Social Networks in Economic Geography

University of Warsaw

February 17-19 2025

Balázs Lengyel

Agglomeration, Networks, and Innovation Research Lab
HUN-REN Centre for Economic- and Regional Studies
Corvinus University of Budapest

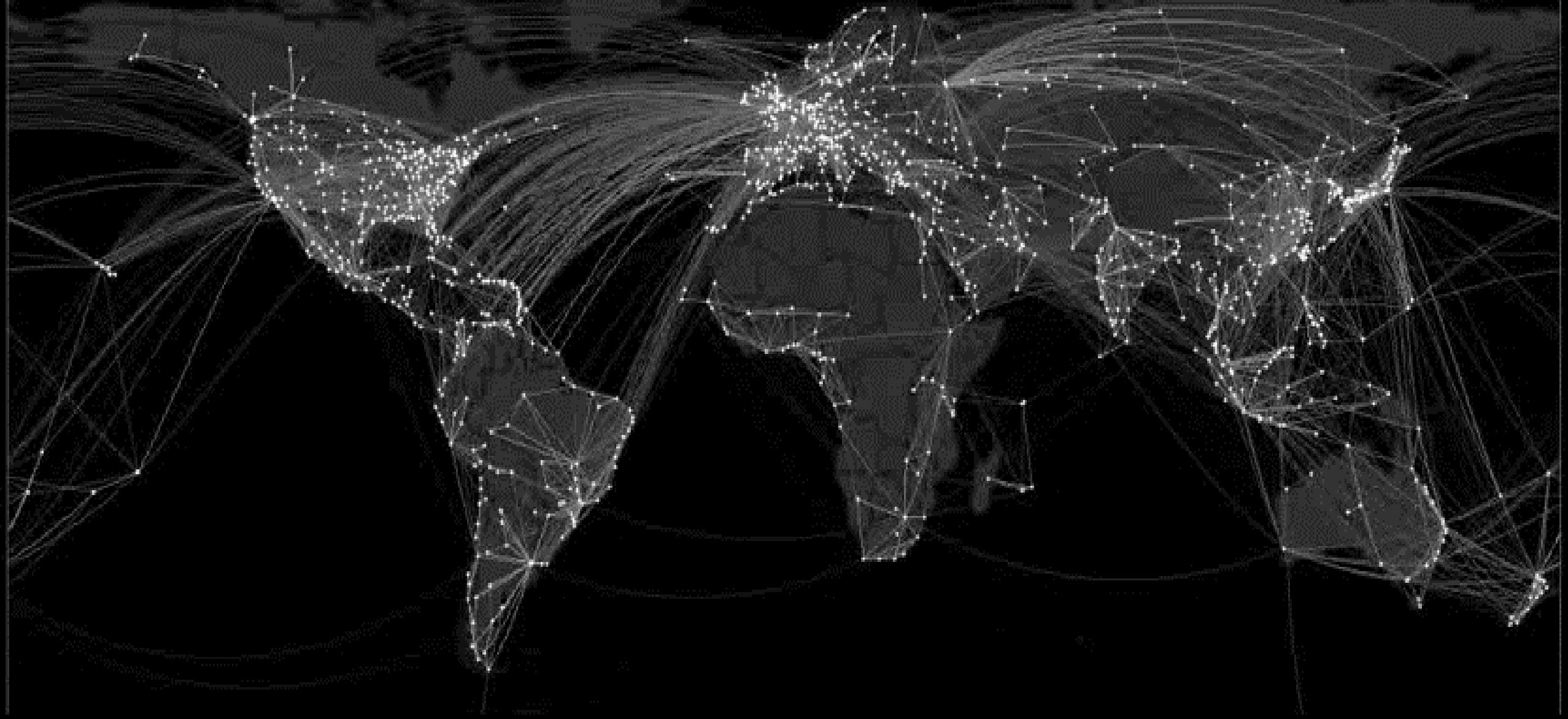
HUN REN | Centre for Economic and Regional Studies

ANETILAB
HUN-REN CERS • Corvinus University

CORVINUS UNIVERSITY of BUDAPEST

We live in small world network.

Diffusion of viruses and information is fast



Brockman D., Helbing D (2013) Science

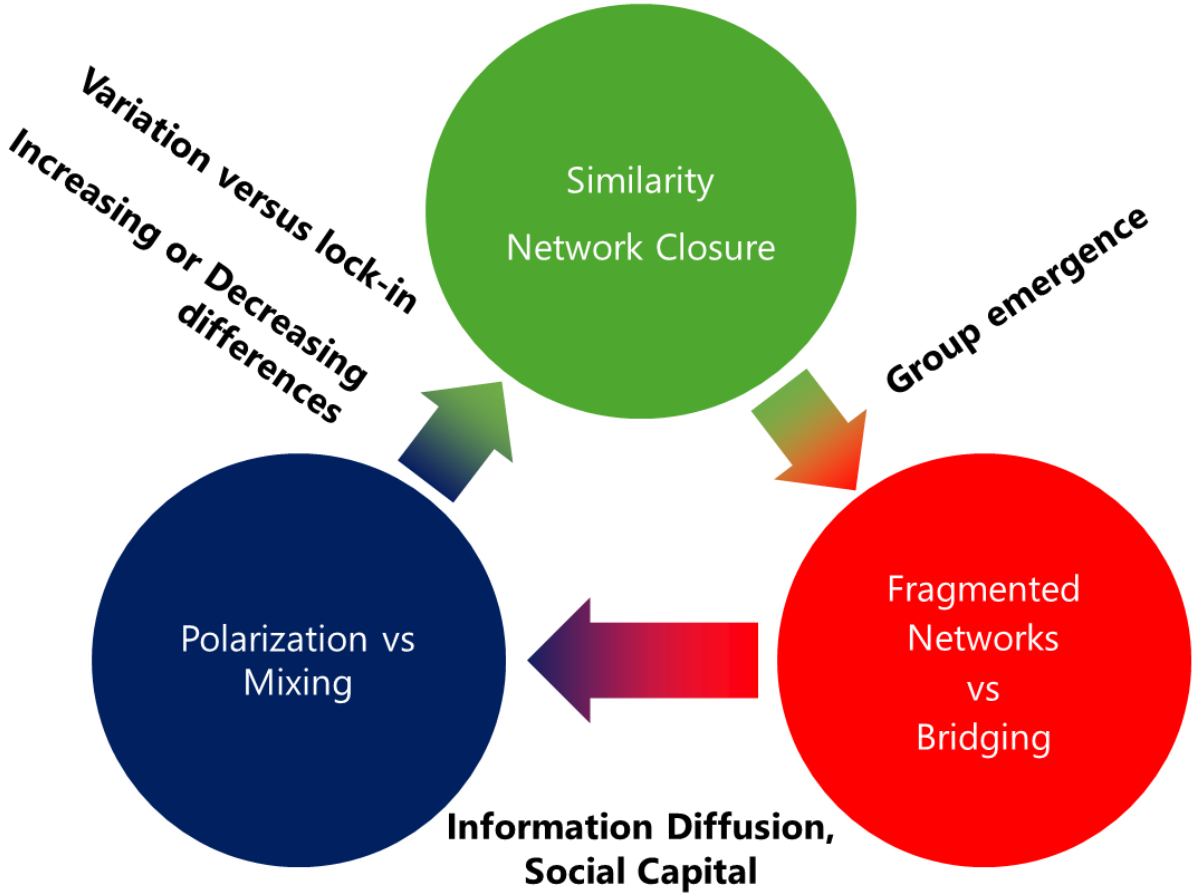
The global society is segregated

Why do we observe growing inequalities
if diffusion of ideas and knowledge
is very quick
in small-world networks?

Prosperous areas next to a slum in Mumbai.

Network fragmentation and inequalities

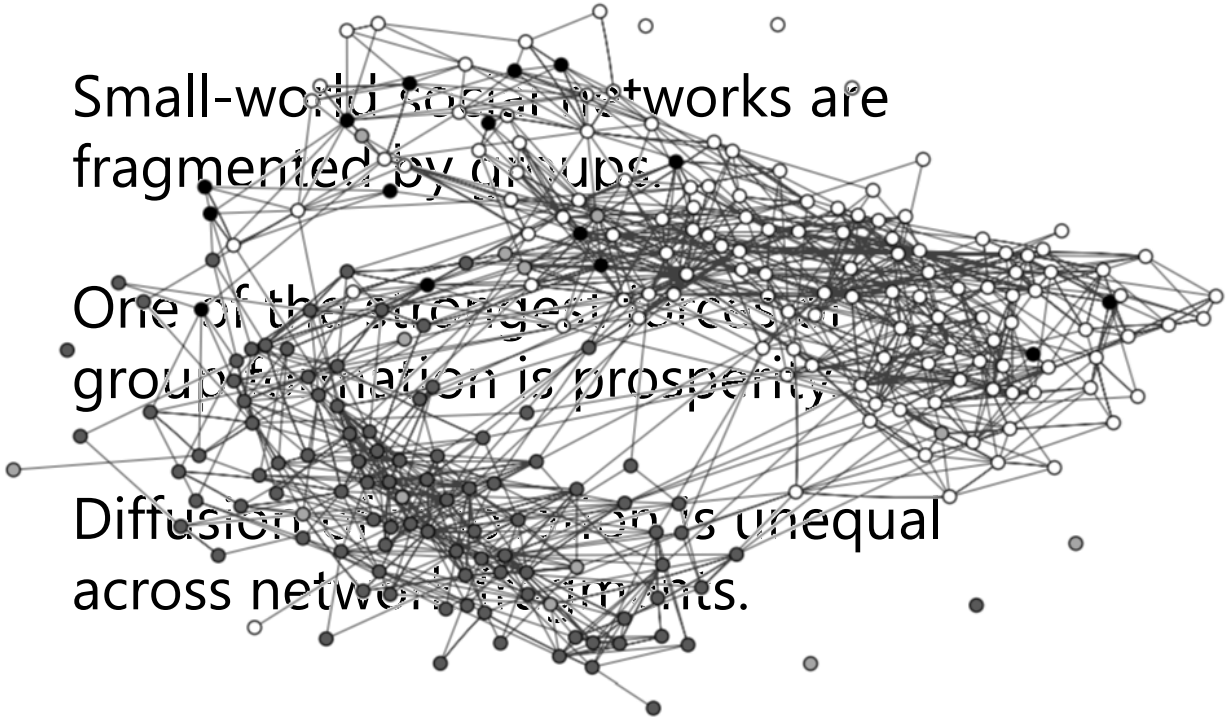
DiMaggio-Garip (2012) Annual Review of Sociology



Small-world social networks are fragmented by groups.

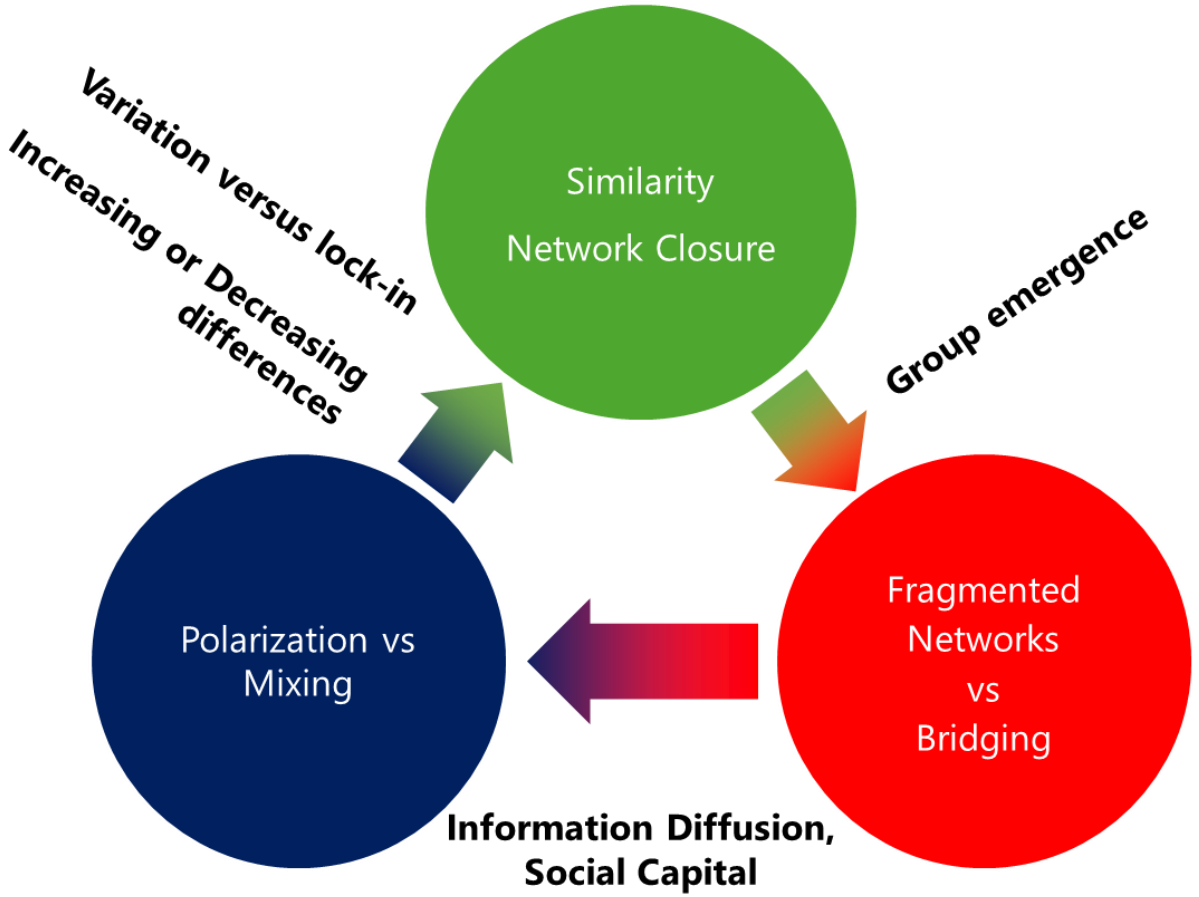
One of the strongest forces of group formation is prosperity.

Diffusion of information is unequal across network segments.



Currarini, Pin, Jackson (2009) Econometrica

Network fragmentation and innovation



Similar specialization in knowledge increases link probability.

- Increases understanding and specialization

Local collaboration networks can become too cohesive and locked-in into technologies – Grabher (1993), Boschma & Frenken (2010), Giuliani (2013), Balland et al. (2016)

Bridging is on purpose

- To collect new knowledge.

In this course

Day 1

1. Network science: concepts and tools to study
2. Social science: Economic relevance and the geography of social and collaboration networks
3. Network science: Community detection
4. Data Lab: network data and visualization in Gephi

Day 2

1. Social science: Social capital
2. Data Lab: Coding basics in R
3. Social science: Social and collaboration networks and regional development

Day 3

1. Network science: Diffusion in networks
2. Network and social science: Spatial diffusion
3. Data Lab: R diffusion with ABM on networks

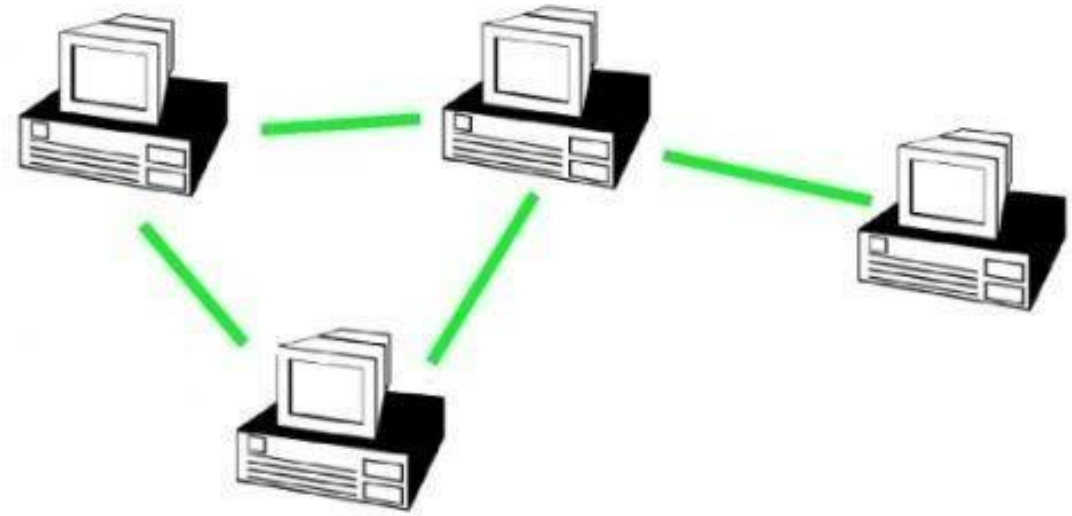
Social Networks and Economic Geography

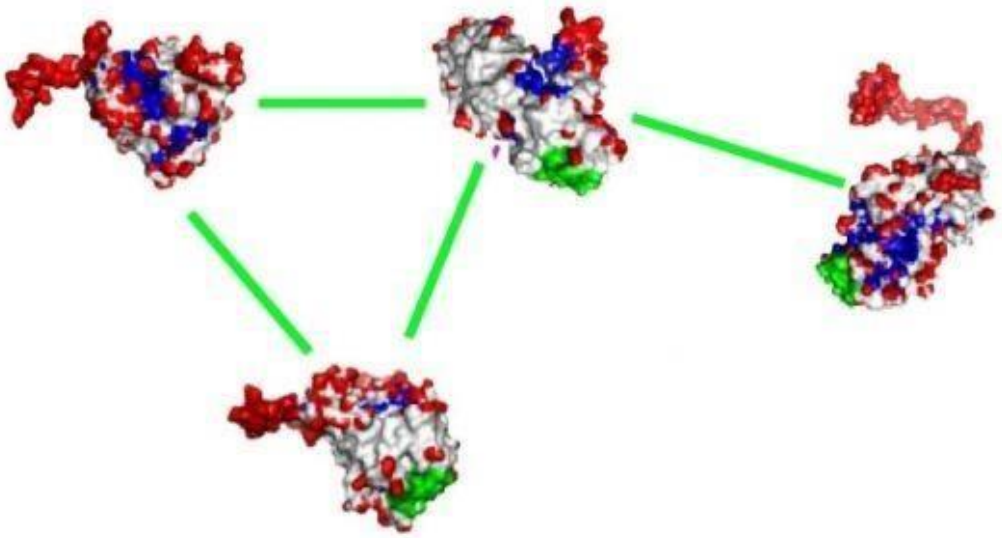
Balázs Lengyel
lengyel.balazs@krtk.hun-ren.hu

