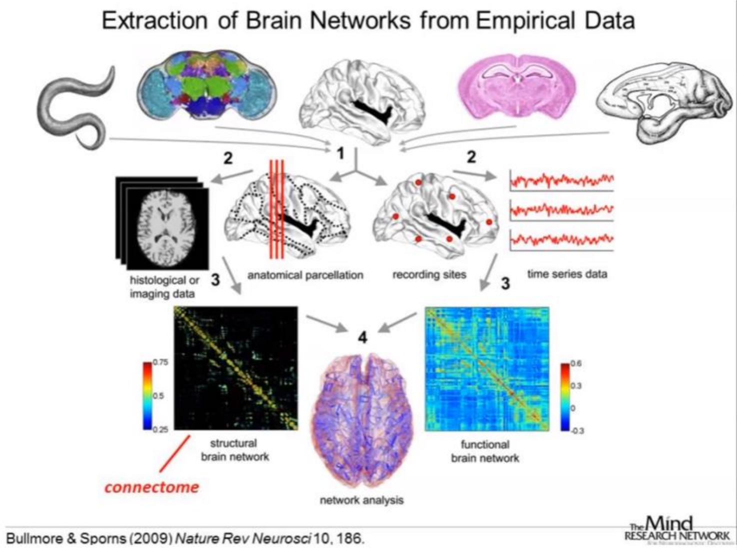
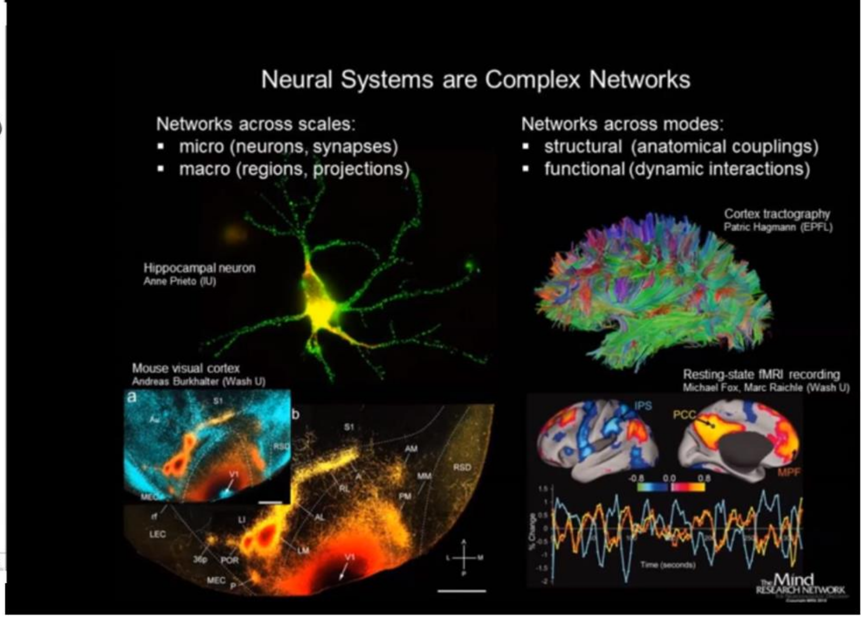




Lecture on Graph Theoretical Network Analysis

Prof. Maria Papadopouli

CS – 590.21 Analysis and Modeling of Brain Networks

Department of Computer Science
University of Crete

Greek Diaspora Fellowship Program

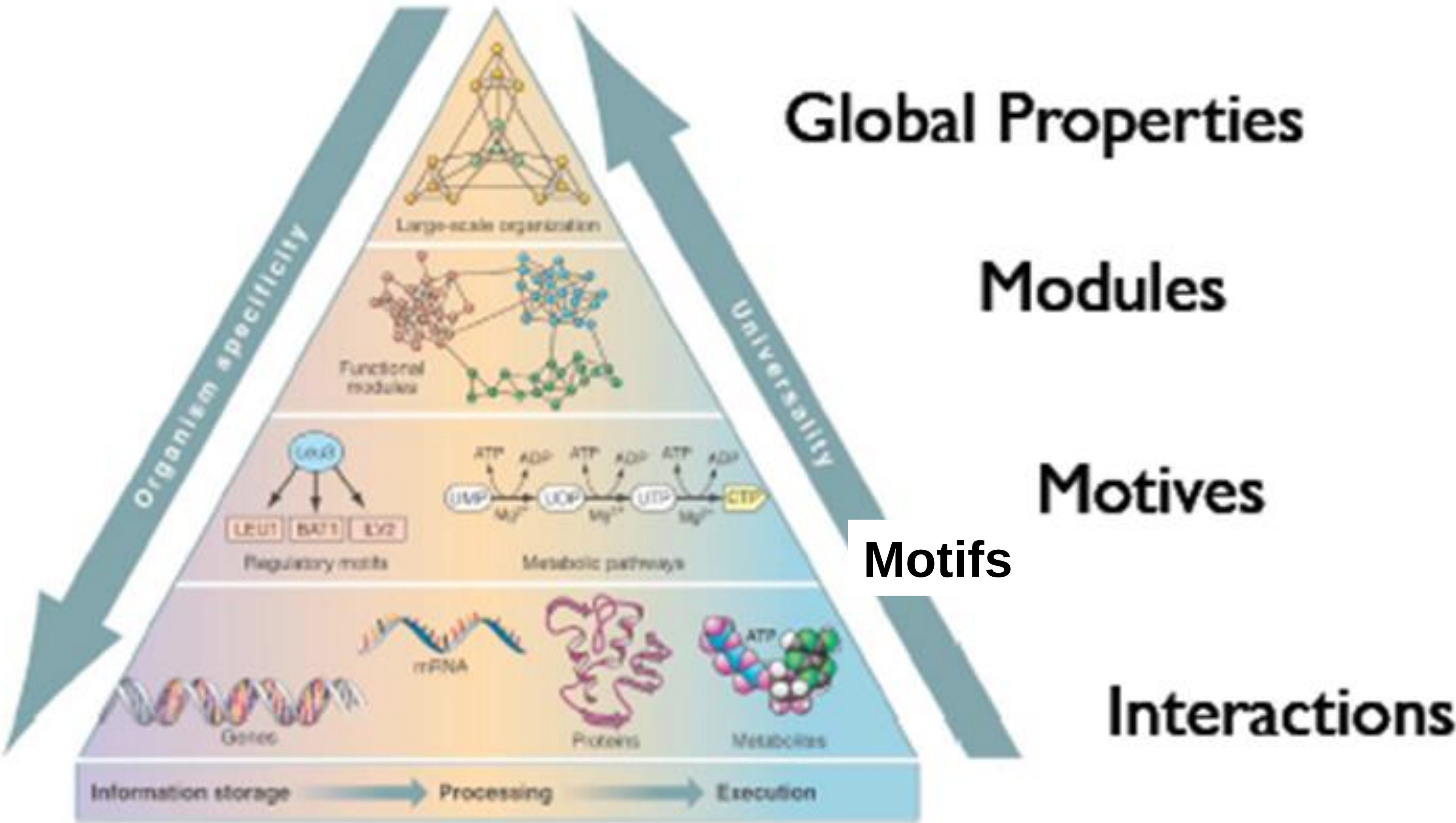
ΙΣΝ / SNF
ΙΔΡΥΜΑ ΣΤΑΥΡΟΣ ΝΙΑΡΧΟΣ
STAVROS NIARCHOS FOUNDATION

ACKNOWLEDGEMENTS

Most of the slides are based on text by
ALBERT-LÁSZLÓ BARABÁSI
NETWORK SCIENCE

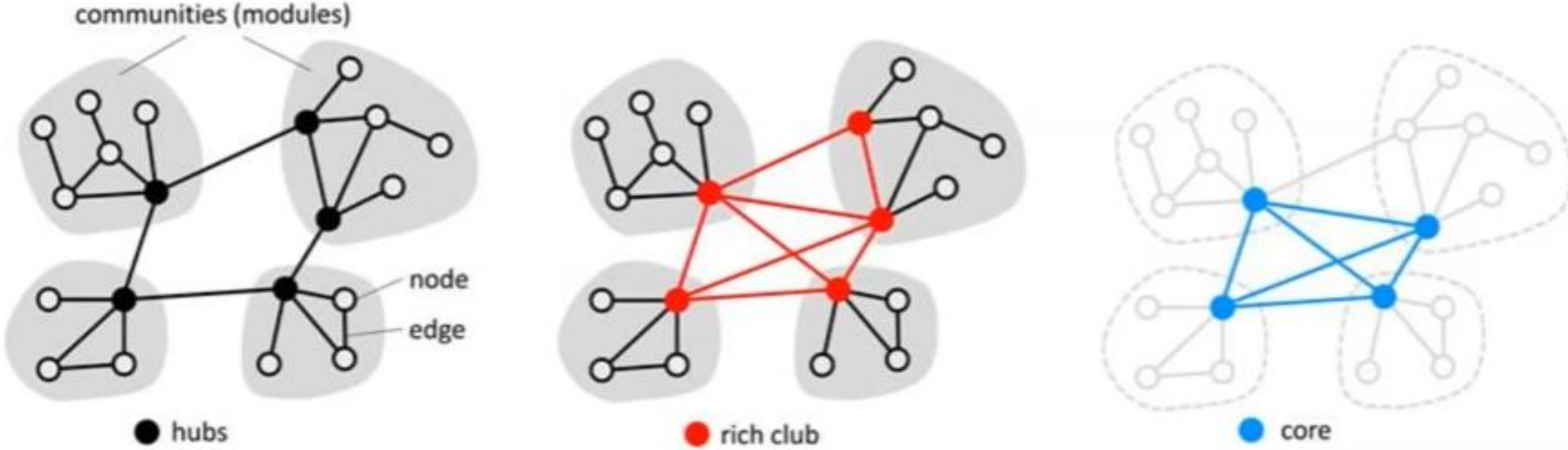
THE RANDOM NETWORKS (Chapter 3) &
THE SCALE-FREE PROPERTY (Chapter 4)

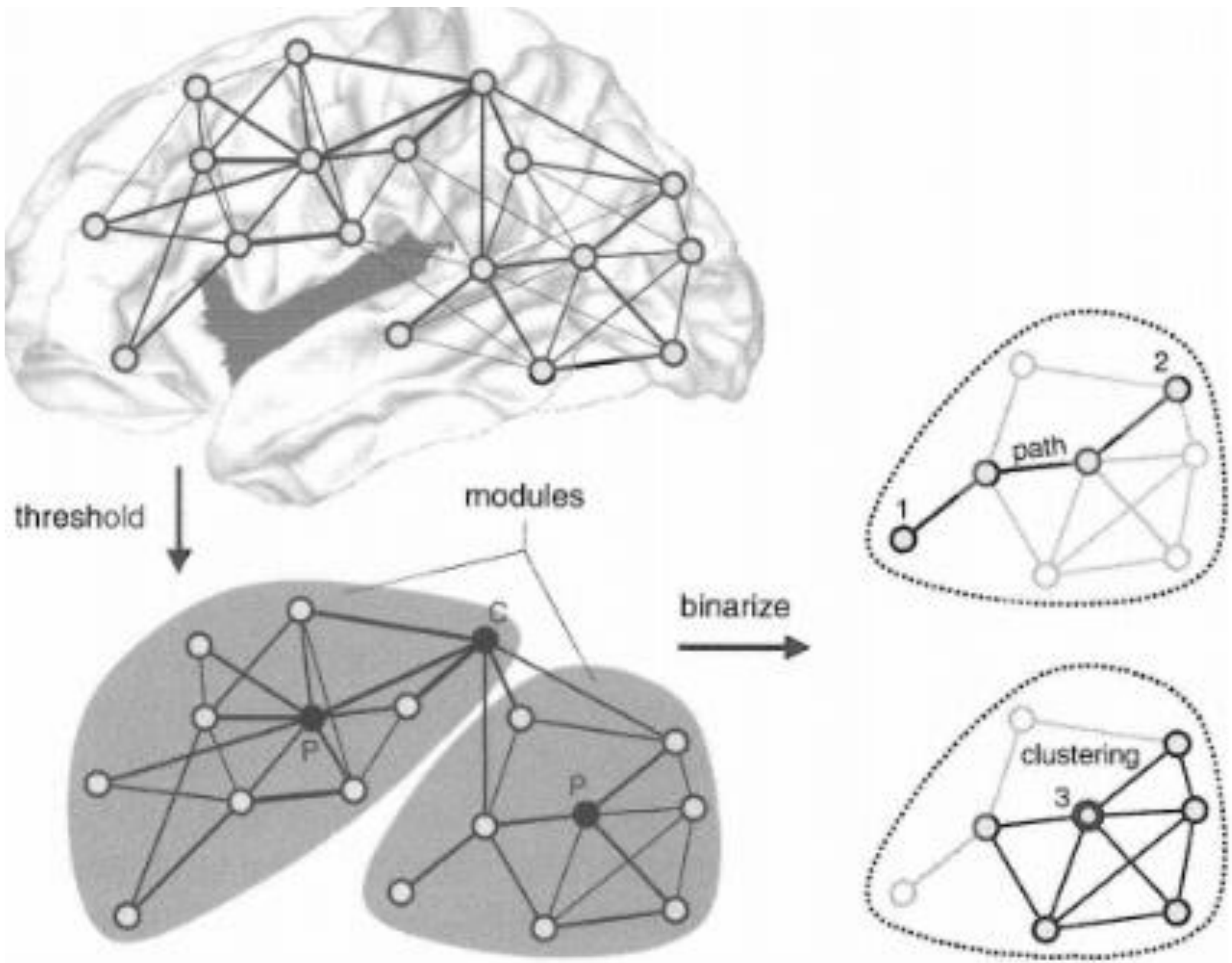
Networks can be analyzed at different levels of detail.



Modules, Cores, and Rich Clubs

In some networks, highly connected/central hub nodes have a tendency to be highly connected to each other ("rich-club" organization).




threshold
modules
binarize
C
P
P
path
1
2
clustering
3

Network Dynamics

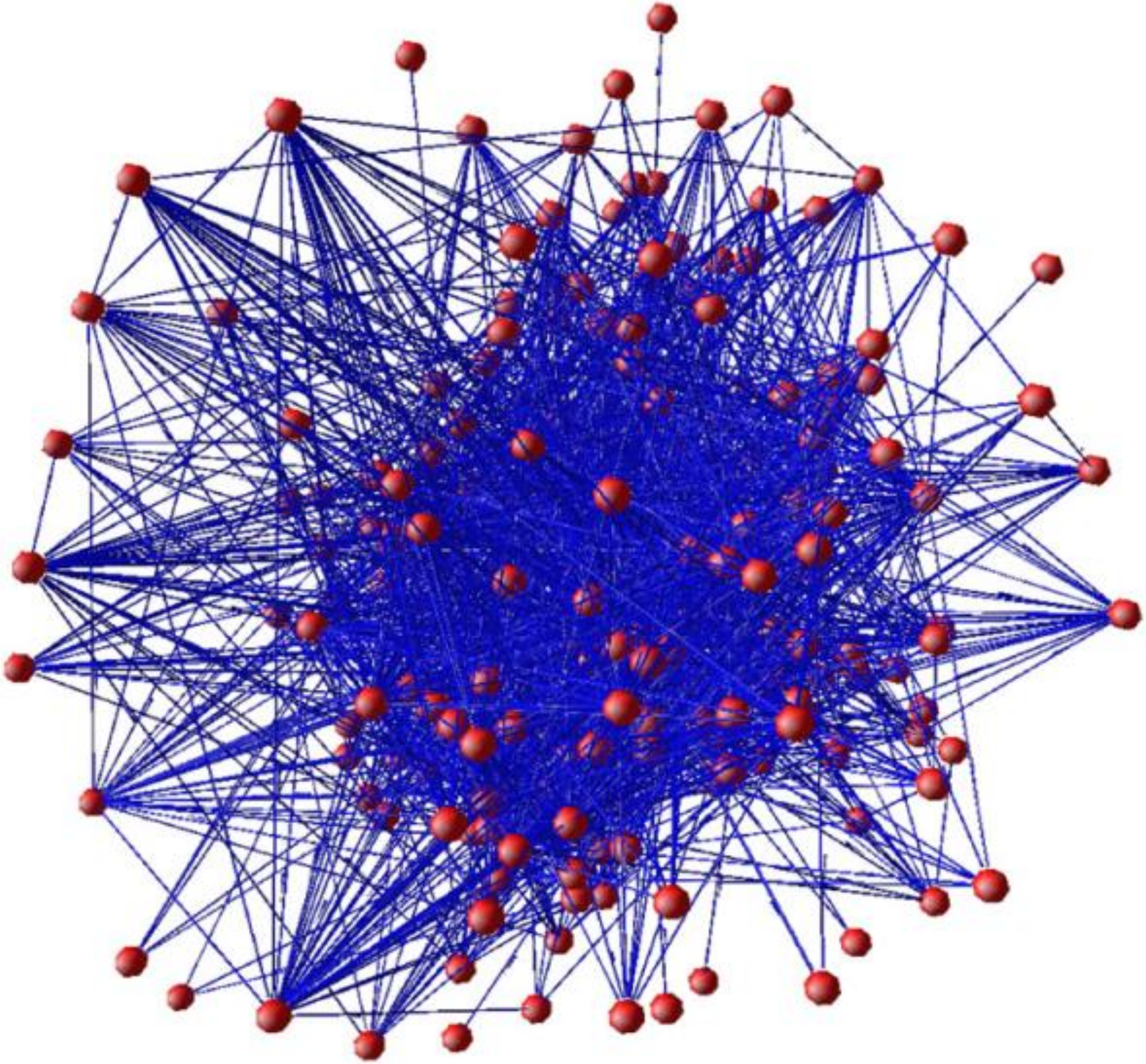
Not all interactions among neurons are active all the time

Not all neurons were born equal!

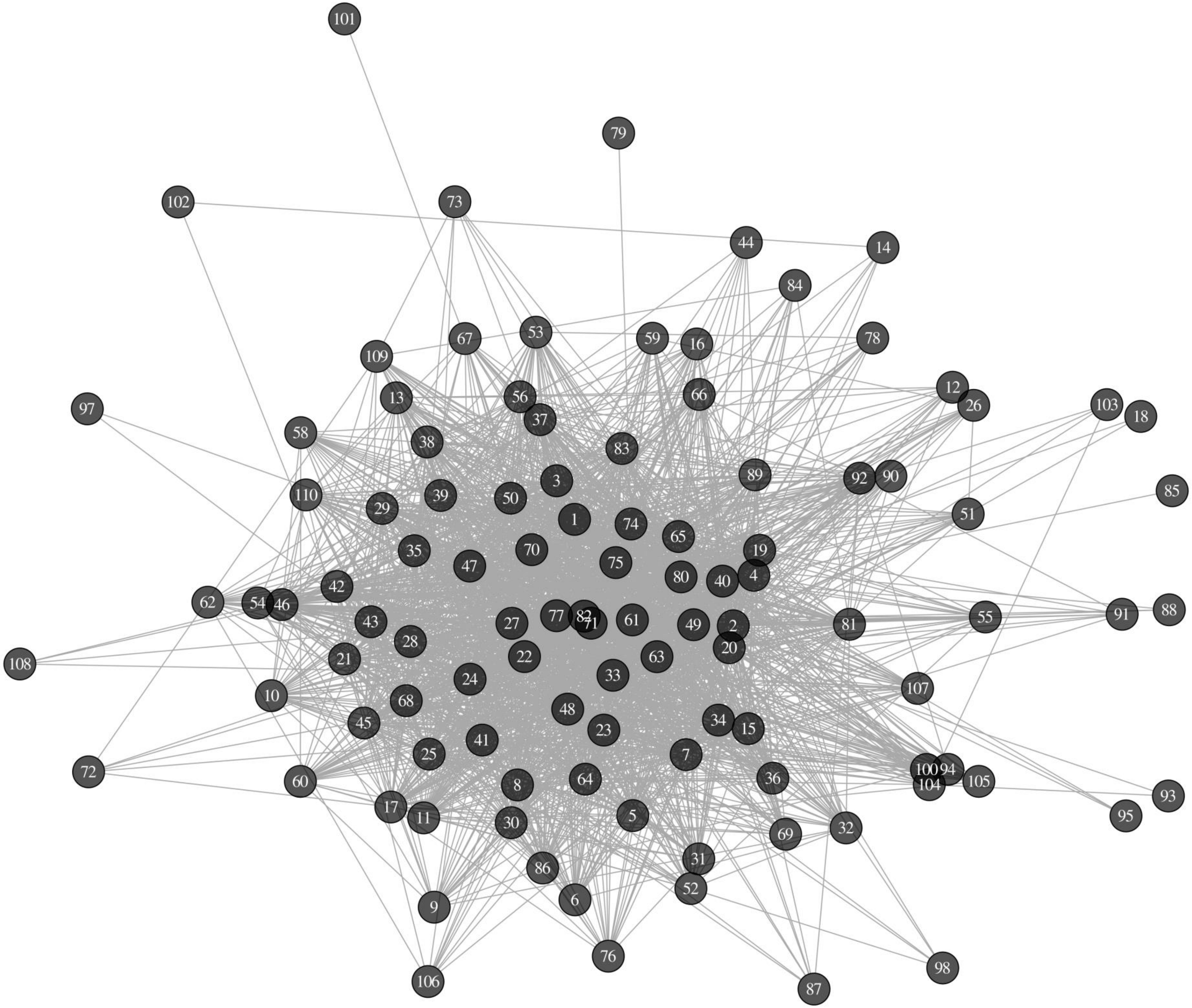
- "Party" hubs: always the same partners (same time & space)
- "Date" hubs: different partners in different conditions (different time and/or space)

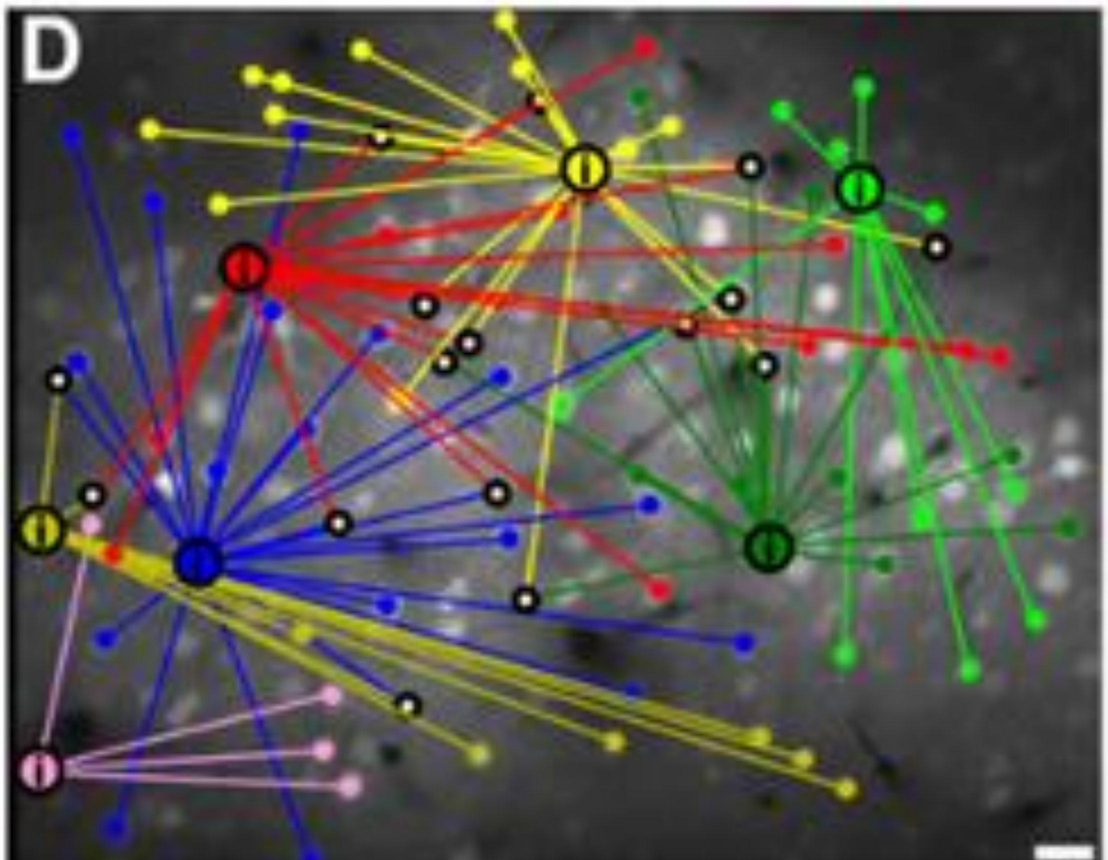
Difference is important for **inter-process communication**

C. elegans neuronal net

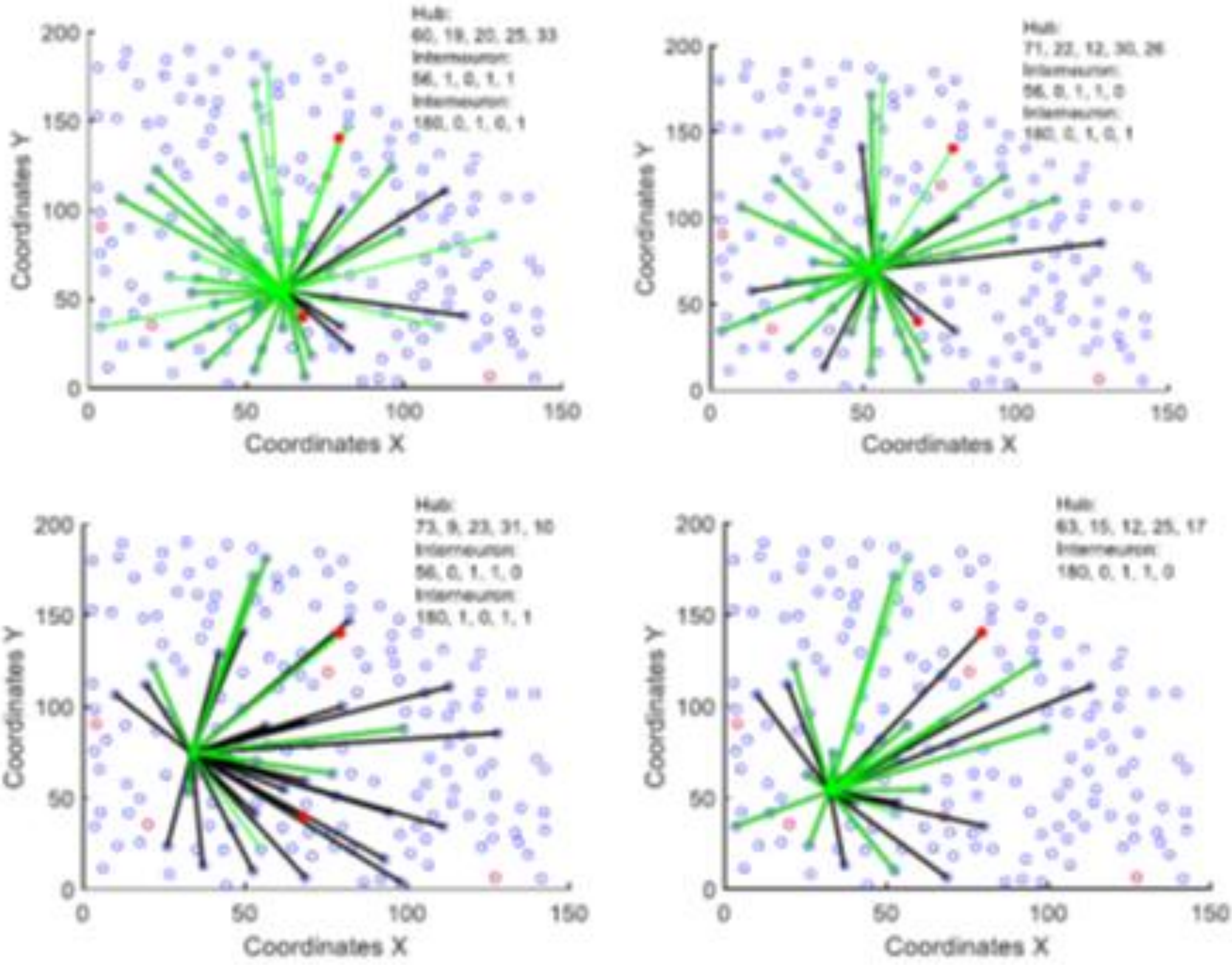


)


101
79
102
73
44
14
84
53
67
59
16
78
109
12
13
56
66
26
103
18
97
37
58
38
83
3
89
92
90
85
110
39
50
29
51
1
74
65
35
70
19
47
75
80
40
4
42
62
54
46
77
82
71
61
55
91
88
43
27
49
2
81
28
20
108
21
22
63
33
24
107
10
68
48
45
34
15
23
41
25
7
72
100
94
60
64
36
104
105
8
93
17
95
11
30
5
32
69
31
86
52
6
9
76
98
106
87

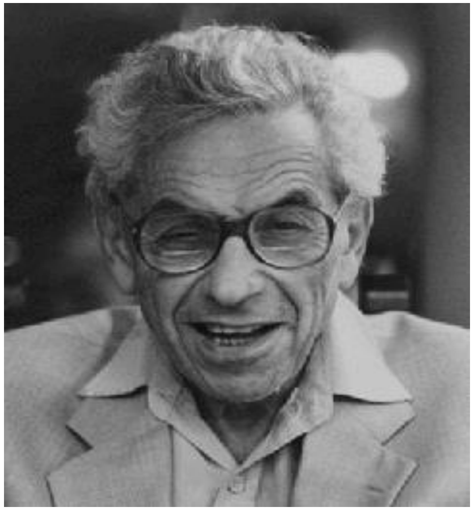

D

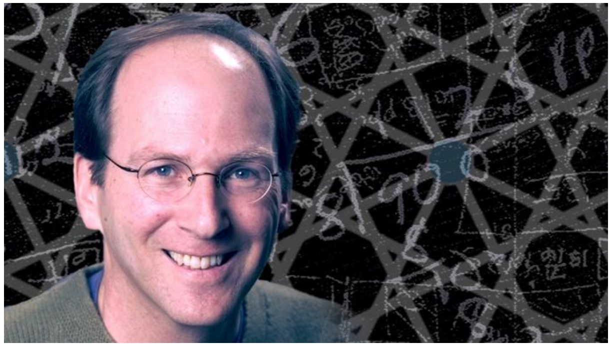
interneurons
pyramidal members
shared pyramidal members


Hub:
60, 19, 20, 25, 33
Interneuron:
56, 1, 0, 1, 1
Interneuron:
180, 0, 1, 0, 1
Coordinates Y
Coordinates X
Hub:
71, 22, 12, 30, 26
Interneuron:
56, 0, 1, 1, 0
Interneuron:
180, 0, 1, 0, 1
Coordinates Y
Coordinates X
Hub:
73, 9, 23, 31, 10
Interneuron:
56, 0, 1, 1, 0
Interneuron:
180, 1, 0, 1, 1
Coordinates Y
Coordinates X
Hub:
63, 15, 12, 25, 17
Interneuron:
180, 0, 1, 1, 0
Coordinates Y
Coordinates X

