Core concepts

COMPLEX NETWORKS: A PRIMER

Shakti N. Menon

The Institute of Mathematical Sciences, Chennai

Dec 11th-13th 2024

Workshop on "Spins, Games & Networks"

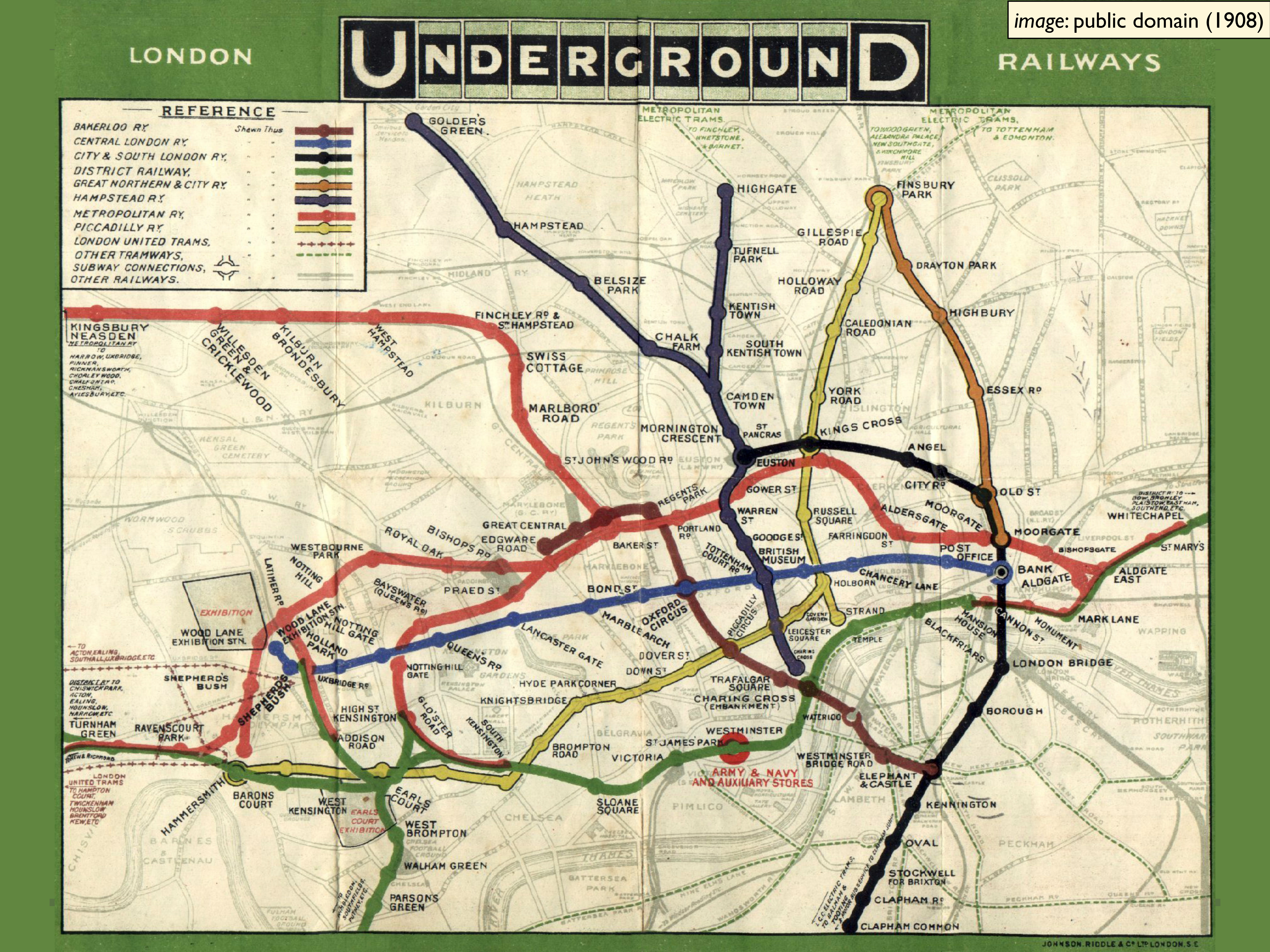


image: public domain (1908)

Harry Beck's Tube map (1933)

image: Rex Features

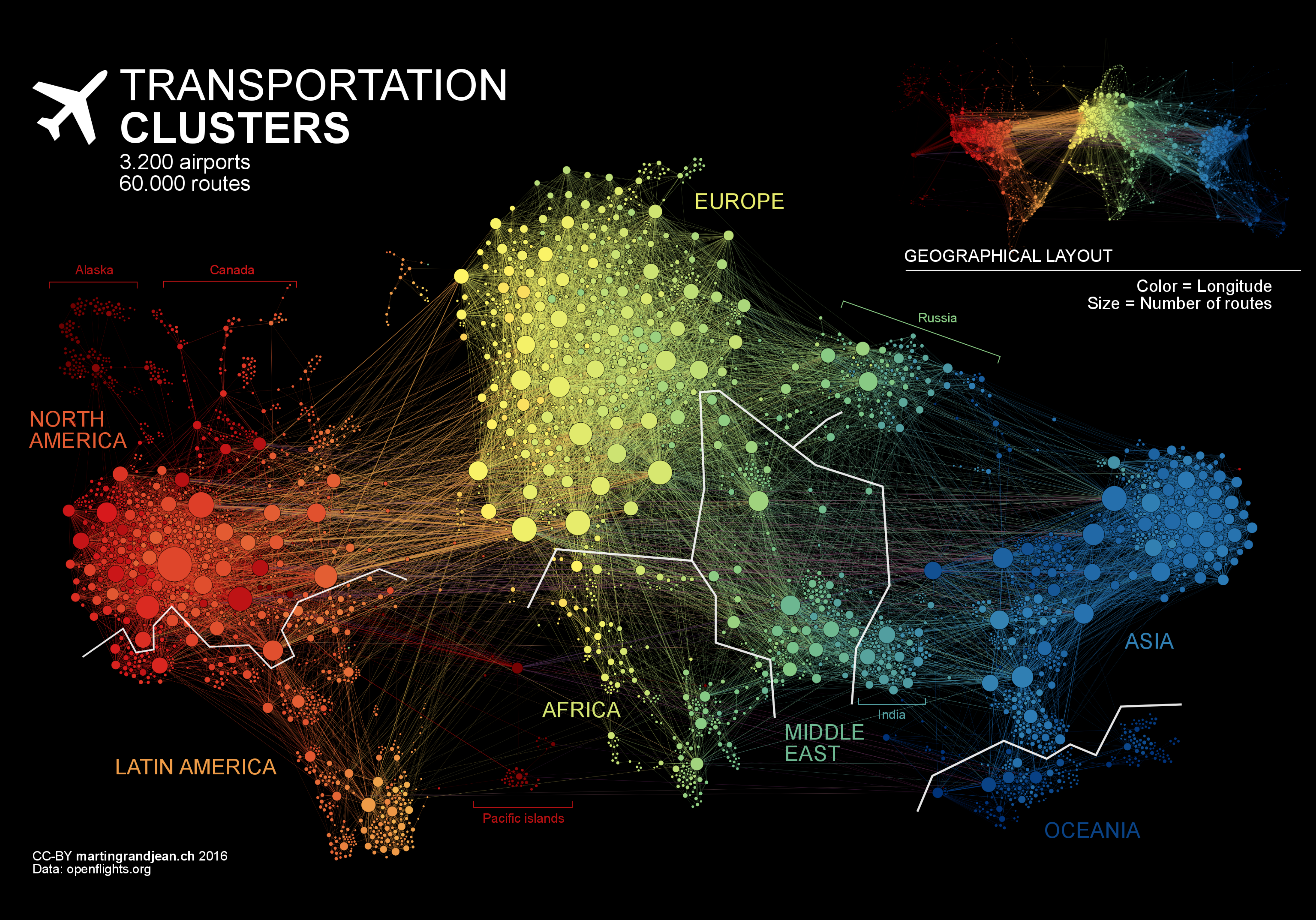


image: https://www.martingrandjean.ch/connected-world-air-traffic-network/

The Political Blogosphere and the 2004 U.S. Election

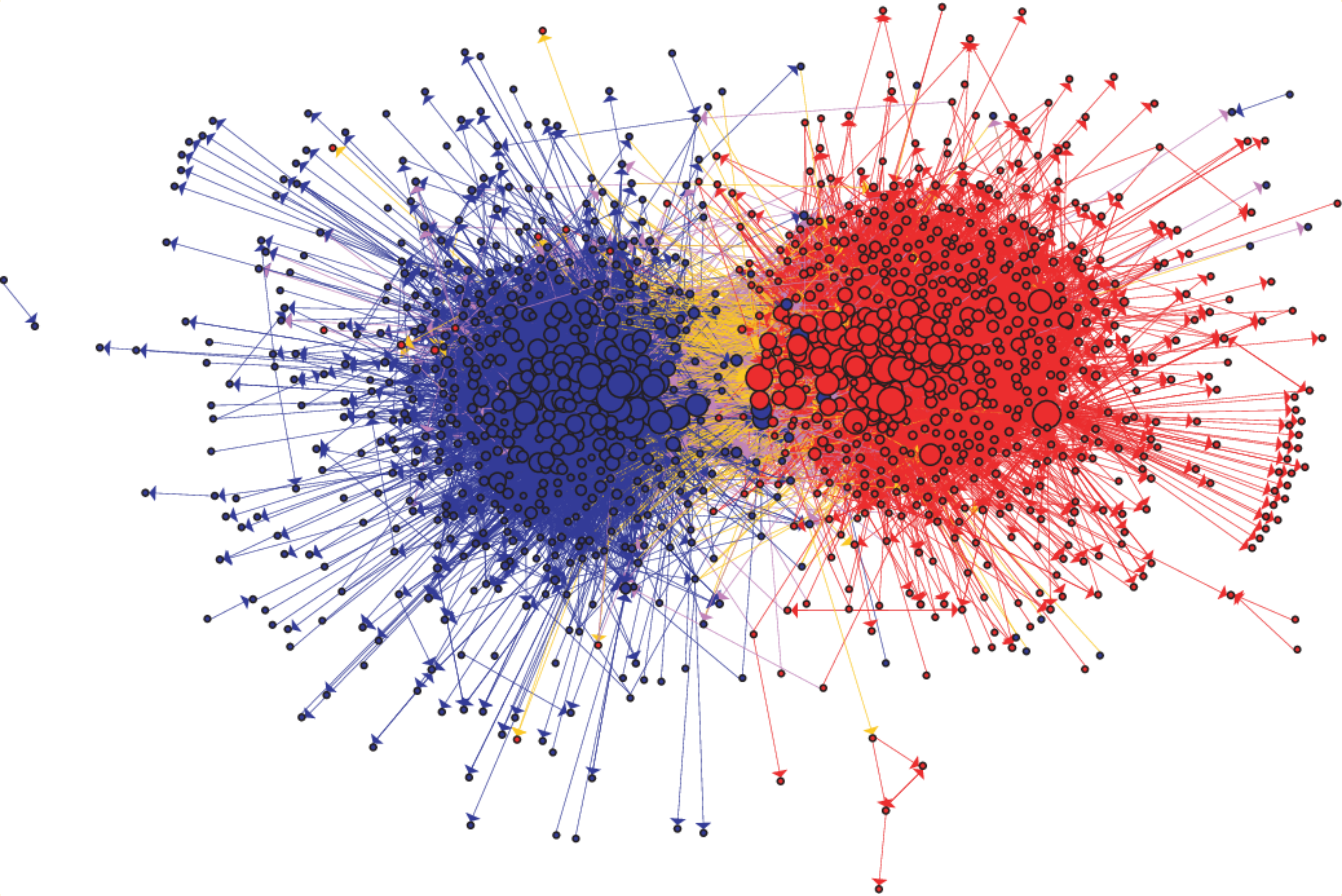
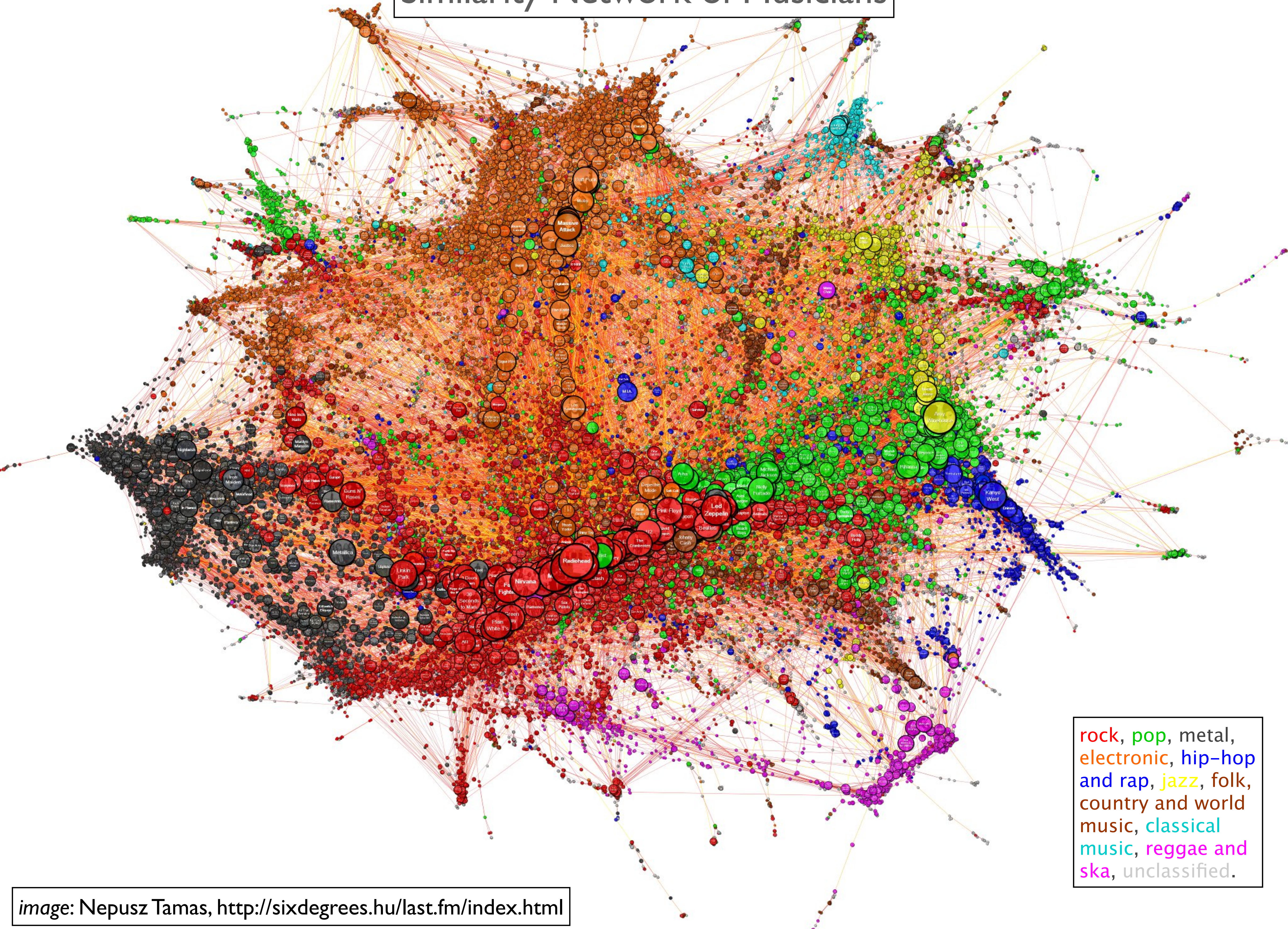


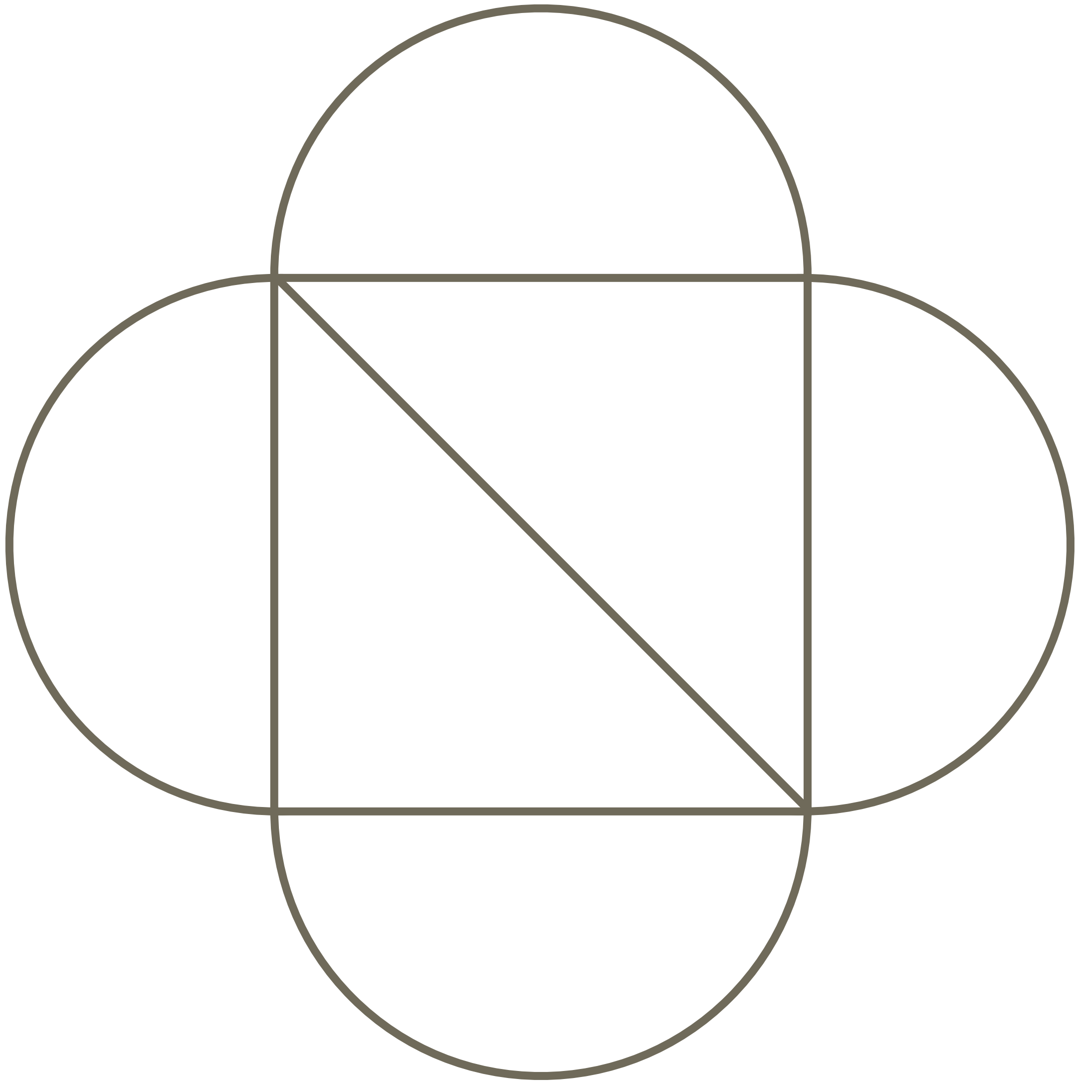
image: Lada Adamic & Natalie Glance, Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem (2005).

Similarity Network of Musicians



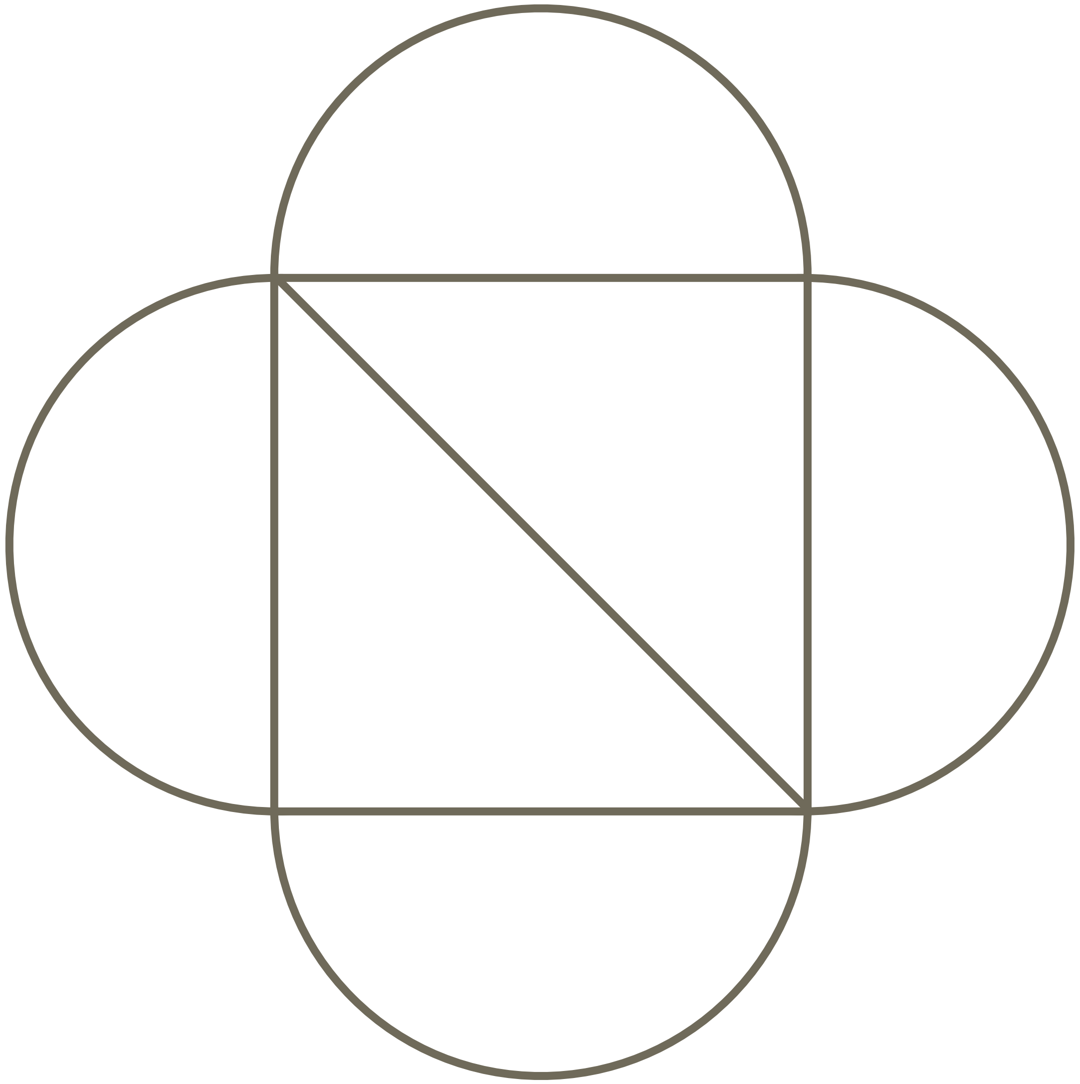
rock, pop, metal, electronic, hip-hop and rap, jazz, folk, country and world music, classical music, reggae and ska, unclassified.

