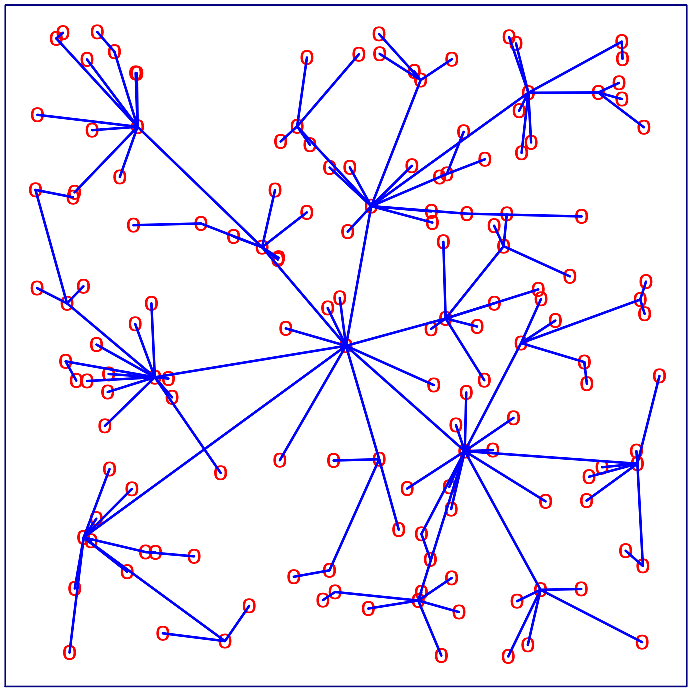
# ECS 253 / MAE 253, Lecture 2

April 5, 2023



"Power laws, Random graphs, phase transitions"

# Class structure

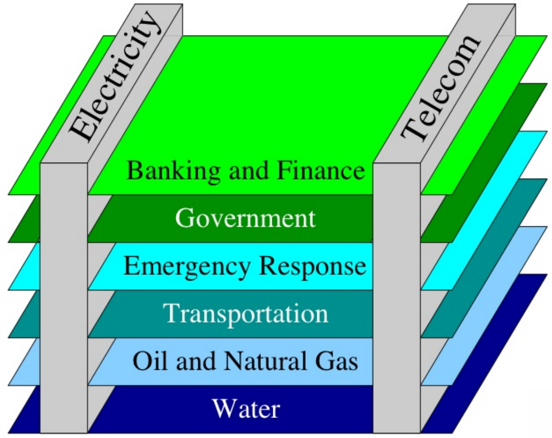
- Two tracks to the class:

  Track A: Project
  (1) Common homeworks (e.g. HW1.pdf, HW2.pdf) and
  (2) HW1a.pdf, HW2a.pdf etc.

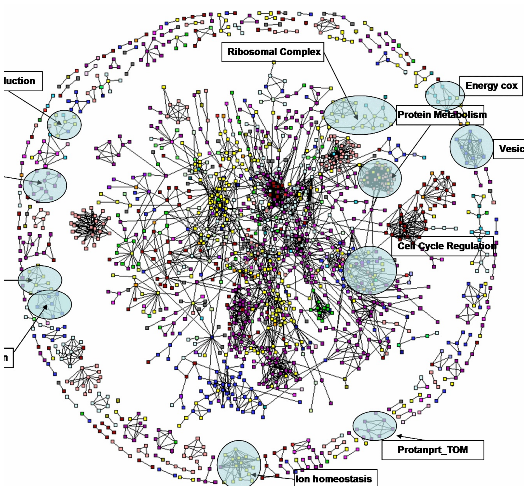
  Track B: Advanced HWs
  (1) Common homeworks (e.g. HW1.pdf, HW2.pdf) and
  (2) HW1b.pdf, HW2b.pdf etc.

- Track A: Project
  - Teams of 5-6 people ideal
  - Negative results are OK
  - Ideally aim to have a result for a journal or conference

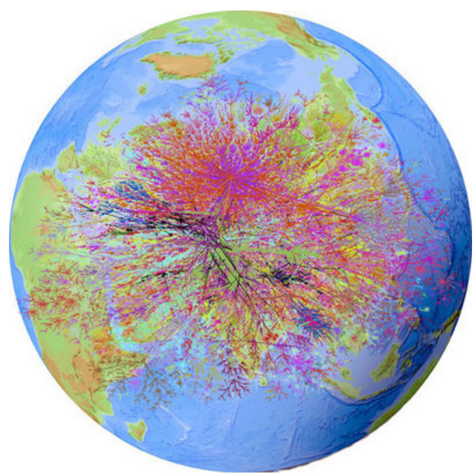
# Complex networks are ubiquitous:



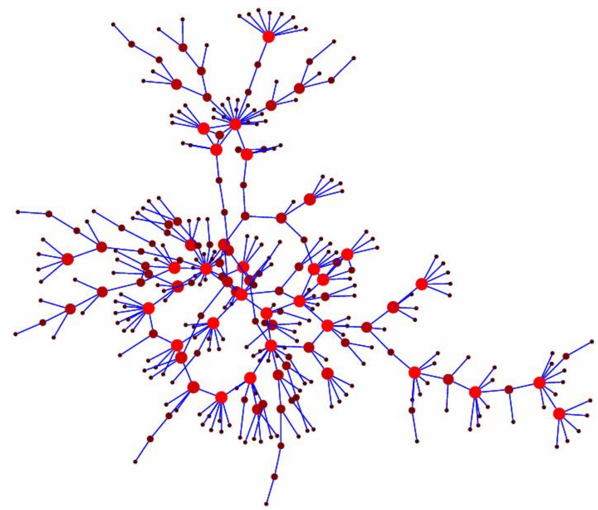

*Critical Infrastructure*




*Biological & Ecological networks*



*Information and Communication technology*



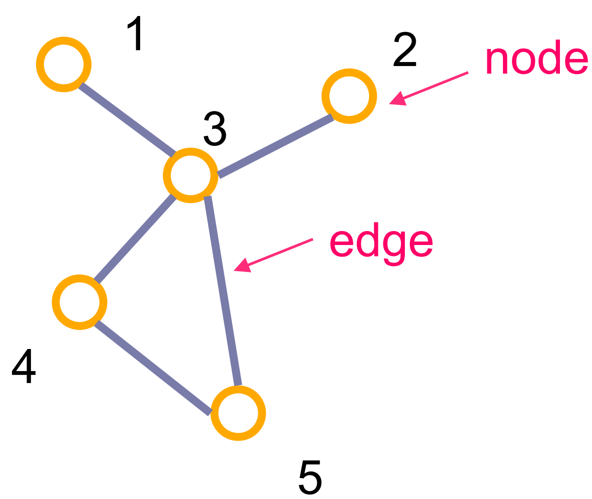
*Social networks: Economics & Epidemics*

## Networks: Physical, Biological, Social, Technological

- **Geometric** versus **virtual** (Internet versus WWW).
- **Natural** /spontaneously arising versus **engineered** /built.
- **Directed** versus **undirected** edges.
- Each network may **optimize** something unique.
- Identifying **similarities** and fundamental **differences** can guide future design/understanding.
- Interplay of **topology** and **function** ?
- Unifying features: – **Broad heterogeneity in node degree.** – **Small Worlds** (Diameter $\sim \log(N)$).

# What are networks?

- Networks are collections of points joined by lines.




"Network" ≡ "Graph"

| points | lines | |
|---|---|---|
| vertices | edges, arcs | math |
| nodes | links | computer science |
| sites | bonds | physics |
| actors | ties, relations | sociology |

*Slide from Adamic's course*

# Subtle details of edges

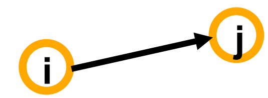
## Network elements: edges

- Directed (also called arcs)
  - A -> B ($E_{BA}$)
    - A likes B, A gave a gift to B, A is B's child
- Undirected
  - A <-> B or A – B
    - A and B like each other
    - A and B are siblings
    - A and B are co-authors
- Edge attributes
  - weight (e.g. frequency of communication)
  - ranking (best friend, second best friend…)
  - type (friend, relative, co-worker)
  - properties depending on the structure of the rest of the graph: e.g. betweenness
- Multiedge: multiple edges between two pair of nodes
- Self-edge: from a node to itself

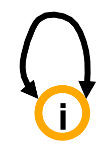
*Slide from Adamic's course*

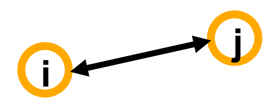
# Adjacency matrices

- Representing edges (who is adjacent to whom) as a matrix
  - $A_{ij}$ = 1 if node i has an edge to node j
    = 0 if node i does not have an edge to j



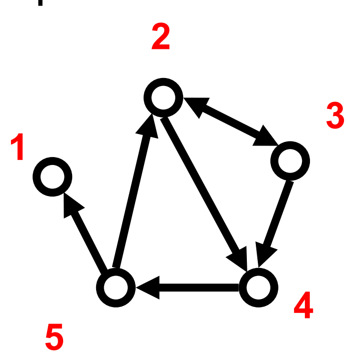

  - $A_{ii}$ = 0 unless the network has self-loops
    - If self-loop, $A_{ii}$=1




  - $A_{ij} = A_{ji}$ if the network is undirected,
    or if i and j share a reciprocated edge

Example:

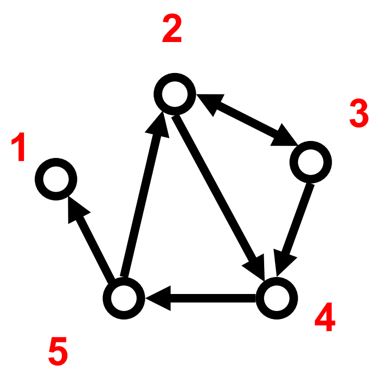



$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

*Slide from Gunes course, UNR*

# Adjacency lists

- Edge list
  - 2 3
  - 2 4
  - 3 2
  - 3 4
  - 4 5
  - 5 2
  - 5 1




- Adjacency list
  - is easier to work with if network is
    - large
    - sparse
  - quickly retrieve all neighbors for a node
    - 1:
    - 2: 3 4
    - 3: 2 4
    - 4: 5
    - 5: 1 2

