Computational Analysis of Big Data

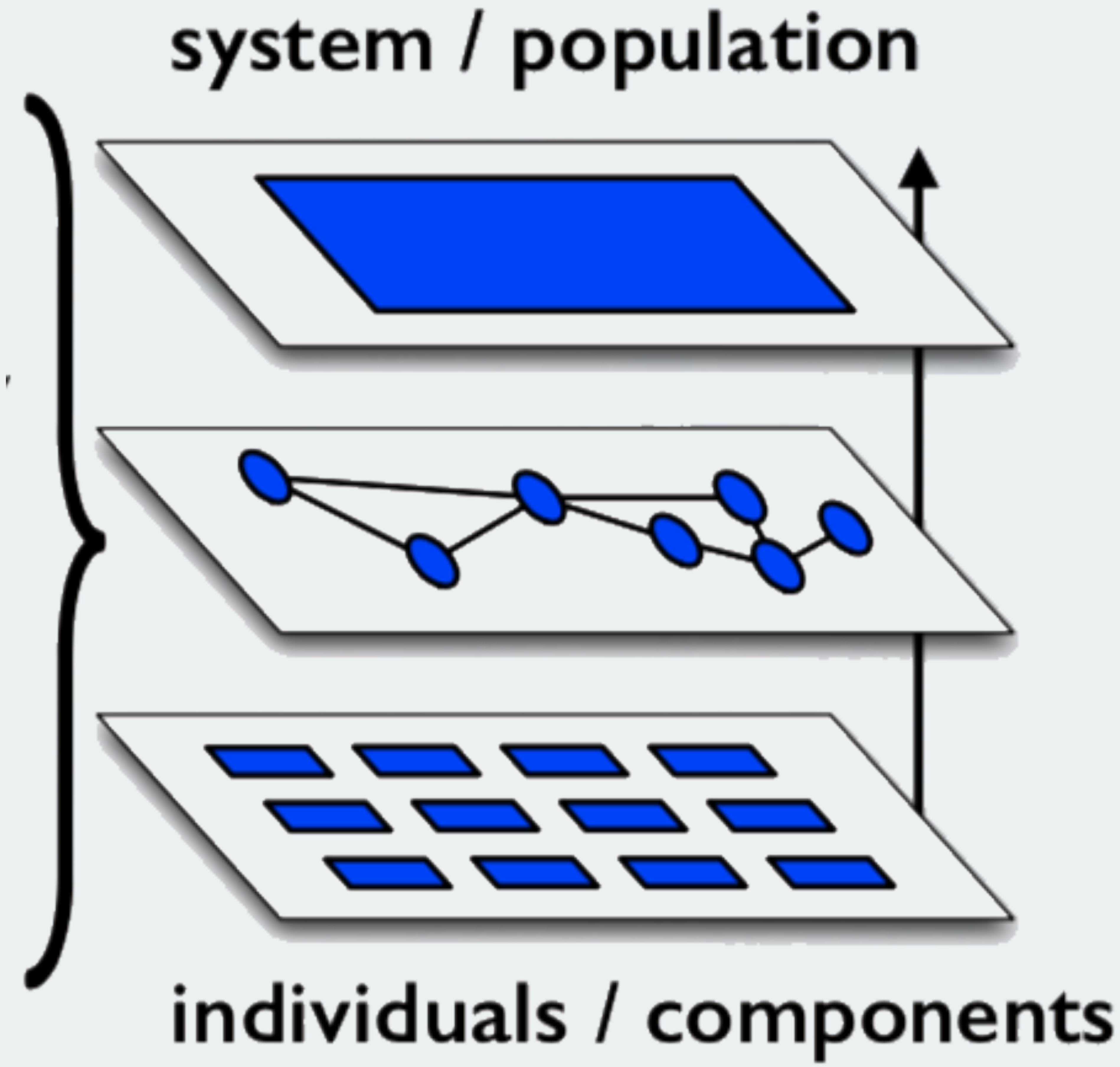
Week 6

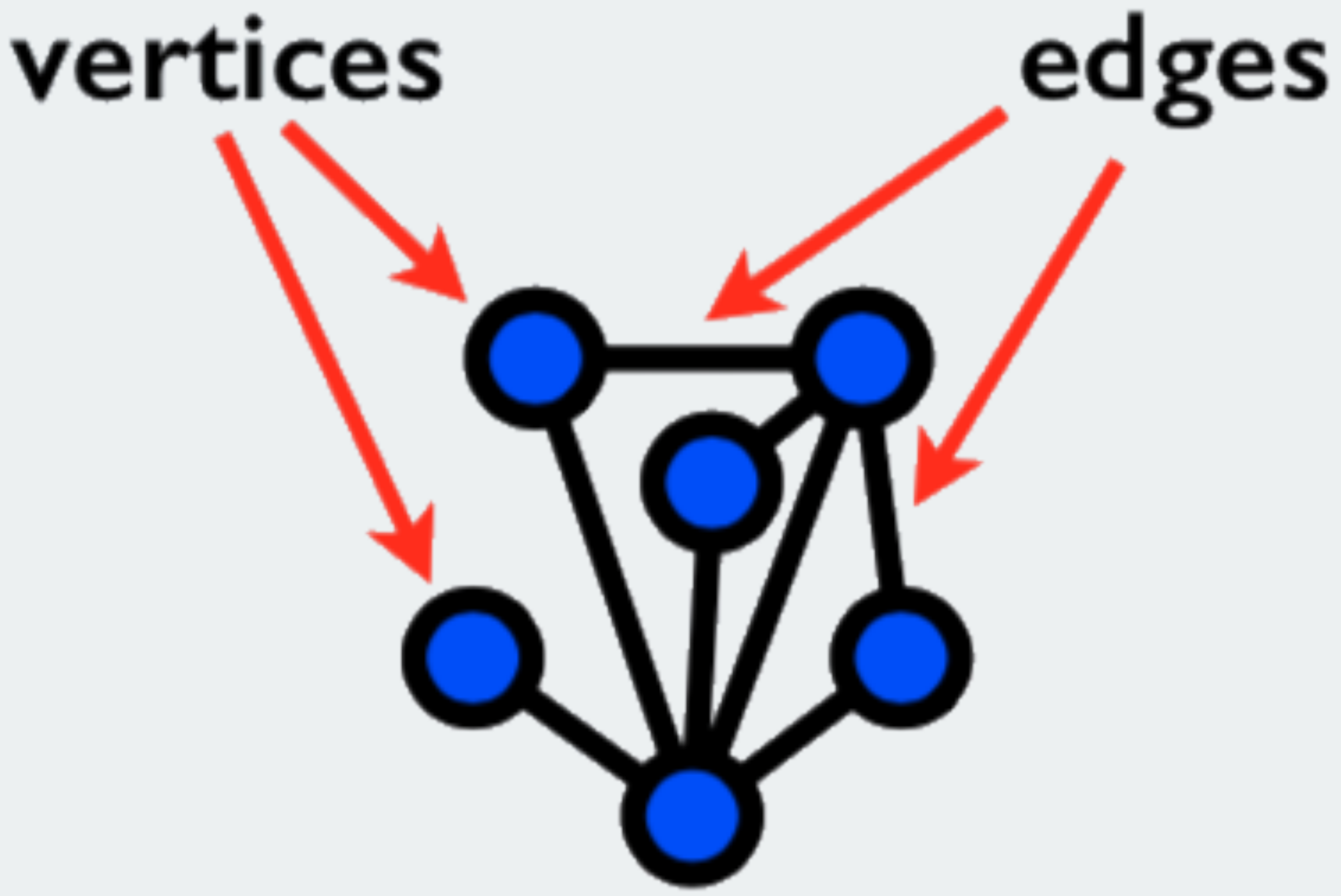
Networks

What are networks?

What are networks?

- An approach
- A mathematical representation
- Provide structure to complexity
- Modeling systems both at individual scale and population scale





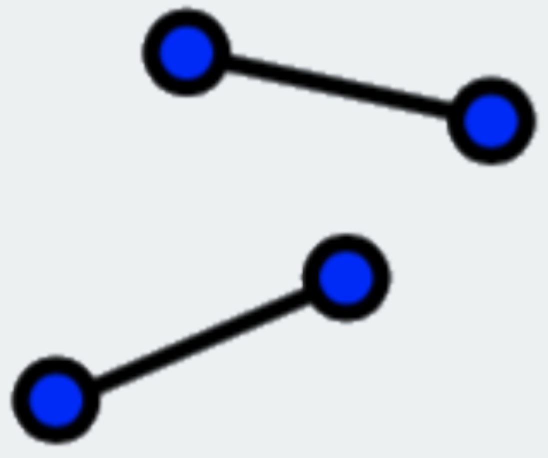
What is a vertex?

V: an component/node in a system

What is an edge?

$E \subseteq V \times V$: a pairwise relation (edges / links / ties)



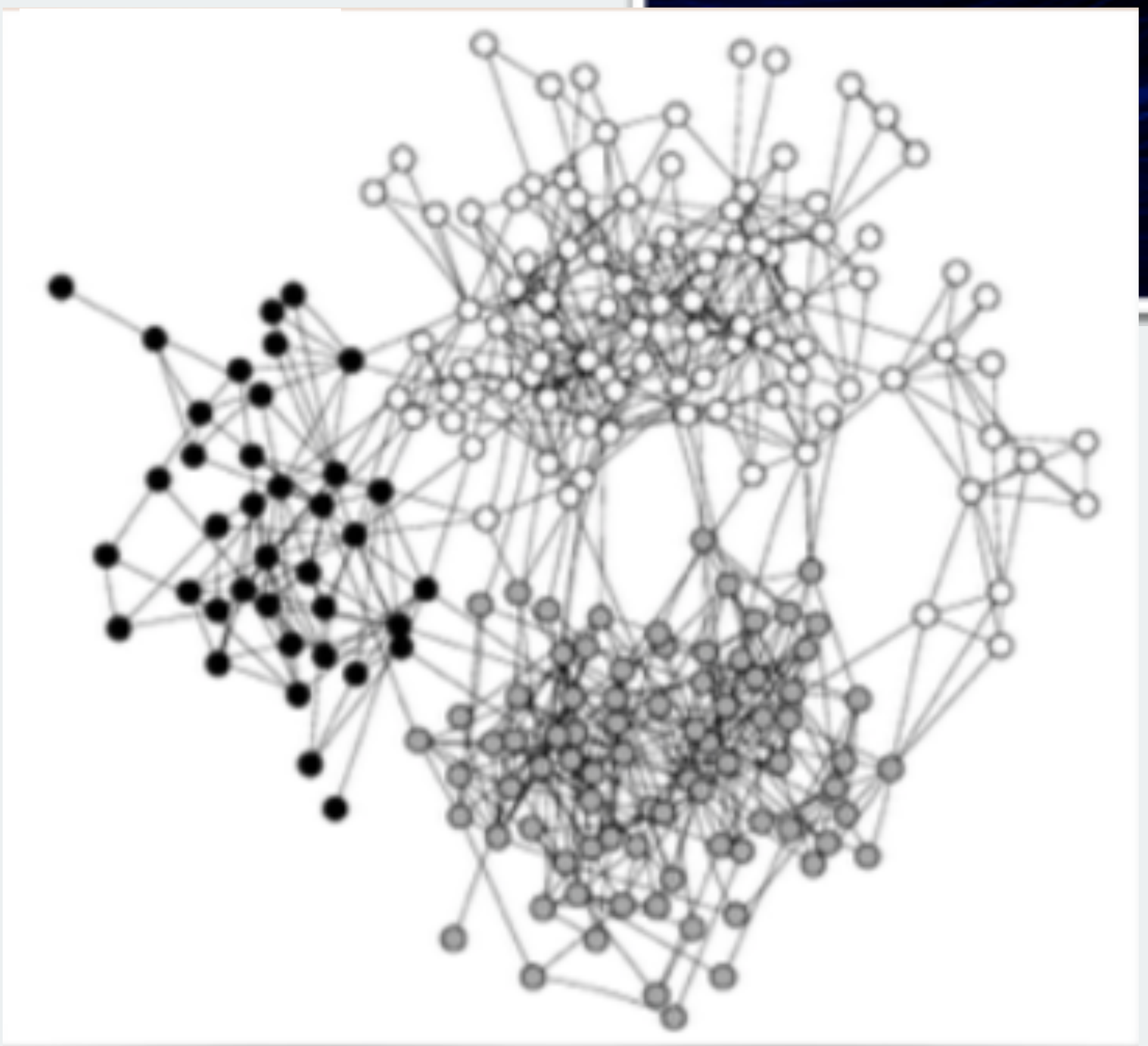


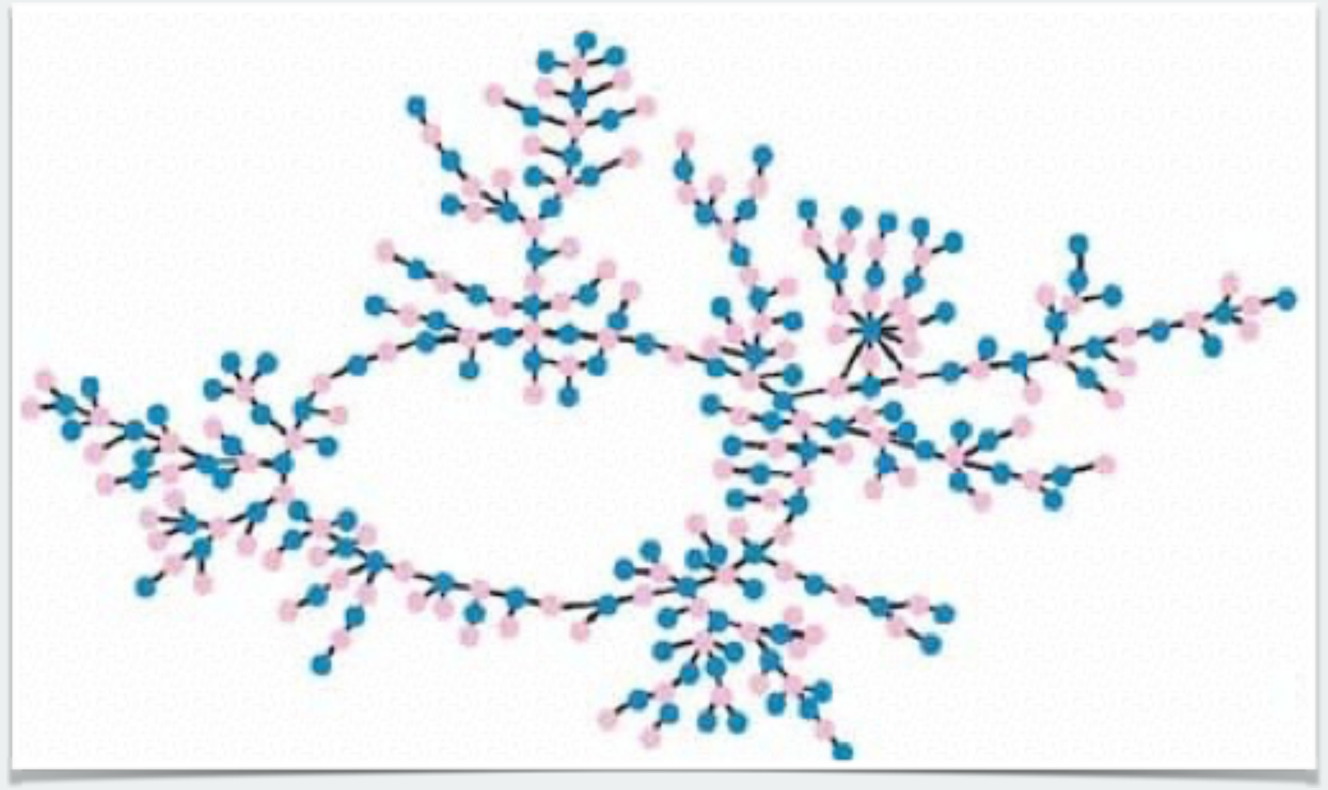
	System	Vertex	Edge
Telecom	Internet(1)	computer	IP network adjacency
	Internet(2)	autonomous system (ISP)	BGP connection
Information	software	function	function call
	World Wide Web	web page	hyperlink
	documents	article, patent, or legal case	citation
Transport	power grid transmission	generating or relay 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, collab., sexual contact, communication, authority, exchange, etc.

