PhD course on Network Science

Giovanni Neglia

Inria – EPI Maestro

University of Pisa

23-27 March 2015

Institut National de Recherche in Informatique et Automatique

- Général De Gaulle in 67

8 research centers in France

4200 persons including

- 1300 permanent researchers
- 1200 PhD students
- 250 postdocs and R&D engineers

900 non-French researchers from 78 countries

General information

- Website
  - www-sop.inria.fr/members/Giovanni.Neglia/complexnetworks15pisa/
- Who are the audience?
- What can you expect from me?
- At home
- Evaluation
- For any question: giovanni.neglia@inria.fr
- First homework
  - send me an email with your name

PhD course on Network Science

Module 1:
Introduction to Network Science

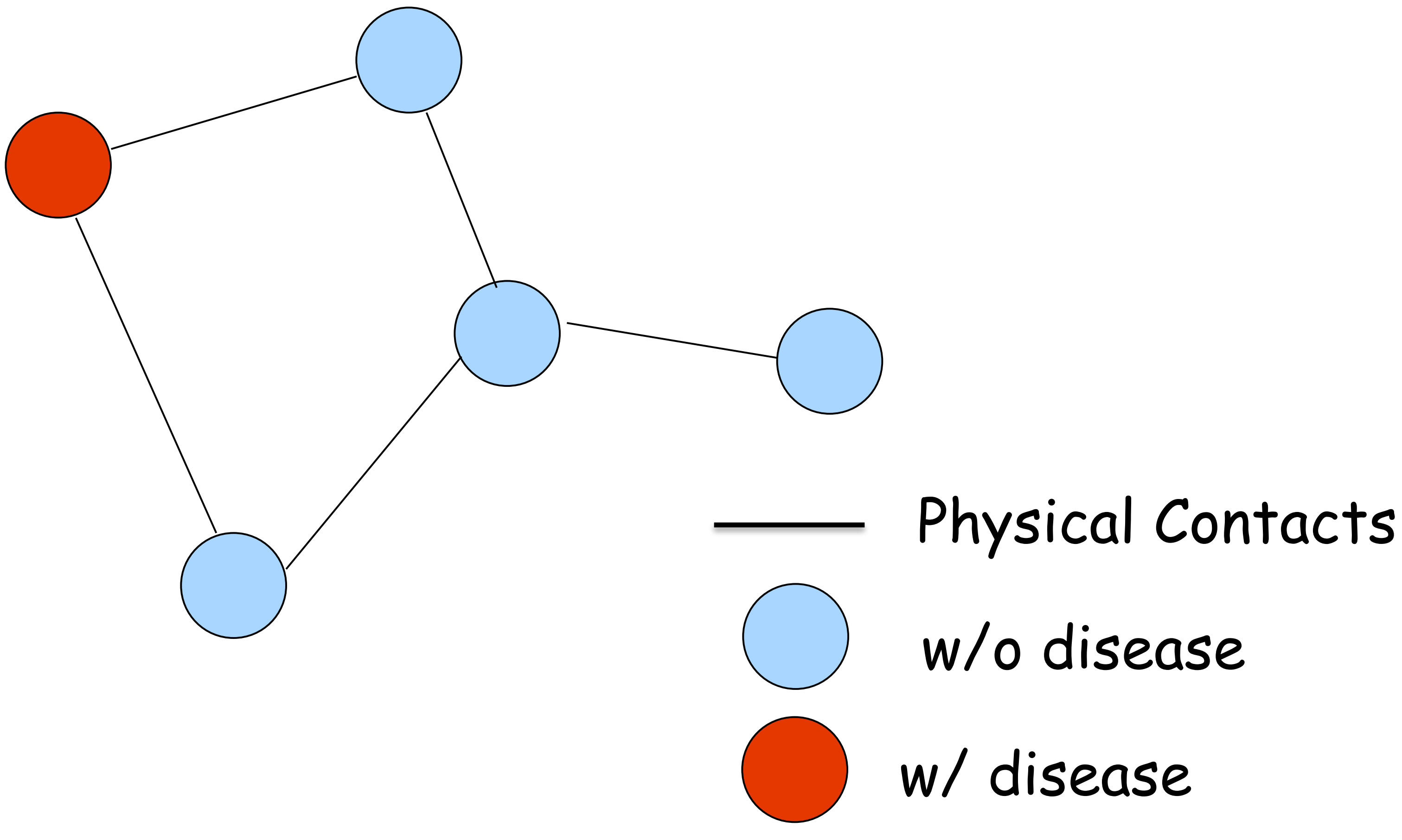
Which network?

Which network?

