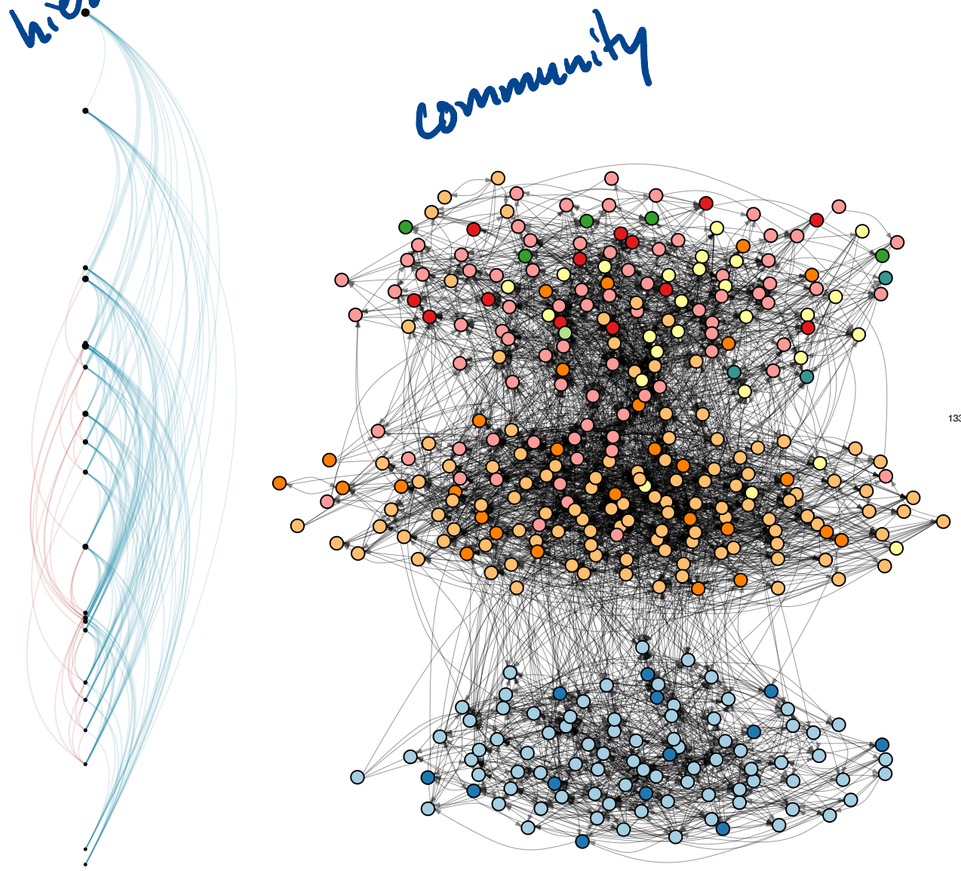
# about me

Assistant Professor, BioFrontiers Institute & Department of Computer Science
Previously: fellowships at Harvard, Santa Fe Institute; PhD Applied Math

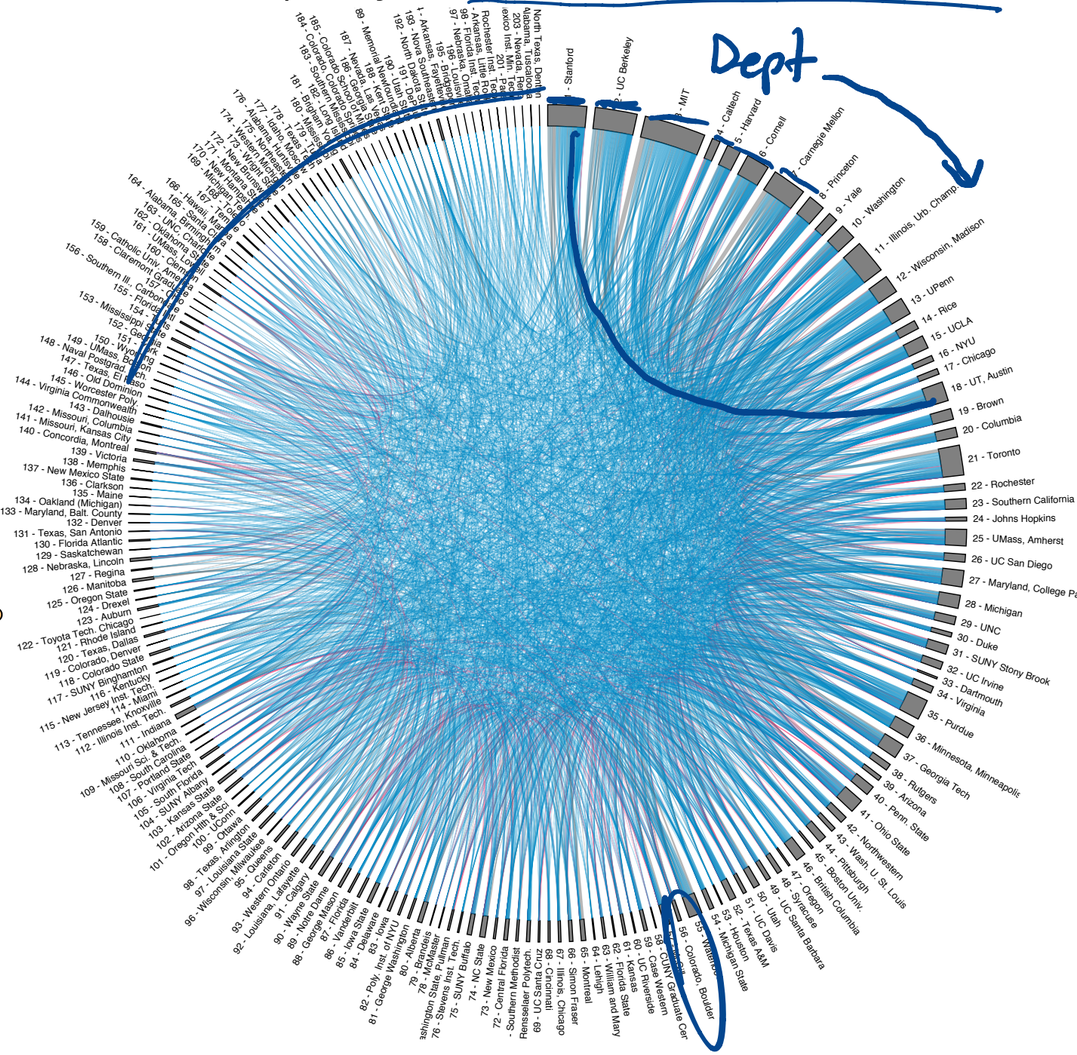
## malaria parasite evolution and epidemiology




## mathematical methods for statistical inference/analysis




## inequality in networked labor markets



# about me

Assistant Professor, BioFrontiers Institute & Department of Computer Science
Previously: fellowships at Harvard, Santa Fe Institute; PhD Applied Math

malaria parasite evolution and epidemiology

mathematical methods for statistical inference/analysis

inequality in networked labor markets

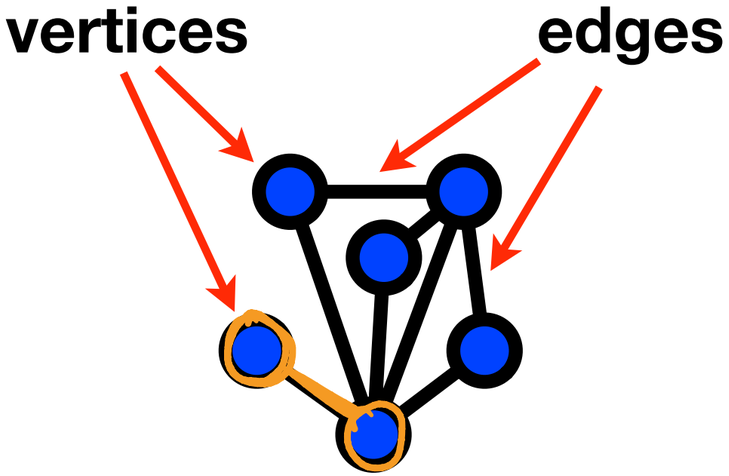
structured hierarchy in online dating markets

provably optimal play in generalized *misere* Connect 4

controllability of neuronal networks

1. defining a network
2. describing a network

# vertices & edges




**what is a vertex?**

$V$ distinct objects (vertices / nodes / actors)

set of vertices

math

application physics

soc. sci.

