

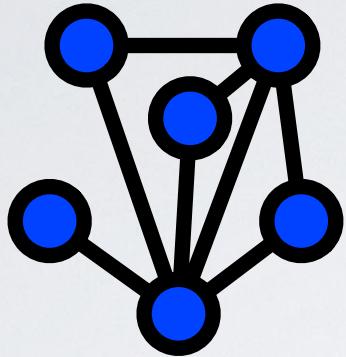
Network Analysis and Modeling

Aaron Clauset

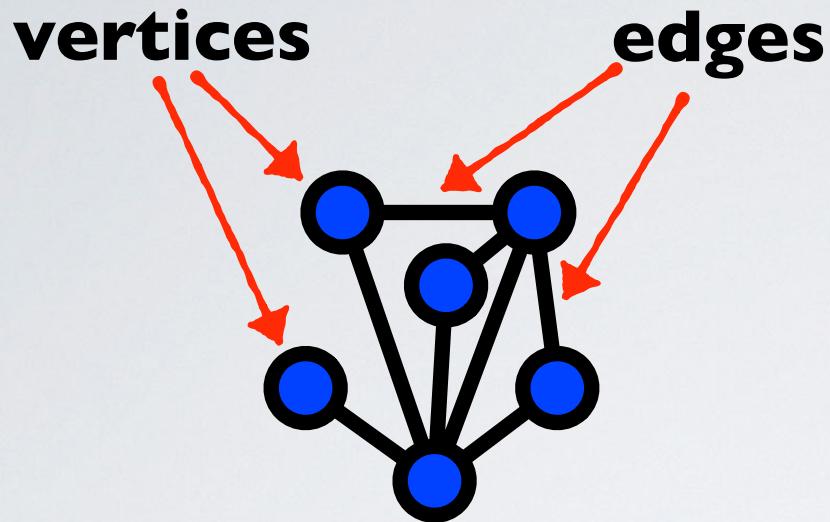
 @aaronclauset

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?



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



what is a vertex?

V distinct objects (vertices / nodes / actors)

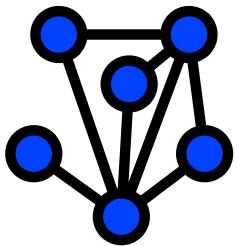
when are two vertices connected?

$$E \subseteq V \times V$$

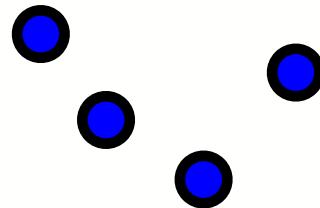
pairwise relations (edges / links / ties)

6 major classes of networks

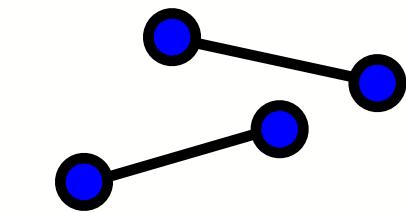
- technological
- information
- transportation
- social
- biological
- economic



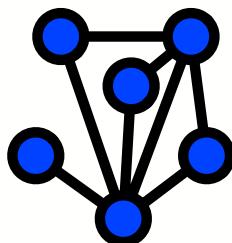
network



vertex



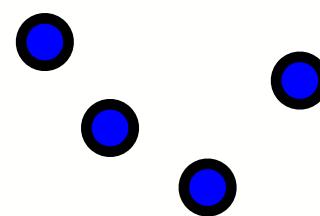
edge



network

Internet(1)

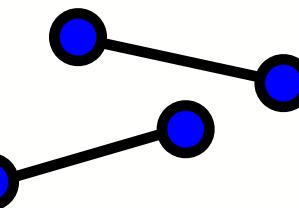
Internet(2)



vertex

computer

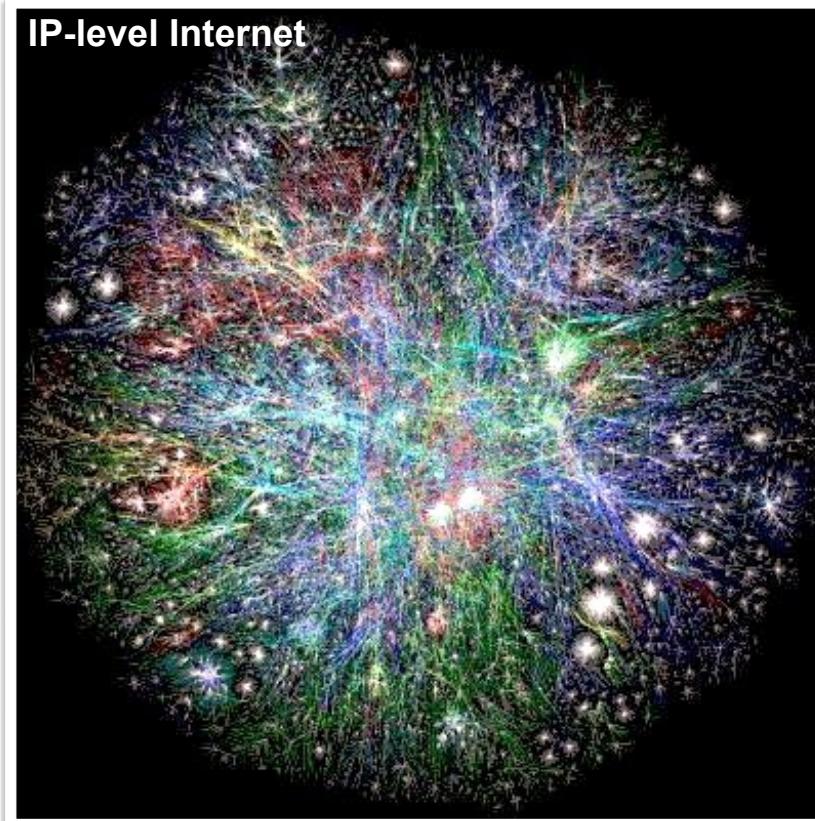
autonomous system (ISP)



