Network Theory and Dynamic Systems

04. Hubs

SOSE 2025

Dr. -Ing. Stefania Zourlidou

Institute for Web Science and Technologies
Universität Koblenz

Recap from Previous Lecture

- Birds of a Feather
- Paths and Distances
- Connectedness and Components
- Trees
- Finding Shortest Paths
- Social Distance
- Six Degrees of Separation
- Small Worlds
- Friend of a Friend

Objectives of this Lecture

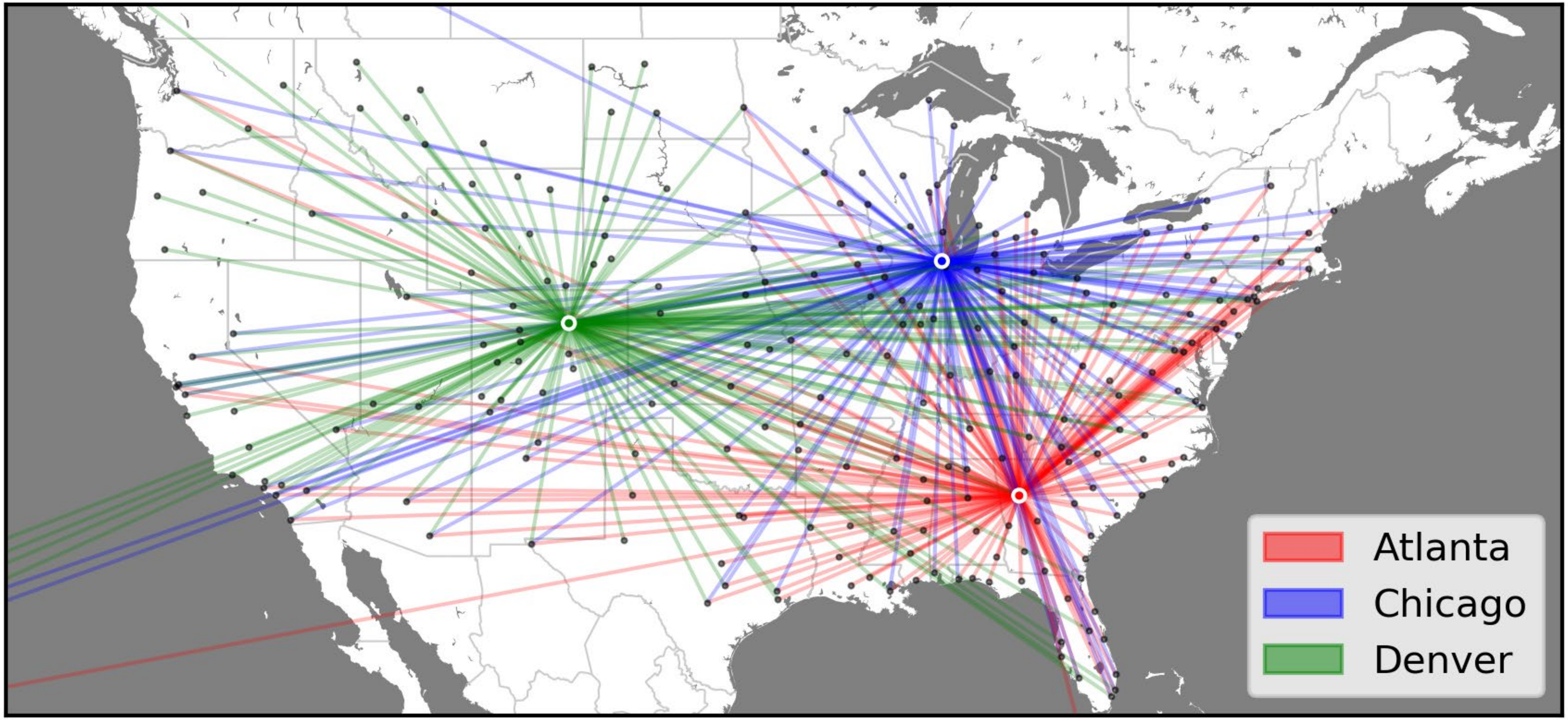
universität koblenz
weiter:denken

- Centrality Measures
- Centrality Distributions
- The Friendship Paradox
- Ultra-Small Worlds
- Robustness
- Core Decomposition

1. Centrality Measures

Real Networks are Heterogeneous (1/2)

universität koblenz
weiter:denken



Some nodes (and links) are much more important (**central**) than others!

Real Networks are Heterogeneous (2/2)

universität koblenz
weiter:denken

Can you think other examples?

Real Networks are Heterogeneous: Examples (1/4)

- **Social Communities**
  - **Nodes**: Individuals
  - **Links**: Social interactions or relationships
  - **Key Influencers**: Individuals who have a vast network of connections or significant influence
  - **Characteristic**: Like major airports, these individuals act as **hubs** in social networks, having more visibility and connections

Real Networks are Heterogeneous: Examples (2/4)

- **The Web**
  - **Nodes**: Websites
  - **Links**: Hyperlinks between websites
  - **Major Sites**: Google.com
  - **Characteristic**: Major sites like Google are visited by millions daily, making them central nodes in the web network, similar to major airport hubs in air travel

Real Networks are Heterogeneous: Examples (3/4)

- **Public Transportation Network**
  - **Nodes**: Bus and train stations
  - **Links**: Bus routes and train lines that connect these stations
  - **Major Hubs**: Central stations in major cities like Grand Central Station in New York or Union Station in Washington D.C.
  - **Characteristic**: These hubs serve thousands of passengers daily and provide connections to numerous local and regional transit routes, much like major airports connect to numerous flight routes

Real Networks are Heterogeneous: Examples (4/4)

- **Scientific Collaboration Network**
  - **Nodes**: Researchers or scientists
  - **Links**: Co-authorships on research papers
  - **Major Hubs**: Highly prolific researchers
  - **Characteristic**: Some researchers are particularly central to the network due to their extensive collaborations on multiple projects, mirroring the role of major hubs in transportation networks. These individuals often influence the direction of research and information flow within their fields

Heterogeneity

- Heterogeneity in networks refers to the variability in the properties and roles of elements within the network, such as nodes and links
  - Air transportation networks (Airports, Flights)
  - Social networks (Individuals, Relationships)
  - The Web (Websites, Hyperlinks)

- This variability reflects the diversity present in complex systems and affects how networks function and interact

The Role of Nodes in Networks

universität koblenz
weiter:denken

- A key source of heterogeneity is the degree of nodes, indicating the number of connections each node has
  - major airports like Atlanta
  - influential social figures like Barack Obama, and
  - dominant websites like Google
- These nodes act as hubs in their respective networks, influencing flow and accessibility significantly

Network Centrality and Its Impact

universität koblenz
weiter:denken

- The importance of a node or a link is estimated by computing its **centrality**
- There are different ways of measuring **centrality** such as **Degree**, **Betweenness**, and **Closeness** centrality
- **Hubs**: High-Degree Nodes
  - These nodes, or hubs, are crucial for understanding network dynamics
  - Facilitate quick dissemination of information or resources across a network
  - Enhance the resilience of networks to certain types of failures but may also pose risks of rapid spread (e.g., diseases in social networks or vulnerabilities in the web)
  - In general: hubs contribute to some striking properties that characterize a broad variety of networks, affecting everything from information flow to network stability