**when are two vertices connected?**

$E \subseteq V \times V$

all possible connections

pairwise relations (edges / links / ties)

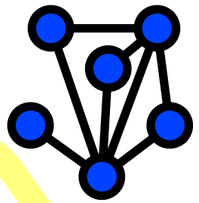
edge set

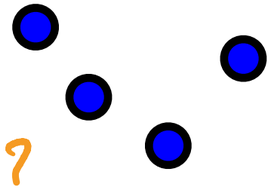
subset of

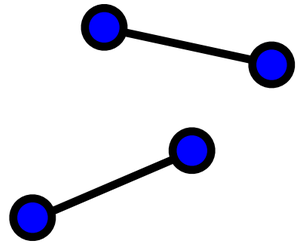
math

phys. application

soc. sci.







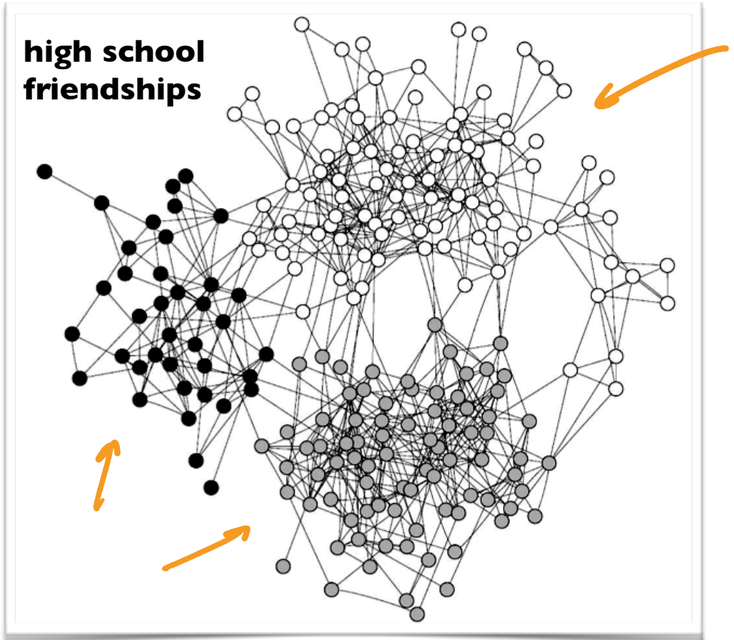
| | network | vertex | edge |
|---|---|---|---|
| telecommunications | Internet(1) | computer | IP network adjacency |
| | Internet(2) | autonomous system (ISP) | BGP connection |
| informational | software | function | function call |
| | World Wide Web | web page | hyperlink |
| | documents | article, patent, or legal case | citation |
| transportation | power grid transmission | generating or relay station | transmission line |
| | rail system | rail station | railroad tracks |
| | road network(1) | intersection | pavement |
| | road network(2) | named road | intersection |
| | airport network | airport | non-stop flight |
| social | friendship network | person | friendship |
| | sexual network | person | intercourse |
| biological | metabolic network | metabolite | metabolic reaction |
| | protein-interaction network | protein | bonding |
| | gene regulatory network | gene | regulatory effect |
| | neuronal network | neuron | synapse |
| | food web | species | predation or resource transfer |

# social networks

**vertex:** a person

**edge:** friendship, collaborations, sexual contacts, communication, authority, exchange, etc.

IRL

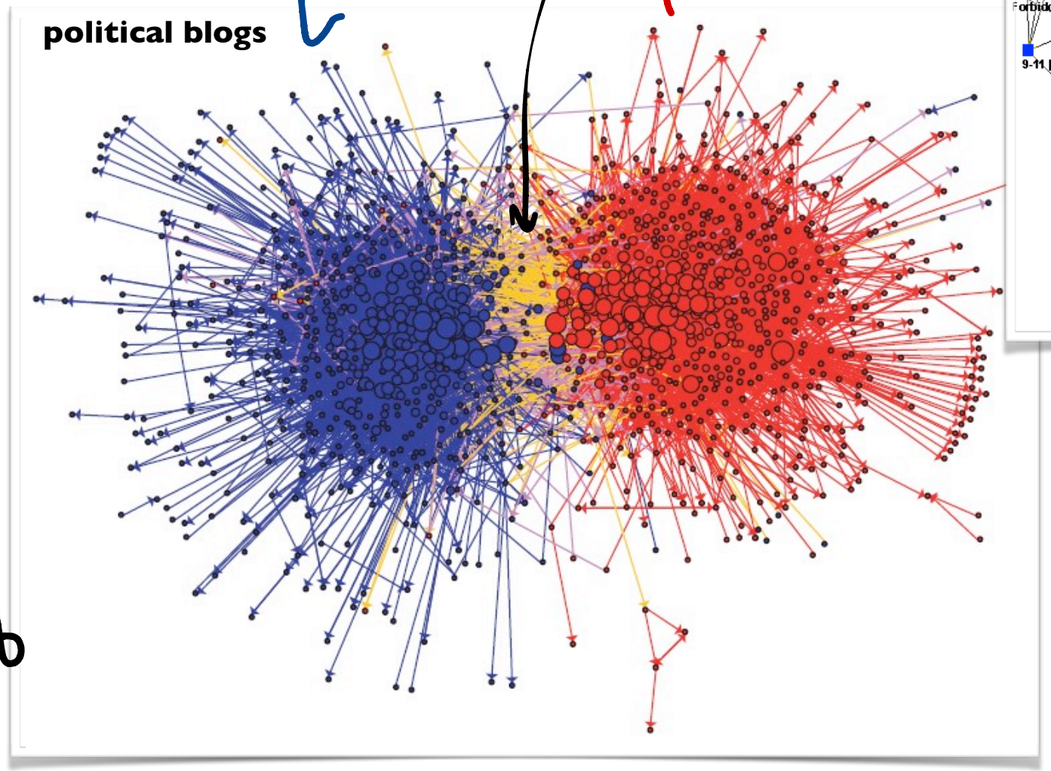


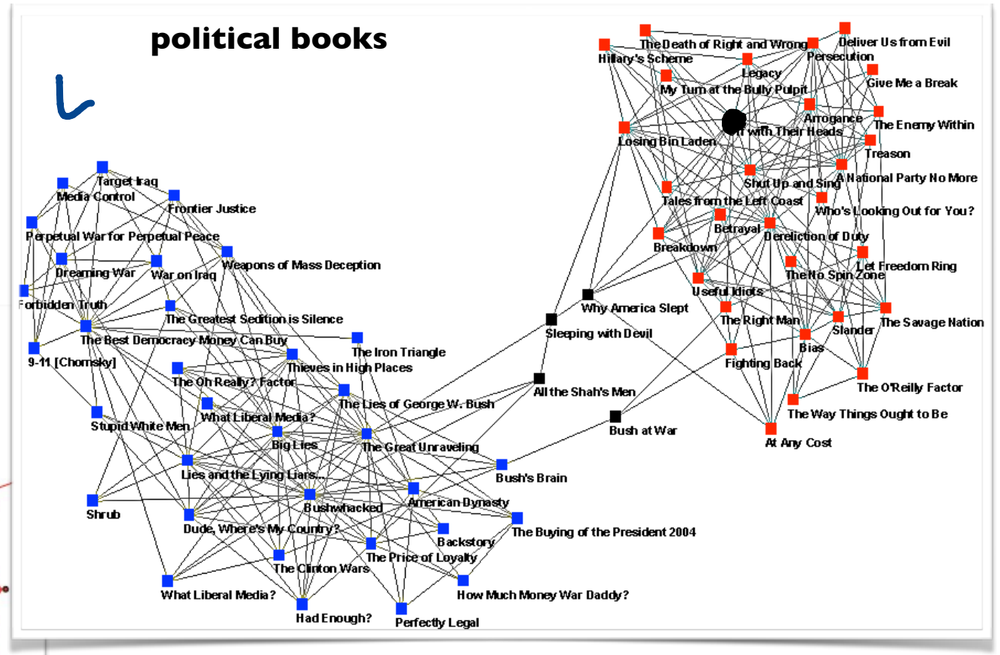

digital

# information networks

**vertex:** books, blogs, webpages, etc.

**edge:** citations, hyperlinks, recommendations, similarity, etc.



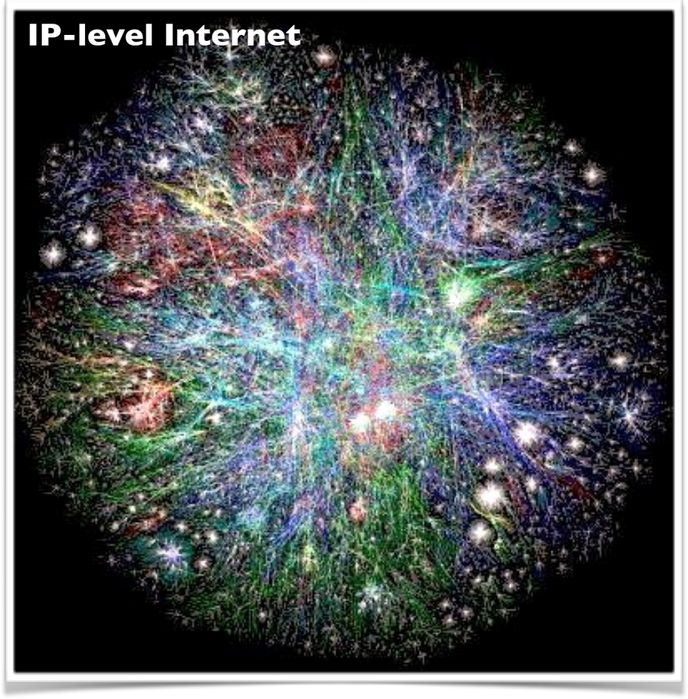

U.S.

