Bangalore School on Statistical Physics XIV, Sept 2023

Statistical Mechanics of Complex Networks

Lecture 4: Meso (Core-Periphery, Modularity & Hierarchy)

Sitabhra Sinha

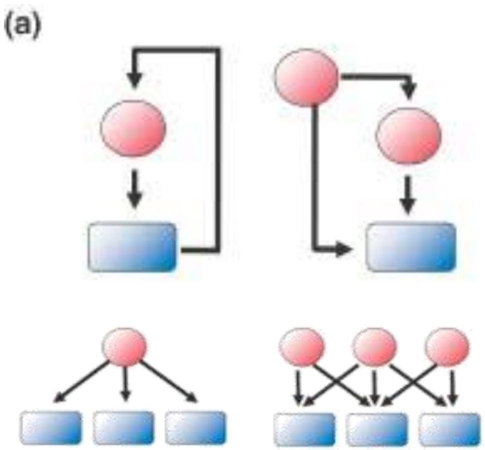
The Institute of Mathematical Sciences, Chennai

How to characterize properties of a network ?

Image: Kim & Park, WIREs Syst Biol & Med, 2010

Local properties

Micro

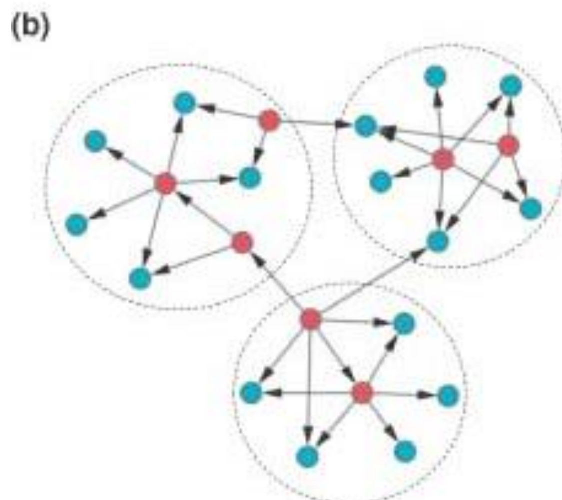


How cliquish is my neighbourhood ?

Clustering

Intermediate-scale structures

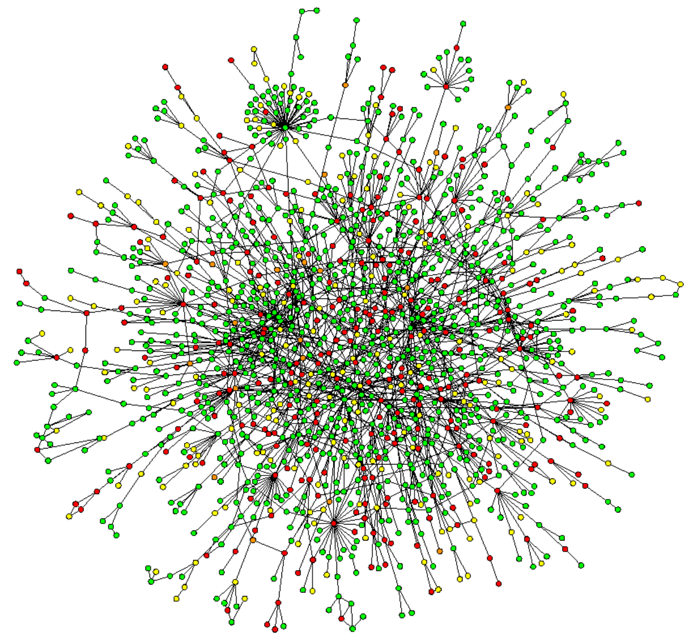
Meso



?

Global properties

Macro

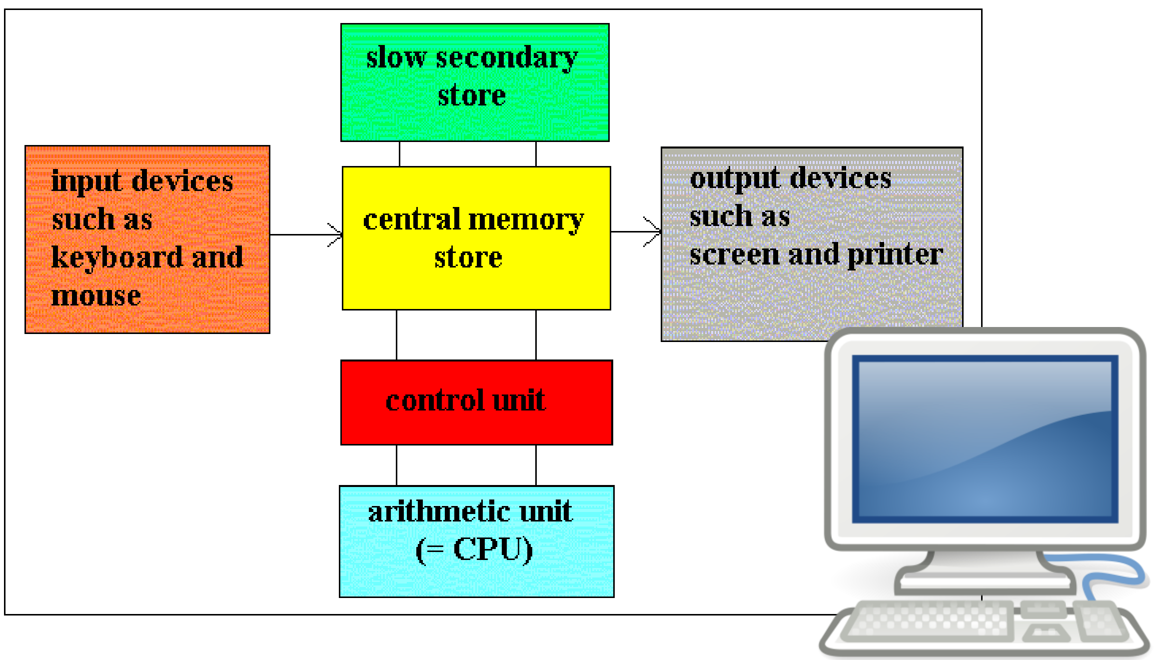


How fast can I travel to the farthest point of my network ?

Path Length

Why meso-scale analysis ?

"Coarse-grain" the complex system so as not to deal with the elementary components but clusters that appear to contribute to a specific function



…analogous to block models in engineering

Degree: from distribution to correlations

Random networks – e.g., having scale-free degree distribution - do not exhibit any correlations between the degrees of connected nodes

The probability a link connecting nodes of degrees k & k' is

$P(k,k') = k\,P(k)\,k'\,P(k') / \langle k \rangle^2$ (degree-uncorrelated network)

Most real-life networks exhibit degree correlations

For example,
high-degree nodes may prefer to connect to other high-degree nodes
(popular folk tend to hang out with each other!)

In other cases, high-degree nodes may preferentially connect to low-degree nodes

Is there a quantitative metric for such correlations ?

Assortativity

When individuals prefer to associate with "similar" individuals

⇒ *Dis*assortativity: when "dissimilar" individuals prefer to associate with each other

"Similarity" of nodes in a network may be in terms of any intrinsic characteristics, including their degree (number of connections)

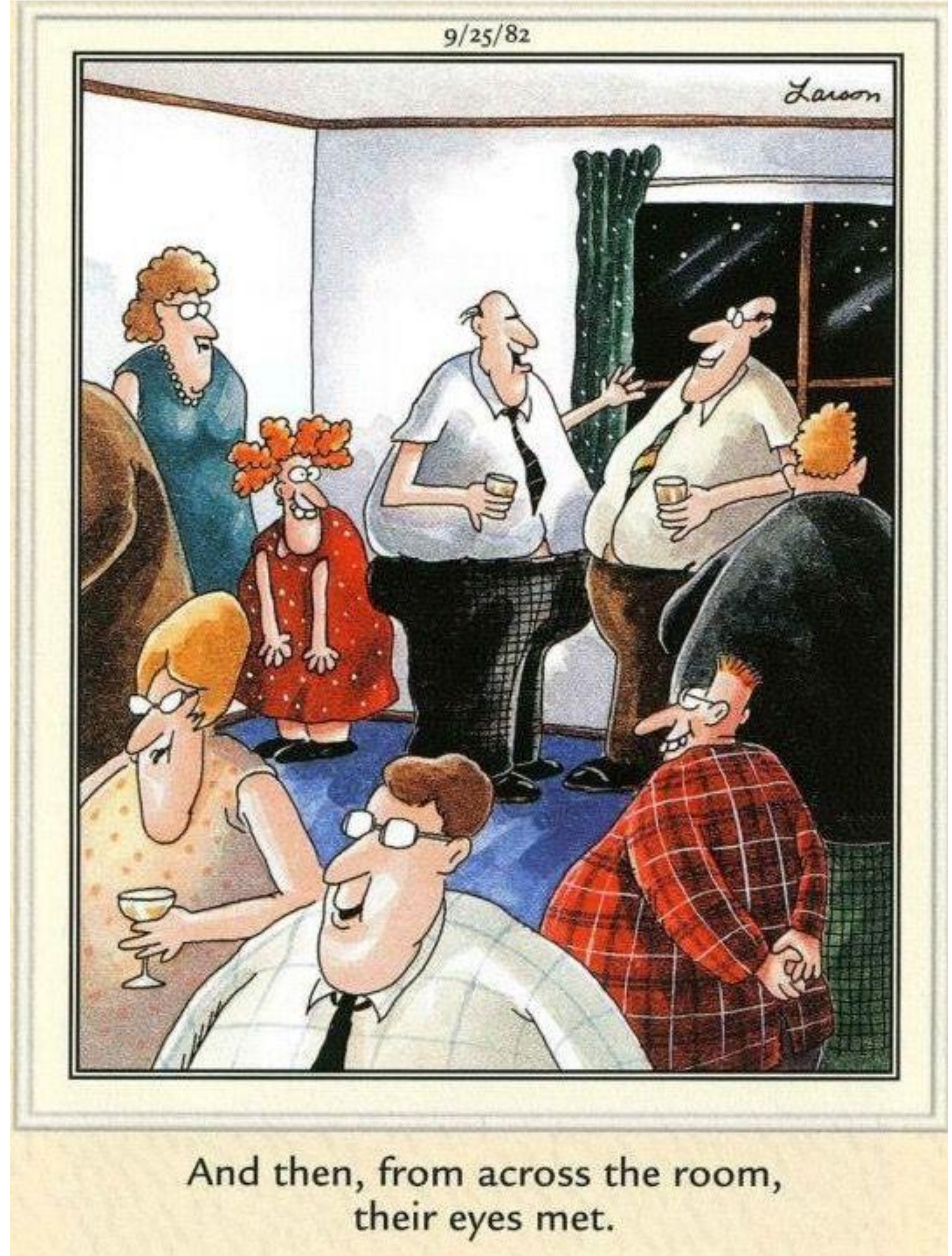


Image: The Far Side by Gary Larson

Degree assortativity

Correlations in degree of connected nodes is measured by

(Newman, 2002)

$$\text{Assortativity, r} = \frac{\frac{1}{L}(\sum_{i=1}^{L} j_i k_i) - (\frac{1}{L}\sum_{i=1}^{L}\frac{1}{2}(j_i + k_i))^2}{\frac{1}{L}(\sum_{i=1}^{L}\frac{1}{2}(j_i^2 + k_i^2) - (\frac{1}{L}(\sum_{i=1}^{L}\frac{1}{2}(j_i + k_i))^2}$$

j_i, k_i: degrees of vertices at ends of the i-th edge   L: total number of links

r < 0: disassortative mixing

Unlike preferred

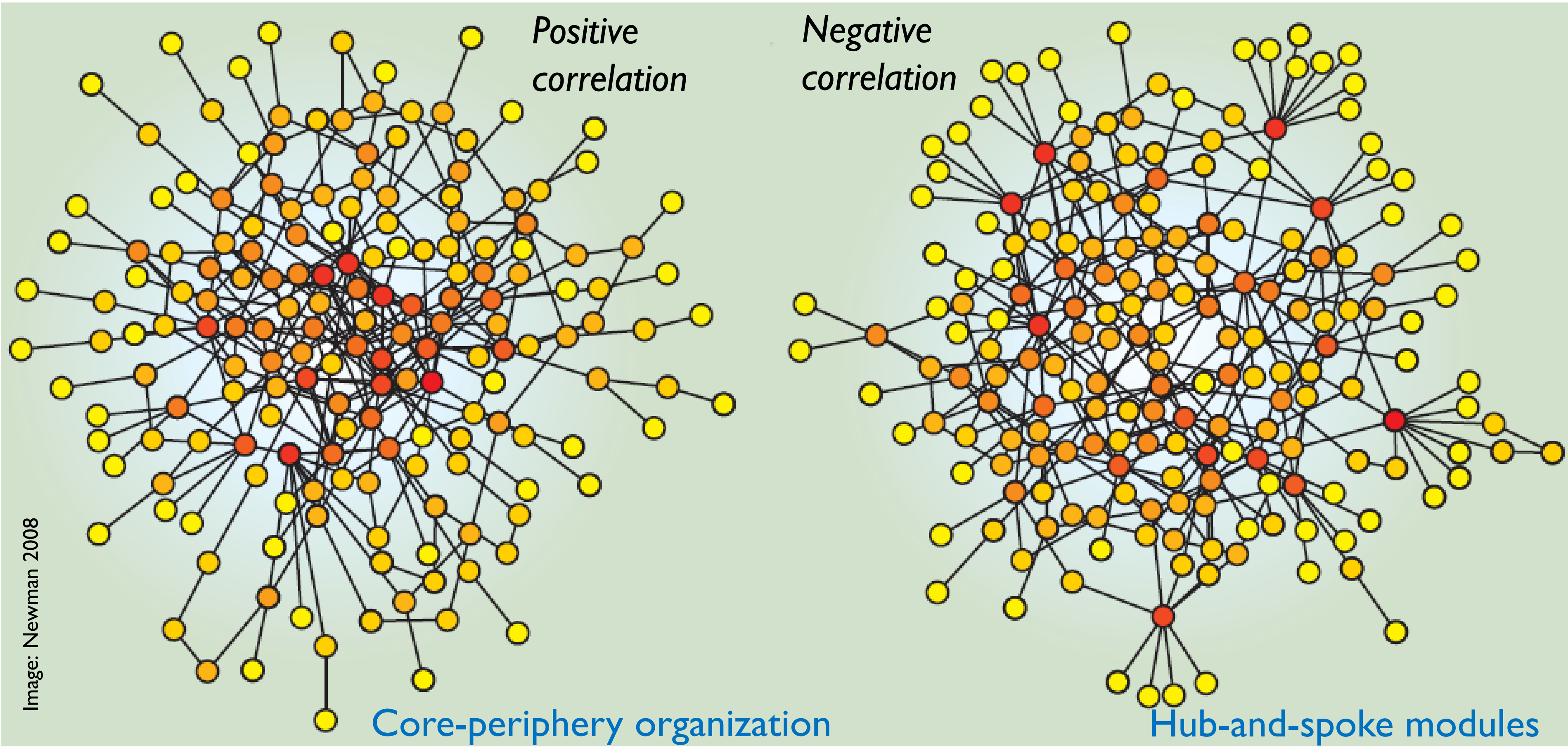
Nodes of high degree mostly have nearest neighbors of low degree
E.g., most biological & technological networks

r > 0: assortative mixing

Like prefers like

Nodes of high degree mostly have nearest neighbors of high degree
E.g., social networks

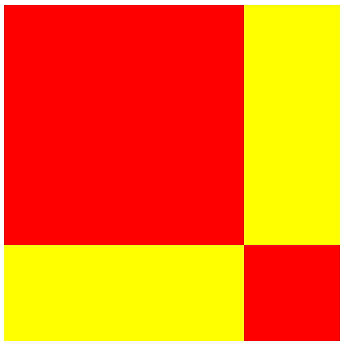
Nodes can prefer to connect to nodes with similar or dissimilar connectivity



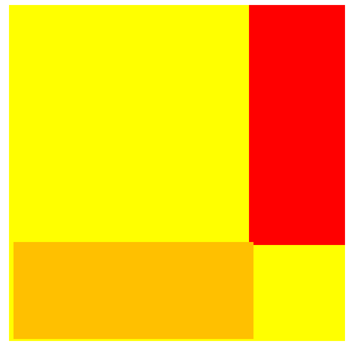
Two networks may have the same degree distribution but different connectivity patterns overall because high-degree nodes may prefer to connect to other high-degree nodes (positive degree correlation) or may want to avoid them (negative degree correlation)

Mesoscopic organization of networks

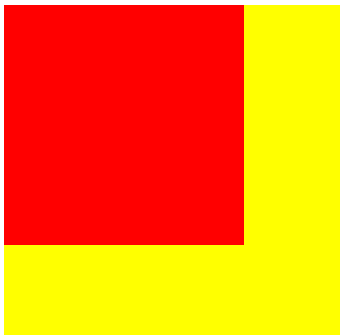
Dense connections   Sparse connections



Modular



Hierarchical



Core-periphery

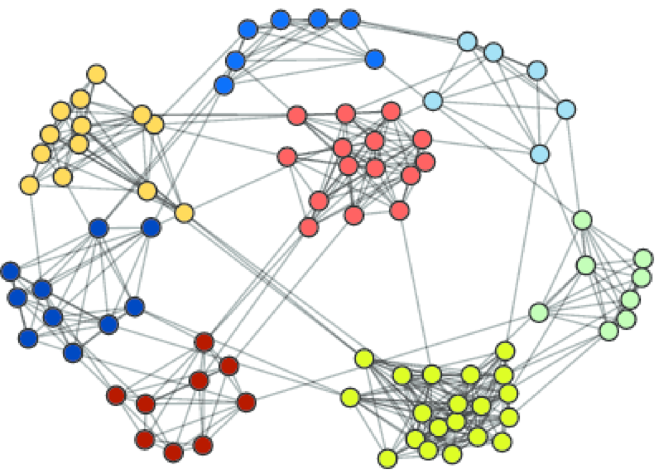


Image: Y Ma et al, EPL 108, 48004 (2014)

