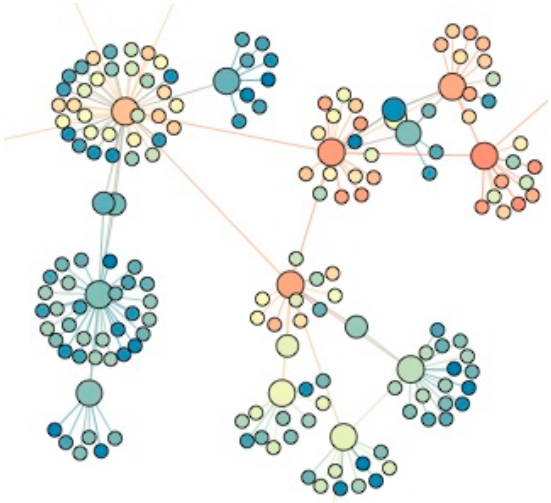


Social Network Analysis

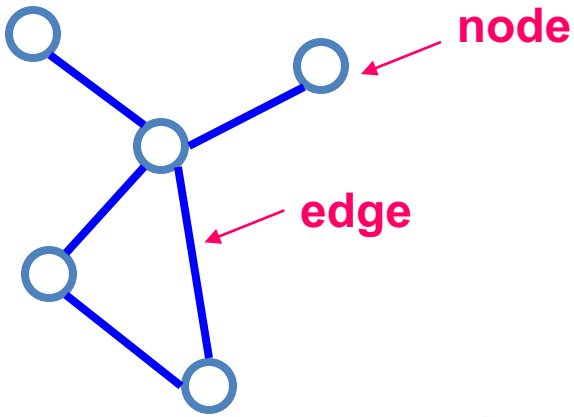
Truc Viet “Joe” Le

Carnegie Mellon University
Heinz College

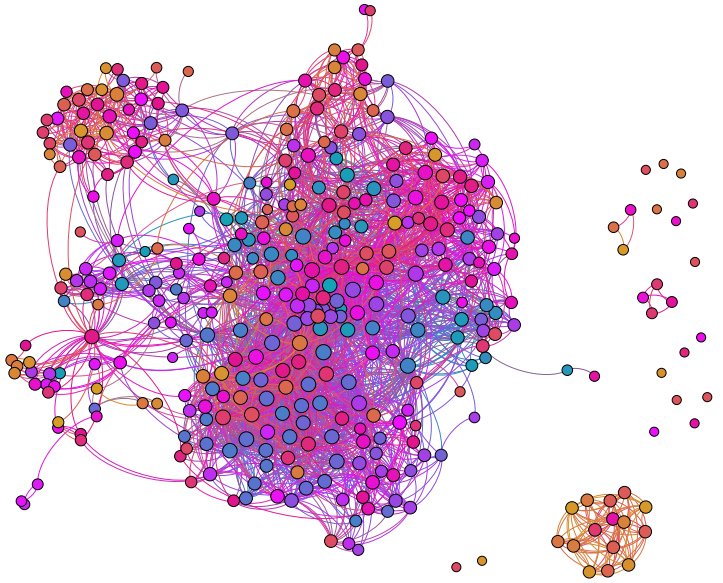
What are networks?

- Networks are sets of **nodes** connected by **edges**.

"Network" = "Graph"



points	lines	context
vertices	edges, arcs	math
nodes	links	computer science
sites	bonds	physics
actors	ties, relations	sociology



Networks model the **complexity** of the **real world**

Network elements: Edges

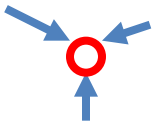
- Directed (also called arcs, links)
  - A -> B
    - A likes B, A gave a gift to B, A is B's child
- Undirected
  - A <-> B or A – B
    - A and B like each other
    - A and B are siblings
    - A and B are co-authors

Edge attributes

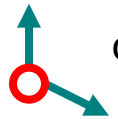
- Examples
  - weight (e.g., frequency of communication)
  - ranking (best friend, second best friend…)
  - type (friend, relative, co-worker)
  - properties depending on the structure of the rest of the graph: e.g., betweenness

