Network Analysis and Modeling

Aaron Clauset
@aaronclauset
Assistant Professor of Computer Science
University of Colorado Boulder
External Faculty, Santa Fe Institute

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

who are network scientists?

- Physicists
- Computer Scientists
- Applied Mathematicians
- Statisticians
- Biologists
- Ecologists
- Sociologists
- Political Scientists

it's a big community!

who are network scientists?

- Physicists
- Computer Scientists
- Applied Mathematicians
- Statisticians
- Biologists
- Ecologists
- Sociologists
- Political Scientists

it's a big community!

- different traditions
- different tools
- different questions

who are network scientists?

- Physicists
- Computer Scientists
- Applied Mathematicians
- Statisticians
- Biologists
- Ecologists
- Sociologists
- Political Scientists

it's a big community!

- different traditions
- different tools
- different questions

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

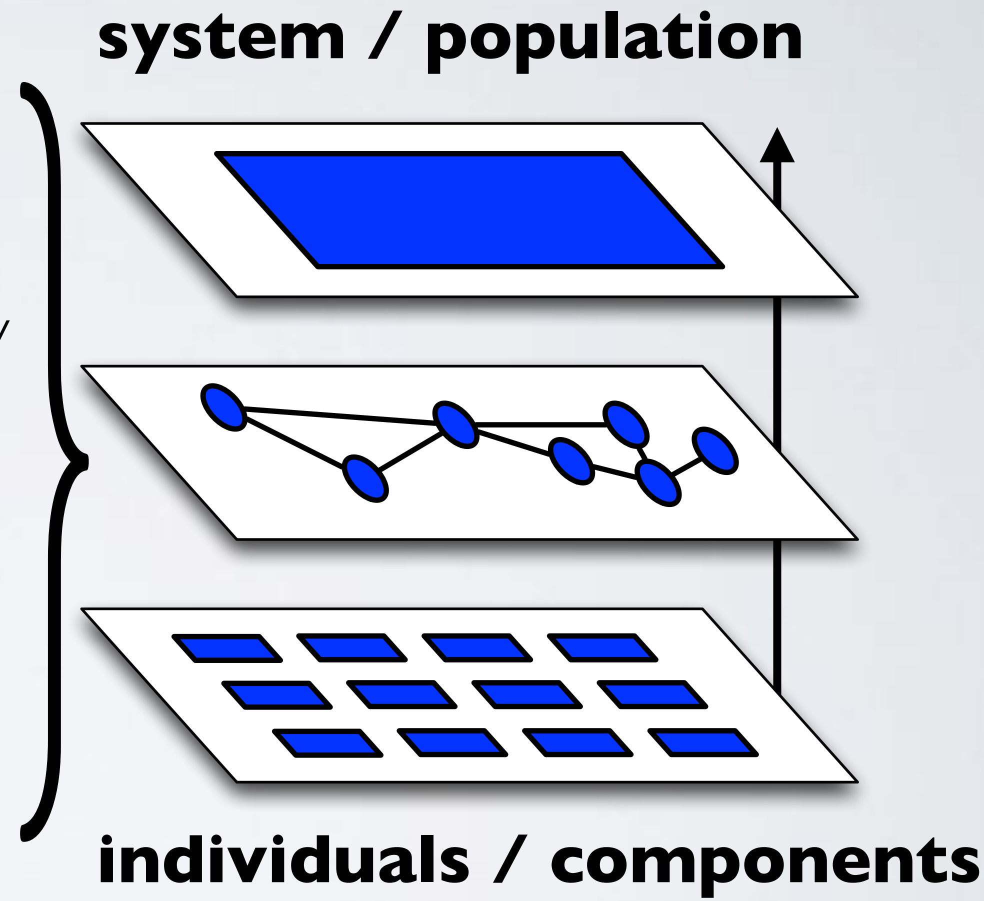
who are network scientists?

Physicists	phase transitions, universality
Computer Scientists	data / algorithm oriented, predictions
Applied Mathematicians	dynamical systems, diff. eq.
Statisticians	inference, consistency, covariates
Biologists	experiments, causality, molecules
Ecologists	observation, experiments, species
Sociologists	individuals, differences, causality
Political Scientists	rationality, influence, conflict

what are networks?

what are networks?

- an approach
- a mathematical representation
- provide structure to complexity
- *structure above*
  individuals / components
- *structure below*
  system / population



CSCI 5352 Network Analysis and Modeling : 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

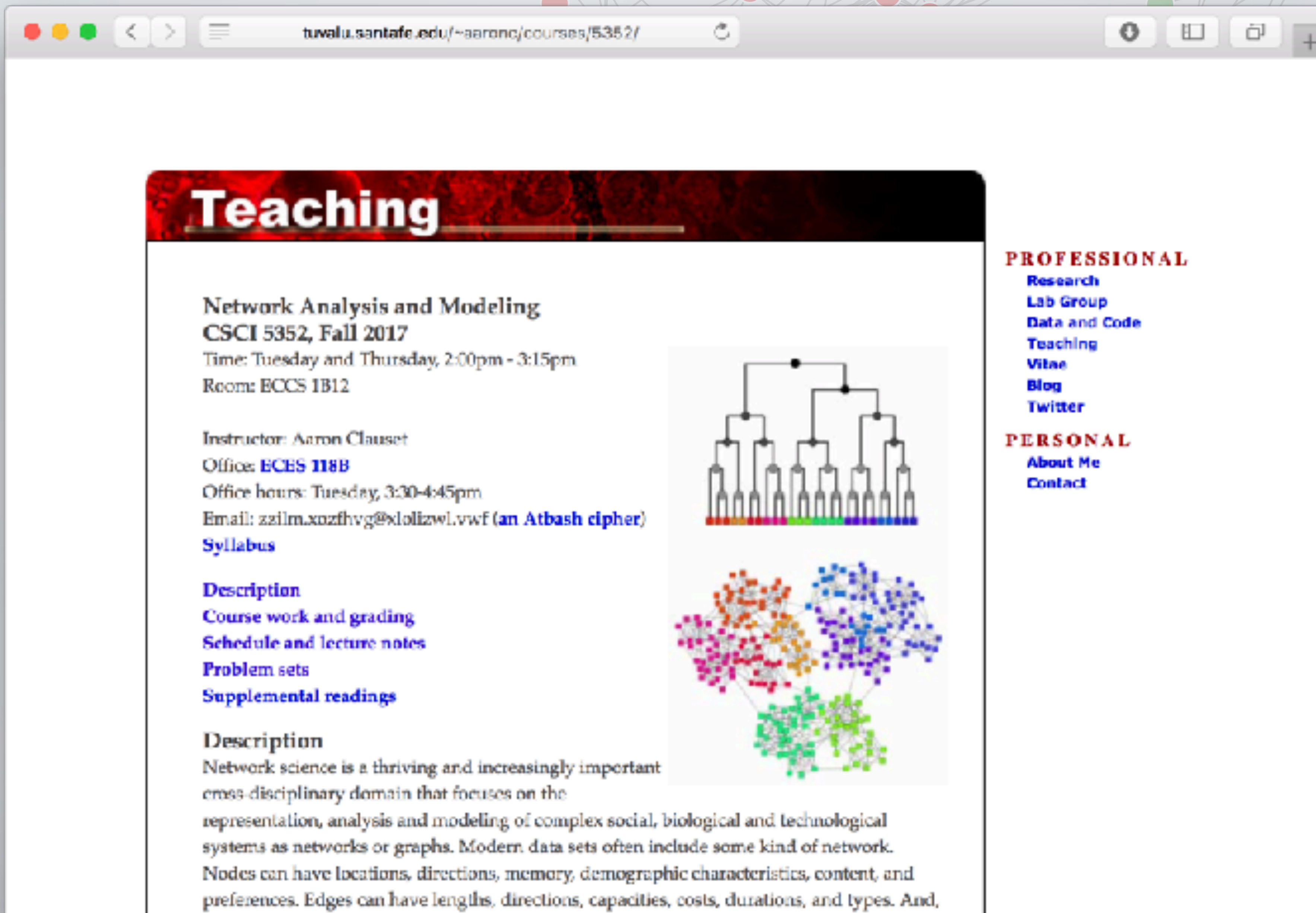
4. build *familiarity with advanced techniques* for *exploring / testing hypotheses* about networks

◆ building intuition
● basic concepts, tools
★ practical tools
■ advanced tools

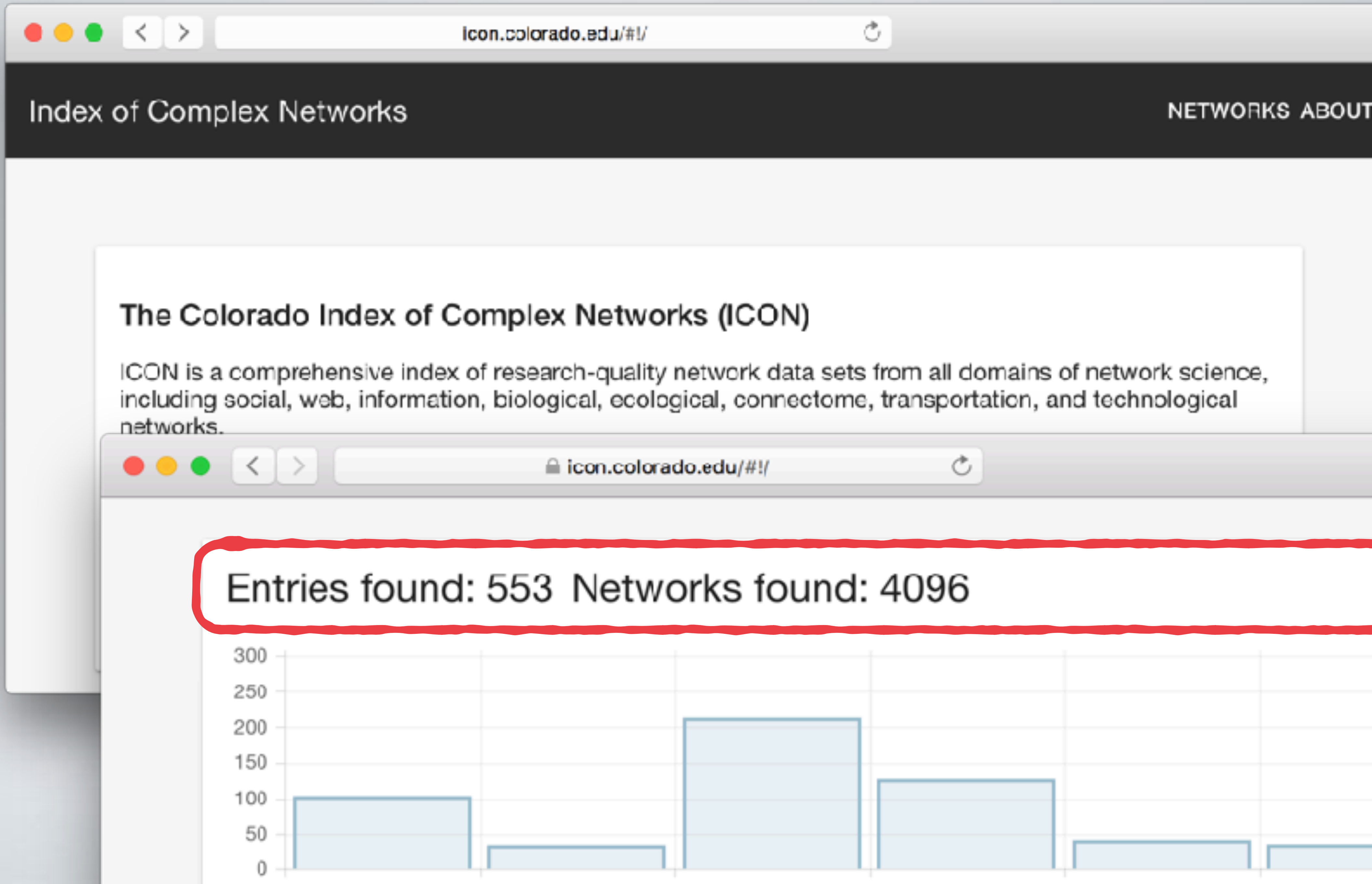
Course schedule (roughly) :

1.	network basics	◆	●		
2.	centrality measures	◆	●	★	
3.	random graphs (simple)	◆	●		
4.	configuration model	◆		★	
5.	large-scale structure (communities, hierarchies, etc.)	◆	●		
6.	probabilistic generative models (SBMs)			★	■
7.	metadata, label and link prediction		●	★	
8.	spreading processes (social, biological, SI-type)	◆	●		
9.	data wrangling + data sampling (artifacts)			★	■
10.	role of statistics in hypothesis generation / testing	◆	●		
11.	spatial networks	◆			■
12.	citations networks, dynamics, preferential attachment		●		■
13.	temporal networks			★	■
14.	student project presentations				■

Course webpage: http://santafe.edu/~aaronc/courses/5352/



Network data for assignments



lessons learned from past instances

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

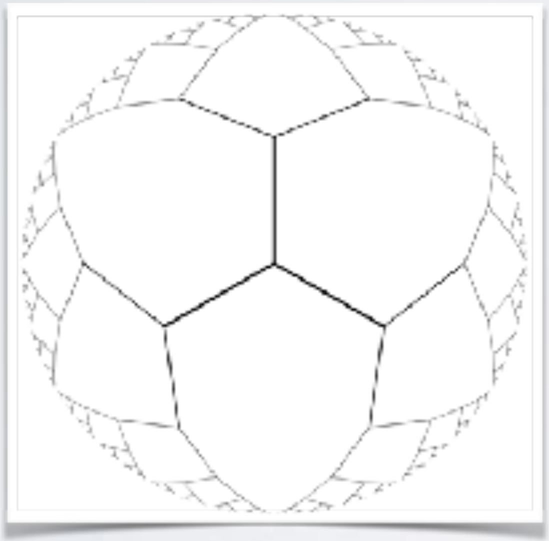
2. can't teach all of these things to all types of students!

