Karlsruhe Institute of Technology

Resilient Networking

Disclaimer: this lecture has been created with very valuable input from Jussi Kangasharju

Module 2 – Background on Graphs (Winter Term 2021)

Thorsten Strufe

Competence Center for Applied Security Technology

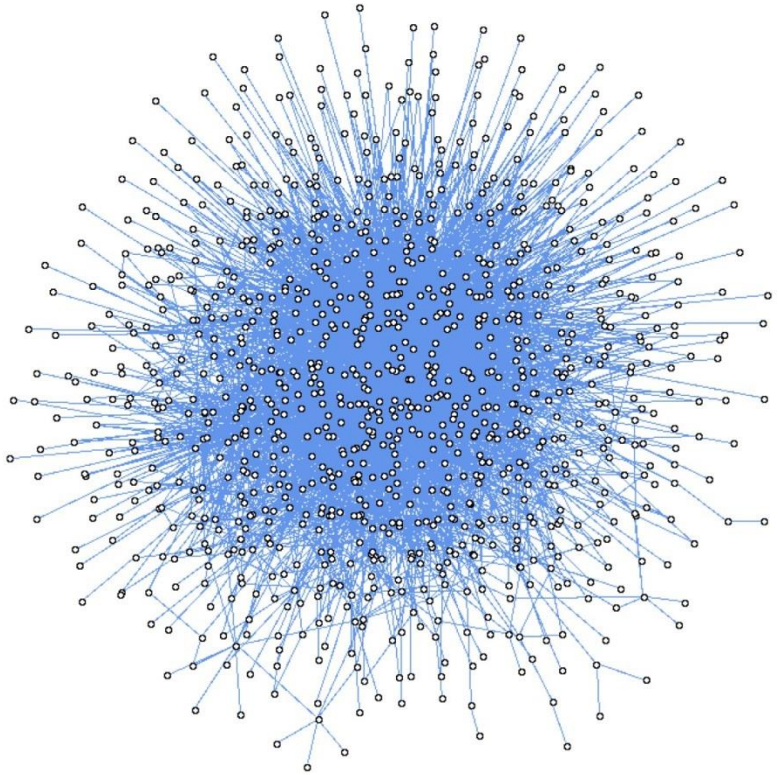
Module Outline

- Background 1: Graph Analysis
  - Why bother with theory?
  - Graphs and their representations
  - Important graph metrics
  - On robustness and resilience

- Background 2: Crypto
  - Stream ciphers and the OTP
  - Block ciphers and their operation modes
  - Key agreement
  - Asymmetric Crypto
  - Integrity

Some questions...

- How robust is the Internet?
- Why do darknets work?
- What do the existing networks actually look like, and why?
- What would an ideal computer network look like?



Gnutella snapshot, 2000

Graphs

- Graph families and models
  - Random graphs
  - Small world graphs
  - Scale-free graphs

- Graph theory and real computer networks
  - How are the graph properties reflected in real systems?
    - Users/nodes are represented by vertices in the graph
    - Edges represent connections in overlay / routing table entries

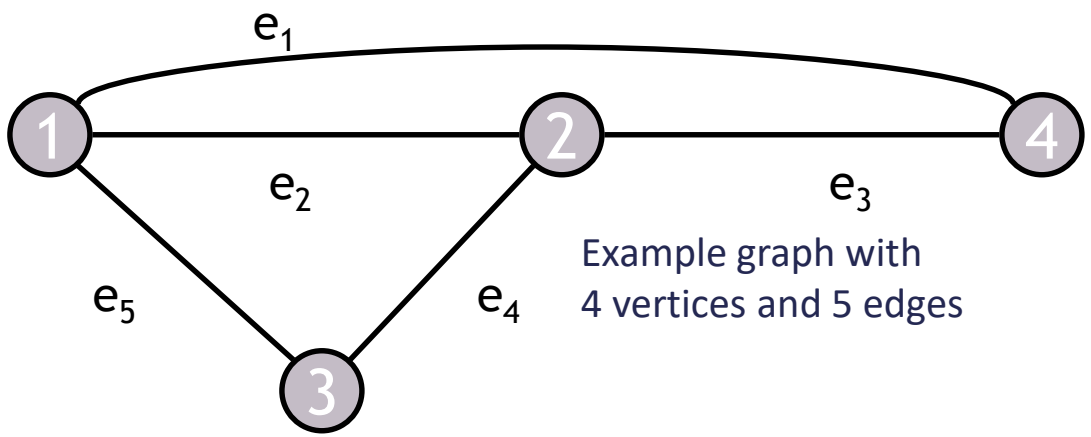
- Concept of self-organization (how/why do they evolve?)
  - Network structures emerge from simple rules
  - E.g. also in social networks, www, actors playing together in movies

What is a Graph?

- Definition of a graph:

  Graph $G = (V, E)$ consists of two finite sets, set V of vertices (nodes) and set E of edges (arcs, links) for which the following apply:

  1. If $e \in E$, then exists $(v, u) \in V \times V$, such that $v \in e$ and $u \in e$
  2. If $e \in E$ and above (v, u) exists, and further for $(x, y) \in V \times V$ applies $x \in e$ and $y \in e$, then $\{v, u\} = \{x, y\}$



Example graph with
4 vertices and 5 edges

Side note:

Edges can have (multiple) "weights" $w : E \rightarrow R$

Properties of Graphs

- An edge e $\in$ E is directed if the start and end vertices in condition 2 above are identical: $v = x$ and $y = u$

- An edge e $\in$ E is undirected if $v = x$ and $y = u$ as well as $v = y$ and $u = x$ are possible

- A graph G is directed (undirected) if the above property holds for all edges

- Graph $G_1 = (V_1, E_1)$ is a subgraph of $G = (V, E)$, if $V_1 \subseteq V$ and $E_1 \subseteq E$ (such that conditions 1 and 2 are met)

How are Graphs Implemented?

- Adjacency/Incidence Matrix

	1	2	3
1	0	1	0
2	1	0	1
3	0	1	0

- Adjacency/Incidence List

(1,2)
(2,1),(2,3)
(3,2)

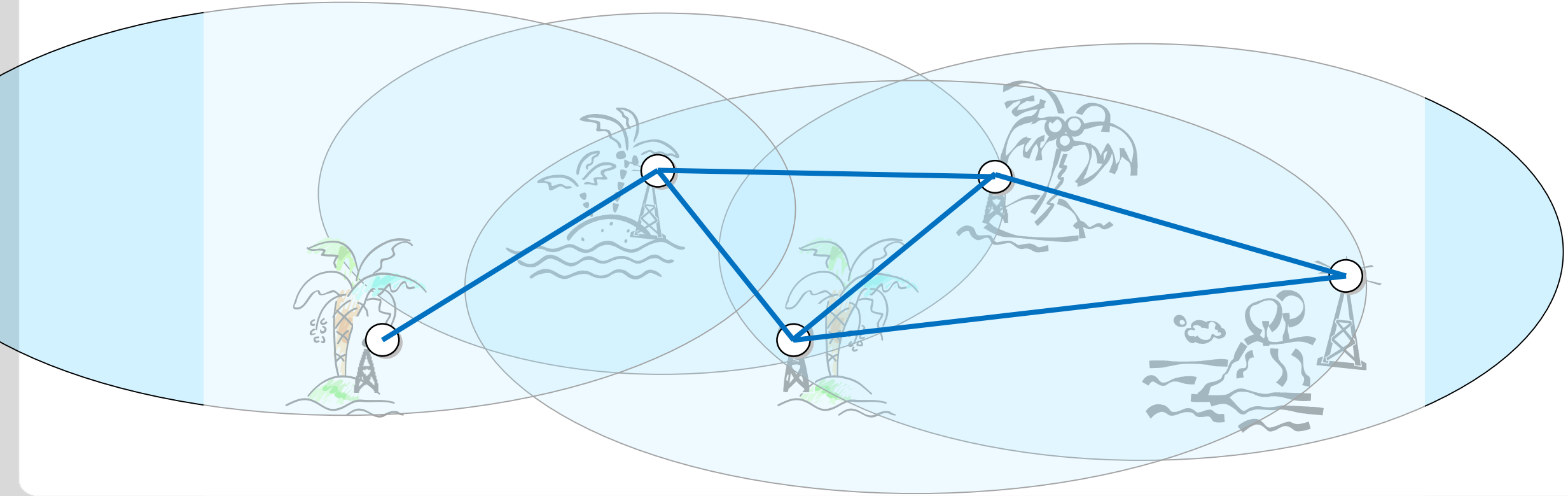
1:2
2:1,3
3:2

- (Plus specialized others..)