Centrality Measures

- **Centrality:** measure of **importance** of a node
- **Measures**
  1. Degree
  2. Closeness
  3. Betweenness

Degree

- **Degree of a node:** number of neighbors of the node

$$k_i = \text{number of neighbors of node } i$$

- High-degree nodes are called **hubs**
- **Average degree of the network:**

$$\langle k \rangle = \frac{\sum_i k_i}{N} = \frac{2L}{N}$$

```
G.degree(2) # returns the degree of node 2
G.degree()  # dict with the degree of all nodes of G
```

Closeness

- **Idea:** a node is the more central the *closer* it is to the other nodes, on average

$$g_i = \frac{1}{\sum_{j \neq i} \ell_{ij}}$$

where ℓ_{ij} is the distance between nodes *i* and *j*

```
nx.closeness_centrality(G, node) # closeness centrality
                                 # of node
```

Betweenness (1/3)

universität koblenz
weiter:denken

- **Idea**: a node is the more central the *more often it is crossed by paths*

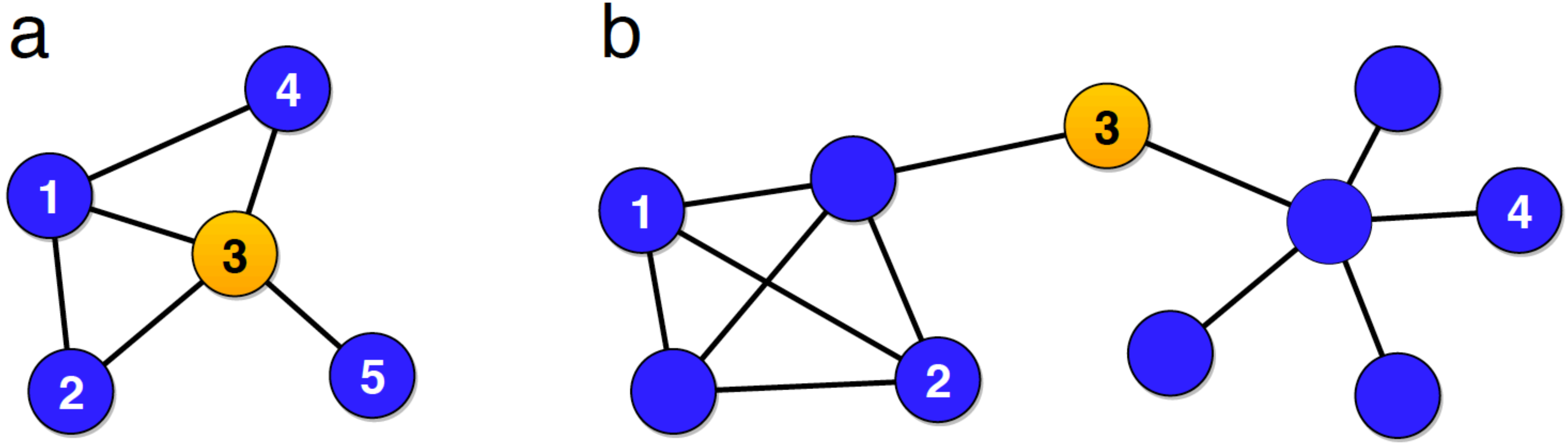
$$b_i = \sum_{h \neq j \neq i} \frac{\sigma_{hj}(i)}{\sigma_{hj}}$$

σ_{hj} = number of shortest paths from h to j

$\sigma_{hj}(i)$ = number of shortest paths from h to j running through i

Betweenness (2/3)

- Hubs usually have high betweenness, but there can be nodes with high betweenness **that are not hubs**



(a) The orange node has high degree (4) as well as high betweenness (3.5). **(b)** The orange node has low degree (2) but keeps the network connected, acting as the only bridge between nodes in the two subnetworks. For example, the shortest path between nodes **1** and **2** does not go through the orange node, but the path between **1** and **4** does. In fact, **all** the shortest paths between the four nodes in one subnetwork and the five nodes in the other subnetwork go through the orange node. Therefore its betweenness is 4x5=20.

Betweenness (3/3)

- Betweenness can be easily extended to links
- **Link betweenness**: fraction of shortest paths among all possible node pairs that pass through the link

```
nx.betweenness_centrality(G)      # dict nodes ->
                                  # betweenness centrality
nx.edge_betweenness_centrality(G) # dict links ->
                                  # betweenness centrality
```

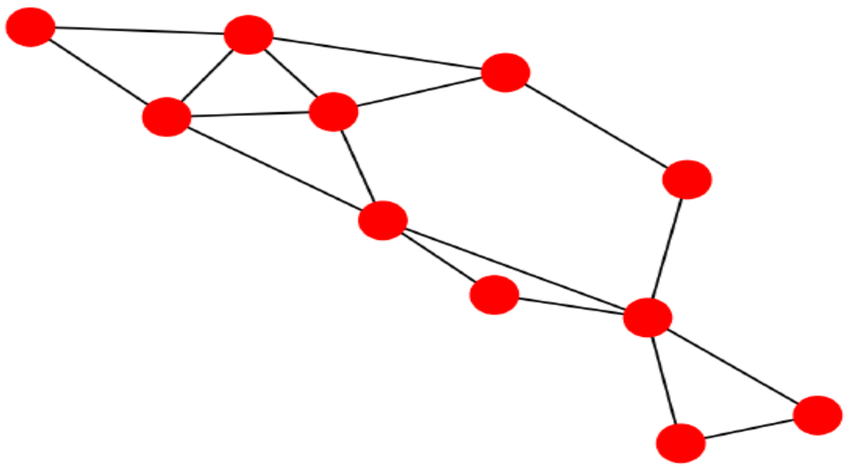
2. Centrality Distributions

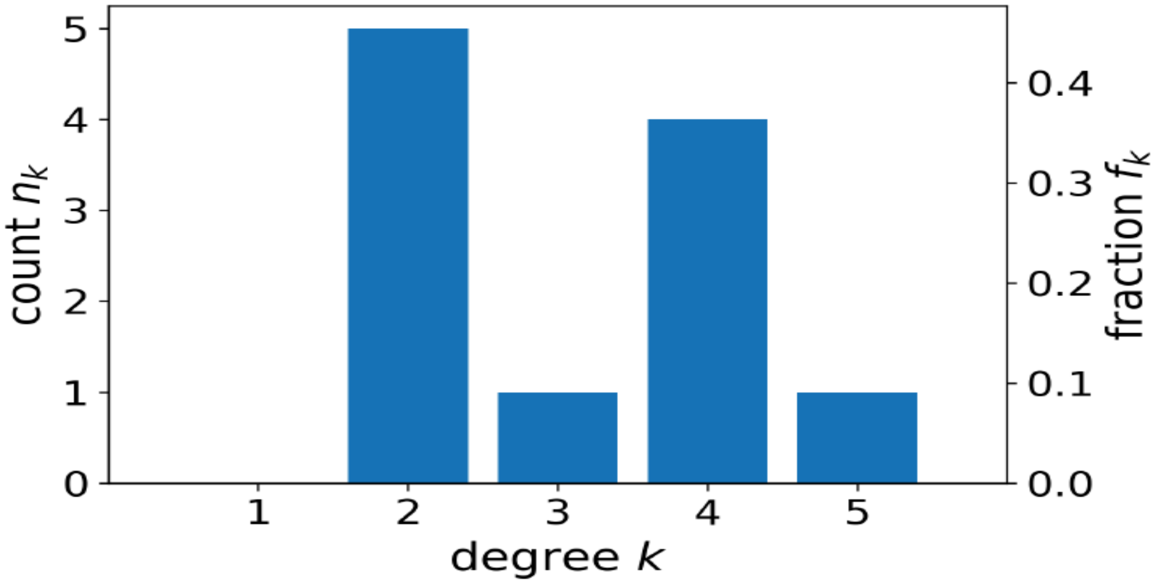
Centrality Distributions (1/3)

