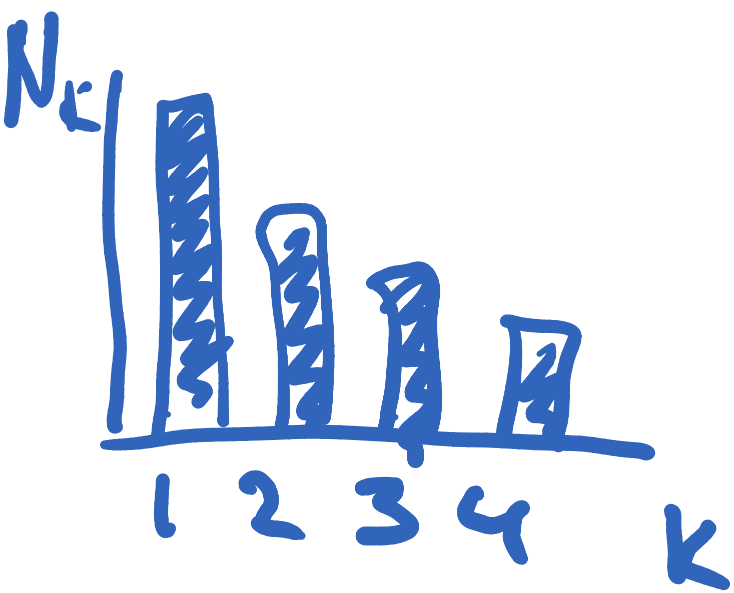
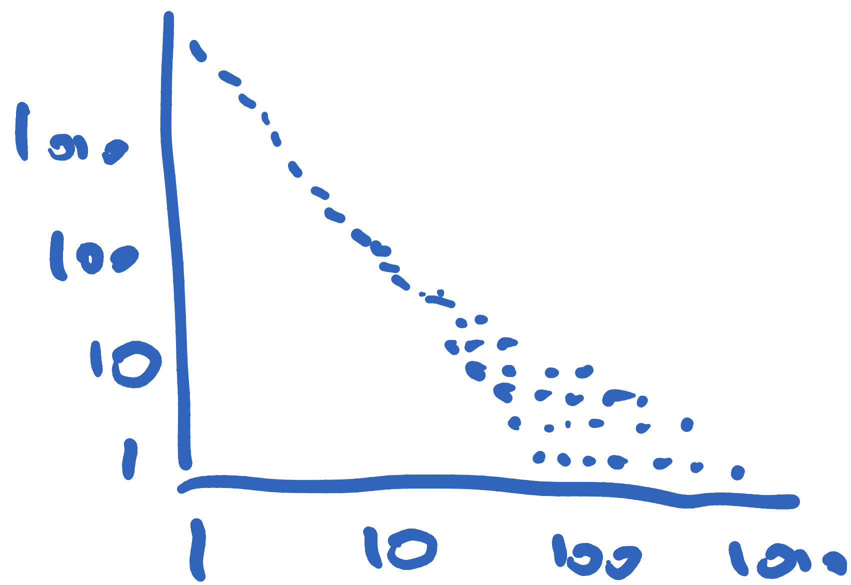
Recap

- Type of graphs
- Connectivity/Giant component
- Diameter
- Clustering coefficient
- Betweenness Centrality
- **Degree distributions**

Degree Distribution

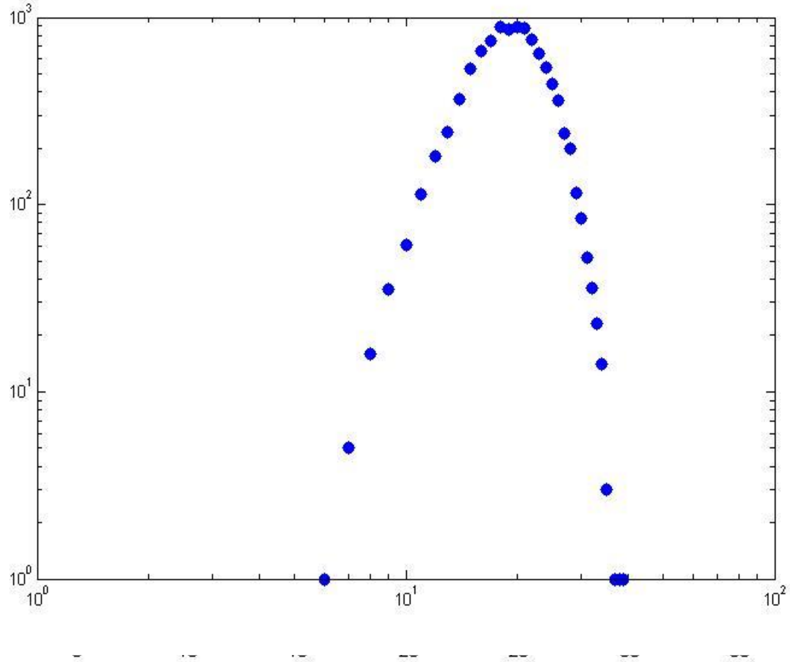
- N_k is the number of nodes with degree *k*
- *P(k)* is the probability that a randomly chosen node has degree *k*.
  - $P(k) = N_k / N$
  - Often power-law distributions (linear in loglog scale)