Important Network Models

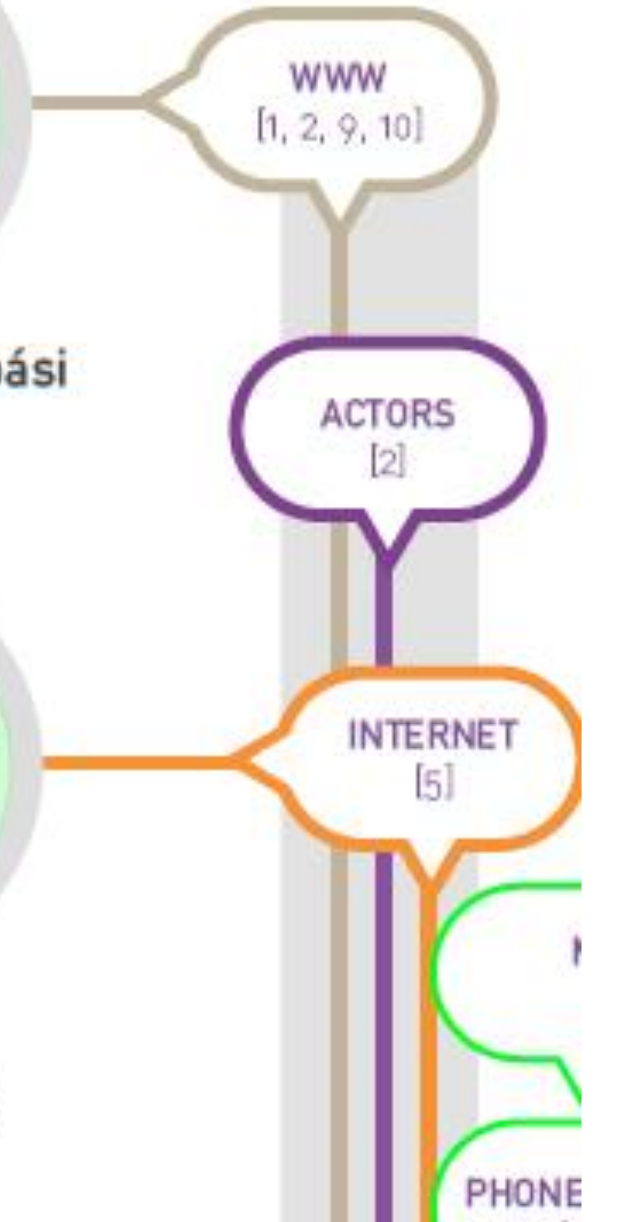
- **Random graph** model (Erdős & Rényi, 1959)
- **Small-world** model (Watts & Strogatz, 1998)
- **Scale-free** model (Barabasi & Alert, 1999)



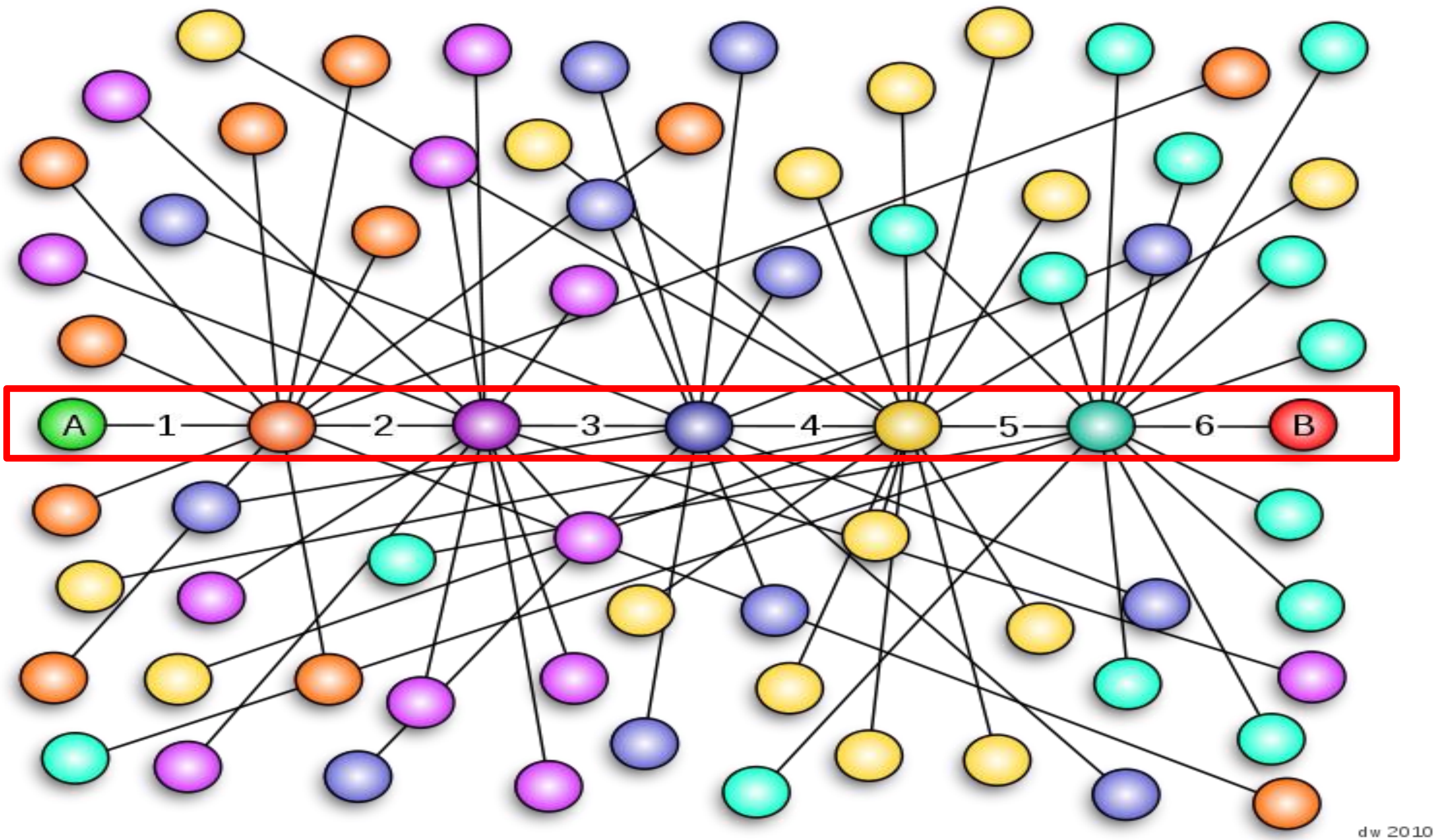




Réka Albert, **Hawoong Jeong**, and **Albert-László Barabási**
discover the power-law nature of the WWW [1]
and introduce scale-free networks [2, 10].

Michalis, **Petros**, and **Christos Faloutsos**
discover the scale-free nature of the internet [15].



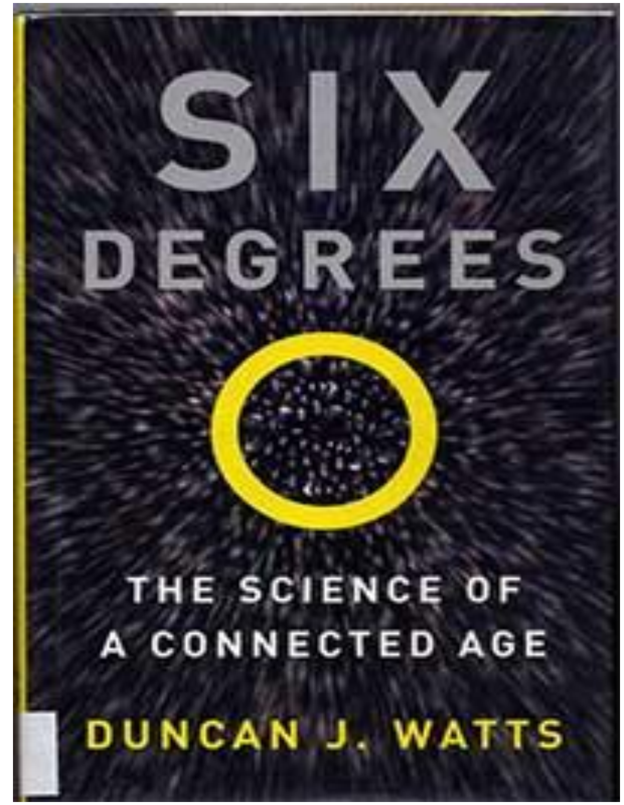
Six Degrees of Separation



Everyone is on **average approximately** **six** **steps away** from any other person on Earth

Stanley Milgram (1933-1984)

- Controversial social psychologist
- Yale & Harvard
- *Small world experiment*, 1967
  - 6 degrees of separation
- *Obedience to authority* - 1963



Modeling a Network as a Graph

- G**raph**: an ordered pair $G = (V, E)$ of a set V of **vertices** (**nodes)** & a set E of **edges** (2-element subsets of V).
- Can be extended to include the set W of the **weights of all edges** in E.

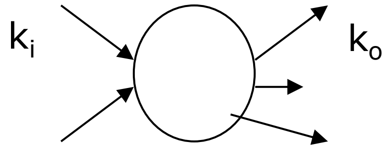
- Edge: models the **interaction between the neurons** it connects.
- The weight of an edge can model the **strength of the interaction**.

- Directed graph: each edge has a direction

  e.g., the edge (a,b) indicates that there is an edge from a to b.

Degree Distribution of a Network

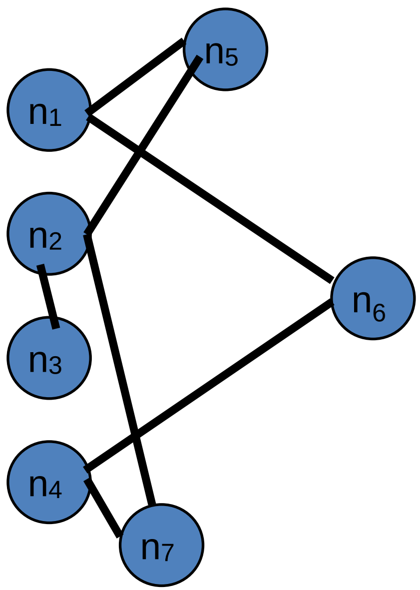
When modeled as a directed graph:



- **in-degree of a vertex** (k_i): number of incoming edges of a vertex
- **out-degree of a vertex** (k_o): number of outgoing edges of a vertex
- **degree** (k): the total number of **connections** $k = k_i + k_o$

Diameter & Paths

Diameter of a graph is the “longest shortest path”.



Path in a graph is a finite or infinite sequence of edges which connect a sequence of vertices which, by most definitions, are all distinct from one another.

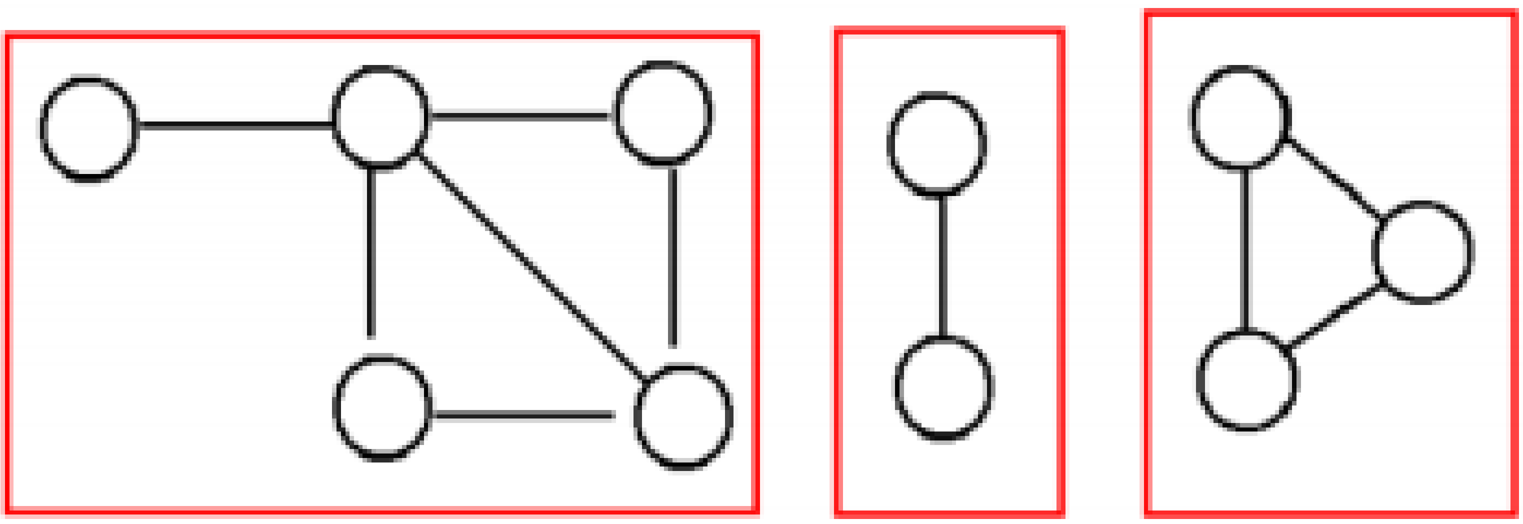
Connectivity

- a graph is ***connected*** if
  - you can get from any node to any other by following a sequence of edges OR
  - any two nodes are connected by a path.

- A directed graph is ***strongly connected*** if there is a directed path from any node to any other node.

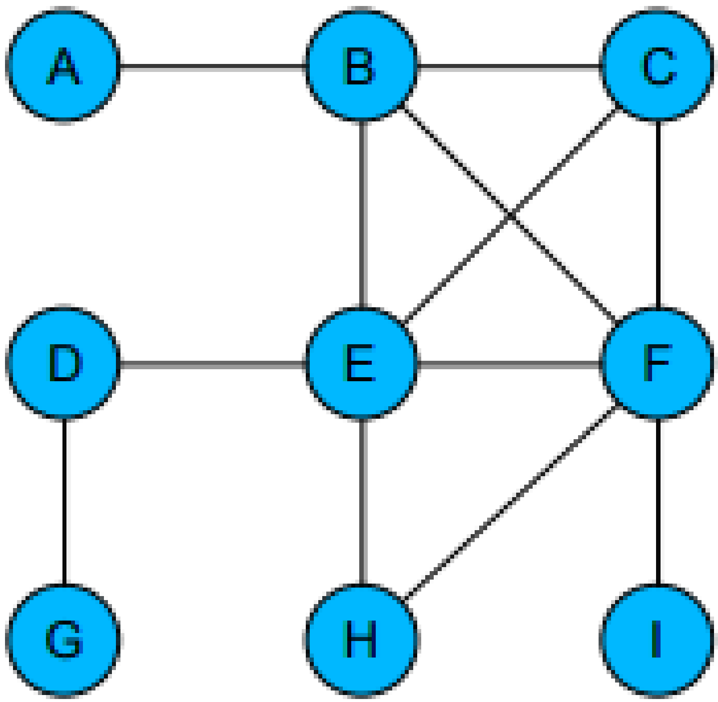
Component

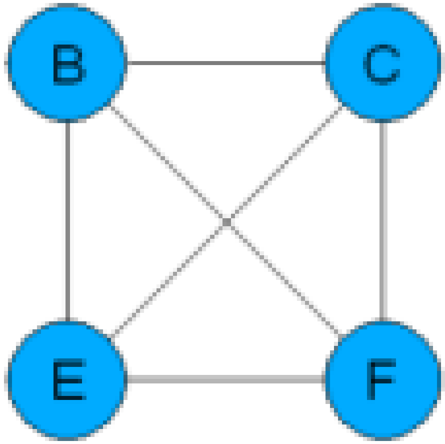
- Every disconnected graph can be split up into a number of connected ***components***.

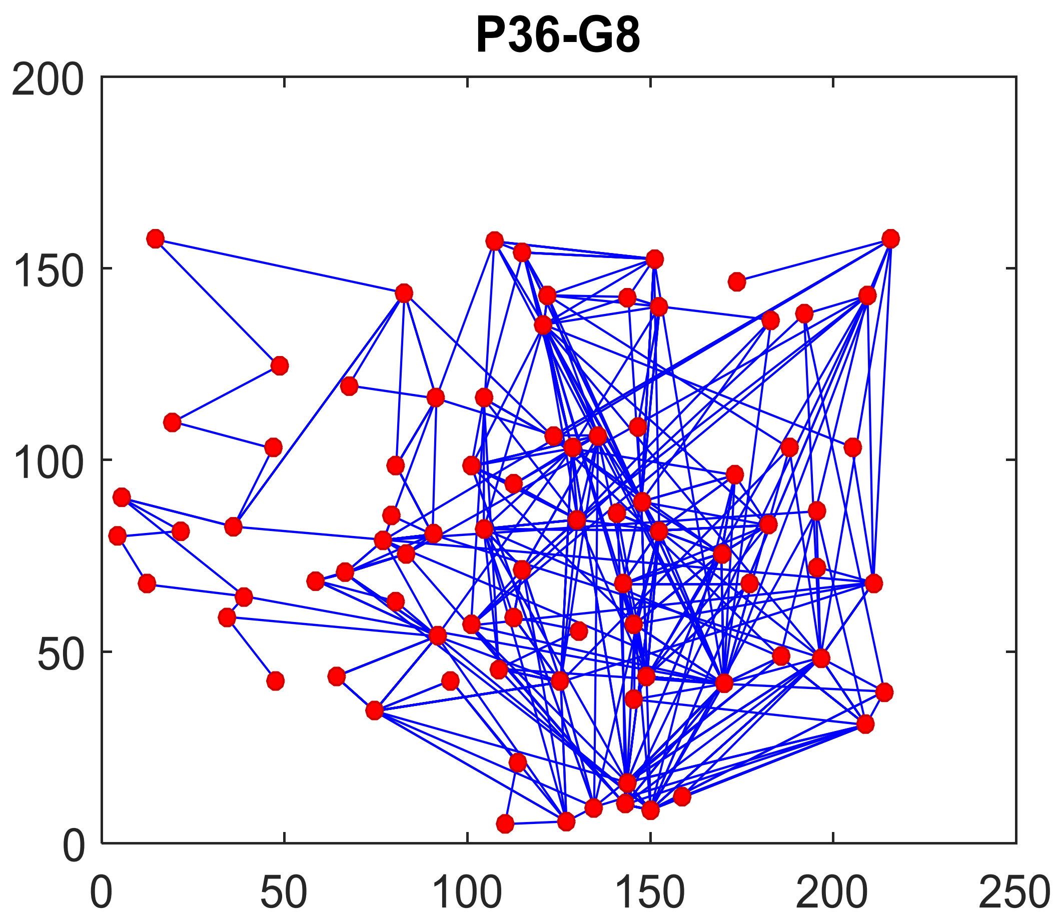


Special Subgraphs: Cliques

A **clique** is a maximum complete connected subgraph.







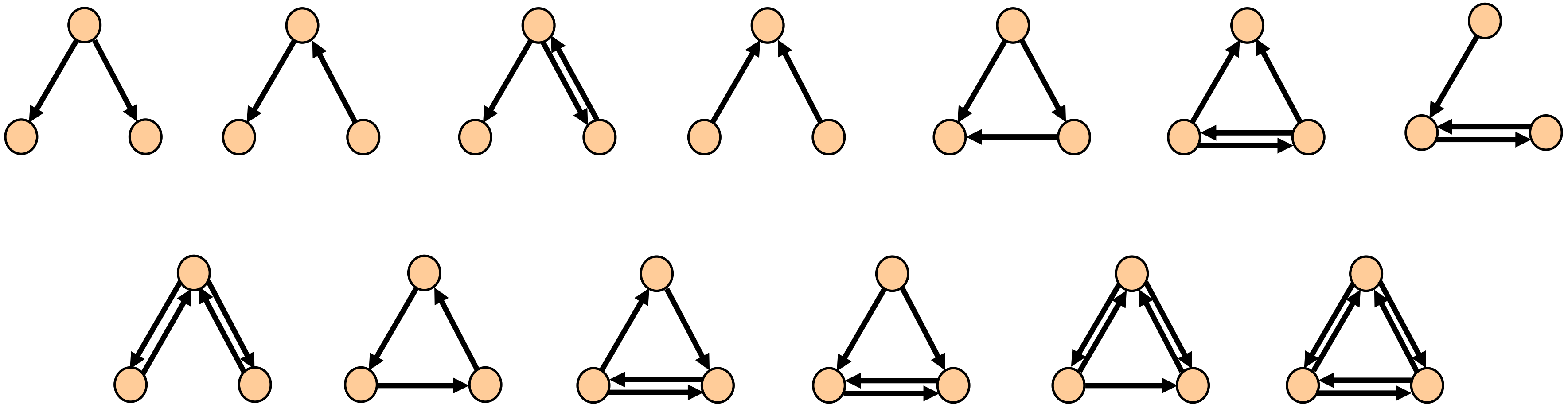
The **big connected component, formed by 83 neurons** (43 neurons were not connected to any other neuron).

Degree of Connectivity				Number (Percentage)		
Average	Median	Max	Min	Hubs	Nodes	Edges
11.8554	8	45	1	9 (10.84%)	83 (65.87%)	492 (6.25%)

Network Motif

Simple Building Blocks of Complex Networks

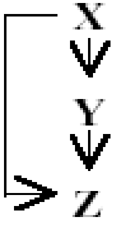
- Focused on directed, cyclic subgraphs of 3 or 4 nodes in yeast (no self-loops)

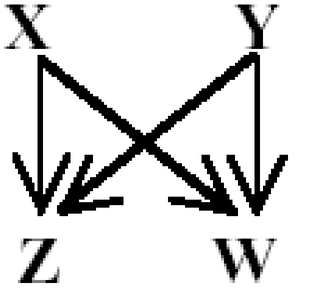


Milo et al. *Science* (2002) *Vol. 298 no. 5594 pp. 824-827*

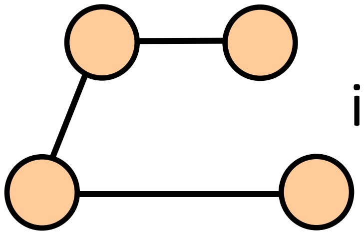
Network motifs

- Biological networks
  - Feed-forward loop
  - Bi-fan motif



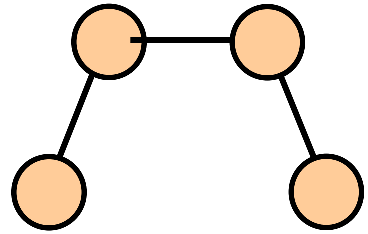


Others ?



h

is isomorphic to



g

Isomorphism

- Bijection, i.e., a one-to-one mapping:

  f : V(G) -> V(H)

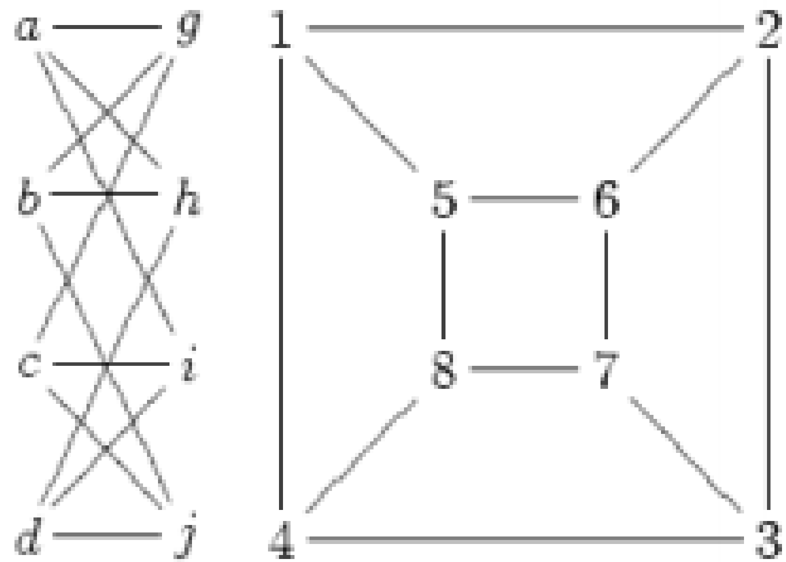
  u and v from G are adjacent if and only if f(u) and f(v) are adjacent in H.
- If an isomorphism can be constructed between two graphs, then we say those graphs are ***isomorphic***.

Isomorphism Problem

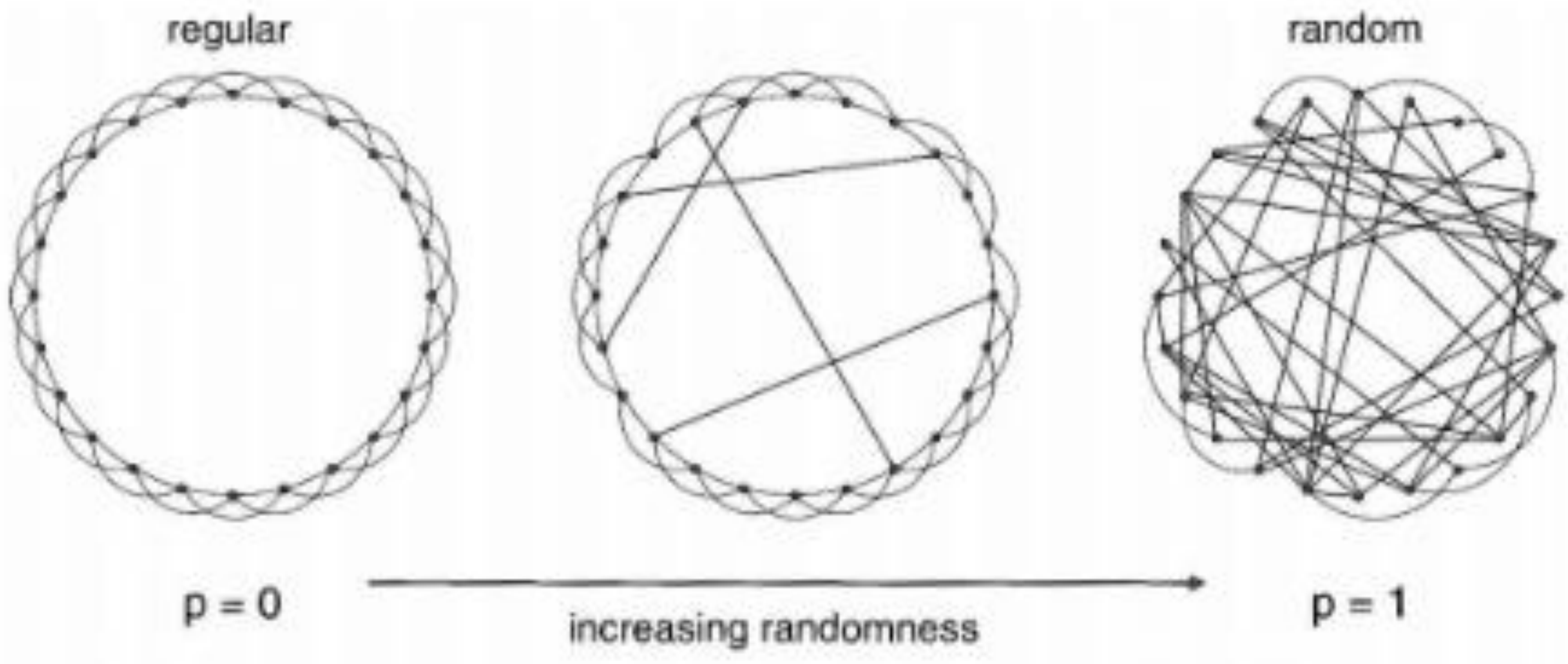
- Determining whether two graphs are isomorphic
- Although these graphs look very different, they are isomorphic; one isomorphism between them is

  f(a)=1 f(b)=6 f(c)=8 f(d)=3

  f(g)=5 f(h)=2 f(i)=4 f(j)=7

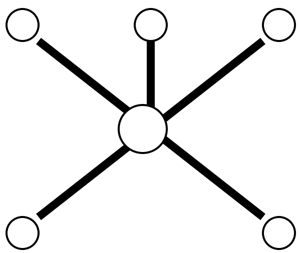


The discovery of the isomorphic subgraphs is a computationally hard task!