Network elements: Nodes

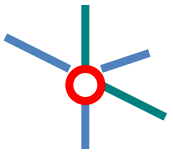
- Node network properties
  - from immediate connections
    - indegree
      how many directed edges (arcs) are incident on a node
    - outdegree
      how many directed edges (arcs) originate at a node
    - degree (in or out)
      number of edges incident on a node



indegree=3



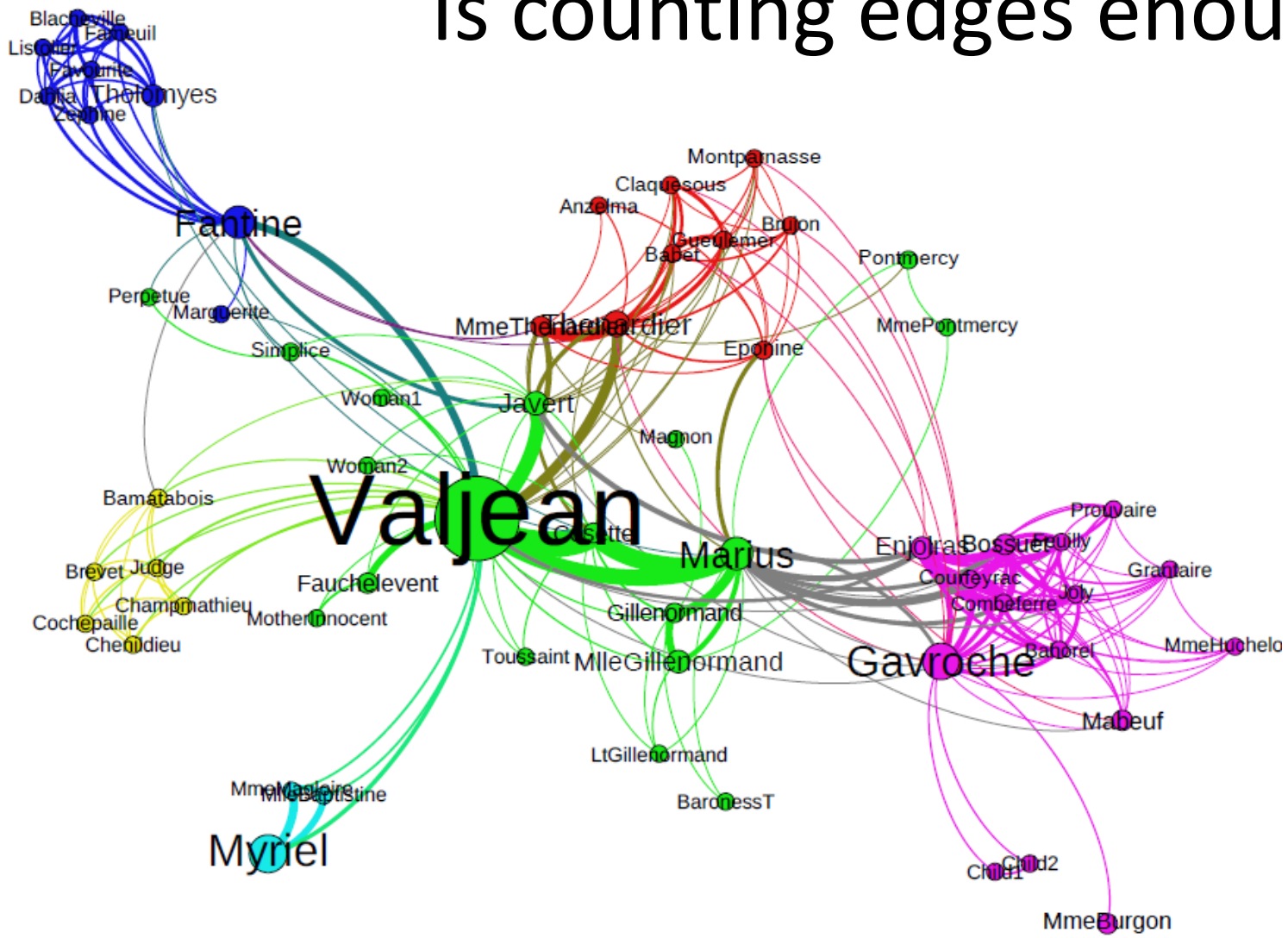
outdegree=2



degree=5

  - from the entire graph
    - centrality (e.g., betweenness, closeness, etc.)

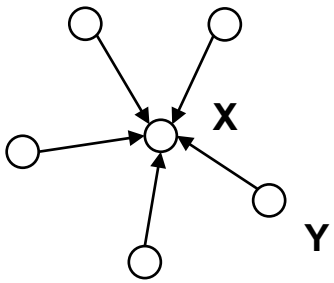
Is counting edges enough?



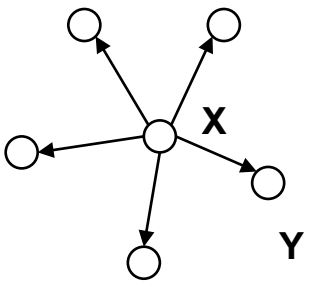
Characters and their relationships in *Les Misérables*

Different notions of **centrality**

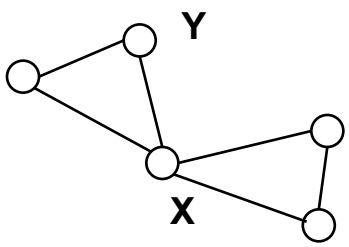
In each of the following networks, X has higher centrality than Y according to a particular measure.