- For small networks, identifying the most important nodes or links is practical
- For large networks, this individual approach is impractical
- Solution: Adopt a statistical perspective
  - Rather than examining individual nodes and links, analyze groups (or classes) of nodes and links sharing similar properties

Centrality Distributions (2/3)

Histogram

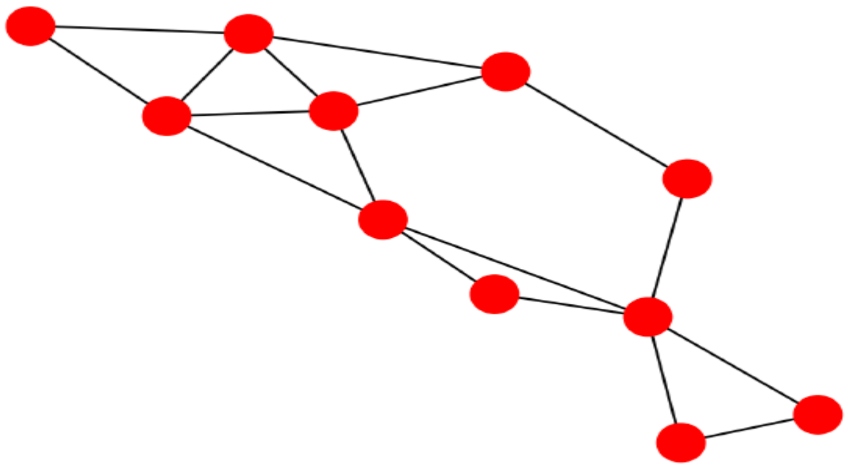




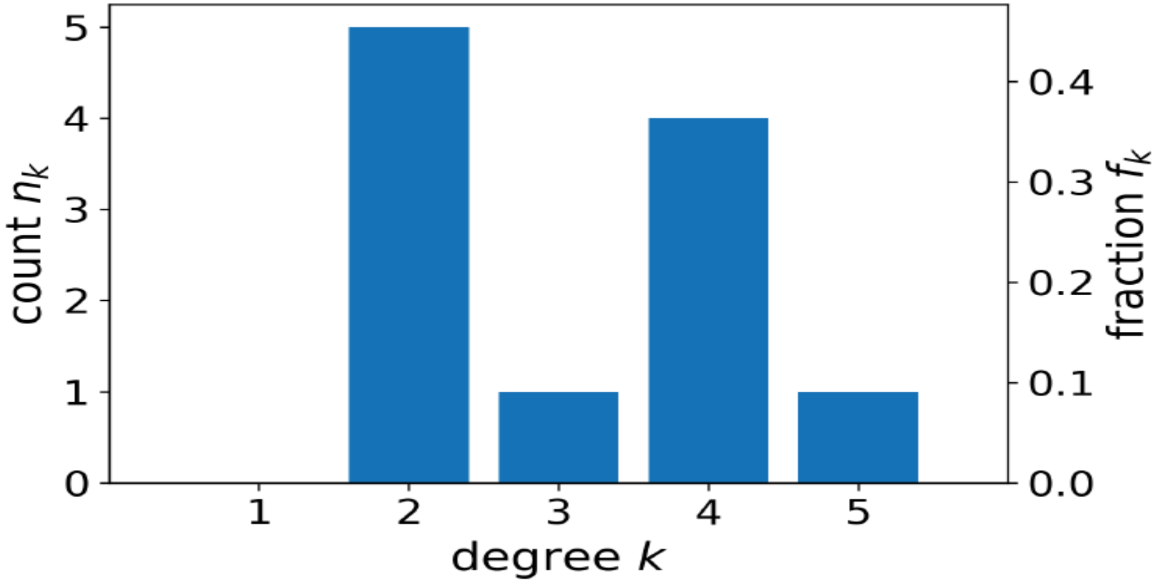
- n_k = **number of nodes with degree** k
- $f_k = \frac{n_k}{N}$ = **frequency of degree** k

Histogram representation of the degree distribution of a small network. First a list with the degree of each node is generated. The heights of the histogram bars are given by the counts of nodes with each degree k. The relative frequency of occurrence is defined as the fraction of all the nodes with degree k.

Centrality Distributions (3/3)



Histogram



- When $N \rightarrow \infty$, f_k becomes the **probability** p_k of having degree k
- p_k versus k is the **probability distribution** of node degree

Cumulative Distributions

universität koblenz
weiter:denken

- For non-integer variables (e.g., betweenness), we divide the range into intervals (bins) and count the values within each bin
- **Cumulative distribution *P(x)*:** The probability that the variable has values *larger* than a given value x
- **How to calculate it:** Sum frequencies of all intervals from x onward:

$$P(x) = \sum_{v \geq x} f_v$$

Logarithmic Scale

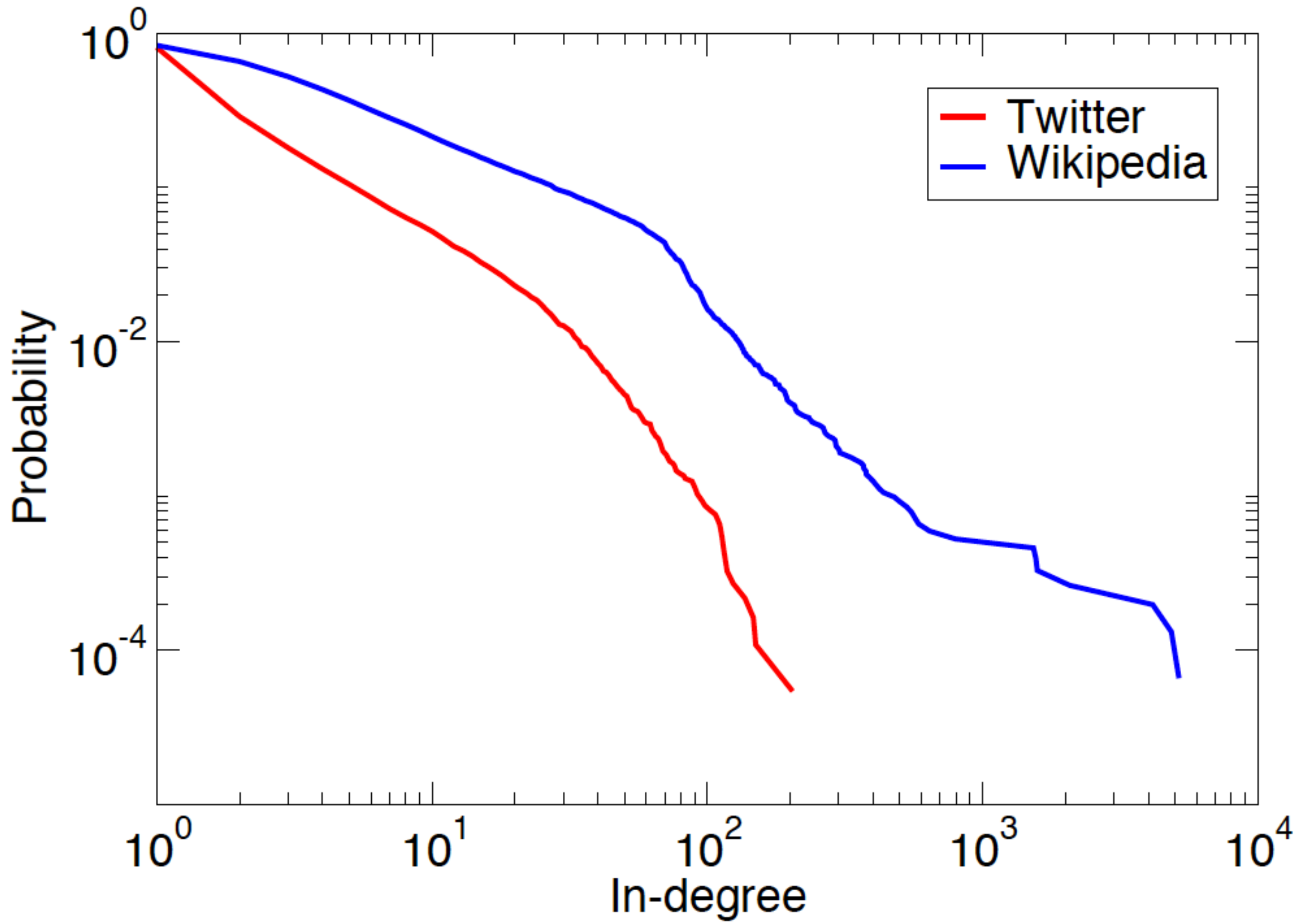
- Issue: How can we visualize probability distributions when the values span a very wide range?
- Solution: Use a logarithmic scale
- Method: Represent the logarithms of values on both the x-axis and the y-axis
- Example