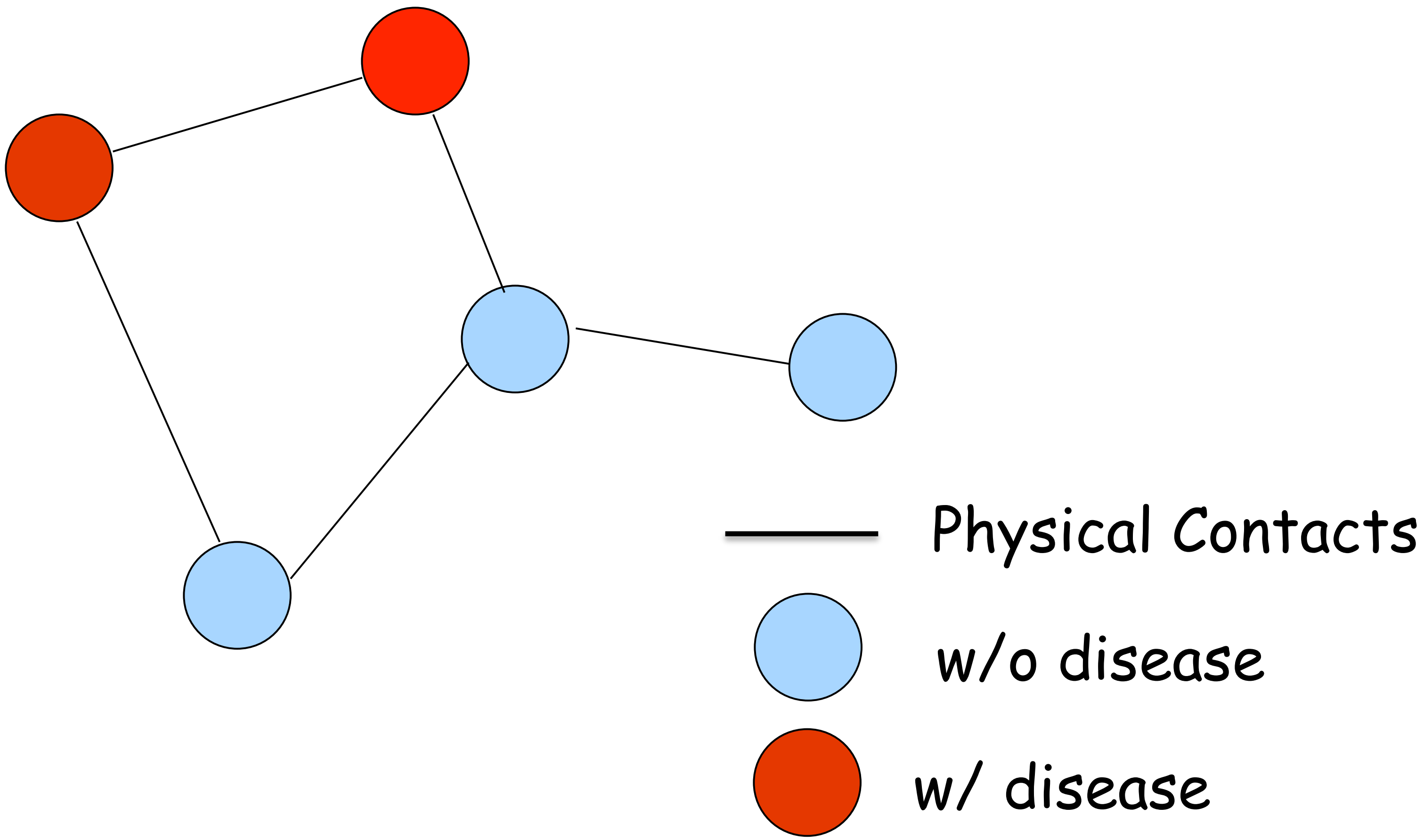
Network Science

1. Common properties to many existing networks
   - Social nets, transportation nets, electrical power grids, Internet AS net, P2P nets, gene regulatory net,
   - These are the "complex networks" that exhibit "non-trivial topological features—features that do not occur in simple networks such as lattices or random graphs but often occur in real graphs" [confusing wikipedia's definition]
2. Important dynamic processes on these networks show the same properties