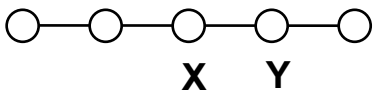
indegree



outdegree



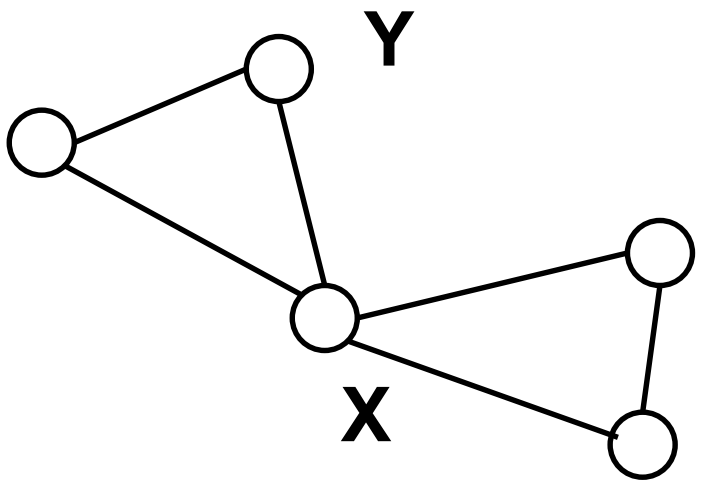
betweenness



closeness

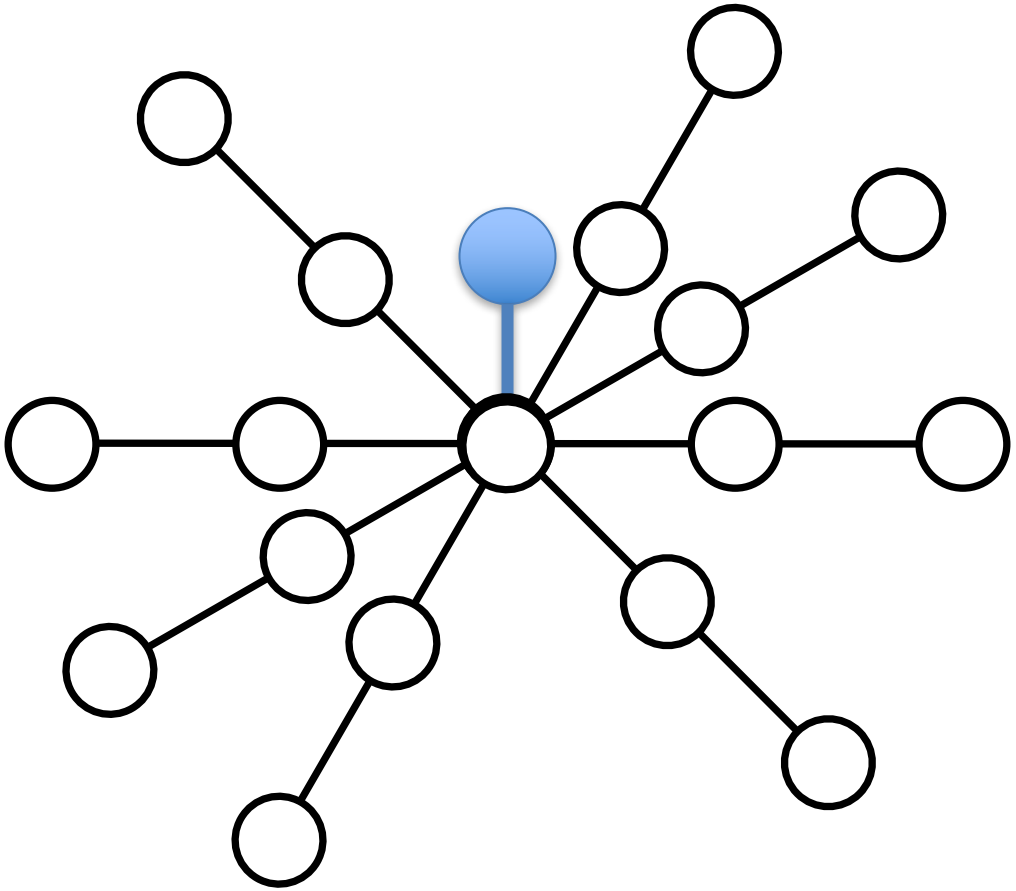
Betweenness: Capturing brokerage

- How many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?