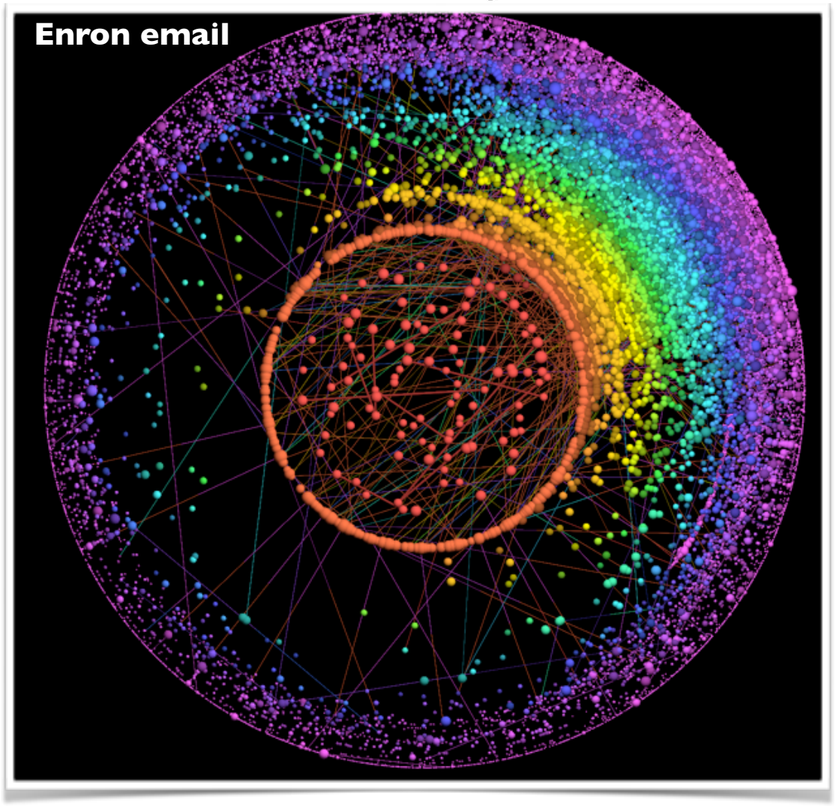
# communication networks

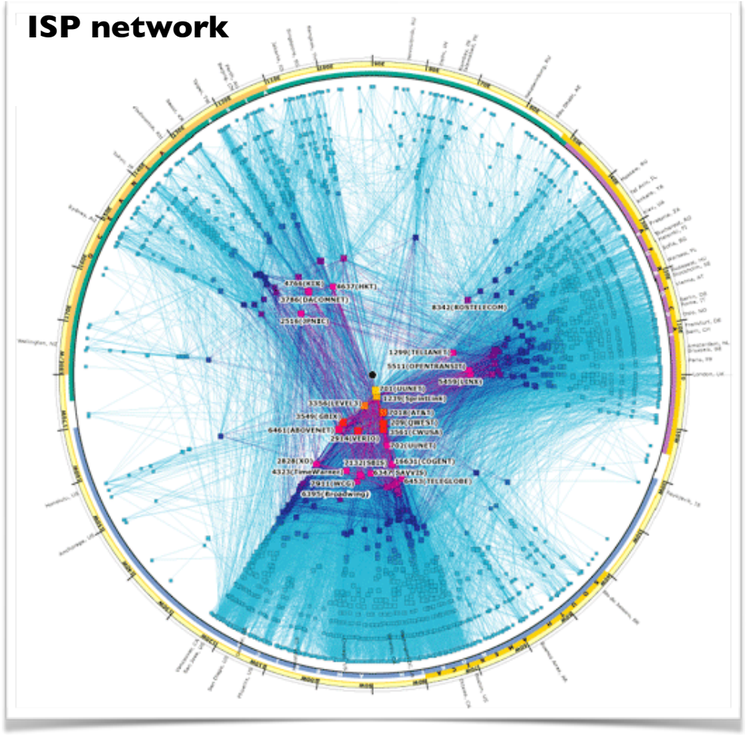
**vertex:** network router, ISP, email address, mobile phone number, etc.

**edge:** exchange of information

Enron



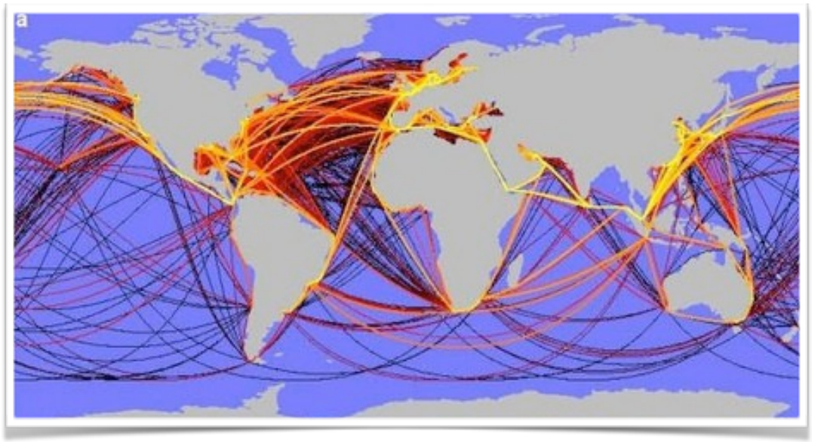
# transportation networks

**vertex:** city, airport, junction, railway station, river confluence, etc.

**edge:** physical transportation of material

US Interstates

spatial network.
planar.