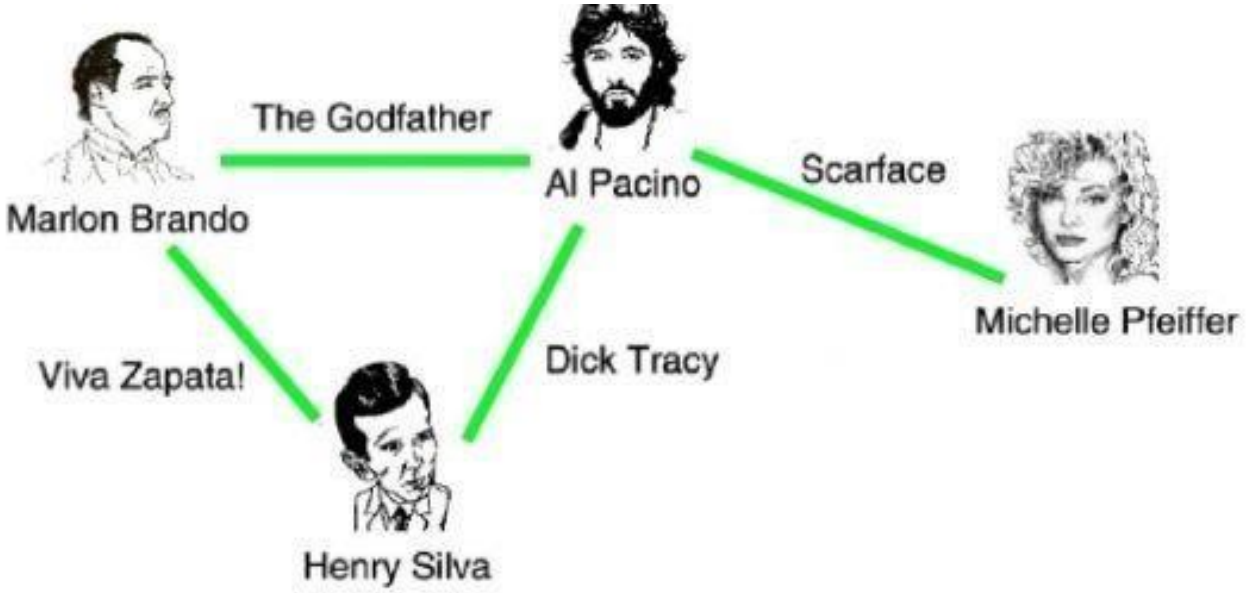
Class 1: Network Science

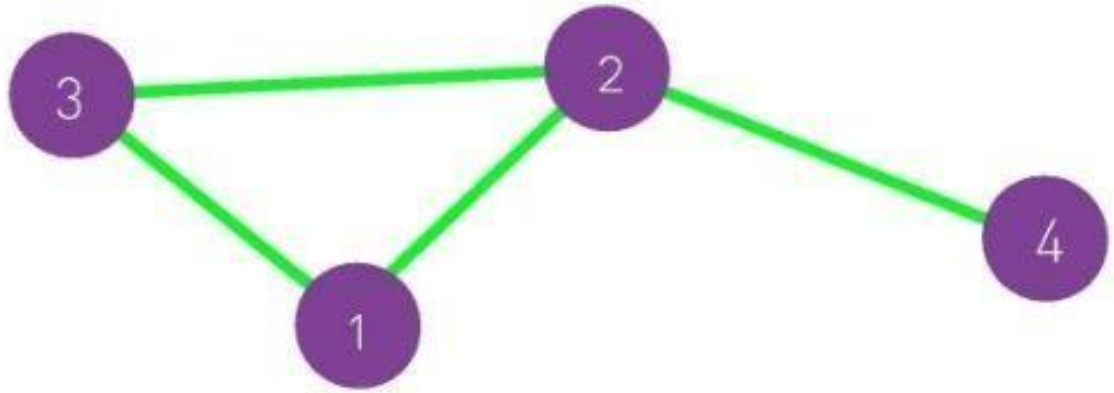
February 17 2024

Networks and Graphs





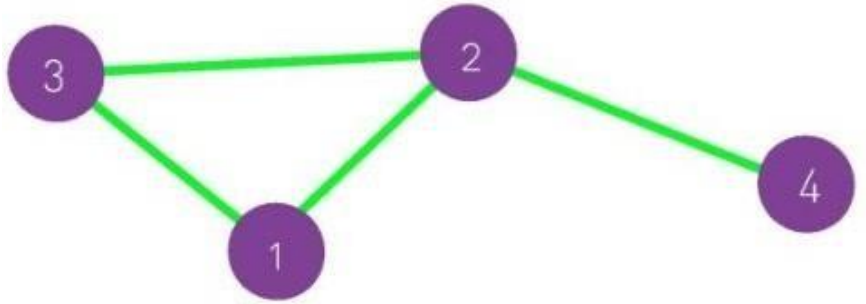




Barabasi (2016) Image 2.2

Network	Nodes	Links	Directed / Undirected	N	L	‹K›
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile-Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorships	Undirected	23,133	93,437	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Papers	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Networks and Graphs



Average Degree

We can express the total number of links, L, as the sum of the node of degrees, $k_1=2$, $k_2=3$, $k_3=2$, $k_4=1$

$$L = \frac{1}{2}\sum_{i=1}^{N} = k_i$$

In undirected networks averagedegree:

$$\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N}$$

In case of directed network the total number of links is a sum of outgoing and ingoing degrees:

$$k_i = k_i^{in} + k_i^{out}$$

$$\langle k^{in} \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i^{in} = \langle k^{out} \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i^{out} = \frac{L}{N}$$

Networks and Graphs

Degree Distribution

Degree Distribution p_k gives the probability that a given node has a degree of k, normalized to 1:

$$\sum_{k=1}^{\infty} p_k = 1$$

,where there are N number of nodes, the normalized distribution can be expressed by:

$$p_k = \frac{N_k}{N}$$

,thus the number of degree-k nodes given by the degree distribution $N_k = Np_k$.

Networks and Graphs

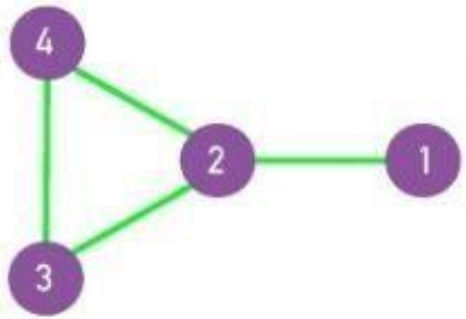
Degree Distribution

Degree Distribution plays a key role in Graph and Network Theory. Most of the structural network properties require the calculation of p_k.

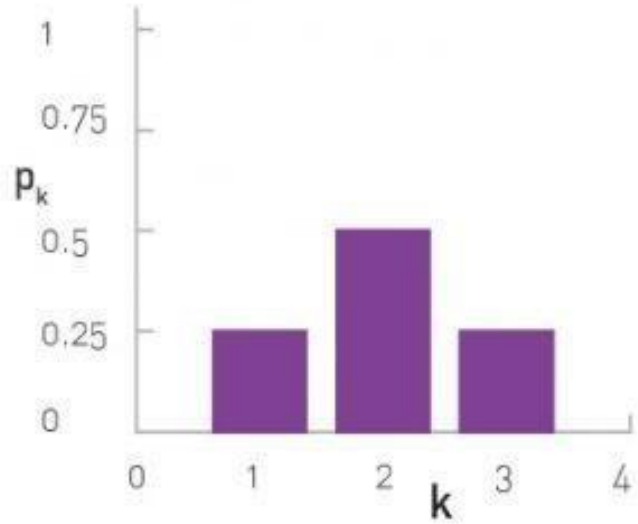
e.g Average Degree can be expressed as a function of Degree Distribution:

$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k$$

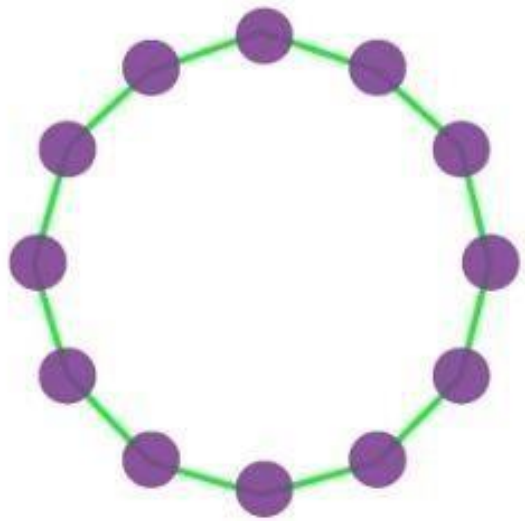
a.



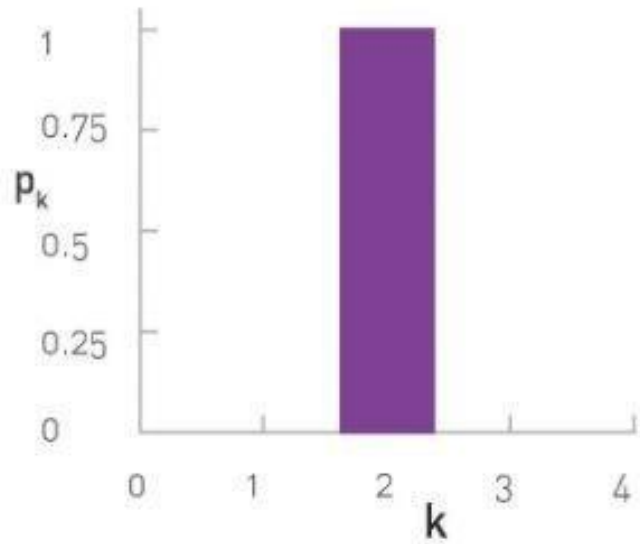
b.

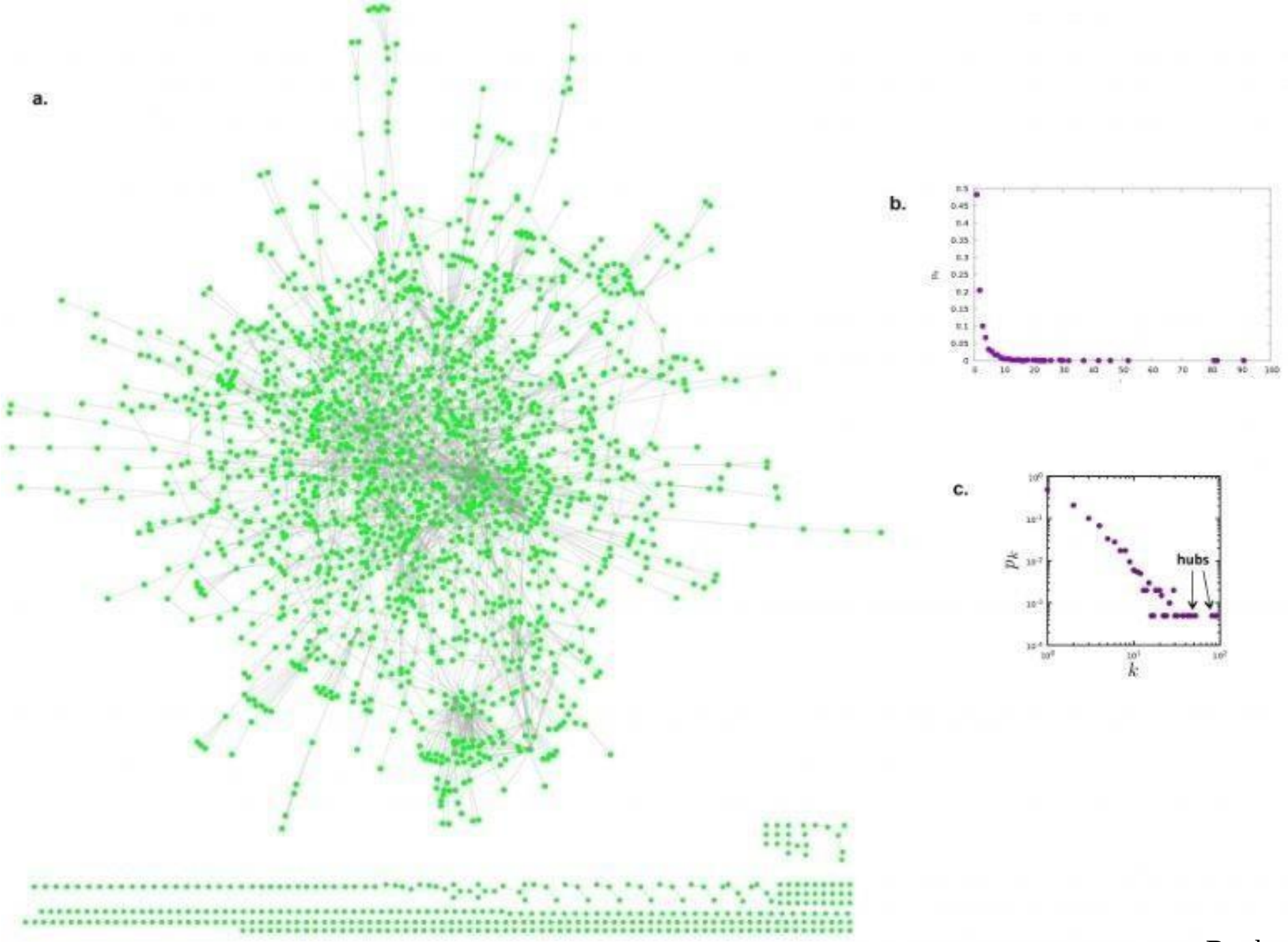


c.



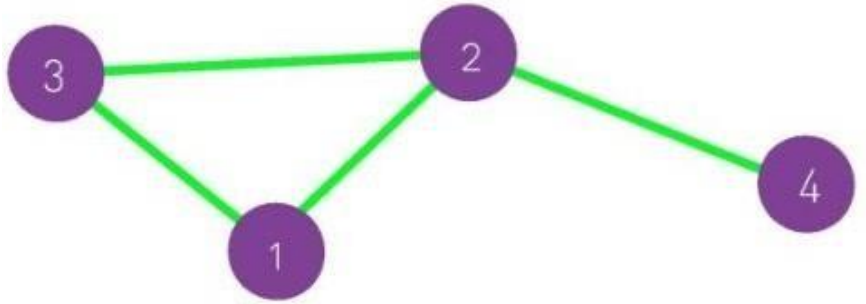
d.





Barabasi (2016) Image 2.4

Adjacency Matrix



To keep track of the links we can provide a complete edge list:
{(1,2);(1,3);(2,3);(2,4)}

We can translate it into an Adjacency Matrix, where network of N nodes has N rows and N columns:

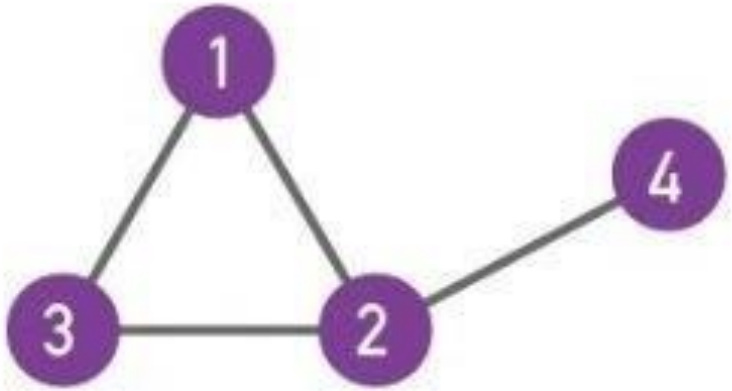
- A_{ij}= 1 if there is a connection between *i* and *j*,
- A_{ij}= 0 no connection between *i* and *j*,
- in an undirected network $A_{ij} = A_{ji}$.