VERY good book is: Sedgewick: Algorithms in C, part 3 (Graph Algorithms)

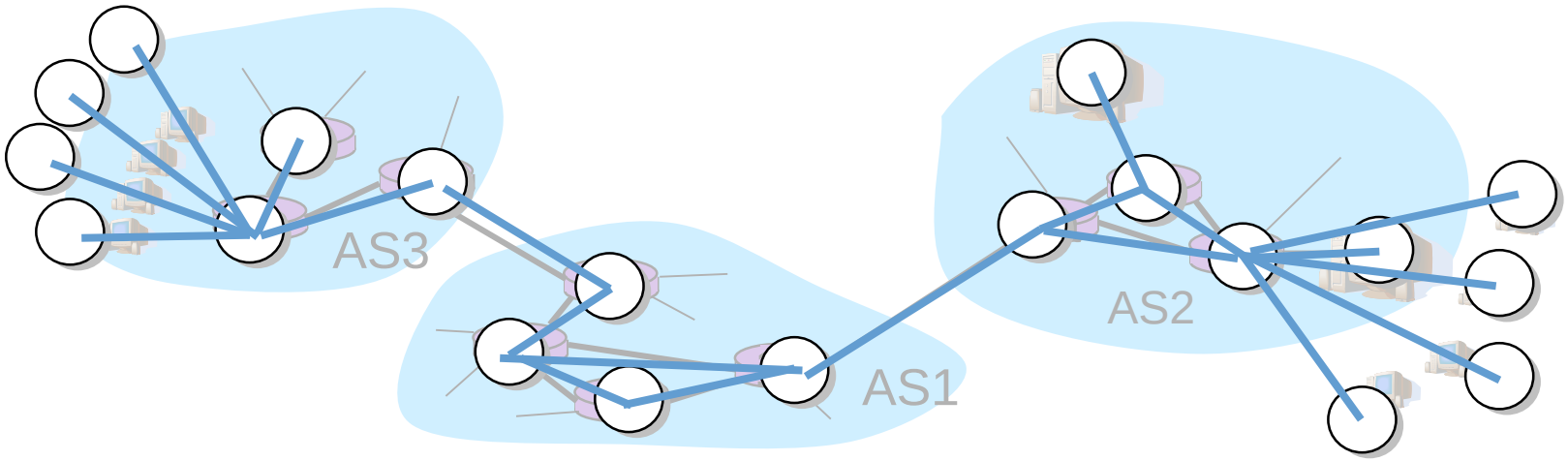
Some Examples of Computer Networks

- Early Computer Networks Aloha (or WSN, for that matters)
- Network Layers 1,2



Examples: The Internet

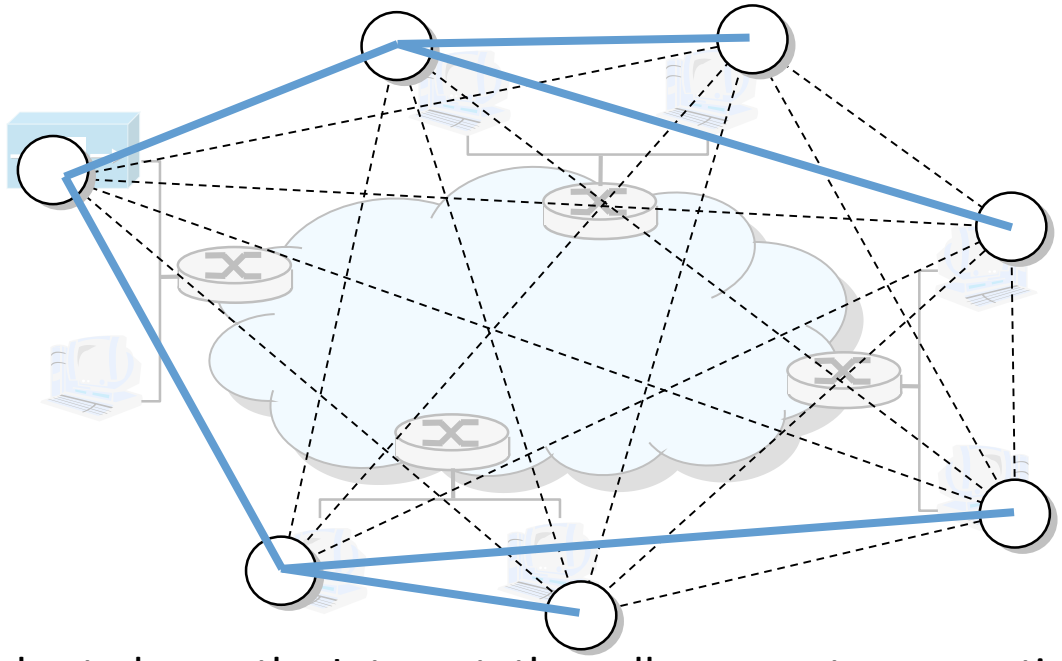
- Globally internetworked computers (Layer 3)



Examples: Overlays (Layer 7 (?))

- A CLIQUE is a graph that is fully connected $(u,v) \in E$ | for all $u \in V$ and $v \in V$, $u \neq v$

- A (P2P) Overlay (V_o,E_o) (in general) is a subgraph such that $V_o=V$ and $E_o \subseteq E$
  (edges are selected edges from a CLIQUE graph)



- ***Why?*** Considering the nodes to be on the Internet, they all can create connections between each other…

Important Graph Metrics

- Order: the number of vertices in a graph: |V|
- Size of the graph is the number of edges |E|

- Distance: $d(v, u)$ between vertices v and u is the length of the shortest path between v and u

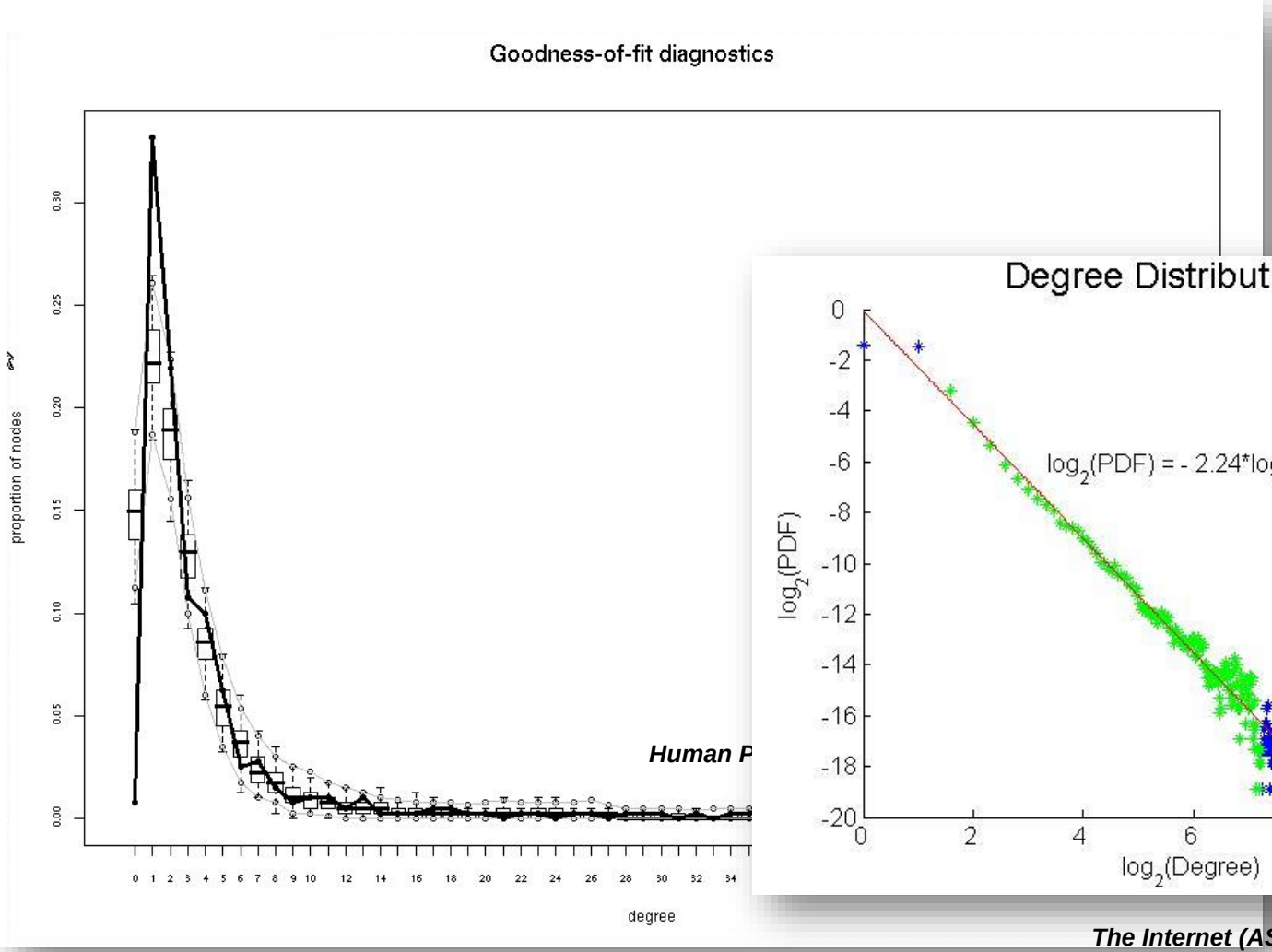
- Diameter: $d(G)$ of graph G is the maximum of $d(v, u)$ for all $v, u \in V$