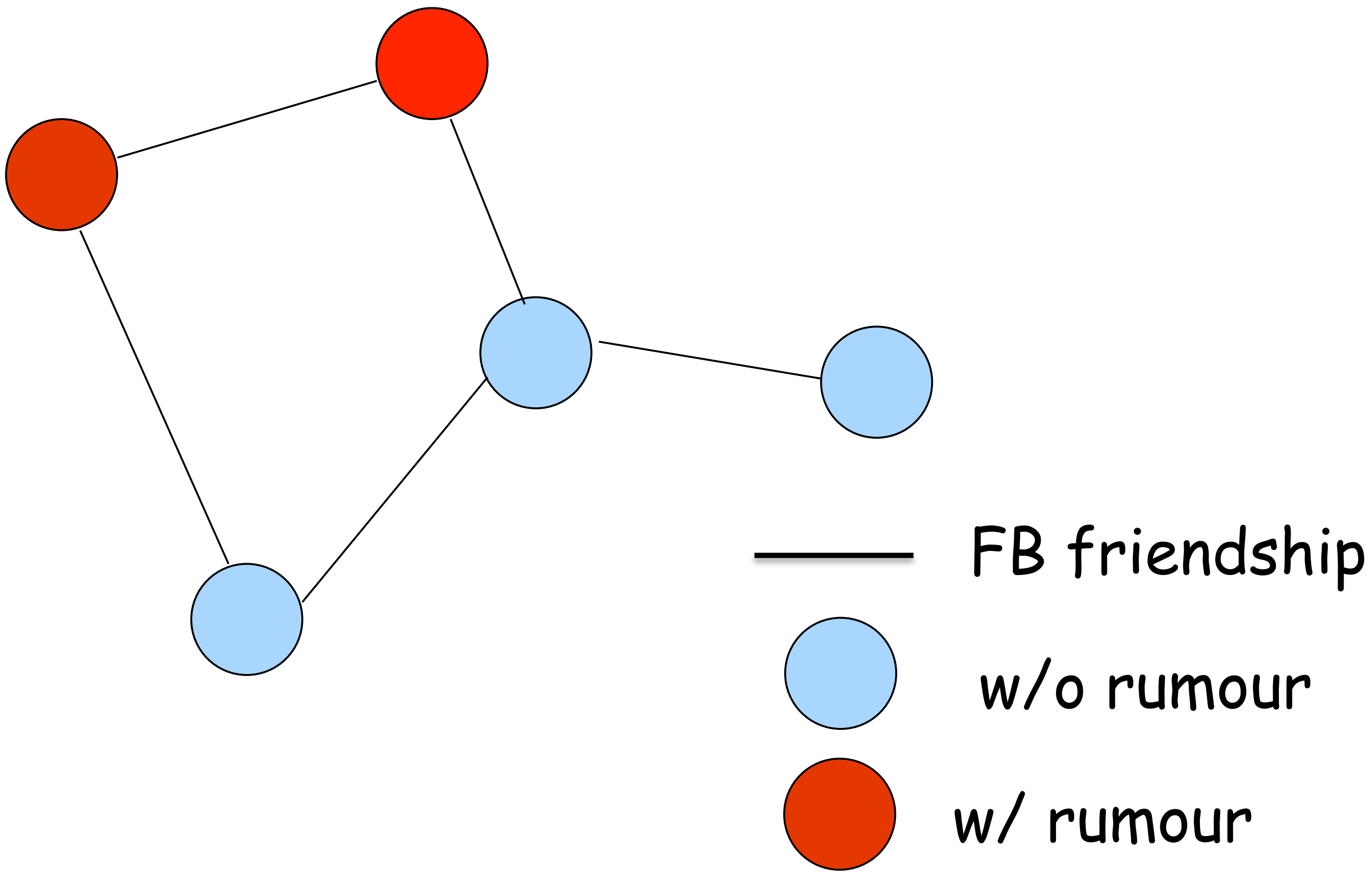
Contagion



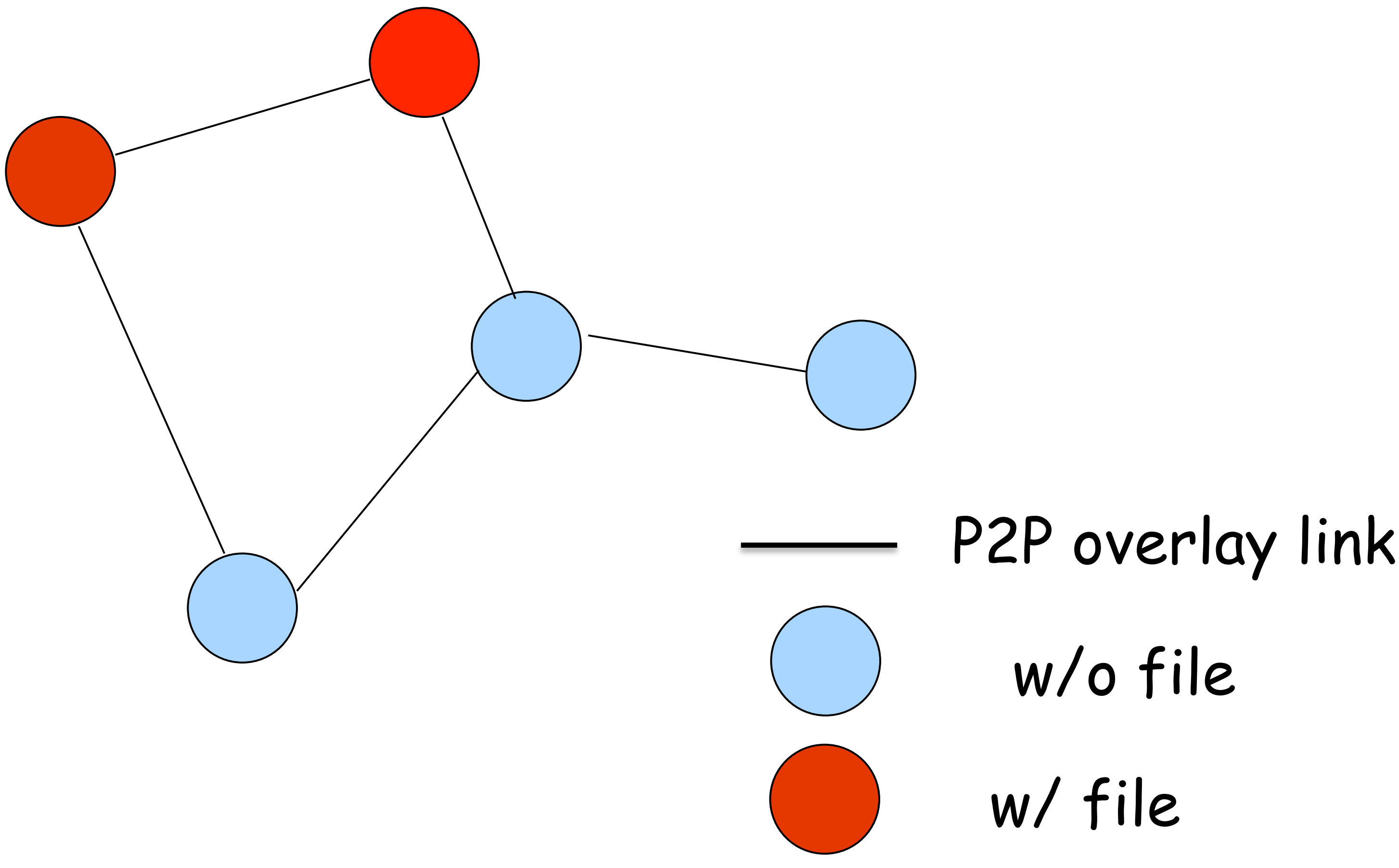
Contagion



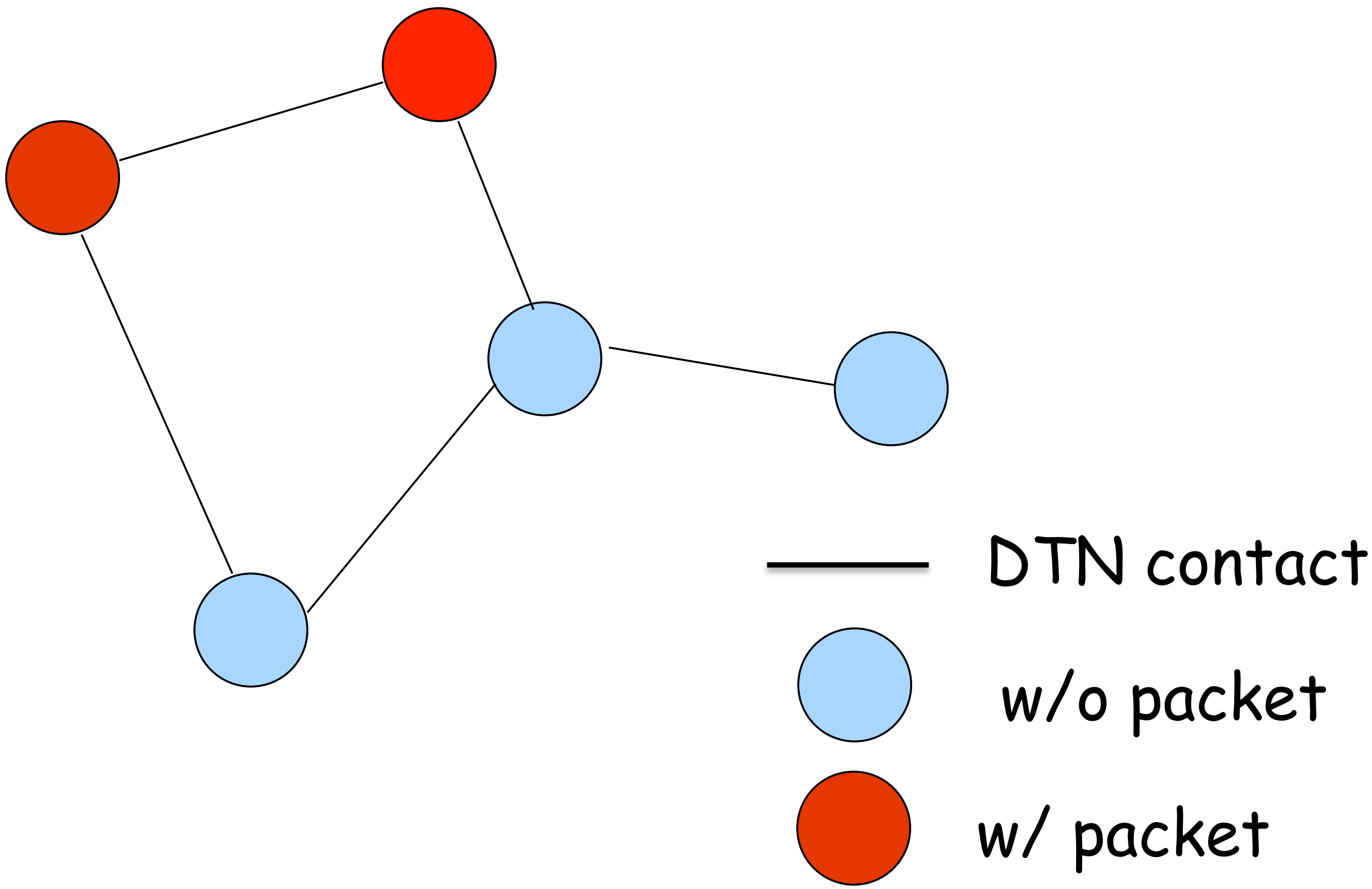
Contagion



Contagion



Contagion



Take Home Lesson

If we understand how topological properties influence contagion

- We can speed-up or slow-down contagion
- We can use these lessons to engineer new protocols (overlay topologies, replication mechanisms,…)

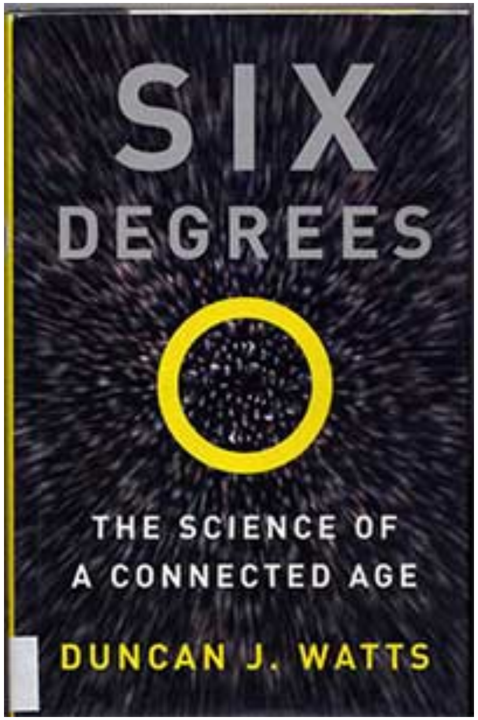
Outline

- Properties of Complex Networks (high-level view)
  - Small diameter
  - High Clustering
  - Hubs and heavy tails
- Physical causes
- What is Network Science?
  - Is it really a new science? Different from graph theory?

Milgram's experiment (1967)

6 degrees of separation

Six degrees of separation is the idea that everyone is on average approximately six steps away, by way of introduction, from any other person in the world, so that a chain of "a friend of a friend" statements can be made, on average, to connect any two people in six steps or fewer.



