

Introduction to social networks and the friendship paradox

Max Pellert

University of Konstanz

Social Media Data Analysis

So Far

- Search behavior and future orientation
- Social trends: the Simmel effect
- Social Impact Theory
- Text and Sentiment Analysis
 - Theoretical basis: measuring emotions and the semantic differential
 - Unsupervised methods based on dictionaries
 - Supervised methods
 - Applications and evaluation

What comes next

Today: social network basics and the friendship paradox

Centrality measures and social resilience

Assortativity and communities, permutation tests

Social issues: polarization, misinformation, representation

Basics of social networks

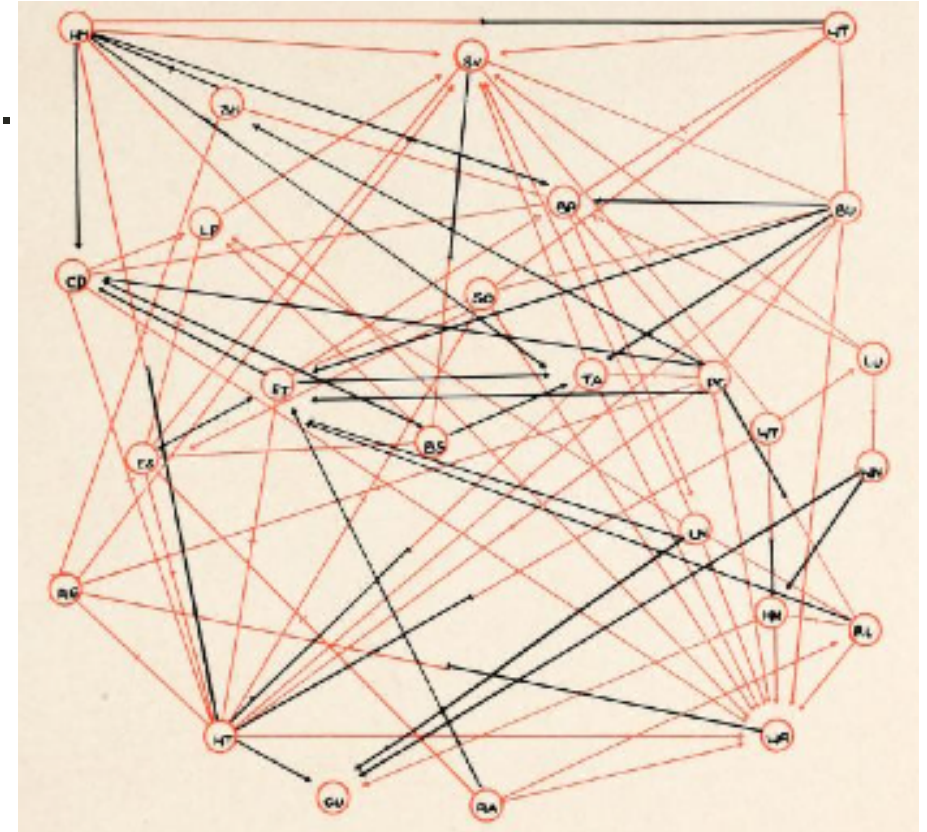
1. Basics of social networks

2. The friendship paradox

Social networks

Social Networks are models to represent individuals and the relationships between them. The minimal components of a social network are:

- **Nodes** (or vertices), which represent individuals. These individuals are social actors, for example humans, animals, fictional characters...
- **Links** (or edges) are relationships between individuals, for example friendship, family ties, interaction, communication...



Jacob Moreno's sociogram

Representing social networks

A graph or network is a tuple $G = (V, E)$

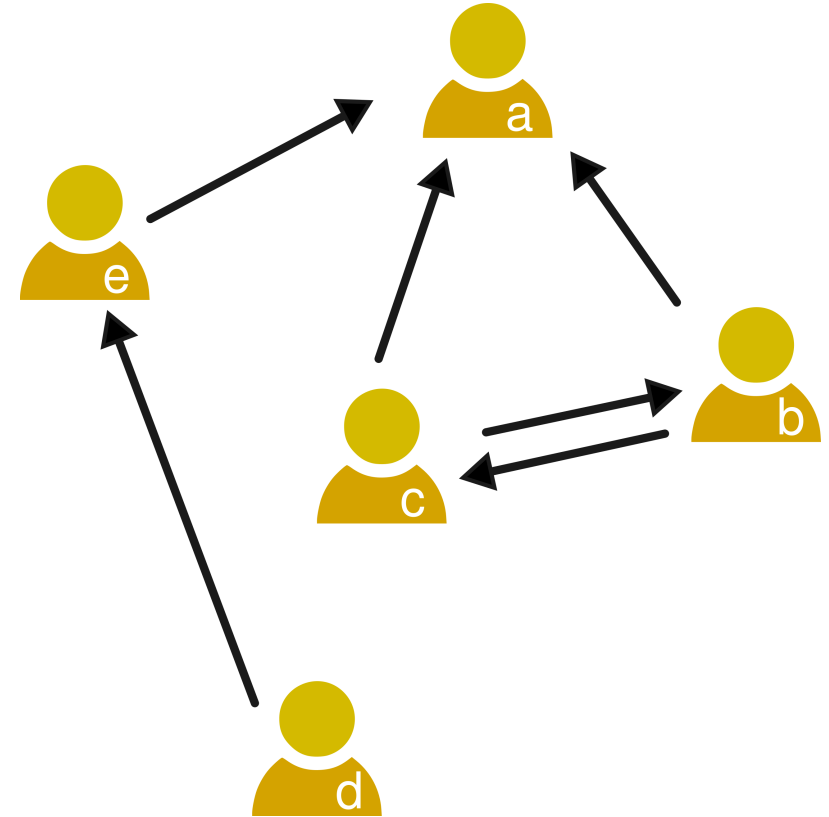
- V is a set of vertices or nodes
- $E \subseteq V \times V$ is a set of edges or links
- $V \times V$ is the Cartesian product (i.e. the set of all possible links)

Edges are denoted as ordered pairs (i, j) , which means that a link points from node i to node j .