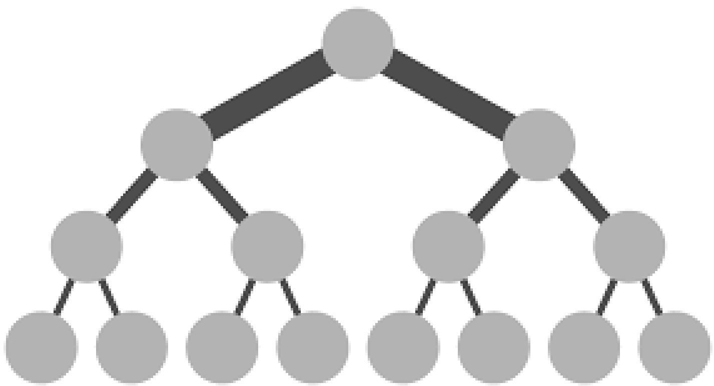


Image: Wikipedia

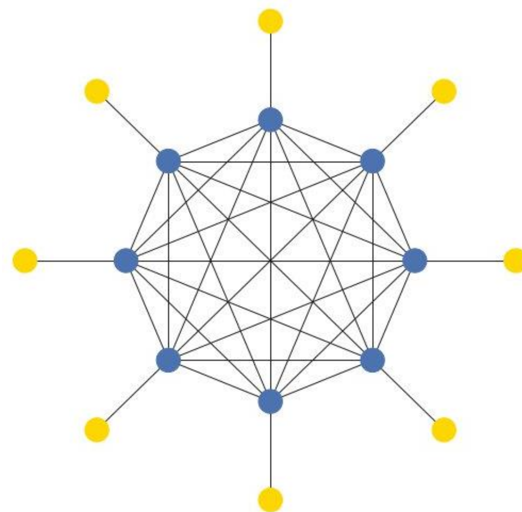
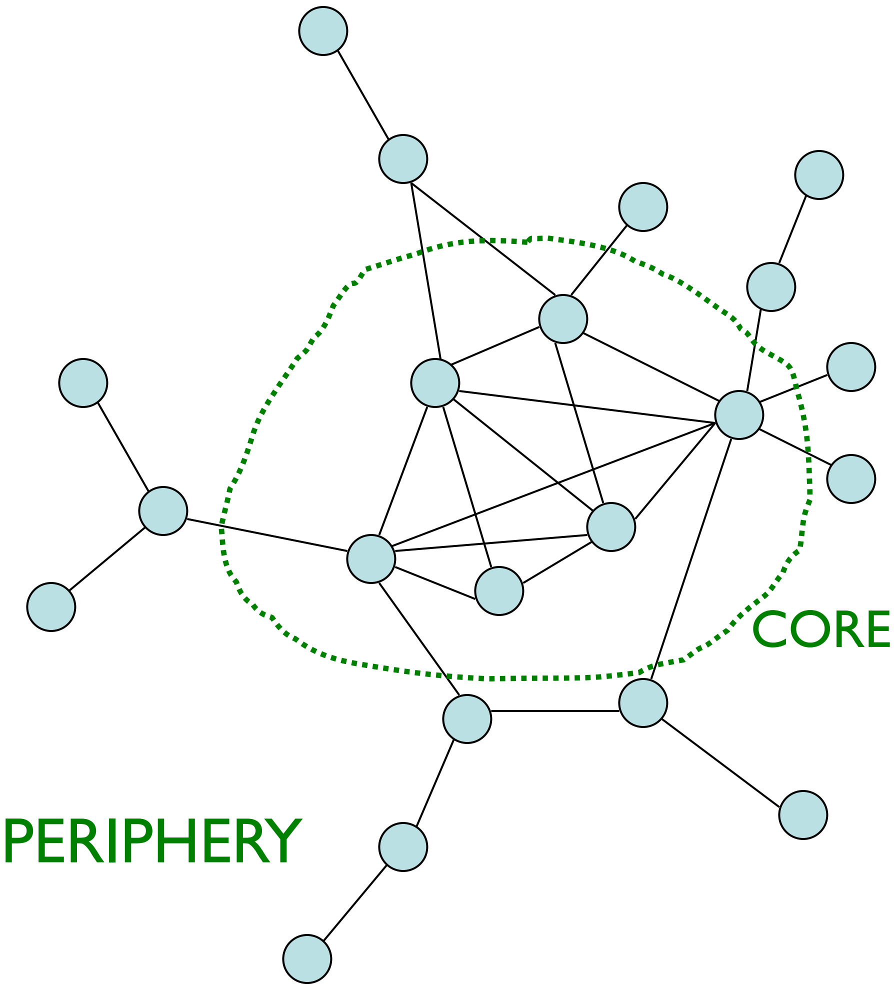


Image: Quora

Also “Rich Club”

Core-periphery organization



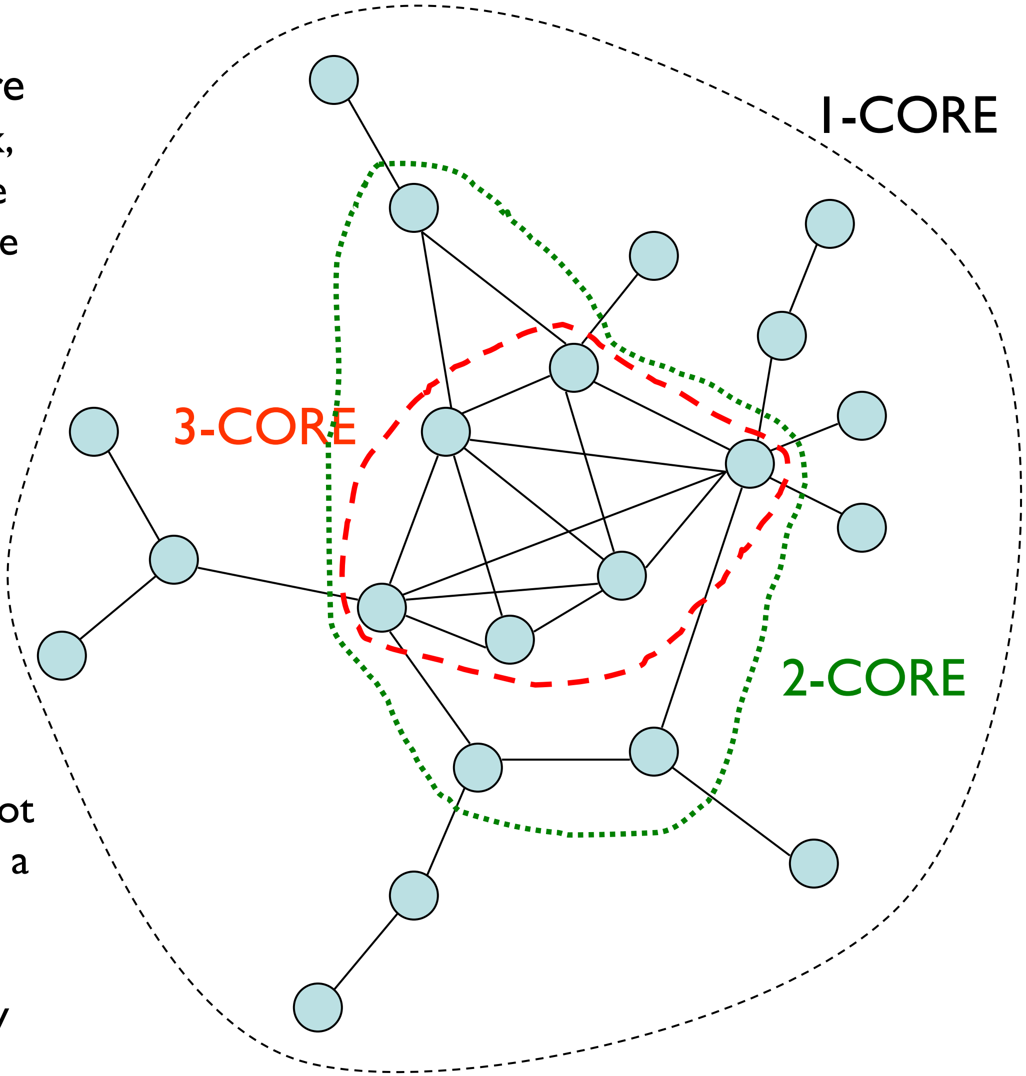
The *k-core* of a network is the subnetwork containing all nodes that have degree *at least* equal to *k*

k-Core Decomposition

Seidman (1983)

iterative process for determining *k*-core

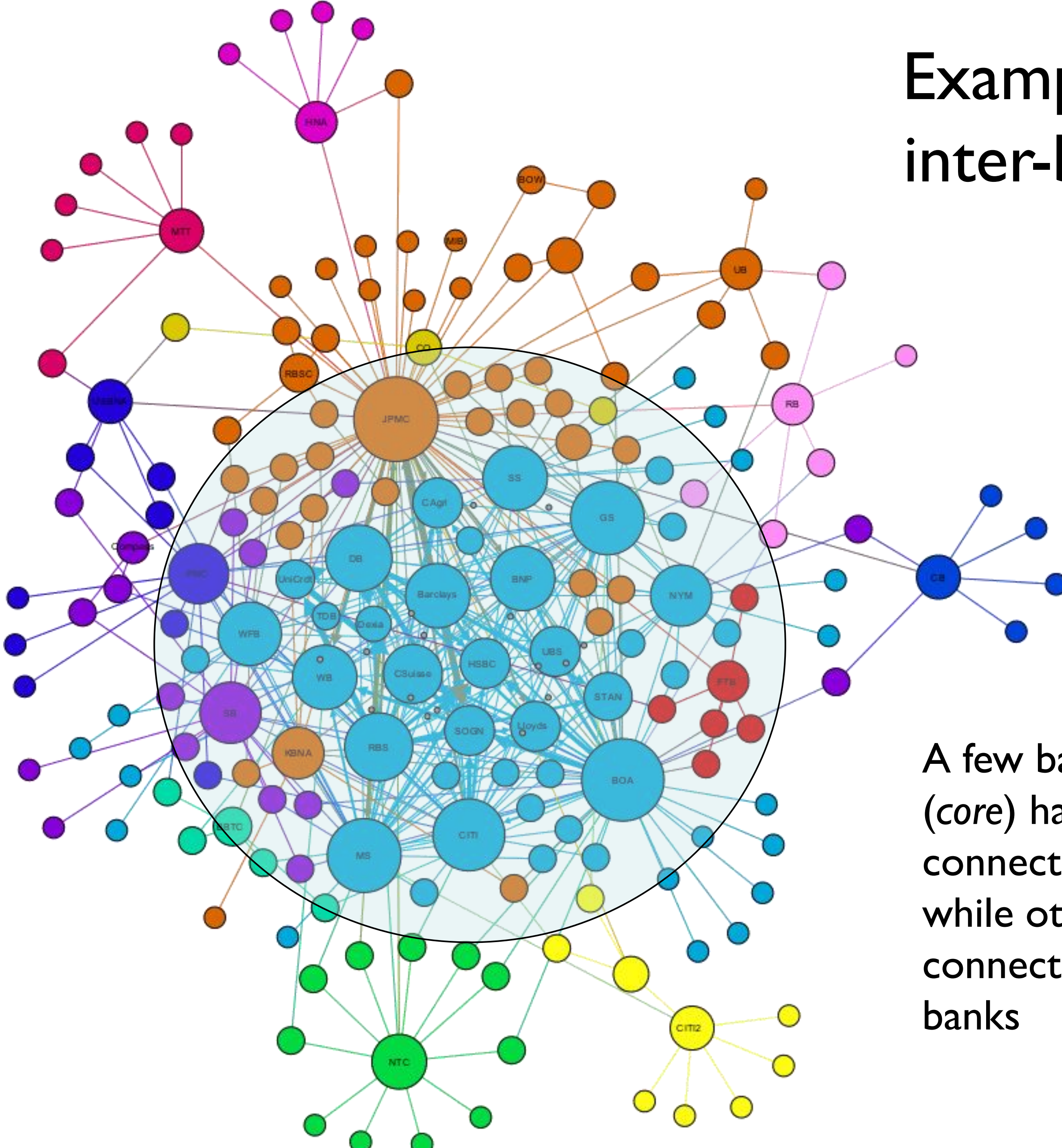
(i) remove all nodes having degree < *k*,
(ii) check the resulting network to see if any of the remaining nodes now have degree < *k* as a result of (i), and if so
(iii) repeat steps (i)-(ii) until all remaining nodes have degree at least equal to *k*.



Example: 2-core
obtained by eliminating all nodes that do not form part of a loop (a closed path through a subset of the connected nodes)

⇒ There exist at least *k* paths between any pair of nodes belonging to *k*-core.

Example: Network of inter-bank borrowing



A few banks having high capital (*core*) have many strong connections with each other, while other banks (*periphery*) connect to one or few of these banks

Brokering between core and periphery

Collaboration network in the Hungarian film industry

2006 collaboration network of movie creators
(actors, directors, cinematographers, editors and writers)

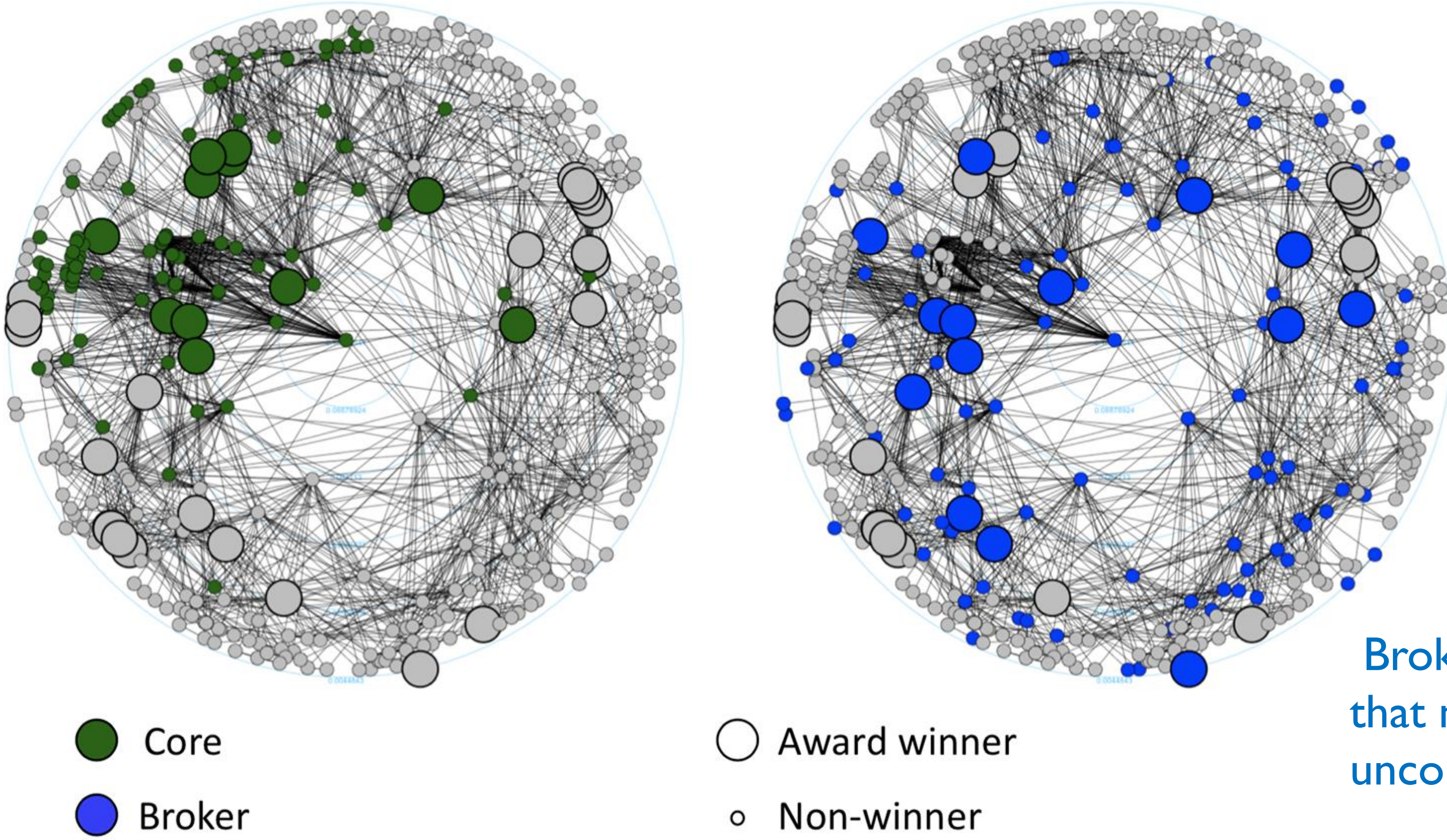
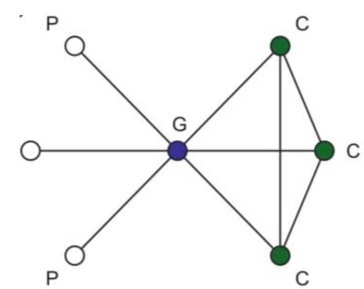


Image: Sándor Juhász et al, *PLoS ONE* **15** (2020) e0229436



Broker nodes link nodes that may be otherwise unconnected

Being in the core of the collaboration network and bridging the core to the periphery significantly helps creative success (measured by awards won)