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

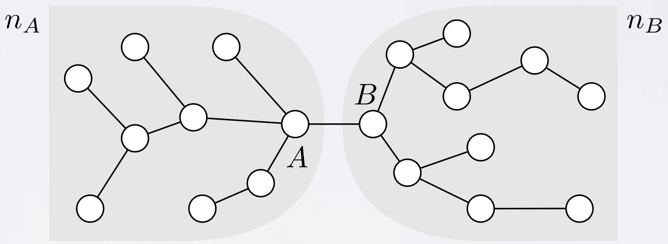
lessons learned from past instances

what works well:

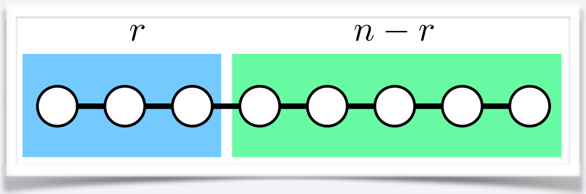
1. simple mathematical problems
   build intuition + practice with concepts



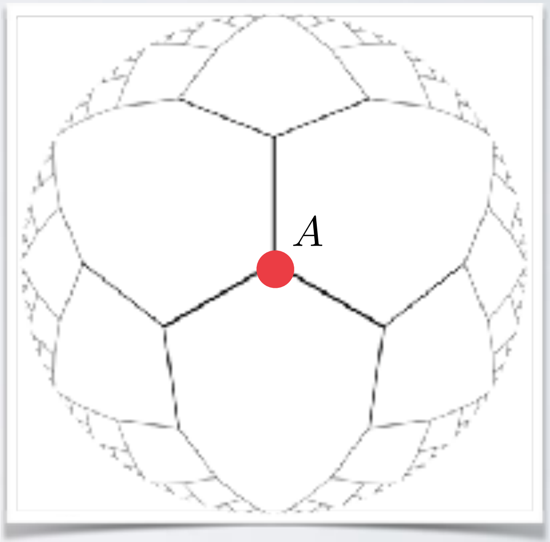
calculate the diameter



closeness centrality



modularity of a line graph $Q(r)$

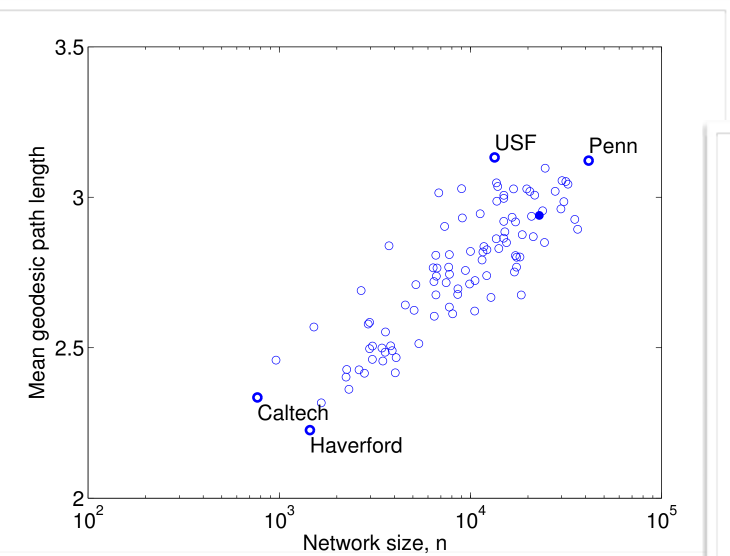


betweenness of A

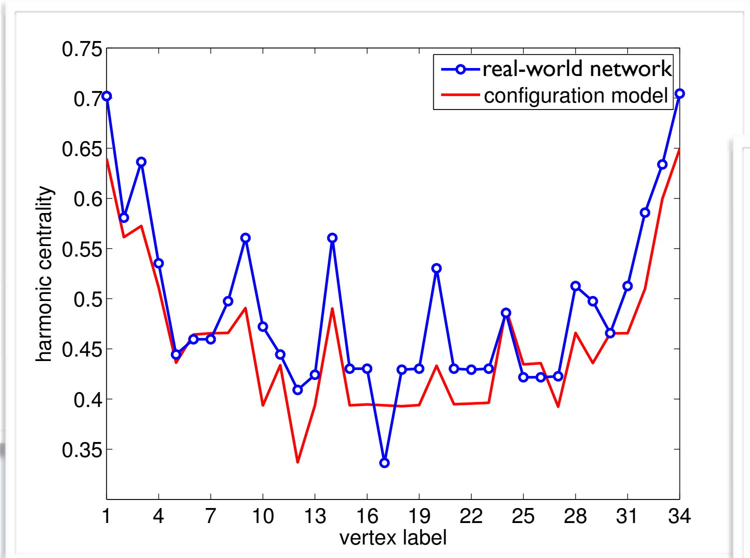
lessons learned from past instances

what works well:

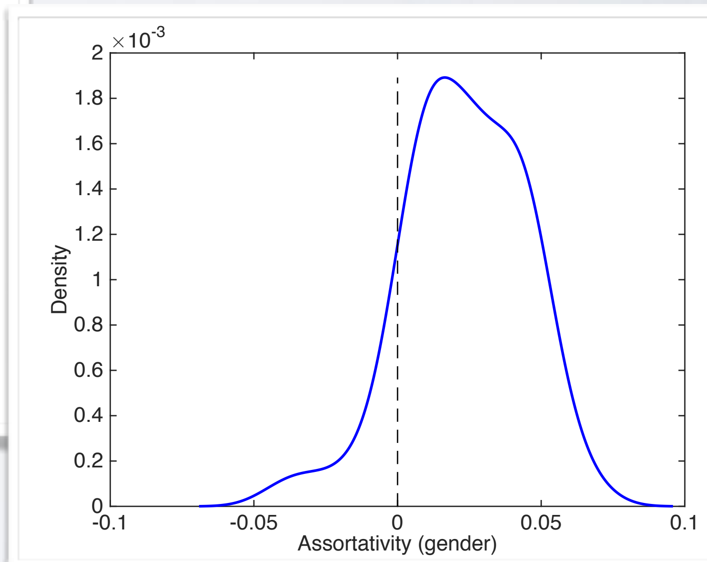
2. analyze real networks
   test understanding + practice with implementing methods



mean geodesics and $O(\log n)$



node centrality vs. configuration model
(when is a pattern interesting?)

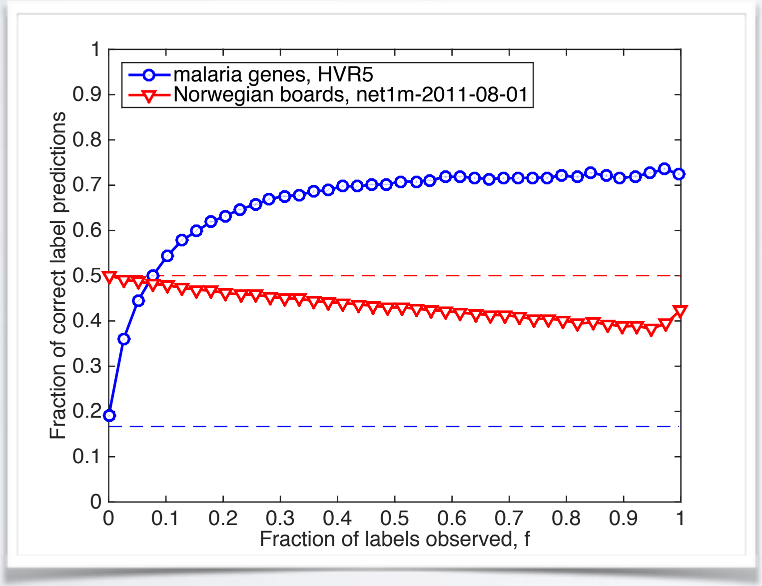


attribute assortativity

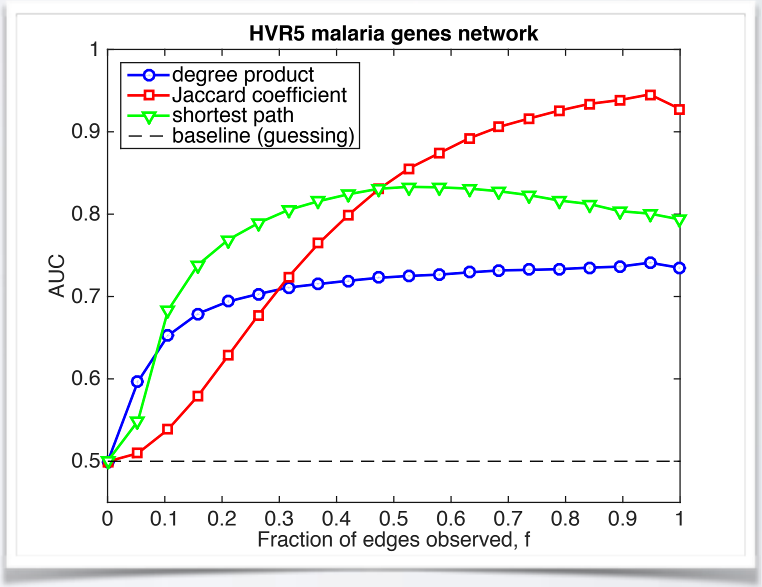
lessons learned from past instances

what works well:

3. simple prediction tasks
   test intuition + run numerical experiments



label prediction via homophily

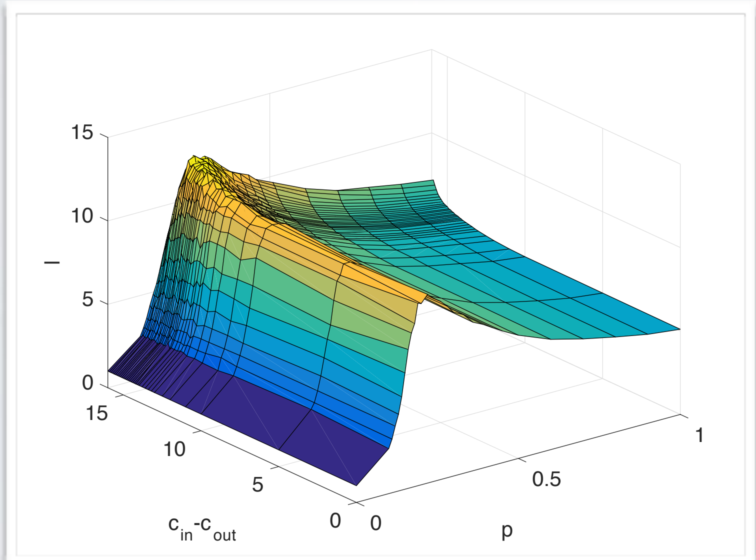


link prediction via heuristic

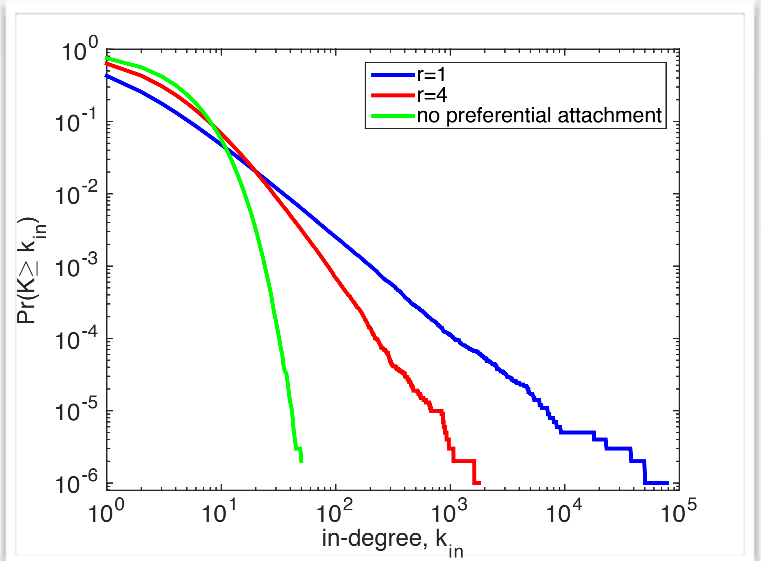
lessons learned from past instances

what works well:

4. simple simulations
   explore dynamics vs. structure + numerical experiments



simulate epidemics (SIR) on planted partitions



simulate Price's model

lessons learned from past instances

what works well:

5. team projects
   teamwork + exploring their own ideas

Identifying Politicians in the Panama Papers Network

CSCI 5352 Final Project
Irene Beckman, Santhanakrishnan Ramani, Ruhi Saraf

Modeling Information Diffusion in Faculty Hiring Networks*

Dimitrios Economou[†] and Allison Morgan[‡]
Department of Computer Science, University of Colorado, Boulder, CO 80303, USA
(Network Analysis & Modeling Final Project)

A Network Science Approach to Harmonic Analysis using annotated Beatles songs

Nicolas Metts
University of Colorado
CSCI 5532

Uncertainty Quantification of Voltage State in Power Systems

Samantha Molnar
CSCI5352 Final Project

Relational Topic Model for Congressional Bills Corpus

You Lu
Department of Computer Science
University of Colorado Boulder

Shudong Hao
Department of Computer Science
University of Colorado Boulder

key takeaways

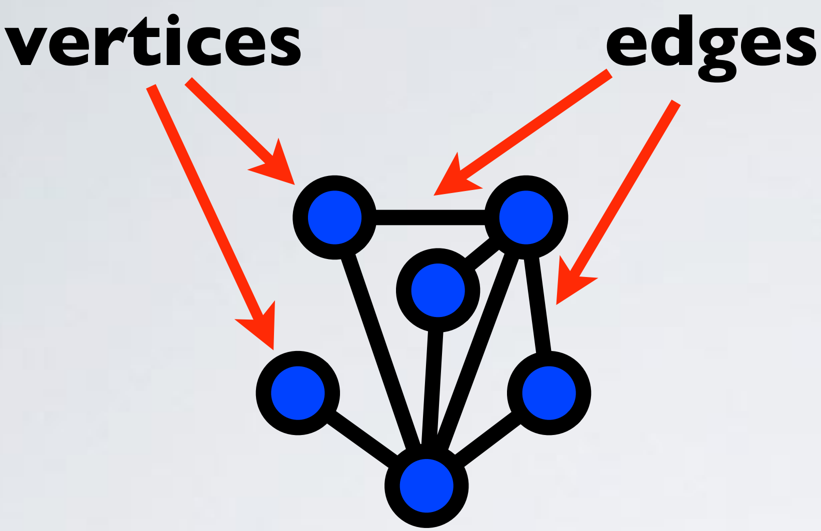
key takeaways

- **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!

1. defining a network
2. describing a network



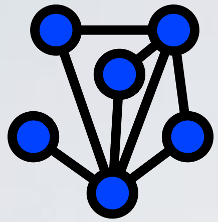
what is a vertex?

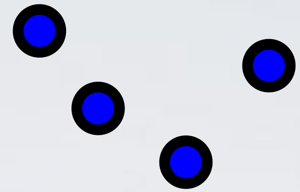
V distinct objects (vertices / nodes / actors)

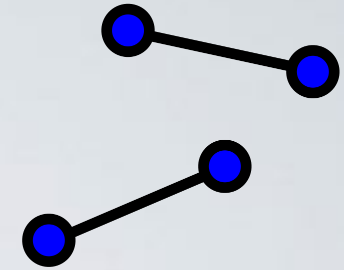
when are two vertices connected?

$E \subseteq V \times V$

pairwise relations (edges / links / ties)





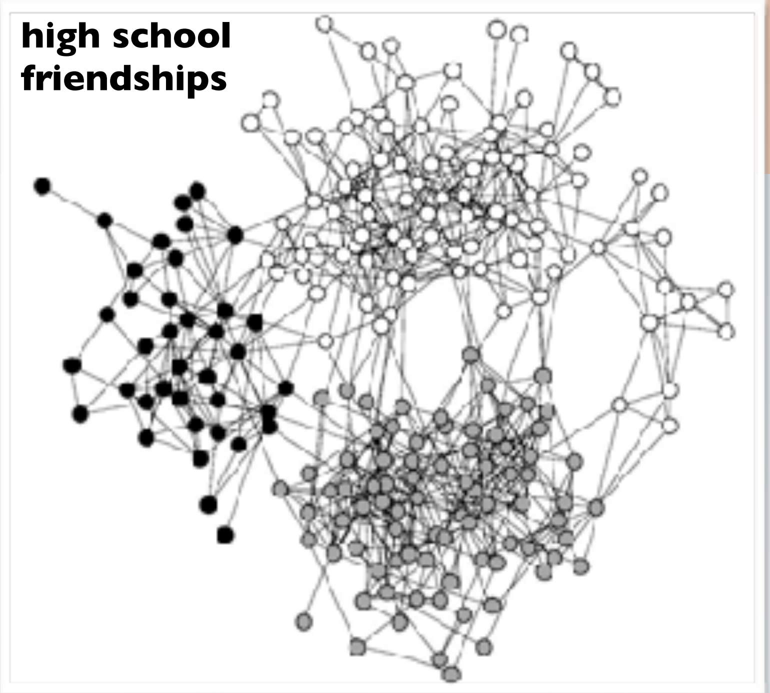


	network	vertex	edge
informationaltelecommunications	Internet(1)	computer	IP network adjacency
	Internet(2)	autonomous system (ISP)	BGP connection
	software	function	function call
	World Wide Web	web page	hyperlink
	documents	article, patent, or legal case	citation
transportation	power grid transmission	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

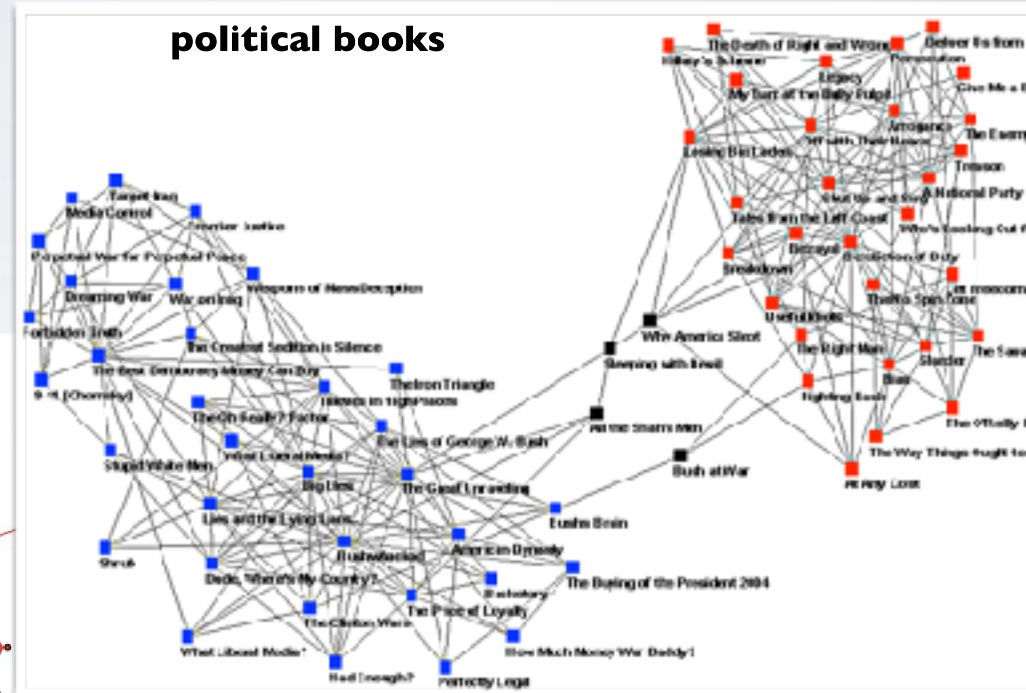
edge: friendship, collaborations, sexual contacts, communication, authority, exchange, etc.

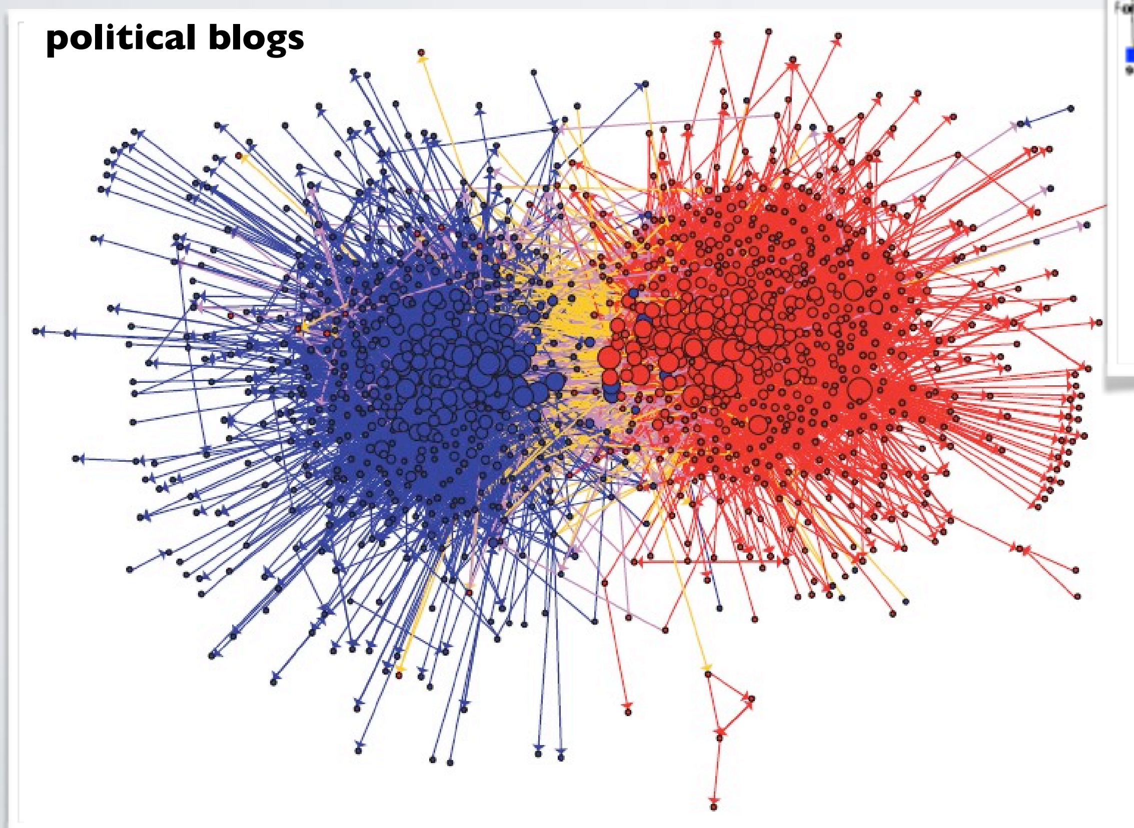


information networks

vertex: books, blogs, webpages, etc.

