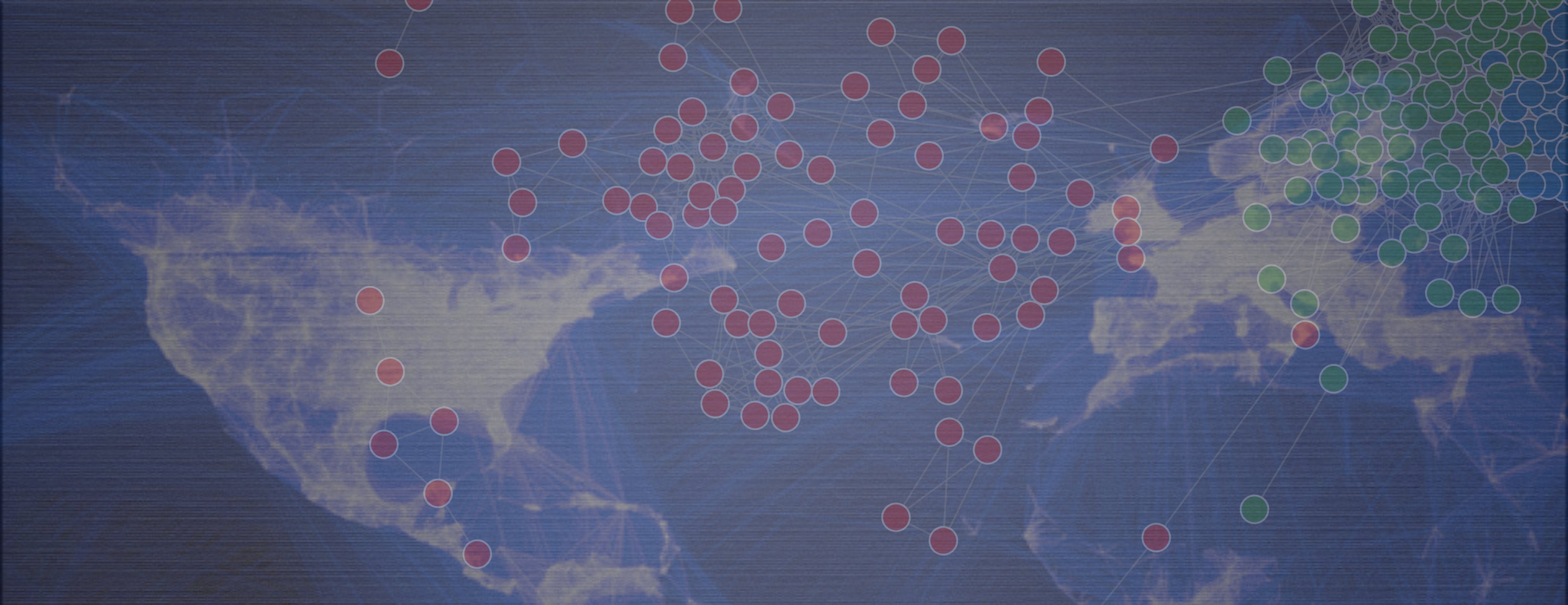


Five Lectures on Networks

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

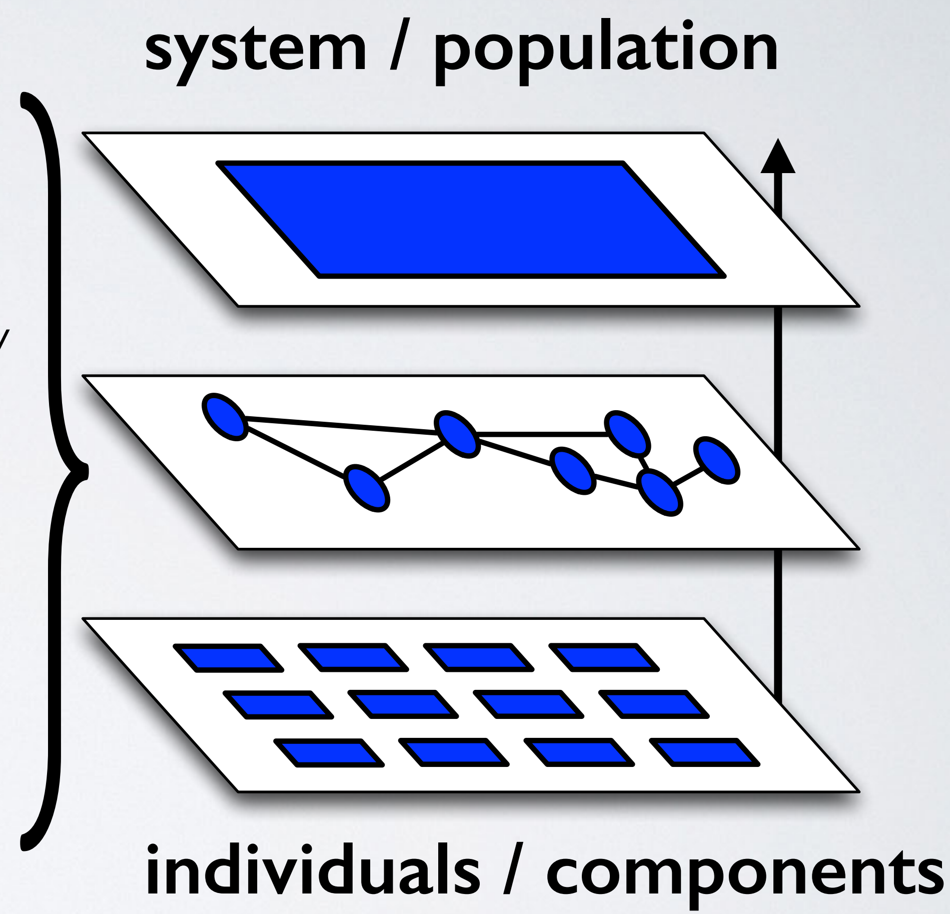
lecture 1

© 2014 Aaron Clauset

what are networks?

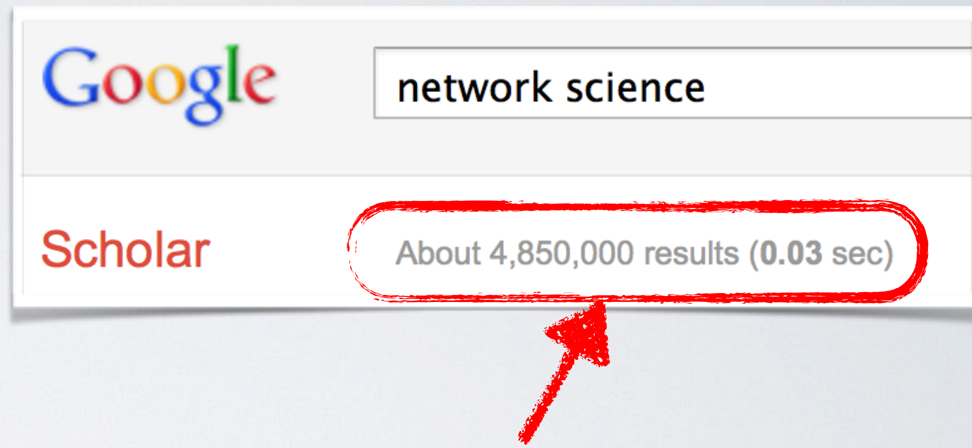
what are networks?

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



these lectures

- build intuition
- expose key concepts
- highlight some big questions
- teach a little math
- provide many examples
- give pointers to further study
- not a substitute for technical coursework



it's a big field now

University of Colorado **Boulder**

Network Analysis and Modeling

Instructor: Aaron Clauset

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 algorithms and methods, and both lectures and assignments will emphasize model interpretability and understanding the processes that generate real data. Applications will be drawn from computational biology and computational social science. No biological or social science training is required. (Note: this is not a scientific computing course, but there will be plenty of computing for science.)

Full lectures notes online (~150 pages in PDF)

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

Software

R
Python
Matlab
NetworkX [python]
graph-tool [python, c++]
GraphLab [python, c++]

Standalone editors

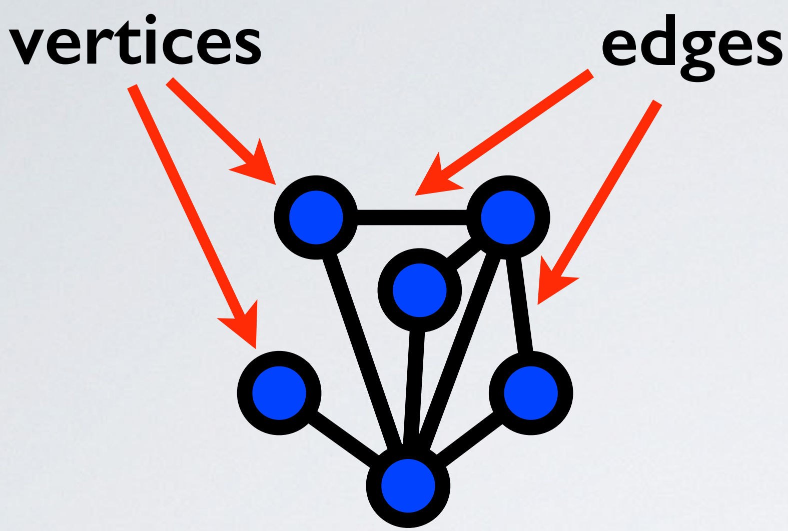
UCI-Net
NodeXL
Gephi
Pajek
Network Workbench
Cytoscape
yEd graph editor
Graphviz

Data sets

Mark Newman's network data sets
Stanford Network Analysis Project
Carnegie Mellon CASOS data sets
NCEAS food web data sets
UCI NET data sets
Pajek data sets
Linkgroup's list of network data sets
Barabasi lab data sets
Jake Hofman's online network data sets
Alex Arenas's data sets

1. defining a network
2. describing a network
3. null models for networks
4. statistical inference
5. network dynamics

1. **defining a network**
2. **describing a network**
3. null models for networks
4. statistical inference



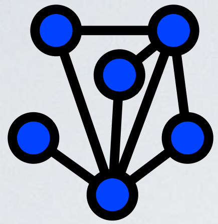
what is a vertex?