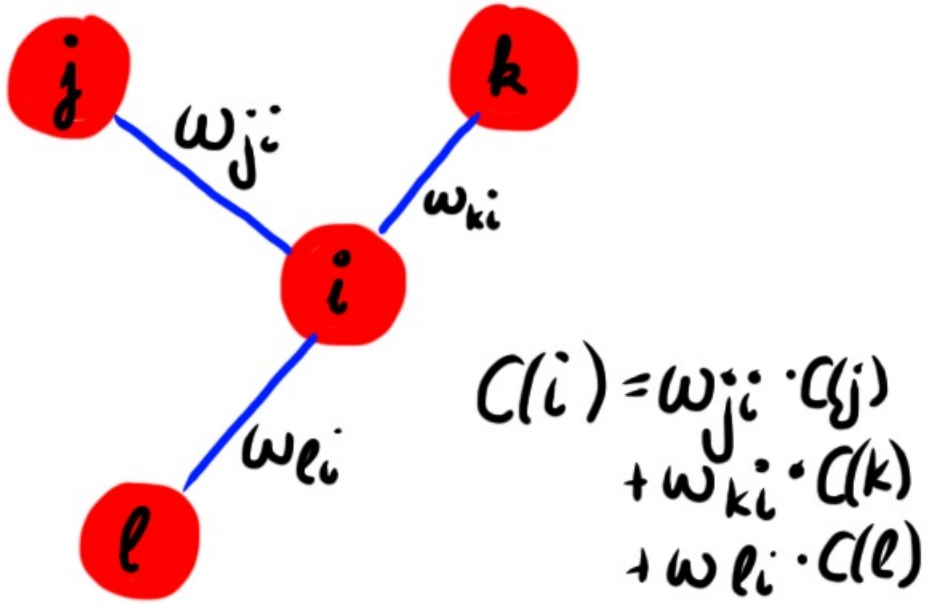
Closeness

- What if it's *not* so important to have many direct friends?
  - Or be "between" others
- But one still wants to be in the "middle" of things, not too far from the center