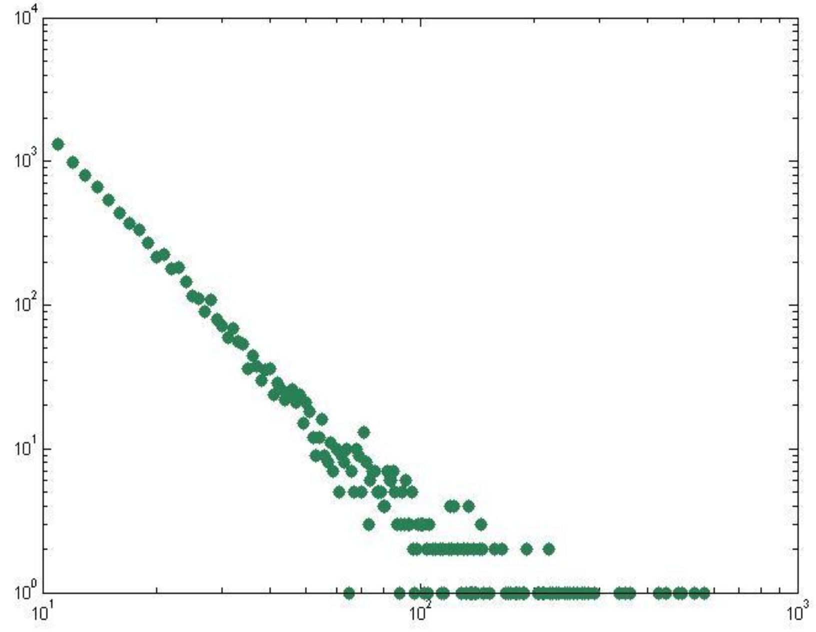


More degree distributions

- Normal vs. power-law distributions
- N=10k nodes, avg d=20,



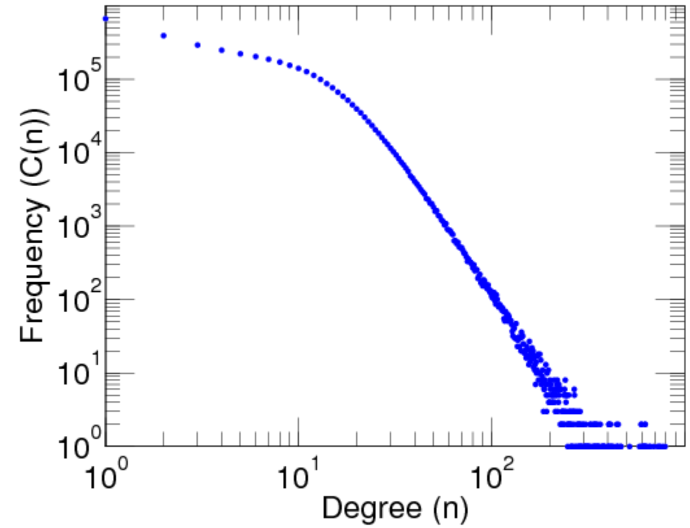
Random Graph

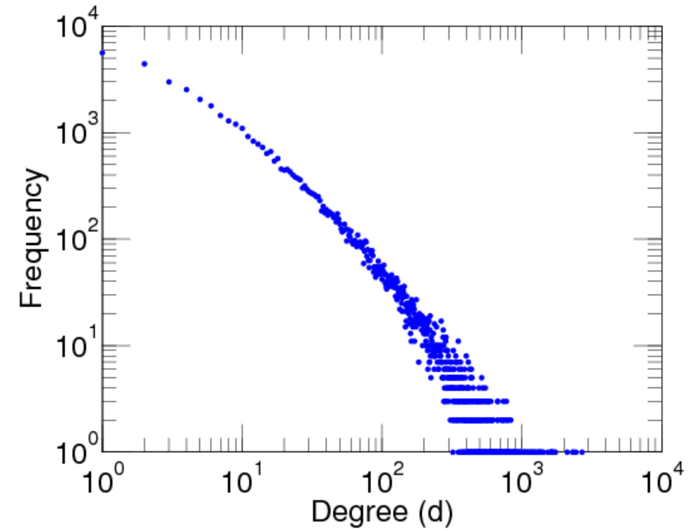


Preferential attachment (power law)

Some real world examples (from http://konect.uni-koblenz.de)

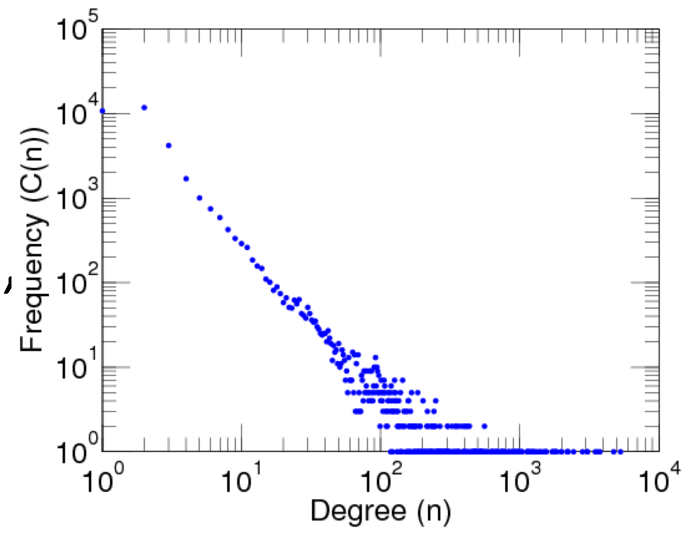
- US patents
  - Patent-patent citation
  - N=3774768
  - E=16522438
  - Effective diameter 9,79
