Machine Learning for Neuroimaging and Neuroscience

Lesson 4

Complex

networks
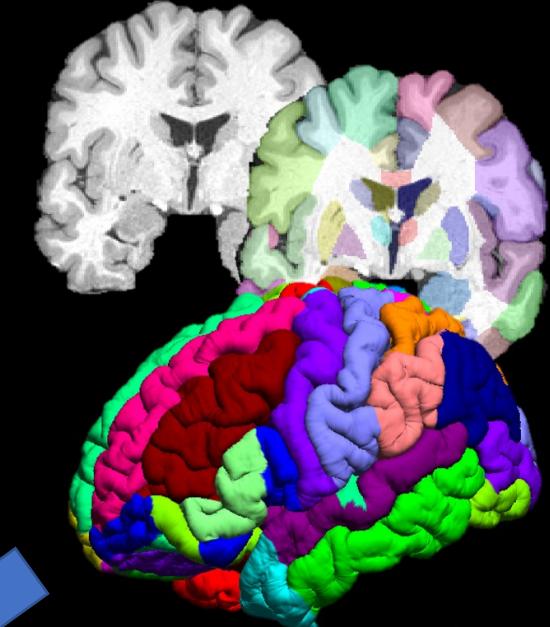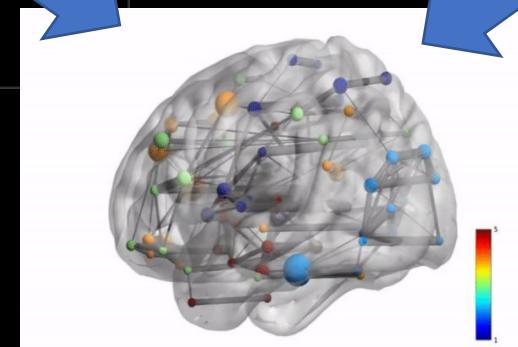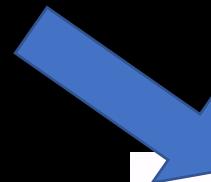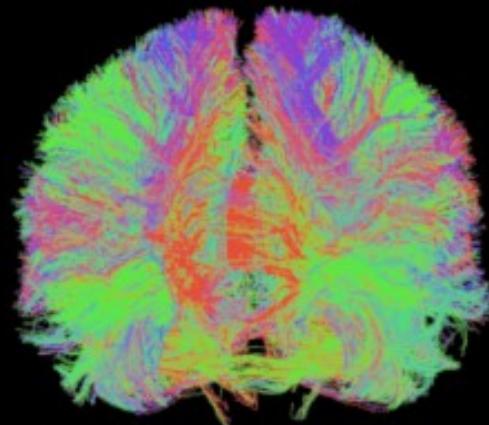
Alessandro Crimi

AGH

AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY
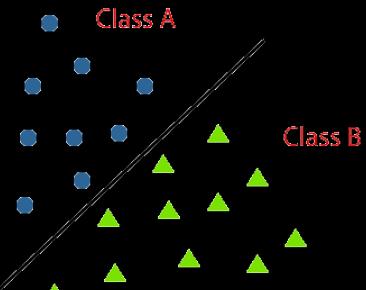
Structural Connectivity
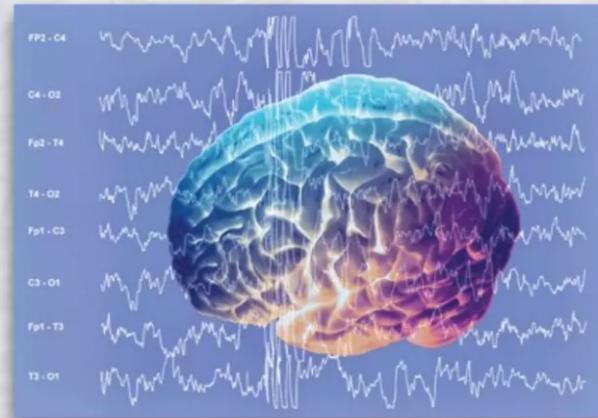
brainsuite.org

Functional Connectivity
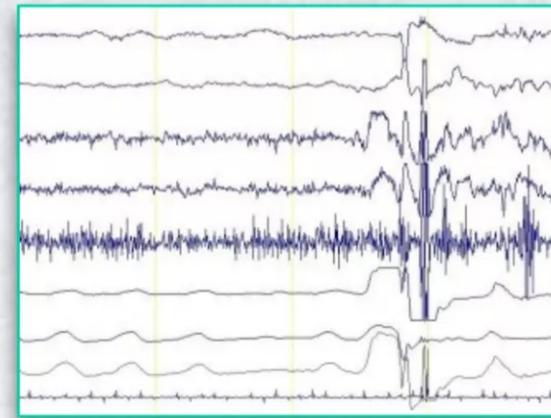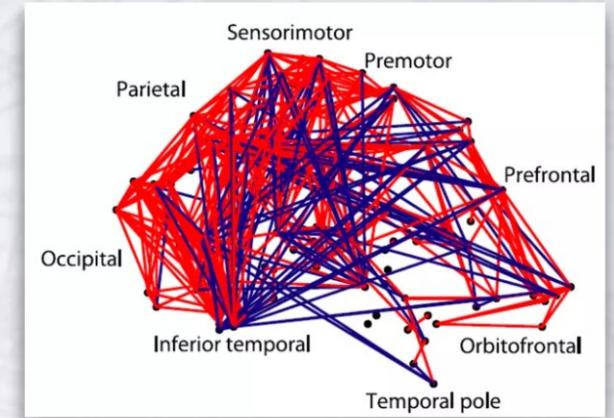
STEP 1	STEP 2	STEP 3
Measuring Brain Activity	Time Series Analysis & Network Construction	Network Analysis
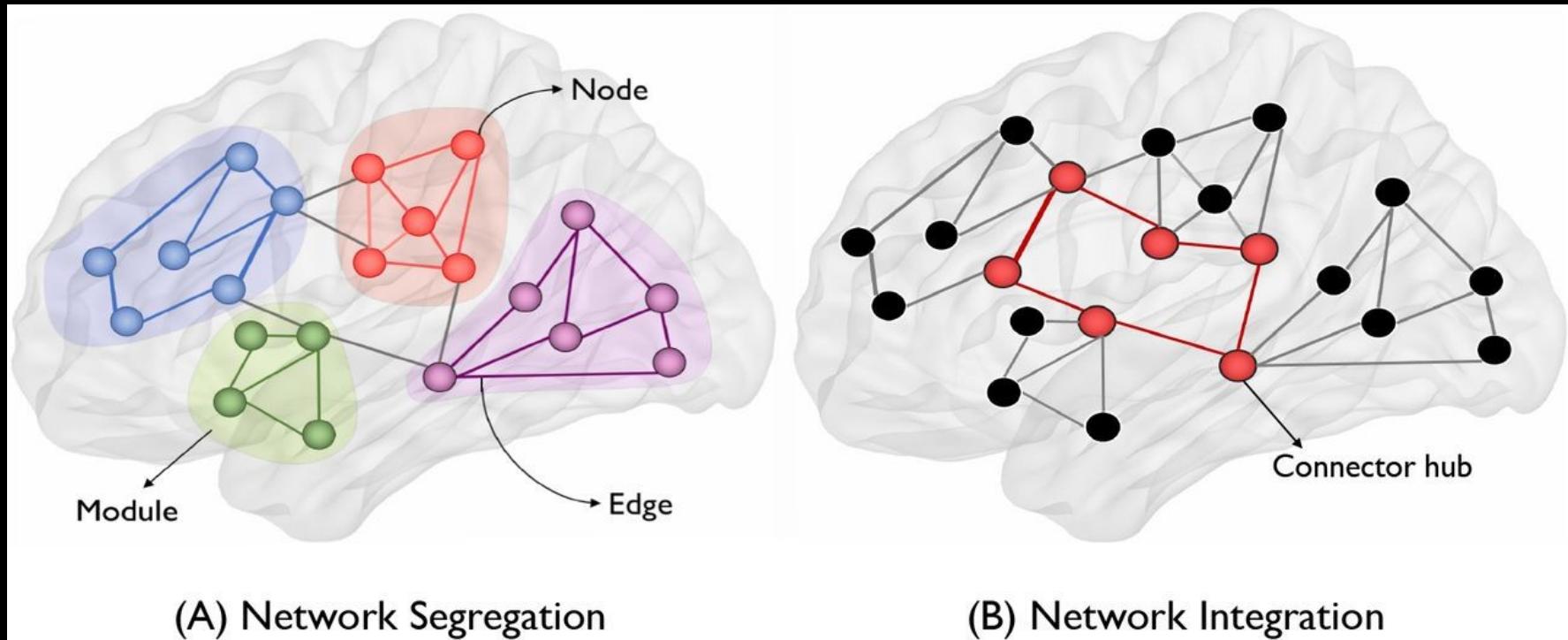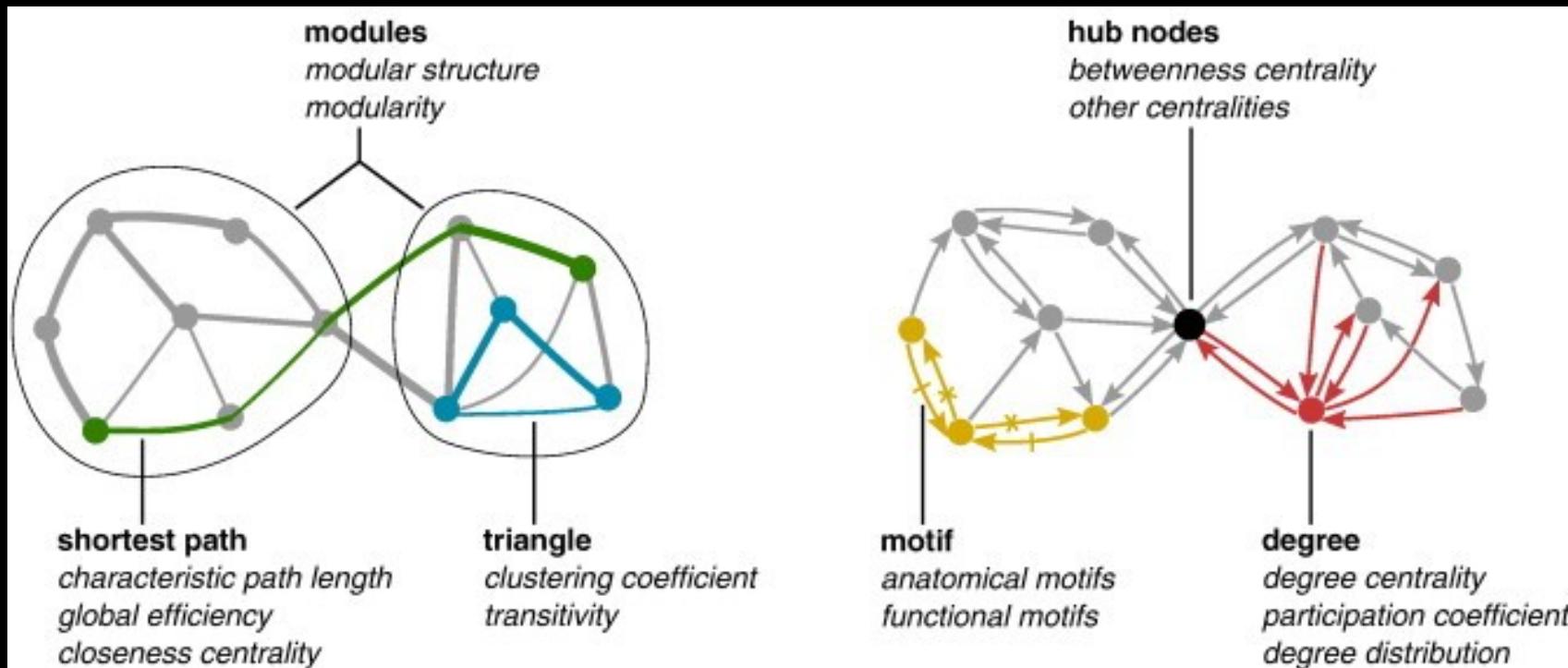
Segregation and integration