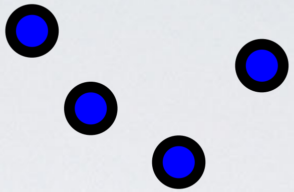
V distinct objects (vertices / nodes / actors)

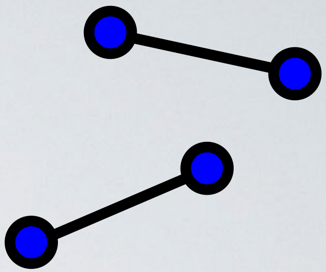
when are two vertices connected?

$E \subseteq V \times V$

pairwise relations (edges / links / ties)





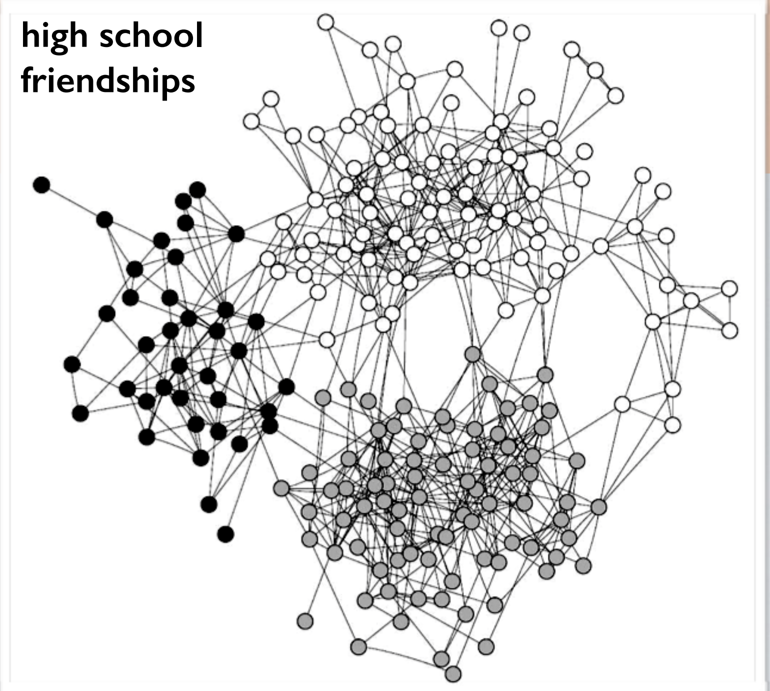


	network	**vertex**	**edge**
telecommunications	Internet(1)	computer	IP network adjacency
	Internet(2)	autonomous system (ISP)	BGP connection
informational	software	function	function call
	World Wide Web	web page	hyperlink
	documents	article, patent, or legal case	citation
transportation	power grid transmission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

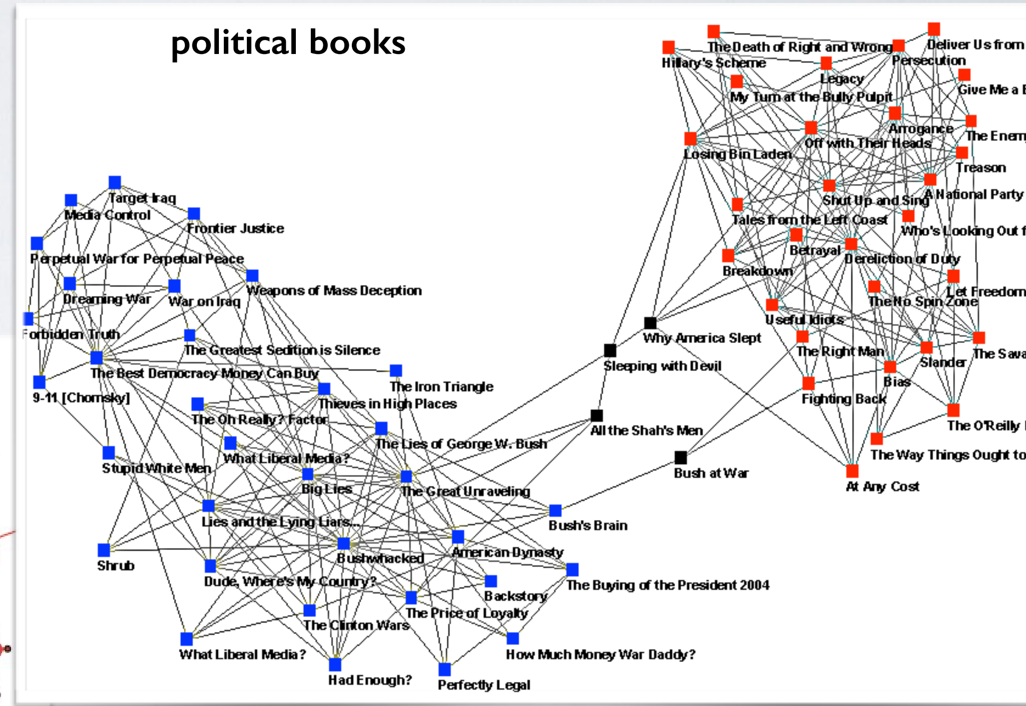
edge: friendship, collaborations, sexual contacts, communication, authority, exchange, etc.

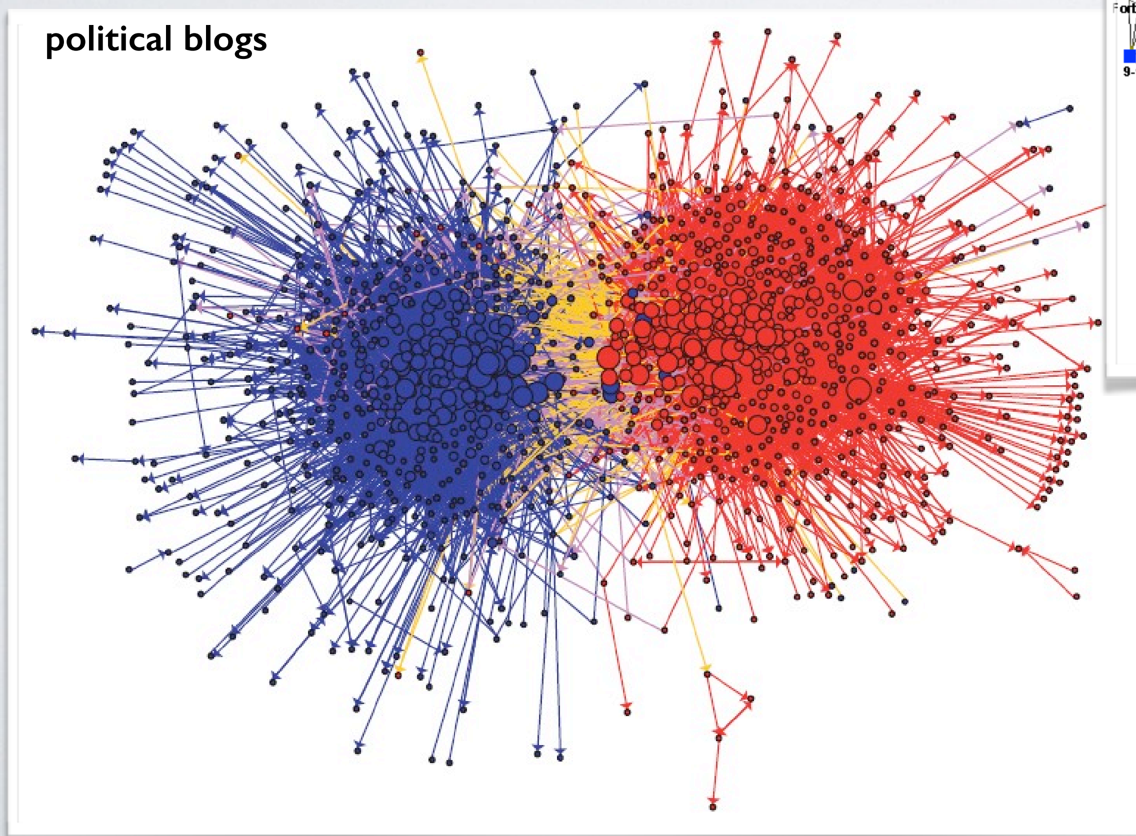


information networks

vertex: books, blogs, webpages, etc.