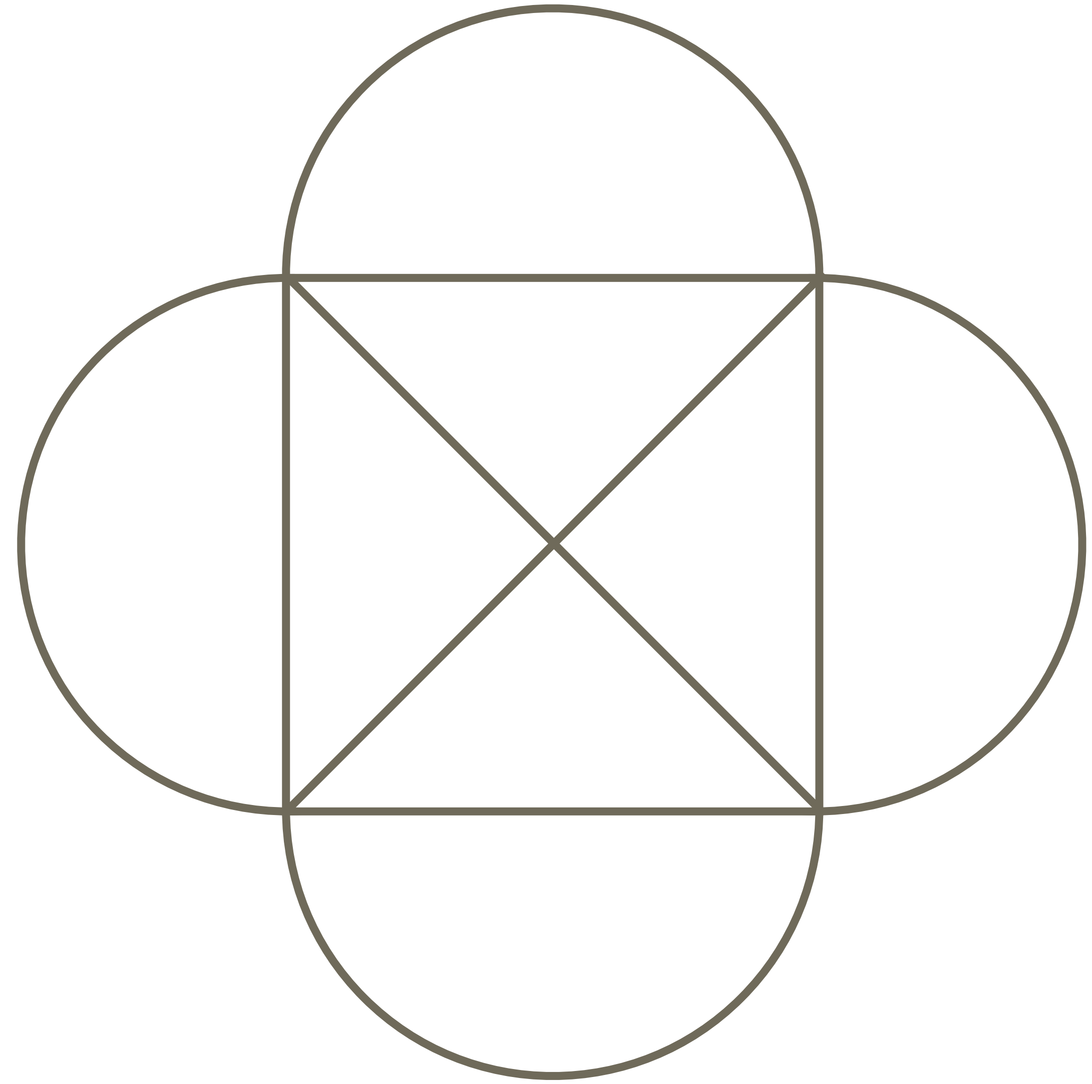
image: Nepusz Tamas, http://sixdegrees.hu/last.fm/index.html



Can you draw this

- without taking your pen off the paper, and
- without crossing any path twice?



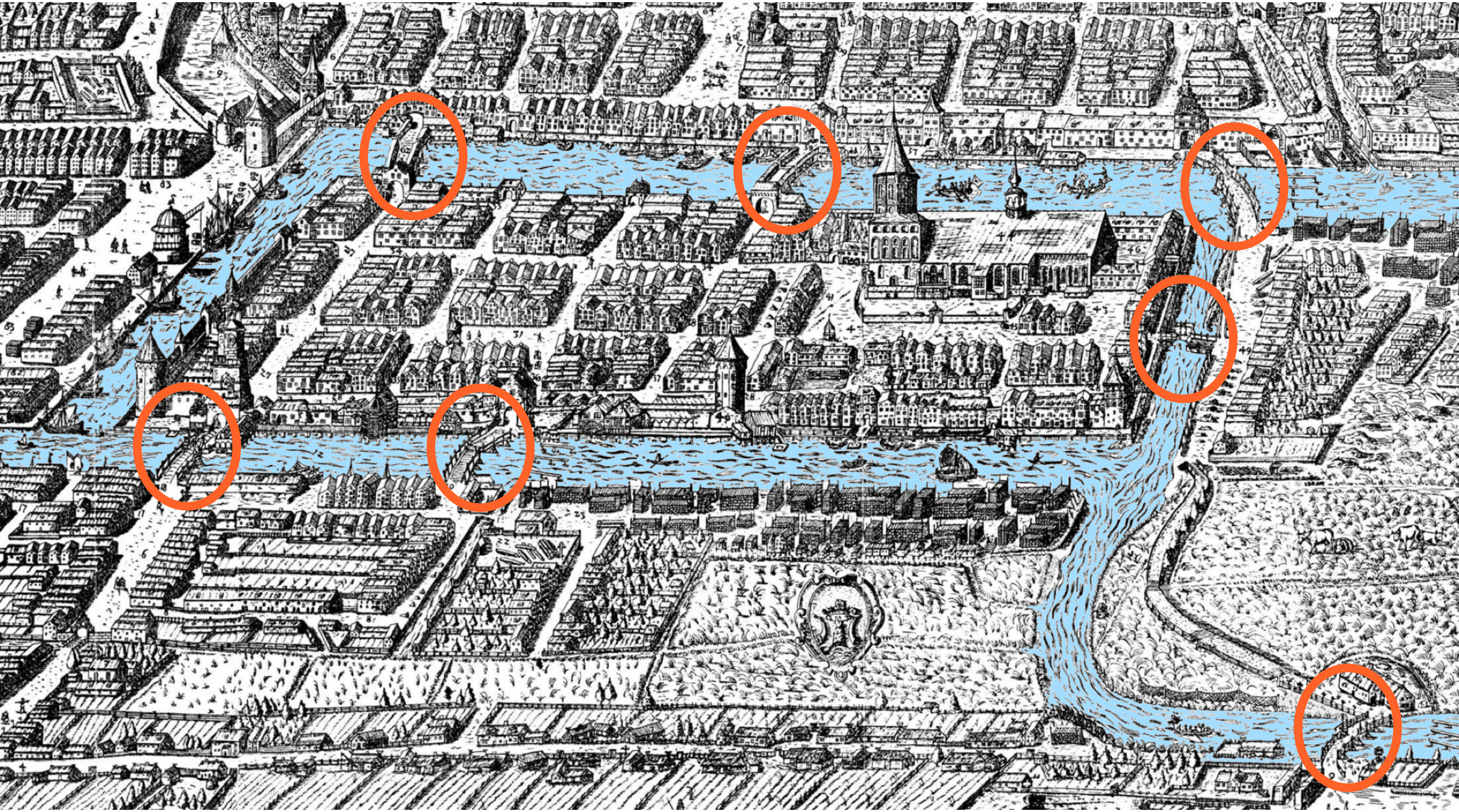


Can you draw this

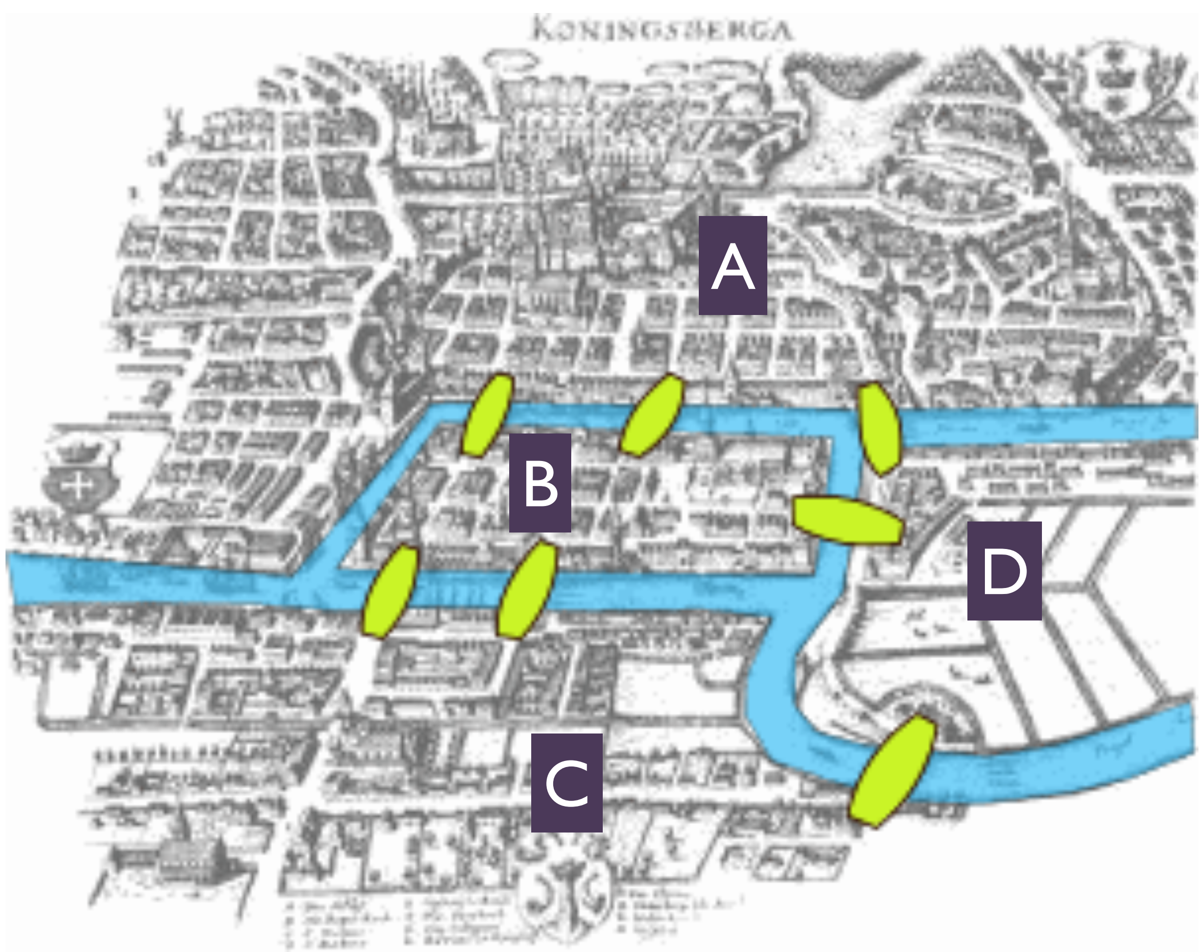
- without taking your pen off the paper, and
- without crossing any path twice?

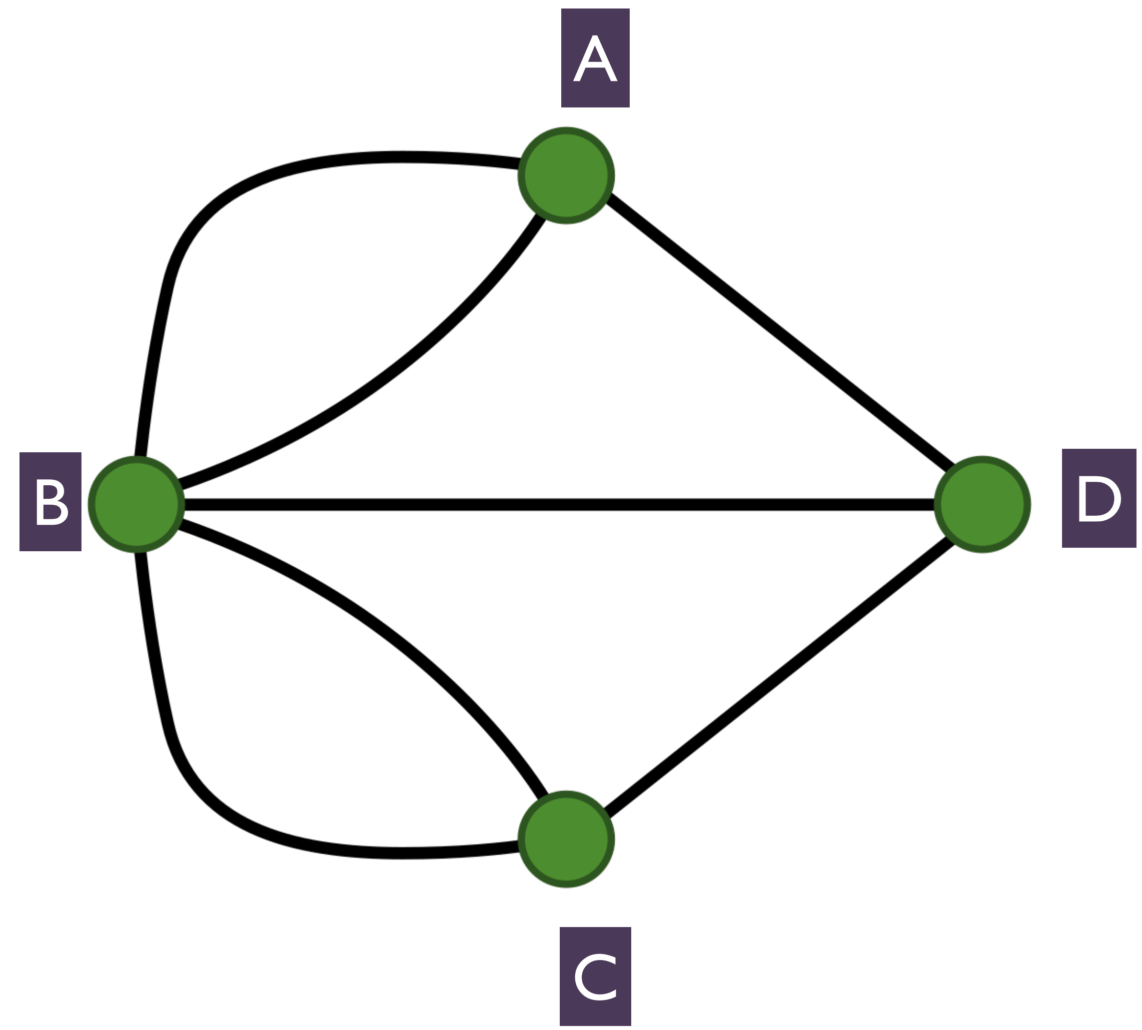
What about this?

THE SEVEN BRIDGES OF KÖNIGSBERG



EULER'S SOLUTION IN 1736





Each land mass can be viewed as a "vertex" and each bridge as a "link".

Only <u>terminal</u> vertices can have an odd number of links.

GRAPHS AND NETWORKS

Euler's work laid the foundation for the field of **graph theory**.

Any network of connections between entities can be analysed by viewing it as a graph that describes the manner in which a set of objects are connected.

A network can simply be thought of as a graph where the objects and relations have certain attributes.

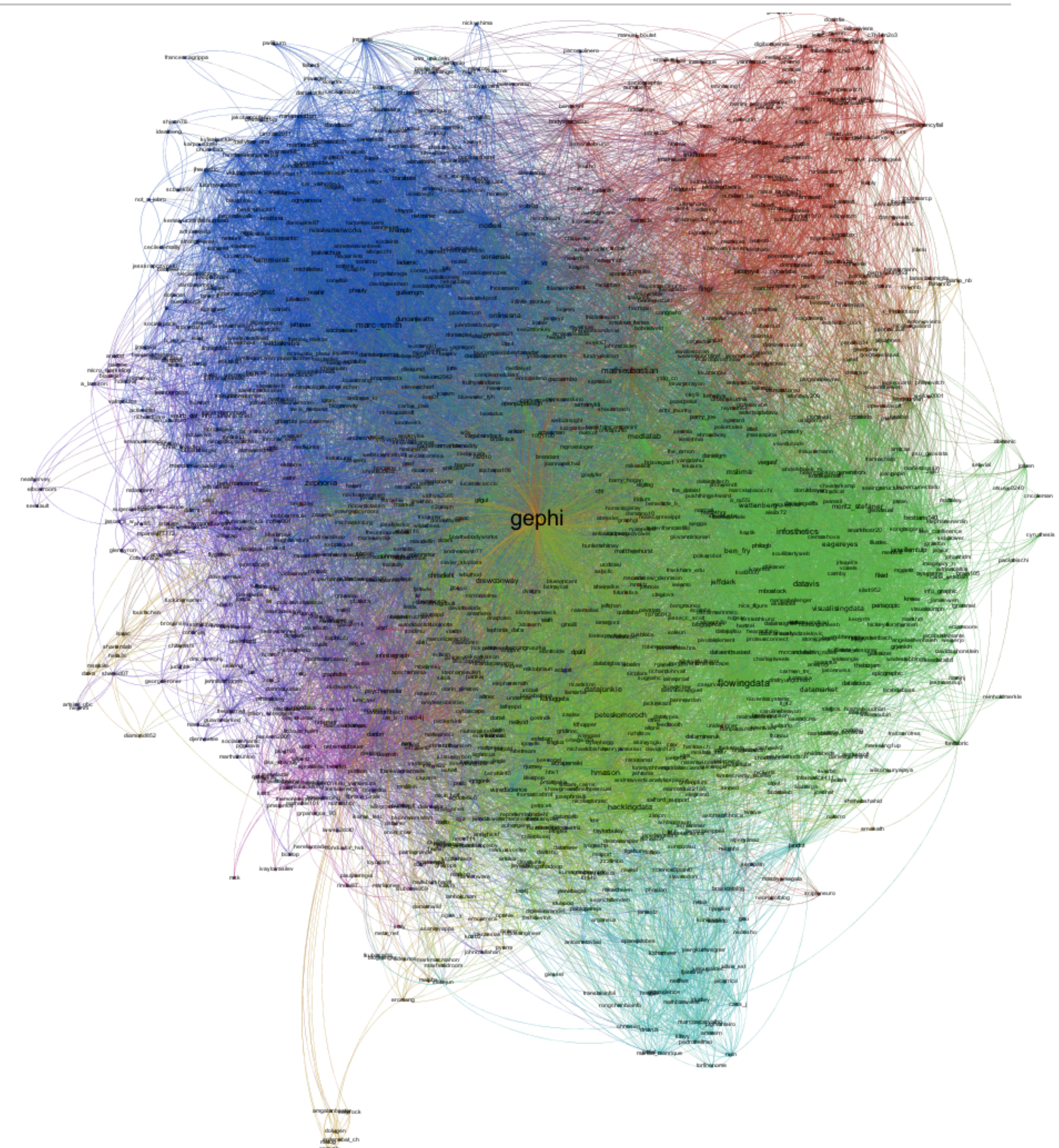
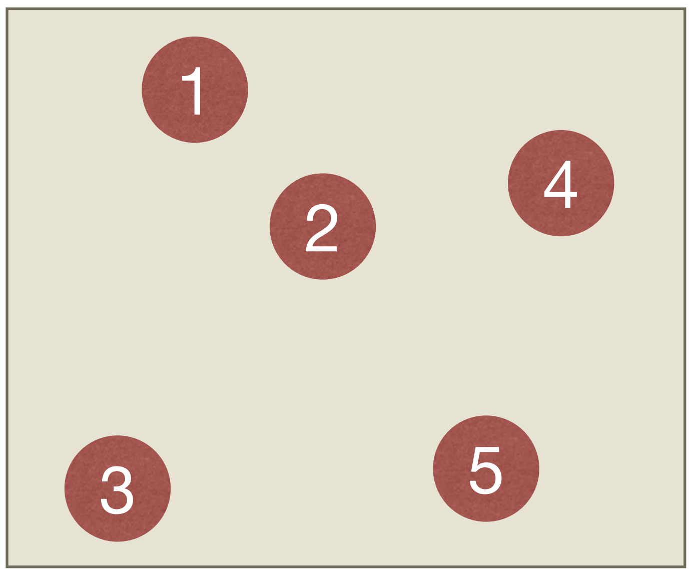
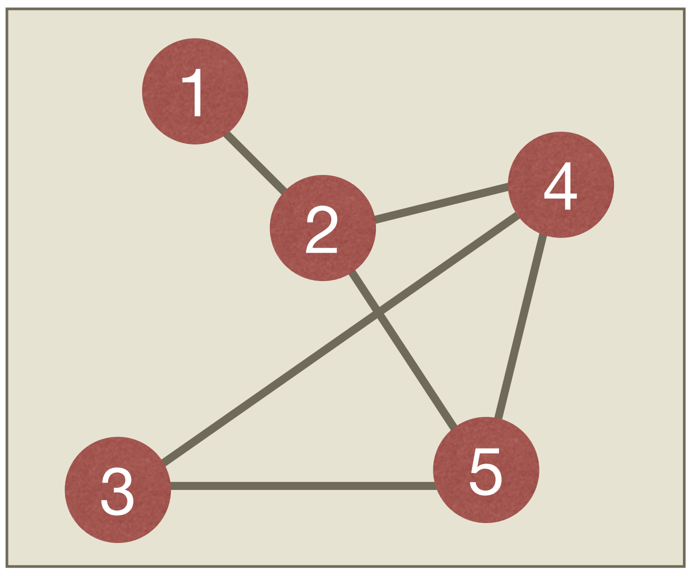


image: http://social-dynamics.org/a-gephi-visualization-of-gephi-on-twitter/

FUNDAMENTAL CONCEPTS: NODES AND LINKS



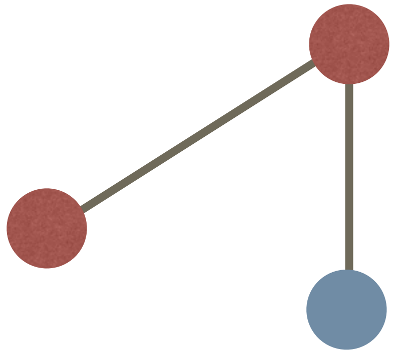
Nodes (or "Vertices")



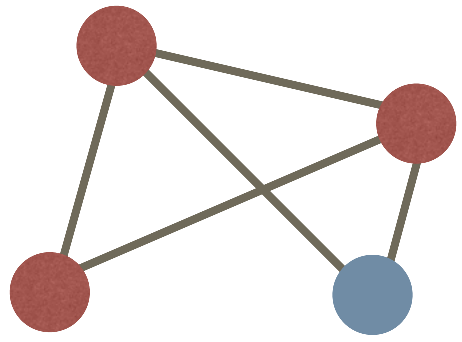
Links (or "Edges")

NODE	DEGREE
1	1
2	3
3	2
4	3
5	3

The total number of links associated with a node it its **degree** (k).