(A) Network Segregation

(B) Network Integration

Segregation and integration
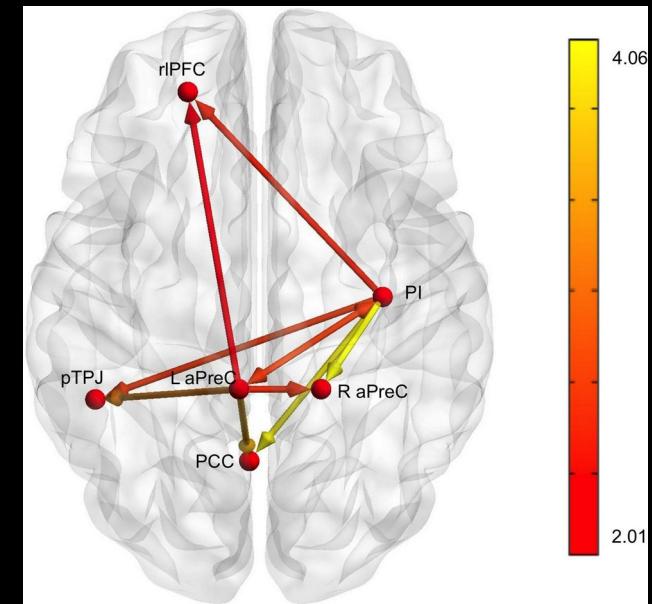
Brain causality

Edges can be directed or undirected

- Undirected: structural and functional connectivity

- Directed: effective connectivity
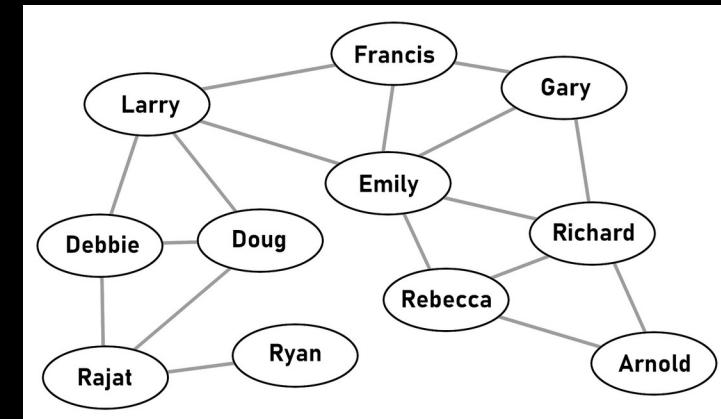
Effective connectivity is related to the activity of a brain region can impact another brain region.

The small-world brain network

A small-world network is a type of network where most nodes are not directly connected to each other but can be reached from every other node by a relatively short number of steps, even though the network might be quite large.

The defining characteristic of a small-world network is its high clustering coefficient (meaning nodes tend to cluster together) combined with short average path lengths between nodes. This phenomenon was popularized by the "six degrees of separation" concept, suggesting that any two people in the world can be connected by a chain of acquaintances with an average of six intermediaries.
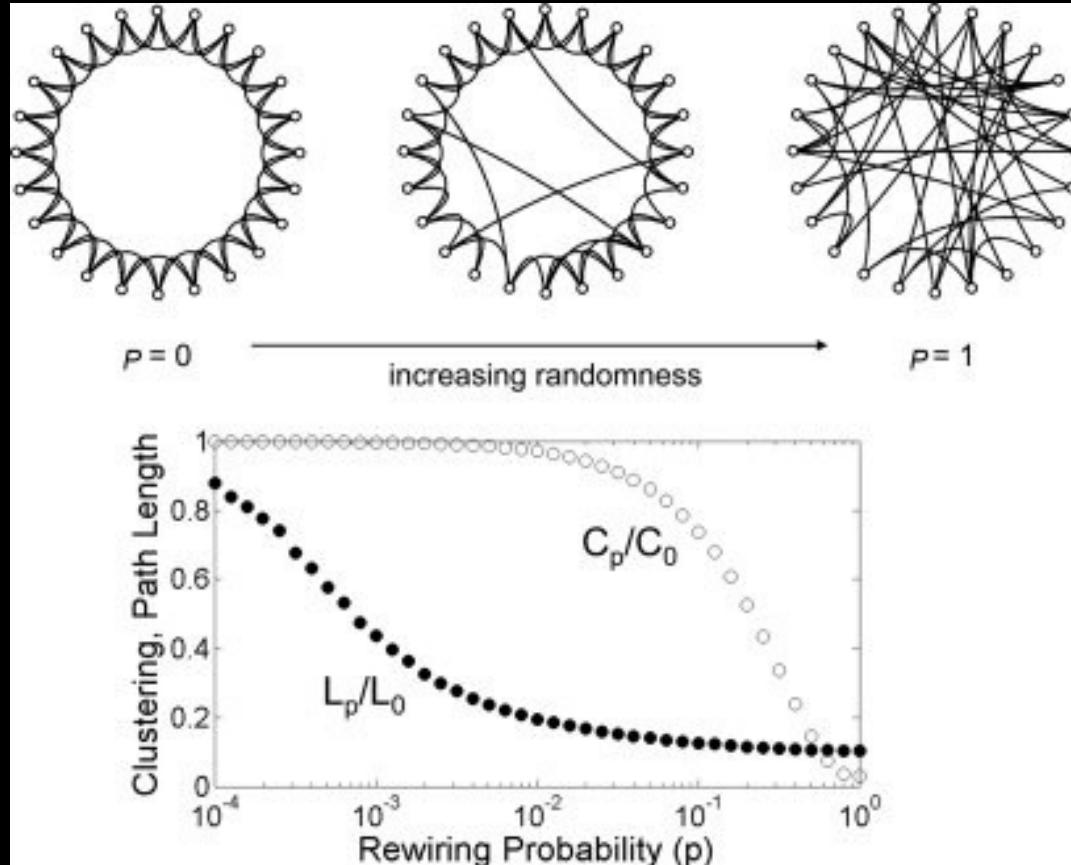
The small-world brain network (Watts&Strogazt)