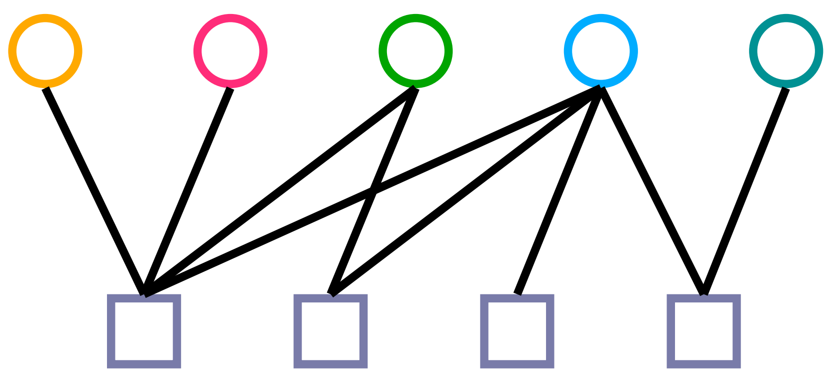
*Slide from Gunes course, UNR*

# Beyond simple networks:

## Bipartite (two-mode) networks

- edges occur only between two groups of nodes, not within those groups

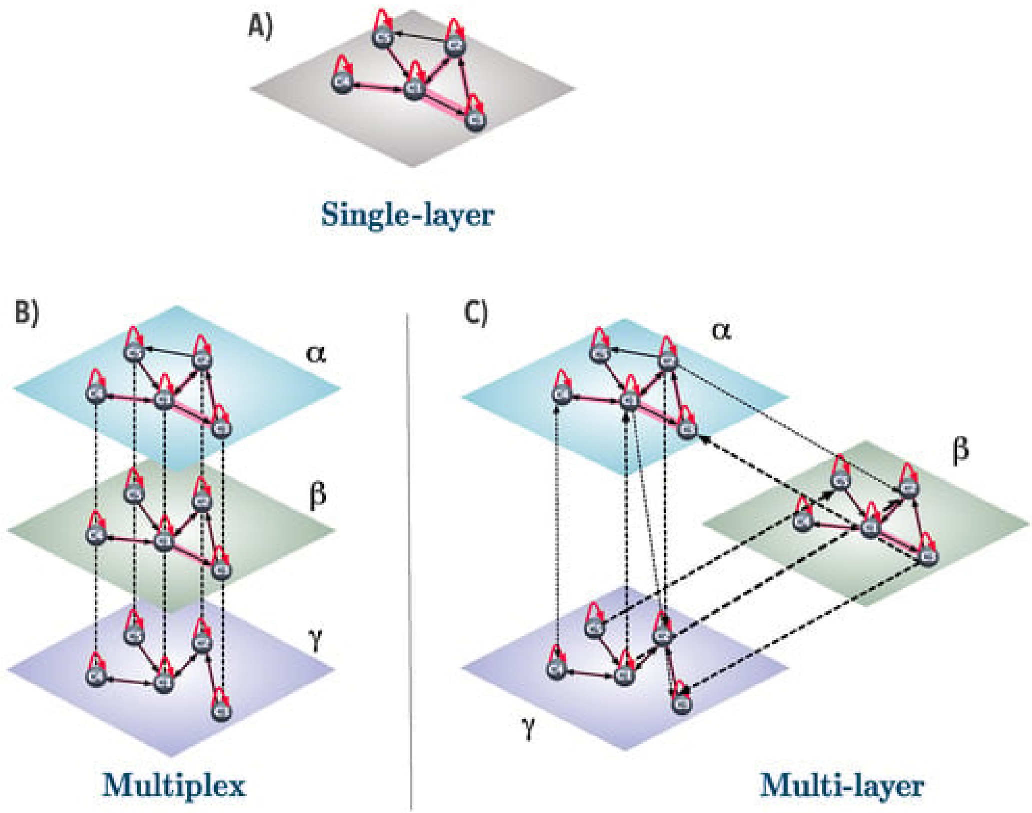
- for example, we may have individuals and *events*
  - directors and boards of directors
  - customers and the items they purchase
  - metabolites and the reactions they participate in



*Slide from Gunes course, UNR*

# Beyond simple networks: Multiplex and multi-layered

*Image from Alves et al,* **Entropy, 2018**

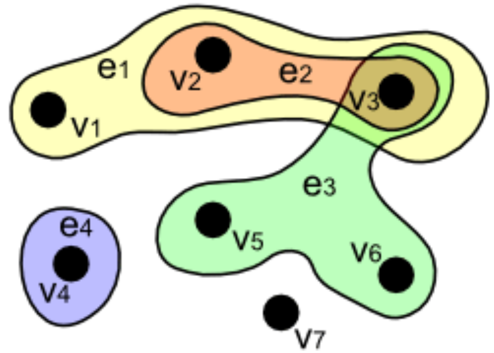



- B) **Multiplex**: the same set of nodes have multiple types of relationships, each one described by a layer.
- C) **Multi-layer**: The nodes in each layer can be distinct.

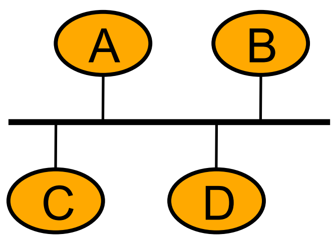
# Beyond simple networks:

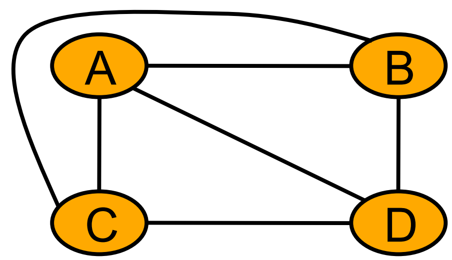
## HyperGraphs

- Edges join more than two nodes at a time (*hyperEdge*)
- Affliation networks
- Examples
  - Families
  - Subnetworks




- Chemical reactions
- co-author networks



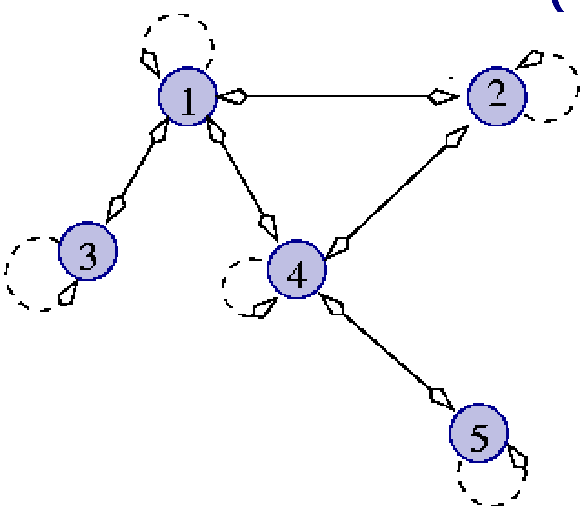

Can be transformed to a *bipartite network*

*Slide from Gunes course, UNR*

# NETWORK TOPOLOGY; simple edges

Binary connectivity matrix, $M$:

$$M_{ij} = \begin{cases} 1 \text{ if edge exists between } i \text{ and } j \\ 0 \text{ otherwise.} \end{cases}$$




$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} = M$$

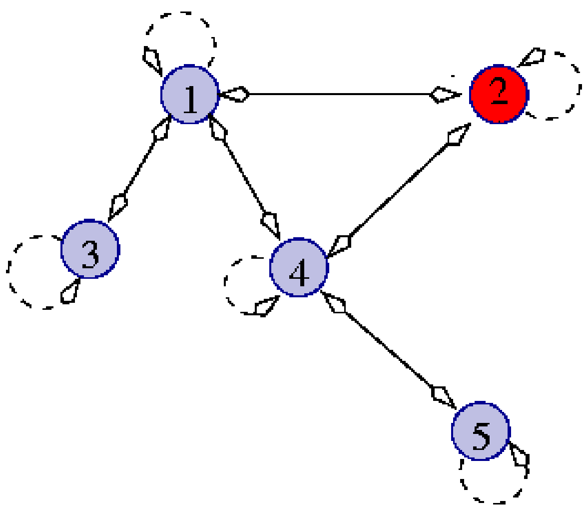
Node **degree** is number of links.