Modules

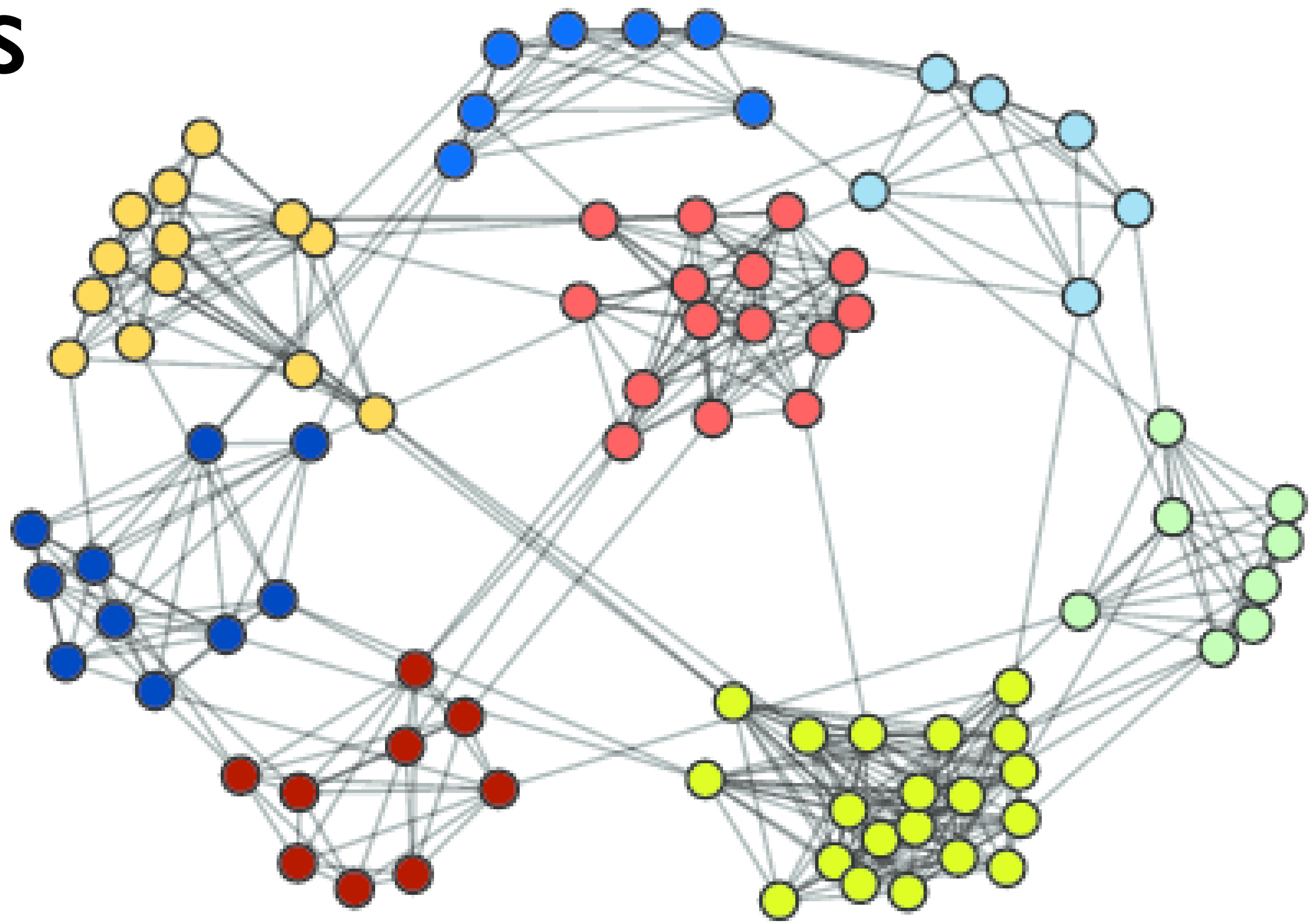


Image: Yifang Ma et al, EPL 108 (2014) 48004

Modular Networks

dense connections *within* certain sub-networks (modules or communities) & relatively few connections *between* modules

Ubiquity of modular networks

Modules in biological networks are often associated with specific functions

Metabolic network of *E coli*

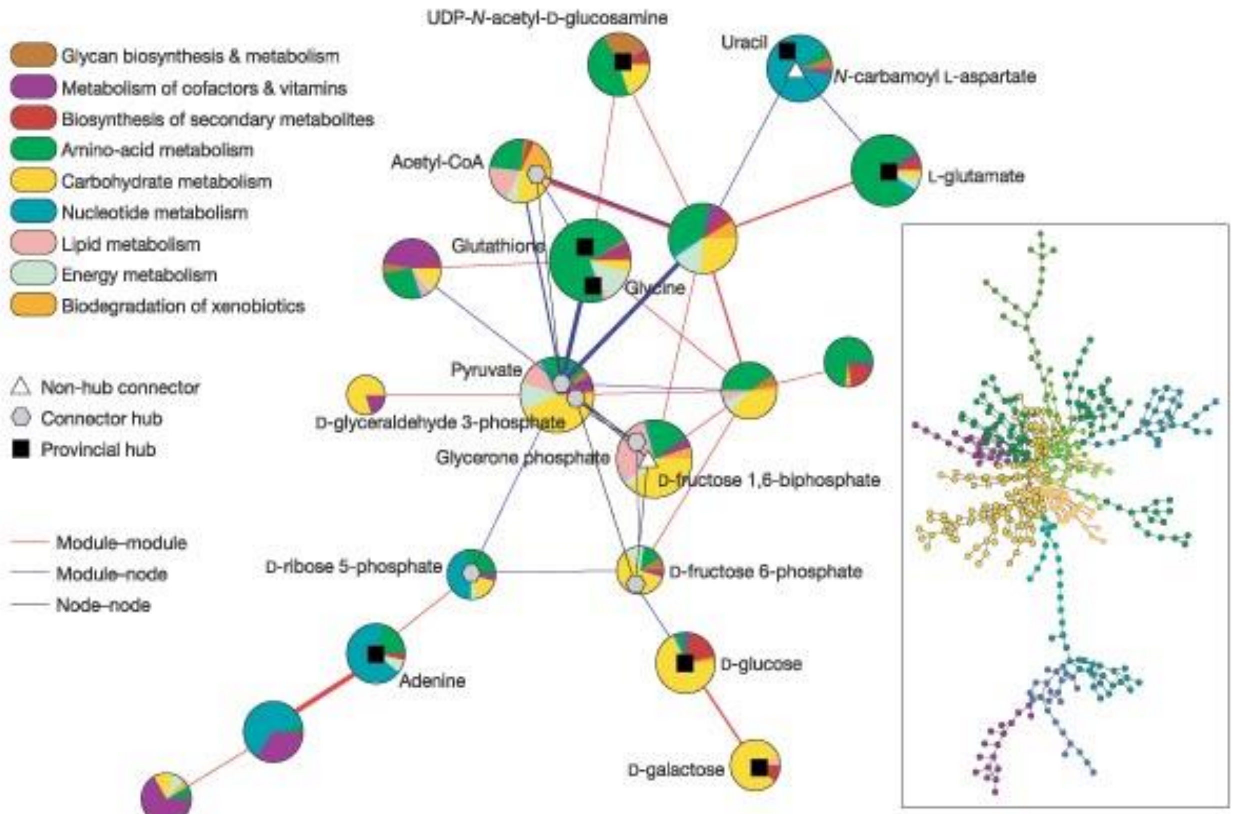
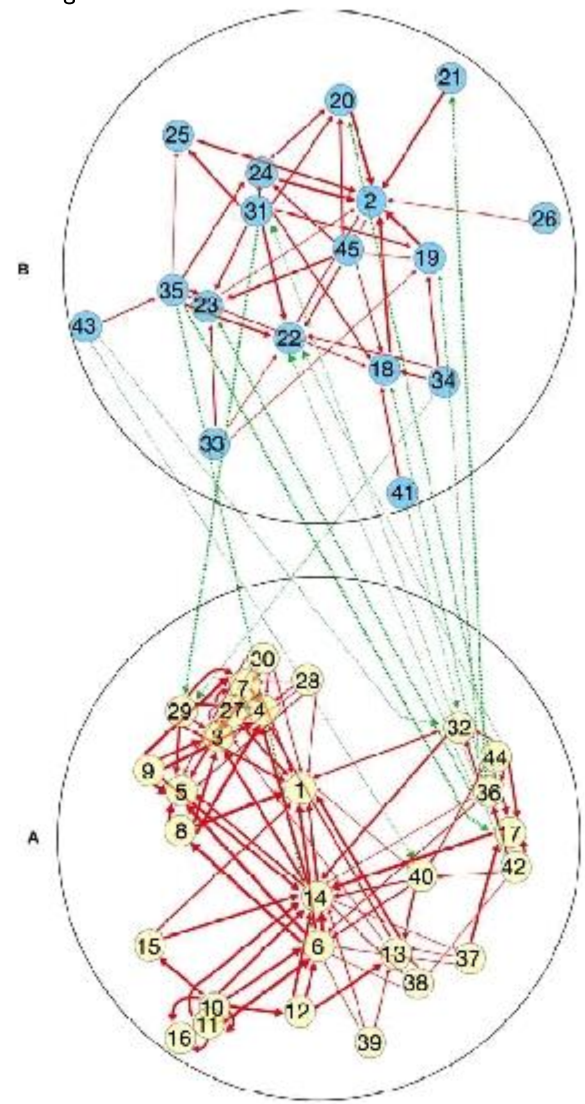


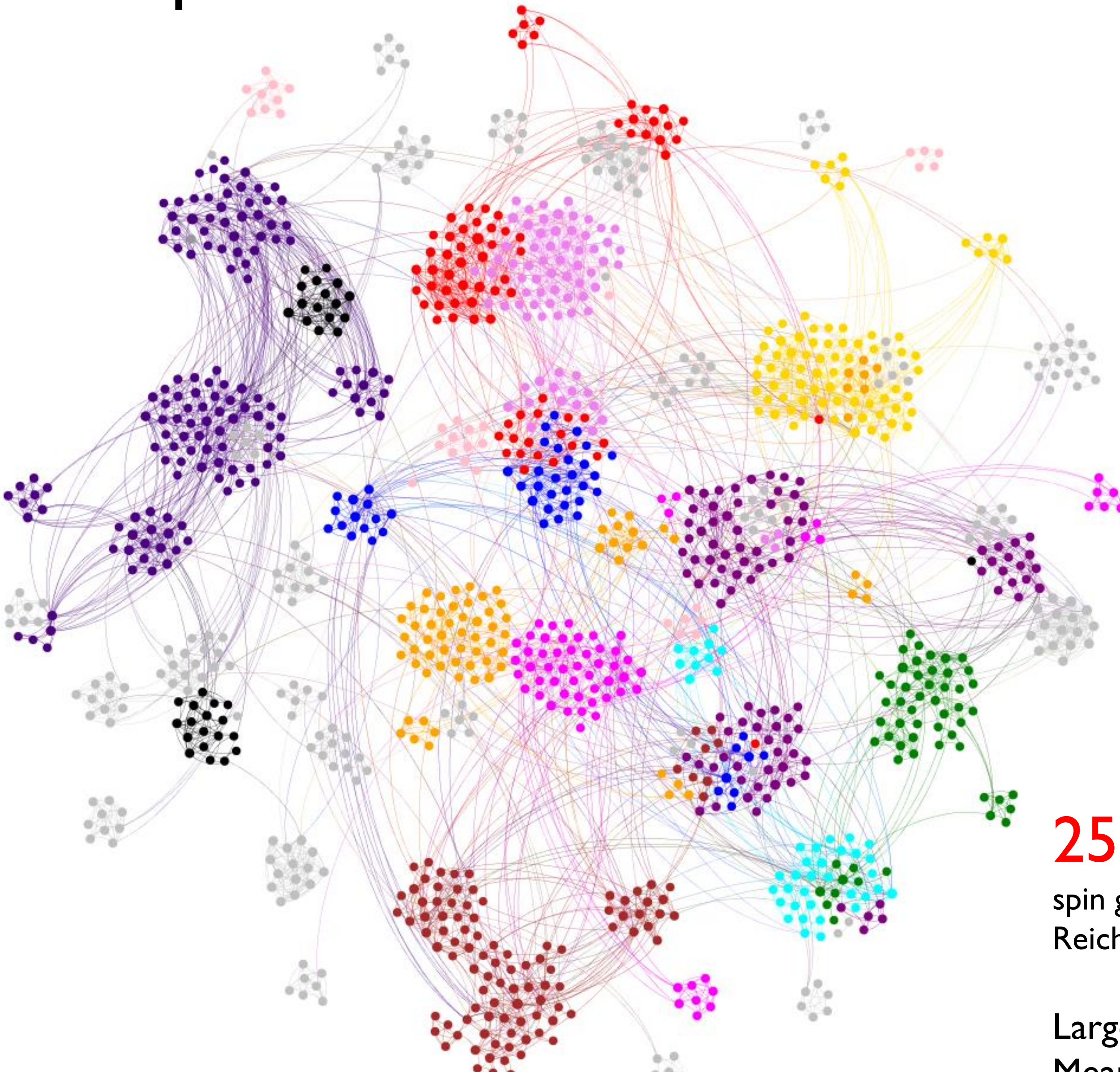
Image: Guimera and Amaral, *Nature* (2005)

Chesapeake Bay foodweb

Image: Ulanowicz et al



Example: Social Network of a Karnataka Village



Data: Bharatha Swamukti Samsthe microfinance institution

Described in A Banerjee et al, *Science* (2013)

Nodes: Individuals

Links: Social relations

Village "55"

Population: 1180 individuals

Largest connected component: 1151 individuals

25 modules

spin glass simulated annealing method

Reichardt & Bornholdt, PRE (2006)

Largest module: 127 nodes

Mean module size: 47 nodes

Node colors represent the community to which they belong

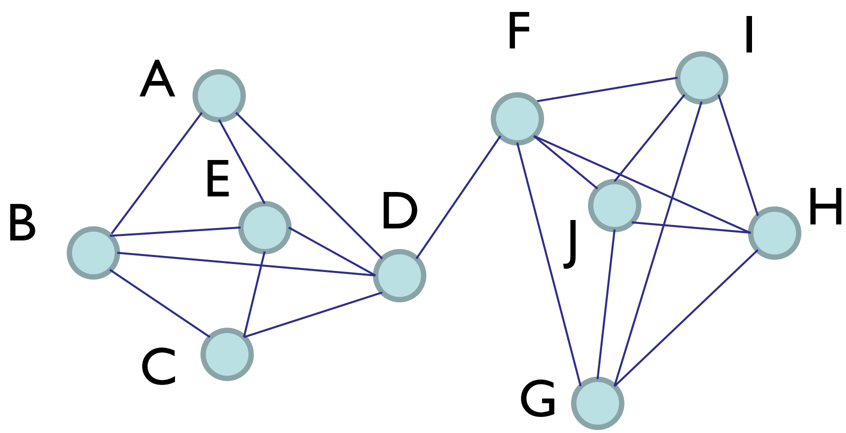
Problem:

Given a network,
how do we find the modules (communities)
into which it can be divided ?

Community Detection in Networks

Also referred to as Graph Partitioning or Module Determination

How to divide the nodes of a network into several groups such that nodes in each group are densely or strongly inter-connected



Example:
visually obvious that node clusters I: {A,B,C,D,E} and II: {F,G,H,I,J} constitute two separate groups highly intra-connected but only a single link connecting the two groups

The corresponding adjacency matrix will have an almost block- diagonal form – the two blocks corresponding to node clusters I & II

However for large networks the modular character may not be visually apparent – and adjacency matrices need to be partitioned

Graph partitioning

A classic problem in computer science from 1960s

How to divide the nodes of a network into a given number of non-overlapping groups of given sizes such that the number of edges between groups is minimized ?

A generalization of this problem,
How to divide the nodes into several groups such that most links are within groups and few links are between groups

referred to as

Community detection

How we define "most" and "few" can vary from one algorithm to another

Measuring modularity

How to quantify the degree of modularity for a given partitioning of a network into communities ?

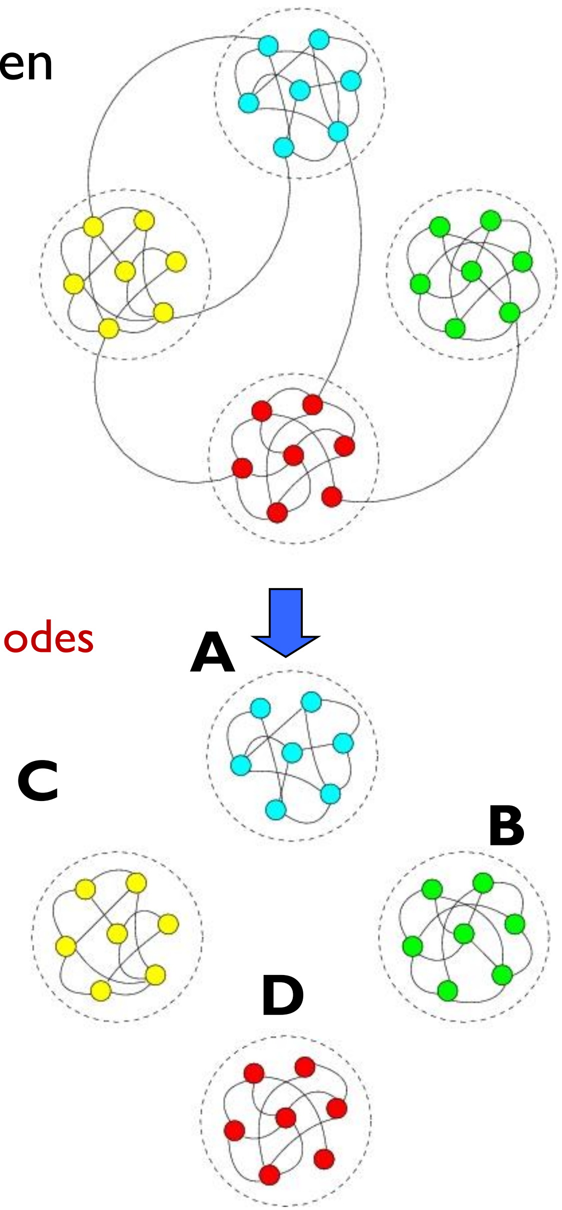
Modularity index (Newman, EPJB, 2004)

$$Q \equiv \frac{1}{2L}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2L}\right]\delta_{c_i c_j}$$

=1 if nodes are in same community

probability of an edge betn 2 nodes proportional to their degrees

A: Adjacency matrix
L : Total number of links
k_i : degree of *i*-th node
c_i : label of module to which *i*-th node belongs



For a random network, Q = 0
i.e., the connection density within a module is no different from that anywhere else in the network

Community detection by maximizing Q

For directed & weighted networks:

$$Q^W \equiv \frac{1}{L^W}\sum_{i,j}\left[W_{ij} - \frac{s_i^{\text{in}} s_j^{\text{out}}}{L^W}\right]\delta_{c_i c_j} \qquad (L^W = \sum_{i,j} W_{ij})$$

W: Weight matrix

s_i : strength of *i*-th node

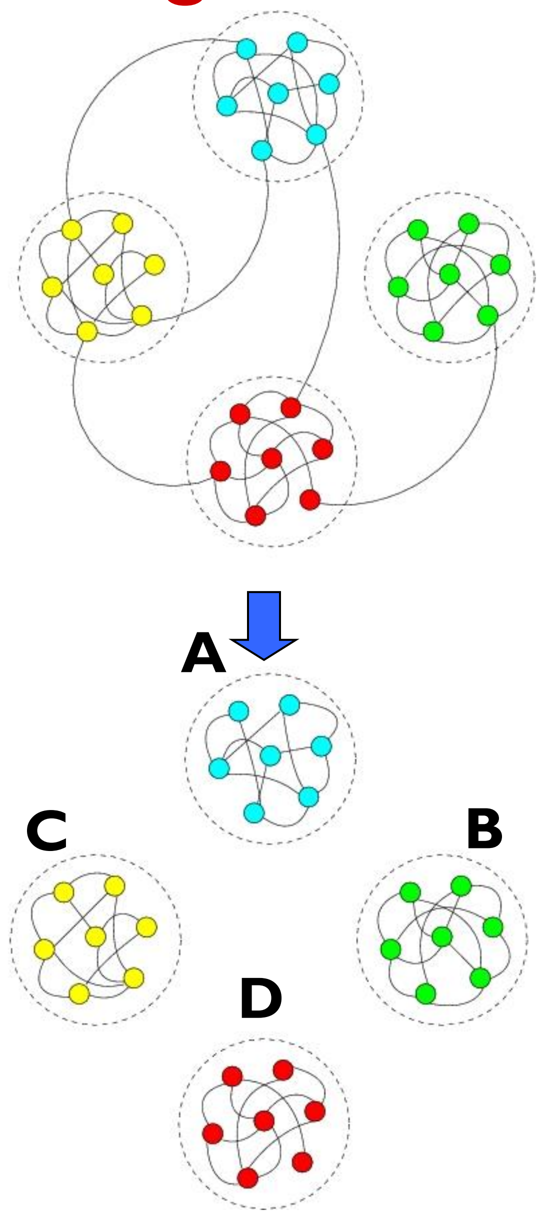
Modules determined through generalization of spectral method (Leicht & Newman, 2008)

Calculate eigenvector corresponding to largest positive eigenvalue of symmetrized modularity matrix $\mathbf{B} + \mathbf{B}^T$ where

$$B_{ij} = W_{ij} - [s_i^{\text{in}} s_j^{\text{out}} / L^W]$$

and then assign communities based on the signs of the elements of the eigenvector.