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

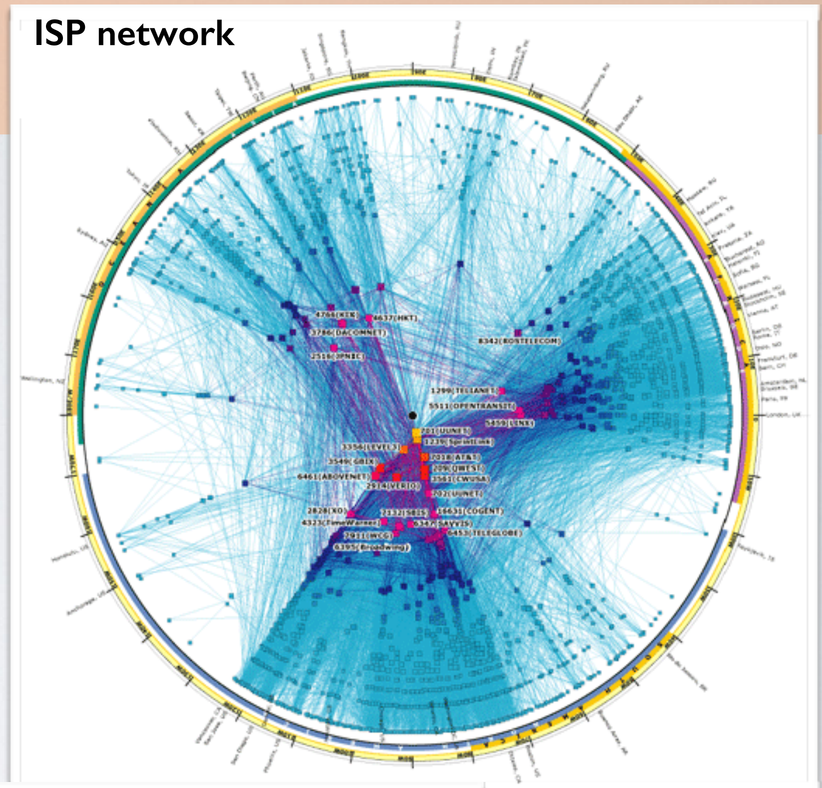


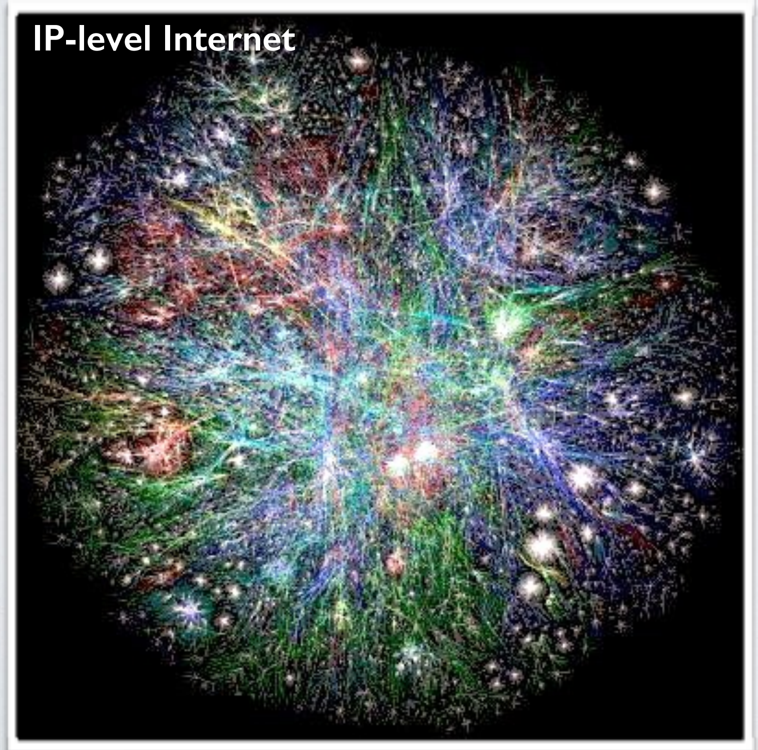


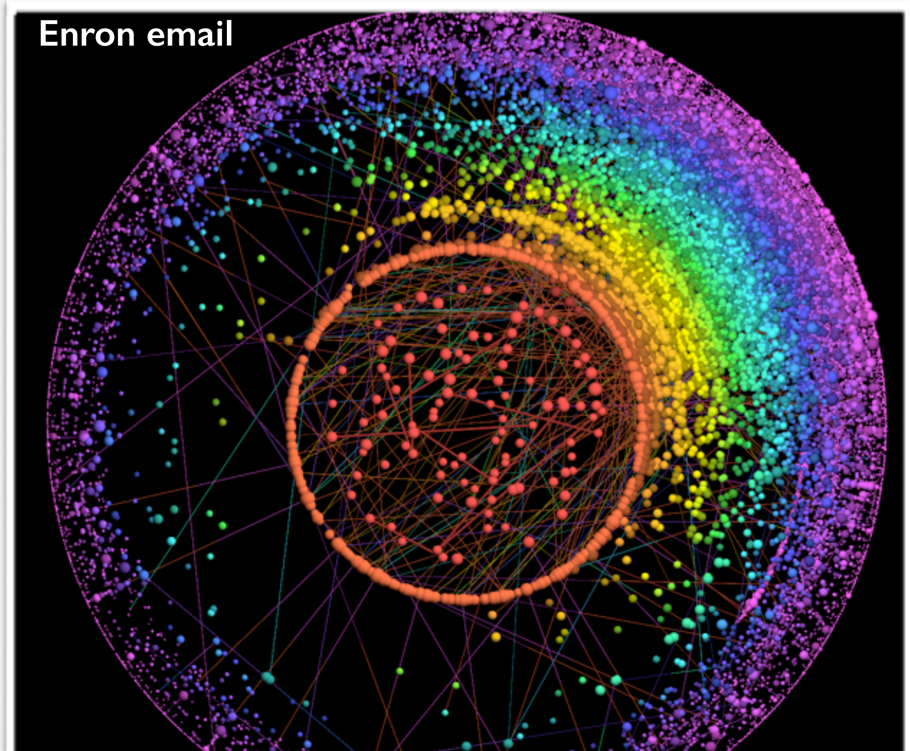
communication networks

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

edge: exchange of information





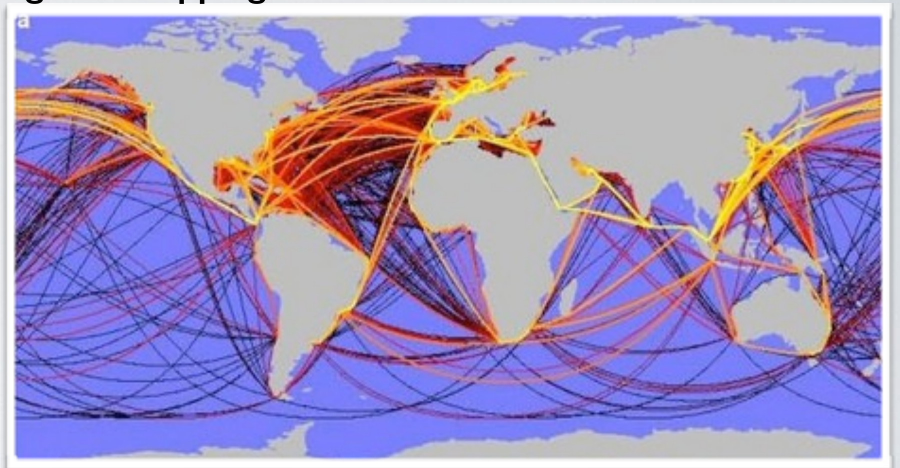


transportation networks

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

edge: physical transportation of material

global shipping



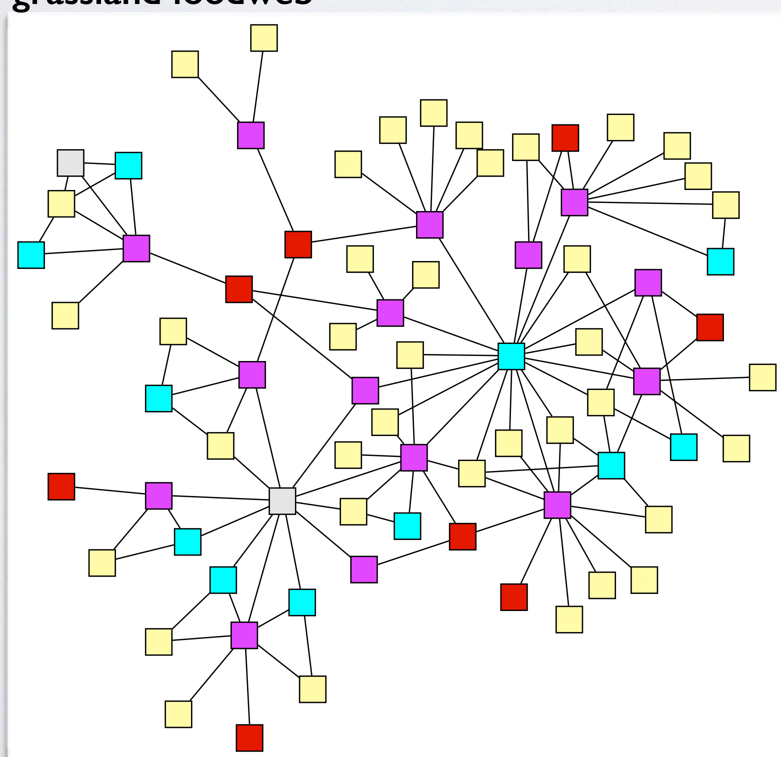