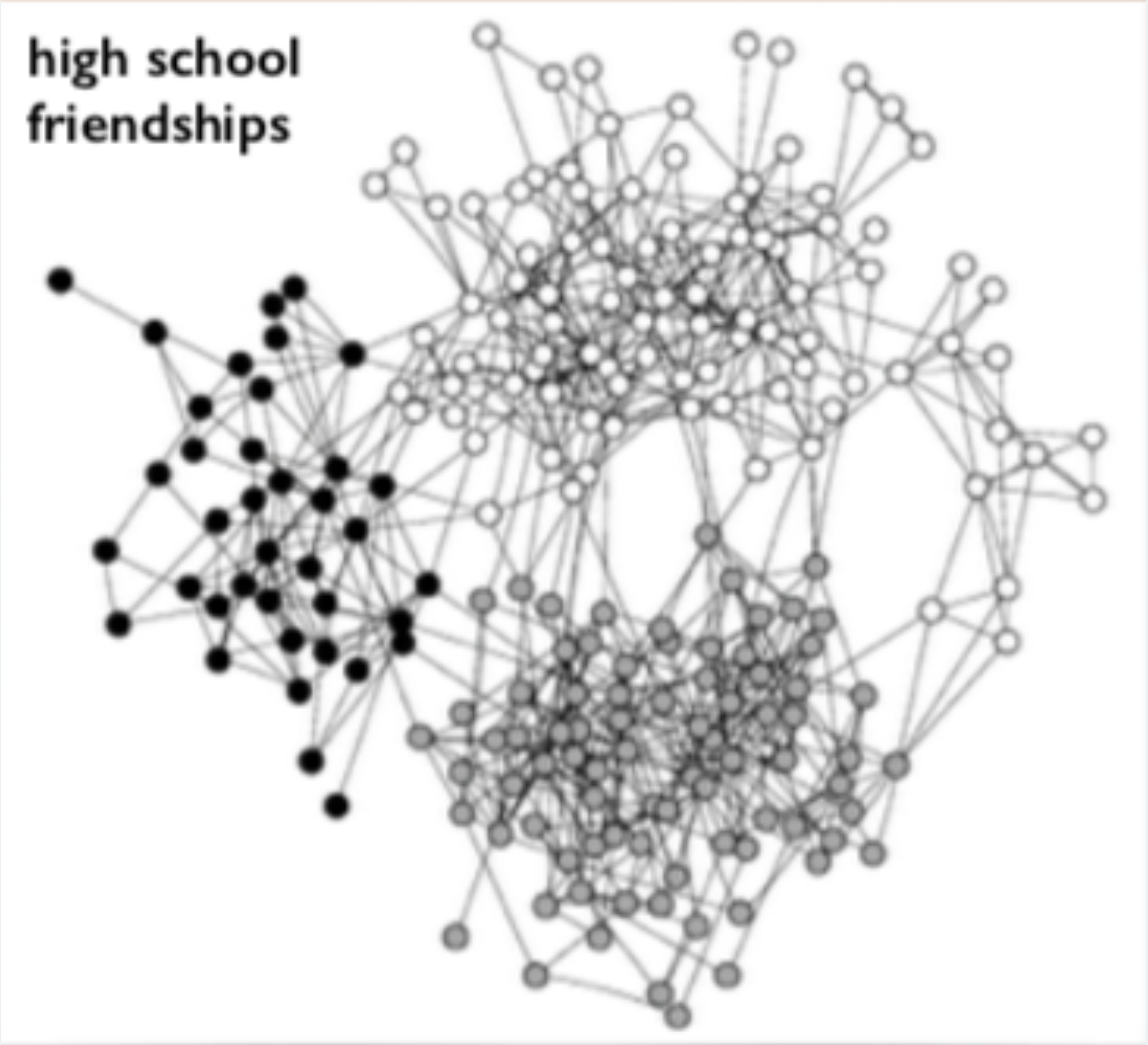


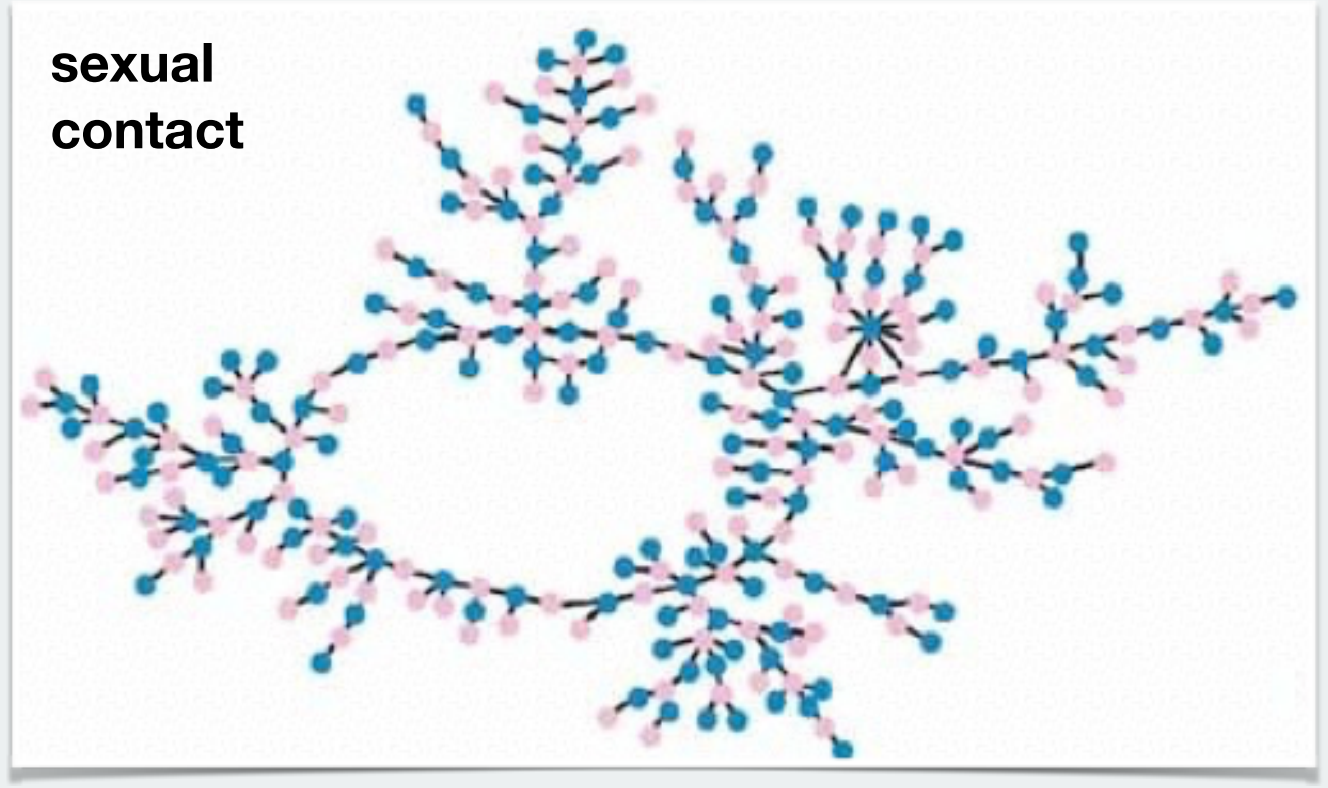


Social networks

Vertex: A person

Edge: Friendship, collab., sexual contact, communication, authority, exchange, etc.

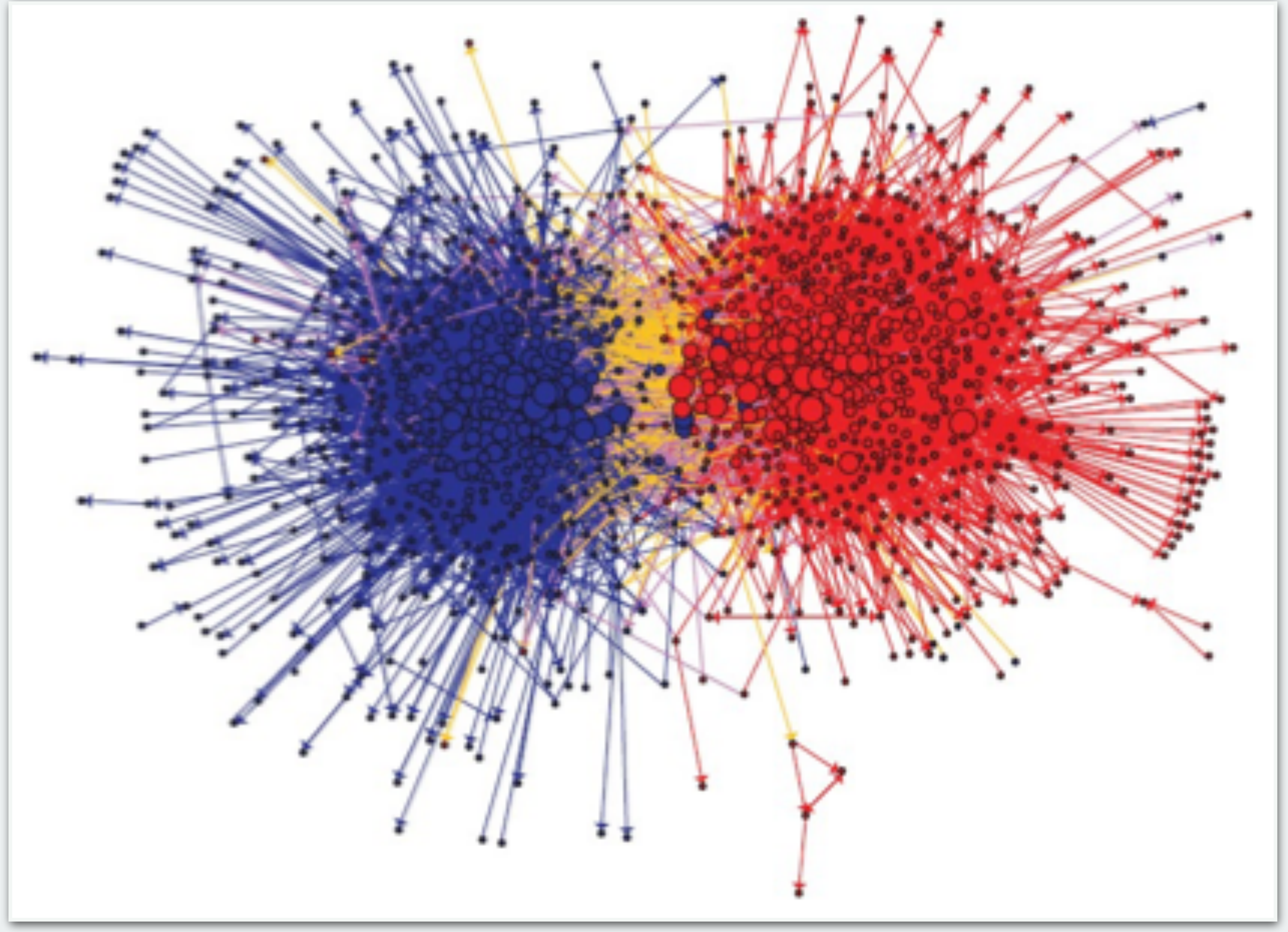




Information networks

Vertex: Books, articles, blogs, webpages, etc.

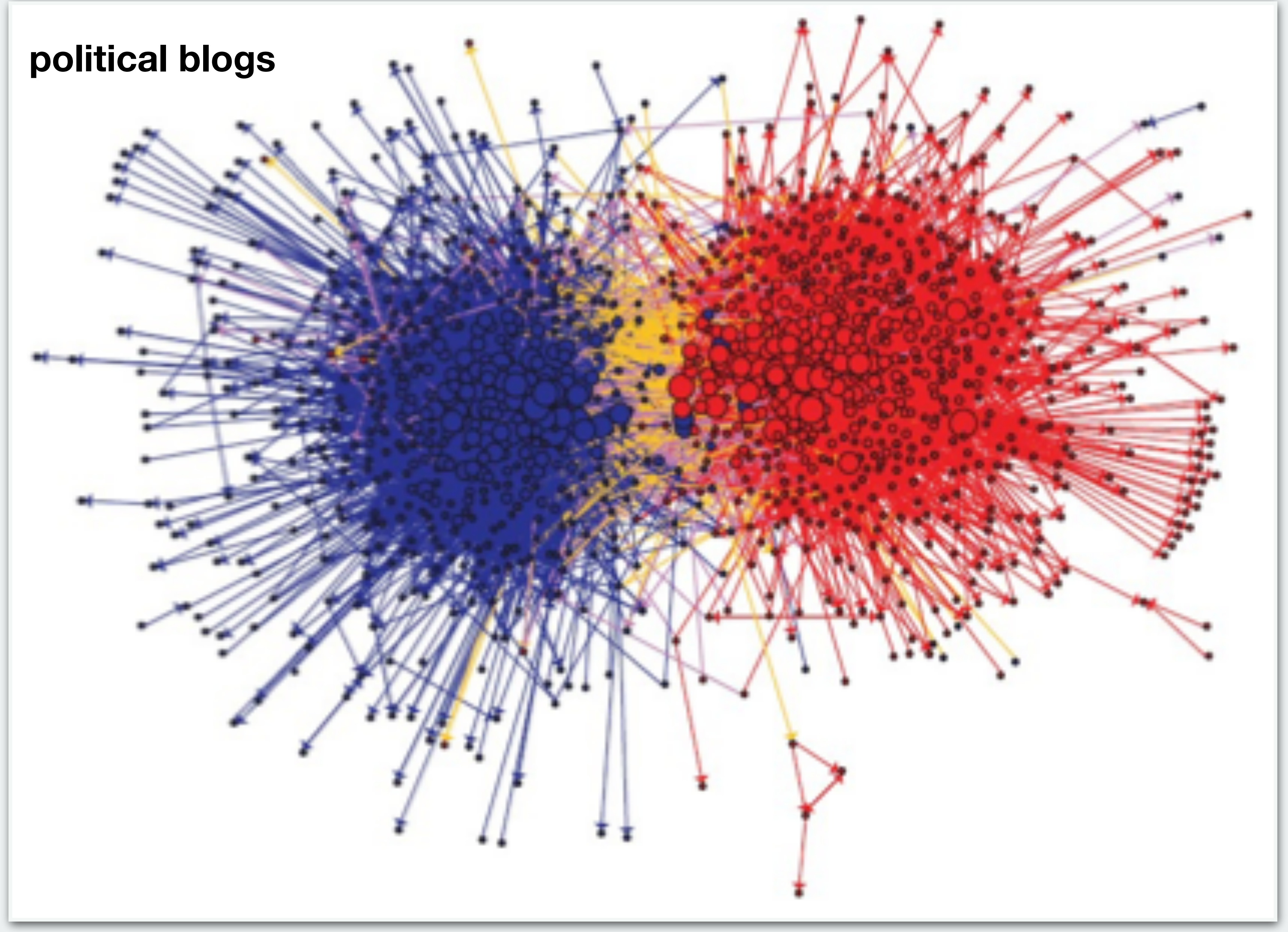
Edge: Citation, hyperlinks, recommendations, similarity, etc.

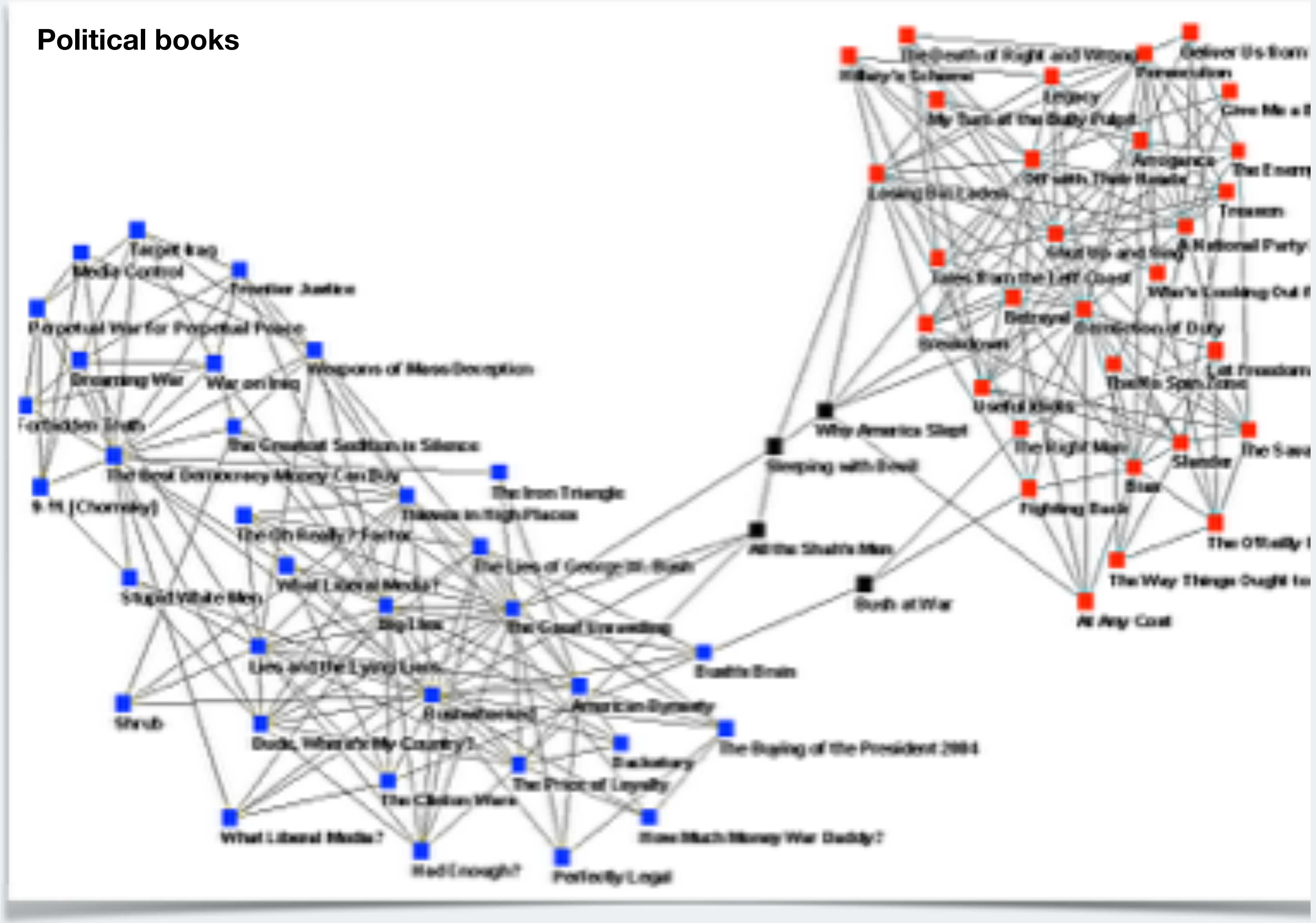


Information networks

Vertex: Books, articles, blogs, webpages, etc.

Edge: Citation, hyperlinks, recommendations, similarity, etc.