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

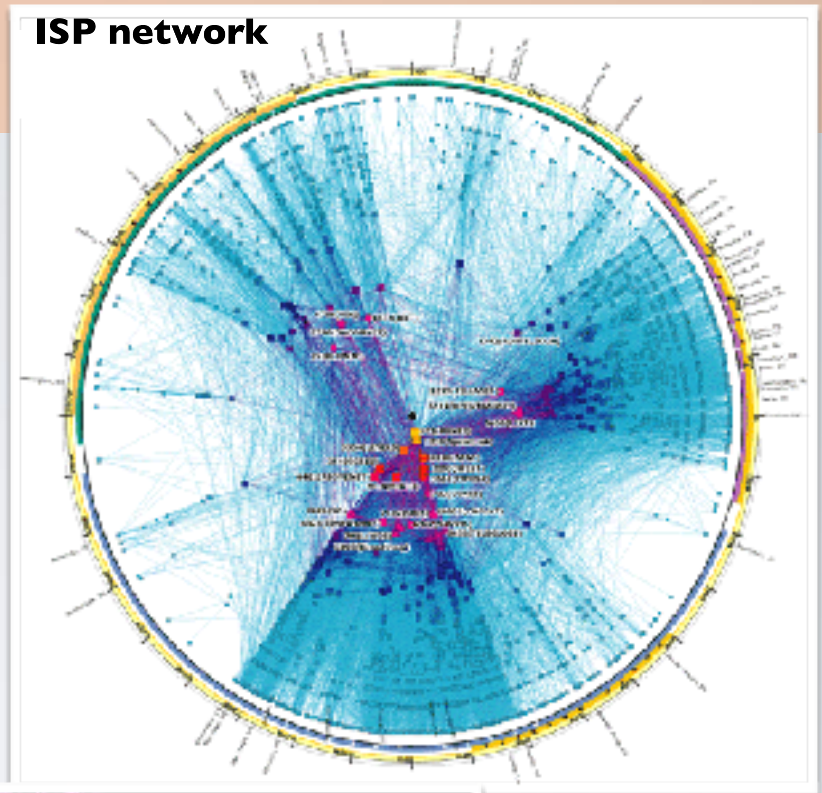


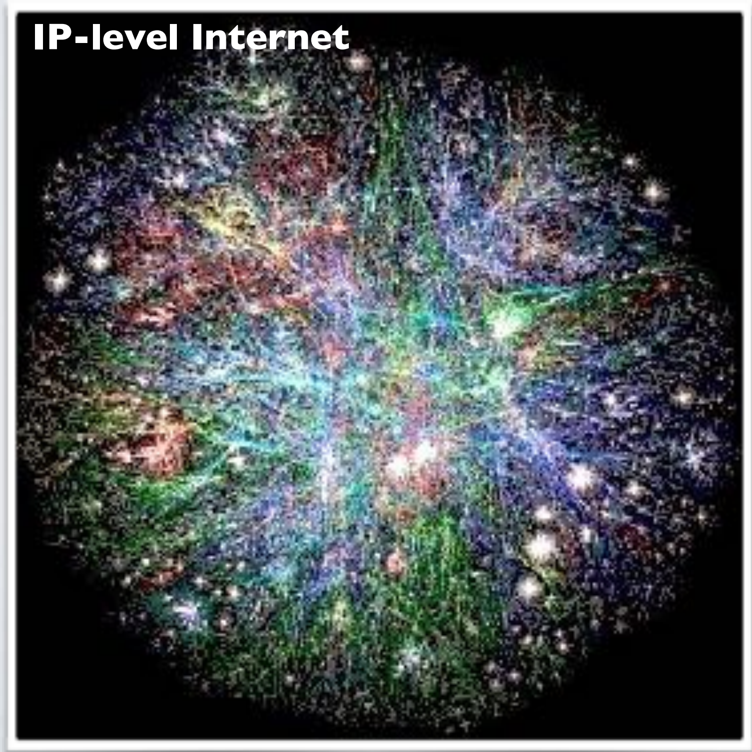


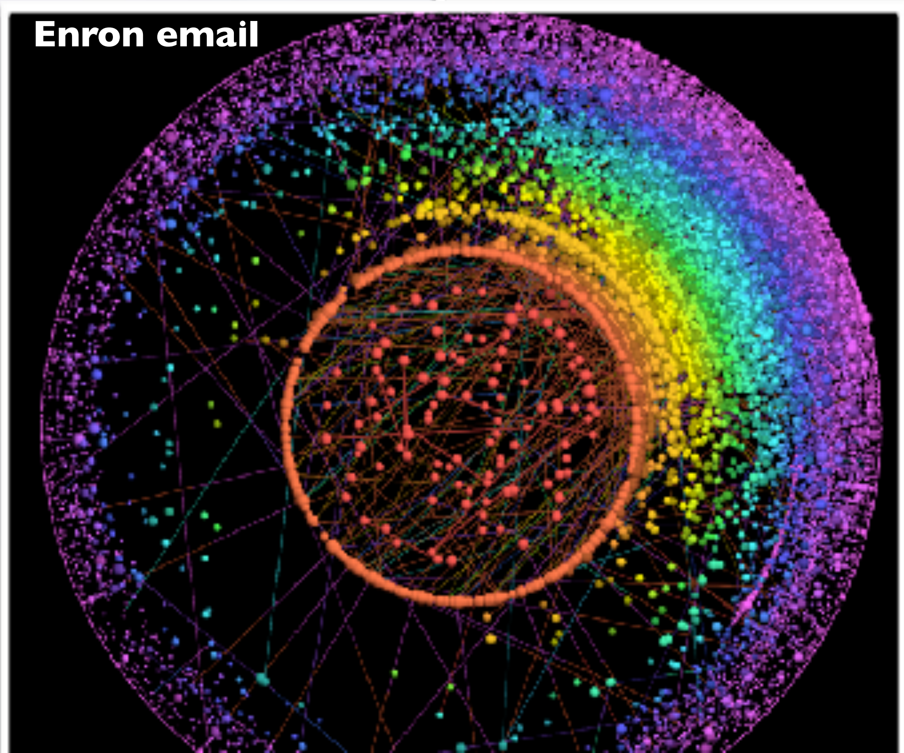
communication networks

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

edge: exchange of information





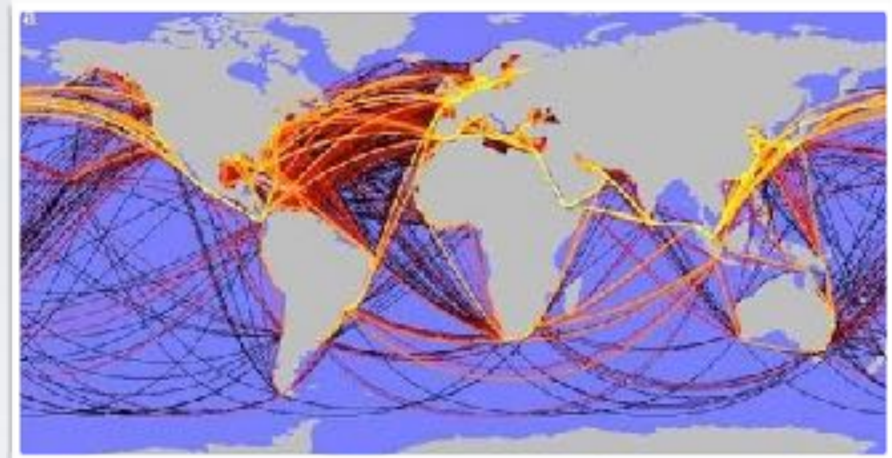


transportation networks

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

edge: physical transportation of material

global shipping



US Interstates

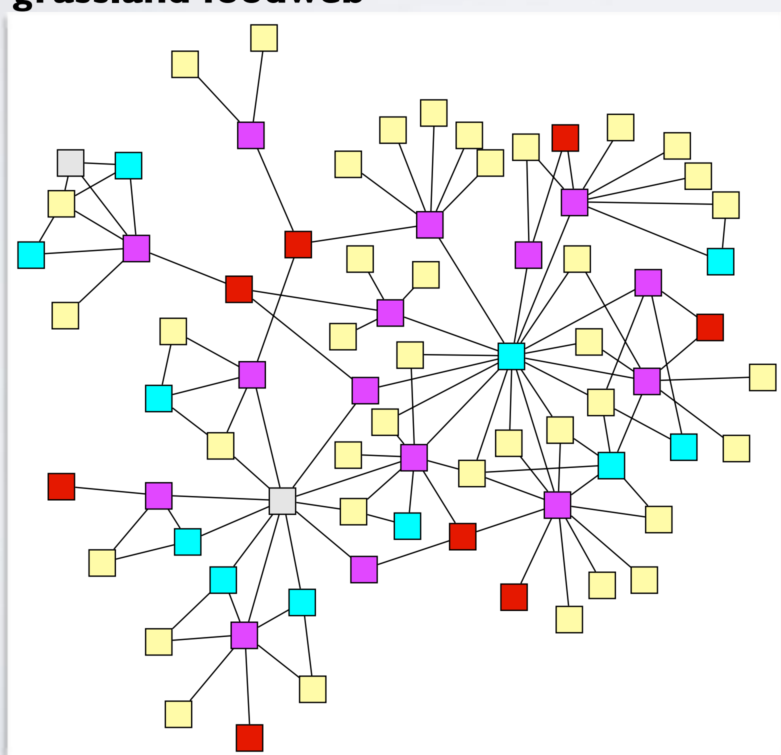
global air traffic

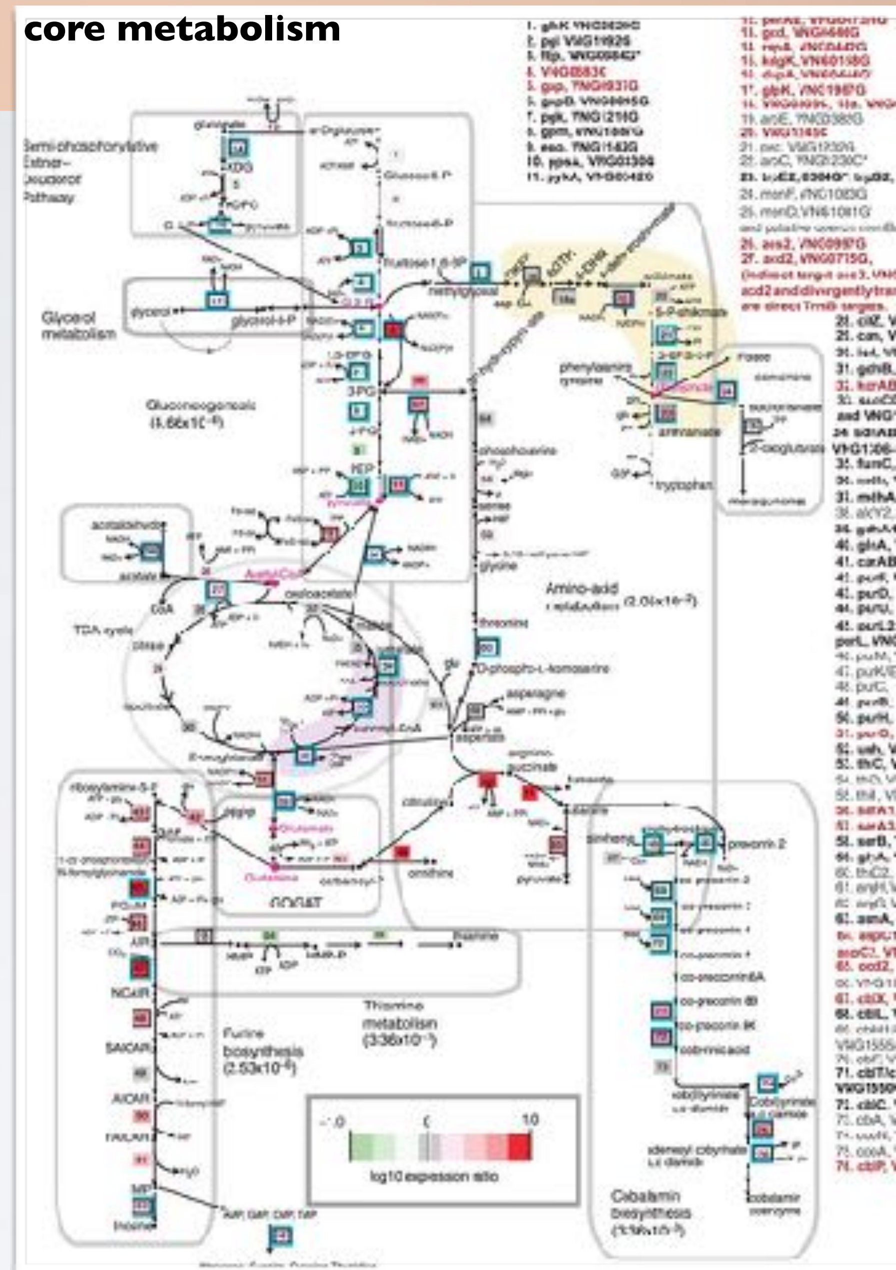
biological networks

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

