

Network Analysis & Modeling

lecture 1a: what are networks and how do we talk about them?

Prof. Aaron Clauset
Computer Science & BioFrontiers Institute
@aaronclauset
aaron.clauset@colorado.edu

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

Computer Scientists

Applied Mathematicians

Statisticians

Biologists

Ecologists

Sociologists

Political 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?

an approach.

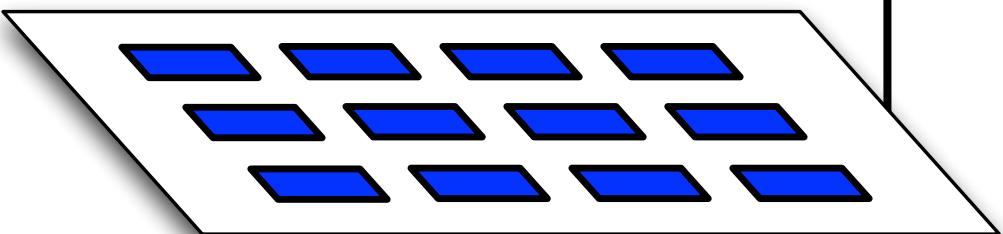
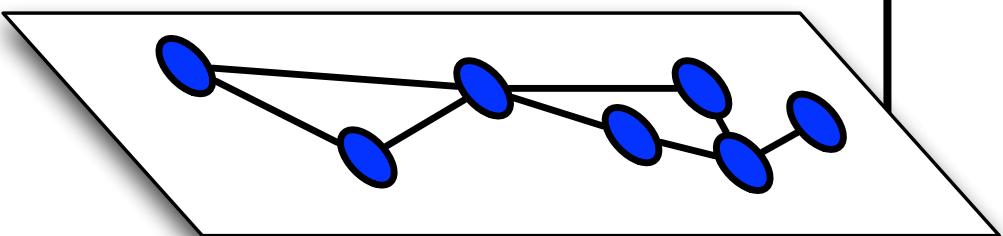
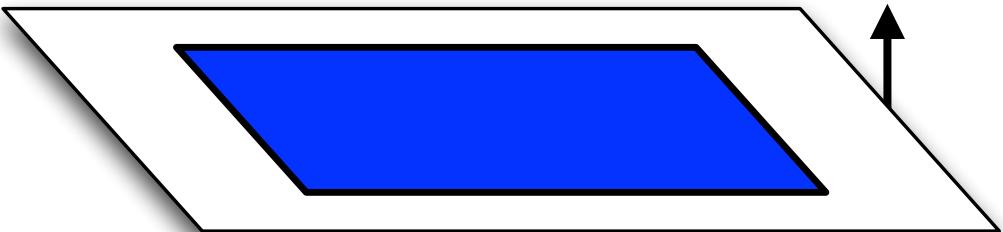
a mathematical representation
provide structure to complexity.

structure that exists *above*
individuals / components

or: structure that exists *below*
system / population



system / population



individuals / components

tools and resources

Software

[R](#)
[Python](#)
[Matlab](#)

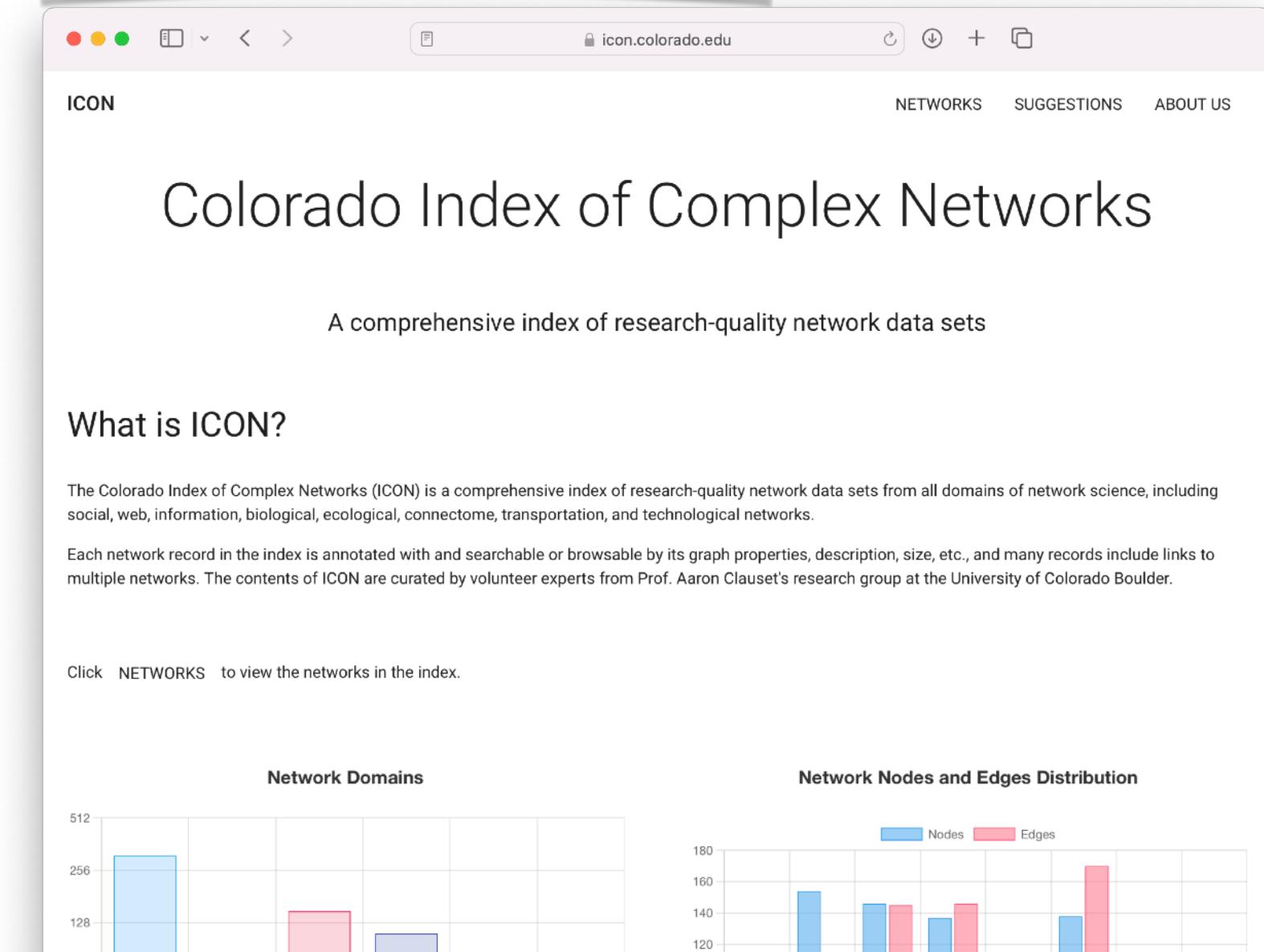
- ★ [NetworkX \[python\]](#)
- ★ [igraph \[python, R, c++\]](#)
- [graph-tool \[python, c++\]](#)
- [graspologic \[python\]](#)

Standalone editors

[UCI-Net](#)
[NodeXL](#)
[Gephi](#)
[Pajek](#)
[Network Workbench](#)
[Cytoscape](#)
[yEd graph editor](#)
[Graphviz](#)

Network data sets

- ★ [Colorado Index of Complex Networks](#)



learning goals

1. develop a **network intuition** for reasoning about network phenomena
2. understand **network representations, terminology, and concepts**
3. learn **descriptive network analysis** and **its interpretation**
4. learn algorithms to **predict missing network information**
5. understand how to **conduct and interpret** numerical network experiments, to **explore and test hypotheses** about networks
6. analyze and model **real-world network data**, using math and computation

course format

- course meets in-person in JSCBB E125 & on Zoom
- lectures 2 times a week, some guest lectures and some class discussions
- deliverables:
 - 5 biweekly problem sets — practice with concepts and tools
 - 1 class project: 1) proposal, 2) presentation, 3) final report — explore *your* ideas
- all content via class Canvas (lecture notes, recordings, problem sets, submissions)
- **see syllabus** for all course policies
- this semester — *I'm revising the content and writing a textbook*
- things will be rough in some places, but please bear with me

course schedule (rough)

week by week

- ◆ ● 1. what are networks & why are they special?
- ◆ ● 2. descriptive network analysis
- ◆ ● ★ ■ 3. ranking in networks
- ◆ ● 4. network intuition & random graphs
- ★ ■ 5. network null models
- ◆ ● ★ ■ 6. predicting missing data, node attributes
- ★ ■ 7. predicting missing data, links
- ◆ ● 8. community structure and mixing patterns
- ★ ■ 9. community structure models
- ◆ ● 10. network dynamics, spreading processes and cascades
- ★ 11. network dynamics, with structure
- ★ ■ 12. advanced topics
- ★ ■ 13. advanced topics
- 14. student project presentations

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

lessons learned

what's difficult

1. students need to do many different things:

some probability

some mathematics

some statistics

some machine learning

some programming

random graphs, distributions, expectations

counting things, some physics-style calculations

summary statistics, correlations, distributions

prediction, likelihoods, features, estimation algorithms

data wrangling, coding up measures and algorithms

2. best results come from forming mixed-background study groups

lessons learned

what works well

1. solve simple math problems – build intuition & practice with concepts

lessons learned

what works well

1. solve simple math problems – build intuition & practice with concepts
2. analyze *real* data – improve understanding by getting messy

lessons learned

what works well

1. solve simple math problems – build intuition & practice with concepts
2. analyze *real* data – improve understanding by getting messy
3. simple prediction tasks – test intuition & build practical skills

lessons learned

what works well

1. solve simple math problems – build intuition & practice with concepts
2. analyze *real* data – improve understanding by getting messy
3. simple prediction tasks – test intuition & build practical skills
4. learn to simulate – explore using numerical experiments

lessons learned

what works well

1. solve simple math problems – build intuition & practice with concepts
2. analyze *real* data – improve understanding by getting messy
3. simple prediction tasks – test intuition & build practical skills
4. learn to simulate – explore using numerical experiments
5. develop advanced tools – reach past intuition to complexity

lessons learned

what works well

1. solve simple math problems – build intuition & practice with concepts
2. analyze *real* data – improve understanding by getting messy
3. simple prediction tasks – test intuition & build practical skills
4. learn to simulate – explore using numerical experiments
5. develop advanced tools – reach past intuition to complexity
6. team projects! – explore your own ideas

