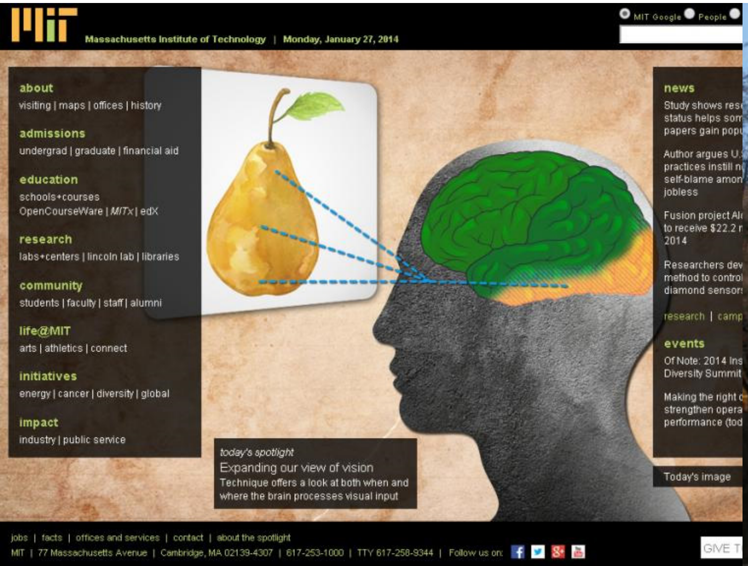
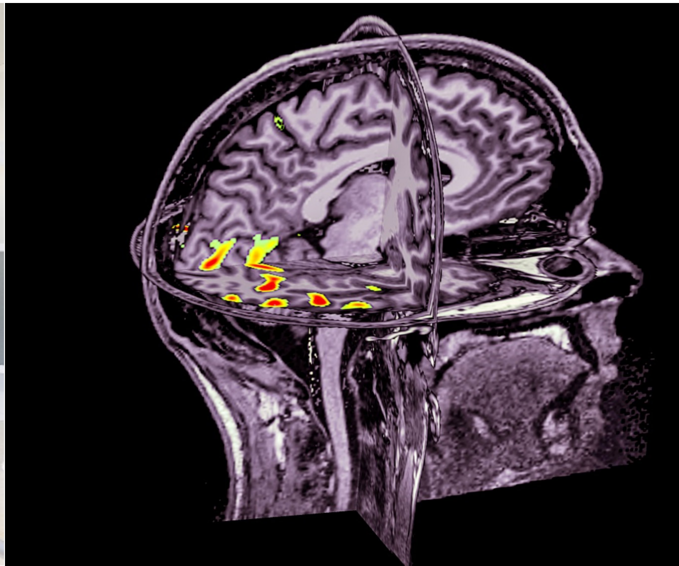
CuttingEEG - Los Angeles Garden
October 19, 2023

# Graph representation learning of MEG signals opens a window to aging trajectories and Alzheimer's disease


Dimitrios Pantazis
http://meglab.mit.edu
Principal Research Scientist
Director, Magnetoencephalography Laboratory
Massachusetts Institute of Technology

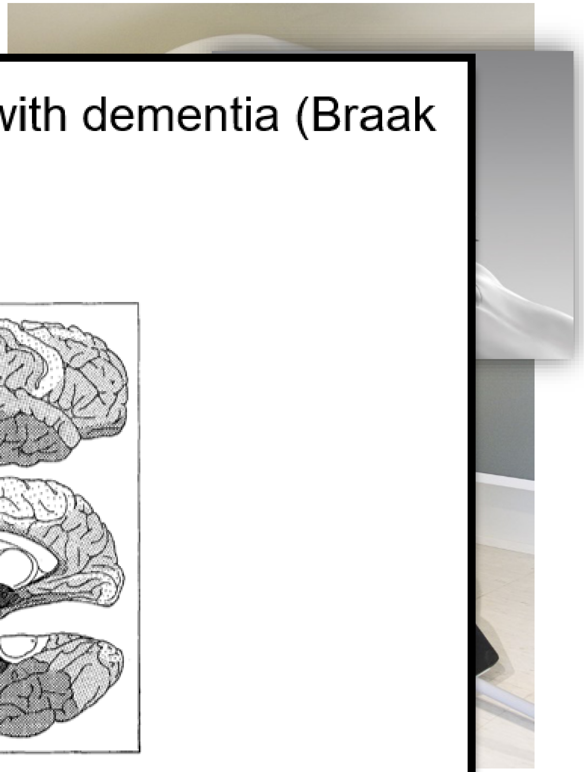



# Alzheimer's disease pathology

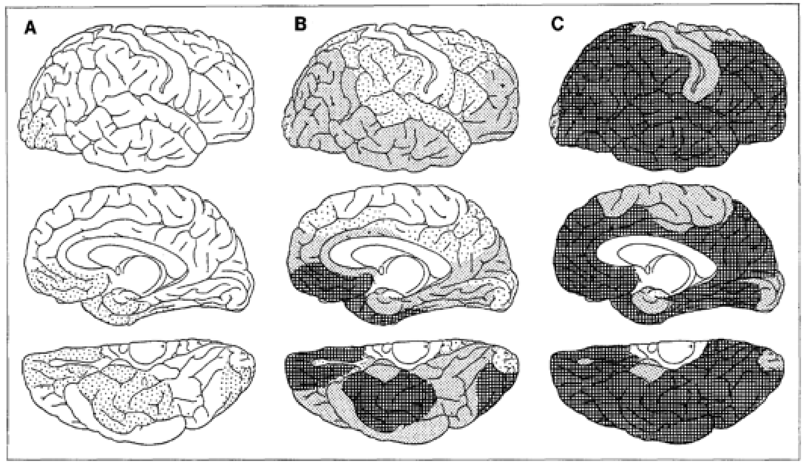
**Alzheimer's disease (AD)** is a brain network disease (connectopathy) with complex etiology and multiple pathogeneses.

**Magnetoencephalography (MEG)**



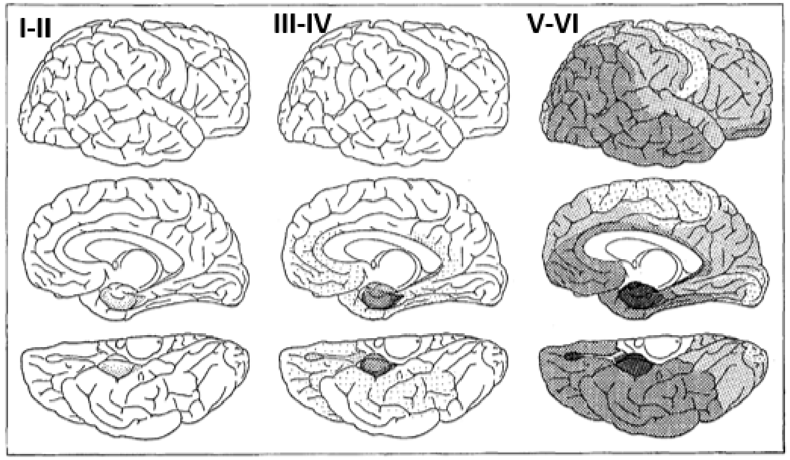
Neuropathological staging of AD with autopsy of 83 brains of individuals with dementia (Braak and Braak, 1991)

**Amyloid deposits**




Stage A: basal portions of the cortex
Stage B: all cortical association areas
Stage C: all areas of the cortex

**Tau deposits**




Stage I-II: transentorhinal region
Stage III-IV: severe involvement of entorhinal and transentorhinal
Stage V-VI: cortical destruction

Genetic mutation and risk factors

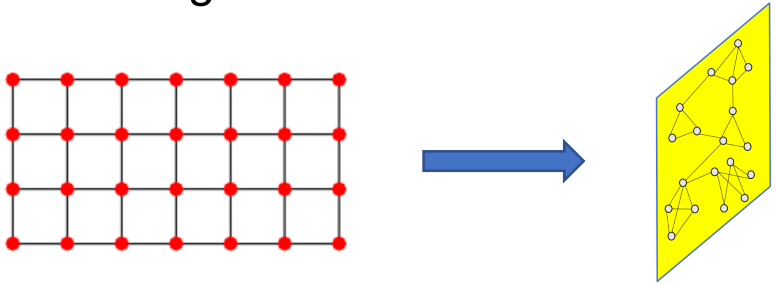
Misfolding of Aβ, and tau followed by plaques and tangles

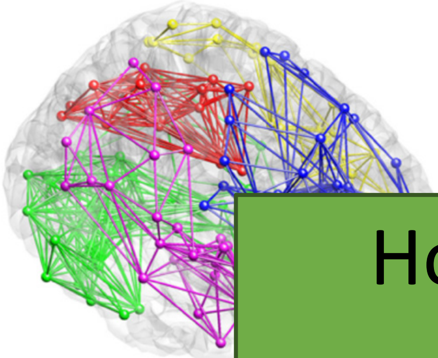
Pathogenic neuronal activity

autopsy

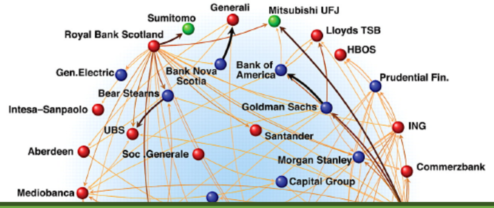
# Geometric (graph) deep learning

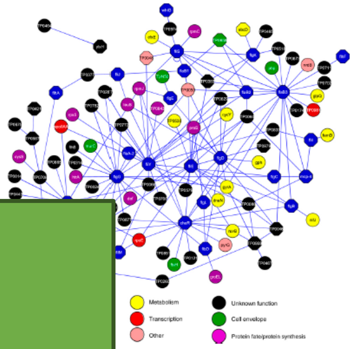
Increasingly more fields have to deal with geometric non-Euclidean structured data such as graphs.





