Lecture 1: Introduction to real-world networks and their properties

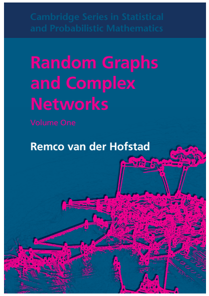
Remco van der Hofstad

MSRI Summer Graduate School on Random Graphs

July 5 – July 15, 2022

Material

▷ Intro random graphs:
Random Graphs and Complex Networks Volume 1
http://www.win.tue.nl/~rhofstad/NotesRGCN.html
Volume 2: in preparation on same site



Treat selected parts of Chapters I.1, I.6–I.8 and II.2–II.8, as well as related material.

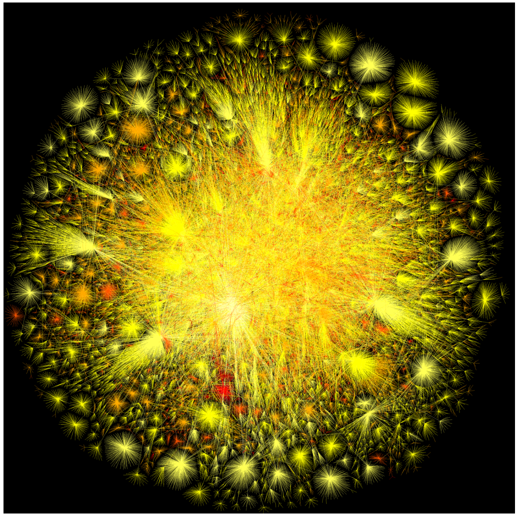
Argument are probabilistic, using
▷ first and second moment method;
▷ branching process approximations.

Will also use KONECT to show statistics of network statistics[a]

[a] KONECT project http://konect.cc

Complex networks

Yeast protein interaction network[a]



Internet 2010[b]

Attention focussing on unexpected commonality.

[a] Barabási & Óltvai 2004

[b] Opte project `http://www.opte.org/the-internet`

Graphs or networks

Network is another word for a graph.
Graphs are mathematical constructs to study

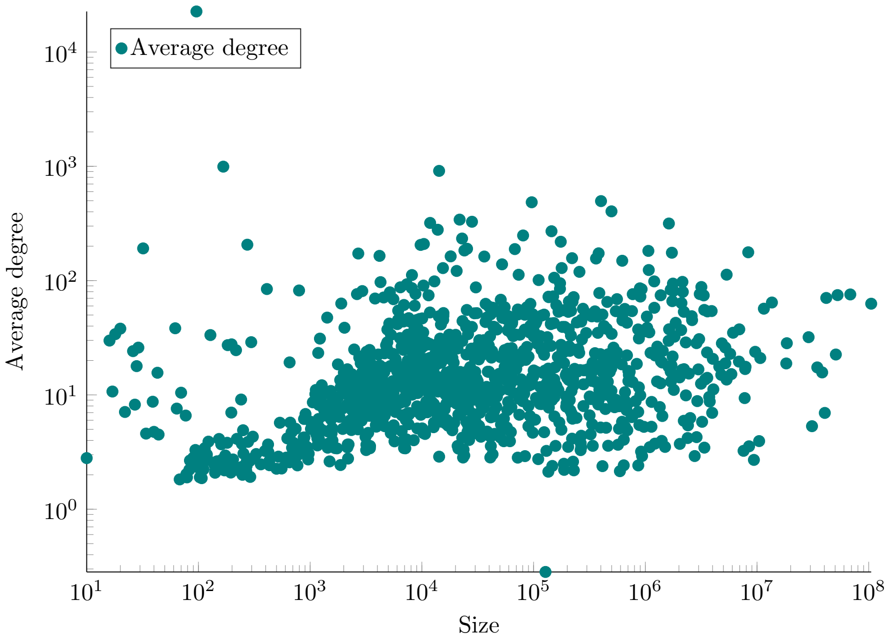
relations between objects.

Graph consists of vertices (= nodes, sites) and edges (= bonds).

Vertices: elements of the graph.
Edges: relations between the elements:
cables, friendships, who eats who, hyperlink,...