$$\log_{10} 10 = 1$$

$$\log_{10} 1{,}000 = \log_{10} 10^3 = 3$$

$$\log_{10} 1{,}000{,}000 = \log_{10} 10^6 = 6$$

Degree Distributions (1/3)



Heavy-tail distributions: the variable goes from small to large values

Degree Distributions (2/3)

To formally define the heterogeneity parameter κ (Greek letter "kappa") of a network's degree distribution, we need to introduce the *average squared degree* $\langle k^2 \rangle$, which is the average of the squares of the degrees:

$$\langle k^2 \rangle = \frac{k_1^2 + k_2^2 + \cdots + k_{N-1}^2 + k_N^2}{N} = \frac{\sum_i k_i^2}{N}. \qquad (3.4)$$

The heterogeneity parameter can be defined as the ratio between the average squared degree and the square of the average degree of the network [Eq. (1.5)]:

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle^2}. \qquad (3.5)$$

For a normal or narrow distribution with a sharp peak at some value, say k_0, the distribution of the squared degrees is concentrated around k_0^2. Therefore $\langle k^2 \rangle \approx k_0^2$ and $\langle k \rangle \approx k_0$, yielding $\kappa \approx 1$. For a heavy-tailed distribution with the same average degree k_0, $\langle k^2 \rangle$ blows up because of the large degree of the hubs, so that $\kappa \gg 1$.

Degree Distributions (3/3)

universität koblenz
weiter:denken

- The **heterogeneity parameter $\boldsymbol{\kappa}$** says how broad the distribution is:

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

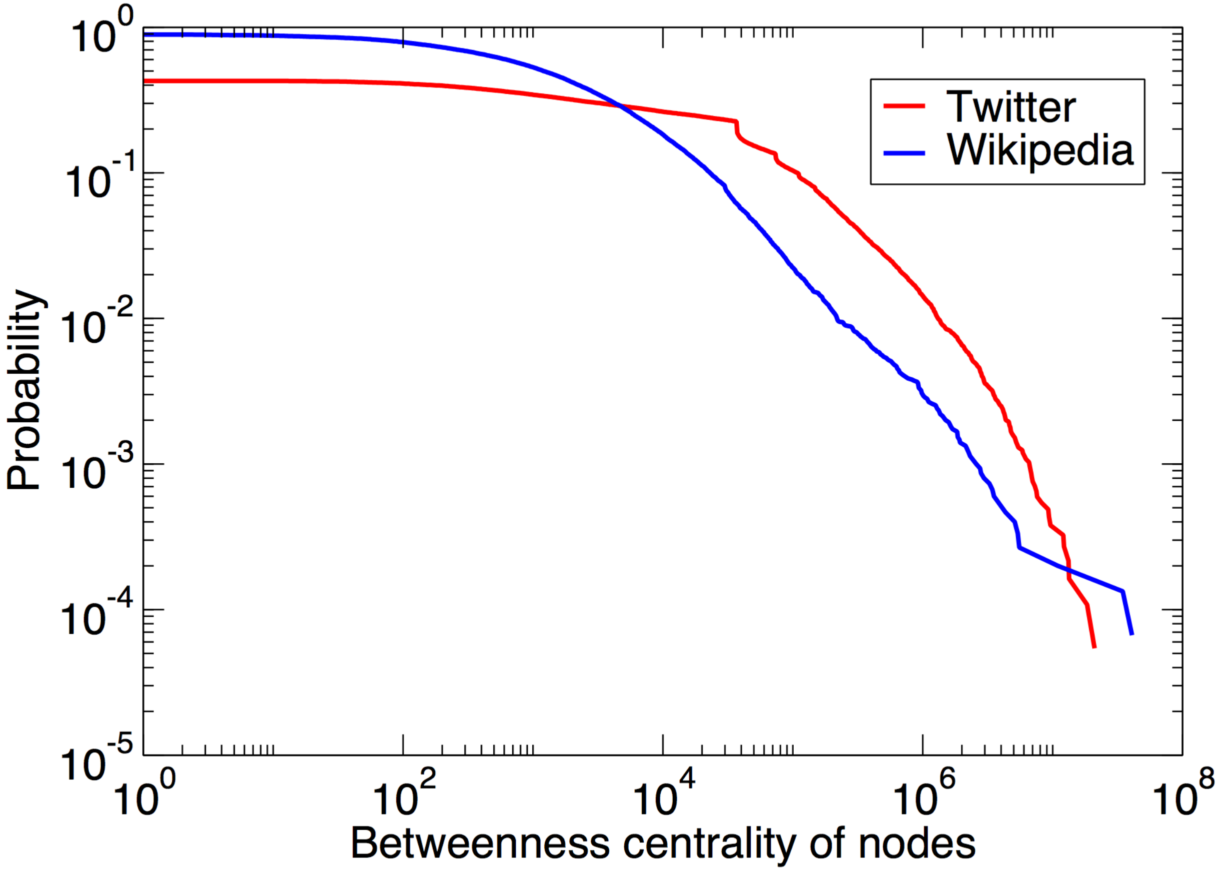
$$\langle k \rangle = \frac{\sum_i k_i}{N} = \frac{2L}{N}; \langle k^2 \rangle = \frac{\sum_i k_i^2}{N}$$

- If most degrees have the same value, say k_0 :

$$\langle k \rangle \approx k_0, \langle k^2 \rangle \approx k_0^2 \Longrightarrow \kappa \approx 1$$