The example of the picture can be written as:

$$V = a, b, c, d, e$$

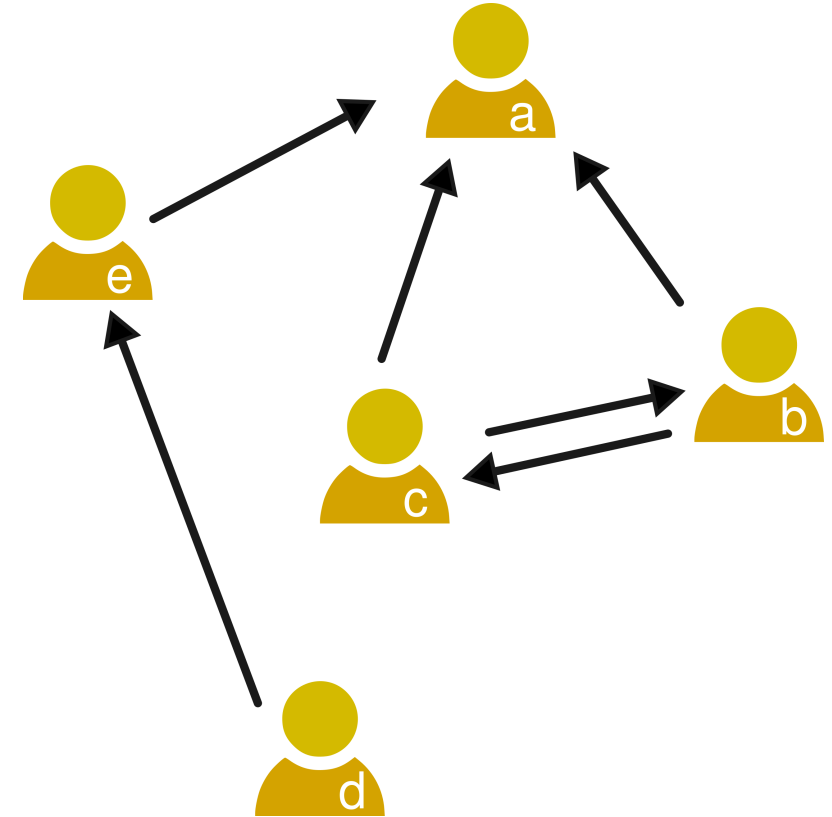
$$E = (b, a), (c, a), (e, a), (d, e), (c, b), (b, c)$$



Adjacency matrix

The above list is what is called an *edge list*, but there are other ways to represent a network. A common one is to use an **adjacency matrix** A with 1 in the entries of nodes connected by an edge and 0 otherwise.

	a	b	c	d	e
a	0	0	0	0	0
b	1	0	1	0	0
c	1	1	0	0	0
d	0	0	0	0	1
e	1	0	0	0	0

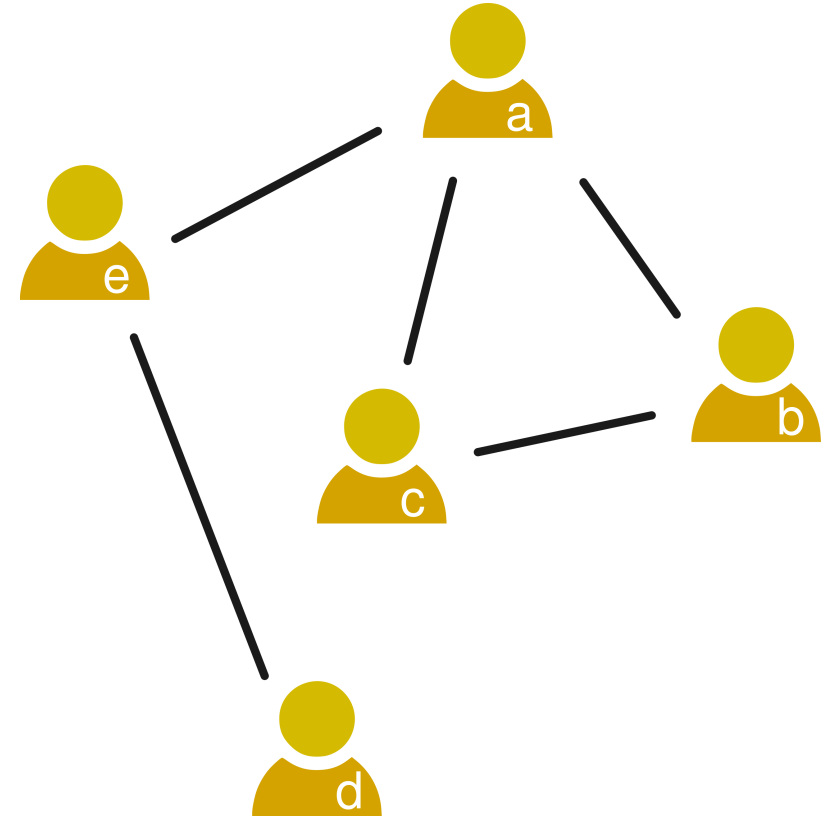


Undirected networks

Networks might be **undirected** if a link between two nodes is always reciprocal.

Their adjacency matrices satisfy the condition $A_{ij} = A_{ji}$.

An example of an undirected network is the network of friendships on Facebook.

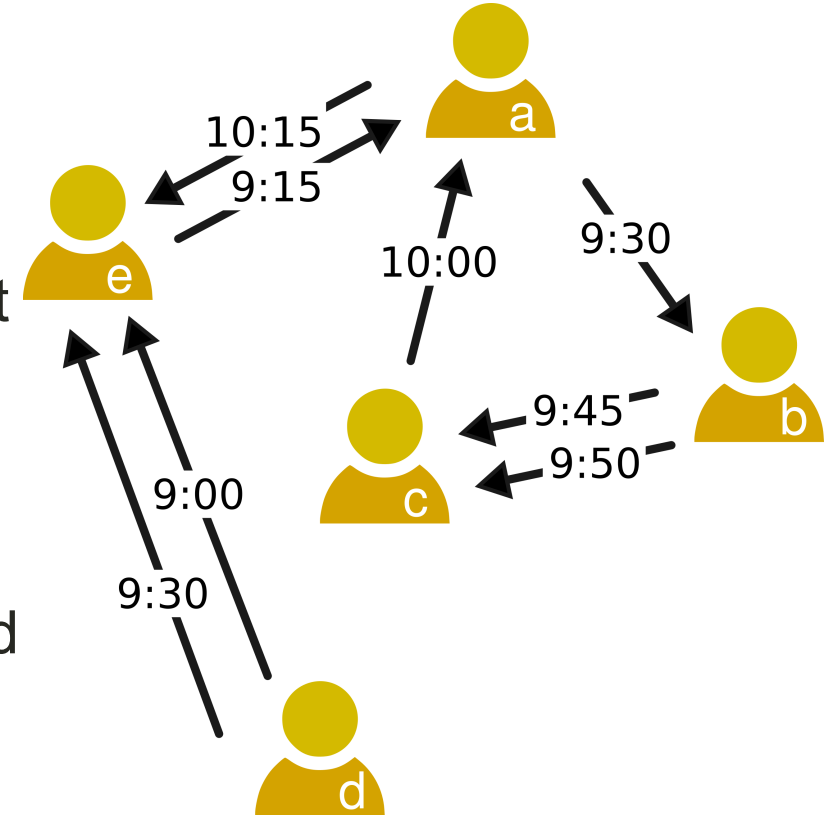


Multit-edge networks

In a **multi-edge network**, multiple links are allowed from one node to another. Multi-edge networks can be both directed and undirected.