Can be refined by using your favorite combinatorial optimization routine, e.g., simulated annealing



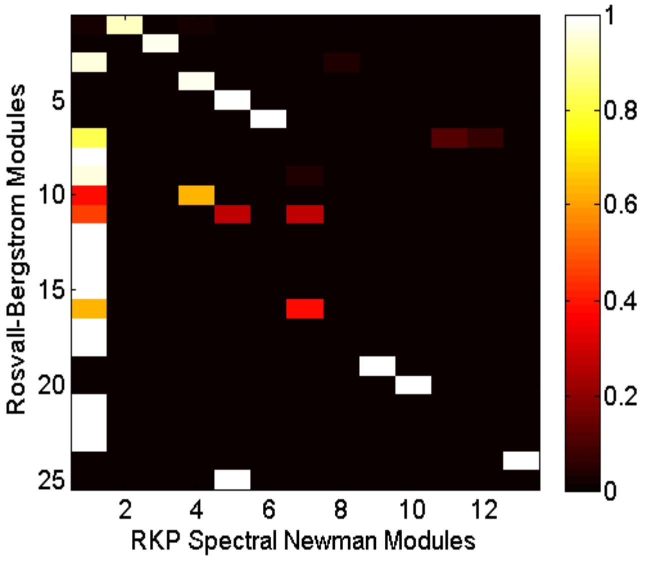
Information Flow Method for Community Detection

M. Rosvall and C. T. Bergstrom, PNAS, 2007
M. Rosvall and C. T. Bergstrom, PNAS, 2008

Fundamental idea: Optimal compression of network topology using the regularities in its structure such as modules

http://www.tp.umu.se/~rosvall/code.html

"… the Infomap method … is the best performing on the set of benchmarks we have examined here."
Lancichinetti and Fortunato, Phys Rev E 80, 056117 (2009)



In general spectral method yields lower number of detected modules compared to Rosvall-Bergstrom method, … but **high degree of overlap between modules identified by the two techniques**

"…modularity optimization may fail to identify modules smaller than a scale which depends on the total size of the network and on the degree of interconnectedness of the modules, even in cases where modules are unambiguously defined." *Fortunato and Barthelemy, PNAS 104, 36 (2007)*

Example: Macaque social network

UAS-GKVK Campus, Bangalore
Data: Anindya Sinha (NIAS, B'lore)
Analysis: Raj K Pan & Sumithra Surendralal (IMSc)

The Bonnet Macaque (*Macaca radiata*) seen widely in southern India

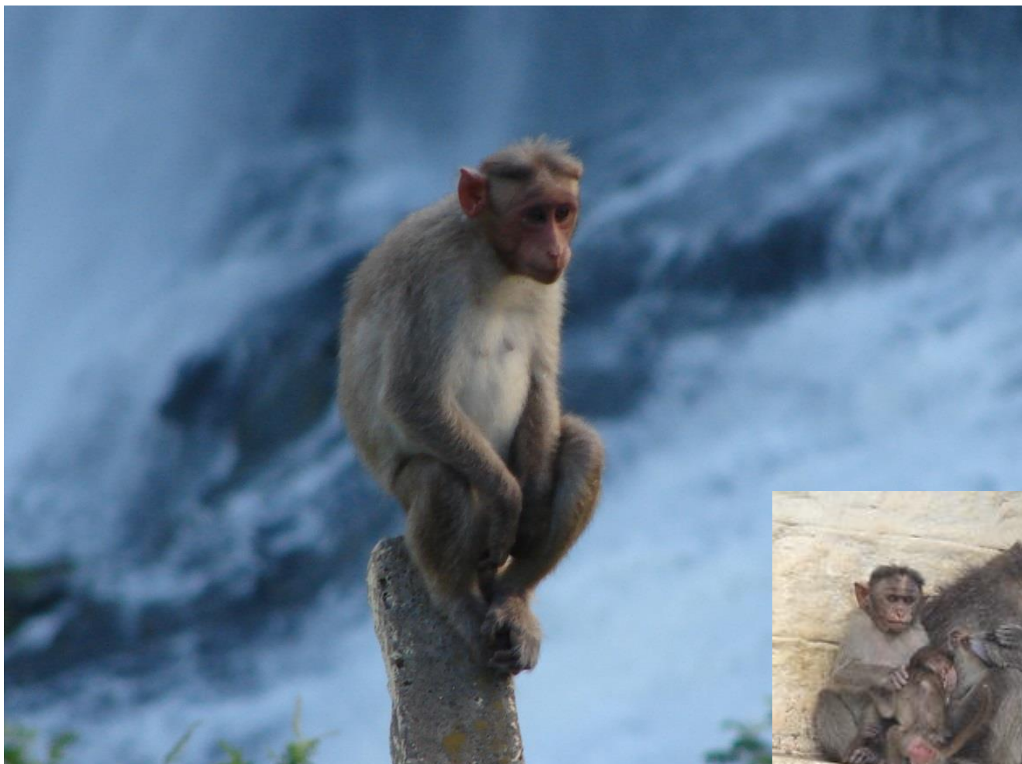


Image: Arunkumar

Usually live in large (~ 40) multi-male, multi-female troops where the adult individuals (~ 10) develop strong affiliative relationships

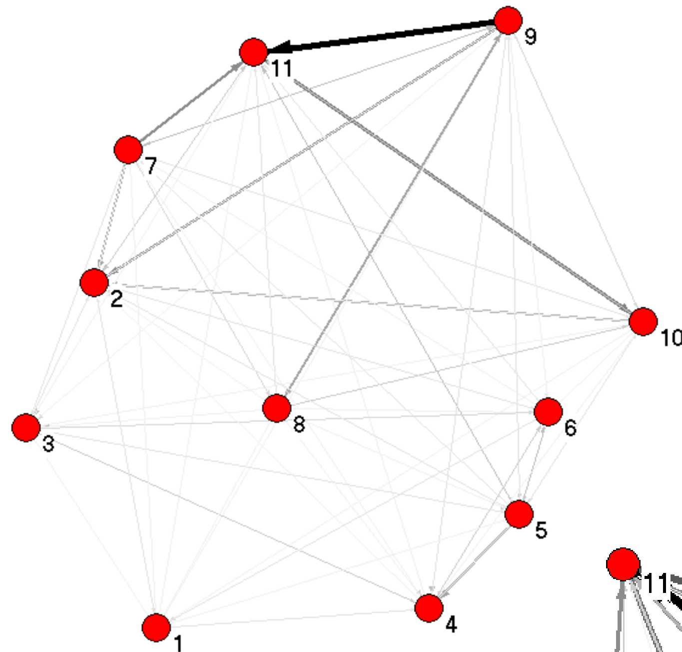
Image: Ramki (www.wildventures.com)

Macaque Social Networks can be defined in terms of

- grooming frequency
- total grooming time
- approach frequency

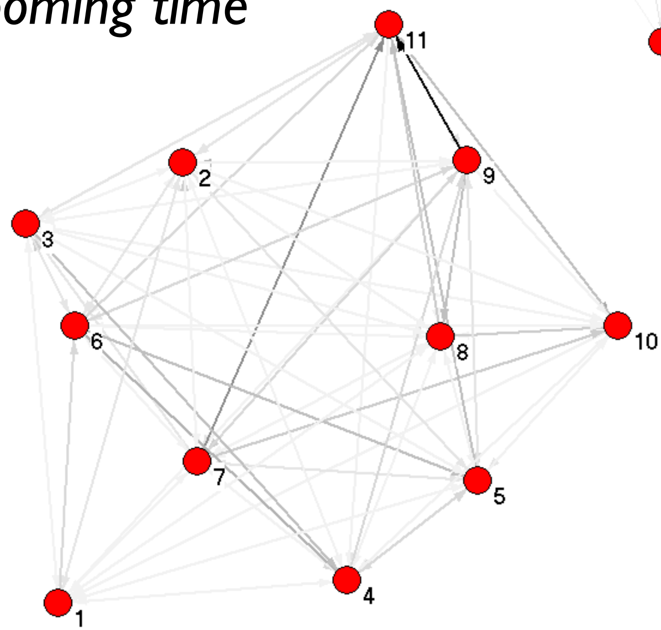
Data: 1993-1997

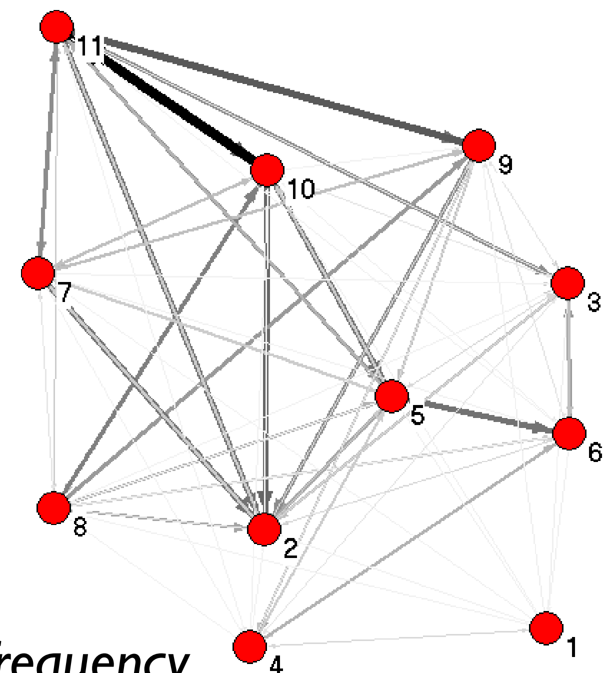
Numbers refer to rank among the adult females from 11 (most dominant) to 1 (least dominant)



grooming frequency

grooming time





approach frequency

Network analysis can predict group dynamics !

Female bonnet macaques
•usually remain in the group throughout their life
•as adults, form strong linear matrilineal dominance hierarchies that are **stable** over time

Male bonnet macaques
•as adults, form **unstable** dominance hierarchies
•occupy low ranks when young, high when mature and at peak of health

Community detection generates consistent partitions for females, not for males

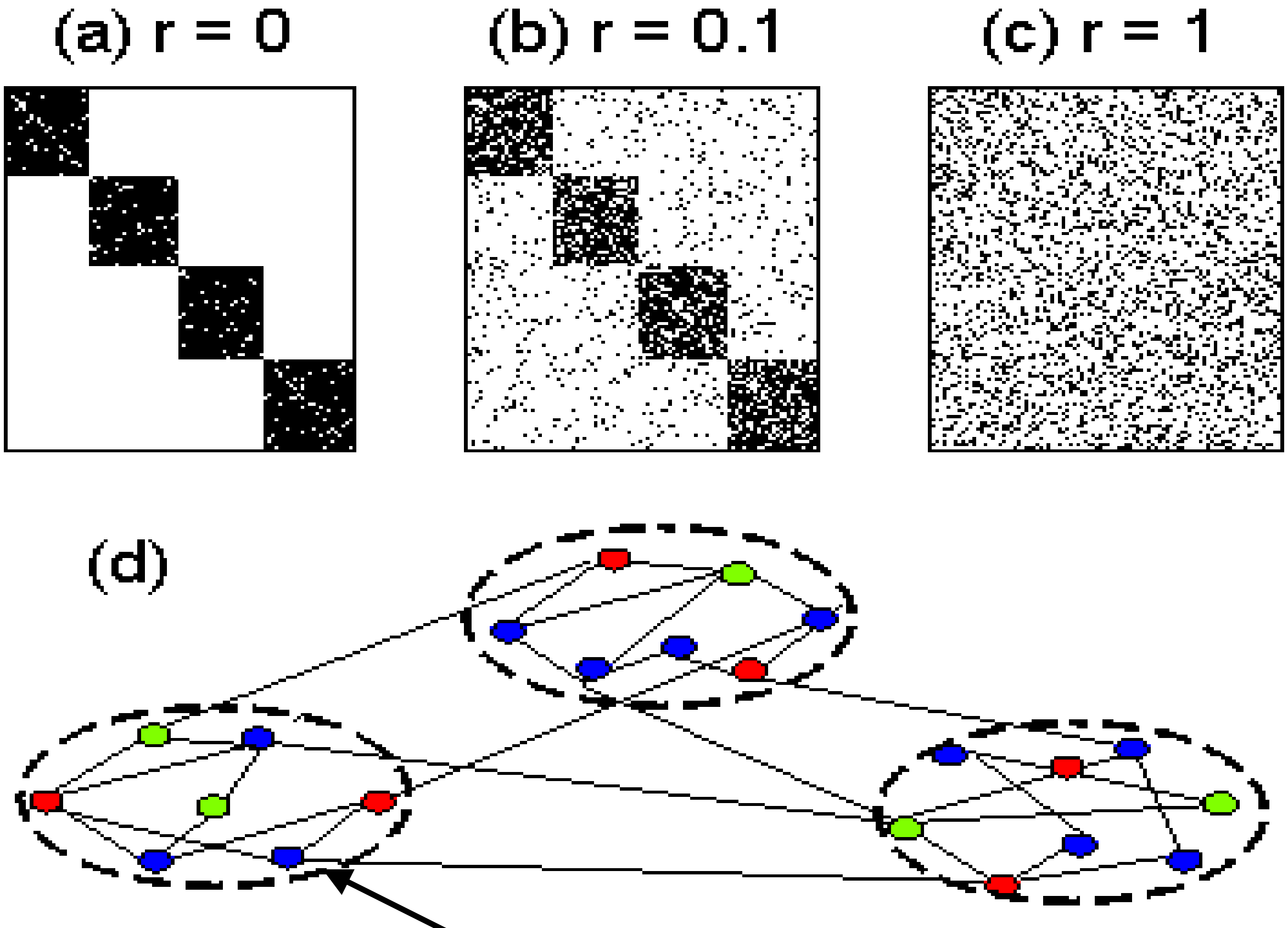
Gender	Type	Q	N.comm	Q_{random}±std	Modules
Female	AG.Freq	0.1205	2	0.0812±0.0173	[1 3 4 5 6] [2 7 8 9 10 11]
	AG.Time	0.1397	2	0.0983±0.0209	[1 3 4 5 6] [2 7 8 9 10 11]
	AF	0.1095	2	0.0729±0.0197	[1 3 4 5 6] [2 7 8 9 10 11]
Male	AG.Freq	0.0852	2	0.1301±0.0247	[1 4 9 10 11 12] [2 3 5 6 7 8]
	AG.Time	0.1646	4	0.1369±0.0244	[1 2 4 6] [3 5 7] [8 9] [10 11 12]
	AF	0.2398	4	0.1426±0.0253	[1 3] [2 4] [5 8 9][6 7 10 11 12]

Predictive power: Observation in 1998 showed the group had split into two (11,10,9,8,7,2) and (6,5,4,3) [1 had died]

A simple model of modular networks

Model parameter *r* :

Ratio of inter- to intra-modular connection density



Module ≡ random network

Comparison with Watts-Strogatz model

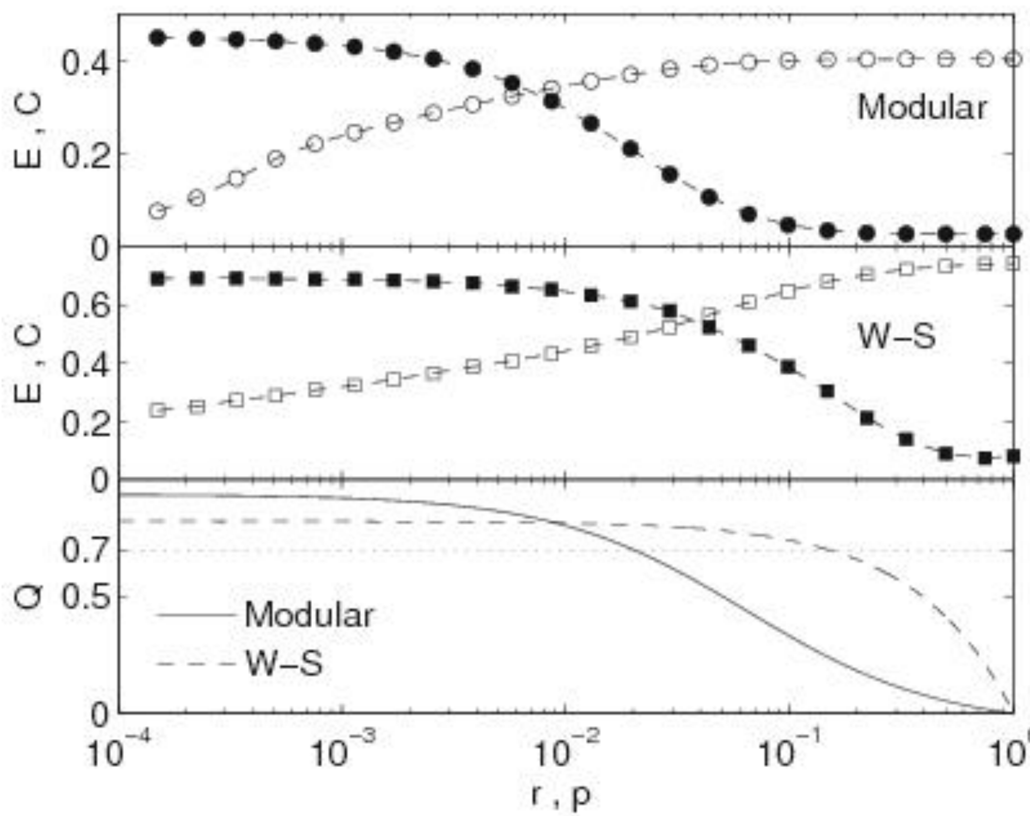
Structural measures used:

Communication efficiency $E = [\text{avg path length}, \ell\,]^{-1} = 2/N(N-1) \sum_{i>j} d_{ij}$

Clustering coefficient C = fraction of observed to potential triads $= (1/N) \sum_i 2n_i / k_i (k_i - 1)$

WS and Modular networks behave similarly as function of *p* or *r* (Also for between-ness centrality, edge clustering, etc)

In fact, for same N and <k>, we can find *p* and *r* such that the WS and Modular networks have the same *"modularity"* Q



Pan and Sinha, EPL (2009)

How can you tell them apart ?