**Network Activity**: FLOWS on NETWORKS

(Spread of disease, routing data, materials transport/flow, gossip spread/marketing)

# FLOWS on NETWORKS : Random walks

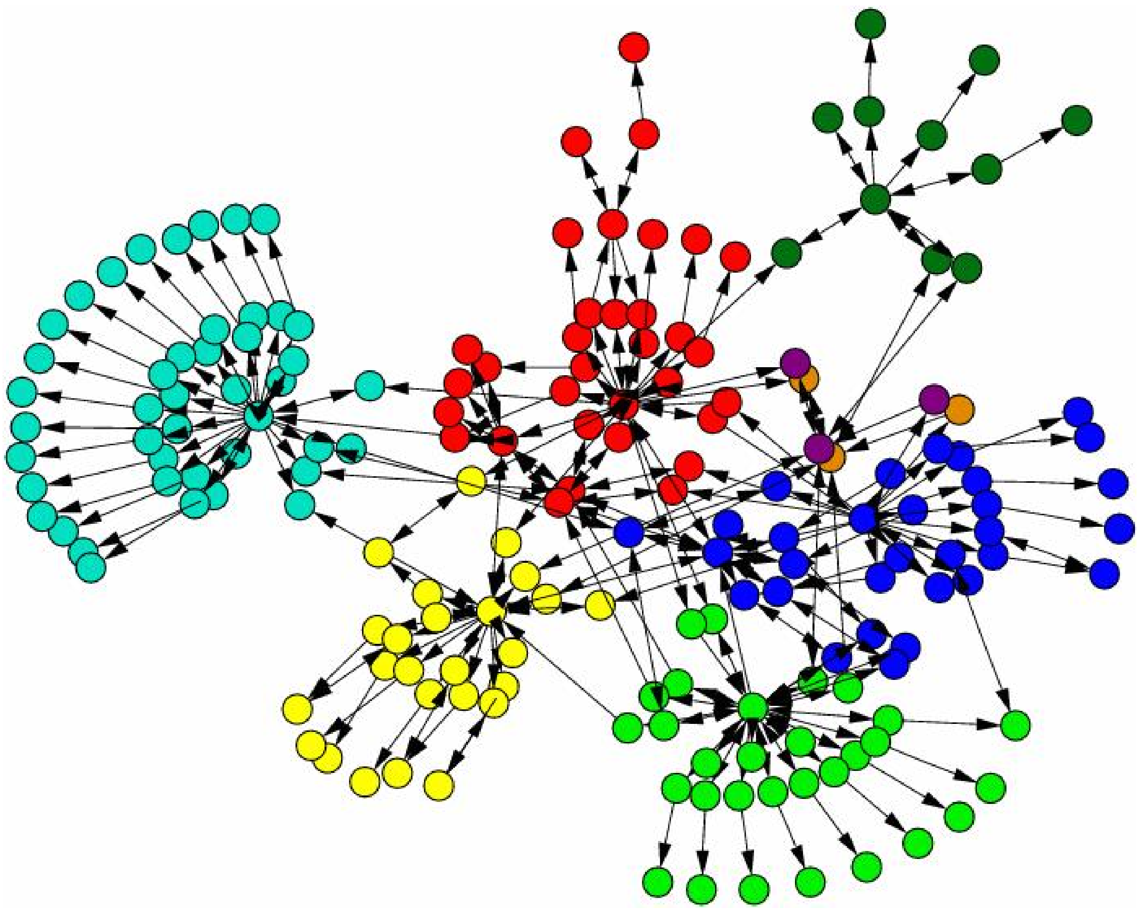
Random walk on the network has state transition matrix, $P$:
(Column normalize the adjacency matrix)




$$\begin{pmatrix} 1/4 & 1/3 & 1/2 & 1/4 & 0 \\ 1/4 & 1/3 & 0 & 1/4 & 0 \\ 1/4 & 0 & 1/2 & 0 & 0 \\ 1/4 & 1/3 & 0 & 1/4 & 1/2 \\ 0 & 0 & 0 & 1/4 & 1/2 \end{pmatrix} = P$$

The eigenvalues and eigenvectors convey much information.
Markov Chains, Spectral Gap.

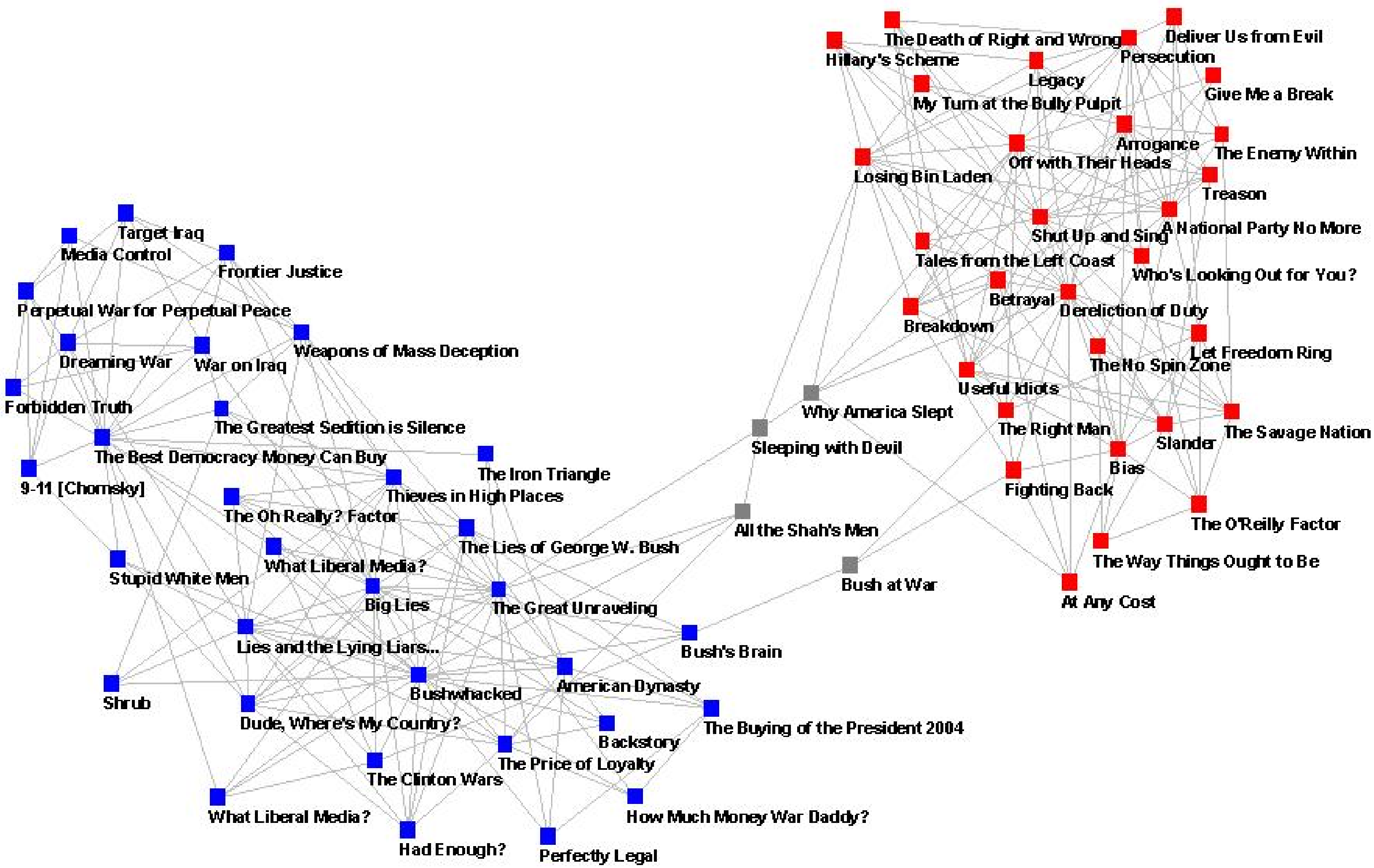
# Random walk on the WWW is the "Page Rank"



Page Rank of a node is the steady-state random walk occupancy probabilty.

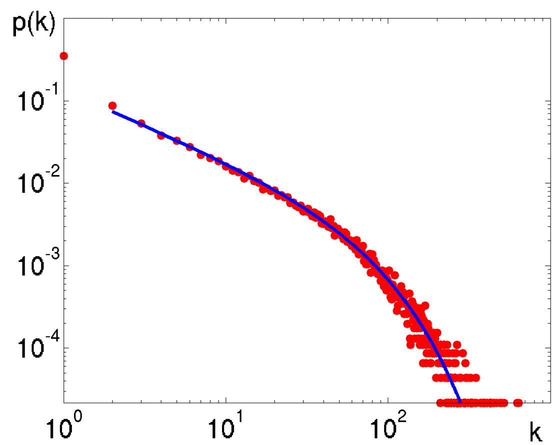
(We will discuss building a search engine in detail later.)

# Example Eigen-technique: Community structure (Political Books 2004)

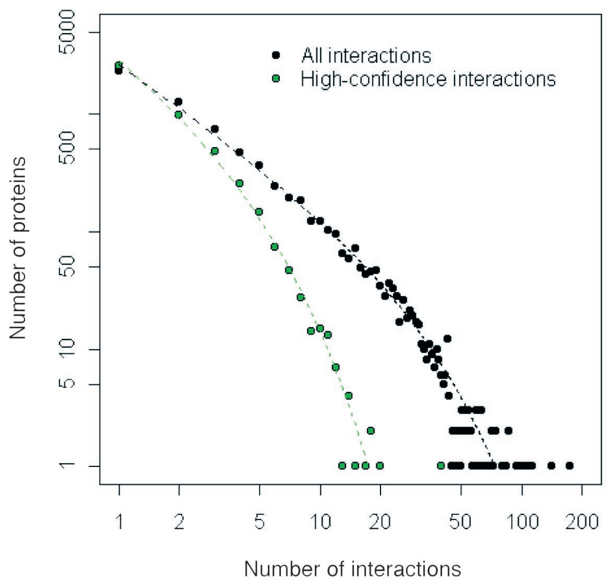



M. Girvan and M. E. J. Newman

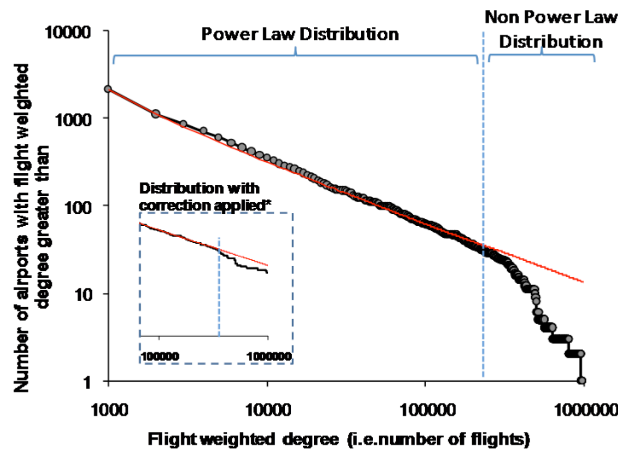
# Back to topology: Broad scale degree distributions




Social contacts
Szendröi and Csányi




Protein interactions
Giot et al Science 2003




Airport traffic
Bounova 2009

- A few hubs, dominated by leaves

- Small data sets, power laws vs log normal, stretched-exponential, etc...
- Exceptions: Power grids? Router-level Internet?