The degree ki can be expressed by the sum of either the rows or the columns of the matrix:

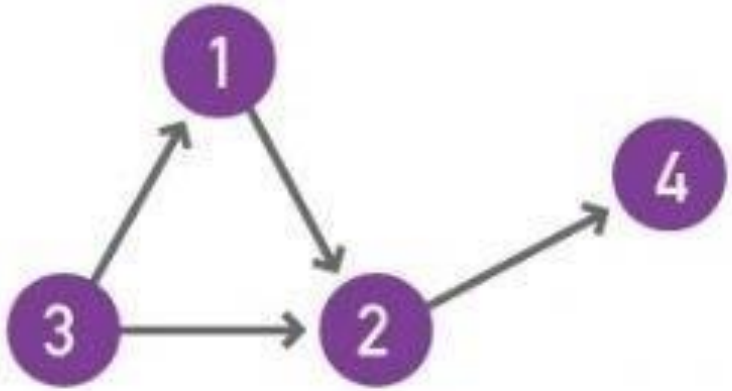
$$k_i = \sum_{i=1}^{N} A_{ij} = \sum_{j=1}^{N} A_{ij}$$

$$2L = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N} k_i^{out} = \sum_{ij}^{N} A_{ij}$$

Adjacency Matrix



$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$



$$A_{ij} = \begin{matrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

Adjacency Matrix

$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

$$k_2 = \sum_{j=1}^{4} A_{2j} = \sum_{i=1}^{4} A_{i2} = 3$$

$$A_{ij} = A_{ji} \qquad A_{ii} = 0$$

$$L = \frac{1}{2}\sum_{i\ i=1}^{N} A_{ij}$$

$$\langle k \rangle = \frac{2L}{N}$$

$$A_{ij} = \begin{matrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

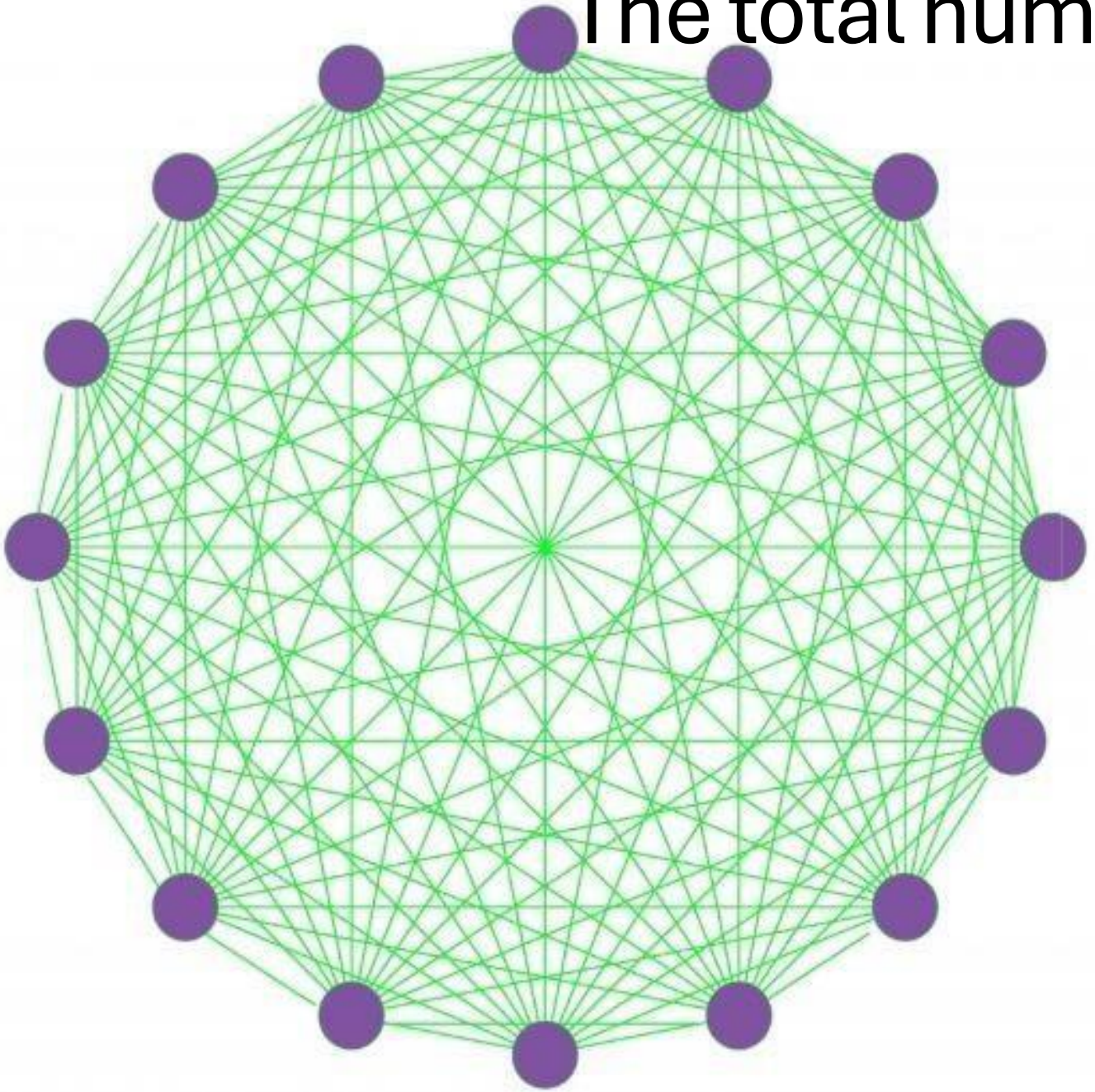
$$k_2^{\text{in}} = \sum_{j=1}^{4} A_{2j} = 2,\ k_2^{\text{out}} = \sum_{i=1}^{4} A_{i2} = 1$$

$$A_{ij} \neq A_{ji} \qquad A_{ii} = 0$$

$$L = \sum_{i,j=1}^{N} A_{ij}$$

$$\langle k^{\text{in}} \rangle = \langle k^{\text{out}} \rangle = \frac{L}{N}$$

The total number of links varies widely.



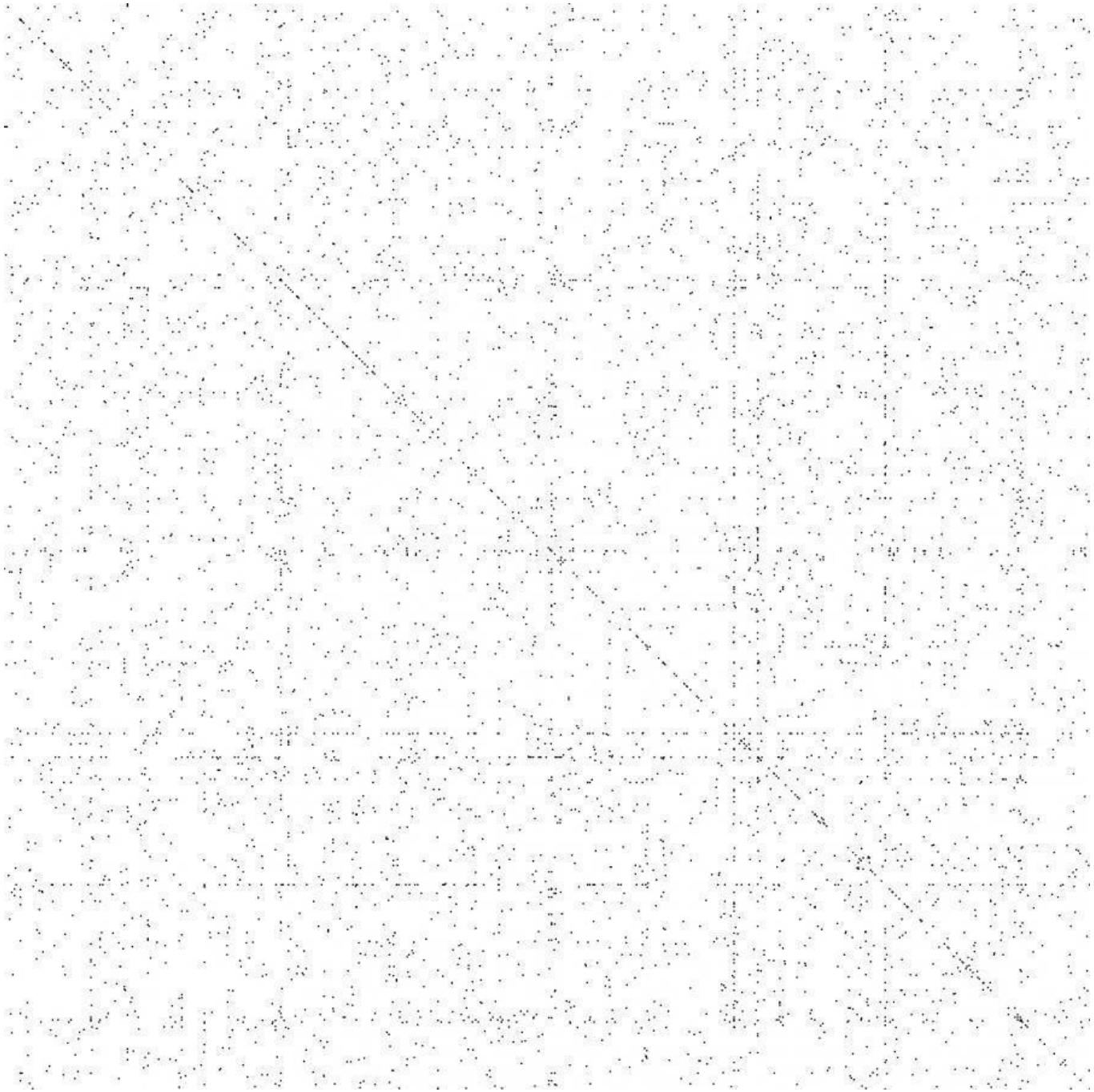
$$L_{min} = 0$$

$$L_{max} = \frac{N(N-1)}{2}$$

Real networks are sparse!

$L << L_{\text{max}}$.

Barabasi (2016) Image 2.6



$$L_{min} = 0$$

$$L_{max} = \frac{N(N-1)}{2}$$

Real networks are sparse!
$L << L_{\max}$.

Overwhelming fraction of elements are zero!

Barabasi (2016) Image 2.7

Weighted Networks

Unweighted Networks

$$A_{ij} = 1$$

A_{ij} = 1 if there connection between *i* and *j*

- is there a phone call?
- any export/import between *i* and *j*?

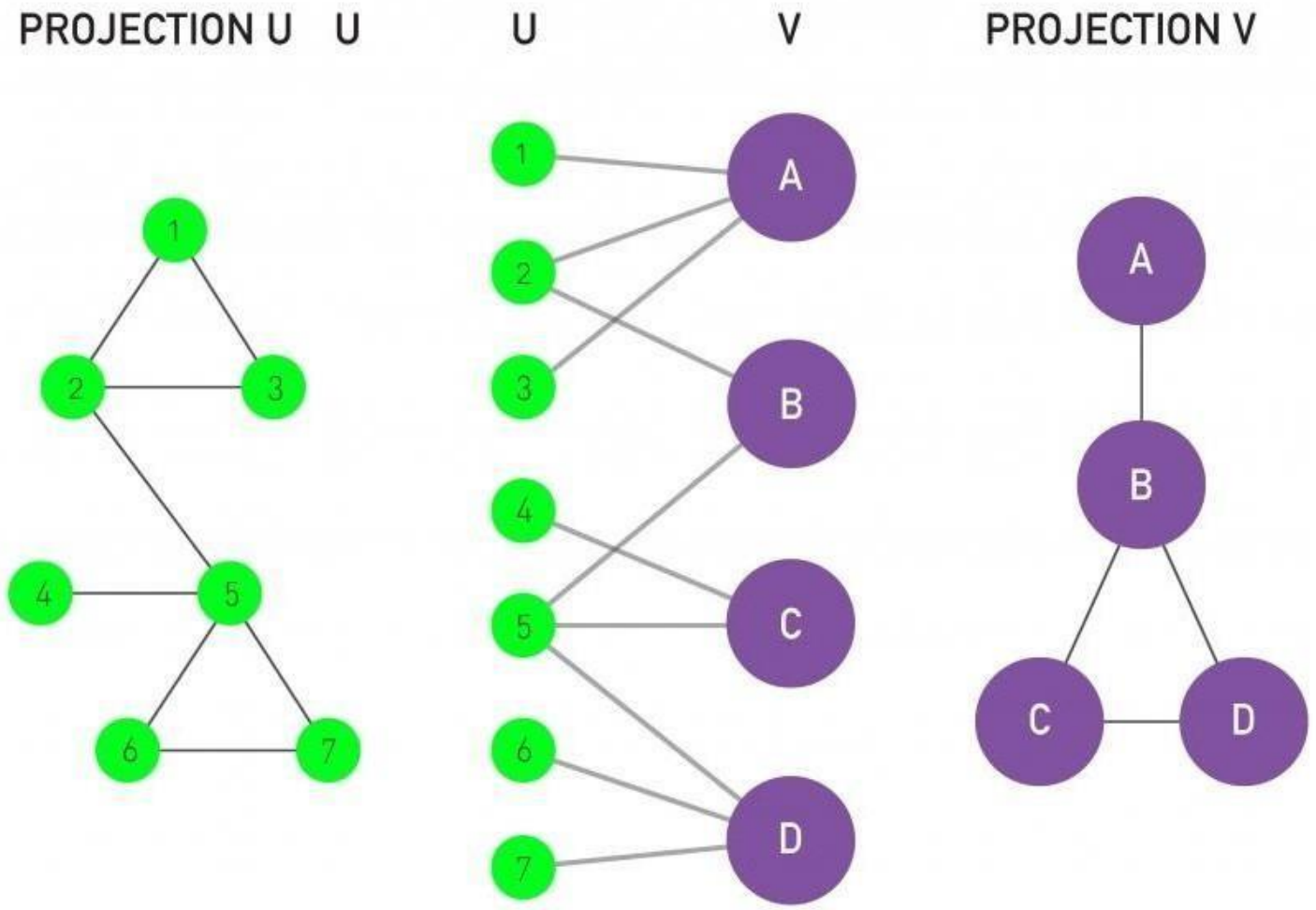
Weighted Networks

$$A_{ij} = w_{ij}$$

A_{ij} = w_{ij} the extent of connection between *i* and *j*

- how many phone calls have happened?
- the extent of export/import between countries?

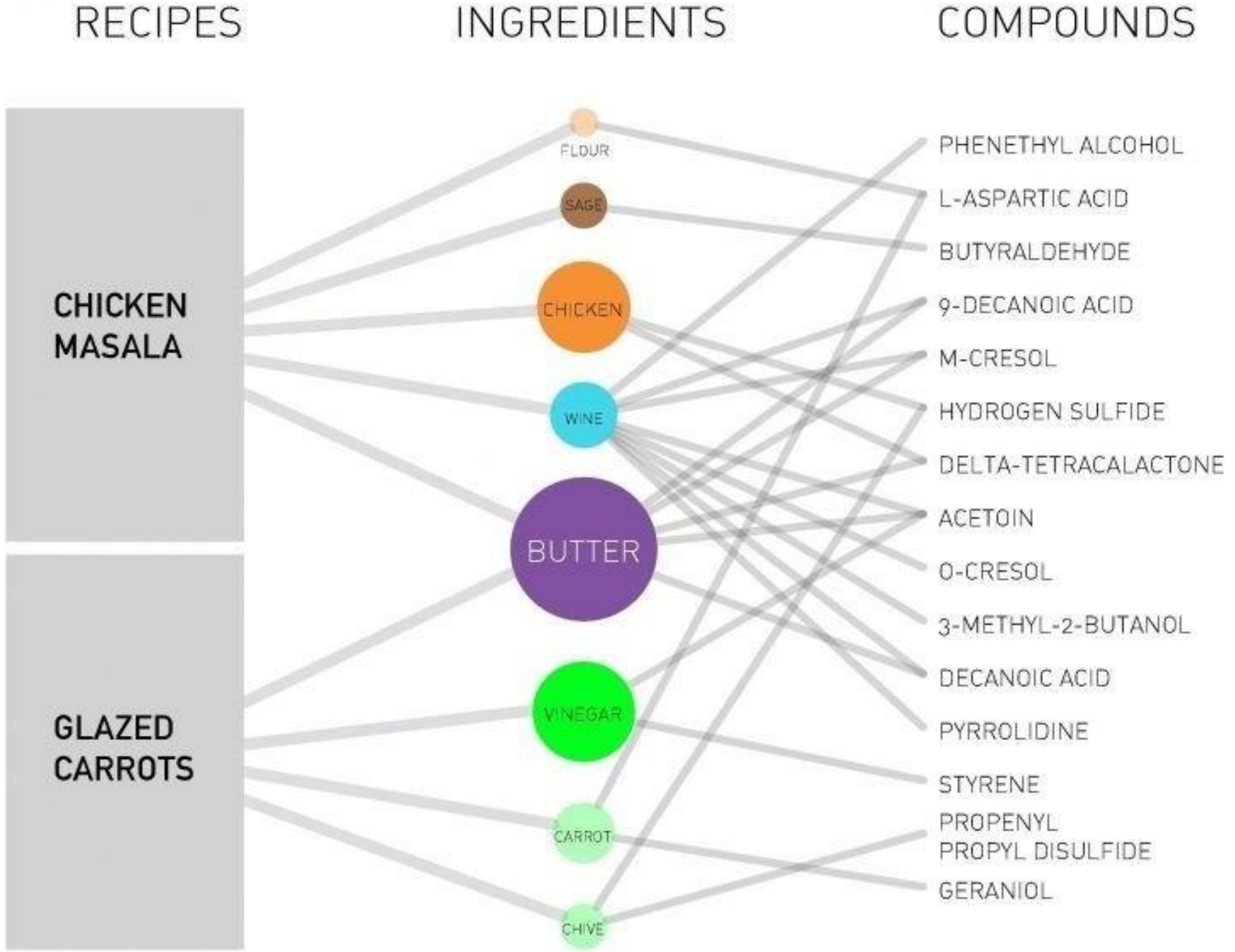
Bipartite Networks



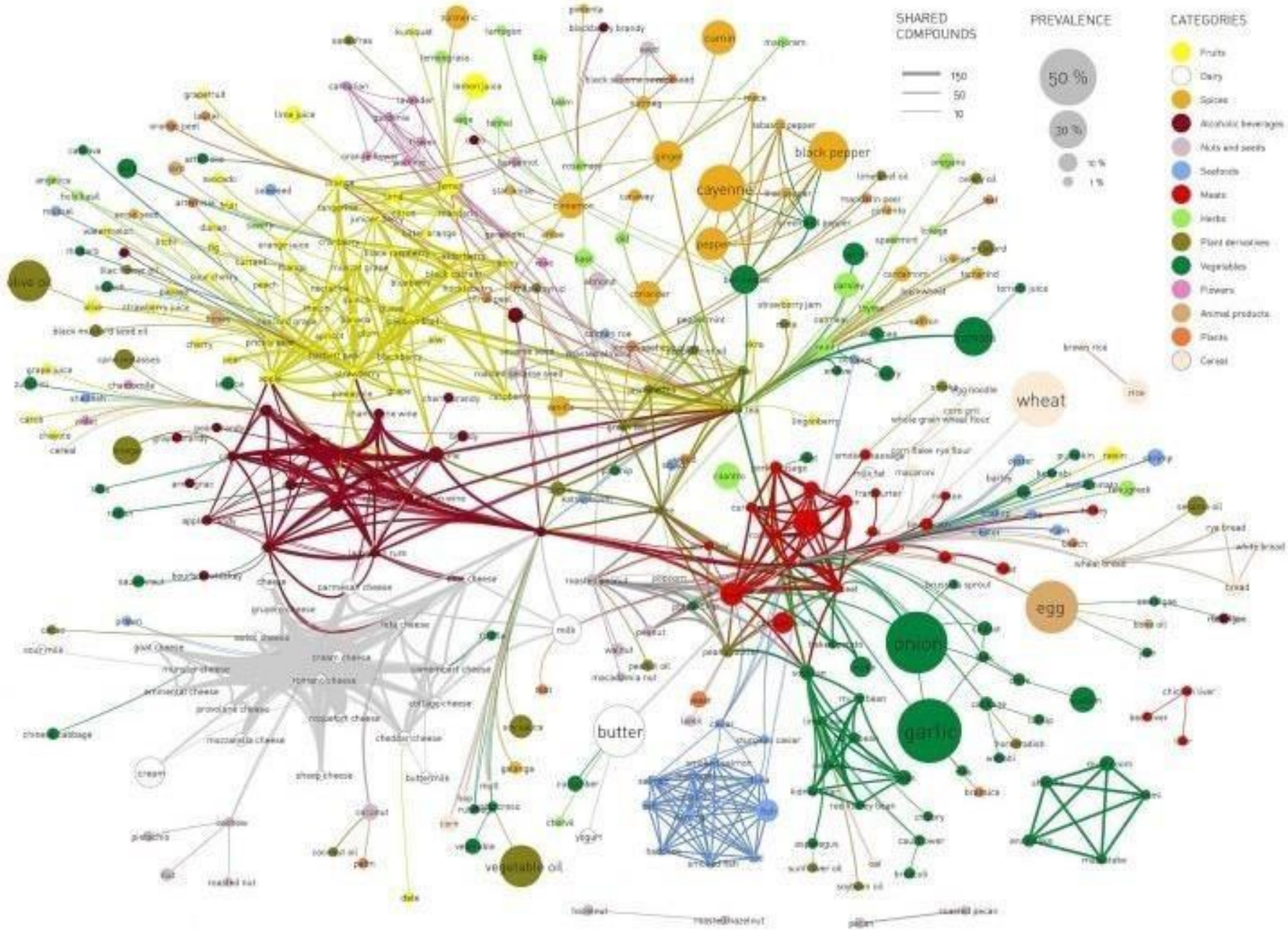
Nodes can be divided into two disjoint sets *U* and *V*