Human bra

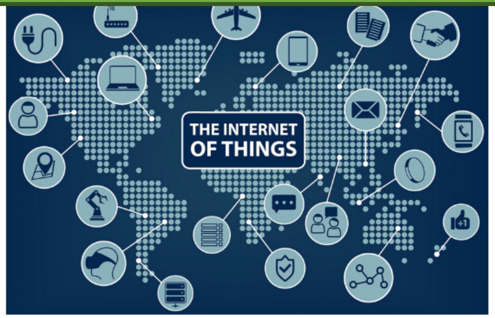


n networks of
rial motility

How do we take advantage of the relational structure for better prediction?

Social networks

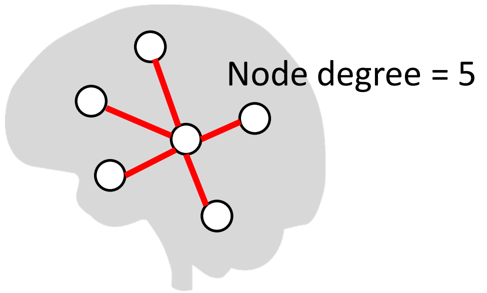



Internet of things

Boston MBTA traffic network

# Graph theory

Prior studies primarily focused on handcrafted, domain-specific (ad-hoc) graph topological properties of brain networks constructed by MEG (node degree, node centrality, clustering coefficient,…).




Box 2 | **Network measures**

A network is defined in graph theory as a set of nodes or vertices and the edges or lines between them. Graph topology can be quantitatively described by a wide variety of measures, some of which are discussed here. It is not yet established which measures are most appropriate for the analysis of brain networks. The figure shows a schematic diagram of a brain network drawn as a directed (left) and an undirected (right) graph; both structural and functional networks can be either directed or undirected (BOX 1).

**Node degree, degree distribution and assortativity**

The degree of a node is the number of connections that link it to the rest of the network — this is the most fundamental network measure and most other measures are ultimately linked to node degree. The degrees of all the network's nodes form a degree distribution[15]. In random networks all connections are equally probable, resulting in a Gaussian and symmetrically centred degree distribution. Complex networks generally have non-Gaussian degree distributions, often with a long tail towards high degrees. The degree distributions of scale-free networks follow a power law[90]. Assortativity is the correlation between the degrees of connected nodes. Positive assortativity indicates that high-degree nodes tend to connect to each other.

**Clustering coefficient and motifs**

If the nearest neighbours of a node are also directly connected to each other they form a cluster. The clustering coefficient quantifies the number of connections that exist between the nearest neighbours of a node as a proportion of the maximum number of possible connections[18]. Random networks have low average clustering whereas complex networks have high clustering (associated with high local efficiency of information transfer and robustness). Interactions between neighbouring nodes can also be quantified by counting the occurrence of small motifs of interconnected nodes[150]. The distribution of different motif classes in a network provides information about the types of local interactions that the network can support[48].

**Path length and efficiency**

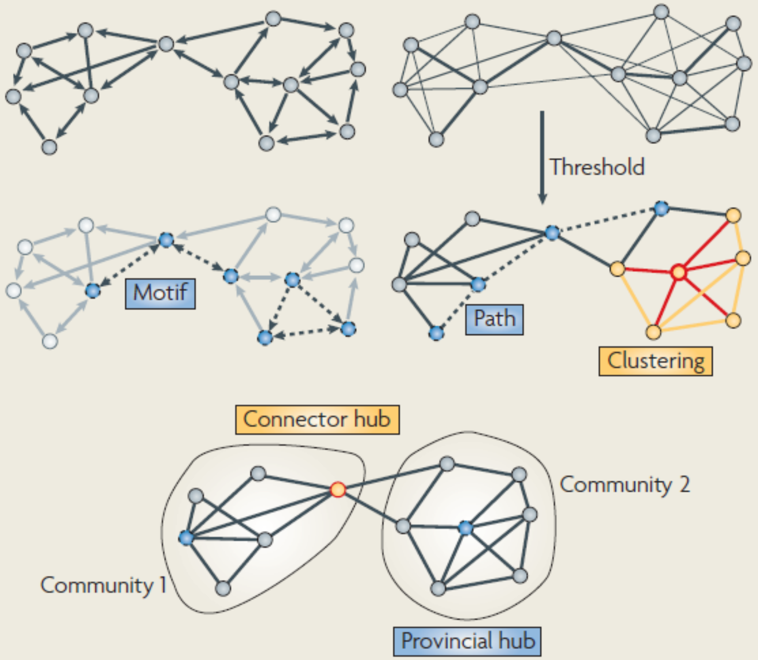
Path length is the minimum number of edges that must be traversed to go from one node to another. Random and complex networks have short mean path lengths (high global efficiency of parallel information transfer) whereas regular lattices have long mean path lengths. Efficiency is inversely related to path length but is numerically easier to use to estimate topological distances between elements of disconnected graphs.

**Hubs, centrality and robustness**

Hubs are nodes with high degree, or high centrality. The centrality of a node measures how many of the shortest paths between all other node pairs in the network pass through it. A node with high centrality is thus crucial to efficient communication[151]. The importance of an individual node to network efficiency can be assessed by deleting it and estimating the efficiency of the 'lesioned' network. Robustness refers either to the structural integrity of the network following deletion of nodes or edges or to the effects of perturbations on local or global network states.

**Modularity**

Many complex networks consist of a number of modules. There are various algorithms that estimate the modularity of a network, many of them based on hierarchical clustering[23]. Each module contains several densely interconnected nodes, and there are relatively few connections between nodes in different modules. Hubs can therefore be described in terms of their roles in this community structure[27]. Provincial hubs are connected mainly to nodes in their own modules, whereas connector hubs are connected to nodes in other modules.



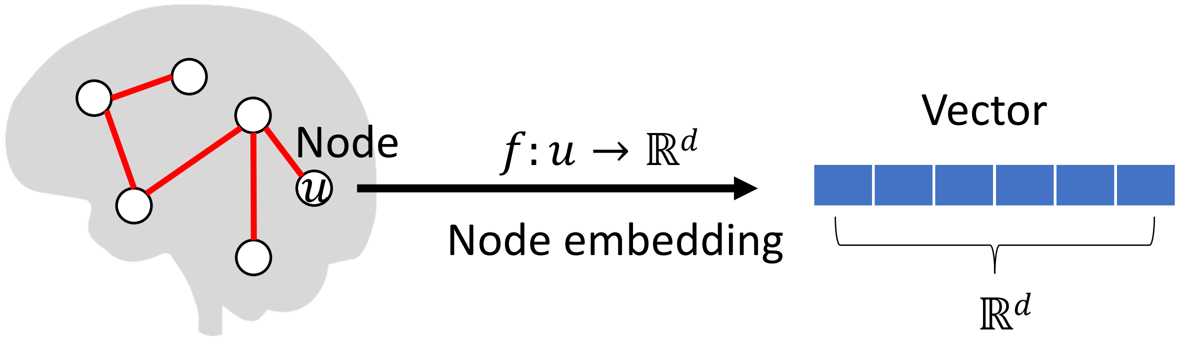

E. Bullmore and O. Sporns, *Nature Reviews Neuroscience*, 2009

# Graph representation learning

Inductive and automatic network representation learning from raw MEG functional imaging data remains an open problem.

Recently, emerging graph embedding techniques (e.g., deepWalk, node2vec, Graph2Gauss, etc.) enabled automatic learning of hierarchical, heterogeneous and latent network representations from original complex and high-dimensional graphs in irregular domains.