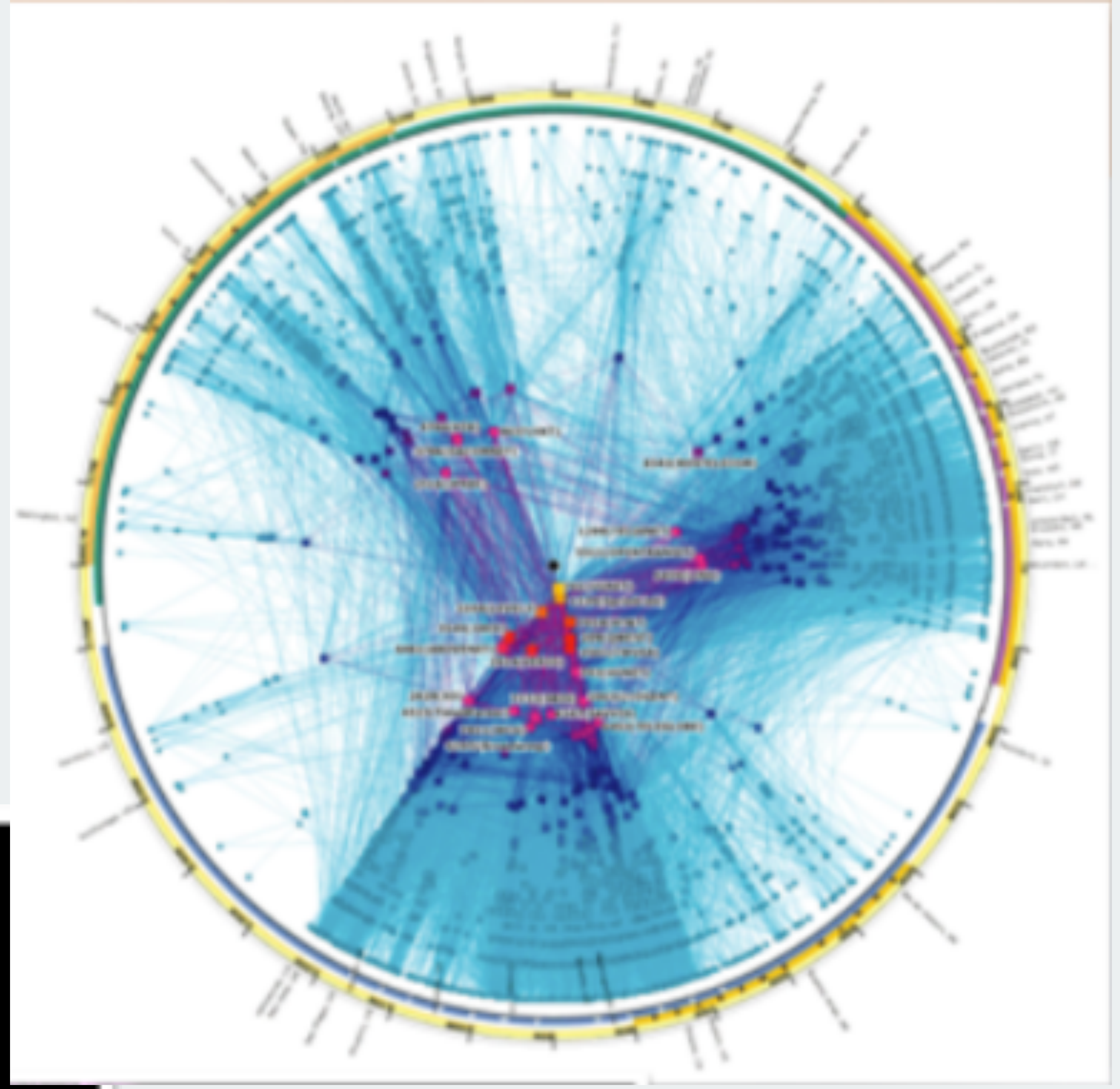


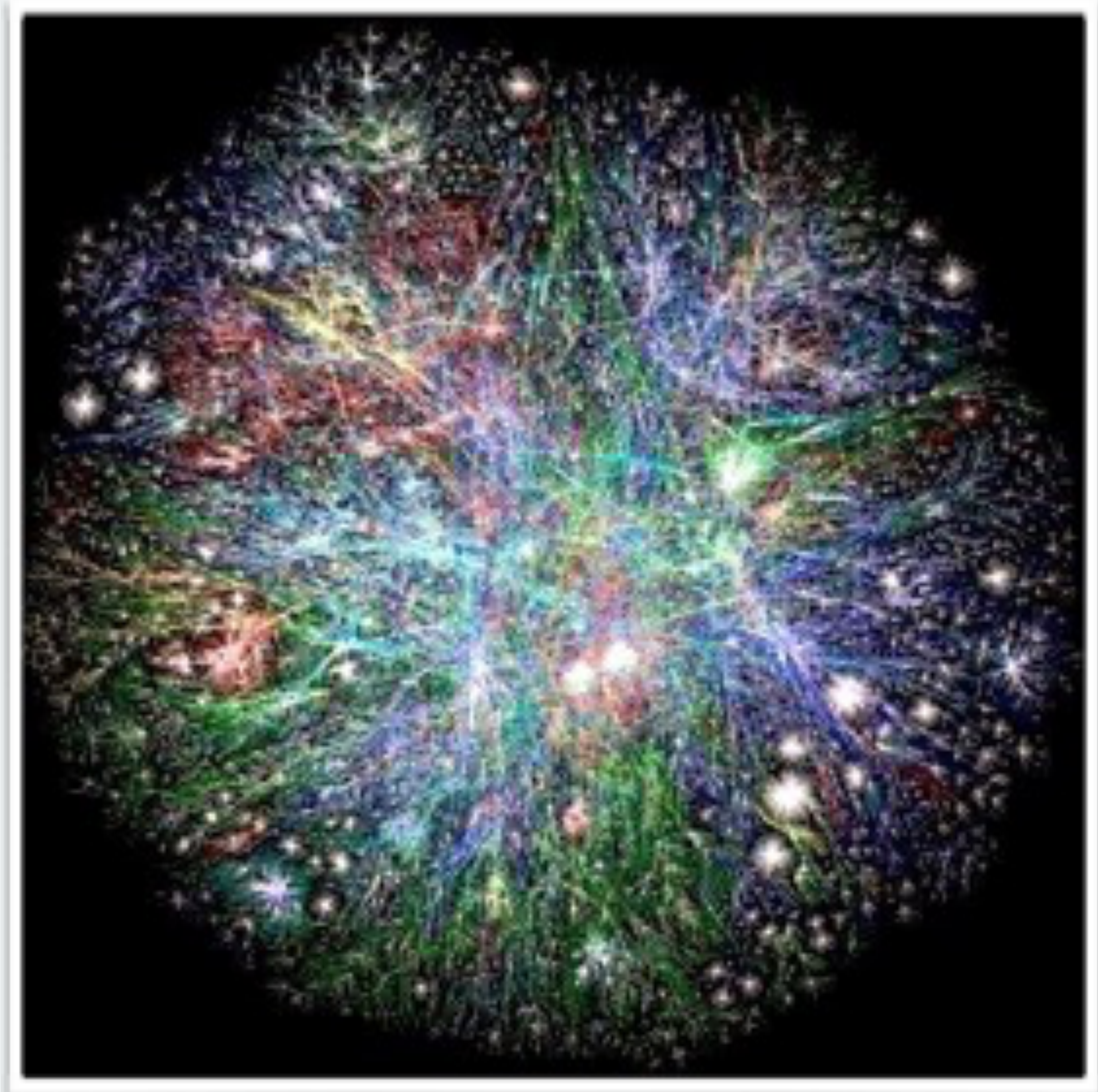


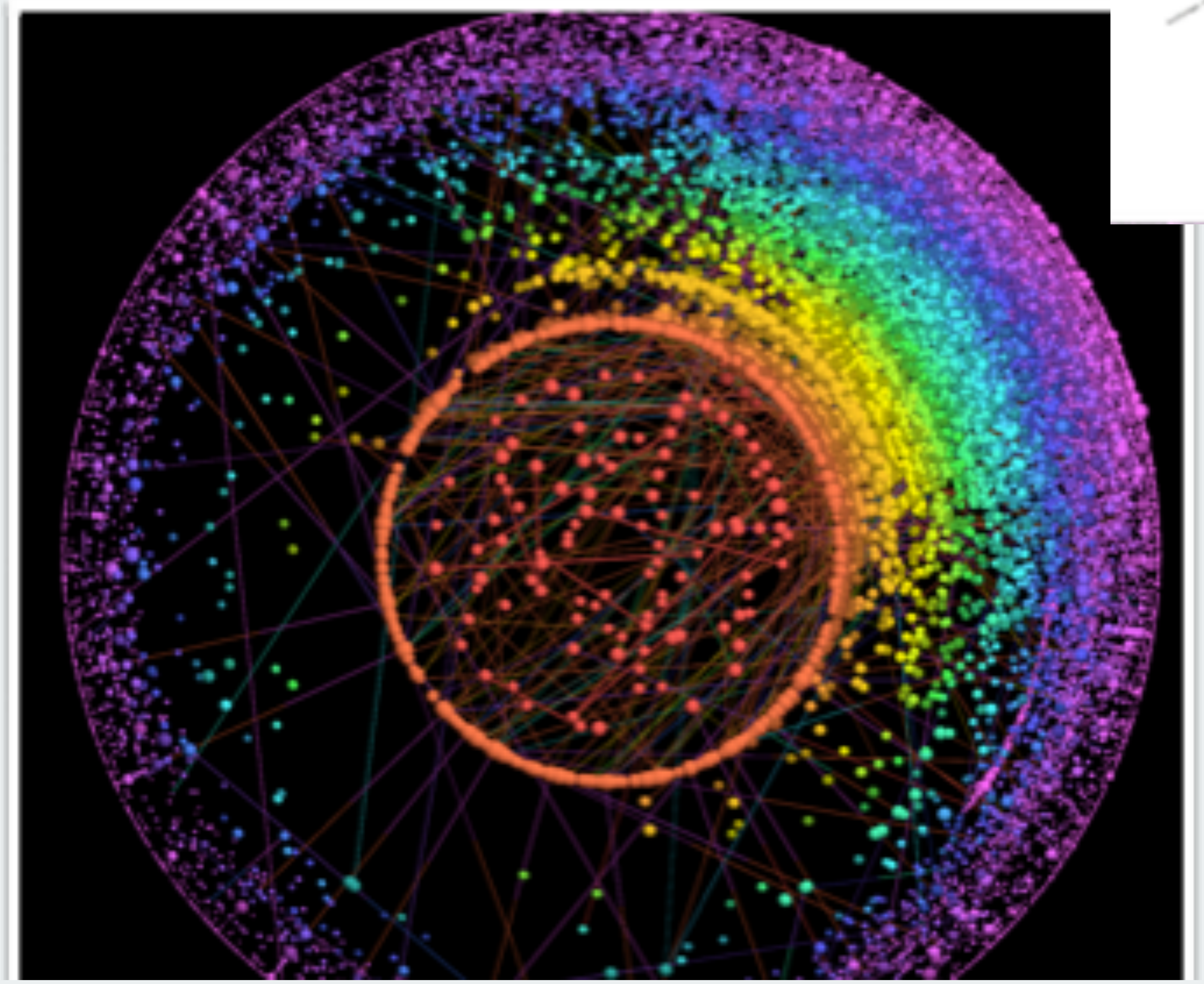
Communication networks

Vertex: Network router, ISP, email address, phone number, etc.

Edge: Information exchange



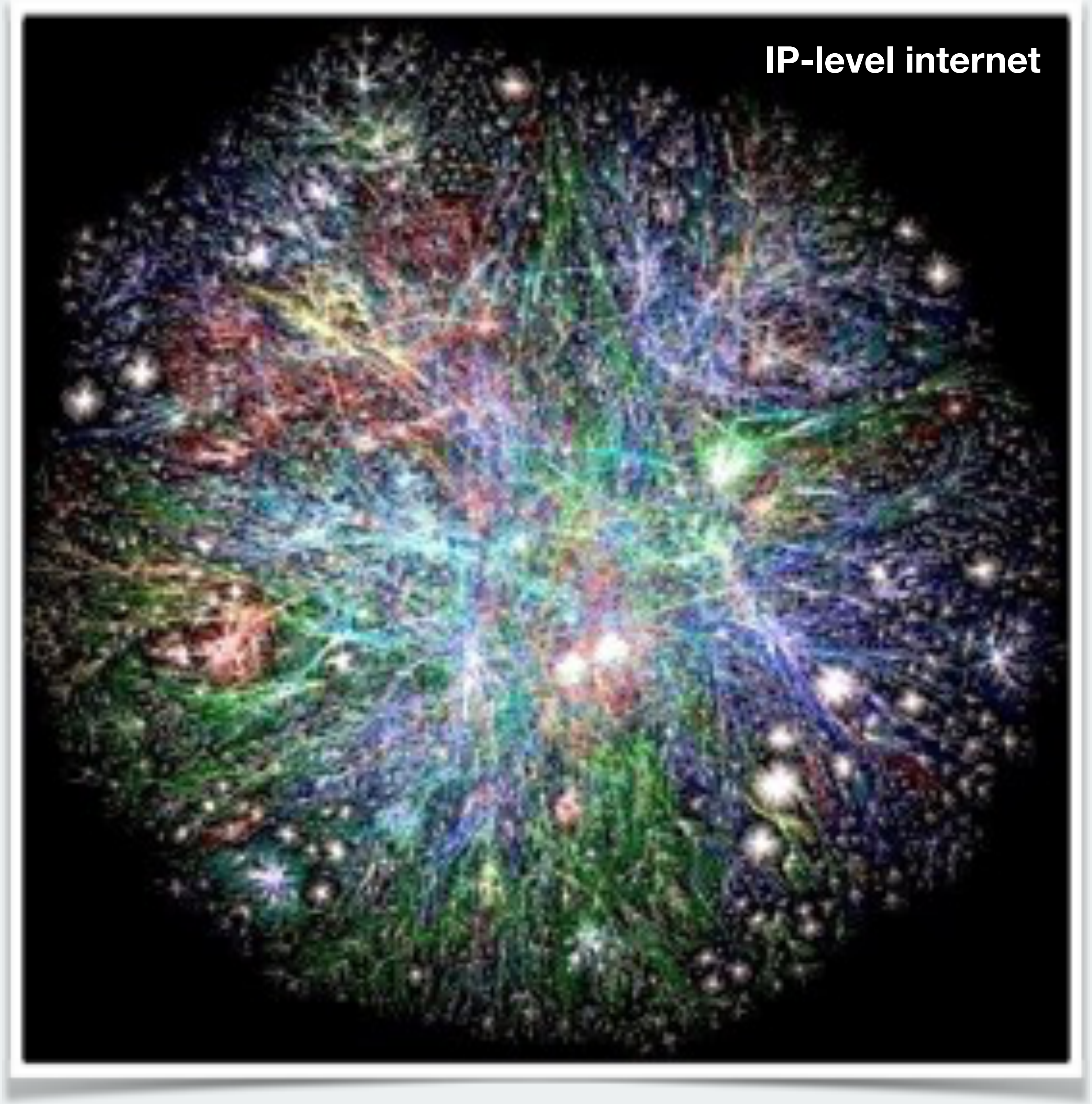


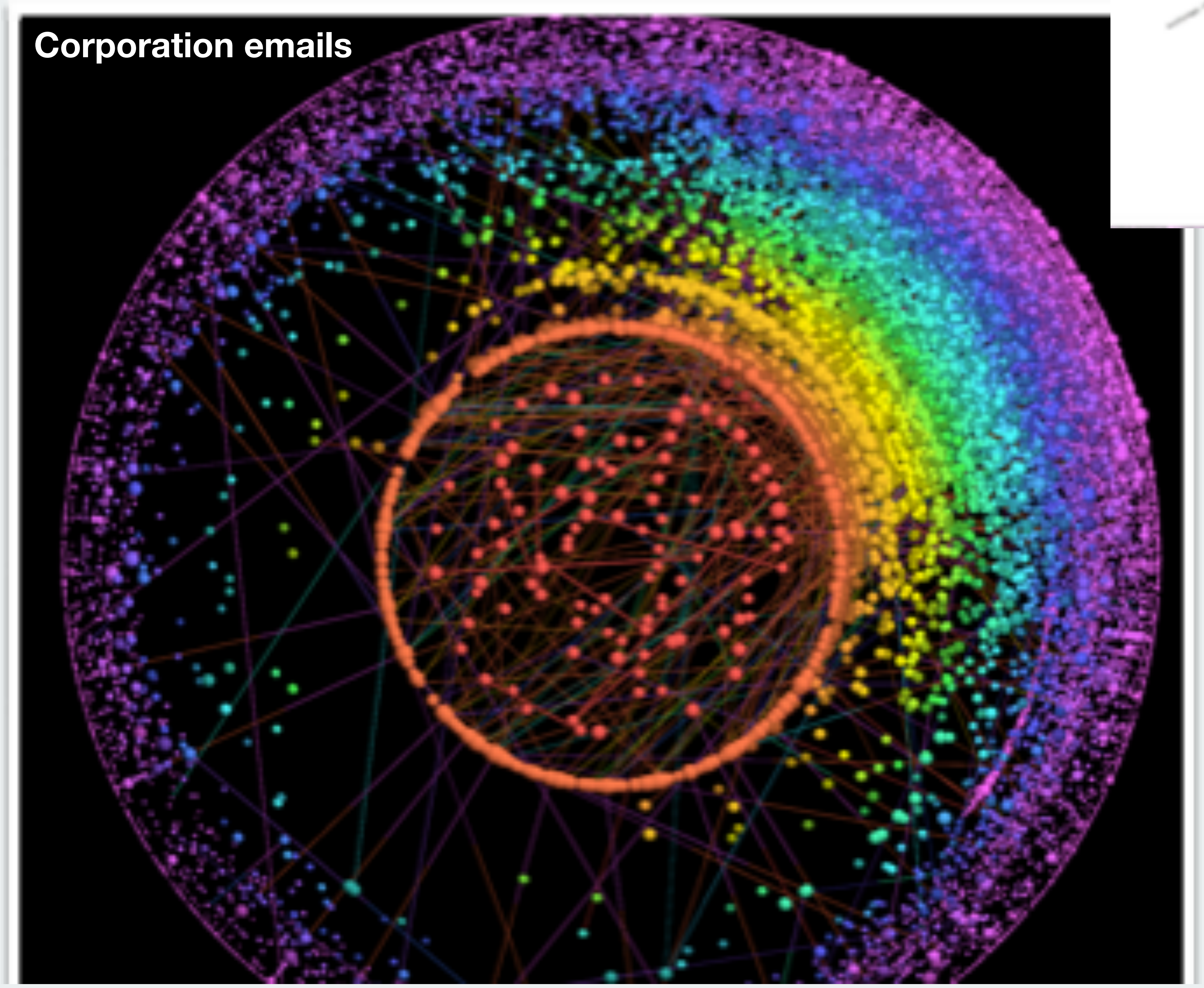


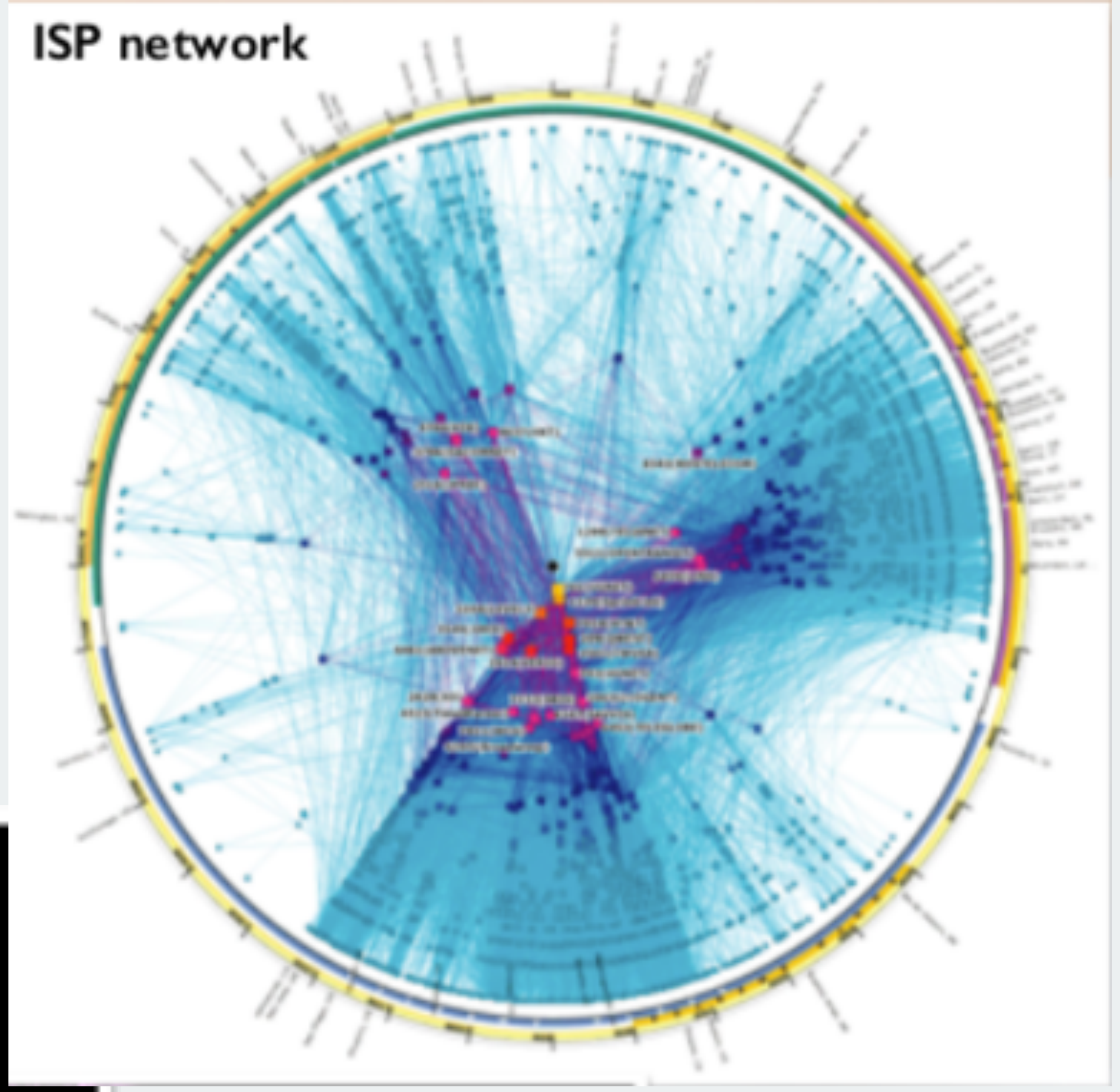
Communication networks

Vertex: Network router, ISP, email address, phone number, etc.