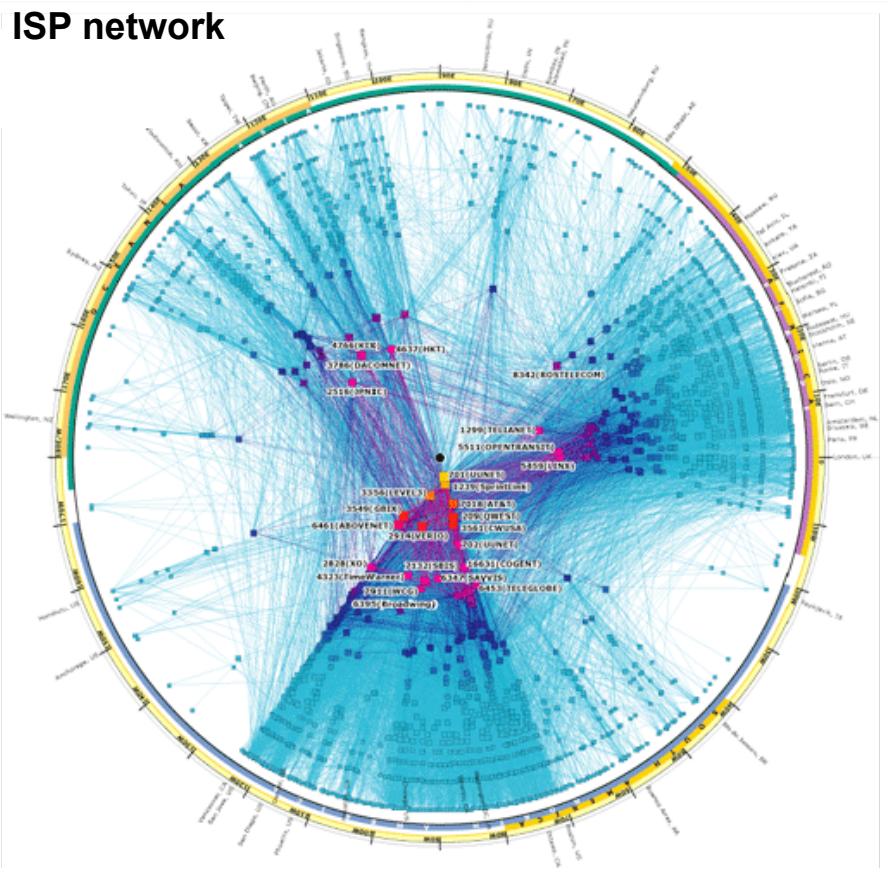
edge

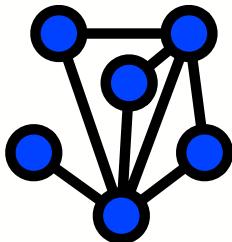
IP network adjacency

BGP connection



ISP network



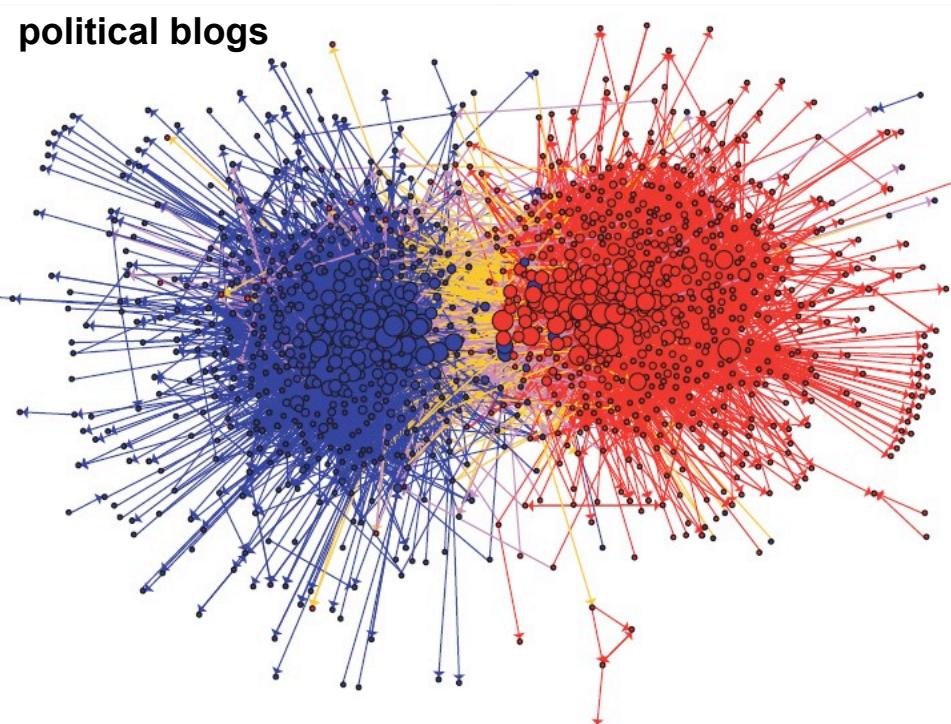


network

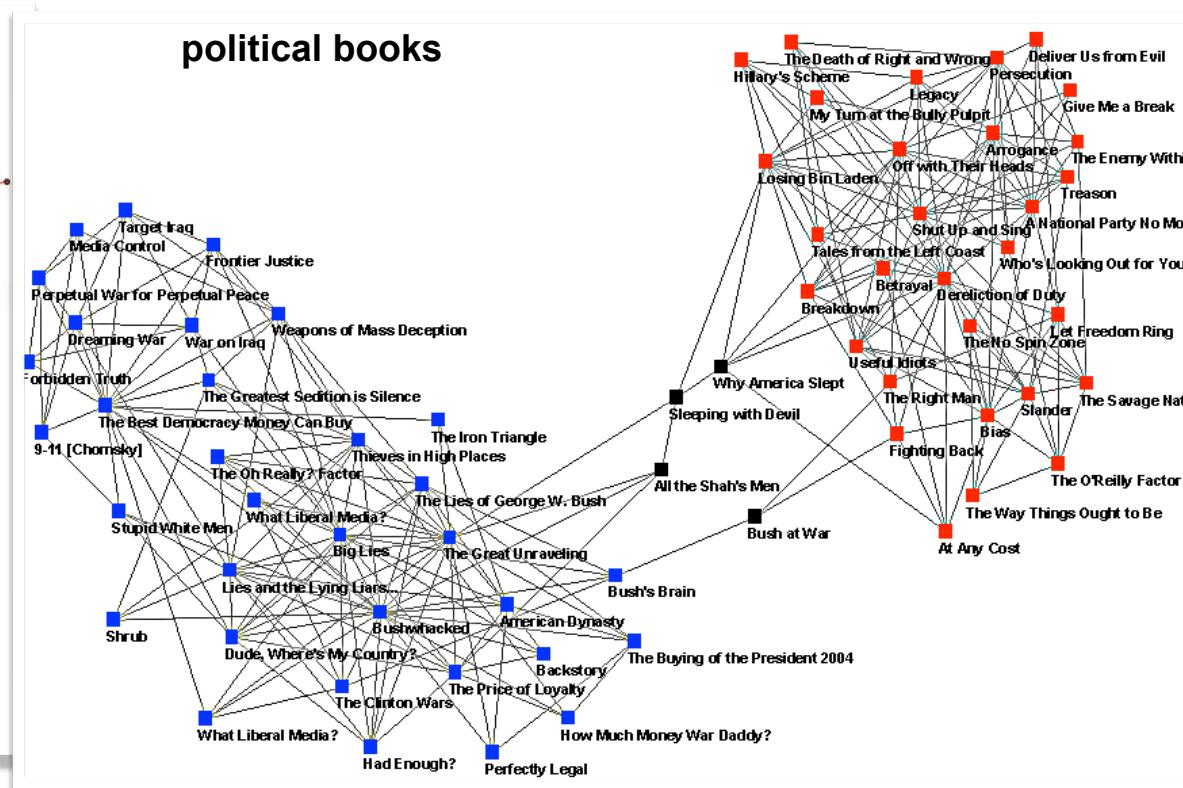
Internet(1)	computer	IP network adjacency
Internet(2)	autonomous system (ISP)	BGP connection
software	function	function call
World Wide Web	web page	hyperlink
documents	article, patent, or legal case	citation

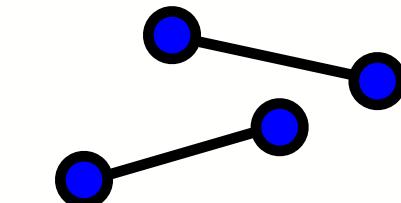
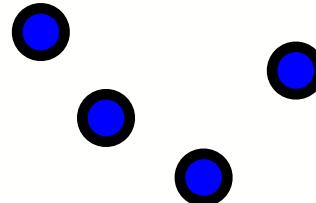
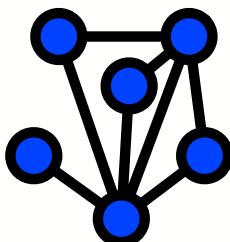


vertex



political books





network

vertex

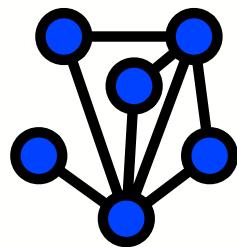
edge

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

technological

information

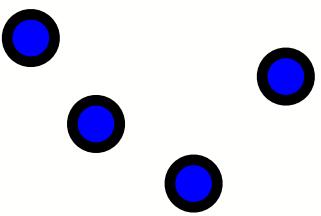
transportation



network

road network(1)

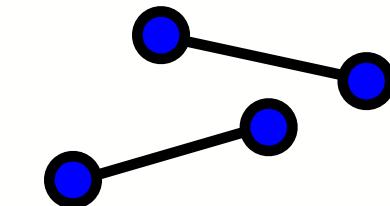
road network(2)



vertex

intersection

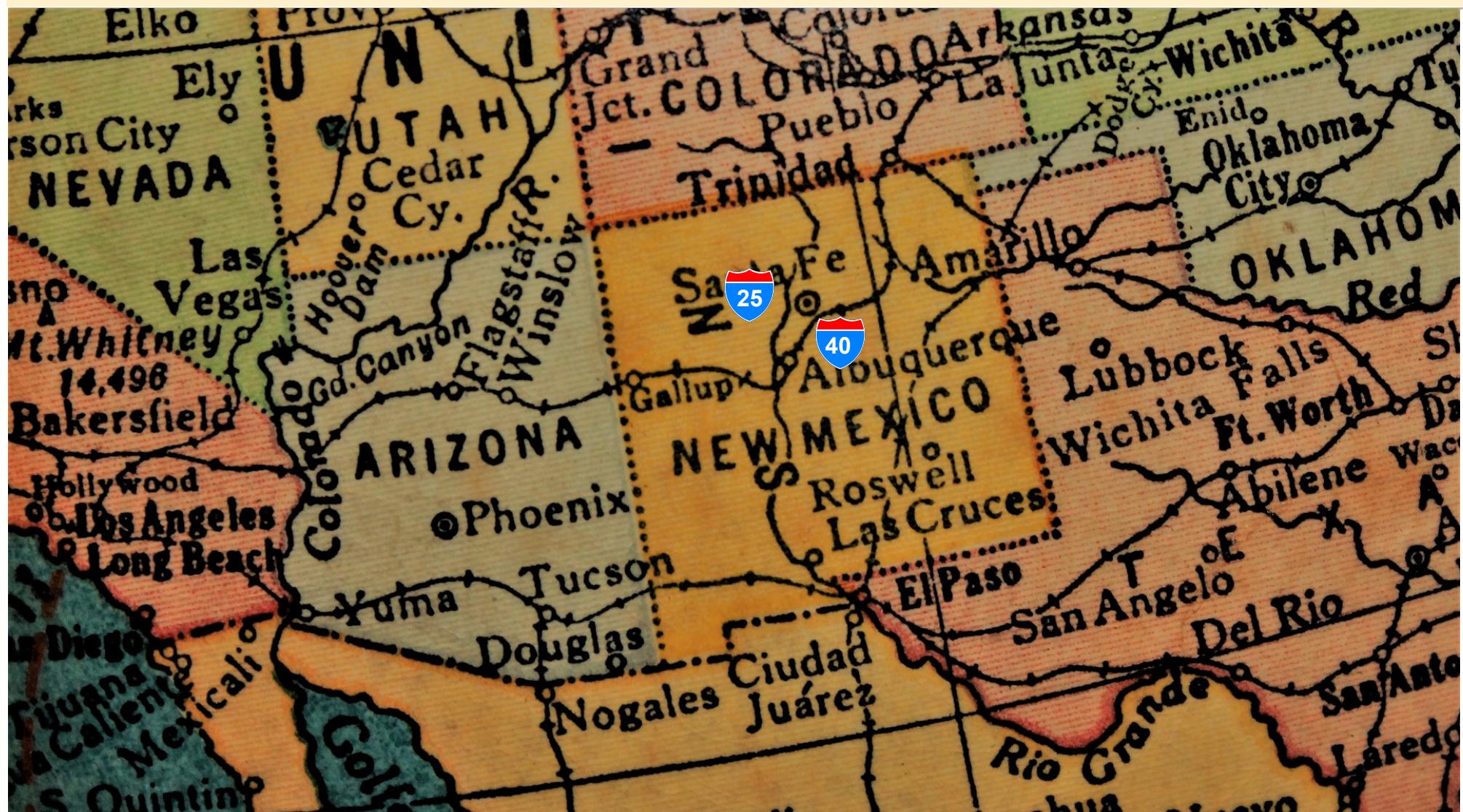
named road

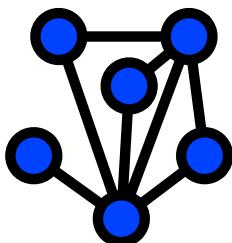


edge

pavement

intersection





network

road network(1)

vertex

road network(2)

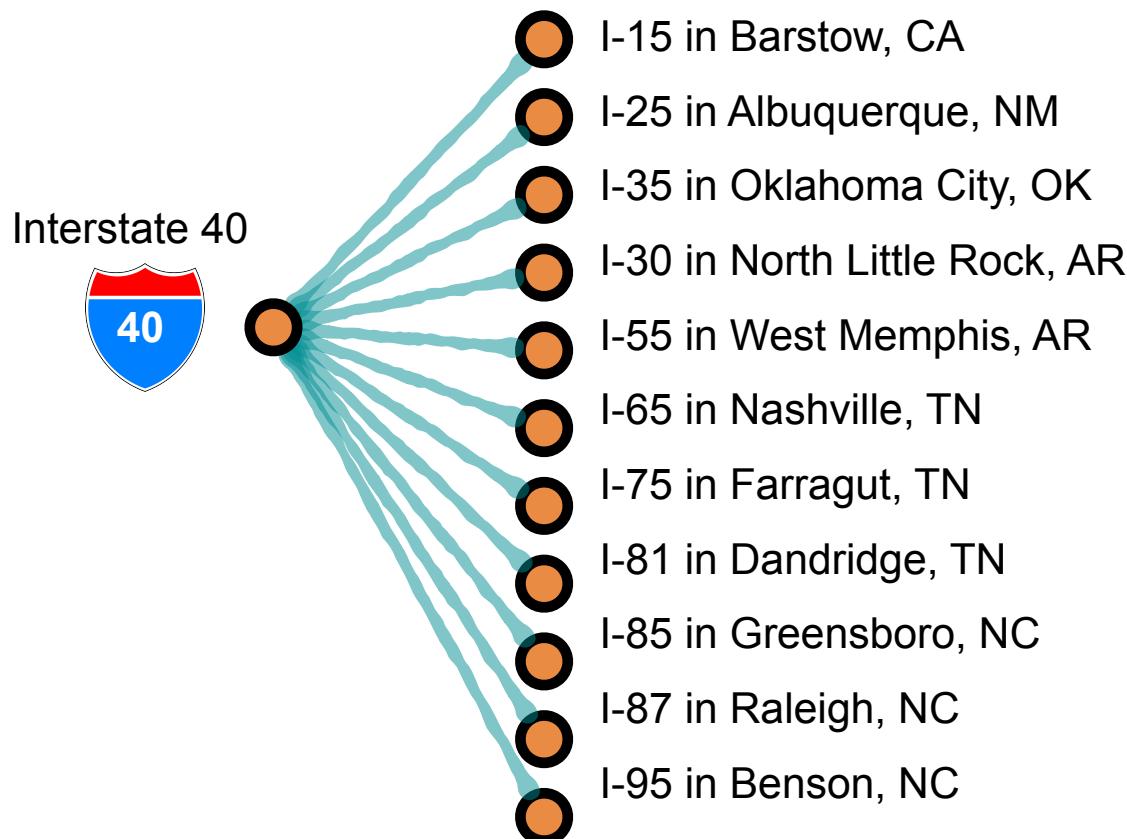
intersection

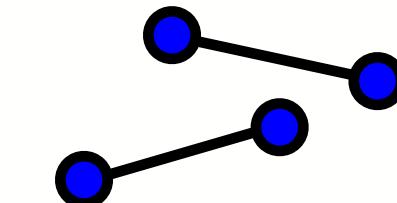
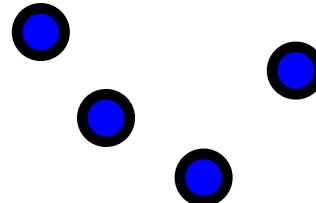
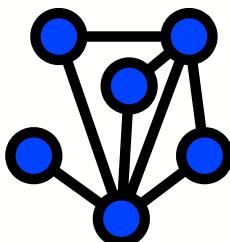
named road