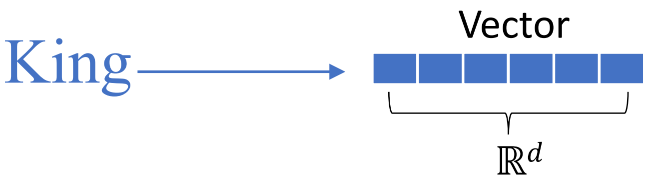
## Vector embedding




(e.g., deepWalk, node2vec)

# NLP Example: Word embeddings

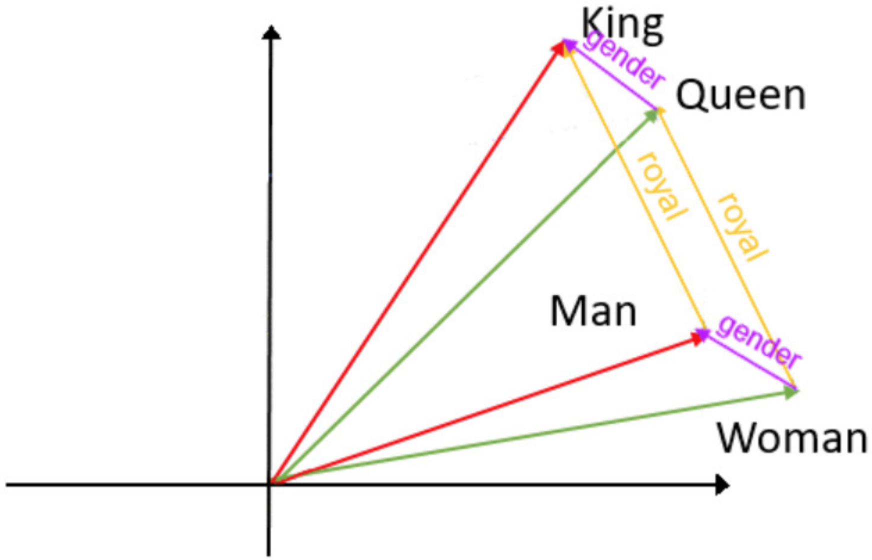
**Natural Language Processing:** For a computer to perform any "reasoning“ with words, they need to be represented numerically as vectors of numbers termed "embeddings".




Intuitively, if words are similar in some respect, this can be reflected by certain values in their embeddings being similar.

In recent years, astonishingly successful algorithms have been developed, that learn word embeddings by extracting information from huge text sources such as Wikipedia. State of the art technology: chatGPT.

**King - man + woman = ?  Queen**




**But… unfortunately there are machine learning biases (gender, racial biases, etc)**

Doctor – man + woman = ?  Nurse

Programmer – man + woman = ?  Homemaker

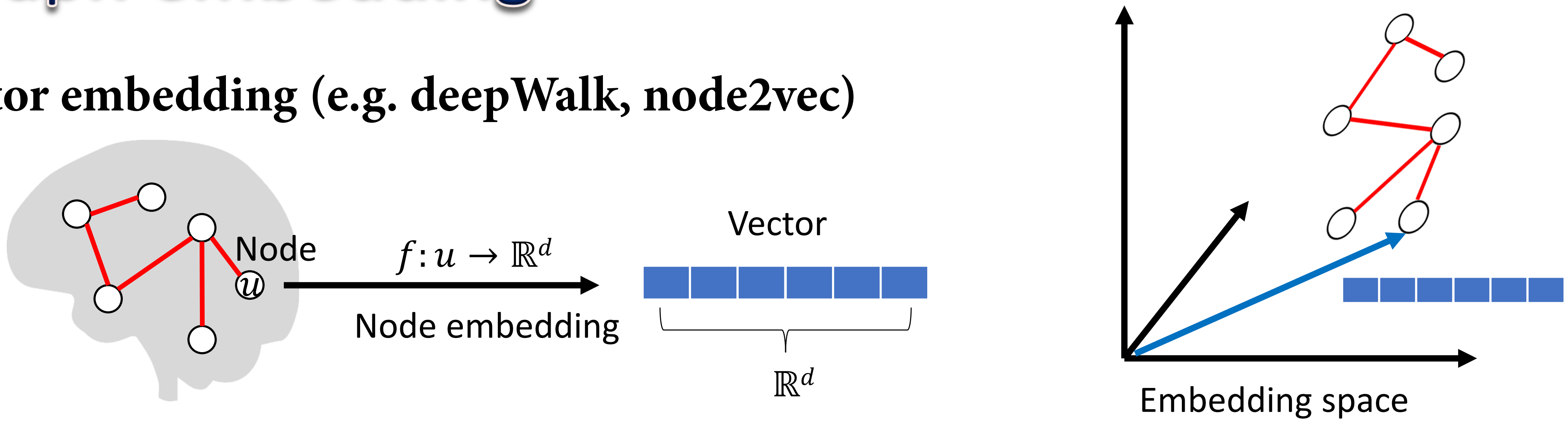
Neighbors of *man* and *woman* (K-nearest embedding)



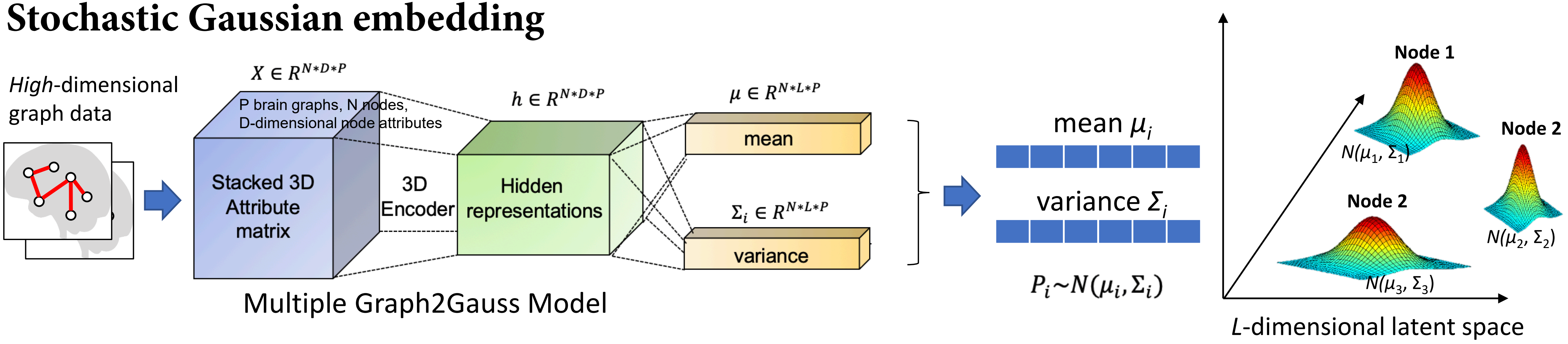
https://towardsdatascience.com/gender-bias-word-embeddings-76d9806a0e17

# Graph embedding

## Vector embedding (e.g. deepWalk, node2vec)




## Stochastic Gaussian embedding





Bojchevski, A., & Günnemann, S. (2017). Deep Gaussian Embedding of Graphs: Unsupervised Inductive Learning via Ranking. *ArXiv:1707.03815 [Cs, Stat]*.

Xu, M., Wang, Z., Zhang, H., Pantazis, D., Wang, H., & Li, Q. *PLOS Computational Biology* (2020). Gaussian embedding-based functional brain connectomic analysis for amnestic mild cognitive impairment patients with cognitive training.

Xu, M., Sanz, D.L., Garces, P., Maestu, F., Li, Q., and Pantazis, D. (2021). A Graph Gaussian Embedding Method for Predicting Alzheimer's Disease Progression With MEG Brain Networks. IEEE Transactions on Biomedical Engineering *68*, 10.

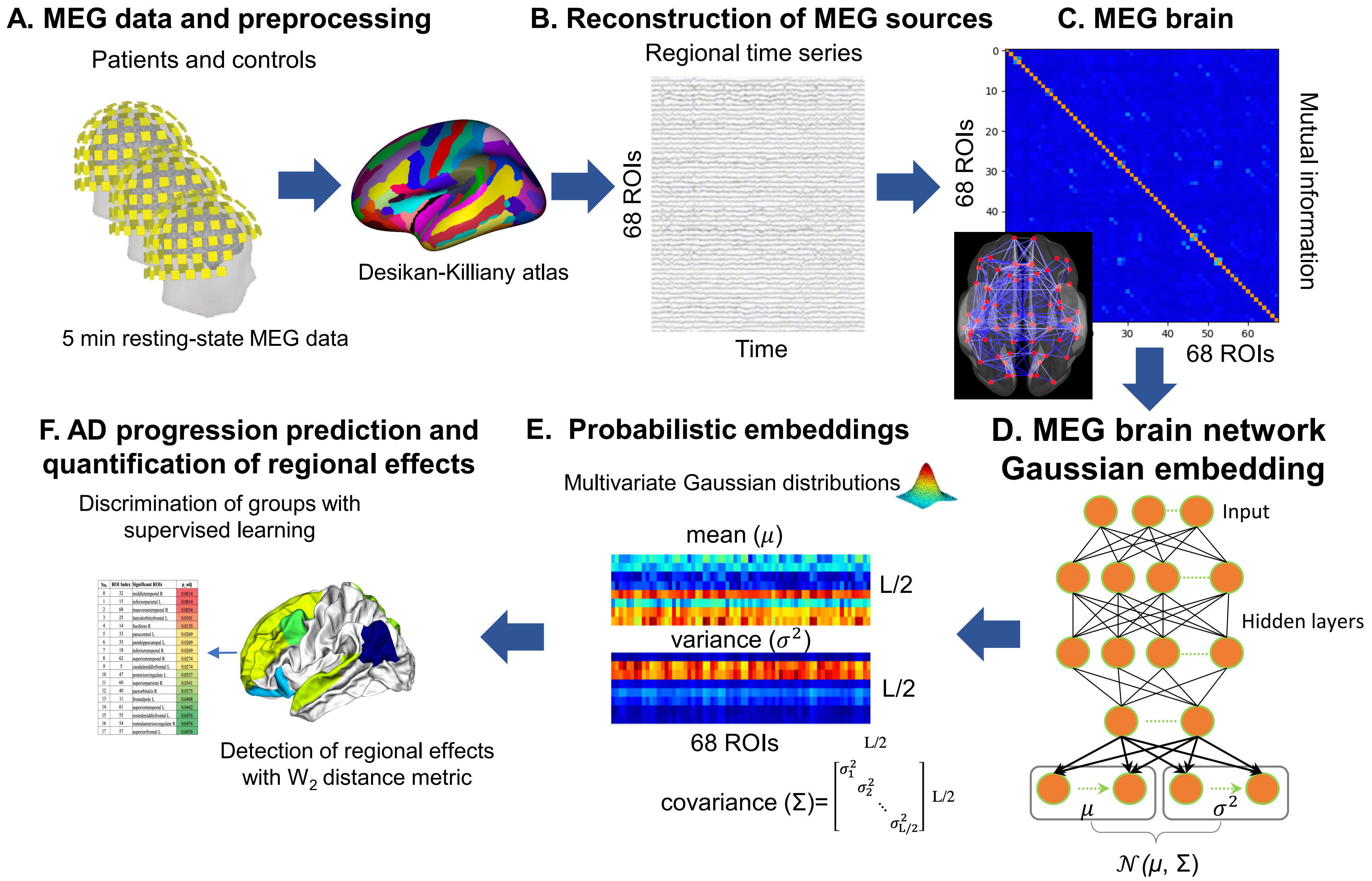
# Predicting AD using MEG brain networks