- The density of a graph is the ratio of the number of edges and the number of possible edges.

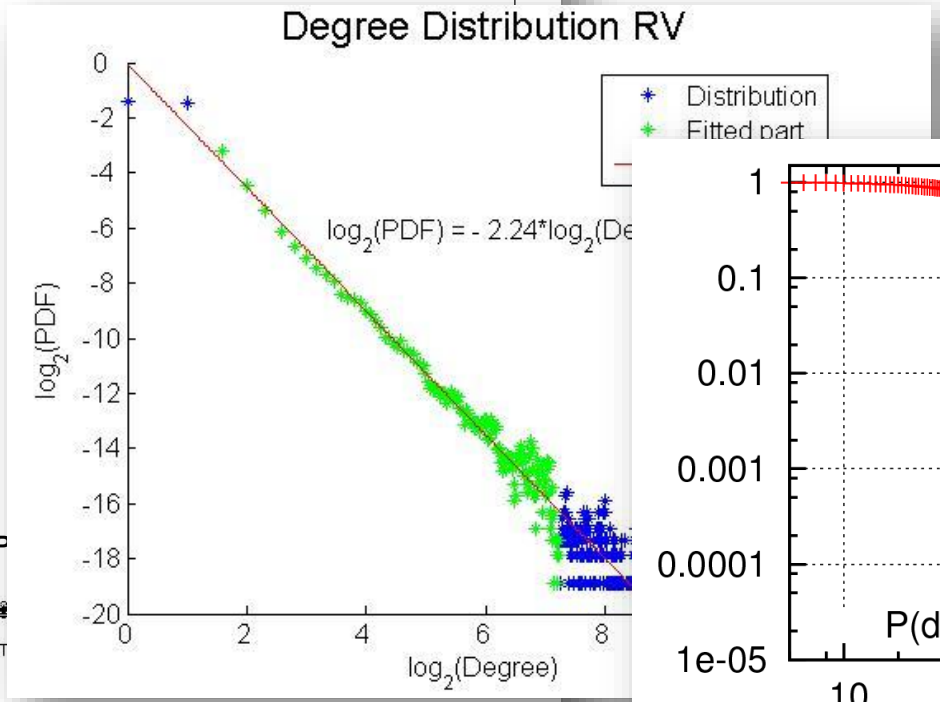
Graph Metrics: Vertex Degree

- In graph $G = (V, E)$, the degree of vertex $v \in V$ is the total number of edges $(v, u) \in E$ and $(u, v) \in E$
  - Degree is the number of edges incident to a vertex
- For directed graphs, we distinguish between in-degree and out-degree
  - In-degree is number of edges with the vertex as end-point
  - Out-degree is number of edges going with the vertex as starting point
- The degree of a vertex can be obtained as:
  - Sum of the elements in its row in the incidence matrix
  - Length of its vertex incidence list
- The ***degree distribution*** is the distribution over all node degrees
  (given as a frequency distribution or (often) complementary cumulative distribution function CCDF (Komplement der Verteilungsfunktion))

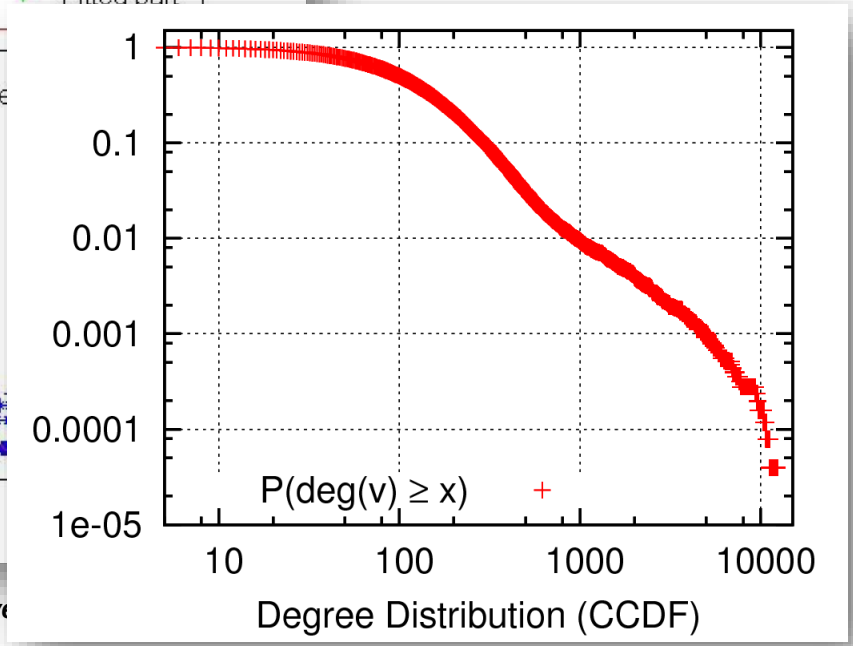
Graph Metrics: Degree Distribution (Examples)



Human P



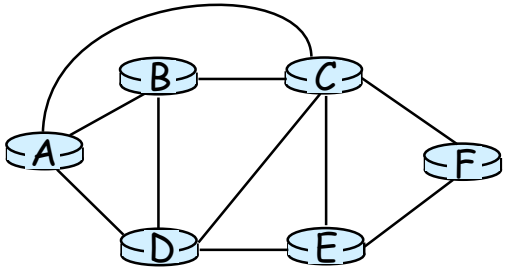
The Internet (AS-leve



Online Social Network (xing crawl) [strufe10popularity]

Routing and Graph Metrics on Path Length

- Routing: Define strategy to find path from *s* to *d*
  commonly local strategy, based on address/distances,
  usually "greedy"



- All pairs shortest paths (APSP): $d(v, u)$ | all $v,u \in V$

- Hop Plot: Distance distribution over all distances *Hist(APSP(G))*

- Average/characteristic path length (CPL): Sum of the distances over all pairs of nodes divided by the number of pairs

- *For defined routings (usually greedy) on directed graphs:*

Characterisic Routing Length (CRL): average length of paths *found* (potentially stochastic…)

Important Graph Metrics: Connectivity

- Edge connectivity: is the minimum number of edges that have to be removed to separate the graph into at least two components

- Vertex connectivity: the minimum number of nodes..