The adjacency matrix of multi-edge networks is not well-defined, as each entry of the matrix would have to contain more than just a number.

A common example is a network of phone calls between people, where edges can be differentiated by the timestamp when the call was initiated.

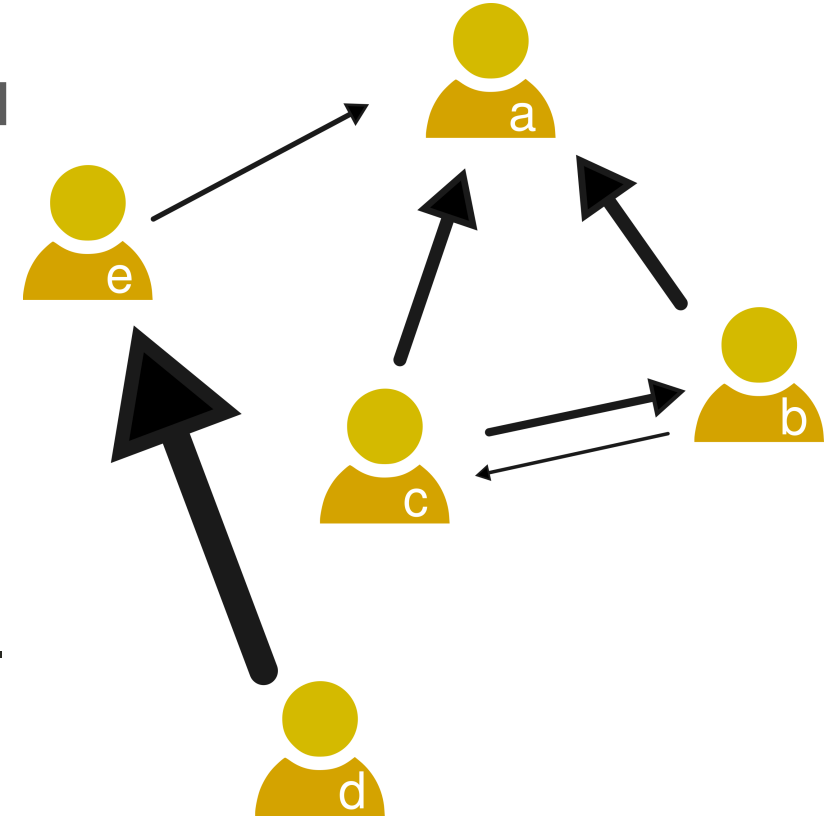


Weighted networks

The nodes of a network can have weights associated to them, then we talk about a **weighted network**.

The weighting function w maps links to weights:
 $w : E \rightarrow \mathbb{R}$.

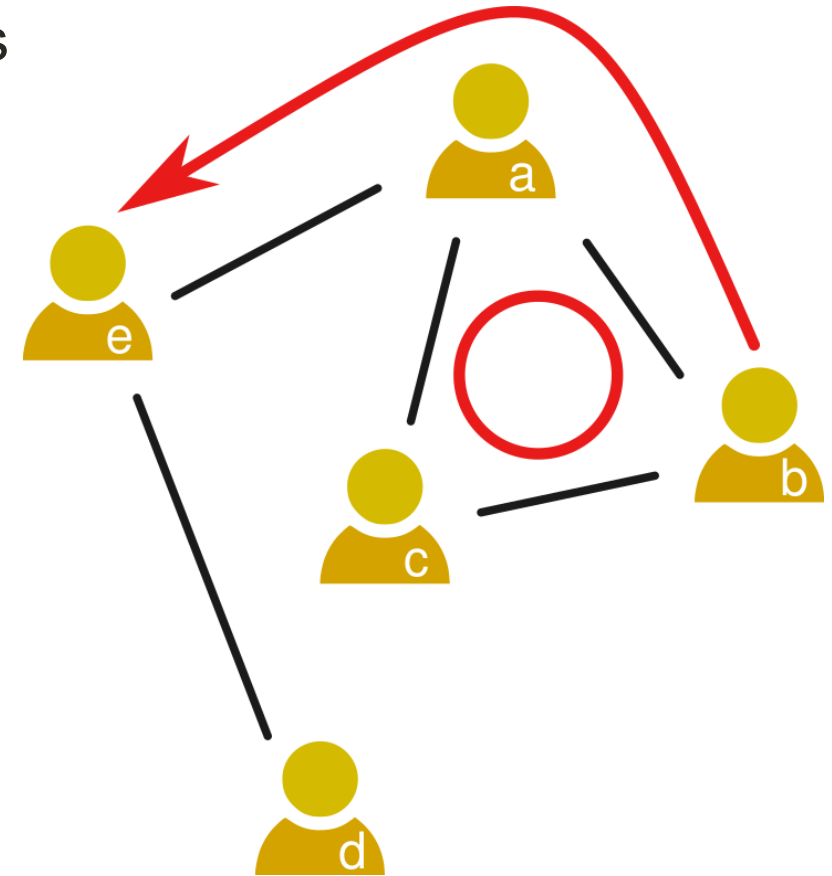
Weighted networks can be represented by adjacency matrices with real values in their entries that correspond to the link weights: $A_{ij} = w(i, j)$.



Paths and cycles

One of the most important concepts in a network is that of a **path**. A path is defined as a sequence of nodes, where any pair of consecutive nodes is connected by a link.

A **cycle** is a path starting and ending in the same node. A **triangle** is a cycle of length 3, in the example the nodes *a*, *b* and *c* form a triangle.