Pan and Sinha, EPL (2009)

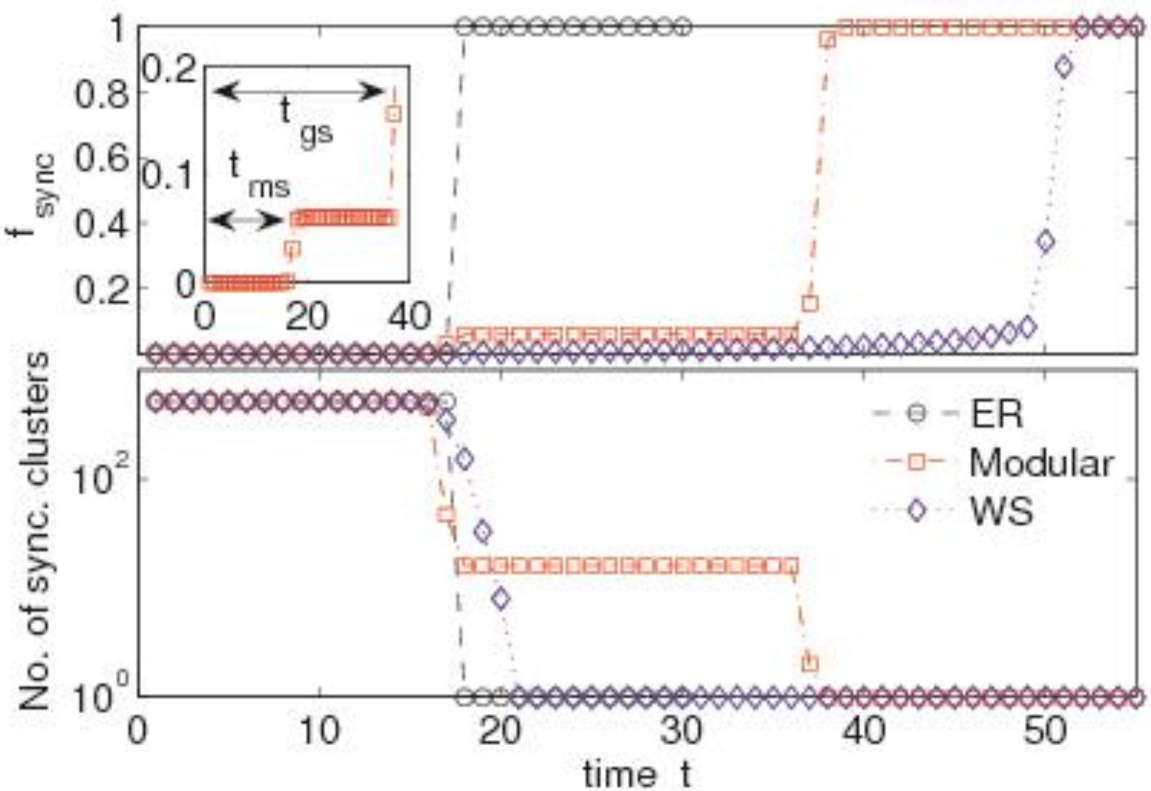
Dynamics on modular networks different from that on Watts-Strogatz small-world networks

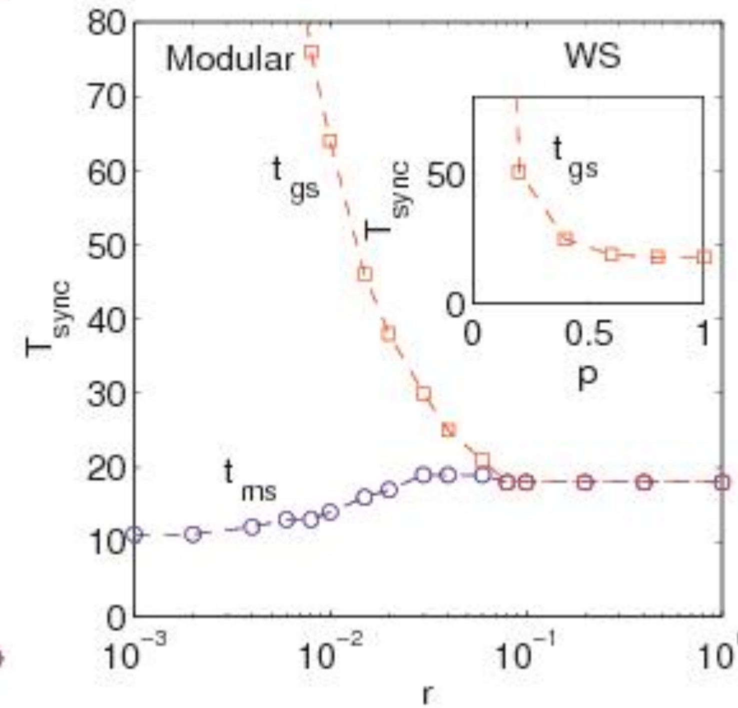
Consider synchronization on modular networks

e.g., phase oscillators: $d\theta_i /dt = \omega_i + (1/k_i)\sum K_{ij} \sin(\theta_j - \theta_i)$

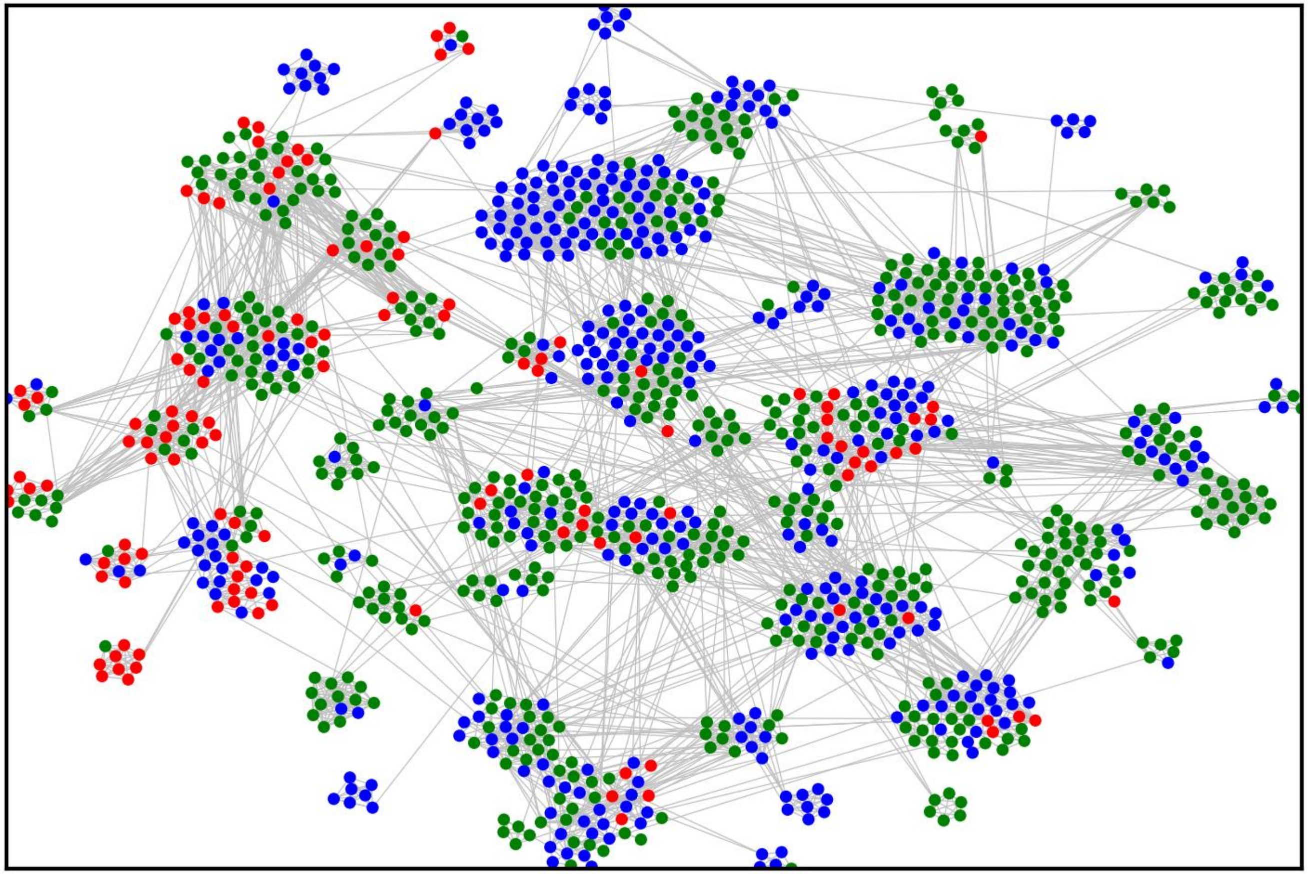
Network topology e.g., $K_{ij} = \kappa A_{ij}$

2 distinct time scales in Modular networks: $t_{modular}$ & t_{global}





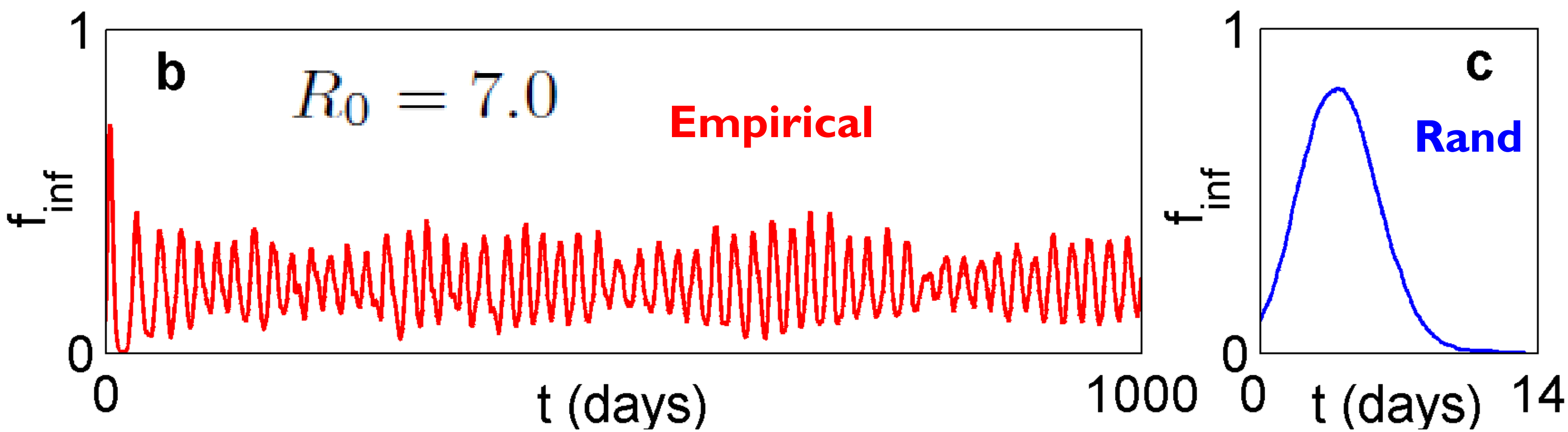
Progress of an epidemic in a modular social network



Susceptible Infected Recovered

Modularity promotes disease persistence

Contagia in empirical **modular** social contact network are **surprisingly persistent** compared to degree-preserved **randomized** networks which do not have community organization



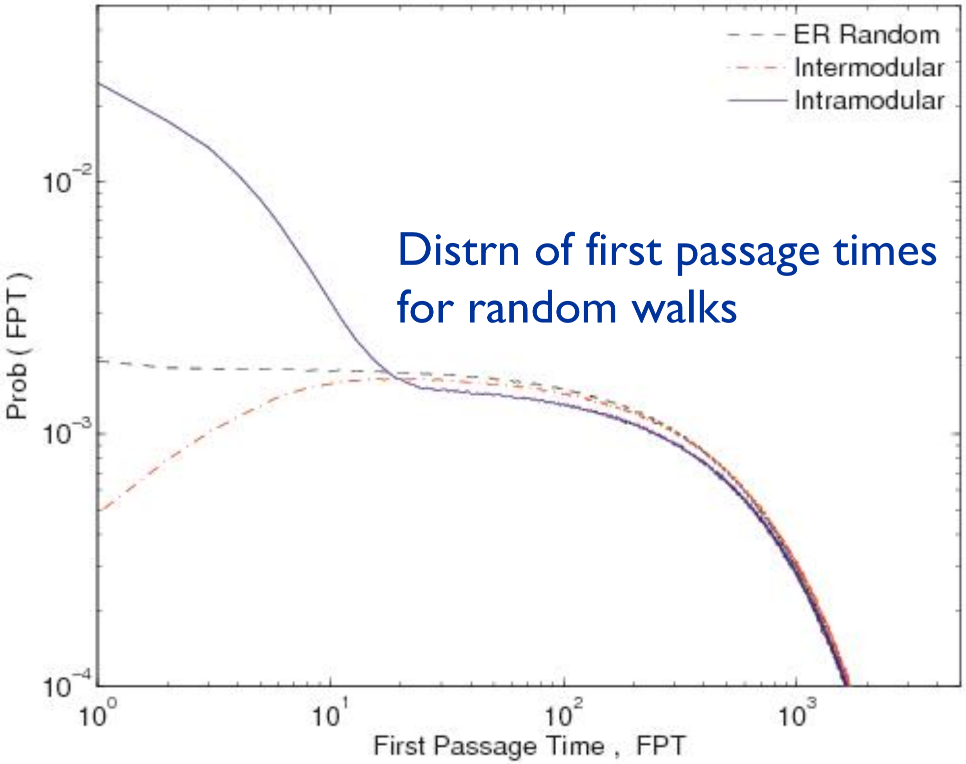
Difference even more pronounced if modularity is enhanced by selectively decreasing inter-modular connectivity

Figure: Pan & Sinha, EPL 2009

The presence of disease persistence in contact networks with community organization can be understood by analyzing

Diffusion process on modular networks

E.g., Random walker moving from one node to randomly chosen neighboring node



Shows the existence of two *distinct* time scales:

- fast intra-modular diffusion
- slower inter-modular diffusion

while random networks show a continuous range of time scales

In modular networks, the disease spreads slowly from module to module, allowing parts of the network to recover before spreading !