movie network:
U(actors)
V(movies)

co-publication network:
U(inventors/researchers)
V(patents/publications)



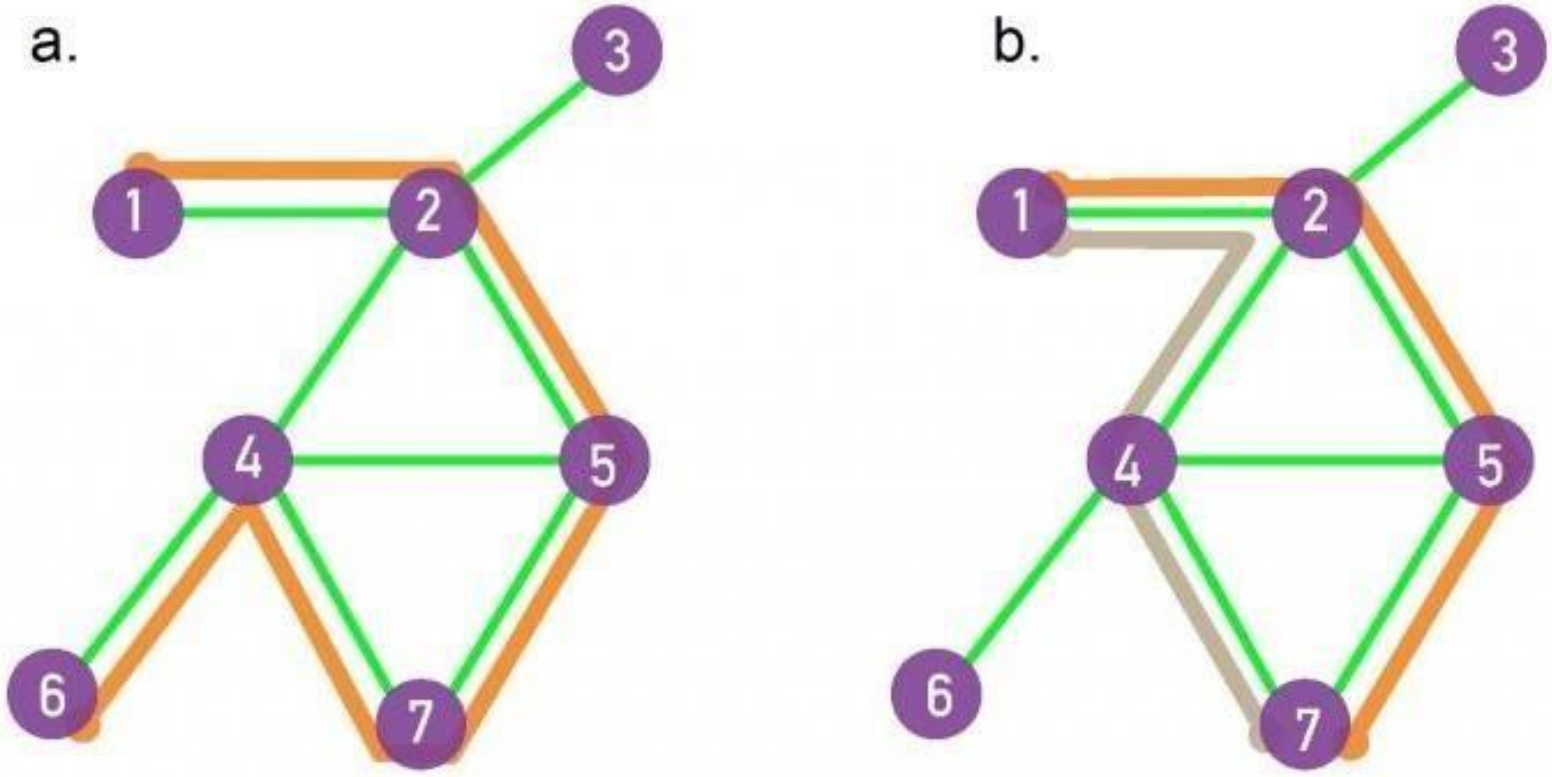
Barabasi (2016) Image 2.11



Barabasi (2016) Image 2.11

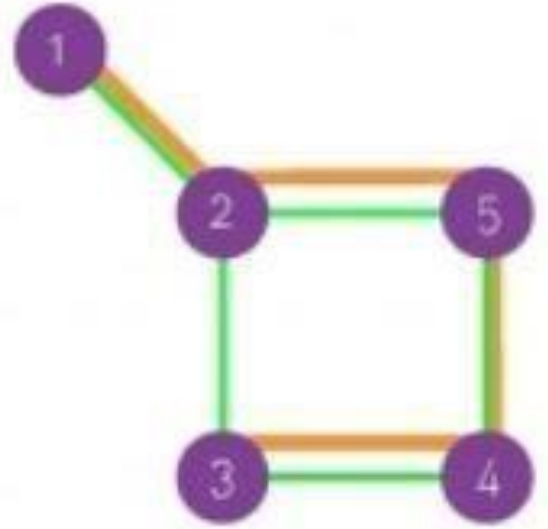
Paths and Distances

To determine the probability of interaction between two components of a system we usually use the physical distance between the agents, e.g. distance between two atoms or distance between two planets.

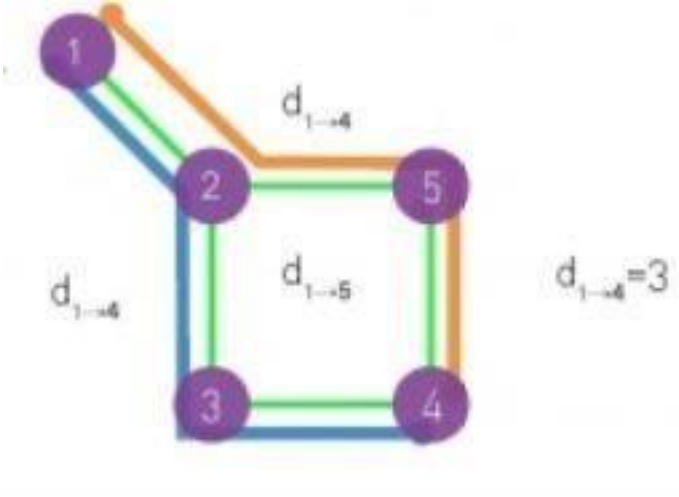


Paths and Distances

- **Path** is a sequence of connected nodes



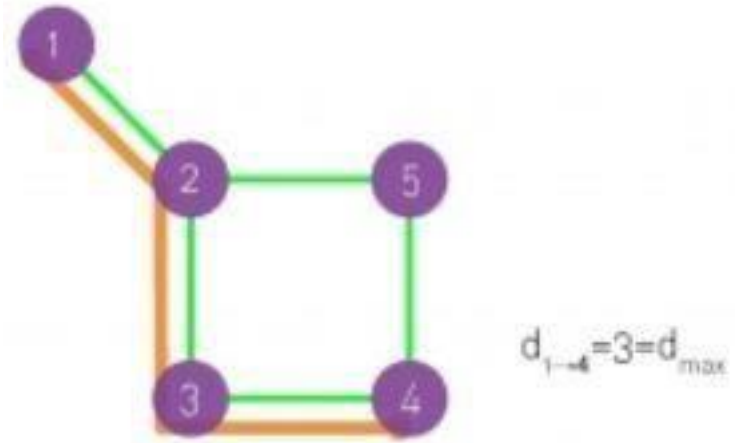
Paths and Distances



- **Path** is a sequence of connectednodes

- **Shortest path** (*d*) is the shortest distance between two nodes.

Paths and Distances



- **Path** is a sequence of connectednodes

- **Shortest path** (d) is the shortest distance between two nodes.

- **Diameter** (d_{max}) is the shortest path between the two furthest points of the graph.
  d_{max} (1,4)=3

Paths and Distances

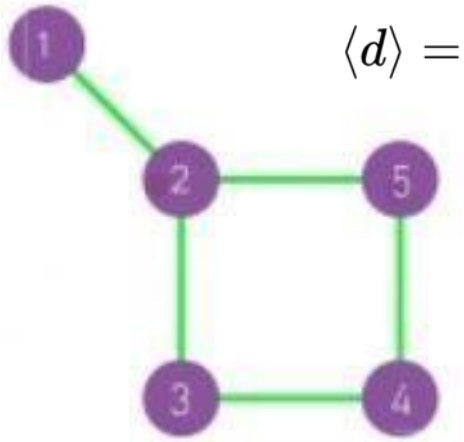
- **Path** is a sequence of connectednodes

- **Shortest path** (*d*) is the shortest distance between two nodes.

- **Diameter** (d_{max}) is the shortest path between the two furthest points of the graph.
  d_{max} (1,4)=3

- **Average Path Length** (<*d*>) is the average of the shortest path between all pair of nodes.
  <d>=1.6

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i,j=1;i\neq j} d_{ij}$$



$\langle d \rangle = [d_{1\to2} + d_{1\to3} + d_{1\to4} + d_{1\to5} + d_{2\to3} + d_{2\to4} + d_{2\to5} + d_{3\to4} + d_{3\to5} + d_{4\to5}]/10 = 1.6$

Clustering

Clustering coefficient shows the degree to which the neighbors of a given node connected to each other.

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

,where L is the number of links between k_i neighbors of node i.

$C_i = 0$ if there is no connection between the neighbors of i.
$C_i = 1$ if the neighbors of i form a complete graph.
$C_i = 0.5$ implies that there is a 50% chance that two neighbors of i are connected to each other.

Average Clustering Coefficient captures the degree of clustering of the whole network, or in other words it shows the probability that two neighbors of a randomly selected node connected to each other:

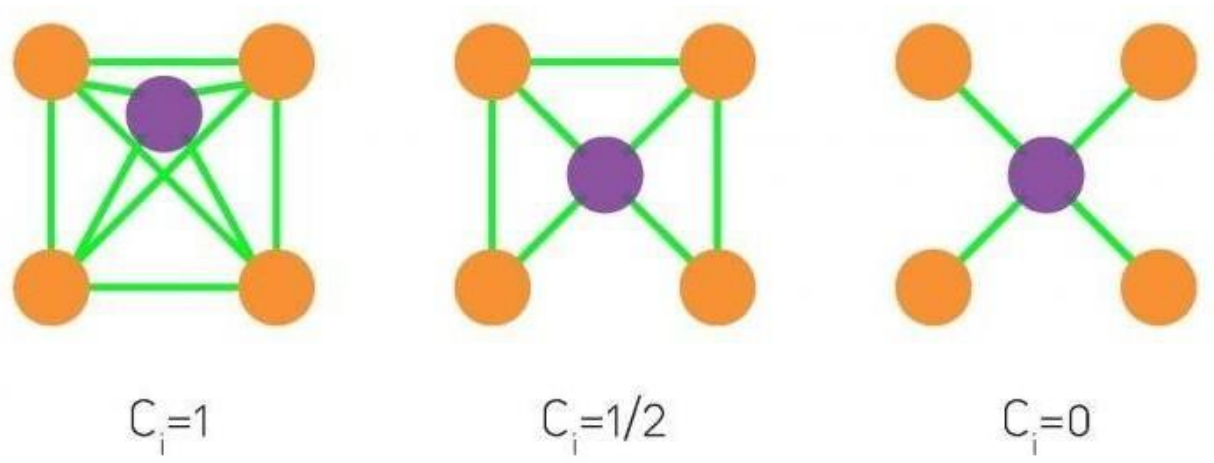
$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

Clustering

Clustering Coefficient

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

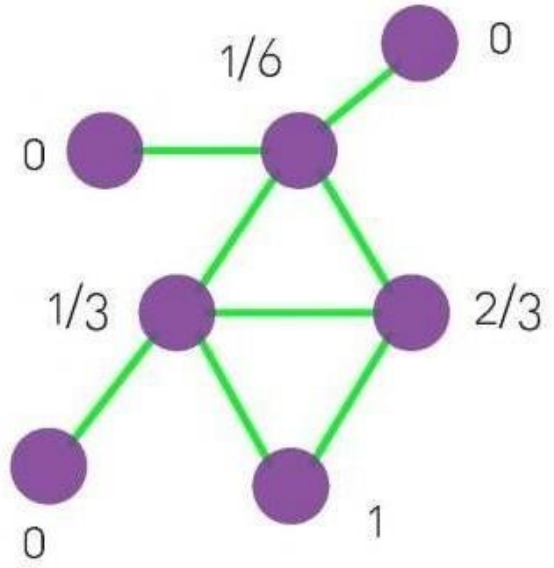
a.



Average Clustering Coefficient

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

b.



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

Global Clustering Coefficient

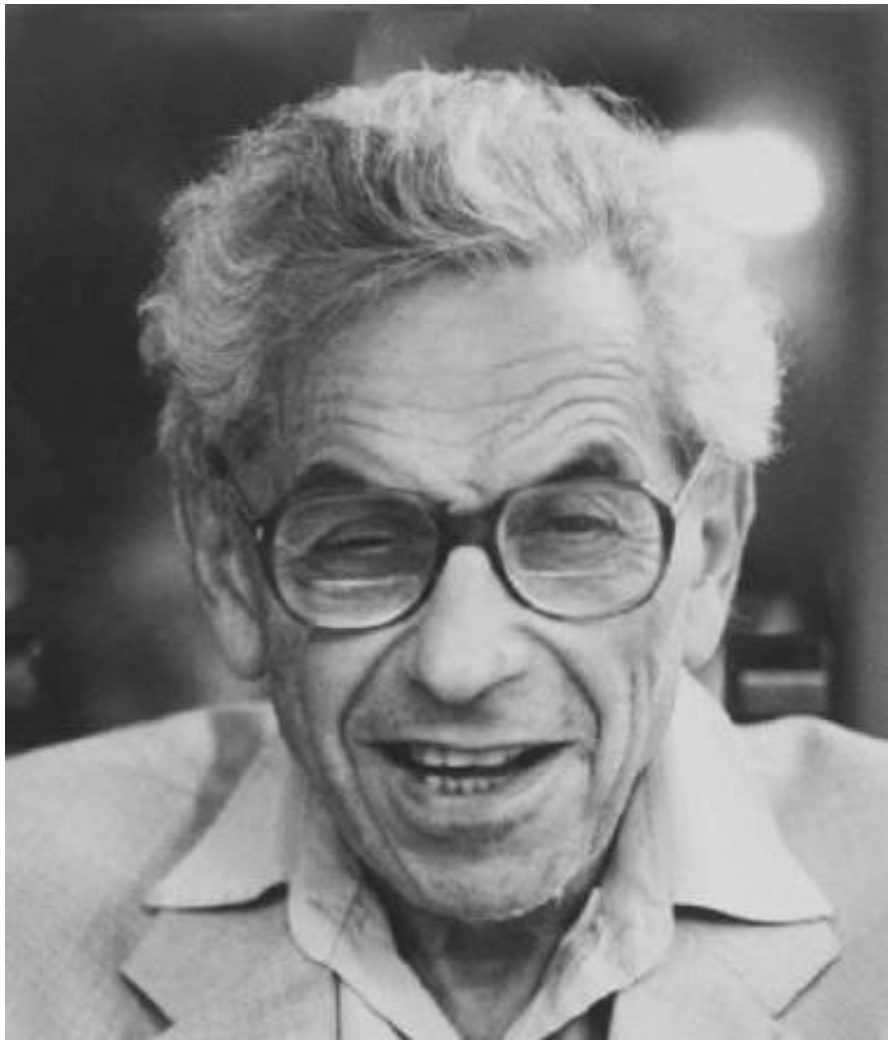
$$C_\Delta = \frac{3 \times \text{Number of } \Delta}{\text{Number of Connected Triples}}$$

$$C_\Delta = \frac{3}{8} = 0.375$$

Pál Erdős

Alfréd Rényi

"A mathematician is a device for turning coffee into theorems"