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,‡}

¹Department of Computer Science, University of Colorado, Boulder, CO, USA

EPJ Data
a SpringerOpen

<https://doi.org/10.1140/epjds/s13688-018-0166-4>



REGULAR ARTICLE

Identifying Politicians in the Panama Papers Network

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

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

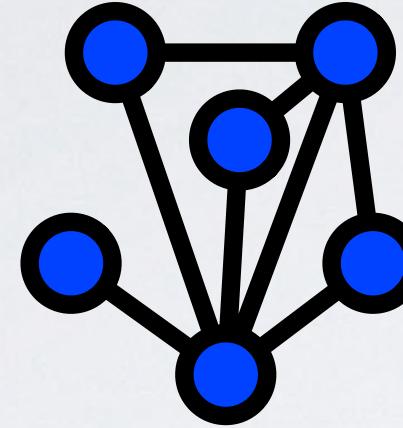
K. Hunter Wapman¹ and Daniel B. Larremore^{1, 2}

¹Department of Computer Science, University of Colorado, Boulder, CO, USA

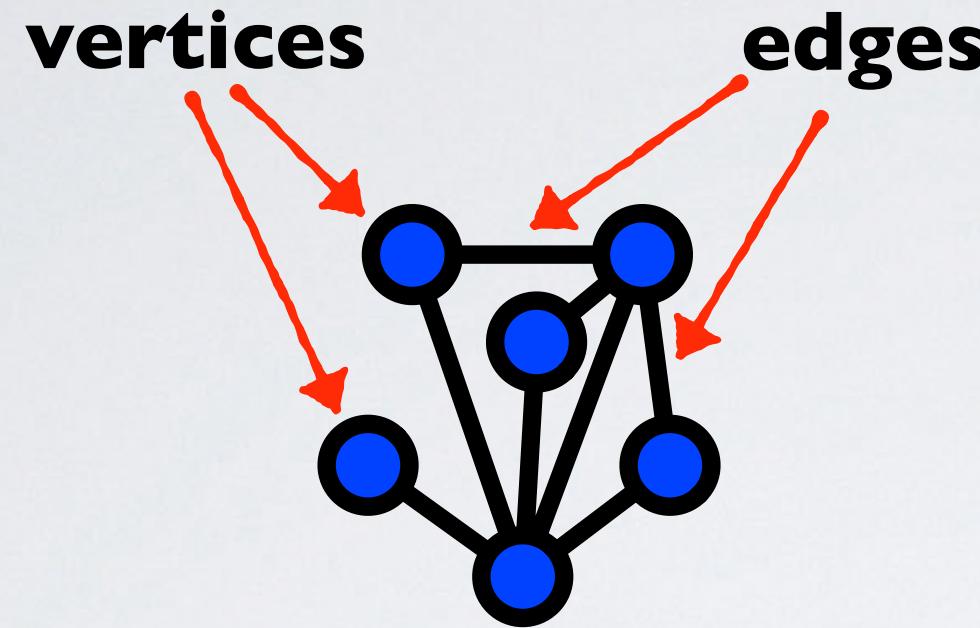
Prestige drives epistemic inequality in the diffusion of scientific ideas

Allison C. Morgan^{1,*}, Dimitrios J. Economou¹, Samuel F. Way¹, and Aaron Clauset^{1,2,3}

questions?



**the two most fundamental questions
in network science**



what is a vertex?

V distinct objects (vertices / nodes / actors)

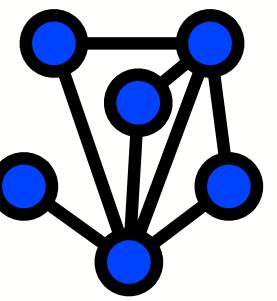
when are two vertices connected?

$$E \subseteq V \times V$$

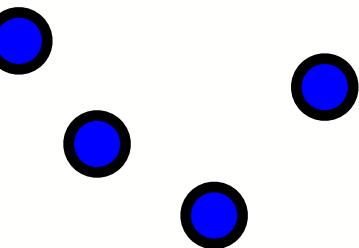
pairwise relations (edges / links / ties)

6 major classes of networks

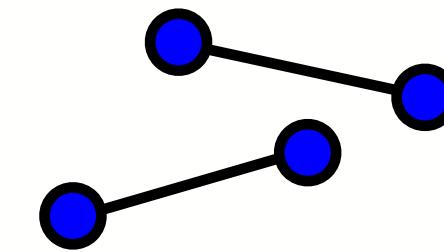
- technological
- information
- transportation
- social
- biological
- economic



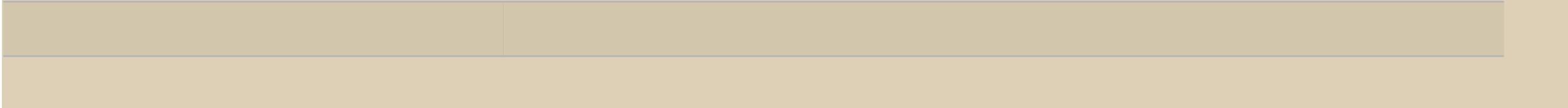
network



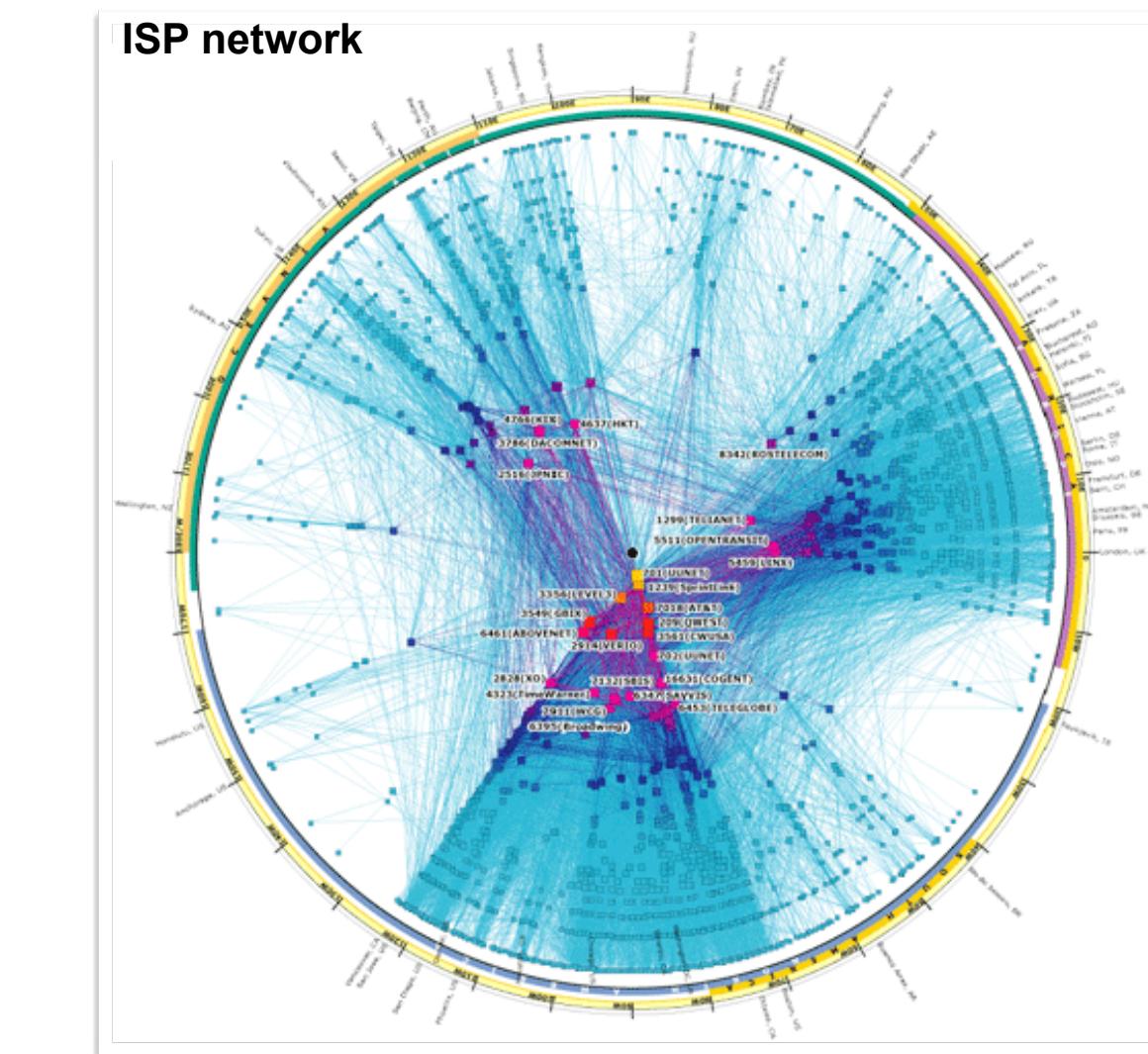
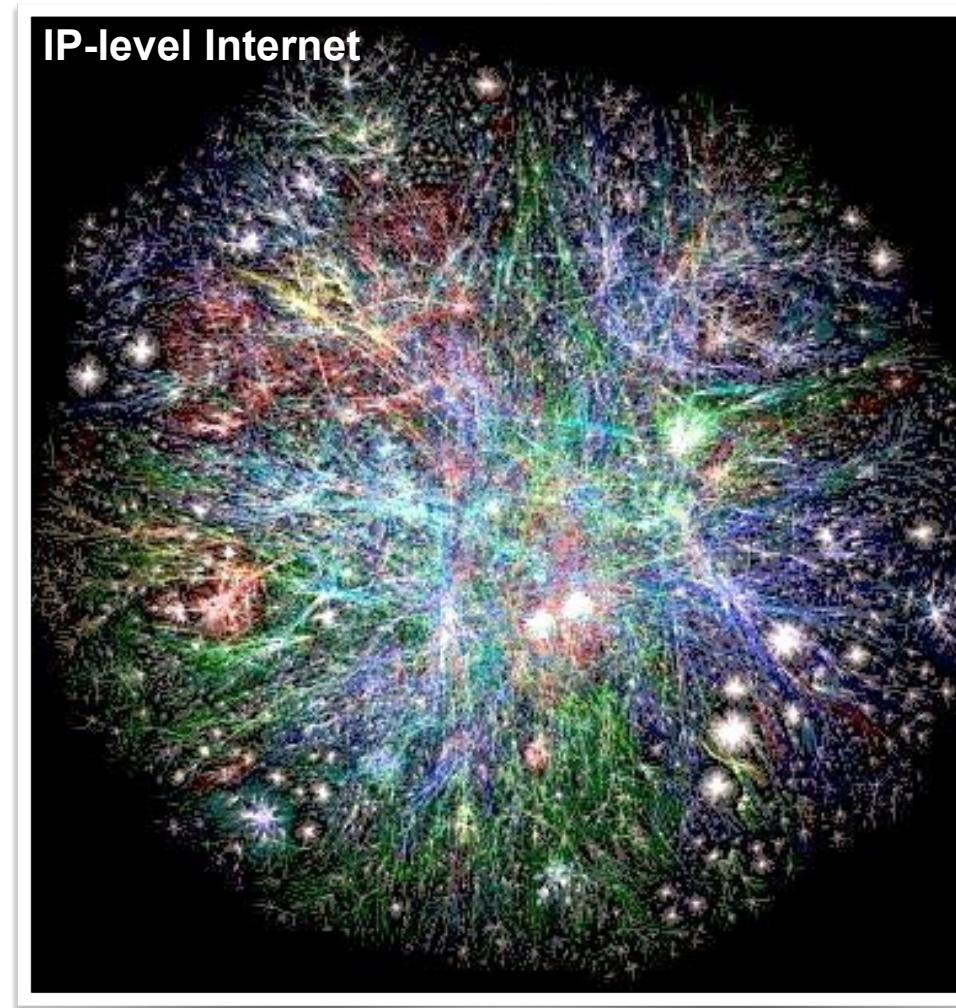
vertex