- If the distribution is very heterogeneous: $\kappa \gg 1$

Betweenness Distributions



- Cumulative distribution of node betweenness centrality for Twitter and Wikipedia, shown on a log–log plot
- We considered both networks as undirected
- For Wikipedia we computed the betweenness only on its giant component, which includes over 98% of the nodes

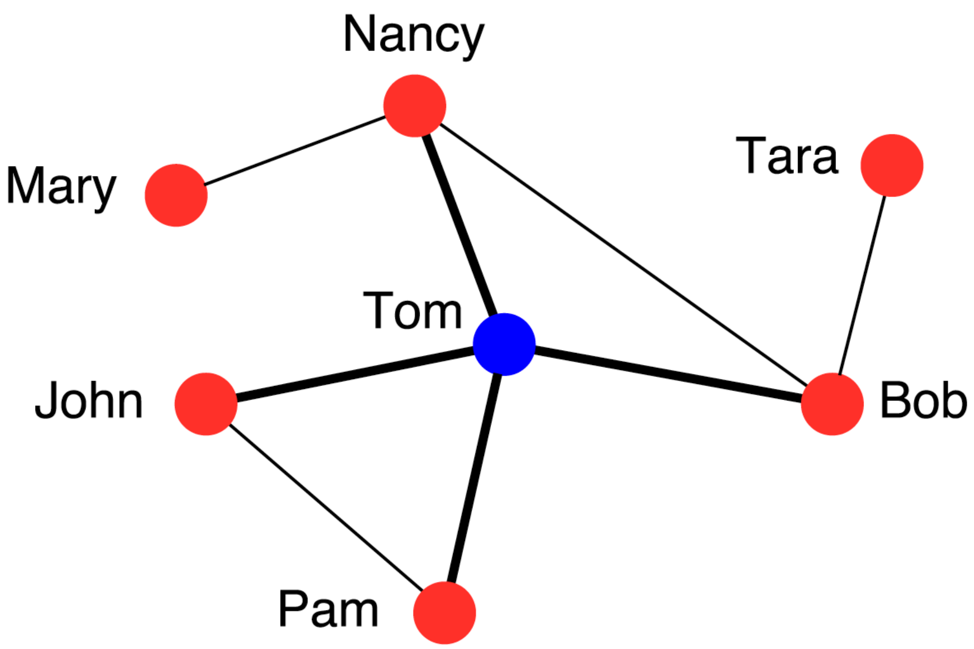
Heavy-tail distribution: the variable goes from small to large values

Degree Centrality

Network	Nodes (N)	Links (L)	Average degree ($\langle k \rangle$)	Maximum degree (k_{max})	Heterogeneity parameter (κ)
Facebook Northwestern Univ.	10,567	488,337	92.4	2,105	1.8
IMDB movies and stars	563,443	921,160	3.3	800	5.4
IMDB co-stars	252,999	1,015,187	8.0	456	4.6
Twitter US politics	18,470	48,365	2.6	204	8.3
Enron Email	36,692	367,662	10.0	1,383	14.0
Wikipedia math	15,220	194,103	12.8	5,171	38.2
Internet routers	190,914	607,610	6.4	1,071	6.0
US air transportation	546	2,781	10.2	153	5.3
World air transportation	3,179	18,617	11.7	246	5.5
Yeast protein interactions	1,870	2,277	2.4	56	2.7
C. elegans brain	297	2,345	7.9	134	2.7
Everglades ecological food web	69	916	13.3	63	2.2

3. The Friendship Paradox

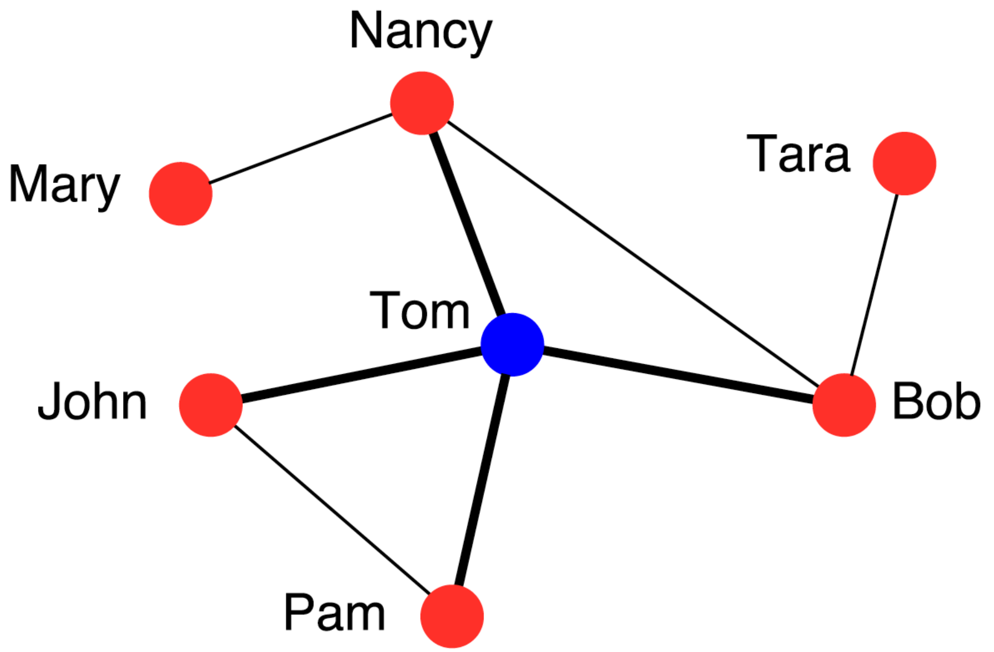
Friendship Paradox (1/4)



- If we pick a person (node) at random, everyone — including Tom — has equal probability of being selected
- If we pick a friendship (link) at random and then look at one of the people it connects, Tom is more likely to be chosen because he has more connections
- Nodes with more links are overrepresented when sampling through connections — this is the core of the **friendship paradox**