**Participants:**

- 76 mild cognitive impairment (MCI) patients (48 stable and 28 progressive)
- 53 age-matched clinically normal subjects (all from the *Madrid cohort* database).

**MEG data:**

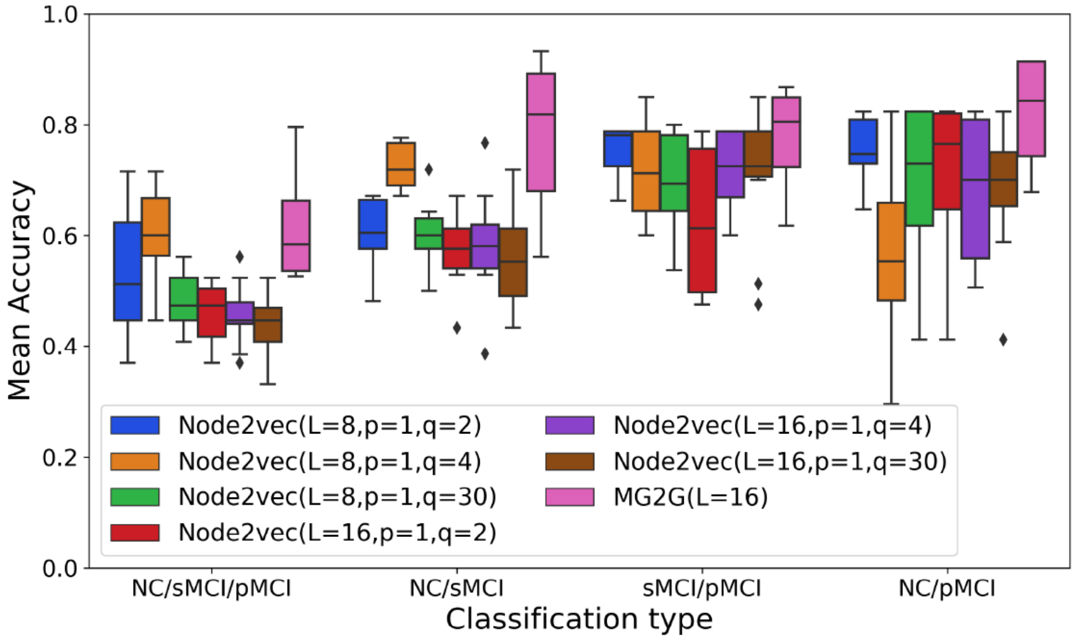
- 5 minutes of eyes-closed resting-state data in a 306-channel MEG system.




Xu et al., 2021

# AD early detection and affected regions

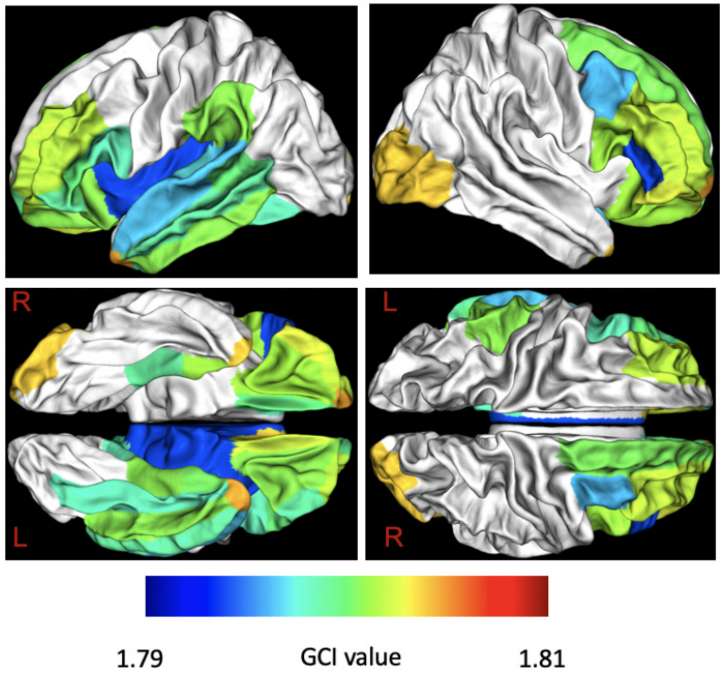
**Supervised classification**




Discriminate: Normal controls (NC),
Stable mild cognitively impairment (sMCI)
Progressive mild cognitively impairment (pMCI)

**79%** NC/sMCI
**78%** sMCI/ pMCI
**82%** NC/pMCI
**61%** NC/sMCI/pMCI

**Brain regions with significant MEG network alterations**

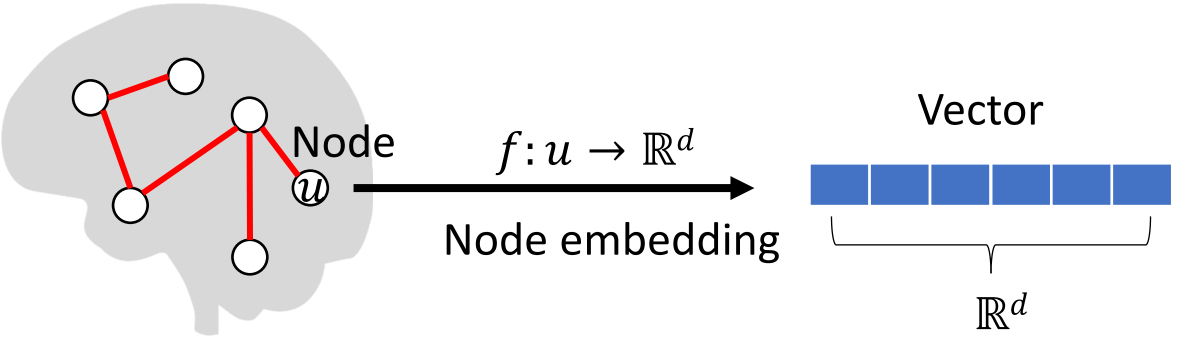


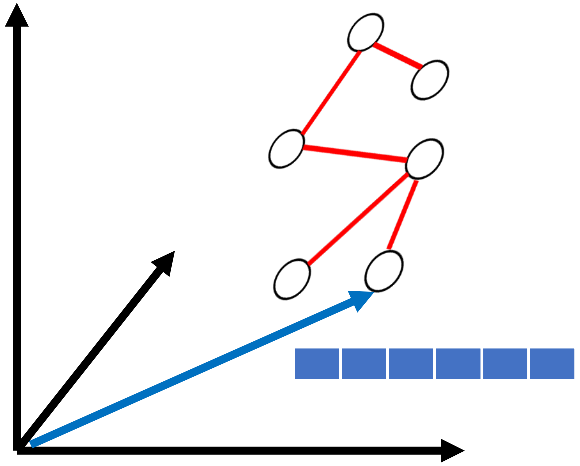

Regions largely include temporal and frontal regions in the cortex, consistent with previous studies using different neuroimaging data.