The Watts and Strogatz model for small-world networks has been applied to study brain connectivity to understand how neurons communicate and form networks within the brain.
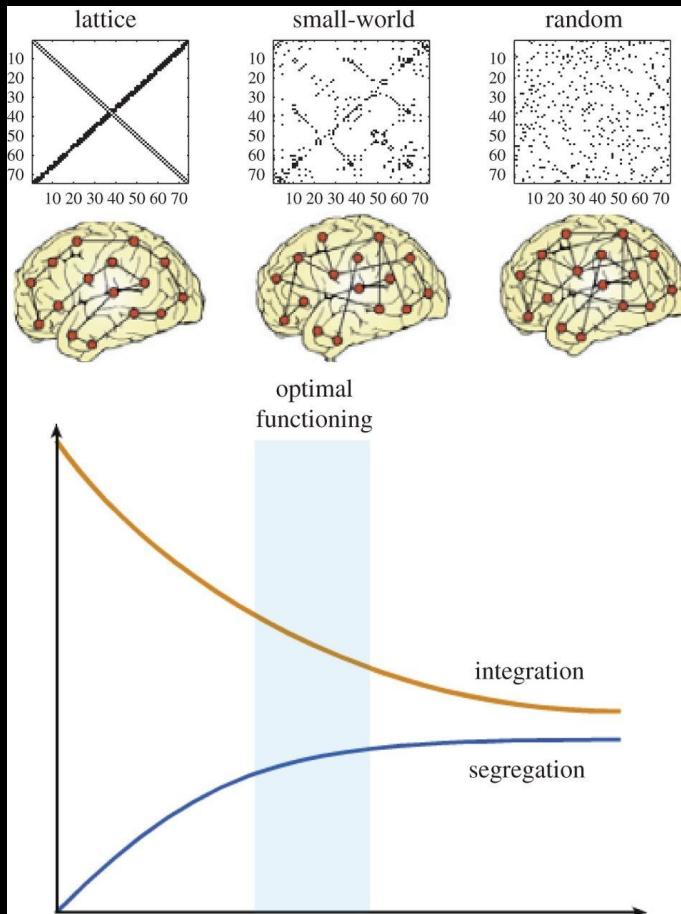
In their model, Watts and Strogatz proposed a way to create small-world networks by starting with a regular lattice structure (where each node is connected to its nearest neighbors) and introducing a small number of random connections. This model demonstrated that by adding a few random connections to an otherwise ordered network, it was possible to achieve a balance between high local clustering (as seen in regular networks) and short average path lengths (characteristic of random networks).

The small-world brain network (Watts&Strogazt)

Sporns 2011, Frontiers in Computational Neuroscience
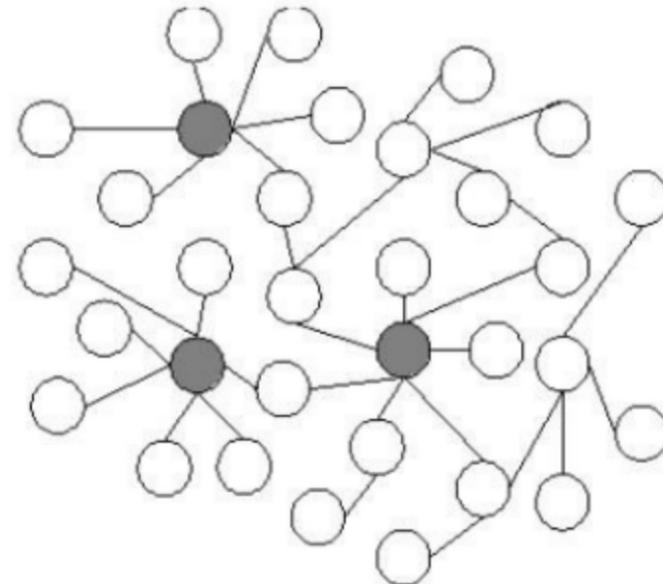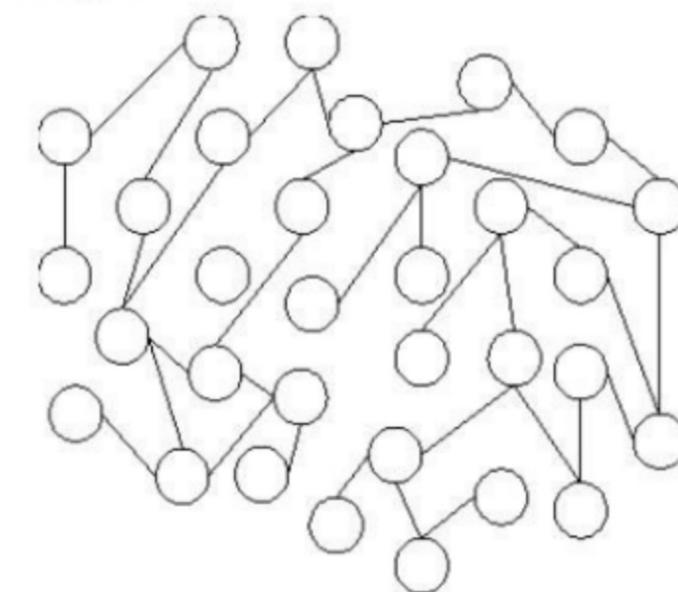
The small-world brain network (Watts&Strogazt)

The brain evolved to maximize efficiency and minimize wiring costs
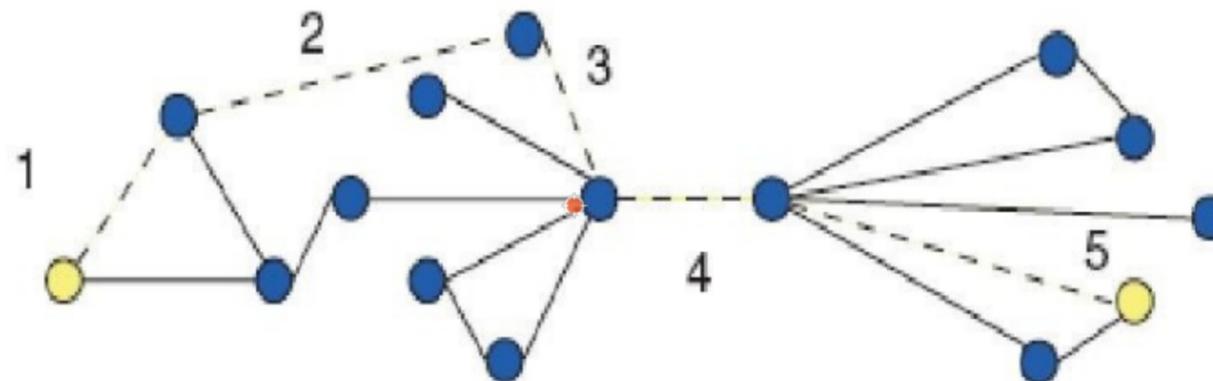
"Understanding principles of integration and segregation using whole-brain computational connectomics: implications for neuropsychiatric disorders" Lord et al. 2017, Philosophical transactions of the royal society

Random vs Chaotic and Complex

Shortest-path-length
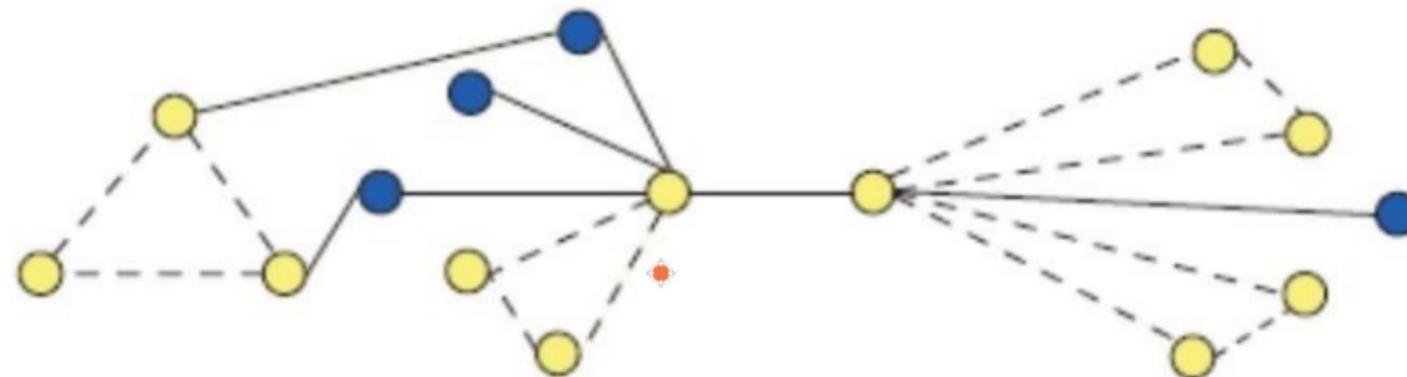
A, The path length between the two yellow nodes is defined as the fewest number of edges that must be traversed to get from one to the other. In this case, five edges must be followed, and therefore the path length between these two nodes is five.