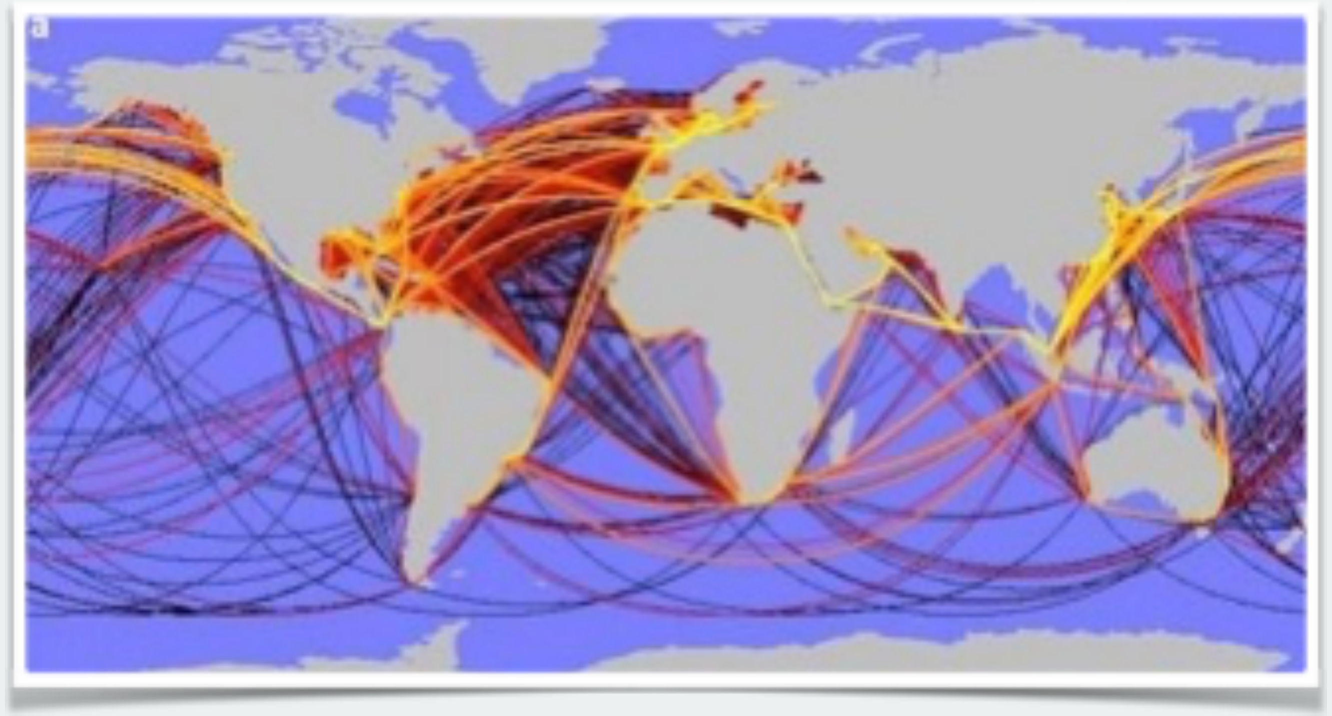
Edge: Information exchange







Transportation networks



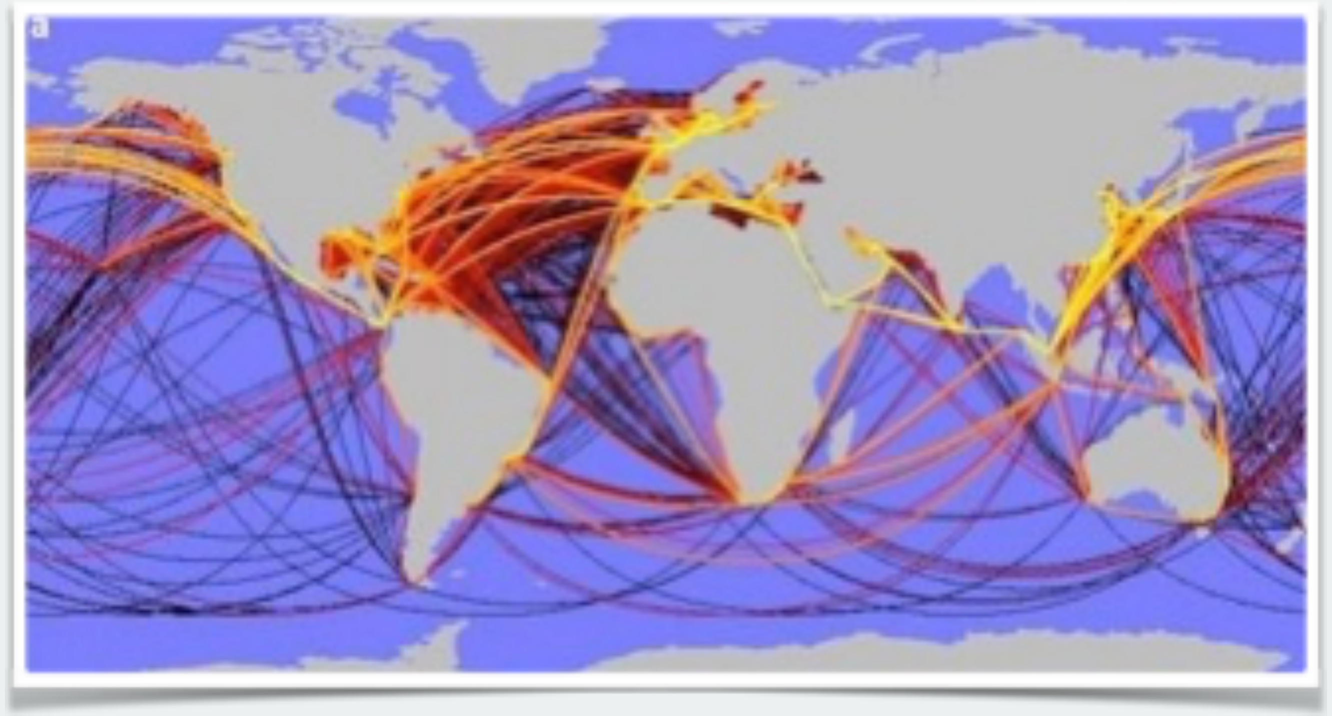
Vertex: city, airport, junction, railway, station, river confluence

Edge: physical material being transported

Transportation networks

Vertex: city, airport, junction, railway, station, river confluence

Edge: physical material being transported



global shipping

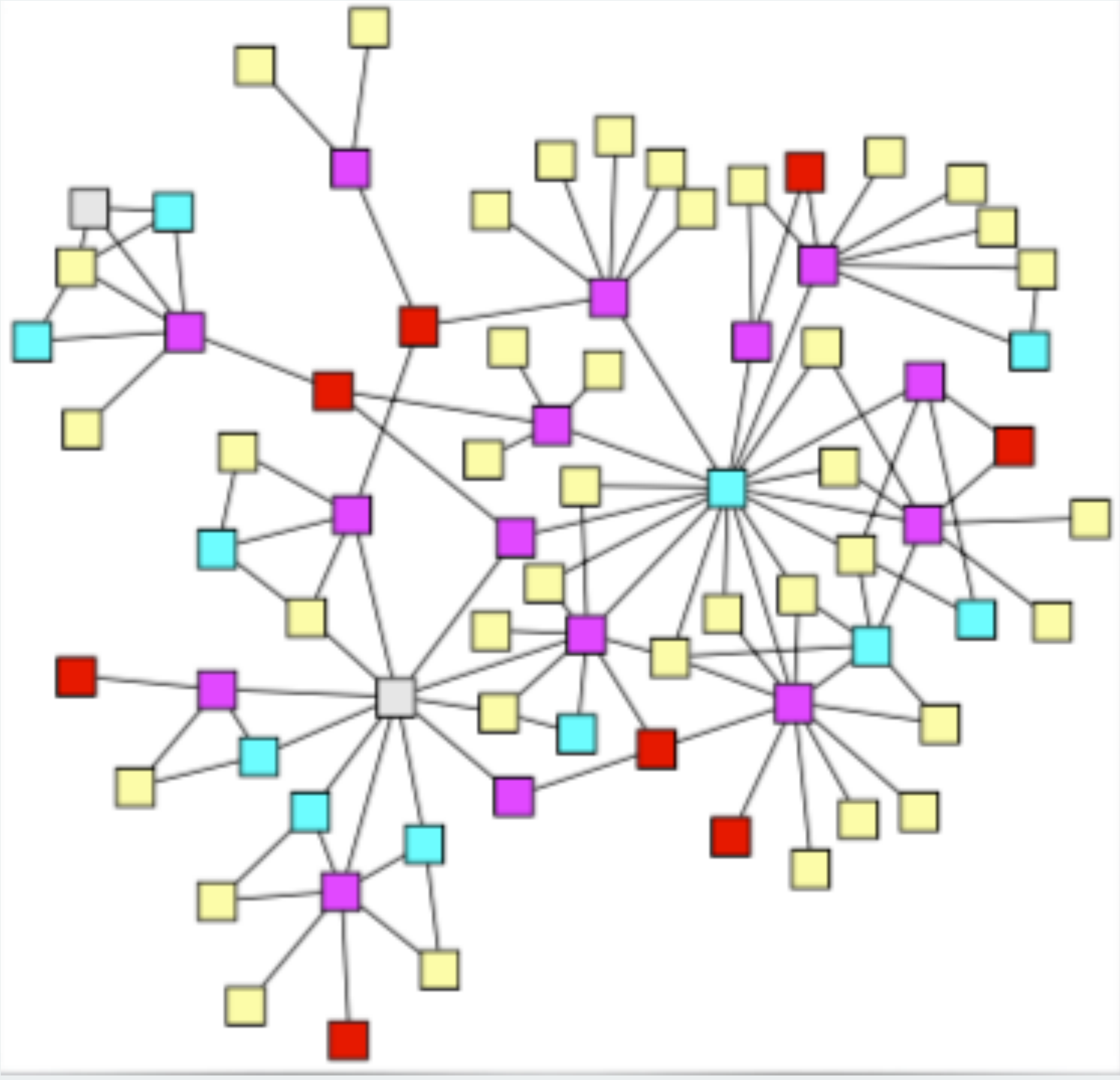
global air traffic

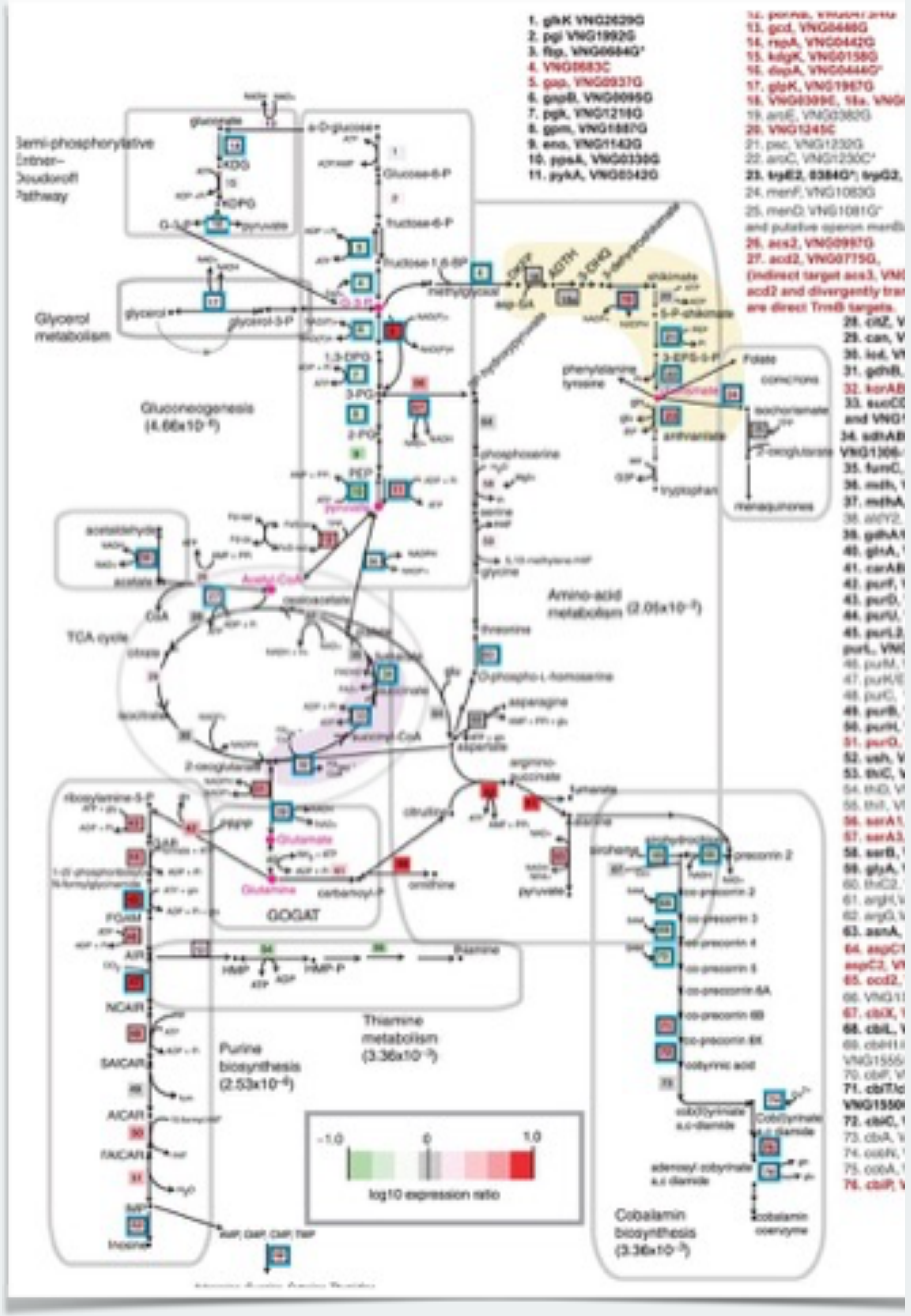
US interstates

Biological networks

Vertex: species, metabolite, protein, gene, neuron, etc.