global shipping

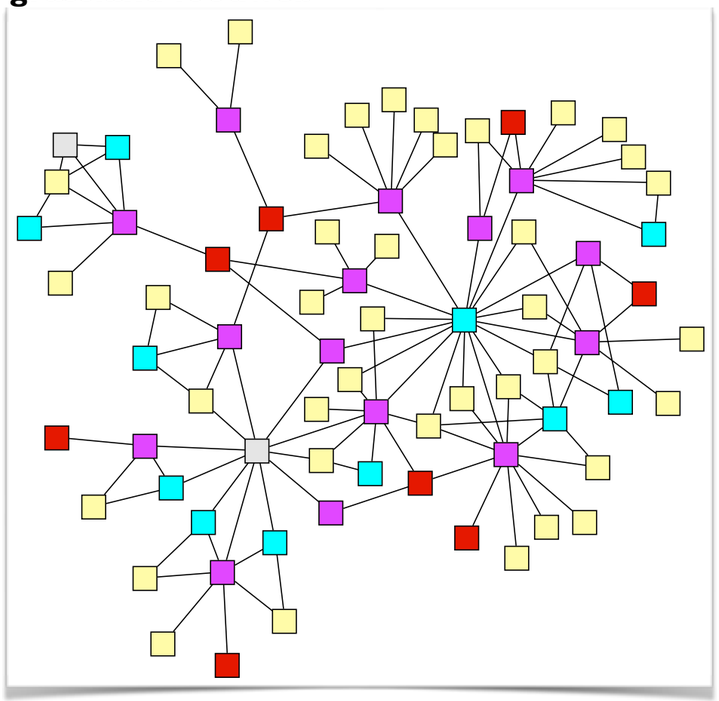
global air traffic

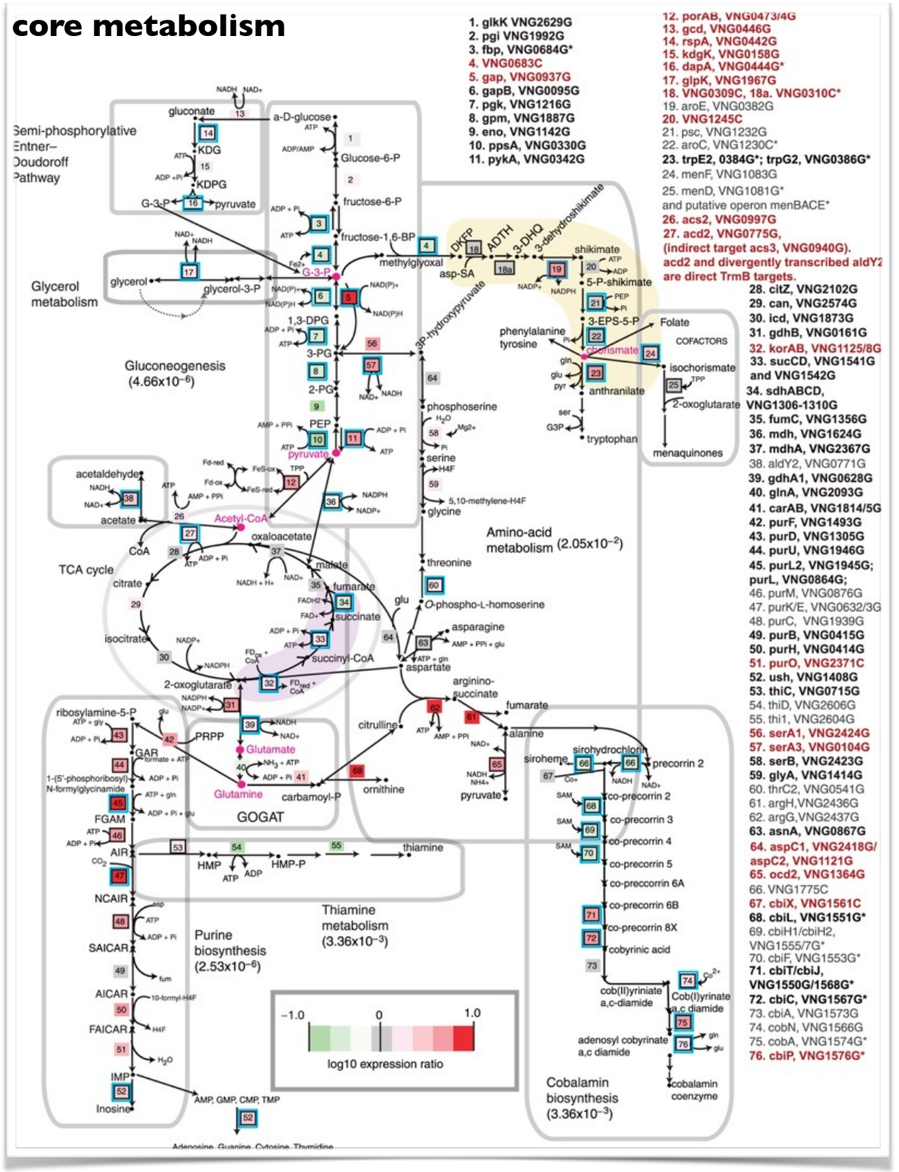
# biological networks

**vertex:** species, metabolic, protein, gene, neuron, etc.

**edge:** predation, chemical reaction, binding, regulation, activation, etc.

**grassland foodweb**





# what’s a network?

pop quiz

# what's a network?

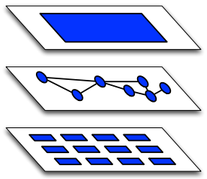
not a network:

- stars all belong to the same galaxy.
- what is an edge in this?
- what is a node?

Andromeda galaxy

yes, a network?

- nodes: celestial bodies
- edges: gravity connections, e.g.



# what's a network?

not a network!

What even is a spiral?

Are they distinct?

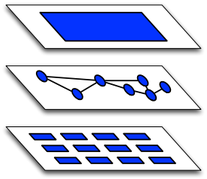
cauliflower fractal

yes, a network!

as it grows, each spiral has a parent spiral.

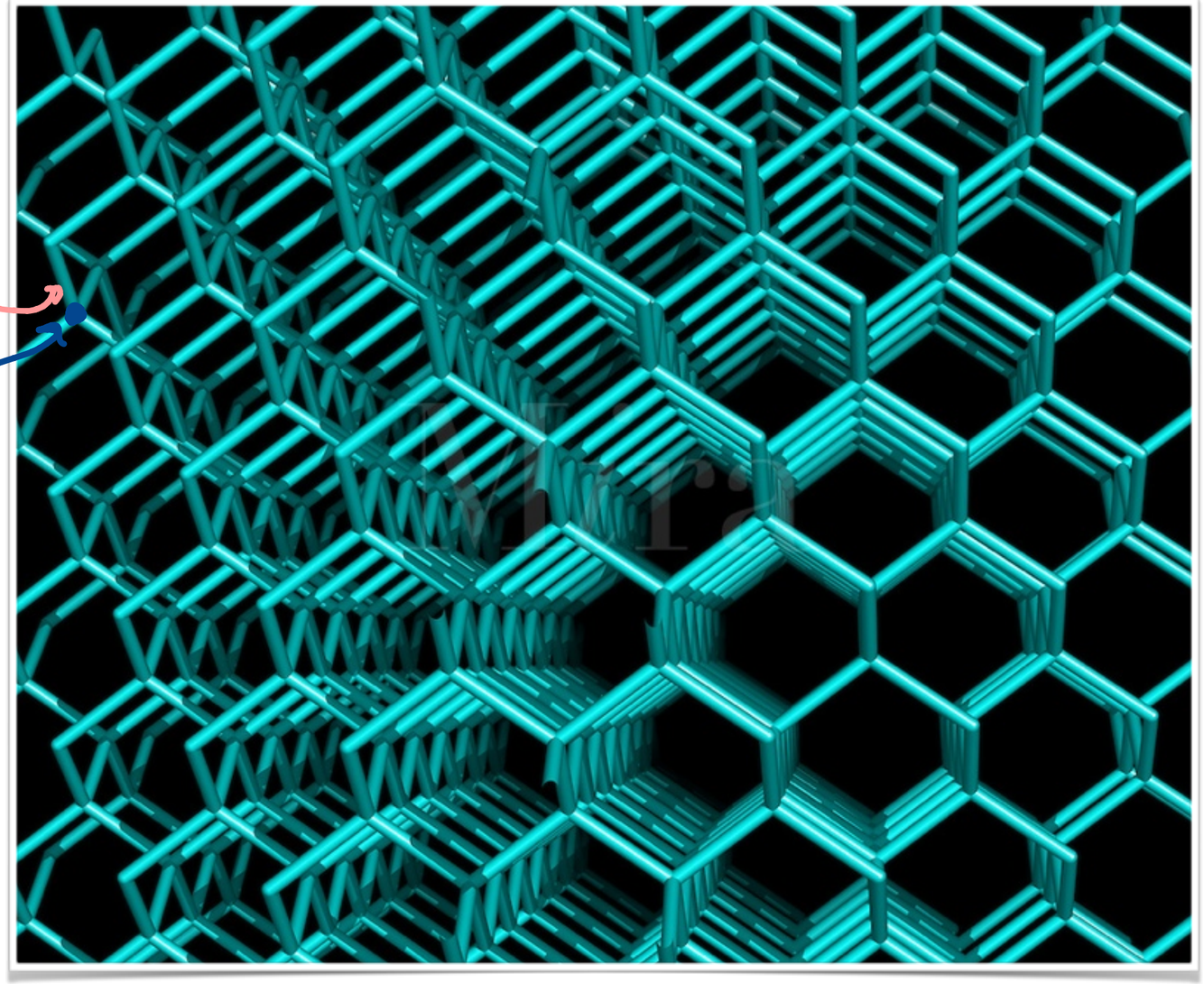
vertex: spiral

edge: spiral connects to its parent.



# what's a network?

yes, a network



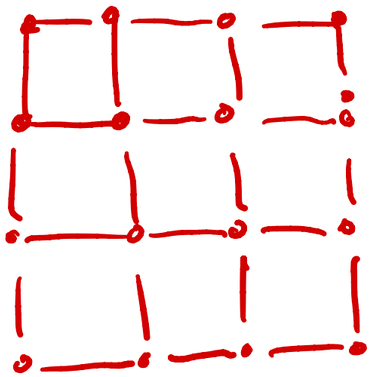

diamond lattice

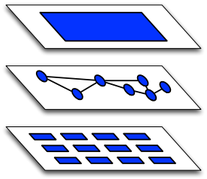
not a network !

why ?