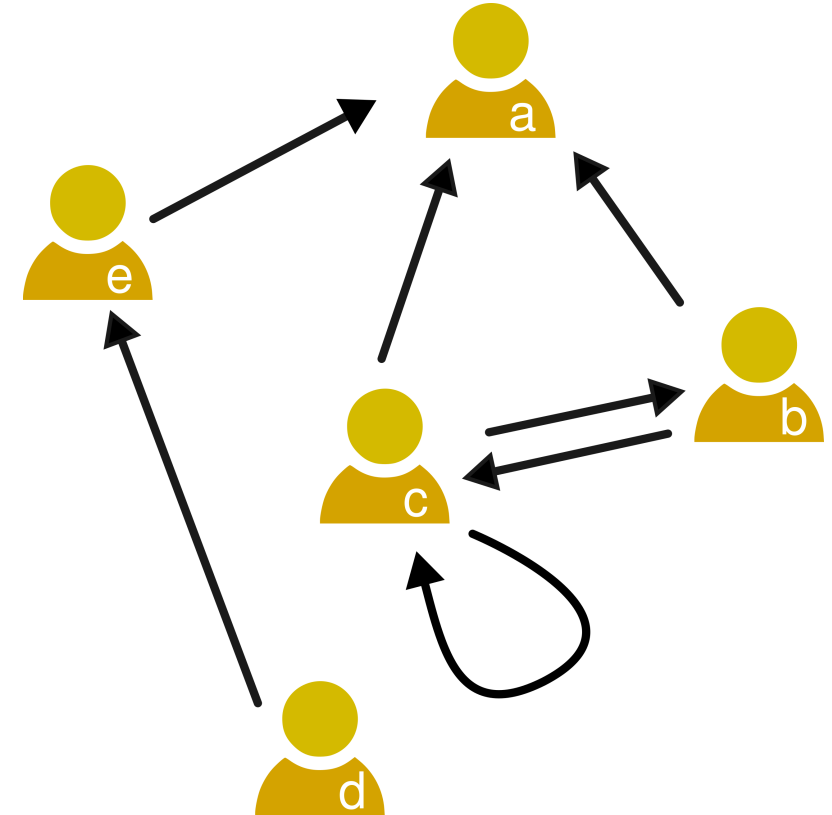


Self-loops

Some networks might contain **self-loops**, which are links that start and end in the same node, i.e. they have the form (i, i) .

Self-loops appear as ones in the diagonal of the adjacency matrix.

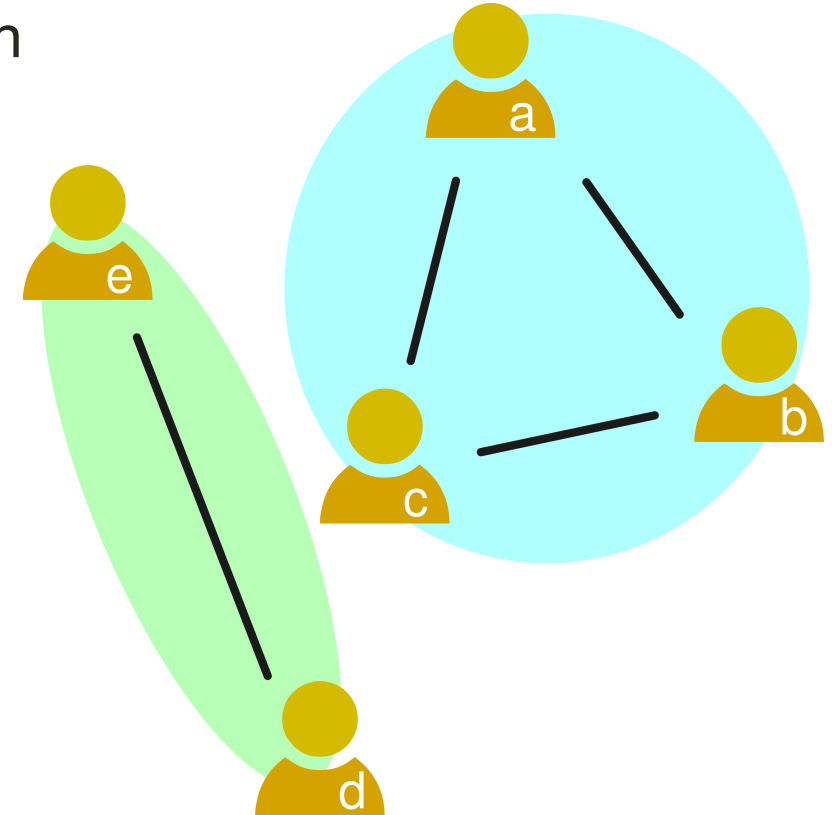
Self-loops are rare in social networks. They can appear in some communication networks, for example when a Twitter user retweets their own tweet or sends an email to themselves.



Connected components

A network is **connected** if, for each pair of nodes in the network, there is at least one path connecting them.

If a network is not connected, it can be divided in **connected components**, which are maximally connected subgraphs. The example is not a connected network, as it has two connected components.



Distance

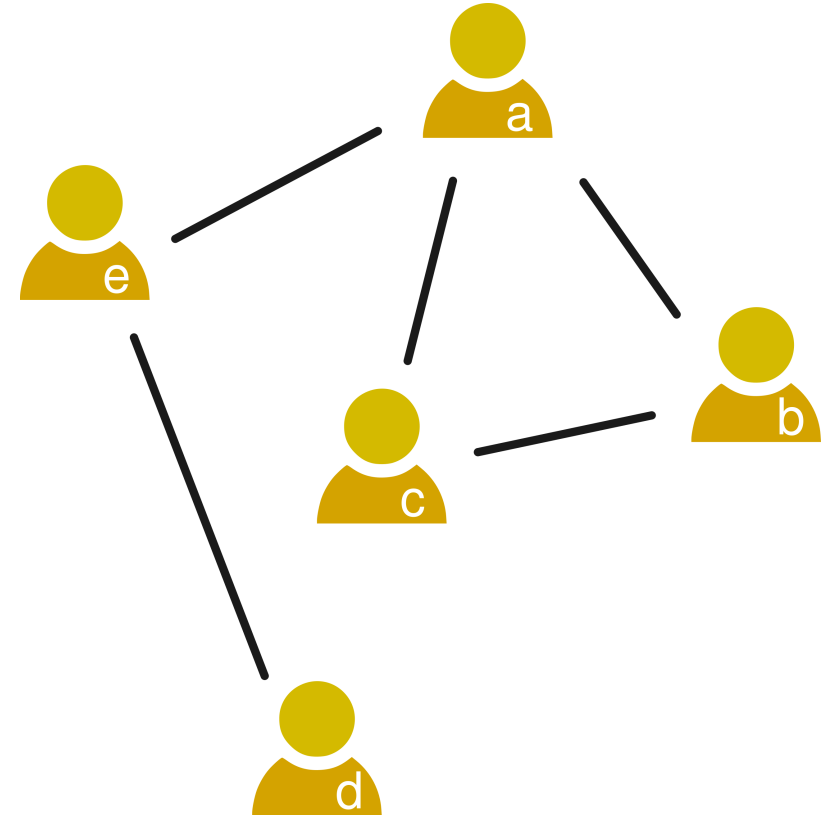
The **distance** between nodes v and w is denoted as $dist(v, w)$ and measures the minimum length among all the paths connecting v and w .