Xu et al., 2021

# Euclidean embeddings are limited for brain networks

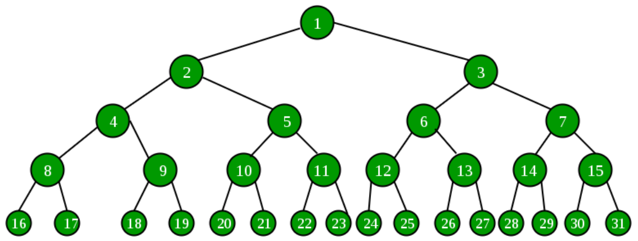
## Vector embedding

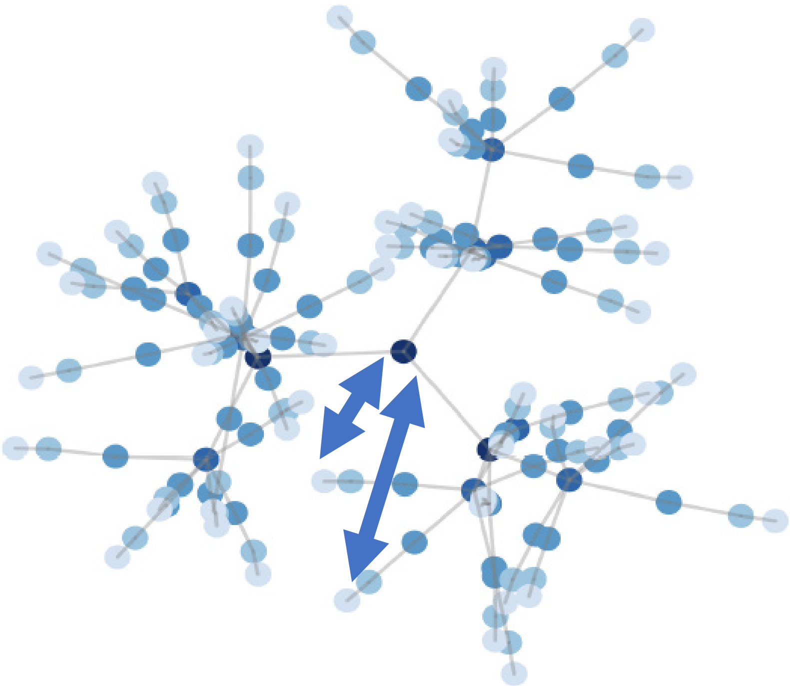


Euclidean embedding space

Brain networks are scale-free graphs ⇒ tree-like structure. The graph volume grows *exponentially* as a function of radius

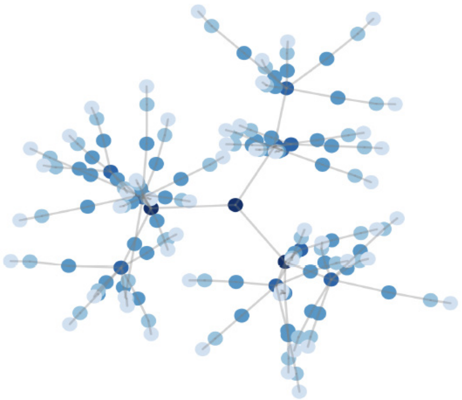


**Graph volume:** number of nodes within some radius to a center node

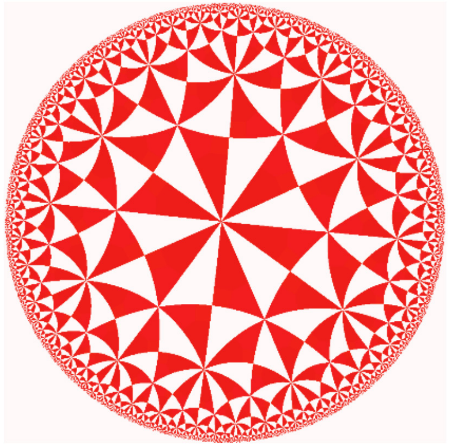
# Hyperbolic embeddings (negative curvature)

Embedding of a tree in **Euclidean space:**

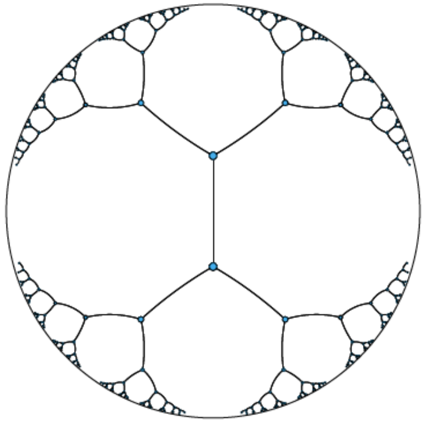


*Polynomial expansion*: Space increases by $x^2$

Embedding of tree in **Hyperbolic (Poincaré) space:**



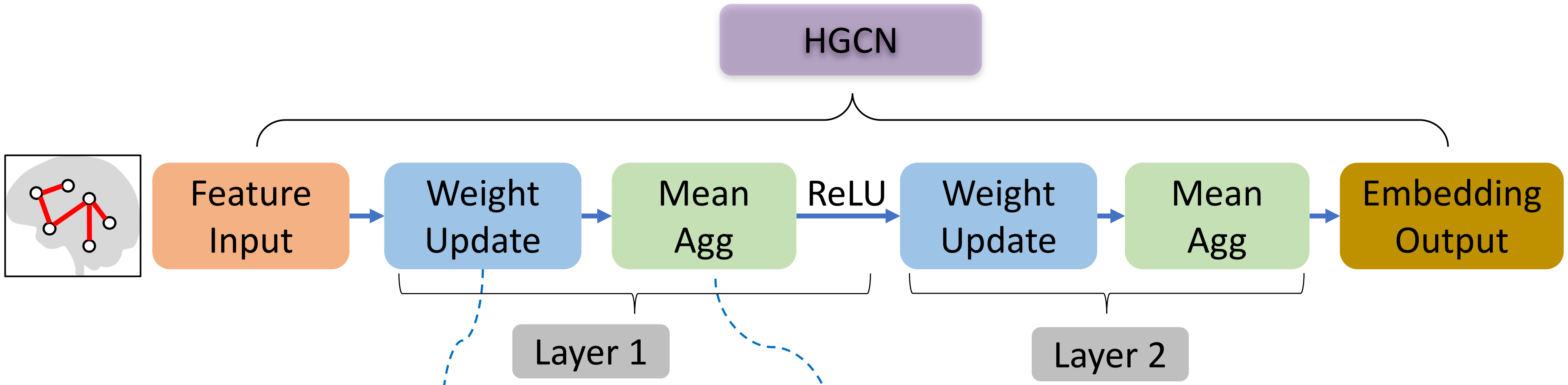
Poincaré model

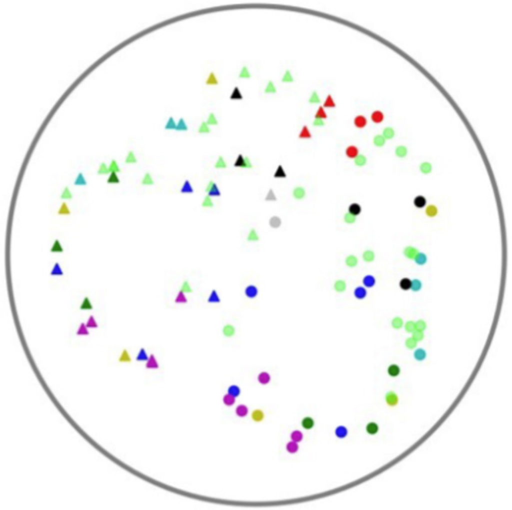


Embedding of tree

*Exponential expansion*: Space increases by $e^x$

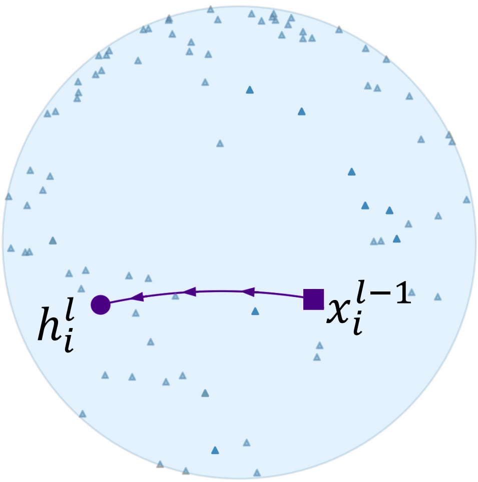
# Hyperbolic graph convolutional network (HGCN)