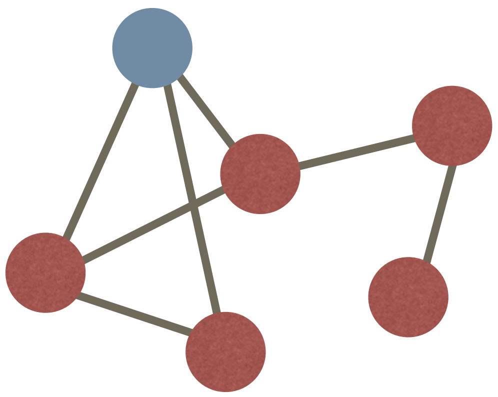


$k = 1$



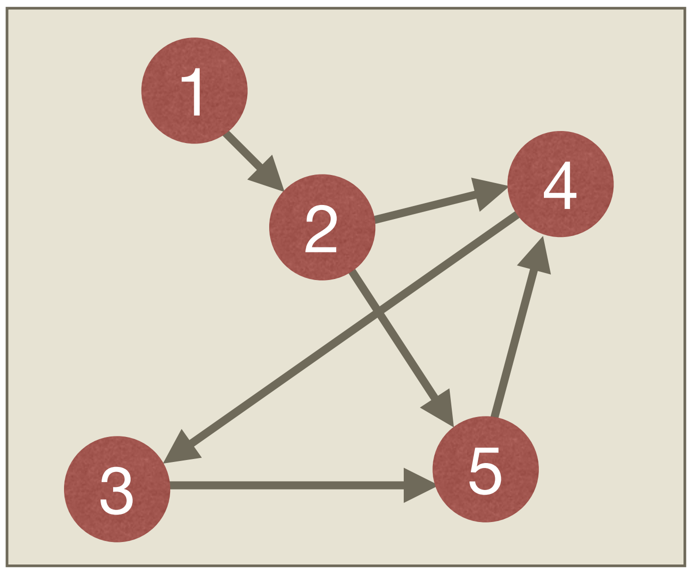
$k = 2$

$k = 3$

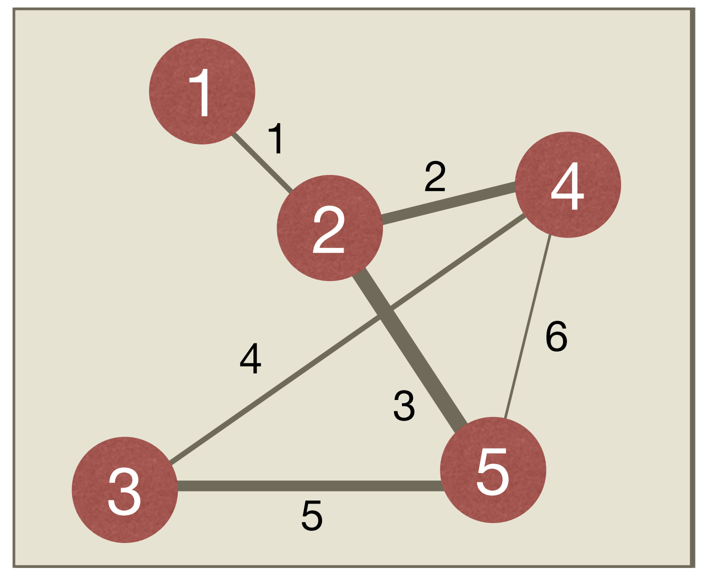


FUNDAMENTAL CONCEPTS: DIRECTED AND WEIGHTED NETWORKS

Directed network



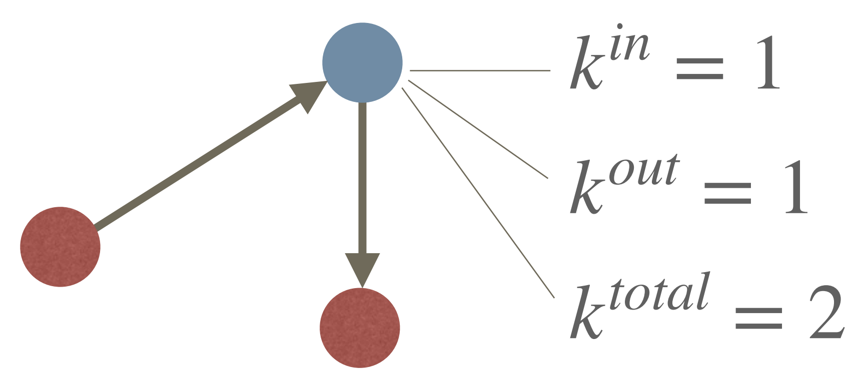
Weighted network



EDGE	WEIGHT
1	1
2	2
3	3
4	1
5	2
6	0.5

NODE	IN-DEGREE	OUT-DEGREE	TOTAL DEGREE
1	0	1	1
2	1	2	3
3	1	1	2
4	2	1	3
5	2	1	3

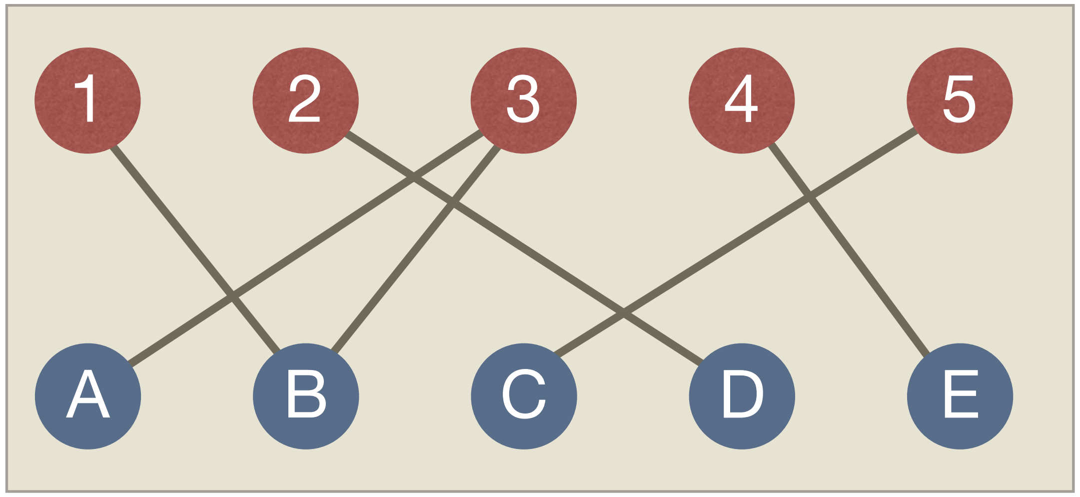
In a directed network a node can have an in-degree different to its out-degree



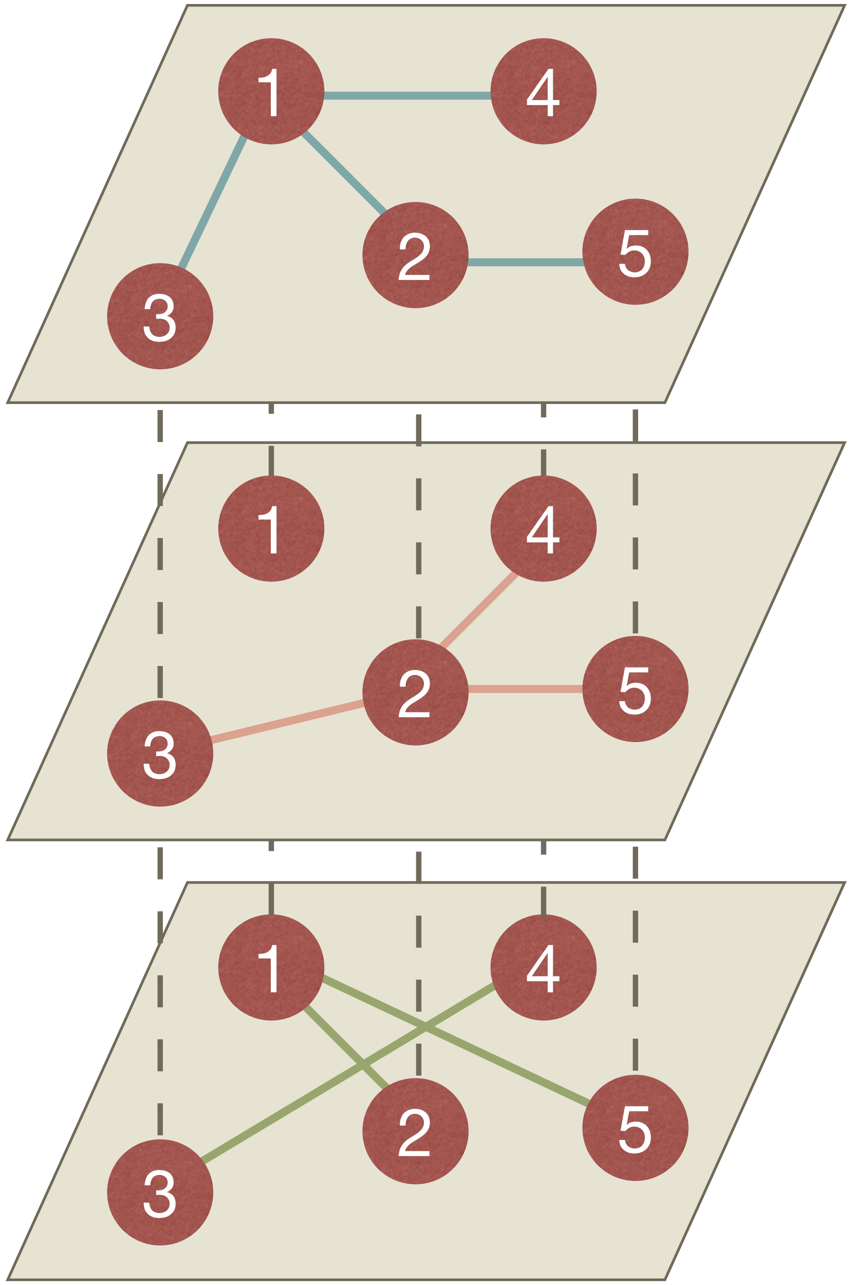
SOME OTHER TYPES OF NETWORKS

Networks that describe relations between two different classes of objects are known as Bipartite networks.

Networks in which there may be different types of links between nodes are known as Multiplex networks.

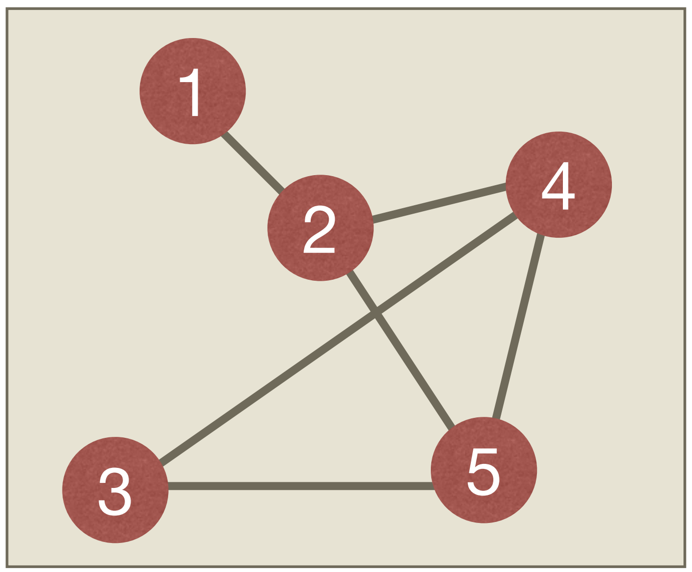


Bipartite network



Multiplex network

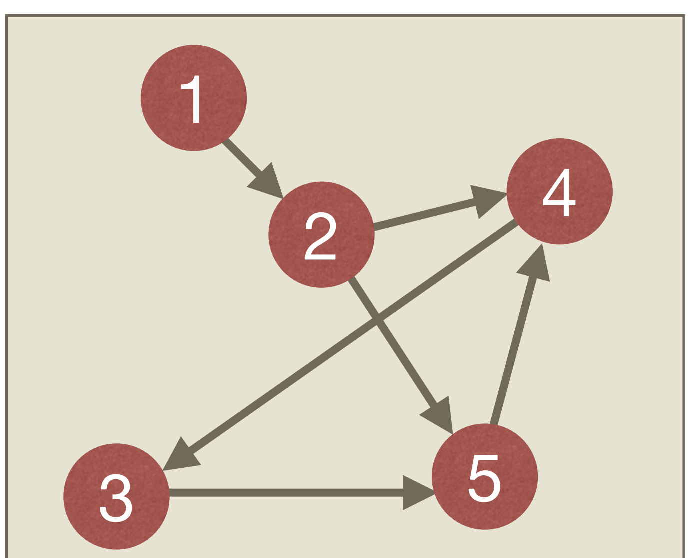
FUNDAMENTAL CONCEPTS:
ADJACENCY MATRIX



Adjacency matrix

	1	2	3	4	5
1	0	1	0	0	0
2	1	0	0	1	1
3	0	0	0	1	1
4	0	1	1	0	1
5	0	1	1	1	0

The Adjacency matrix $\mathbf{A}$ specifies all connections in the network. If nodes i and j are connected then $A_{ij} = 1$ else $A_{ij} = 0$.



target →

source ↓

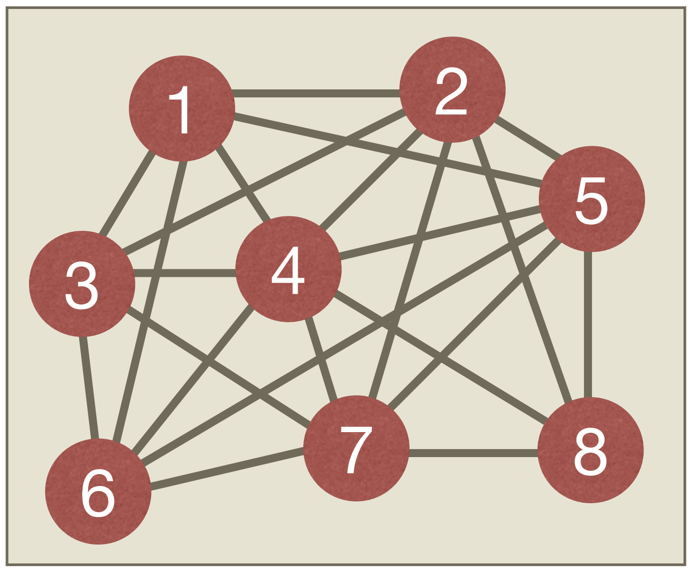
	1	2	3	4	5
1	0	1	0	0	0
2	0	0	0	1	1
3	0	0	0	0	1
4	0	0	1	0	0
5	0	0	0	1	0

In an <u>undirected</u> network, the degree k_i of a node i can be obtained via:

$$k_i = \sum_i A_{ij} = \sum_j A_{ij}$$

FUNDAMENTAL CONCEPTS: DENSITY AND SPARSITY

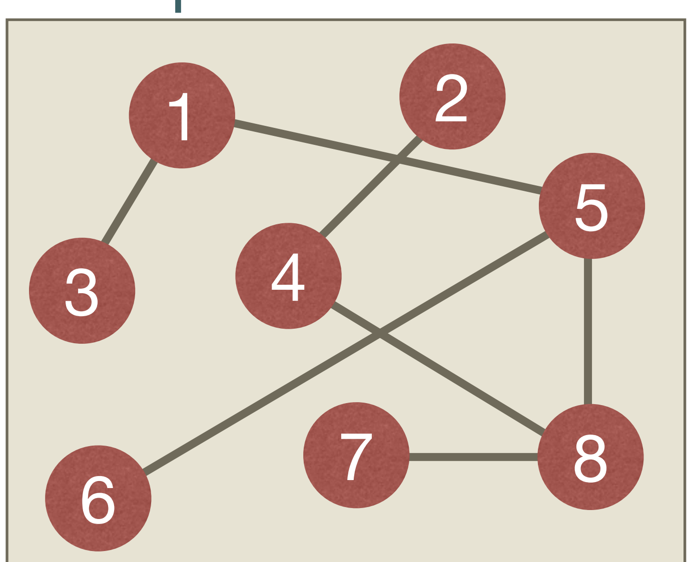
Dense network



	1	2	3	4	5	6	7	8
1	0	1	1	1	1	1	0	0
2	1	0	1	1	1	0	1	1
3	1	1	0	1	0	1	1	0
4	1	1	1	0	1	1	1	1
5	1	1	0	1	0	1	1	1
6	1	0	1	1	1	0	1	0
7	0	1	1	1	1	1	0	1
8	0	1	0	1	1	0	1	0

The density ρ can be understood as the probability that there exists a link between a pair of nodes (i, j).

Sparse network

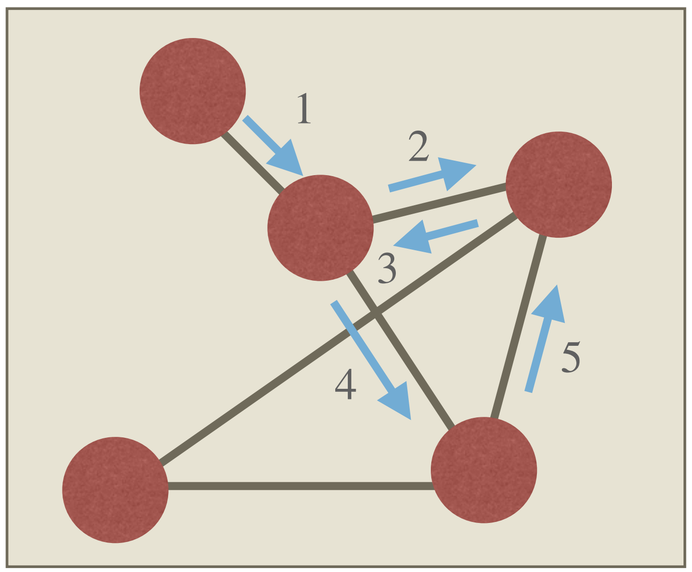


	1	2	3	4	5	6	7	8
1	0	0	1	0	1	0	0	0
2	0	0	0	1	0	0	0	0
3	1	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	1
5	1	0	0	0	0	1	0	1
6	0	0	0	0	1	0	0	0
7	0	0	0	0	0	0	0	1
8	0	0	0	1	1	0	1	0

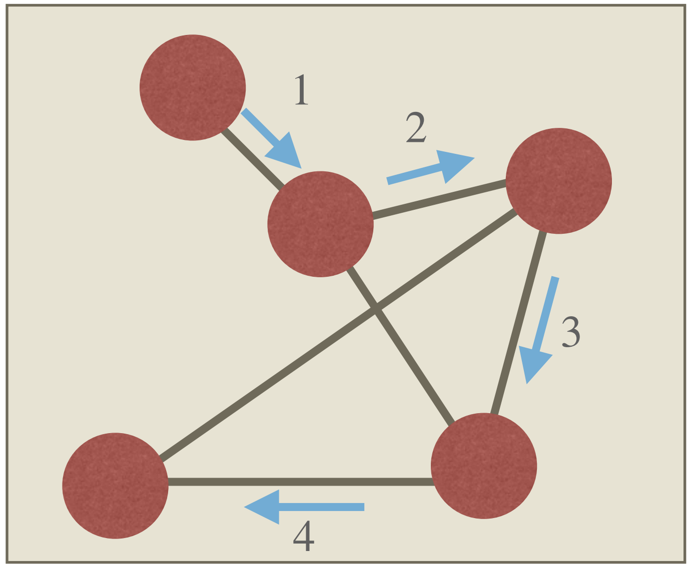
A network is said to be dense if "most" of the possible links are present, and sparse if "most" are absent.

FUNDAMENTAL CONCEPTS: WALKS AND PATHS

Walk



Path

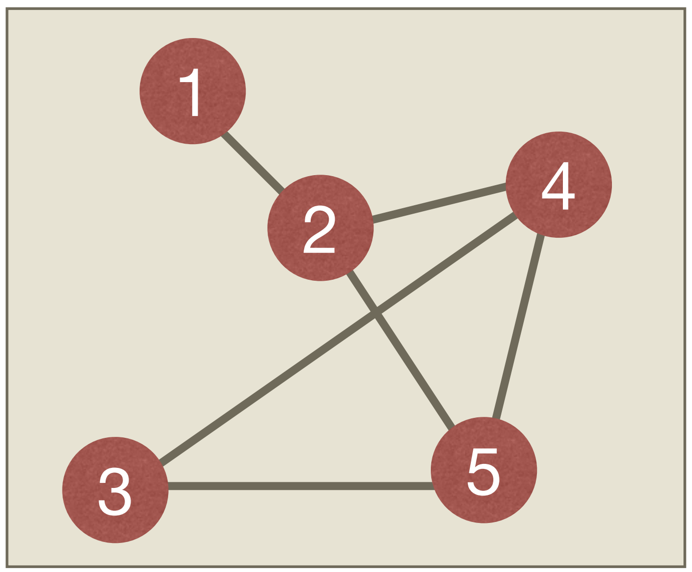


A walk is a route along the edges of a network. In an undirected network, an edge can be crossed in either direction.

The length of a walk is the number of hops taken along the route.

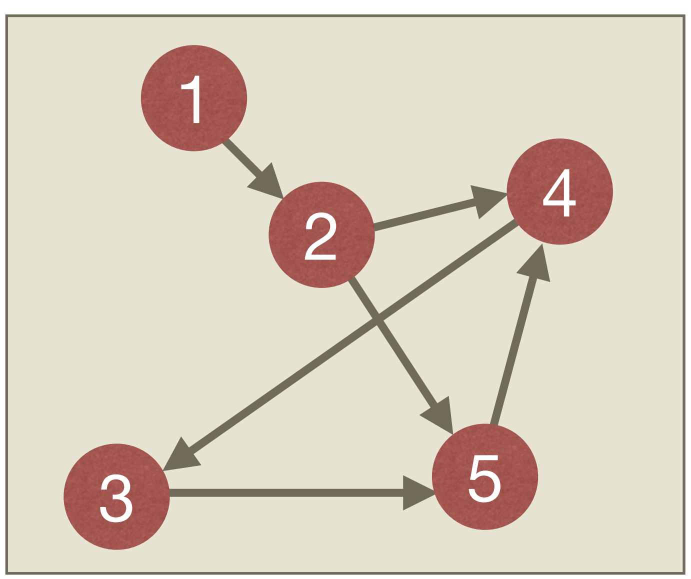
A path is a self-avoiding walk, i.e. one in which no edge is traversed twice.