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

grassland foodweb

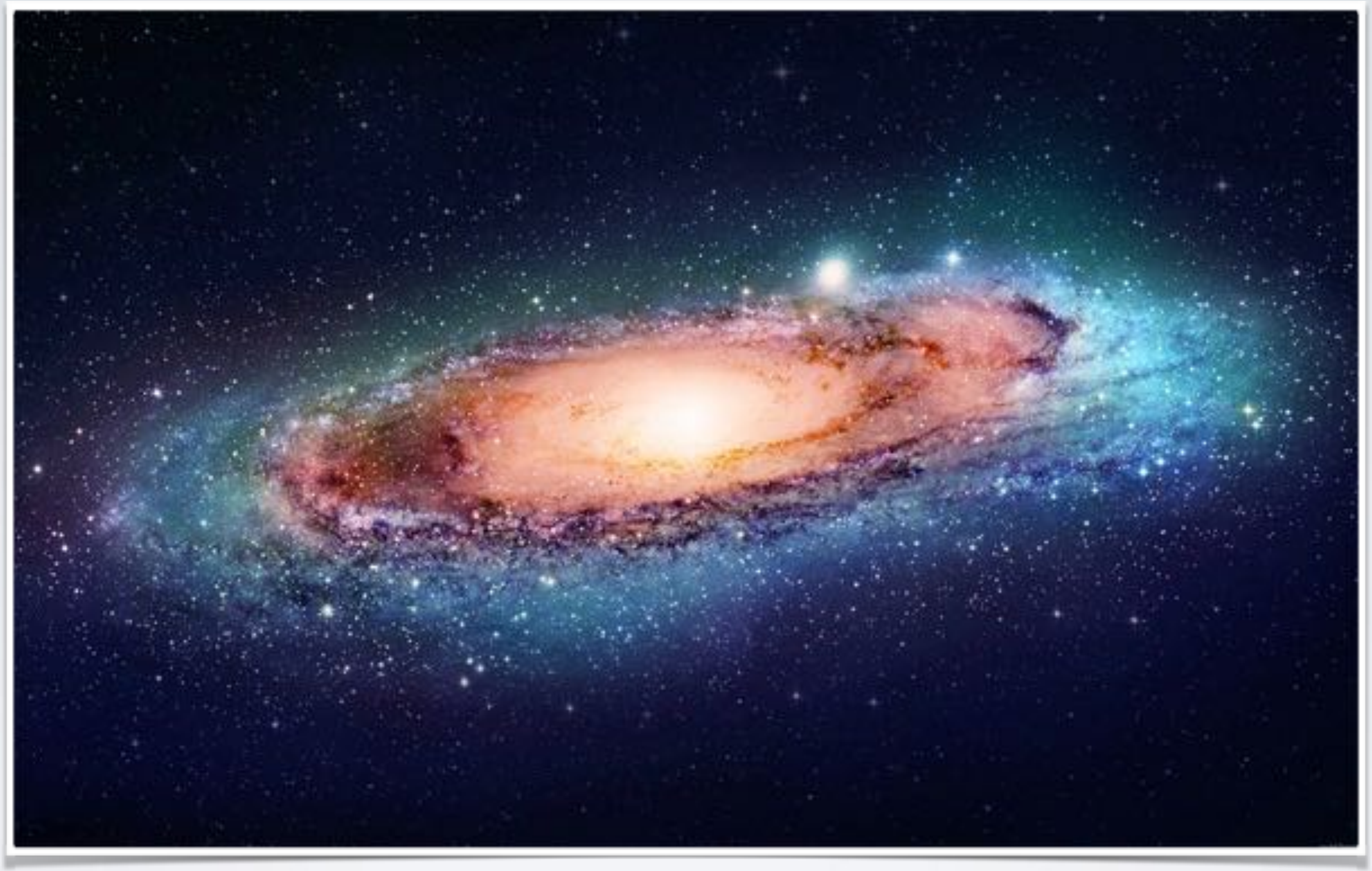




what’s a network?

pop quiz

what's a network?

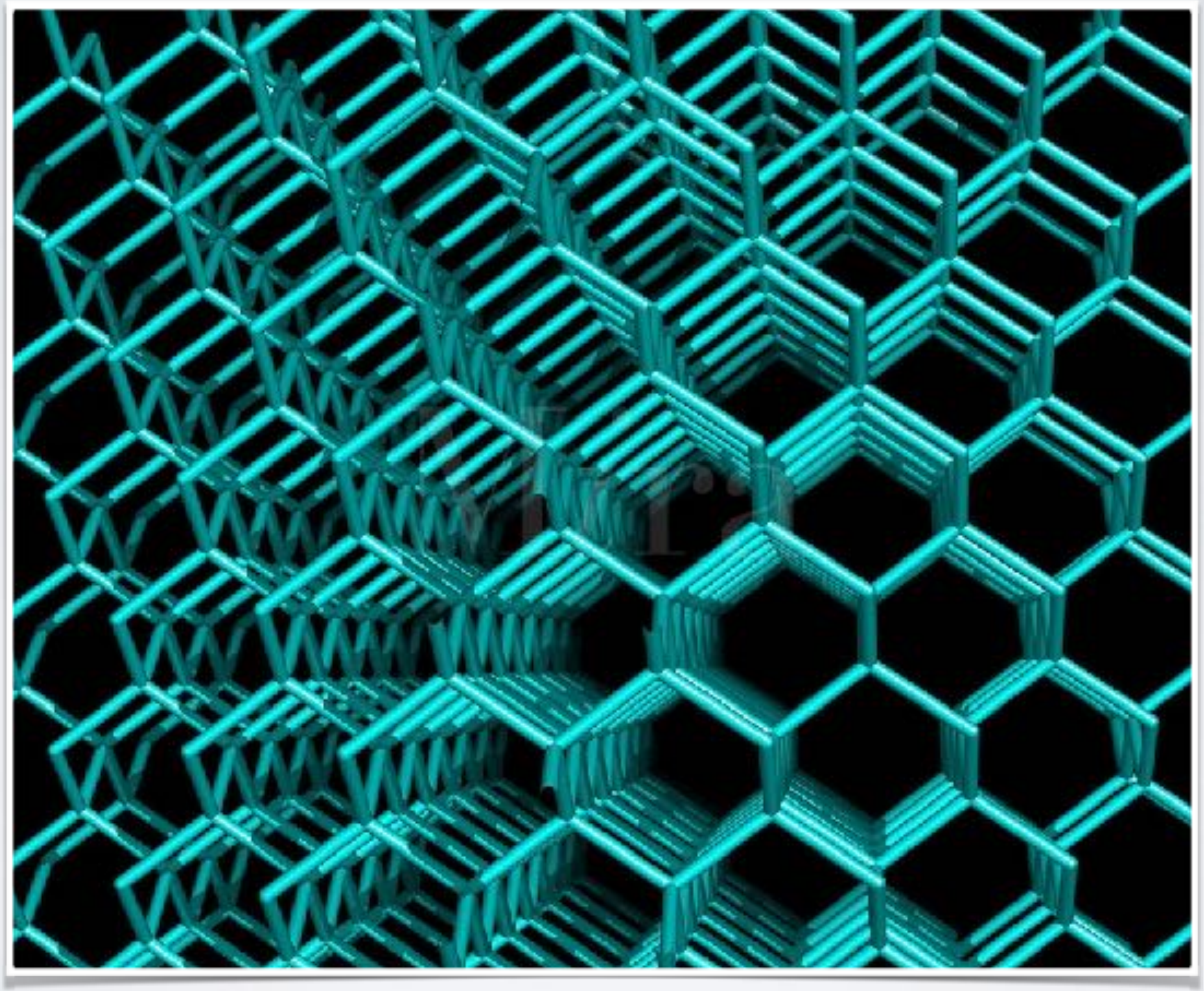


Andromeda galaxy

what's a network?

cauliflower fractal

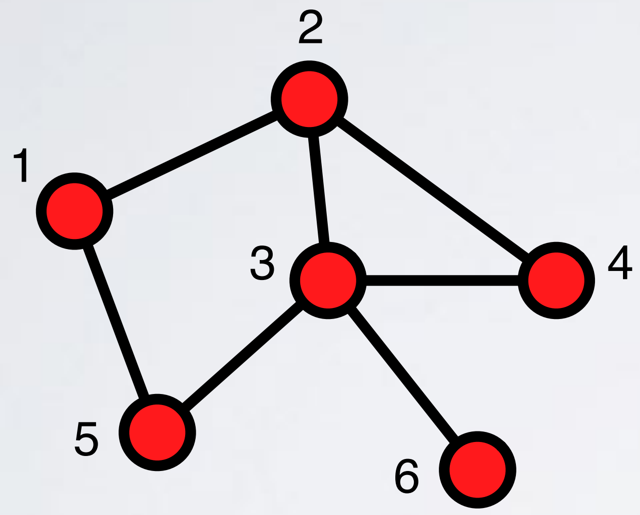
what's a network?



diamond lattice

representing networks

a *simple* network

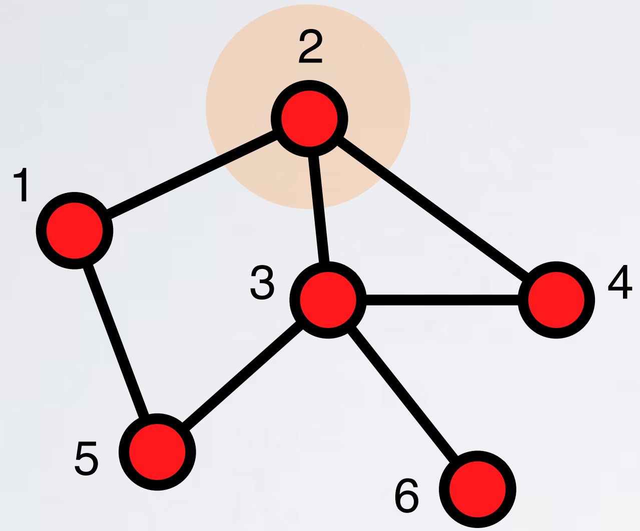


undirected

unweighted

no self-loops

a *simple* network



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

adjacency list

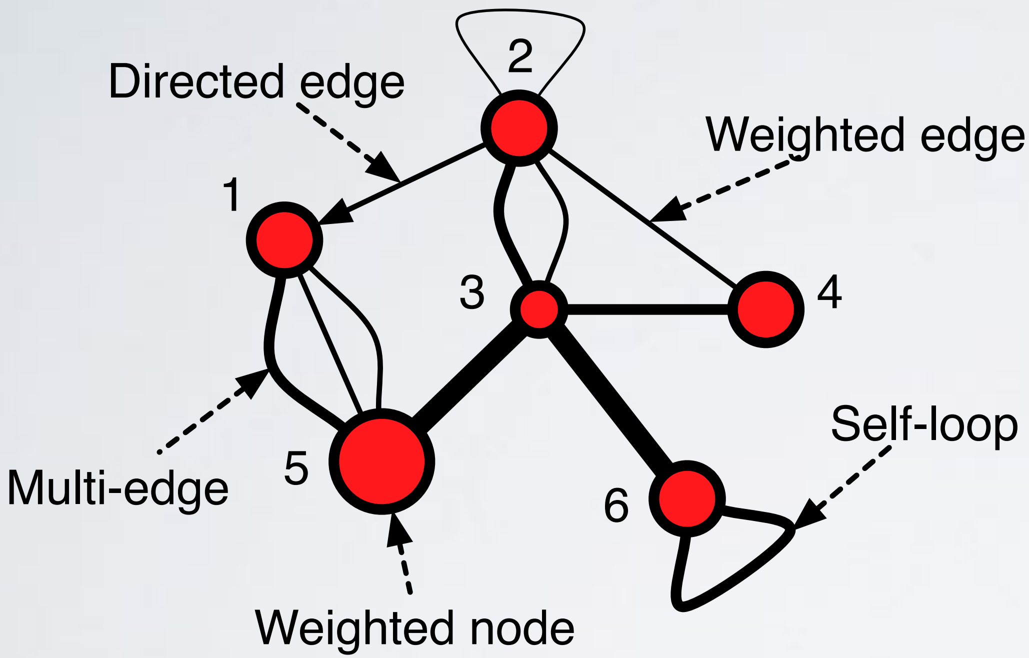
A
$1 \to \{2, 5\}$
$2 \to \{1, 3, 4\}$
$3 \to \{2, 4, 5, 6\}$
$4 \to \{2, 3\}$
$5 \to \{1, 3\}$
$6 \to \{3\}$