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

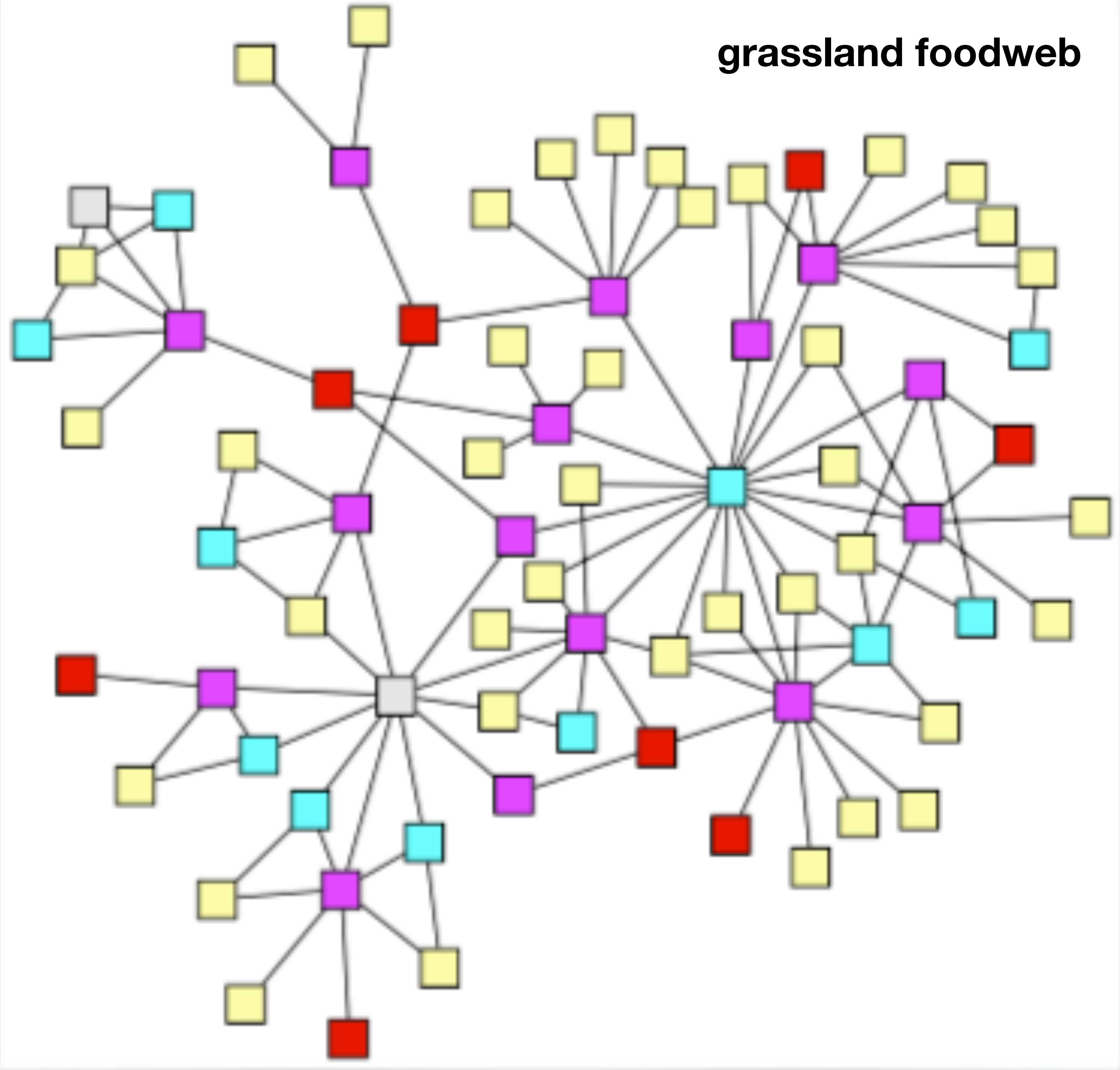


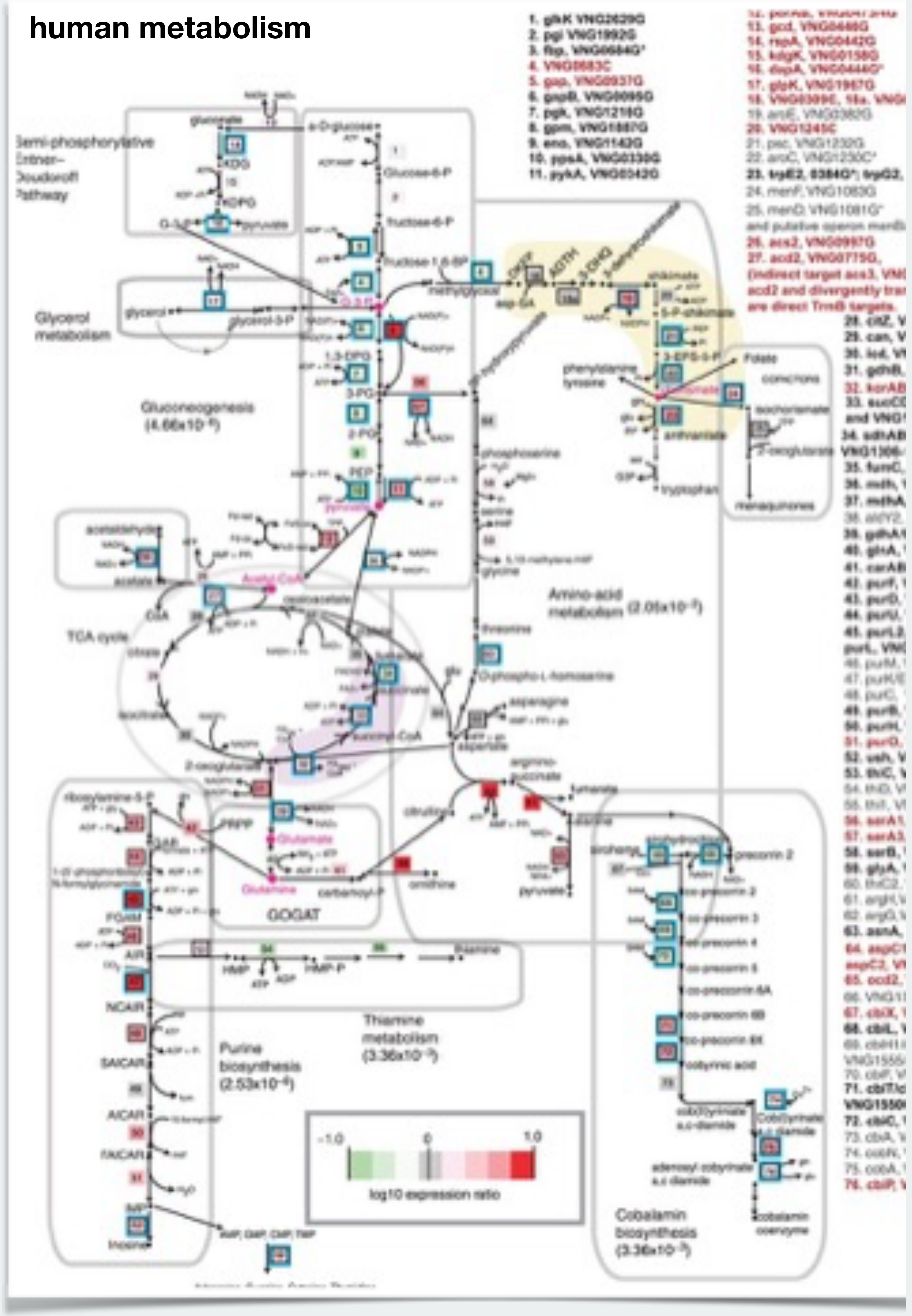


Biological networks

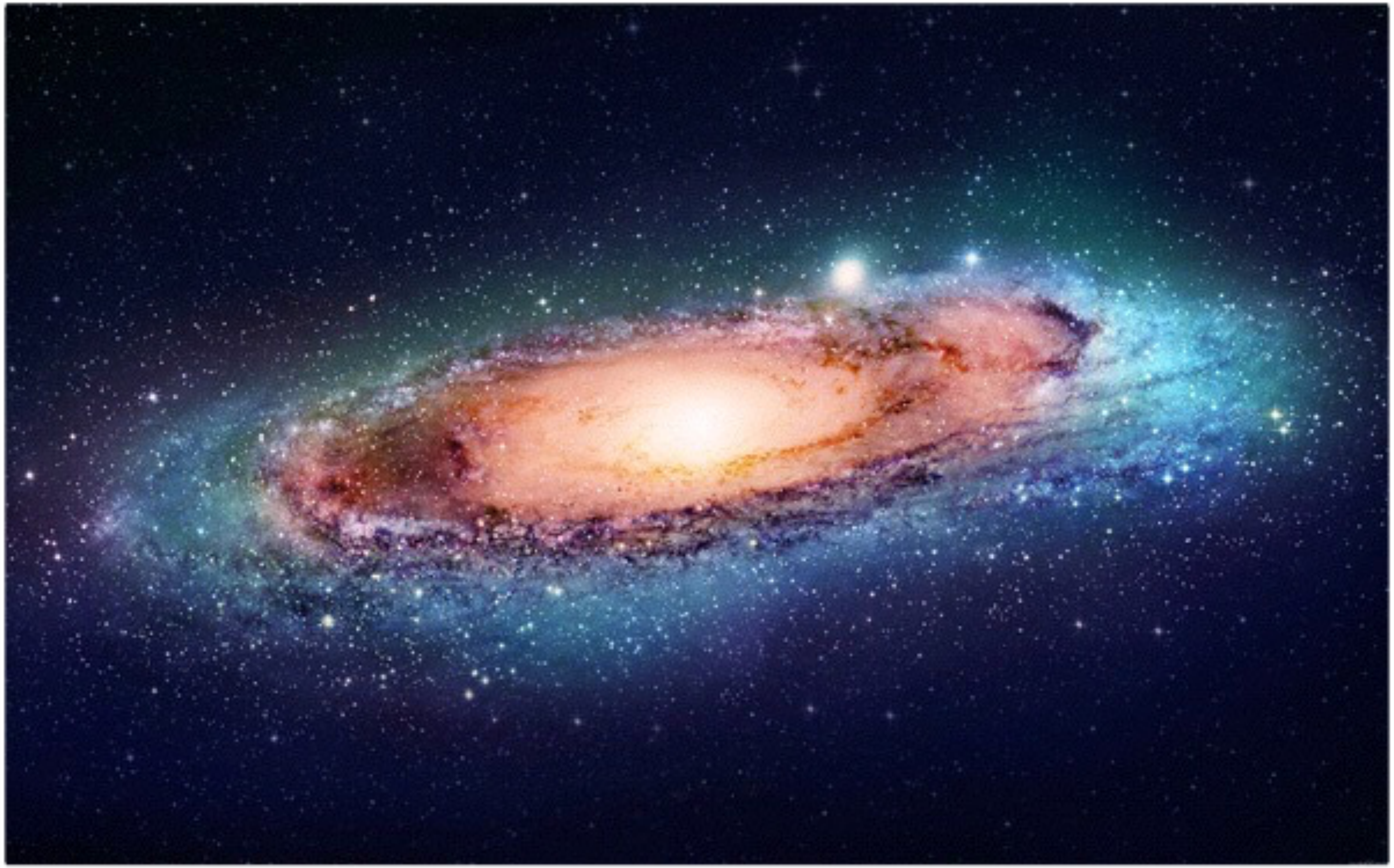
Vertex: species, metabolite, protein, gene, neuron, etc.

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





Quiz



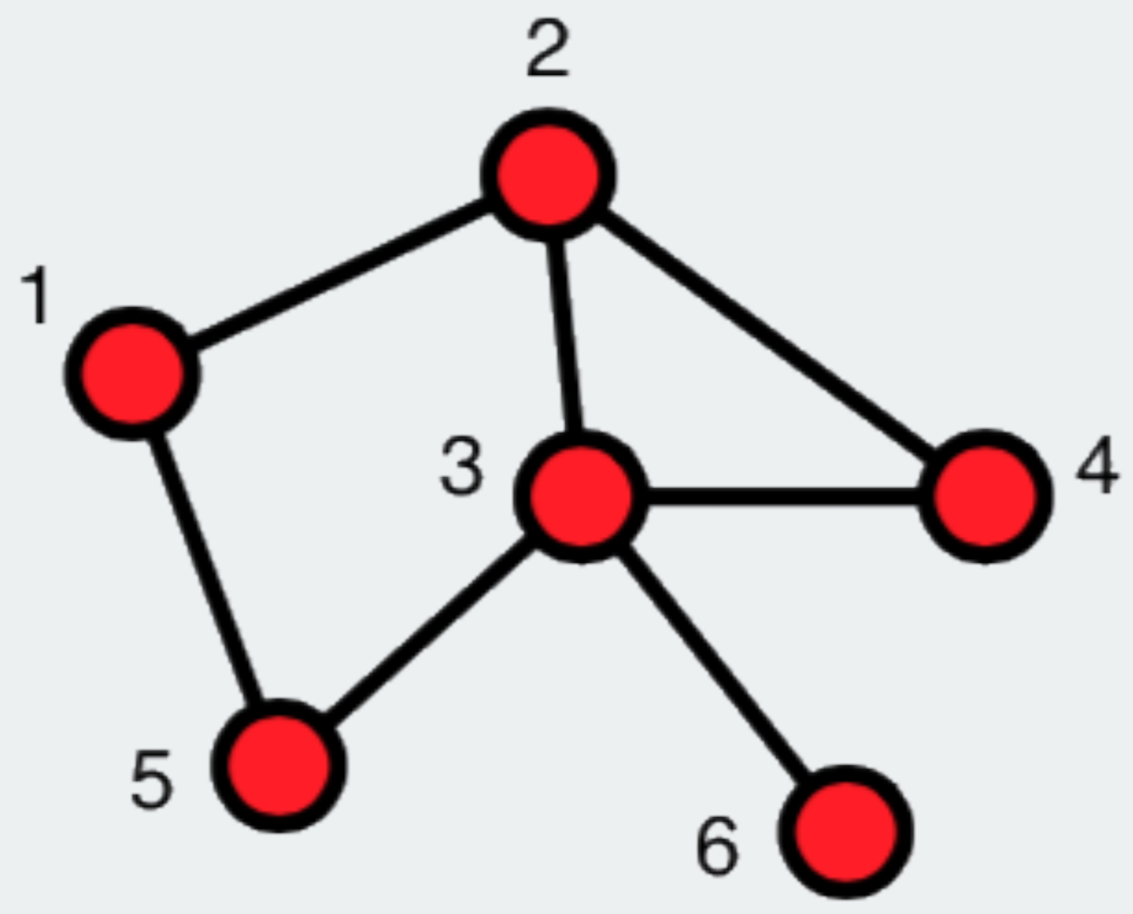
Quiz

Representing networks

Representing networks

- Simple networks
- Directed and weighted networks
- Bipartite networks
- Temporal networks