- How can we calculate them?
- Which of both is higher?
- In which cases are they the same?
- So where do you attack, naively? ;-)

- Homework: check maxflow, Menger‘s Theorem

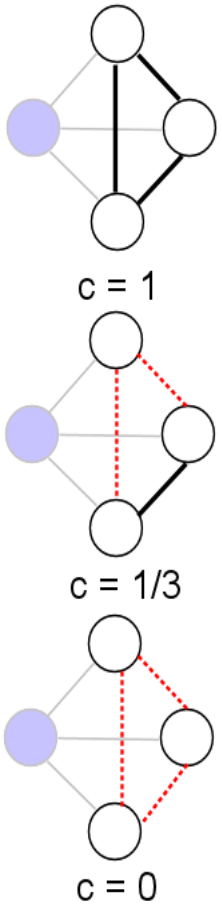
Graph Metrics: Network Clustering

- Clustering coefficient: number of edges between neighbors divided by maximum number of edges between them
  - k neighbors: k(k-1)/2 possible edges between them

$$C(i) = \frac{2E(N(i))}{d(i)(d(i)-1)}$$

E(N(i)) = number of edges between neighbors of *i*

d(i) = degree of *i*



Source: Wikipedia

- What if: a node has only one neighbor? ☺

- Variations exist: *local, average, global CC*

Classes of Graphs

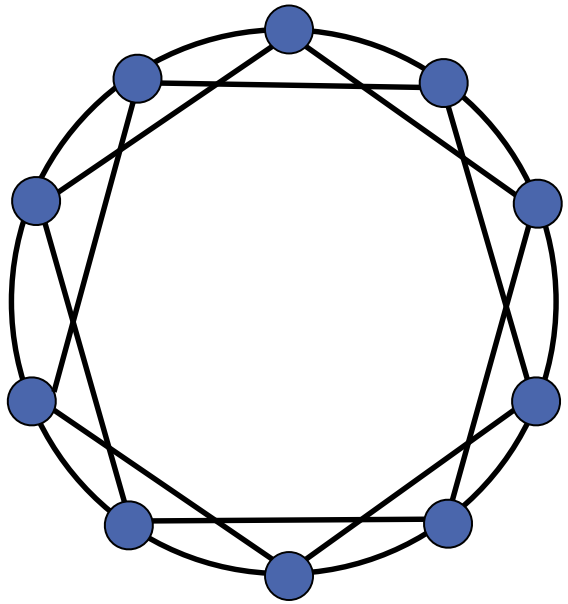
- Regular graphs
- Random graphs
- Graphs with Small-World characteristic
- Scale-free graphs

- …Graphs with plenty more characteristics
  - (dis-) assortativity
  - Rich-club connectivity
  - …

Regular Graphs

- Regular graphs have traditionally been used to model networks, they have
  - constant node degree (discrete degree distribution of a single value)
  - potentially different topologies
- However, the model does not reflect real nets well

Regular Graph



Random Graphs

- Random graphs are first widely studied graph family
  - Many overlay networks choose neighbors more or less randomly

- Two different generators generally used:
  - Erdös and Renyi
  - Gilbert

- Gilbert's definition: Graph $G_{n,p}$ (with n nodes) is a graph where the probability of an edge $e = (v, w)$ is p

Construction algorithm:

- For each possible edge, draw a random number in (0,1)
- If the number is smaller than p, then the edge exists
- (p can be function of n or constant)

Basic Properties of Random Graphs

Karlsruhe Institute of Technology

Giant Connected Component

Let c > 0 be a constant and p = c/n.

If c < 1 every component of $G_{n,p}$ has order O(log N) with high probability.

*If c > 1 then there is one component of order **n*(f(c) + O(1))** where f(c) > 0, with high probability. All other components have order O(log N)*

- ***English***: Giant connected component emerges with high probability when average degree is about 1

Node degree distribution

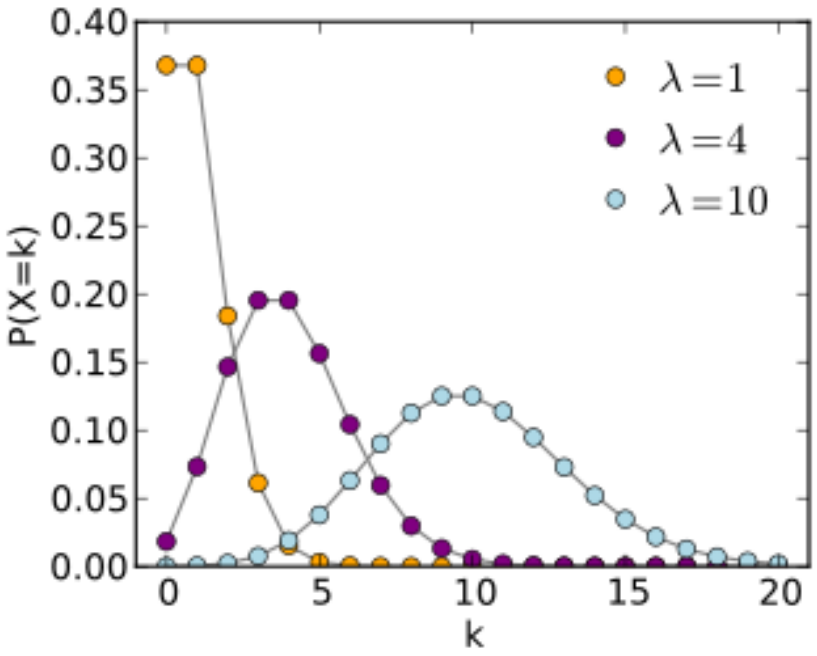
- If we take a random node, how high is the probability *P(k)* that it has degree *k*?
- Node degree is Poisson distributed
  - Parameter c = expected number of occurrences

$$P(k,c) = \frac{c^k e^{-c}}{k!}$$

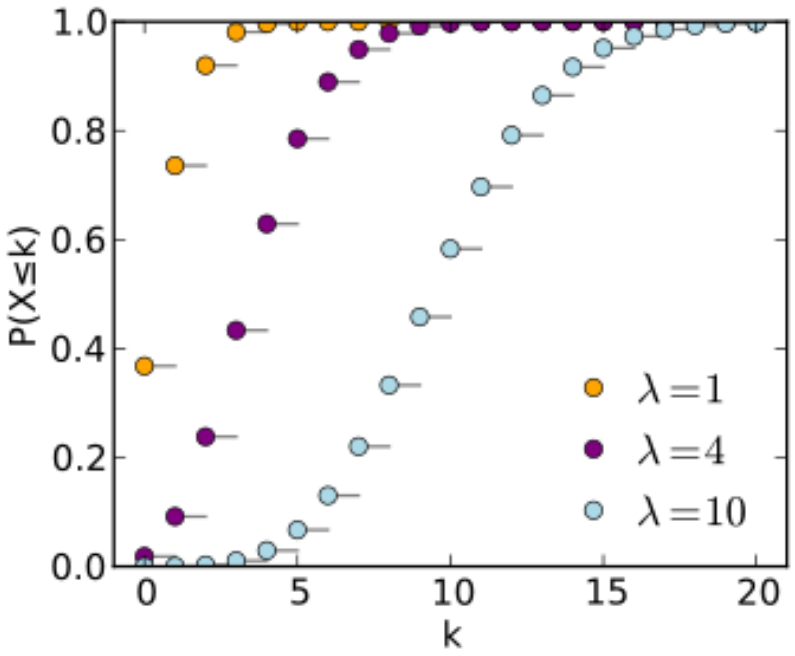
Clustering coefficient

- Clustering coefficient of a random graph is asymptotically equal to *p* with high probability