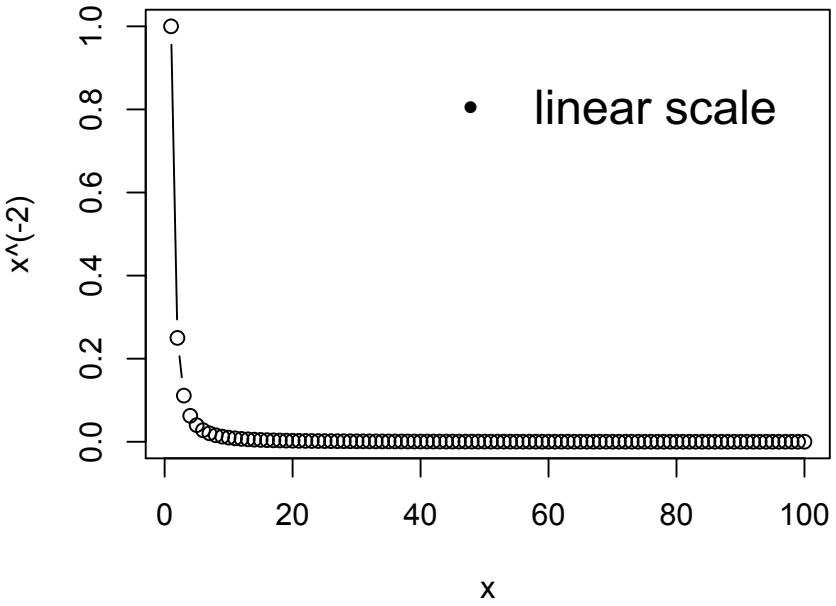


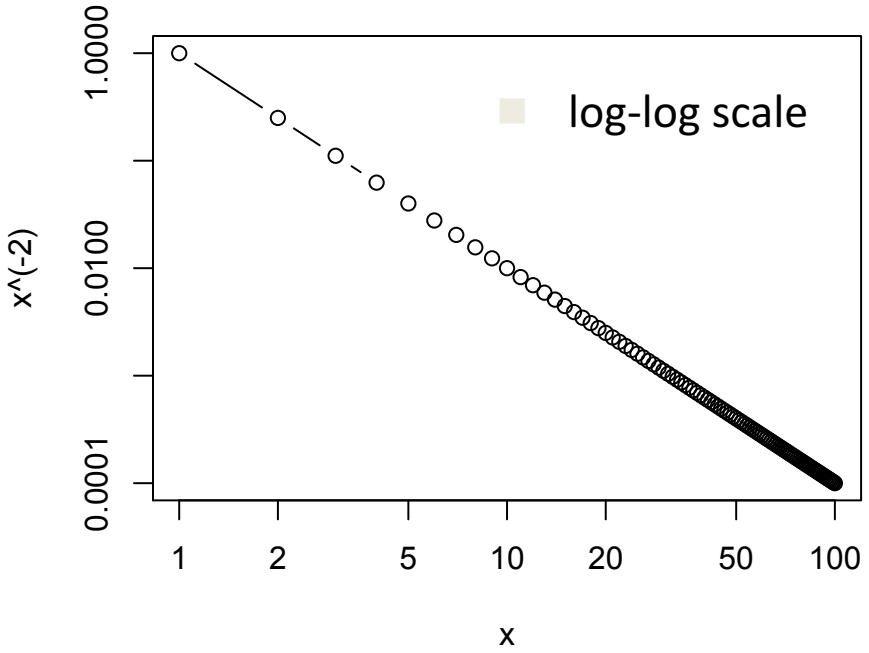
Eigenvector centrality

- How central you are depends on how central your neighbors (or friends) are
  - Ingenuity of the (Google's) PageRank algorithm → PageRank centrality



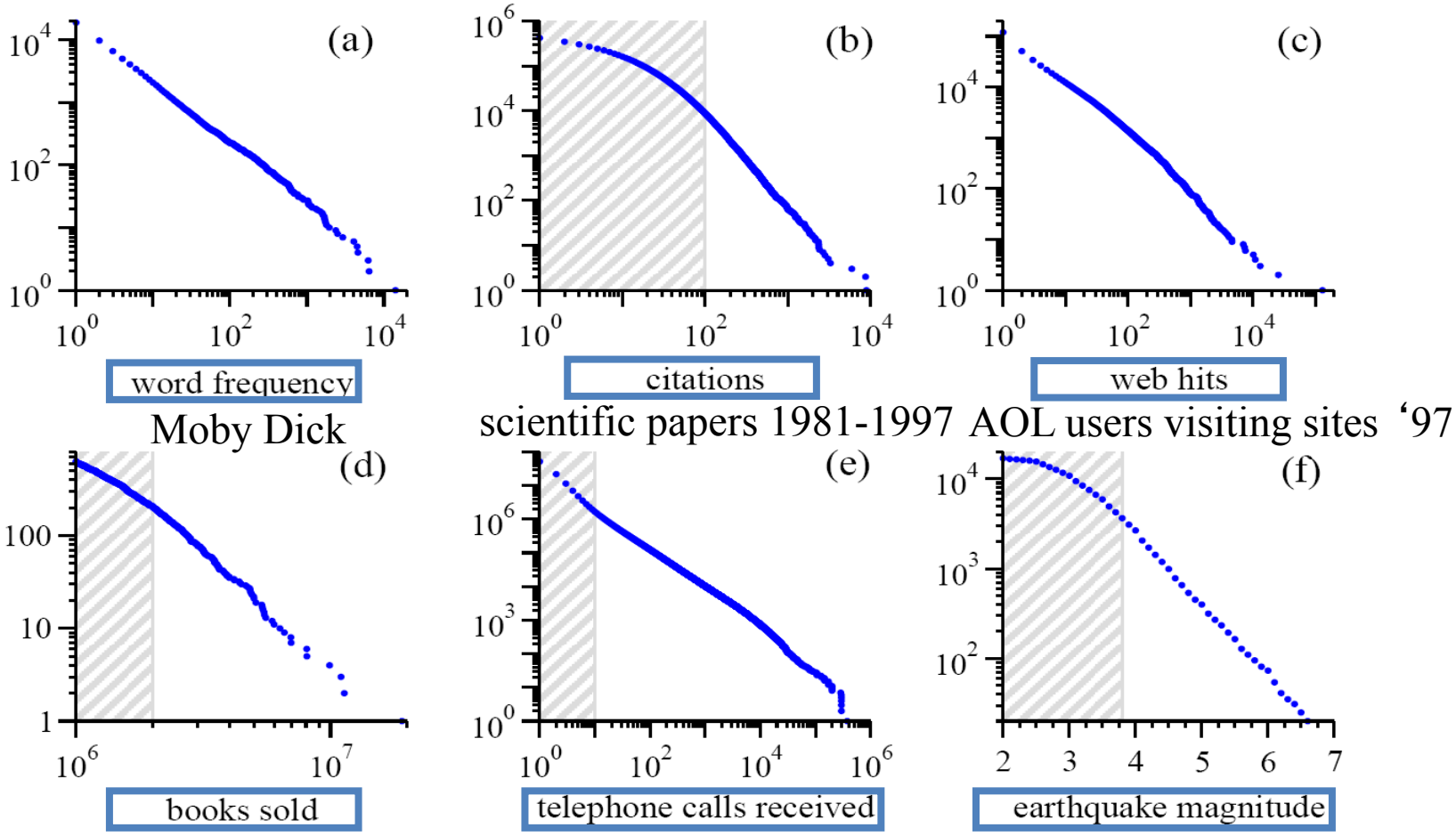
Power-law distribution





- High skew (asymmetry)
- Long tail
- Straight line on a log-log plot

Power laws are seemingly everywhere!