If there is no path between v and w , the distance between them is defined as $dist(v, w) := \infty$.

Example: $dist(b, e) = 2$

In directed networks, it might happen that $dist(v, w) \neq dist(w, v)$.

proximity: $prox(v, w) = \frac{1}{dist(v, w)}$.



Node degree

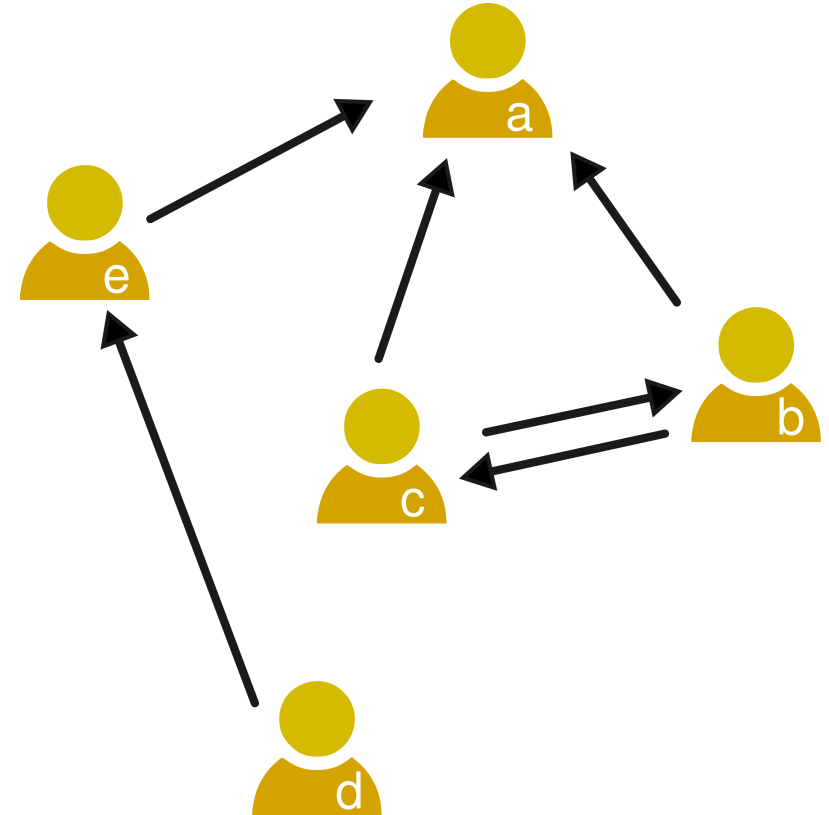
A node's **degree centrality** measures the number of links connected to it.

In directed networks:

- **in-degree** $d_{in}(i)$ that is the number of edges ending in i , i.e. (j, i)
- **out-degree** $d_{out}(i)$ that is the number of edges leaving from i , i.e. (i, j)

$$d_{in}(c) = 1 \text{ and } d_{out}(c) = 2$$

If importance on Twitter is the number of followers of an account, in-degree centrality is a way to measure it.



