# Graph Learning
# SD212
# 1. Graph Structure

Thomas Bonald

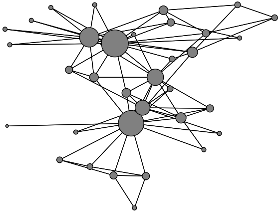
2023 – 2024

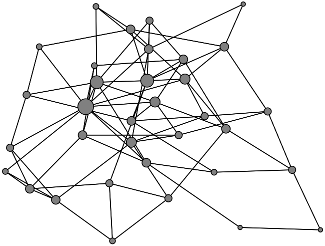
TELECOM Paris

IP PARIS

# Motivation

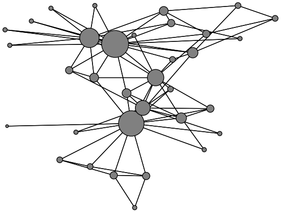
How do (large) **real graphs** look like?

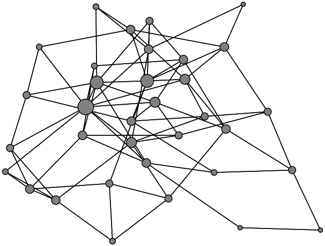




# Outline

1. **The friendship paradox**
2. Scale-free property
3. Small-world property
4. Clustering property





# The friendship paradox

You have less friends than your friends have.

# Sampling bias

Consider a graph of $n$ nodes and $m$ edges

Let $D$ be the degree of a random node

## Uniform sampling

$$\mathrm{E}(D) = \frac{2m}{n}$$

# Sampling bias

Consider a graph of $n$ nodes and $m$ edges

Let $D$ be the degree of a random node

## Uniform sampling

$$\mathrm{E}(D) = \frac{2m}{n}$$

## Edge sampling

$$\mathrm{E}'(D) = \frac{\mathrm{E}(D^2)}{\mathrm{E}(D)} \geq \mathrm{E}(D)$$

## Sampling bias

Consider a graph of $n$ nodes and $m$ edges

Let $D$ be the degree of a random node

### Uniform sampling

$$\mathrm{E}(D) = \frac{2m}{n}$$

### Edge sampling

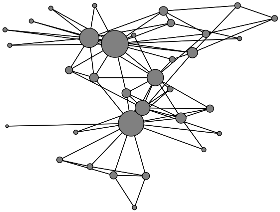
$$\mathrm{E}'(D) = \frac{\mathrm{E}(D^2)}{\mathrm{E}(D)} \geq \mathrm{E}(D)$$

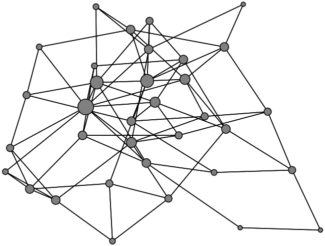
### Neighbor sampling

$$\mathrm{E}''(D) \geq \mathrm{E}(D)$$

# Outline

1. The friendship paradox
2. **Scale-free property**
3. Small-world property
4. Clustering property





# Power law

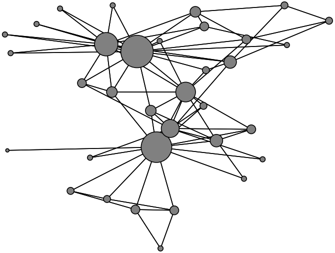
In real graphs, the degrees typically have a **power law** (= Pareto)

The degree $D$ of a random node satisfies

$$\mathrm{P}(D \geq k) \approx \left(\frac{k_{\min}}{k}\right)^{\alpha}$$

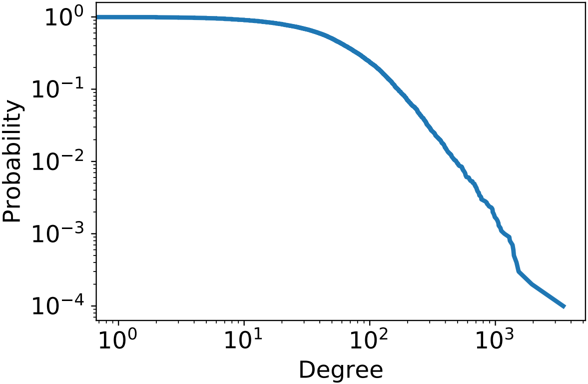
where

- $k_{\min}$ is the **minimum degree**
- $\alpha$ is the **exponent**, typically between 1 and 2