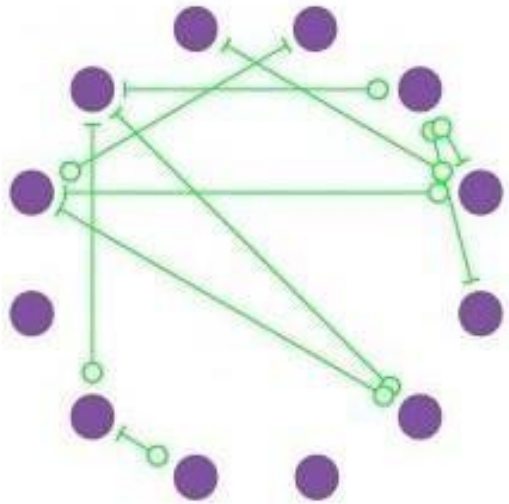


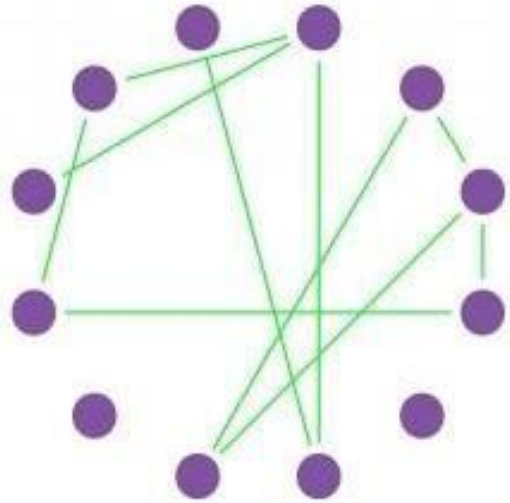
Random Network

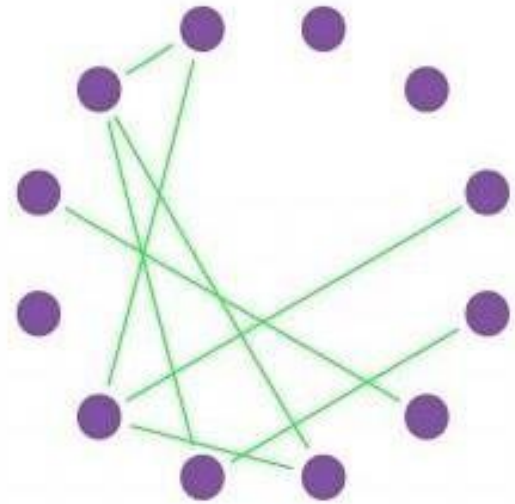
G(N,L): have N number of nodes connected by L number of random connections.
Erdős-Rényi network

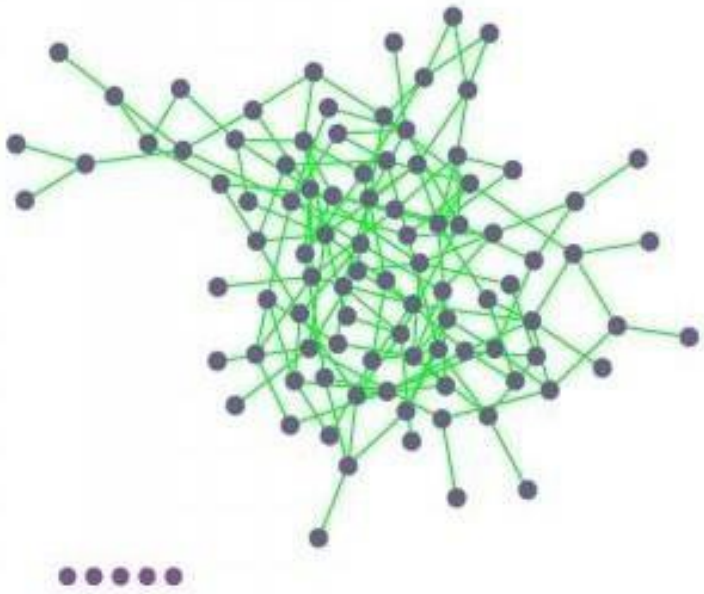
G(N,*p*): Between each pair of N_i and N_j there is a probability (*p*) of connection.
Gilbert

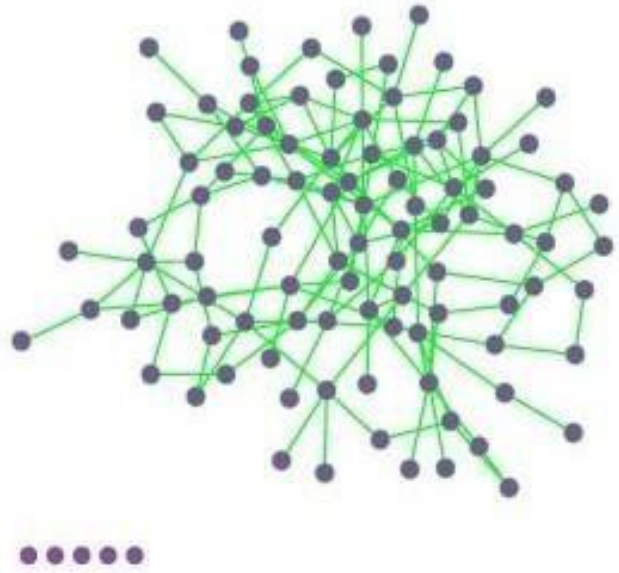
G(N,*p*) model fixes the probability of connections, while the G(N,L) fixes the total number of connections in the network.
In G(N,L) model the average degree easily calculable: $\langle k \rangle = \frac{2L}{N}$, but the number of connections are rarely fixed, so in other cases we use the G(N,*p*) model.

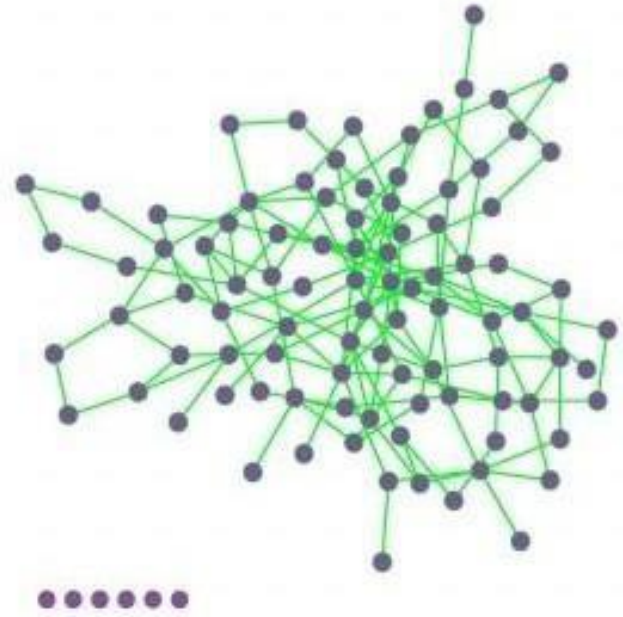












Evolution of a Random Network

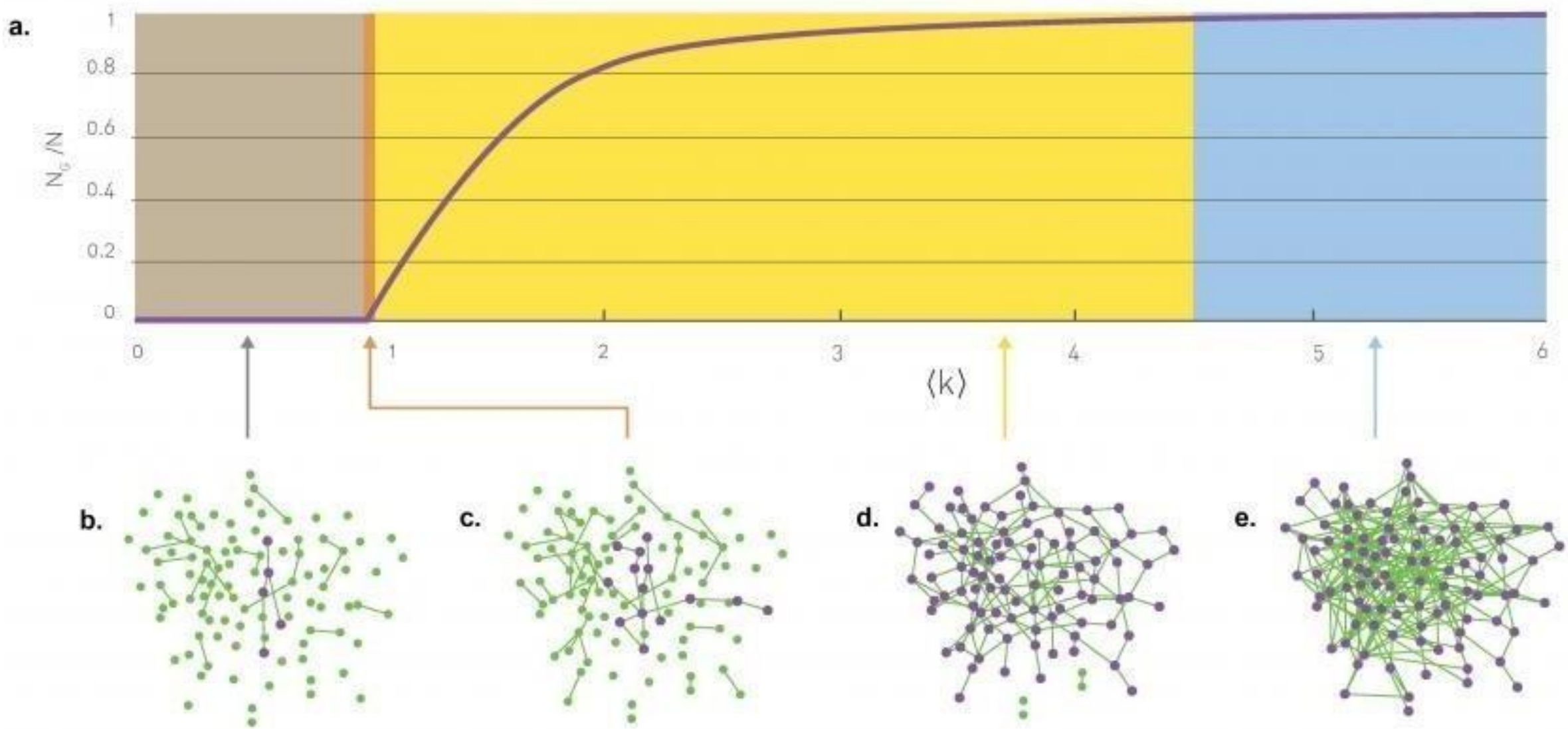
We have two extreme cases:

- $p = 0$, all the nodes are isolated, then the largest component $N_G = 1$ and N_G/N tends to 0.

- $p = 1$, therefore $\langle k \rangle = N-1$, then the network is a complete graph and all nodes belong to the giant component $N_G = N$, thus $N_G/N = 1$.

Once $\langle k \rangle$ exceeds a critical value N_G/N increases, which means that a large cluster - what we call as a Giant Component - emerges.

Critical point: $\langle k \rangle = 1$
Subcritical regime : $0 < \langle k \rangle < 1$
Supercritical: $\langle k \rangle > 1$
Connected Regime: $\langle k \rangle > \ln N$

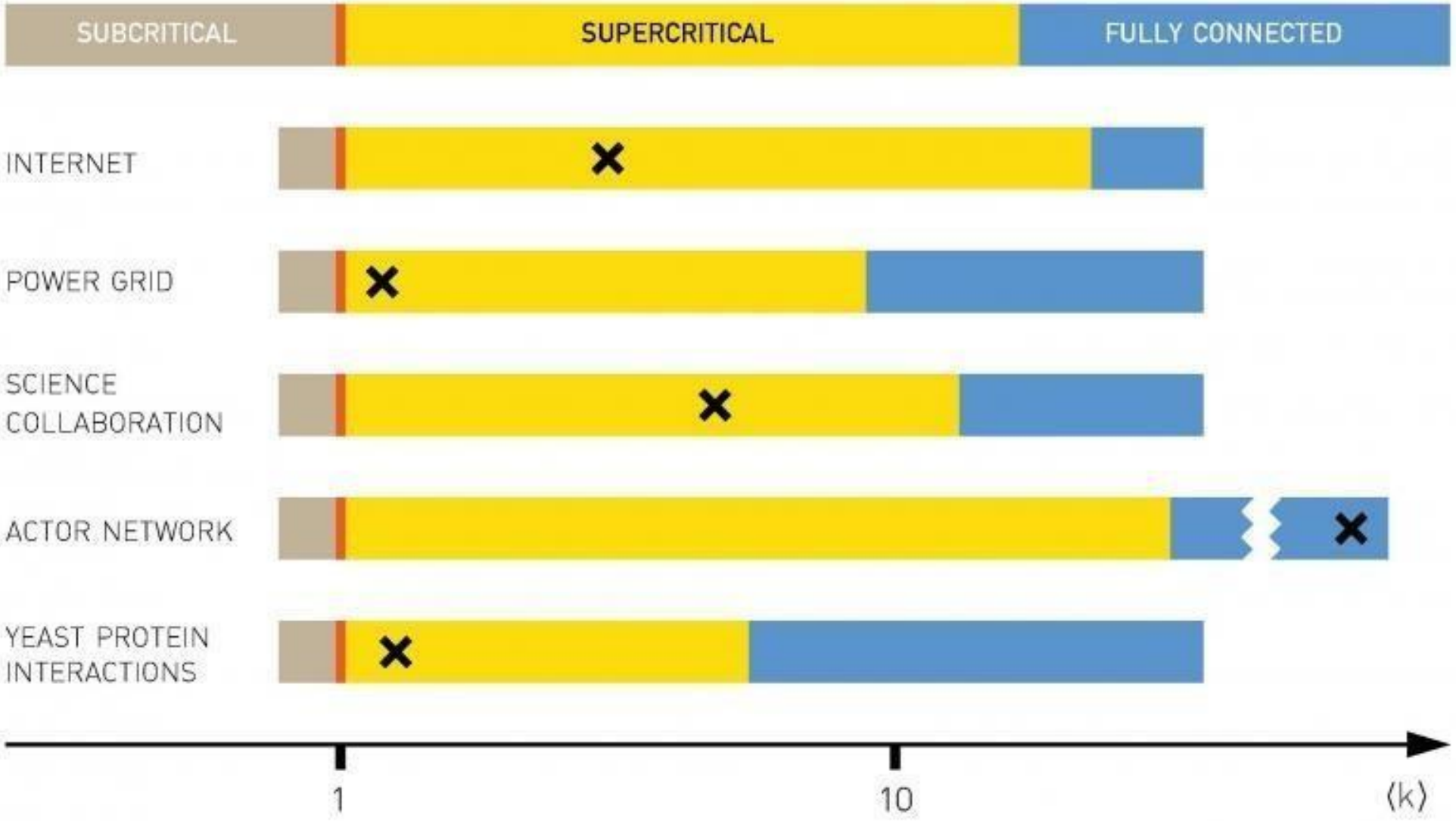

a.
1
0.8
0.6
0.4
0.2
0
N_G/N
0
1
2
3
4
5
6
⟨k⟩
b.
c.
d.
e.

Real Networks are Supercritical

Critical point: <k> =1
Subcritical regime : 0 < <k> <1
Supercritical: <k> > 1
Connected Regime: <k> > lnN

Network	N	L	‹K›	lnN
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	94,437	8.08	10.05
Actor Network	702,388	29,397,908	83.71	13.46
Protein Interactions	2,018	2,930	2.90	7.61

Real Networks are Supercritical



Small World Networks

Six degrees of separation.

The distance between two randomly chosen node is short.

- What does short mean?
- Short compare to what?

Consider a random network with average degree of <k>, then a node has on average:

- <k> nodes at distance d=1,
- $\langle k \rangle^2$ nodes at distance d=2

  ...
- $\langle k \rangle^d$ nodes at distance d.
-

So if we assume that a social network has a <k> ~ 1000, then we know 10^6 people at d=2, and the whole earth population is at d=3.

"The worker knows the manager in the shop, who knows Ford; Ford is on friendly terms with the general director of Hearst Publications, who last year became good friends with Árpád Pásztor, someone I not only know, but to the best of my knowledge a good friend of mine."

Karinthy, 1929

"Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice. It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought. How every person is a new door, opening up into other worlds."

Guare, 1991

Manfred Kochen

Ithiel de Sola Pool

Stanley Milgram

John Guare

Duncan J. Watts

Steven Strogatz

4-DEGREE OF SEPARATION

6-DEGREE OF SEPARATION

DISCOVERY

PUBLISHED 20 YEARS LATER

MILESTONES

PUBLICATION DATE

WWII

XXI

1929 1935 1940 1945 1950 1958 1960 1967 1970 1978 1980 1985 1991 1998 1999 2000 2005 2011

Frigyes Karinthy (1887-1938)
Hungarian writer, journalist and playwright, the first to describe the small world property. In his short story entitled 'Láncszemek' (Chains) he links a worker in Ford's factory to himself [26, 27].

Manfred Kochen (1928-1989), **Ithiel de Sola Pool** (1917-1984)
Scientific interest in small worlds started with a paper by political scientist Ithiel de Sola Pool and mathematician Manfred Kochen. Written in 1958 and published in 1978, their work addressed in mathematical detail the small world effect, predicting that most individuals can be connected via two to three acquaintances. Their paper inspired the experiments of Stanley Milgram.