US Interstates

biological networks

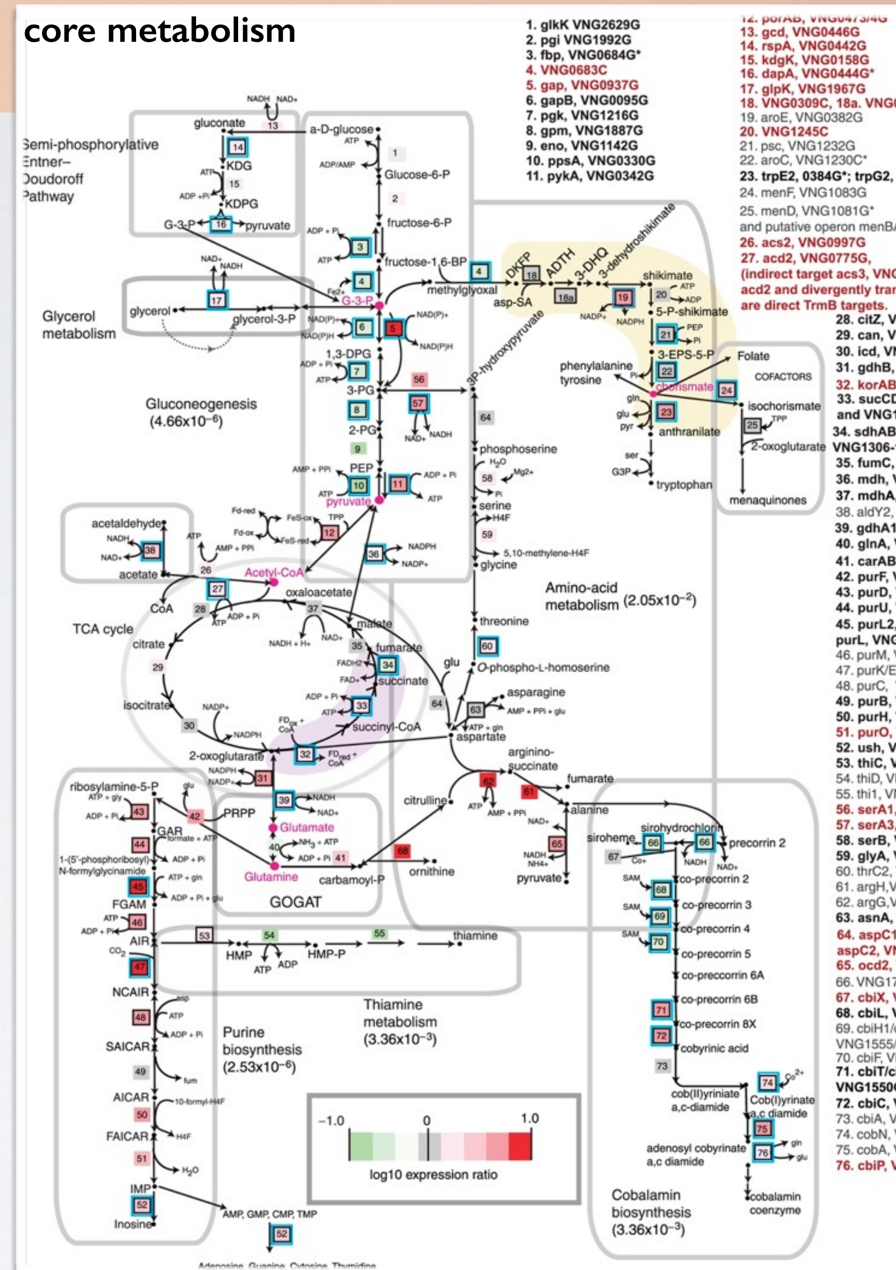
vertex: species, metabolic, protein, gene, neuron, etc.

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

grassland foodweb



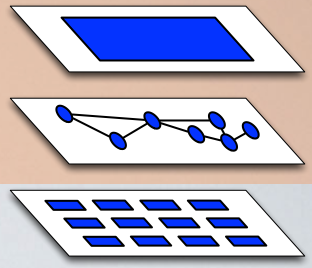
core metabolism



what's a network?

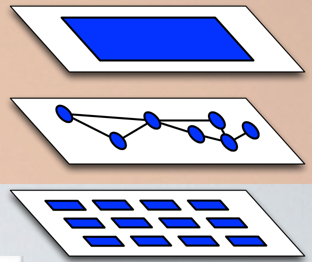
pop quiz

what's a network?



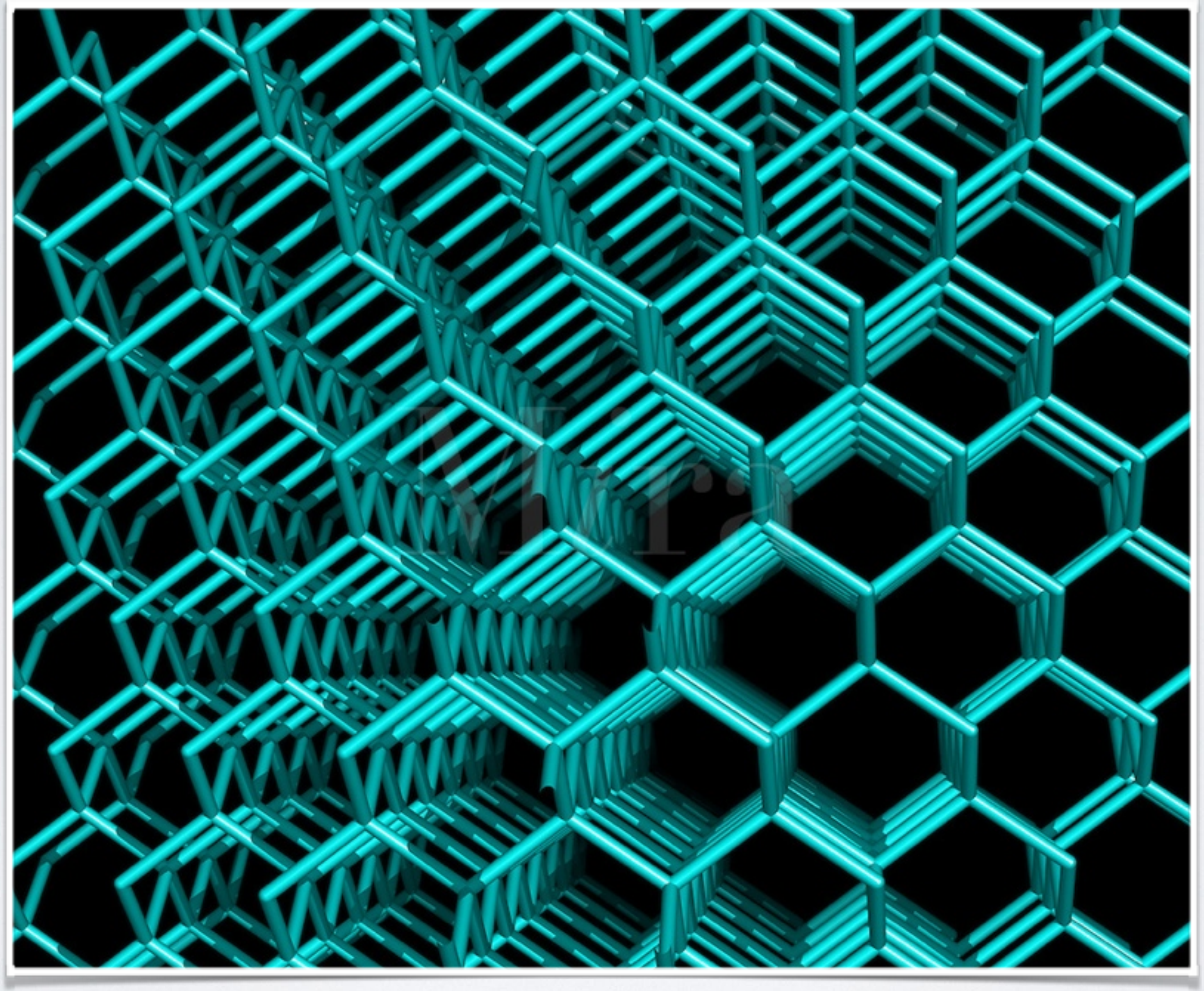
Andromeda galaxy

what's a network?



cauliflower fractal

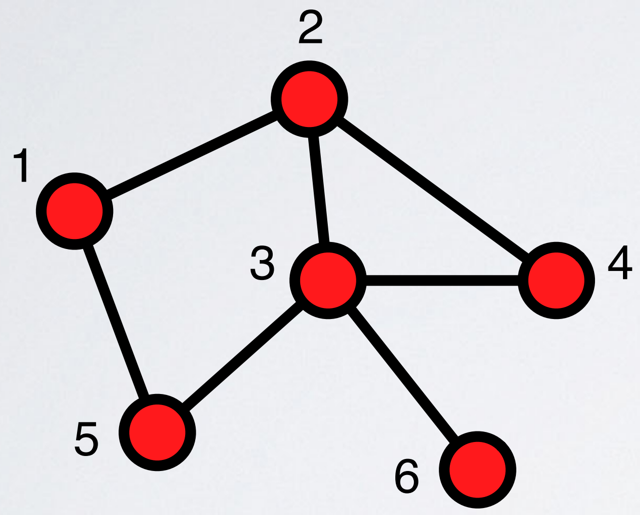
what’s a network?