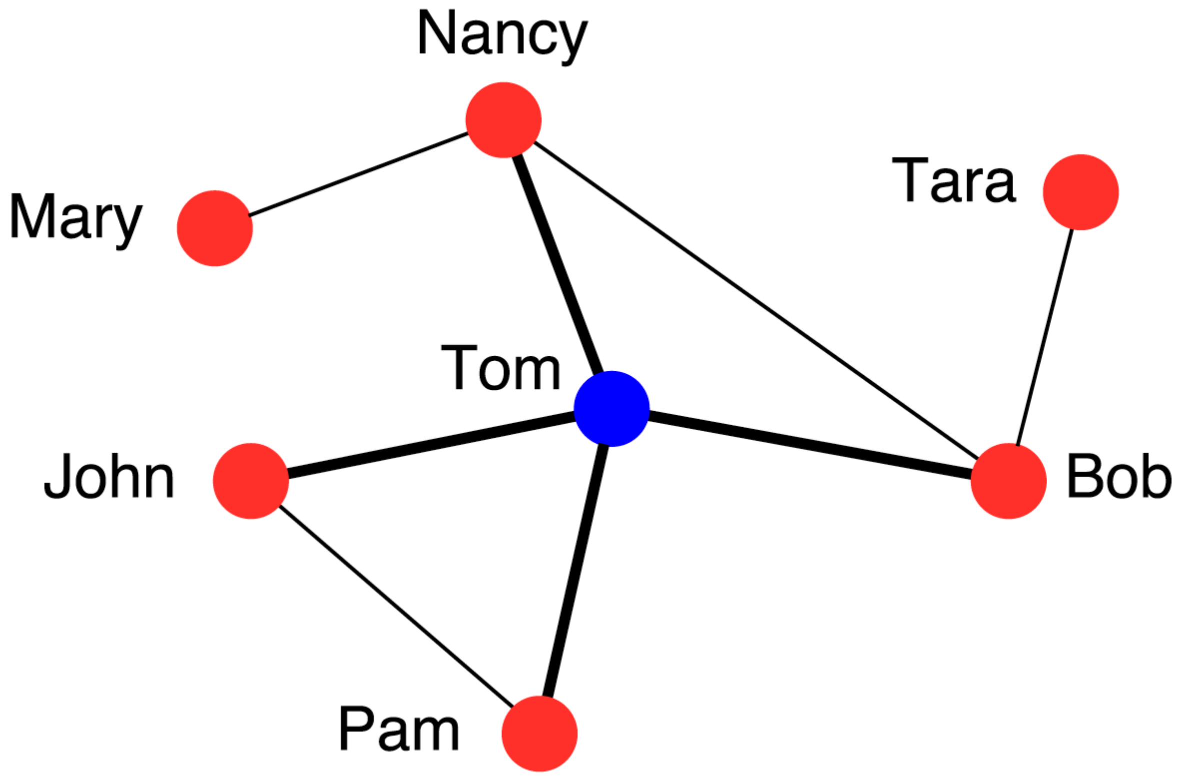
Friendship Paradox (2/4)

universität koblenz
weiter:denken



- When you follow a random link in the network, you are more likely to reach a highly connected node (a hub)
- This is because hubs are connected to many links, so they are more likely to be reached by random traversal
- In other words, the more friends someone has, the more likely you are to encounter him/her by following a random friendship

Friendship Paradox (3/4)



- Average degree of a node = 2.29
- Average degree of the neighbors of a node = 2.83 > 2.29
- Our friends have more friends than we do, on average: **friendship paradox**

Friendship Paradox (4/4)

- **Question:** Why does the friendship paradox happen?
- **Answer**
  1. When we average over all people (nodes), we treat each person equally
  2. When we average over friends (neighbors), we sample through links
     - A person with k friends appears k times — this **overrepresents highly connected nodes** and **raises the average**
- The more **hubs** there are in the network, the **stronger the paradox**

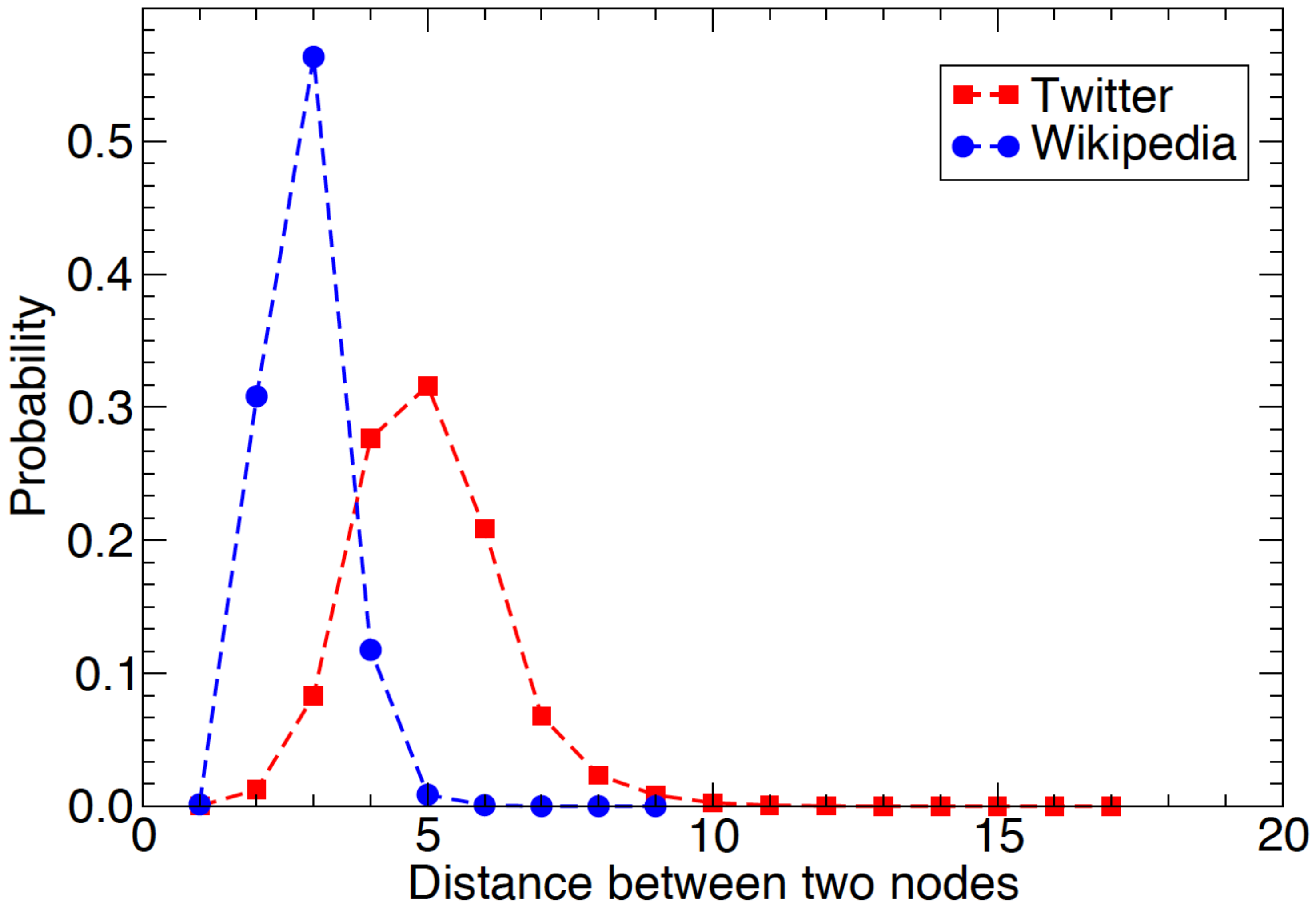
4. Ultra-Small Words

Ultra-Small Worlds (1/2)

- In real networks, many shortest paths go through hubs
- **Example**: air travel
  - Small airports A and B might not be directly connected,
  - But you can still travel from A to B through a hub airport like C
- In networks with hubs, shortest paths become extremely short — this is known as the **ultra-small world** effect
  - Hubs dramatically reduce the average distance between nodes

Ultra-Small Worlds (2/2)