Stanley Milgram (1933-1984)
American social psychologist who carried out the first experiment testing the small-world phenomena. (BOX 3.7).

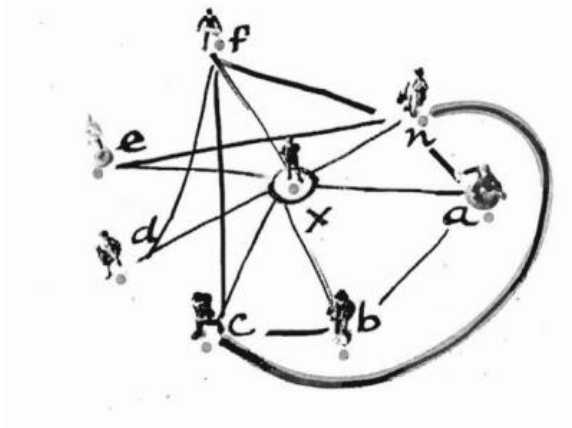
19 Degrees of the WWW
Measurements on the WWW indicate that the separation between two randomly chosen documents is 19 [21] (Box 3.6).

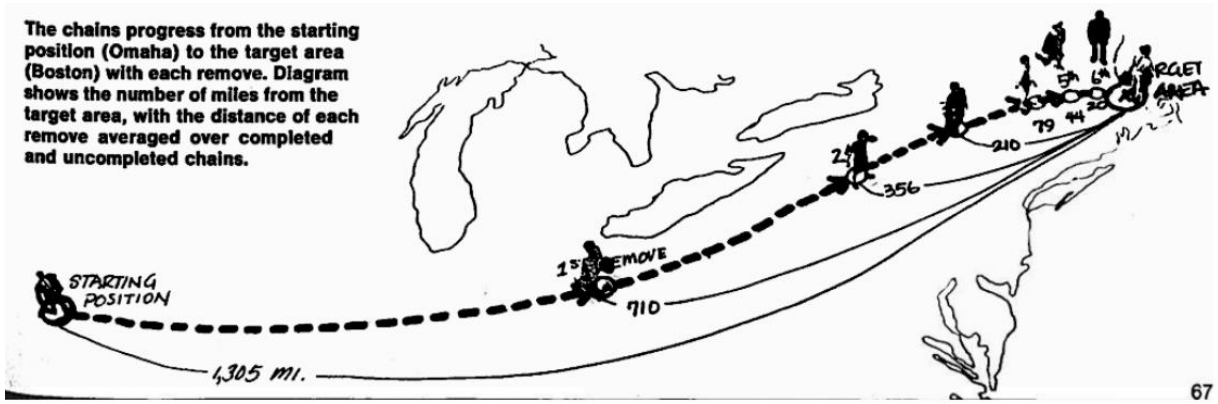
John Guare (1938)
The phrase 'six degrees of separation' was introduced by the playwright John Guare, who used it as the title of his Broadway play [28].

The **Facebook Data Team** measures the average distance between its users, finding "4 degrees" (BOX 3.7).

Duncan J. Watts (1971), **Steven Strogatz** (1959)
A new wave of interest in small worlds followed the study of Watts and Strogatz, finding that the small world property applies to natural and technological networks as well [29].

Small-world networks





Milgram, S. (1967) The Small-world Problem. *Psychology Today*, 1, 1, 61-67.

Small World Networks

Clustering Coefficient

Degree of a node does not include information about the neighbors of the node. Local clustering measures the density of connections between the immediate neighbors of the ego. In a random network to calculate C_i first we need to know the expected number of a links of the ego network of i:

$$\langle L_i \rangle = p\frac{k_i(k_i-1)}{2}$$

Thus, the local clustering coefficient in a random random network can be expressed by:

$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i-1)} = p = \frac{\langle k \rangle}{N}$$

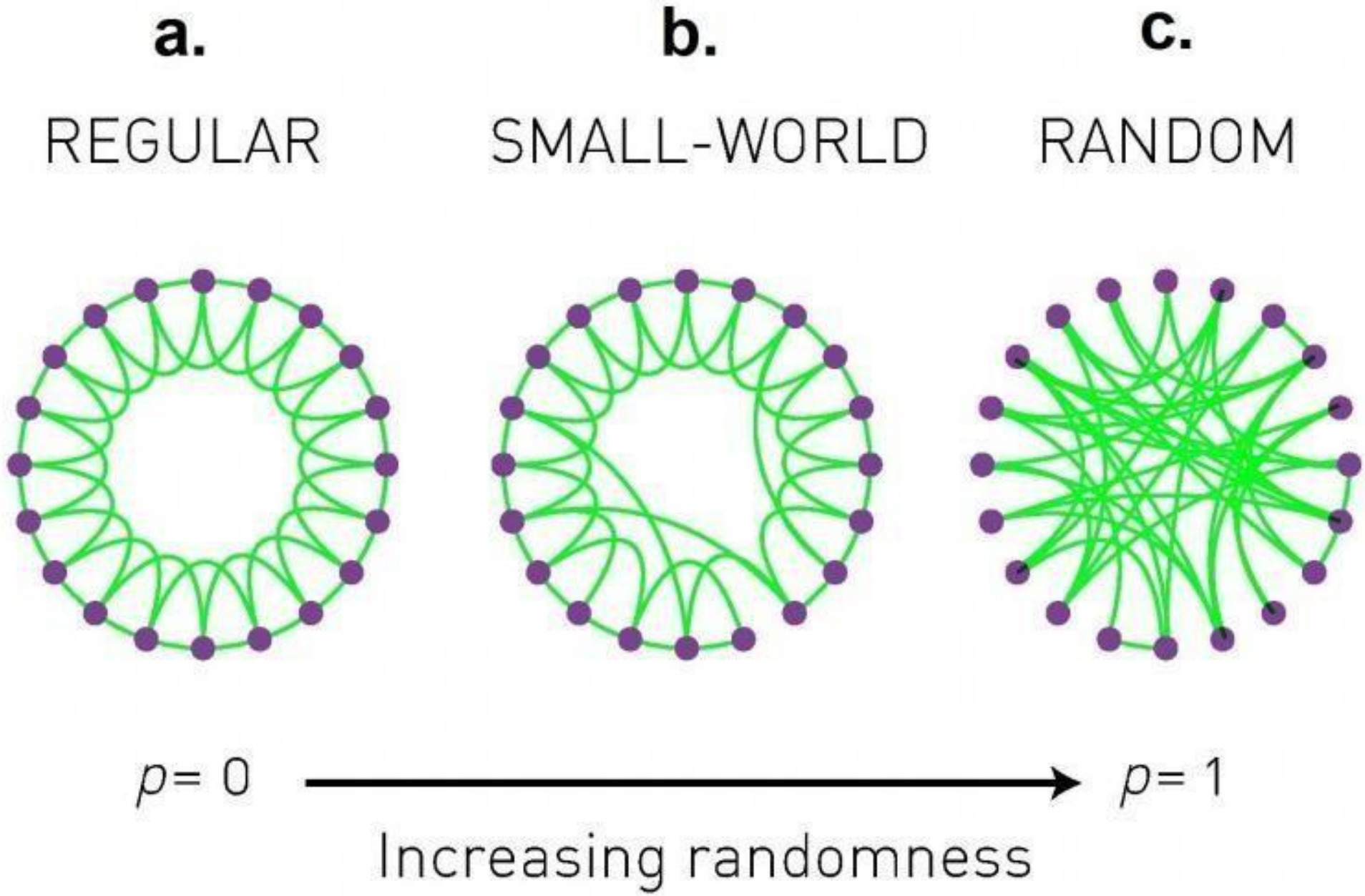
- If we fix <k>, the larger the network, the smaller is the node's local clustering.
- Local clustering coefficient of a node is independent from the number of connections of the node.

Small World Networks

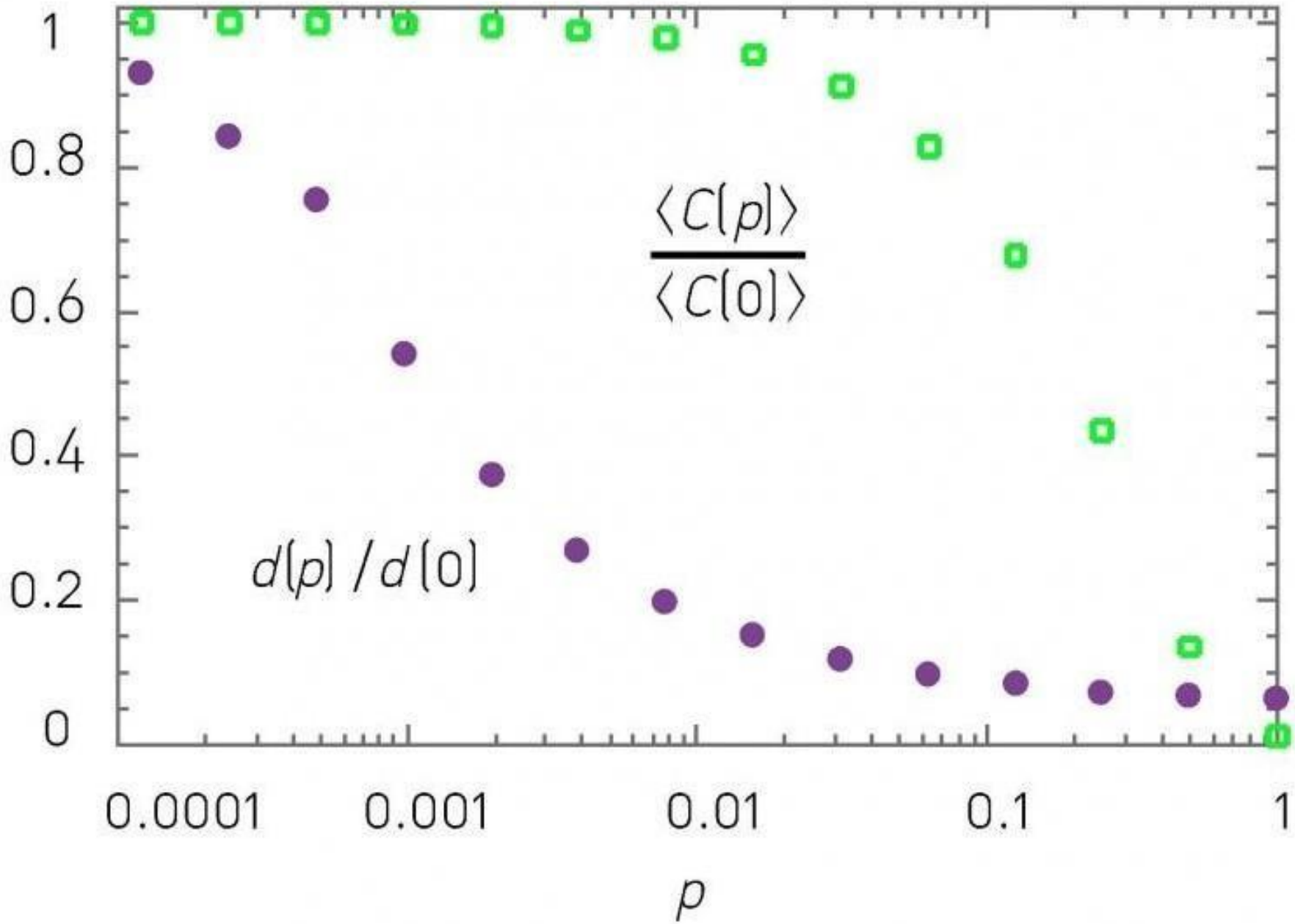
Watts-Strogatz Model

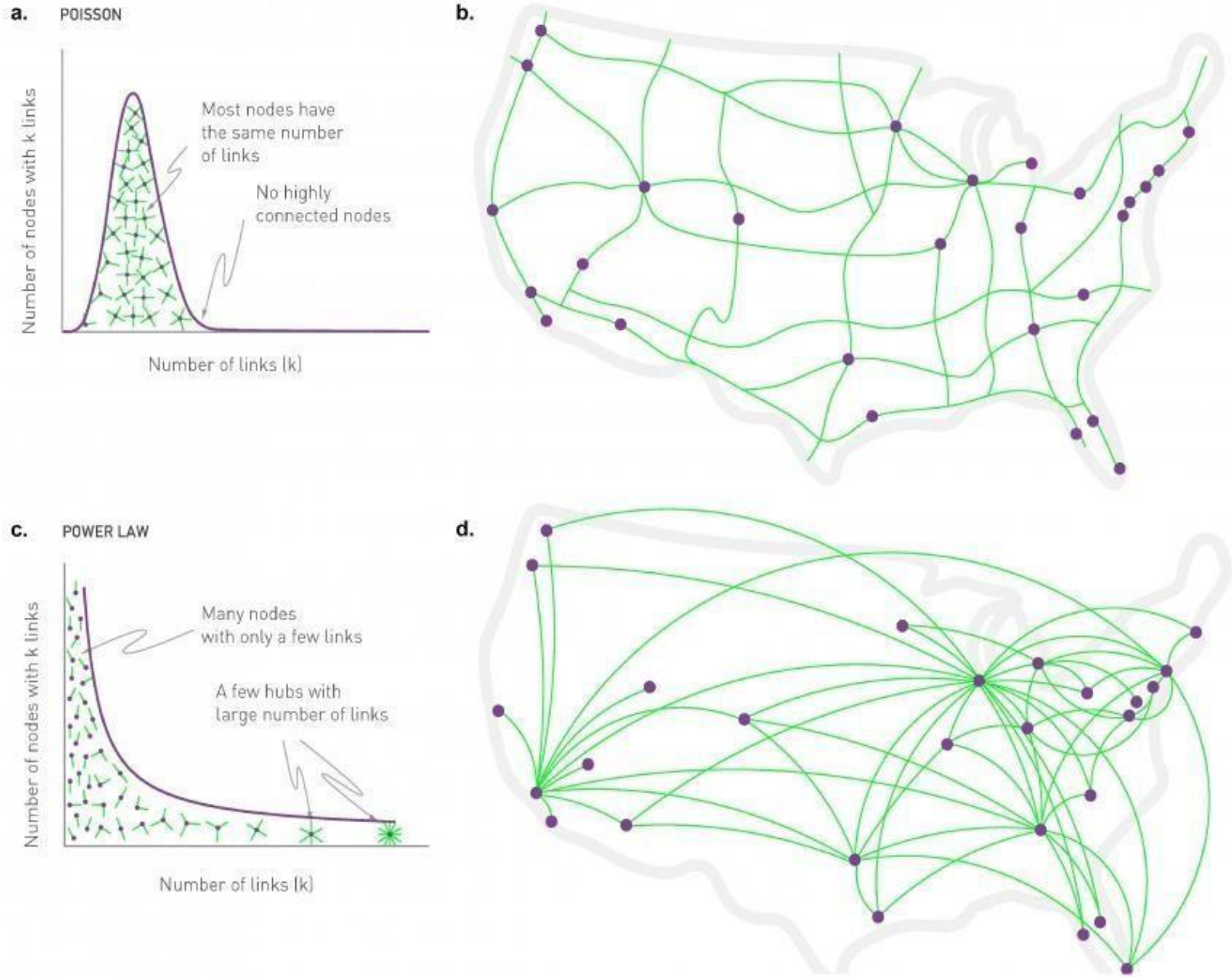
What gives the word *small* in Small World Networks?
In real networks the average distance between two points is a logarithmic function of N.

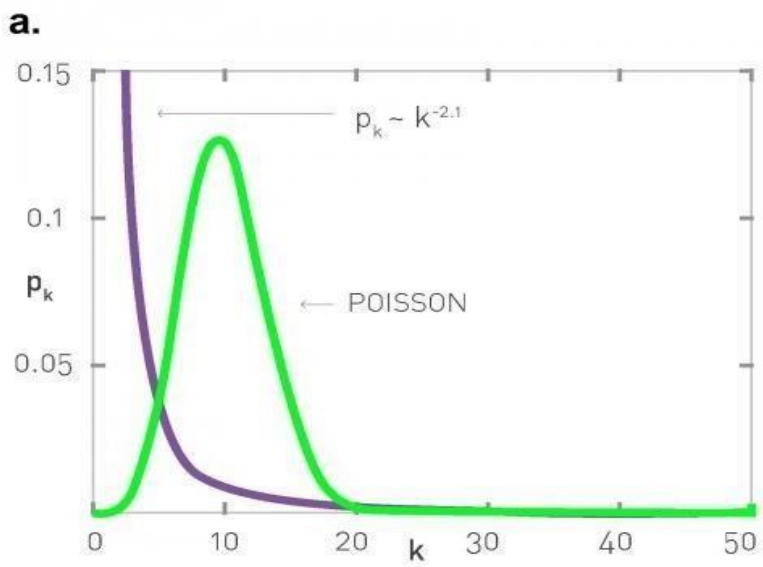
What makes it a real life network?
The average clustering coefficient of a real network is much higher than we would expect from a random network with the same properties of N and L.