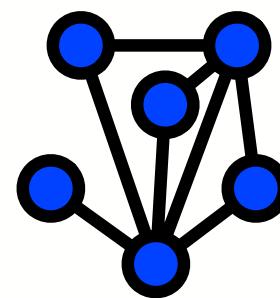


edge

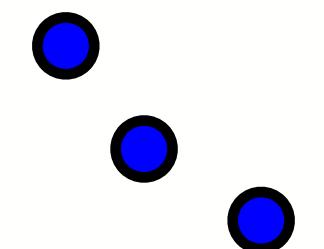


	network	vertex	edge
technological	Internet(1)	computer	IP network adjacency
	Internet(2)	autonomous system (ISP)	BGP connection

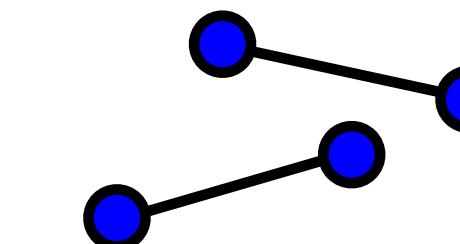




network

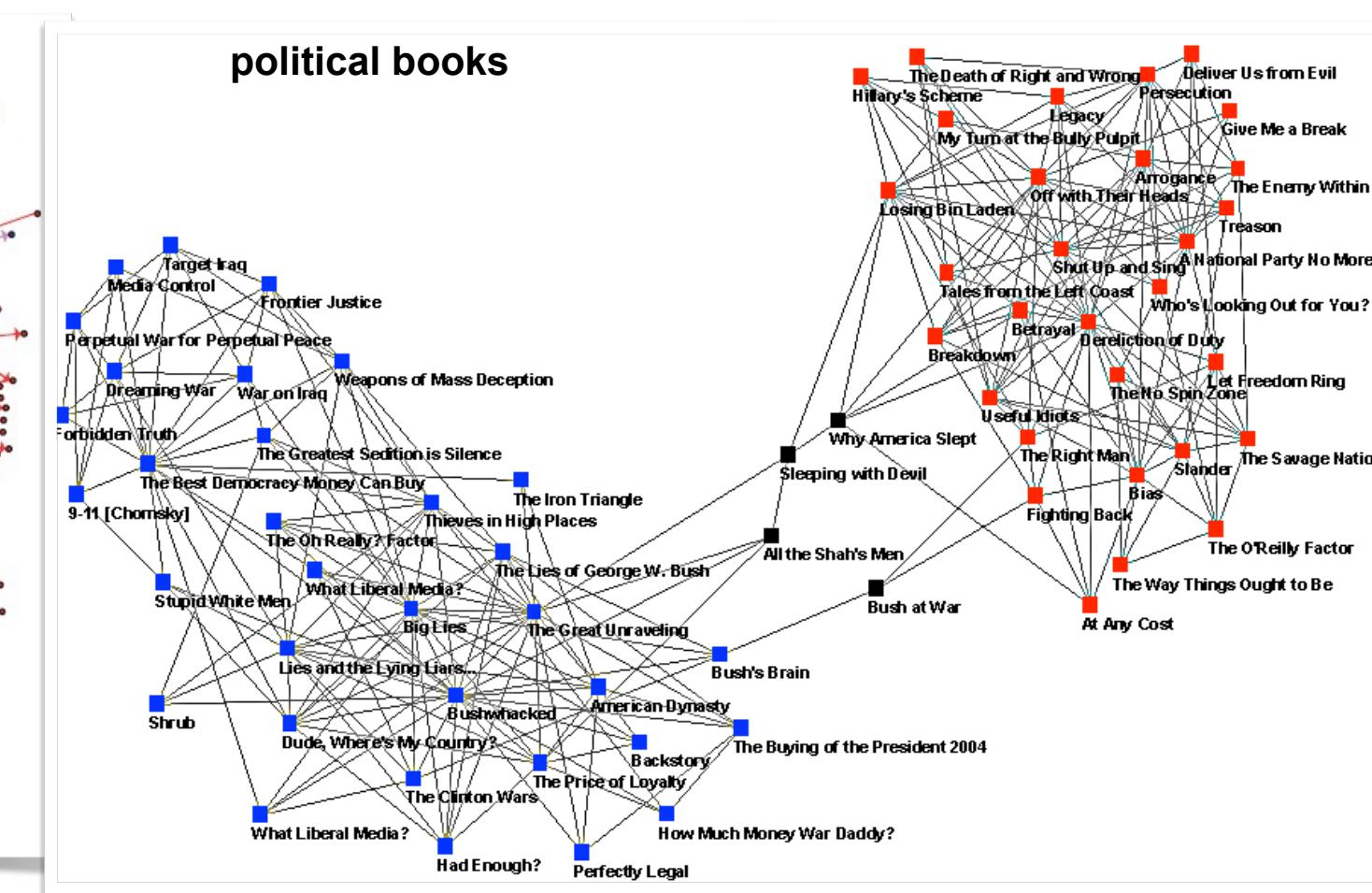
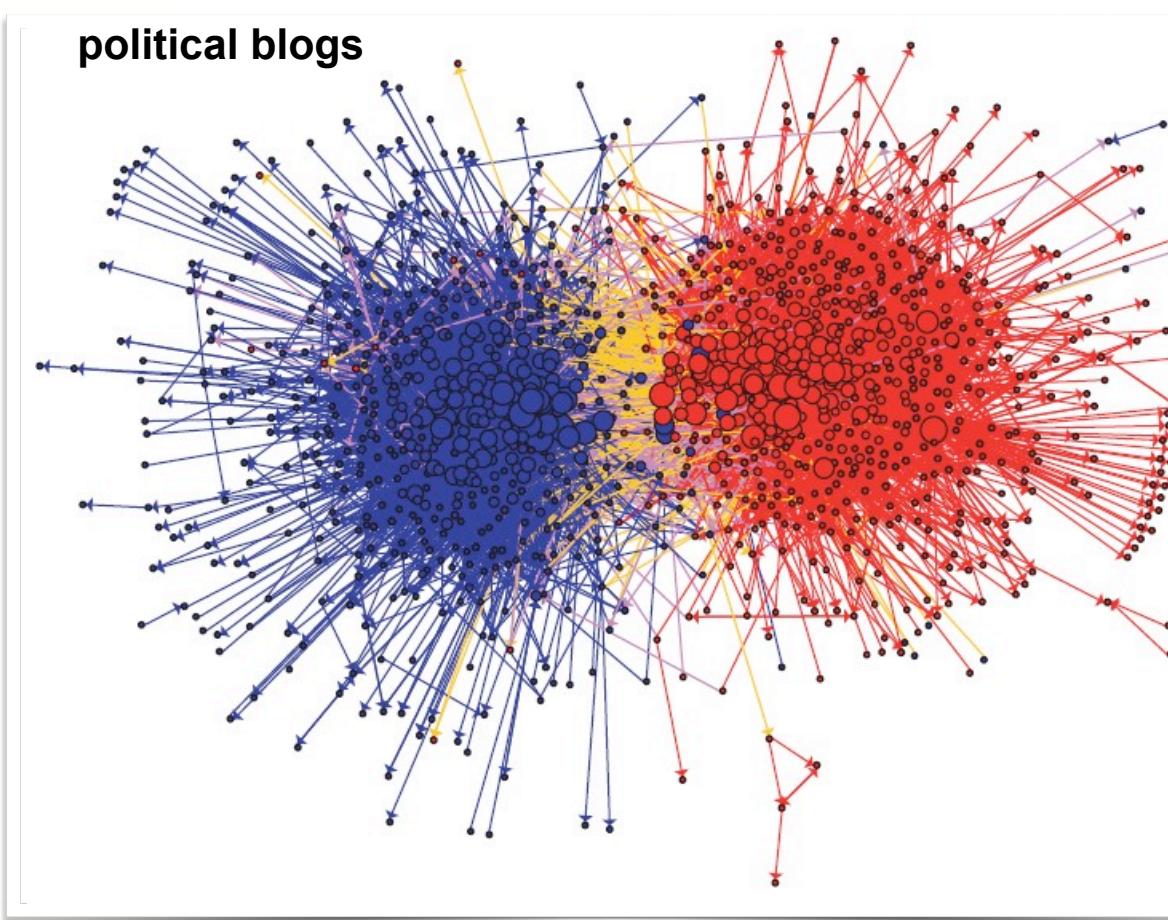