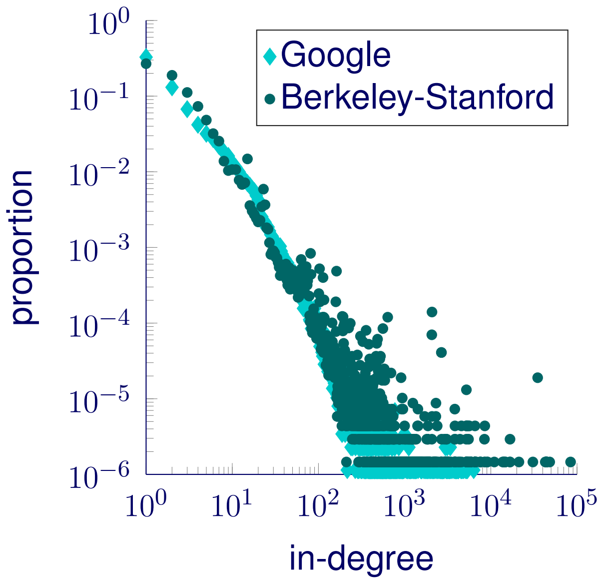
Edge is building block of relational data

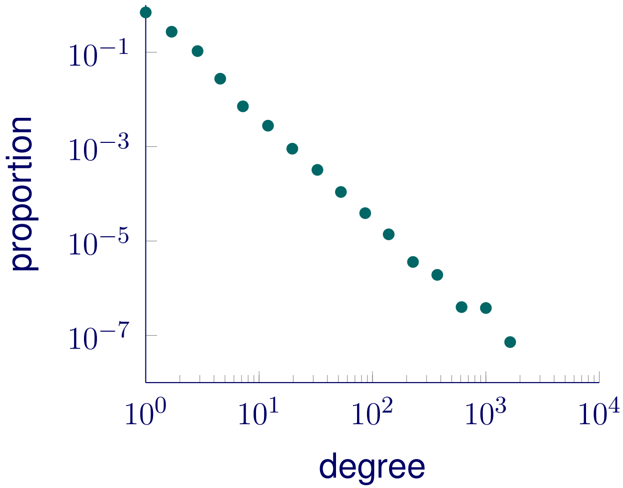
Networks are sparse



Average degrees of 1203 networks in KONECT

Scale-free paradigm





Loglog plot degree sequences WWW in-degree and Internet

- ▷ Straight line: proportion p_k of vertices of degree k satisfies $p_k = ck^{-\tau}$.
- ▷ Empirical evidence: Often $\tau \in (2,3)$ reported.

Scale-free paradigm

Degree sequence $(n_1, n_2, n_3, \ldots)$ of graph:

n_1 is number of elements with degree 1,
n_2 is number of elements with degree 2,
...
n_k is number of elements with degree k.