Existence of distinct time-scales in modular networks

Pan and Sinha, EPL (2009)

Consider linearized dynamics around synchronized state

$d\theta_i /dt = -(\kappa /k_i)\sum_j L_{ij}\, \theta_j$, (i = 1,…,N)

L is the Laplacian (**L** ≡ **D** – **A**)
κ: coupling strength of oscillators

Focus on the normal modes:

$\phi_i(t) = \sum_j B_{ij}\, \theta_j = \phi_i(0) \exp(-\lambda_i t)$, (i = 1,…,N)

B: matrix of eigenvectors
λ_i :eigenvalues

of **L'** = $\mathbf{D}^{-1}$ **L**,
D: diagonal matrix s.t. $D_{ii}=k_i$

$\mathbf{L'} \rightarrow \mathcal{L} = \mathbf{D}^{1/2}\, \mathbf{L'}\, \mathbf{D}^{-1/2}$ is symmetric, normalized Laplacian ⇒ λ_i real

Differences in time-scales of modes ⇒ gap in spectrum of $\mathcal{L}$

Mode for smallest λ_i : associated with global synchronization
Other modes : synchronization within different groups of oscillators

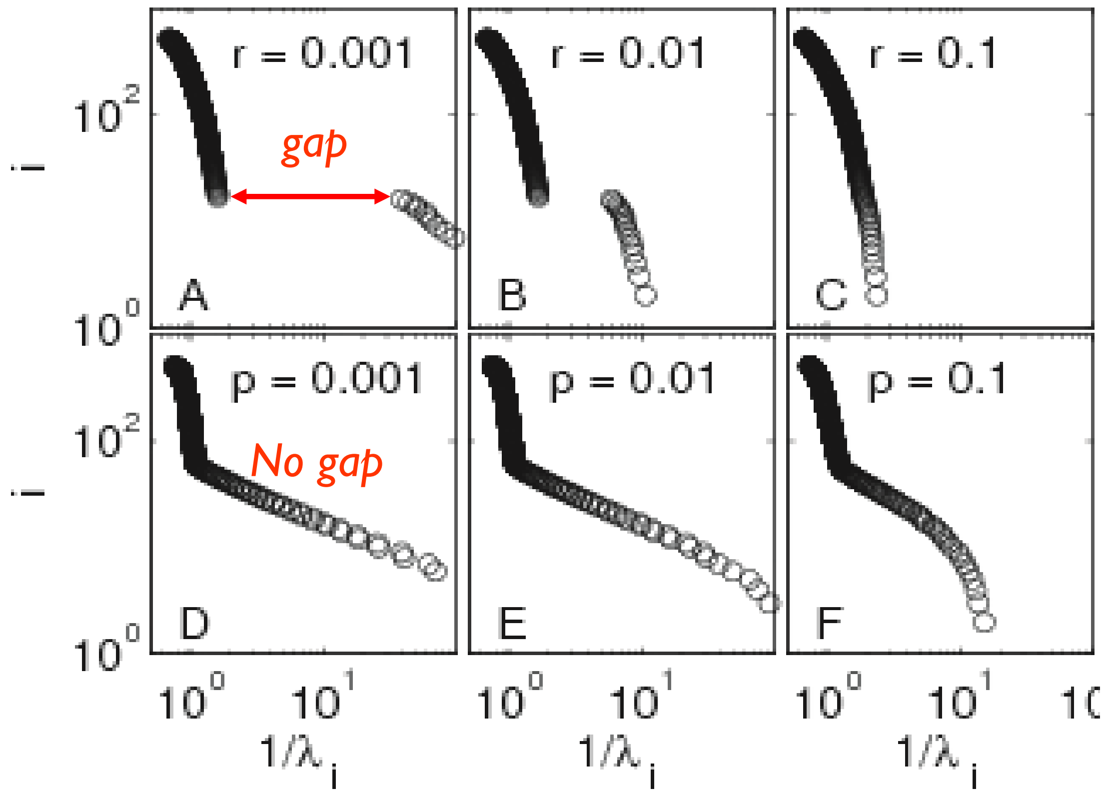
To relate this with diffusion, note that the transition probability from node i to j at each step of a random walk is $P_{ij} = A_{ij}/k_i$. The transition matrix **P** is related to the normalized Laplacian of the network as $\mathcal{L} = \mathbf{I} - \mathbf{D}^{1/2}\, \mathbf{P}\, \mathbf{D}^{-1/2}$ where **I** is identity matrix
Eigenvalues of **P** which are all real, the largest being 1 while the others related to different diffusion time-scales also show a gap ⇒ existence of distinct time-scales in the system

Pan and Sinha, EPL (2009)

Eigenvalue spectra of the Laplacian

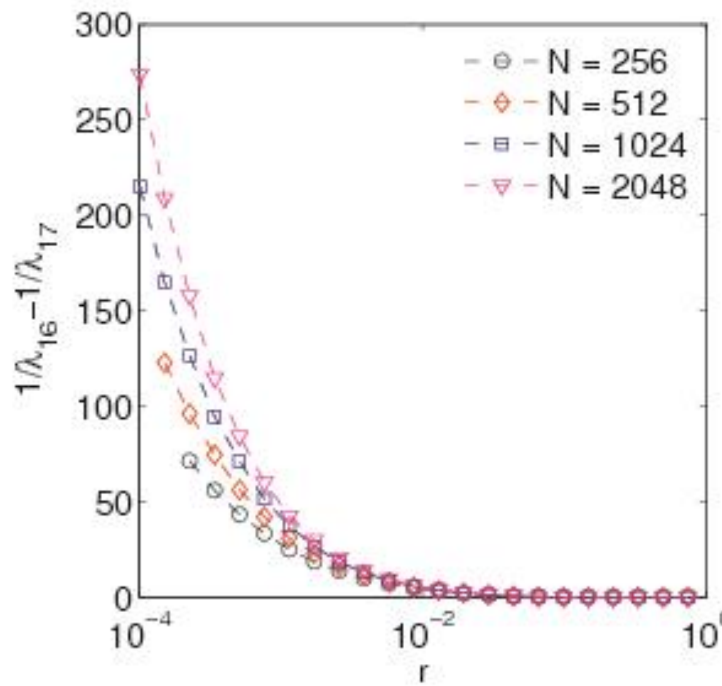
Shows the existence of spectral gap ⇒ distinct time scales

Modular network Laplacian spectra



WS network Laplacian spectra

Spectral gap in modular networks diverges with decreasing r



Existence of distinct time-scales in Modular networks

No such distinction in Watts-Strogatz small-world networks

How about real SW networks ?

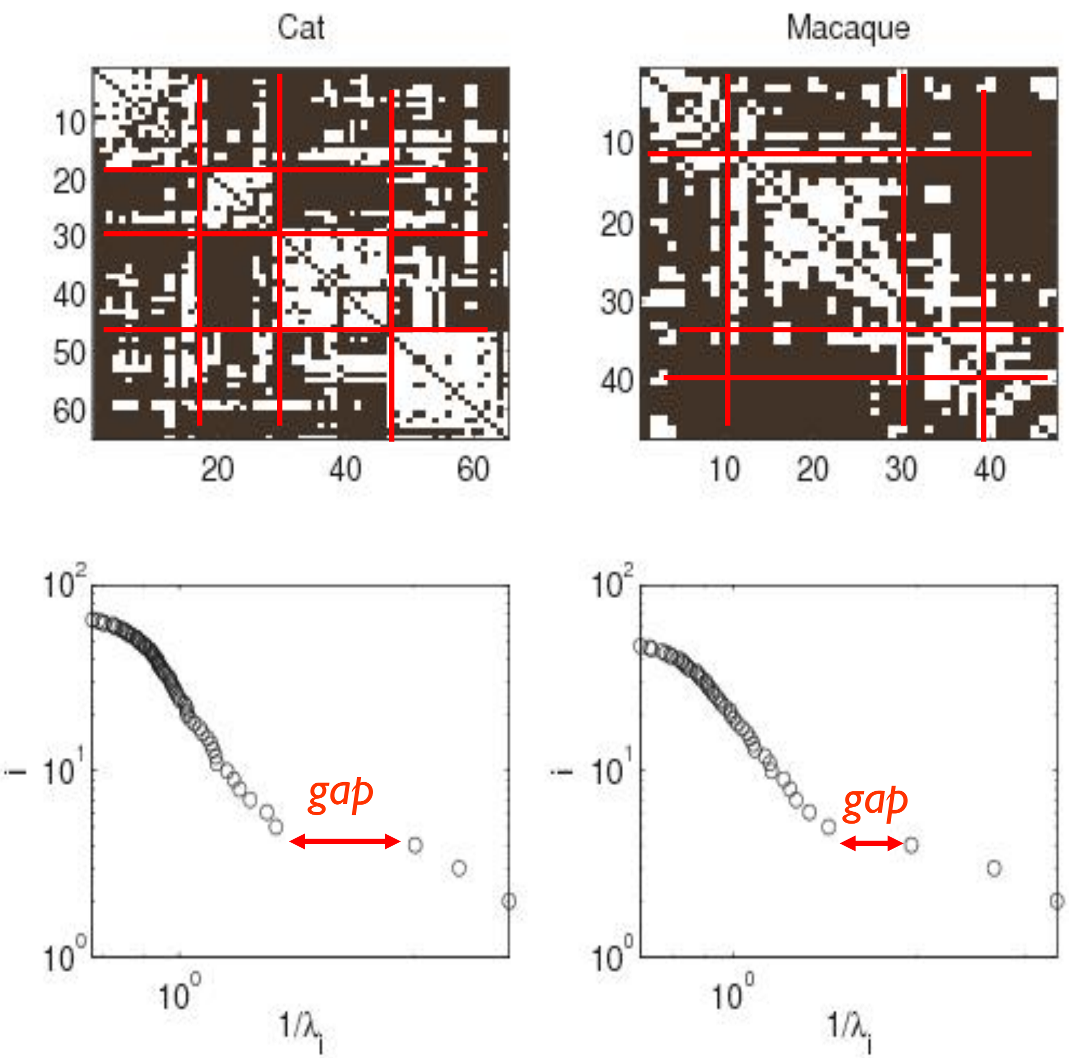


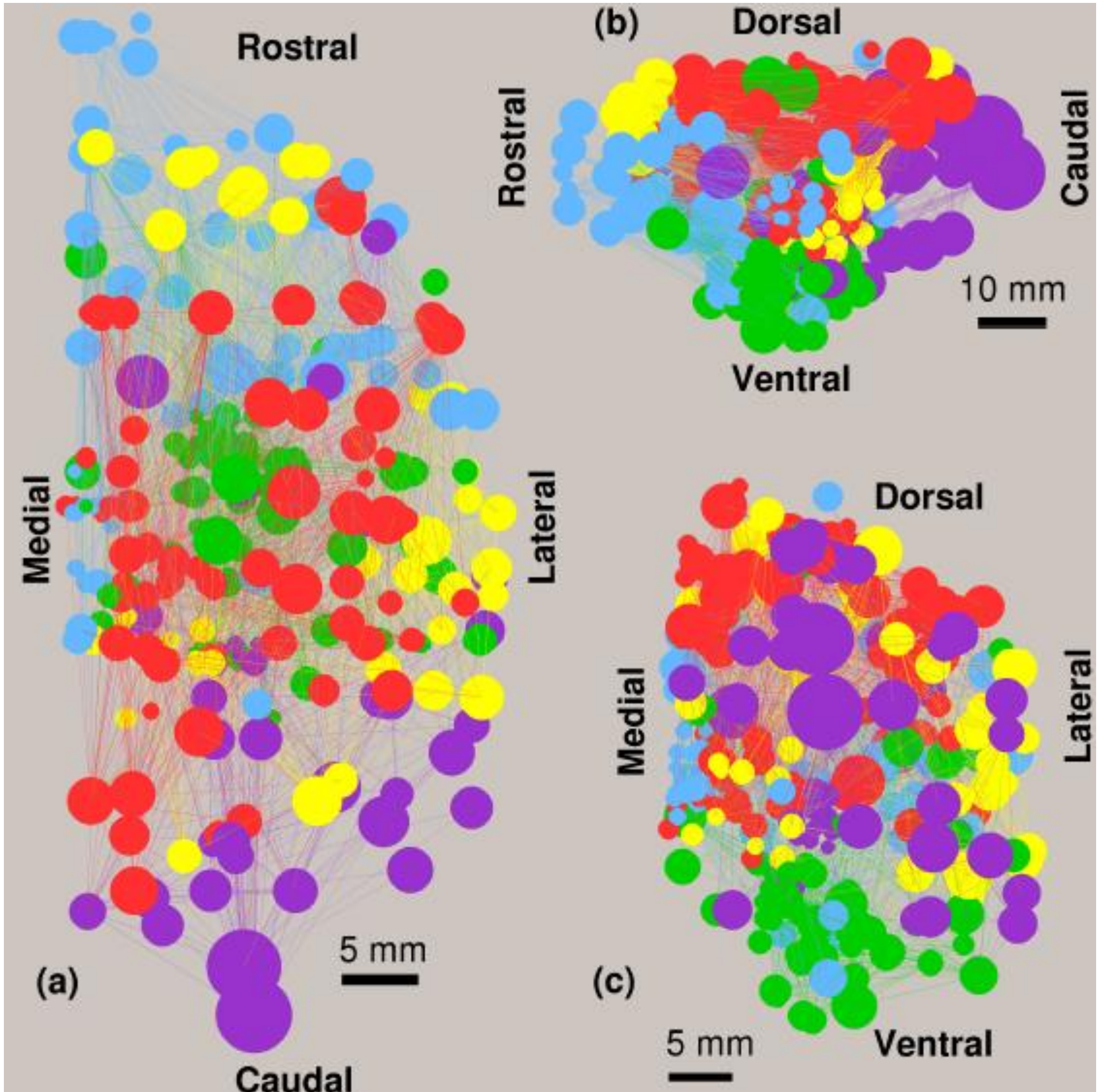
Figure: Pan and Sinha, EPL (2009)

The networks of cortical connections in mammalian brain have been shown to have *small-world* structural properties

The Laplacian spectral gap suggests their dynamical properties are consistent with modular organization

Fast local synchronization of neuronal activity (necessary for information processing) but preventing abnormal large-scale activation (as in seizures)

Modular organization of Macaque brain



Representation of the brain network from horizontal, sagittal and coronal views

CoCoMac : links between areas of macaque brain

(Revised from Modha & Singh, 2010)

Showing spatial positions of brain areas (circles), their relative volumes (different circle sizes) and the fibre tracts connecting them (shown as links).

The network consists of 5 distinct densely connected communities (different colored nodes). The communities appear to be localized in space with some exceptions.

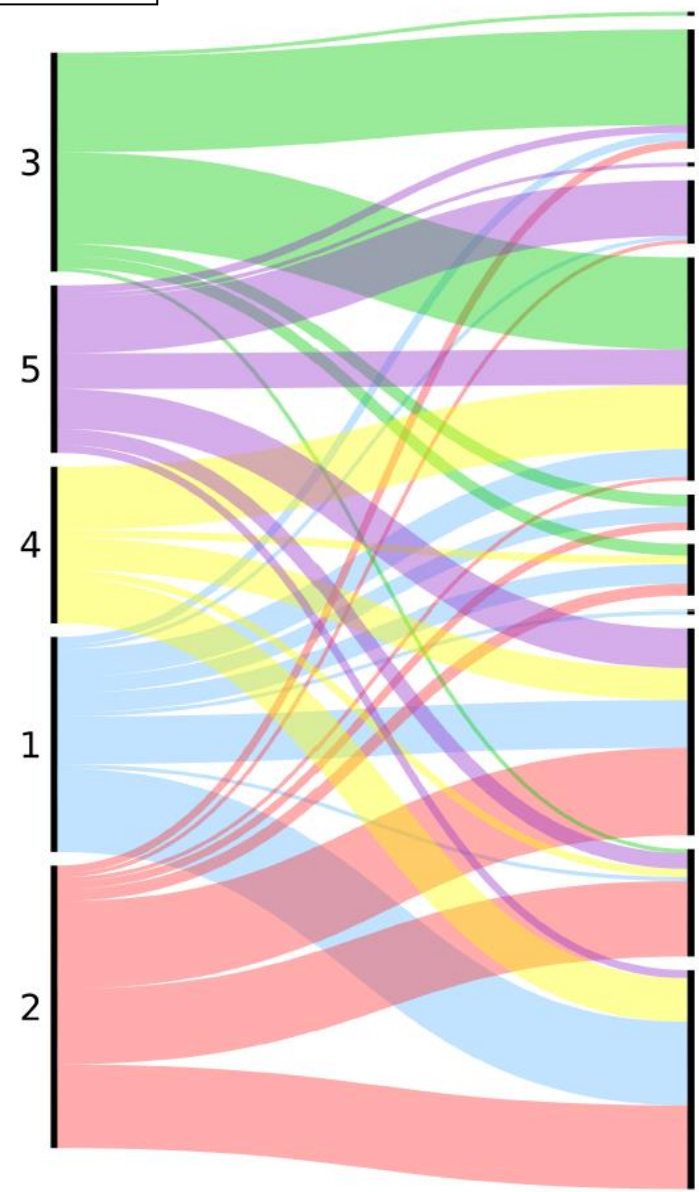
Image: Pathak et al, Phys Rev E, 2022

Relating Modules to Brain Structure & Function

Major cortical & sub-cortical divisions

Functional Modalities

Modules



emotional responses
formation of memory

visual

auditory

olfactory & gustatory

somatosensory

Hyp. Hypothalamus
BG Basal Ganglia
MB Mid-brain
OL Occipital Lobe
TL Temporal Lobe
Ins. Insula
Cing. Cingulate gyrus
OFC Olfactory complex
Thal. Thalamus
PL Parietal Lobe
FL Frontal Lobe

Image: Pathak et al, Phys Rev E, 2022

Information spreads in Macaque brain faster due to specific pattern of intra-/inter-modular links

Importance of connector hubs: possibly integrating local activity for coherent response