Simple networks

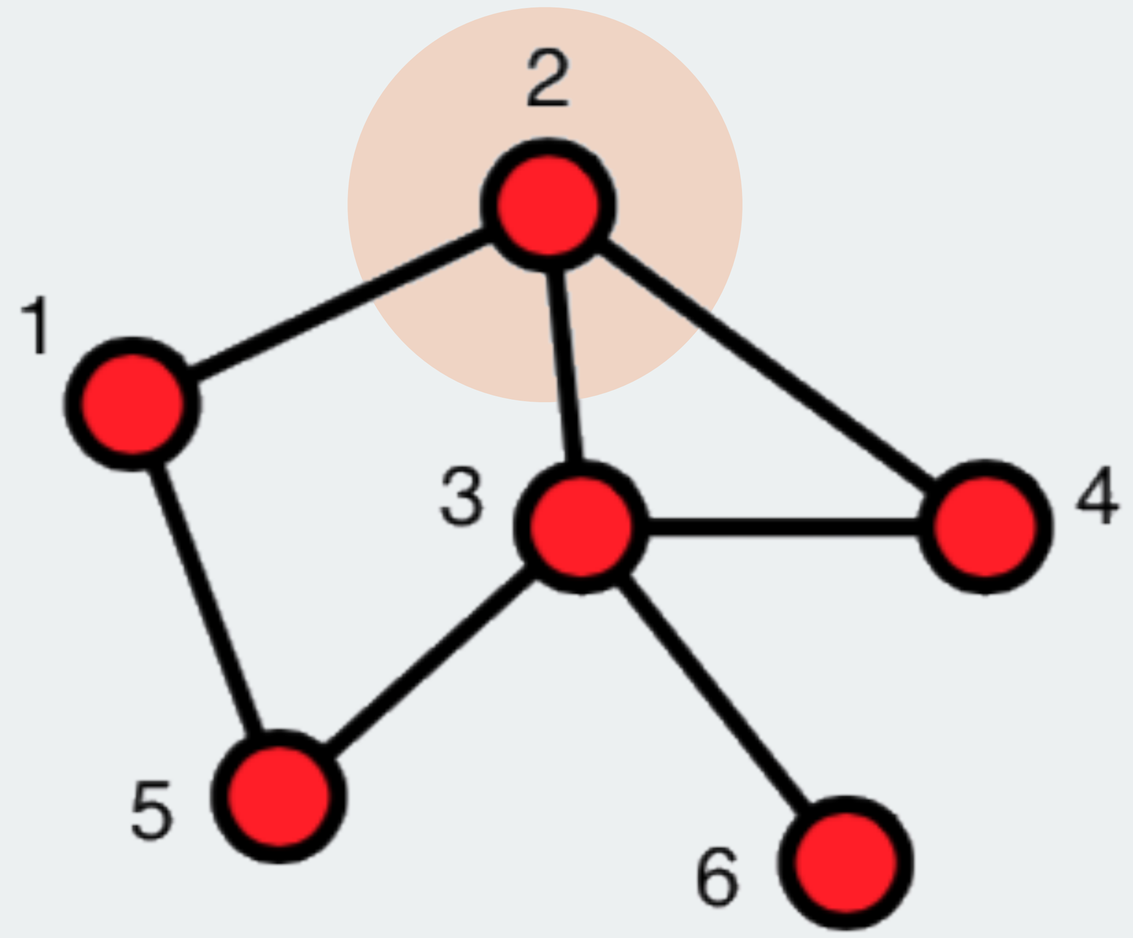


Undirected

Unweighted

No self-loops

Simple networks



Undirected

Unweighted

No self-loops

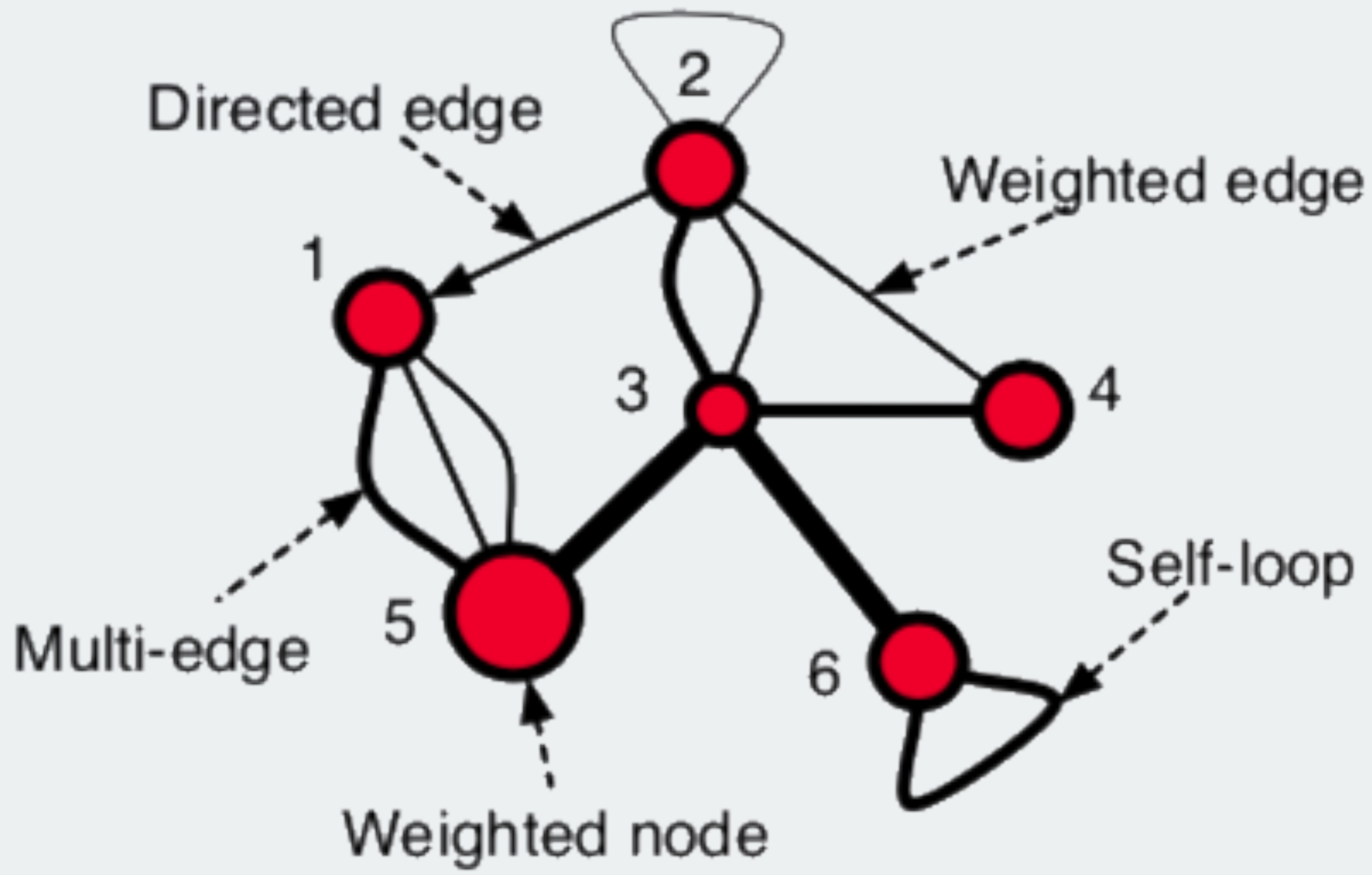
adjacency matrix, A

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

Directed and weighted networks

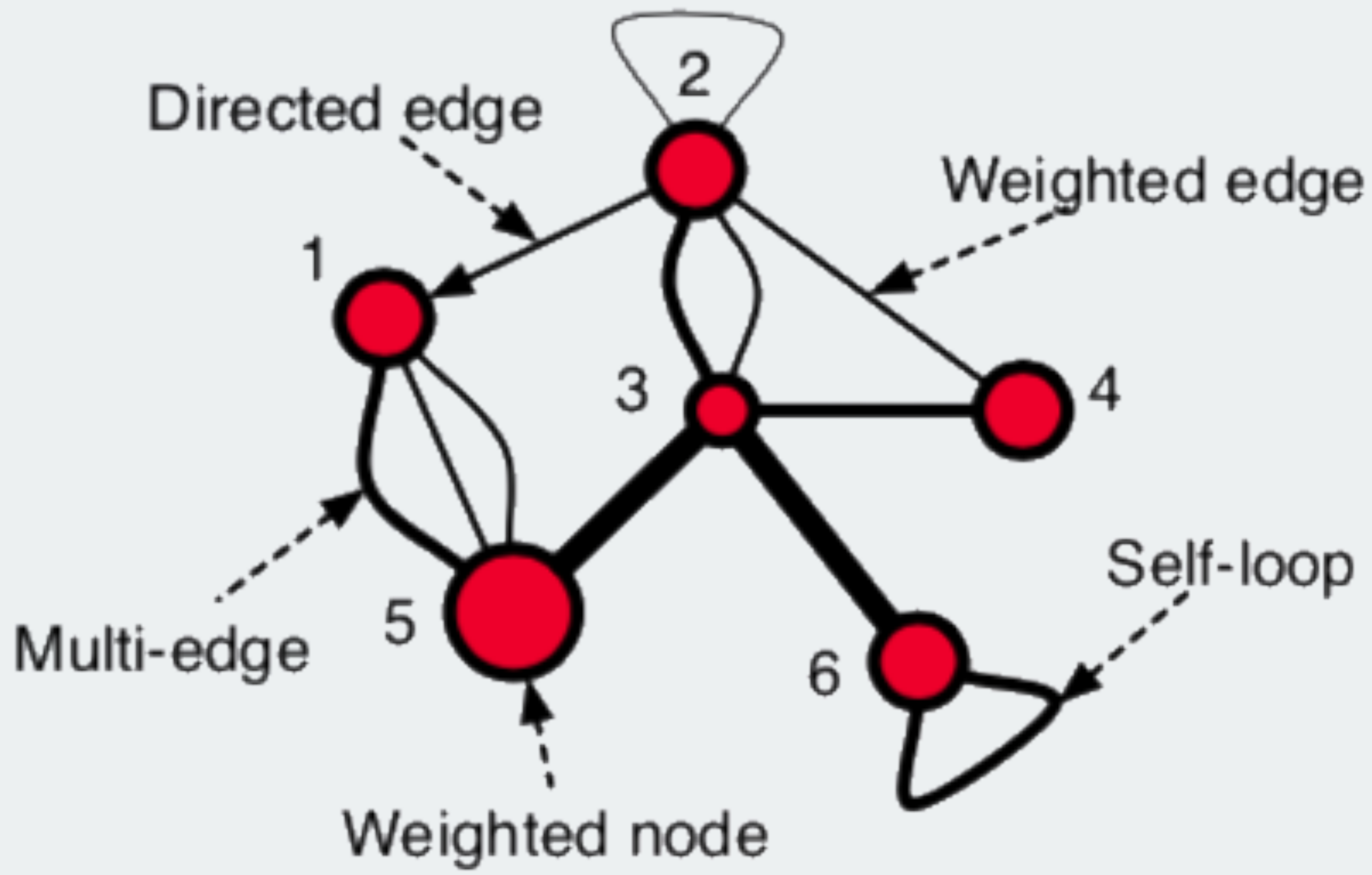


~~Undirected~~

~~Unweighted~~

~~No self-loops~~

Directed and weighted networks



~~Undirected~~

~~Unweighted~~

~~No self-loops~~

adjacency matrix, A

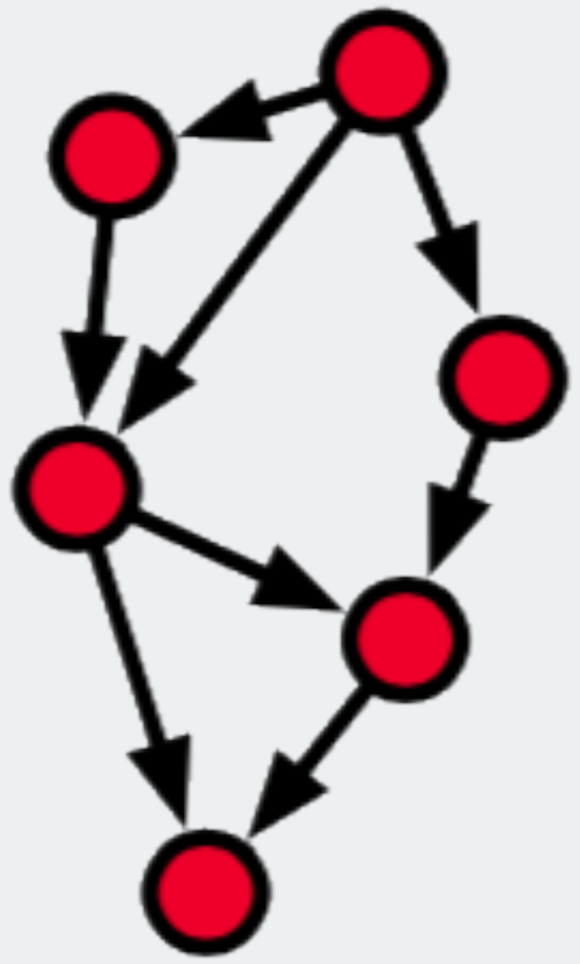
A	1	2	3	4	5	6
1	0	0	0	0	$\{1,1,2\}$	0
2	1	$\frac{1}{2}$	$\{2,1\}$	1	0	0
3	0	$\{2,1\}$	0	2	4	4
4	0	1	2	0	0	0
5	$\{1,1,2\}$	0	4	0	0	0
6	0	0	4	0	0	2

adjacency list

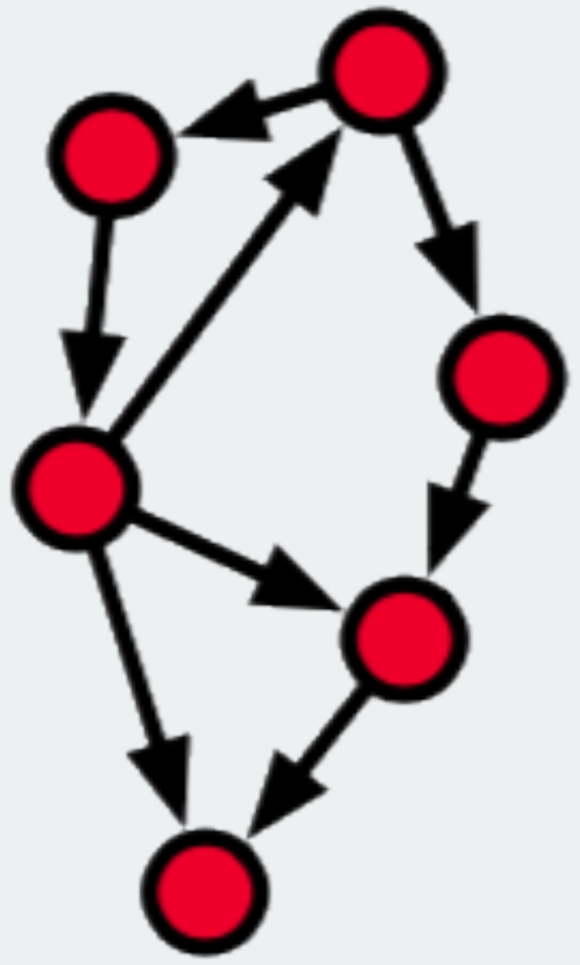
A	
1	$\rightarrow \{(5,1),(5,1),(5,2)\}$
2	$\rightarrow \{(1,1),(2,\frac{1}{2}),(3,2),(3,1),(4,1)\}$
3	$\rightarrow \{(2,2),(2,1),(4,2),(5,4),(6,4)\}$
4	$\rightarrow \{(2,1),(3,2)\}$
5	$\rightarrow \{(1,1),(1,1),(1,2),(3,4)\}$
6	$\rightarrow \{(3,4),(6,2)\}$

Directed networks

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



directed acyclic graph

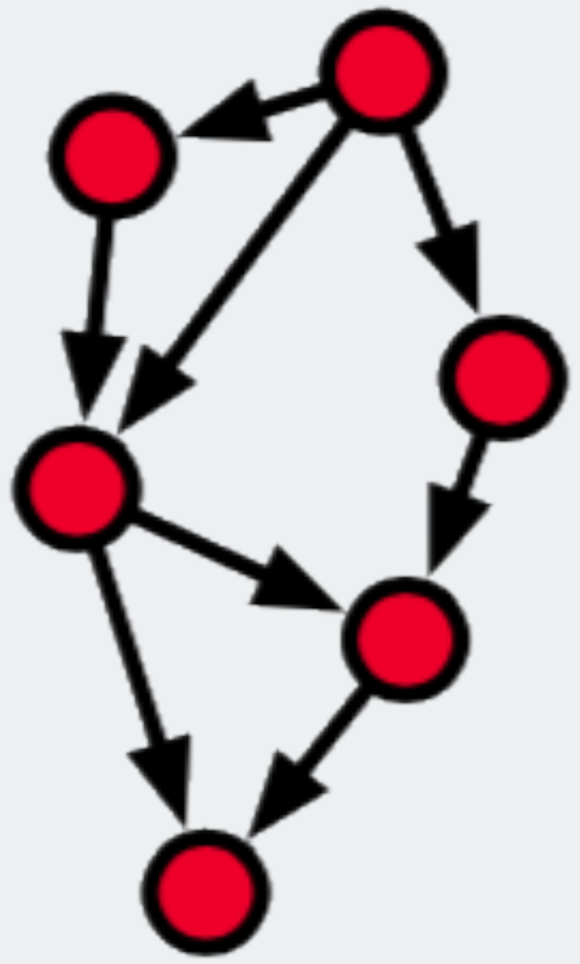


directed graph

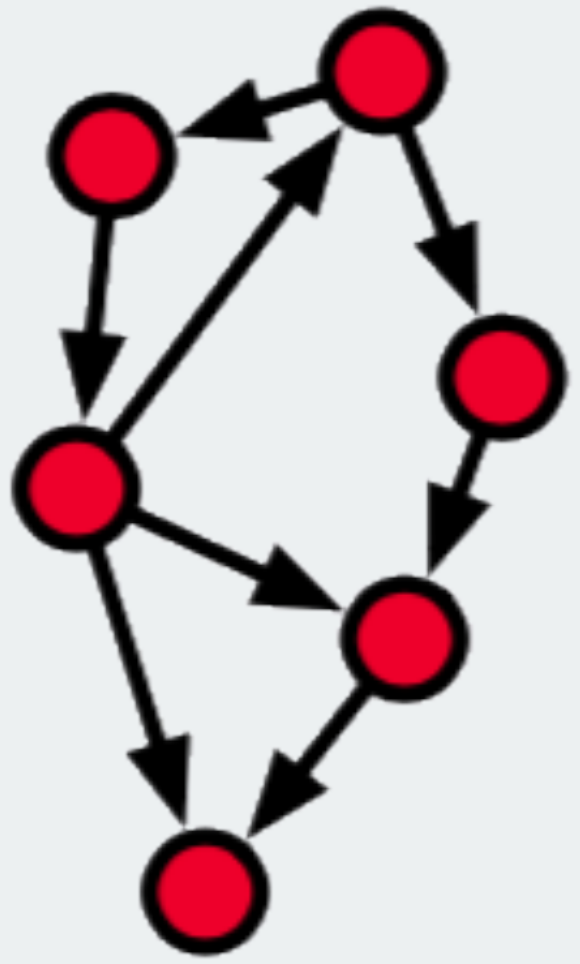
Directed networks

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

- Citation networks
- (some) Foodwebs
- Epidemiological
- Family trees
- IOTA “tangle” tree



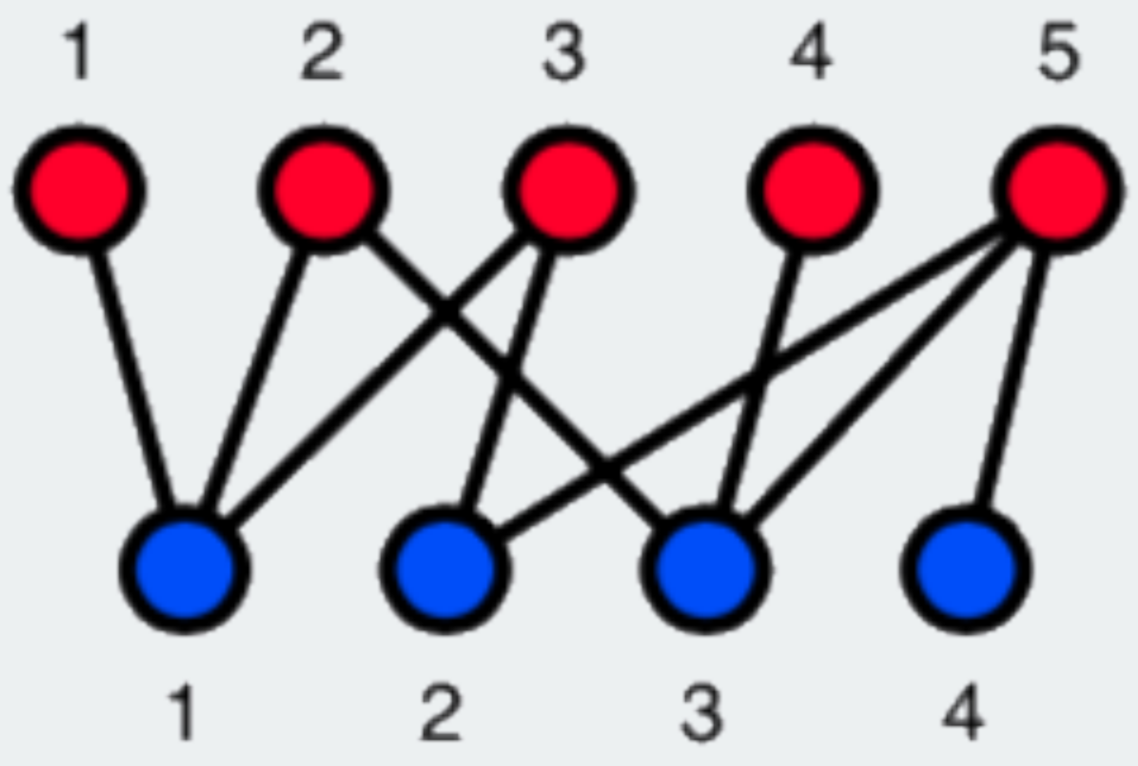
directed acyclic graph



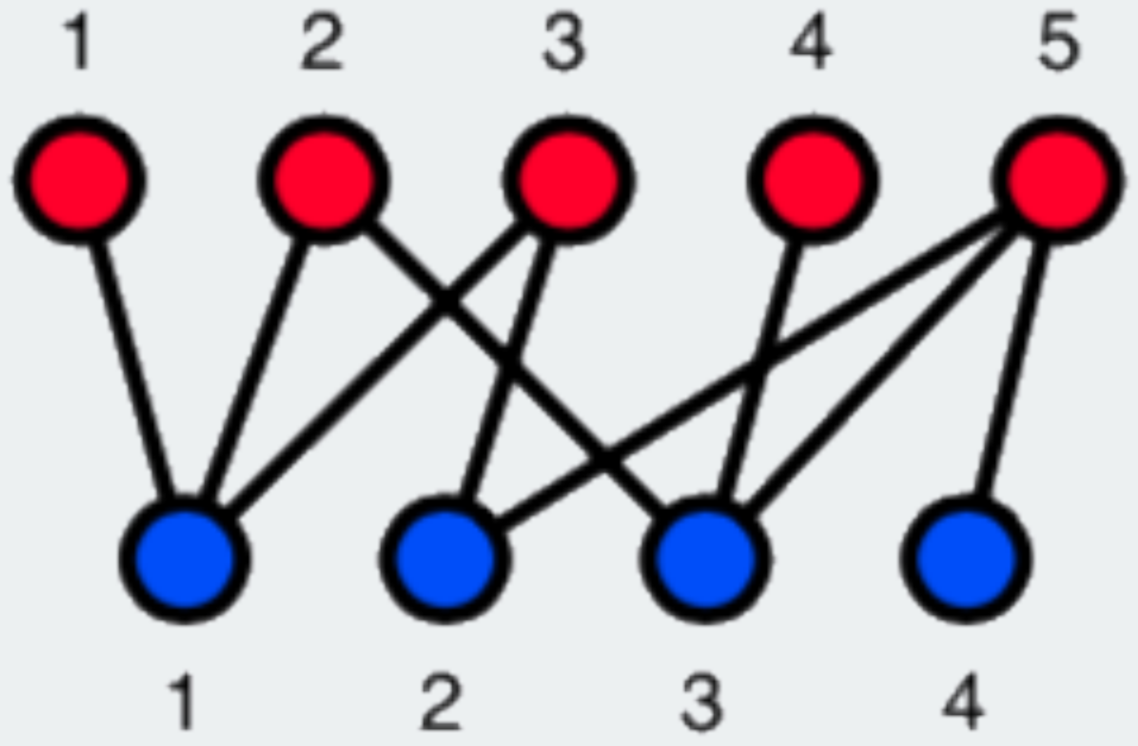
directed graph

- WWW
- Friendship?
- Flow of goods and information
- Payments
- Dominance
- Neuronal activity