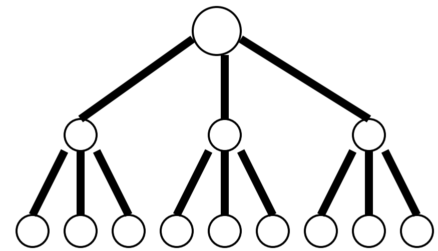

regular
random
p = 0
increasing randomness
p = 1

Regular Network Topologies

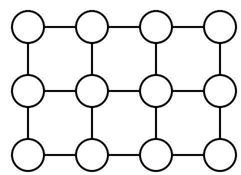
STAR



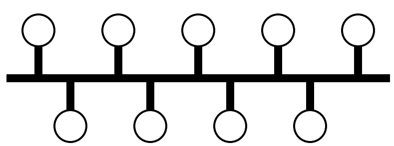
TREE



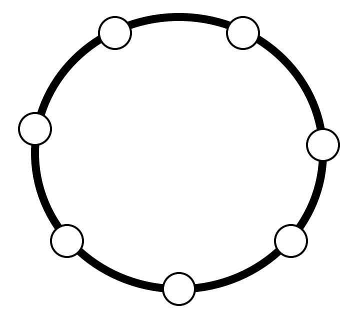
GRID



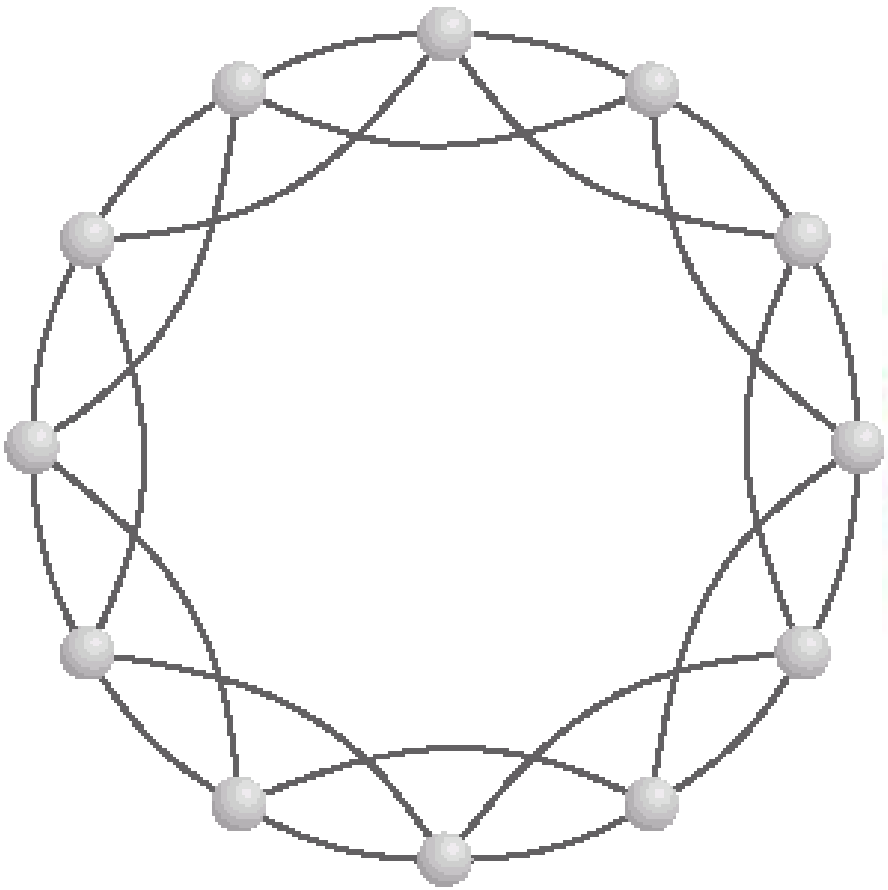
BUS



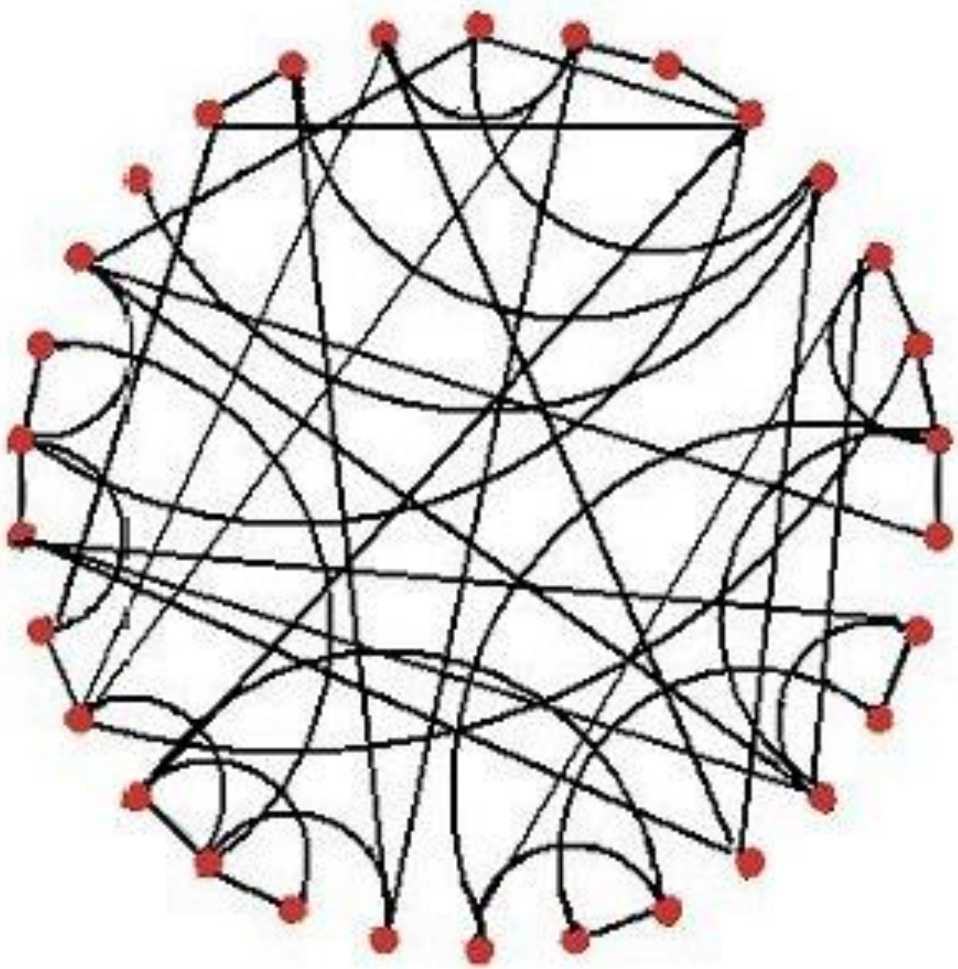
RING



Regular Network



Random Network



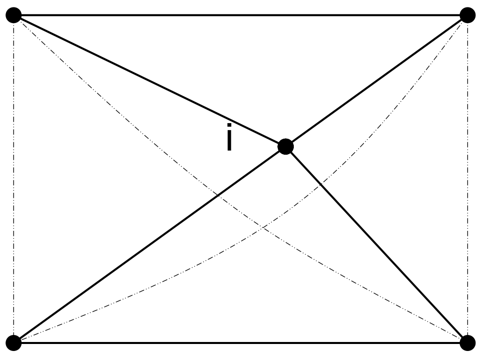
Clustering Coefficient of a Network

- The clustering coefficient characterizes the “connectedness” of the environment close to a node.

$$C_i = \frac{n_i}{\frac{k_i(k_i - 1)}{2}}$$

n_i: number of connections among the neighbors

$k_i(k_i-1)/2$: number of **possible connections** among the neighbors



Clustering coefficient of a network

The average clustering coefficient value $\overline{C}$ reflects **how connected are the neighboring nodes**

$$\overline{C} = \frac{1}{N}\sum_i C_i$$

also shows the **"density" of small loops of length 3**

$\overline{C}$ of a tree is 0

$\overline{C}$ of a fully connected graph (clique) is 1

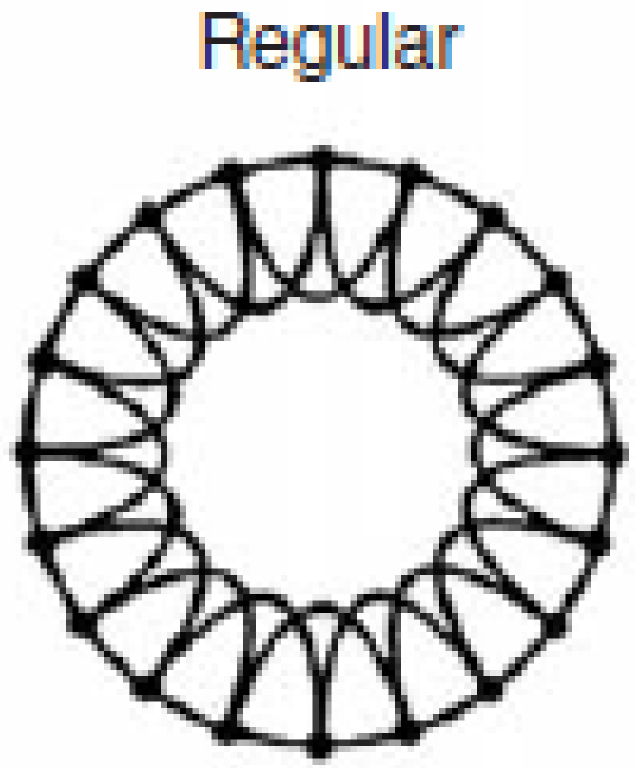
Clustering Coefficient

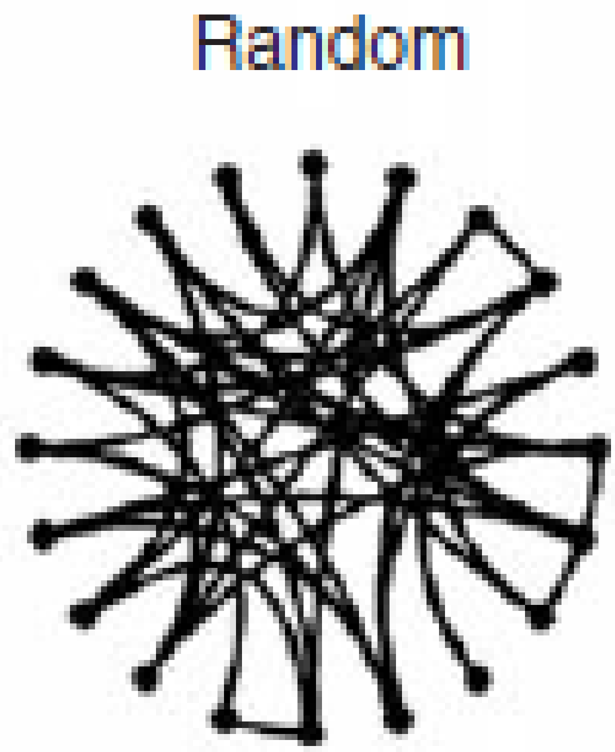
$$C_g^{\triangle} = \frac{3 \text{ x } number\ of\ closed\ triangles}{number\ of\ paths\ of\ length\ 2}$$

Average shortest Path Length

– Smallest number of steps to travel from node u to node v

$$L_g = \frac{2}{n(n-1)} \sum d(u, v)$$





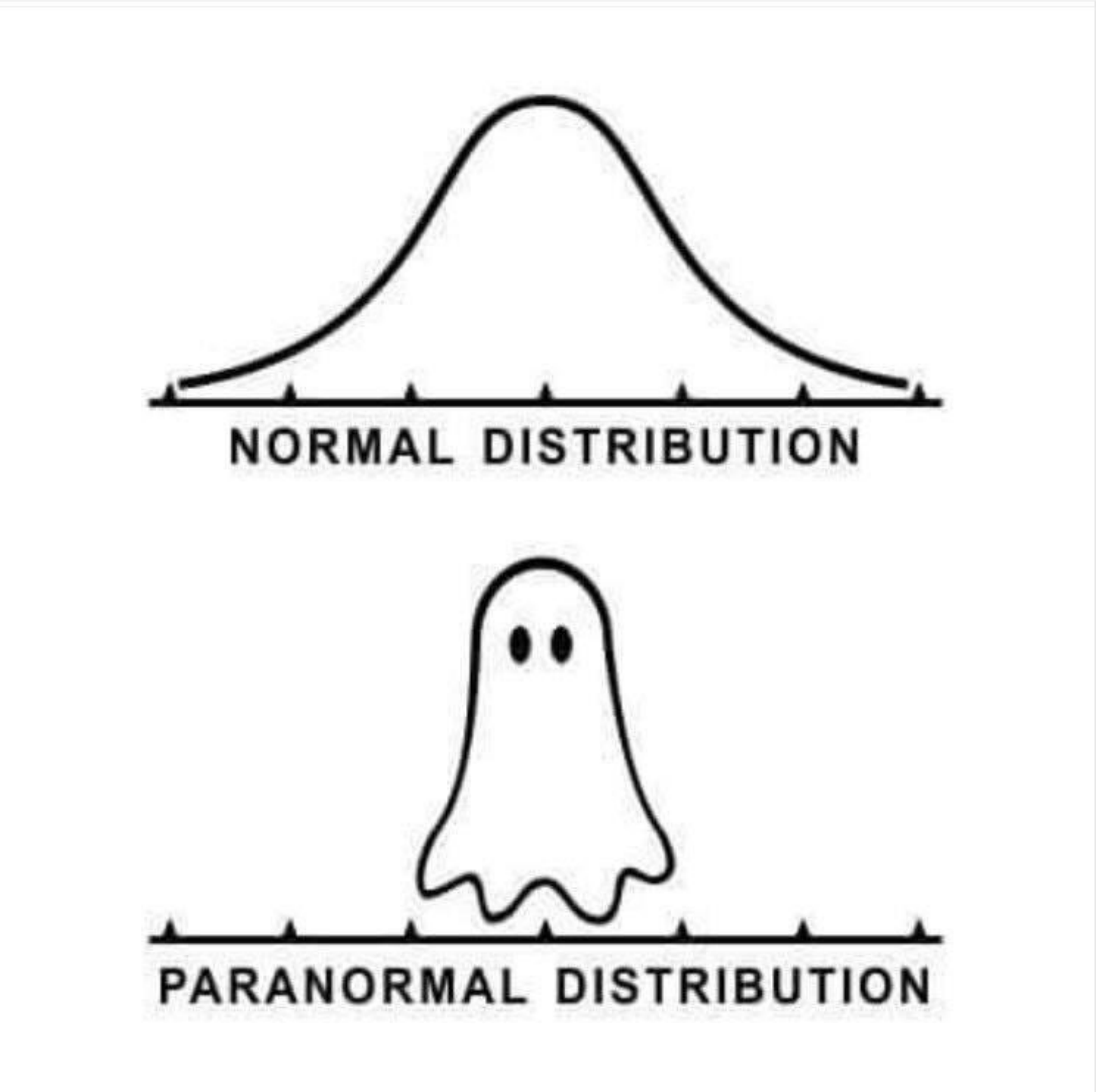
- **High clustering** coefficient
- **High average shortest path length**

- **Low clustering** coefficient
- **Average shortest path length close to one**

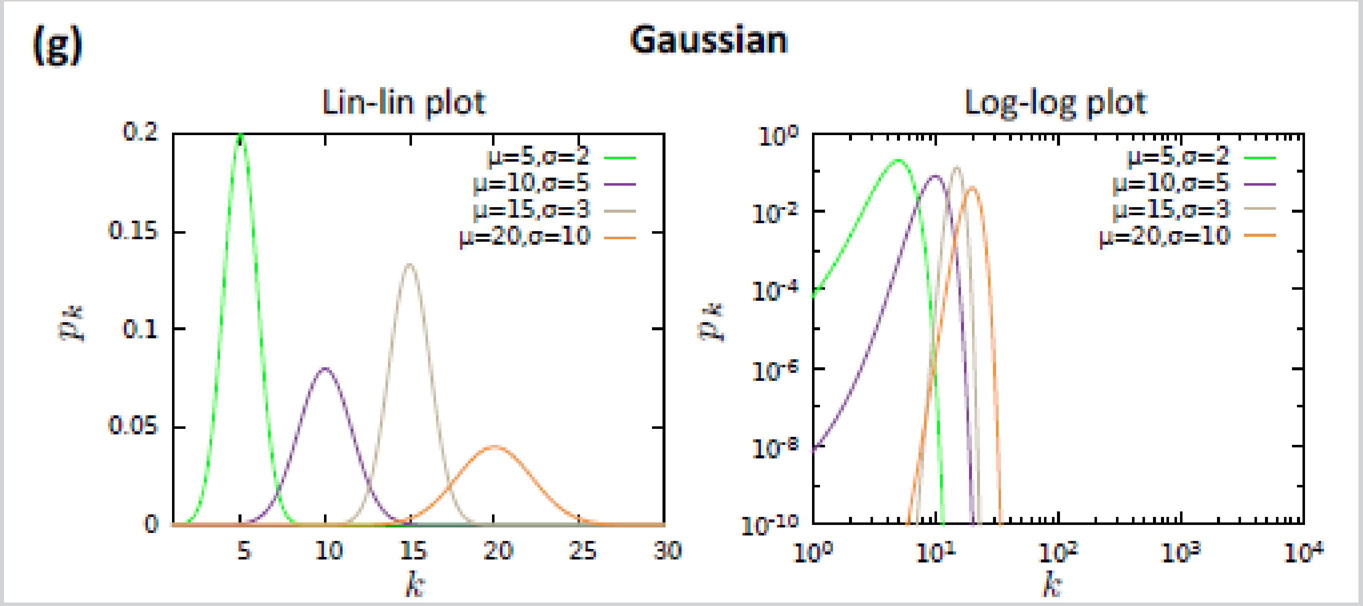
Nearby nodes have a large numbers of interconnections but "distant" nodes have few.

The randomness makes it **less likely that nearby nodes will have lots of connections**, but introduces **more links that connect one part of the network to another**.

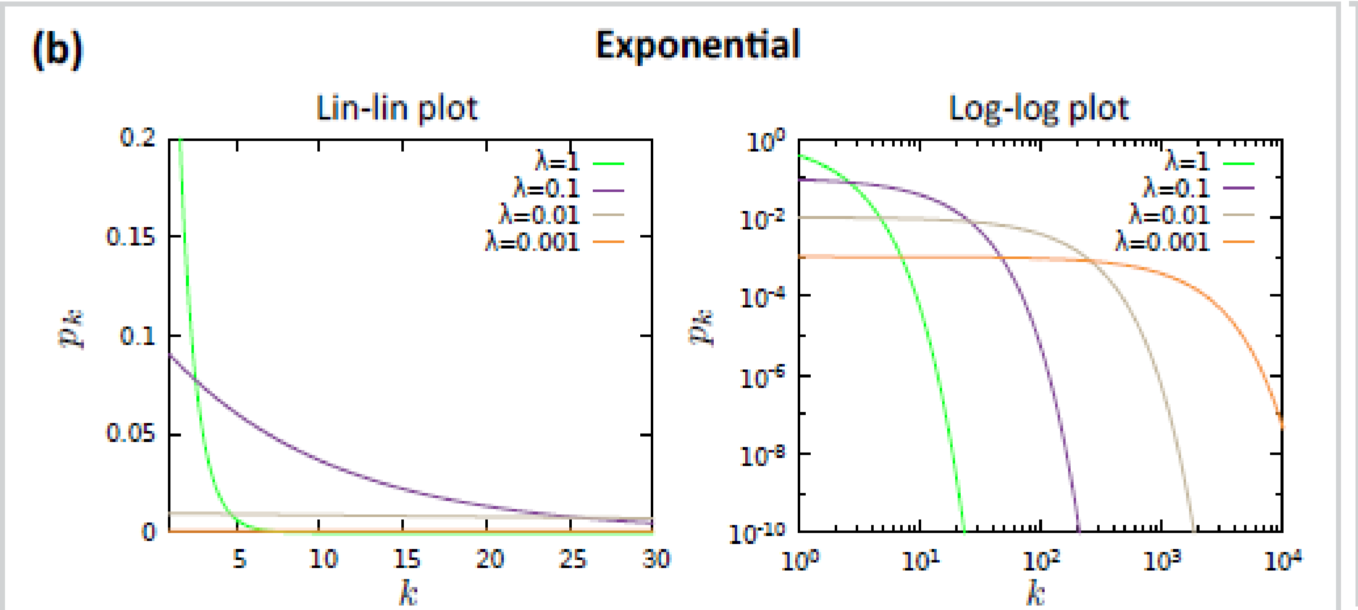
Less variance, more scariance



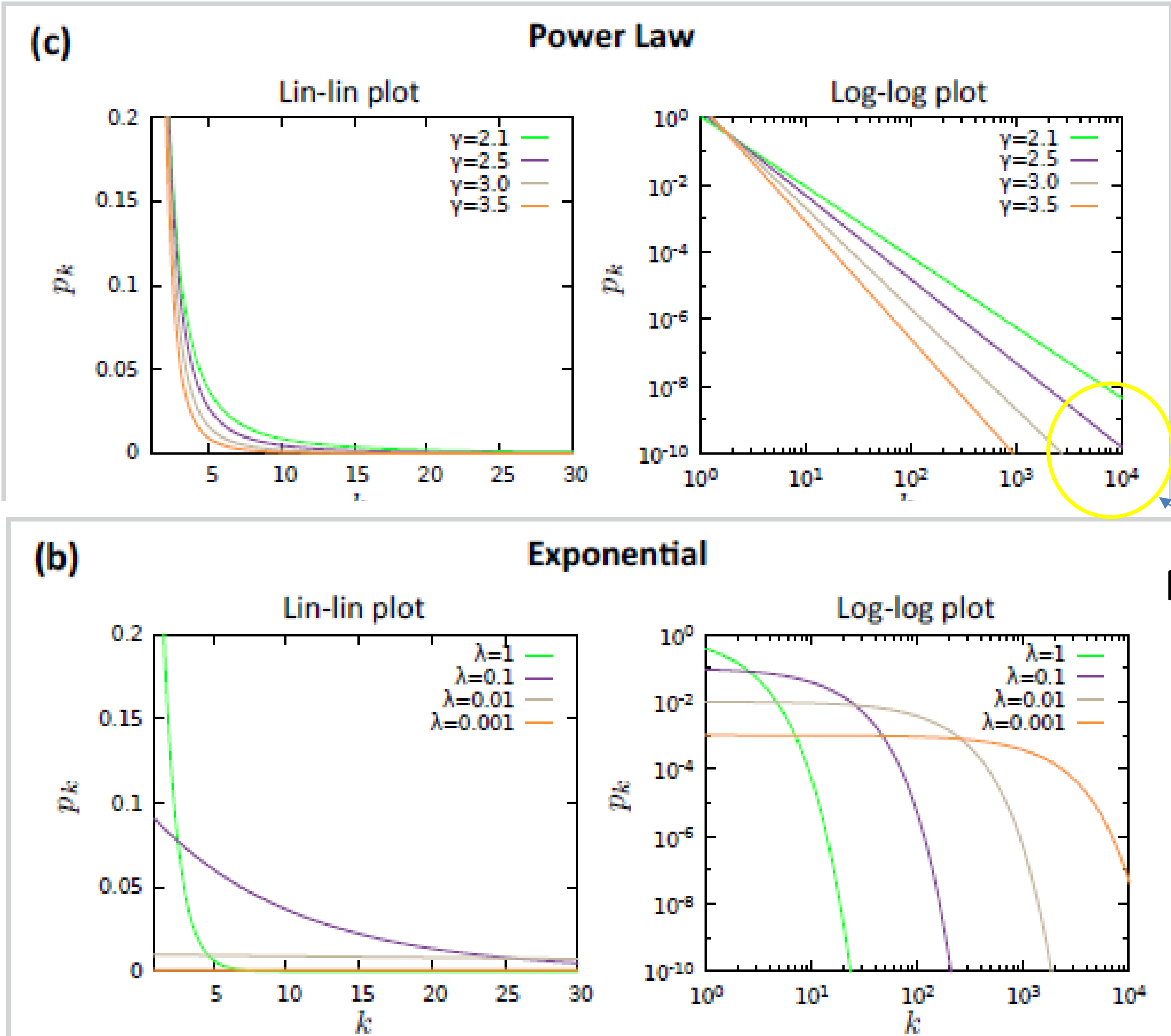
NAME	$p_x/p(x)$	$\langle x \rangle$	$\langle x^2 \rangle$
Poisson (discrete)	$e^{-\mu}\mu^x/x!$	μ	$\mu(1+\mu)$
Exponential (discrete)	$(1-e^{-\lambda})e^{-\lambda x}$	$1/(e^\lambda - 1)$	$(e^\lambda + 1)/(e^\lambda - 1)^2$
Exponential (continuous)	$\lambda e^{-\lambda x}$	$1/\lambda$	$2/\lambda^2$
Power law (discrete)	$x^{-\alpha}/\zeta(\alpha)$	$\begin{cases} \zeta(\alpha-2)/\zeta(\alpha), & \text{if } \alpha > 2 \\ \infty, & \text{if } \alpha \le 1 \end{cases}$	$\begin{cases} \zeta(\alpha-1)/\zeta(\alpha), & \text{if } \alpha > 1 \\ \infty, & \text{if } \alpha \le 2 \end{cases}$
Power law (continuous)	$\alpha x^{-\alpha}$	$\begin{cases} \alpha/(\alpha-1), & \text{if } \alpha > 2 \\ \infty, & \text{if } \alpha \le 1 \end{cases}$	$\begin{cases} \alpha/(\alpha-2), & \text{if } \alpha > 1 \\ \infty, & \text{if } \alpha \le 2 \end{cases}$
Power law with cutoff (continuous)	$\frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} x^{-\alpha} e^{-\lambda x}$	$\lambda^{-1} \frac{\Gamma(2-\alpha)}{\Gamma(1-\alpha)}$	$\lambda^{-2} \frac{\Gamma(3-\alpha)}{\Gamma(1-\alpha)}$
Normal (continuous)	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$	μ	$\mu^2 + \sigma^2$



Most frequently encountered distribution



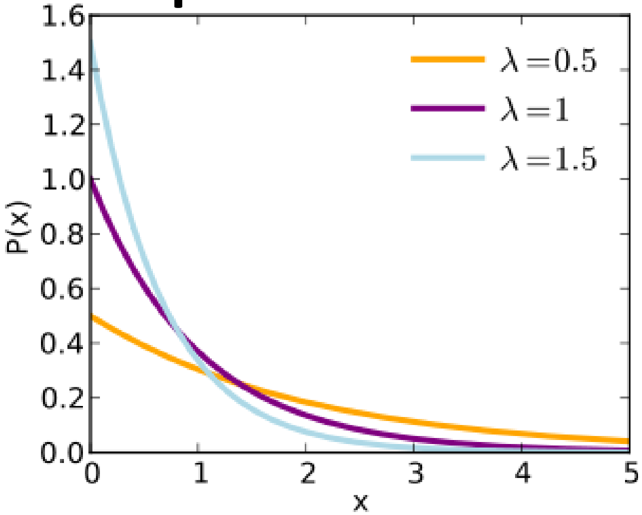
Describes the degree of distribution of a **random network**



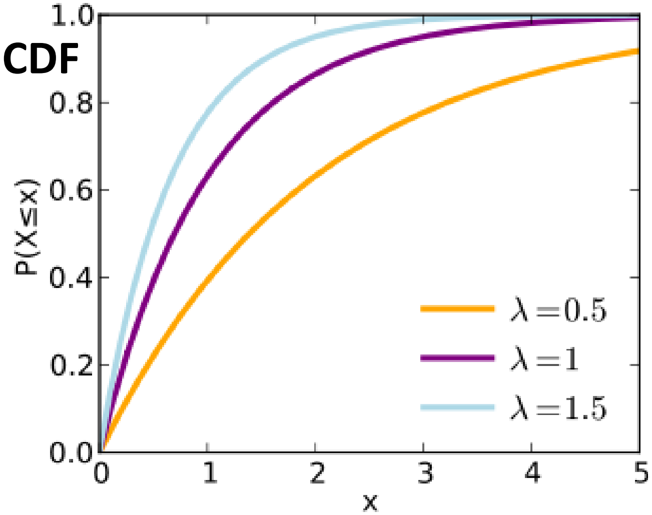
Heavy tailed:
Whose decay a
large x
is slower than
exponential

Rare events

Exponential Distribution



CDF

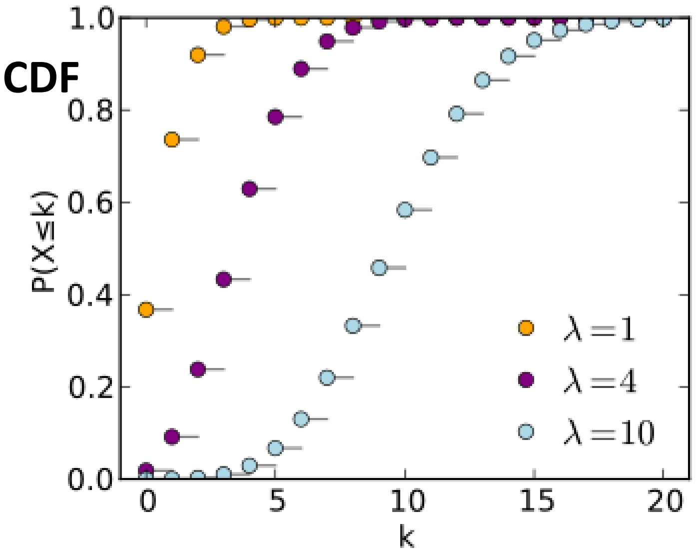


PDF: $\lambda e^{-\lambda x}$

Cumulative Distribution Function (CDF) $\mathbf{1 - e^{-\lambda x}}$

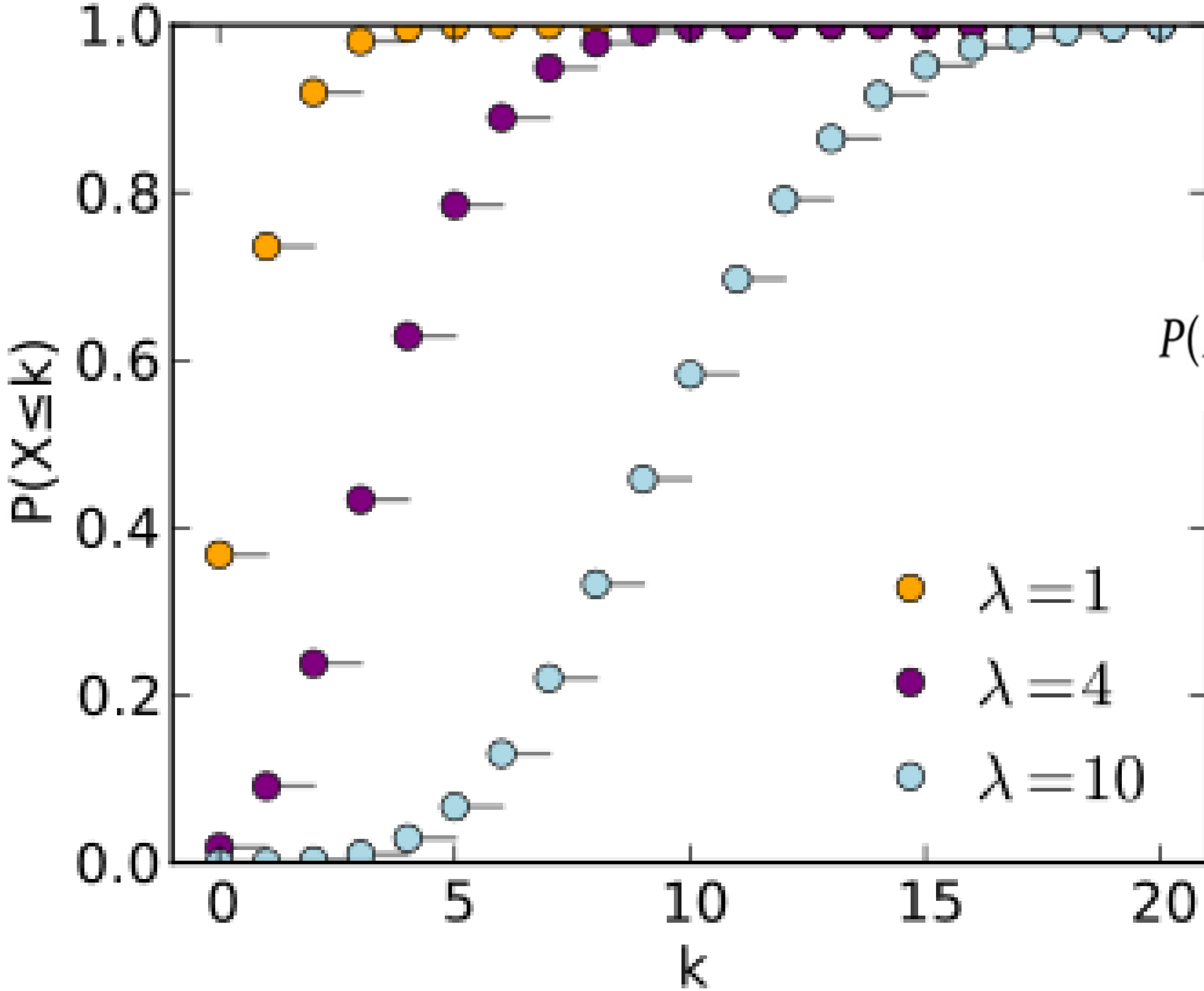
The exponential distribution describes the time between events in a **Poisson process**

CDF



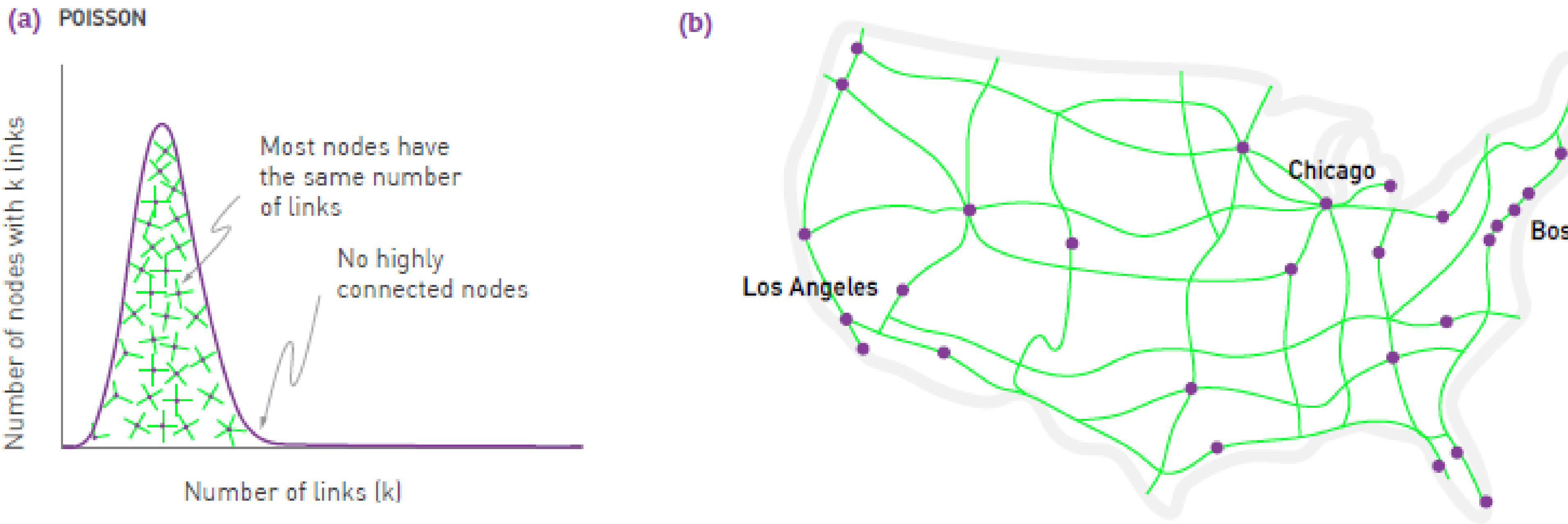
Number of occurrence: index k
The CDF is discontinuous at the integers of k
λ: expected number of occurrences
PMF: $\lambda^k e^{-\lambda} / k!$

Poisson Distribution
Cumulative Distribution Function (CDF)



$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

The CDF is discontinuous at the integers of *k* and flat everywhere else because a variable that is Poisson distributed takes on only integer values.