diamond lattice

representing networks

a *simple* network

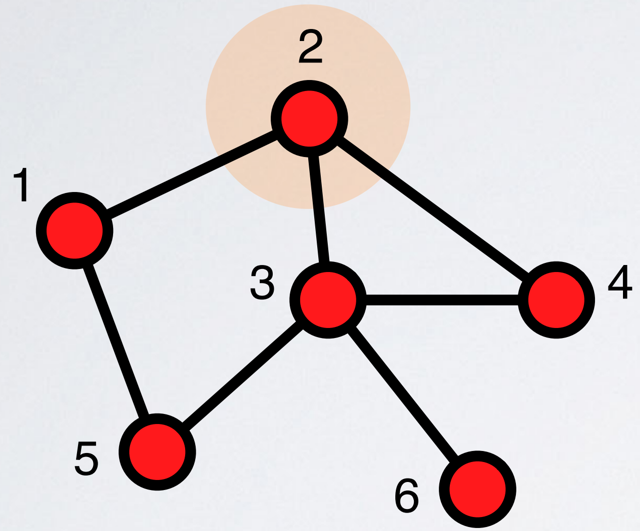


undirected

unweighted

no self-loops

a *simple* network



undirected

unweighted

no self-loops

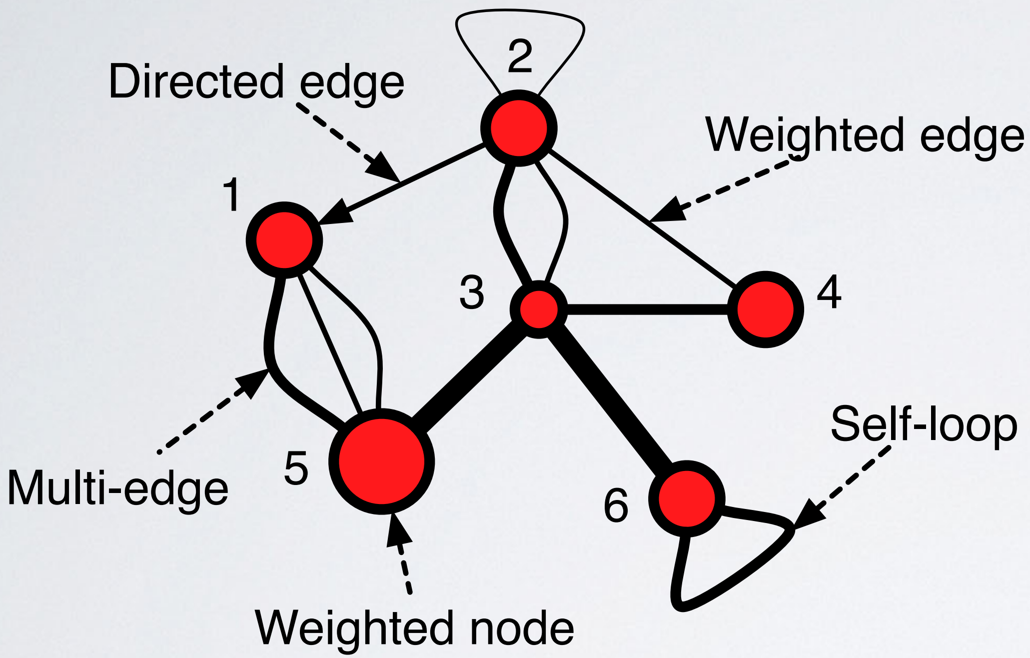
adjacency matrix

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

adjacency list

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

a less simple network

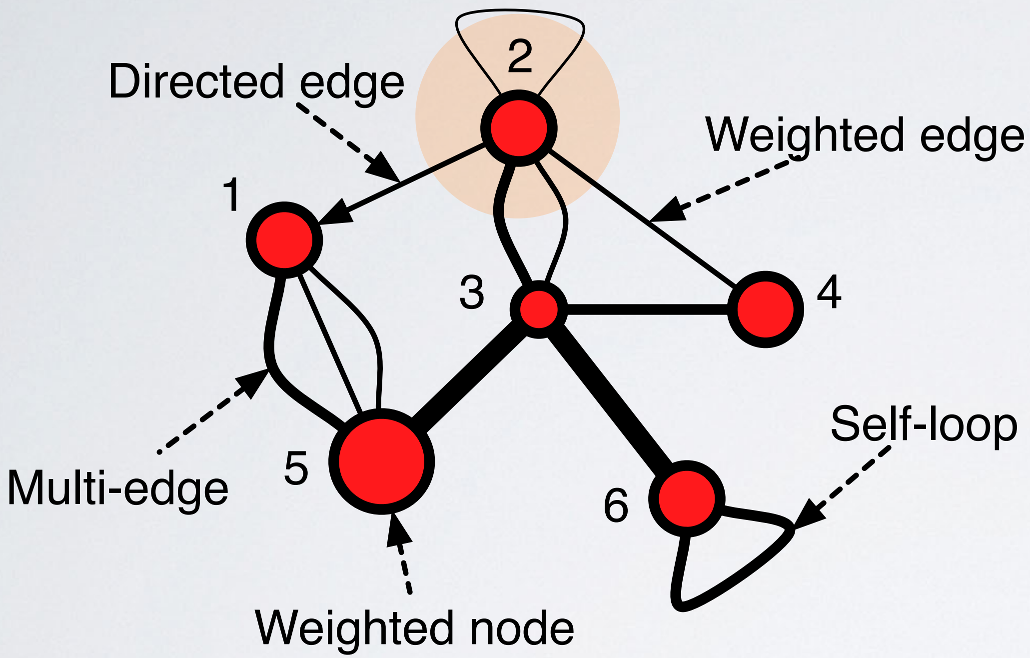


~~undirected~~

~~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

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	$\rightarrow \{(3,4),(6,2)\}$

directed networks

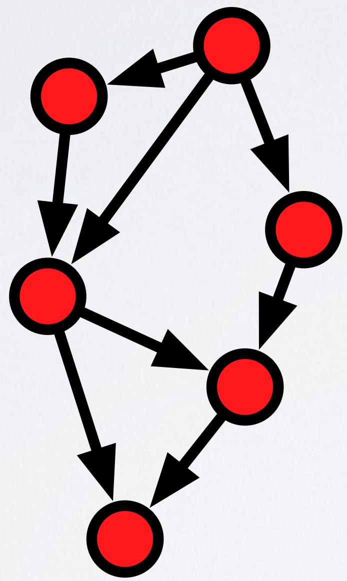
$$A_{ij} \neq A_{ji}$$

citation networks

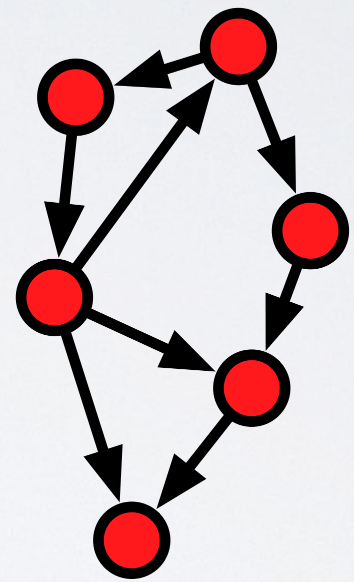
foodwebs*

epidemiological

others?



directed acyclic graph



directed graph

WWW

friendship?