Bipartite networks

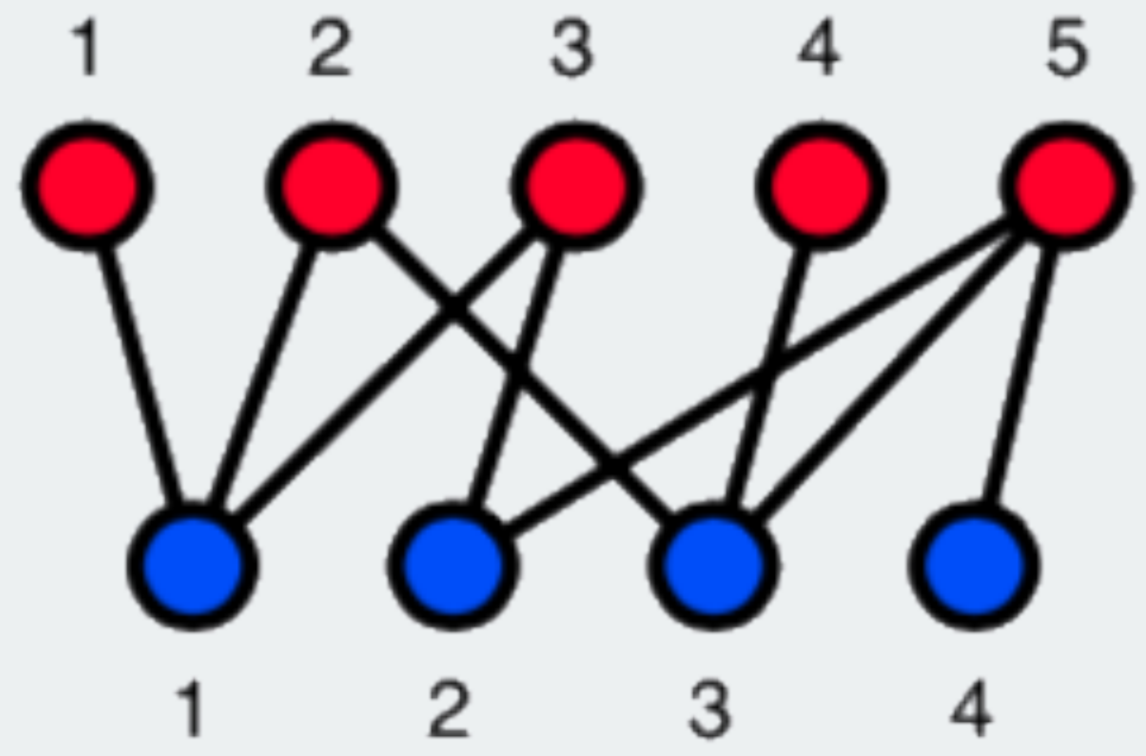


Bipartite networks



no within-type edges

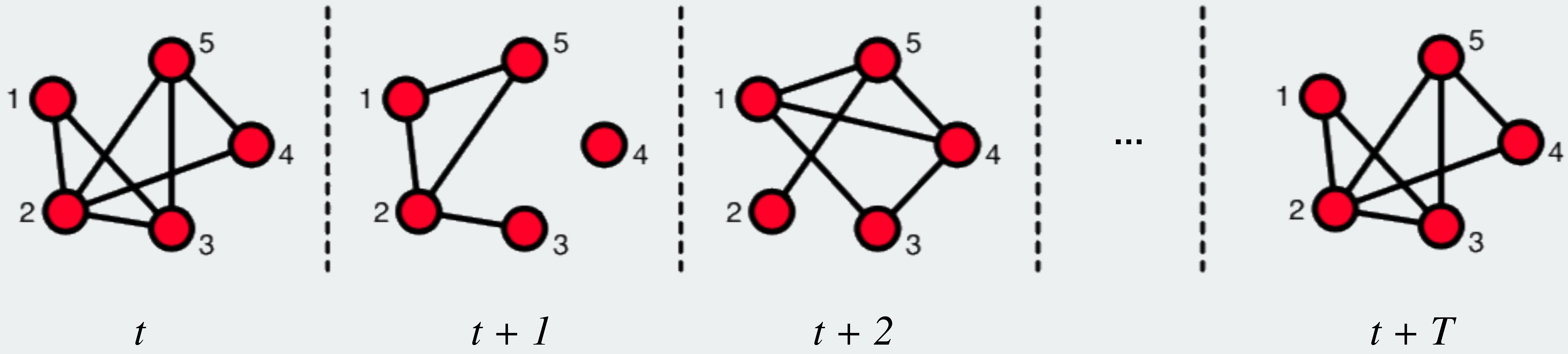
Bipartite networks



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

Temporal networks

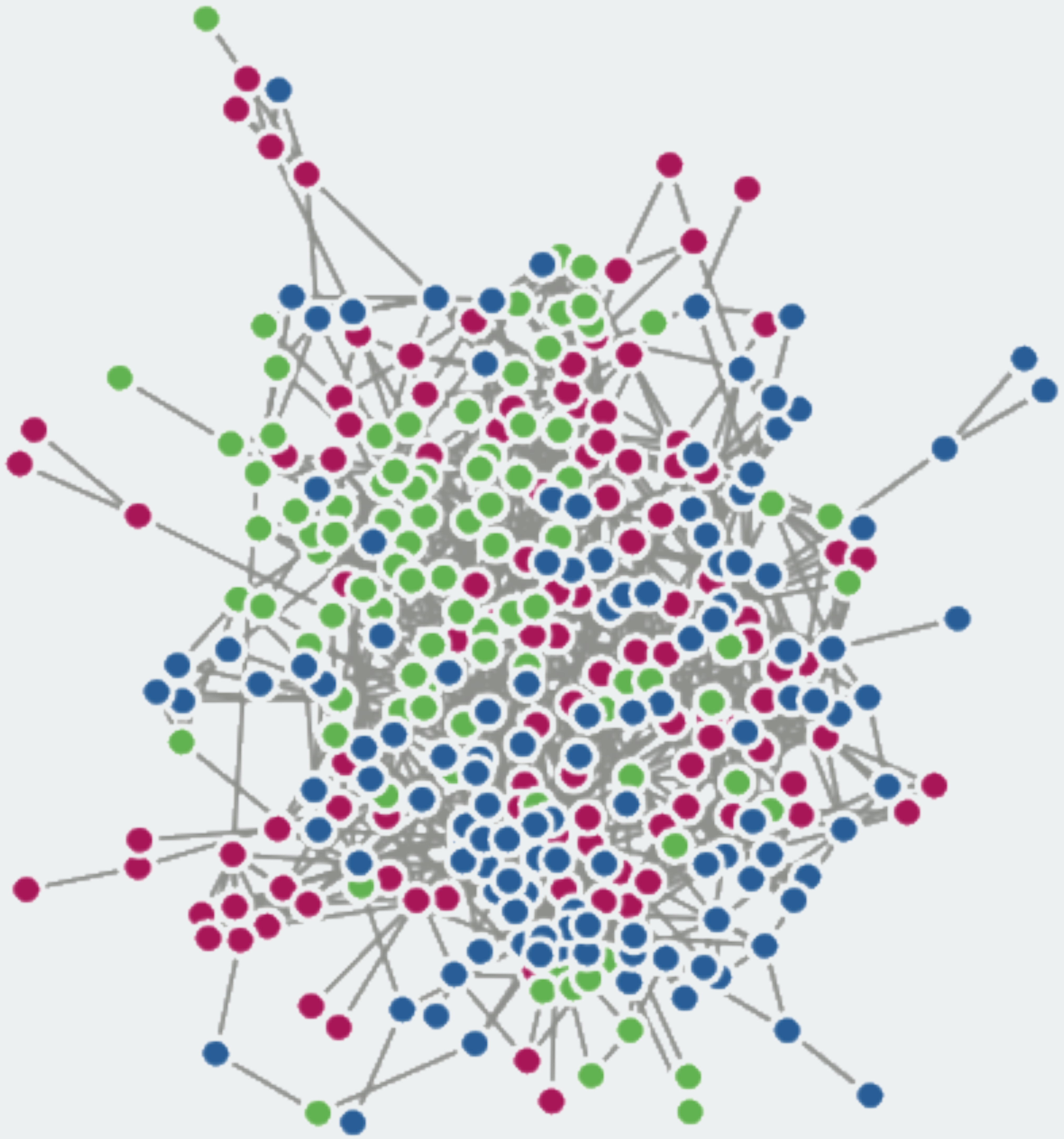


Examples: Most (if not all) networks, are in reality temporal

Describing networks

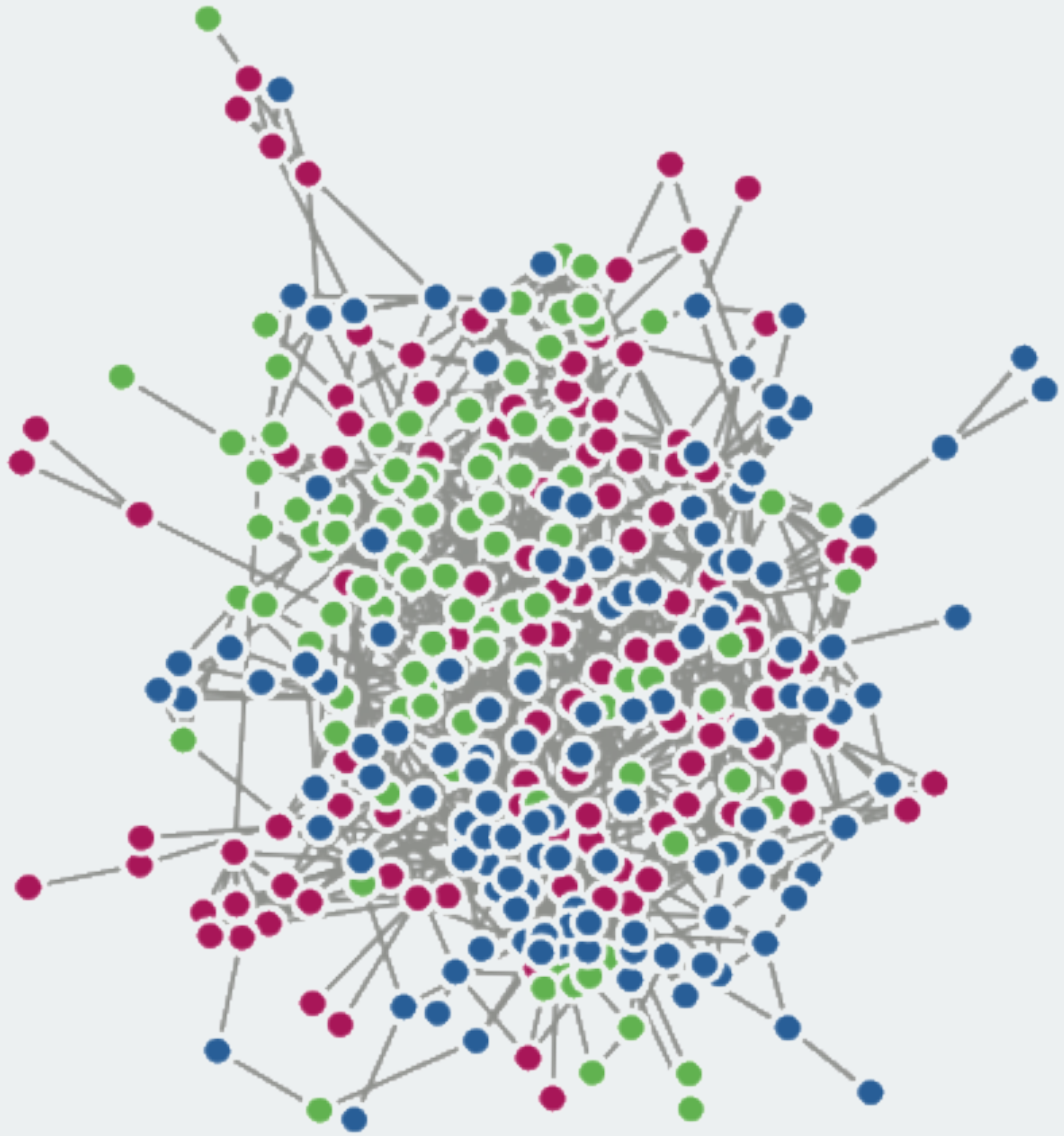
Describing networks

What does a network "look like"?



Describing networks

What does a network "look like"?



Questions:

- **How are the edges organized?**
- **How do vertices differ?**
- **Do locations in the network differ?**
- **Are there any underlying patterns?**

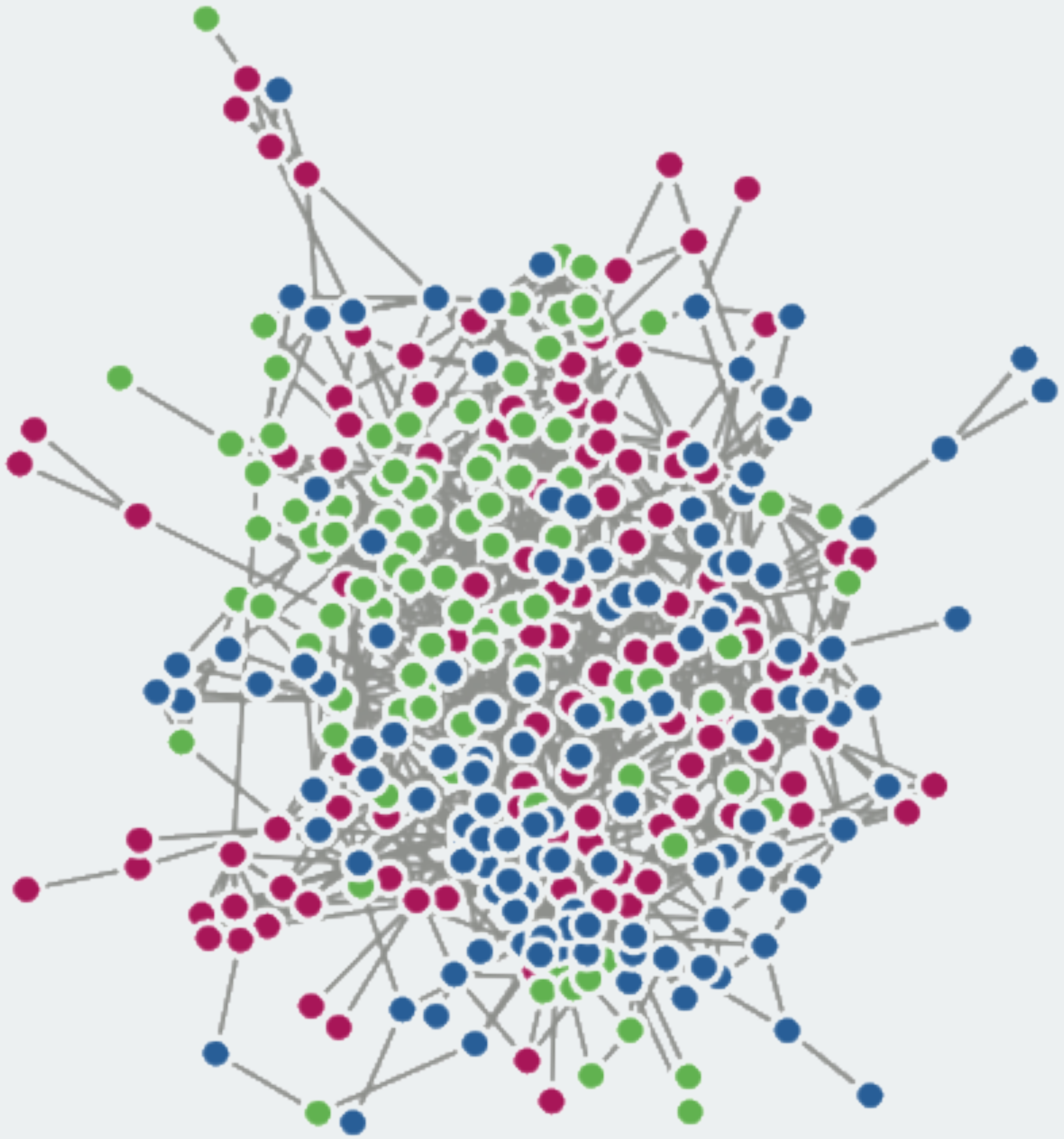
Describing networks

First step: Describe its features

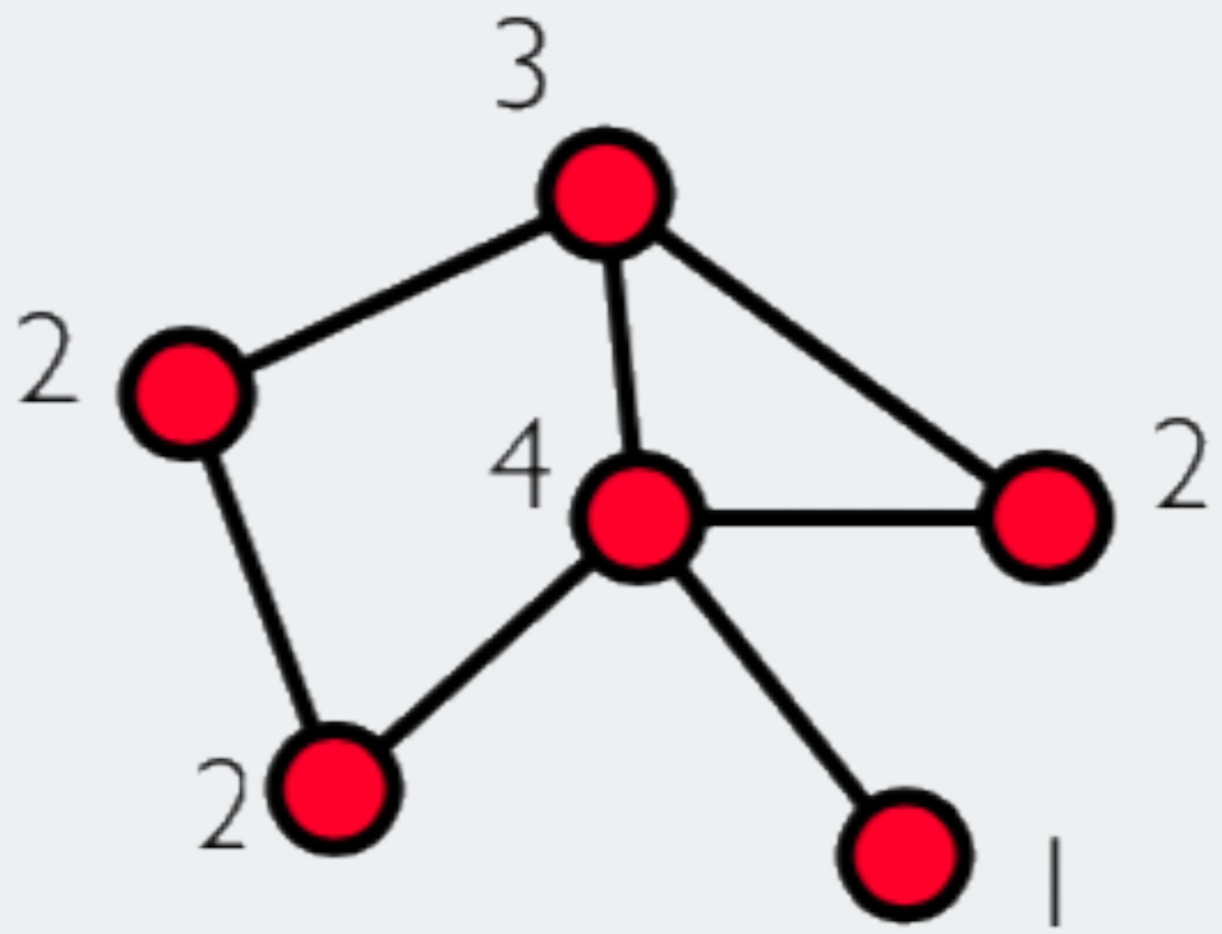
- **Degree distribution**
- **Short-loop density (triangles, etc.)**
- **Shortest paths (diameter, etc.)**
- **Vertex positions (centrality, etc.)**