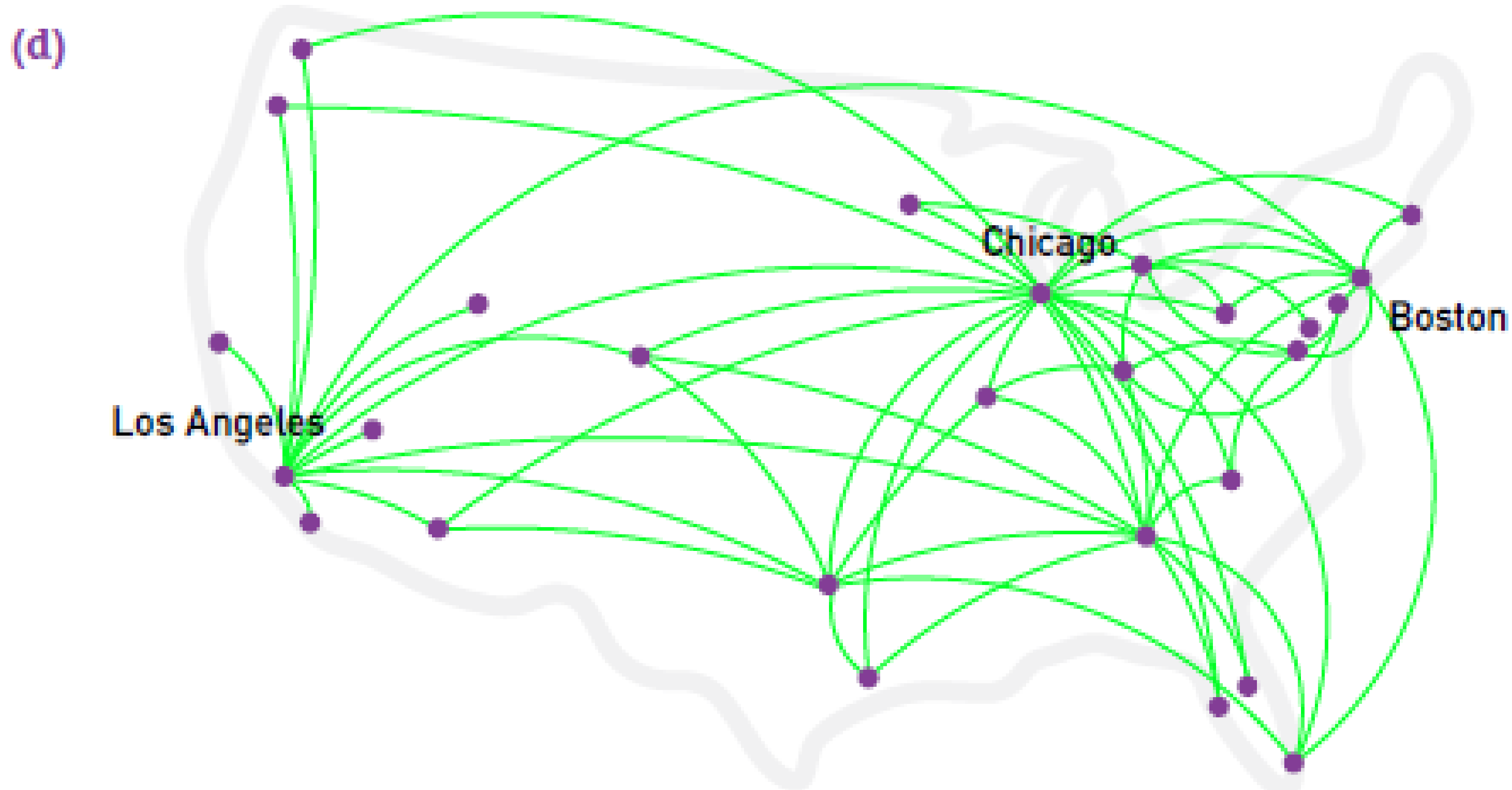


National highway network

Nodes are cities, links are major highways

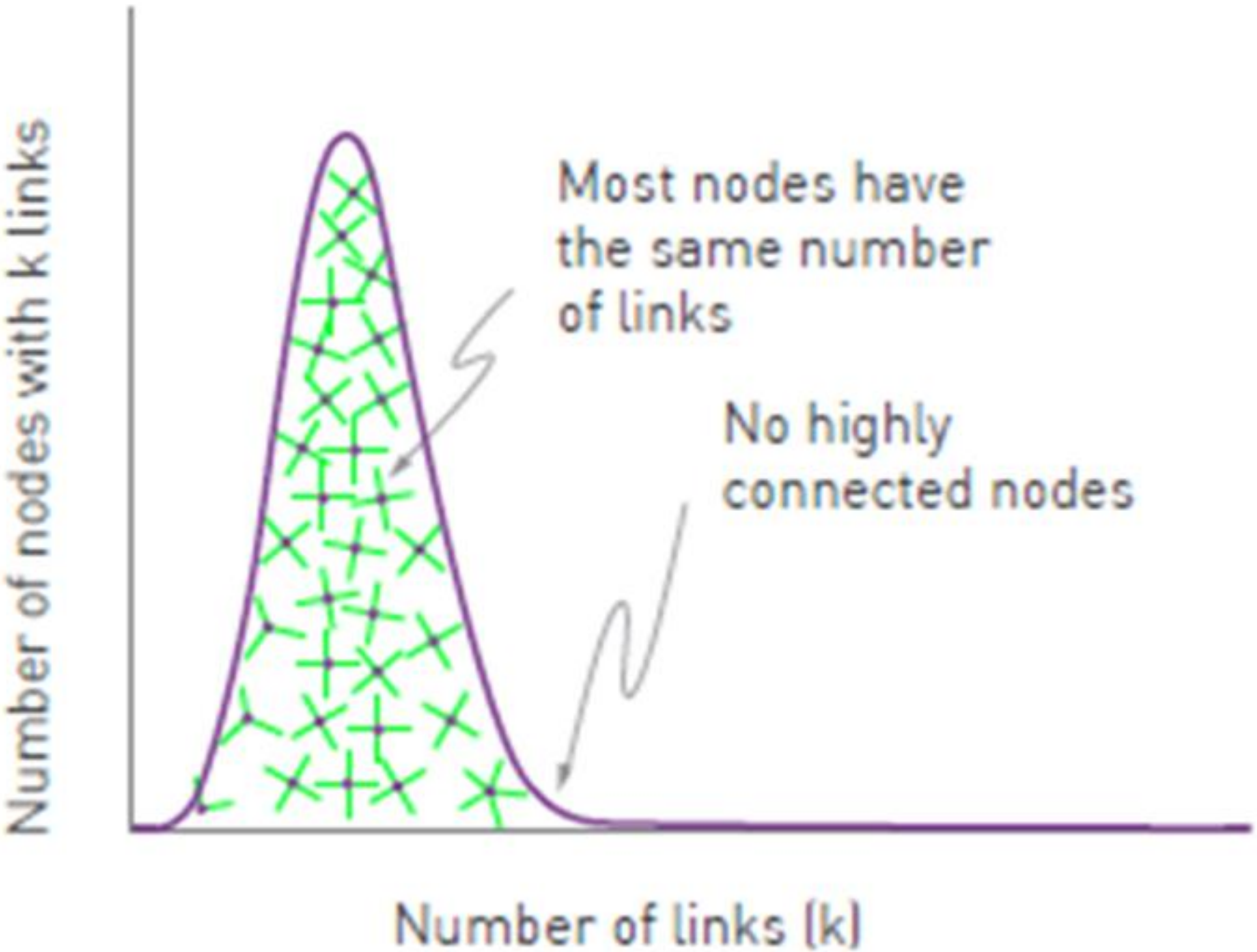
No cities with hundreds of highways

No city disconnected from the highway system

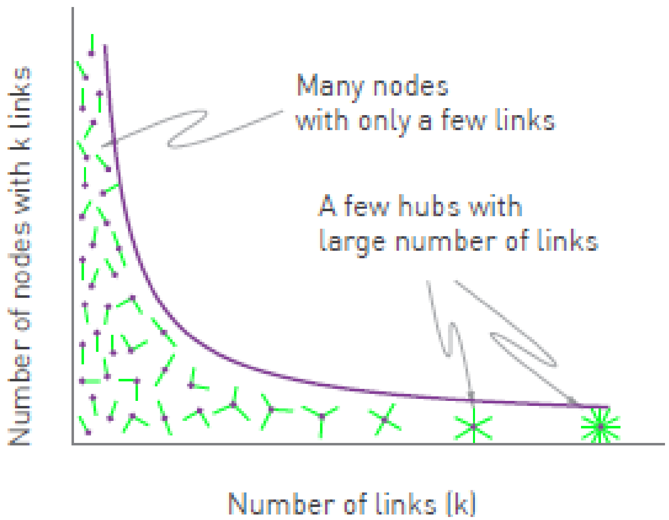


Nodes are airports & links are direct flights between them.
Most airports have only a few flights.
Yet, we have a few very large airports, acting as major hubs, connecting many smaller airports.
Once hubs are present, they change the way we navigate the network. E.g. if we travel from Boston to Los Angeles by car, we must drive through many cities. On the airplane network, however, we can reach most destinations via a single hub, like Chicago.

Random Networks have a degree of connectivity that follows Poisson distribution

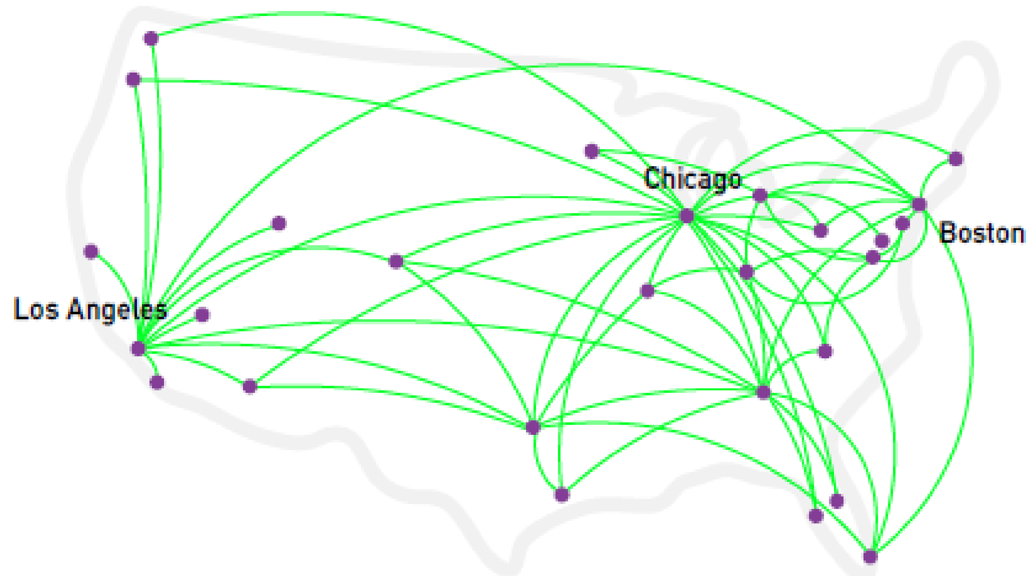


(c) POWER LAW



Many nodes
with only a few links
A few hubs with
large number of links
Number of nodes with k links
Number of links (k)

(d)



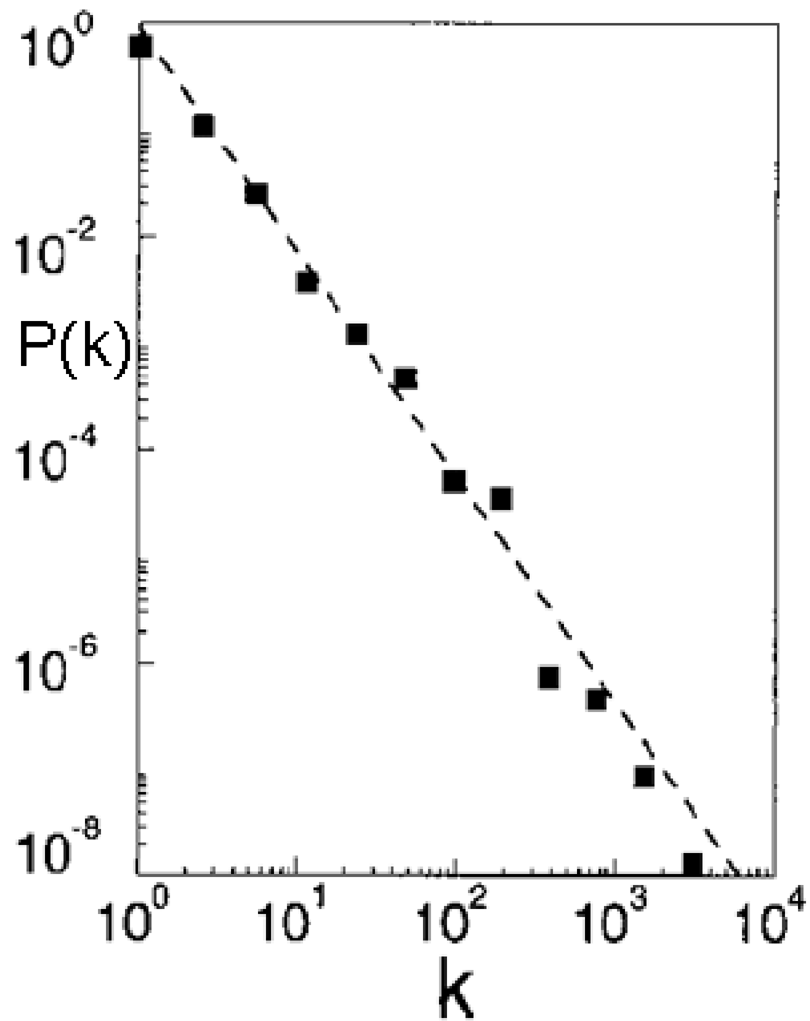
Chicago
Boston
Los Angeles

Power Law Distribution

In a log-log scale, data points form an approximate straight line, suggesting that the distribution is well approximated with

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

degree exponent γ



The 80/20 Rule & the top one percent

- A few wealthy individuals earned most of the money, while the majority of the population earned rather small amounts
- Incomes follow a power law
- 80/20 rule: Roughly 80 percent of money is earned by only 20 percent of the population
- US 1% of the population earns a disproportionate 15% of the total US income

Vilfredo Federico Damaso Pareto (1848 – 1923)

The emergence of the 80/20 rule in various areas:

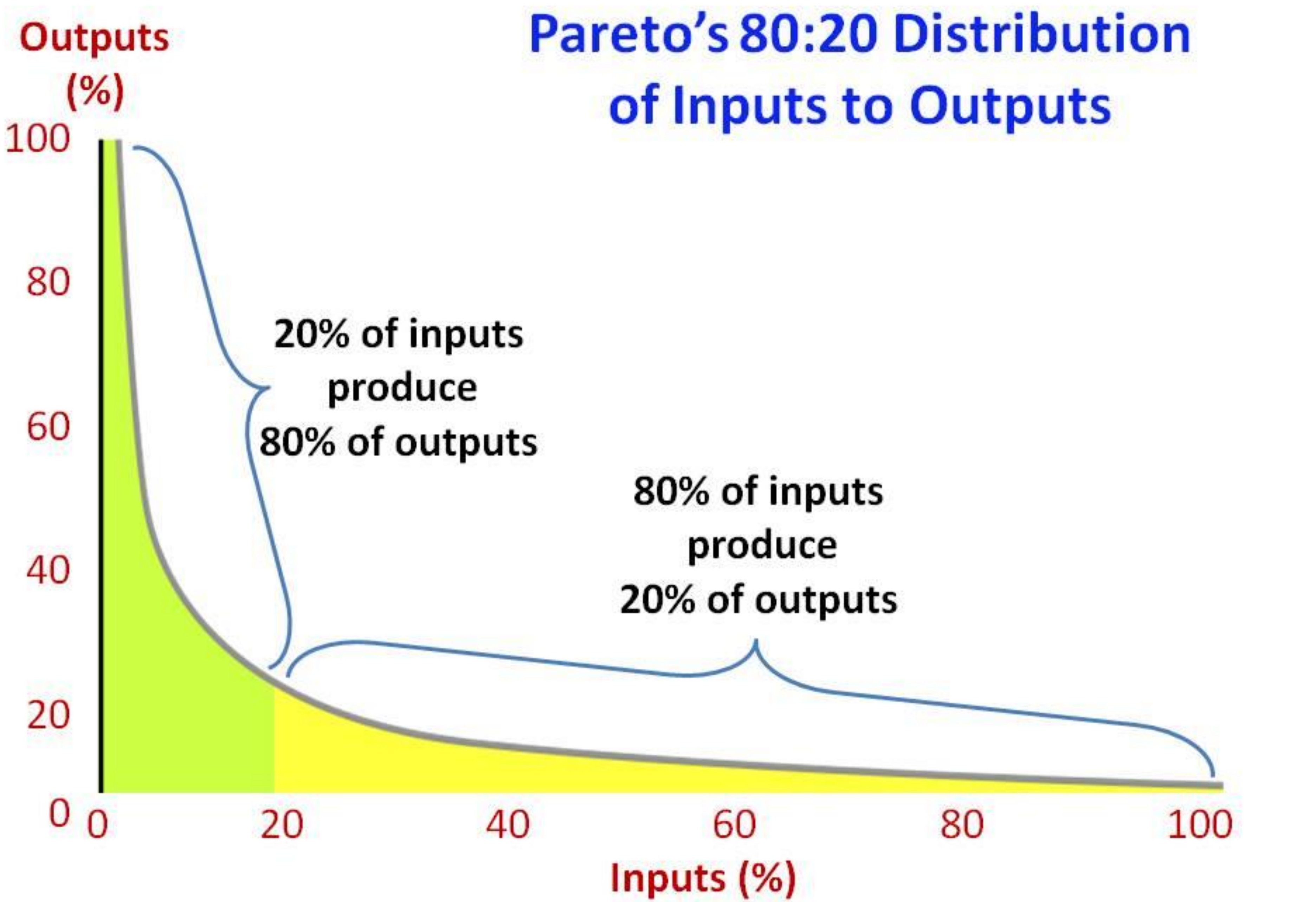
Management

i. 80% of profits are produced by only 20% of the employees
ii. 80% of decisions are made during 20% of meeting time

Networks

i. 80% of links on the Web point to only 15% of webpages
ii. 80% of citations go to only 38% of scientists
iii. 80% of links in Hollywood are connected to 30% of actors

Most quantities following a power law distribution obey the 80/20 rule



Pareto's 80:20 Distribution
of Inputs to Outputs
Outputs
(%)
100
80
60
40
20
0
20% of inputs
produce
80% of outputs
80% of inputs
produce
20% of outputs
0
20
40
60
80
100
Inputs (%)

Internet

- Link between routers in Boston and Budapest would require to lay a cable between North America and Europe: prohibitively expensive
- The degree distribution of the Internet is well approximated by a power law
- Few high-degree routers hold together a large number of routers with only a few links

History: first map of the WWW

Objective: **To understand the structure of the network behind it.**

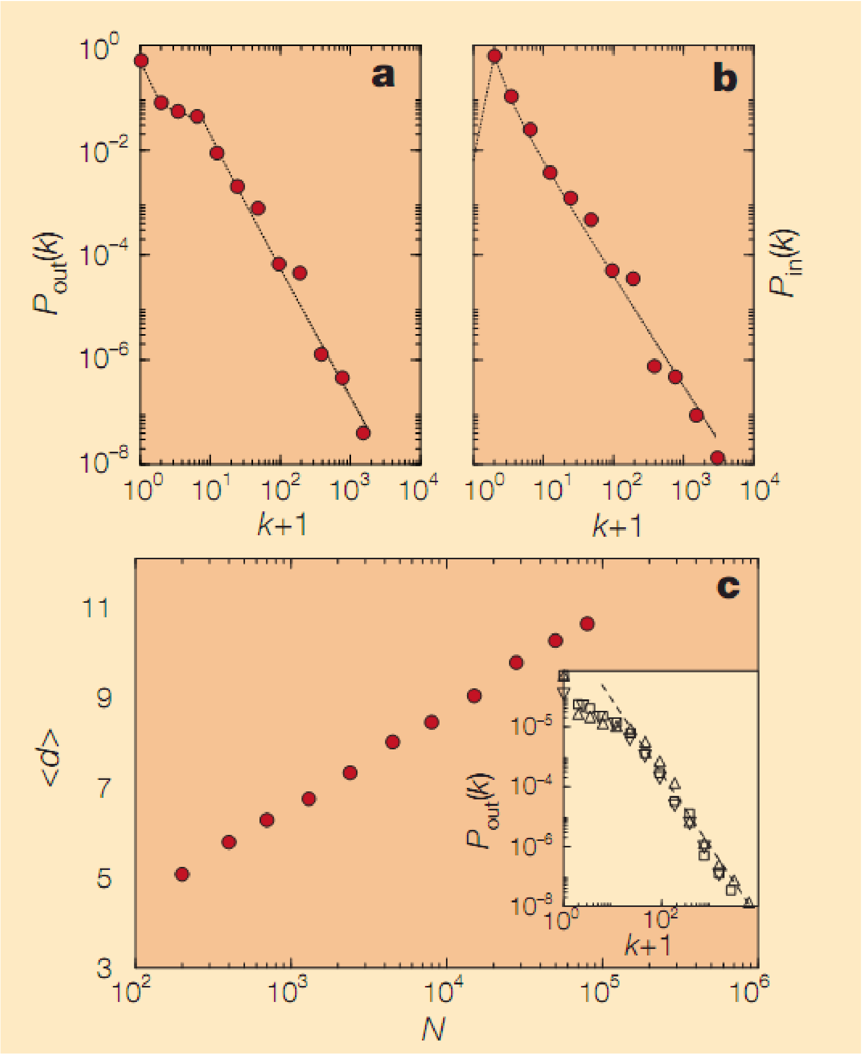
- Generated by Hawoong Jeong at University of Notre Dame.
- Mapped out the nd.edu domain, consisting of about 300,000 documents and 1.5 million links.
- Compared the properties of the Web graph to the random network model.

The web played an important role in the development of network theory.

- WWW: network whose **nodes are documents** & **links are the URLs**
- With an estimated size of over one trillion documents ($N \approx 10^{12}$), the Web is the largest network humanity has ever built
- Exceeds in size even the human brain ($N \approx 10^{11}$ neurons)

Standard testbed for most network measures

WWW has power-law degree distribution



The degree distribution scales as a power-law

Outgoing links

The tail of the distributions follows $P(k) \approx k^{-r}$, with $r^{out}=2.45$

Incoming links: $r^{in}=2.1$

Average of the shortest path between two documents as a function of system size

R. Albert, H. Jeong, A.-L. Barabási, Nature 401, 130 (1999)

WWW follows a power law

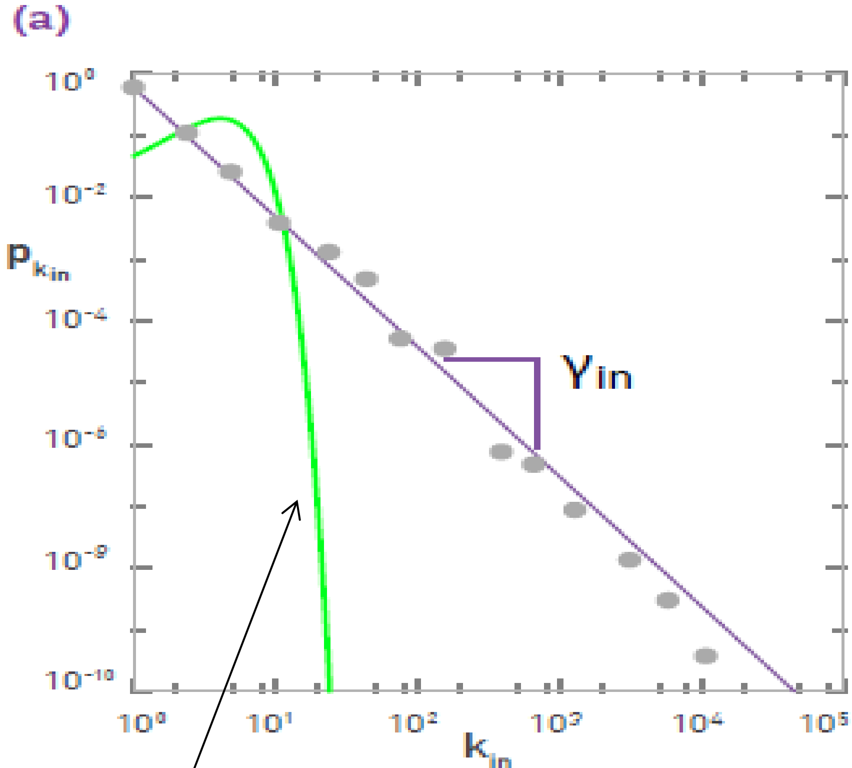
- If the WWW were to be a random network, the degrees of the Web documents would follow a Poisson distribution
- **Poisson form offers a poor fit for the WWW's degree distribution**
- Instead of a log-log scale data points form an approximate straight line, suggesting that the degree distribution of the WWW is well approximated with

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

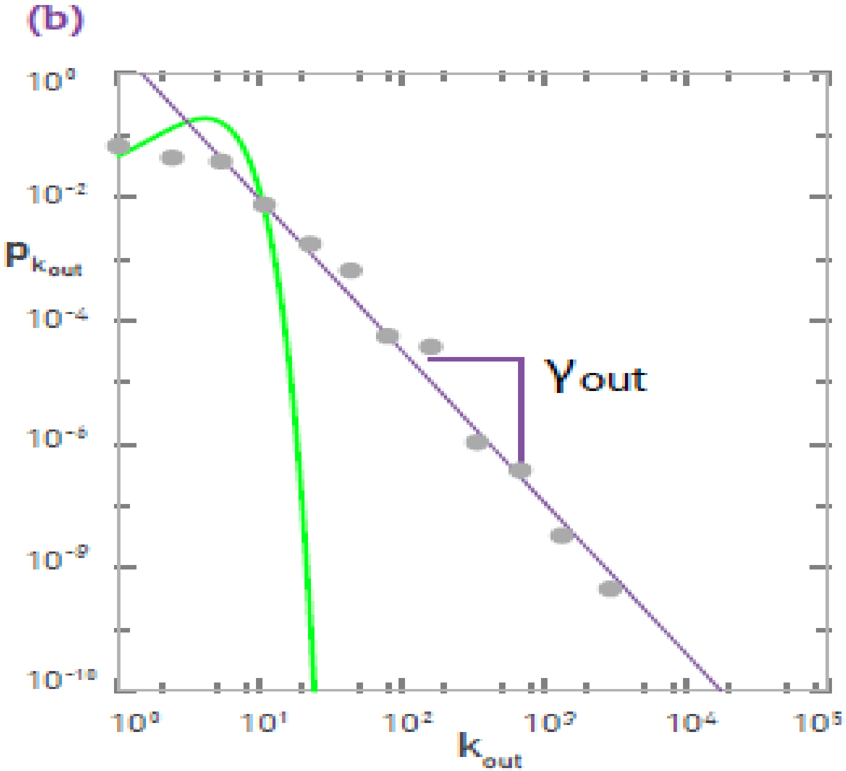
$$\log p_k \sim -\gamma \log k.$$

Power law distribution *(*exponent γ is its *degree exponent)*

The Degree Distribution of the WWW



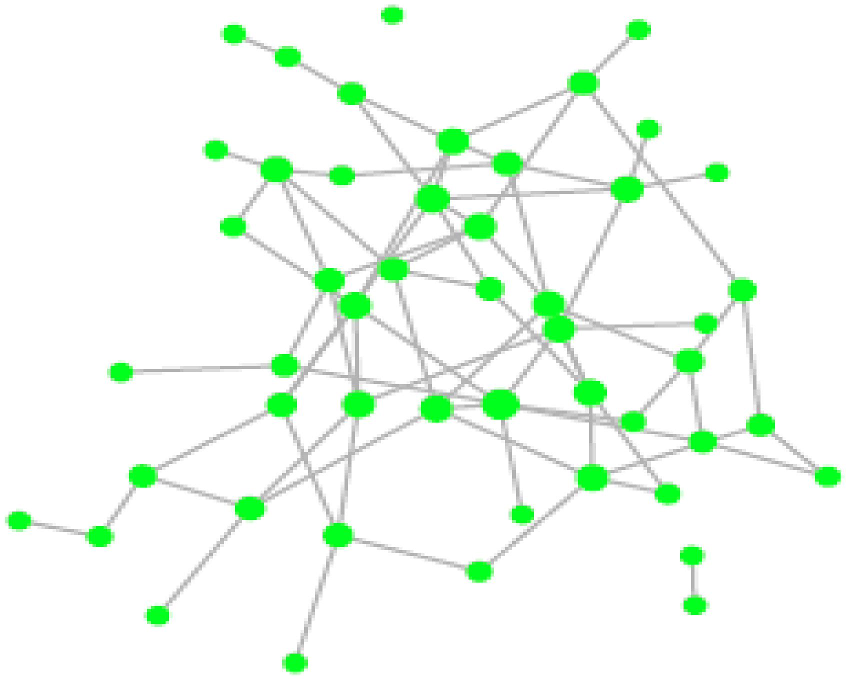
The incoming degree distribution



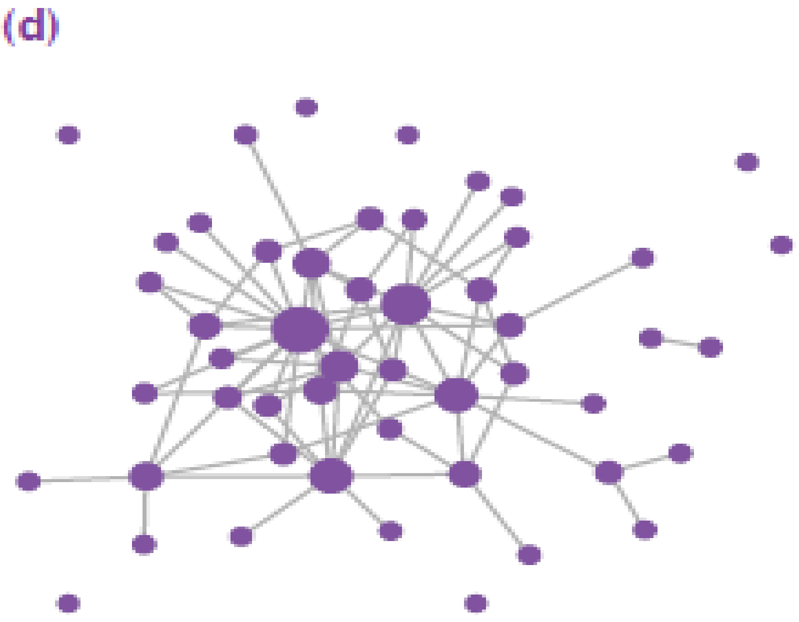
The outgoing degree distribution

Degree distribution predicted by a Poisson function with the average degree $\langle k_{in} \rangle = \langle k_{out} \rangle = 4.60$ of the WWW sample (green line).