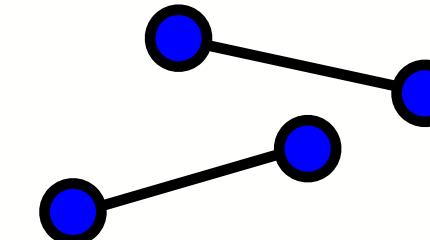
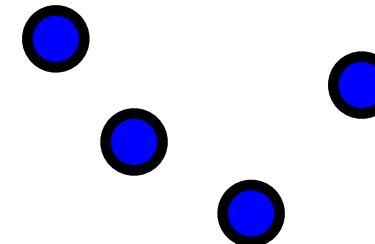
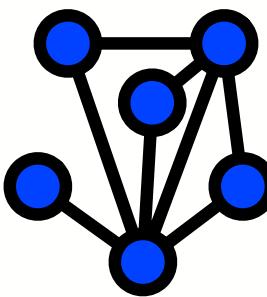
vertex



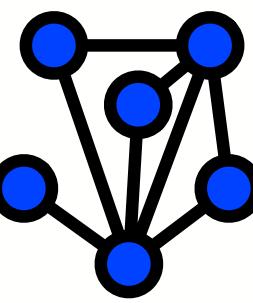
ed

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

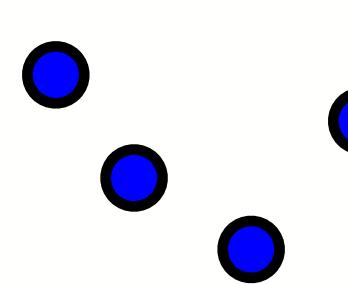




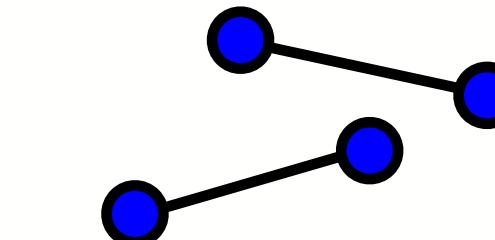
	network	vertex	edge
technological information	Internet(1)	computer	IP network adjacency
	Internet(2)	autonomous system (ISP)	BGP connection
	software	function	function call
	World Wide Web	web page	hyperlink
	documents	article, patent, or legal case	citation
	power grid transmission	generating or relay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



network



vertex



edge

road network(1)

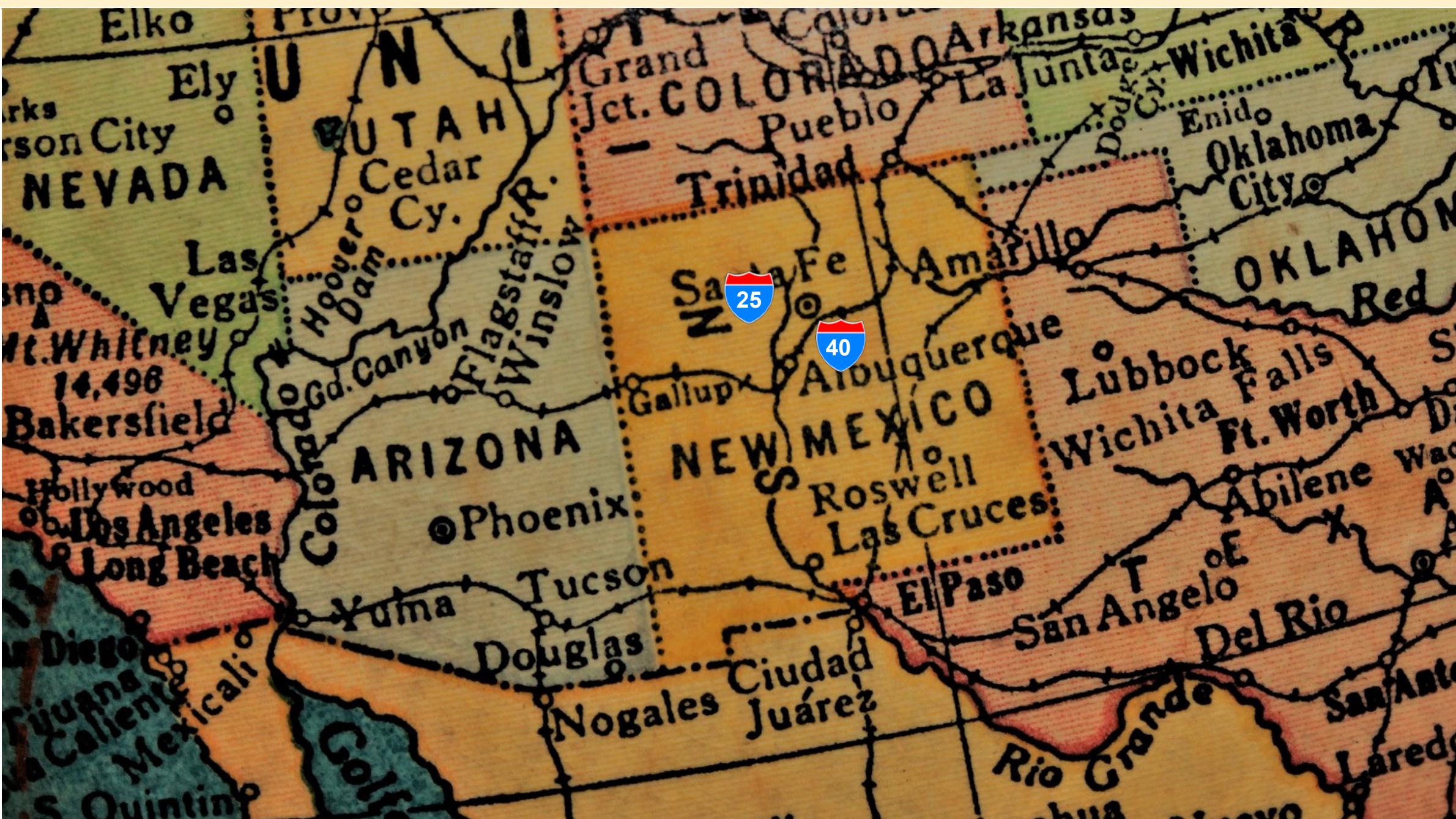
intersection

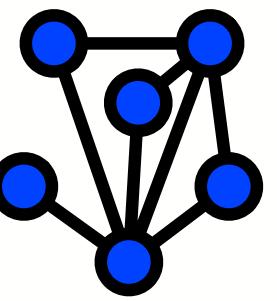
pavement

road network(2)

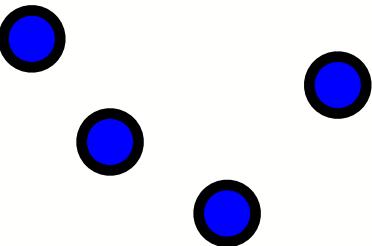
named road

intersection

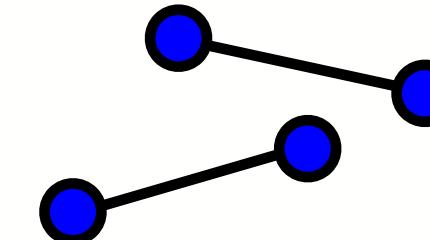




network



vertex



edge

road network(1)

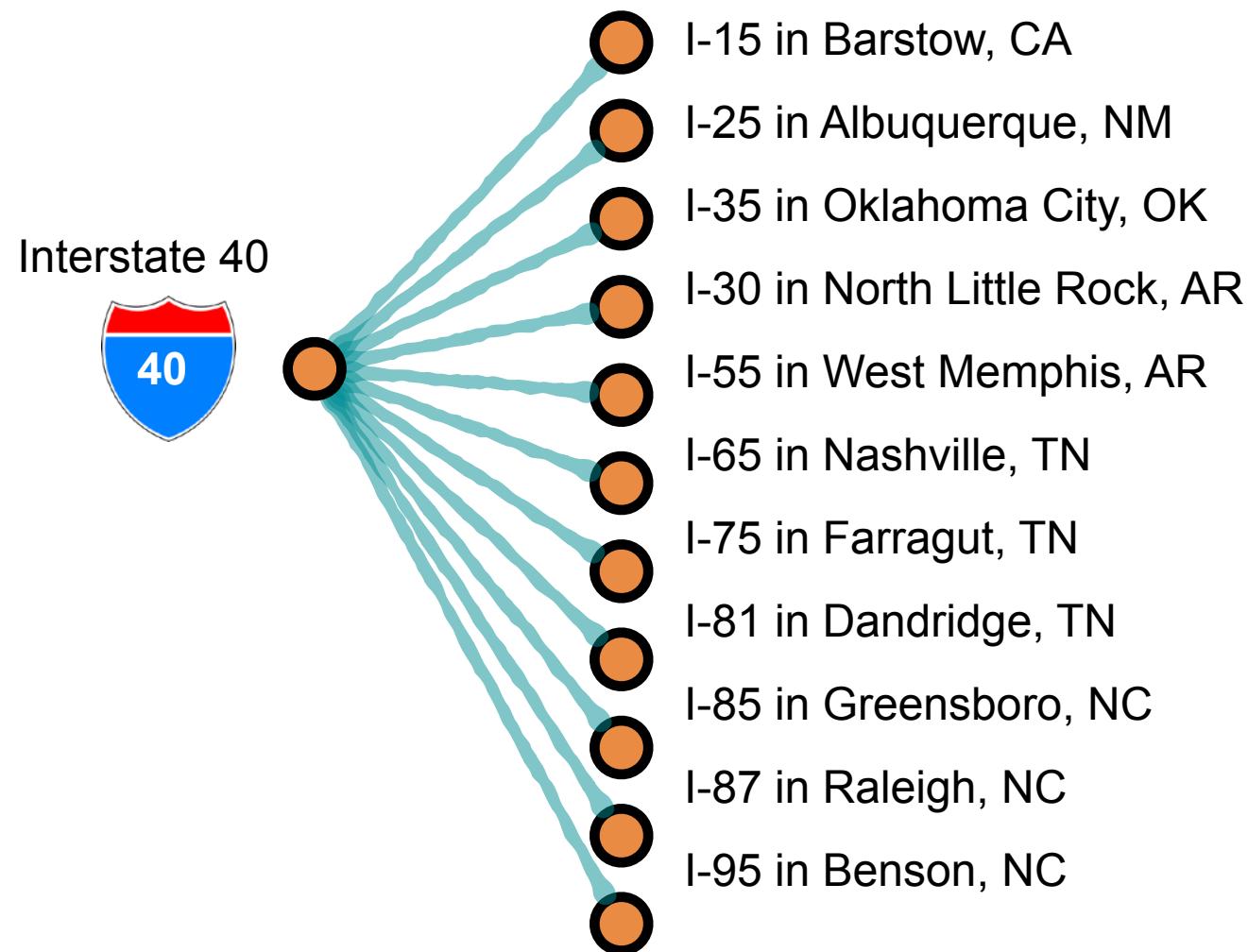
intersection

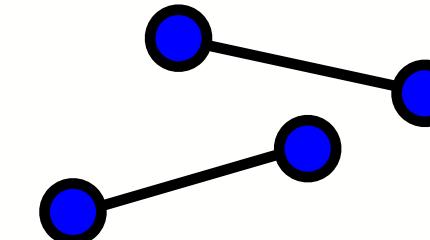
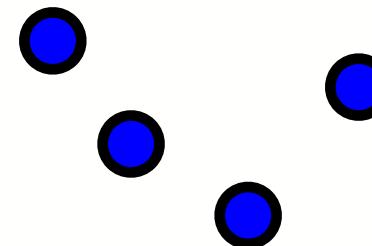
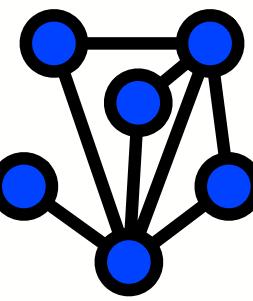
pavement

road network(2)