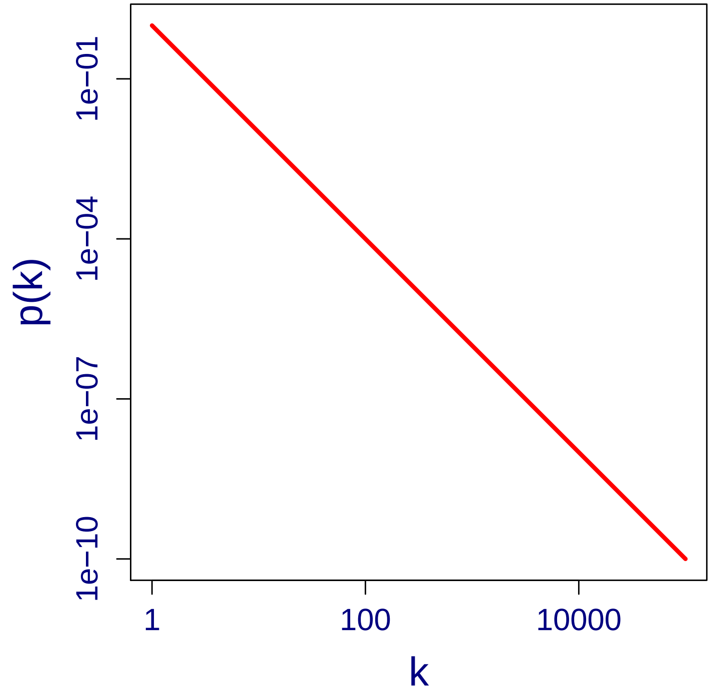
# Degree distribution

- Often observe **"heavy-tailed"** / **"broad-scale"** degree distributions.

- The simplest example of such a distribution is a power law (Pareto distribution).

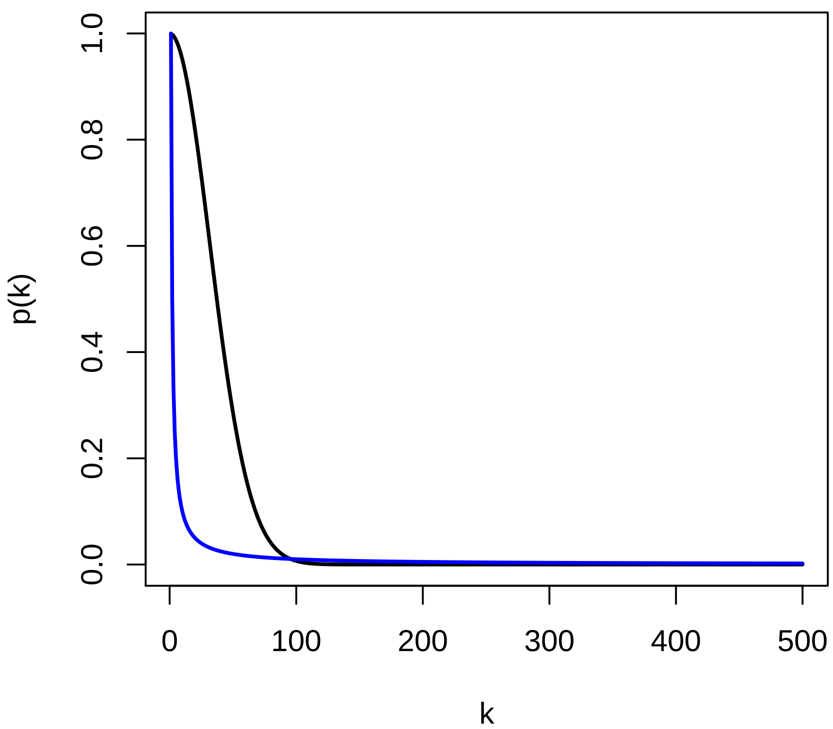
$$p_k \sim k^{-\gamma}$$

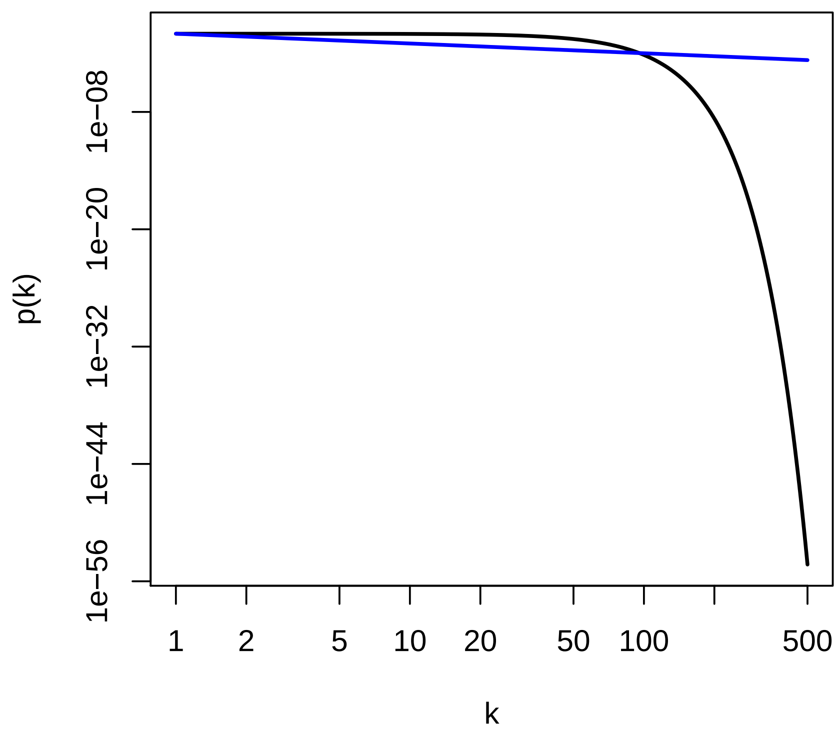
$$\ln p_k \sim -\gamma \ln k$$



# Power Laws versus Bell Curves: "Heavy tails"

- Power law distribution: $p_k \sim k^{-\gamma}$.
- Gaussian distribution: $p_k \sim \exp(-k^2/2\sigma^2)$.



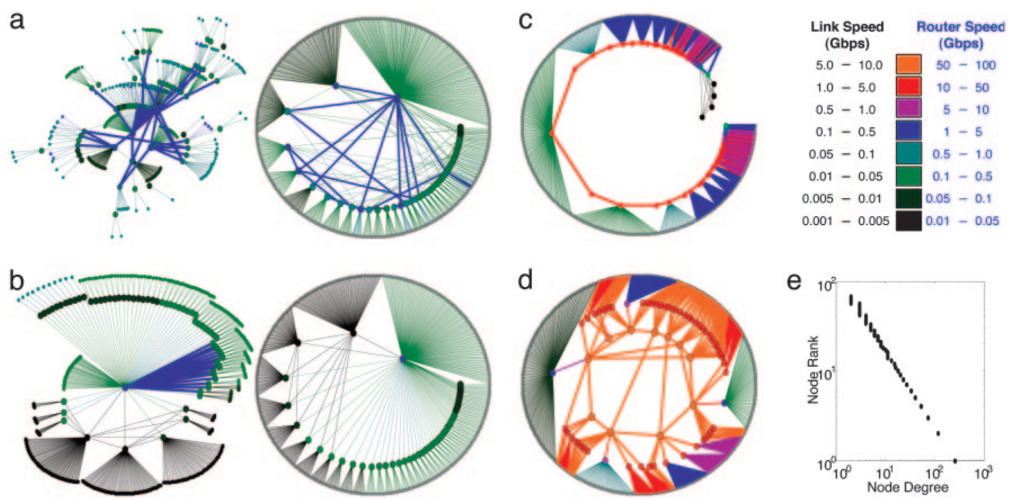

- Most nodes have low degree
- But a few nodes are hubs, with massive degree

# Many network growth models produce power law degree distribution (we will study some of these)

- Preferential attachment
- Copying models (WWW, biological networks, ...)
- Optimization models

## Degree distribution misses other structure.

- Doyle, et. al.,
  PNAS **102** (4)2005.



**Power law probability distributions:** $p_k = Ak^{-\gamma}$ **with** $\gamma > 0$

- $0 \leq p_k \leq 1 \quad \forall k$ which are valid degrees (typically $k \in \mathbb{Z}^+$).
- Must be properly normalized:

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} \frac{A}{k^\gamma} = 1$$

- Approximating discrete sum by integral:

$$1 = \int_{k=1}^{\infty} \frac{A}{k^\gamma} = -\left(\frac{A}{\gamma - 1}\right) \frac{1}{k^{(\gamma-1)}}\bigg|_{k=1}^{\infty}$$

$$= \left(\frac{A}{1-\gamma}\right)\left(\frac{1}{\infty^{(\gamma-1)}} - 1\right)$$

- Finite requirement means $\gamma > 1$, in which case $A = (\gamma - 1)$ .

# The first moment (the mean)

Recall, $p_k = \frac{A}{k^\gamma}$

- Mean degree:

$$\langle k \rangle = \sum_{k=1}^{\infty} k p_k \approx \int_{k=1}^{\infty} k p_k dk = \int_{k=1}^{\infty} \frac{A}{k^{(\gamma-1)}} dk$$

Diverges (i.e., $\langle k \rangle \to \infty$) if $\gamma \leq 2$.