Poisson vs. Power-law Distributions



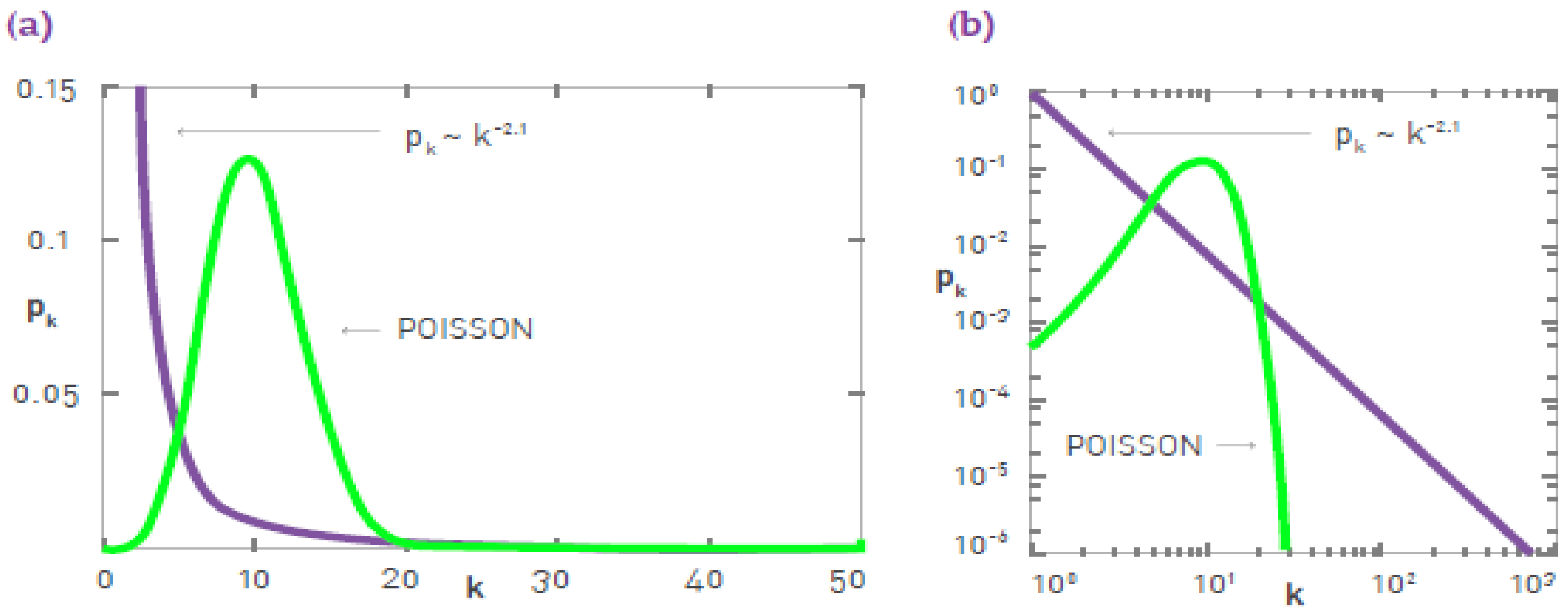
A random network with $\langle k \rangle$= 3 & N = 50, illustrating that **most nodes have comparable degree** $k \approx \langle k \rangle$.



A scale-free network with γ=2.1 & $\langle k \rangle$= 3, illustrating that **numerous small-degree nodes** coexist with **a few highly connected hubs**.

The size of each node is proportional to its degree.

Poisson vs. Power-law Distributions



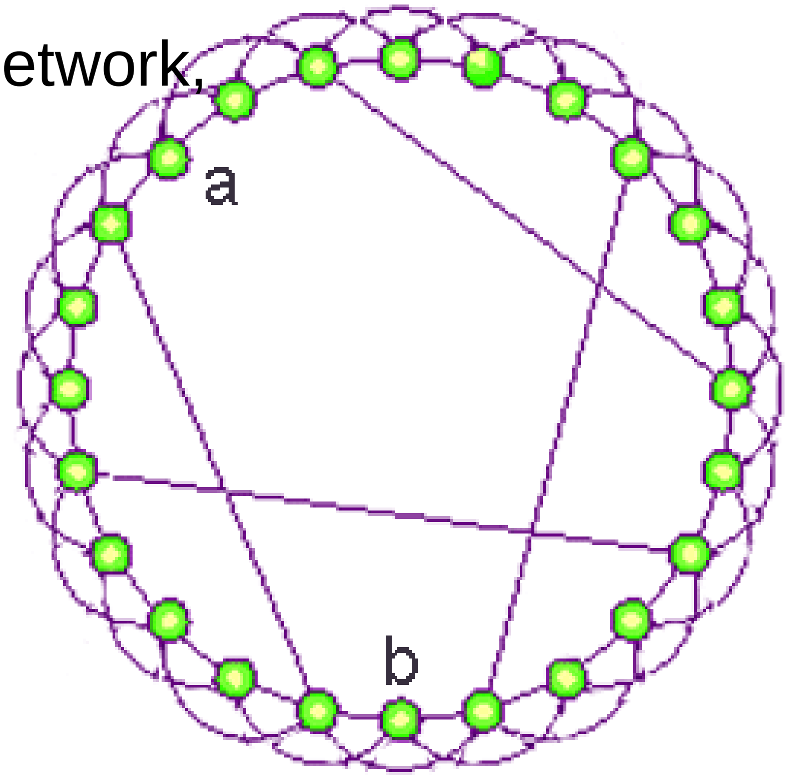
Comparing a Poisson function with a power-law function ($\gamma = 2.1$) on a linear plot.
Both distributions have $\langle k \rangle = 11$

Generation of small-world networks

A small-world network can be generated from a regular one by

1. **randomly disconnecting a few points,** &
2. **randomly reconnecting them elsewhere**.

For the creation of this small world network, some '**shortcut**' links are added to a regular network.



Shortcuts because they allow travel from node a to node b to occur in only 3 steps, instead of 5 without the shortcuts.

Small-World Phenomenon

Any two nodes of a complex & **high clustered network** would be connected by **a relatively small paths distances**.

Watts & Strogatz define simple network models by **rewiring regular lattice networks** with a probability

Such networks have:

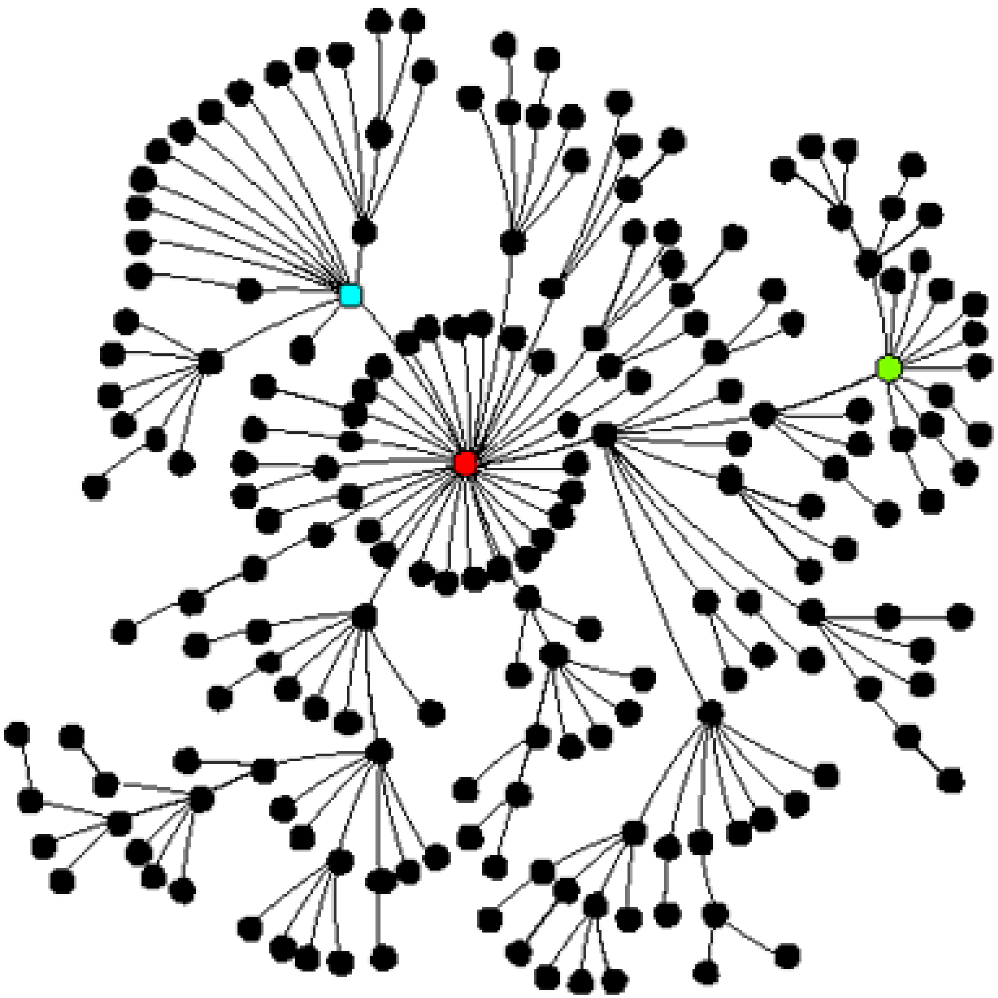
- **Highly clustered** like lattice
- **Very small path length** like random graphs

Scale-free Networks

Barabasi *et al.* found that the structure of the **WWW** **did not** conform to the then-accepted model of **random connectivity**.
Instead, their experiment yielded a connectivity that they named "scale-free."

Scale-free means there is **no characterizing degree** in the network

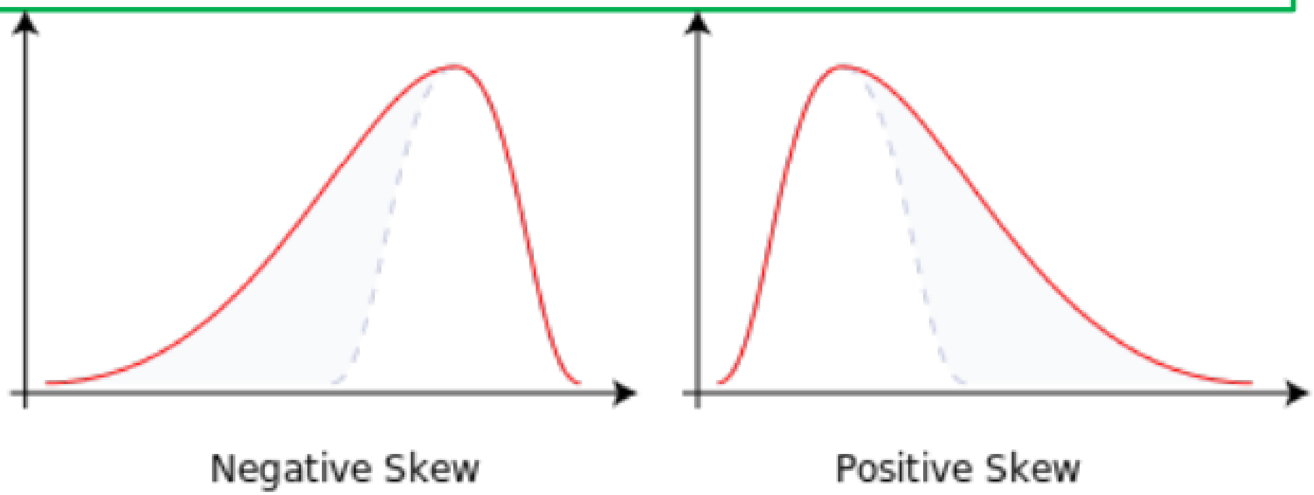
In a scale-free network, the **characteristic clustering** is maintained even as the networks themselves grow arbitrarily large.



THE MEANING OF SCALE-FREE

Let's first talk about **moments**!

- n^{th} moment $\langle k^n \rangle = \sum_{k_{min}}^{\infty} k^n p_k \approx \int_{k_{min}}^{\infty} k^n p(k) dk.$
- n=1, **mean** $\langle k \rangle$
- n=2, ***variance*** $\sigma^2 = \langle k^2 \rangle$
- n=3: **skewness** (how symmetric is the distribution around the mean)



The meaning of scale-free

This divergence indicates that **fluctuations around the average** can be **arbitrary large**.
A degree of a randomly selected node, could be tiny or arbitrarily large.
Hence networks with $\gamma < 3$ do **not** have a **meaningful internal scale**, but are **"scale-free"**

In a random network, nodes have comparable degrees:
The average degree serves as the scale of a random network

Small world vs. Scale-free Networks

Often small-world networks are also **scale-free.**

Some **small-world networks** of **modest size** do **not follow a power law but are exponential**.

This point can be significant when trying to understand the rules that underlie the network.

Random networks vs. Scale-free networks

- Main difference between a random vs. a scale-free network comes in the ***tail* of the degree distribution**
- For small *k,* power law is above Poisson function, indicating that a **scale-free network has a large number of small degree nodes**, most of which are absent in a random network
- For ***k* in the vicinity of ⟨*k*⟩,** Poisson distribution is above power law, indicating that in a random network there is an excess of nodes with degree $k \approx \langle k \rangle$

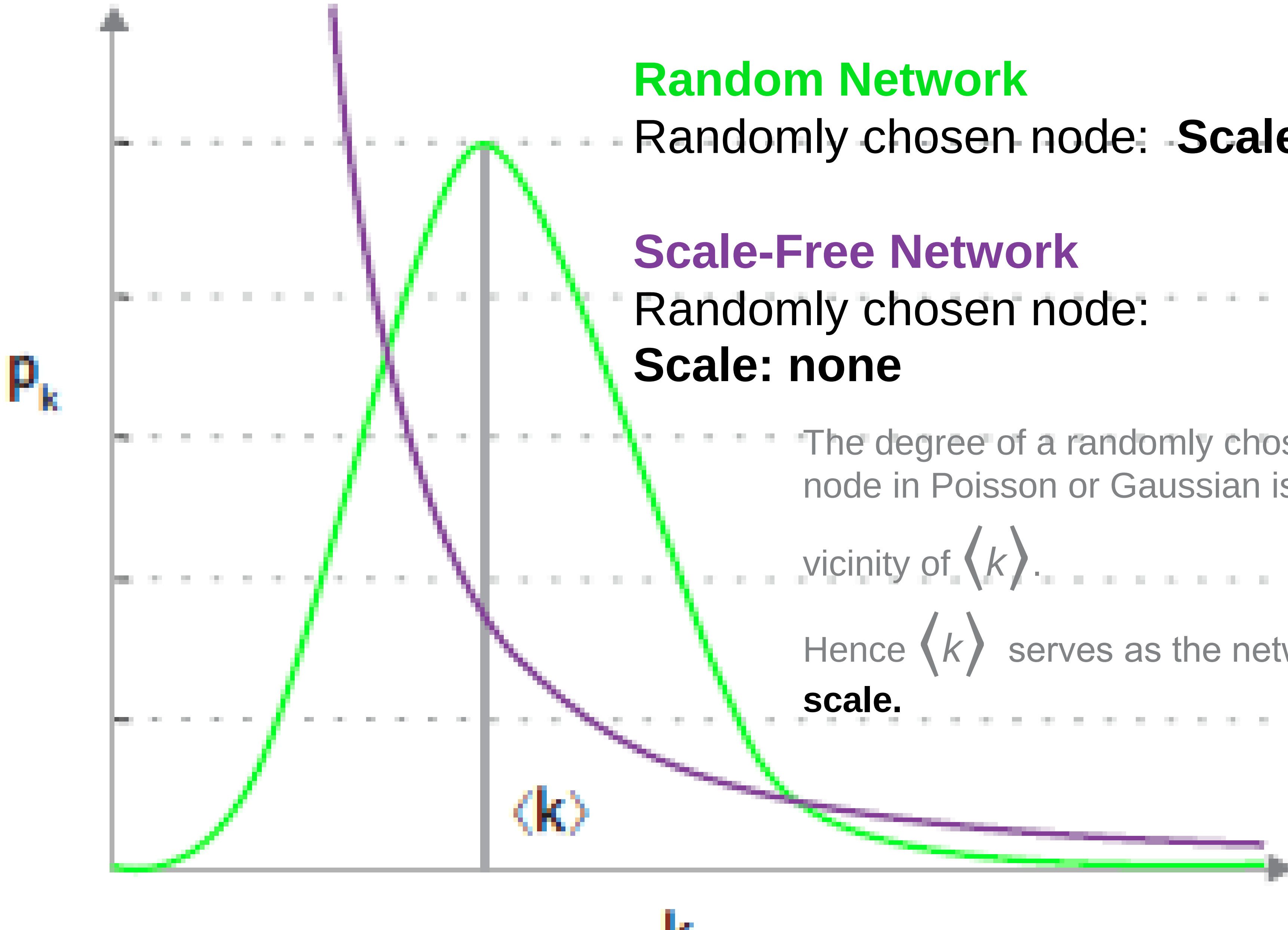
Scale-free Networks Lack a Scale

- Scale-free name captures the lack of an internal scale, a consequence of the fact that nodes with **widely different degrees coexist in the same network**

This feature distinguishes scale-free networks from

- lattices, in which **all nodes have exactly the same degree (σ = 0),**
- random networks, whose degrees vary in **a narrow range** (σ = $\langle k^{1/2} \rangle$)

Scale-free Networks Lack a Scale



Random Network

Randomly chosen node: **Scale: ⟨*k*⟩**

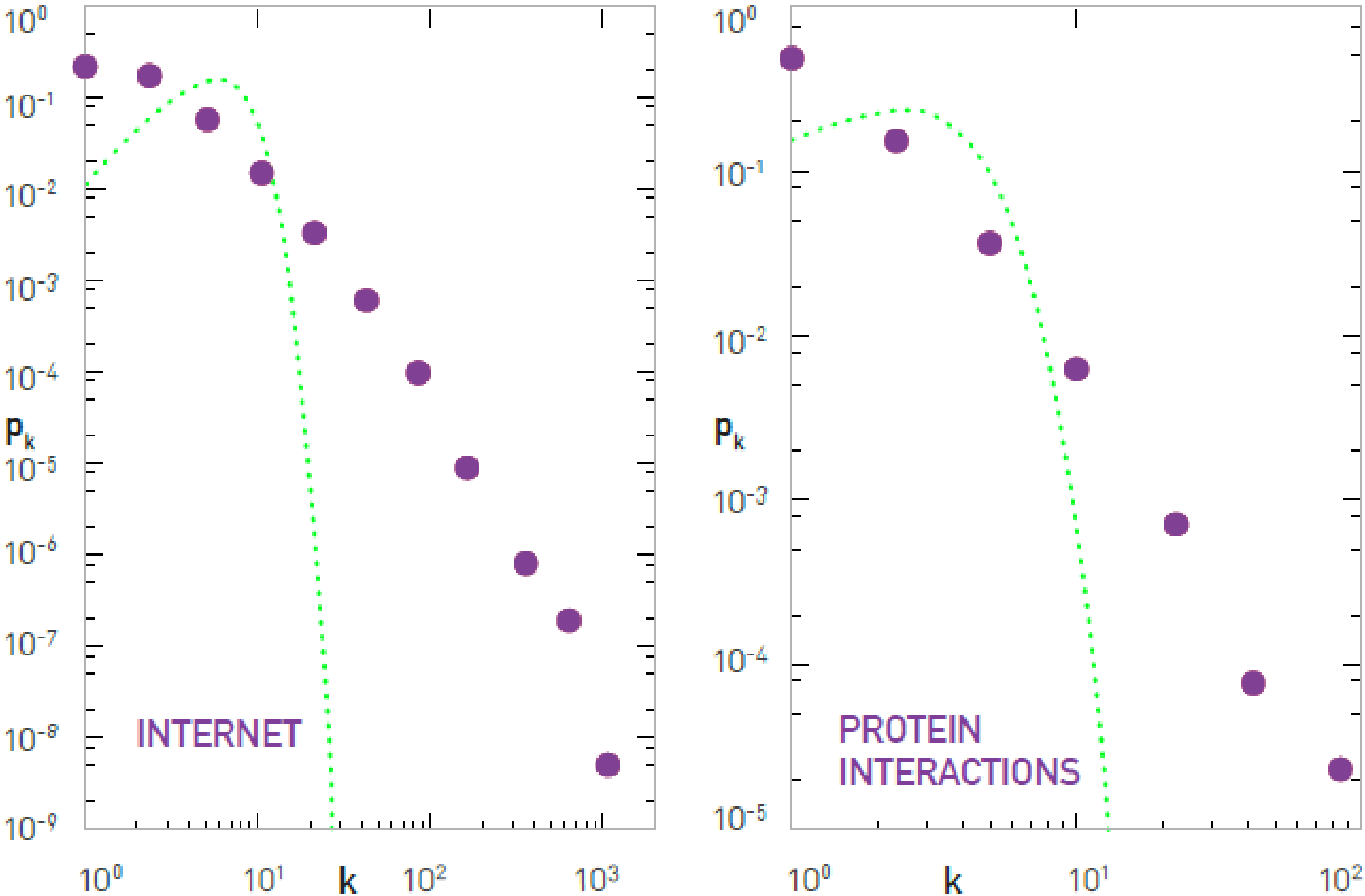
Scale-Free Network

Randomly chosen node:
Scale: none

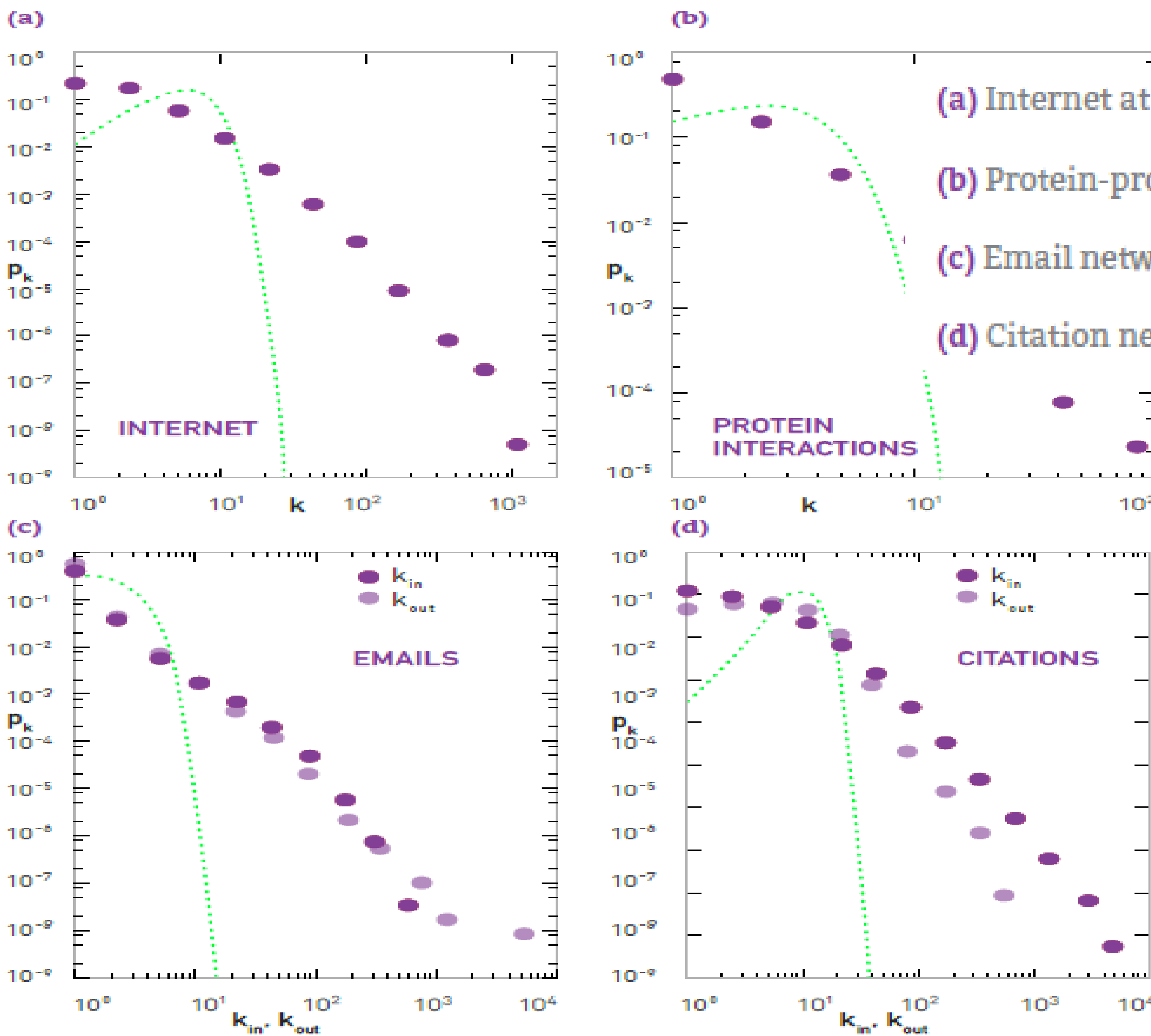
The degree of a randomly chosen node in Poisson or Gaussian is in the vicinity of $\langle k \rangle$.

Hence $\langle k \rangle$ serves as the network's **scale.**

Degree of Distribution of two Scale-free Networks



Examples of Scale-free Networks



(a) Internet at the router level.

(b) Protein-protein interaction network.

(c) Email network.

(d) Citation network.

THE MEANING OF SCALE-FREE (con’td)

For a scale-free network, the ***n*-th moment** of the degree distribution is

$$\langle k^n \rangle = \int_{k_{min}}^{k_{max}} k^n p(k) dk = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1}.$$

- For many scale-free networks, the degree exponent $\gamma \in [2, 3]$.
- For these in the $N \to \infty$ limit, the mean is finite, but the 2^{nd} & higher moments (e.g., $\langle k^2 \rangle$, $\langle k^3 \rangle$) go to infinity

This divergence indicates that **fluctuations around the average** can be **arbitrary large**.
A degree of a randomly selected node, could be tiny or arbitrarily large.
Hence networks with $\gamma < 3$ do **not** have a **meaningful internal scale**, but are **“scale-free”**

Random Networks Have a Scale

- For a random network with a Poisson degree distribution $\sigma_k = \langle k^{1/2} \rangle$, which is always smaller than $\langle k \rangle$

  Network's nodes have degrees in the range $k = \langle k \rangle \pm \langle k \rangle^{1/2}$

- Nodes have **comparable degrees:**

  the **average degree** $\langle k \rangle$ serves as the "***scale***" of a random network

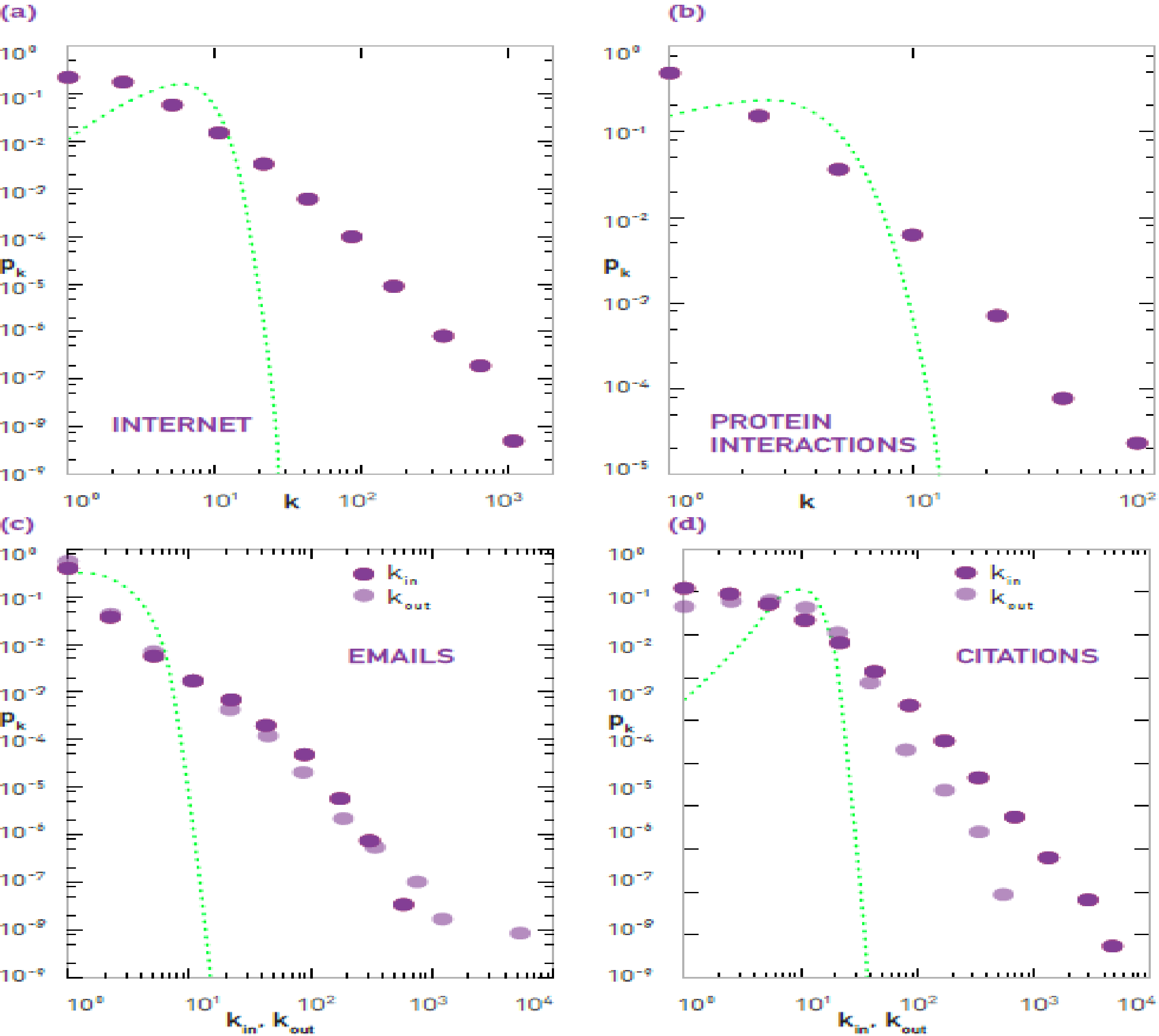
Is the network scale-free?