Source: MEJ Newman, ' Power laws, Pareto distributions and Zipf' s law' , *Contemporary Physics* **46**, 323–351 (2005)

The power law

- Straight line on a log-log plot

$$\ln(p(x)) = c - \alpha \ln(x)$$

- Exponentiate both sides to get that *p(x)*, the probability of observing an item of size ‘x’ is given by

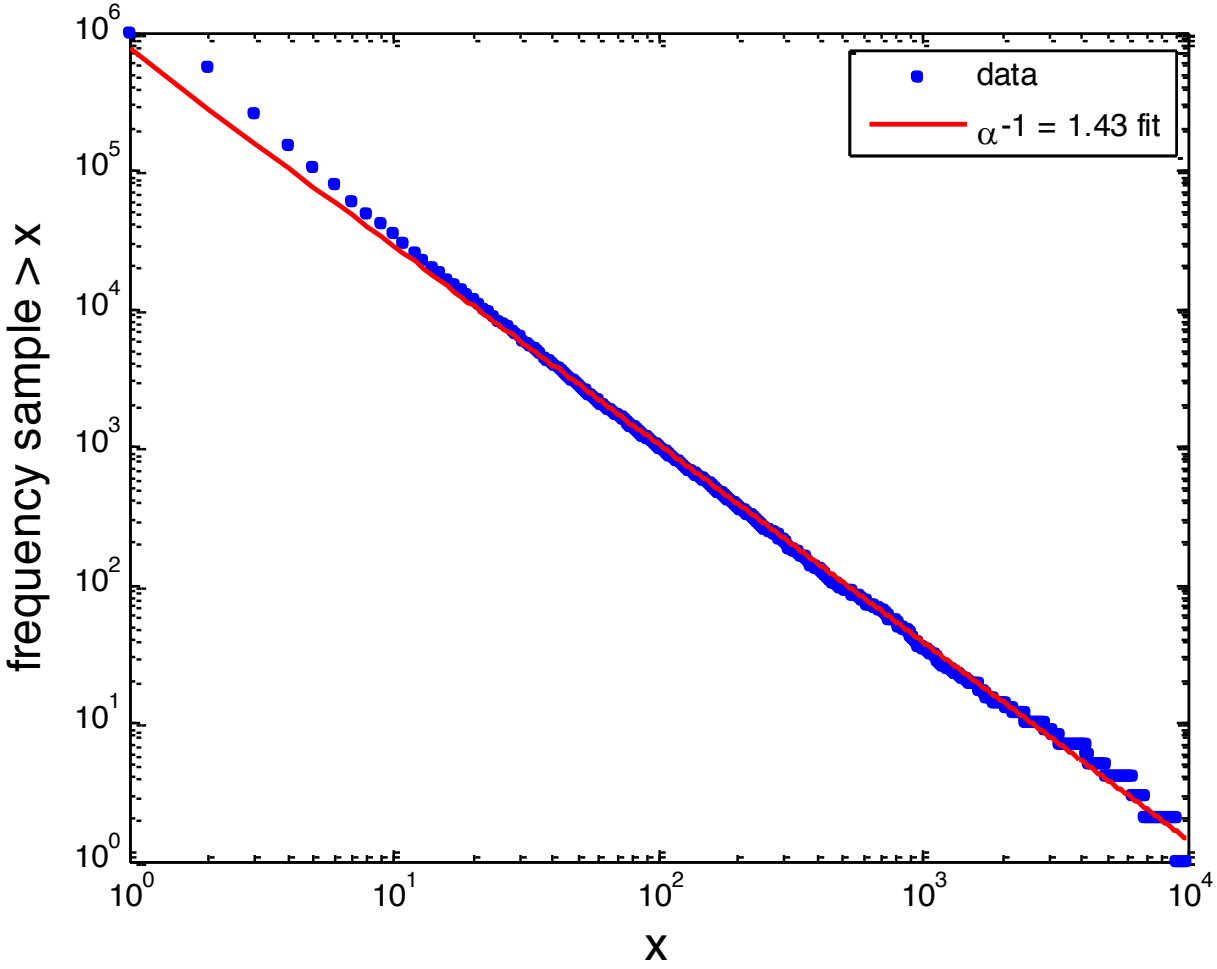
$$p(x) = Cx^{-\alpha}$$

normalization **constant** (probabilities over all *x* must sum to 1)