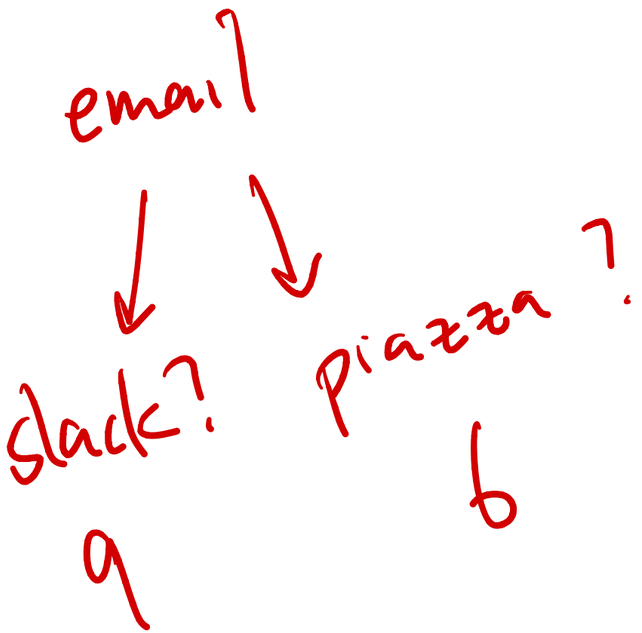
do we need all the overhead?

not complex

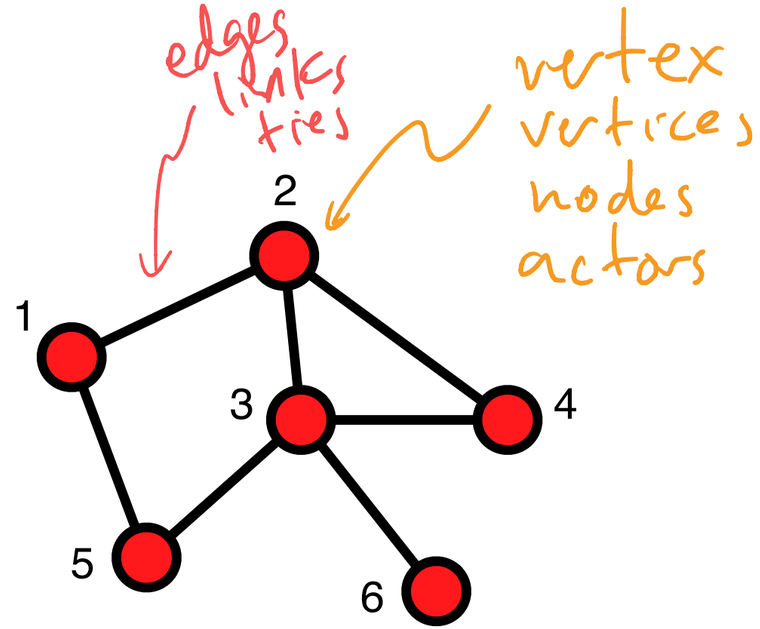




# representing networks



# a *simple* network



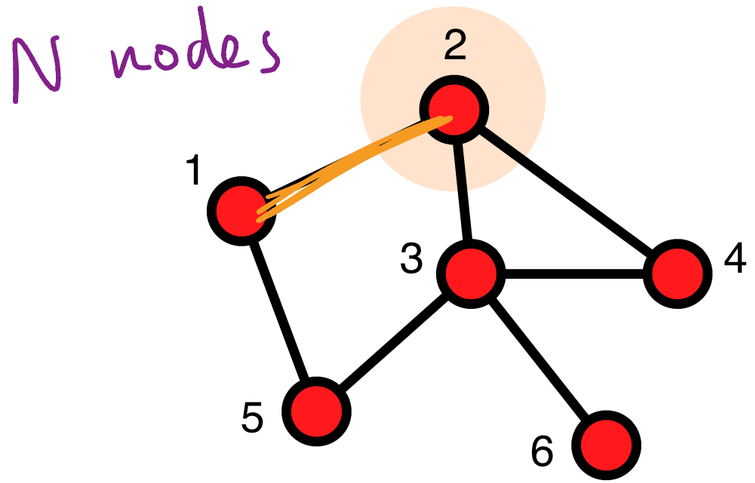

① Slack! (email)

② HW#1 posted.
Due Friday 9/4.
3 questions

simple ← undirected ⟷ e.g. Facebook.
unweighted — edge = presence/absence
no self-loops — no self loops = no self connections

# a *simple* network

N




N

undirected

unweighted

no self-loops

adjacency matrix

| $A$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 2 | 1 | 0 | 1 | 1 | 0 | 0 |
| 3 | 0 | 1 | 0 | 1 | 1 | 1 |
| 4 | 0 | 1 | 1 | 0 | 0 | 0 |
| 5 | 1 | 0 | 1 | 0 | 0 | 0 |
| 6 | 0 | 0 | 1 | 0 | 0 | 0 |

$A_{ij} = 1$ if $i \leftrightarrow j$

$A_{ij} = 1 = A_{ji}$

convenient

zeros!

space.

$N = 10^9$

adjacency list

| $A$ |
|---|
| $1 \rightarrow \{2, 5\}$ |
| $2 \rightarrow \{1, 3, 4\}$ |
| $3 \rightarrow \{2, 4, 5, 6\}$ |
| $4 \rightarrow \{2, 3\}$ |
| $5 \rightarrow \{1, 3\}$ |
| $6 \rightarrow \{3\}$ |

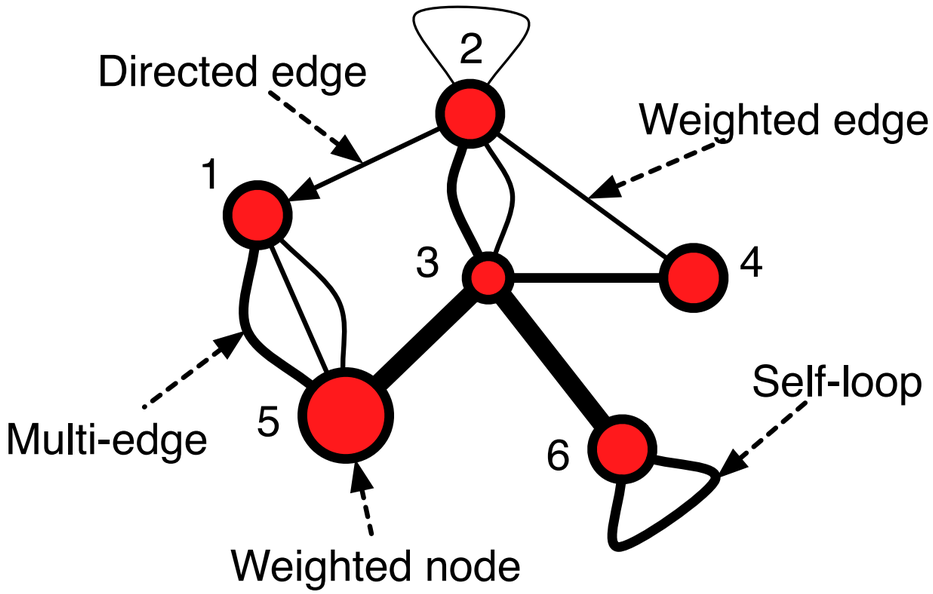
compact.

less manipulable.

nodes

their connections

# a less simple network



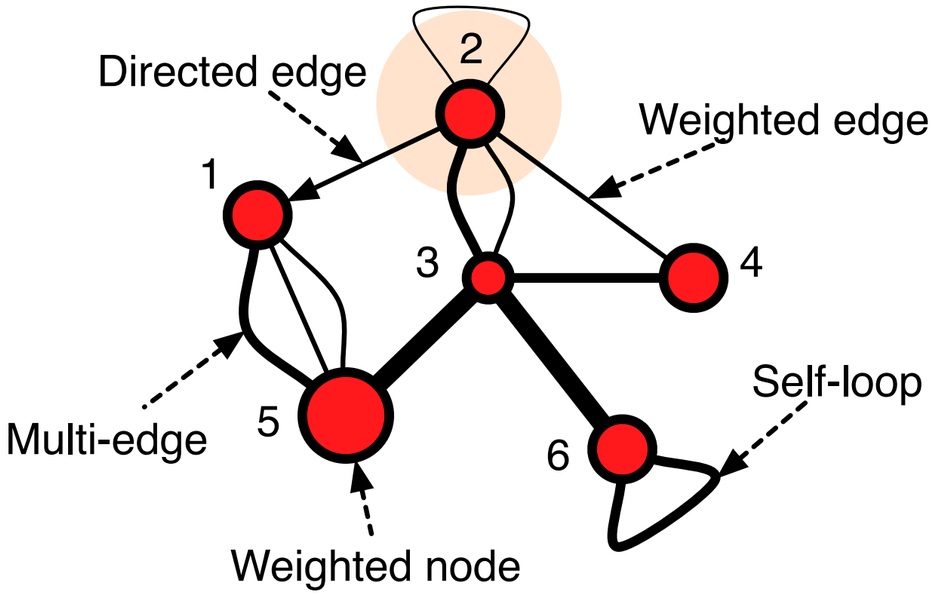

~~undirected~~

~~unweighted~~

~~no self-loops~~

# a less simple network

most of the time




adjacency matrix

| $A$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | $\{1,1,2\}$ | 0 |
| 2 | 1 | $\frac{1}{2}$ | $\{2,1\}$ | 1 | 0 | 0 |
| 3 | 0 | $\{2,1\}$ | 0 | 2 | 4 | 4 |
| 4 | 0 | 1 | 2 | 0 | 0 | 0 |
| 5 | $\{1,1,2\}$ | 0 | 4 | 0 | 0 | 0 |
| 6 | 0 | 0 | 4 | 0 | 0 | 2 |