Clustering

B, A high clustering coefficient means that if two nodes are both connected to a third node, then they are probably also connected to each other. The calculation of the clustering coefficient takes into account the number of connected triangles (shown here with yellow nodes and dashed edges).

Computing the small-worldness σ=γ/λ

Gamma (γ) measures the clustering coefficient of a network **compared to a random network** with the same number of nodes and edges. Clustering coefficient refers to how much nodes in a network tend to cluster together. A higher gamma (**γ** > 1) indicates a network where nodes are more clustered than expected in a random network, suggesting a higher degree of local connectivity. For example, in social networks, a high gamma means that friends of your friends are more likely to be friends with each other too.

Lambda (λ) is a metric that quantifies the average shortest path length in a network **compared to an equivalent random network**. Shortest path length is the fewest number of steps needed to go from one node to another. A smaller lambda (**λ** < 1) indicates that the average path length in the network is shorter than what's typically expected in a random network. This means there are relatively short distances between nodes, allowing for efficient communication across the network.

There are other metrics   (Rubinov&Sporns 2011)

| Global efficiency	Global efficiency of the network (Latora and Marchiori, 2001), $$E = \frac{1}{n}\sum_{i \in N} E_i = \frac{1}{n}\sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}^{-1}}{n - 1},$$ where E_i is the efficiency of node i.	Weighted global efficiency, $E^{\mathrm{w}} = \frac{1}{n}\sum_{i \in N} \frac{\sum_{j \in N, j \neq i}\left(d_{ij}^{\mathrm{w}}\right)^{-1}}{n - 1}$. Directed global efficiency, $E^{\rightarrow} = \frac{1}{n}\sum_{i \in N} \frac{\sum_{j \in N, j \neq i}\left(d_{ij}^{\rightarrow}\right)^{-1}}{n - 1}$.
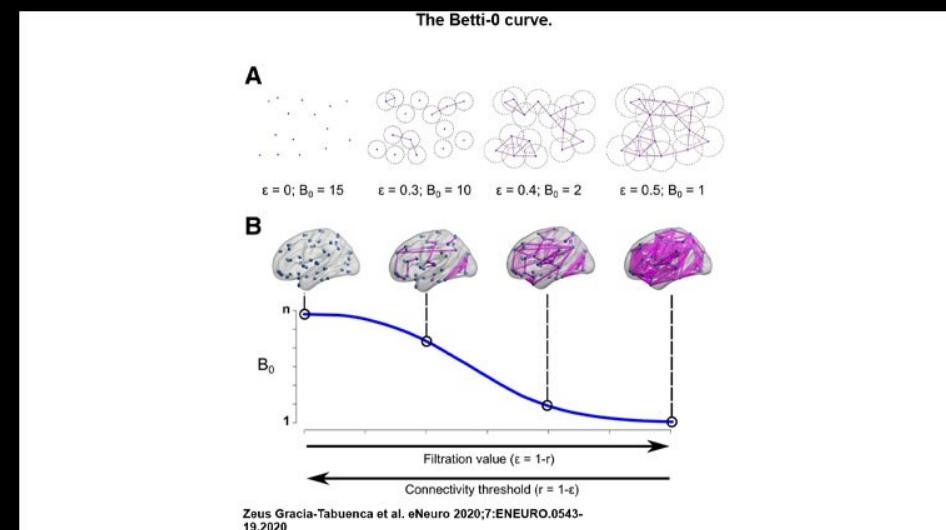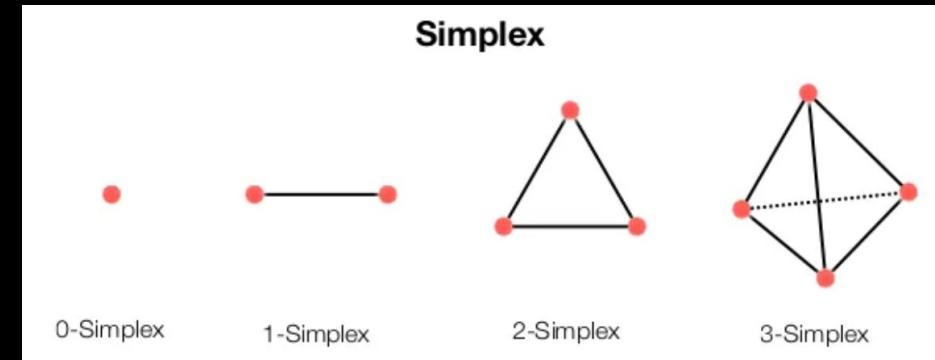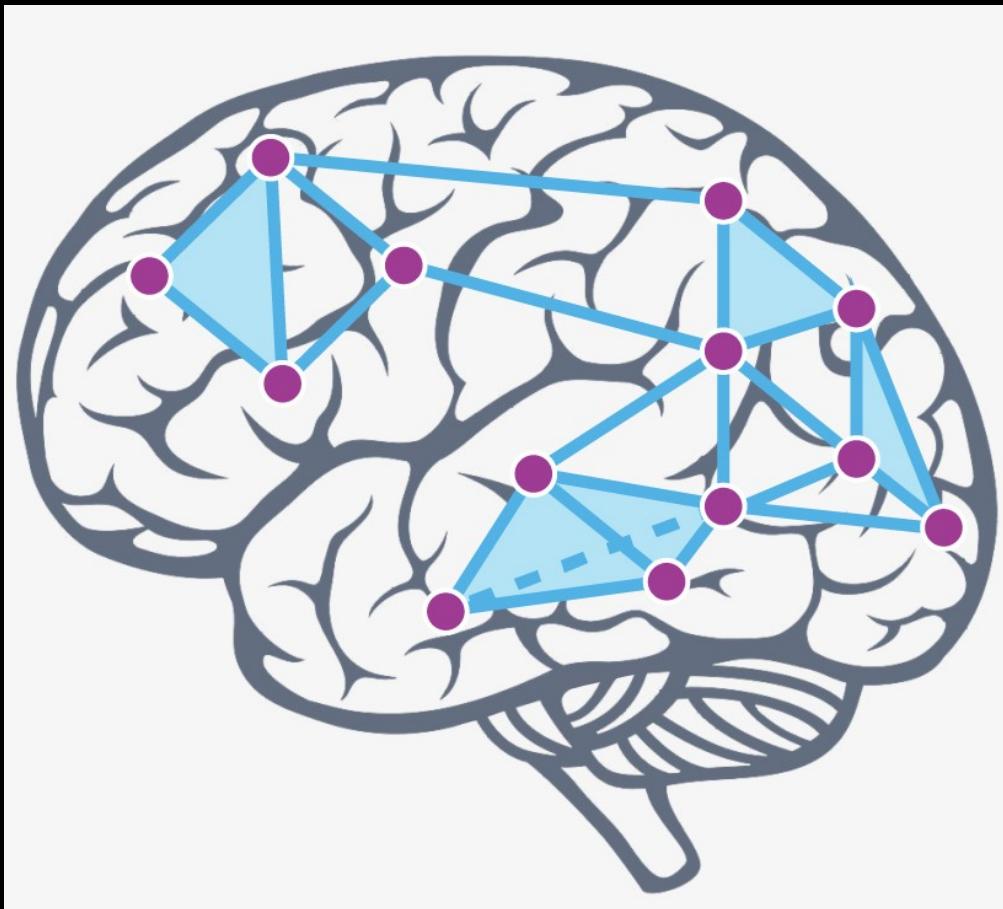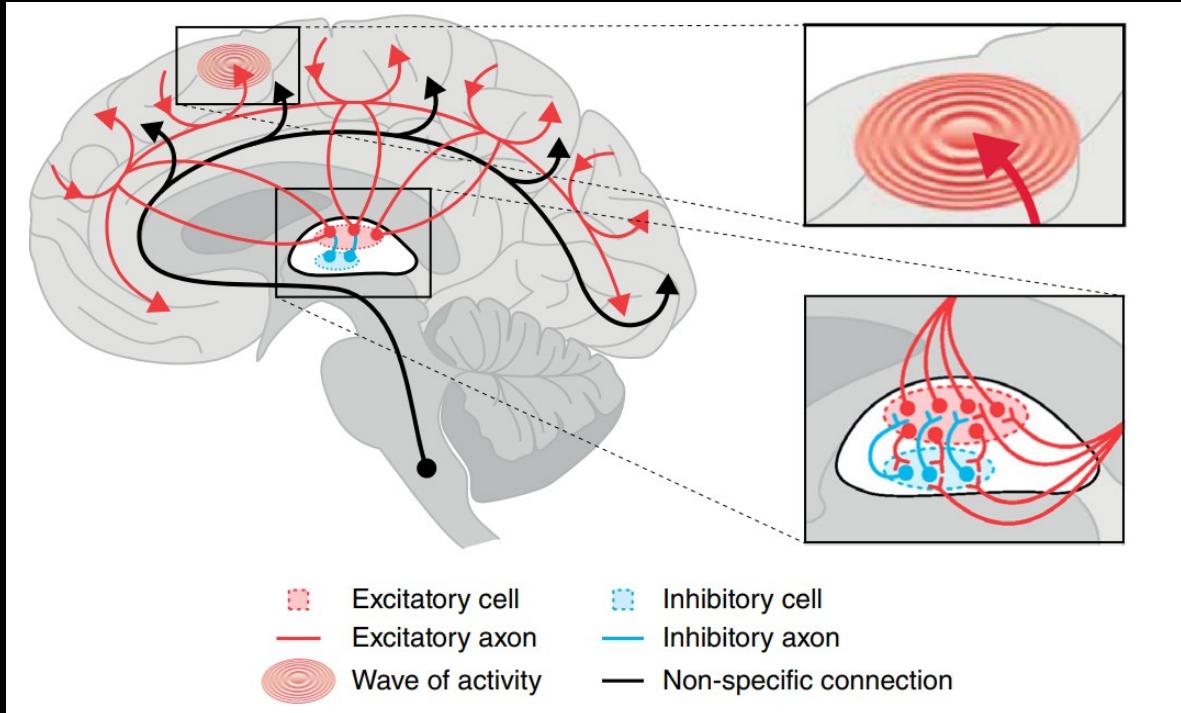
Measure	Binary and undirected definitions	Weighted and directed definitions
Measures of resilience Degree distribution	Cumulative degree distribution of the network (Barabasi and Albert, 1999), $$P(k) = \sum_{k' \geq k} p(k'),$$ where $p(k')$ is the probability of a node having degree k'.	Cumulative weighted degree distribution, $P(k^{\mathrm{w}}) = \sum_{k' \geq k^{\mathrm{w}}} p(k')$, Cumulative out-degree distribution, $P(k^{\mathrm{out}}) = \sum_{k' \geq k^{\mathrm{out}}} p(k')$. Cumulative in-degree distribution, $P(k^{\mathrm{in}}) = \sum_{k' \geq k^{\mathrm{in}}} p(k')$.
Average neighbor degree	Average degree of neighbors of node i (Pastor-Satorras et al., 2001), $$k_{\mathrm{nn},i} = \frac{\sum_{j \in N} a_{ij} k_j}{k_i}.$$	Average weighted neighbor degree (modified from Barrat et al., 2004), $$k_{\mathrm{nn},i}^{\mathrm{w}} = \frac{\sum_{j \in N} w_{ij} k_j^{\mathrm{w}}}{k_i^{\mathrm{w}}}.$$ Average directed neighbor degree*, $$k_{\mathrm{nn},i}^{\rightarrow} = \frac{\sum_{j \in N}\left(a_{ij} + a_{ji}\right)\left(k_i^{\mathrm{out}} + k_i^{\mathrm{in}}\right)}{2\left(k_i^{\mathrm{out}} + k_i^{\mathrm{in}}\right)}.$$