edge

pavement

intersection





network

vertex

edge

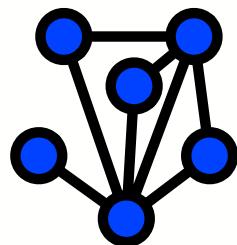
Internet(1)	computer	IP network adjacency
Internet(2)	autonomous system (ISP)	BGP connection
software	function	function call
World Wide Web	web page	hyperlink
documents	article, patent, or legal case	citation
power grid transmission	generating or relay station	transmission line
rail system	rail station	railroad tracks
road network(1)	intersection	pavement
road network(2)	named road	intersection
airport network	airport	non-stop flight
friendship network	person	friendship
sexual network	person	intercourse

technological

information

transportation

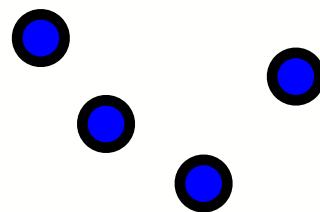
social



network

friendship network

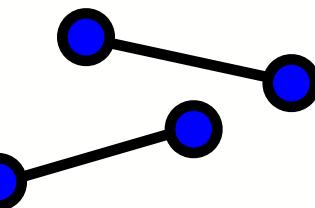
sexual network



vertex

person

person

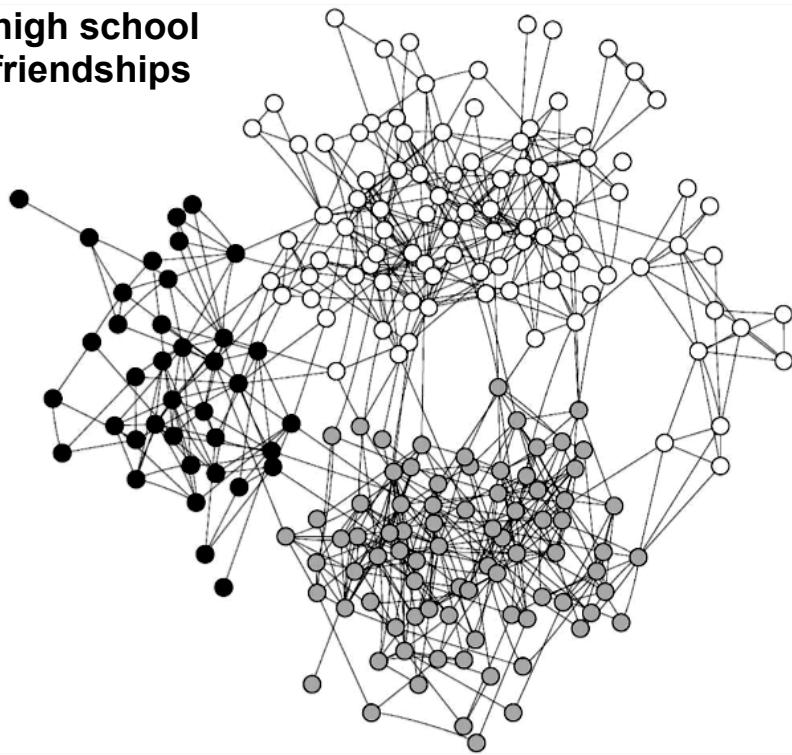


edge

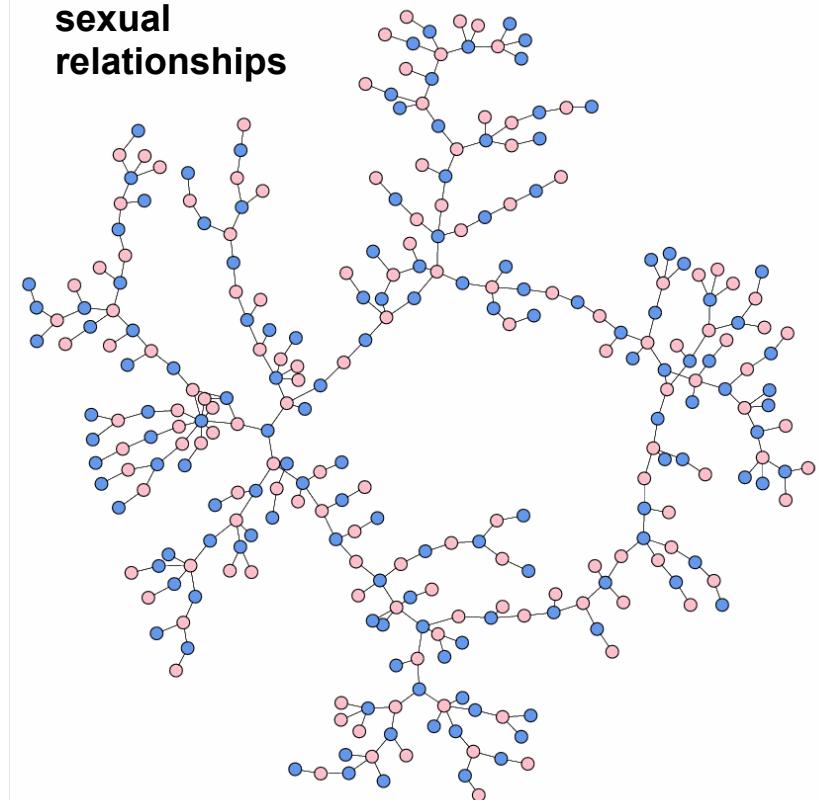
friendship

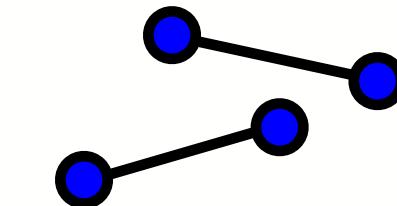
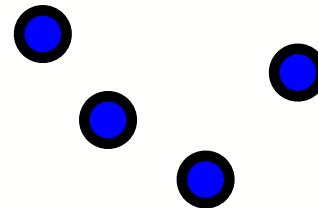
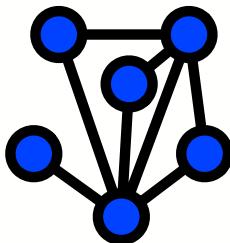
intercourse

**high school
friendships**

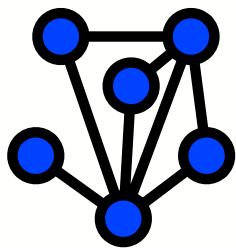


**sexual
relationships**





	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	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)	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