Recall: Poisson PDF and CDG



$$P(k,c) = \frac{c^k e^{-c}}{k!}$$



Source: Wikipedia

Utility of Random Graphs

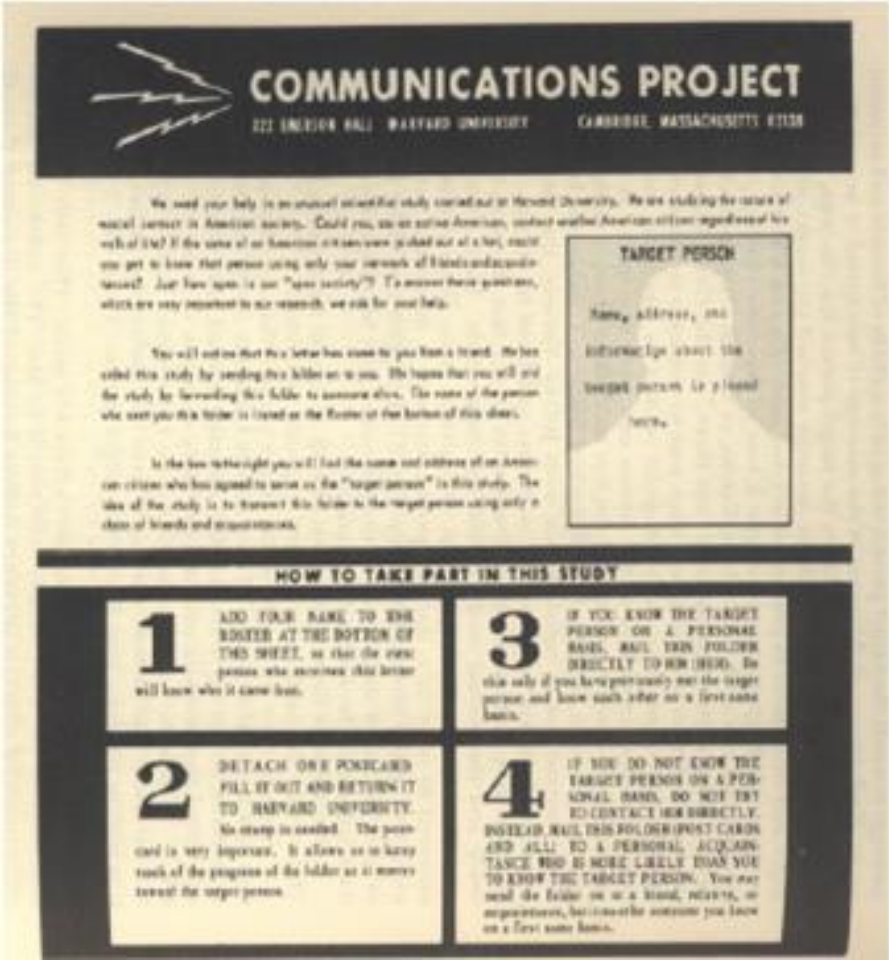
- *All models are wrong, but some are useful.*

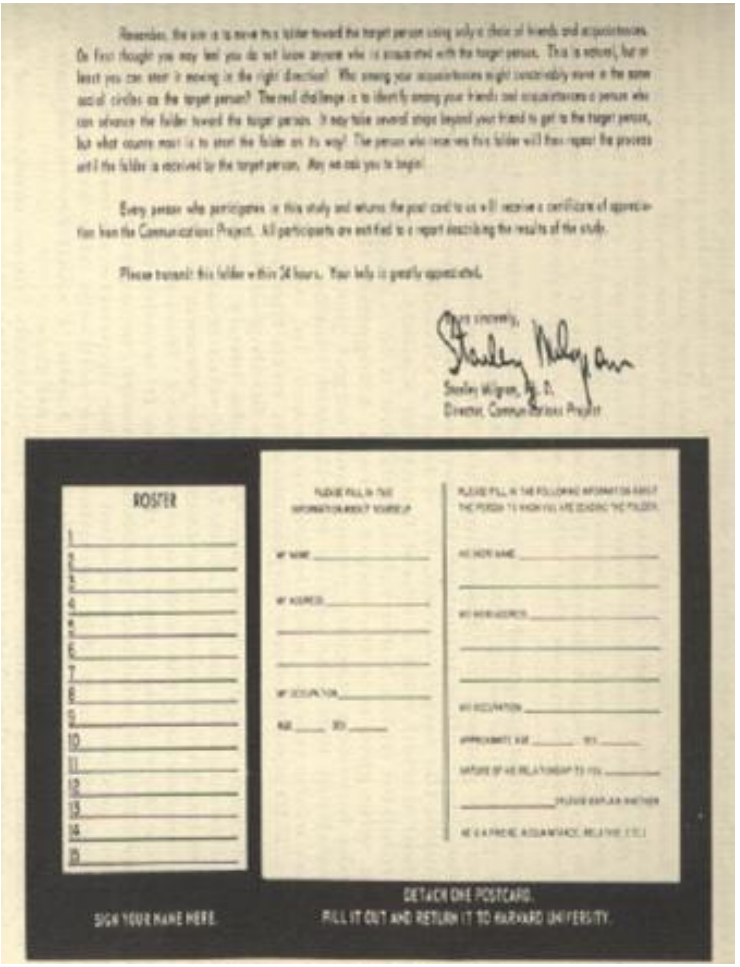
  -- George Box

- Random Graphs are useful!
  - Approximate reality with regards to some properties
  - Easy to analyze

- Random Graphs are useless!
  - Don‘t model reality well!
  - Many properties of real-world networks diverge considerably from this random case

Milgram's Small World Experiment





Six Degrees of Separation

- Famous experiment from 1960's (S. Milgram)

- Send a letter to random people in Kansas and Nebraska and ask people to forward letter to a person in Boston
  - Person identified by name, profession, and city
- ***Rule***: Give letter only to people you know by first name and ask them to pass it on according to same rule
  - Note: ***Some*** letters reached their goal

- Letter needed six steps on average to reach the destination

- Graph theoretically: Social networks have dense local structure, but (apparently) small diameter
  - Generally referred to as "small world effect"
  - Usually, small number of persons act as "hubs"

Small-World Network Model

- Developed/discovered by Watts and Strogatz (1998)
  - Over 30 years after Milgram's experiment!

- Watts and Strogatz looked at three networks
  - Film collaboration between actors, US power grid, Neural network of worm *Caenorhabditis elegans ("C. elegans")*

- Measured characteristics:
  - Clustering coefficient as a measure for 'regularity', or 'locality' of the network
    - If it is high, short edges exist with high probability
  - The average path length between vertices

- Results:
  - Grid-like networks:
    - High clustering coefficient $\Rightarrow$ high average path length
      (edges are not 'random', but rather 'local')
  - Most real-world (natural) networks have a **high clustering coefficient** (0.3-0.4), but nevertheless a **low average path length**

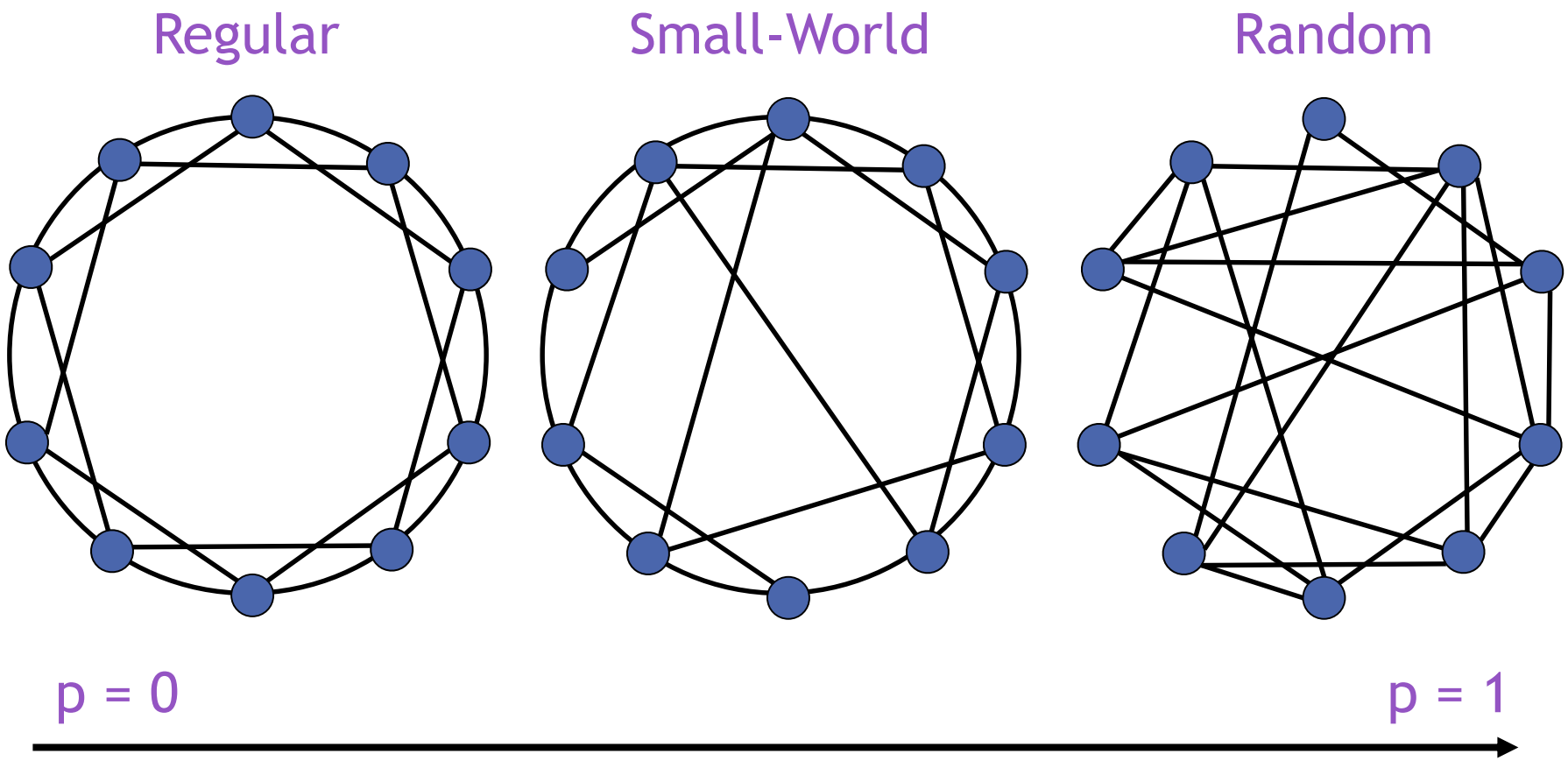
Small-World Graph Generator (W&S)