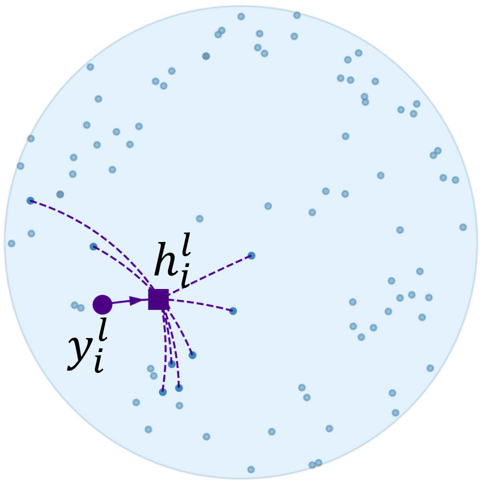


$$h_i^l = W^l \otimes_{c_{l-1}} x_i^{l-1} \oplus_{c_{l-1}} \boldsymbol{b}^l$$

$$y_i^l = \mathrm{FrechetAgg}_{j \in \{N(i),\, i\}}(h_i^l)$$



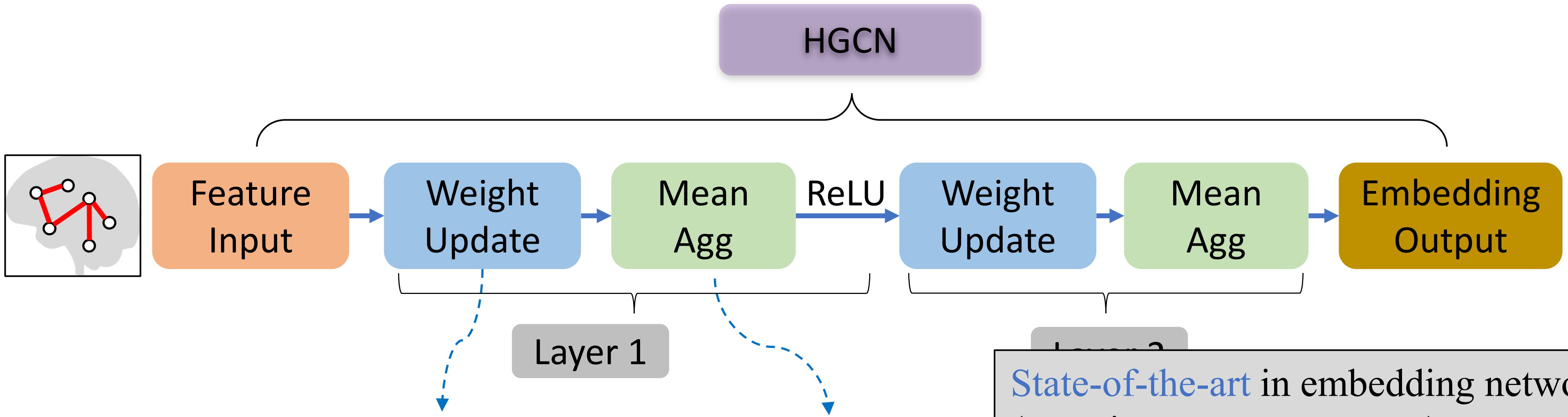


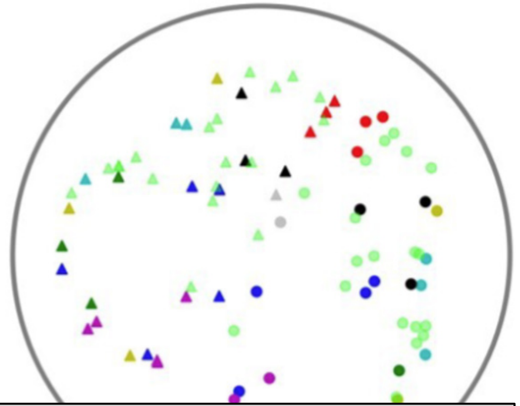
Matrix-Vector Multiplication: $Wx \to W \bigotimes_C x$

Bias Addition: $x + b \to x \bigoplus_C b$

Modified from: Chami, I., Ying, R., Re, C., and Leskovec, J. (2019). Hyperbolic Graph Convolutional Neural Networks. NeurIPS 2019

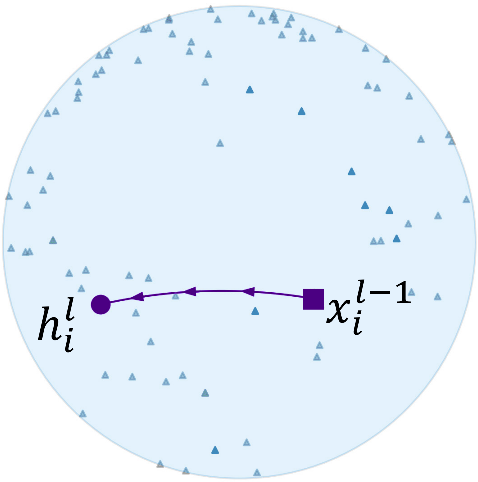
# Hyperbolic graph convolutional network (HGCN)

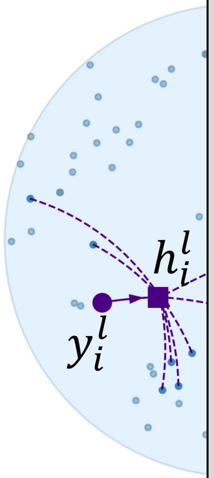


$$h_i^l = W^l \otimes_{c_{l-1}} x_i^{l-1} \oplus_{c_{l-1}} \boldsymbol{b}^l$$

$y_i^l$ = Freche




State-of-the-art in embedding networks in hyperbolic spaces (negative curvature spaces)

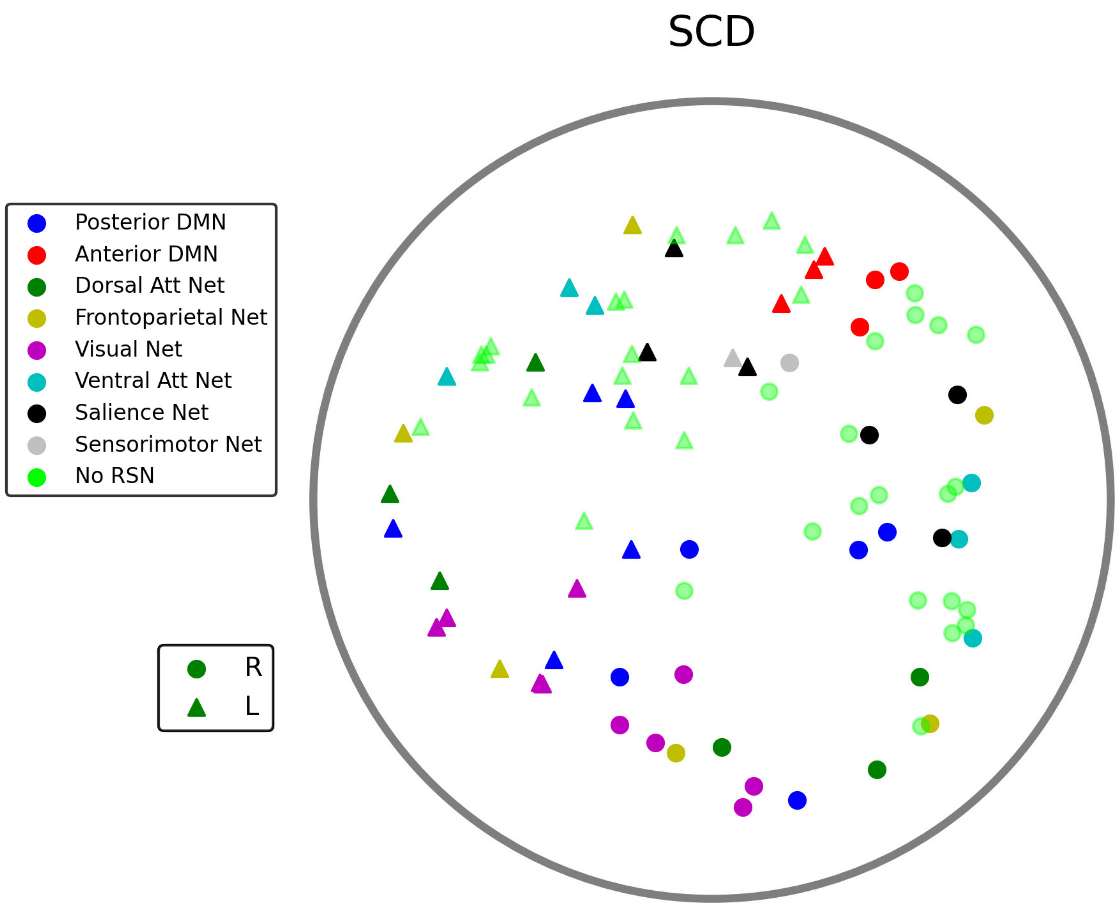
1. Yields the least distortion for scale-free data as in the case of brain networks
2. Due to the low distortion, brain networks can be mapped in low-dimensional spaces (even 2D spaces; MG2G encoder required 16-dimensions), which facilitates the downstream task of classification of AD staging
3. Important! Provides additional information on the hierarchy of brain networks

Matrix-Vector Multiplication: $Wx \rightarrow W \bigotimes_C x$

Bias Addition: $x + b \rightarrow x \bigoplus_C b$

Modified from: Chami, I., Ying, R., Re, C., and Leskovec, J. (2019). Hyperbolic Graph Convolutional Neural Networks. NeurIPS 2019

# Hyperbolic graph embedding of MEG brain networks to study brain alterations in patients with subjective cognitive decline




**Participants:**

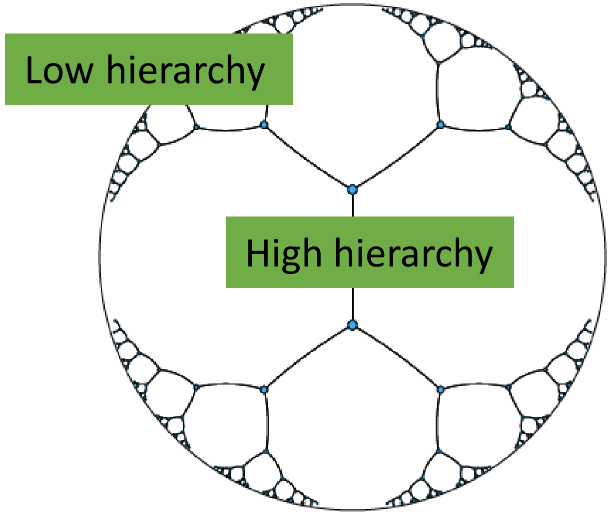
- 25 subjective cognitive decline (SCD) patients
- 19 age-matched clinically normal subjects (all from the *Madrid cohort* database).