metabolic network

food web

vertex

metabolite

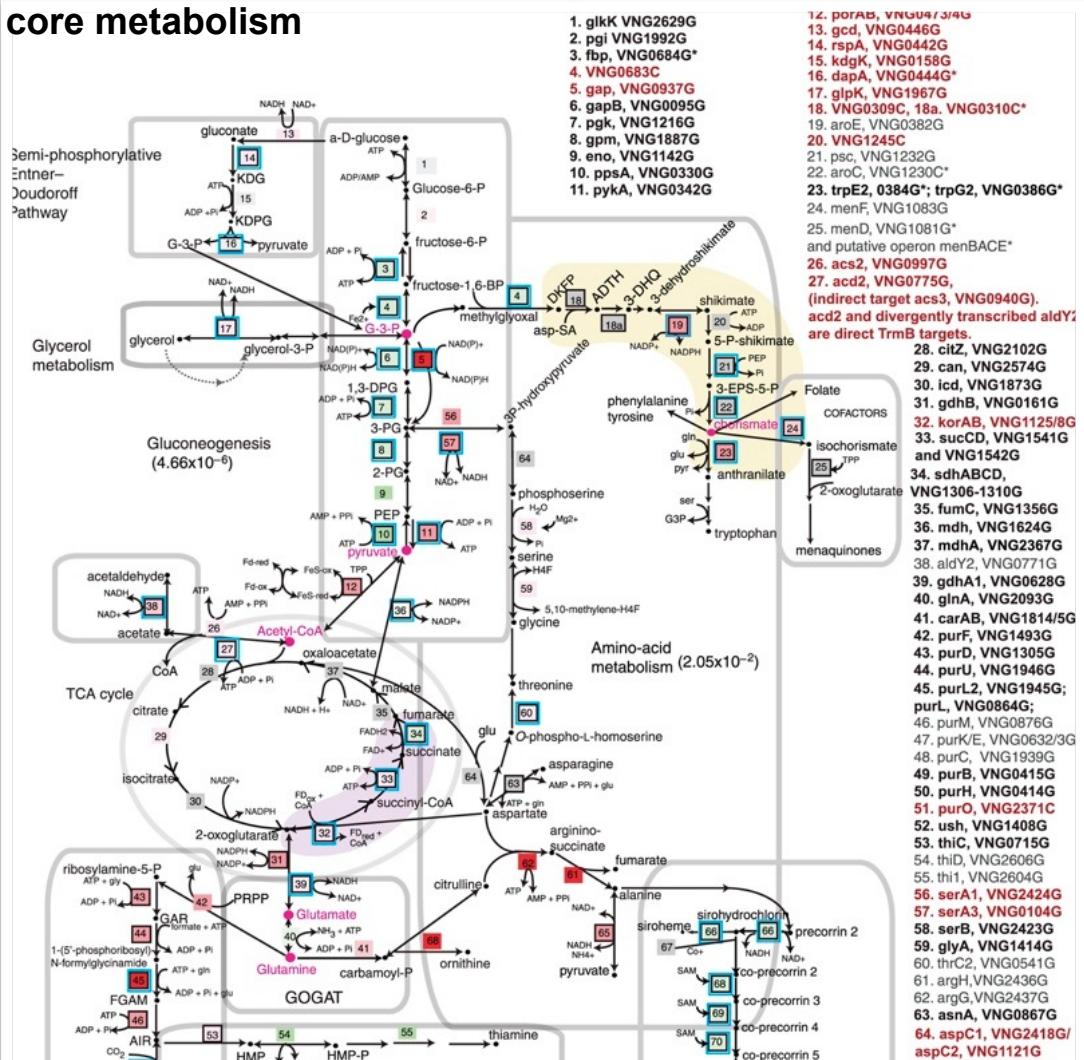
species

edge

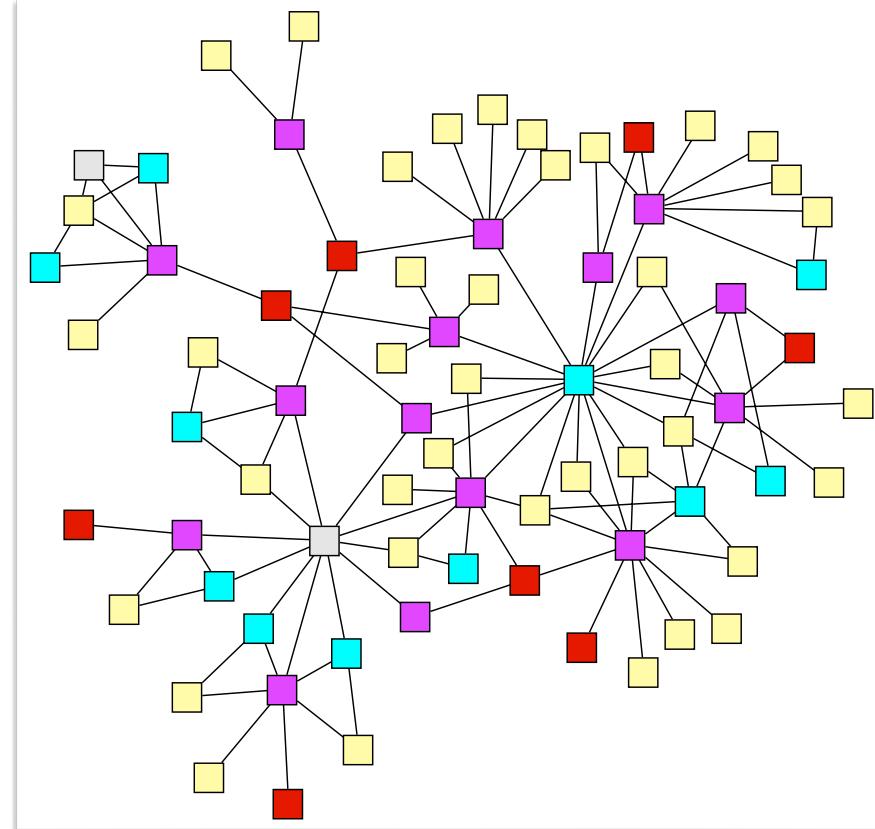
metabolic reaction

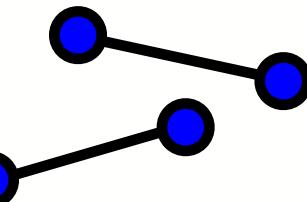
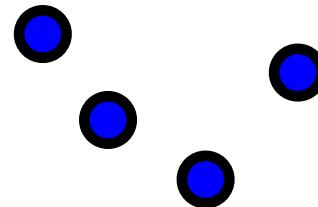
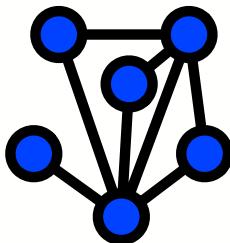
predation or resource transfer

core metabolism



grassland foodweb



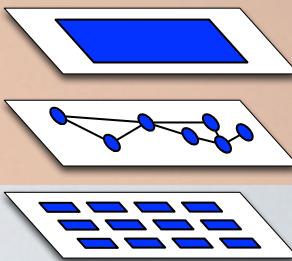


	network	vertex	edge
technological	Internet(1)	computer	IP network adjacency
	Internet(2)	autonomous system (ISP)	BGP connection
information	software	function	function call
	World Wide Web	web page	hyperlink
documents	documents	article, patent, or legal case	citation
	power grid transmission	generating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

what's a network?

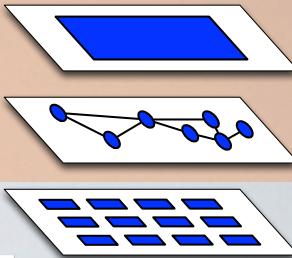
pop quiz

what's a network?



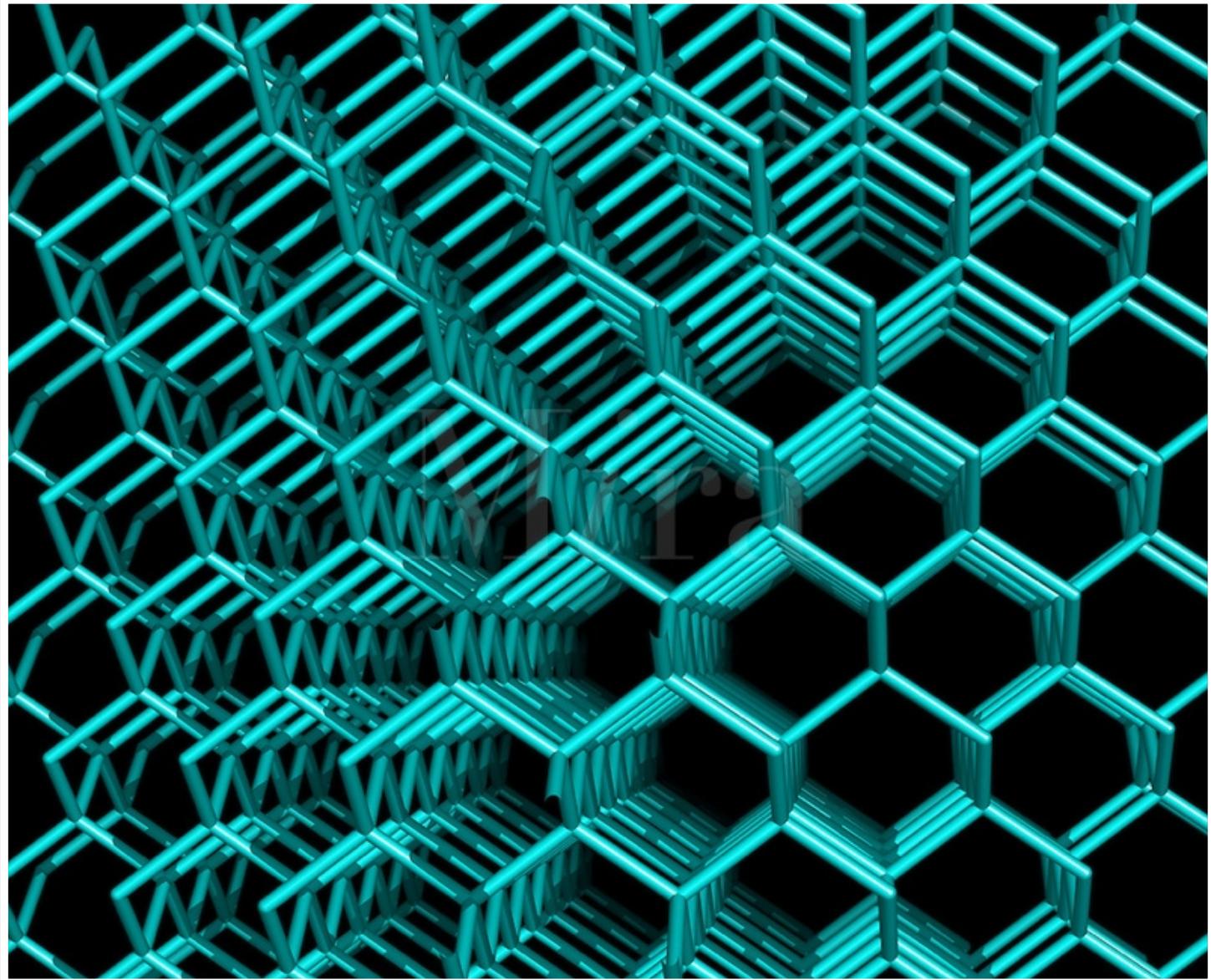
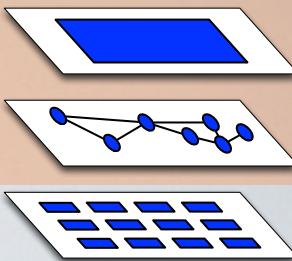
Andromeda galaxy

what's a network?



cauliflower fractal

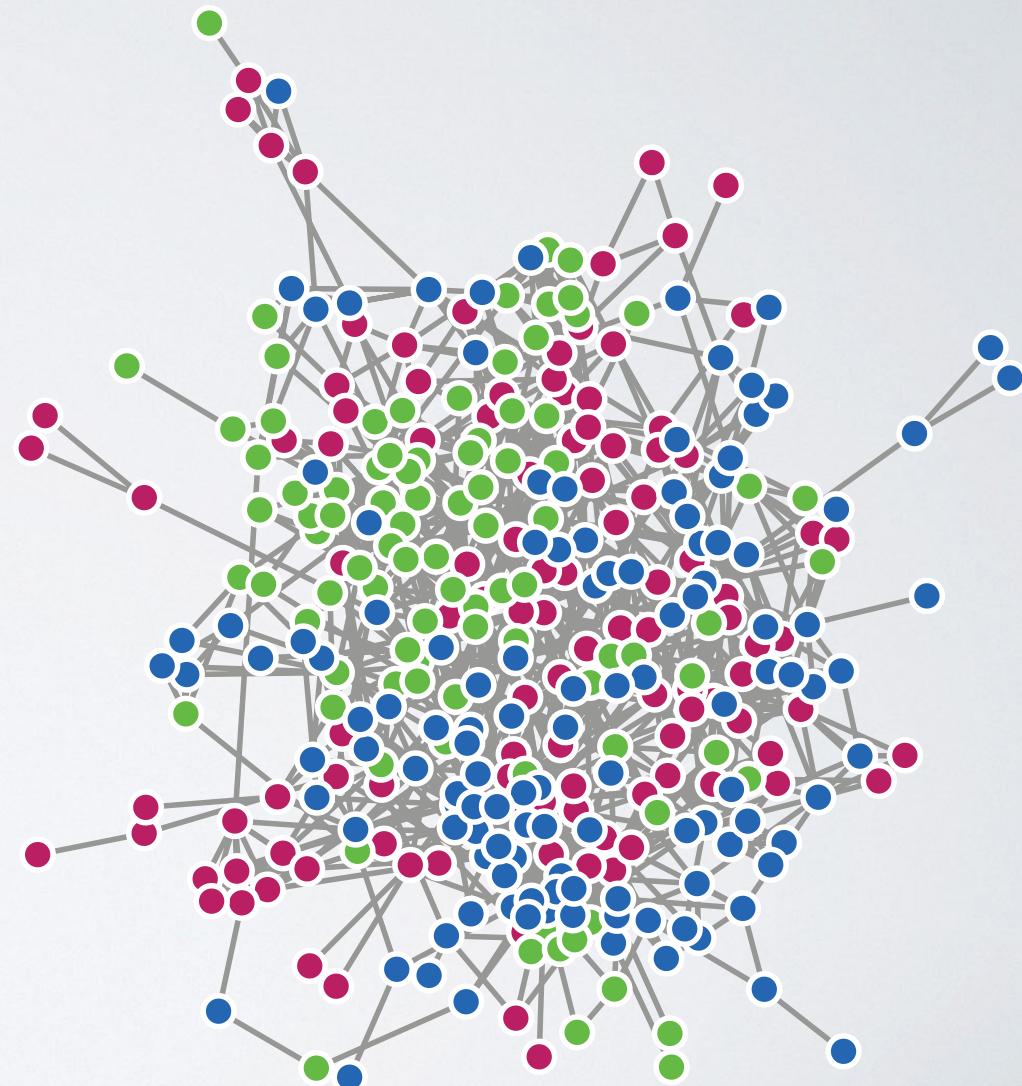
what's a network?