## The second moment and the variance

- Second moment:

$$\langle k^2 \rangle = \sum_{k=1}^{\infty} k^2 p_k \approx \int_{k=1}^{\infty} k^2 p_k dk = \int_{k=1}^{\infty} \frac{A}{k^{(\gamma-2)}} dk$$

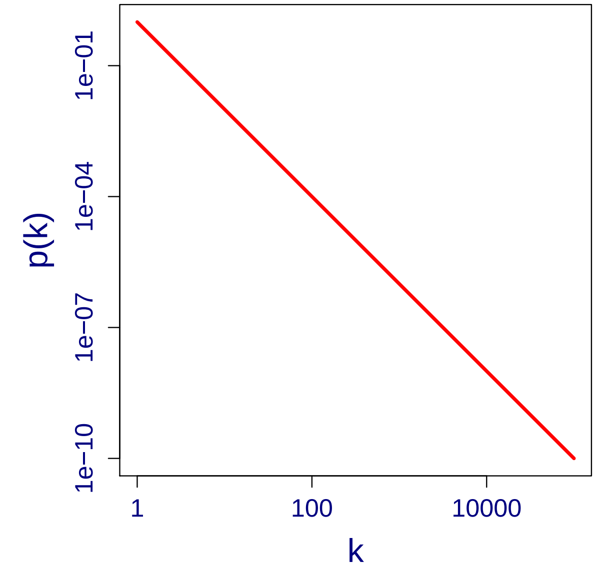
Diverges (i.e., $\langle k^2 \rangle \to \infty$) if $\gamma \leq 3$.

- Variance $= \langle k^2 \rangle - \langle k \rangle^2$, likewise diverges if $\gamma \leq 3$.

# Properties of a power law PDF (Summary)

(PDF = probability density function)

- To be a properly defined probability distribution need $\gamma > 1$.
- For $1 < \gamma \leq 2$, both the average $\langle k \rangle$ and variance $\sigma^2$ are infinite!
- For $2 < \gamma \leq 3$, average $\langle k \rangle$ is finite, but variance $\sigma^2$ is infinite!
- For $\gamma > 3$, both average and variance finite.



# Why a power law is “scale-free”

- Power law for “x”, means “scale-free” in x:

$$p(bx) = (bx)^{-\gamma} = b^{-\gamma} p(x)$$

$$\boxed{\frac{p(bk)}{p(k)} = b^{-\gamma}}$$ regardless of $k$.

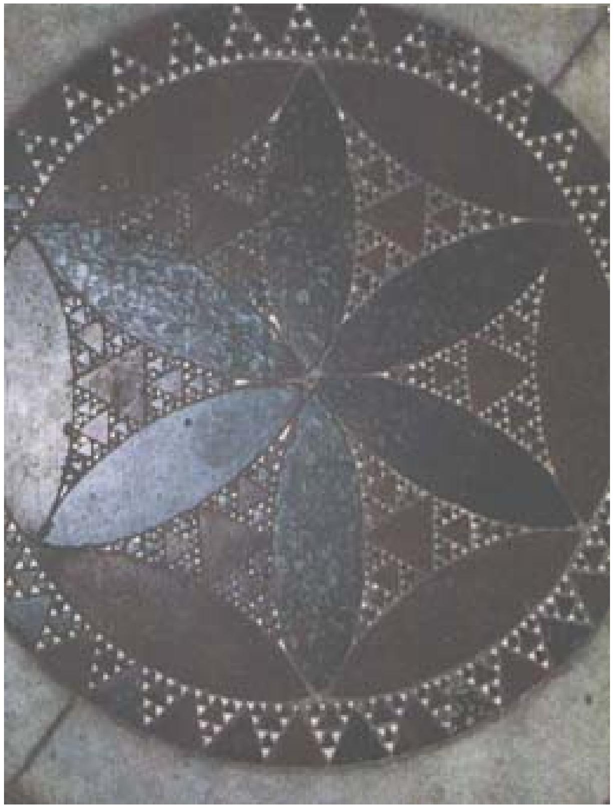
---

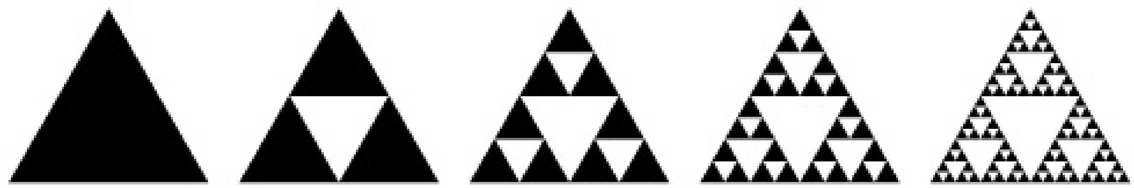
In contrast consider: $p(k) = A\exp(-k)$.

So $p(bk) = A\exp(-bk)$.

$$\boxed{\frac{p(bk)}{p(k)} = \exp[-k(b-1)]}$$ dependent on $k$

## Self-similar/scale-free fractal structures





Sierpinski Sieve/Gasket/Fractal, $N \sim r^d$.

When $r$ doubles, $N$ triples: $3 = 2^d$

$d = \log N / \log r = \log 3 / \log 2$

# Power laws in the real world

Confusion

- Power law
- Log normal
- Weibull
- Stretched exponential

All of these distributions can look the same! (Especially when we are dealing with finite data sets — not enough data to get good statistics).

## How to deal with real data

- Can adjust bin size: increase exponentially with degree.
- Consider the Cumulative PDF (the CDF): $P_k = \sum_{l=k}^{\infty} p_l$.

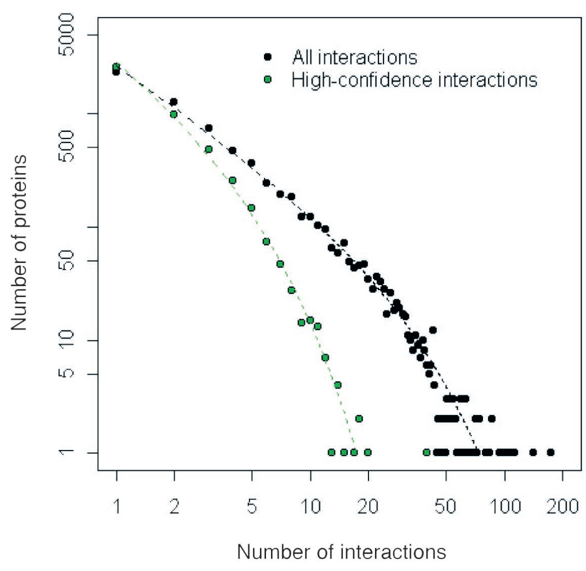
Good reviews:

- Aaron Clauset, Cosma R. Shalizi, M. E. J. Newman. “Power-Law Distributions in Empirical Data”, *SIAM Review*, Vol. 51, No. 4. (2009), pp. 661-703.
- A Brief History of Generative Models for Power Law and Lognormal Distributions Michael Mitzenmacher, *Internet Math*. Vol 1 (2003), 226-251.

# Power law with exponential tail

Ubiquitous empirical measurements:

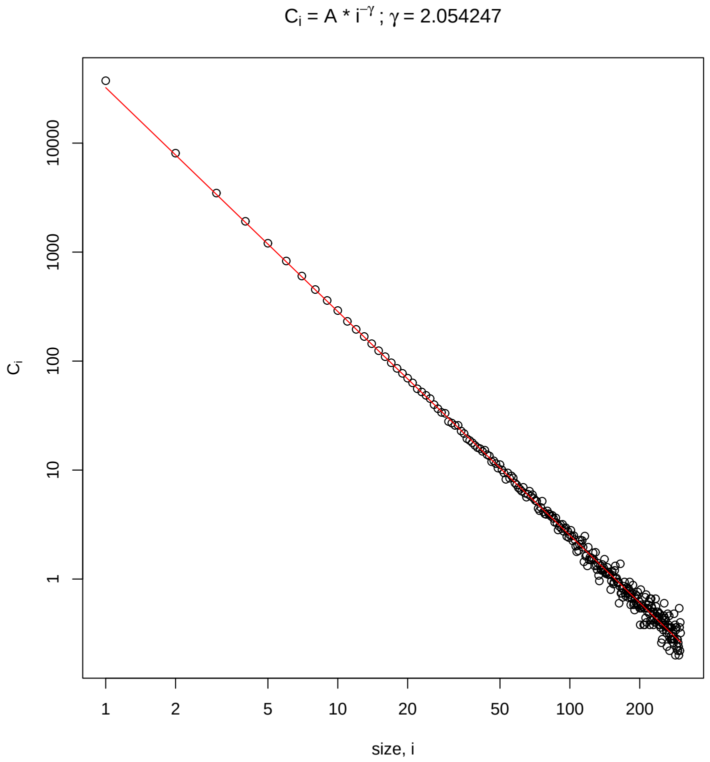
| **System with:** $p(x) \sim x^{-B} \exp(-x/C)$ | $B$ | $C$ |
|---|---|---|
| Full protein-interaction map of *Drosophila* | $1.20$ | $0.038$ |
| High-confidence protein-interaction map of *Drosophila* | $1.26$ | $0.27$ |
| Gene-flow/hydridization network of plants as function of spatial distance | 0.75 | $10^5$ m |
| Earthquake magnitude | 1.35 - 1.7 | $\sim 10^{21}$ Nm |
| Avalanche size of ferromagnetic materials | 1.2 - 1.4 | $L^{1.4}$ |
| ArXiv co-author network | $1.3$ | $53$ |
| MEDLINE co-author network | $2.1$ | $\sim 5800$ |
| PNAS paper citation network | 0.49 | 4.21 |



# True power laws are observed in many systems

- Signature of a system at the "critical point" of a phase transition.
  Foundation of renormalization group approach to critical phenomena

- Random graphs at critical point;
  component sizes: $N_k \sim k^{-5/2}$
  (Note, $\gamma = 2.5$)



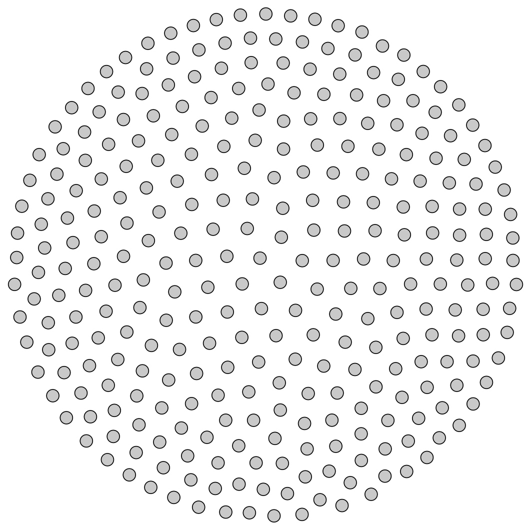
## The origins of network theory: Random graphs

# What does a "typical" graph with $n$ vertices and $m$ edges look like?

- P. Erdös and A. Rényi, "On random graphs", *Publ. Math. Debrecen.* **6**, 1959.
- P. Erdös and A. Rényi, "On the evolution of random graphs", *Publ. Math. Inst. Hungar. Acad. Sci.* **5**, 1960.
- E. N. Gilbert, "Random graphs", *Annals of Mathematical Statistics* **30**, 1959.