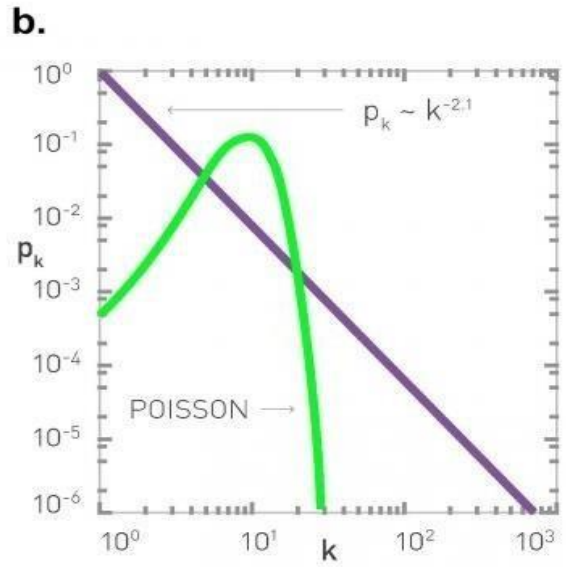

a.
b.
c.
REGULAR
SMALL-WORLD
RANDOM
p= 0
p= 1
Increasing randomness

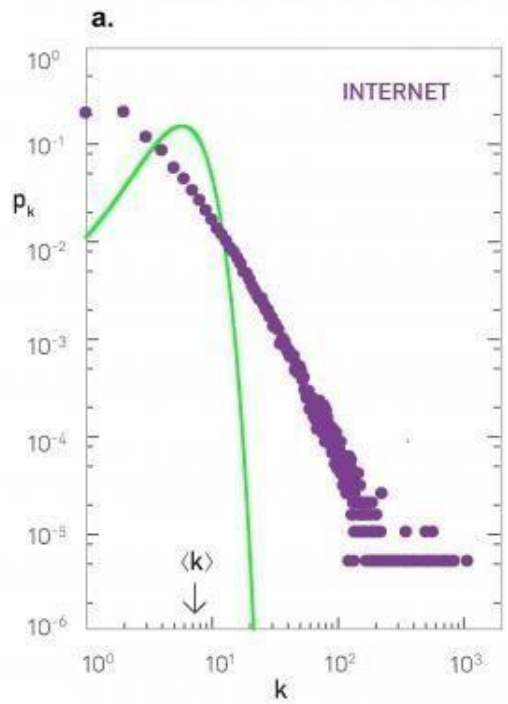
d.

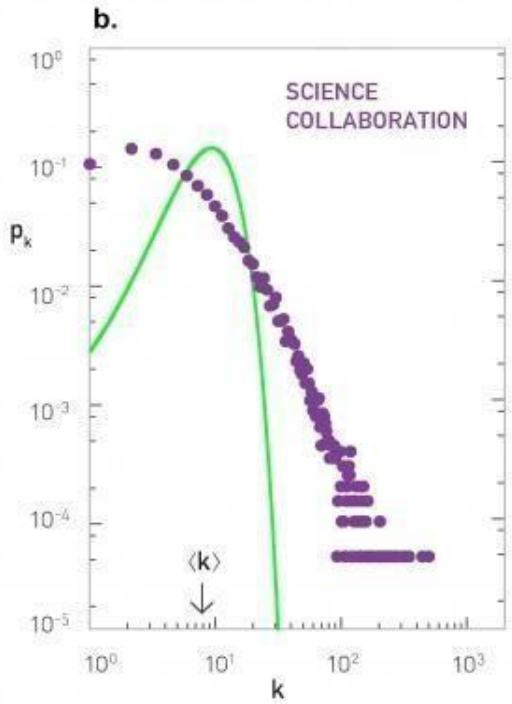


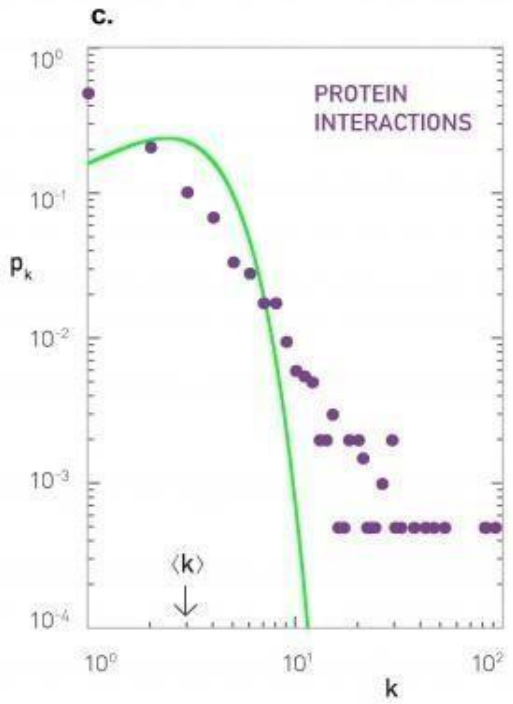

a.
POISSON
Number of nodes with k links
Most nodes have the same number of links
No highly connected nodes
Number of links (k)
b.
c.
POWER LAW
Number of nodes with k links
Many nodes with only a few links
A few hubs with large number of links
Number of links (k)
d.

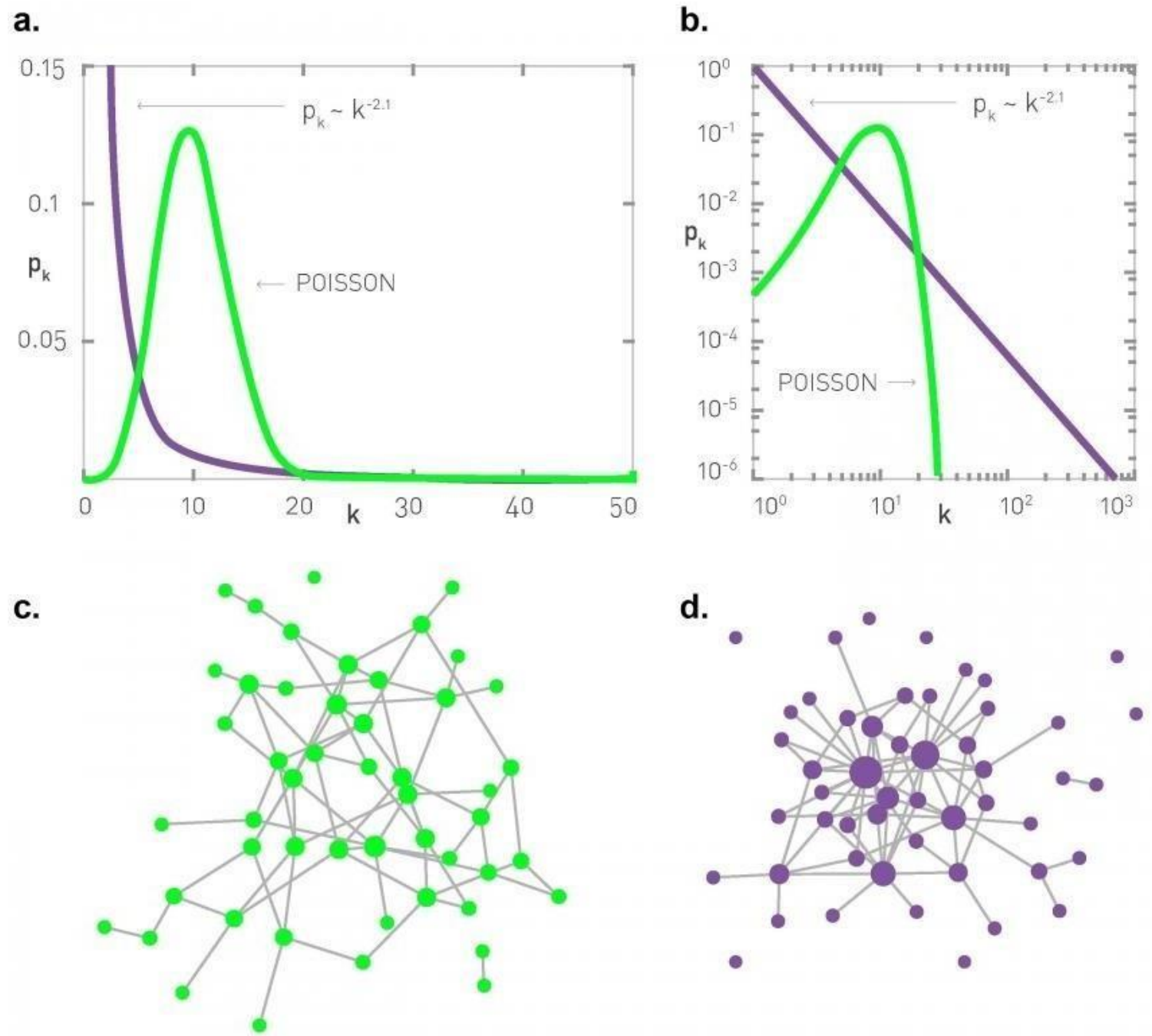

a.
0.15
0.1
0.05
p_k
$p_k \sim k^{-2.1}$
POISSON
0
10
20
k
30
40
50

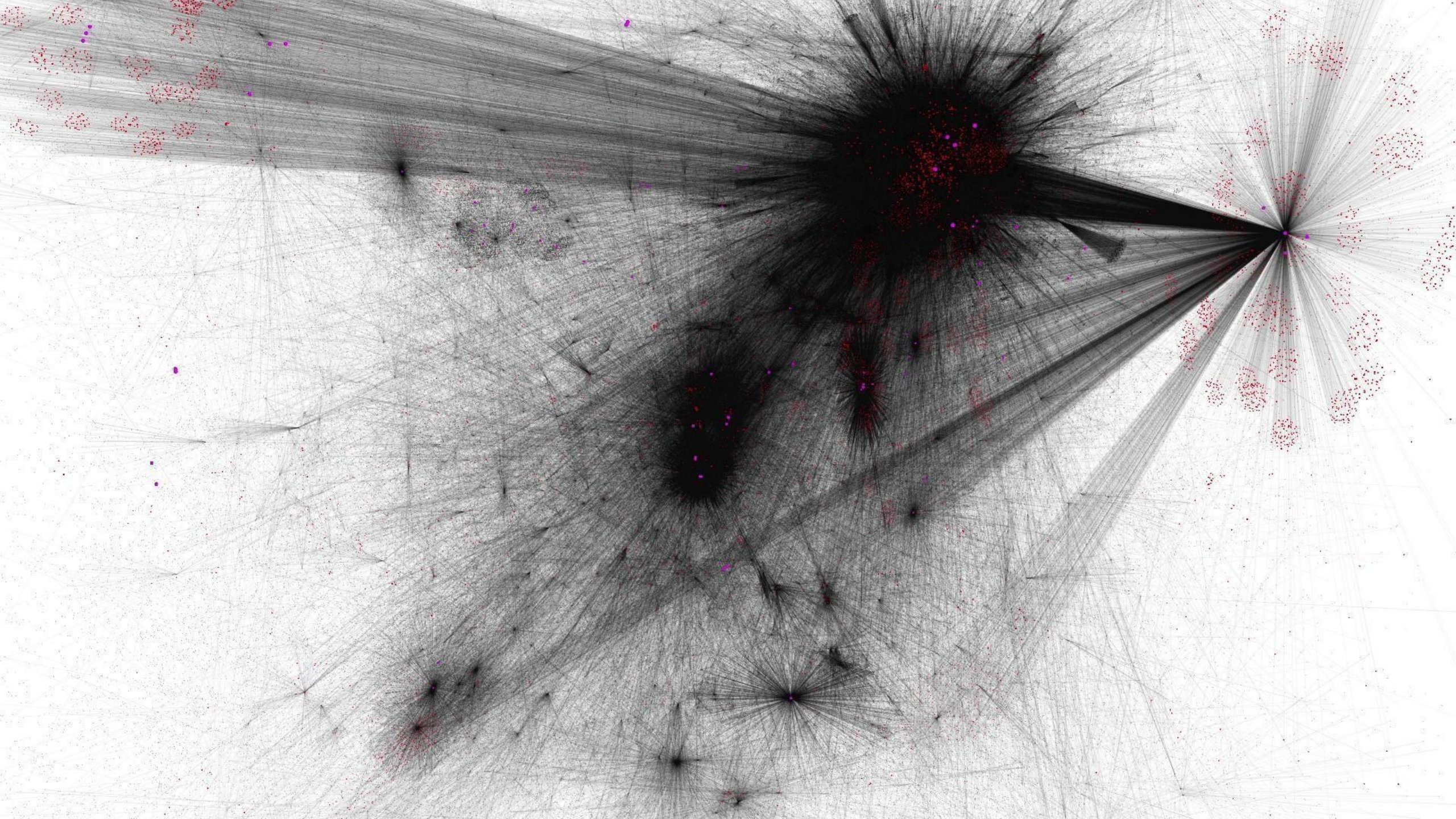

b.
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
p_k
$p_k \sim k^{-2.1}$
POISSON
10^0
10^1
k
10^2
10^3


a.
INTERNET
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
p_k
⟨k⟩
10^0
10^1
10^2
10^3
k


b.
SCIENCE
COLLABORATION
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
p_k
⟨k⟩
10^0
10^1
10^2
10^3
k


c.
PROTEIN
INTERACTIONS
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
p_k
⟨k⟩
10^0
10^1
10^2
k


a.
0.15
0.1
0.05
p_k
$p_k \sim k^{-2.1}$
POISSON
0
10
20
k
30
40
50
b.
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
p_k
$p_k \sim k^{-2.1}$
POISSON
10^0
10^1
k
10^2
10^3
c.
d.

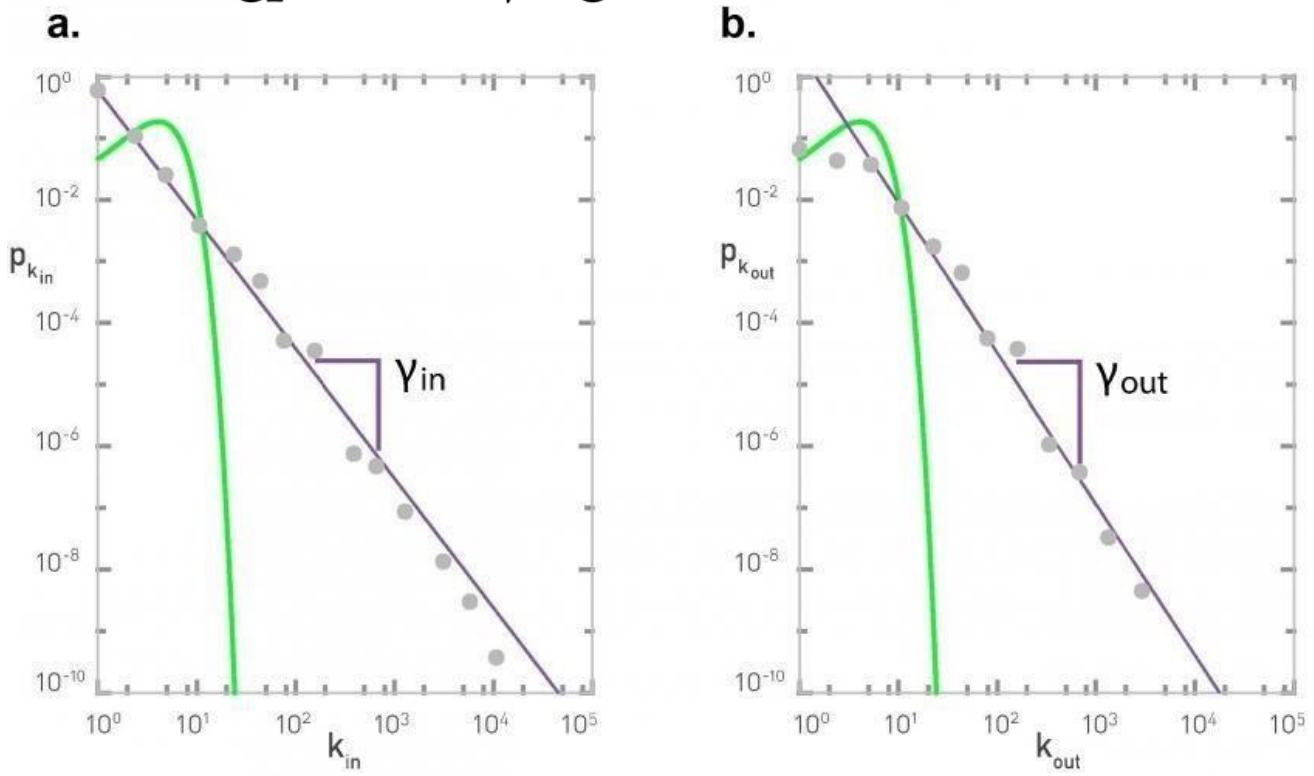


Scale-Free Networks

If the network of WWW was to be a random network than the degree distribution should follow Poisson distribution.

But it follows a power law distribution, with the exponent of gamma: $p_k = k^{-\gamma}$

Take the logarithm, we obtain: $\log p_k = -\gamma \log k$



Why do we have Hubs?

The Barabasi-Albert Model

For to every one who has will more be given, and
he will have abundance; but from him who has not,
even what he has will be taken away.
— Matthew 25:29, RSV.


https://www.maszol.ro/

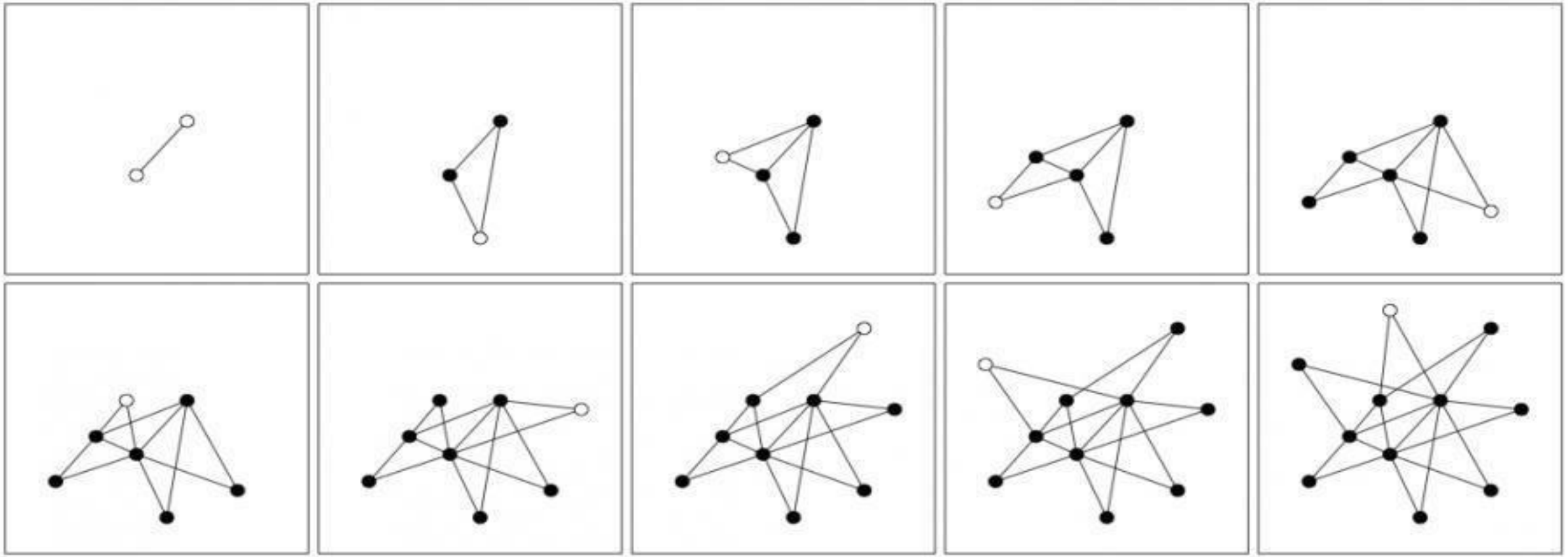
Growth
Real networks are getting larger and larger of a result of a growth process. While in random networks we usually have fix N.