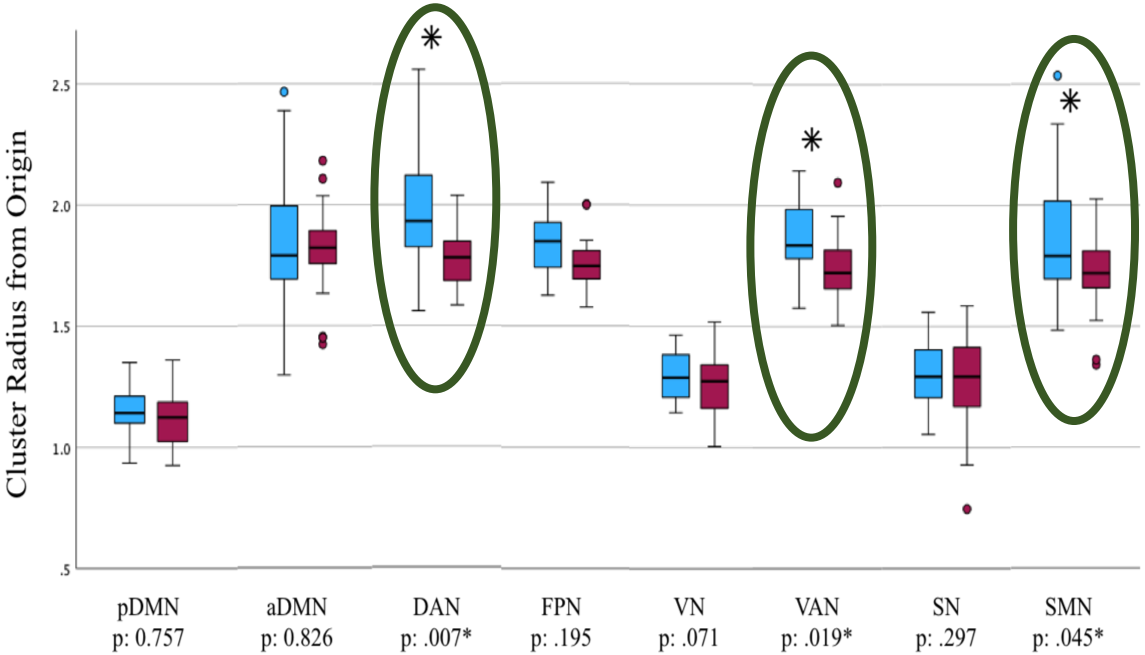
**Mean Average Precision (in link prediction task)**

| Method | # Embedding dimensions ($L$) | | | | |
|---|---|---|---|---|---|
| | 2 | 3 | 4 | 6 | 8 |
| GCN (Euclidean) | 0.563 | 0.701 | 0.791 | 0.818 | 0.827 |
| Poincare embedding | 0.723 | 0.748 | 0.75 | 0.751 | 0.755 |
| HGCN (w/ fixed C = -1) | 0.799 | 0.939 | 0.949 | 0.951 | 0.951 |
| HGCN (w/ learned C) | 0.828 (C=-.98) | 0.952 (C=-.58) | 0.954 (C=-.48) | 0.96 (C=-.44) | 0.964 (C=-.52) |

We chose to use the HGCN with **3 dimensions and learned C**, in order to maximize embedding quality while minimizing dimensionality.

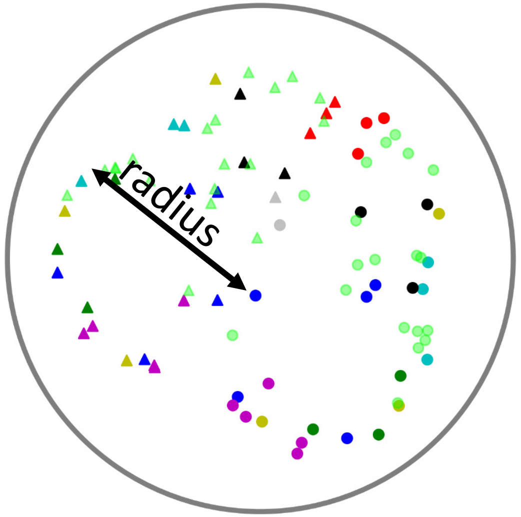
# Subnetwork hierarchy analysis




**Resting State Network Summary**

| RSN Name | Abrv. | Ex. ROI | ROI Abrv. |
|---|---|---|---|
| Posterior Default Mode Network | pDMN | Precuneus | Precu |
| Anterior Default Mode Network | aDMN | Cingulate gyrus, Anterior part | ACC |
| Dorsal Attention Network | DAN | Precentral gyrus | PreCG |
| Fronto-parietal Network | FPN | Angular gyrus | Ang |
| Visual Network | VN | Middle Occipital Lobe | MOccL |
| Ventral Attention Network | VAN | Inferior Frontal gyrus, Triangular | IFGt |
| Salience Network | SN | Amygdala | Amyg |
| Sensorimotor Network | SMN | Supplementary Motor area | SMA |




**Radius of the node embeddings**: A unique metric which effectively proxies the hierarchical organization of the brain.

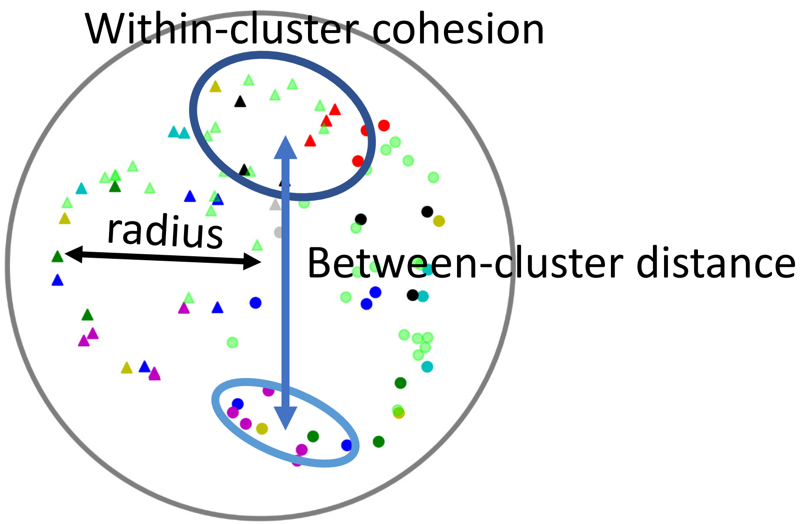
**Hierarchy alterations:**

- Participants with SCD had reduced distance from origin (increase in hierarchical centrality) in brain subnetworks: *Dorsal Attention Network (DAN), Ventral Attention Network (VAN), Sensorimotor Network (SMN)*

**Relationship of hierarchy alterations to cognition:**

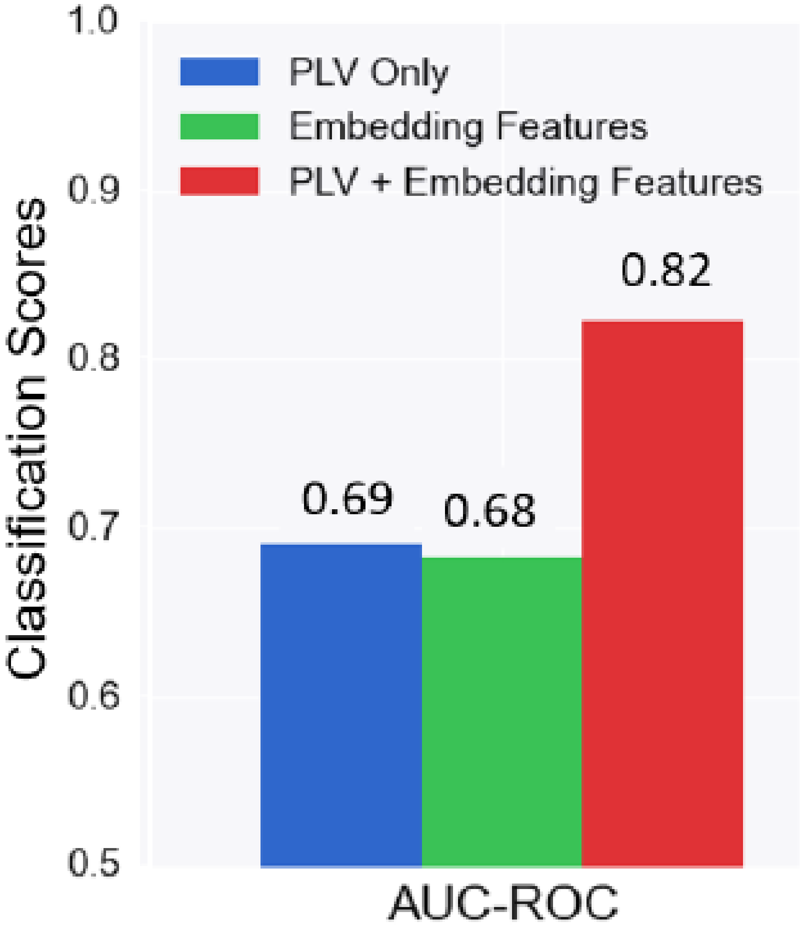
- DAN, VAN, SMN: Higher hierarchy correlated to higher scores on geriatric depression scale
- SMN: greater impairment in completing daily functional activities
- DAN: higher subjective rating of memory failures