flows of goods, information

economic exchange

dominance

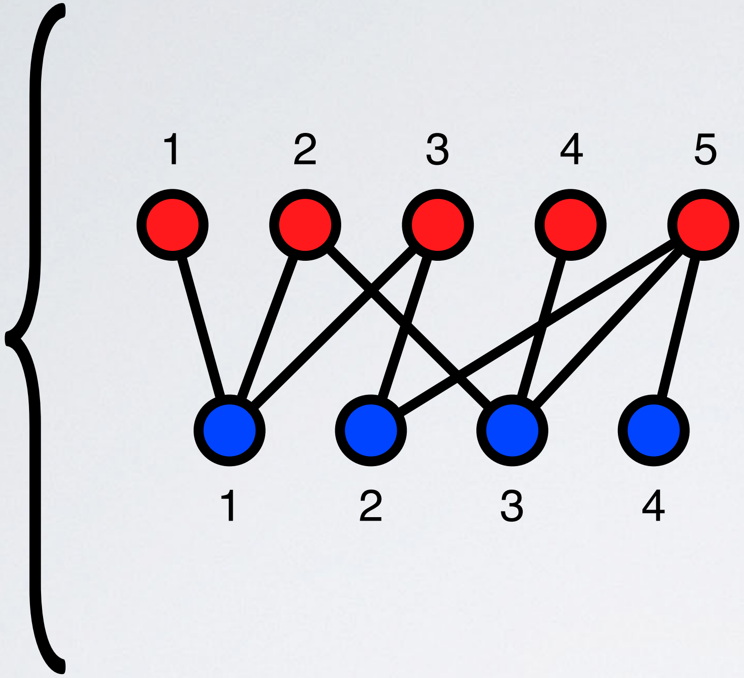
neuronal

transcription

time travelers

bipartite networks

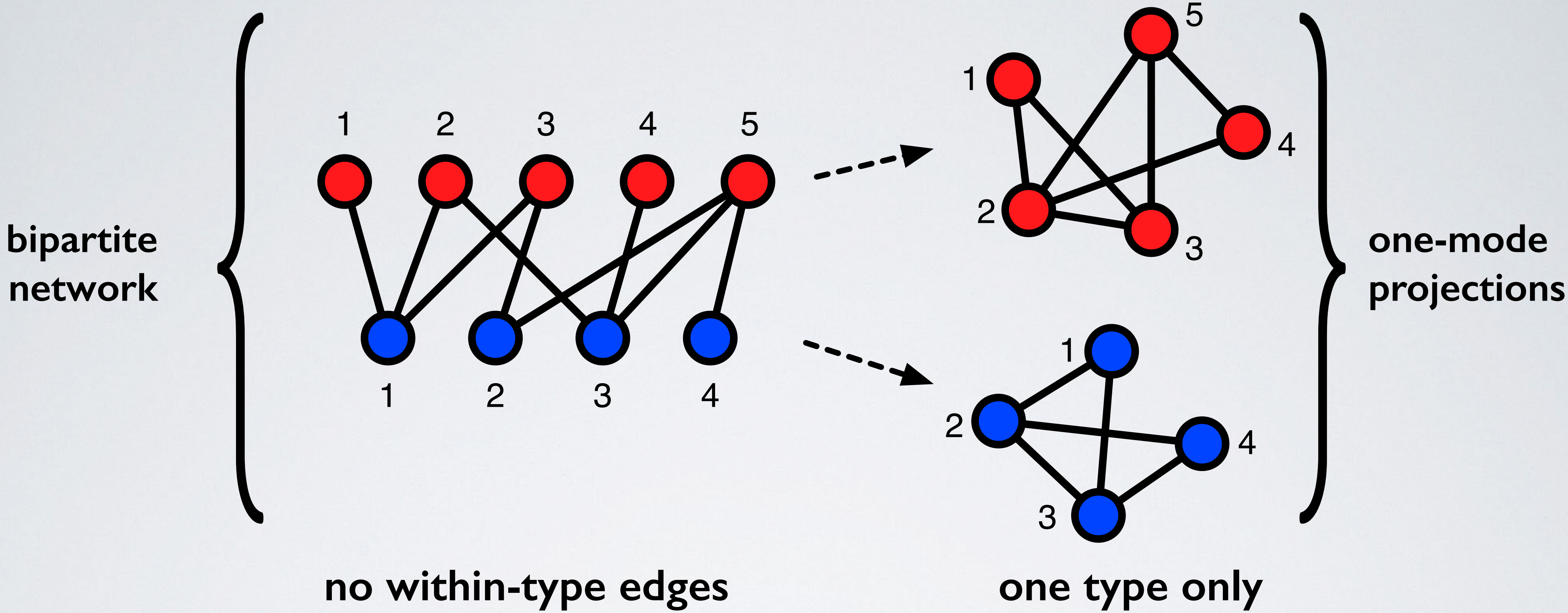
bipartite network



no within-type edges

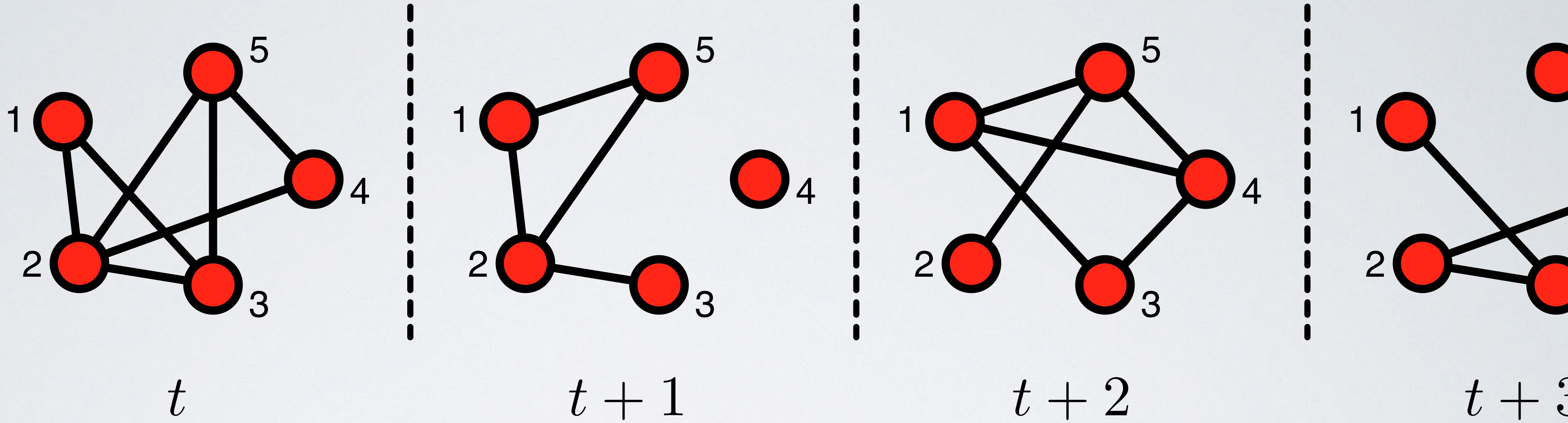
authors & papers	people & locations (checkins)
actors & movies/scenes	metabolites & reactions
musicians & albums	genes & substrings
people & online groups	words & documents
people & corporate boards	plants & pollinators

bipartite networks



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

temporal networks



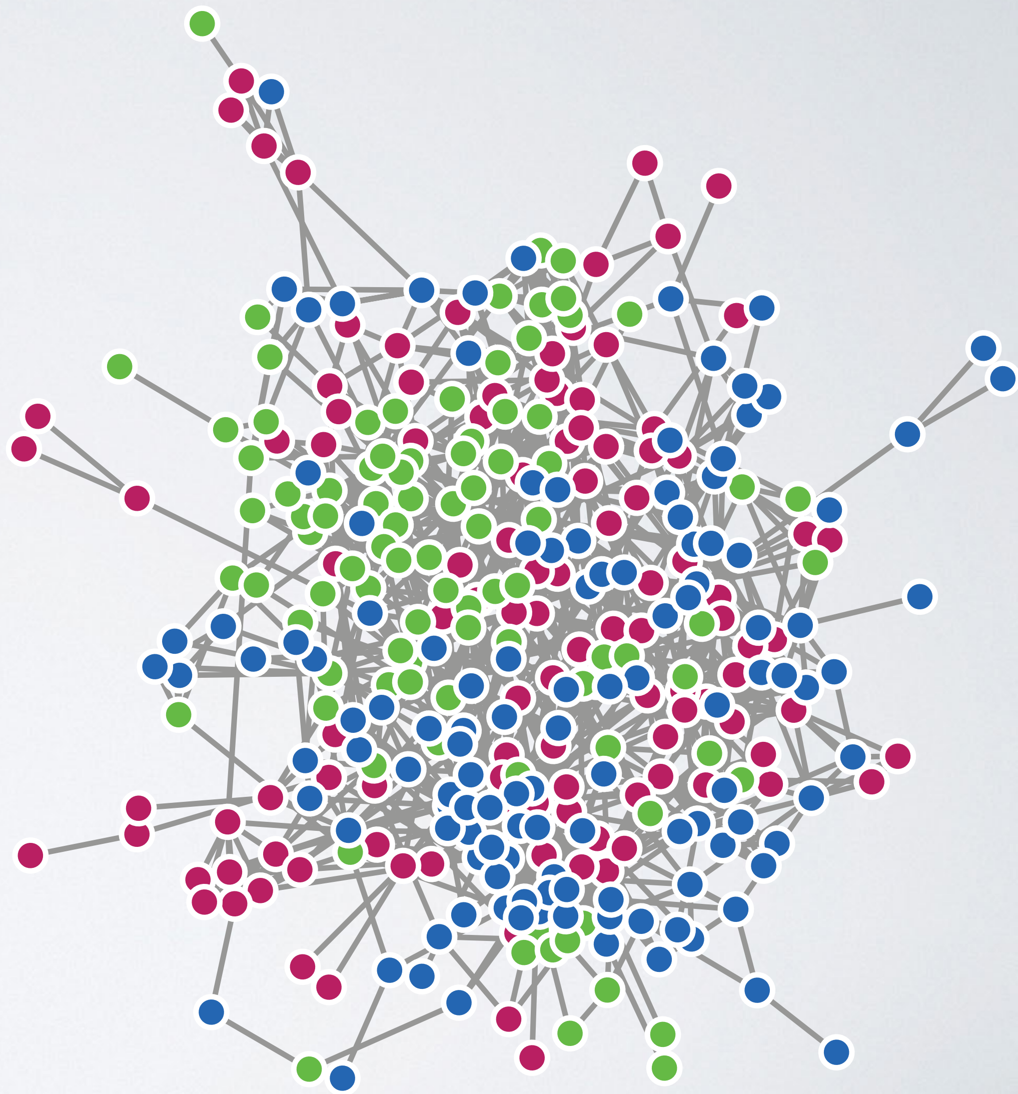
any network over time

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

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

describing networks

what networks look like



describing networks

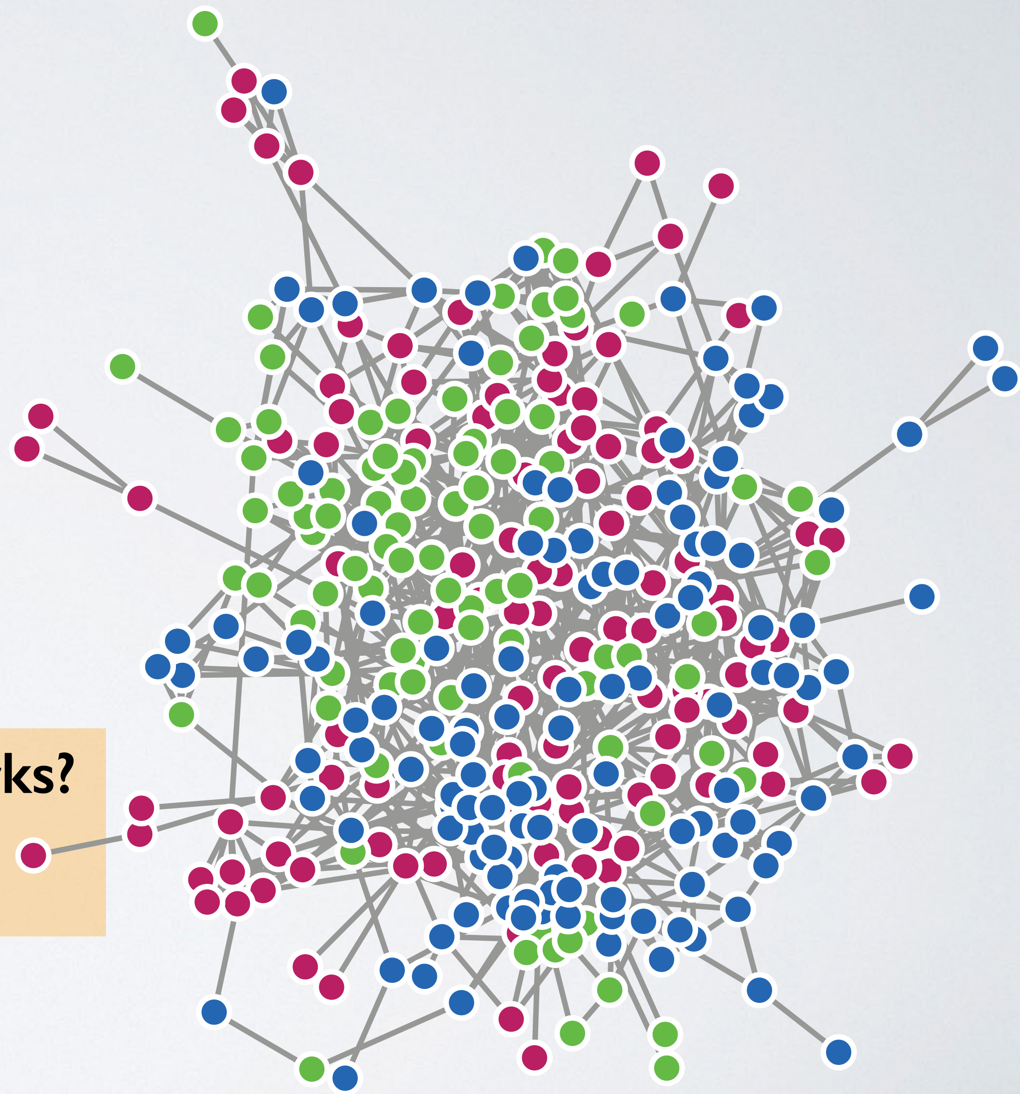
what networks look like

questions:

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

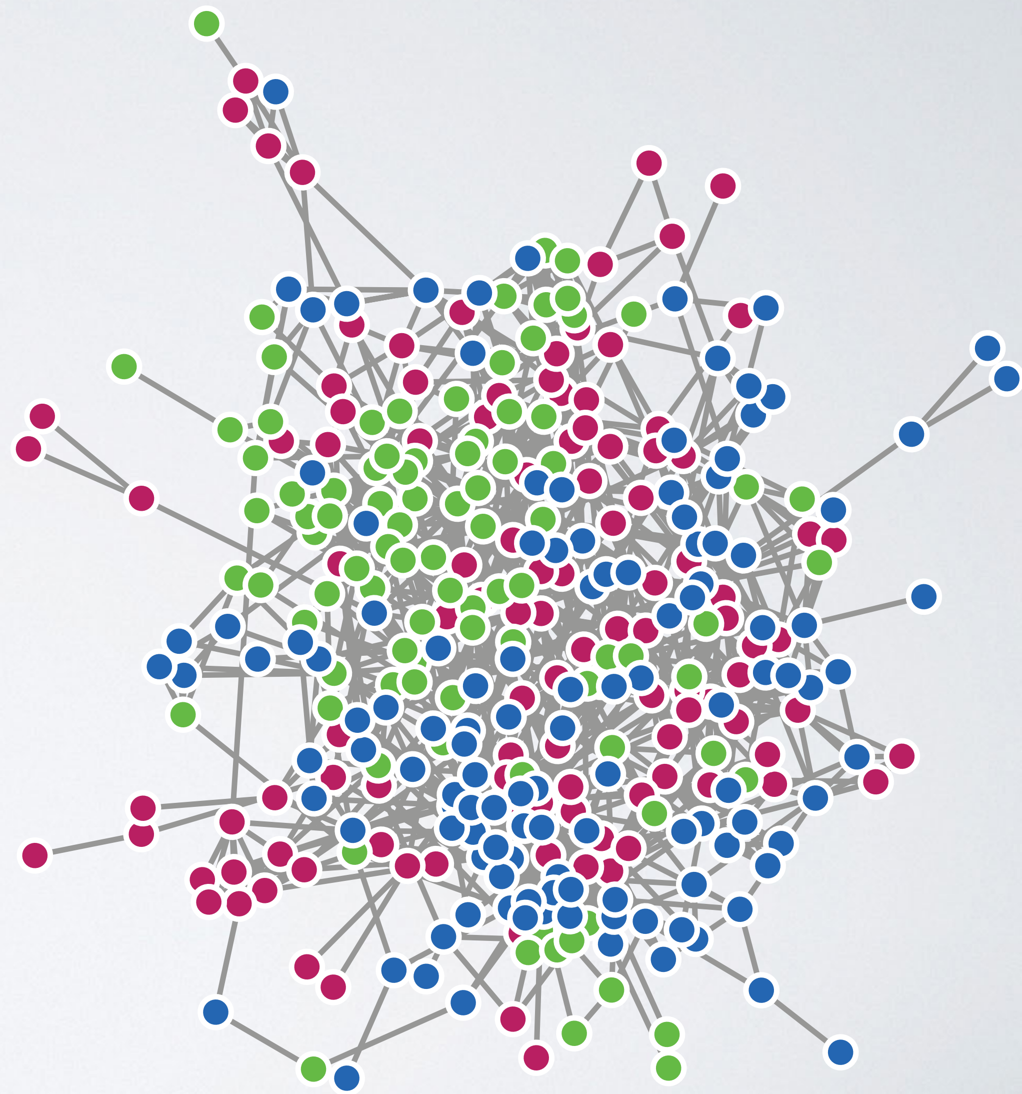
what we want to know

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



describing networks

a first step : **describe its features**

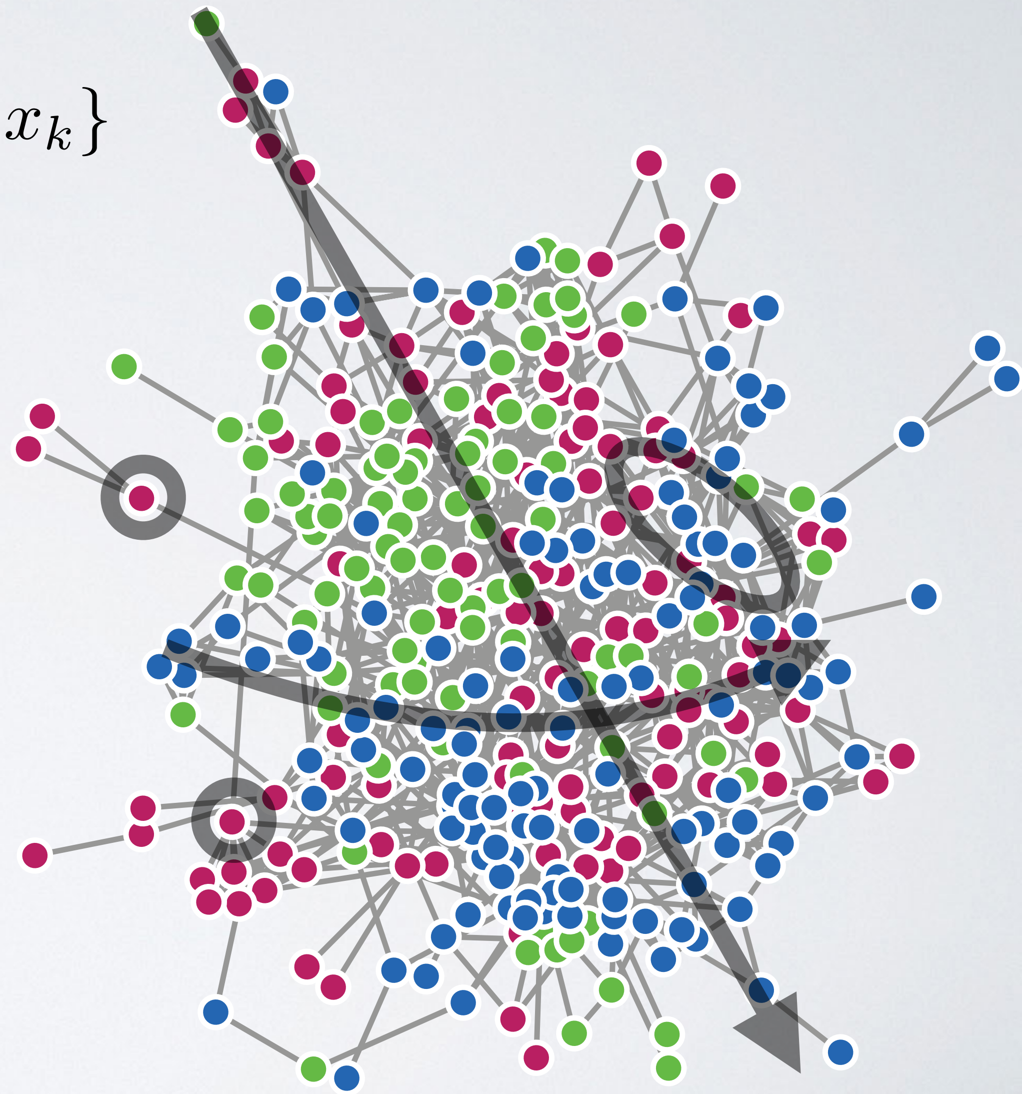


describing networks

a first step : **describe its features**

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

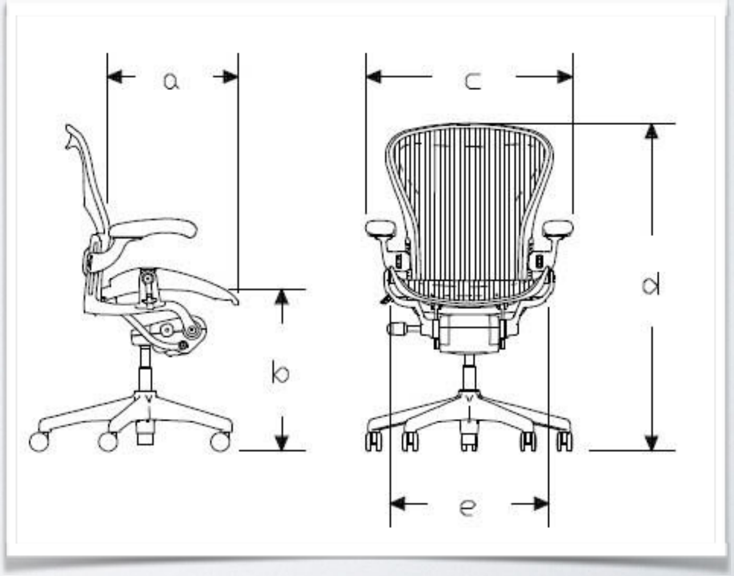
- **degree distributions**
- **short-loop density (triangles, etc.)**
- **shortest paths (diameter, etc.)**
- **vertex positions**
- **correlations between these**