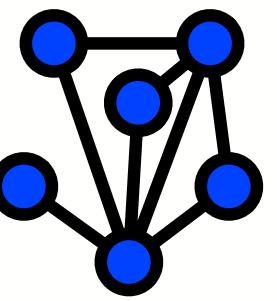
named road

intersection



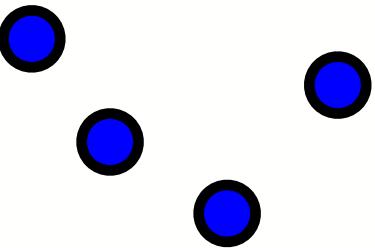


	network	vertex	edge
technological	Internet(1)	computer	IP network adjacency
	Internet(2)	autonomous system (ISP)	BGP connection
information	software	function	function call
	World Wide Web	web page	hyperlink
	documents	article, patent, or legal case	citation
	power grid transmission	generating or relay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



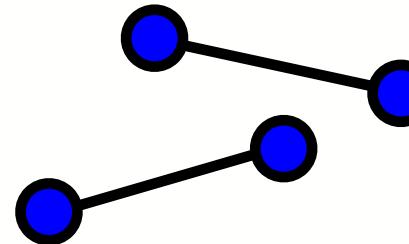
network

friendship network



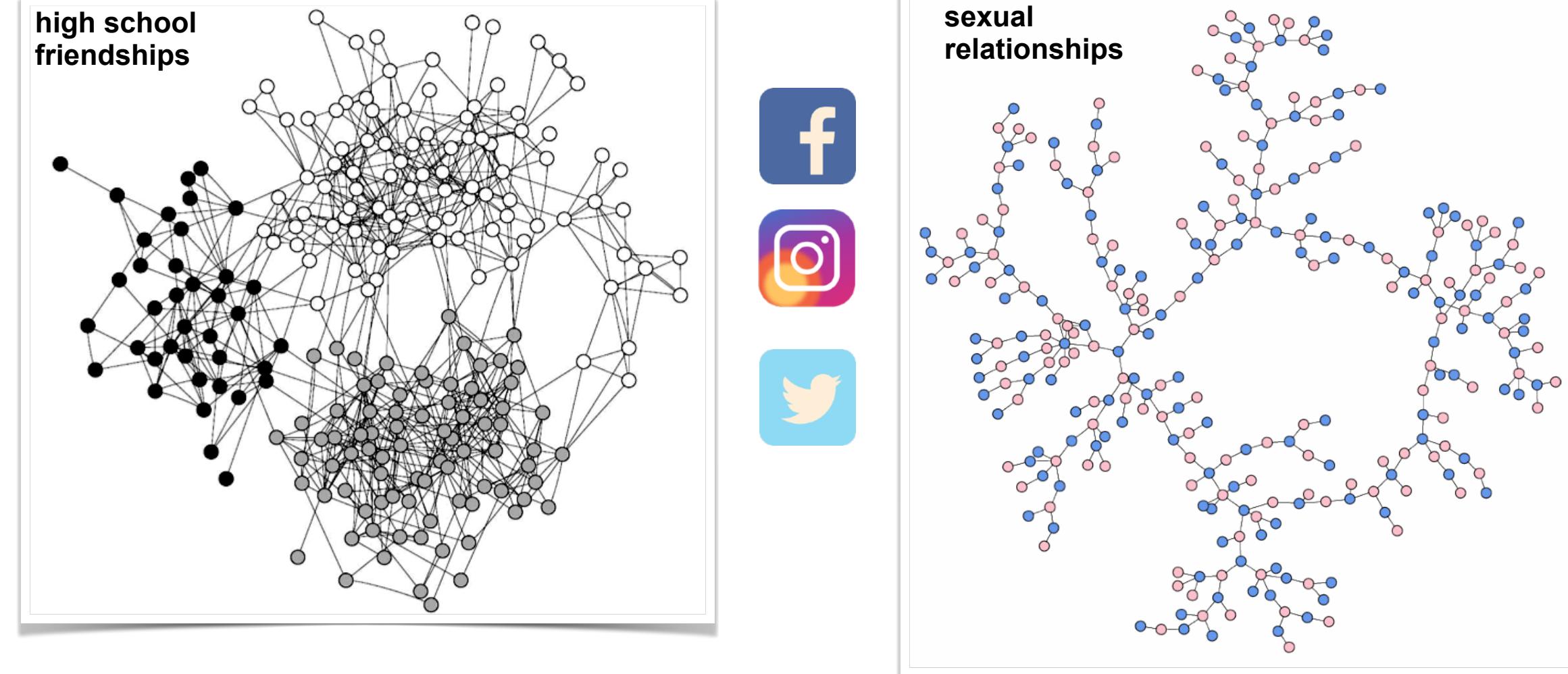
vertex

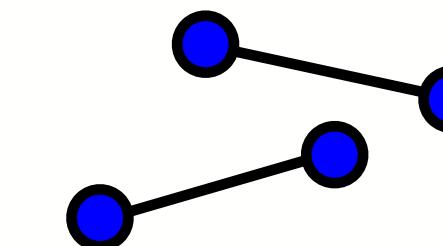
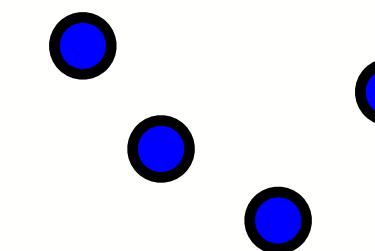
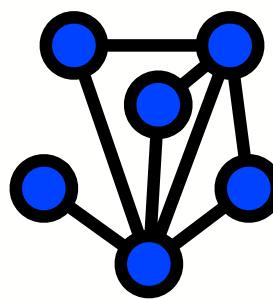
person