Shortest-path length distribution



- Ultra-small worlds
- The distributions of distance between nodes are peaked at very low values for both Twitter and Wikipedia
- This is due to the presence of hubs, which shrink the distance between most pairs of nodes, as shortest paths run through them
- Distances are computed by ignoring the direction of the links

universität koblenz
weiter:denken

5. Robustness

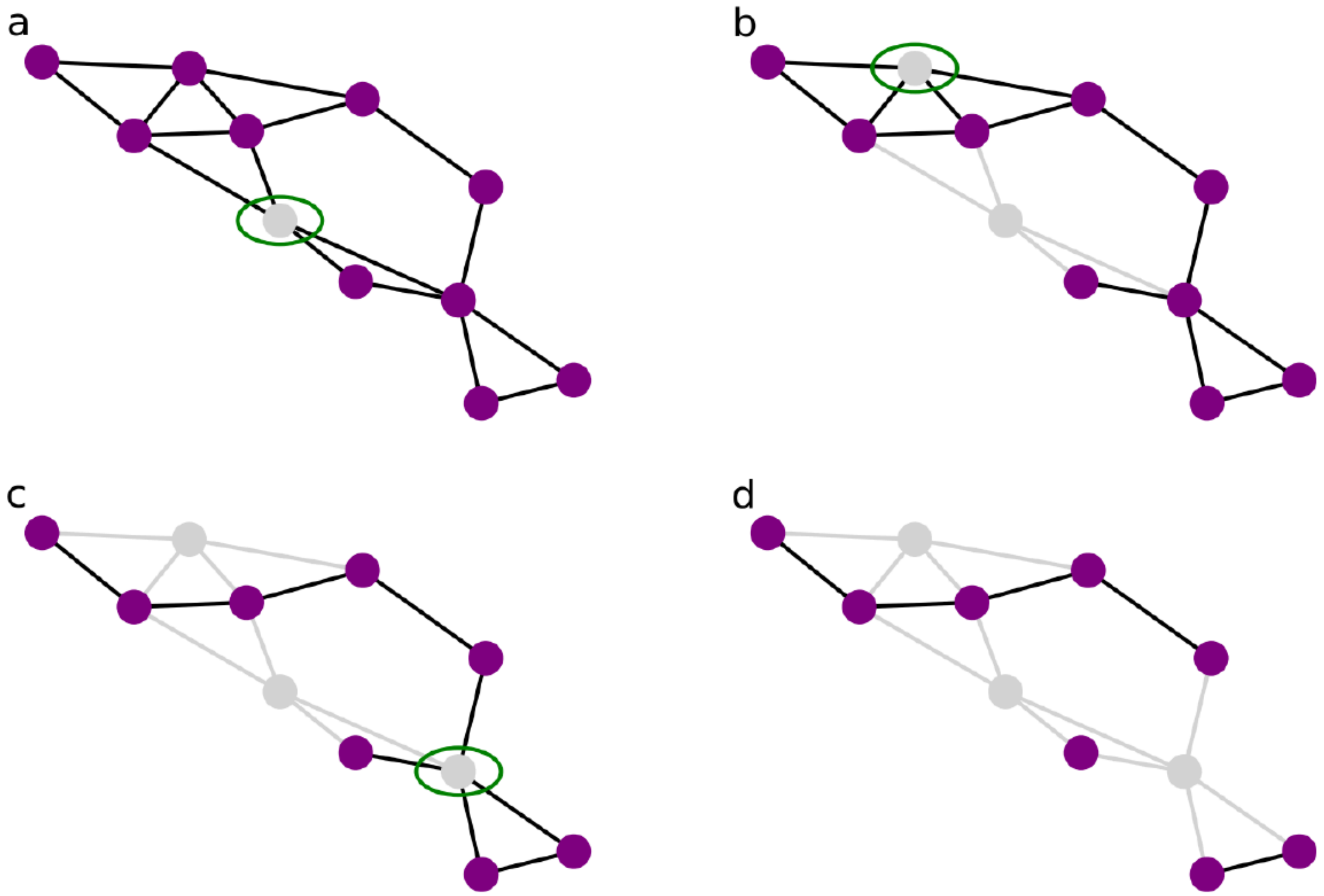
Robustness (1/5)

- A system is **robust** if it continues to function even when some components fail
- **Question**: How do we measure the *robustness* of a network?
- **Answer**: We simulate failures by removing nodes or links and observe the structural impact
- **Key point:** *connectedness*
- If the Internet were not connected, it would be impossible to transmit signals (e.g., emails) between routers in different components

Robustness (2/5)

- **Robustness test:** Measure how network connectivity changes as we remove more nodes
- **Method:** Track the size S of the largest connected component as a function of the fraction of nodes removed
- We suppose that the network is initially connected: there is only one component and $S = 1$
- As more and more nodes (and their links) are removed, the network is progressively broken up into components and S goes down
- This helps us understand how resilient the network is to failures or attacks

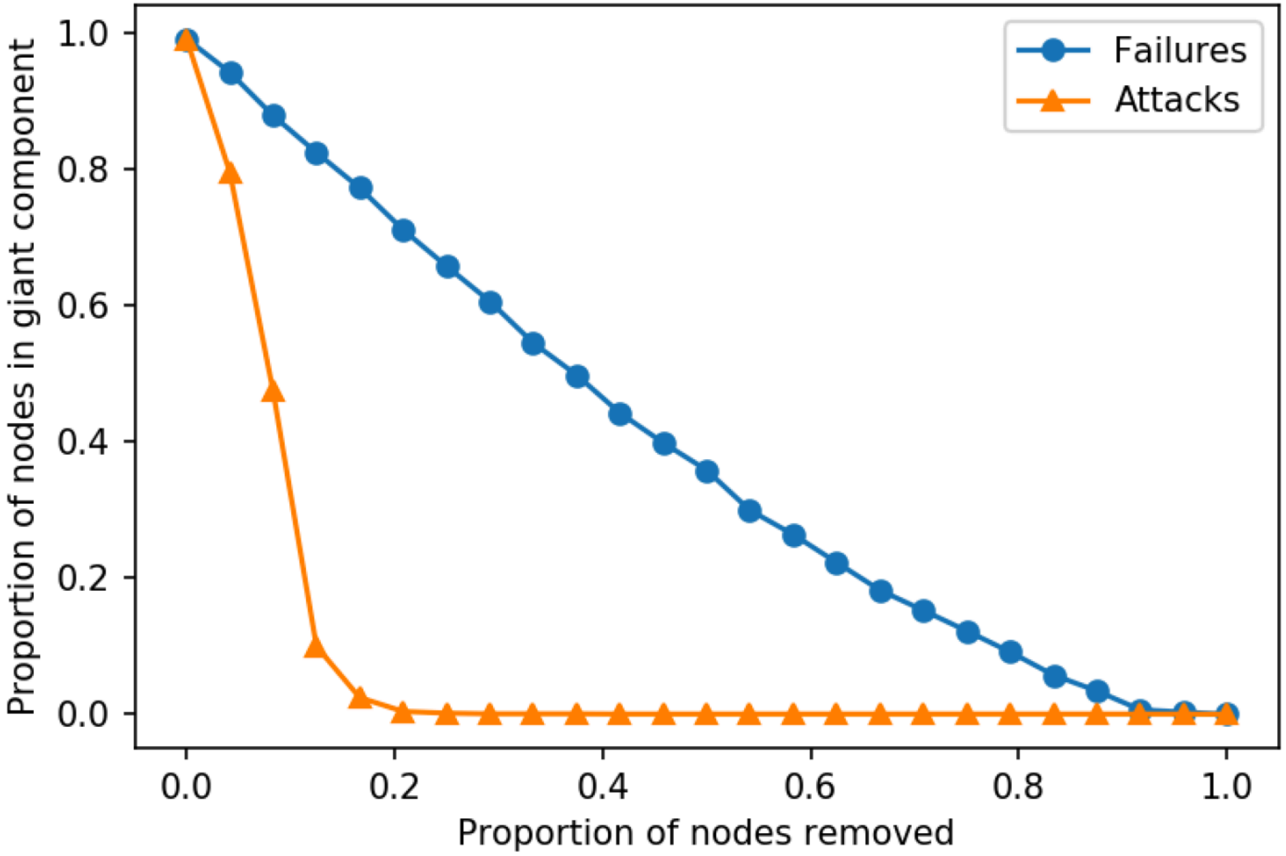
Robustness (3/5)