describing networks

a first step : **describe its features**

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

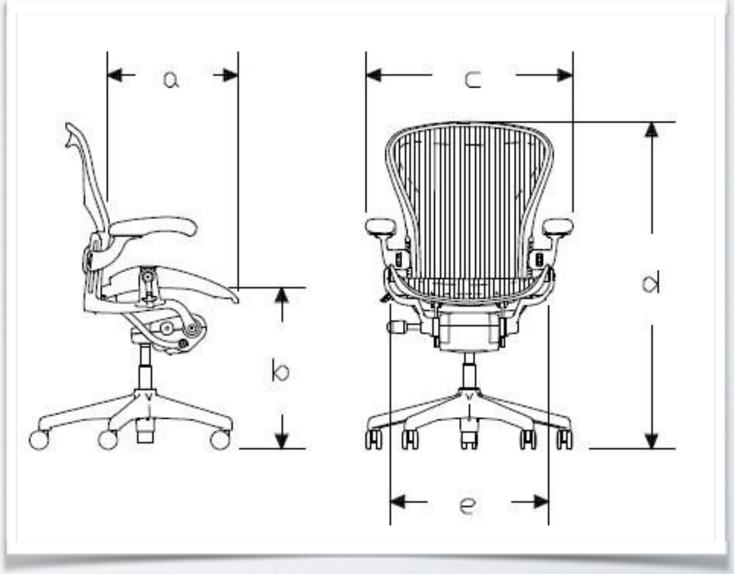


describing networks

a first step : **describe its features**

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

- **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**

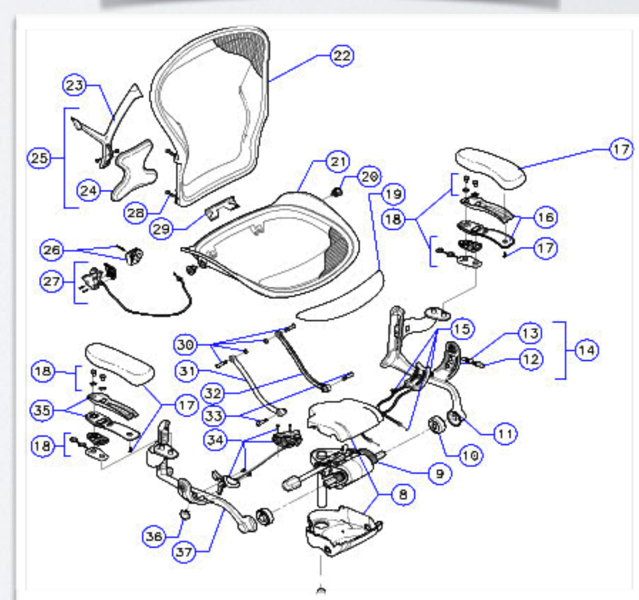
$$f : \text{object} \rightarrow \{\theta_1, \ldots, \theta_k\}$$

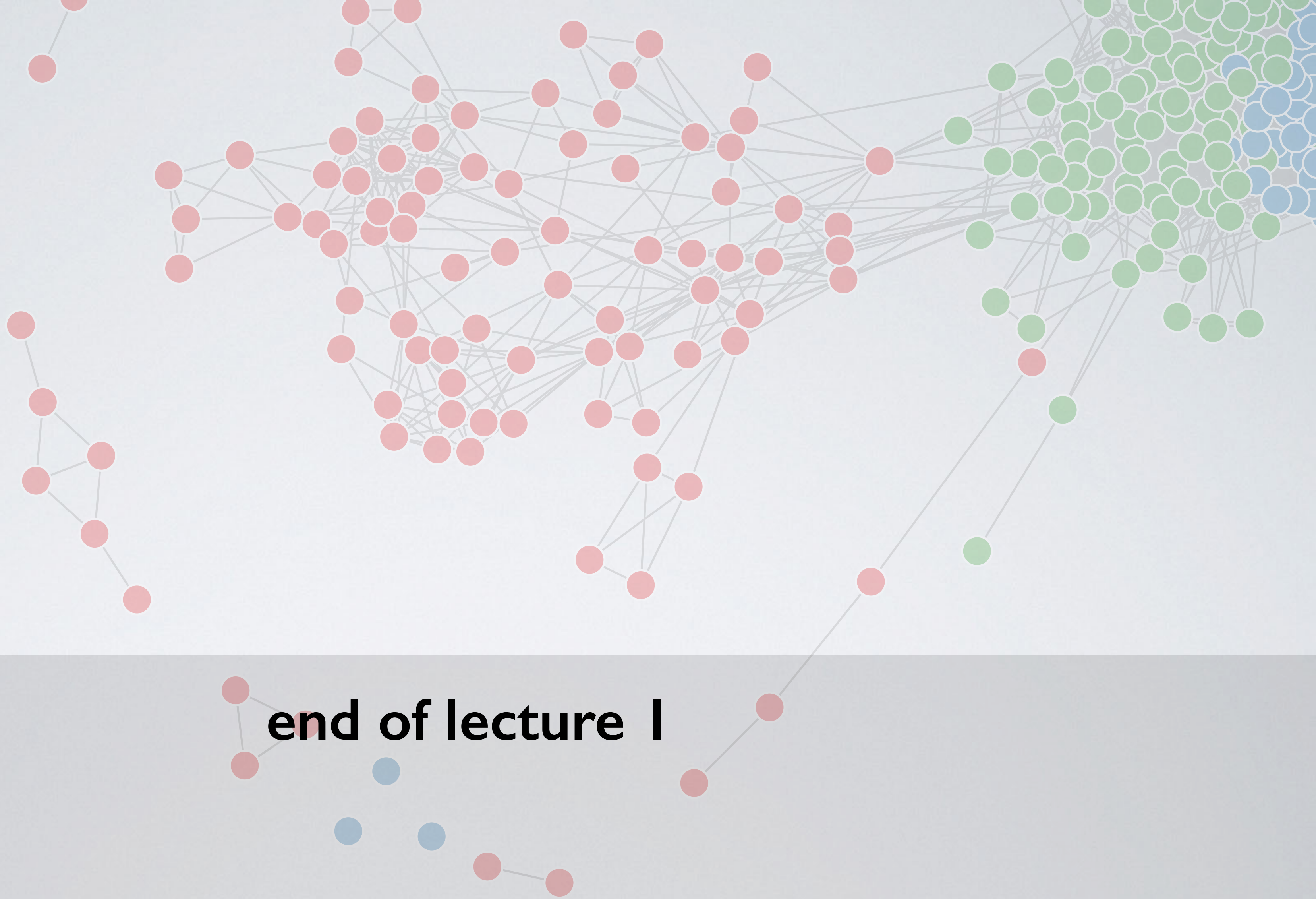
- **what are the fundamental parts?**
- **how are these parts organized?**
- **where are the degrees of freedom** $\vec{\theta}$ **?**
- **how can we define an abstract class?**
- **structure — dynamics — function?**

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

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

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





end of lecture 1

selected references

- The structure and function of complex networks. M. E. J. Newman, *SIAM Review* **45**, 167–256 (2003).
- *The Structure and Dynamics of Networks*. M. E. J. Newman, A.-L. Barabási, and D. J. Watts, Princeton University Press (2006).
- Hierarchical structure and the prediction of missing links in networks. A. Clauset, C. Moore, and M. E. J. Newman, *Nature* **453**, 98–101 (2008).
- Modularity and community structure in networks. M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **103**, 8577–8582 (2006).
- Why social networks are different from other types of networks. M. E. J. Newman and J. Park, *Phys. Rev. E* **68**, 036122 (2003)
- Random graphs with arbitrary degree distributions and their applications. M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
- Comparing community structure identification. L. Danon, A. Diaz-Guilera, J. Duch and A. Arenas. *J. Stat. Mech.* P09008 (2005).
- Characterization of Complex Networks: A Survey of measurements. L. daF. Costa, F. A. Rodrigues, G. Travieso and P. R. VillasBoas. arxiv:cond-mat/050585 (2005).
- Evolution in Networks. S.N. Dorogovtsev and J. F. F. Mendes. *Adv. Phys.* **51**, 1079 (2002).
- Revisting "scale-free" networks. E. F. Keller. *BioEssays* **27**, 1060-1068 (2005).
- Currency metabolites and network representations of metabolism. P. Holme and M. Huss. arxiv:0806.2763 (2008).
- Functional cartography of complex metabolic networks. R. Guimera and L. A. N. Amaral. *Nature* **433**, 895 (2005).
- Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. J. Leskovec, J. Kleinberg and C. Faloutsos. *Proc. 11th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining* 2005.
- The Structure of the Web. J. Kleinberg and S. Lawrence. *Science* 294, 1849 (2001).
- Navigation in a Small World. J. Kleinberg. *Nature* 406 (2000), 845.
- Towards a Theory of Scale-Free Graphs: Definitions, Properties and Implications. L. Li, D. Alderson, J. Doyle, and W. Willinger. *Internet Mathematics 2(4), 2006*.
- A First-Principles Approach to Understanding the Internet's Router-Level Topology. L. Li, D. Alderson, W. Willinger, and J. Doyle. *ACM SIGCOMM* 2004.
- Inferring network mechanisms: The Drosophila melanogaster protein interaction network. M. Middendorf, E. Ziv and C. H. Wiggins. *Proc. Natl. Acad. Sci. USA* 102, 3192 (2005).
- Robustness Can Evolve Gradually in Complex Regulatory Gene Networks with Varying Topology. S. Ciliberti, O. C. Martin and A. Wagner. *PLoS Comp. Bio*. 3, e15 (2007).
- Simple rules yield complex food webs. R. J. Williams and N. D. Martinez. *Nature* 404, 180 (2000).
- A network analysis of committees in the U.S. House of Representatives. M. A. Porter, P. J. Mucha, M. E. J. Newman and C. M. Warmbrand. *Proc. Natl. Acad. Sci. USA* 102, 7057 (2005).
- On the Robustness of Centrality Measures under Conditions of Imperfect Data. S. P. Borgatti, K. M. Carley and D. Krackhardt. *Social Networks* **28**, 124 (2006).