Topological data analysis

Simplex

0-Simplex   1-Simplex   2-Simplex   3-Simplex

The Betti-0 curve.

A

$\varepsilon = 0; B_0 = 15$ $\varepsilon = 0.3; B_0 = 10$ $\varepsilon = 0.4; B_0 = 2$ $\varepsilon = 0.5; B_0 = 1$

B

n

B_0

1

Filtration value ($\varepsilon = 1-r$)

Connectivity threshold ($r = 1-\varepsilon$)

Zeus Gracia-Tabuenca et al. eNeuro 2020;7:ENEURO.0543-19.2020
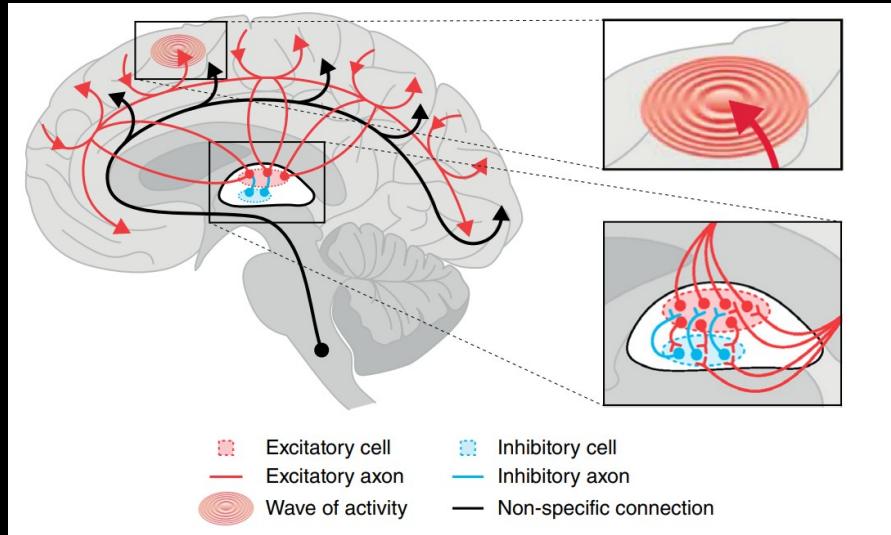
And there are other things than structure and function

Excitatory cell
Excitatory axon
Wave of activity

Inhibitory cell
Inhibitory axon
Non-specific connection

- Receptors (e.g. GABA)