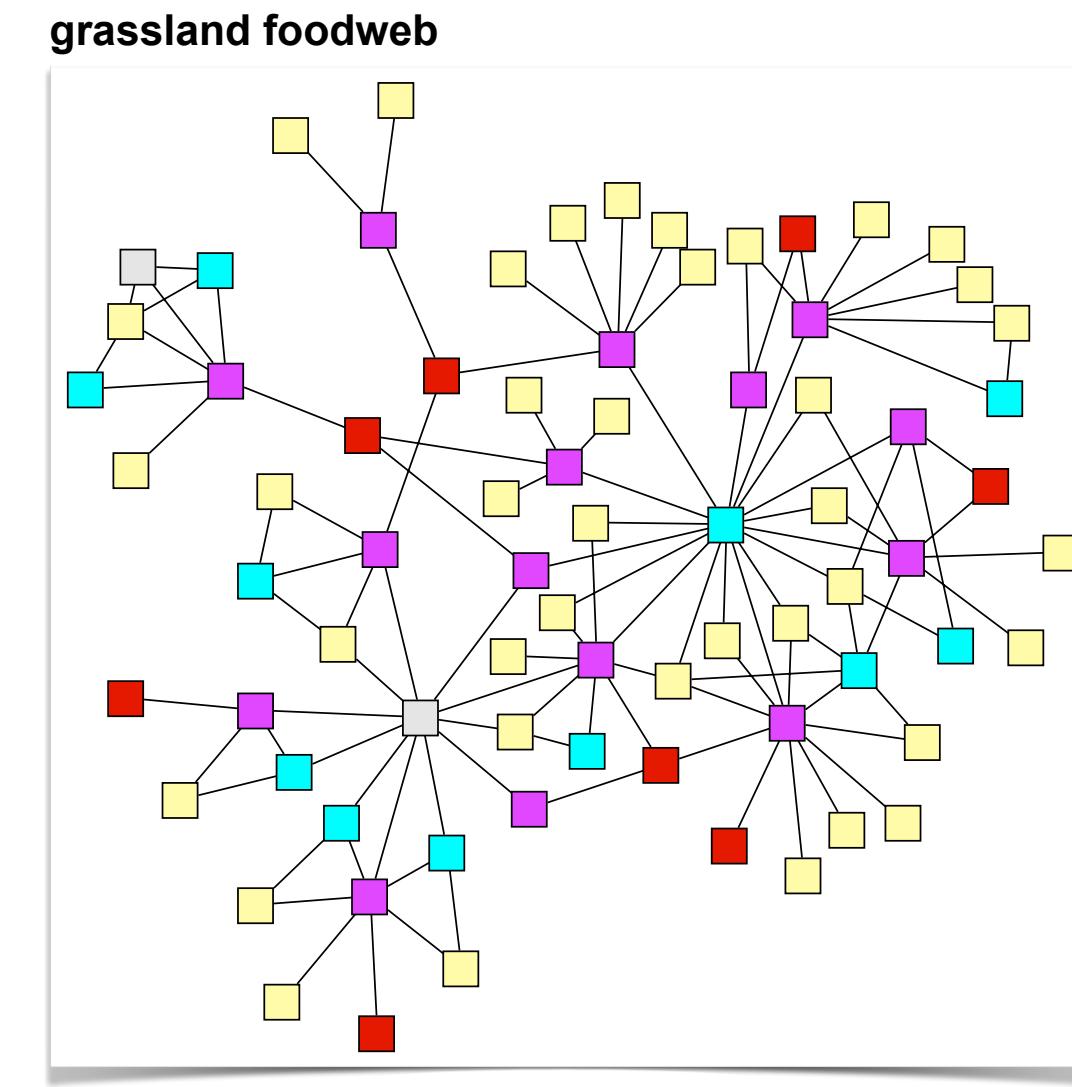
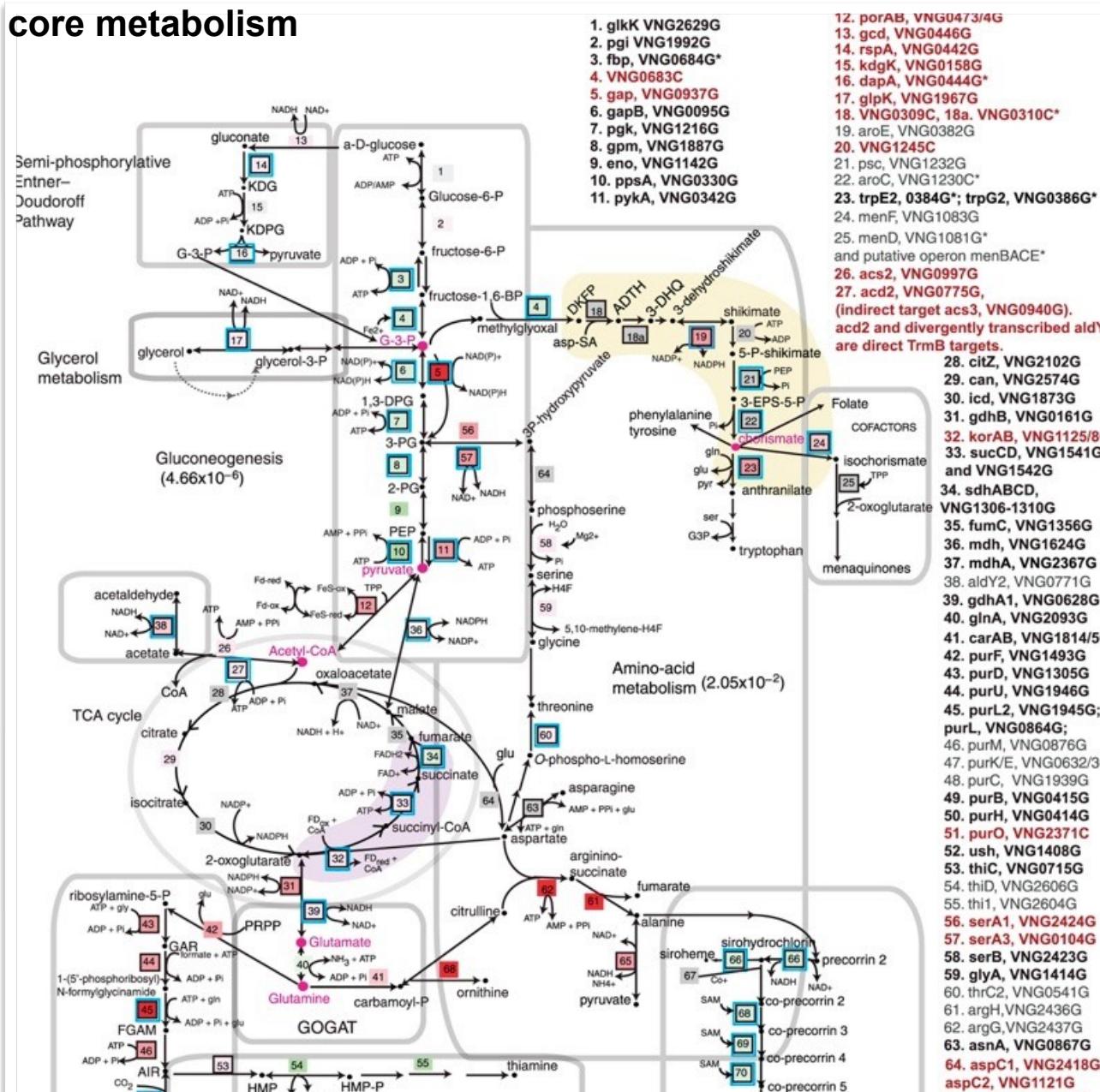
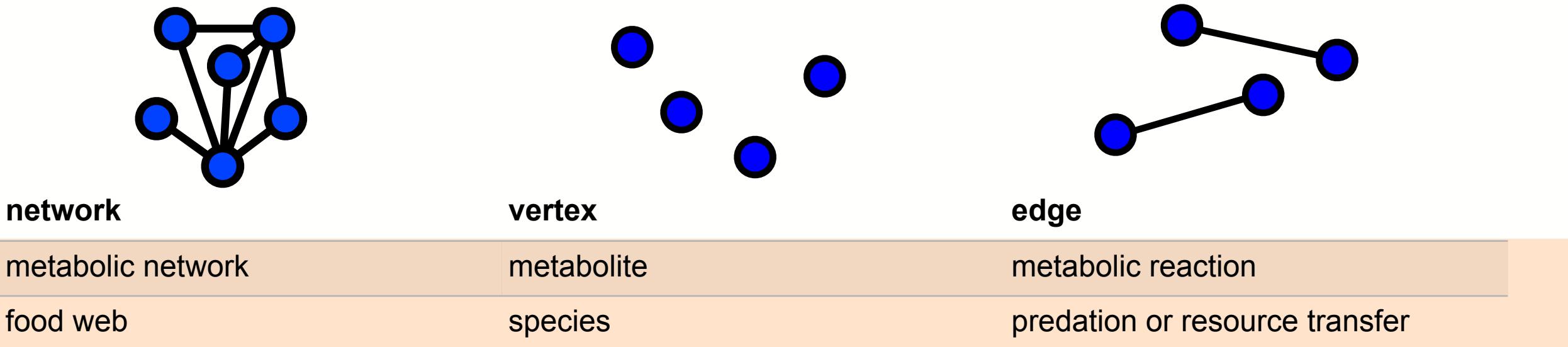
edge

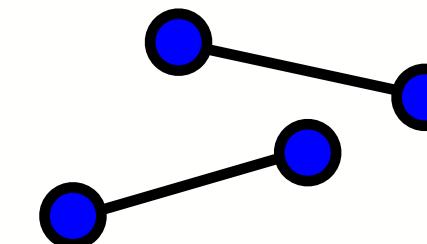
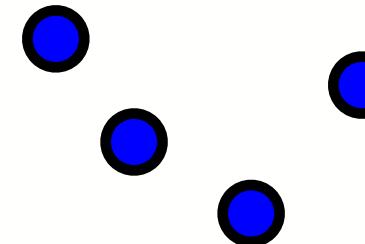
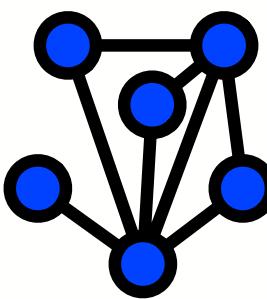
friendship





	network	vertex	edge
technological	Internet(1)	computer	IP network adjacency
	Internet(2)	autonomous system (ISP)	BGP connection
information	software	function	function call
	World Wide Web	web page	hyperlink
	documents	article, patent, or legal case	citation
	power grid transmission	generating or relay station	transmission line
transportation	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
	gene regulatory network	gene	regulatory effect
	neuronal network	neuron	synapse
	food web	species	predation or resource transfer

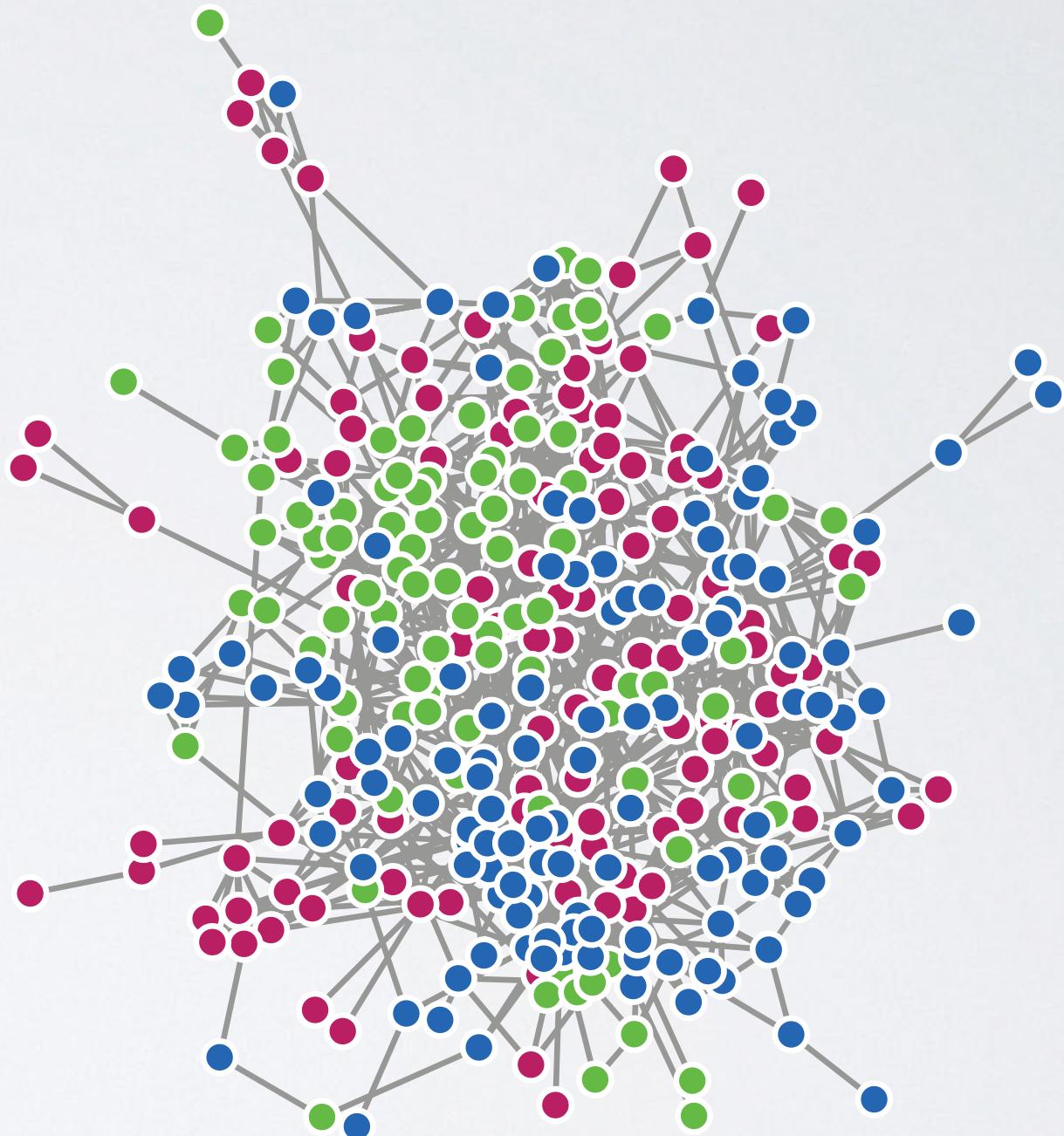




	network	vertex	edge
technological	Internet(1)	computer	IP network adjacency
	Internet(2)	autonomous system (ISP)	BGP connection
information	software	function	function call
	World Wide Web	web page	hyperlink
	documents	article, patent, or legal case	citation
	power grid transmission	generating or relay station	transmission line
transportation	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
	gene regulatory network	gene	regulatory effect
	neuronal network	neuron	synapse
economic	food web	species	predation or resource transfer
	faculty hiring	universities	faculty hiring