# Erdös-Rényi random graphs

- Consider a *labelled* graph. Each vertex has a label ranging from $[1, 2, 3, \cdots n]$, for a set of $n$ vertices. (This will make counting and analysis easier.)



- Let $E$ denote the total number of edges possible:

$$E = \binom{N}{2} = \frac{N!}{2!(N-2)!} = \frac{N(N-1)}{2}$$

(If directed edges, we would not divide by $2$).

## Two formulations

- 1) $\mathcal{G}(n,p)$: The *ensemble* of graphs constructed by putting in edges with probability $p$, independent of one another. (An edge is present with probability $p$ and absent with probability $[1-p]$.)

  Let $G(n,p)$ denote a random realization of $\mathcal{G}(n,p)$.

- 2) $\mathcal{G}(n,m)$: The ensemble of all graphs with $n$ nodes and exactly $m$ edges.

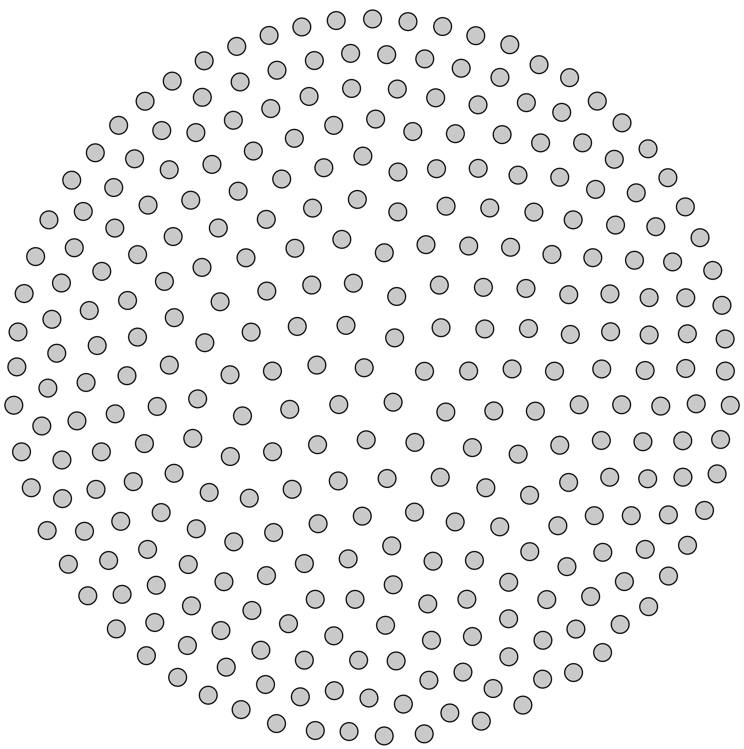
  Let $G(n,m)$ denote a random realization of $\mathcal{G}(n,m)$.

- The two are almost interchangeable with $m = pE$.
  (Recall, $E$ is total number of edges possible).

- We will focus on $G(n,p)$.

# The “classic” random graph, $G(N, p)$
# (The Null Model)

- P. Erdös and A. Rényi, “On random graphs”, *Publ. Math. Debrecen.* 1959.
- P. Erdös and A. Rényi, “On the evolution of random graphs”, *Publ. Math. Inst. Hungar. Acad. Sci.* 1960.
- E. N. Gilbert, “Random graphs”, *Annals of Mathematical Statistics*, 1959.



- Start with $N$ isolated vertices.
- Add random edges one-at-a-time.

  $N(N-1)/2$ total edges possible.
- After $E$ edges, probability $p$ of any edge is $p = 2E/N(N-1)$

## What does the resulting graph look like?

(Typical member of the ensemble)

## Explicitly building $G(n, p)$

- Build a realization of $G(n, p)$ by the following graph process:

- Start with $n$ isolated vertices.

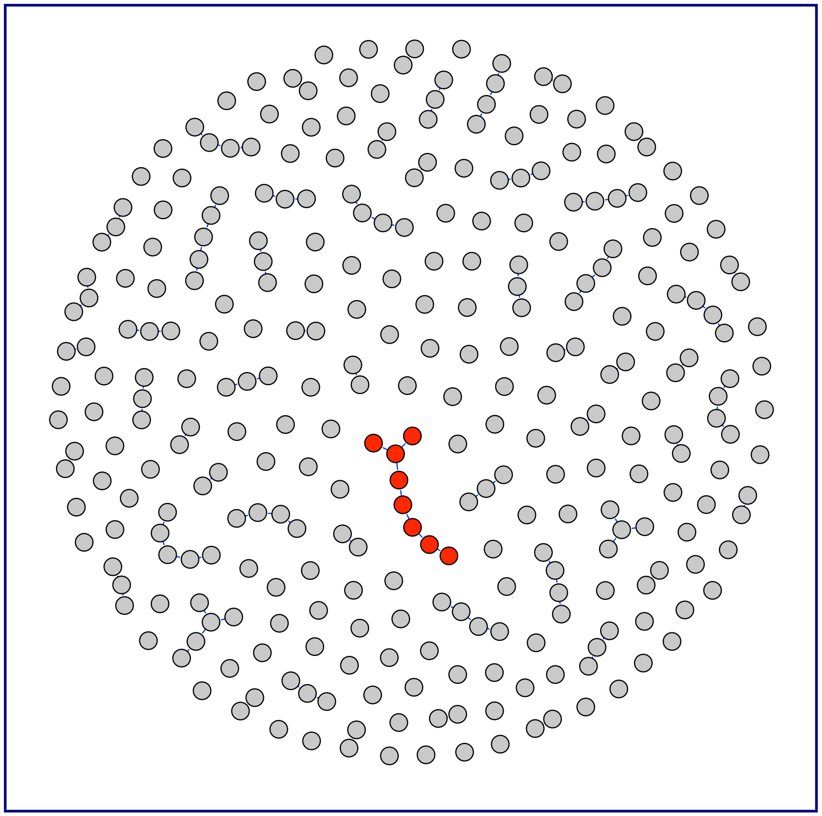
- At each discrete time step, add one edge chosen at random from edges not yet present on the graph.

- At "time" $t$ (i.e., at the addition of $t$ edges), we have built a realization of $G(n, p)$ where $p = t/E$.

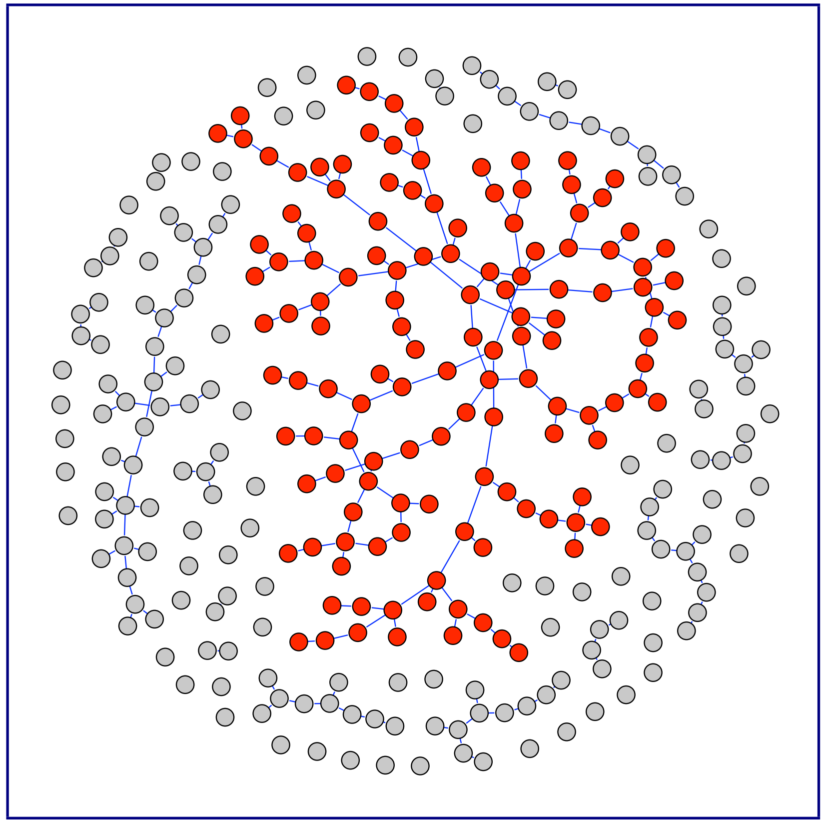
- This is a Markov process (build graph at time $t + 1$ from graph at time $t$).