2003

Small Diameter, more formally

- A linear network has diameter N-1 and average distance $\Theta(N)$
  - How to calculate it?
- A square grid has diameter and average distance $\Theta(\mathrm{sqrt}(N))$
- Small Diameter: diameter $O((\log(N))^a)$, $a>0$
- Lessons from model: a few long distance random connections are enough

Erdös-Rényi graph

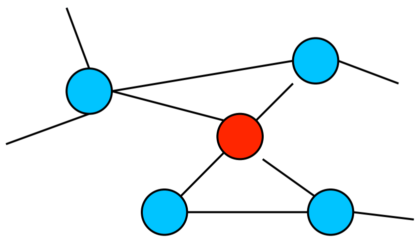
- A ER graph $G(N,q)$ is a stochastic process
  - N nodes and edges are selected with prob. q
- Purpose: abstract from the details of a given graph and reach conclusions depending on its average features

Erdös-Rényi graph

- A ER graph $G(N,q)$ is a stochastic process
  - N nodes and edges are selected with prob. q
  - Degree distribution: $P(d)=C^d_{N-1}\, q^d(1-q)^{N-1-d}$
    - Average degree: $\langle d\rangle=q\,(N-1)$
    - For $N\to\infty$ and Nq constant: $P(d)=e^{-\langle d\rangle}\langle d\rangle^d/d!$
      - $\langle d^2\rangle=\langle d\rangle(1+\langle d\rangle)$
  - Average distance: $\langle l\rangle\approx \log N/\log\langle d\rangle$
    - Small diameter

Clustering

- "The friends of my friends are my friends"
- Local clustering coefficient of node i
  - (# of closed triplets with i at the center) / (# of triplets with node i at the center) = (links among i's neighbors of node i)/(potential links among i's neighbors)



$C_i=2/(4*3/2)=1/3$

- Global clustering coefficient
  - (total # of closed triplets)/(total # of triplets)
    - # of closed triplets = 3 # of triangles
  - Or $1/N \Sigma_i C_i$

Clustering

- In ER
  - $C \approx q \approx \langle d \rangle / N$

Clustering

- In real networks

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-author[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible], 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	[illegible]	2.12	[illegible]	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co[illegible]hip	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]01b, 2001c	7
Math. co-authorship	70 975	[illegible]	9.5	8.2	0.59	[illegible]$.4\times10^{-5}$	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Good matching for avg distance,
Bad matching for clustering coefficient

How to model real networks?

"Geometric" Graphs have a high clustering coefficient

but also a high diameter

Random Graphs have a low diameter

but a low clustering coefficient

--> Combine both to model real networks: the Watts and Strogatz model

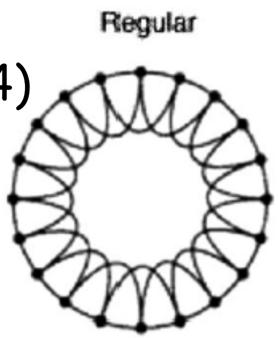
Regular Graph (k=4)

Long paths

$L = n/(2k)$

Highly clustered

$C=3/4$



Random Graph (k=4)

Short path length

$L=\log_k N$

Almost no clustering

$C=k/n$

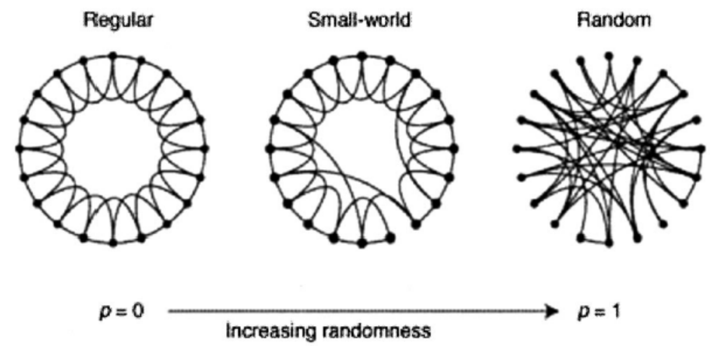
Regular ring lattice

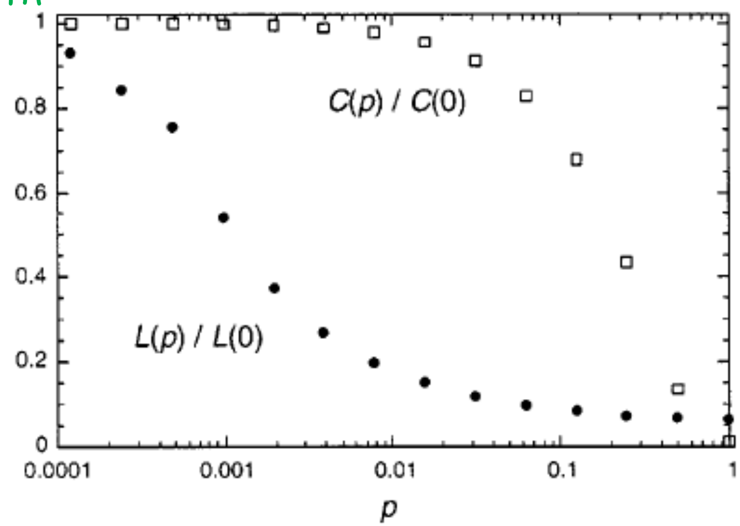
R. Albert and A.-L. Barabasi: Statistical mechanics of complex networks, Rev. Mod. Phys., Vol. 74, No. 1, January 2002