FUNDAMENTAL CONCEPTS: SHORTEST PATH LENGTH



i-j	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
d_{ij}	1	3	2	2	2	1	1	1	1	1

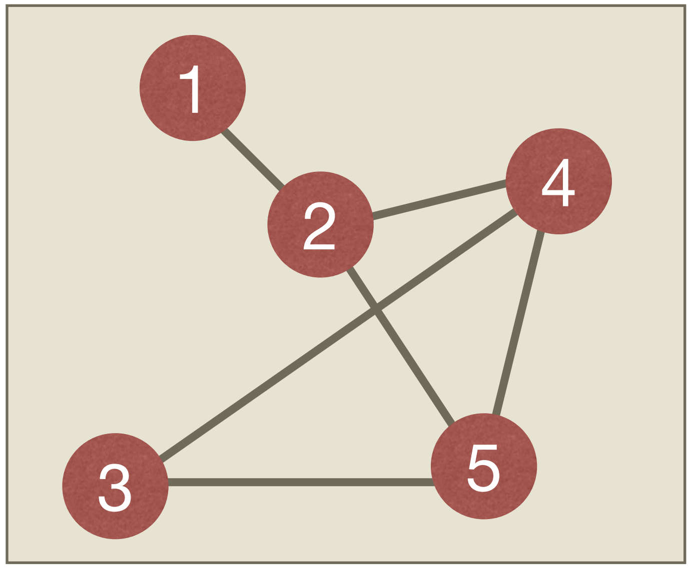
The shortest path length d_{ij} between two nodes i and j is the minimum number of links one has to cross to travel between them.



Can you work out what the shortest path length is between every pair of nodes of this directed network?

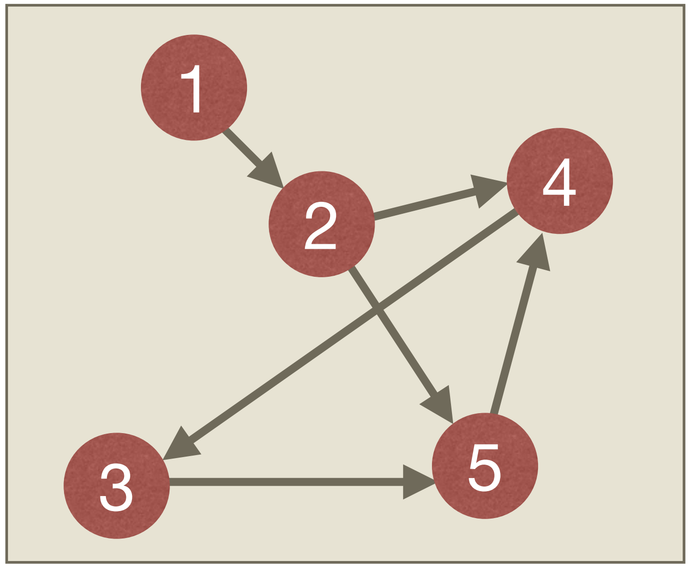
FUNDAMENTAL CONCEPTS: DIAMETER

$d_{\max} = 3$



The diameter $d_{\max}$ of a network is the "longest shortest path" between all pairs of nodes i and j in the network : $\max_{(i,j)}(d_{ij})$.

$d_{\max} = ?$



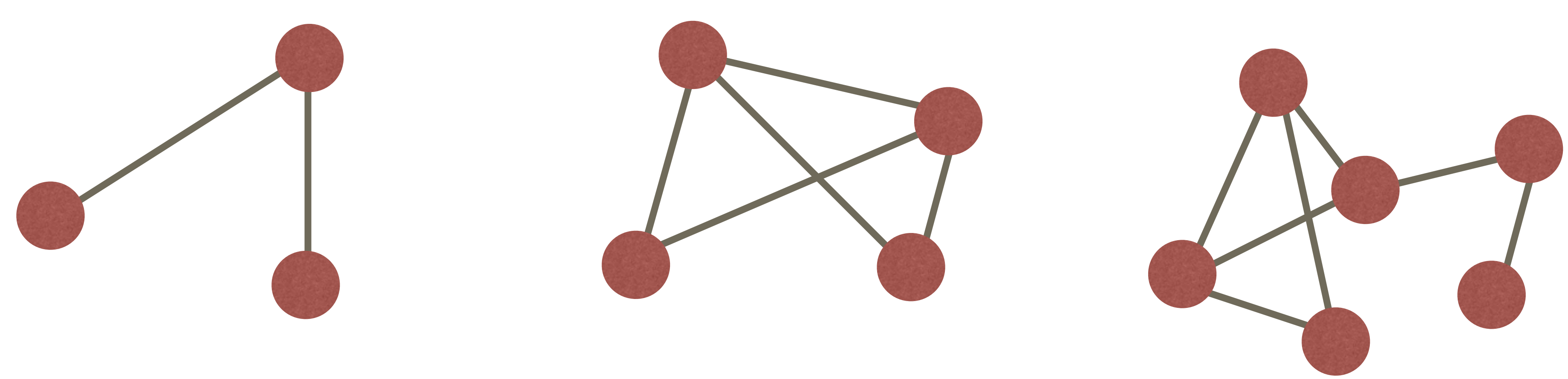
Can you work out the diameter of this directed network?

FUNDAMENTAL CONCEPTS: AVERAGE PATH LENGTH

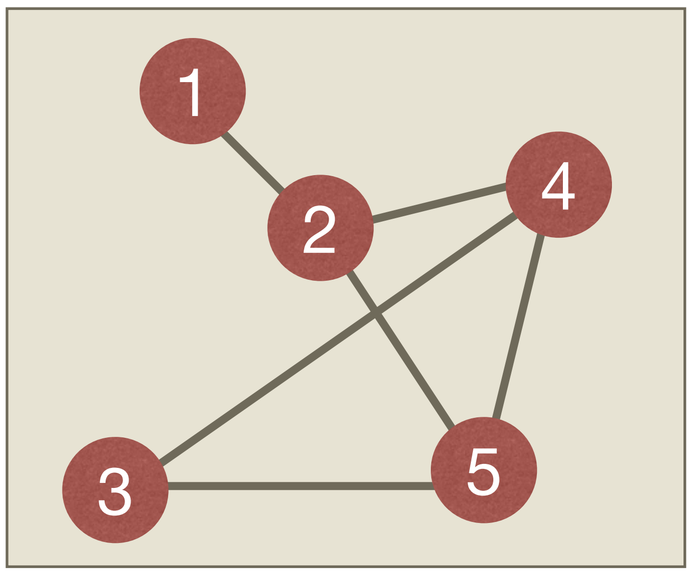
- The average path length is the average of the shortest path lengths between every pair of nodes in the network.
- For a network comprising N nodes, if $d(i, j)$ is the shortest number of steps between nodes i and j, then the average path length is:

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d(i, j)$$

What is the average path length of these networks?



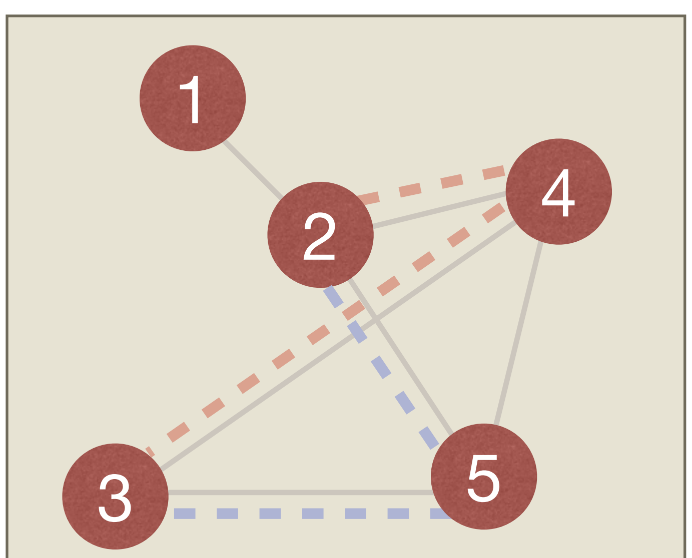
MORE ON PATH LENGTHS: TOTAL NUMBER OF WALKS OF A GIVEN LENGTH



#walks of length 1

	1	2	3	4	5
1	0	1	0	0	0
2	1	0	0	1	1
3	0	0	0	1	1
4	0	1	1	0	1
5	0	1	1	1	0

The total number of walks $N_{ij}^{(1)}$ of length 1 between a pair of nodes (i, j) is just A_{ij}

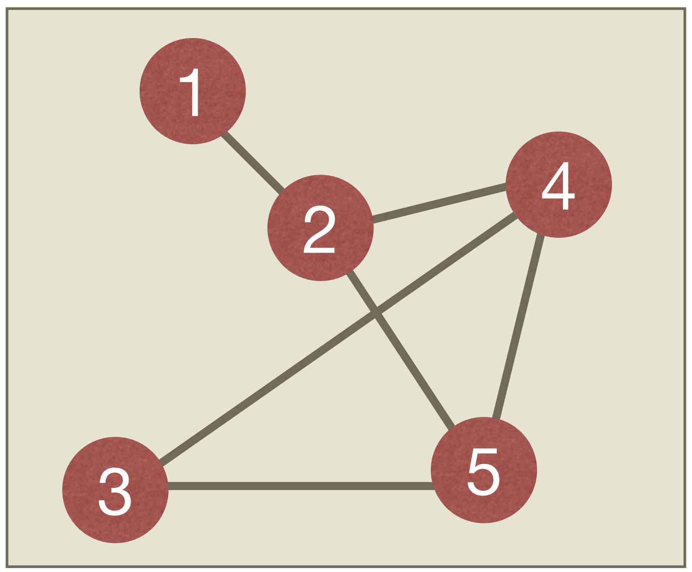


#walks of length 2

	1	2	3	4	5
1	1	0	0	1	1
2	0	3	2	1	1
3	0	2	2	1	1
4	1	1	1	3	2
5	1	1	1	2	3

How does one find the total number of walks $N_{ij}^{(d)}$ of length d between pair (i, j)?

MORE ON PATH LENGTHS:
TOTAL NUMBER OF WALKS OF A GIVEN LENGTH



	1	2	3	4	5
1	0	1	0	0	0
2	1	0	0	1	1
3	0	0	0	1	1
4	0	1	1	0	1
5	0	1	1	1	0

	1	2	3	4	5
1	0	1	0	0	0
2	1	0	0	1	1
3	0	0	0	1	1
4	0	1	1	0	1
5	0	1	1	1	0

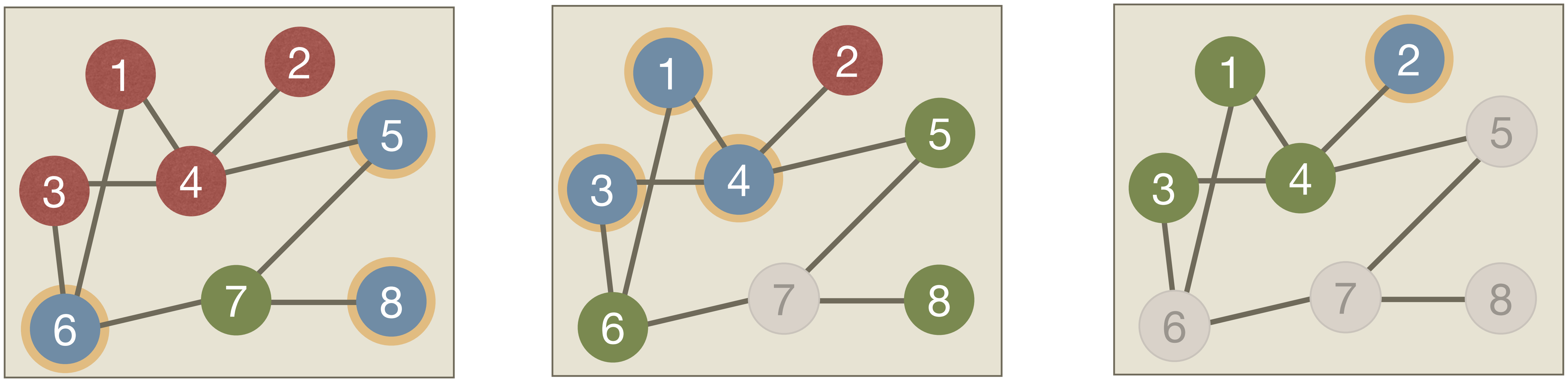
In order to have a walk of length 2 between nodes $(2,3)$, we consider all the nodes of distance 1 from node #2, and count how many of them are distance 1 from node #3.

$$N_{23}^{(2)} = \sum_k A_{2k} A_{k3}$$

i.e. $N^{(d)} = \underbrace{A\,A\ldots\,A}_{d} = A^d$

MORE ON PATH LENGTHS: BREADTH-FIRST SEARCH

distance 1 from node #7 | distance 2 from node #7 | distance 3 from node #7

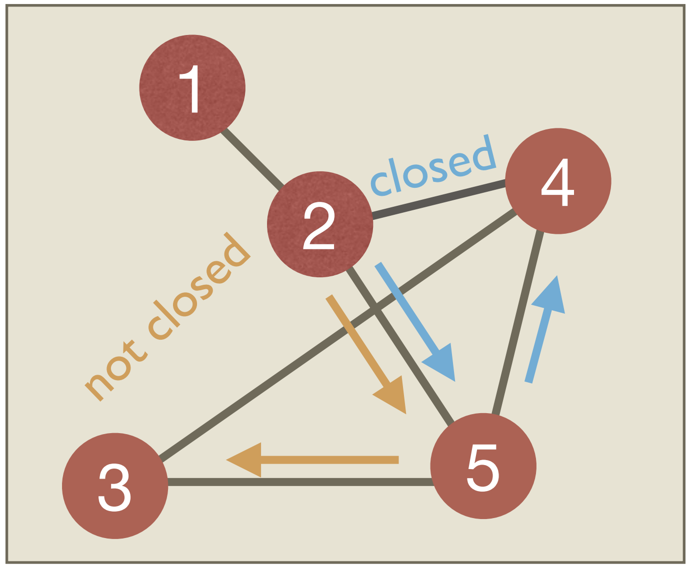


To find the shortest path between nodes (i, j), we can follow the breadth-first search algorithm:

1. Find the neighbours (blue) of node i (green) from the adjacency matrix A.
2. Remove the green node and make the blue nodes green.
3. Find the neighbours of the green nodes (excluding removed ones).
4. Repeat as long as there are neighbours.

FUNDAMENTAL CONCEPTS: CLUSTERING COEFFICIENT

- In real networks, one often finds that nodes that form links with one another also form links with those that the neighbour link to.
- This can be measured by the (global) clustering coefficient: the fraction of paths of length 2 that are "closed" (the three nodes of the path are all connected).
- A triangle of nodes connected to each other contain 3 closed paths.



Thus, the global clustering coefficient is:

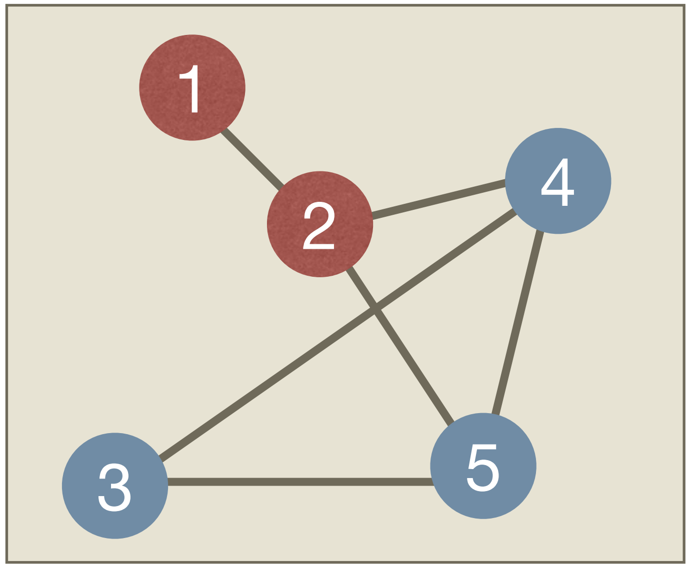
$$C = \frac{\#\text{triangles} \times 3}{\#\text{connected triples}}$$

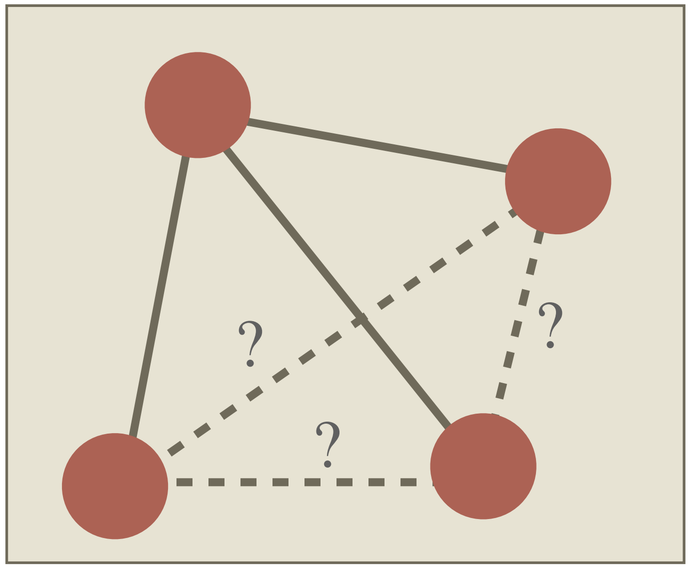
where a connected triple is a path of length 2 (either closed or not closed).

FUNDAMENTAL CONCEPTS: LOCAL CLUSTERING COEFFICIENT

- The (local) clustering coefficient of a node measures the extent of connectivity of its local neighbourhood, i.e. how close they are to being a "clique" or a complete subgraph.
- If a node i in an undirected network has k_i neighbours, there can be a maximum of $k_i(k_i - 1)/2$ links between them.
- The local clustering coefficient C_i of node i is the fraction of these links that exist.

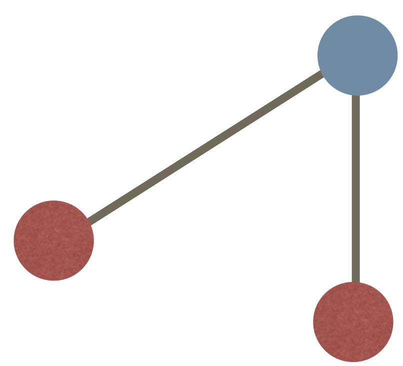
blue nodes form a clique

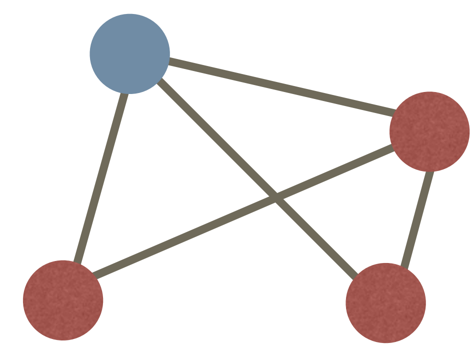


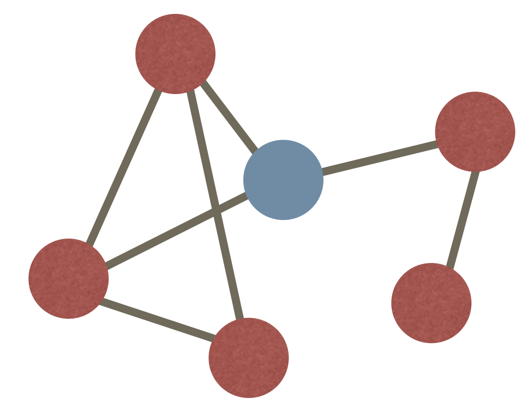


FUNDAMENTAL CONCEPTS: CLUSTERING COEFFICIENT

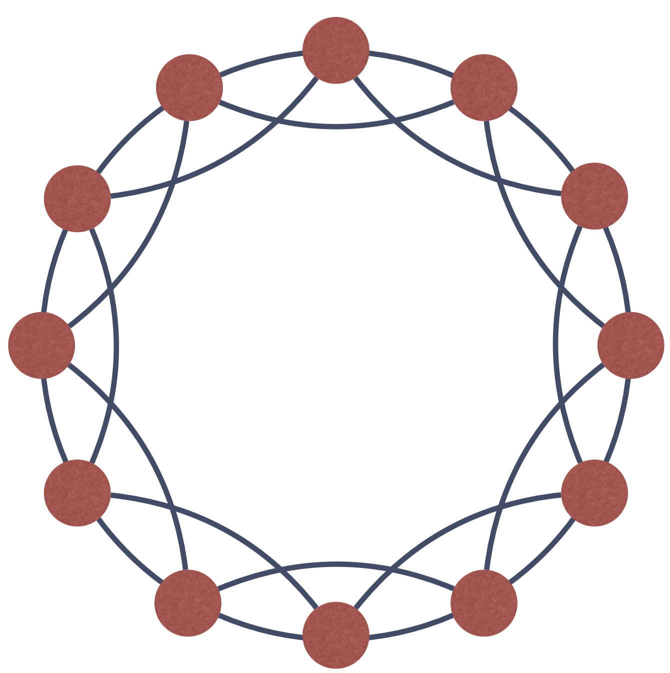
What is the clustering coefficient of the blue nodes?