# Example

In-degree distribution of Wikipedia Vitals
(10,011 nodes, 824,999 edges)



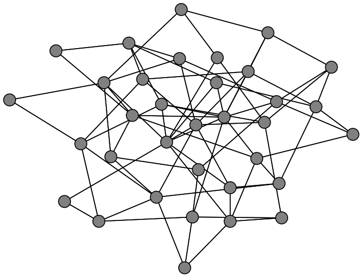
# A random graph

Erdős-Rényi (1959)

Consider $n$ nodes, with pairs connected with probability $p$

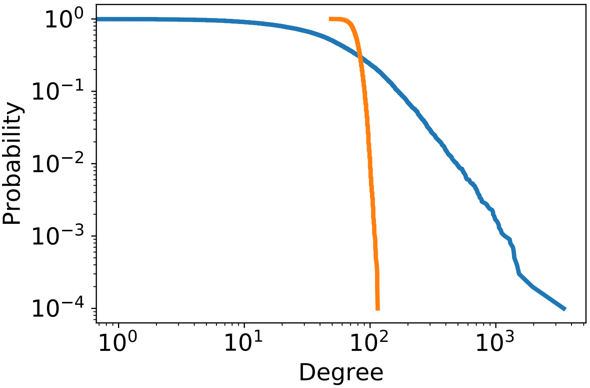
Adjacency matrix = symmetric matrix with

$$A_{ij} \sim \text{Bernoulli}(p) \text{ for } i < j$$



# Example

Wikipedia Vitals vs. random graph
(10,011 nodes, 824,999 edges)

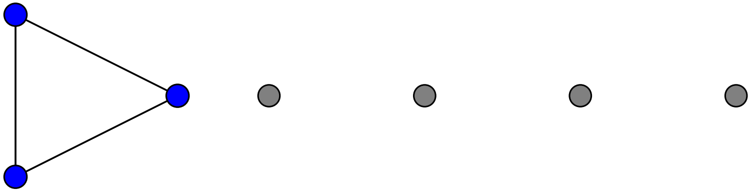


# Why the power law?

Barabasi-Albert model (1999)

Start from a clique of $d$ nodes (with $d \geq 1$)

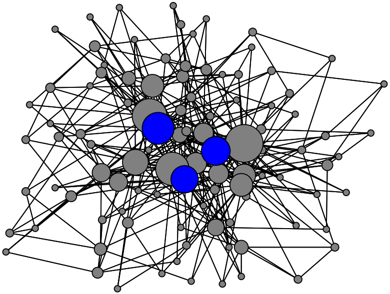
Add new nodes one at a time, each of degree $d$ and with **preferential attachment**



**"rich get richer"**

# Example

Graph generated from the **Barabasi-Albert model**
$(n = 100, d = 3)$



# The scale-free property

Let $D$ be the degree distribution:

$$\mathrm{P}(D \geq k) = \left(\frac{k_{\min}}{k}\right)^{\alpha}$$

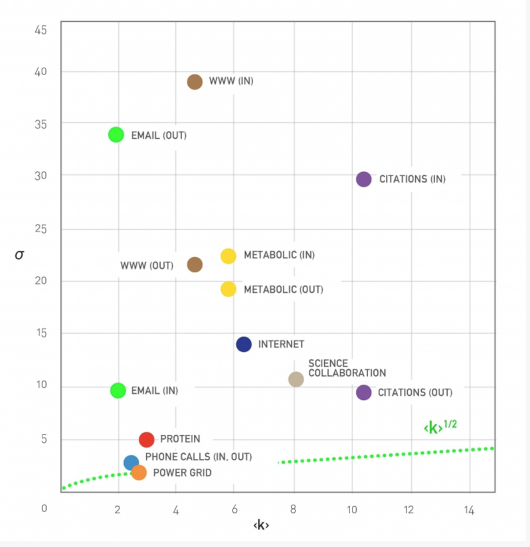
For a typical power exponent $\alpha \in (1, 2]$, we have:

$$\mathrm{E}(D) = \frac{\alpha}{\alpha - 1} k_{\min}$$

$$\mathrm{var}(D) = +\infty$$

$\rightarrow$ The average degree is **not** informative!