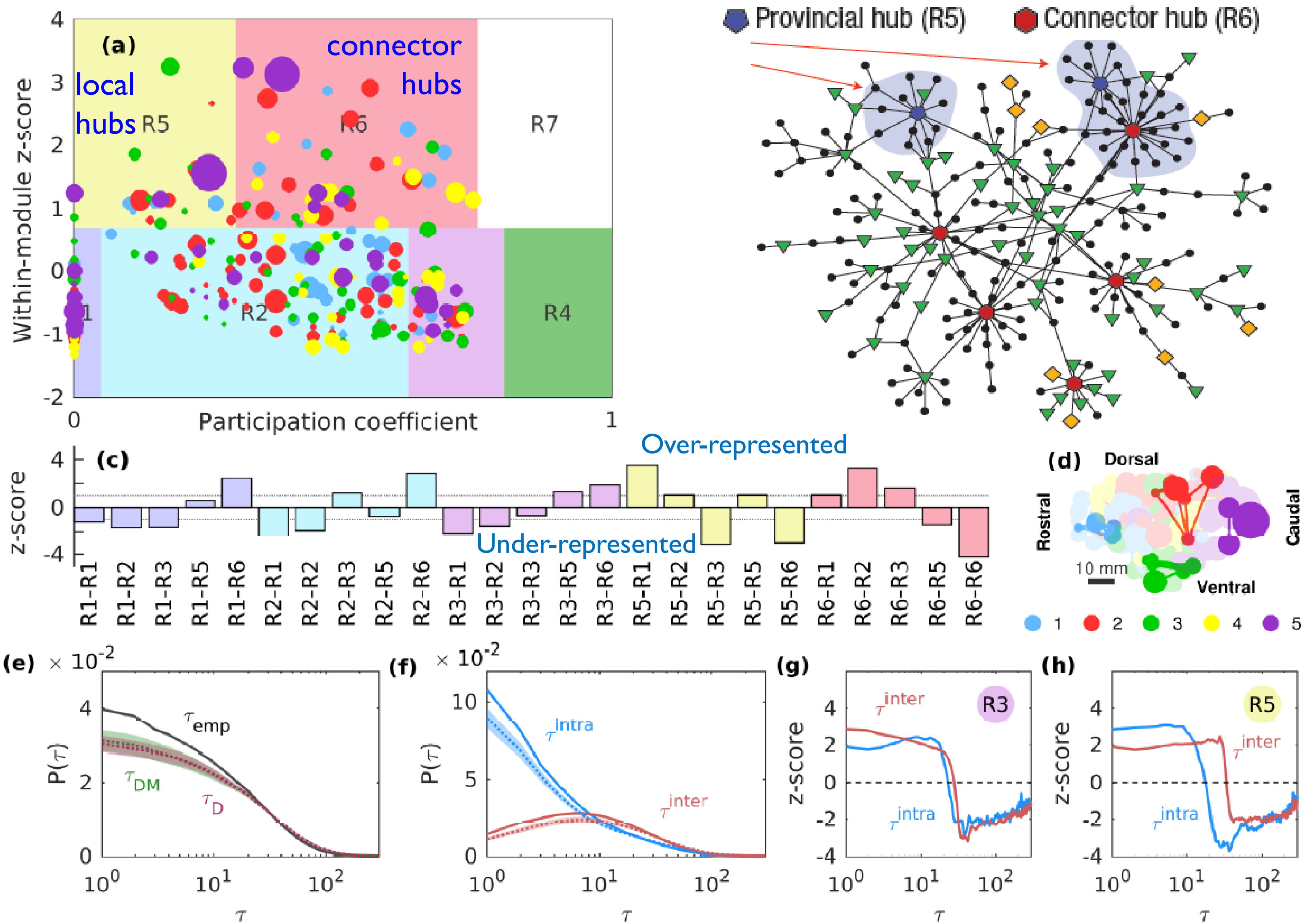
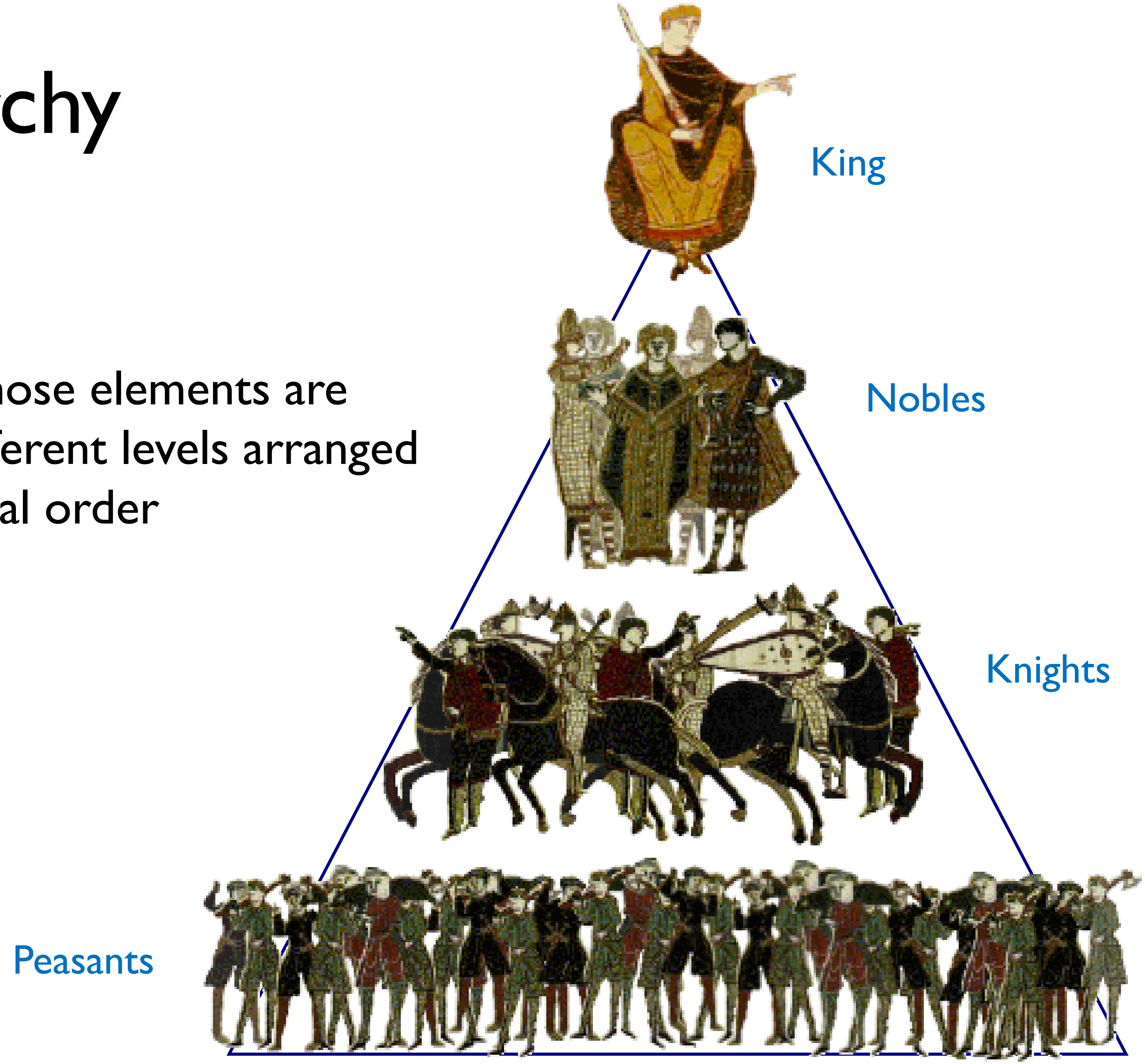


Figure: Pathak, Menon aand Sinha (2022)

Figure: Guimera & Amaral (2005)

Hierarchy

A system whose elements are placed at different levels arranged in a sequential order



Example: Food Webs

Approximately directed acyclic (i.e., no cycles) networks

$\Rightarrow$ intrinsic hierarchy among species such that, those higher up in the hierarchy prey on those lower down, but not vice versa

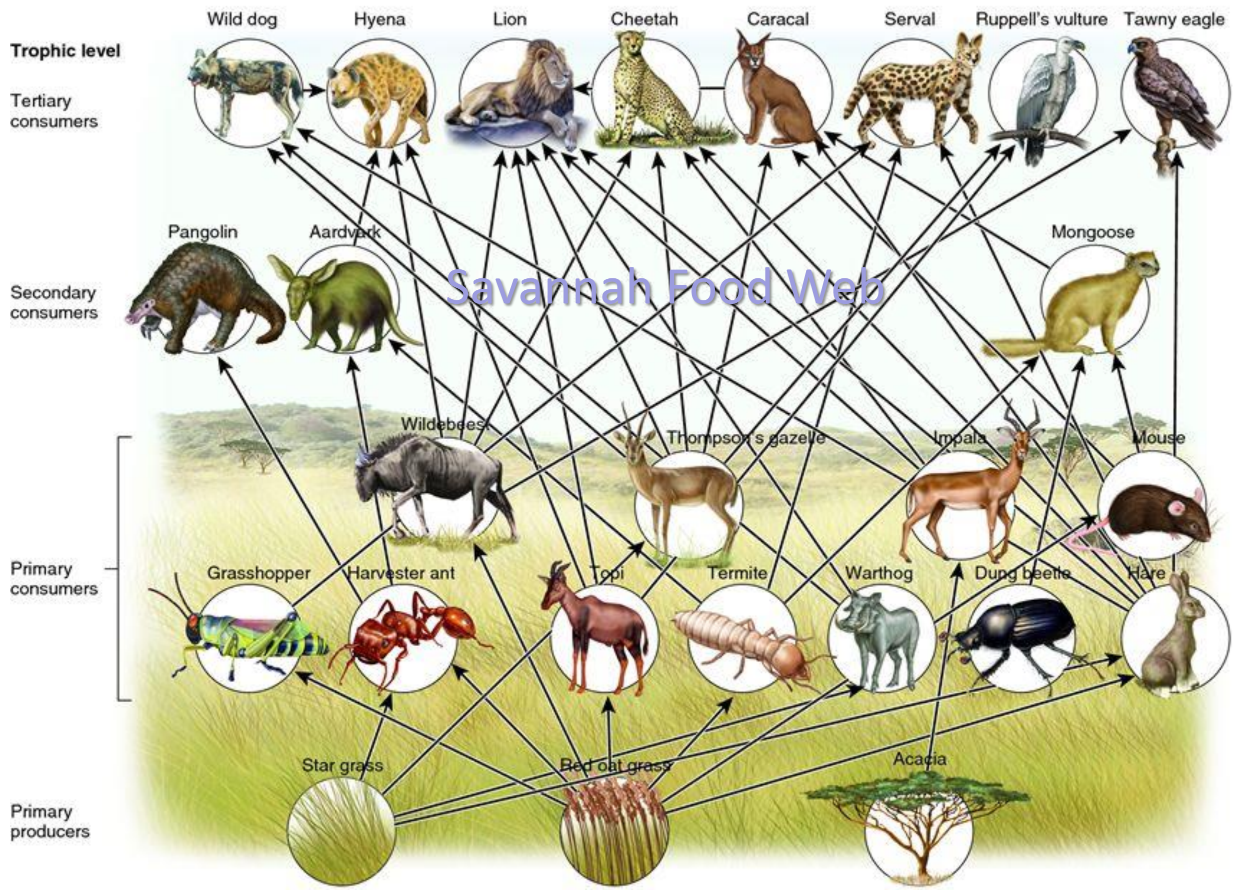
Rank of a species in the hierarchy $\rightarrow$ trophic level

Number of loops $L_r = \sum_i \lambda_i^r$

λ : eigenvalues of A

acyclic networks (no cycles) have *nilpotent* adjacency matrix $\rightarrow$ all eigenvalues zero

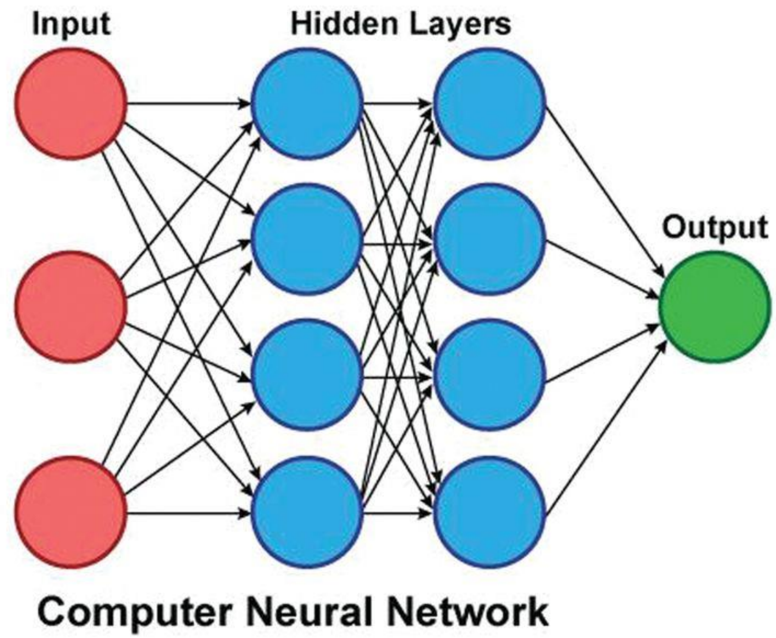
Image: savannabiomeassignment.weebly.com/



Example: Brain

Neural networks used in deep learning are inspired by the layered network organization of the brain

Input layer ↔ sensory neurons
Hidden layers ↔ interneurons
Output layer ↔ motor neurons



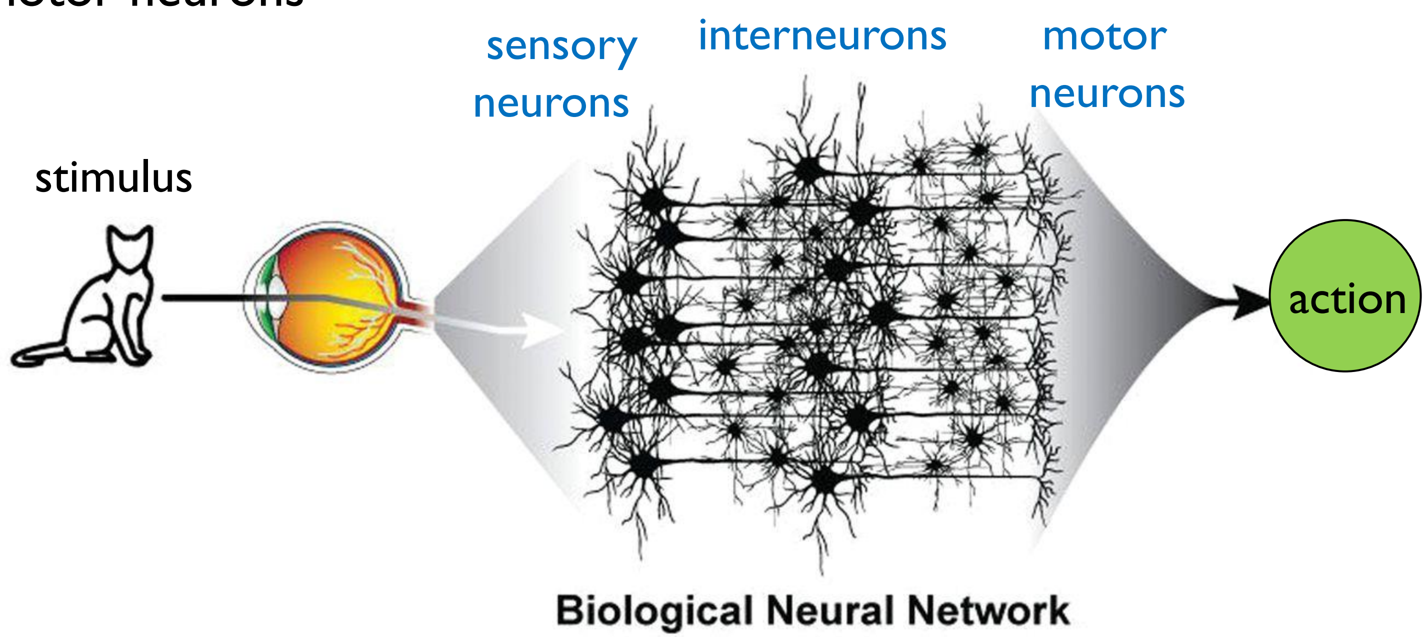
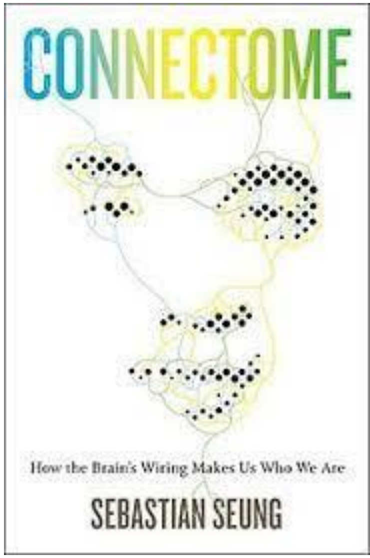
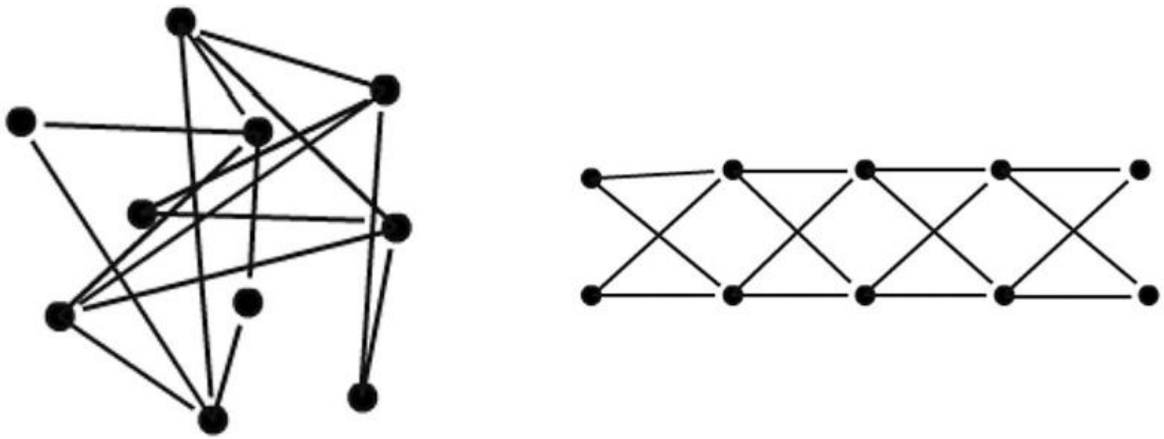


Image: G. Zaharchuk et al, Am J Neuroradiol **39** (2018) 1776

The Problem



"[In birds, a network of HVC neurons] retains the memory of a song. Whenever the bird sings, the memory is recalled by converting it into sequential spiking. … we could simply examine the connectome to find out whether it's organized like a synaptic chain …
[Unfortunately] it's not obvious whether a connectome contains a chain unless the sequential ordering of the neurons is known. To see why…



both [networks] have exactly the same connectivity. The neurons on the left have been scrambled to hide the chain. To reveal it, we must unscramble the neurons to yield the diagram on the right."

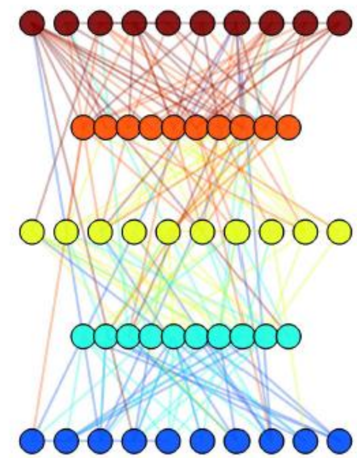
Can't be done by hand for any reasonably large network!