power law **exponent** α

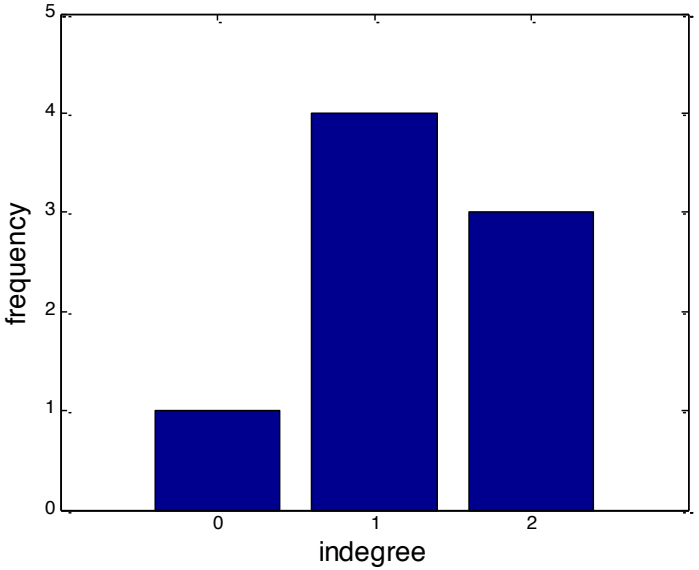
Fitting PL via regression to the **Cumulative** distribution

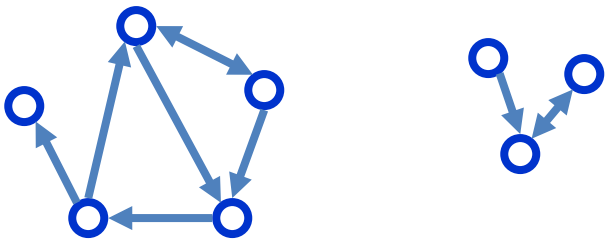
- Fitted exponent (2.43) is close to actual (2.5)



Degree distribution

- Degree distribution: A frequency count of the occurrences of each degree in the entire graph.
  - Indegree distribution:
    - [(2,3) (1,4) (0,1)]
  - Outdegree distribution:
    - [(2,4) (1,3) (0,1)]
  - (Undirected) distribution:
    - [(3,3) (2,2) (1,3)]





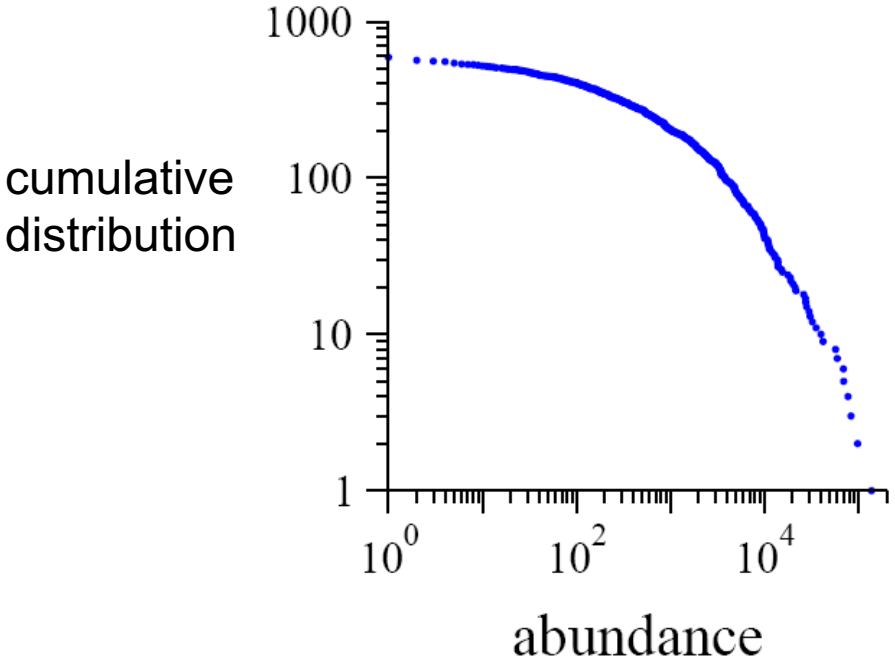
Two connected components

Many real-world **networks** are power law!