Then

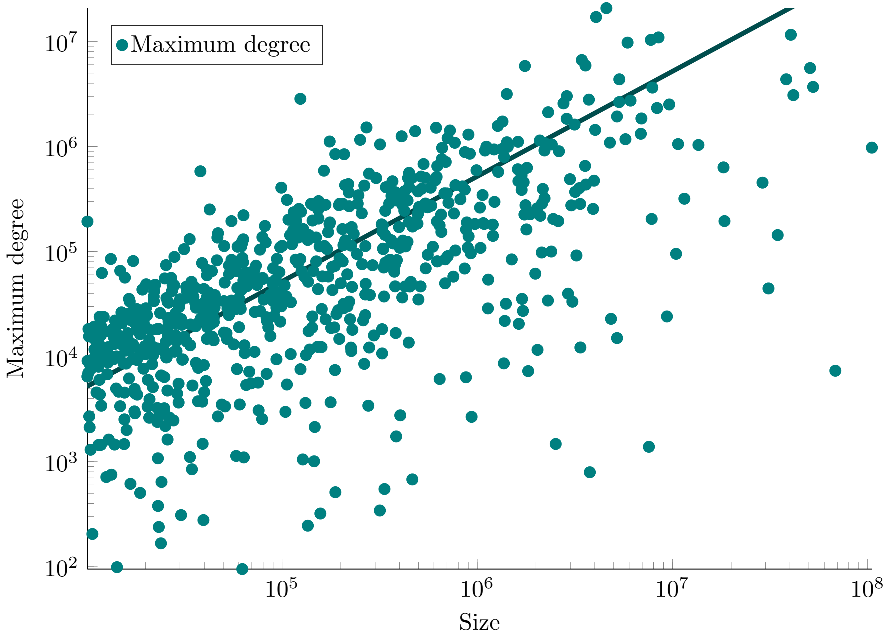
$$n_k \approx Ck^{-\tau},$$

precisely when

$$\log n_k \approx \log C - \tau \log k.$$

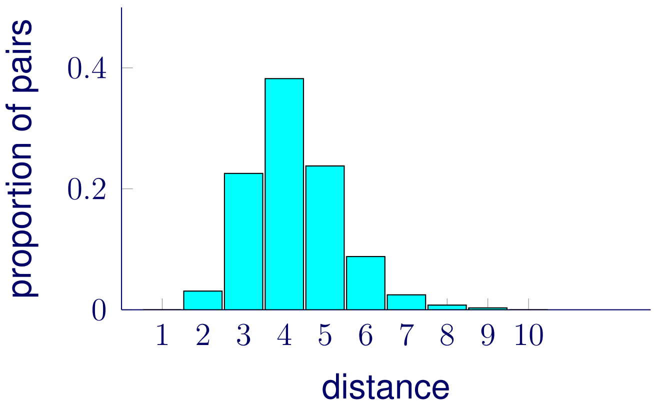
Approximate linear relationship $\log n_k$ **and** $\log k$

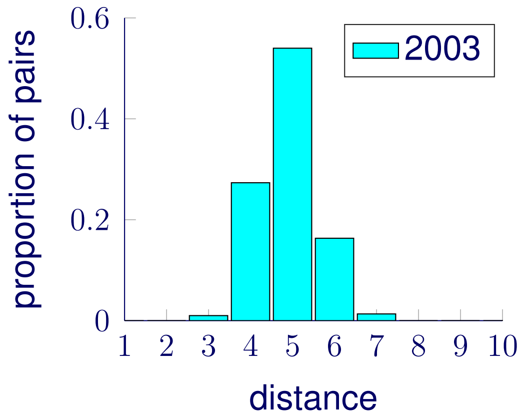
Network inhomogeneity



Maximal degrees in 727 networks larger than 10000 from KONECT
Linear regression gives $\log d_{\max} = 0.742 + 0.519 \log n.$

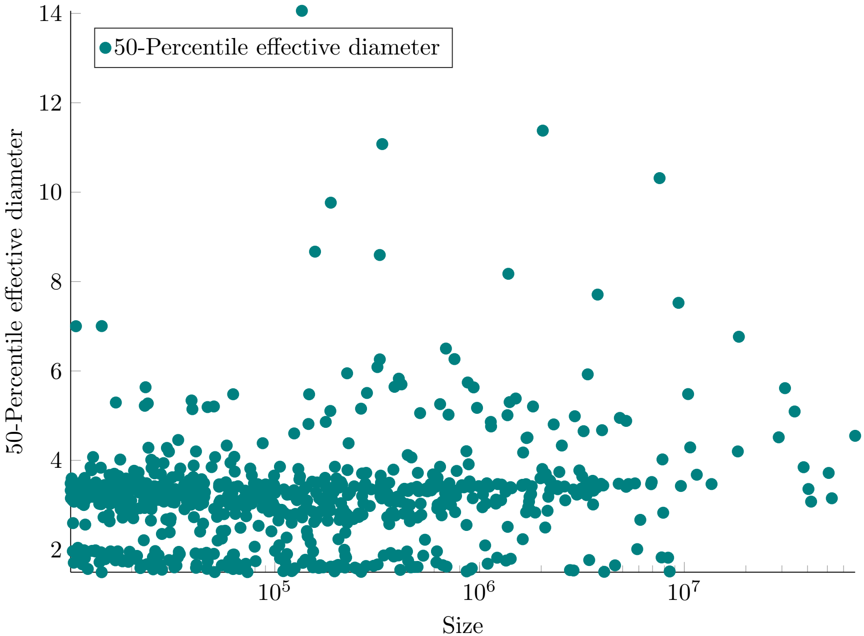
Small-world paradigm





Distances in Strongly Connected Component WWW and IMDb.

Network are small-worlds



Median typical distances in 727 networks larger than 10,000 in KONECT

Facebook

Largest virtual friendship network:
721 million active users,
69 billion friendship links.