- Facebook (user-user wall posts)
  - Directed
  - N=63891, E=876993
  - CC=19,1% Effective diameter=7,25

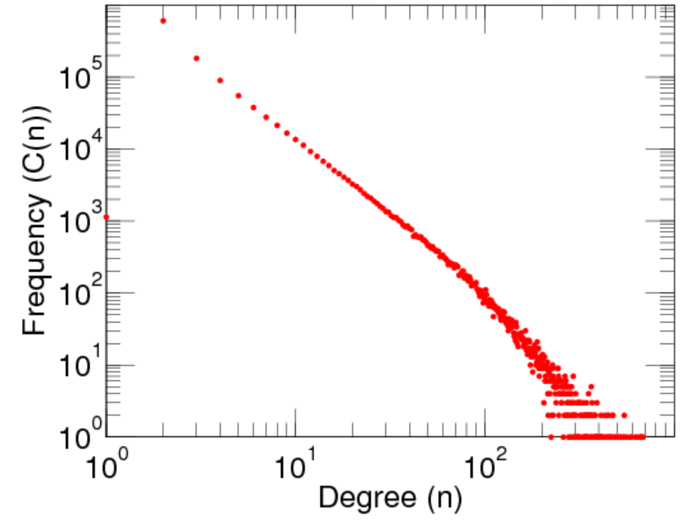




Some real world examples (cont.)

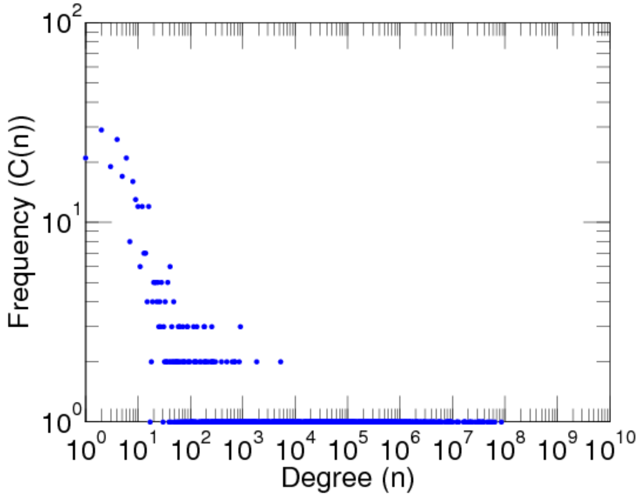
- Internet Topology
  - AS-AS connection
  - N=34761, E=171403, CC=4,851%,
  - Mean shortest path 3,77
- DBLP
  - Author-publication authorship
  - Bipartate
  - N=4337293
  - E=6651968
  - Effective diameter 15,3