Ben-Naim, **Krapivsky**, "Kinetic theory of random graphs", *PRE*, 2005.

## N=300



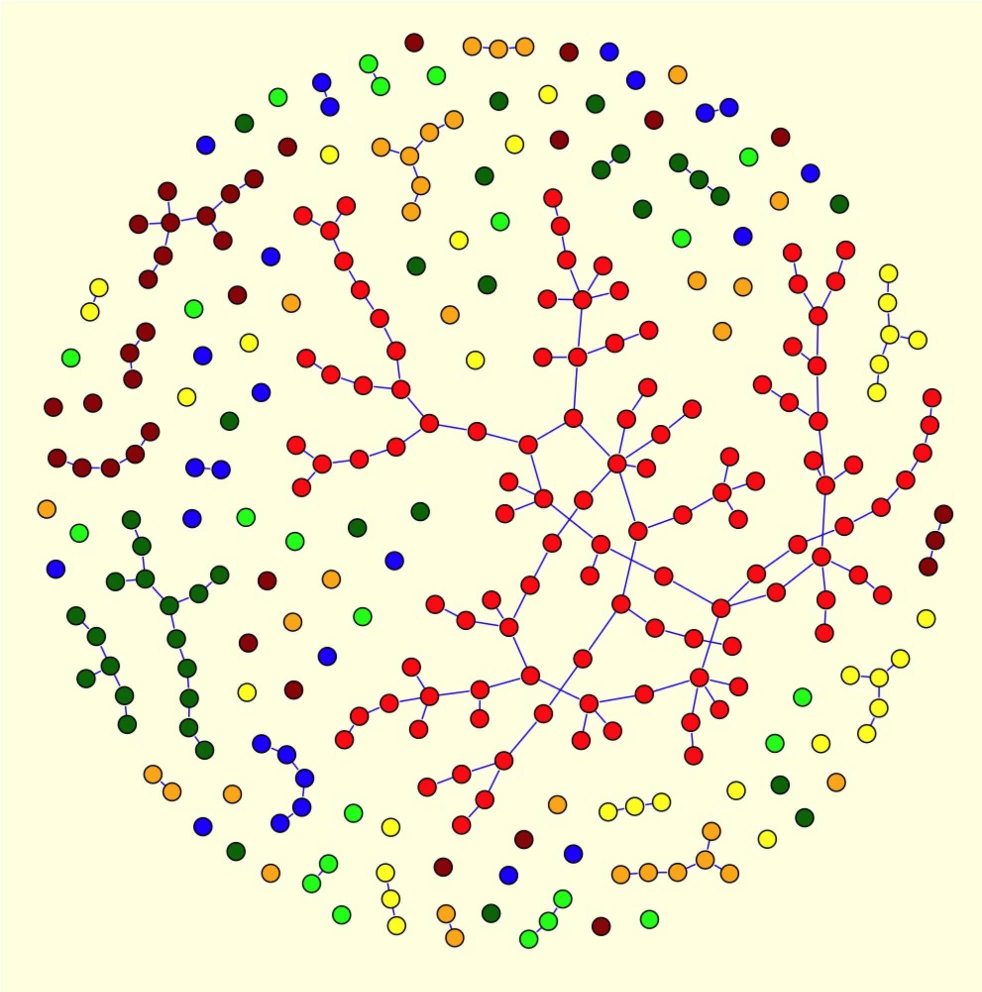
$p = 1/400 = 0.0025$



$p = 1/200 = 0.005$

# Component

A component is a subset of vertices in the graph each of which is reachable from the other by some path through the network.



# Behavior for small $p$

- Consider a realization $G(n, p)$ for $0 < p < 1$ and $n \to \infty$.

  (A number of interesting properties of random graphs can be proven in this limit.

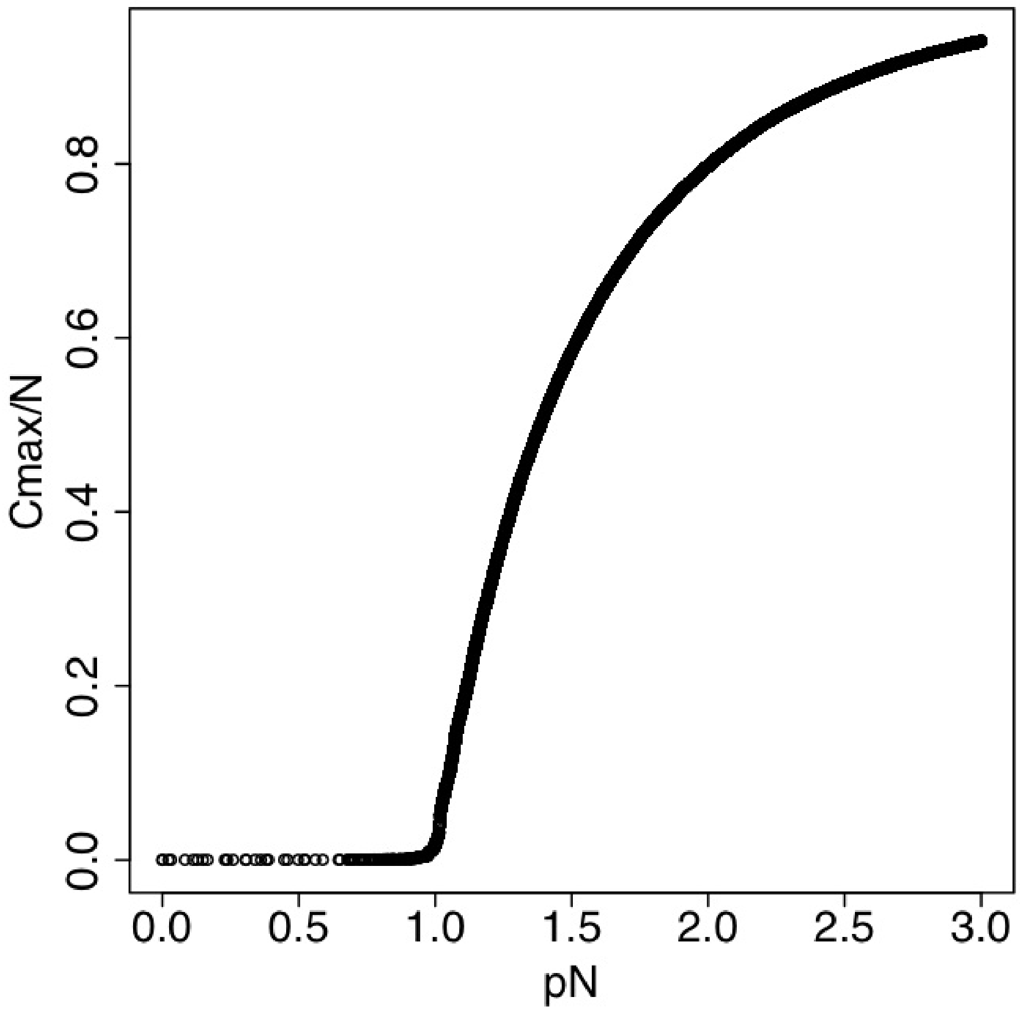
  The $n \to \infty$ limit is also called the “thermodynamic limit”. )

- Let $C_{max}(p)$ denote the size of the largest component of $G(n, p)$ as a function of $p$.

- For small $p$, few edges on the graph. Almost all vertices disconnected. The components are small, with size $O(\log n)$, independent of $p$.

- Keep increasing $p$ (or equivalently $t$ in our model).
  At $p = 1/n$ (i.e. $t = E/n$), something surprising happens:

# Emergence of a "giant component"




- $p_c = 1/N$.
- $p < p_c$, $C_{\max} \sim \log(N)$
- $p > p_c$, $C_{\max} \sim A \cdot N$

(Ave node degree $t = pN$ so $t_c = 1$.)

Branching process (Galton-Watson); "tree"-like at $t_c = 1$.

# A Phase Transition!

An abrupt sudden change in one or more physical properties, resulting from a small change in a external control parameter.

Examples from physical systems:

- Magnetization
- Superconductivity
- Liquid/Gas
- Bose-Einstein condensation

## Giant component observed in real-world networks

- Formation reminiscent of many real-world networks.
  "Gain critical mass".
- Lower bound on emergence of epidemic outbreak.
- The giant component/Strongly Connected Component used extensively to categorize networks.

# Phase transition in connectivity

- Below $p = 1/n$, only small disconnected components.
- Above $p = 1/n$, one large component, which quickly gains more mass. All other components remain sub-linear.
- Note the average node degree, z:

$$\begin{aligned} z &= (2 \times \#edges)/\#vertices \\ &= (2pn(n-1)/2)/n = pn(n-1)/n = (n-1)p \approx np. \end{aligned}$$

(Factor of $2$ since each edge contributes degree to two vertices – each end of the edge contributes.)

Recall, expected number of edges, is pn(n-1)/2 .

- At the phase transition, $z = np = 1$. The phase transition occurs when the average vertex degree is one!

## Erdős-Rényi, a continuous, second order transition: Mean-field scaling behaviors

- Divergence of susceptibility: $\chi = \frac{\partial m}{\partial h} \sim |T - T_c|^{-\gamma}$
- Random graph "susceptibility" (second moment of the component sizes): $\chi = \sum_{i=1}^{\infty} i^2 n_i$
- For Erdős-Rényi, $\chi \sim |t_c - t|^{-\gamma}$, with $\gamma = 1$.