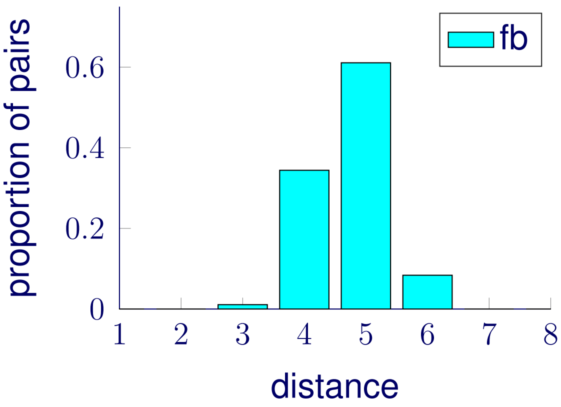
Typical distances on average four:

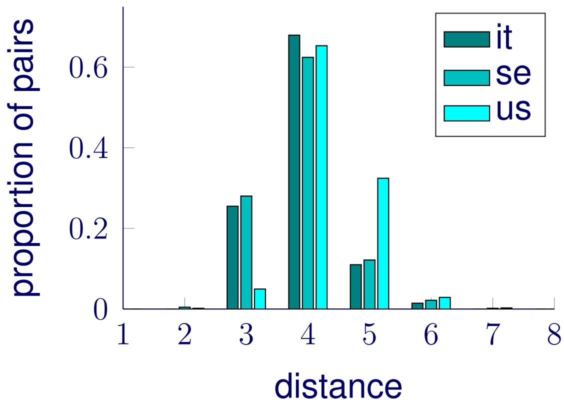
Four degrees of separation!

Fairly homogeneous (within countries, distances similar).

Recent studies:
Ugander, Karrer, Backstrom, Marlow (2011): topology
Backstrom, Boldi, Rosa, Ugander, Vigna (2011): graph distances.

Four degrees of separation





Distances in FaceBook in different subgraphs

Backstrom, Boldi, Rosa, Ugander, Vigna (2011)

Network statistics

▷ Clustering:

$$C = \frac{3 \times \text{ number of triangles}}{\text{number of connected triplets}}.$$

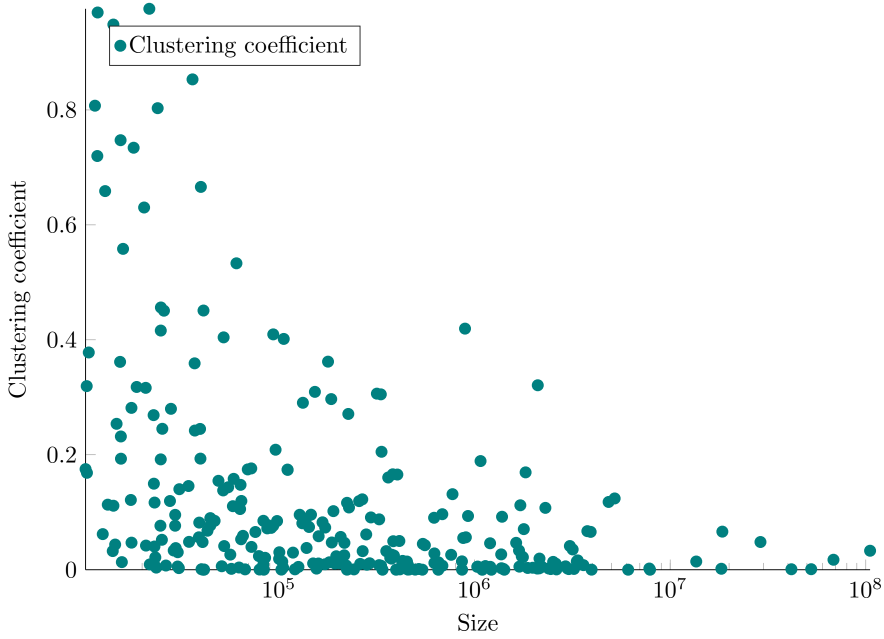
Proportion of friends that are friends of one another.

▷ Assortativity:

$$\rho = \frac{\frac{1}{|E_n|}\sum_{ij \in E_n} d_i d_j - \left(\frac{1}{|E_n|}\sum_{ij \in E_n} d_i\right)^2}{\frac{1}{|E_n|}\sum_{ij \in E_n} d_i^2 - \left(\frac{1}{|E_n|}\sum_{ij \in E_n} d_i\right)^2}.$$

Correlation between degrees at either end of edge.