- Hormones

- other things Like "weird" proteins
-etc
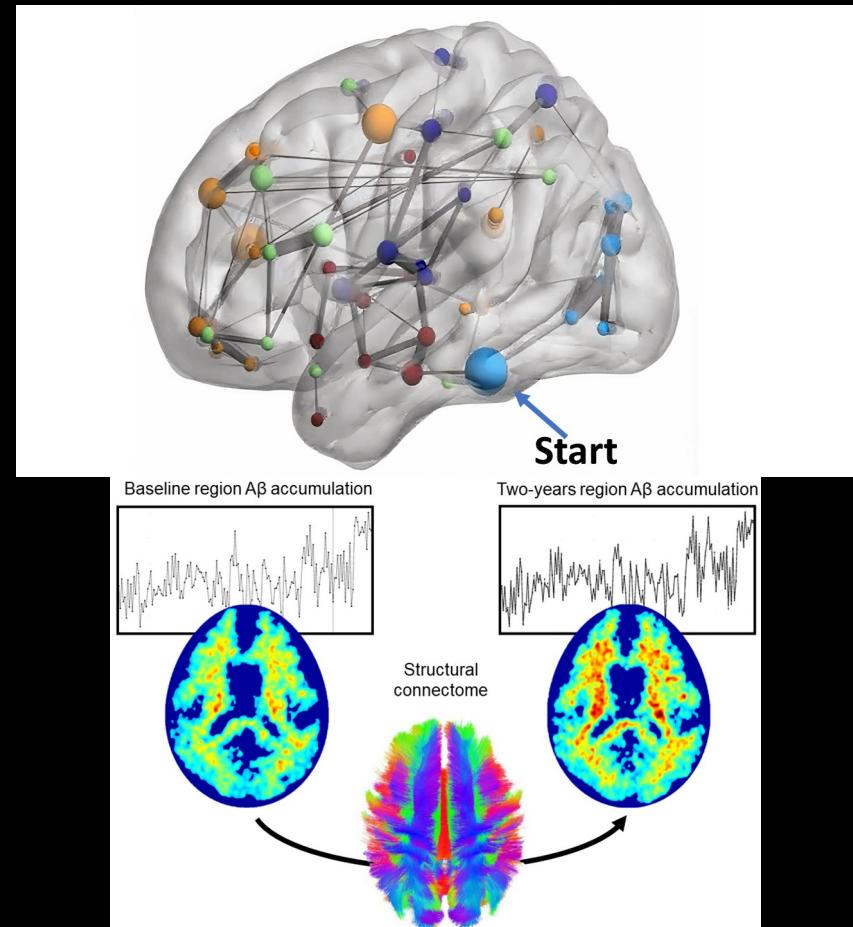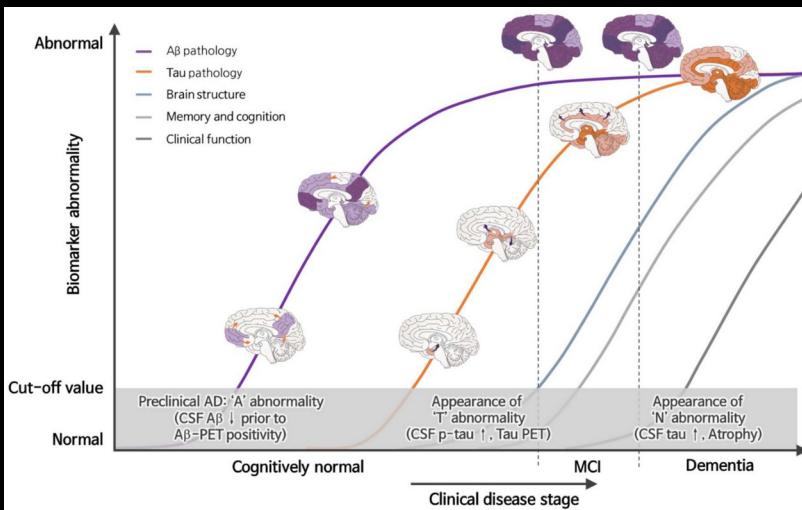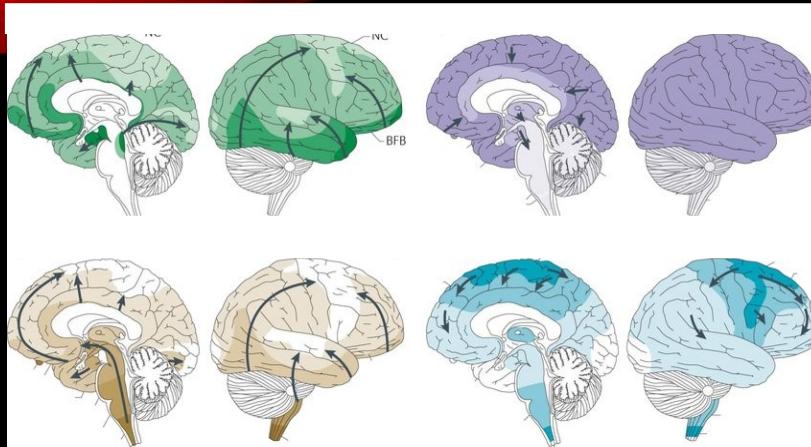
The brain is a dynamic system

Excitatory cell   Inhibitory cell
Excitatory axon   Inhibitory axon
Wave of activity   Non-specific connection

Systems theory is the transdisciplinary study of systems, i.e. cohesive groups of interrelated, interdependent components that can be natural or artificial.

Every system has causal boundaries, is influenced by its context, defined by its structure, function and role, and expressed through its relations with other systems. A system is "more than the sum of its parts" by expressing synergy or emergent behavior.
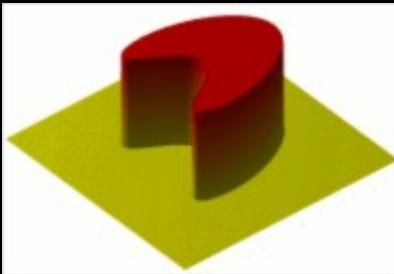
Misfolded Protein spreading

Misfolded Protein spreading

Heat equation

$$\dot{u} = \Delta u.$$
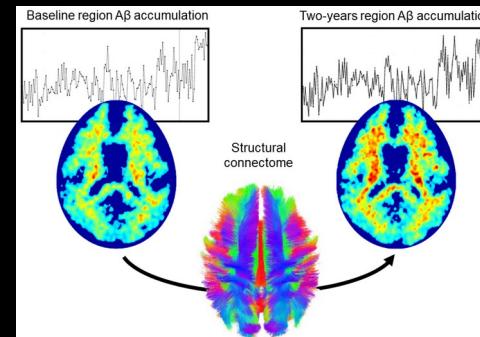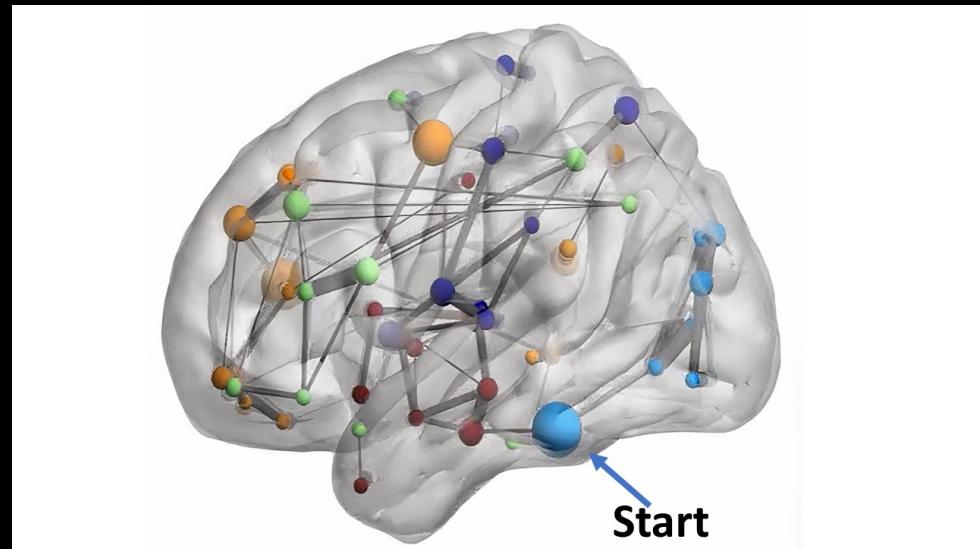
Solution for finite elements
(the nodes of the connectome)
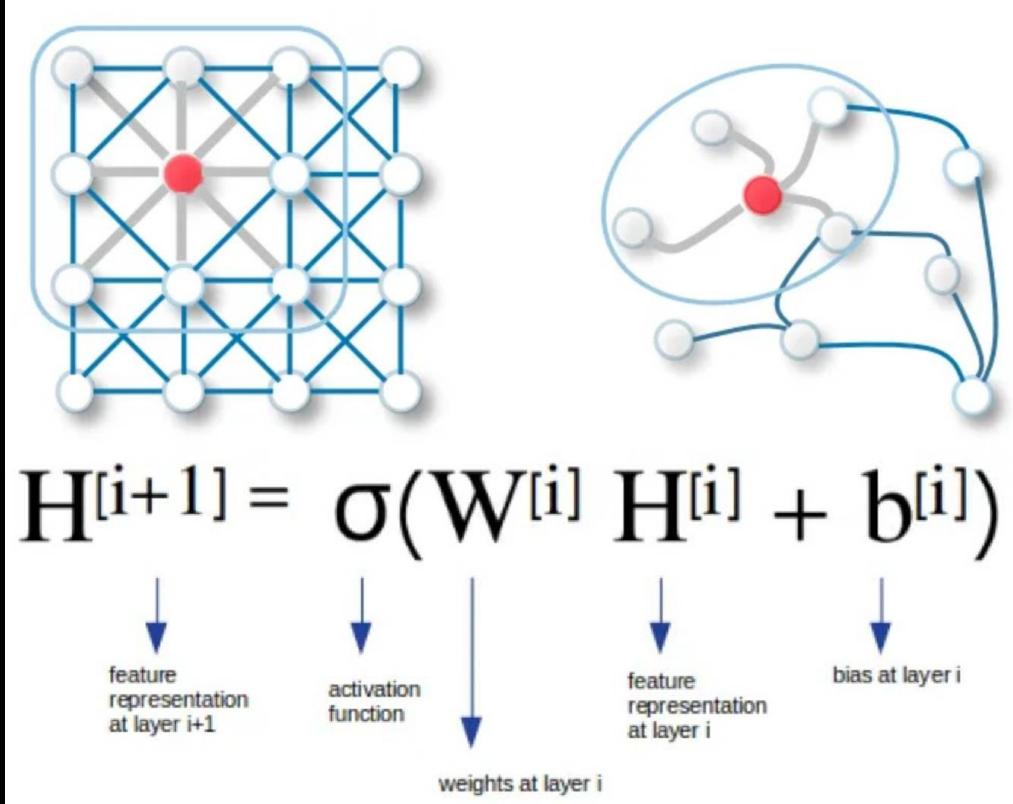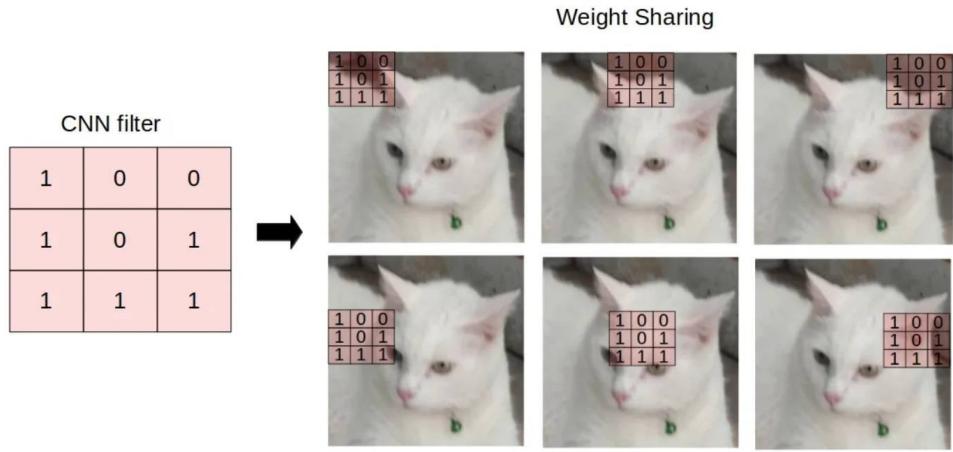
Network Dynamics Model
(Raj et al. Neuron 2012)