adjacency list

| $A$ | |
|---|---|
| 1 | $\rightarrow \{(5,1),(5,1),(5,2)\}$ |
| 2 | $\rightarrow \{(1,1),(2,\frac{1}{2}),(3,2),(3,1),(4,1)\}$ |
| 3 | $\rightarrow \{(2,2),(2,1),(4,2),(5,4),(6,4)\}$ |
| 4 | $\rightarrow \{(2,1),(3,2)\}$ |
| 5 | $\rightarrow \{(1,1),(1,1),(1,2),(3,4)\}$ |
| 6 | $\rightarrow \{(3,4),(6,2)\}$ |

no 100% universal standard.

BUT

Adj Mtx + Edge list common.

# directed networks

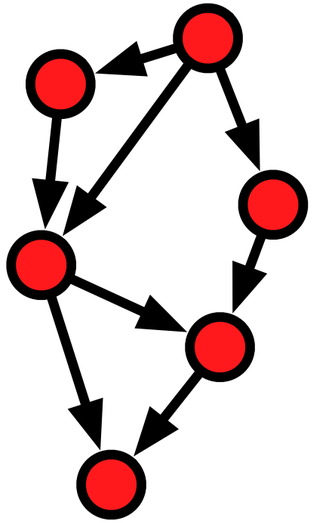
$$A_{ij} \neq A_{ji}$$

citation networks

foodwebs*

epidemiological

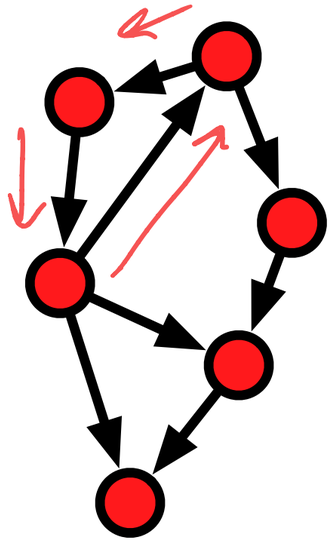
others?



"DAG"

directed acyclic graph

no cycles



directed graph

WWW

friendship?

flows of goods, information

economic exchange

dominance

neuronal

transcription

time travelers

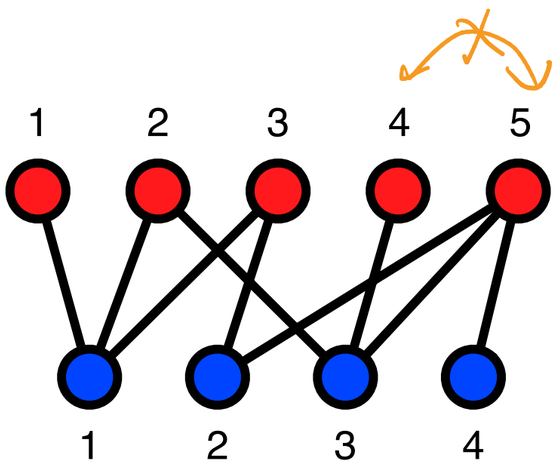
arXiv

# bipartite networks

multi-partite networks.

2 parts

**bipartite network**




authors

paper

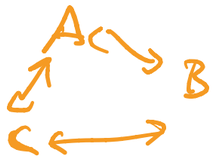
**no within-type edges**

- authors & papers
- actors & movies/scenes
- musicians & albums
- people & online groups
- people & corporate boards
- people & locations (checkins)
- metabolites & reactions
- genes & substrings
- words & documents
- plants & pollinators

bi: rule: connect to other class of node. not your own.

tri: can't connect to your own type...

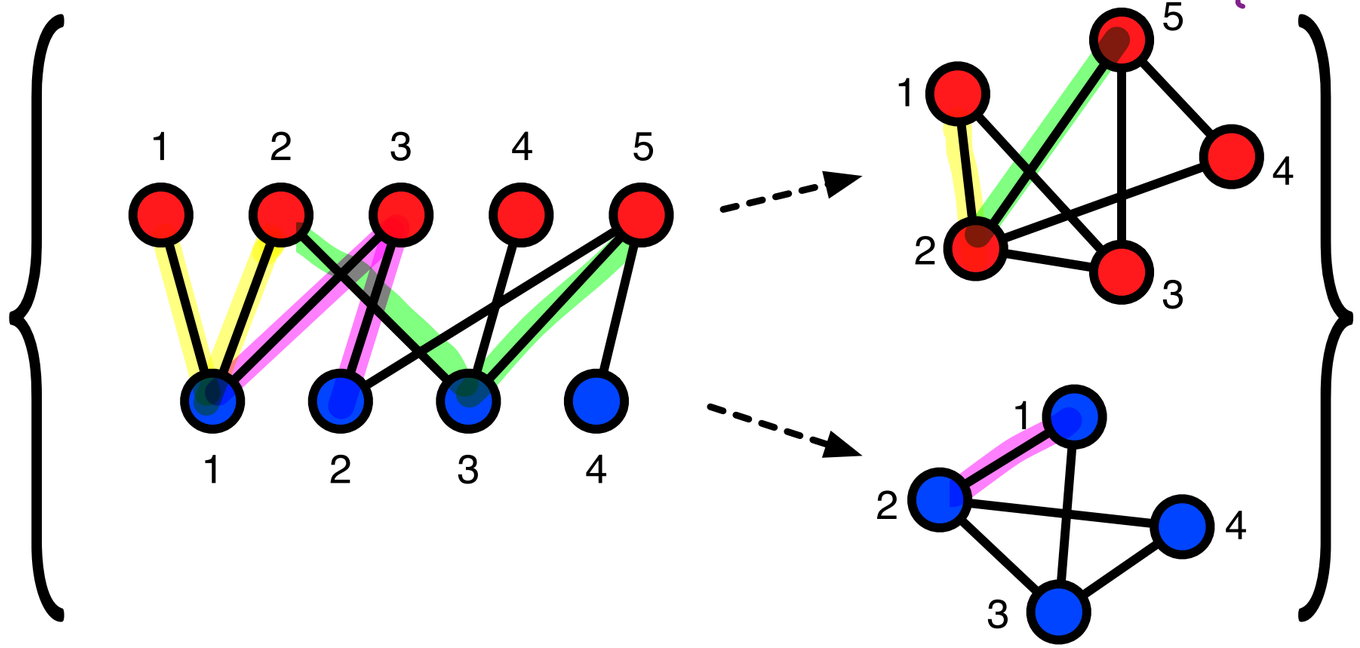
A ↔ B ↔ C



# bipartite networks

two-mode networks.

two nodes connected if they share a common neighbor in bip. net.