Watts and Strogatz model

Random rewiring of regular graph

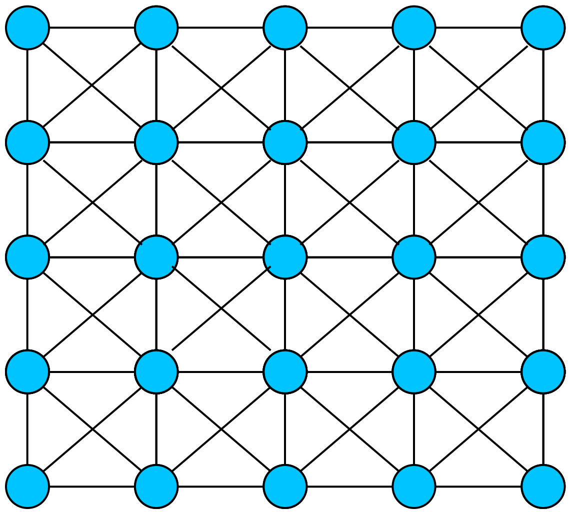
- With probability *p* rewire each link in a regular graph to a randomly selected node
- Resulting graph has properties both of regular and random graphs
  - --> High clustering and short path length



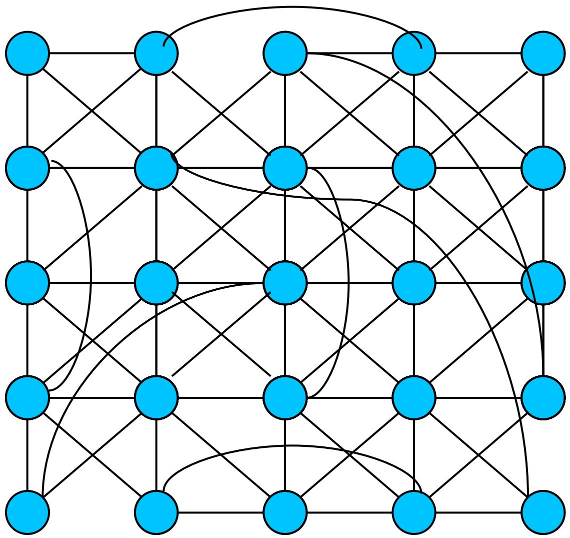


R. Albert and A.-L. Barabasi: Statistical mechanics of complex networks, Rev. Mod. Phys., Vol. 74, No. 1, January 2002

The 2D case



The 2D case



Small World

- to denote
  1. Small diameter
  2. **Small diameter + high clustering**
  3. Small diameter + navigability

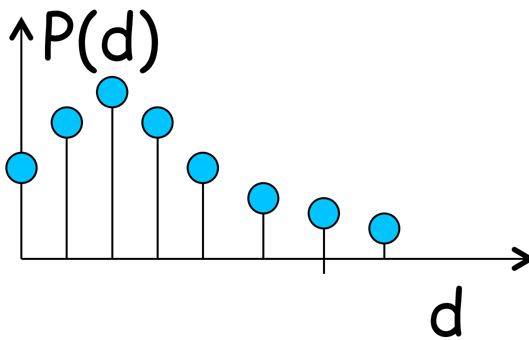
- Cause
  - Nodes are embedded in some multidimensional space (e.g. geography, jobs, hobbies)
  - There are some random far-away links

Intermezzo: navigation

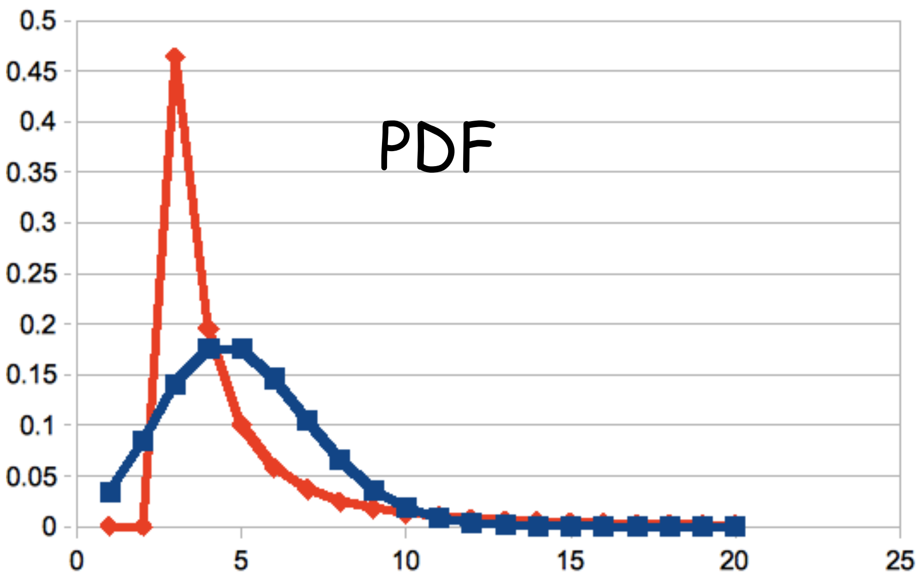
- In Small world nets there are short paths $O((\log(N))^a)$
- But can we find them?
  - Milgram's experiment suggests nodes can find them using only local information
  - Standard routing algorithms require O(N) information!
  - The answer will arrive in a later module