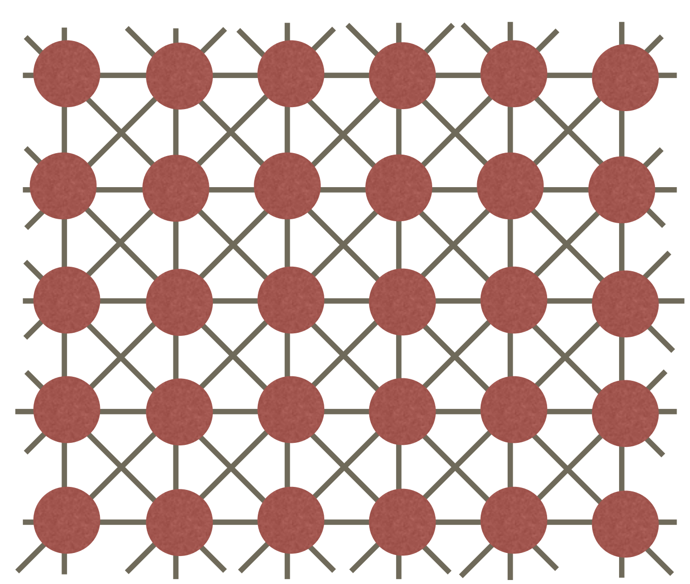




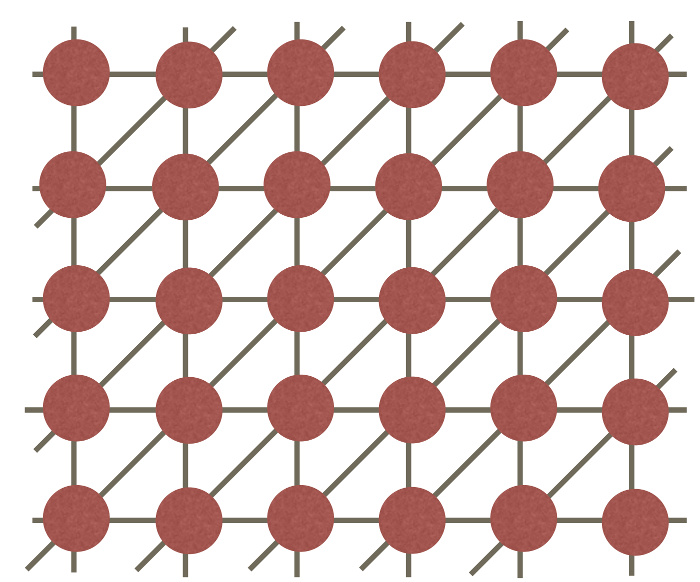
Calculate the clustering coefficients for a node in the following networks:



$k_i = 4$

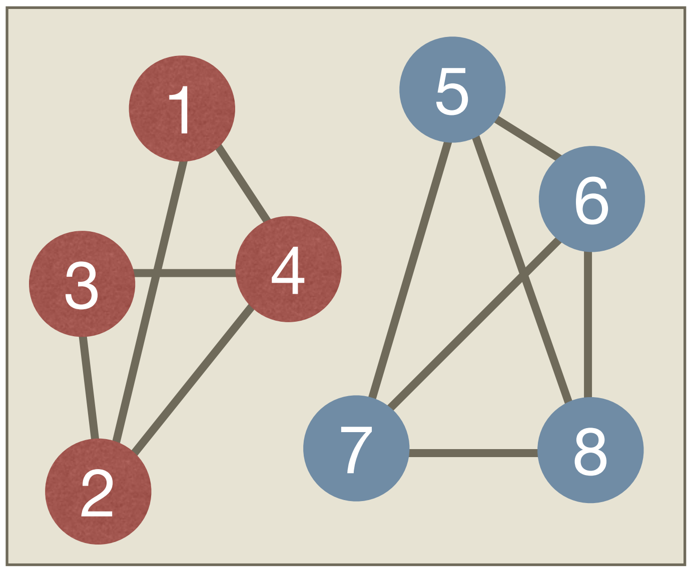


$k_i = 8$

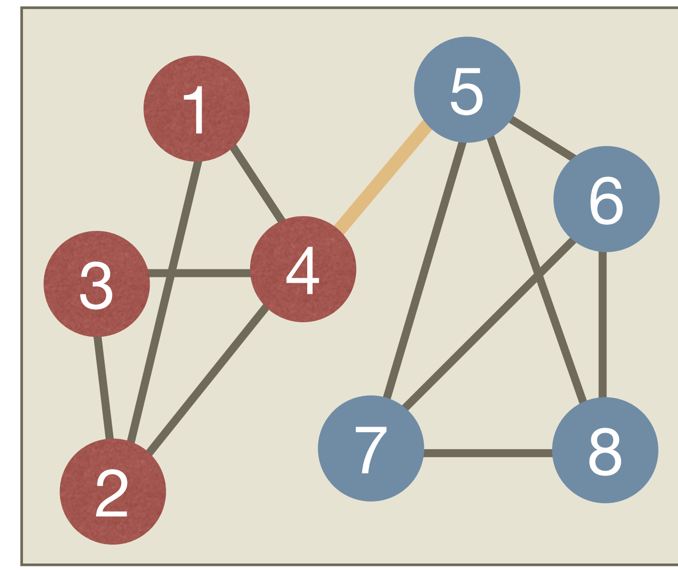


$k_i = 6$

FUNDAMENTAL CONCEPTS: COMPONENTS



	1	2	3	4	5	6	7	8
1	0	1	0	1	0	0	0	0
2	1	0	1	1	0	0	0	0
3	0	1	0	1	0	0	0	0
4	1	1	1	0	0	0	0	0
5	0	0	0	0	0	1	1	1
6	0	0	0	0	1	0	1	1
7	0	0	0	0	1	1	0	1
8	0	0	0	0	1	1	1	0

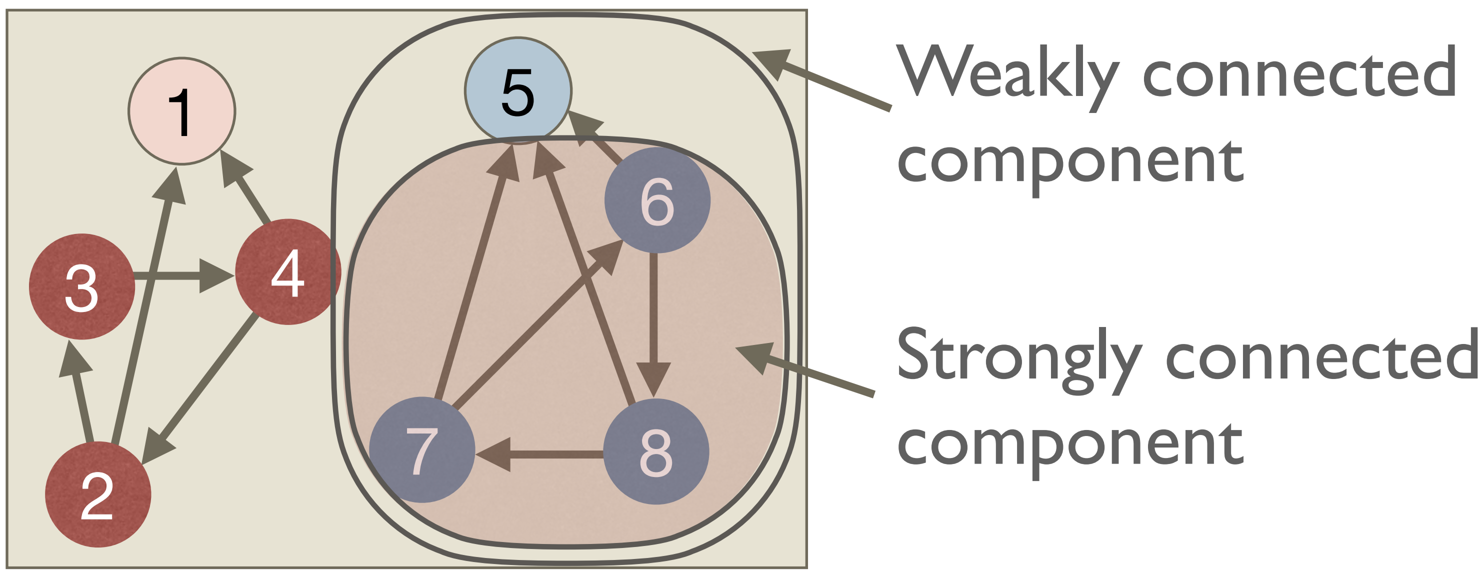


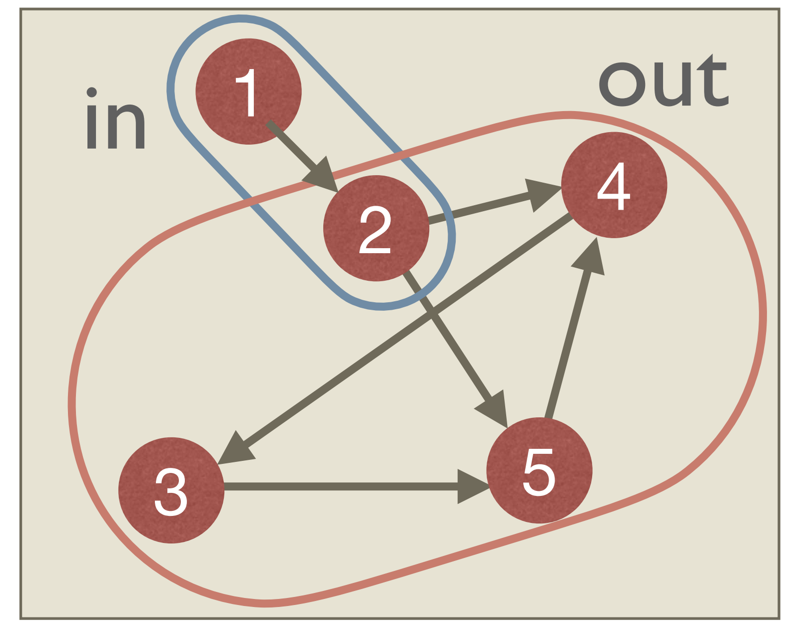
In an undirected network, a pair of nodes (i, j) are connected if there exists a path (of any length) between them.

A component is a subset of the network in which all nodes are connected.

A bridge is a link that, when cut, causes the network to be disconnected.

FUNDAMENTAL CONCEPTS: COMPONENTS



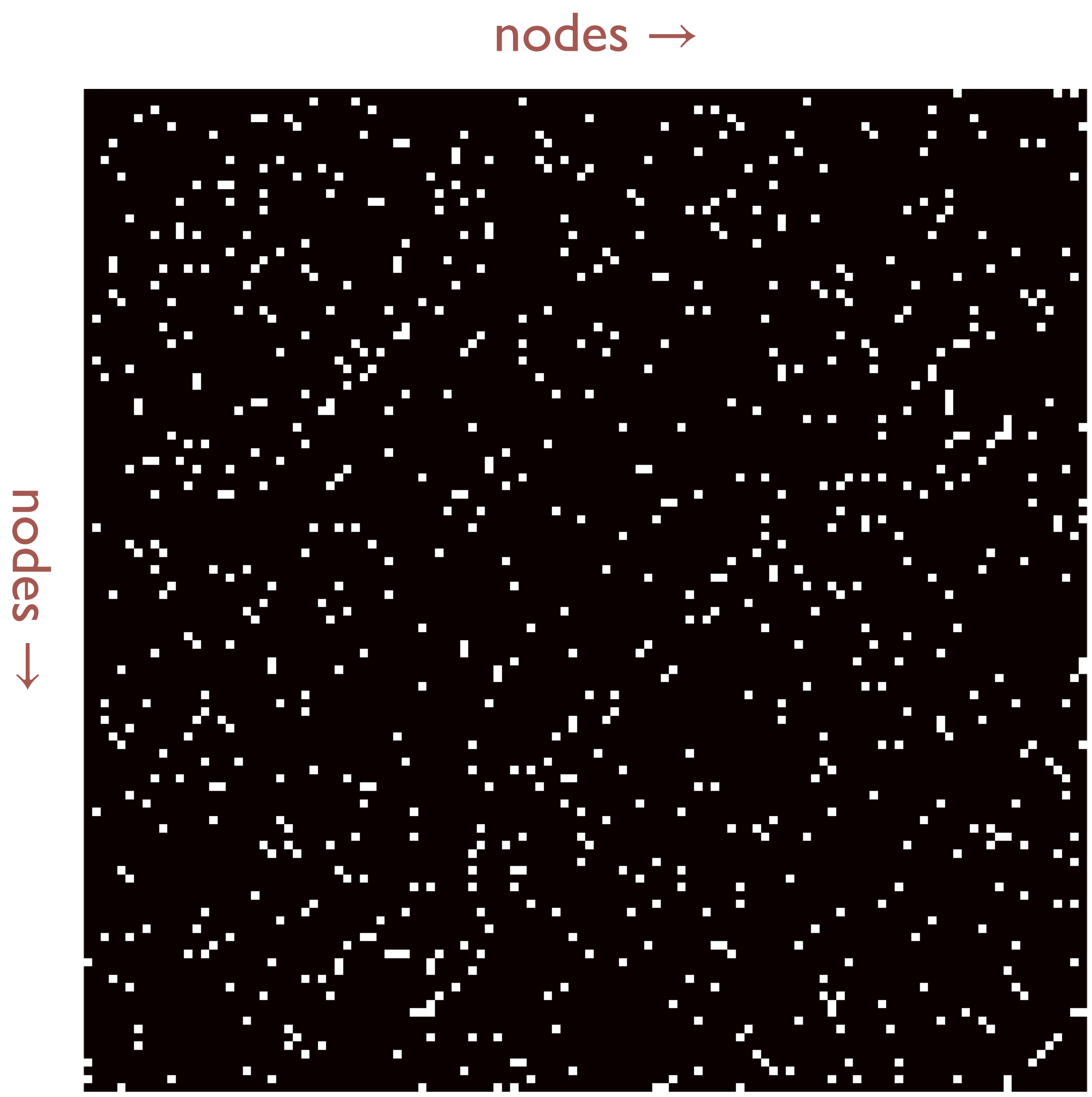


In an directed network, a strongly connected component is one where exists a path between all constituent nodes.

A weakly connected component is a connected component that exists if one were to ignore the directed nature of the edges.

The in-component of a node in a directed network is the set that can reach it, and its out-component is the set that can be reached from it.

FUNDAMENTAL CONCEPTS: COMPONENTS

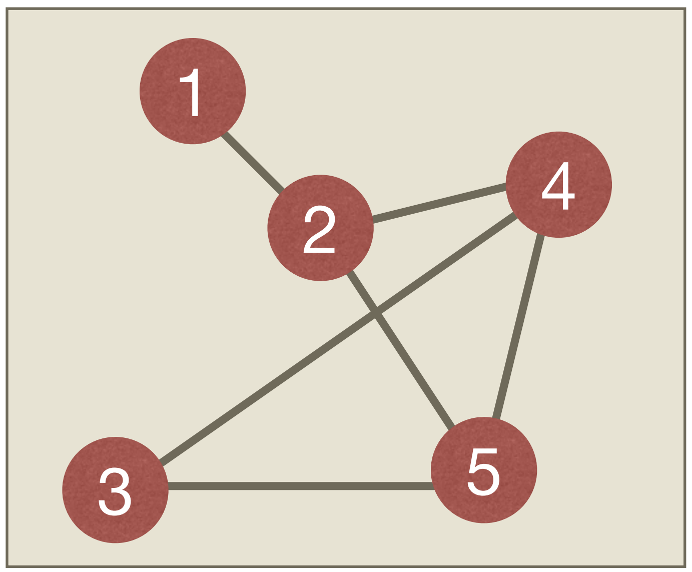


The image on the right displays the adjacency matrix of a large undirected network.

White squares represent connections between nodes, while black represents the absence of a link.

Can you guess the number of connected components of this network? 😳

FUNDAMENTAL CONCEPTS: THE GRAPH LAPLACIAN



Adjacency matrix

	1	2	3	4	5
1	0	1	0	0	0
2	1	0	0	1	1
3	0	0	0	1	1
4	0	1	1	0	1
5	0	1	1	1	0

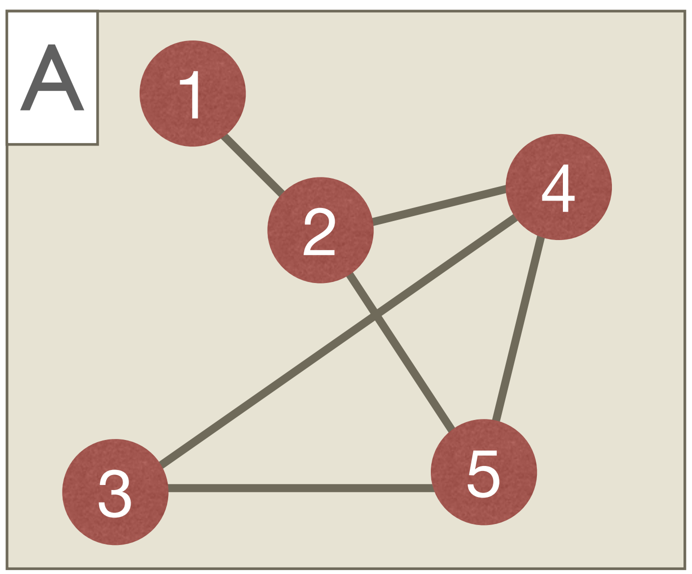
graph Laplacian

	1	2	3	4	5
1	1	-1	0	0	0
2	-1	3	0	-1	-1
3	0	0	2	-1	-1
4	0	-1	-1	3	-1
5	0	-1	-1	-1	3

For the case of undirected networks with no self-edges, one can define the graph Laplacian $\mathbf{L}$ as follows: $L_{ij} = k_i \delta_{ij} - A_{ij}$ or $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where the degree matrix $D_{ij} = k_i \delta_{ij}$ contains the degree along the diagonal and $0s$ elsewhere.

If the network is weighted, the definition is as follows: $L_{ij} = \sum_j A_{ij} \delta_{ij} - A_{ij}$

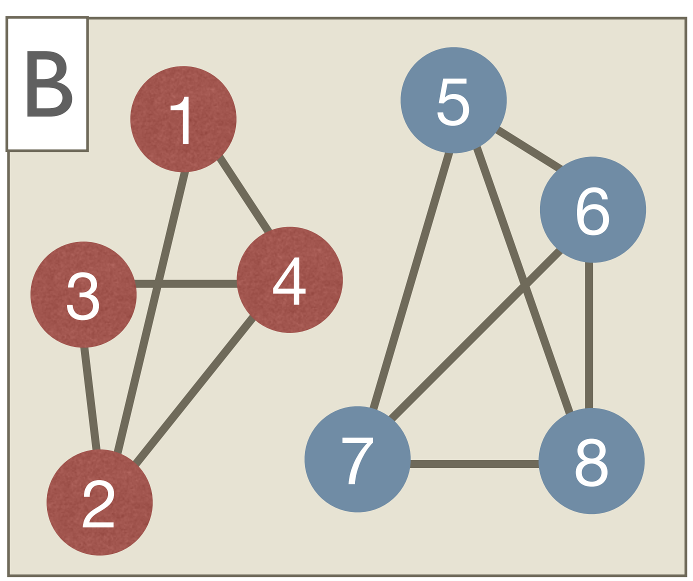
FUNDAMENTAL CONCEPTS: THE GRAPH LAPLACIAN



L_A

	1	2	3	4	5
1	1	-1	0	0	0
2	-1	3	0	-1	-1
3	0	0	2	-1	-1
4	0	-1	-1	3	-1
5	0	-1	-1	-1	3

The number of <u>zero eigenvalues</u> of the laplacian indicate the number of connected components of the network.



L_B

	1	2	3	4	5	6	7	8
1	2	-1	0	-1	0	0	0	0
2	-1	3	-1	-1	0	0	0	0
3	0	-1	2	-1	0	0	0	0
4	-1	-1	-1	3	0	0	0	0
5	0	0	0	0	3	-1	-1	-1
6	0	0	0	0	-1	3	-1	-1
7	0	0	0	0	-1	-1	3	-1
8	0	0	0	0	-1	-1	-1	3

$$\lambda_A = eig(L_A) = \{0, 0.83, 2.69, 4, 4.48\}$$

$$\lambda_B = eig(L_B) = \{0, 0, 2, 4, 4, 4, 4, 4\}$$

FUNDAMENTAL CONCEPTS: THE GRAPH LAPLACIAN

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 3 & 0 & -1 & -1 \\ 0 & 0 & 2 & -1 & -1 \\ 0 & -1 & -1 & 3 & -1 \\ 0 & -1 & -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

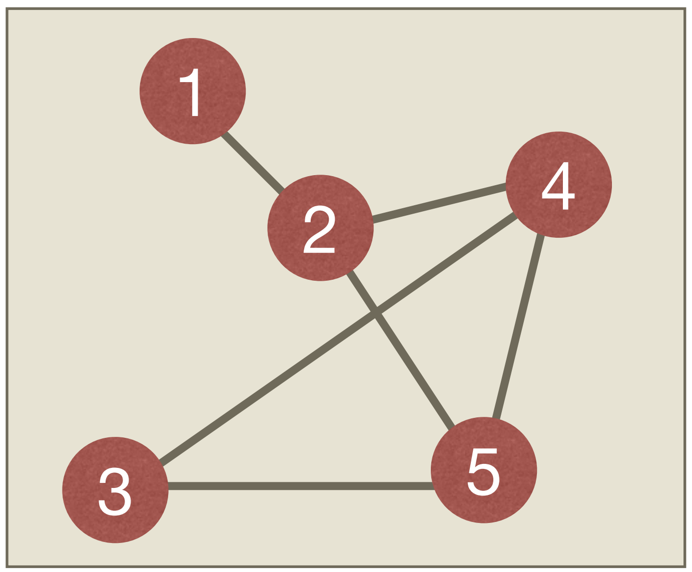
If $\mathbf{v}$ is an eigenvector of the Laplacian and λ is its associated eigenvalue, then

$$\mathbf{Lv} = \lambda \mathbf{v}$$

A Laplacian with a single component, has an eigenvector $\mathbf{v} = [1,1,\ldots]^T$ with $\lambda = 0$.

A Laplacian with a two components, has $\mathbf{v} = [1,1,\ldots,0,0,\ldots]^T$ and $\mathbf{v} = [0,0,\ldots,1,1,\ldots]^T$ both with $\lambda = 0$.

FUNDAMENTAL CONCEPTS: BETWEENNESS CENTRALITY



On the right are the list of all possible shortest paths between every pair of nodes in the above network.

SHORTEST PATHS	
1-2	{1,2}
1-3	{1,2,5,3}, {1,2,4,3}
1-4	{1,2,4}
1-5	{1,2,5}
2-3	{2,4,3}, {2,5,3}
2-4	{2,4}
2-5	{2,5}
3-4	{3,4}
3-5	{3,5}
4-5	{4,5}

To find the betweenness centrality of a node, we count the fraction of times it appears in the shortest paths between other nodes.