- authors & papers
- actors & movies/scenes
- musicians & albums
- people & online groups
- people & corporate boards
- people & locations (checkins)
- metabolites & reactions
- genes & substrings
- words & documents
- plants & pollinators

co-starring network. — movestars-to-movestars.

cast-overlap network — movies-to-movies

# temporal networks

**any network over time**

discrete time (snapshots), edges $(i, j, t)$

continuous time, edges $(i, j, t_s, \Delta t)$

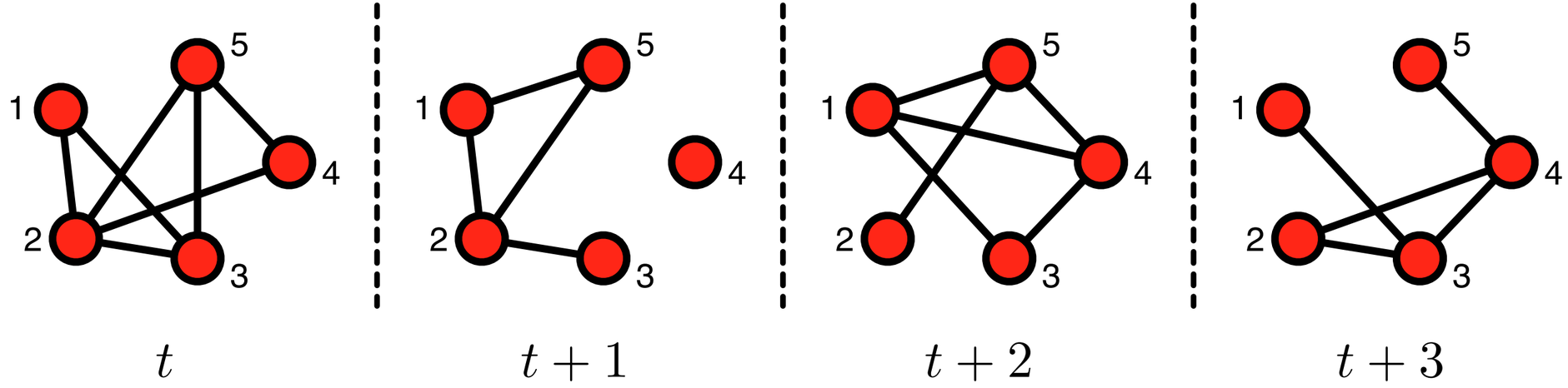
(i, j) typical

(i, j, t) time-varying

time slot edge exists in.

start time

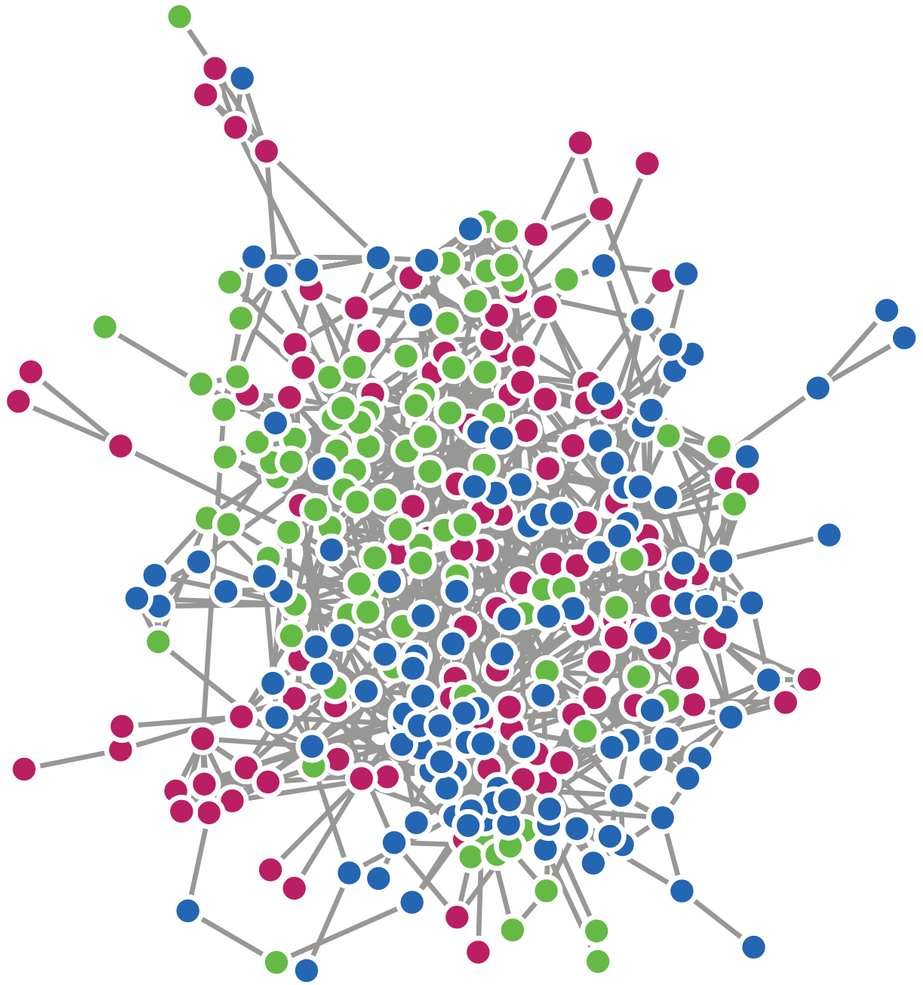
duration




ex: trophylaxis network

# describing networks

what networks look like



# describing networks

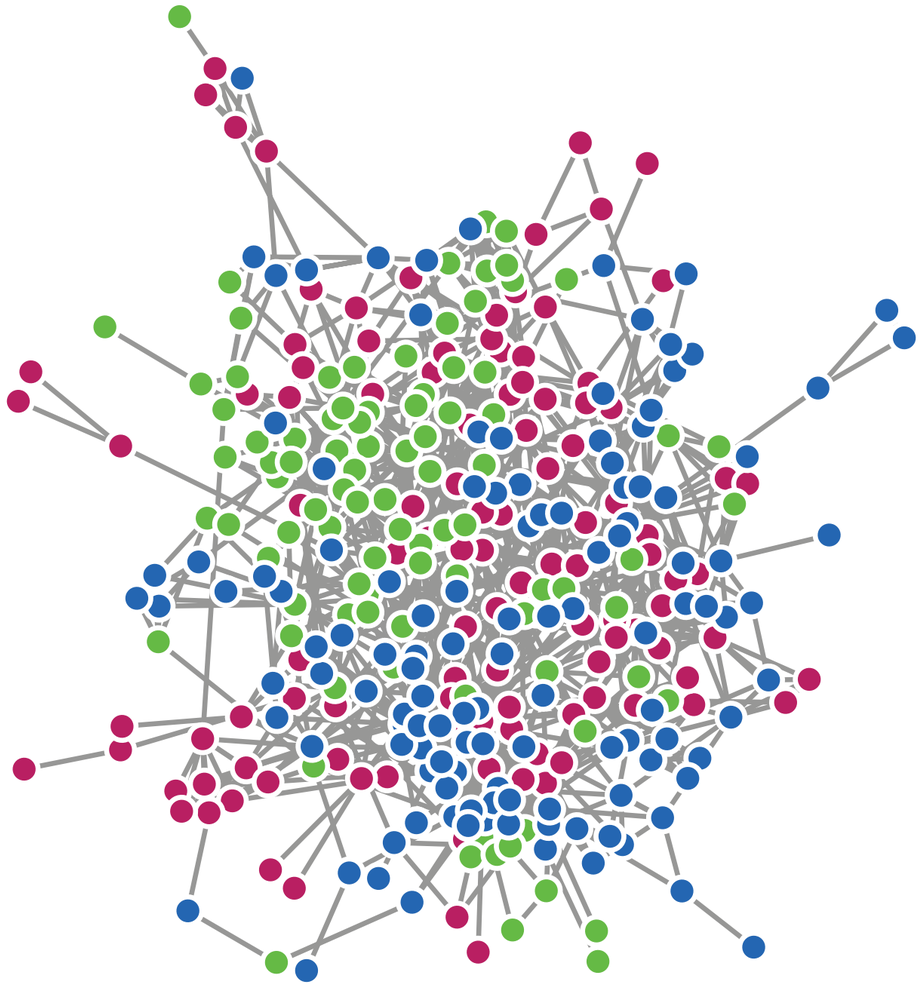
what networks look like

questions:

- how are the edges organized?
- how do vertices differ?
- does network location matter?
- are there underlying patterns?