- Put all n nodes on a ring, number them consecutively from 1 to n
- Connect each node with its k clockwise neighbors
- Traverse ring in clockwise order

- For every edge
  - Draw random number r
  - If $r < p$, then re-wire edge by selecting a random target node from the set of all nodes (no duplicates)
  - Otherwise keep old edge

- Different values of p give different graphs
  - If p is close to 0, then original structure mostly preserved
  - If p is close to 1, then new graph is random
  - Interesting things happen when p is somewhere in-between

Regular, Small-World, Random



Utility of Small World Property

- Small world property
  - Explains why short paths exist

- Does not explain, how and why they are found?

Kleinberg's Small-World Navigability Model

- Small-world model explains why short paths exist

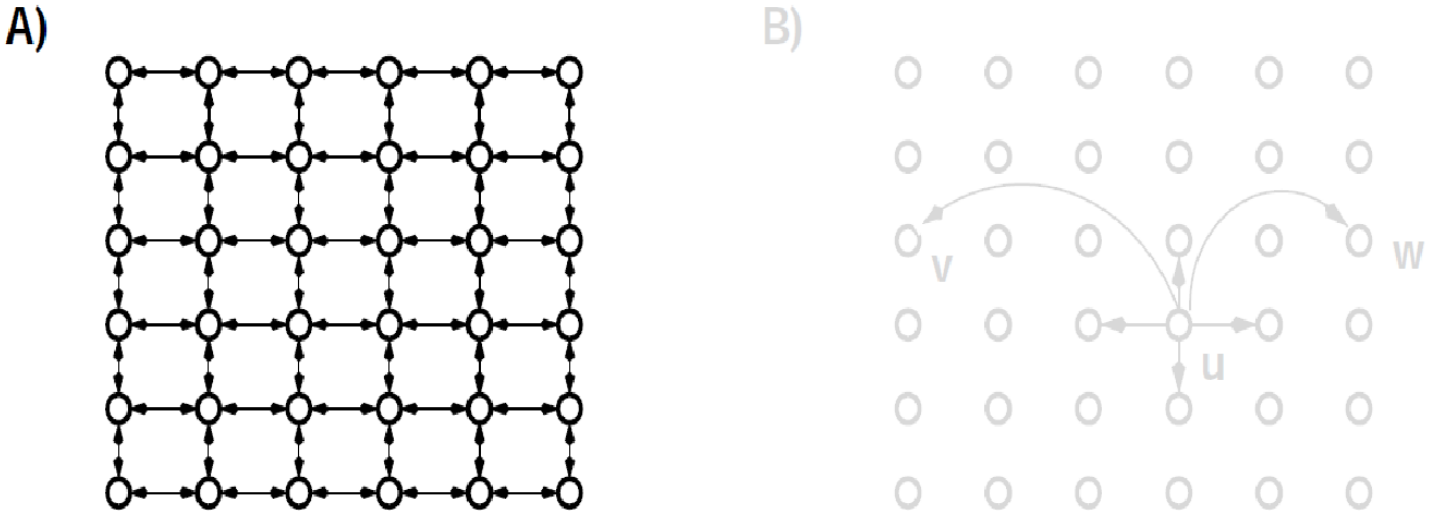
- Missing piece in the puzzle: why can we find these paths?
  - Each node has only local information
  - Even if a shortcut exists, how do people know about it?
  - Milgram's experiment:
    - Some additional information (profession, address, hobbies etc.) is used to decide which neighbor is "closest" to recipient
    - Results showed that first steps were the largest

- Kleinberg's Small-World Model
  - Set of points in an n x n grid
  - Distance is the number of "steps" separating points
    - $d(i, j) = |x_i - x_j| + |y_i - y_j|$

Kleinberg‘s Topologies

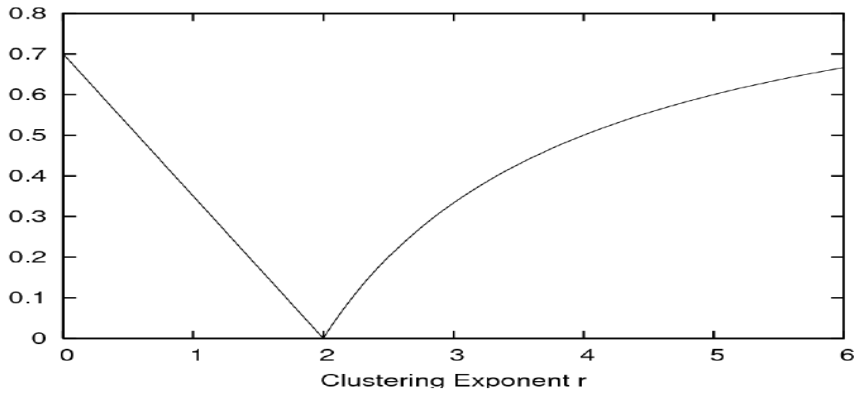
- Take *d*-dimensional grid in which all nodes are connected to all neighbors along each axis
- *Additionally* connect nodes in higher distance with probability decreasing with distance

iow: the probability that node *j* is selected as neighbor for i is proportional to $d(i, j)^{-r}$, with clustering exponent *r*



Intuition of Navigation in Kleinberg's Model

- ***Simple greedy routing***: nodes only know local links and target position, always use the link that brings message closest to target
  - If r=2, expected lookup time is $O(\log^2 n)$
  - If r≠2, expected lookup time is $O(n^{\varepsilon})$, where ε depends on r



- Kleinberg has shown: Number of messages needed is proportional to

  $O(\log^2 n)$ *iff* $r=s$ (s = number of dimensions)
  - Idea behind proof: for any r > s there are too few long edges to make paths short
  - For r < s there are too many random edges ⇒ too many choices for passing message, greedy may not deterministically converge to destination
    - The message will make a (long) random walk through the network

Problems with Small-World Graphs

Small-world graphs explain why:

- Highly clustered graphs can have short average path lengths (“short cuts”)

Small-world graphs do *NOT* explain why:

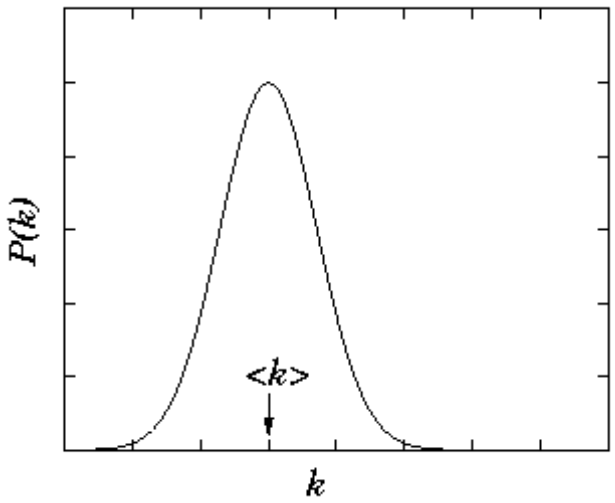
- This property emerges in real networks
  - Real networks are practically never ring-like