undirected

unweighted

no self-loops

a less simple network

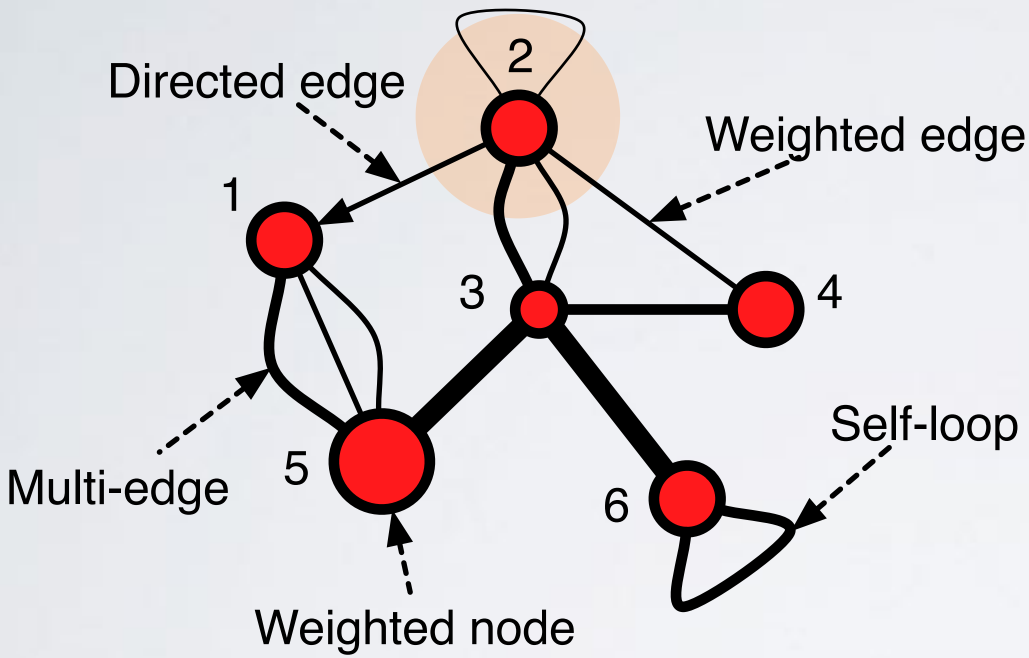


~~undirected~~

~~unweighted~~

~~no self loops~~

a less simple network



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

$$
\begin{array}{ll}
A & \\
\hline
1 & \to \{(5,1),(5,1),(5,2)\} \\
2 & \to \{(1,1),(2,\frac{1}{2}),(3,2),(3,1),(4,1)\} \\
3 & \to \{(2,2),(2,1),(4,2),(5,4),(6,4)\} \\
4 & \to \{(2,1),(3,2)\} \\
5 & \to \{(1,1),(1,1),(1,2),(3,4)\} \\
6 & \to \{(3,4),(6,2)\}
\end{array}
$$

directed networks

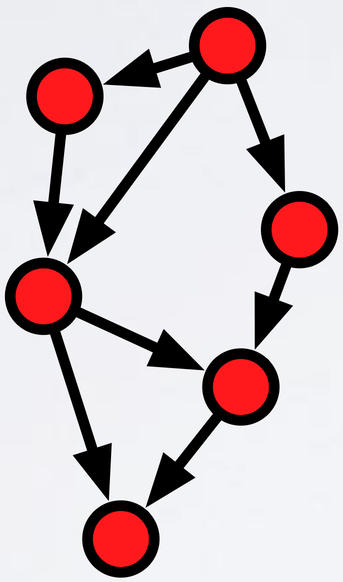
$$A_{ij} \neq A_{ji}$$

citation networks

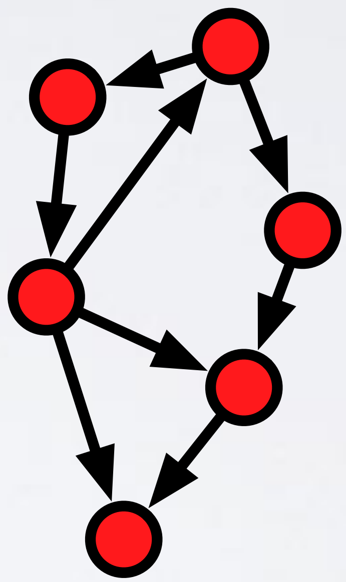
foodwebs*

epidemiological

others?



directed acyclic graph



directed graph

WWW

friendship?