Network are clustered

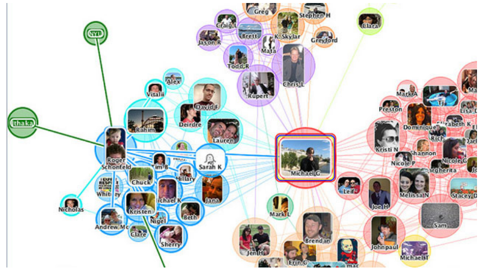


Clustering coefficient in 727 networks larger than 10,000 in KONECT

Friendship paradox

Average number of friends random individual equals

$$\sum_k k\mathbb{P}(D = k) = \frac{2|E|}{n},$$



where $|E|$ is number of edges.

Wikipedia: The average number of friends that a typical friend has can be modeled by choosing, uniformly at random, an edge of the graph and an endpoint of that edge, and again calculating the degree of the selected endpoint.

With $D^\star$ degree vertex in random edge,

$$\mathbb{E}[D^\star] = \mathbb{E}[D] + \frac{\mathrm{Var}(D)}{\mathbb{E}[D]} > \mathbb{E}[D].$$

Your friends have more friends than you do!

Friendship paradox

Take vertex uniformly at random, then take one of its neighbors and inspect its degree. Denote degrees at both sides (D_1, D_2). Then,

- $\triangleright$ D_1 has same distribution as D, but
- $\triangleright$ D_2 does not have same distribution as $D^\star$!

Still

$$\mathbb{E}[D_2] > \mathbb{E}[D]!$$

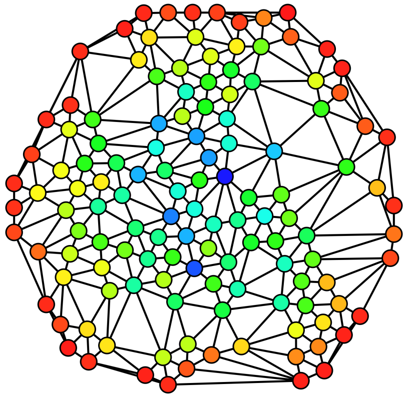
Your friends have more friends than you do!

Centrality measures

▷ Closeness centrality:
Measures to what extent vertex can reach others using few hops.
Vertices with low closeness centrality are central in network.

▷ Betweenness centrality:
Measures extent to which vertex connects various parts of network.

Betweenness large for bottlenecks.



▷ PageRank:
Measures extent to which vertex is visited by random walk.
Used in Google to rank importance in web pages.

Network science

▷ Complex networks modelled using

random graphs.

▷ Network functionality modelled by stochastic processes on them.

▷ A plethora of examples:

Disease spread	Synchronization
Information diffusion	Robustness to failures
Consensus reaching	Information retrieval
Percolation	Random walks...

▷ Also algorithms on networks important: PageRank, assortativity, community detection,...

▷ Prominent part of applied math for decades to come.

Models complex networks

▷ Inhomogeneous Random Graphs:
Static random graph, independent edges with inhomogeneous edge occupation probabilities, yielding scale-free graphs.
(Chapters I.6, II.2 and II.5)

[Extensions of Erdős-Rényi random graphs Chapters I.4 and I.5.]

▷ Configuration Model:
Static random graph with prescribed degree sequence.
(Chapters I.7, II.3 and II.6)

▷ Preferential Attachment Model:
Dynamic model, attachment proportional to degree plus constant.
(Chapters I.8, II.4 and II.7)

Universality??

Erdős-Rényi

Erdős-Rényi random graph is random subgraph of complete graph on $[n] := \{1, 2, \ldots, n\}$ where each of $\binom{n}{2}$ edges is occupied independently with prob. p.

Simplest imaginable model of a random graph.

$\triangleright$ Attracted tremendous attention since introduction 1959, mainly in combinatorics community:

Probabilistic method (Spencer, Erdős et al.).

$\triangleright$ Average degree equals $(n-1)p \approx np$, so choose $p = \lambda/n$ to have sparse graph.

$\triangleright$ Egalitarian: Every vertex has equal connection probabilities. Misses hub-like structure of real networks.

Conclusion networks

Networks useful to interpret real-world phenomena:
centrality and friendship paradox.

Many real-world networks share important features:

scale-free and small-world paradigms.

Often, suggestion of infinite-variance degrees.

This course focuses on models invented to describe properties:

Configuration model, generalized random graph
preferential attachment model.