	exponent α (in/out **degree**)
film actors – porn industry w/ high centrality	2.3
telephone call graph	2.1
email networks	1.5/2.0
sexual contacts	3.2
WWW	2.3/2.7
internet	2.5
peer-to-peer	2.1
metabolic network	2.2
protein interactions	2.4

But not everything is a power law…

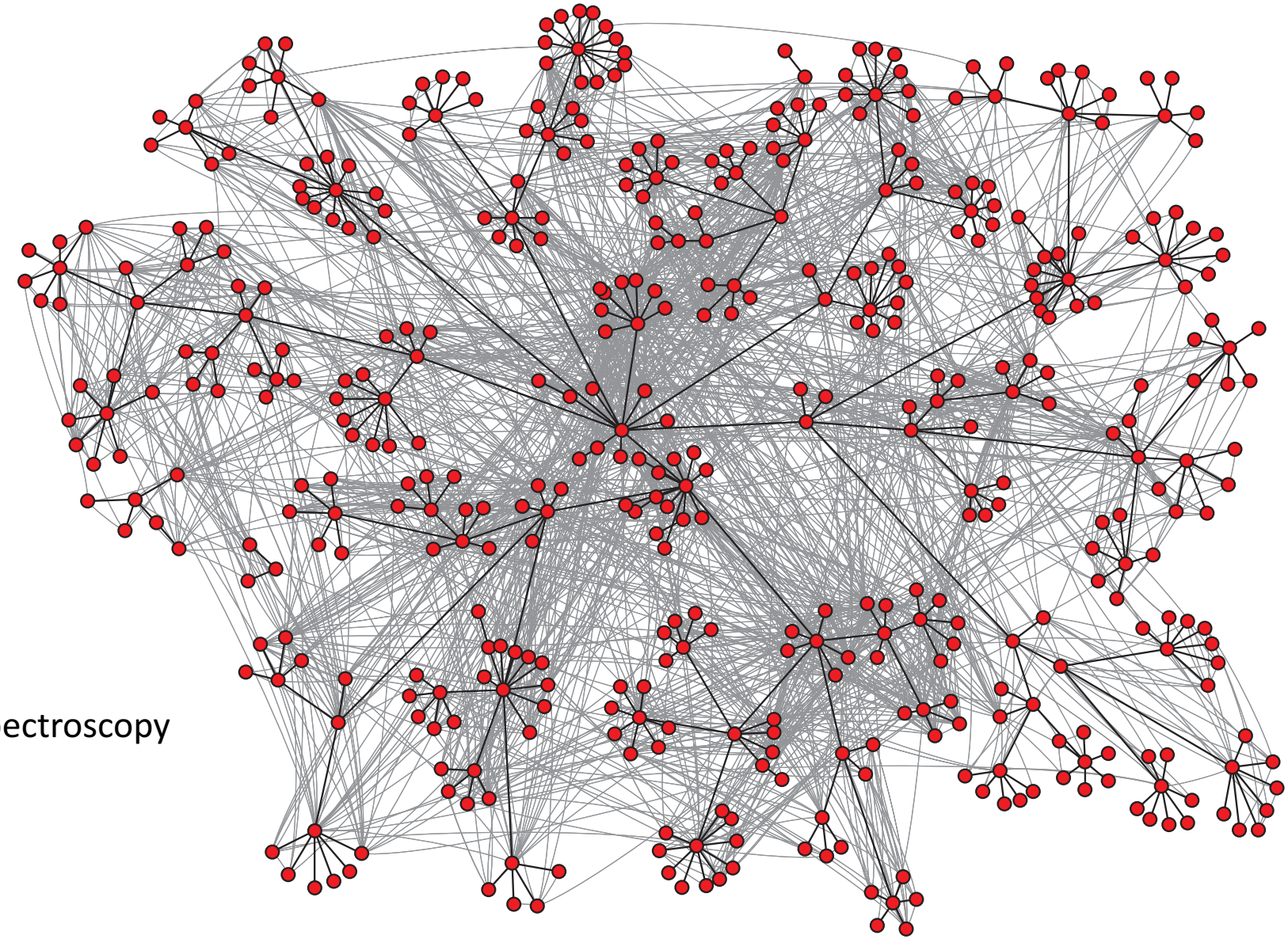
- Number of sightings of 591 bird species in the North American Bird survey in 2003.



- another example:
  - size of wildfires (in acres)

Source: MEJ Newman, ' Power laws, Pareto distributions and Zipf' s law' , *Contemporary Physics* **46**, 323–351 (2005)

Network communities



Example:

- Email spectroscopy

Community detection

- Social and other networks have a natural community structure
- We want to discover this structure rather than impose a certain size of community or fix the number of communities
  - Without "looking", can we **discover** community structure in an automated way?
  - **Modularity** algorithm