Hubs

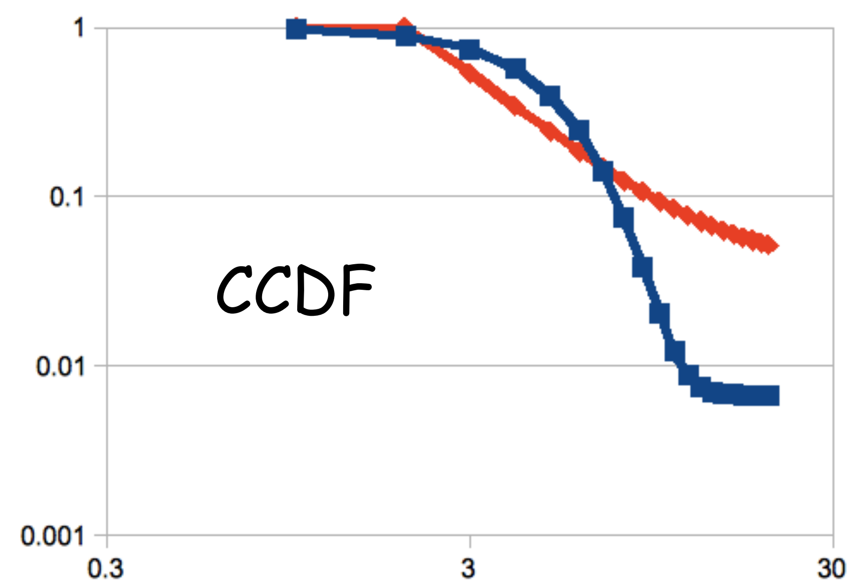
- 80/20 rule
  - few nodes with degree much higher than the average
  - a lot of nodes with degree smaller than the average
  - (imagine Bill Clinton enters this room, how representative is the avg income)
- ER with $N=1000$, $\langle d\rangle=5$, $P(d) \approx e^{-\langle d\rangle}\langle d\rangle^d/d!$
  - #nodes with $d=10$: $N*P(10) \approx 18$
  - #nodes with $d=20$: $N*P(20) \approx 2.6\ 10^{-4}$