Preferential Attachment
In real networks nodes tend to connect to the more connected nodes.

Why do we have Hubs?

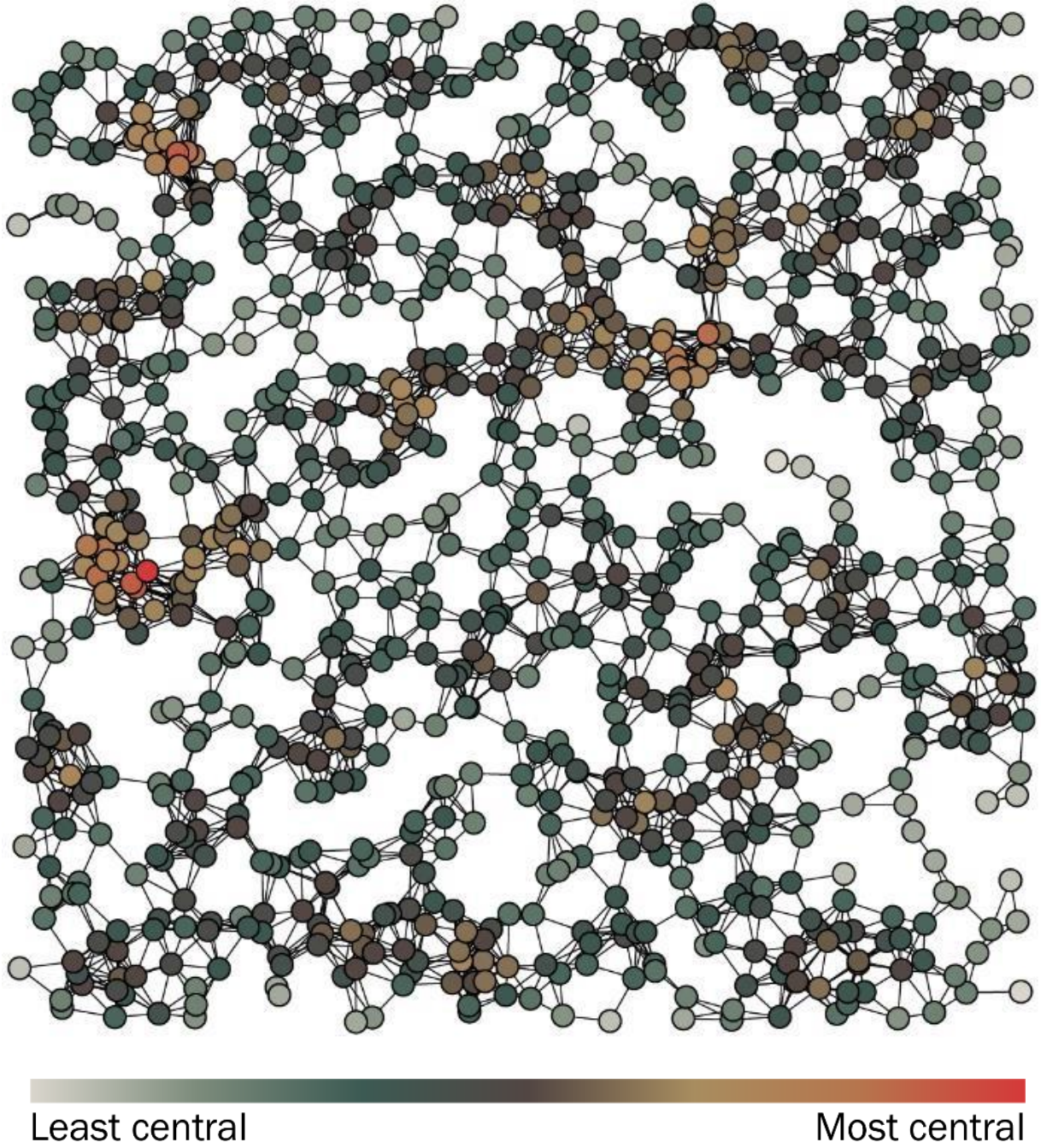
At each step we add new node to the network, with links connect to nodes already in the network.

The probability of that a link of a new node connected to a present node is depends on the degree of the node: $\prod(k_i) = \frac{k_i}{\sum_j k_j}$



Which node is a hub?

Node centrality
- Degree centrality



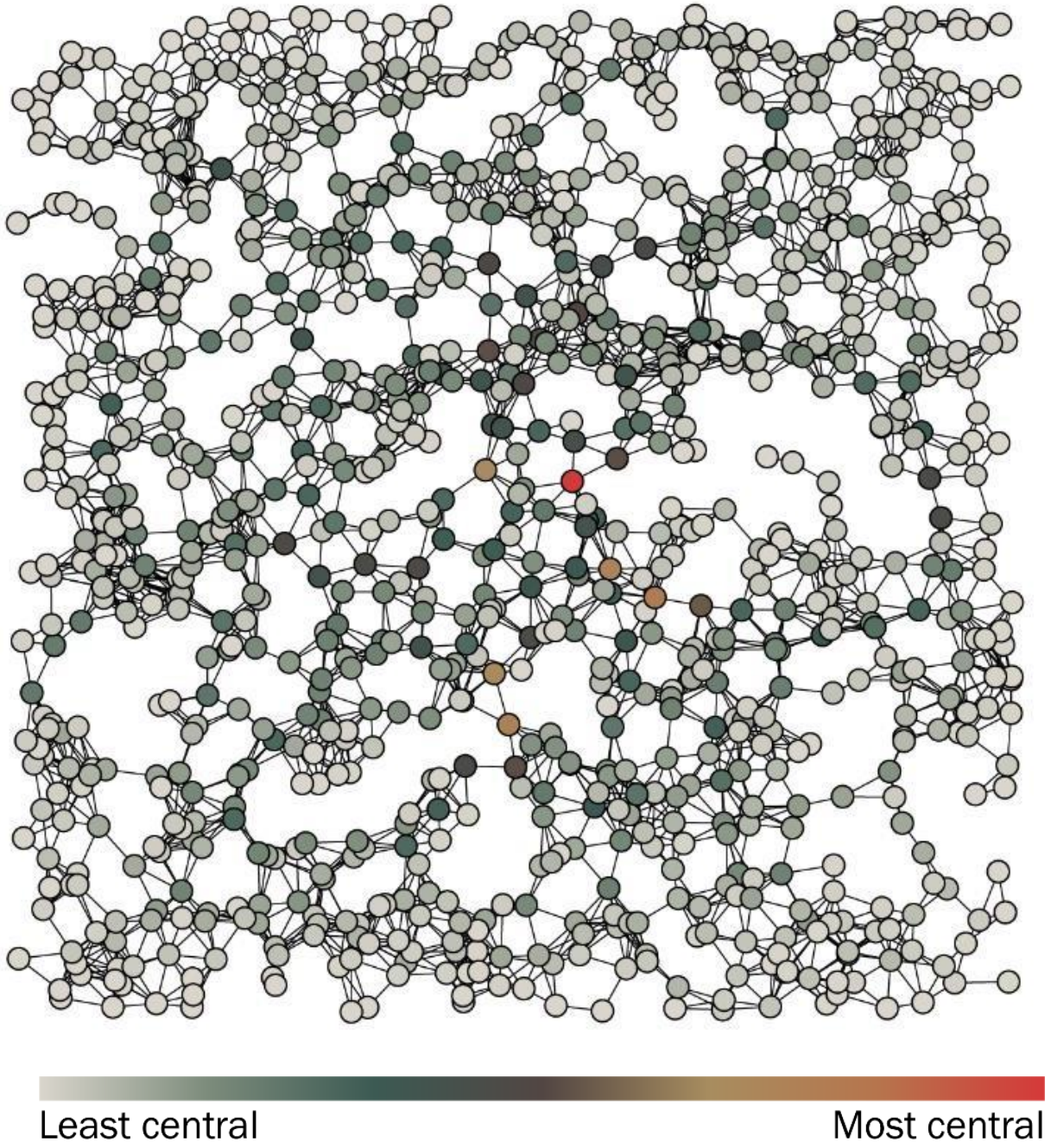
Which node is a hub?

Node centrality

- Betweenness centrality

Captures the node's role as a bridge between groups of nodes. Betweenness in about how critical a bode is to the networks functioning as a bridge point between other parts of the network.

$$C_B(i) = \sum_{i \neq j \neq k} \sigma_{jk}(i)/\sigma_{jk}$$



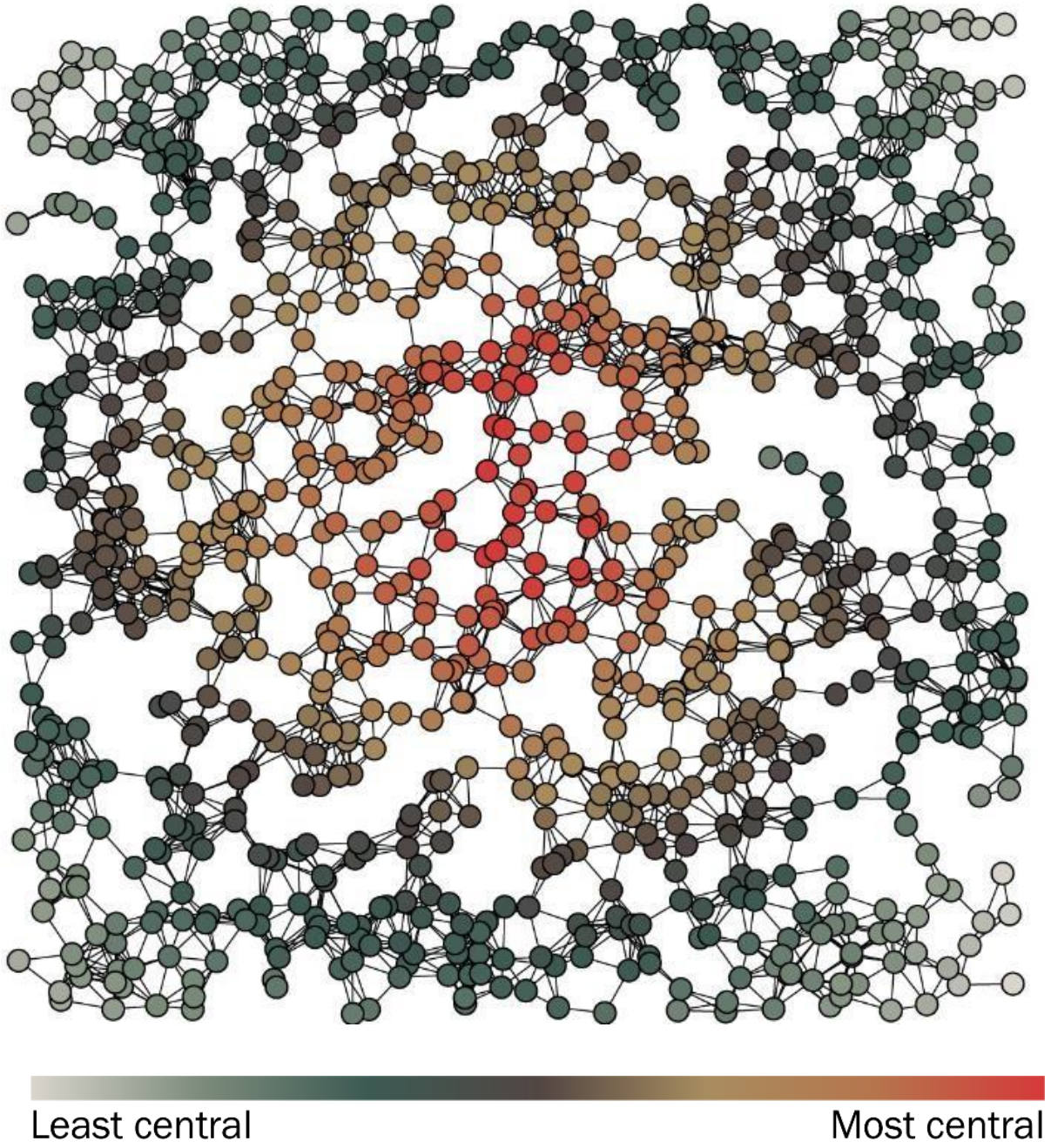
Which node is a hub?

Node centrality

- Closeness centrality

Closeness captures how close a node is to any other randomly selected node in the network. That is how quickly can the mode reach other nodes in the network.

$$C_C(i) = 1/\sum_{i \neq j} d_{ij}$$



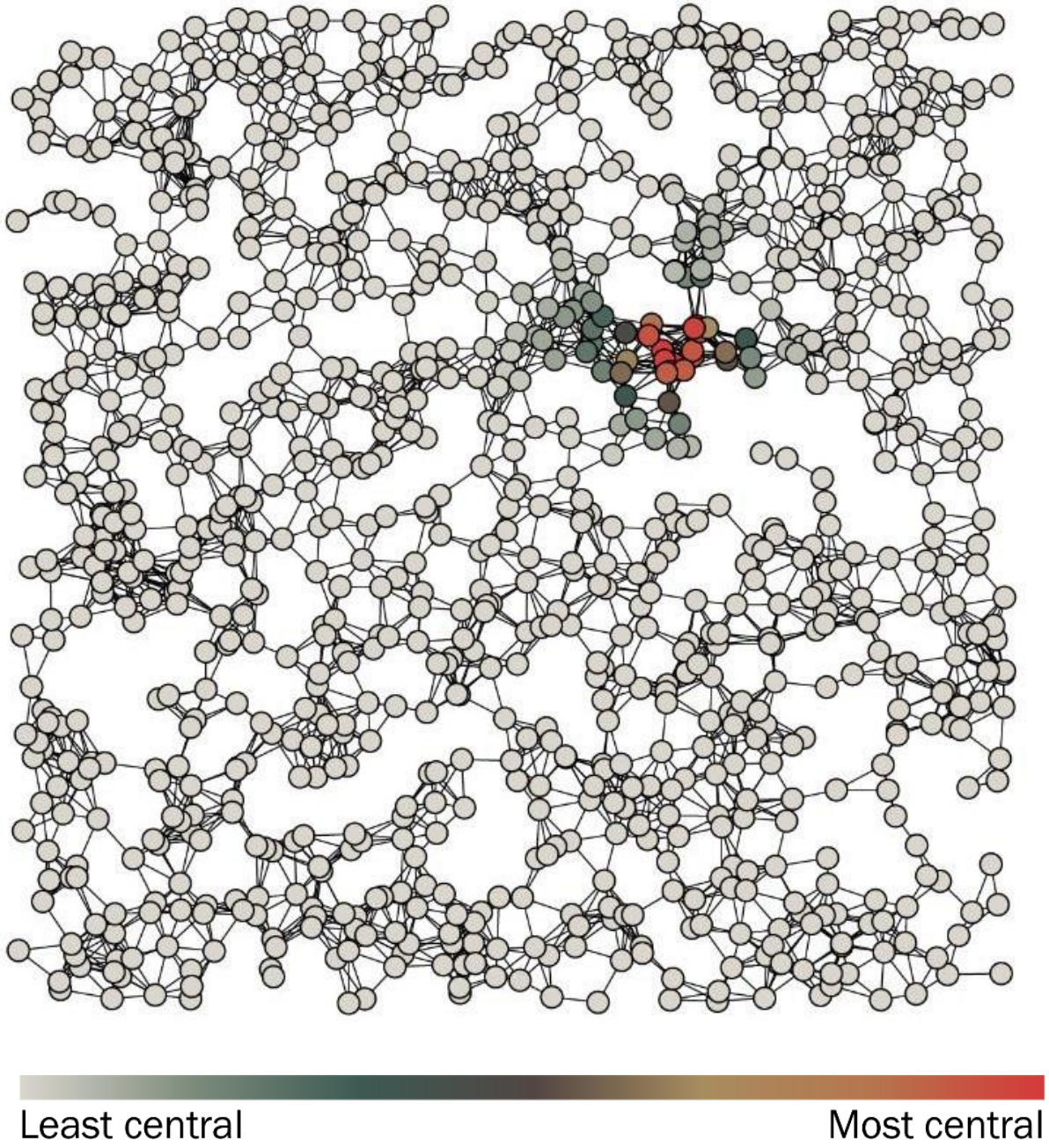
Which node is a hub?

Node centrality

- Eigenvector centrality

How important a node in a network based on the significance of the nodes that given node is connected to. So instead of looking on the number of connections it represents the value of the connections.

$$C_E(i) = x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j \in G} A_{ij} x_j$$



Which node is a hub?

Betweenness

Captures the node's role as a bridge between groups of nodes. Betweenness in about how critical a bode is to the networks functioning asa bridge point between other parts of the network.

Closeness

Closeness captures how close a node is to any other randomly selected node in the network. That is how quickly can the mode reach other nodes in the network.

Eigenvector centrality

How important a node in a network based on the significance of the nodes that given node is connected to. So instead of looking on the number of connections it represents the value of the connections.

Degree centrality