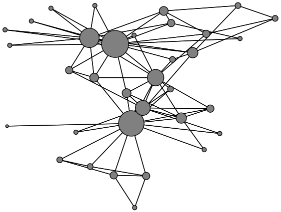
# Scale-free graphs

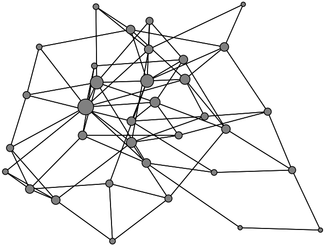



Source: Barabasi, Network Science, 2016

# Outline

1. The friendship paradox
2. Scale-free property
3. **Small-world property**
4. Clustering property





# Small world

Which fraction of the articles are **accessible in $k$ clicks** from Plato on Wikipedia?

# Small world

Which fraction of the articles are **accessible in $k$ clicks** from Plato on Wikipedia?

Using Wikipedia Vitals (10,011 articles):

| # clicks | # nodes | proportion |
|---|---|---|
| 1 | 392 | 4% |
| 2 | 5866 | 59% |
| 3 | 9939 | 99% |
| 4 | 9990 | 99.8% |

# The six degrees of separation

- First mention in *Chains*, a short story by Karinthy in 1929

# The six degrees of separation

- First mention in *Chains*, a short story by Karinthy in 1929
- Verified experimentally by Milgram in 1967


Source: Wikipedia

# The six degrees of separation

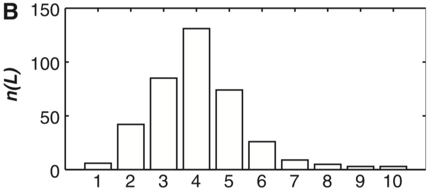
- First mention in *Chains*, a short story by Karinthy in 1929
- Verified **experimentally** by Milgram in 1967


Source: Wikipedia

# Emails

Dodds, Muhamad, Watts 2003

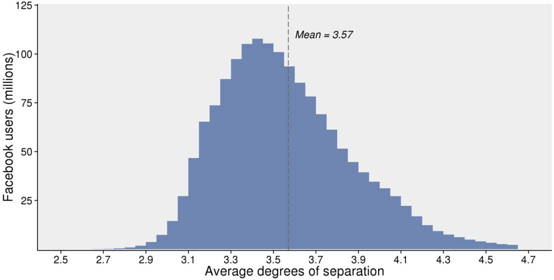
- 18 target people from all over the world
- 24,163 volunteers
- 384 successful chains
  Length of successful chains



# Facebook

Bhagat, Burke, Diuk, Filiz, Edunov 2016

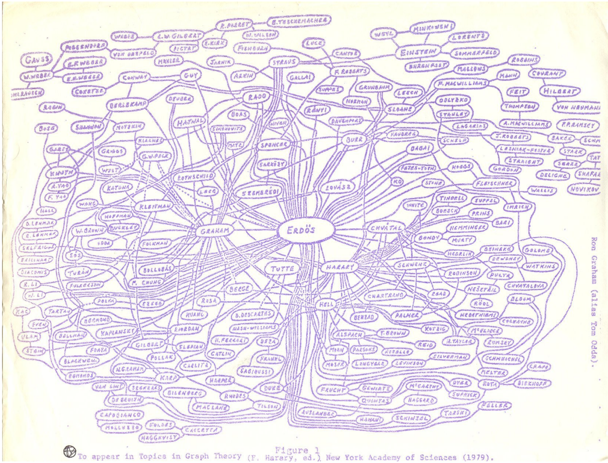
- Based on the 1.6 billion people active on Facebook
- Compute the average path length to any other people




The 3.5 degrees of separation of Facebook

# Erdős number

- Graph of co-authors of scientific papers
- Distance to Erdős (1913-1996)