Images: S Seung, *Connectome*, Houghton Mifflin Harcourt, NY, 2012

Analogous to modularity, we can define

The Hierarchy Index

Using the intuitive notion of a hierarchical network having its nodes arranged in *multiple levels* having a specific sequence, with a preference for nodes in *neighboring* levels to be connected



For a directed network, hierarchy index for a given arrangement of a network into hierarchical levels

$$H = \frac{1}{L}\sum_{i,j}\left[A_{ij} - \frac{k_i^{in} \cdot k_j^{out}}{L}\right] \cdot \left(\delta_{l_i,l_j+1} + \delta_{l_i+1,l_j}\right)$$

probability of a directed edge between 2 nodes proportional to the product of their in- and out-degrees

=1 if nodes belong to successive levels in the hierarchy

The method

$$H = \frac{1}{L}\sum_{i,j}\left[A_{ij} - \frac{k_i^{in}\cdot k_j^{out}}{L}\right]\cdot\left(\delta_{l_i,l_j+1} + \delta_{l_i+1,l_j}\right)$$

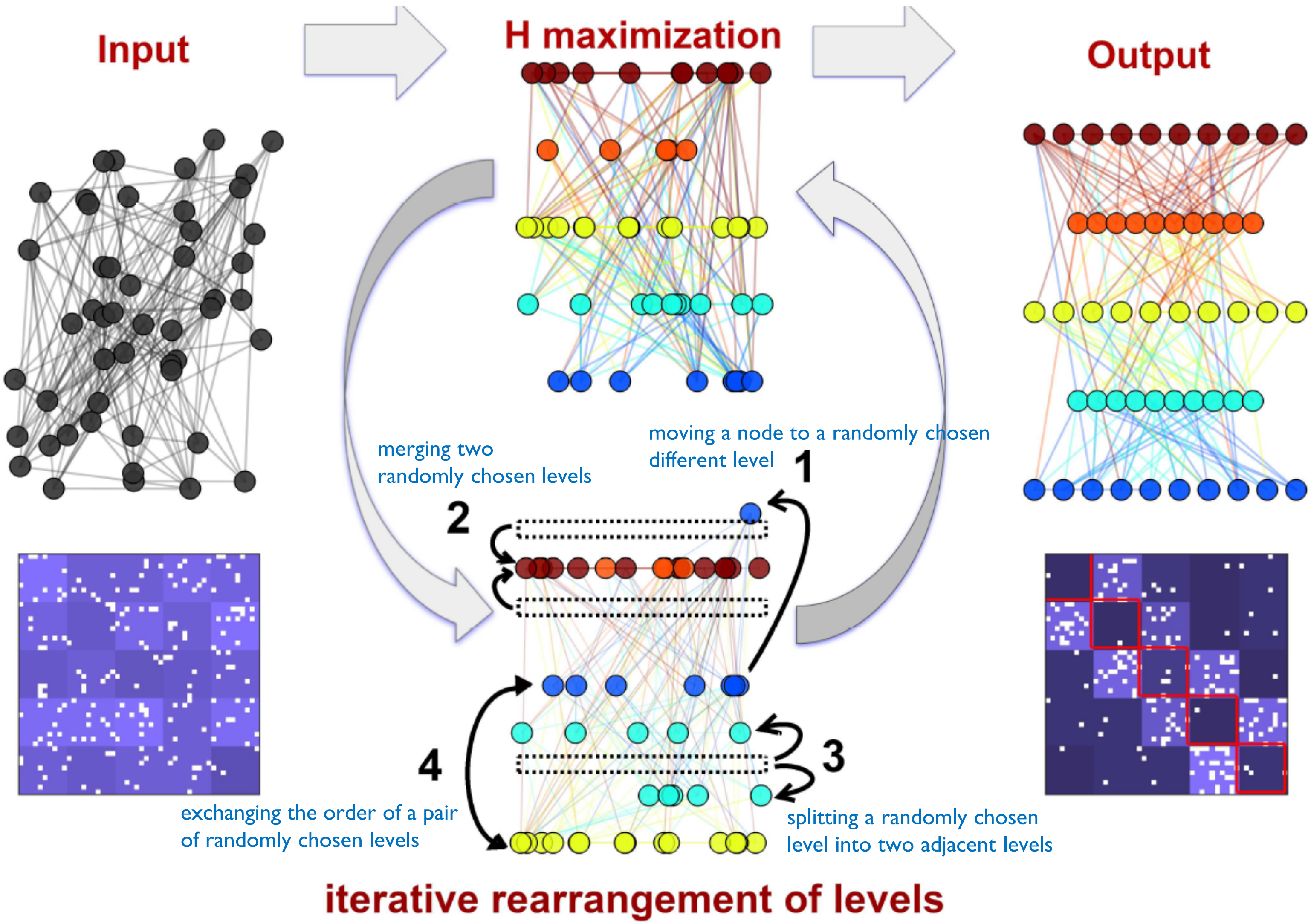
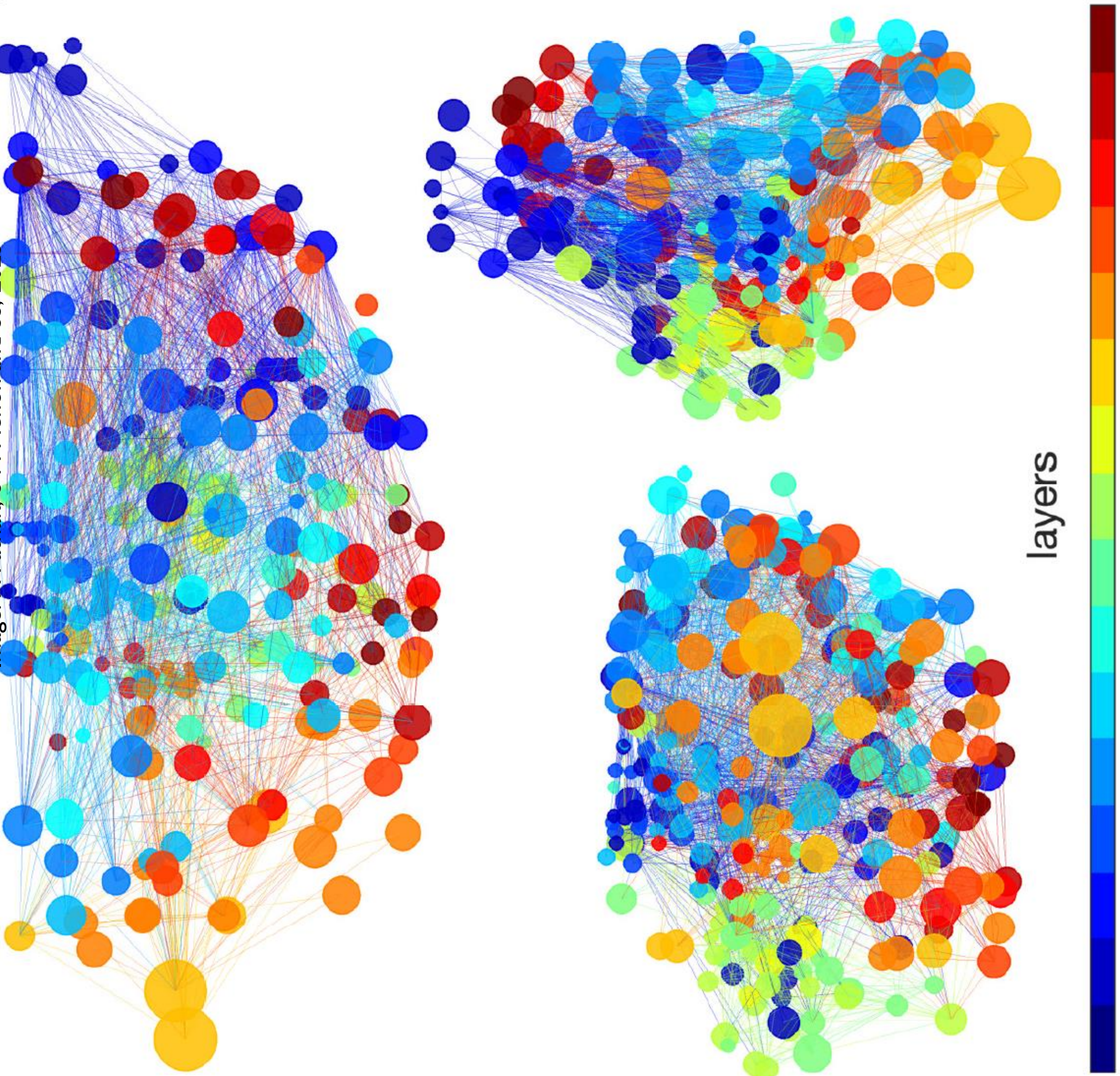


Image: A Pathak, S N Menon and SS, 2023

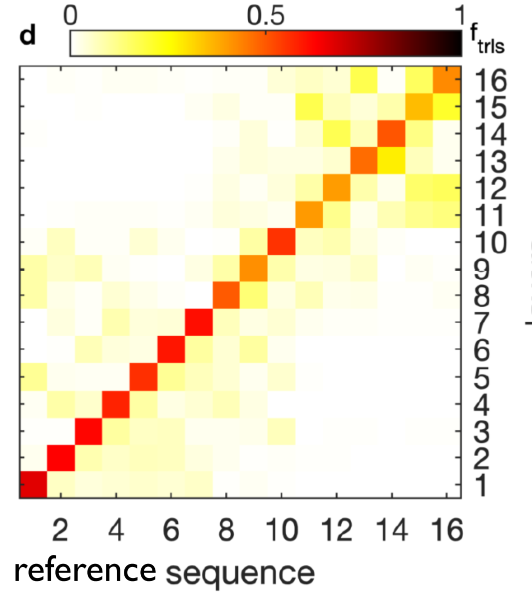
Back to the Macaque

Image: A Pathak, S N Menon and SS, 2023



The reference sequence: Representation of the brain network from horizontal, sagittal and coronal views

robust sequential arrangement of layers following a reference sequence over 10^3 rlzns



invariance of hierarchical partitioning of brain areas over 10^3 different rlzns

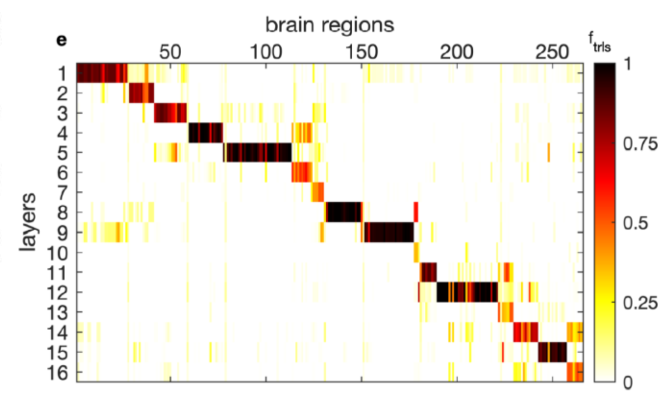
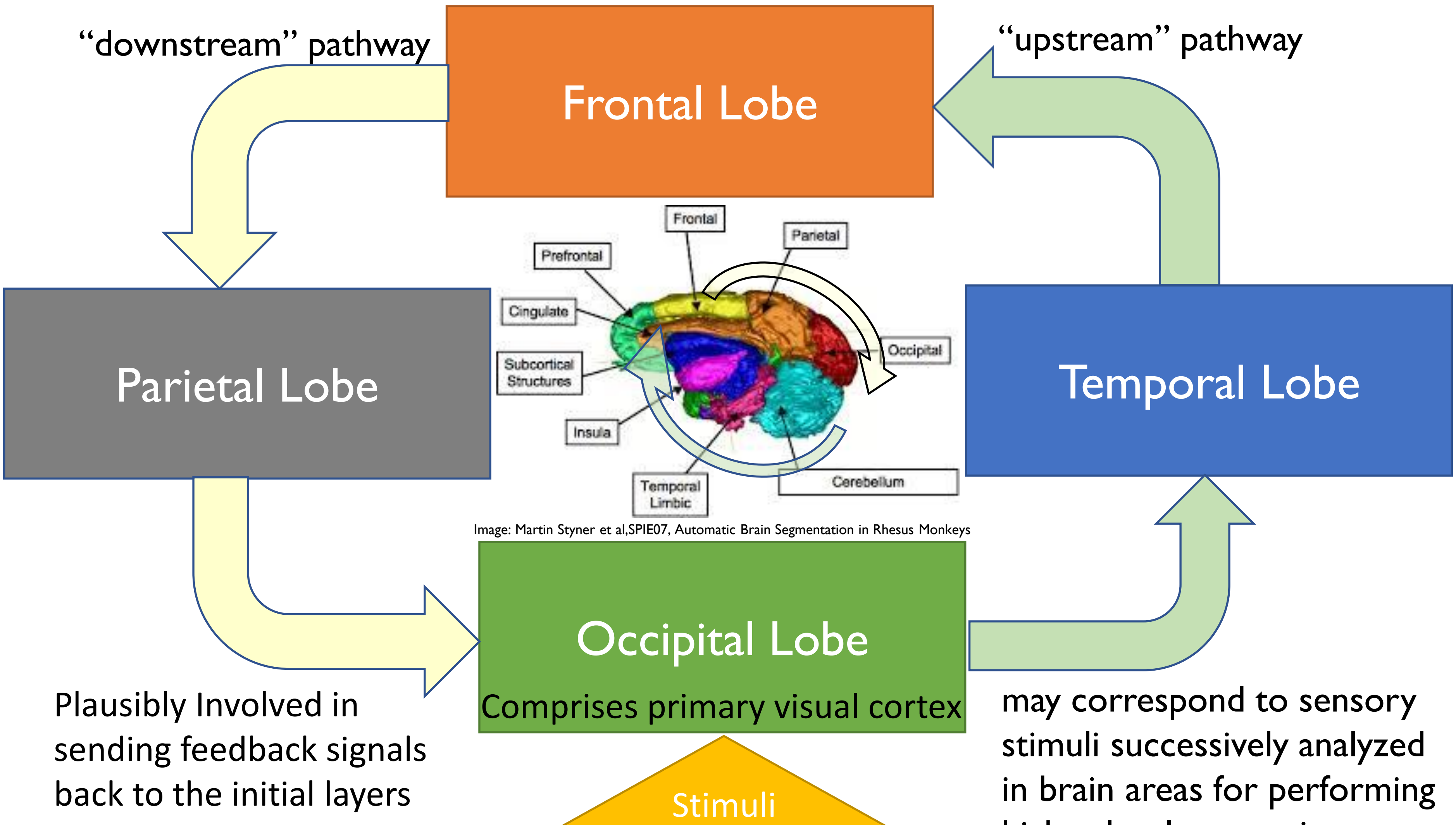


Image: A Pathak, S N Menon and SS, 2023

Implication

two distinct streams of signals propagation parallel to anteroposterior axis



analogous model of bottom-up and top-down processing working in conjunction has been proposed for vision

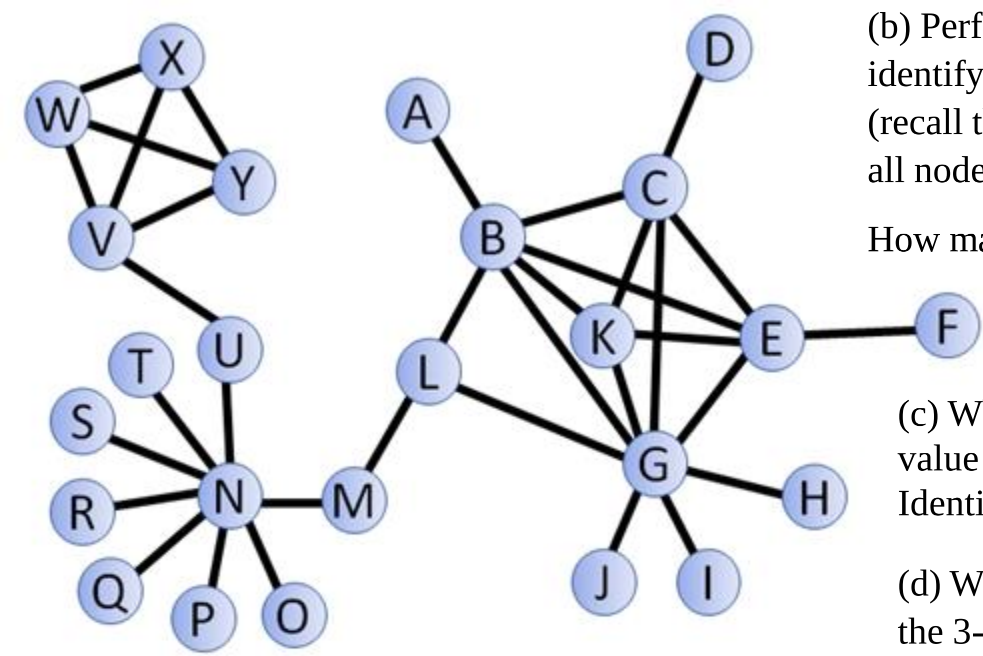
Assignment

	A	B	C	D	E	F	G	H
A	0	1	0	0	0	1	1	0
B	1	0	0	0	0	1	1	0
C	0	0	0	1	1	1	0	1
D	0	0	1	0	1	0	0	1
E	0	0	1	1	0	0	0	1
F	1	1	1	0	0	0	1	0
G	1	1	0	0	0	1	0	0
H	0	0	1	1	1	0	0	0

(a) Consider a network comprising 8 nodes (labelled A,…, H) whose adjacency matrix is shown at right. Partition the network into two modules – i.e., show that the nodes can be divided into two groups such that the number of links *between* the nodes belong to the two groups is *minimized*.

Mention which nodes will form one group and which nodes will form the other.

What is least number of links that occur between the two modules ? Identify these links, i.e., mention the pair of nodes (belonging to the different groups) that each link connects.



(b) Perform the recursive core decomposition technique to identify the k-order cores of the network where $k = 1,2,3,\ldots$ (recall that the k-core of a network is the subnetwork containing all nodes that have degree at least equal to k).

How many nodes are in (i) 1-core, (ii) 2-core, and (iii) 3-core ?

(c) What is the order of the innermost core, i.e., the highest value of k for which the corresponding core is non-empty ? Identify the nodes in the innermost core.

(d) What is the size of the largest *connected* component of the 3-core ?