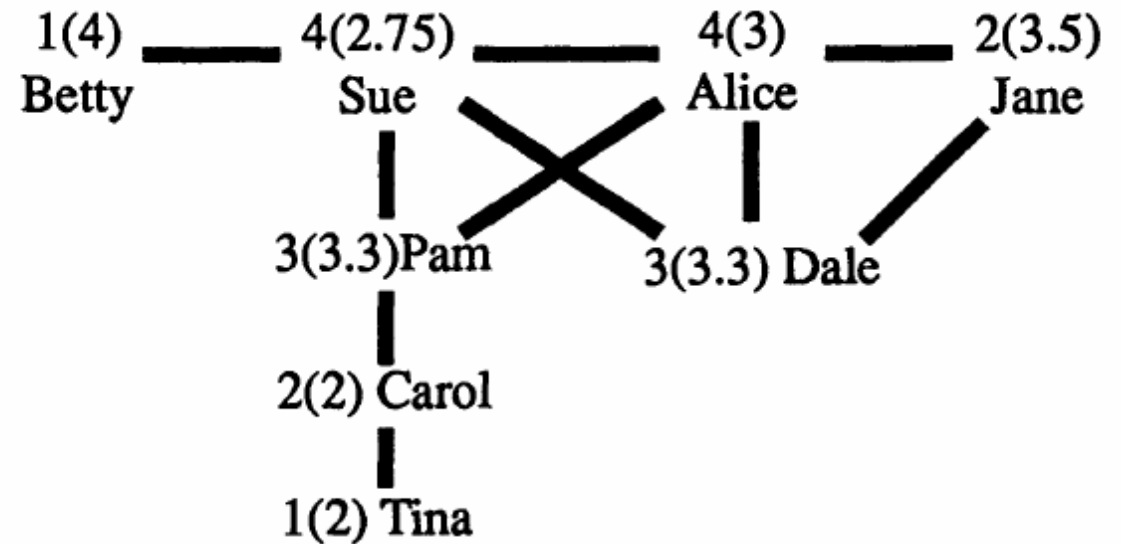
The friendship paradox

1. Basics of social networks

2. The friendship paradox

Friendship and social comparison

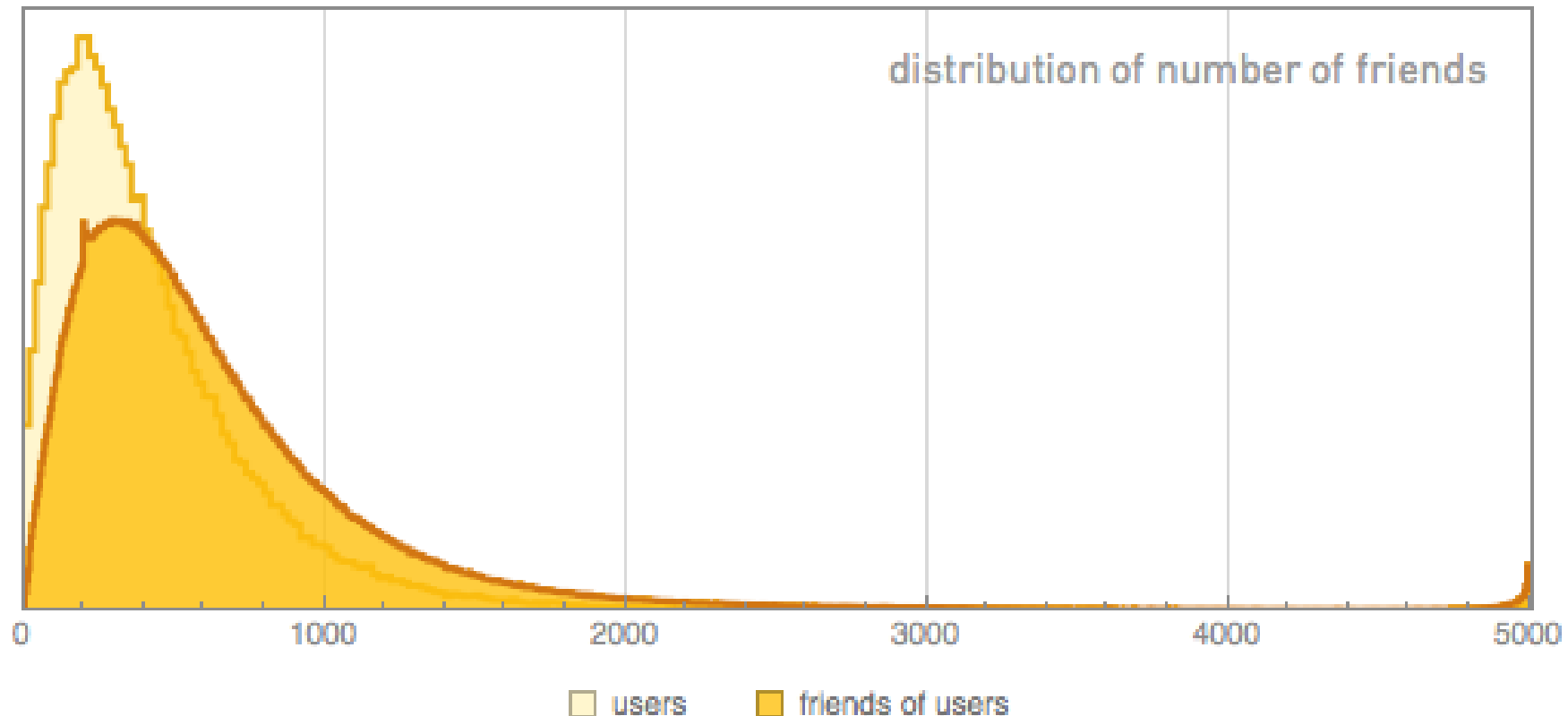
- **five** girls have fewer friends than the average of their friends (Betty, Jane, Pam, Dale, Tina), two have more friends
- **two** have more friends than the average of their friends (Sue and Alice)
- **one** has as many friends as the average of her friends (Carol)



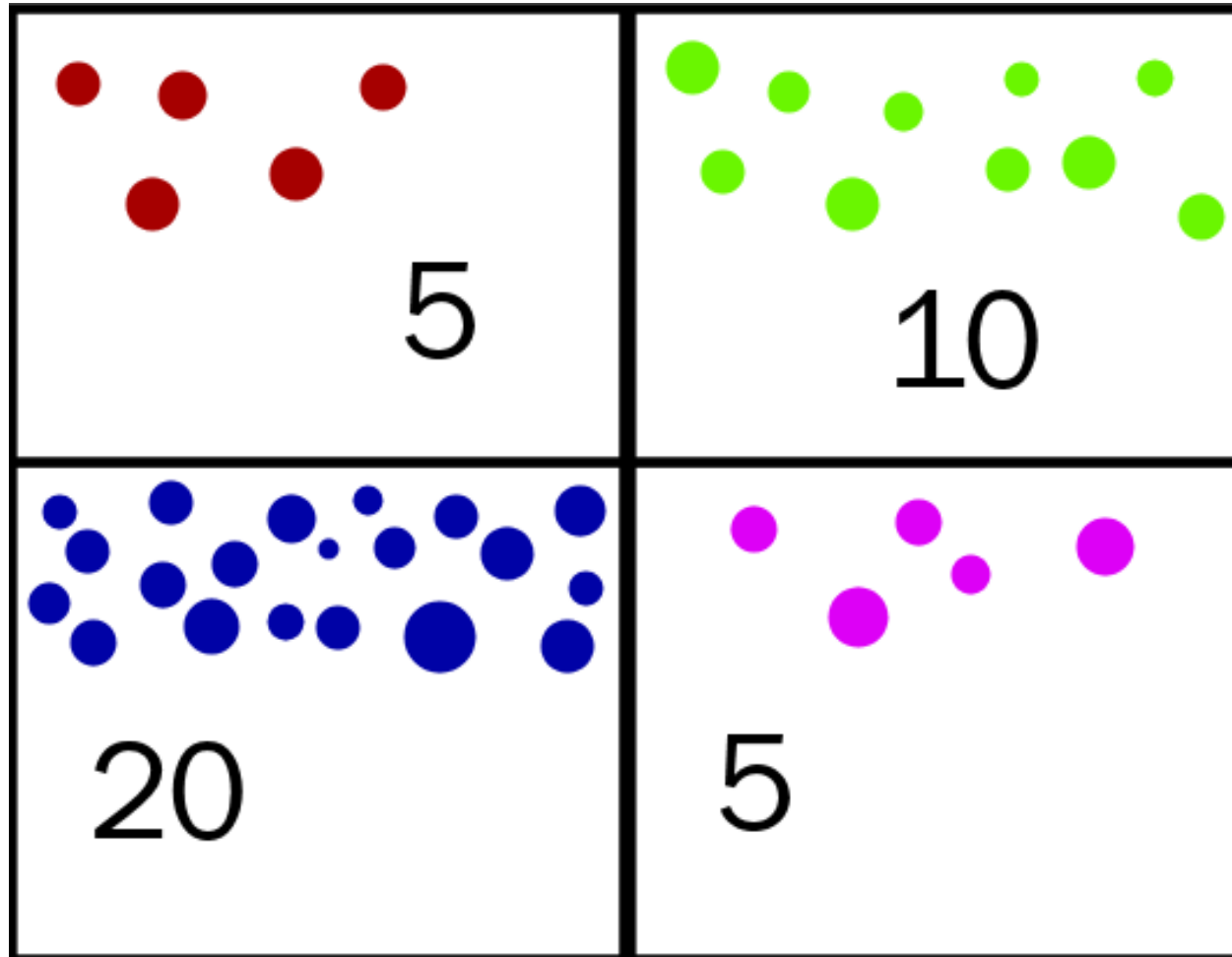
Most girls will think that they have less friends than the norm! If we average all numbers, the average number of friends is **2.5** and the average number of friends of friends is **2.99**.