Robustness (4/5)

- **Two types of node removal strategies:**
  1. **Random failures:** Nodes are removed randomly — each has the same chance of failure
  2. **Attacks:** hubs are deliberately targeted — the more connections a node has, the more likely it is to be targeted
- In the random case, we remove a fraction f of nodes, chosen at random
- In the attack case, we remove the top f-fraction of highest-degree nodes, starting from the most connected

Robustness (5/5)



Conclusion: Real networks are robust to random failures, but very vulnerable to targeted attacks on high-degree nodes

6. Core Decomposition

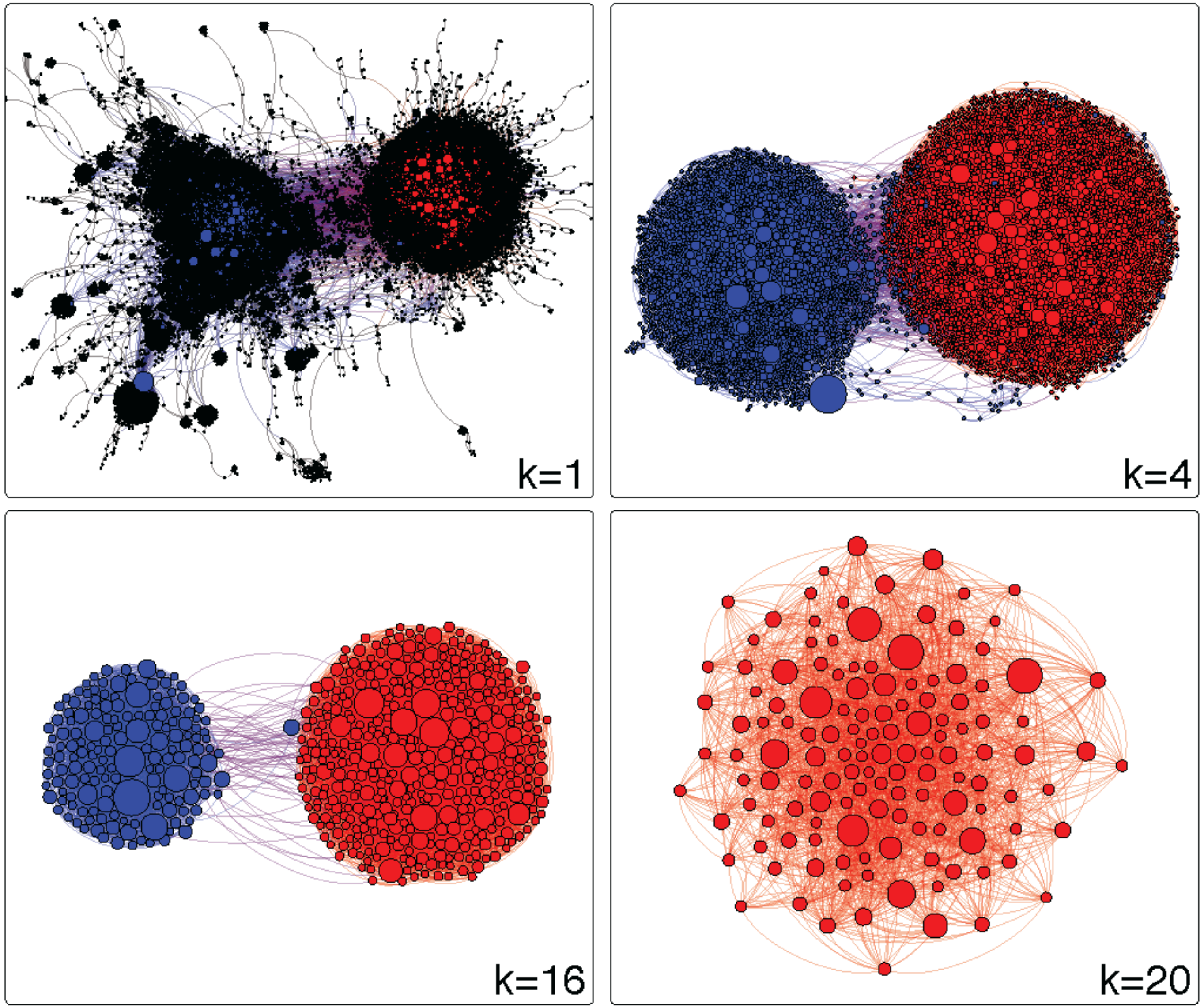
Core Decomposition (1/3)

- **Core:** A dense region in the network made of high-degree nodes
- **Core decomposition:** A method to find increasingly dense subgraphs by removing low-degree nodes step by step
- A **k-core** is what remains after removing all nodes with degree ≤ k−1
- ***k*-core decomposition procedure:**

1. Start with k=0
2. Remove all nodes with degree k, repeatedly, until none remain
3. The set of removed nodes is the ***k*-th shell**, while the remaining ones form the **(*k*+1)-core**
4. If nodes remain, increase k and repeat

Core Decomposition (2/3)

- We apply k-core decomposition to a Twitter political retweet network
- Starting at k=1, the full network is included
- As k increases, nodes with fewer than k neighbors are removed
- The result:
  - The network becomes smaller and denser
  - Only highly connected nodes remain
- At k=20, the core consists only of red nodes (conservative accounts), each connected to at least 20 neighbors



Core Decomposition (3/3)

- Core decomposition makes it easier to explore large networks by removing low-degree nodes and keeping only the densest regions

```
nx.core_number(G)   # return dict with core number of each node
nx.k_shell(G,k)     # subnetwork induced by nodes in k-shell
nx.k_core(G,k)      # subnetwork induced by nodes in k-core
nx.k_core(G)        # innermost (max-degree) core subnetwork
```

7. Summary

Summary (1/2)

- **Node Degree:** Counts the number of links connected to a node. It's a basic measure of how connected a node is within a network
- **Node Betweenness:** Measures how often a node appears on shortest paths. It identifies nodes acting as bridges or bottlenecks for information flow
- **Statistical Tools:** Tools like histograms help analyze large networks by showing the distribution of attributes (e.g., degree)
- **Heterogeneity in Centrality:** Real networks often show heavy-tailed distributions, where most nodes have few connections, but a few hubs have many

Summary (2/2)

- **Friendship Paradox:** On average, a person's friends are more connected than they are—caused by the higher chance of sampling hubs through neighbors
- **Role of Hubs:** Hubs reduce path lengths and improve flow, but also make networks more vulnerable to targeted failures
- **Core-Periphery Structure:** Networks can be decomposed into a dense core and a sparse periphery using core decomposition, which helps analyze resilience and connectivity

References

[1] Menczer, F., Fortunato, S., & Davis, C. A. (2020). A First Course in Network Science Cambridge: Cambridge University Press.

- Chapter 3 Hubs

[2] OLAT course page: https://olat.vcrp.de/url/RepositoryEntry/4669112833

Further Readings

- Marianov, V., Serra, D. and ReVelle, C., 1999. Location of hubs in a competitive environment. European journal of operational research, 114(2), pp.363-371.