diamond lattice

describing networks

what networks look like



describing networks

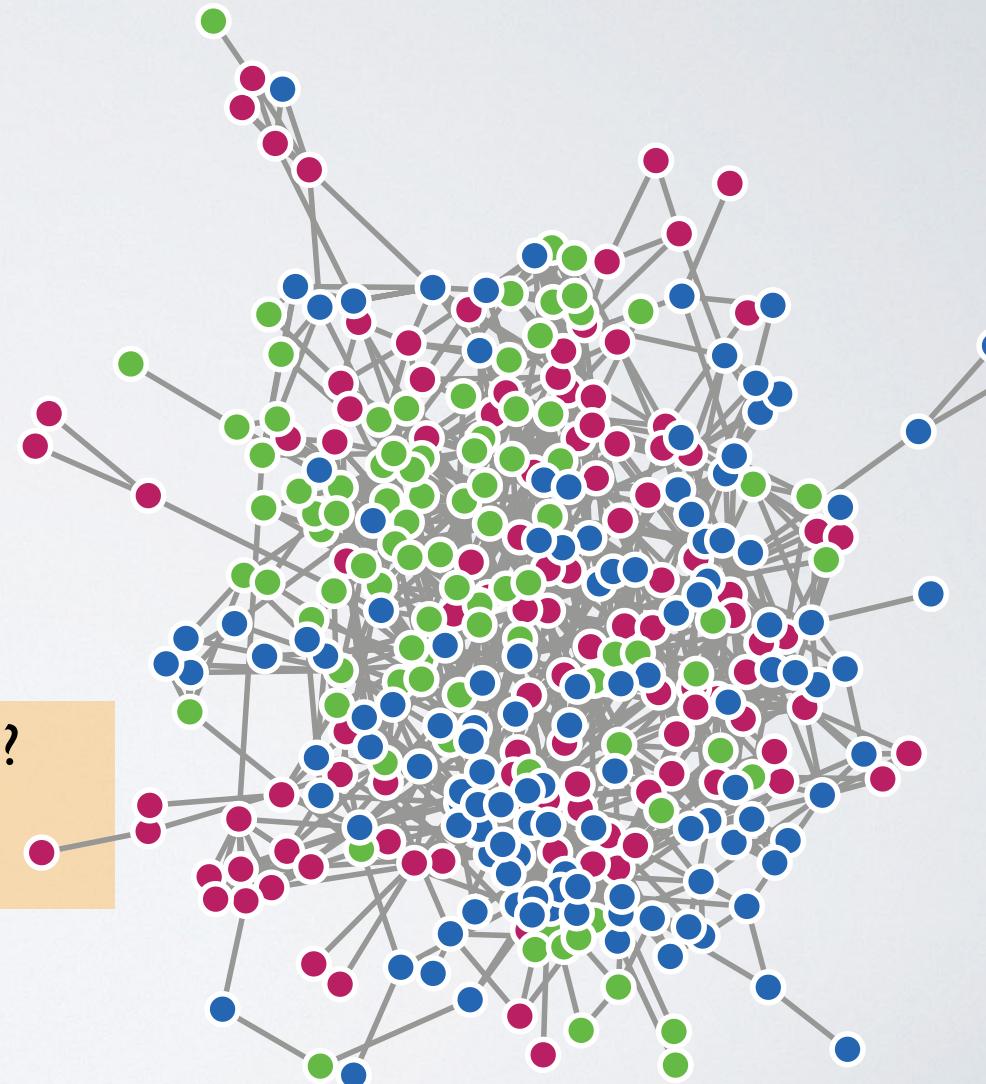
what networks look like

questions:

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

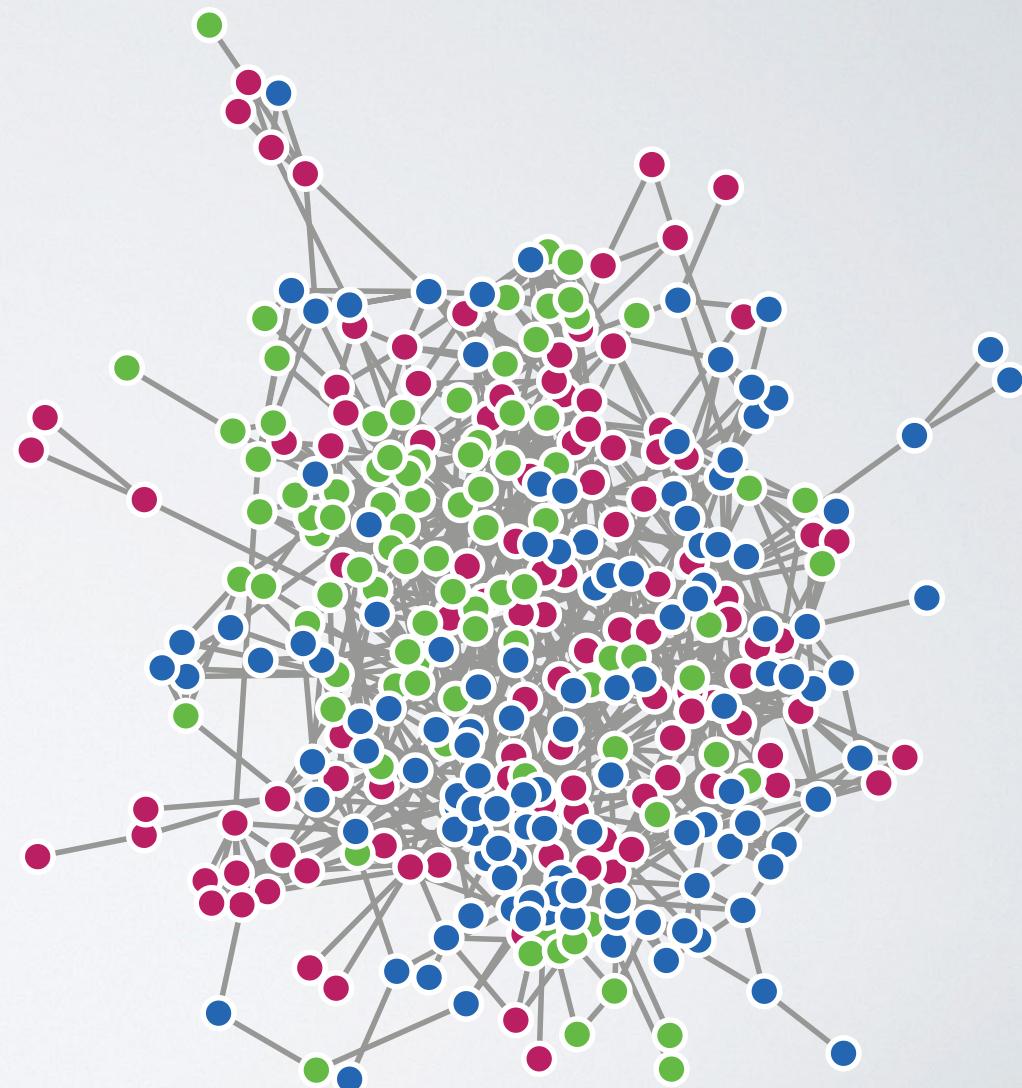
what we want to know

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



describing networks

a first step : describe its features

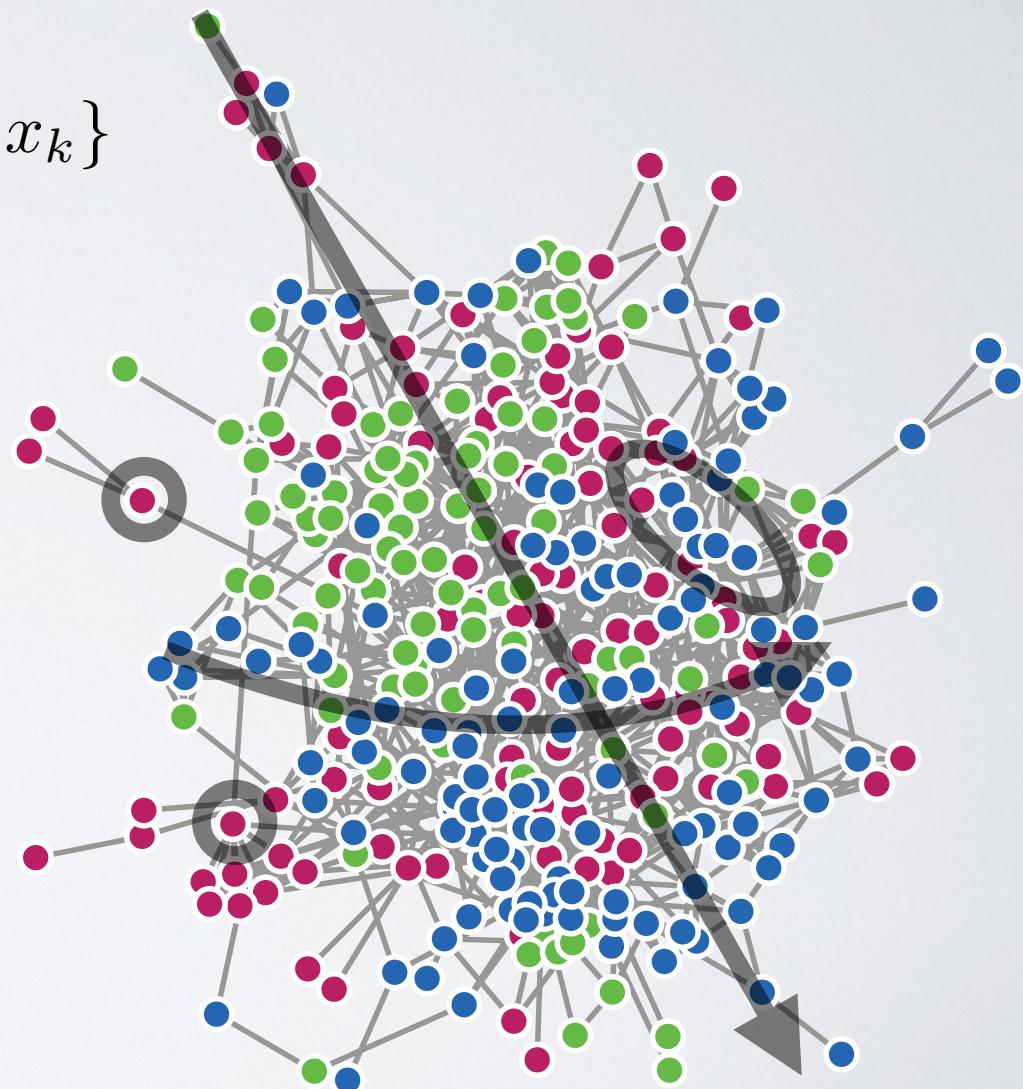


describing networks

a first step : describe its features

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

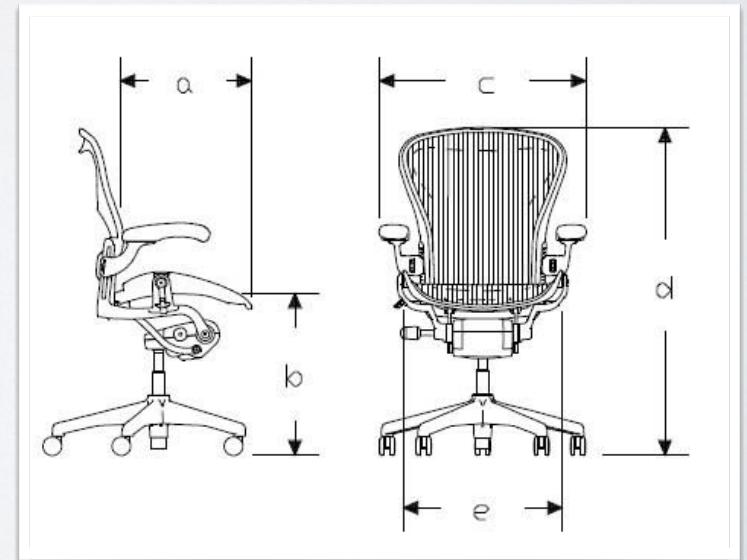
- degree distributions
- short-loop density (triangles, etc.)
- shortest paths (diameter, etc.)
- vertex positions
- correlations between these



describing networks

a first step : describe its features

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



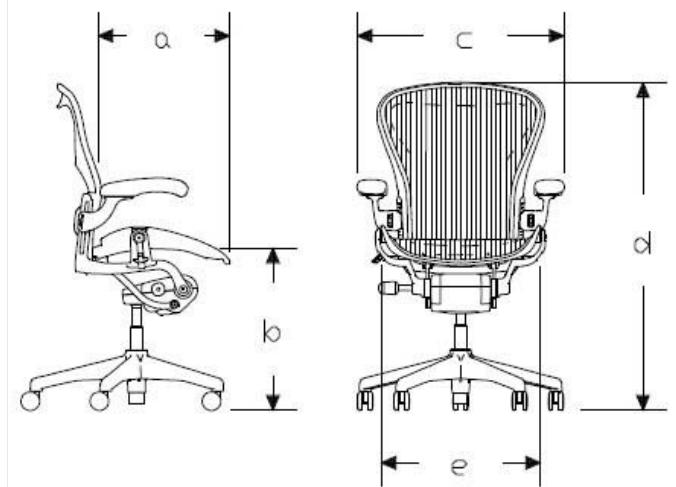
describing networks

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

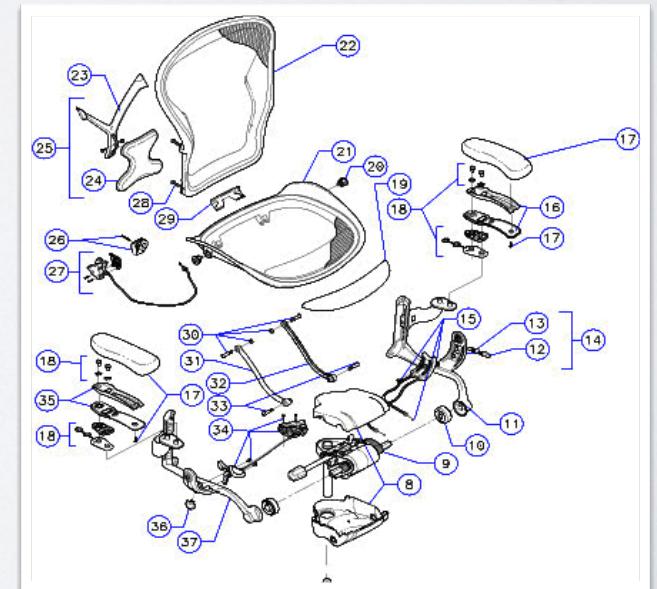
$$f : \text{object} \rightarrow \{\theta_1, \dots, \theta_k\}$$

- 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

6 major approaches

- I. **exploratory network analysis:** count & compare all the things (degree distributions, centrality scores, community detection, etc.)

analyzing networks

6 major approaches

1. **exploratory network analysis:** count & compare all the things (degree distributions, centrality scores, community detection, etc.)
2. **explanatory network analysis:** convert network structure into node-level features, and do traditional explanatory modeling (eg, regression)

analyzing networks

6 major approaches

1. **exploratory network analysis:** count & compare all the things (degree distributions, centrality scores, community detection, etc.)
2. **explanatory network analysis:** convert network structure into node-level features, and do traditional explanatory modeling (eg, regression)
3. **null models:** use some kind of random graph to identify non-random patterns as deviations from the null

analyzing networks

6 major approaches

1. **exploratory network analysis:** count & compare all the things (degree distributions, centrality scores, community detection, etc.)
2. **explanatory network analysis:** convert network structure into node-level features, and do traditional explanatory modeling (eg, regression)
3. **null models:** use some kind of random graph to identify non-random patterns as deviations from the null
4. **mechanisms / simulations:** explain structural or dynamical patterns as caused by specific process, analyzed mathematically or simulated

analyzing networks

6 major approaches

1. **exploratory network analysis:** count & compare all the things (degree distributions, centrality scores, community detection, etc.)
2. **explanatory network analysis:** convert network structure into node-level features, and do traditional explanatory modeling (eg, regression)
3. **null models:** use some kind of random graph to identify non-random patterns as deviations from the null
4. **mechanisms / simulations:** explain structural or dynamical patterns as caused by specific process, analyzed mathematically or simulated
5. **predictive models:** fit parametric model of network structure & use it to predict missing or future data (edges, labels, etc.)

analyzing networks

6 major approaches

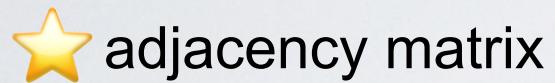
- ★ 1. **exploratory network analysis:** count & compare all the things (degree distributions, centrality scores, community detection, etc.)
- ★ 2. **explanatory network analysis:** convert network structure into node-level features, and do traditional explanatory modeling (eg, regression)
- ★ 3. **null models:** use some kind of random graph to identify non-random patterns as deviations from the null
- ★ 4. **mechanisms / simulations:** explain structural or dynamical patterns as caused by specific process, analyzed mathematically or simulated
- ★ 5. **predictive models:** fit parametric model of network structure & use it to predict missing or future data (edges, labels, etc.)
- ★ 6. **network experiments:** manipulate structure and measure node-level or graph-level behavior as function of changes (incl. causal inference)