Degree distribution will immediately reveal:

- In scale-free networks, the degrees of the smallest & the largest nodes are **widely different**, often spanning several orders of magnitude

  These nodes have **comparable degrees in a random network**

Networks of major scientific, technological & societal importance are scale-free.



Their diversity is remarkable!

Internet:

man-made, with a history ~2 decades

protein interaction:

product of four billion years of evolution

Power Laws and Scale-Free Networks

- The integral of *p(k)* encountered in the continuum formalism

$$\int_{k_1}^{k_2} p(k)\,dk$$

is the **probability that a randomly chosen node has degree between k_1 and k_2.**

Hubs

- Main difference between a random and a scale-free network comes in the *tail* of the degree distribution
- high-k region of p_k
- For small k power law is above Poisson function, indicating that a scale-free network has a large number of small degree nodes, most of which are absent in a random network
- For k in the vicinity of $\langle k \rangle$ Poisson distribution is above power law, indicating that in a random network there is an excess of nodes with degree $k \approx \langle k \rangle$

Hubs

- For large *k,* the power law is again above the Poisson curve
- The probability of observing a high-degree node, or **hub**, is several orders of magnitude higher in a scale-free than in a random network
- if the WWW were to be a random network with $\langle k \rangle = 4.6$ & size $N \approx 10^{12}$, we would expect $N_{k \geq 100}$ nodes with at least 100 links:

$$N_{k\geq 100} = 10^{12} \sum_{k=100}^{\infty} \frac{(4.6)^k}{k!} e^{-4.6} \simeq 10^{-82}$$

But we have more than four billion nodes with degree *k* ≥100 …

How does the network size affect the size of its hubs?

To answer this we calculate the maximum degree, k_{max}, called the **natural cutoff of the degree distribution p_k.**

It represents the **expected size of the largest hub in a network**.

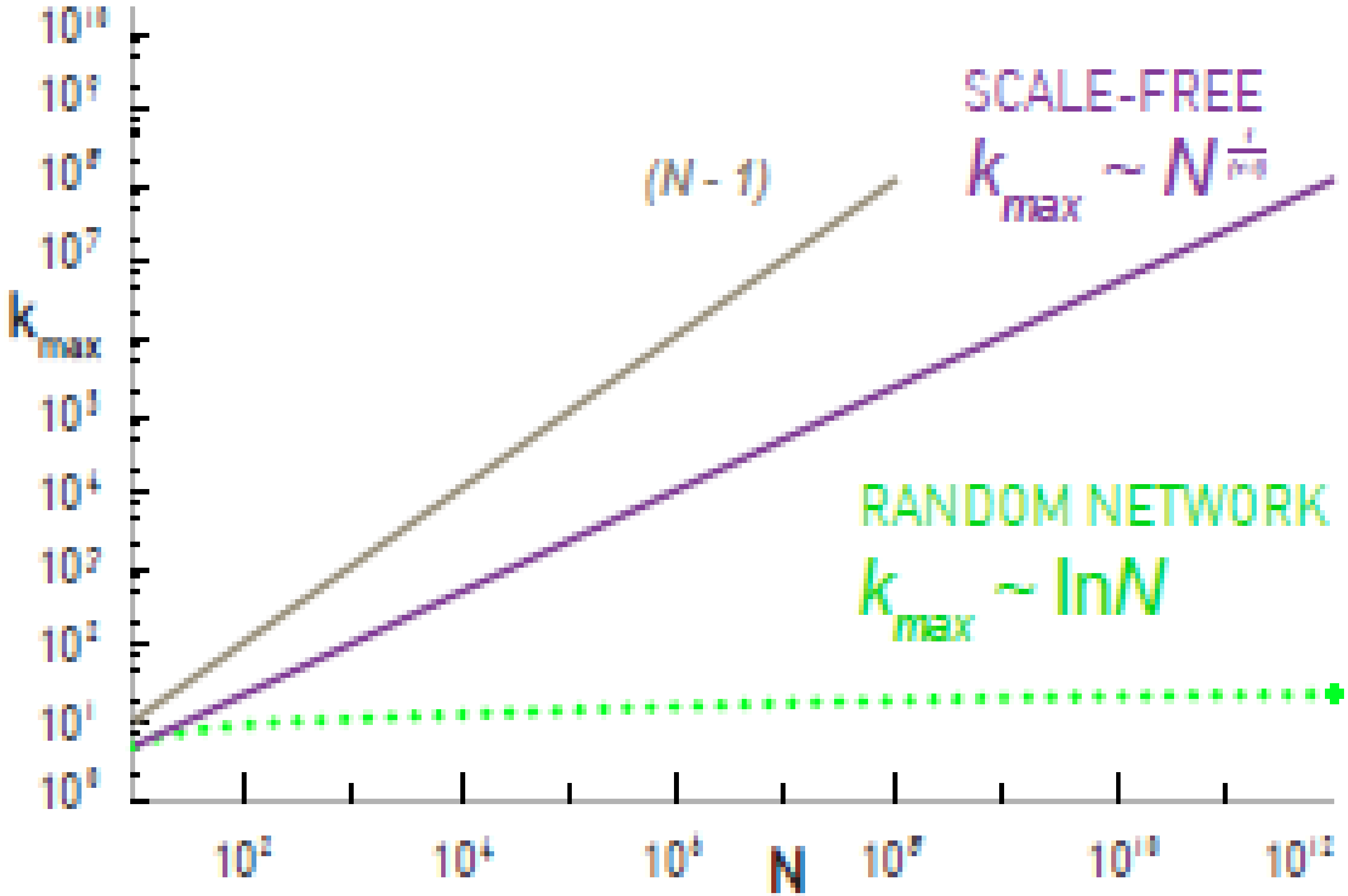
Random networks

$$k_{\max} = k_{\min} + \frac{\ln N}{\lambda}$$

Scale-free networks

$$k_{max} = k_{min} \, N^{\frac{1}{\gamma - 1}}$$

Hubs in Scale-free Networks vs. Random Networks



Hubs in a scale-free network are several orders of magnitude larger than the biggest node in a random network with the same N and $\langle k \rangle$.

Random vs. Scale-free Networks

- Random network most nodes have comparable degrees

- The more nodes a scale-free network has, the larger are its hubs

  The **size of the hubs grows polynomially with network size**: they can grow quite large in scale-free networks.

- In contrast, in a random network the size of the largest node grows logarithmically or slower with N, implying that hubs will be tiny even in a very large random network

Is the network scale-free?

- **Degree distribution** will immediately reveal
- In scale-free networks, the degrees of the smallest & the largest nodes are **widely different**, often spanning **several orders of magnitude**
- **In random networks, the nodes have comparable degrees**

  Random networks have a scale

The behavior of scale-free networks is sensitive to the value of the degree exponent γ.

Scale-Free Regime ($2 < \gamma < 3$)

- k_{max} grows with the size of the network with exponent $1/(\gamma - 1)$, which is smaller than one.
- The market share of the largest hub, representing the fraction of nodes that connect to it, decreases as $\sim N^{-(\gamma-2)/(\gamma-1)}$

Random Network Regime ($\gamma > 3$)

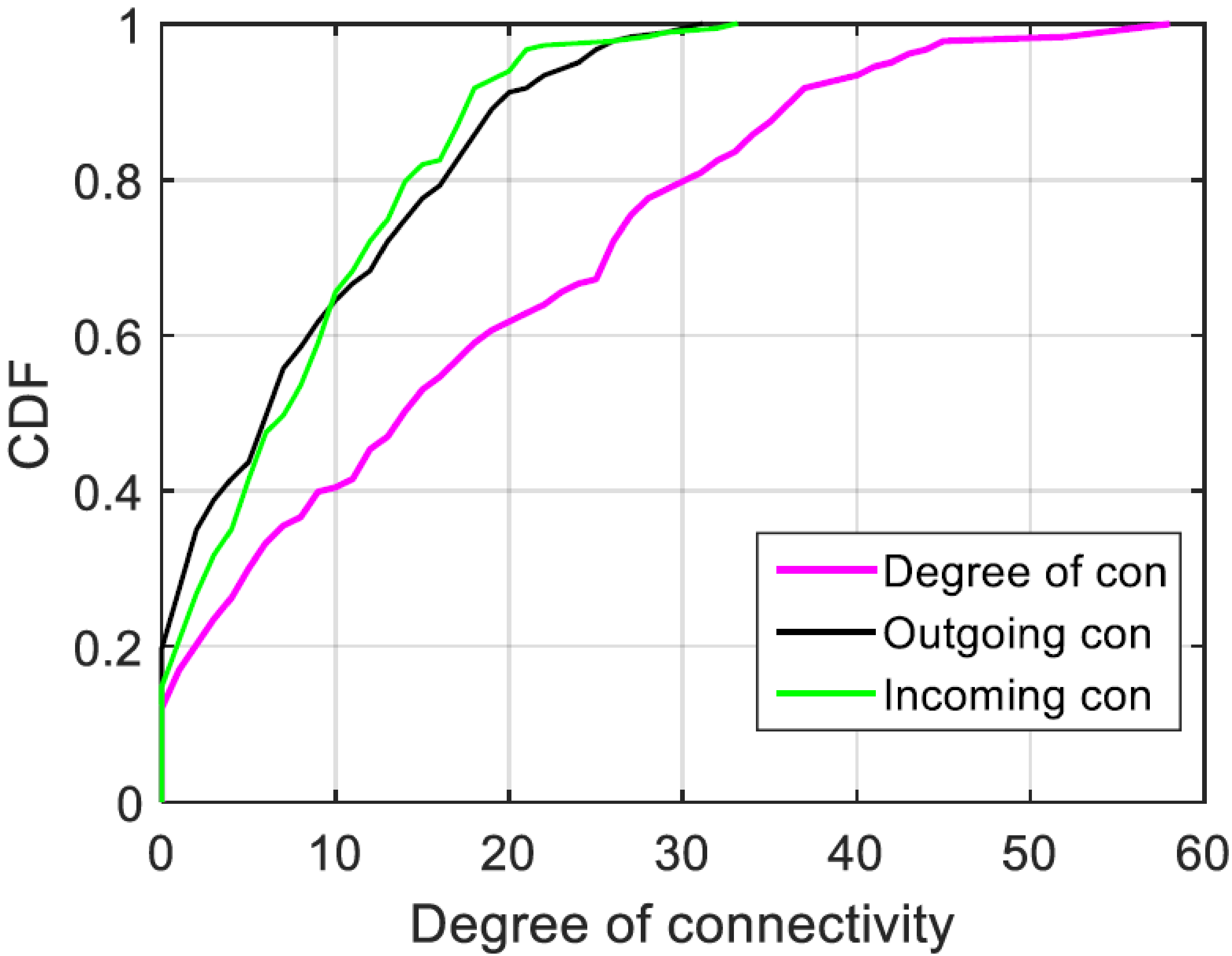
- For all practical purposes the properties of a scale-free network in this regime are difficult to distinguish from the properties a random network of similar size.
- The reason is that for large γ the degree distribution p_k decays sufficiently fast to make the hubs small and less numerous.

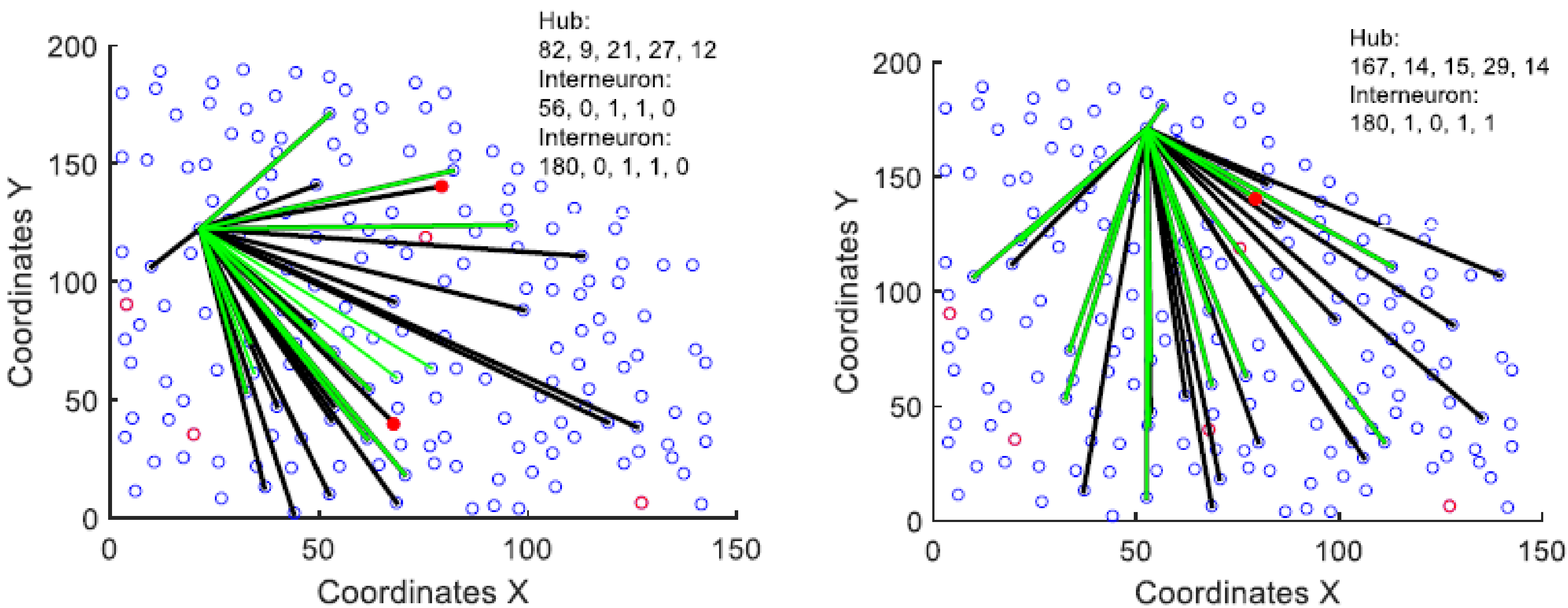
The behavior of scale-free networks is sensitive to the value of the degree exponent γ.

Anomalous Regime ($\gamma \leq 2$)

- The exponent $1/(\gamma - 1)$ is larger than one, hence the number of links connected to the largest hub grows faster than the size of the network:
- For sufficiently large N, the degree of the largest hub must exceed the total number of nodes in the network, hence it will run out of nodes to connect to.
- Similarly, for $\gamma < 2$ the average degree $\langle k \rangle$ diverges in the $N \to \infty$ limit.
- These odd predictions are only two of the many anomalous features of scale-free networks in this regime.
- Large scale-free network with $\gamma < 2$, that lack multi-links, cannot exist

Example – Degree of connectivity considering the significant directional STTC edges (before eye opening mouse)





Green: incoming edges

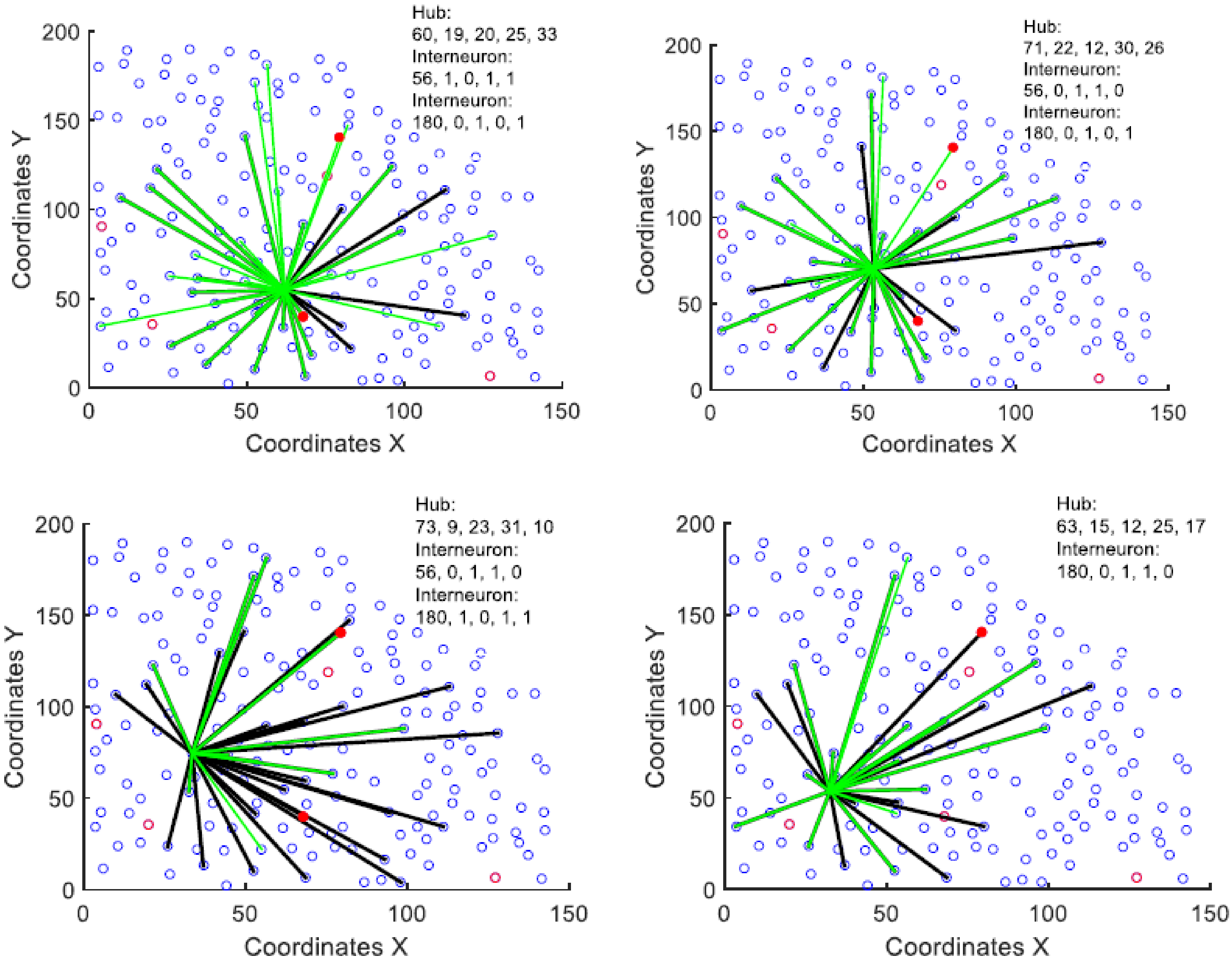
Black: outgoing edges

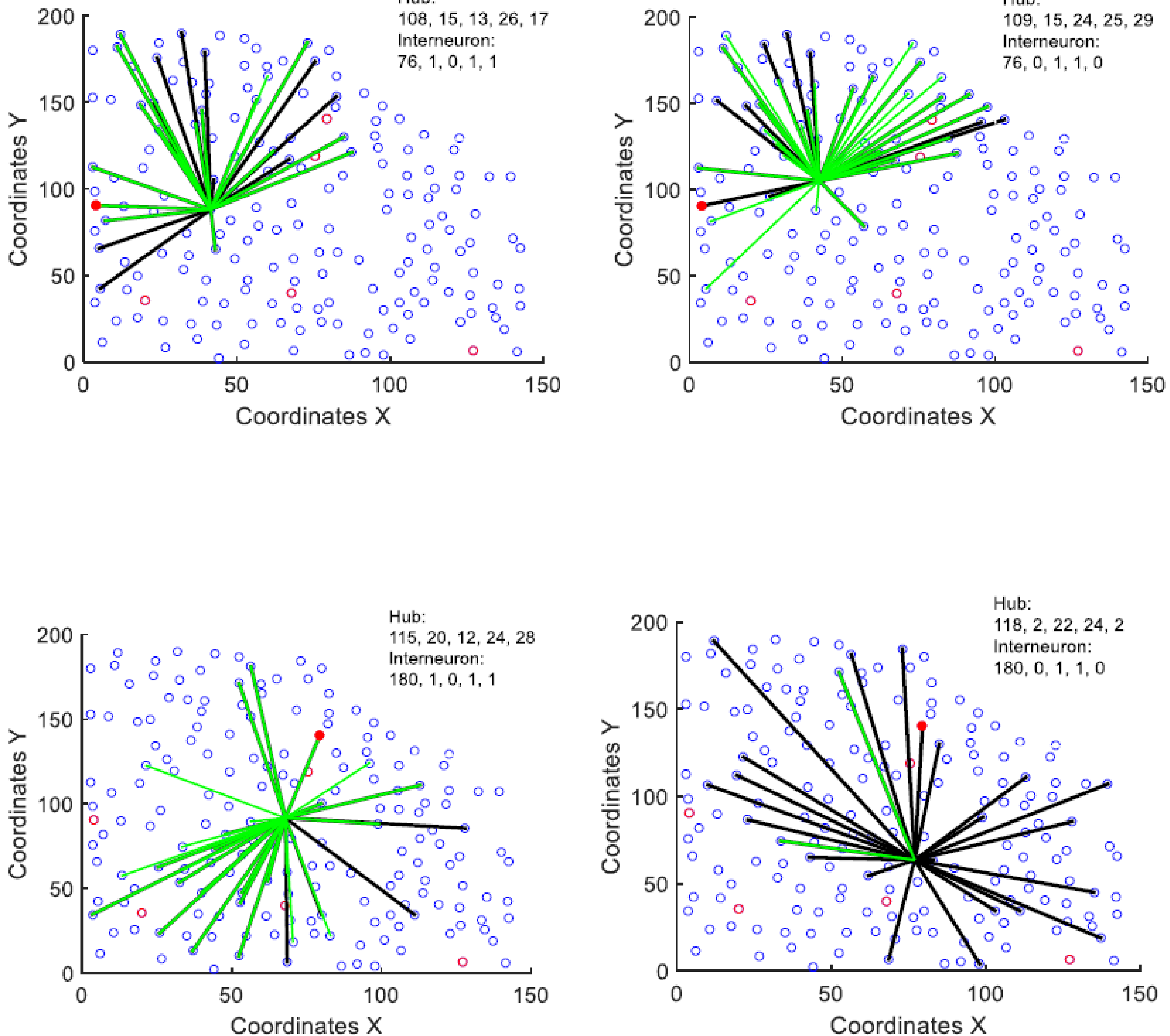
Legend: **Hub id**, number of bidirectional edges, one-way edges, outgoing edges, incoming edges

Red: interneurons

Red (filled) when they have edges with the hub ---- Red (empty) no edges with hub

Before eye-opening mouse


Hub:
60, 19, 20, 25, 33
Interneuron:
56, 1, 0, 1, 1
Interneuron:
180, 0, 1, 0, 1
Coordinates X
Coordinates Y
Hub:
71, 22, 12, 30, 26
Interneuron:
56, 0, 1, 1, 0
Interneuron:
180, 0, 1, 0, 1
Coordinates X
Coordinates Y
Hub:
73, 9, 23, 31, 10
Interneuron:
56, 0, 1, 1, 0
Interneuron:
180, 1, 0, 1, 1
Coordinates X
Coordinates Y
Hub:
63, 15, 12, 25, 17
Interneuron:
180, 0, 1, 1, 0
Coordinates X
Coordinates Y


Hub:
108, 15, 13, 26, 17
Interneuron:
76, 1, 0, 1, 1
Coordinates Y
Coordinates X
Hub:
109, 15, 24, 25, 29
Interneuron:
76, 0, 1, 1, 0
Coordinates Y
Coordinates X
Hub:
115, 20, 12, 24, 28
Interneuron:
180, 1, 0, 1, 1
Coordinates Y
Coordinates X
Hub:
118, 2, 22, 24, 2
Interneuron:
180, 0, 1, 1, 0
Coordinates Y
Coordinates X

Influence and Centrality

- Hubs: highly or densely connected to the rest of the network
- They facilitate global integrative processes

- A **node is central,** if it has great control over the flow of information within the network
  This control results from its participation **in many of the network's short paths**

- **Closeness centrality** of a node: **inverse of** the **average path length** between that node & all other nodes in the network
- **Betweenness centrality** of a node: **fraction of all shortest paths** in the network that **pass through the node**

Influence and Centrality (cont.)

- A node with **high betweenness centrality can control information flow** because it is at the intersection of many short paths
- Centrality **measures identify elements that are highly interactive** and/or carry a **significant proportion of signal traffic**
- A highly central node in a structural network has the potential to participate in a large number of functional interactions
- A node that is not central is unlikely to be important in network-wide integrative processes
- **Loss of highly central nodes have a larger impact on the functioning of the remaining network**

NOT ALL NETWORK ARE SCALE-FREE

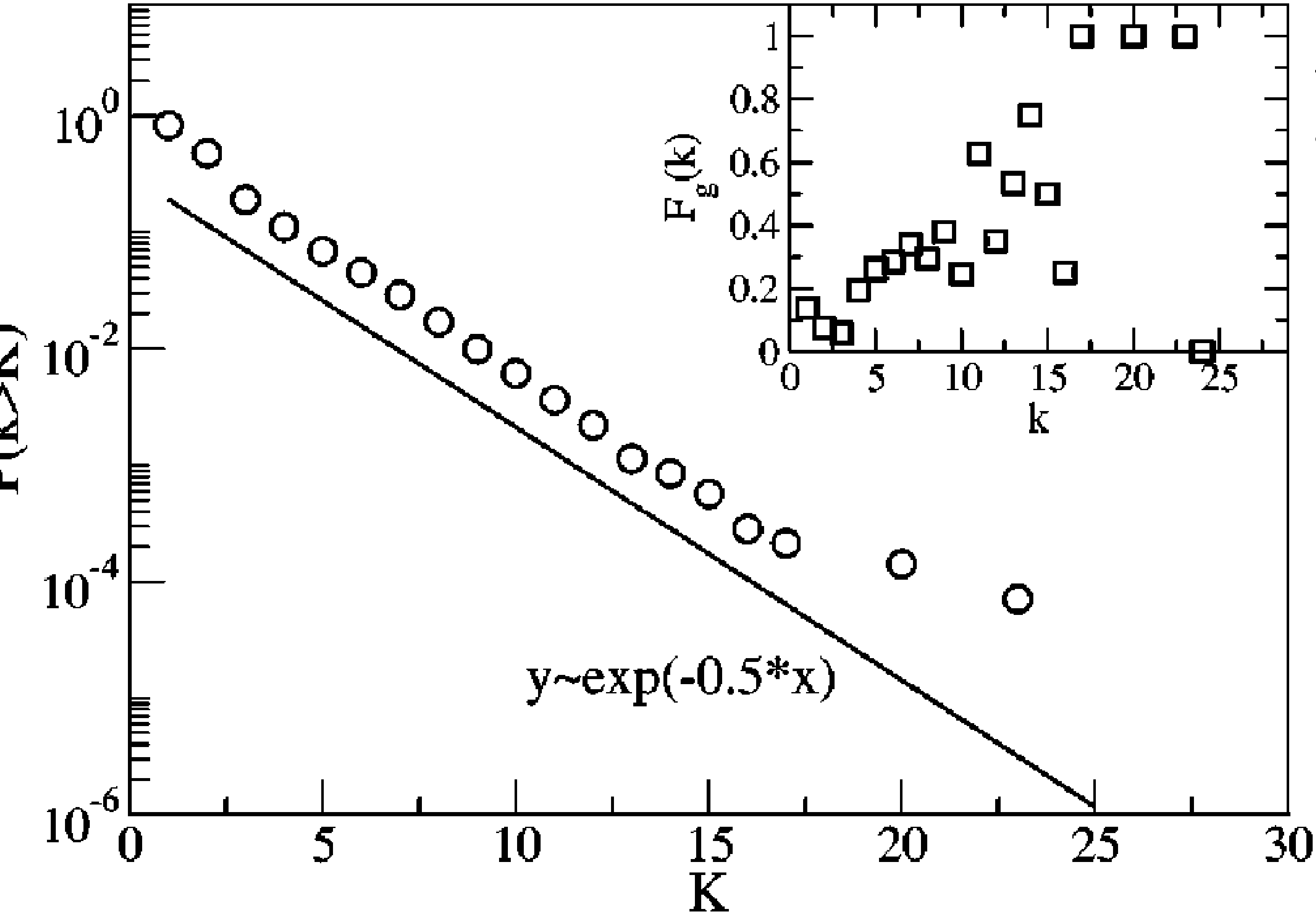
- Networks appearing in material science, describing the bonds between atoms in crystalline or amorphous materials:

  Each node in these networks has exactly the same degree, determined by chemistry

- The neural network of the *C. elegans* worm

- The power grid, consisting of generators & switches connected by transmission lines

Power grid has **exponential degree** distribution.



fraction of generating substations among substations with degree k.

The probability that a substation has more than K transmission lines.

R. Albert et al, Phys. Rev. E 69, 025103(R) (2004)

- Scale-free property to emerge: **nodes need to have the capacity to link to an arbitrary number of other nodes**.