# The Bacon number

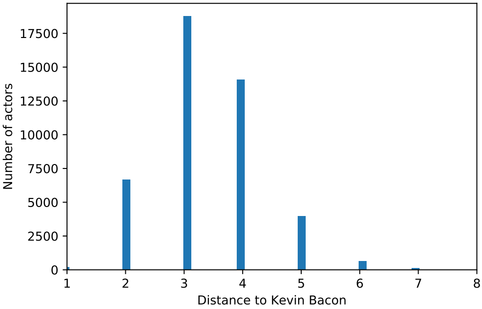
See The Oracle of Bacon

- Originated from an interview of Kevin Bacon by Premiere Magazine in 1994

# The Bacon number

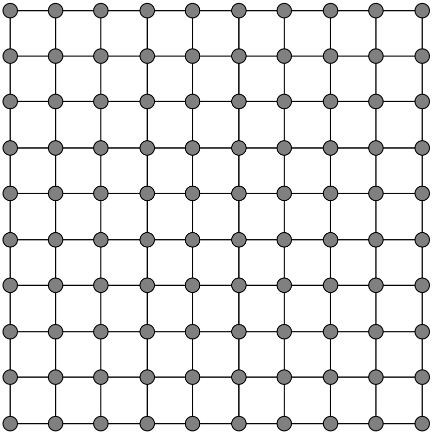
See The Oracle of Bacon

- Originated from an interview of Kevin Bacon by Premiere Magazine in 1994
- Graph of co-starring in movies

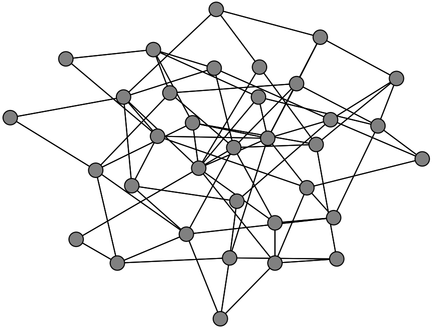



Results from YAGO database (44,586 actors)

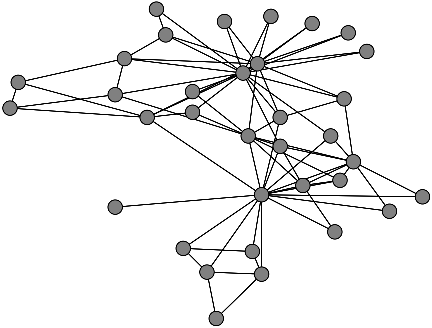
# Planar graphs



# Random graphs

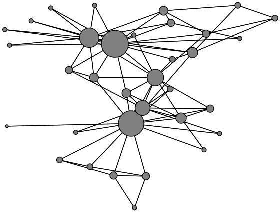


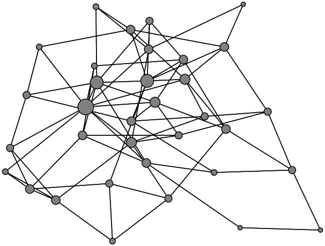
# Power-law graphs



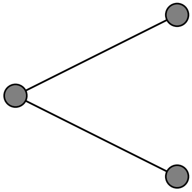
# Outline

1. The friendship paradox
2. Scale-free property
3. Small-world property
4. **Clustering property**

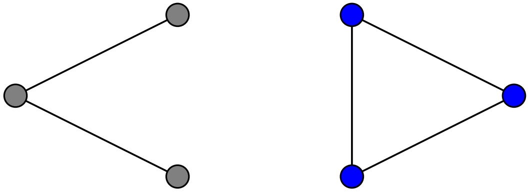




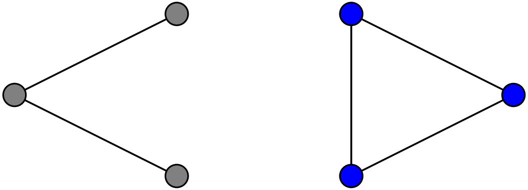
# Clustering coefficient



# Clustering coefficient



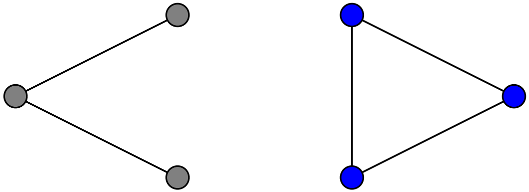
# Clustering coefficient



Fraction of **closed** triangles:

$$C = \frac{3\ \#\text{triangles}}{\sum_i \binom{d_i}{2}}$$

# Clustering coefficient



| Graph | $C$ |
|---|---|
| Karate Club | 0.26 |
| Les Miserables | 0.50 |
| Openflights | 0.25 |
| WikiVitals | 0.21 |

# Summary

## Graph structure

1. **Friendship paradox** (sampling)
2. **Power law** (degrees)
3. **Small world** (distances)
4. **Clustering** (triangles)