describing networks

what networks look like



describing networks

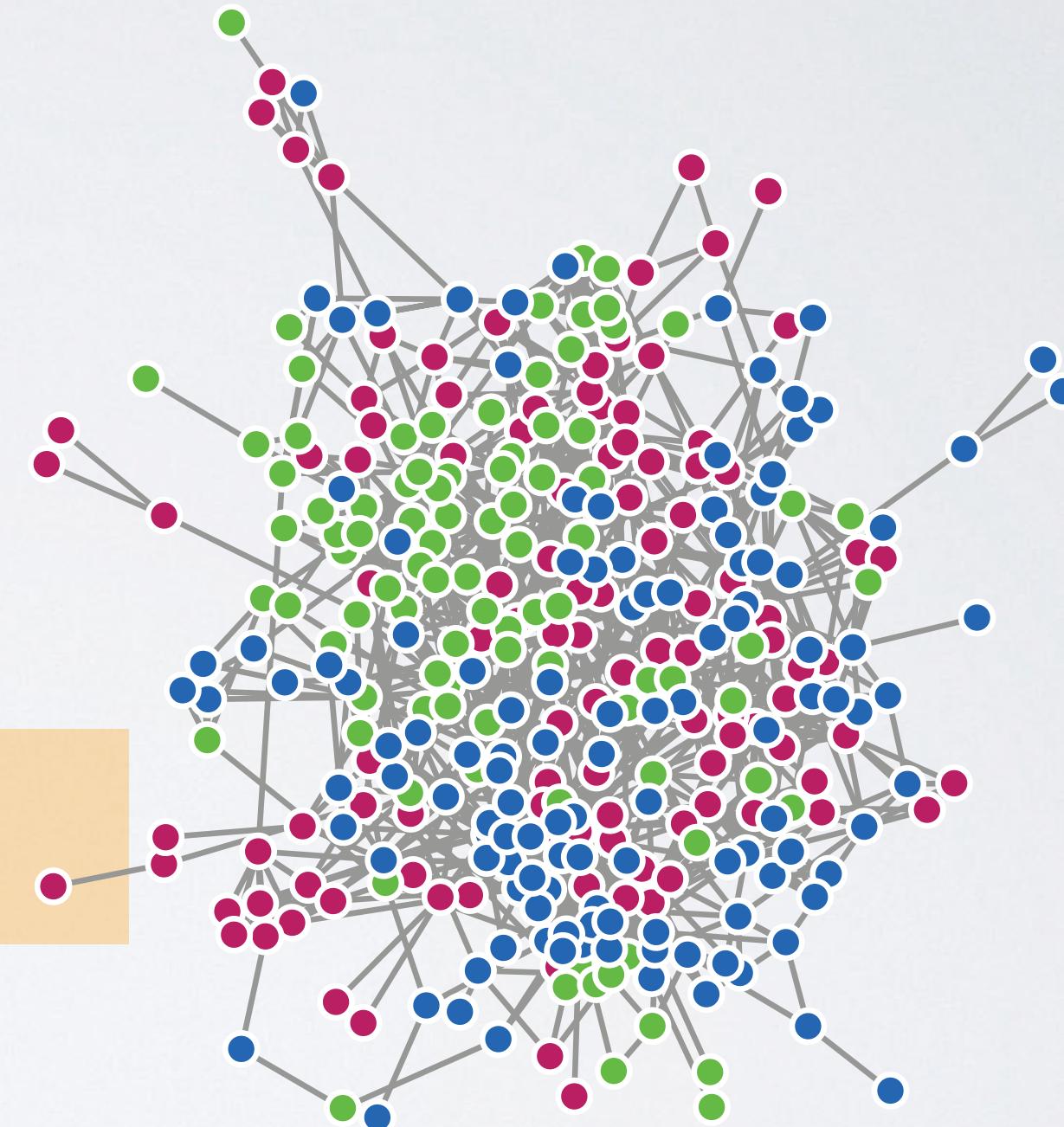
what networks look like

questions:

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

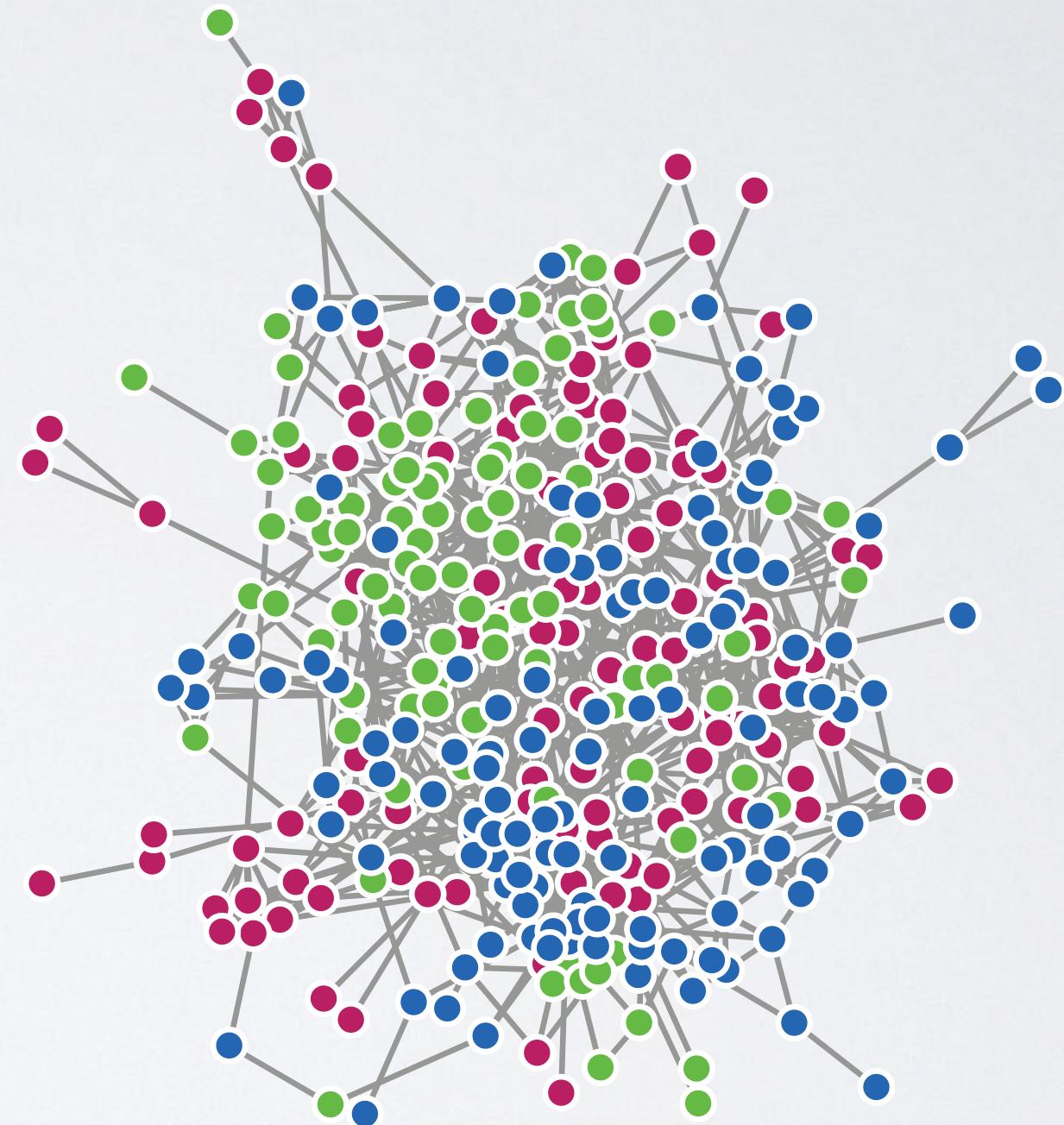
what we want to know

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



describing networks

a first step : describe its features

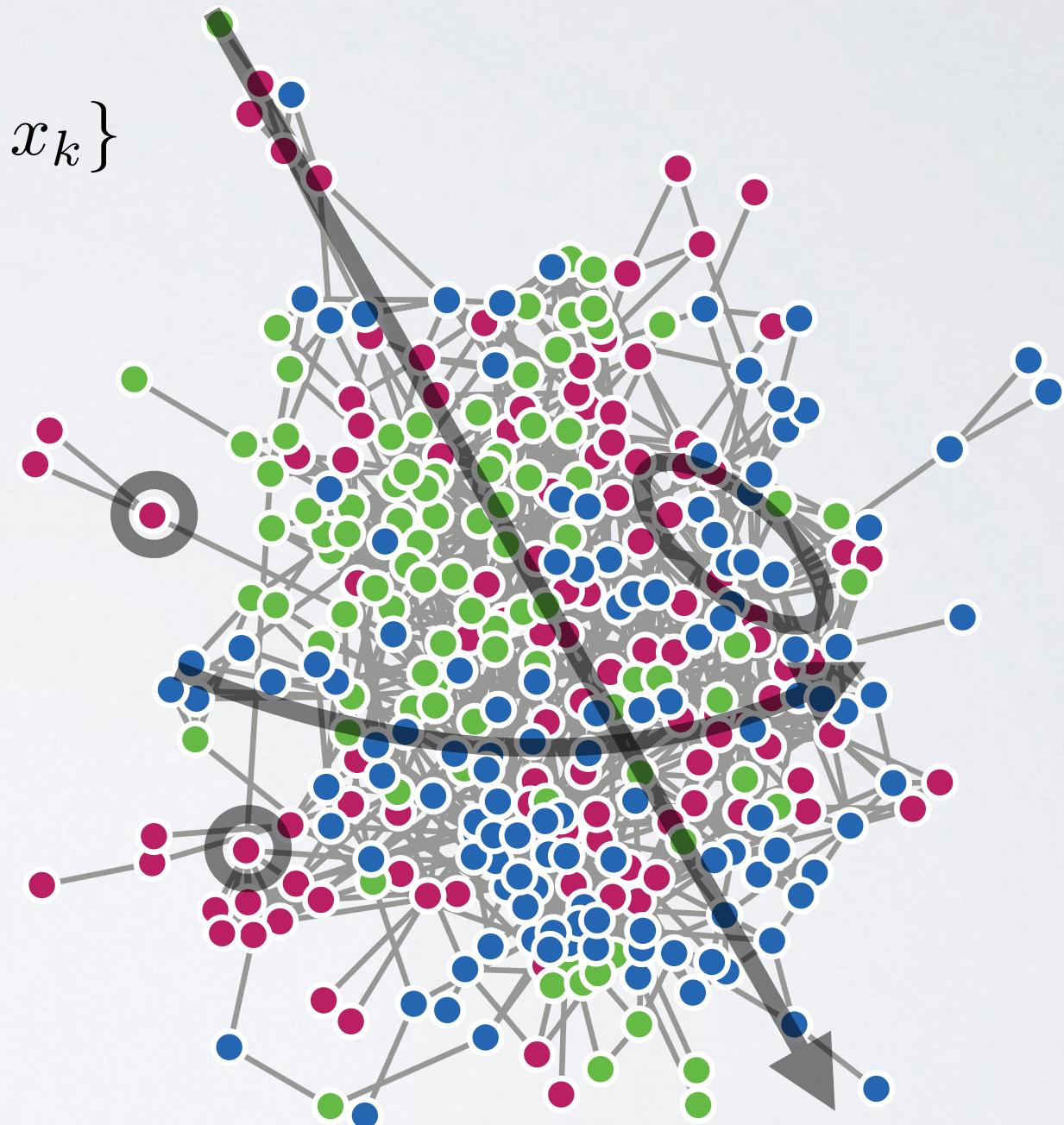


describing networks

a first step : describe its features

$$f : G \rightarrow \{x_1, \dots, x_k\}$$

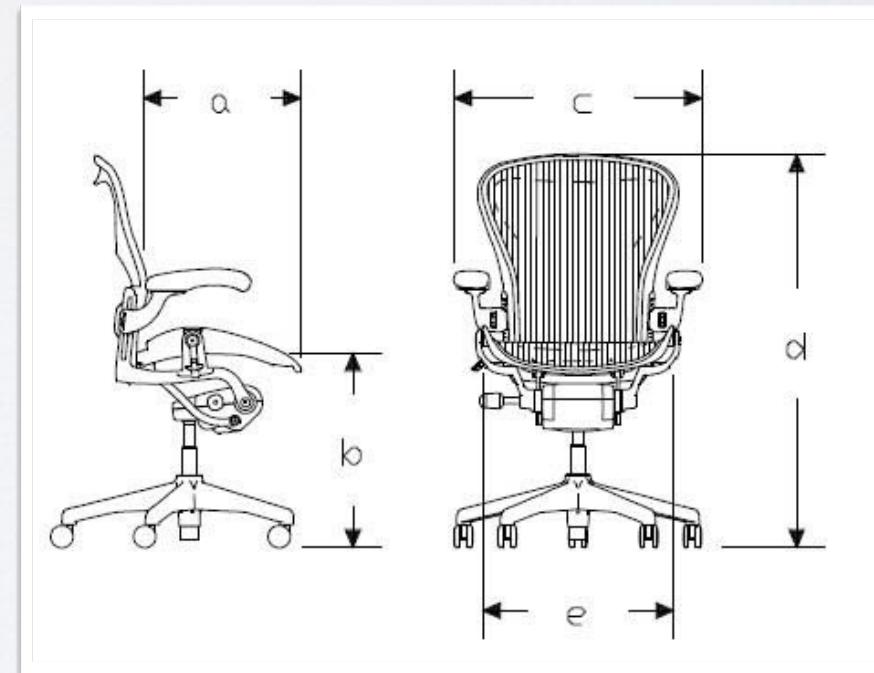
- degree distributions
- short-loop density (triangles, etc.)
- shortest paths (diameter, etc.)
- vertex positions
- correlations between these



describing networks

a first step : describe its features

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



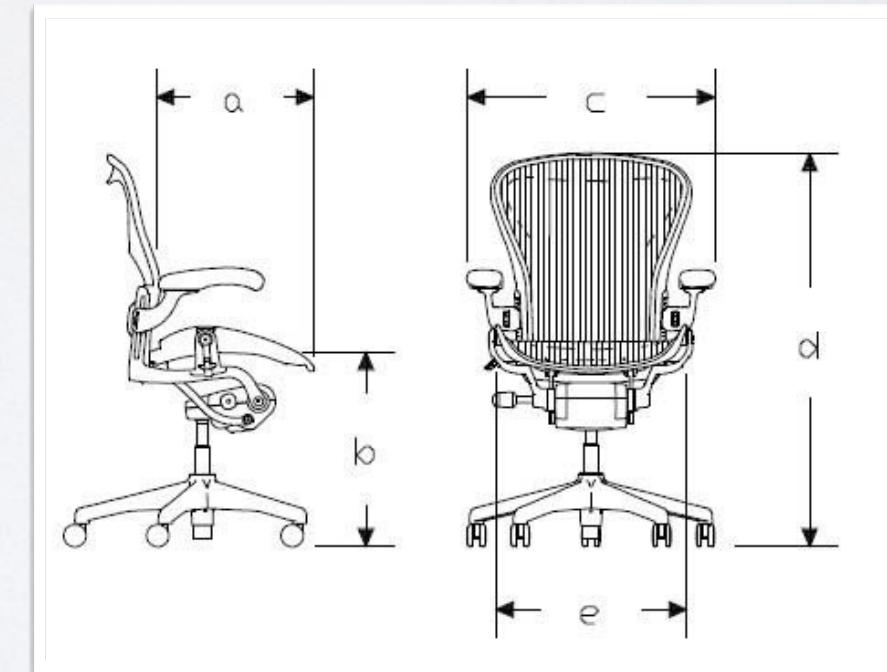
describing networks

a first step : describe its features

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

- 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

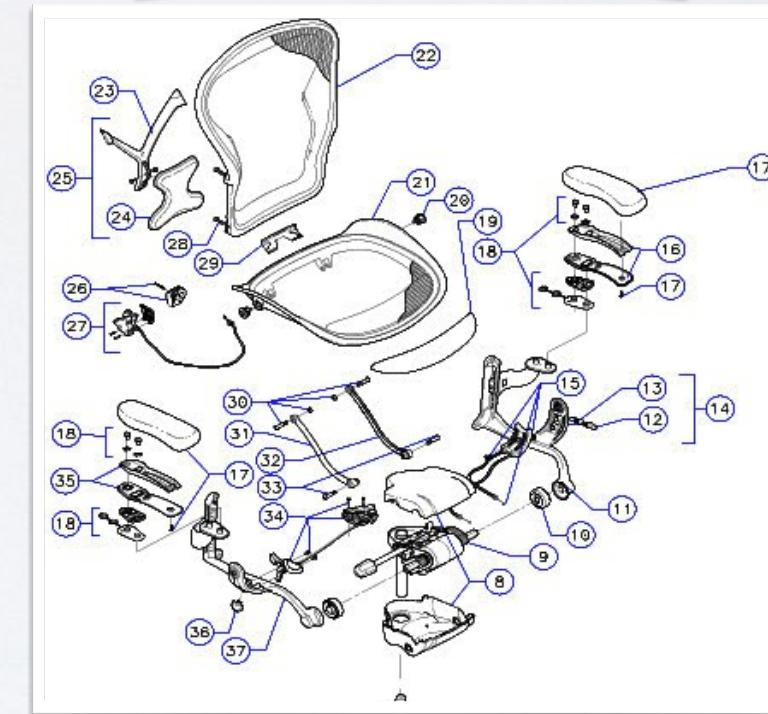
$$f : \text{object} \rightarrow \{\theta_1, \dots, \theta_k\}$$

- 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?



analyzing networks

4 major approaches

I. **exploratory network analysis:** descriptive! count & compare
(degree distributions, centrality scores, community detection, etc.)

null models: use a parameterized random graph to identify non-random patterns as deviations from the null

analyzing networks

4 major approaches

1. **exploratory network analysis:** descriptive! count & compare
(degree distributions, centrality scores, community detection, etc.)
null models: use a parameterized random graph to identify non-random patterns as deviations from the null
2. **explanatory network analysis:** relates node attributes to a node's local connectivity, via explanatory modeling (eg, regression)

analyzing networks

4 major approaches

1. **exploratory network analysis:** descriptive! count & compare
(degree distributions, centrality scores, community detection, etc.)
null models: use a parameterized random graph to identify non-random patterns as deviations from the null
2. **explanatory network analysis:** relates node attributes to a node's local connectivity, via explanatory modeling (eg, regression)
3. **predictive modeling:** fit a network model to data & predict missing or future data (edges, labels, etc.)

analyzing networks

4 major approaches

- ★ 1. **exploratory network analysis:** descriptive! count & compare
(degree distributions, centrality scores, community detection, etc.)
- ★ **null models:** use a parameterized random graph to identify non-random patterns as deviations from the null
- 2. **explanatory network analysis:** relates node attributes to a node's local connectivity, via explanatory modeling (eg, regression)
- ★ 3. **predictive modeling:** fit a network model to data & predict missing or future data (edges, labels, etc.)
- 4. **causal modeling:** identify causes-and-effects in networks
 - ★ **mechanisms / simulations:** explain structural or dynamical patterns as caused by specific process, analyzed mathematically or simulated
 - network experiments:** manipulate structure and measure node-level or graph-level behavior as function of changes (incl. causal inference)