☞ These links **do not need to be concurrent**

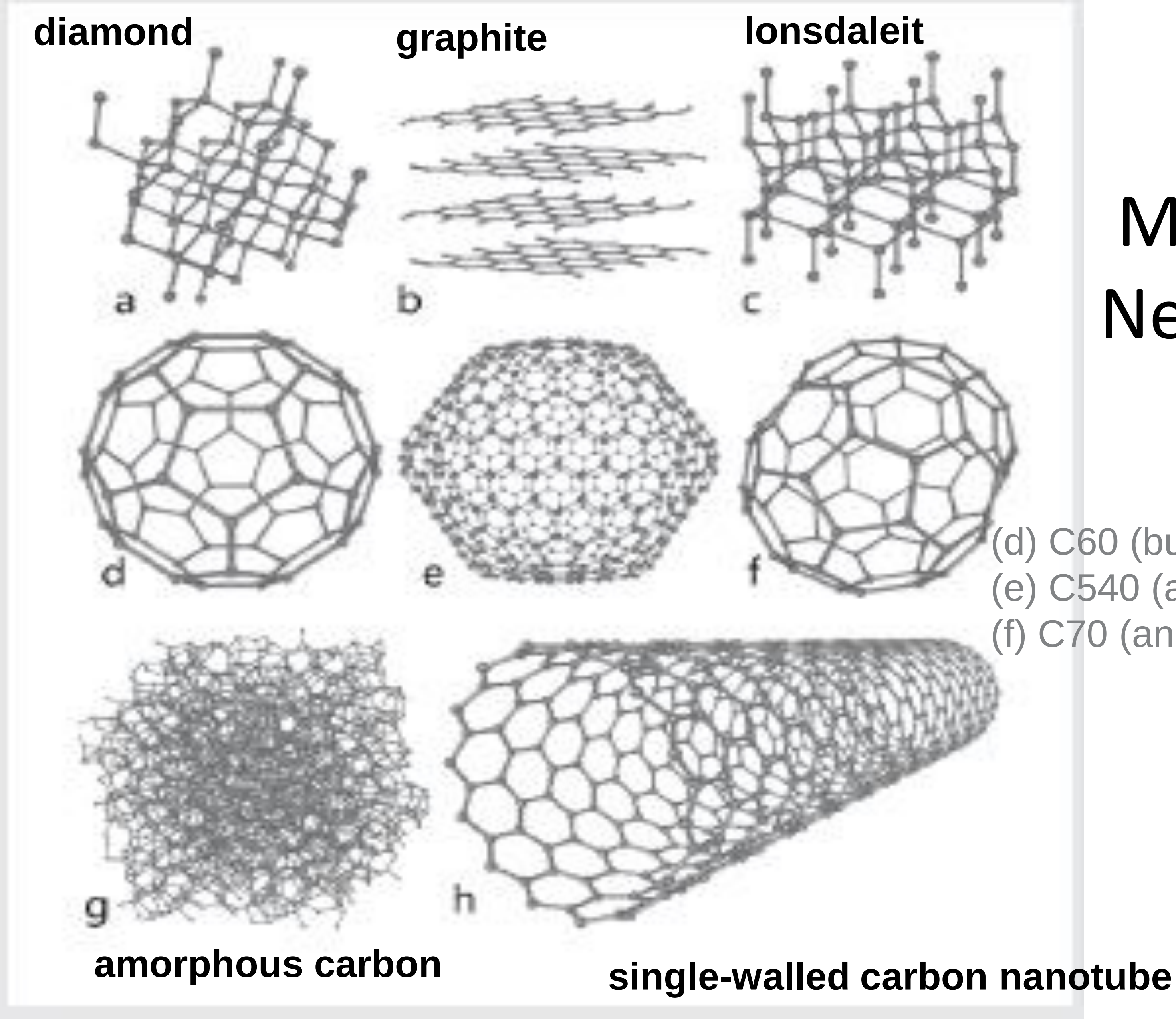
We do not constantly chat with each of our acquaintances

A protein in the cell does not simultaneously bind to each of its potential interaction partners

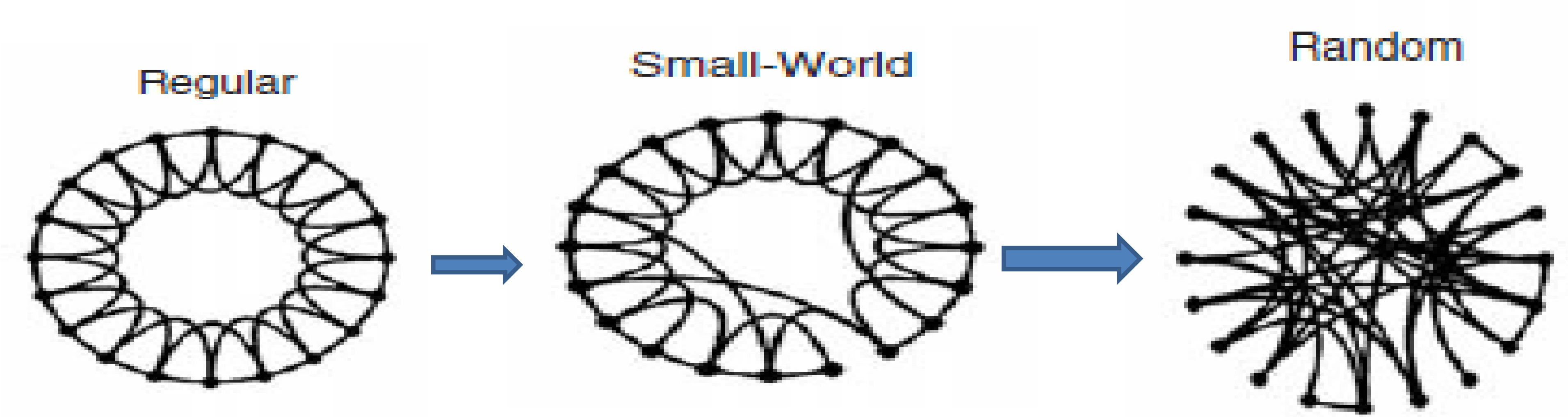
- The **scale-free property is absent in systems** that **limit the number of links a node can have**, effectively restricting the maximum size of the hubs.
- **Such limitations are common in materials**

(explaining why they cannot develop a scale-free topology)

Material Networks



(d) C60 (buckminsterfullerene)
(e) C540 (a fullerene)
(f) C70 (another fullerene)



Rewire with a probability p

For p = 1, we have a random graph

Lattice and random graphs should have:

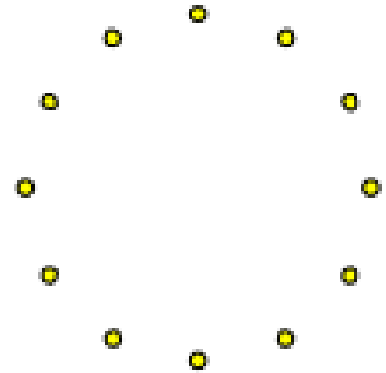
- Same number of nodes
- Same number of edges

Random Graphs

$N = 12$

Erdős and Renyi (1959)

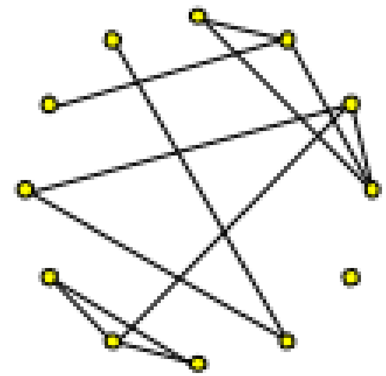
$p = 0.0$; $k = 0$



N nodes

A pair of nodes has probability p of being connected.

$p = 0.09$; $k = 1$

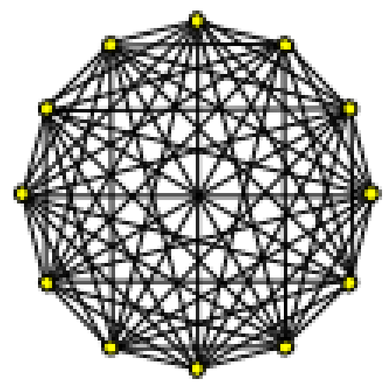


Average degree, $k \approx pN$

What interesting things can be said for different values of p or k ?

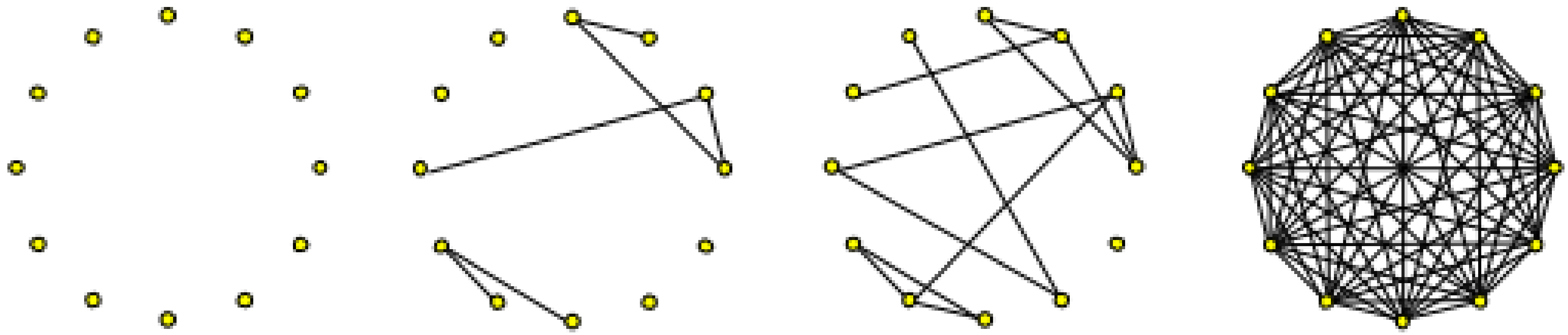
(that are true as $N \rightarrow \infty$)

$p = 1.0$; $k \approx \frac{1}{2}N^2$



Random Graphs

Erdős and Renyi (1959)



	$p = 0.0$; $k = 0$	$p = 0.045$; $k = 0.5$	$p = 0.09$; $k = 1$	$p = 1.0$; $k \approx ½N^2$
Size of largest component	1	5	11	12
Diameter of largest component	0	4	7	1
Average path length between nodes	0.0	2.0	4.2	1.0

Random Graphs

– Erdős and Renyi showed that average path length between connected nodes is $\frac{\ln N}{\ln k}$

Erdős and Renyi (1959)

If $k < 1$:

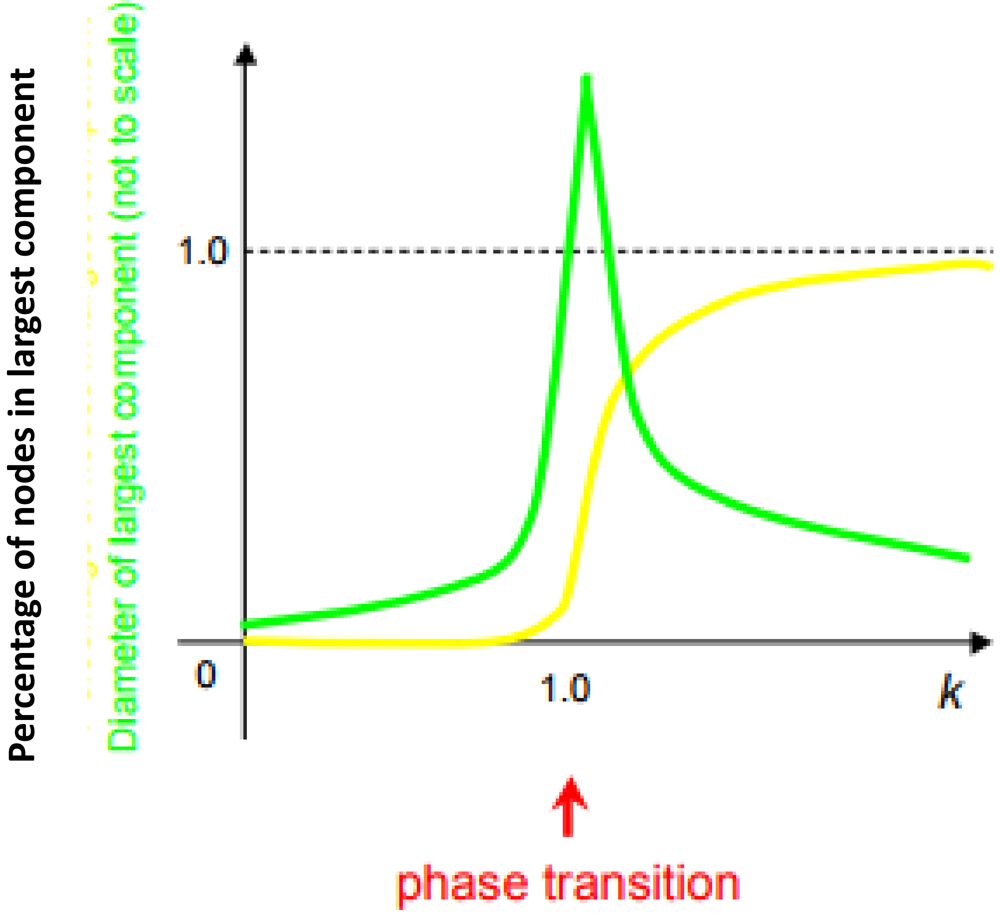
- small, isolated clusters
- small diameters
- short path lengths

At k = 1:

- a *giant component* appears
- diameter peaks
- path lengths are high

For k > 1:

- almost all nodes connected
- diameter shrinks
- path lengths shorten



Construction of Random Networks & Lattice

They can follow different approaches:

1. Erdös-Rényi models

   G(N,p)

   G(N,m)

1. **Sporns Erdös-like model**
2. Sporns real-based model

They use different input.

Erdös-Rényi Models: G(N,p) & G(N,M)
Graphs of N nodes:

In Erdös-Rényi G(N,p)

- A graph is constructed by connecting nodes randomly.
- Each edge is included in the graph with **probability *p*** **independent** from every other edge.

  p can be thought of as a weighting function;

p is estimated from the observed graph = E/(n*(n-1)/2)

In Erdös-Rényi *G*(*N*, *M*)

- A graph is chosen uniformly at random from the collection of all graphs with *N* nodes and *M* edges.

  E.g., in *G*(3, 2) model, each of the three possible graphs on three vertices and two edges are included with probability 1/3.

- G(N, p) model fixes the probability p that two nodes are connected
- G(N, M) model fixes the total number of edges M.
- While in G(N, M) model, the average degree of a node is simply <k> = 2L/N, other network characteristics are easier to calculate in the G(N, p) model.

A random network consists of N nodes where each node pair is connected with probability p.

To construct a random network we follow these steps:

1) Start with N isolated nodes.

2) Select a node pair and generate a random number between 0 and 1. If the number exceeds p, connect the selected node pair with a link, otherwise leave them disconnected.

3) Repeat step (2) for each of the $N(N-1)/2$ node pairs.

Erdös-Rényi Randomization

Start from a **lattice network** and **rewire an edge** with a probability p.

- **N** : number of nodes
- **p** : **rewiring probability**
- **k** : **average degree of connectivity** (it must be an even number)

– **Random network, p = 1** All the edges are rewired
– **Lattice network, p = 0** No edge is rewired

Watts-Strogatz Randomization

k: average degree of connectivity of observed graph G=(V, E)

1. **Create a lattice** by connecting the **k/2 nodes** closer to **the left & right** neighbours of each node.
2. For the creation of the random graph (p=1)
   i. **Disconnect all edges**
   ii. For |E| iterations
      a) Select **two different nodes randomly**
      b) **Create a new edge** between these two nodes

Note that the random graph may have a smaller number of edges than the lattice

Generation of randomized version of given network

Full Randomization

Generates a random (Erdős–Rényi) network with the same N and E as the original network.

1. Select randomly a source node (S1) & two target nodes T1, T2, where T1 is linked to S1 & T2 is not.
2. Destroy the edge (S1,T1) link & create the edge (S1,T2) .
3. Perform this procedure **once for each link in the network**.

Degree-Preserving Randomization (according to Sporns Real-based)

Generates a network in which each node has exactly the same degree as in the original network but the network's wiring has been randomized.

1. Select two edges (S1, T1) & (S2, T2).
2. Destroy them & create new edges (S1,T2) & (S2, T1).

   The swap leaves the degree of each node unchanged.

1. Repeat this procedure **until we rewire each link at least once**.

Sporns Erdös-like – Creation of Lattice

Input: **N : number of nodes**, **K : total number of edges**

1. Place the nodes at the periphery of a circle
2. Connect each node with its immediate left & right neighbour
3. Compute the total number of edges (E)
   i. If E=K, the lattice has been constructed
   ii. If E>K, **randomly disconnect** (E-K) edges
   iii. If E< K, connect each node with its second degree neighbours (left & right) in the circle

   Repeat the step (3)

Sporns Erdös-like – Creation of Random Network

Input: N : number of nodes, **K** : total number of edges

1. Place the nodes at the periphery of a circle
2. Repeat the following steps for **K** iterations
   i. Select two different nodes randomly
   ii. Connect them with an edge

End

Sporns real-based – Creation of Lattice

Input: G: graph (V, E) of real network, R: number of iterations

1. Repeat the following steps for R iterations
2. Repeat the following steps for |E| iterations
3. Select **randomly two different edges** from G, e.g., (A,B), (C,D)
   If these 4 nodes are not different, return to step 3
4. If (A,D) ϵ E or (B,C) ϵ E, return to step 3
   Otherwise
   If ||A,B||+||C,D||> ||A,D||+||B,C||
   create the new edges (A,D) & (B,C)
   destroy the (A,B), (C,D)
   end
   Return to step 3

||A,B||: denotes the Euclidian distance between A &B

Sporns real-based – Creation of Random Network

Input: G: graph (V, E) of real network, R: number of iterations

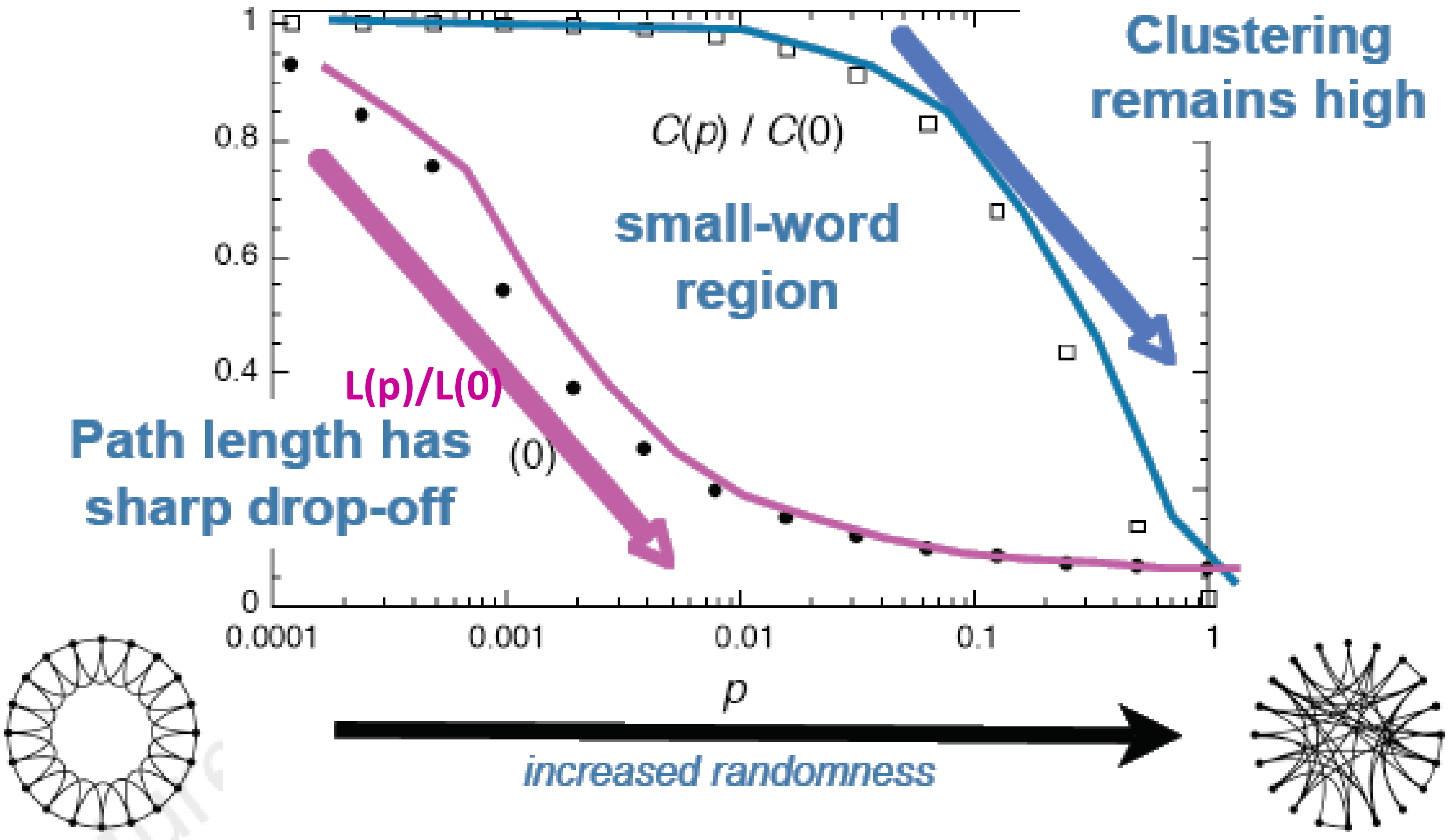
1. Repeat the following steps for R iterations
2. Repeat the following steps for |E| iterations
3. Select randomly **two different edges** from G, e.g., (A,B), (C,D)
   If these 4 nodes are not different, return to step 3
4. If (A,D) ϵ E or (B,C) ϵ E, return to step 3
   Otherwise
   create the new edges (A,D) & (B,C)
   destroy the (A,B), (C,D)
   end

Small-world networks

L = characteristic path length **C = clustering coefficient**

- A small-world network is much more highly clustered than an equally sparse random graph **($C \gg C_{random}$)** & its characteristic path length L is close to the theoretical minimum shown by a random graph **($L \sim L_{random}$)**.
- The reason a graph can have small L despite being highly clustered is that a few nodes connecting distant clusters are sufficient to lower L.
- Because C changes little as small-worldliness develops, it follows that small-worldliness is a global graph property that cannot be found by studying local graph properties.

Small-World



[Watts & Strogatz (1998) *Nature* 393: 440]

Small-world Criteria

- **Small-worldness S^{Δ}** > 1

$$S^{\Delta} = \frac{\gamma_g^{\Delta}}{\lambda_g} \qquad \gamma_g^{\Delta} = \frac{C_g^{\Delta}}{C_{random}^{\Delta}} \qquad \lambda_g = \frac{L_g}{L_{random}}$$

- **Small-world propensity (ϕ)** close to 1 (suggested reference value 0.6)

$$\Delta_C = \frac{C_{lattice}^{\Delta} - C_g^{\Delta}}{C_{lattice}^{\Delta} - C_{random}^{\Delta}} \qquad \Delta_L = \frac{L_g - L_{random}}{L_{lattice} - L_{random}}$$

$$\phi = 1 - \sqrt{\frac{\Delta_C^2 + \Delta_L^2}{2}}$$

g: real network
rand: random network

Small-world Criteria (cont'd)

$$\sigma = \frac{C}{L} \times \frac{L_r}{C_r}$$

C≫Cr and L≈Lr, or **σ>1**

Compares the network's clustering coefficient & average shortest path length to the random reference graph, *ignoring the corresponding lattice*

$$\omega = \frac{L_r}{L} - \frac{C}{C_l}$$

values around 0 considered small world

Recommended when the degree distribution is maintained

$$SWI = \frac{L - Ll}{Lr - Ll} \times \frac{C - Cr}{Cl - Cr}$$

class	#	network	n	m	$\langle k \rangle$	ξ	L	C^{Δ}	C^{ws}	S^{Δ}	S^{ws}
Biological	25	metabolic network	765	3686	9.65	0.0126	2.56	0.09	0.67	8.18	60.89
	26	yeast protein interactions	2115	2240	0.001	2.12	6.8	0.072	0.071	107.85	106.35
	27	marine food web	135	598	4.43	0.0661	2.05	0.16	0.23	7.84	11.27
	28	freshwater food web	92	997	10.84	0.2382	1.9	0.2	0.087	1.7	0.74
	29	C.Elegans†	277	1918	13.85	0.05	2.64	0.2	0.28	3.21	4.51
	30	Macaque cortex†	95	1522	32.04	0.34	1.78	0.7	0.77	1.53	1.69
	31	E. Coli substrate	282	1036	7.35	0.0261	2.9	-	0.59	-	22.08
	32	E. Coli reaction	315	8915	56.6	0.18	2.62	-	0.22	-	0.67
	33	functional cortical connectivity	90	405	9	0.1	2.49	-	0.53	-	4.32

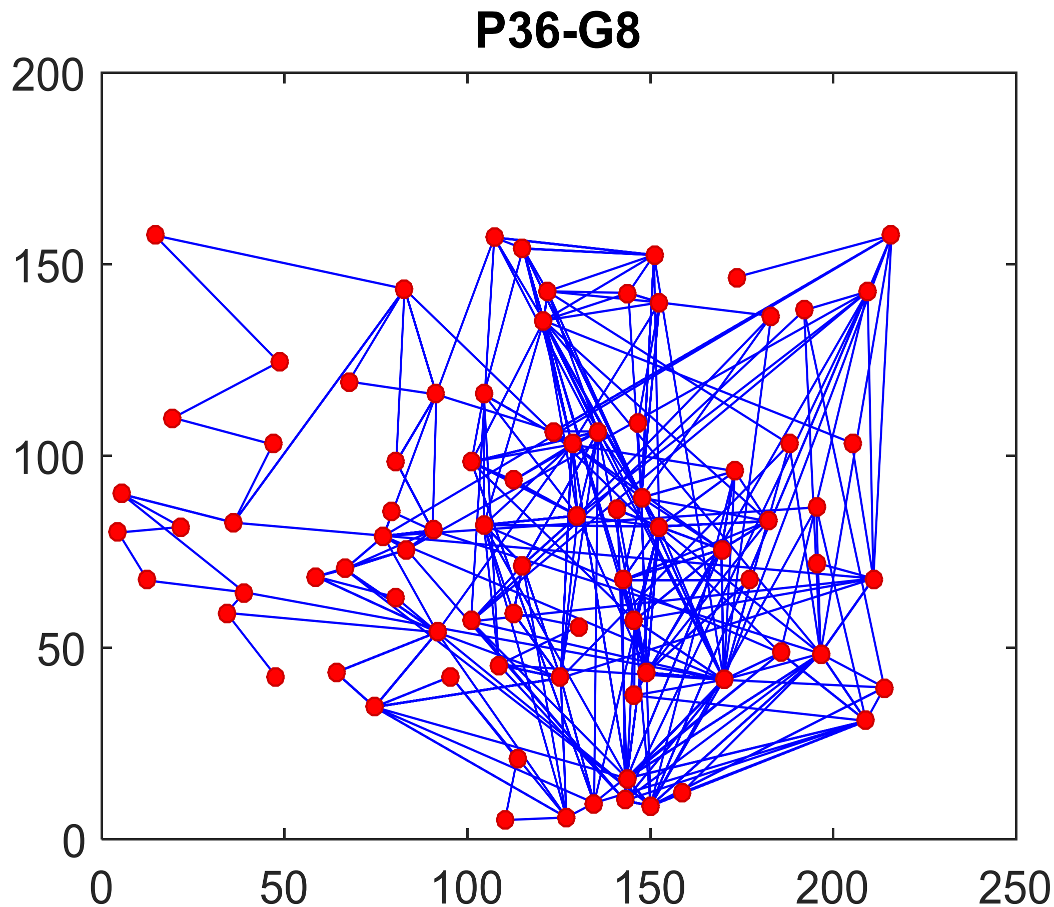
n number of nodes

m number of edges

ξ density of edges

<k> mean degree of connectivity

$$S^{\Delta} = \frac{\gamma_g^{\Delta}}{\lambda_g}$$



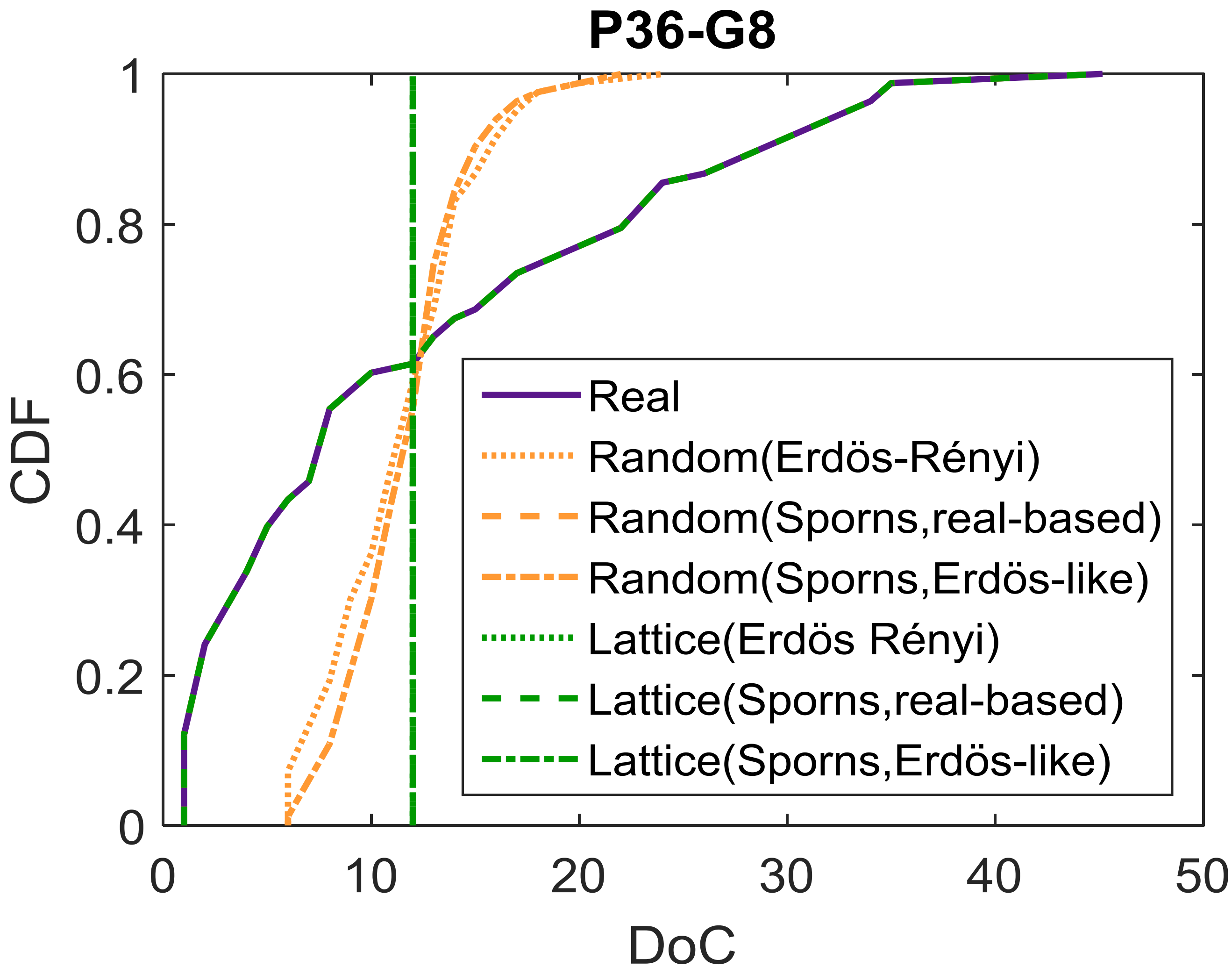
The big connected component, formed by 83 neurons (43 neurons were not connected to any other neuron).