Hubs

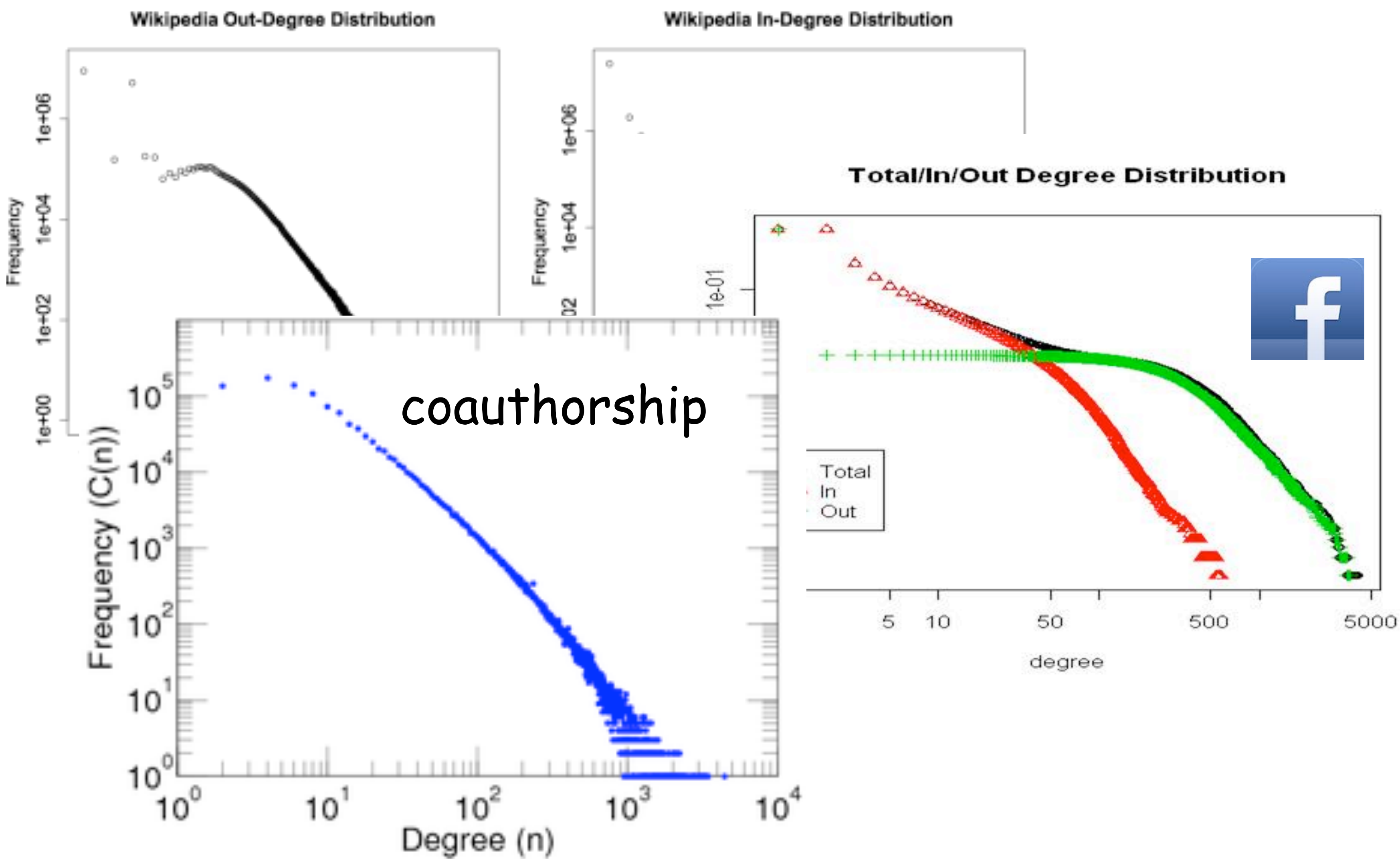


ER
Power law

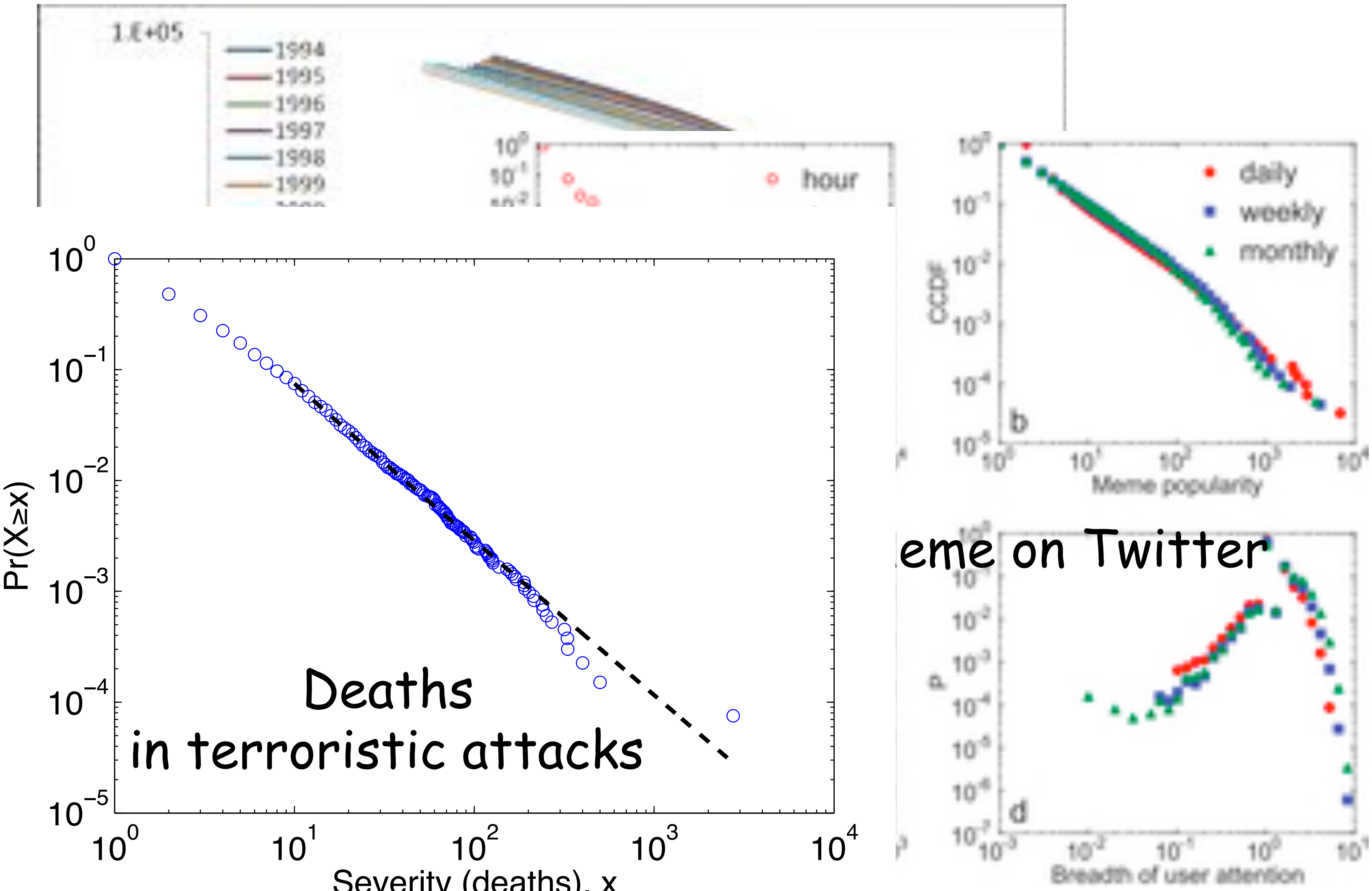


Power law:
$P(d) \sim d^{-\alpha}$

Power law degree distributions



… and more



1.E+05
1994
1995
1996
1997
1998
1999
hour
daily
weekly
monthly
CCDF
Meme popularity
b
eme on Twitter
P
d
Breadth of user attention
Pr(X≥x)
Severity (deaths), x
Deaths
in terroristic attacks

Power Law

- Where does it come from?
  - Albert-Barabasi's growth model
  - Highly Optimized Model
  - And other models
    - See Michael Mitzenmacher, A Brief History of Generative Models for Power Law and Lognormal Distributions

Albert-Barabasi's model

- Two elements
  - Growth
    - m_0 initial nodes, every time unit we add a new node with m links to existing nodes
  - Preferential attachment
    - The new node links to a node with degree k_i with probability

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1,N} k_j}$$

The rich becomes richer

  - It generates power-law

What is Network Science?

- A *natural science*
  - The focus is on existing networks (not graphs in general)
  - Understand observed phenomena
- An interdisciplinary approach, it draws on many different theories and methods
  - graph theory from mathematics, statistical mechanics from physics, data mining and information visualization from computer science, inferential modeling from statistics, social structure from sociology…

What after?

We will

- study more formally these properties
- evaluate the effect of such properties on a specific dynamic process (infection)
- learn about software tools we can use to study complex networks
- see what a specific complex network (Twitter) looks like

What after?

We will

- study more formally these properties
- evaluate the effect of such properties on a specific dynamic process (infection)
- learn about software tools we can use to study complex networks
- see what a specific complex network (Twitter) looks like

…***if we have enough time!***

2^{nd} Homework

- Watch “How Kevin Bacon Cured Cancer”
  - http://tiny.cc/kevinbacon
- Send me an email with 2 questions rising from the movie you would like the course to address