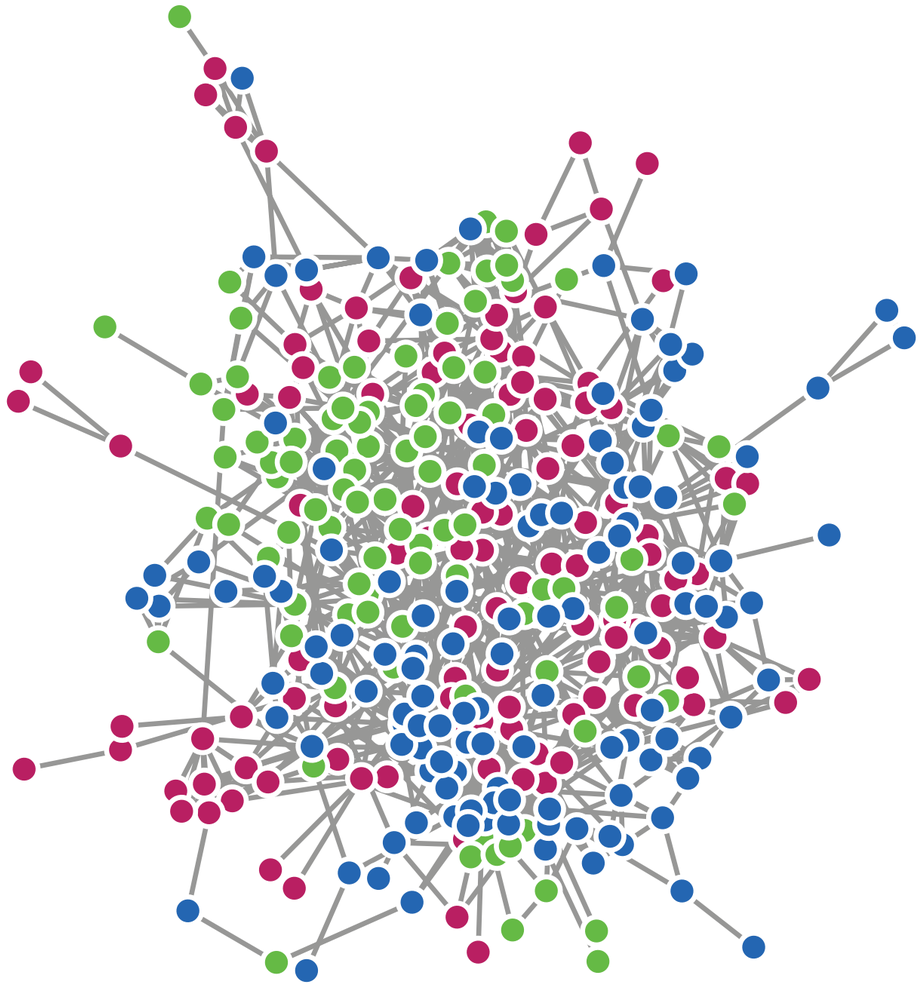
what we want to know

- what processes shape these networks?
- how can we tell?



# describing networks

a first step : **describe its features**

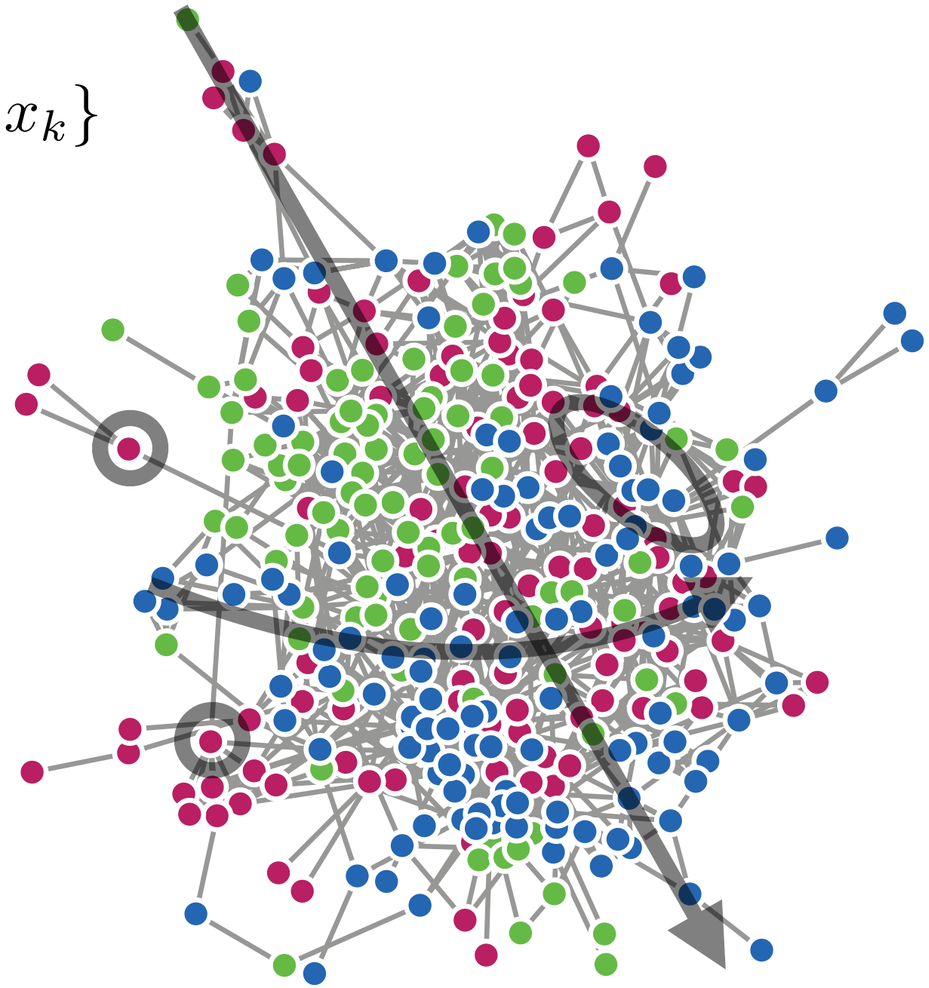


# describing networks

a first step : **describe its features**

$$f : G \to \{x_1, \ldots, x_k\}$$

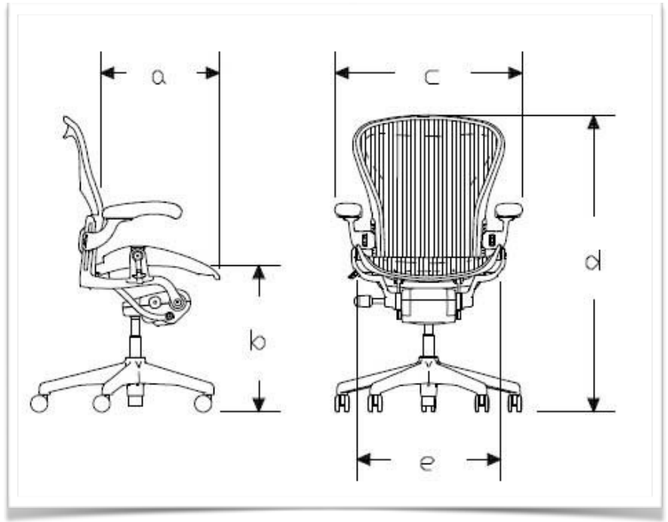
- degree distributions
- short-loop density (triangles, etc.)
- shortest paths (diameter, etc.)
- vertex positions
- correlations between these



# describing networks

a first step : **describe its features**

$$f : \text{object} \rightarrow \{\theta_1, \ldots, \theta_k\}$$



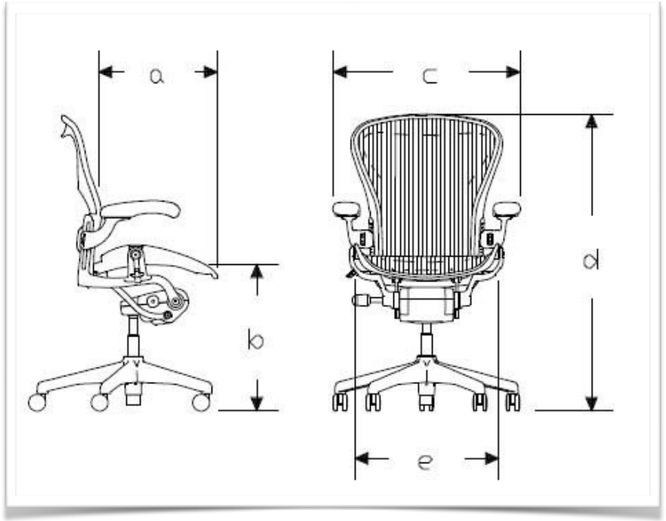
# describing networks

a first step : **describe its features**

$$f : \text{object} \rightarrow \{\theta_1, \ldots, \theta_k\}$$

- physical dimensions
- material density, composition
- radius of gyration
- correlations between these

helpful for exploration, but not what we want…



# describing networks

what we want : **understand its structure**

$$f : \text{object} \rightarrow \{\theta_1, \ldots, \theta_k\}$$

- what are the fundamental parts?
- how are these parts organized?
- where are the degrees of freedom $\vec{\theta}$ ?
- how can we define an abstract class?
- structure — dynamics — function?

what does ***local-level structure*** look like?

what does ***large-scale structure*** look like?

how does ***structure constrain*** function?