$$\vec{y}(t) = U e^{-\vec{\lambda}\beta t} U^{-1} \vec{y}(0)$$

Start

Baseline region Aβ accumulation

Two-years region Aβ accumulation

Structural connectome

Introduction to GNN (message passing NN)

Weight Sharing

CNN filter

- CNN are generally use data on an Euclidean space
- GNN can have irregular connections (and data are on a non-Euclidean)
- Graph convolutional layers aim to capture local neighborhood information for each node in the graph.
- **message passing:** a general nonlinear function dependent on the features of two nodes sharing an edge.

$$H^{[i+1]} = \sigma(W^{[i]} H^{[i]} + b^{[i]})$$

feature representation at layer i+1   activation function   feature representation at layer i   bias at layer i

weights at layer i

$$H^{[i+1]} = \sigma(W^{[i]} H^{[i]} A^{*})$$

Introduction to GNN (message passing NN)

Graph Convolutional Networks, Kipf and Welling [2016]

$$\mathbf{h}_v^{(k)} = \sigma\left(\mathbf{W}^{(k)} \underbrace{\sum_{v \in \mathcal{N}(u) \cup \{u\}}}_{} \frac{\mathbf{h}_v}{\sqrt{|\mathcal{N}(u)||\mathcal{N}(v)|}}\right)$$

Self-loop

Sum of normalized neighbor embeddings

Multi-Layer-Perceptron as Aggregator, Zaheer et al. [2017]

$$\mathbf{m}_{\mathcal{N}(u)} = \boxed{\text{MLP}_\theta}\left(\sum_{v \in N(u)} \text{MLP}_\phi(\mathbf{h}_v)\right)$$

Aggregated message

trainable!

Send states through a MLP

Graph Attention Networks, Veličković et al. [2017]

$$\mathbf{m}_{\mathcal{N}(u)} = \sum_{v \in \mathcal{N}(u)} \alpha_{u,v} \mathbf{h}_v \qquad \alpha_{u,v} = \frac{\exp\left(\mathbf{a}^\top[\mathbf{W}\mathbf{h}_u \oplus \mathbf{W}\mathbf{h}_v]\right)}{\sum_{v' \in \mathcal{N}(u)} \exp\left(\mathbf{a}^\top[\mathbf{W}\mathbf{h}_u \oplus \mathbf{W}\mathbf{h}_{v'}]\right)}$$

Attention weights

Gated Graph Neural Networks, Li et al. [2015]

$$\mathbf{h}_u^{(k)} = \text{GRU}(\mathbf{h}_u^{(k-1)}, \mathbf{m}_{\mathcal{N}(u)}^{(k)})$$

Recurrent update of the state

Misfolded Protein spreading

$$\mathbf{y}(t) = \sum_{i=1}^{n} \mathbf{A}_i \mathbf{y}(t - i) + \epsilon$$
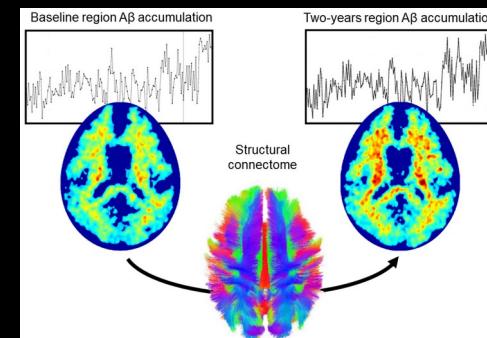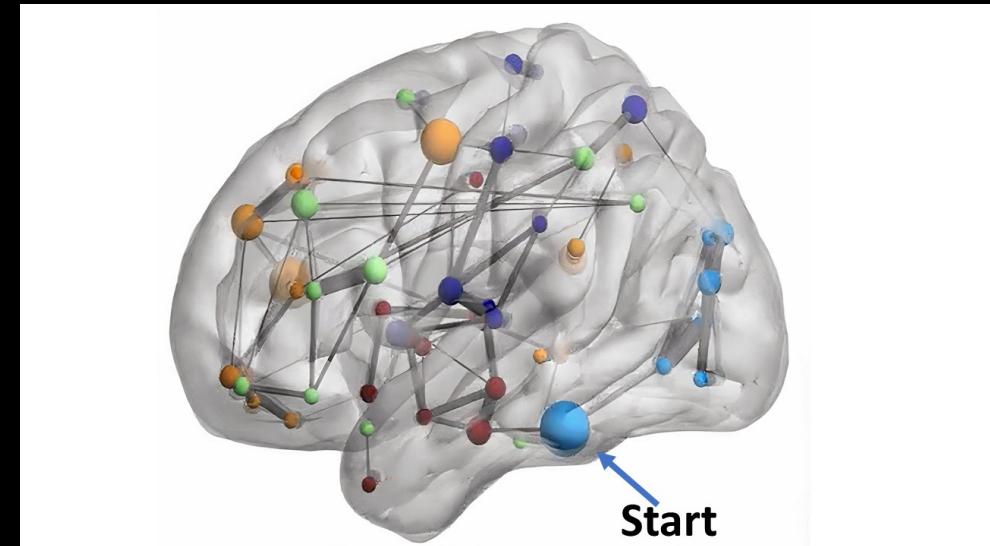
1. Multivariate Autoregressive model (old school ML / statistics)