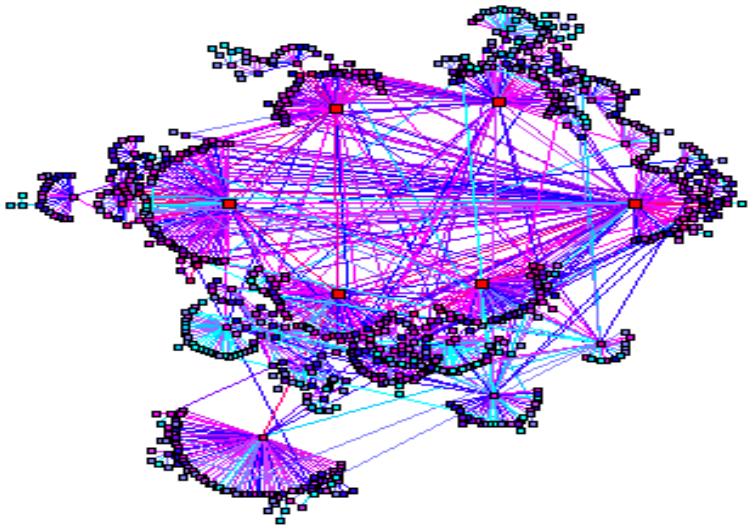
Further problem with small-world graphs:

- Nearly all nodes have same degree
- Not true for random graphs
- What about real networks?

Real World Measurements: World Wide Web

- Links between documents in the World Wide Web
  - 800 Mio. documents investigated (S. Lawrence, 1999)
- What was expected so far?
  - Number of links per web page: $\langle k \rangle \sim 6$
  - Number of pages in the WWW: $N_{WWW} \sim 10^9$

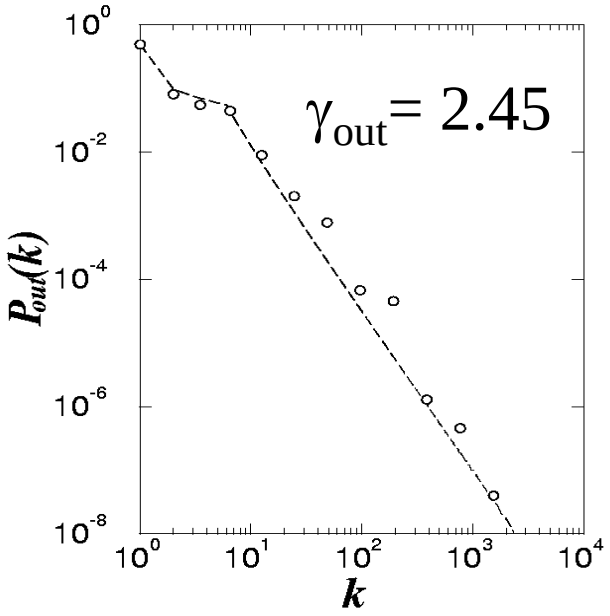




- Probability "page has 500 links": $P(k=500) \sim 10^{-99}$
- Number of pages with 500 links: $N(k=500) \sim 10^{-90}$

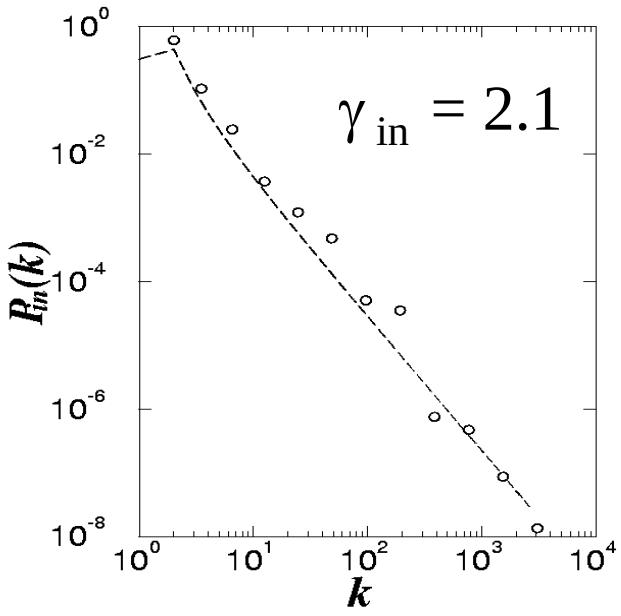
WWW: result of investigation

P(page has k links)



$P_{out}(k) \sim k^{-\gamma_{out}}$

P(k pages link to this page)

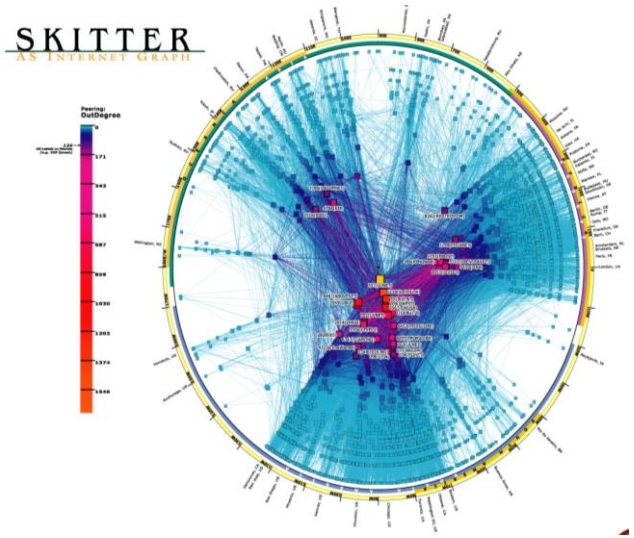


$P_{in}(k) \sim k^{-\gamma_{in}}$

$P(k=500) \sim 10^{-6}$ $N_{WWW} \sim 10^{9}$ → $N(k=500) \sim 10^{3}$

Real-World Measurements: The Internet

- Faloutsos et al. study from 99: Internet topology examined in 1998
  - AS-level topology, during 1998 Internet grew by 45%



- Motivation:
  - What does the Internet look like?
  - Are there any topological properties that don't change over time?
  - How to generate Internet-like graphs for simulations?

- 4 key properties found, each follows a power-law;
- Sort nodes according to their (out)degree
  1. ***Out degree*** *of a node is proportional to its rank to the* ***power of a constant***
  2. *Number of* ***nodes with same out degree*** *is proportional to the out degree to the* ***power of a constant***
  3. ***Eigenvalues*** *of a graph are proportional to the order to the* ***power*** *of a* ***constant***
  4. *Total* ***number of pairs*** *of nodes within a distance d is proportional to d to the* ***power of a constant***

Conclusion: Power Law Networks

- "Power Law" relationships
- For the Internet…
- For Web pages
  - The probability P(k) that a ***page has k*** links (or k other pages link to this page) is ***proportional*** to the number of links ***k to the power of y***

- General "Power Law" Relationships
  - A certain property k is – independent of the growth of the system – always proportional to $\mathbf{k^{-a}}$, where a is a constant (often $2 < a < 4$)

- Power laws very common ("natural")
  - power law networks exhibit small-world-effect (always?)
  - E.g. WWW: 19 degrees of separation
    (R. Albert et al., Nature (99); S. Lawrence et al., Nature (99))

- Also termed: ***scale-free*** networks

Barabasi-Albert-Model

How do power law networks emerge?

- In a network where new vertices (nodes) are added and new nodes tend to connect to well-connected nodes, the vertex connectivities follow a power-law

Barabasi-Albert-Model: power-law network is constructed with two rules

1. Network grows in time
2. New node has preferences to whom it wants to connect

...and here we finally model the generating process!