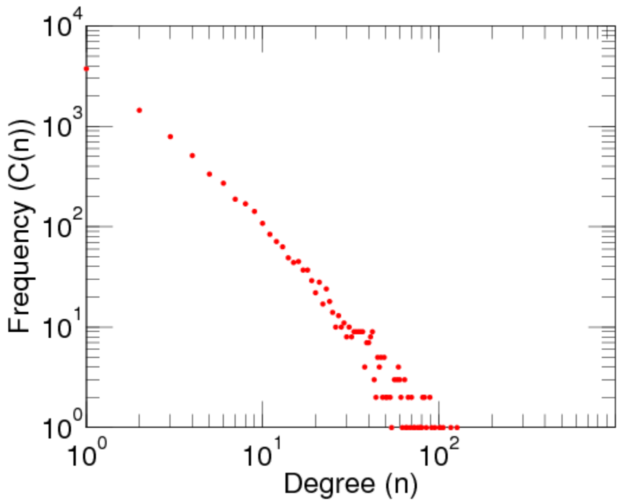


Some real world examples (cont.)

- USA airports (airport-airport flights)
  - N=1574, E=28236, CC=38,4%, dim=8



- Sexual escorts (Buyer–escort contact)
  - bipartate graph
  - N=16730, E=50632, dim=6,05



- **What do we see?**
  - ***Sparse*** networks, with ***low diameter***, ***hubs*** and ***non-random clustering***.

How do we Approach Networks

- **Observations**
  - Structure, properties, patterns, evolution of the networks
  - Our empirical observations often find:
    - Sparse networks
    - Small diameter
    - Large clustering coefficients
    - Power law degree distributions
    - One giant component
  - How to explain them?
- **Models**
  - How do we model edge attachments, epidemics, communities etc?
  - We will start from “simple” to more “sophisticated”.
- **Algorithms (applications)**
  - How to route and search, how to find clusters, how to do efficient information collection/dissemination, gossip learning etc.

Models of Graphs

- **G(n,m) model**
  - Start with n isolated vertices
  - Place m edges among them at random.
  - G(n,m) defines a family of graphs (not a particular graph)
- **G(n, p) model (Erdos-Renyi random graph)**
  - Start with n isolated vertices
  - We place and edge between each distinct vertice pair with probability p.
  - n and p do not uniquely determine the graph!
- **Q1: What's a degree of the above networks?**
- **Q2: Which family is bigger?**

Erdos-Renyi random graph

- What can you say about the graph when we move p from 0 to 1?
  - Diameter?
  - How big is a giant component when p=0 and when p=1?
  - How does the giant component grow inbetween those p values?
    - Network undergoes "***phase transition***"

Erdös-Renyi random graph (cont.)

as N becomes large:

If p < 1/N, probability of a giant component goes to 0

If p > 1/N, probability of a giant component goes to 1, and all other components will have size at most log(N)