Second step: Understand the system

- **Correlations between features**
- **Correlations between features and other variables**
- **Testing network structure against null models**
- **Finding clusters and other mesoscale structures**



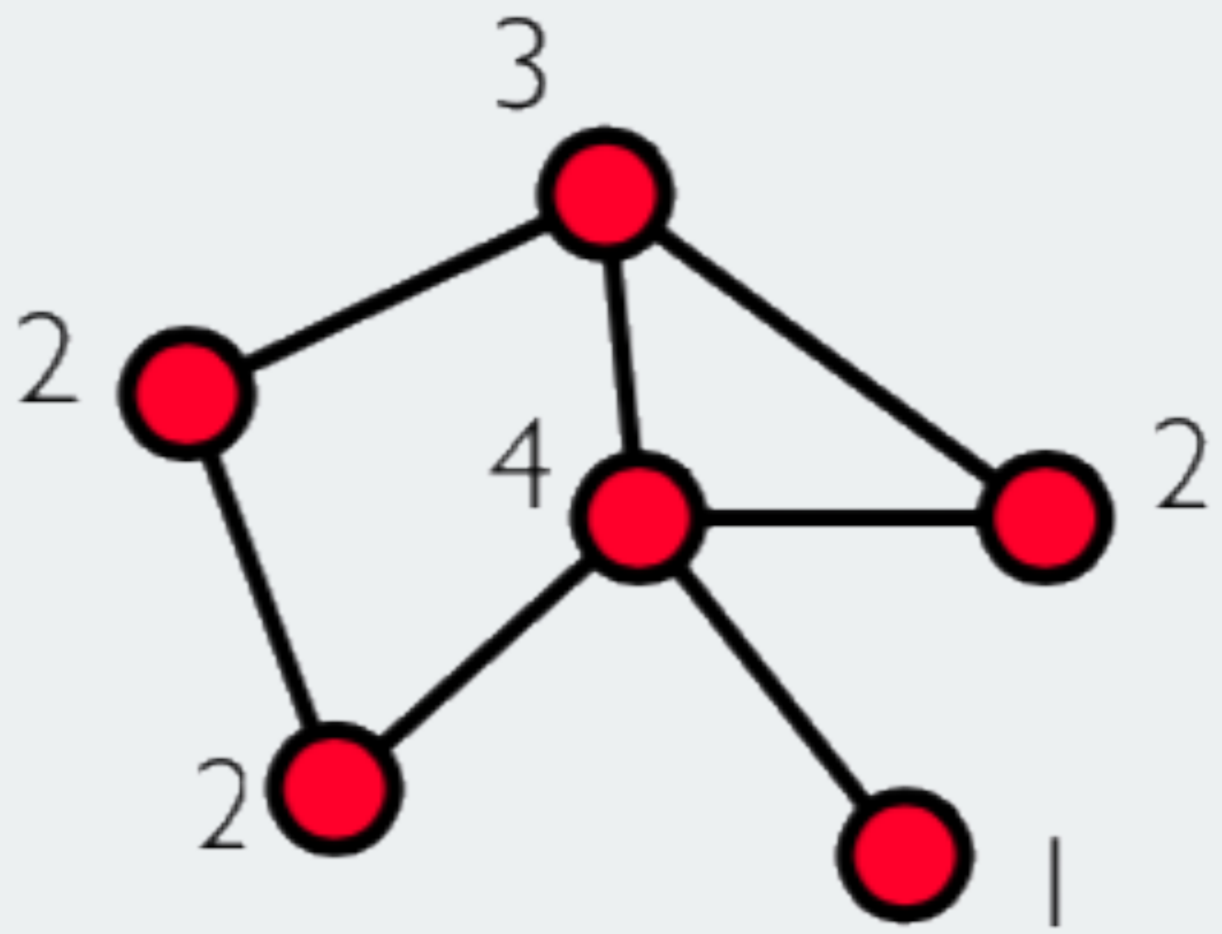
Degree



degree:
number of connections k

$$k_i = \sum_j A_{ij}$$

Degree



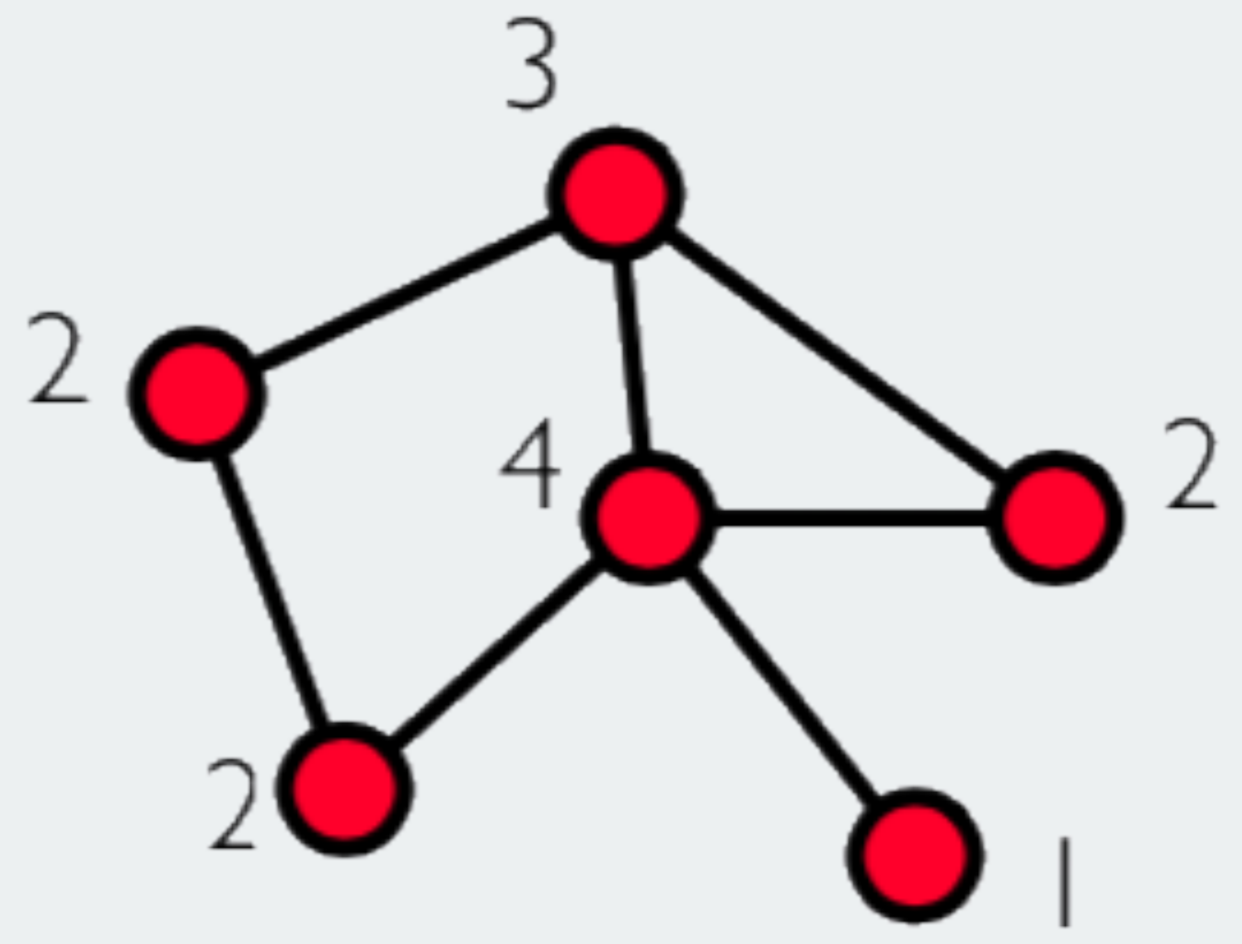
degree:
number of connections k

$$k_i = \sum_j A_{ij}$$

number of edges $$m = \frac{1}{2}\sum_{i=1}^{n} k_i$$

mean degree $$\langle k \rangle = \frac{1}{n}\sum_{i=1}^{n} k_i = \frac{2m}{n}$$

Degree



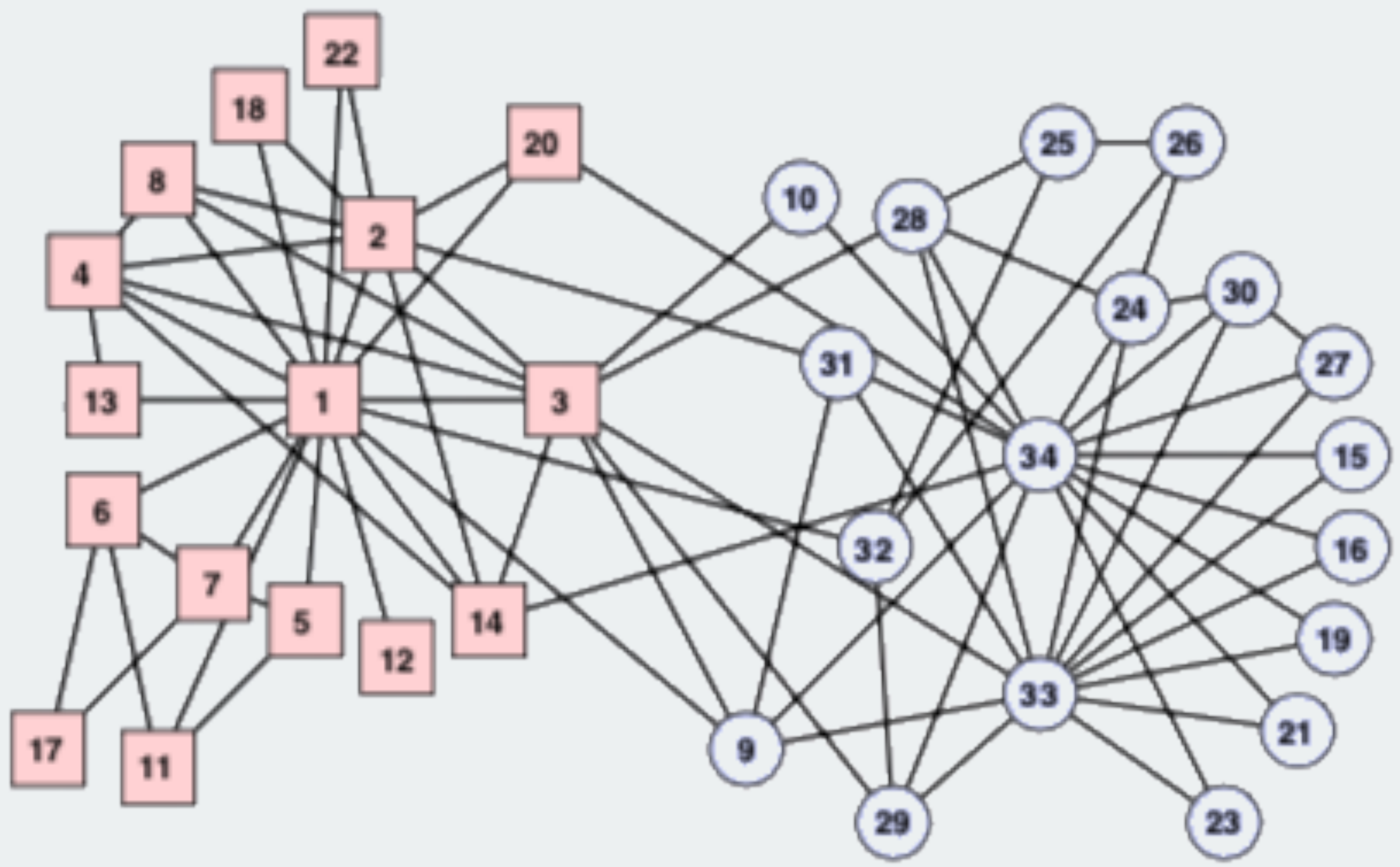
degree:
number of connections k

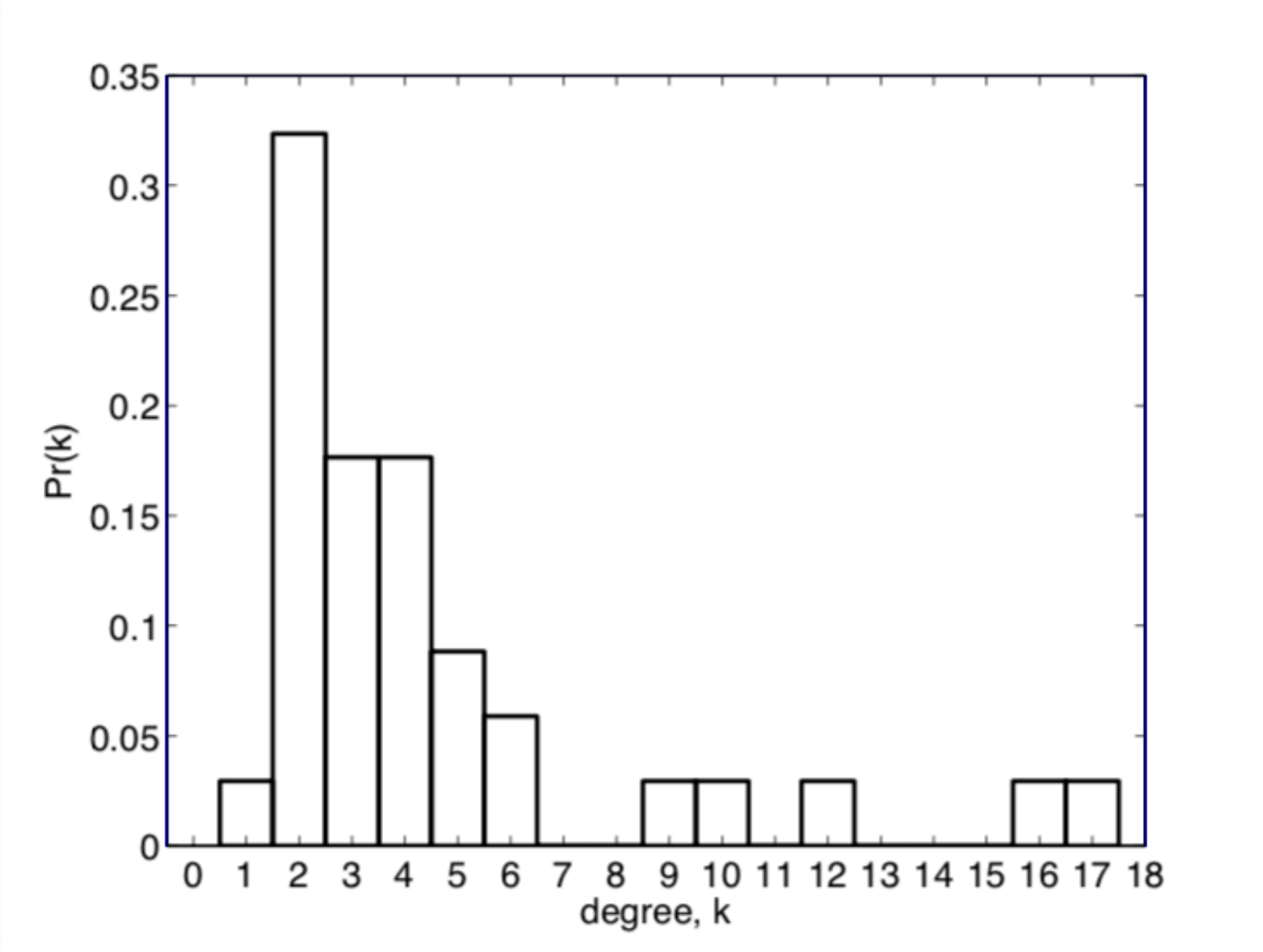
$$k_i = \sum_j A_{ij}$$

degree sequence $\{1, 2, 2, 2, 3, 4\}$

degree distribution $\Pr(k) = \left[\left(1, \frac{1}{6}\right), \left(2, \frac{3}{6}\right), \left(3, \frac{1}{6}\right), \left(4, \frac{1}{6}\right)\right]$

Degree distribution





Degree distribution