Preferential connectivity modeled as

- Each new node wants to connect to *m* other nodes
- Probability that an existing node *j* gets one of the *m* connections is proportional to its degree *d(j)*

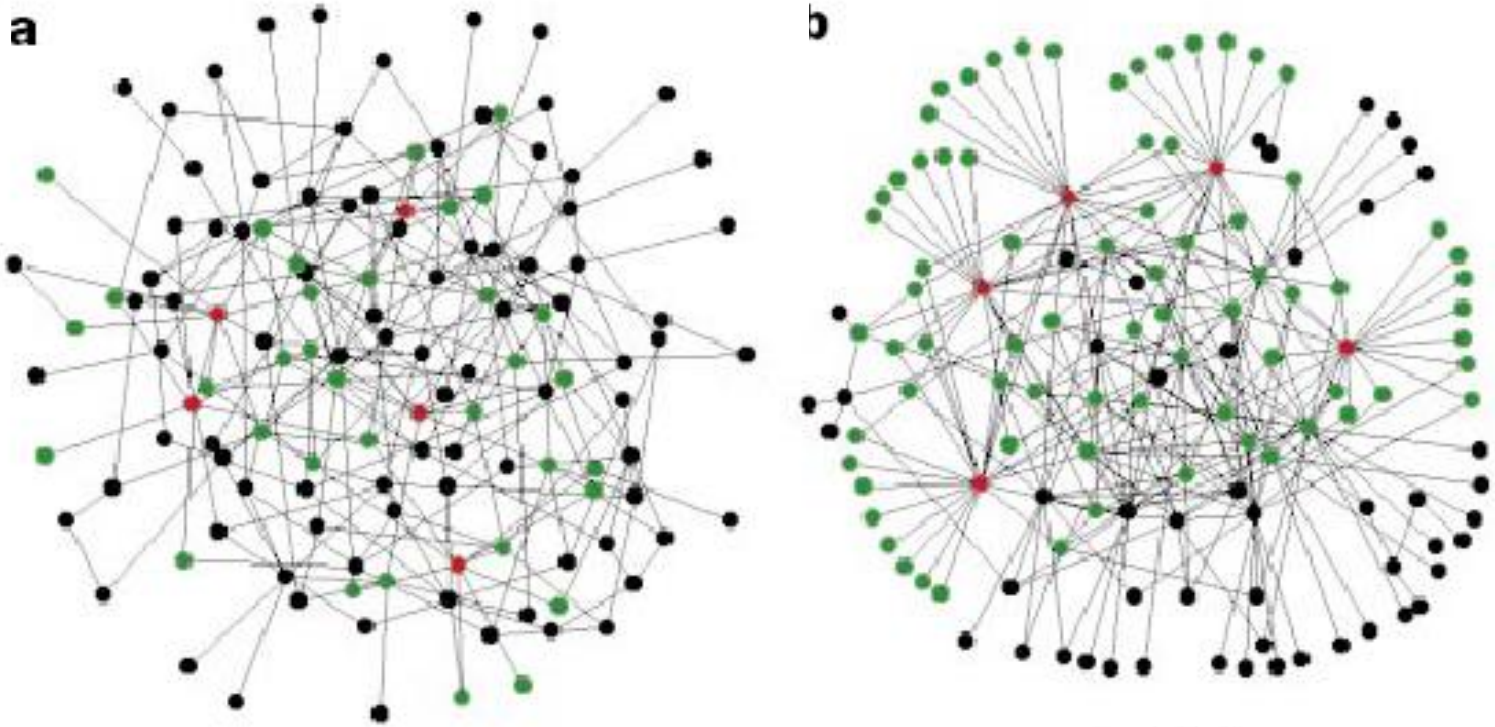
New nodes tend to connect to well-connected nodes

$$\prod(\deg_i) = \frac{\deg_i}{\sum_j \deg_j}$$

Another way of saying this: "the rich get richer"

Resilience of Scale-Free Networks

- Random failures vs. directed attacks



Random Graph **„Power Law“ Graph**

Resilience of Scale Free Networks

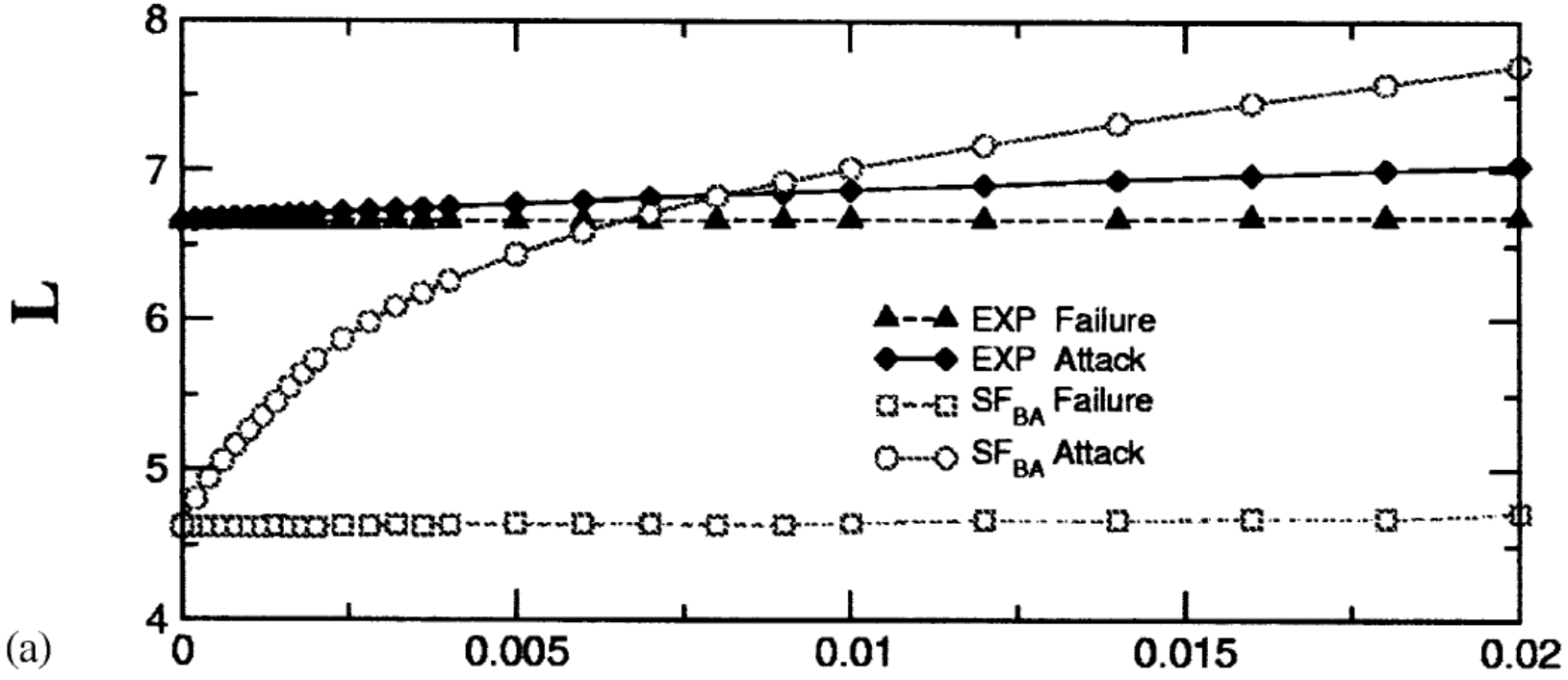
- Experiment: take network of 10000 nodes (random and power-law) and remove nodes randomly

- Random graph:
  - Take out 5% of nodes: Biggest component 9000 nodes
  - Take out 18% of nodes: No biggest component, all components between 1 and 100 nodes
  - Take out 45% of nodes: Only groups of 1 or 2 survive

- Power-law graph:
  - Take out 5% of nodes: Only isolated nodes break off
  - Take out 18% of nodes: Biggest component 8000 nodes
  - Take out 45% of nodes: Large cluster persists, fragments small

- Networks with power law exponent < 3 are very robust against random node failures
  - ONLY true for random failures!

The consequence…



L: „Average Connected Distance“

Summary of Graph Analyses…

- The network structure of a networks influences:
  - average necessary number of hops (path length)
  - possibility of greedy, decentralized routing algorithms
  - stability against random failures
  - sensitivity against attacks
  - redundancy of routing table entries (edges)
  - many other properties of the system build onto this network
- Important measures of a network structure are:
  - average path length
  - the degree distribution
  - clustering coefficient
  - Various resilience metrics (with differing foci)