representing networks

4 representations

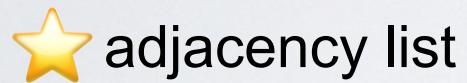
- ridiculogram

nice pictures, best for small networks



adjacency matrix

mathematically convenient & useful mental model



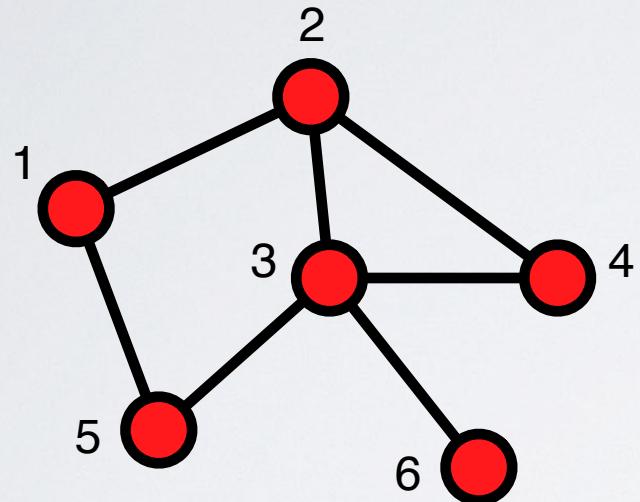
adjacency list

efficient computation

- edge list

efficient storage

a simple network

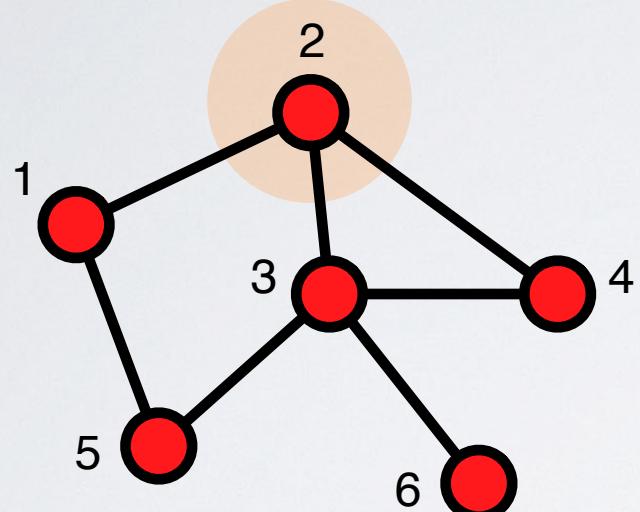


undirected

unweighted

no self-loops

a *simple* network



undirected

unweighted

no self-loops

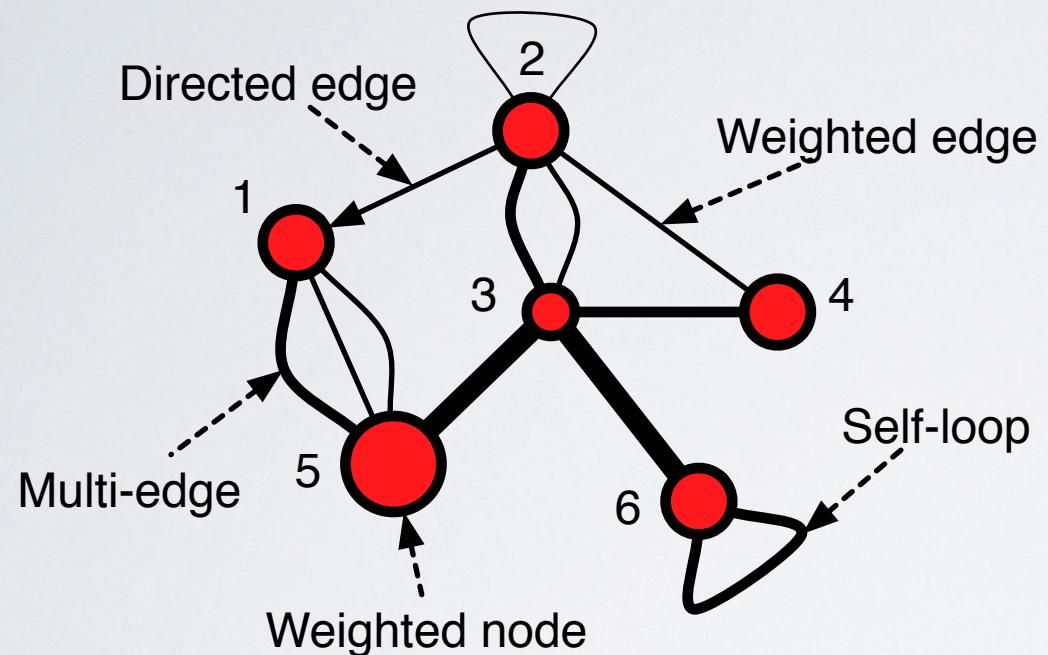
adjacency matrix

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

beyond simple graphs

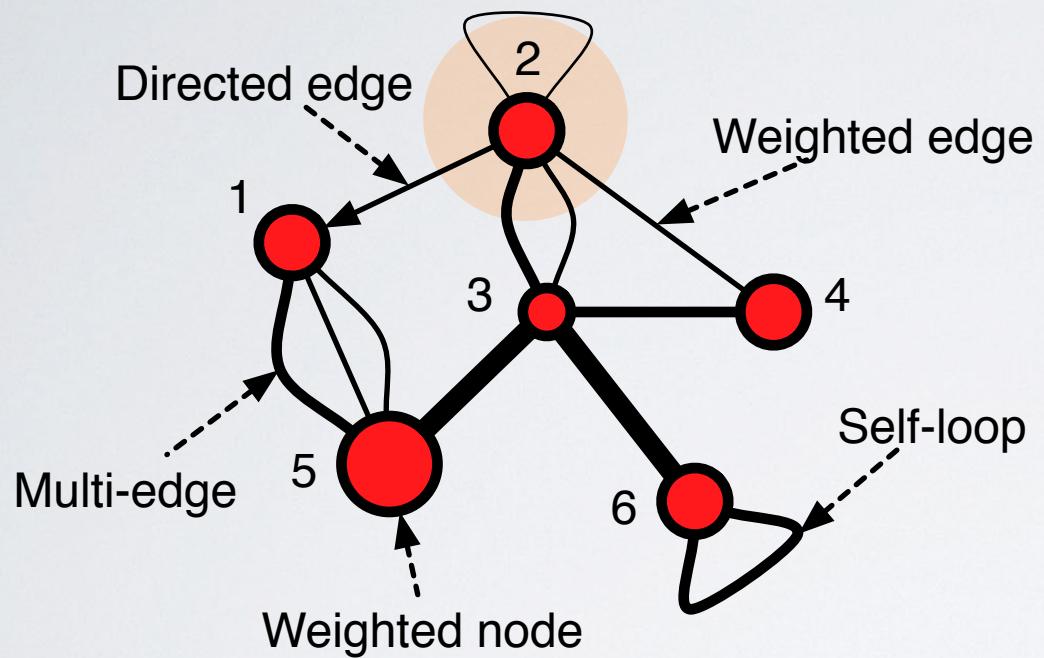


~~undirected~~—

~~unweighted~~—

~~no self loops~~—

beyond simple graphs



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

beyond simple graphs

attributes of

edges	nodes	network
unweighted		
weighted		
signed		
undirected		
directed		
multigraph		
timestamps		

beyond simple graphs

attributes of

edges	nodes	network
unweighted	metadata	
weighted	attributes	
signed	locations	
undirected	state variables	
directed		
multigraph		
timestamps		

beyond simple graphs

attributes of

edges	nodes	network
unweighted	metadata	sparse
weighted	attributes	dense
signed	locations	bipartite
undirected	state variables	projection
directed		acyclic
multigraph		temporal
timestamps		multiplex

directed networks

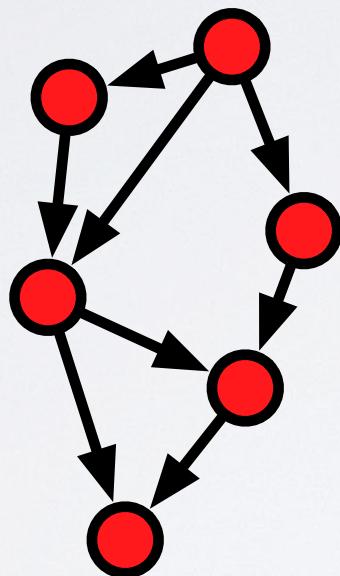
$$A_{ij} \neq A_{ji}$$

citation networks

foodwebs*

epidemiological

others?

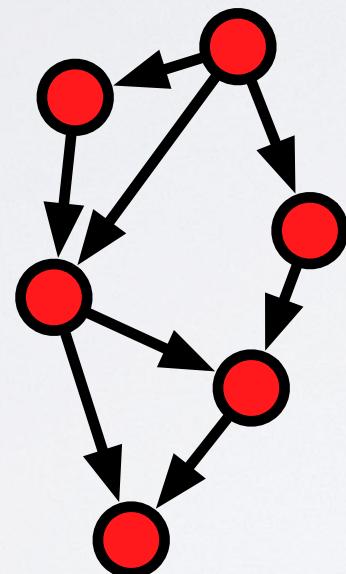


directed acyclic graph

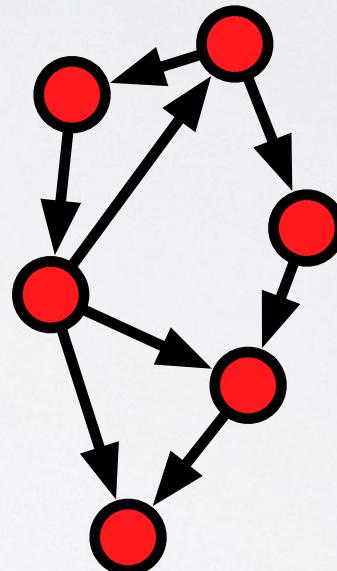
directed networks

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

citation networks
foodwebs*
epidemiological
others?



directed acyclic graph



directed graph

WWW
friendship?
flows of goods,
information
economic exchange
dominance
neuronal
transcription
time travelers

weighted networks

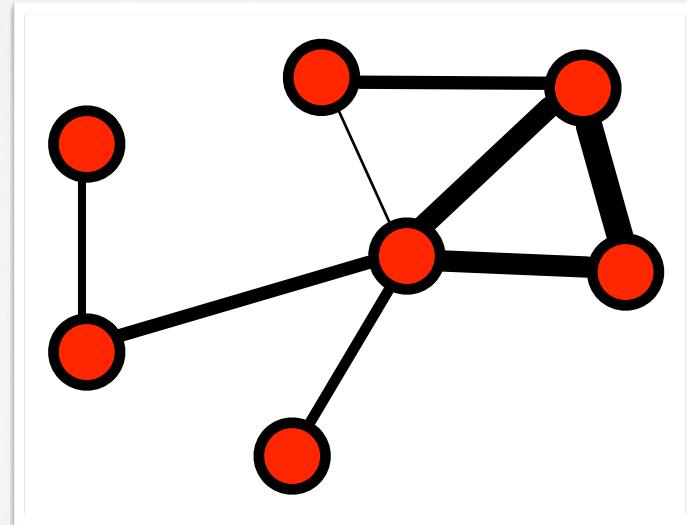
edge type:

- correlations
- similarities
- strengths
- frequencies / counts
- activation / inhibition

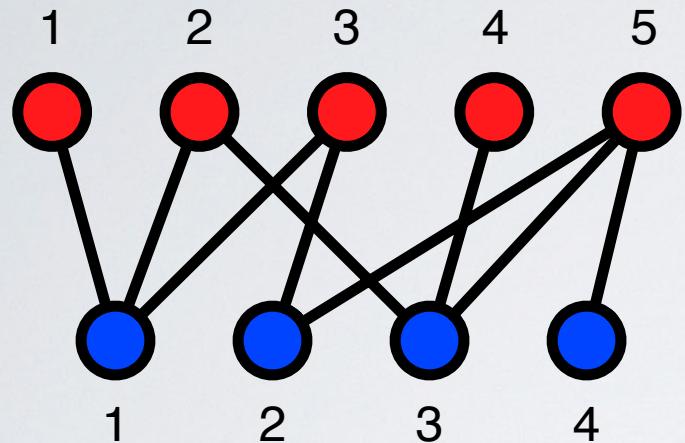
$$A_{ij} \in \mathbb{R} \quad (\text{real valued})$$

$$A_{ij} \in \mathbb{Z}^* \quad (\text{natural numbers})$$

$$A_{ij} \in \{-1, 0, +1\} \quad (\text{signed})$$



bipartite networks



2 types of nodes, with edges only between types

$$V = V_1 \cup V_2$$

$$E \subseteq V_1 \times V_2$$

authors & papers

people & locations (checkins)

actors & movies/scenes

metabolites & reactions

musicians & albums

genes & substrings

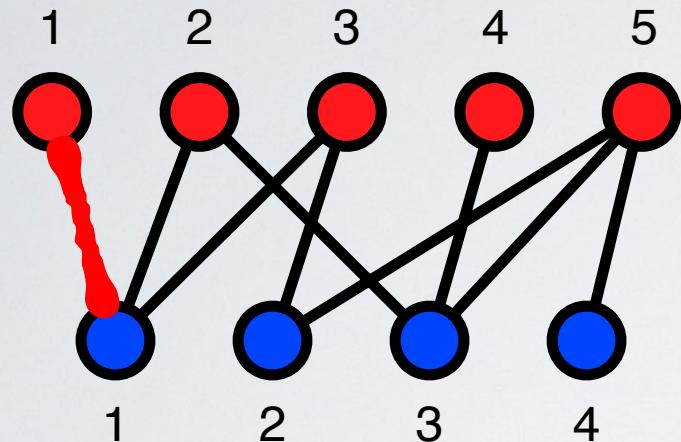
people & online groups

words & documents

people & corporate boards

plants & pollinators

bipartite networks



A	1	2	3	4	5	1	2	3	4
1						1	0	0	0
2						1	0	1	0
3						1	1	0	0
4						0	0	1	0
5						0	1	1	1
	1	2	3	4	5	1	2	3	4
1	1	1	1	0	0				
2	0	0	1	0	1				
3	0	1	0	1	1				
4	0	0	0	0	1				

authors & papers

actors & movies/scenes

musicians & albums

people & online groups

people & corporate boards

people & locations (checkins)