The small world analysis has been done for the connected component.

Degree of Connectivity				Number (Percentage)		
Average	**Median**	**Max**	**Min**	**Hubs**	**Nodes**	**Edges**
11.8554	8	45	1	9 (10.84%)	83 (65.87%)	492 (6.25%)

Example: Degree of Connectivity (DoC) using the big connected component, formed by 83 neurons (43 neurons were not connected to any other neuron).

The small world analysis has been done for the connected component.



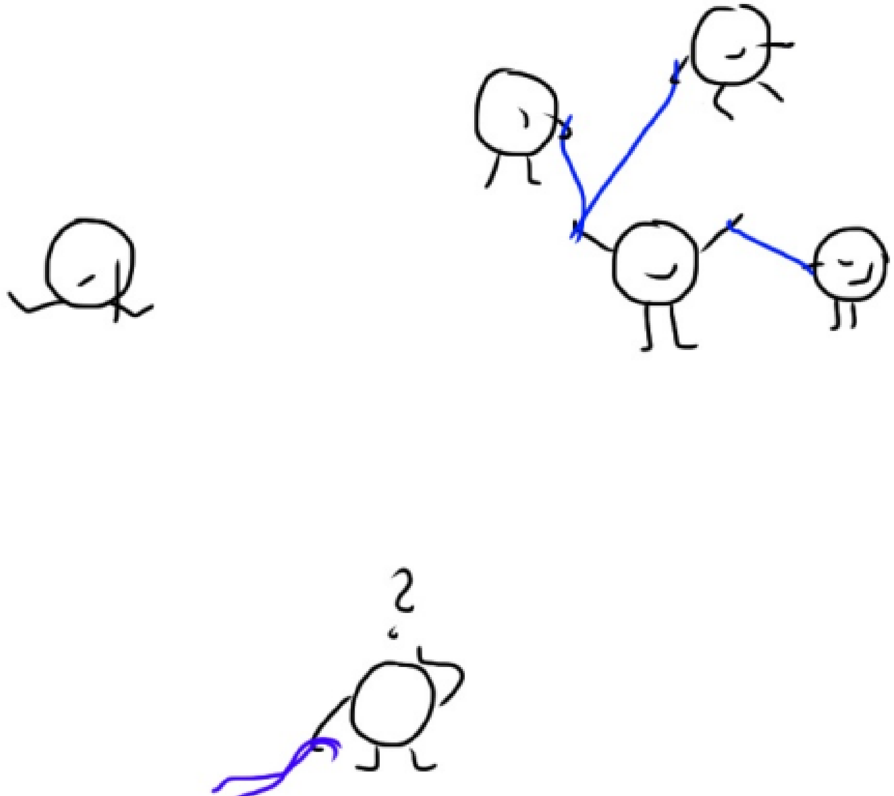
		Clustering Coeff	Shortest Path
Real-life		0.43	2.22
Erdös Rényi	**Random**	**0.15**	**2.01**
	Lattice	0.68	3.92
Sporns Erdös-like	**Random**	**0.14**	**1.98**
	Lattice	0.68	3.92
Sporns real-based	Random	0.38	2.10
	Lattice	0.41	2.17

	$\gamma^{\Delta}_{g\,cc}$	$\lambda_{g\,cc}$	$S^{\Delta} = \frac{\gamma^{\Delta}_{g\,cc}}{\lambda_{g\,cc}}$	Δ_C	Δ_L	$\phi = 1 - \sqrt{\frac{\Delta_C{}^2 + \Delta_L{}^2}{2}}$
Erdös Rényi	2.78	1.11	2.50	0.47	0.11	0.66
Sporns Erdös-like	3.08	1.12	2.75	0.46	0.12	0.66
Sporns real-based	1.11	1.06	1.05	0	1	0.29

Animal P36-G8	Small-Worldness	Small-World Propensity
Erdös Rényi	✓	✓
Sporns Erdös-like	✓	✓
Sporns real-based	✓	✗

Preferential attachment models the growth of a network

- nodes prefer to attach to nodes with many connections (preferential attachment, cumulative advantage)



Preferential attachment models the growth of a network

- **Add a new node**
- **Probability of linking a node** is **proportional to its degree**

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

- The preferential attachment process generates a "long-tailed" distribution following a Pareto distribution or power law in its tail.

- Based on Herbert Simon’s result
  - **Power-laws** arise from “Rich get richer” (cumulative advantage)

Examples

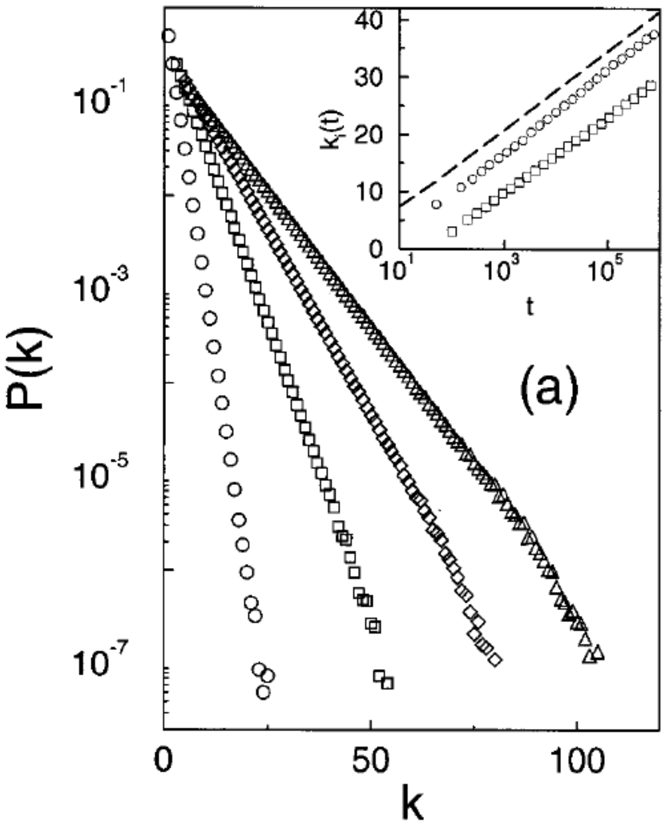
1. Citations: new citations of a paper are proportional to the number it already has [Price 1965]
2. Growth of the WWW [Albert & Barabasi 1999]

Preferential attachment

- Leads to power-law degree distributions

$$p_k \propto k^{-3}$$

- There are many generalizations & variants, but the preferential selection is the **key ingredient that leads to power-laws**

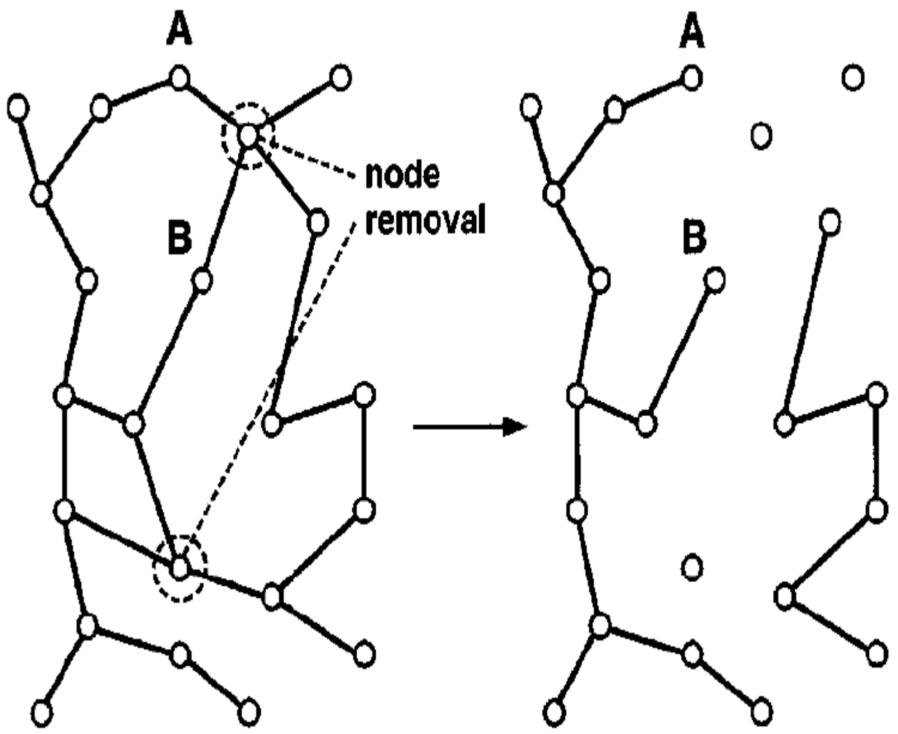


Network resilience (1)

How does the connectivity (length of the paths) of the network changes as the vertices get removed?

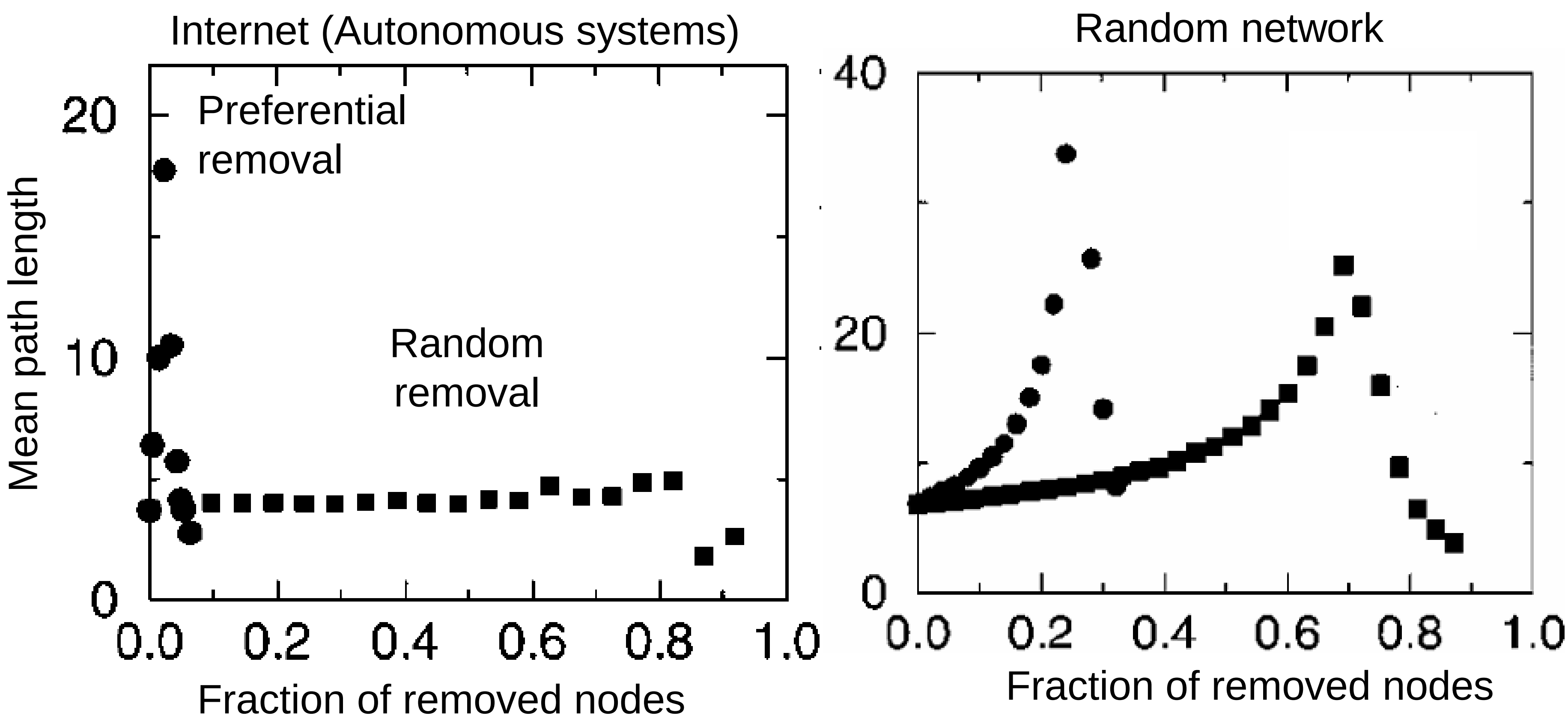
- Removal of vertices
  - Random
  - Targeted
  - According to a **systematic process**
- Important for epidemiology

  e.g., removal of vertices corresponds to vaccination



Network resilience (2)

- Real-world networks are resilient to random attacks
  - One has to remove all web-pages of degree > 5 to disconnect the web
  - But this is a very small percentage of web pages
- **Random** network has better resilience to **targeted** attacks



Dynamical Process

- Starting with N isolated nodes, the links are added gradually through randomly placed edges between nodes.
- This corresponds to a gradual increase of p, with striking consequences on the network topology.

First examine how the size of the largest connected cluster within the network (NG), varies with the average degree of connectivity ‹k›:

Two extreme cases are easy to understand:

1. For p = 0, ‹k› = 0 (all nodes are isolated)

   → NG = 1 & NG/N→0 for large N

1. For p = 1, ‹k›= N-1 (network is a complete graph & all nodes belong to a single component

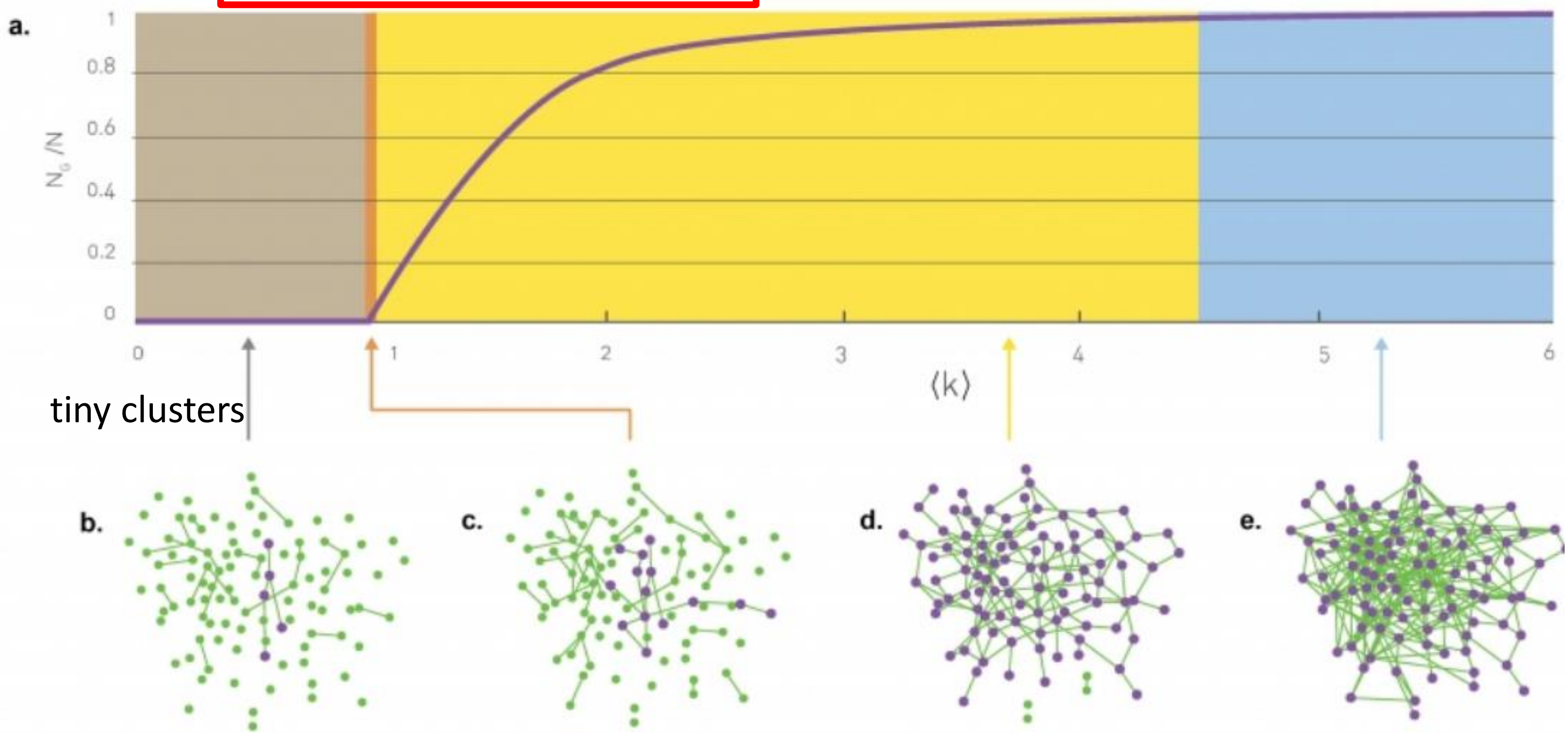
   → NG = N & NG/N = 1

The Evolution of a Random Network

For ‹k› ‹ 1, the largest cluster is a tree with size NG ~ lnN , hence its size increases much slower than the size of the network

The critical point separates the regime where there is **not yet** a giant component (‹k› ‹ 1) from the regime **where there is one** (‹k› › 1).

Critical Point: ‹k› = 1 (p = 1/N)



Absence of a giant component for **small *p*** & its sudden emergence **once *p* reaches a critical value**

Emergence of a Giant Component

- One would expect that the largest component grows gradually from $N_G = 1$ to $N_G = N$, if ‹*k*› increases from 0 to *N*-1.

  Yet, this is not the case!
- N_G/N is zero for small ‹*k*›, indicating the lack of a large cluster.
- Once ‹*k*› exceeds a critical value, N_G/N increases, **signaling the rapid emergence of a large cluster (i.e., the *giant component*)**

The condition for the emergence of the giant component is ⟨k⟩=1

(Erdős & Rényi in their classical 1959 paper)

A giant component exists if and only if **each node has on average more than one link**

The fact that we need at least one link per node to observe a giant component is not unexpected. Indeed, for a giant component to exist, each of its nodes must be linked to at least one other node.

That one link is *sufficient* for its emergence!

Emergence of a Giant Component (con’td)

The condition for the emergence of the giant component is ⟨k⟩=1 is equivalent with

$$p_c = \frac{1}{N-1} \approx \frac{1}{N}$$

Therefore the larger a network, **the smaller p is sufficient for the giant component**.

For ‹k› ‹ 1, the largest cluster is a tree with size **NG ~ lnN**
hence its size increases much slower than the size of the network

At the critical point, the size of the largest component is **NG ~ $N^{2/3}$**

Consequently NG grows much slower than the network’s size, so its relative size decreases as **NG/N ~ $N^{-1/3}$** in the $N \to \infty$ limit.

Note, however, that in *absolute terms,* there is a significant jump in the size of the largest component at ‹k› = 1. For example, for a random network with $N = 7 \times 10^9$ nodes, comparable to the globe’s social network, for ‹k› ‹ 1, the largest cluster is of the order of NG ≃ lnN = ln (7×10^9) ≃ 22.7. In contrast at ‹k› = 1 we expect NG ~ $N^{2/3} = (7 \times 10^9)^{2/3} \simeq 3 \times 10^6$, a jump of about five orders of magnitude. Yet, both in the subcritical regime and at the critical point the largest component **contains only a vanishing fraction of the total number of nodes in the network**.

Network Evolution: Different Regimes

- Subcritical Regine (<k> < 1)
- Critical Point (<k> = 1 or $p_c = \frac{1}{N-1} \approx \frac{1}{N}$)
- Supercritical Regine (<k> > 1)
- Connected Regime (‹*k*› › ln*N* or *p* › ln*N*/*N*)

Network Evolution: Supercritical regime (<k> > 1)

- In the supercritical regime numerous isolated components coexist with the giant component.
- These small components are trees, while the giant component contains loops and cycles.
- The **supercritical regime lasts until all nodes are absorbed by the giant component.**

Network Evolution: Connected Regime

$\langle k \rangle > \ln N$ ($p > \ln N / N$)

- For sufficiently large *p,* the giant component absorbs all nodes and components, hence $N_G \simeq N$
- In the absence of isolated nodes, the network becomes connected
- The average degree at which this happens depends on *N*
- When we enter the connected regime, the network is still relatively sparse, as $\ln N / N \rightarrow 0$ for large *N*
- The network turns into a **complete graph** only at $\langle k \rangle = N - 1$

Phase Transitions: Transitions from Disorder to Order

The emergence of the giant component at ‹*k*›=1 in the random network model is reminiscent of a *phase transition*

Examples from physics & chemistry

- **Water-Ice Transition**:

At high temperatures, the H_2O molecules engage in a diffusive motion, forming small groups & then breaking apart to group up with other water molecules.

If cooled, at 0˚C, the molecules suddenly stop this diffusion, forming **an ordered rigid** ice crystal.

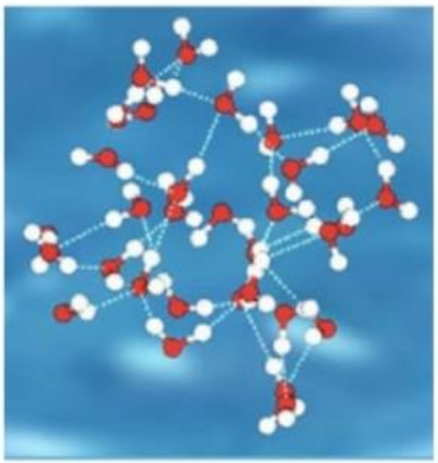
- **Magnetism**:

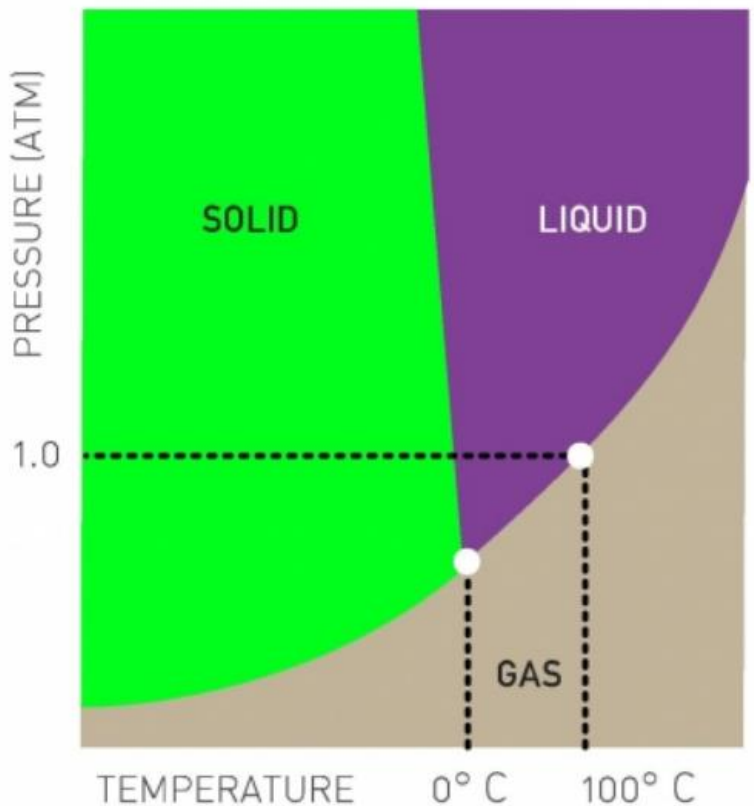
In ferromagnetic metals, like iron, at high temperatures the spins point in randomly chosen directions.

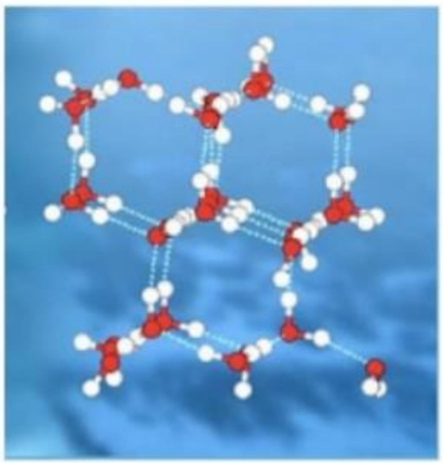
Under some critical temperature T_c all atoms orient their spins in the **same direction** and the metal turns into a magnet.

Water-Ice Transition

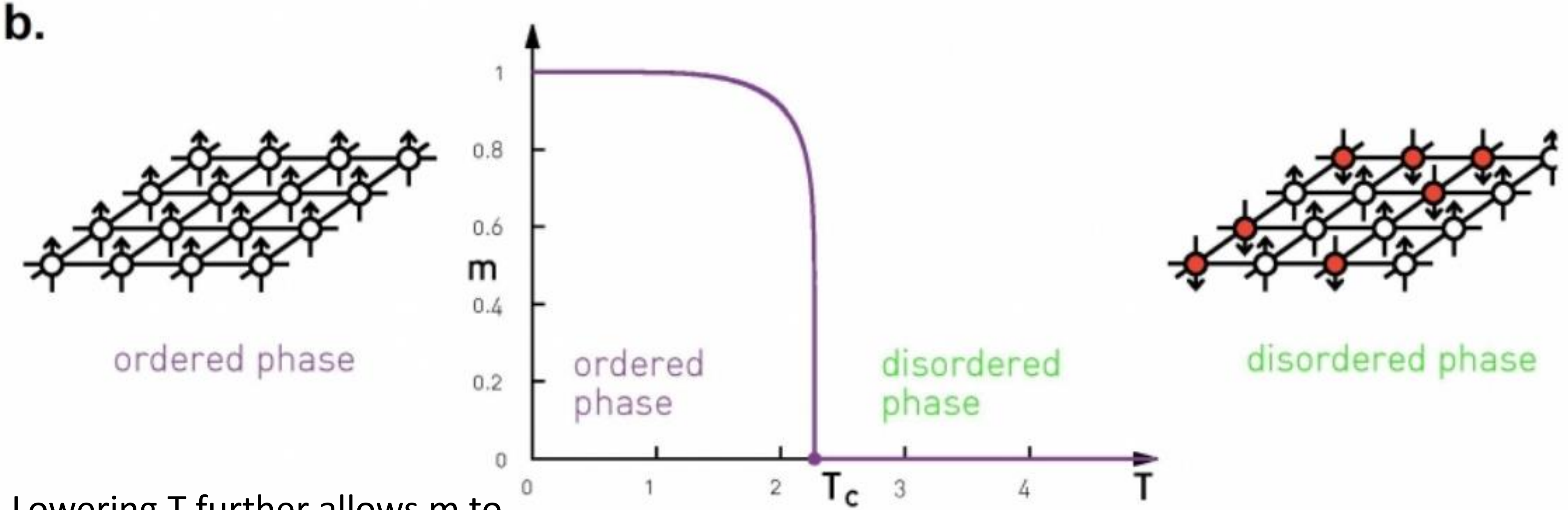
a.







Magnetic Phase Transition

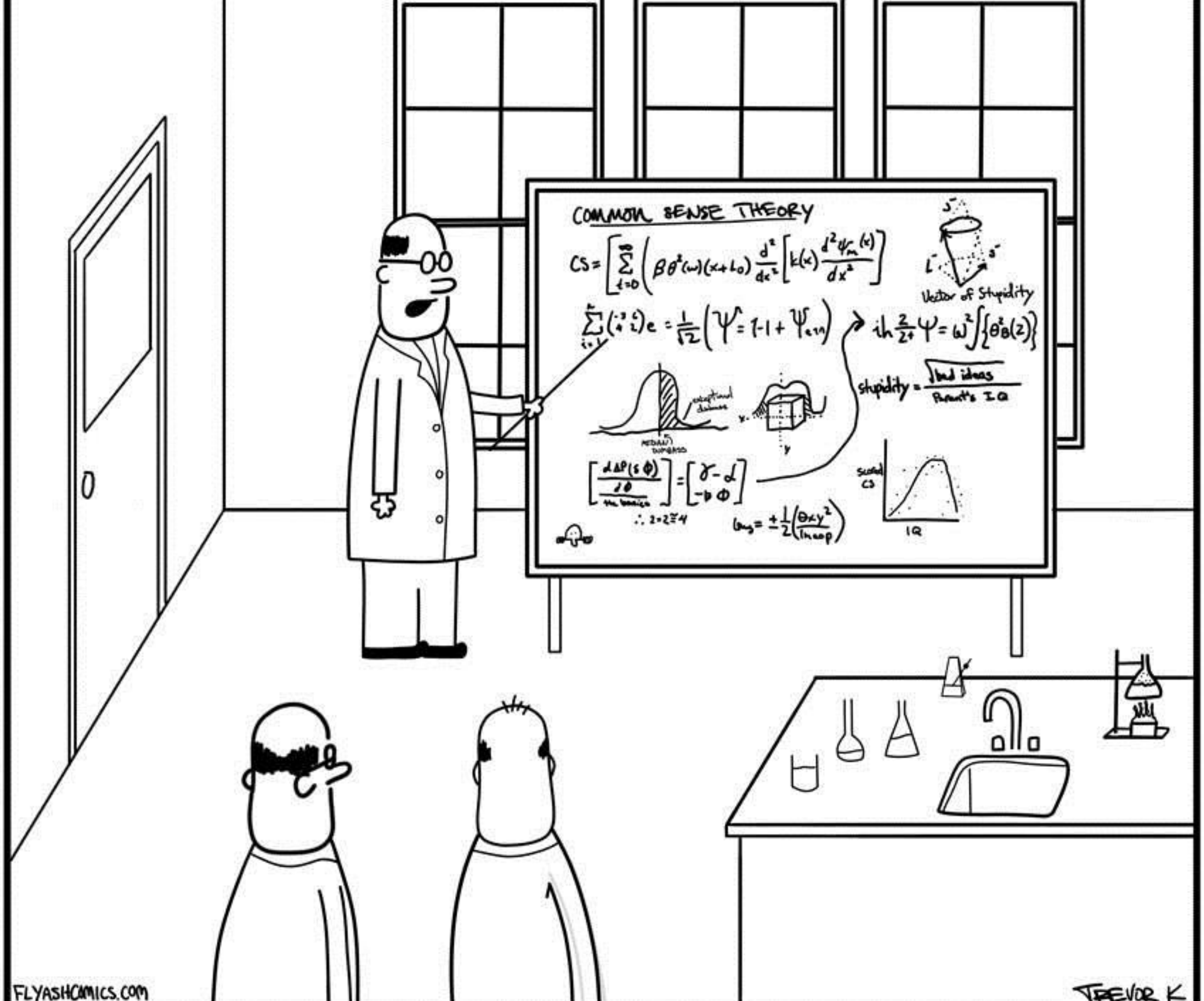


Lowering T further allows m to converge to one. In this ordered phase all spins point in the same direction

by lowering the temperature T, the system undergoes a phase transition at T= Tc, when a nonzero magnetization emerges

In ferromagnetic materials the magnetic moments of the individual atoms (spins) can point in two different directions.

At high temperatures they choose randomly their direction (right panel). In this disordered state the system's total magnetization ($m = \Delta M/N$, where ΔM is the number of up spins minus the number of down spins) is zero.



"Now, while in theory common sense is relatively simple, our real world tests have resulted in abject failure"