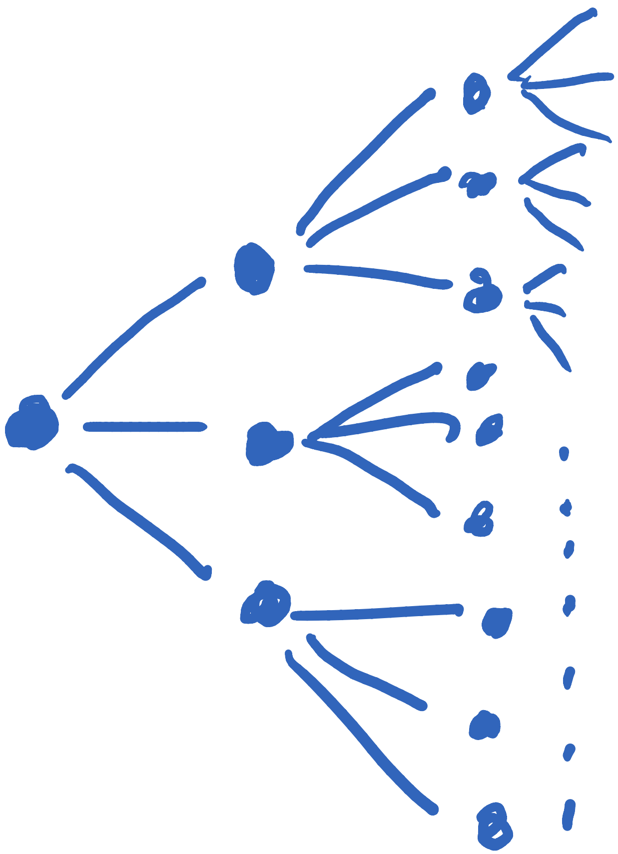
- at p ~ 1/N, average degree is ~ 1 (sparse graph)
  - For regular Random graphs when d is >=3 the graph is connected a.a.s.
- Any monotone property exhibits **Threshold phenomena** in Erdos-Renyi with respect to p.

– E.g., network has a cycle of at least K vertices.

- Diameter: approx log(N)/log(d)
- Clustering coefficient: ?

http://ccl.northwestern.edu/netlogo/models/GiantComponent

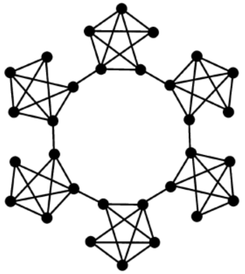
Intuition on Diameter Calculation



- Take a tree with degree d
- Nodes reached in
  - 1^{st} step 1+d
  - 2^{nd} step $1+d+d^2$
  - 3^{rd} step $1+d+d^2+d^3$
  - k^{th} step $1+d+d^2+d^3+\ldots+d^k \sim d^k$
- When do we reach N nodes?
  - $N=d^k$
  - $k=\log_d N$
- Intuition for Random graphs
  - Most of the degrees in random graph are between 3d and d/3 (chernoff bounds), so log(3d) almost equal to log(d/3) and equal to log(d)
  - Very few neighbors are neighbors themselves, so we do not hit many “covered” nodes.
    - Most of the nodes are hit on the last step.

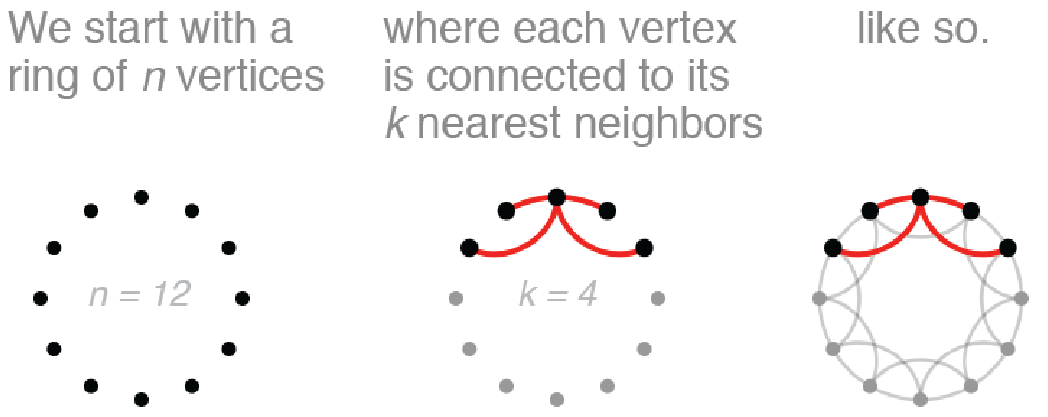
Random graphs and real world

- **Erdos-Renyi can explain small diameter, connectivity under low density**
- Does not explain high clustering
  - No bias picking friends of a friend
- Does not explain power-law degree distributions
- We need a model that can explain biases towards selecting friend-of-a-friend
  - Any ideas?
    - E.g., caveman world



Watts-Strogatz Model

- Regular graph with degree k connected to nearest neighbors



- Can be also a grid, torus, or any other "geographical" structure which has high clusterisation and high path length
- ***With probability p rewire each edge in the network to a random node.***
  - ***Q: What happens when p=1?***