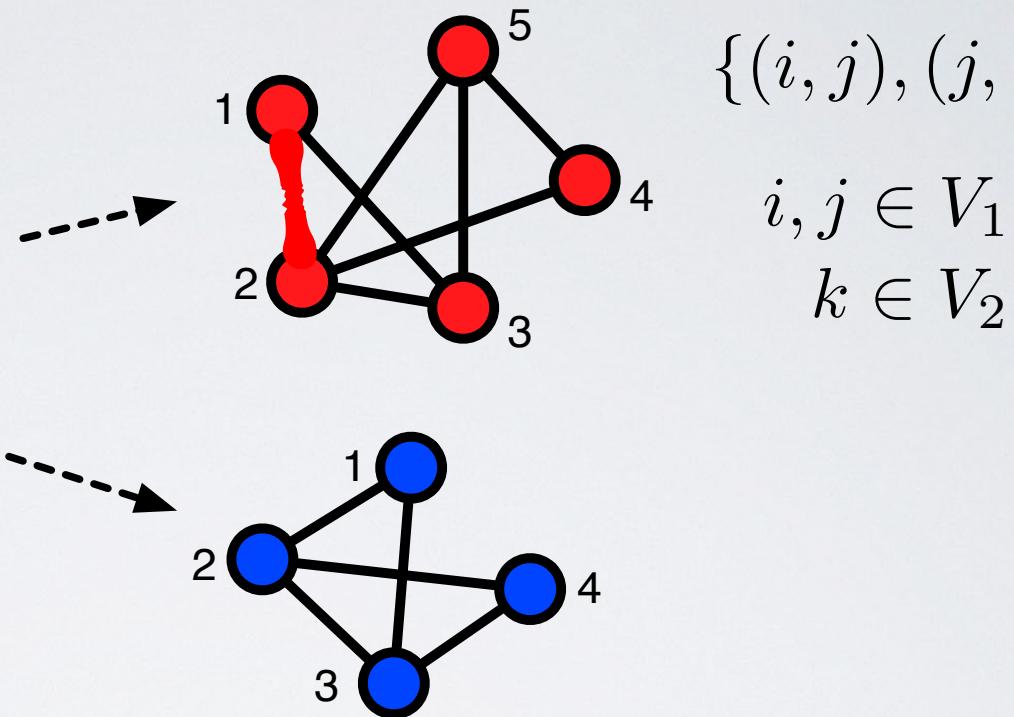
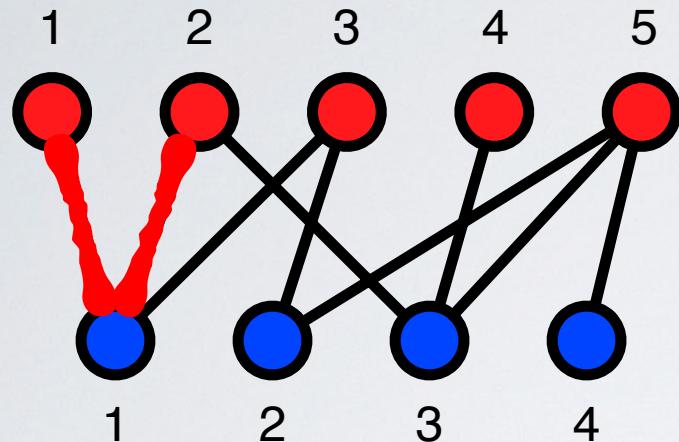
metabolites & reactions

genes & substrings

words & documents

plants & pollinators

bipartite networks



$$\{(i, j), (j, k)\} \in E$$

$$i, j \in V_1 \\ k \in V_2$$

one-mode projections

authors & papers

actors & movies/scenes

musicians & albums

people & online groups

people & corporate boards

people & locations (checkins)

metabolites & reactions

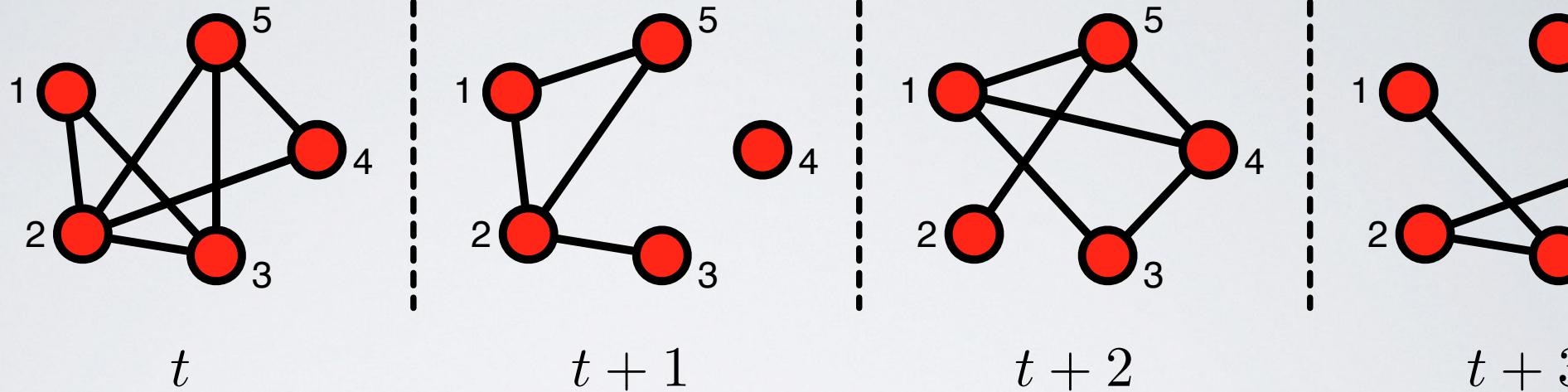
genes & substrings

words & documents

plants & pollinators

* projections always destroy information. Theorem: no weighted projection can always preserve bipartite structure

temporal networks

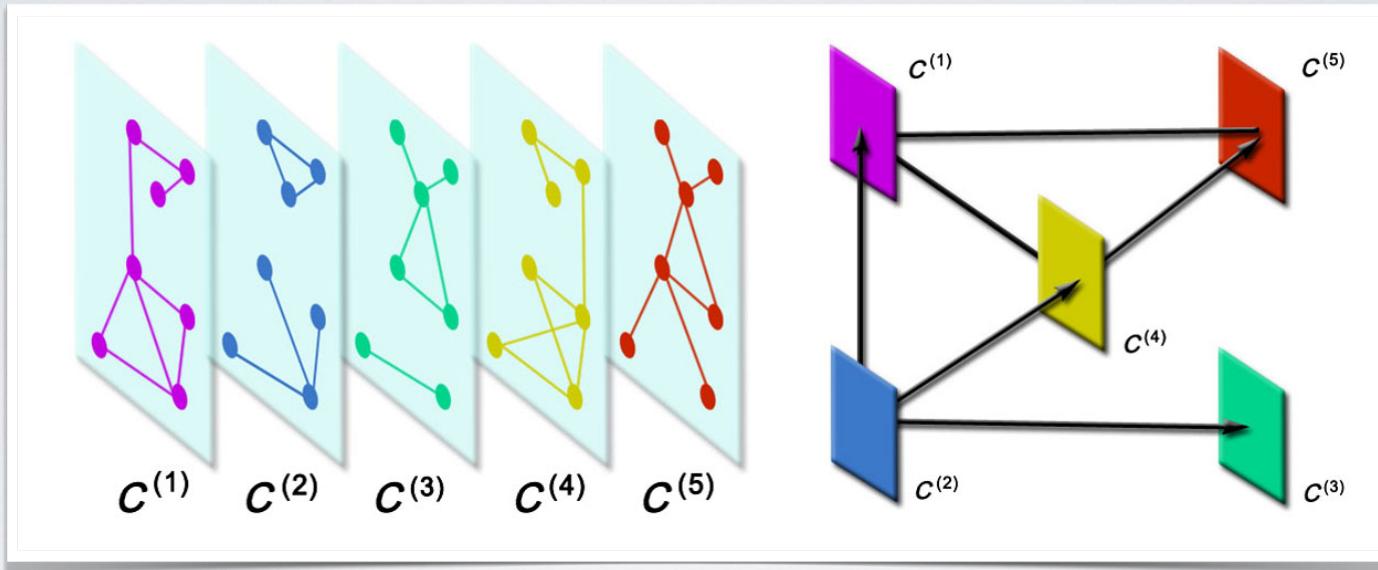


any network over time

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

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

multiplex networks



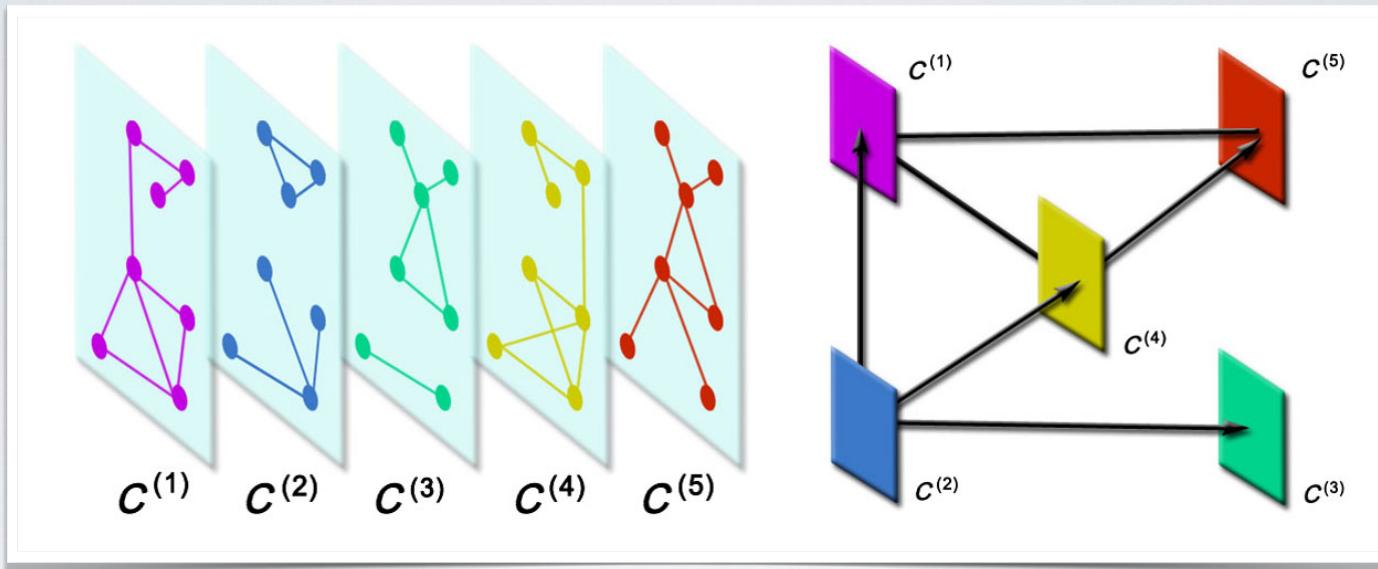
multiple network "layers"

each layer has same set of nodes V

but different sets of edges $\{E_1, E_2, \dots, E_\ell\}$

generalizes "temporal" networks: ℓ index snapshots

multiplex networks



multiple network "layers"

each layer has same set of nodes V

but different sets of edges $\{E_1, E_2, \dots, E_\ell\}$

generalizes "temporal" networks: ℓ index snapshots

different types of transportation within a city

different types of social interactions (trust, socializing, co-located, etc.)

end of Lecture I