Power law correlation lengths and response functions $\rightarrow$

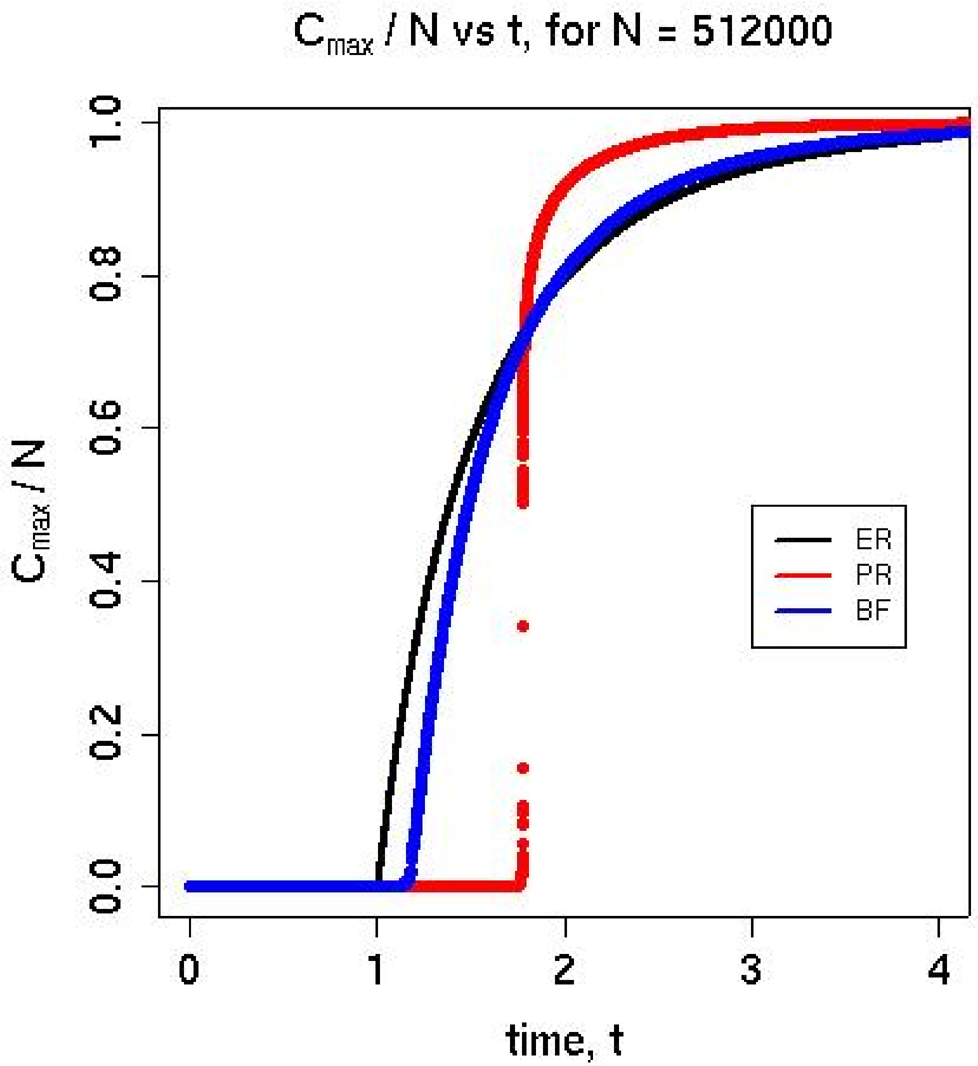
*Potential EARLY WARNING SIGNALS*

(e.g., Scheffer et al. *Nature* 461, 2009)

## Is connectivity a good thing?

- Communication, transportation networks
- Spreading of a virus (human or computer)

# Algorithms for suppressing the emergence of the Giant Component

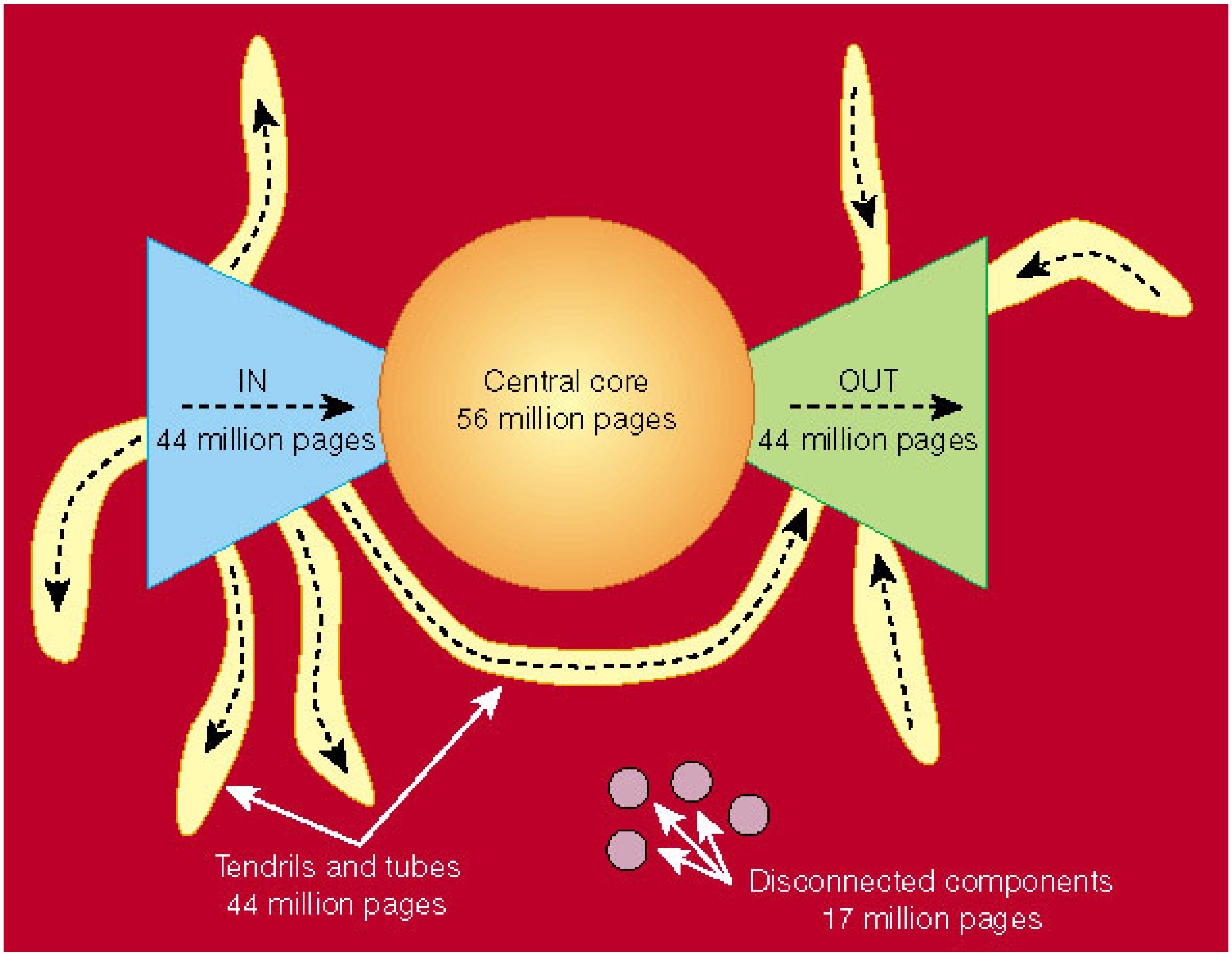



e.g. "Explosive percolation", Achlioptas, D'Souza, Spencer, *Science*, 2009.

## Random graphs as real-world networks?

- What about degree distribution, clustering, assortativity....?
  - – Shown later, Erdos-Renyi yields a Poisson degree distribution, but “configuration” models work around this.
  - – Still need null models to match other properties.
- e.g., “Network Analysis in the Social Sciences”, S. P. Borgatti, A. Mehra, D. J. Brass, G. Labianca, *Science* **323**, 892-895, 2009.
  - – Why would a real network look like a random one?
  - – Local properties of nodes and edges, not statistics of the network.
- Developing the correct null models?

# The giant component/Strongly Connected Component of the WWW




From "The web is a bow tie" Nature **405**, 113 (11 May 2000)

## Summary: Terms introduced today

- Component
- Phase transition
- Degree distribution
- Graph diameter

# Further reading on random graphs

- M. E. J. Newman review, pages 20-25. (Heuristic arguments)
- R. Durrett book, Chaps 1 and 2. (Technical proofs)
- B. Bollobás, *Random Graphs*, 2nd Edition, Cambridge U Press, 2001 (the seminal text on the mathematics of random graphs).

# Class structure

- Two tracks to the class:

  Track A: Project
  (1) Common homeworks (e.g. HW1.pdf, HW2.pdf) and
  (2) HW1a.pdf, HW2a.pdf etc.

  Track B: Advanced HWs
  (1) Common homeworks (e.g. HW1.pdf, HW2.pdf) and
  (2) HW1b.pdf, HW2b.pdf etc.

- Track A: Project
  - Teams of 5-6 people ideal
  - Negative results are OK
  - Ideally aim to have a result for a journal or conference

    e.g., “Latent social structure in open source projects”, C Bird, D Pattison, R D’Souza, V Filkov, P Devanbu, ACM SIGSOFT 2008.

## Project pitch – HW1a

- One page describing your idea. Submitted via Canvas and shared with the class.

- Skill sets to merge:
  Domain specific questions / Methods / Data sets

Jumpstart: In-class on Monday (April 10th) — pitch your idea and build a team!

# Possible topic areas

- Transportation networks and flows; multi-modal transportation
- Open source software – e.g., social and technological networks in github
- Machine learning – e.g., bring network connectivity into binary classifiers
- Power grid modeling
- Opinion dynamics / social unrest / multiplex opinion dynamics
- Ranking in networks; especially temporal, multilayered, higher-order
- Multilayered and temporal macaque monkey networks
- Shocks and tipping points
- Extend standard metrics to multilayered, temporal, or higher-order networks
- Co-author and citation networks
- Food networks
- Recommendation systems
- Biological networks
- Terrorist networks
- See also class homepage "Projects" tab