FUNDAMENTAL CONCEPTS: BETWEENNESS CENTRALITY

	SHORTEST PATHS
1-2	{1,2}
1-3	{1,2,5,3}, {1,2,4,3}
1-4	{1,2,4}
1-5	{1,2,5}
2-3	{2,4,3}, {2,5,3}
2-4	{2,4}
2-5	{2,5}
3-4	{3,4}
3-5	{3,5}
4-5	{4,5}

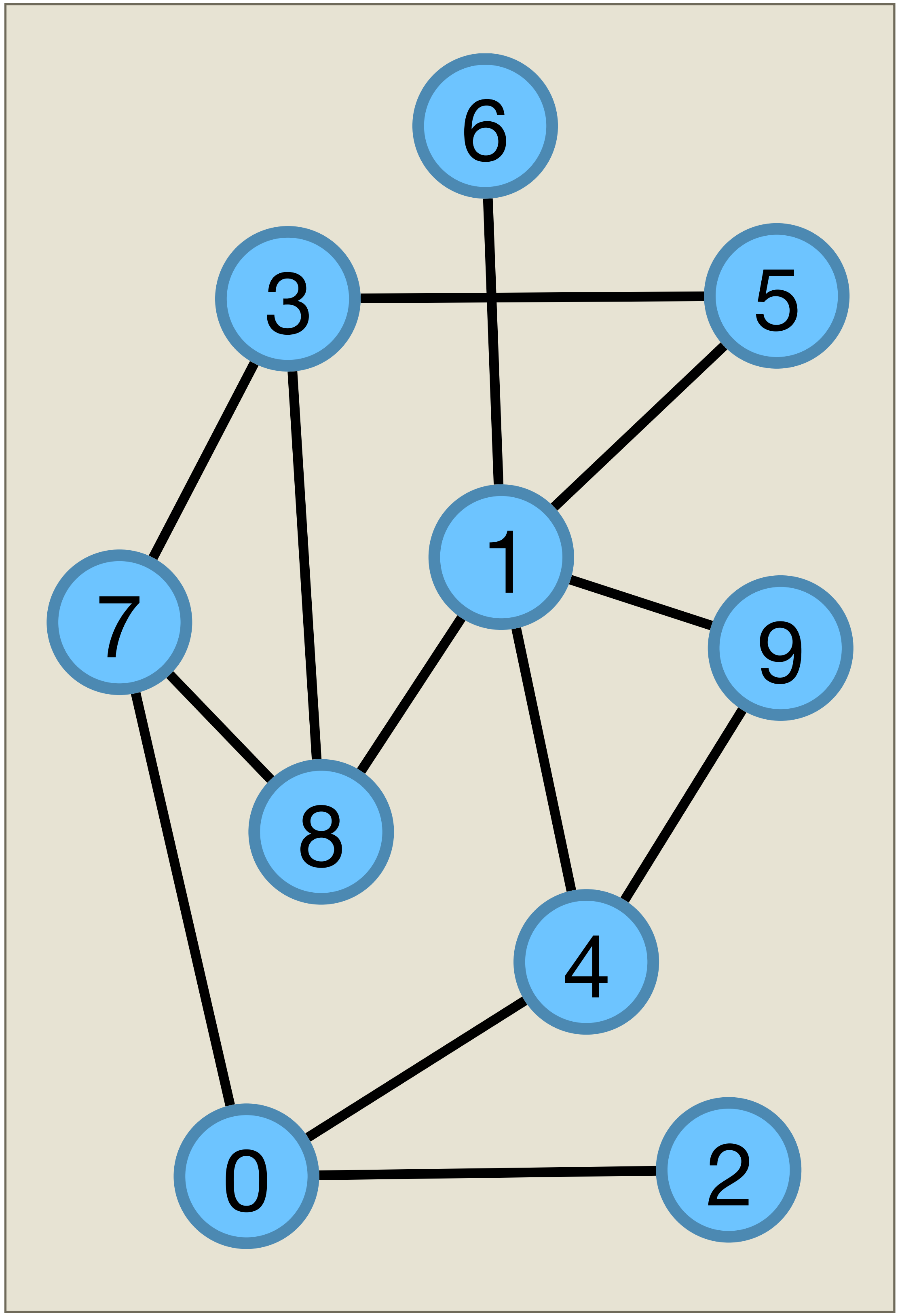
	OCCURRENCES	C_B
1	0	0
2	2/2 + 1 + 1	3
3	0	0
4	1/2 + 1/2	1
5	1/2 + 1/2	1

If σ_{st} is the no. of shortest paths from s to t, and $\sigma_{st}(v)$ is the number of these containing node v, then:

$$C_B(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

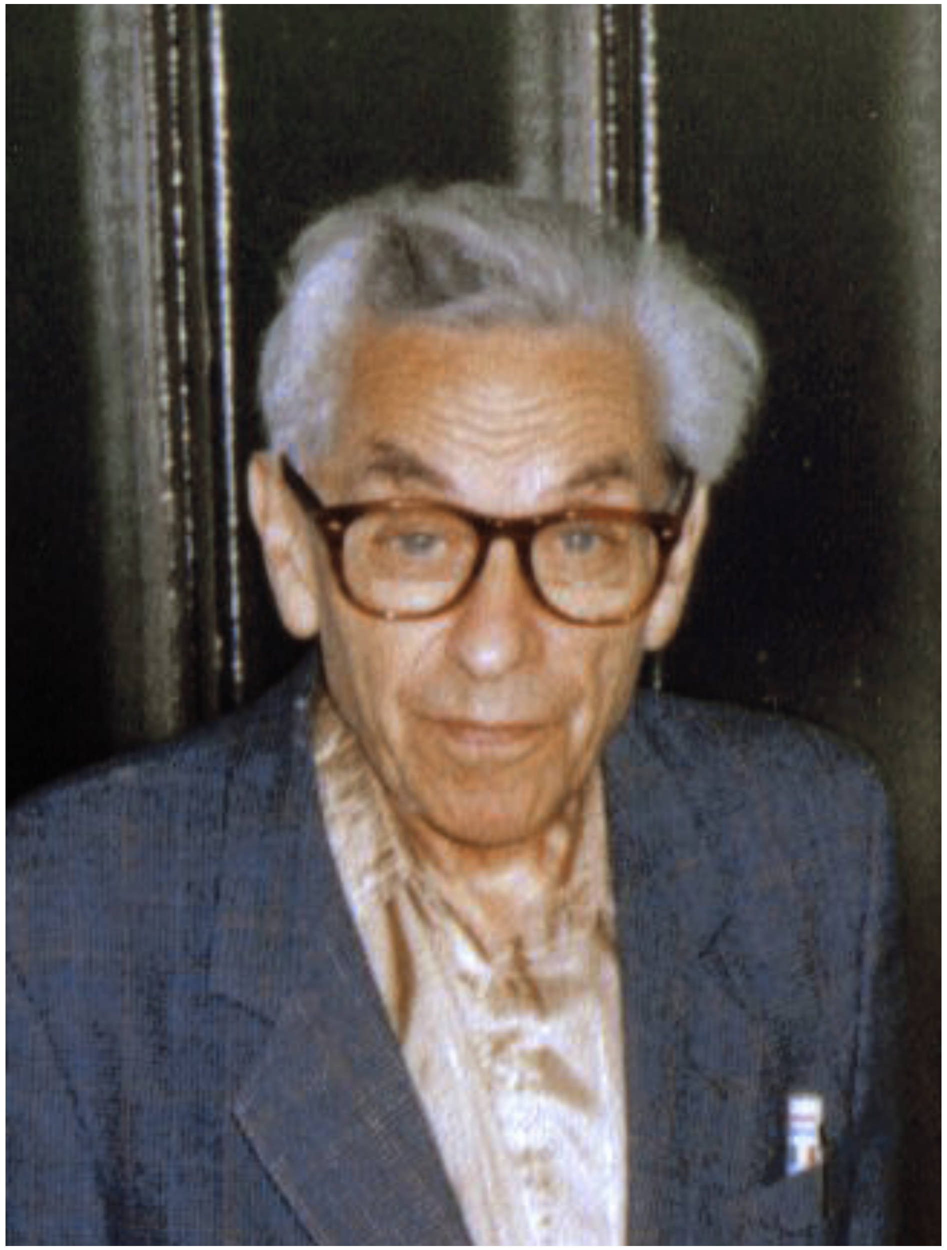
Is the betweenness centrality.

EXERCISES



1. Find the average degree of the network.
2. Find the shortest path length between nodes 3 & 4.
3. Find the diameter of the network
4. Find the number of walks of length 5 between nodes 2 & 3.
5. Find the clustering coefficient of nodes 1 & 8.
6. Find the global clustering coefficient.
7. Find the betweenness centrality of nodes 3 & 4.

WHAT DO THESE TWO HAVE IN COMMON?



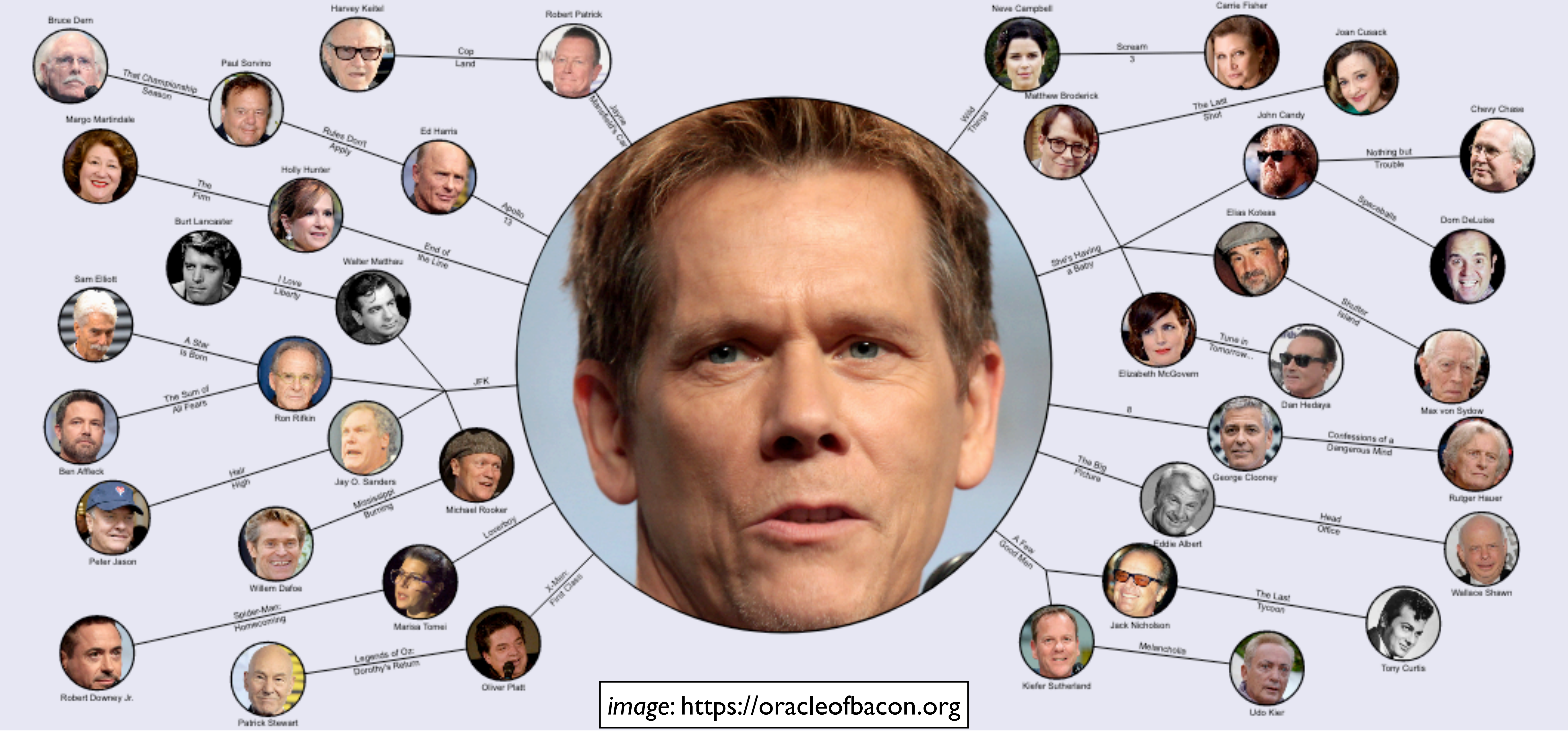
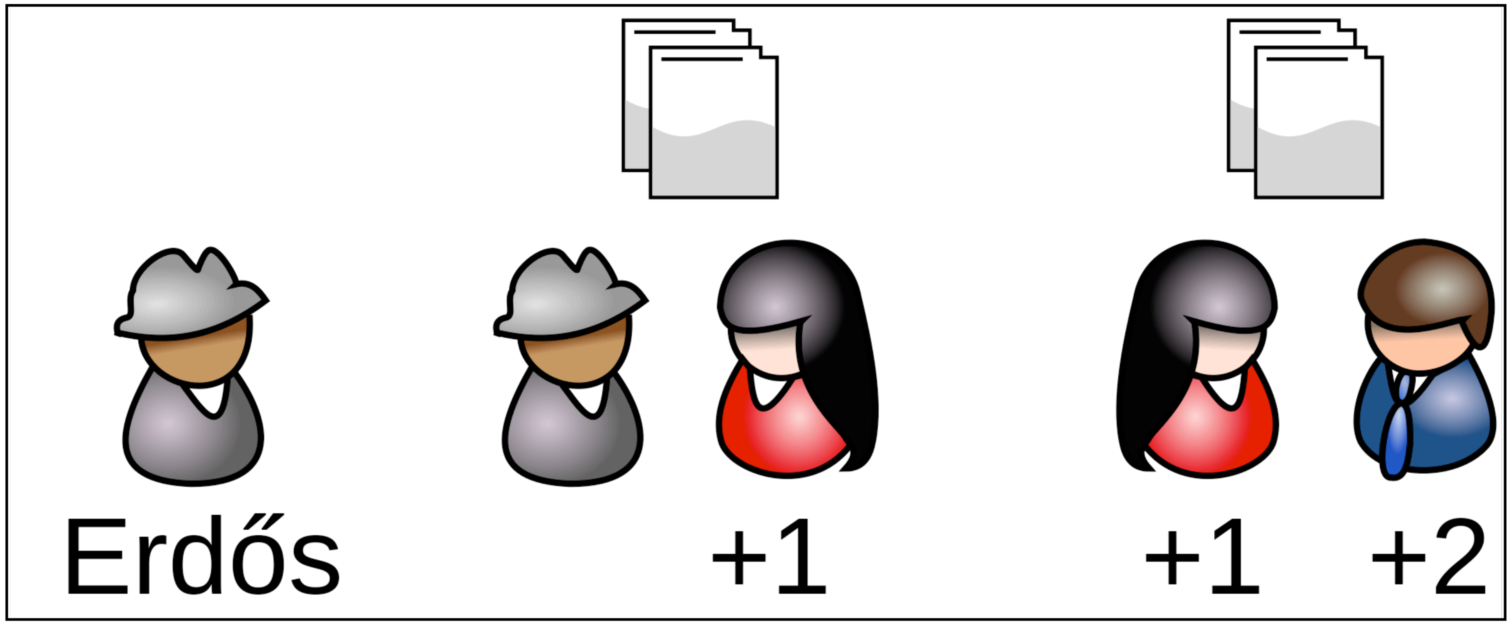
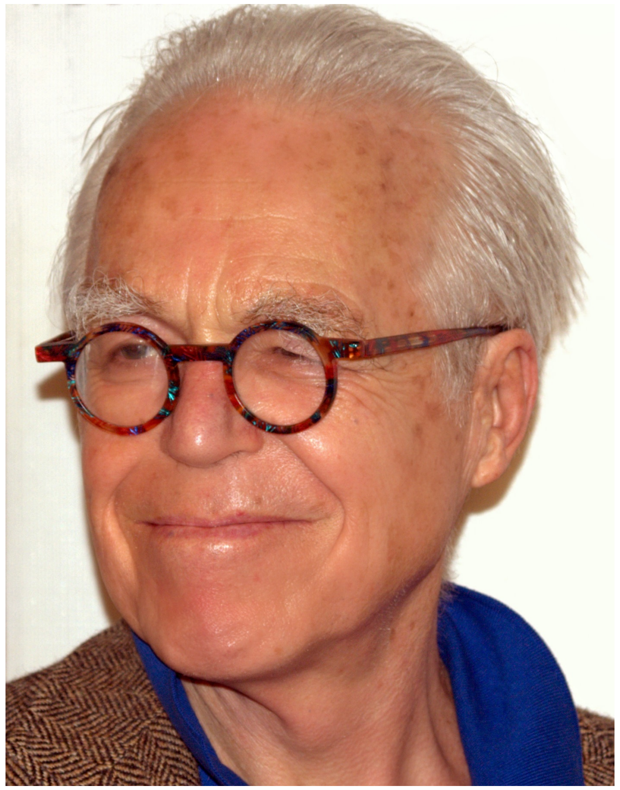


image: https://oracleofbacon.org

Both their names are popularly associated with the concept of "six degrees of separation"





"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 is to the best of my knowledge a good friend of mine - so I could easily ask him to send a telegram via the general director telling Ford that he should talk to the manager and have the worker in the shop quickly hammer together a car for me, as I happen to need one."

Frigyes Karinthy, "*Láncszemek (Chains)*" (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. Fill in the names."

John Guare, "*Six Degrees of Separation*" (1990).

MILGRAM'S LETTER EXPERIMENT

- Information packets were sent to 296 random individuals in Nebraska and Kansas.
- They were asked to forward the letter to someone who they think might know a specified person in Boston (details of the person were mentioned).
- Of the letters that reached the final target, the average path length was ~ 5.2.

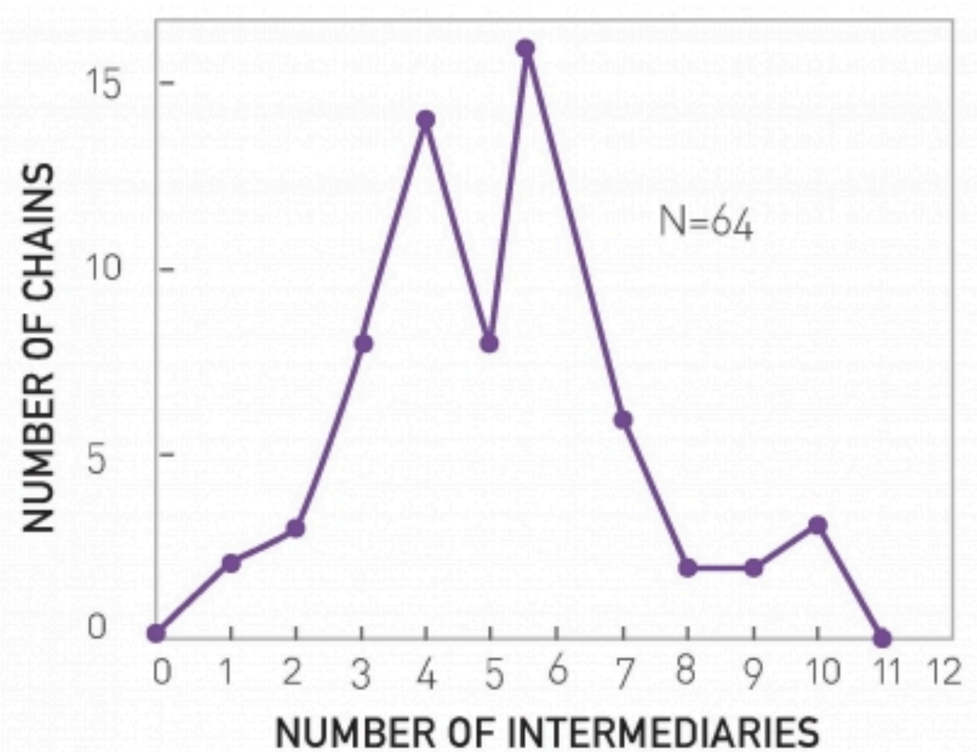
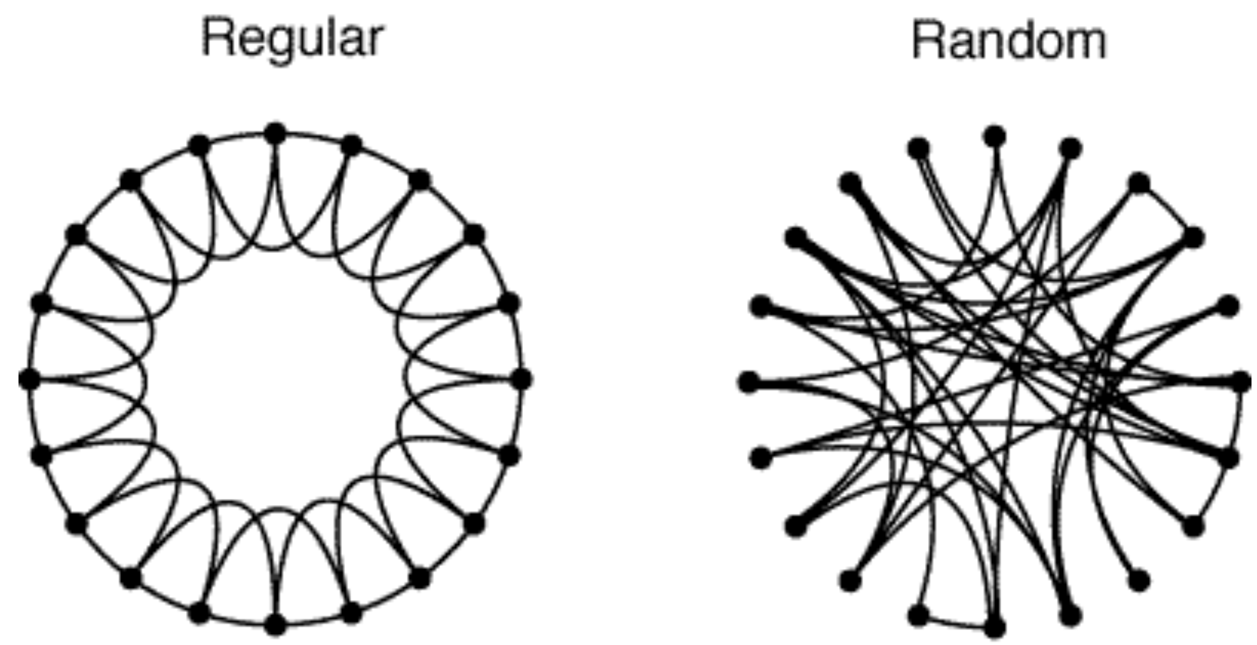


image [left]: Barabási, A.-L., "Network Science" (http://networksciencebook.com/)

SMALL WORLD NETWORKS



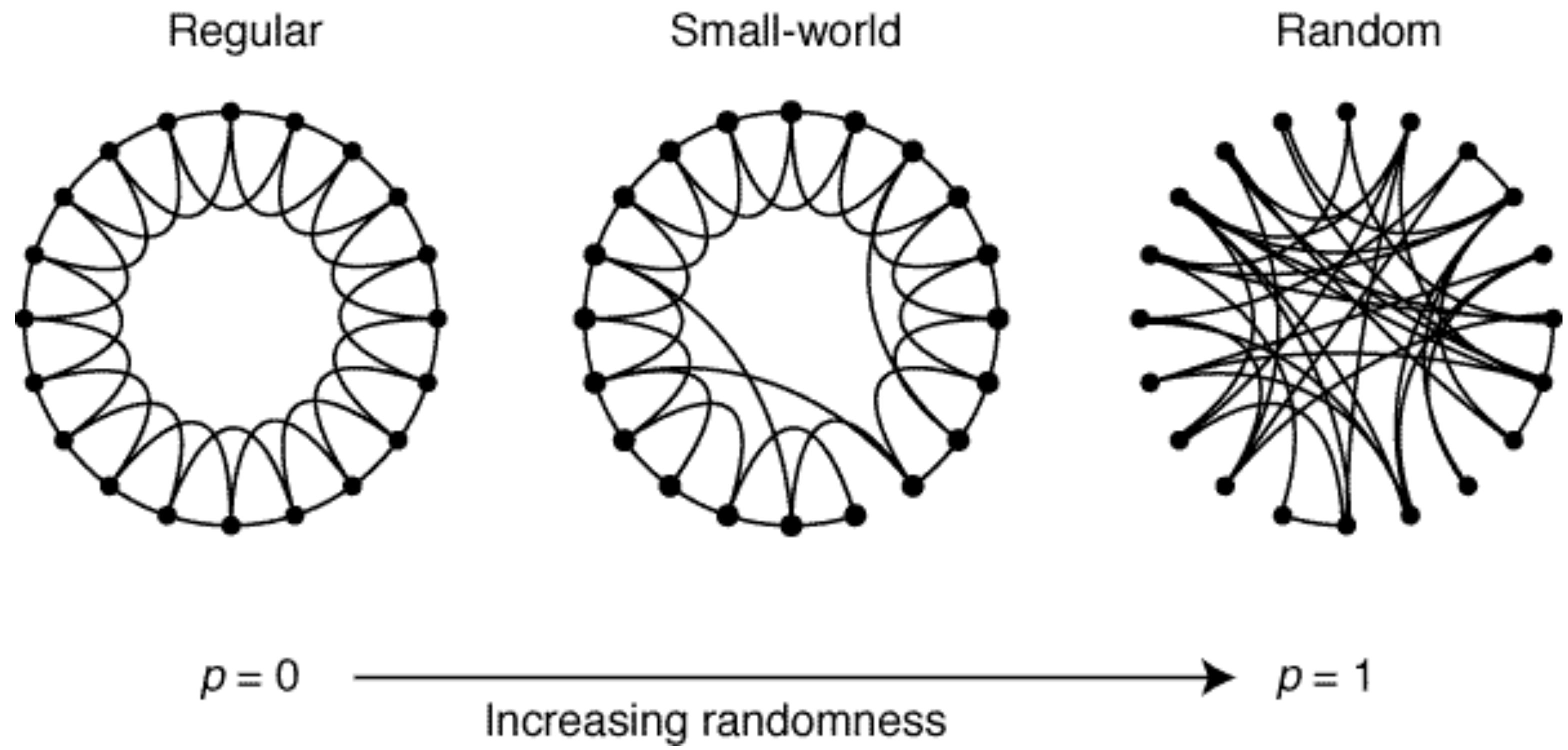
A regular graph has high clustering coefficient C and high average path length L. A random graph has the opposite properties.