flows of goods, information

economic exchange

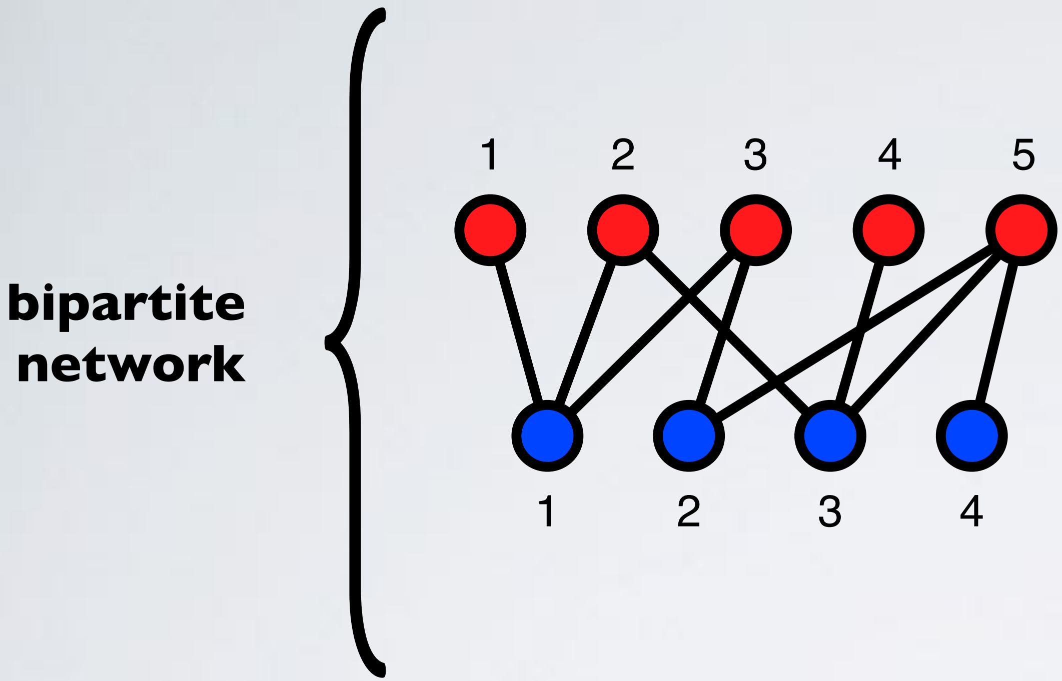
dominance

neuronal

transcription

time travelers

bipartite networks

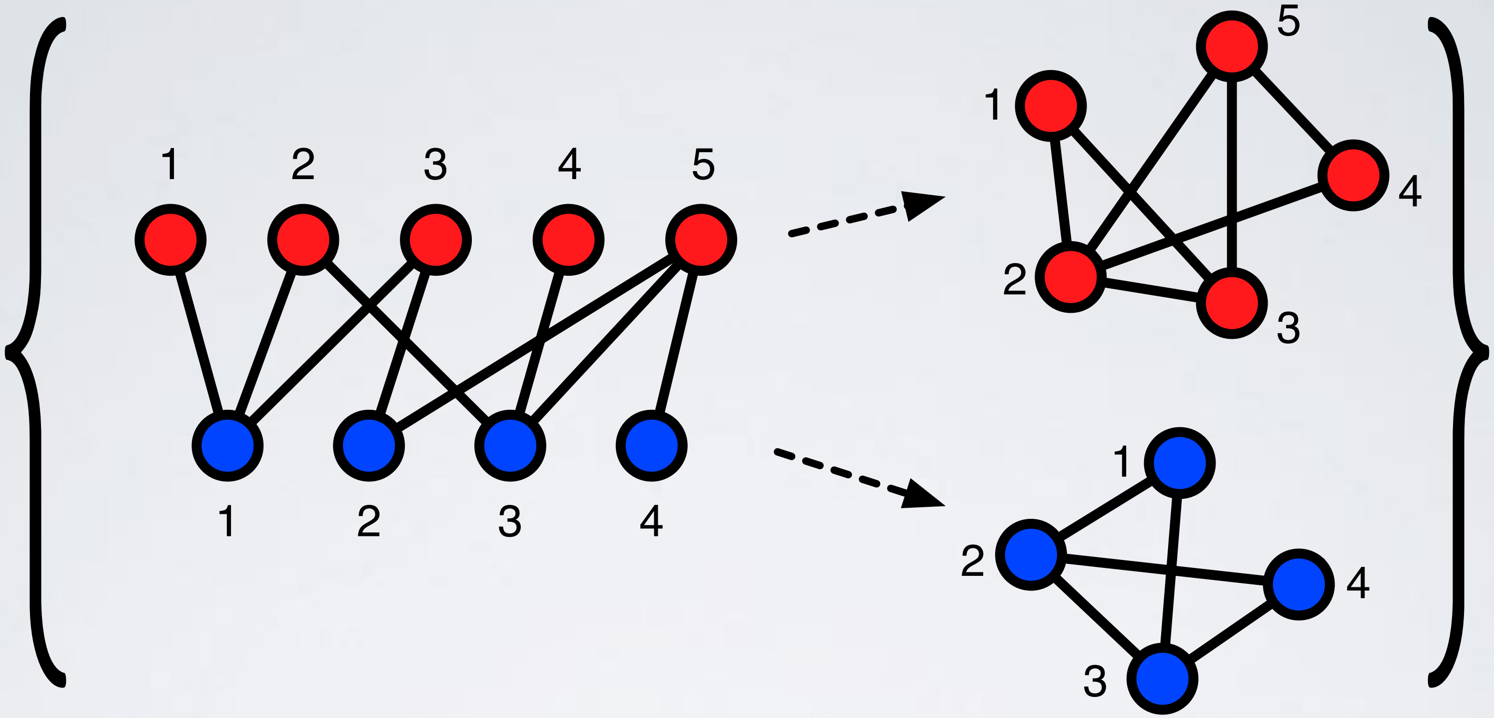


no within-type edges

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

bipartite networks

bipartite network



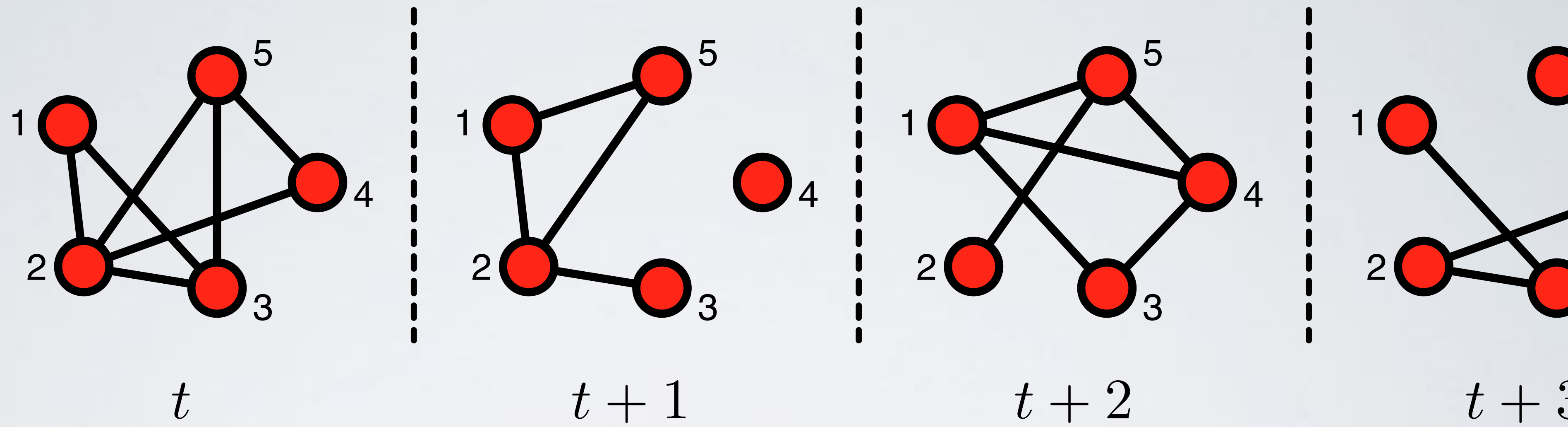
one-mode projections

no within-type edges

one type only

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

temporal networks



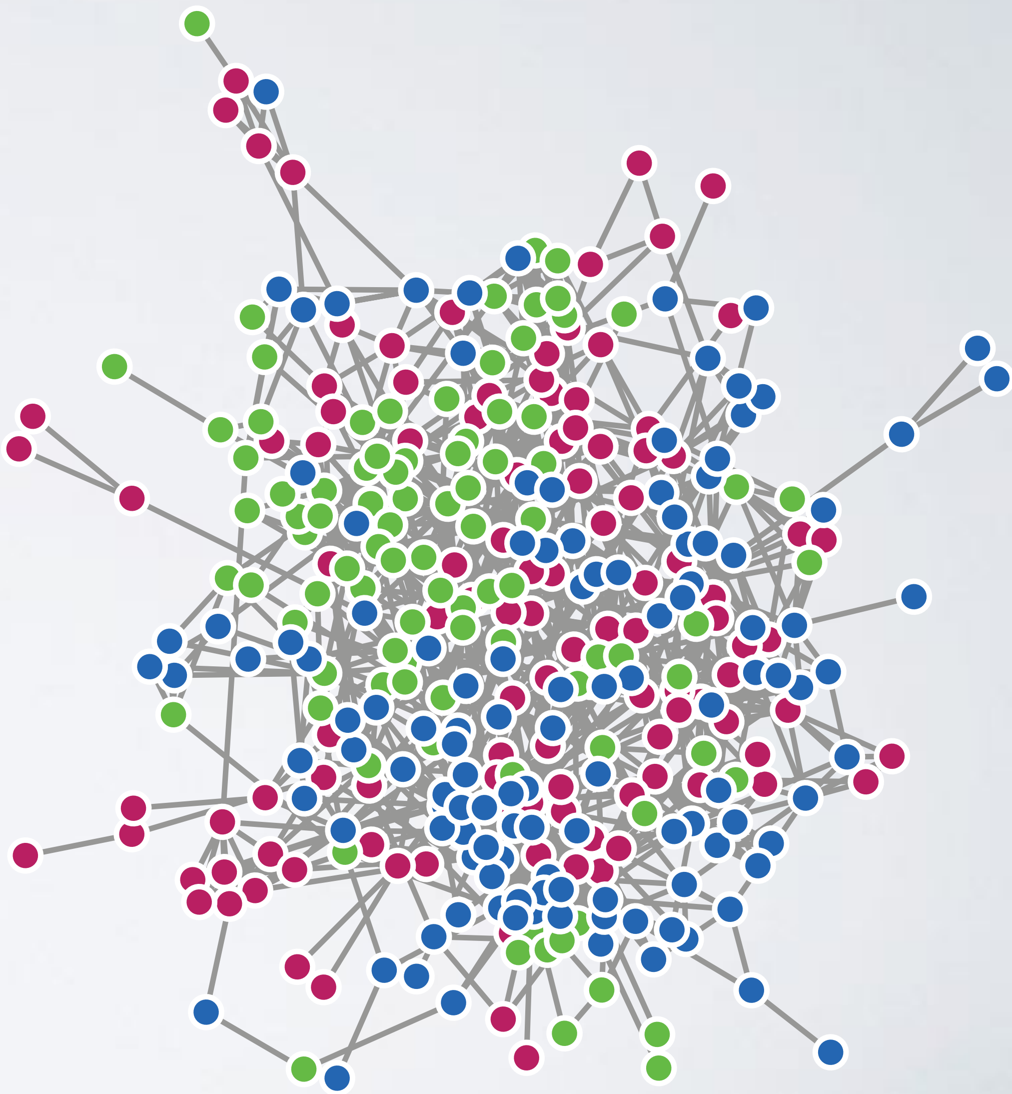
any network over time

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

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

describing networks

what networks look like



describing networks

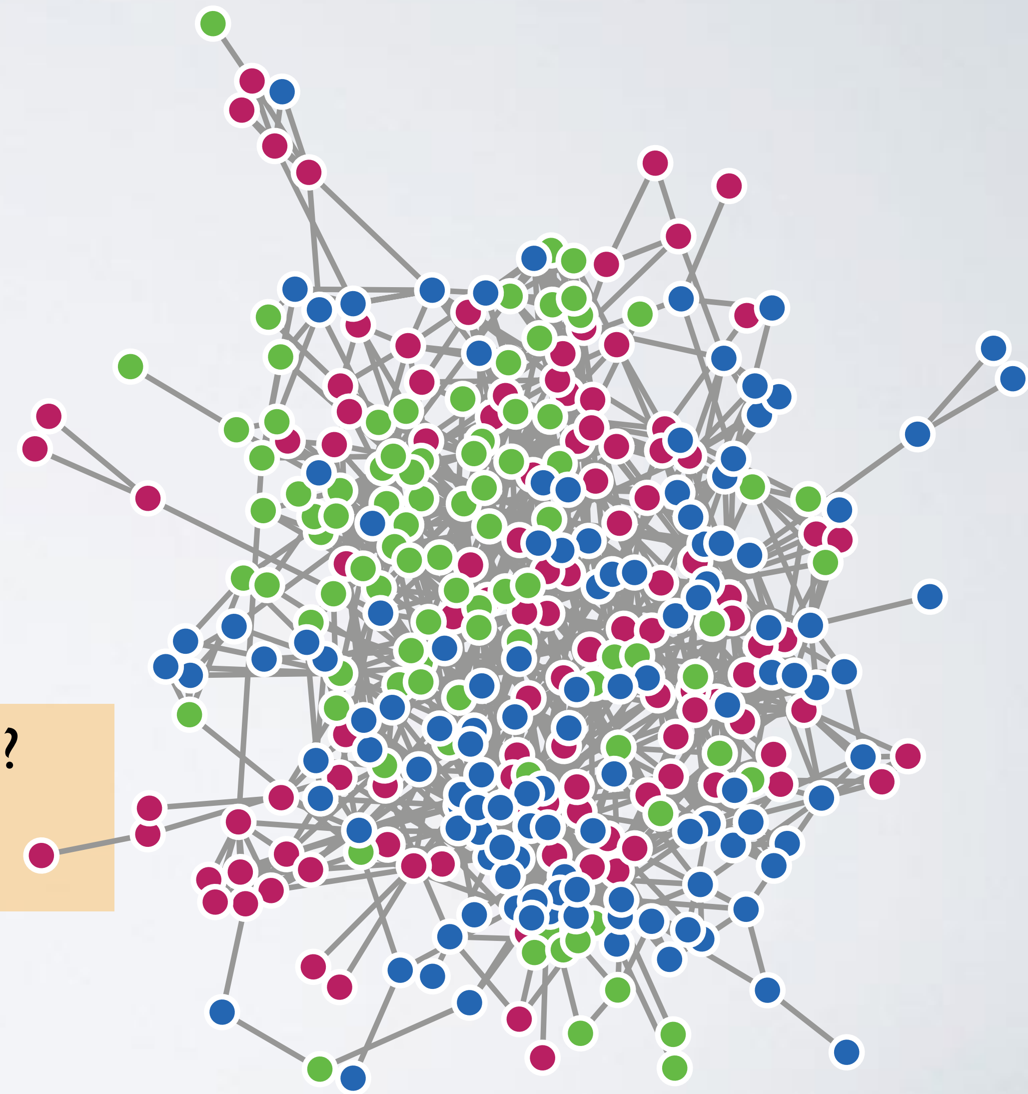
what networks look like

questions:

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

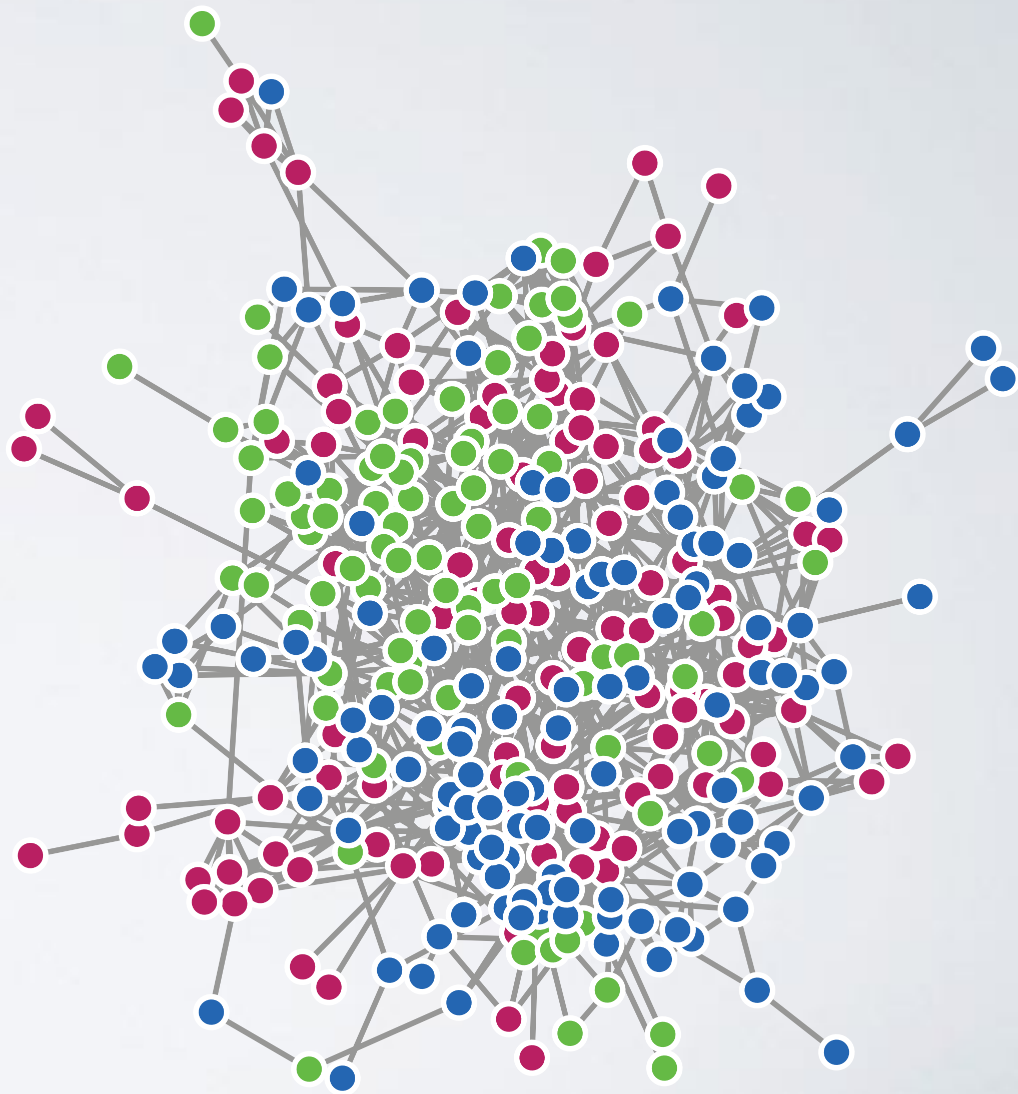
what we want to know

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



describing networks

a first step : **describe its features**

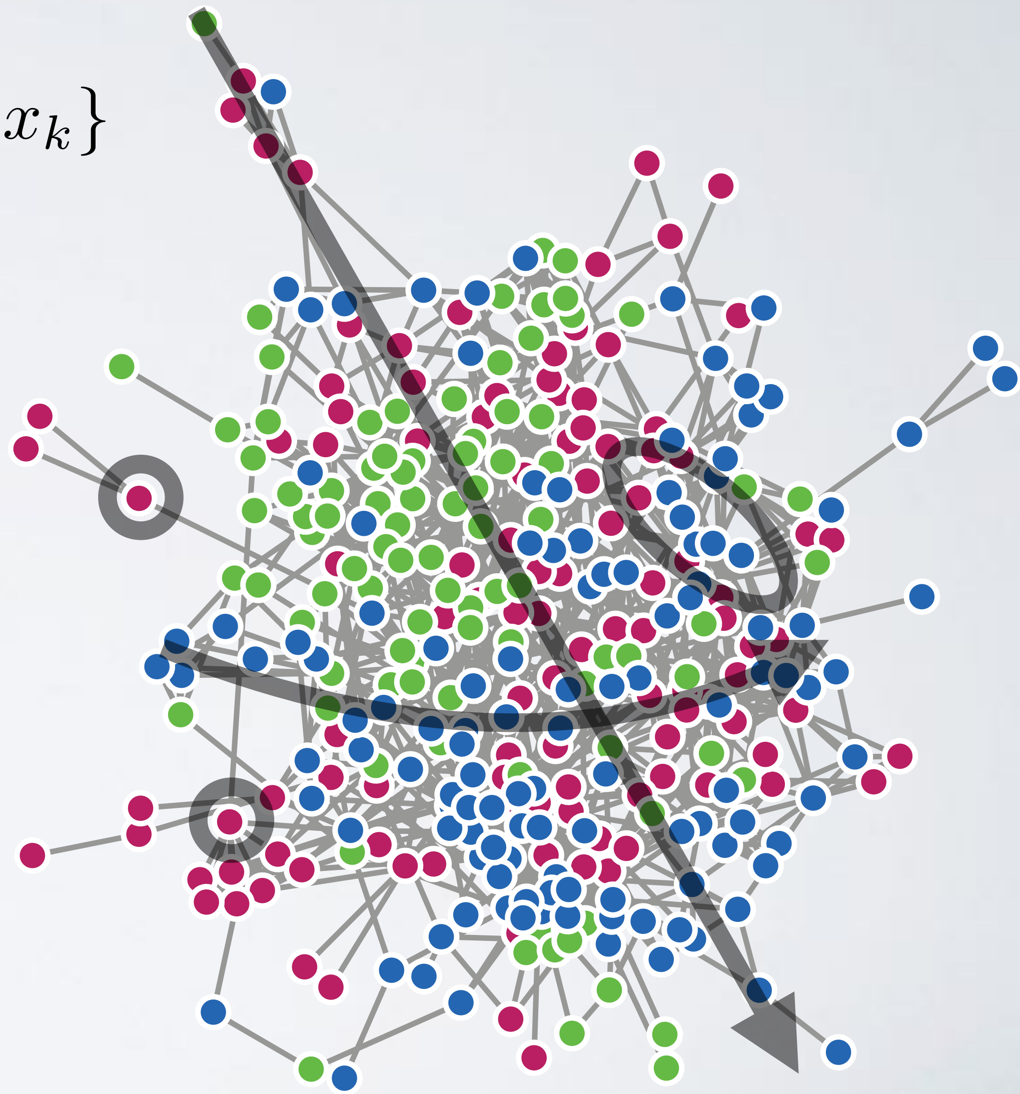


describing networks

a first step : **describe its features**

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

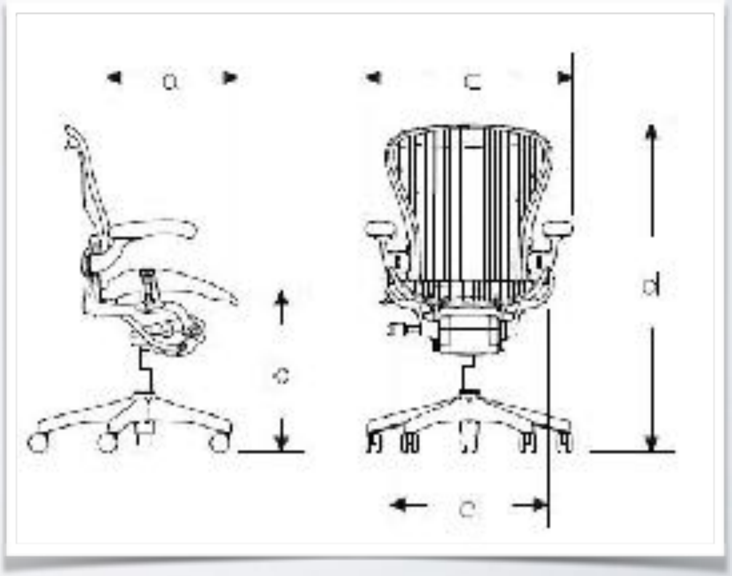
- degree distributions
- short-loop density (triangles, etc.)
- shortest paths (diameter, etc.)
- vertex positions
- correlations between these



describing networks

a first step : **describe its features**

$$f : \text{object} \rightarrow \{x_1, \ldots, x_k\}$$



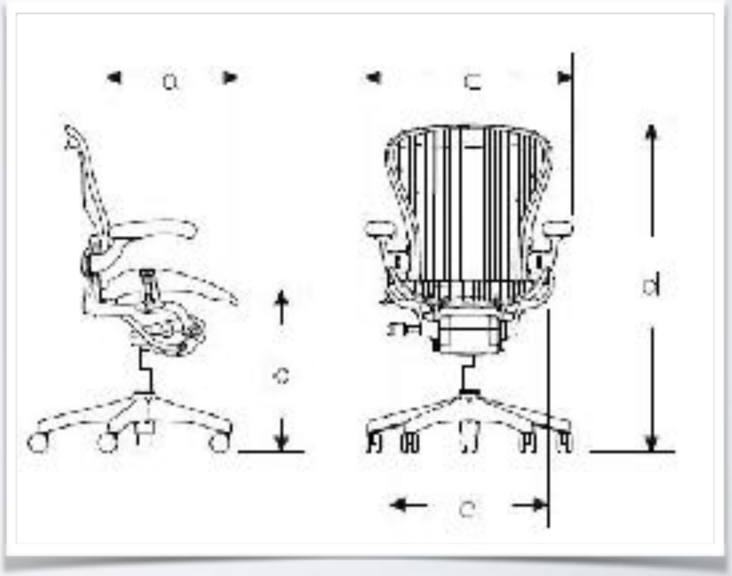
describing networks

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…

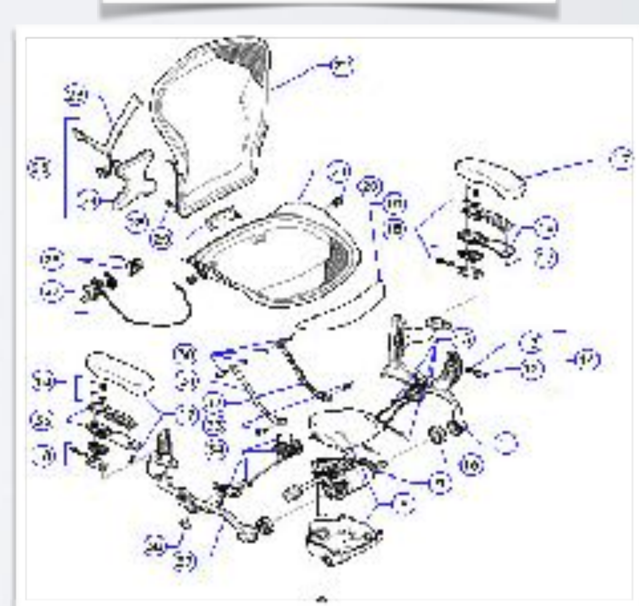


describing networks

what we want : **understand its structure**

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

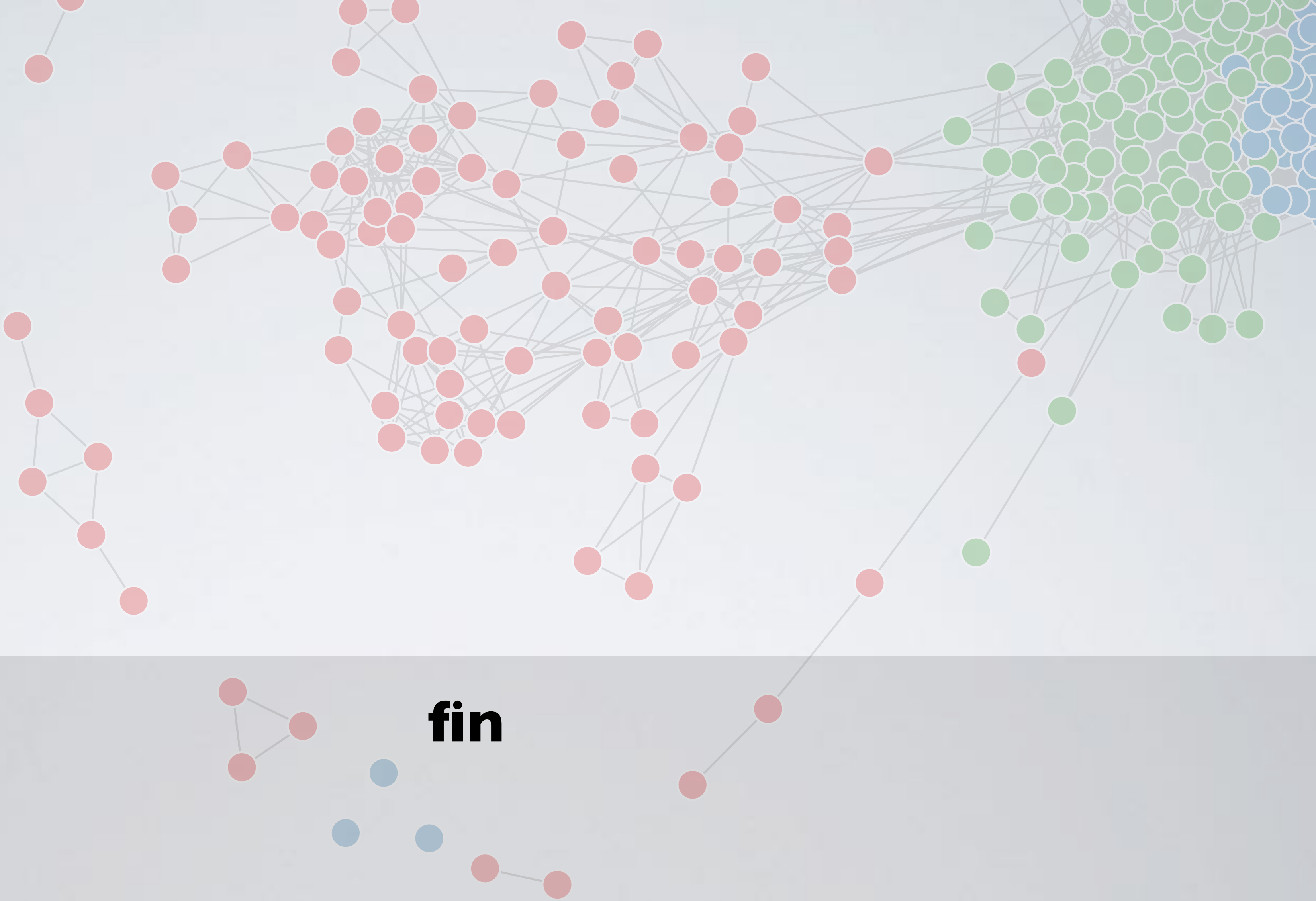
- what are the fundamental parts?
- how are these parts organized?
- where are the degrees of freedom $\vec{\theta}$?
- how can we define an abstract class?
- structure — dynamics — function?



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

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

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



fin