The Friendship paradox

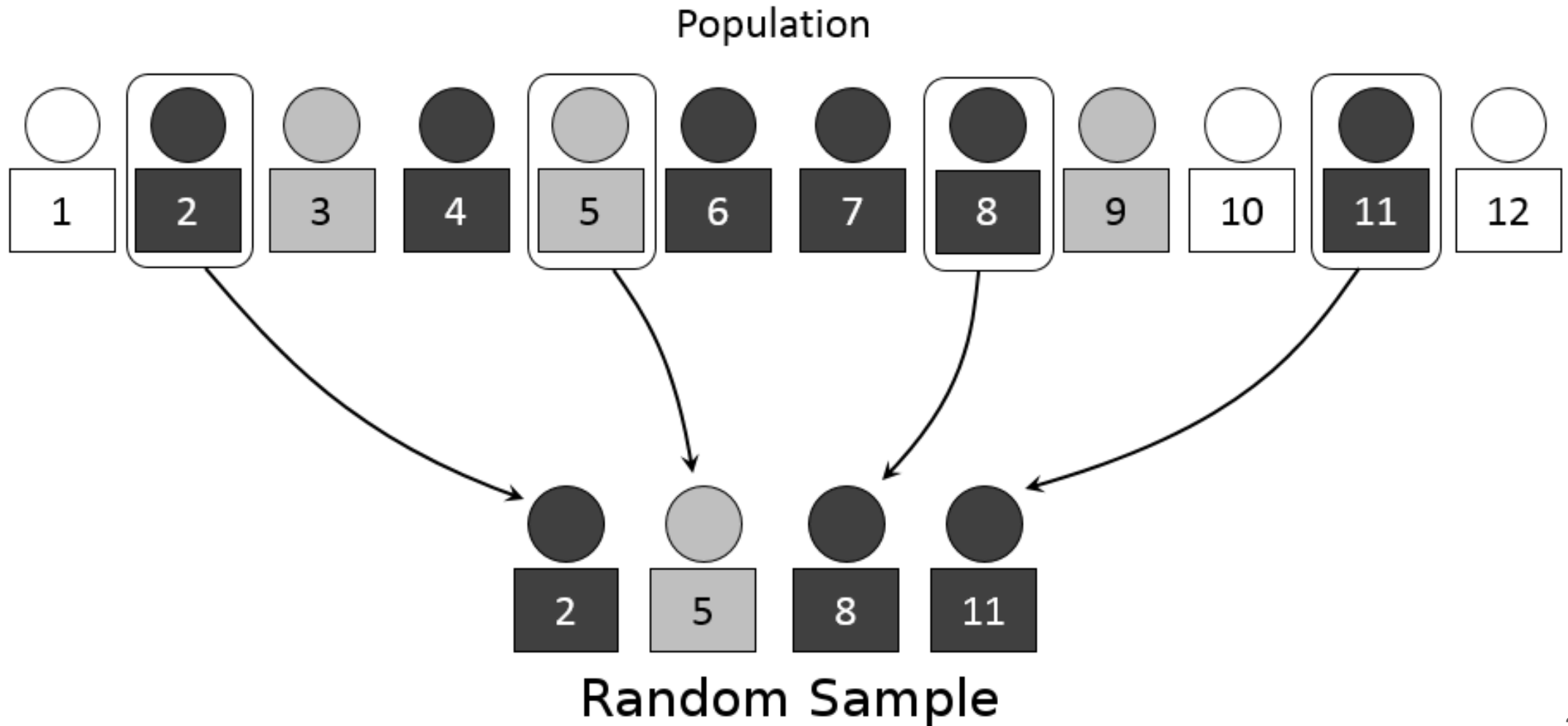
The Friendship Paradox: The fact that most people have less friends than the average number of friends of their friends