Watts and Strogatz (1998) suggested a procedure for obtaining networks with properties of both regular and random networks:

- Start with a regular graph where each node has K neighbours.
- Cycle through each node, and consider the $K/2$ rightward links.
- Randomly rewire each of these links with probability p, avoiding self-loops and duplicate links.

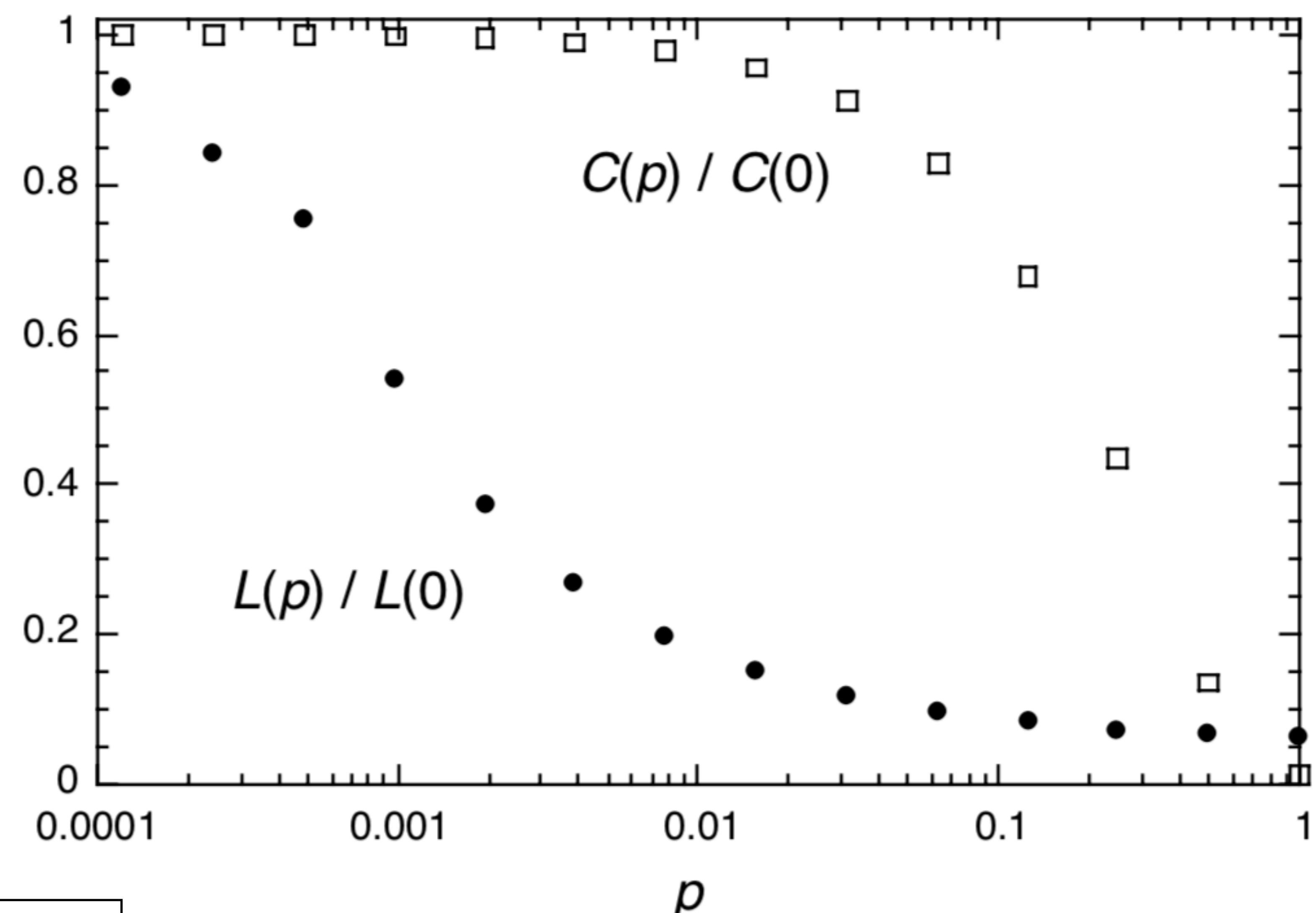
image: Watts, D. & Strogatz, S., *Nature* **393**, 440-442 (1998).

SMALL WORLD NETWORKS



For an intermediate p the networks have a low average path length L and a high clustering coefficient C.

These are referred to as "small-world" networks.

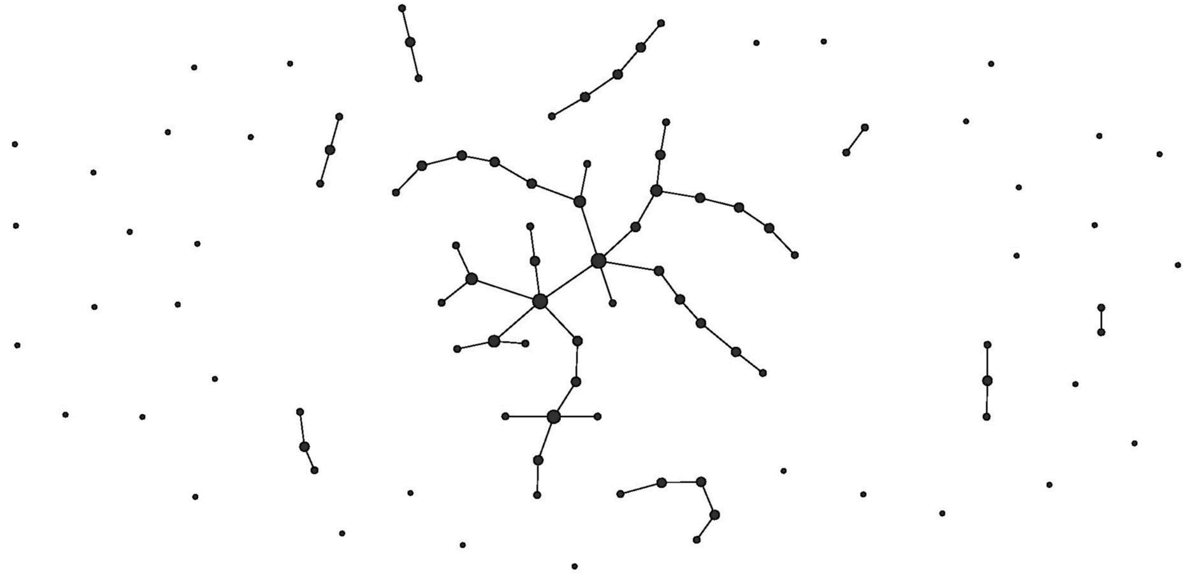


images: Watts, D. & Strogatz, S., *Nature* **393**, 440-442 (1998).

ERDŐS-RÉNYI NETWORKS

- In 1959 two related models for generating random networks were proposed.
- In the more commonly used version of the model ($G(n, p)$), we consider a group of n nodes, and connect each pair with a probability p.

Pál Erdős Alfréd Rényi Edgar Gilbert



- For certain choices of (n, p), the resulting network may have multiple connected components.
- These random networks are commonly referred to as Erdős-Rényi (ER) networks.

image: By Vonfrisch, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=3469734

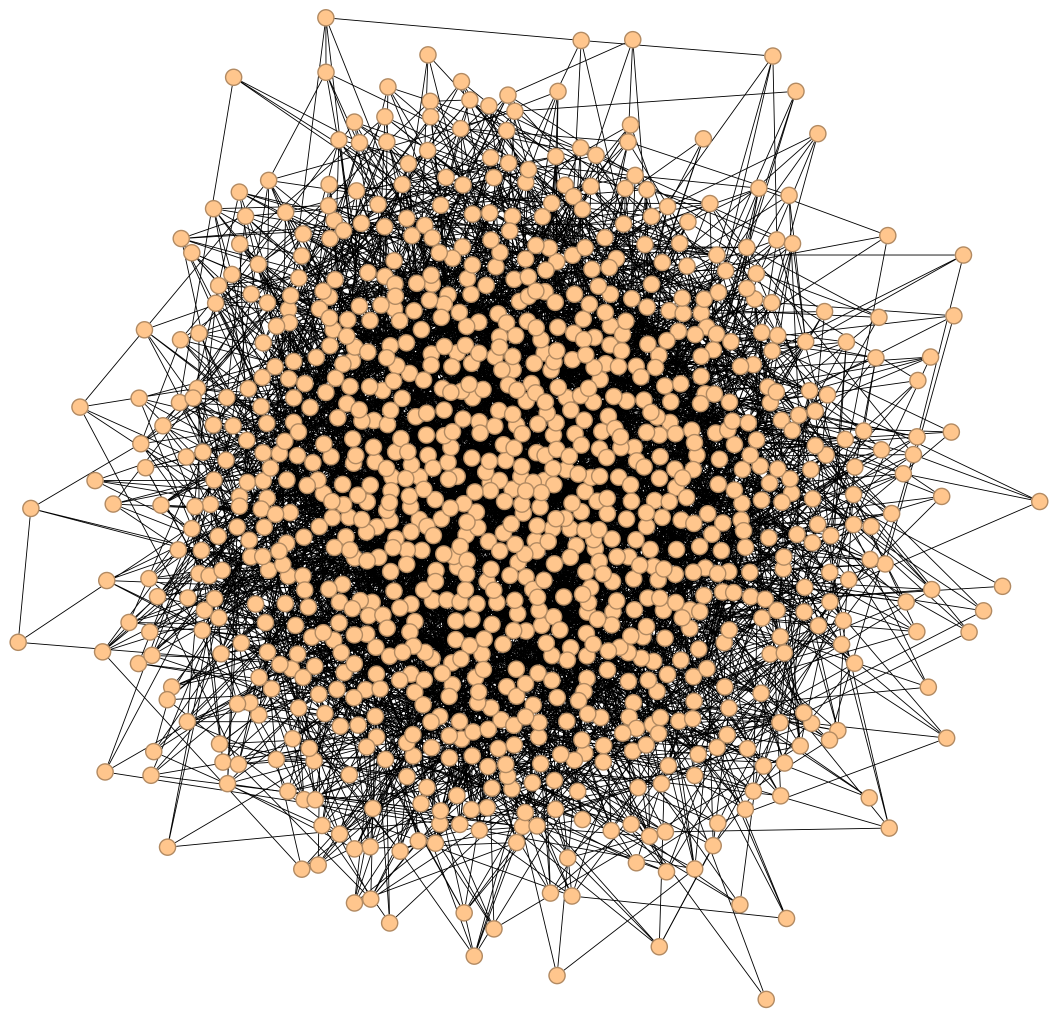
ERDŐS-RÉNYI NETWORKS

- In a network of n nodes, a node has independent probabilities of connecting to each of the other $n-1$ nodes. Thus, the probability of connecting to k chosen nodes and not to the remaining nodes is: $p^k(1-p)^{n-1-k}$.

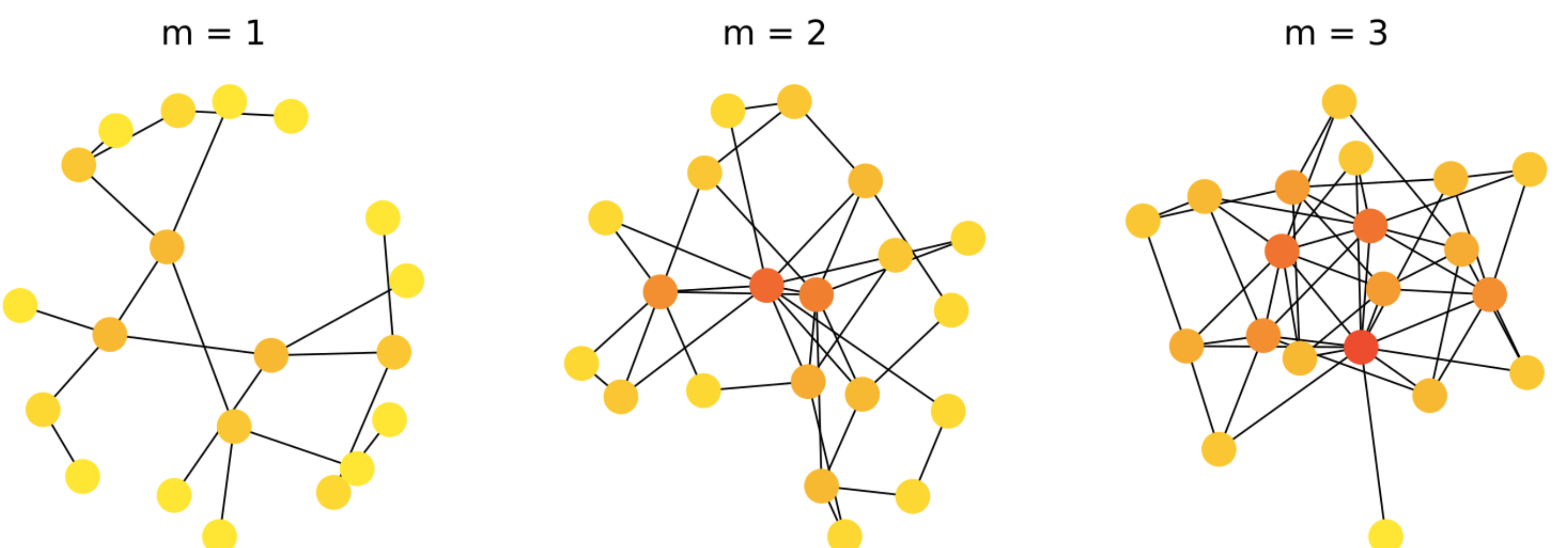
- Given that there are $\binom{n-1}{k}$ ways of choosing k out of $n-1$ nodes, the probability that a node connects to k nodes is:

$$p(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

- Thus, for an ER network the degree distribution (the probability $p(k)$ that a randomly selected node in the network has degree k) is just a binomial distribution.



SCALE-FREE NETWORKS



Albert-László Barabási

Réka Albert

- The Barabási-Albert (BA) model generates random "scale-free" networks using a preferential attachment mechanism.
- Nodes are sequentially added to the network and each connects to m random existing nodes.
- The probability that a new node connects to an existing node i is: $p_i = k_i / \Sigma_j k_j$.

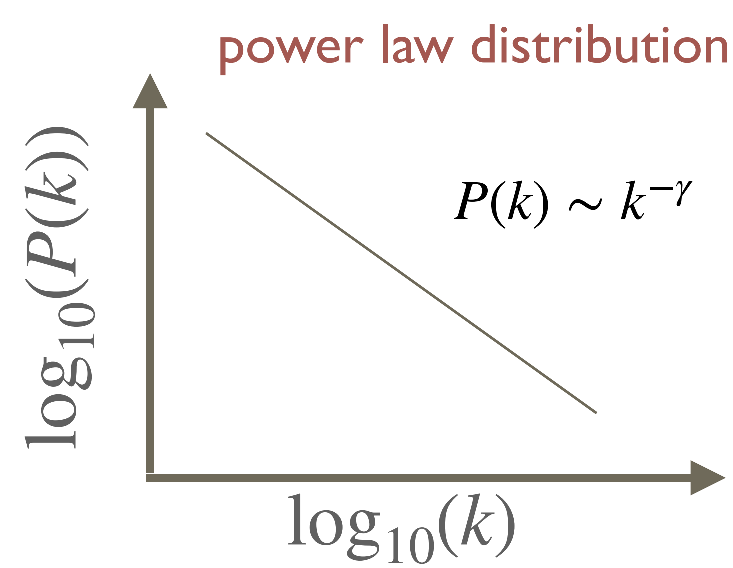
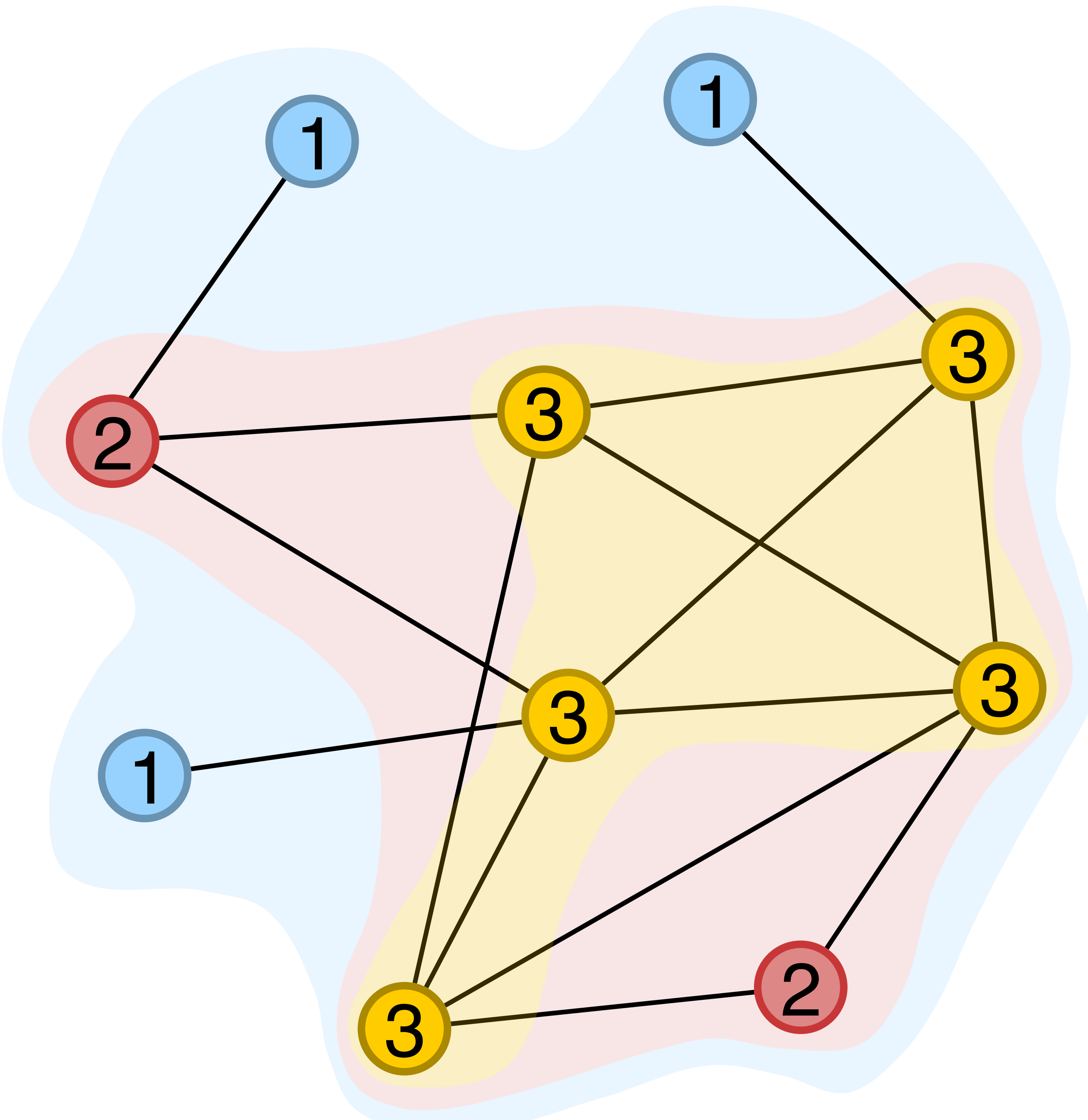


image [top]: By HeMath - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=64122479

STRUCTURAL FEATURES: CORE-PERIPHERY STRUCTURE



One can deconstruct a network in terms of its core-periphery structure.