$$H^{l+1} = \sigma(D^{-1/2} A D^{-1/2} H^l W^l + \vec{b}),$$
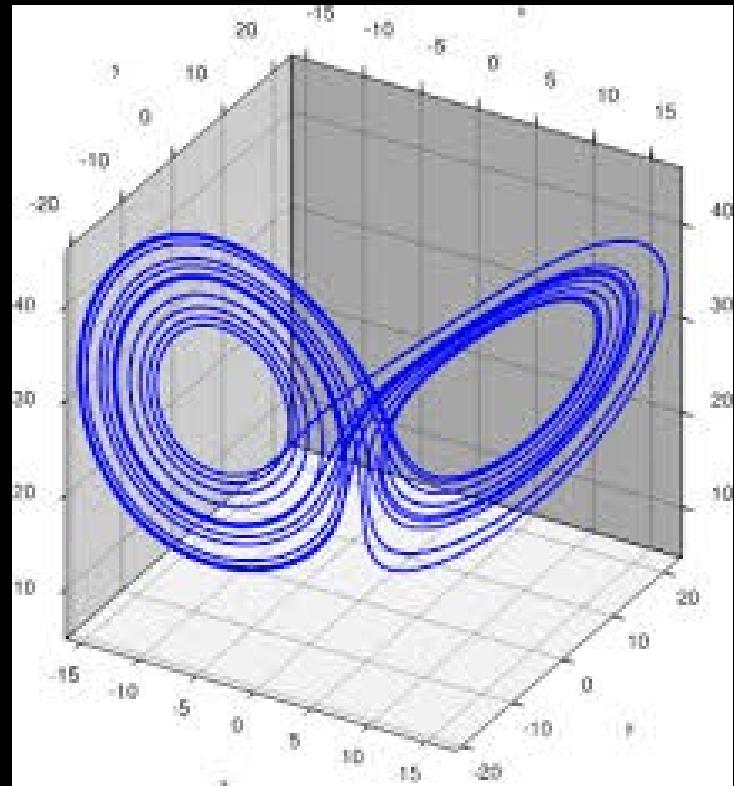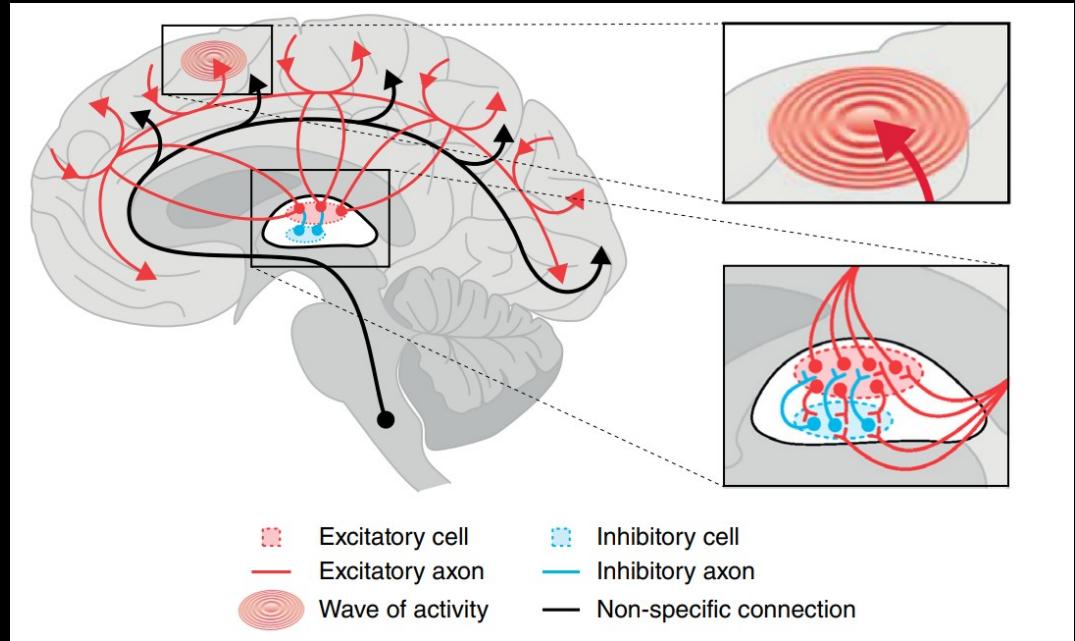
2. Graph convolutional Neural networks
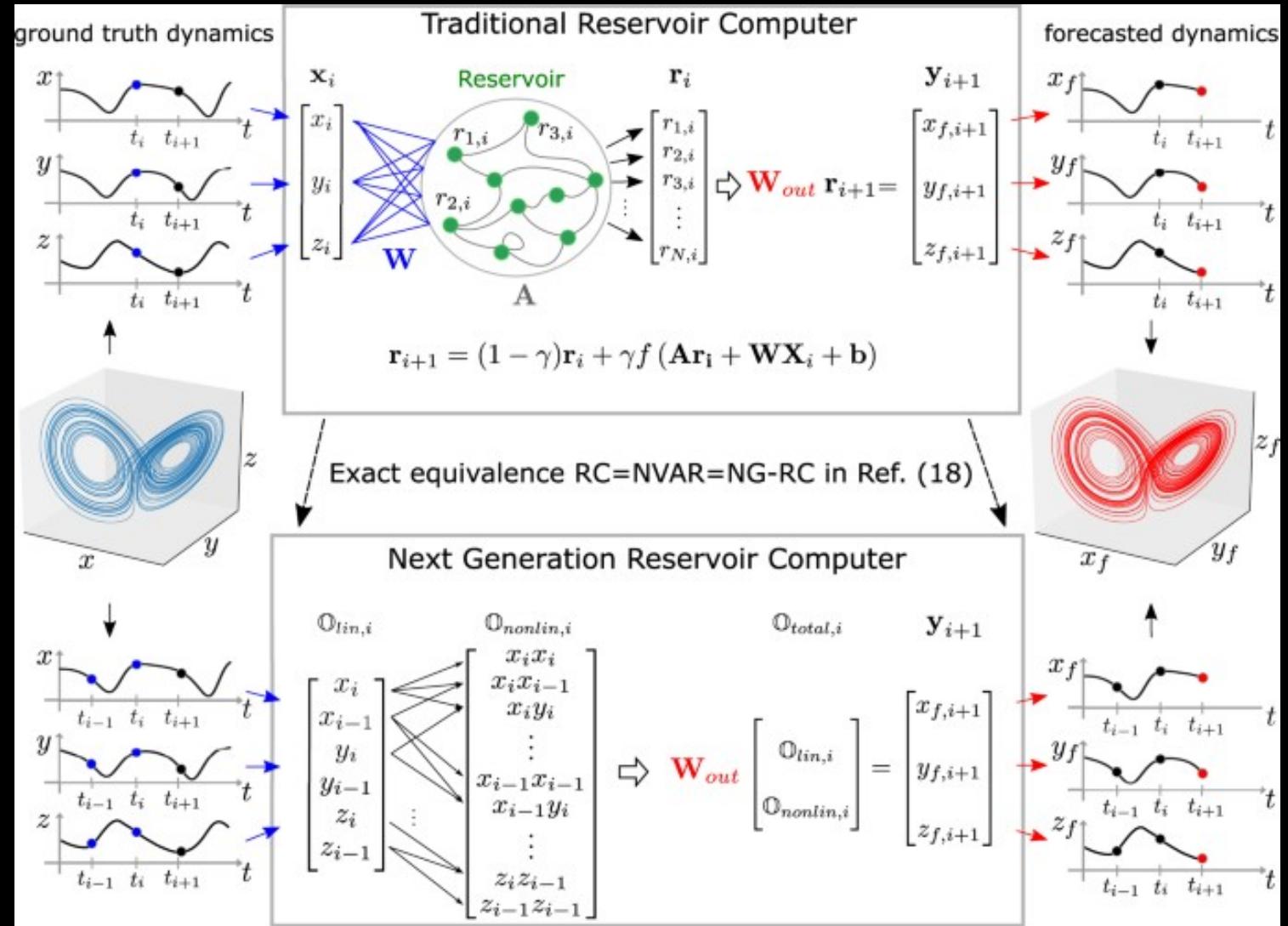
$$\vec{y}(t) = U e^{-\vec{\lambda}\beta t} U^{-1} \vec{y}(0)$$

Network Dynamics Model
(Raj et al. Neuron 2012)

Start

Baseline region Aβ accumulation   Two-years region Aβ accumulation

Structural connectome

Chaotic system theory

ground truth dynamics

Traditional Reservoir Computer

forecasted dynamics

$\mathbf{x}_i$

$\begin{bmatrix} x_i \\ \\ y_i \\ \\ z_i \end{bmatrix}$

Reservoir

$r_{1,i}$ $r_{3,i}$

$r_{2,i}$

$\mathbf{W}$

$\mathbf{A}$

$\mathbf{r}_i$

$\begin{bmatrix} r_{1,i} \\ r_{2,i} \\ r_{3,i} \\ \vdots \\ r_{N,i} \end{bmatrix}$

$\mathbf{W}_{out} \, \mathbf{r}_{i+1} =$

$\mathbf{y}_{i+1}$

$\begin{bmatrix} x_{f,i+1} \\ \\ y_{f,i+1} \\ \\ z_{f,i+1} \end{bmatrix}$

$$\mathbf{r}_{i+1} = (1 - \gamma)\mathbf{r}_i + \gamma f\left(\mathbf{A}\mathbf{r_i} + \mathbf{W}\mathbf{X}_i + \mathbf{b}\right)$$

Exact equivalence RC=NVAR=NG-RC in Ref. (18)

Next Generation Reservoir Computer

$\mathbb{O}_{lin,i}$

$\begin{bmatrix} x_i \\ x_{i-1} \\ y_i \\ y_{i-1} \\ z_i \\ z_{i-1} \end{bmatrix}$

$\mathbb{O}_{nonlin,i}$

$\begin{bmatrix} x_i x_i \\ x_i x_{i-1} \\ x_i y_i \\ \vdots \\ x_{i-1} x_{i-1} \\ x_{i-1} y_i \\ \vdots \\ z_i z_{i-1} \\ z_{i-1} z_{i-1} \end{bmatrix}$

$\mathbb{O}_{total,i}$

$\mathbf{W}_{out} \begin{bmatrix} \mathbb{O}_{lin,i} \\ \\ \mathbb{O}_{nonlin,i} \end{bmatrix} =$

$\mathbf{y}_{i+1}$

$\begin{bmatrix} x_{f,i+1} \\ \\ y_{f,i+1} \\ \\ z_{f,i+1} \end{bmatrix}$

Summary

- We can summarize graph networks to individual scalars
  and use them as features for further machine learning
  tools
- The brain is a complex chaotic system
- There are many phenomena which can be either simulated
  with PDEs or predicted with machine learning
- Dynamic chaotic system generally converges to
  "attractors", RNN and RCC are suitable to model this.

Questions?