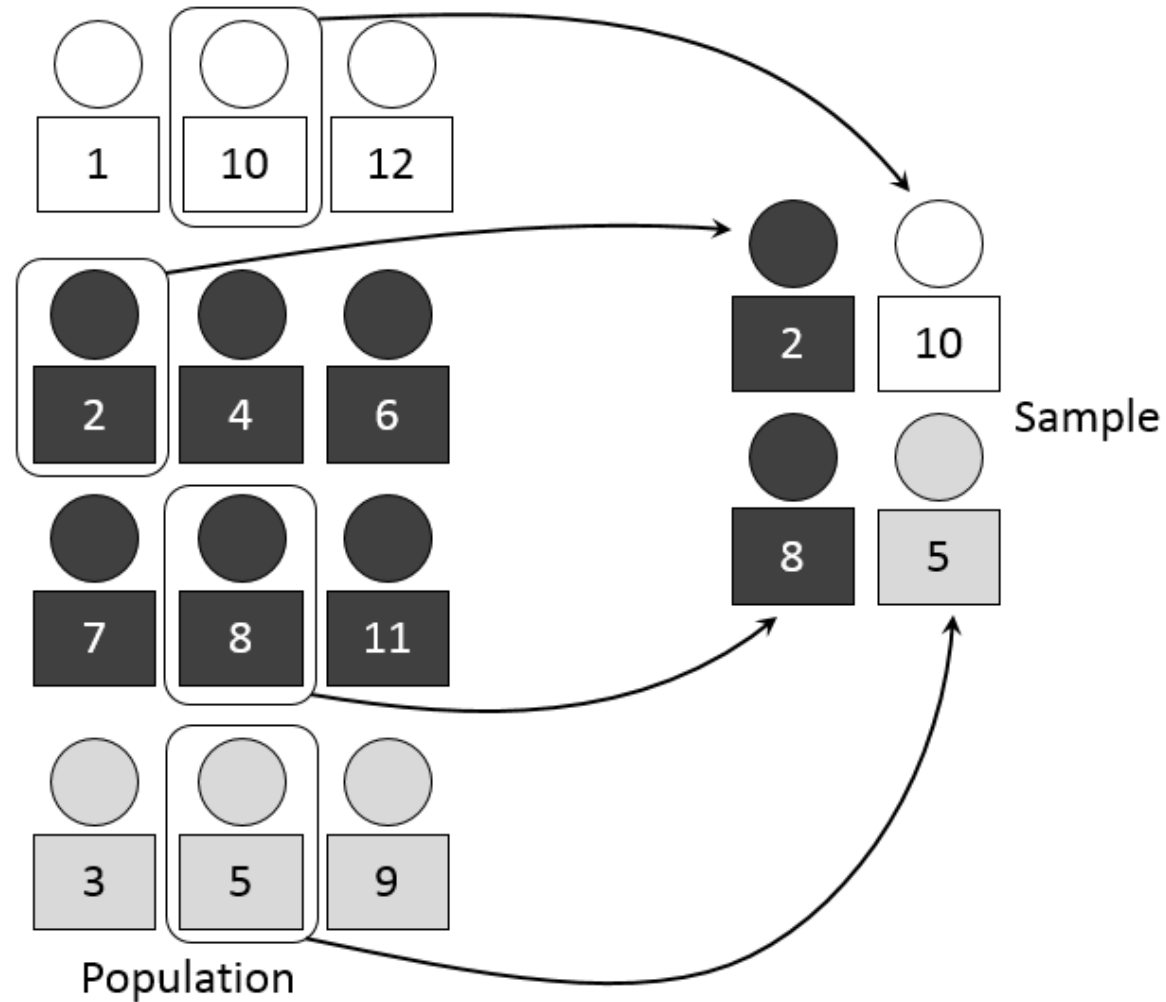
The class size paradox



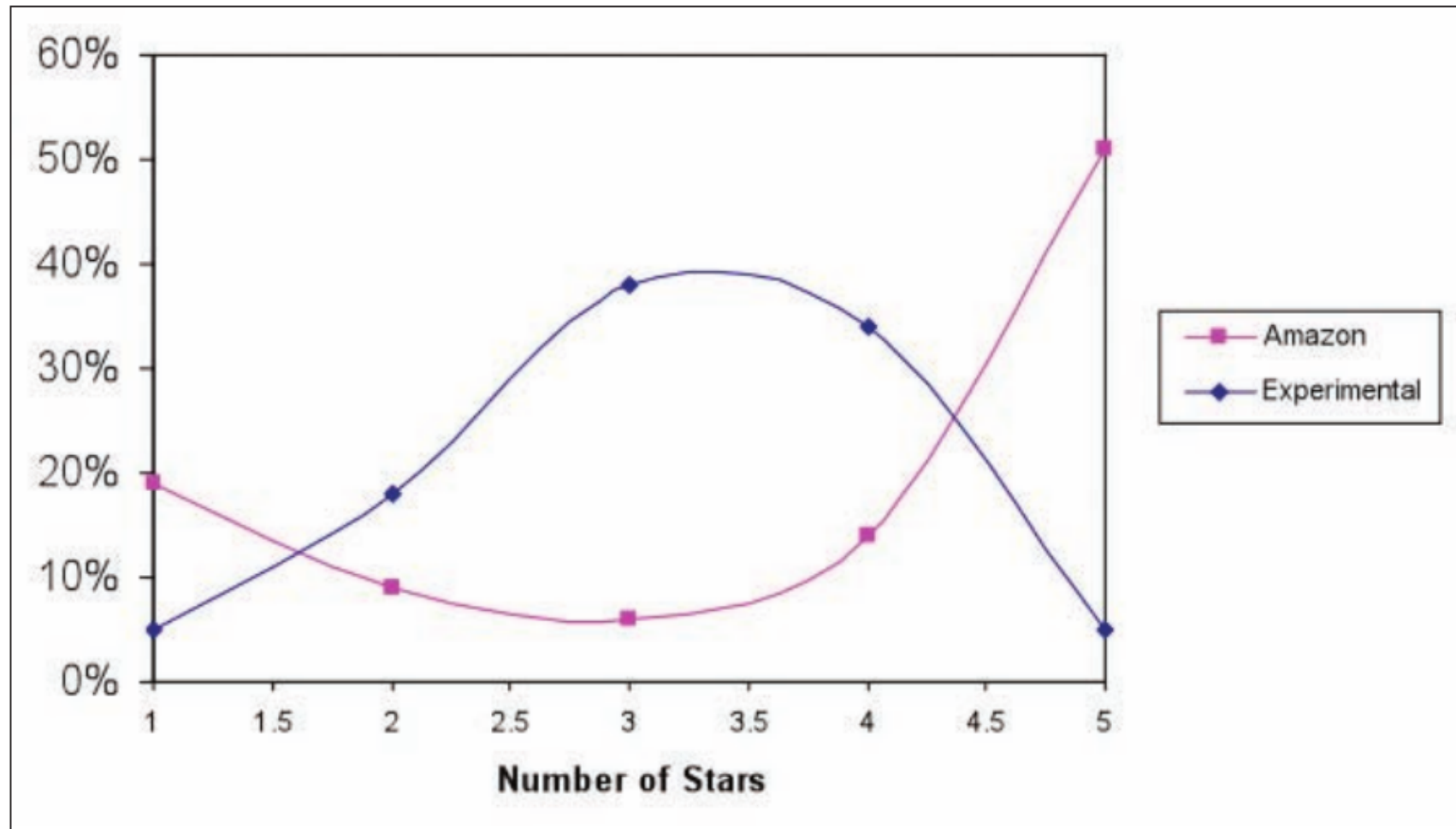
Random sampling



Stratified sampling



Self-selection bias



Summary

- **Projects guide**
 - Don't forget about the project!
- **Introduction to Social Networks**
 - Basic terminology: nodes, edges, directed networks, weights
 - Concepts: paths, distance, self-loops, connected components
- **The Friendship Paradox**
 - Most people have less friends than the average number of friends of their friends
 - Generated by sampling bias in social networks