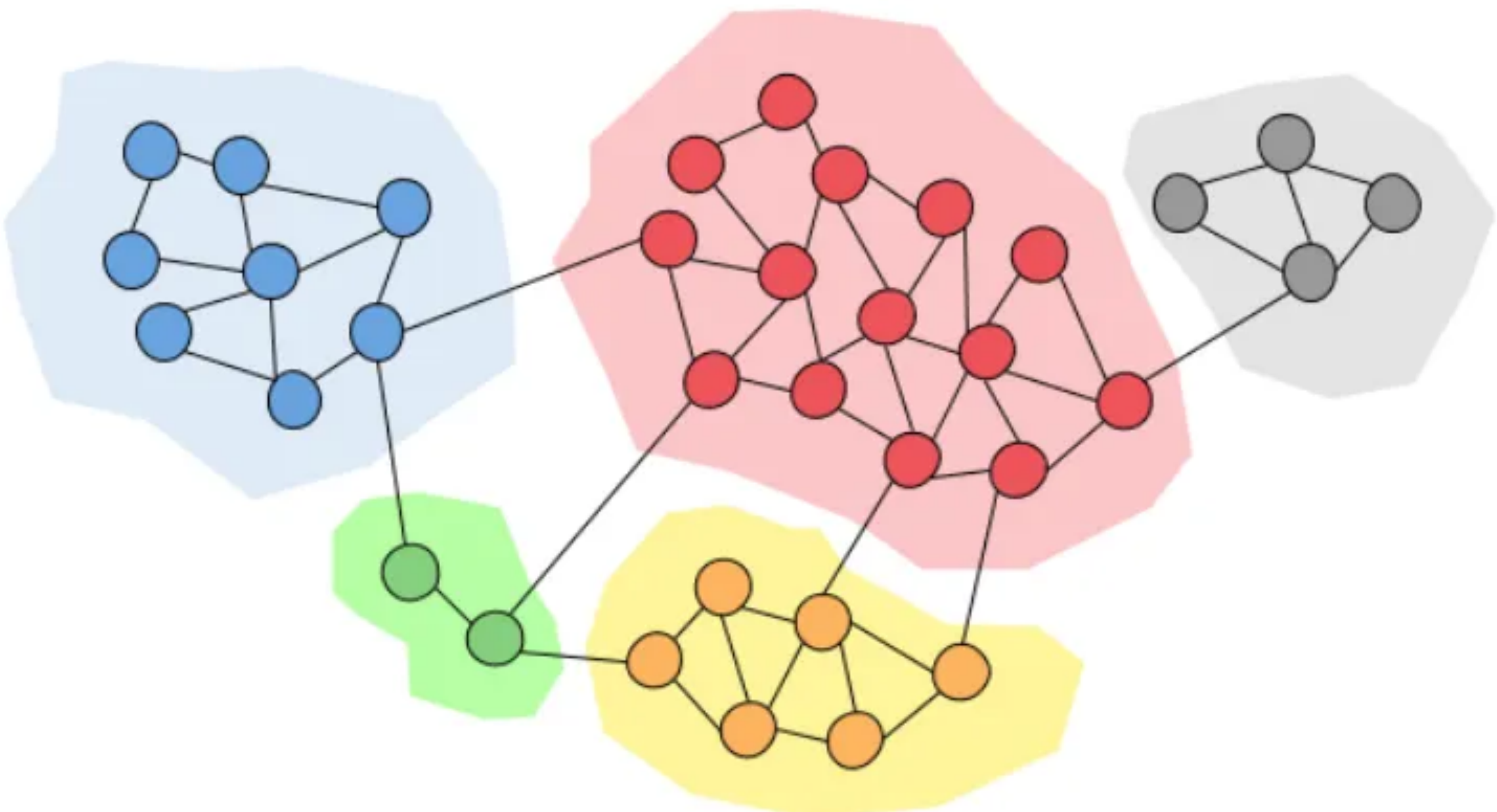
The k-core is the set of nodes of the network, within which each node has k links with each other.

A core need not be a single connected component.

The nodes of the highest k-core are referred to as core nodes and others are peripheral nodes.

STRUCTURAL FEATURES: MODULARITY

A network is said to have a modular structure if there exist groups (or "communities") of nodes that have a higher density of connections than that between groups.



In practice, one has to first specify the modules/communities and then check if the density of intra-connections is more than that of the inter-connections.

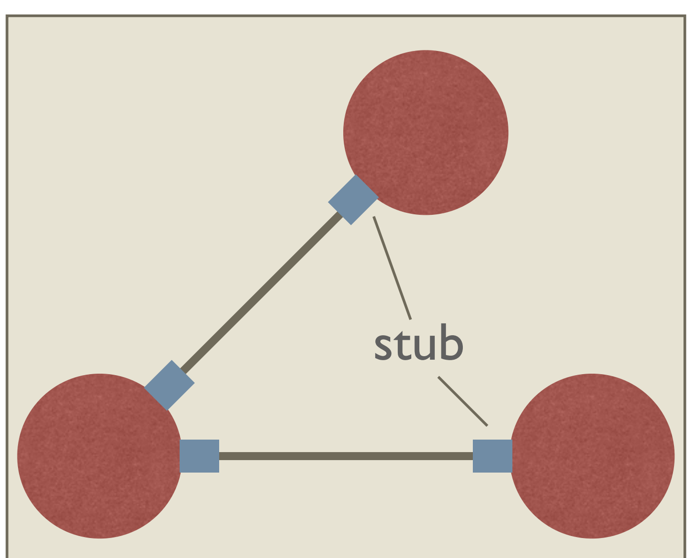
image: https://towardsdatascience.com/community-detection-algorithms-9bd8951e7dae

STRUCTURAL FEATURES: MODULARITY

We can quantify the extent to which communities in a network are segregated as follows.

If g_i represents the community that node i belongs to, then the total number of edges between nodes belonging to the same community is $\frac{1}{2}\sum_{ij} A_{ij}\,\delta_{g_i g_j}$

If the total number of links in the network is m, one can imagine $2m$ "stubs".



If all of the links were random (assuming a given number of stubs k_i per node i), then the expected number of links between stubs of nodes i and j is $k_i k_j / 2m$.

So the *expected* total number of links between nodes belonging to the same community is $\frac{1}{2}\sum_{ij} \frac{k_i k_j}{2m}\,\delta_{g_i g_j}$

STRUCTURAL FEATURES: MODULARITY

Thus, the (normalized) difference between the actual and expected number of edges between nodes of the same group in an undirected network is

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta_{g_i g_j}$$

Here, the quantity Q is referred to as the modularity. This gives us a measure of the *excess* number of links seen within groups than would be expected by chance.

For the case of a directed network, one can similarly derive:

$$Q = \frac{1}{m} \sum_{ij} \left(A_{ij} - \frac{k_i^{in} k_j^{out}}{m} \right) \delta_{g_i g_j}$$

where the factor of 2 is missing since the links are not double-counted.

STRUCTURAL FEATURES: MODULARITY

A number of approaches can be used to obtain the modular structure. One of the most common approaches involves modularity maximization.

For this, we assume that each node i is in one of two groups, and introduce the column vector $\mathbf{s}$ ($s_i = \pm 1$, depending on which group i belongs to). Then we have:

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\frac{s_i s_j + 1}{2} = \frac{1}{4m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)s_i s_j = \frac{1}{4m}\sum_{ij} B_{ij} s_i s_j = \frac{1}{4m}\mathbf{s}^T\mathbf{B}\mathbf{s}$$

Thus the problem can be reduced to the following :

Given the modularity matrix $\mathbf{B}$, find $\mathbf{s}$ such that Q is maximum.

A widely used method is based on the insight that the optimal $\mathbf{s}$ (were it not constrained to take values $s_i = \pm 1$) would be an eigenvector of $\mathbf{B}$ (Newman, 2006; Leicht & Newman, 2008).

Brain networks are modular

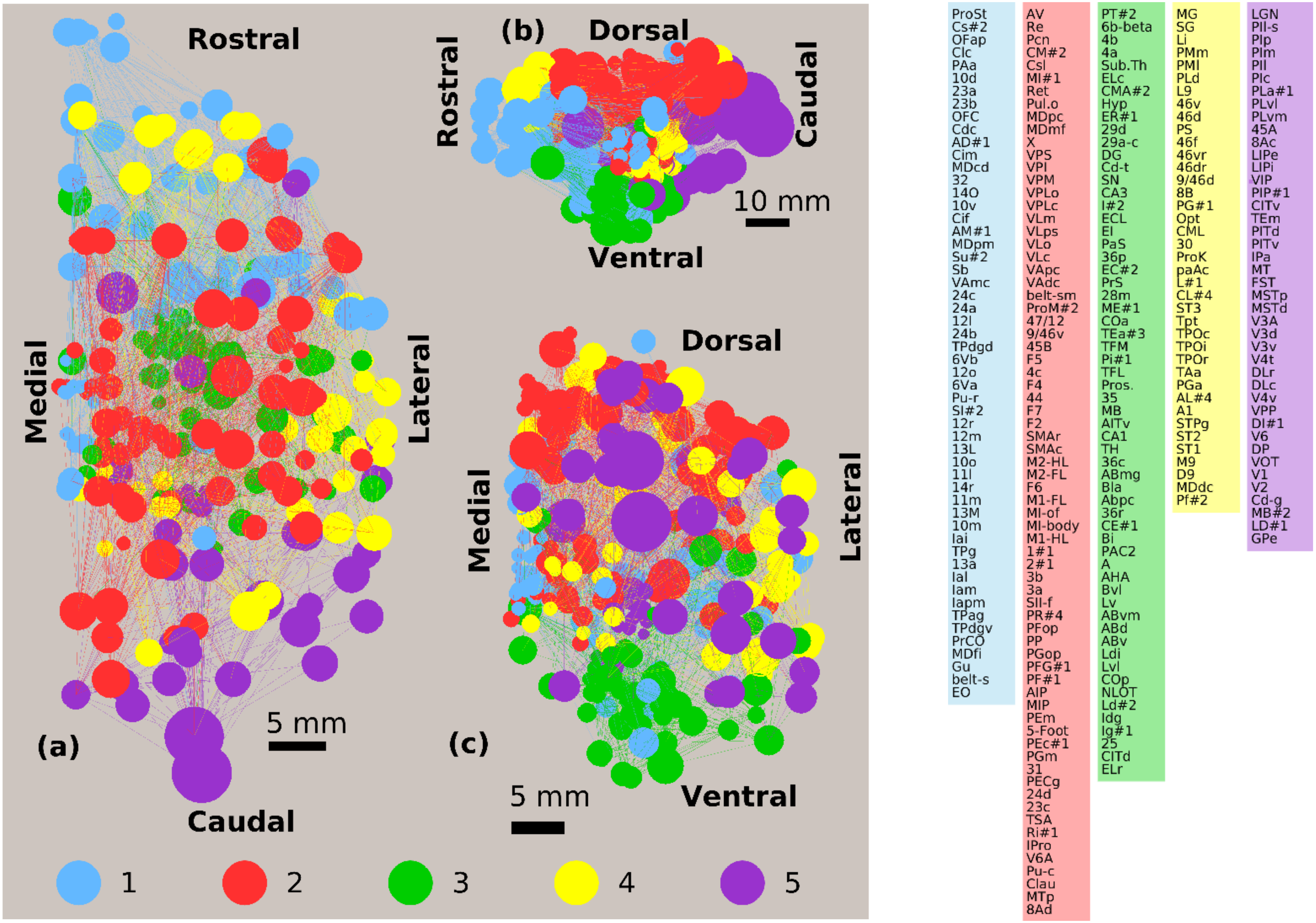


image: Pathak, A., Menon, S. N. and Sinha, S., *Phys. Rev. E* **106**, 054304 (2022).

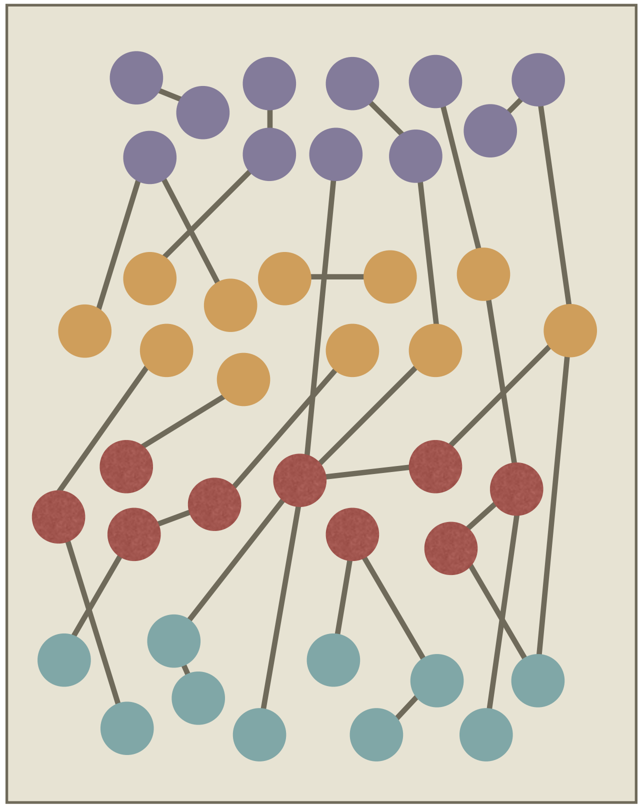
EXTRACTING STRUCTURE: HIERARCHY

A network is said to have a hierarchical structure if there exists "layers" of nodes, such that the density of connections between consecutive layers is higher than that within layers, or between non-consecutive layers.

Just as one can determine the extent to which a network is modular, it is also possible to define a hierarchy index (Pathak et al, 2024) as follows:

$$H = \frac{1}{m} \sum_{ij} \left(A_{ij} - \frac{k_i^{in} k_j^{out}}{m} \right) (\delta_{l_i, l_j+1} + \delta_{l_i+1, l_j})$$

Maximising H leads to the optimal hierarchical decomposition of the network.



reference: Pathak, A., Menon, S. N. and Sinha, S., *PNAS* **121**, e2314291121 (2024).

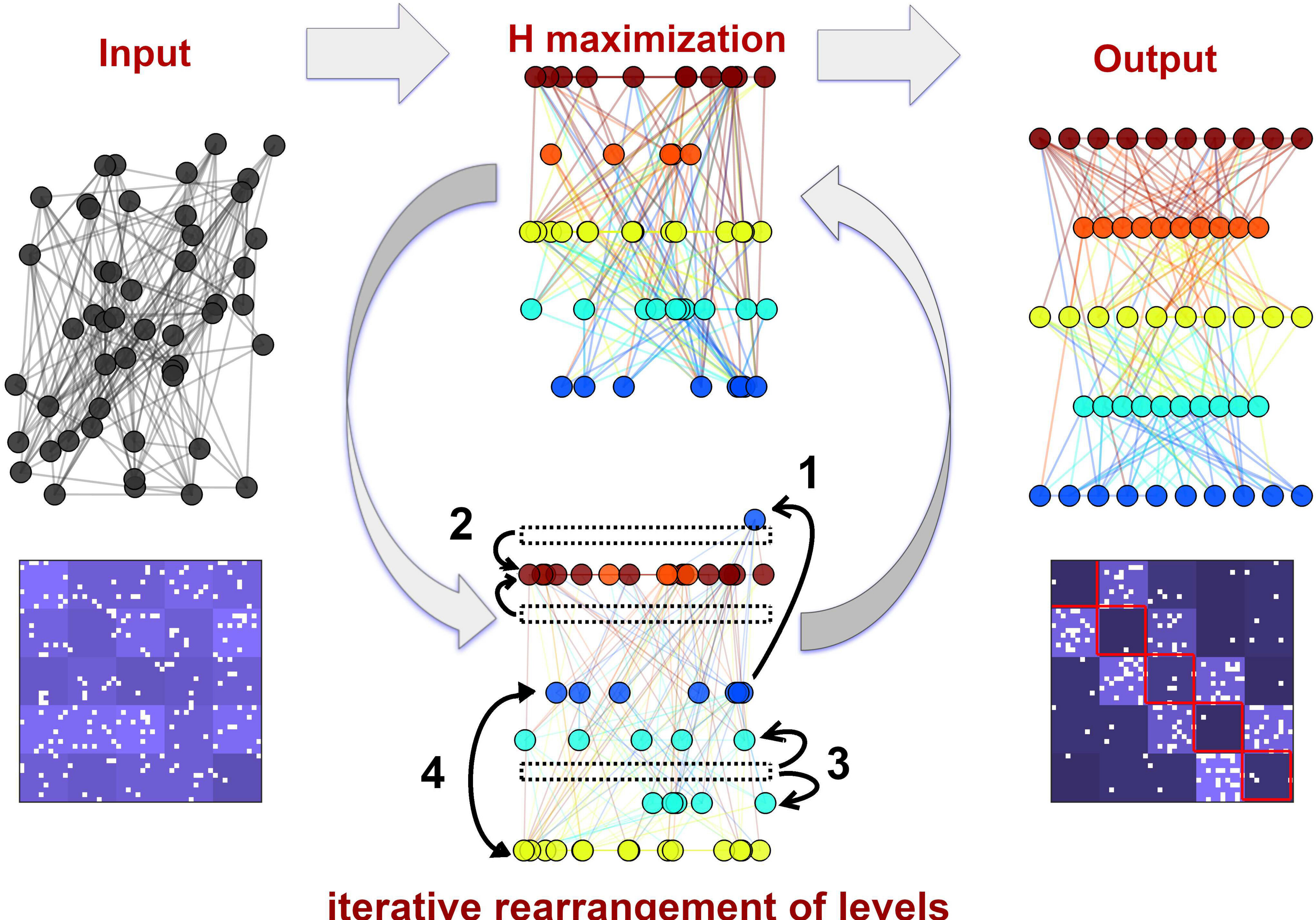


image: Pathak, A., Menon, S. N. and Sinha, S., *PNAS* **121**, e2314291121 (2024).

Brain networks exhibit "modular hierarchy"

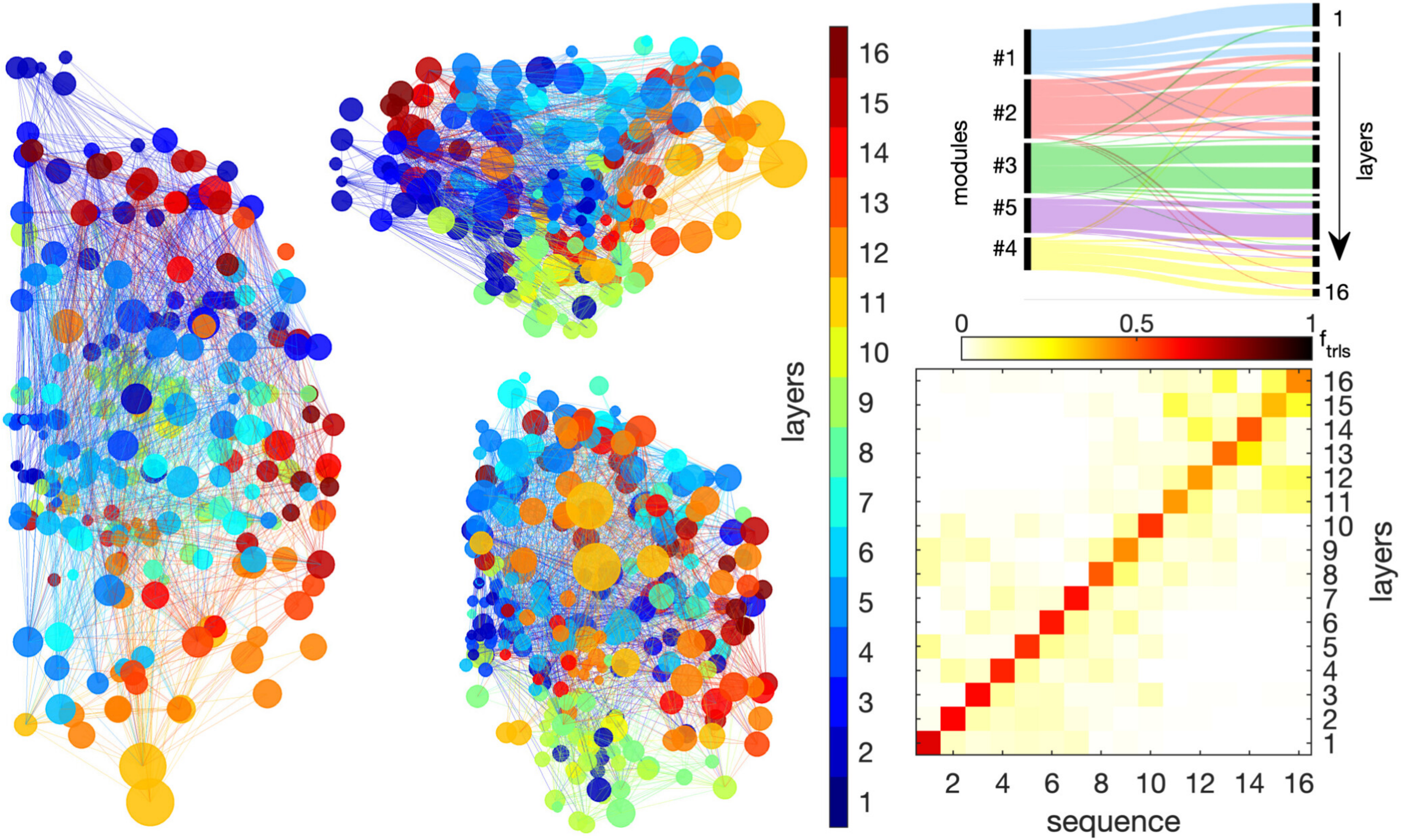


image: Pathak, A., Menon, S. N. and Sinha, S., *PNAS* **121**, e2314291121 (2024).

REFERENCES

ARTICLES:

- Erdős, P. & Rényi, A. *Publicationes Mathematicae* **6**, 290–297 (1959). [Random network model]
- Travers, J. & Milgram, S., *Sociometry* **32**, 425-443 (1969). [Milgram's experiment]
- Watts, D. & Strogatz, S., *Nature* **393**, 440-442 (1998). [Small-world construction]
- Barabási, A.-L. & Albert, R., *Science* **286**, 509–512 (1999). [Scale-free networks]
- Newman, M. E. J., *PNAS* **103**, 8577 (2006). [Spectral method for module detection]
- Leicht, E. A. & Newman, M. E. J., *Phys. Rev. E* **100**, 118703 (2008). [Module detection in directed networks]
- Pathak, A., Menon, S. N. & Sinha, S., *PNAS* **121**, e2314291121 (2024). [Hierarchy index]

BOOKS:

- Barabási, A.-L., "*Network Science*" (http://networksciencebook.com/)
- Newman, M., "*Networks*" (OUP, 2018).