# Predicting SCD using hyperbolic embeddings




Features from hyperbolic embeddings:

- **Radius from Origin -** The average distance from origin for each ROI. The smaller the radius, the more central the RSN is to the network
- **Within Cluster Cohesion -** Average pairwise distance between each node within the same sub-network. Smaller distance signifies a tightly packed network.
- **Between Cluster Distance -** Average pairwise distance between each node in different sub-networks. Smaller distance signifies a tightly linked networks.




Hyperbolic features capture **partially unique features** and improve classification scores over the original PLV features.

NIH
National Institute on Aging

NIH/NIA
1RF1AG074204
2021-2026
John Mosher
Michael Funke
$2,050,969

# A low cost ambulatory test for early detection of Alzheimer's disease

## Sample Selection and Enrollment (5-years)

Recruit **200 patients** with amnesic mild cognitive impairment (aMCI) from the Neurocognitive Disorders Center of the McGovern Medical School of UTHealth in Texas.

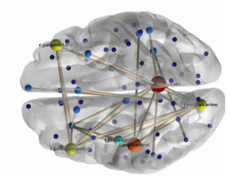
We anticipate an average **17% conversion rate** yearly from aMCI to AD (Luis et al. 2003; Fischer et al. 2007).

## Brain Recordings

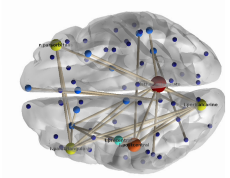
Gold standard: Combined 306-channel (**MEG**) and 128-channel wet-electrode **EEG** measurement in magnetically shielded room, using a whole-head MEG system (MEGIN Triux system).

Ambulatory: A 256-channel dry-electrode EEG (**aEEG**).

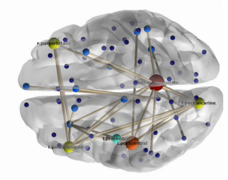
## Functional brain networks



306-channel MEG



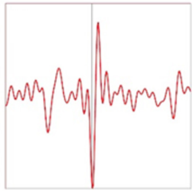
128-channel EEG



256-channel aEEG

## Epileptiform activity

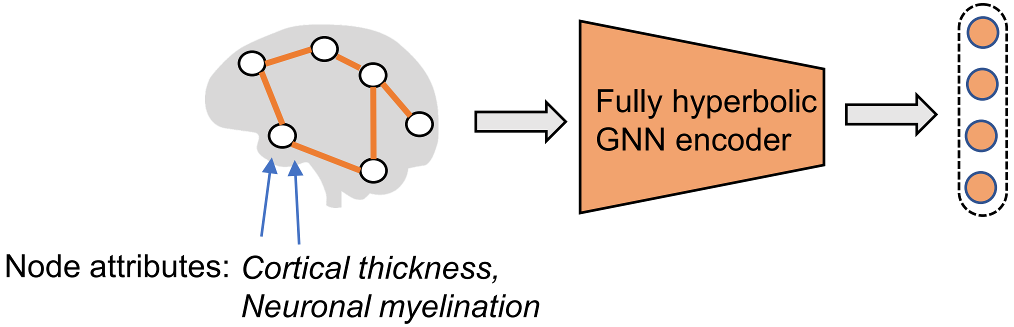
Epileptologist team will identify and characterize EA in the MEG/EEG data



# Ongoing work - Discover aging trajectories

## Cambridge Centre for Ageing and Neuroscience (Cam-CAN) dataset

| Name | N (Raw) Data |
|---|---|
| ***Magnetic Resonance Imaging (MRI) data (conforming to BIDS standard)*** | |
| *Structural Data* | |
| T1 | 653 |
| T2 | 653 |
| Diffusion Weighted Imaging (DWI) | 642 |
| Magnetisation Transfer Imaging (MTI) | 623 |
| *Functional* | |
| Resting State | 652 |
| Movie Watching | 649 |
| Sensori-motor task | 651 |
| ***Magnetoencephalography (MEG) (conforming to BIDS standard)*** | |
| Resting State | 647 |
| Sensori-Motor task | 647 |
| Sensory (passive) task | 639 |



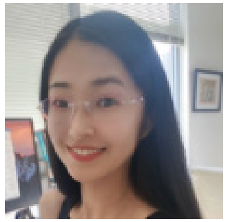

Adapt hyperbolic embedding pipeline to **Cam-CAN** dataset.

Incorporate **cortical thickness** and **neuronal myelination** as node features into brain graphs.

Improve existing HGCN model: convert to **fully hyperbolic (**current models formalize computations in an intermediate tangent Euclidean space, which leads to instabilities and information loss).

$$f^{C_{in},C_{out}}(x) = \exp^{C_{out}}(f(\log^{C_{in}}(x)))$$

Study which features contribute significantly to aging using the **GNNexplainer**, which identifies important subgraphs (network edges) and features (via a node mask) that contribute to model predictions.

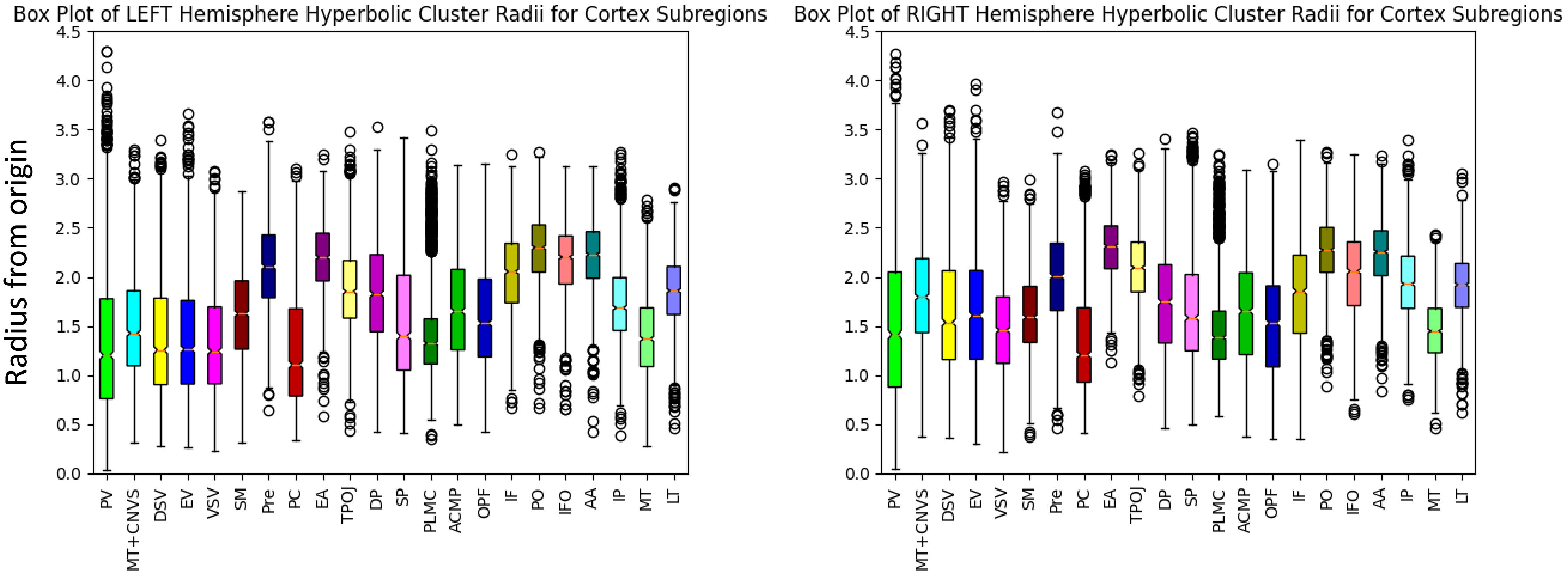


Mengjia Xu

Hugo Ramirez

Davide Tabarelli

# Hyperbolic embeddings and hierarchy of subnetworks

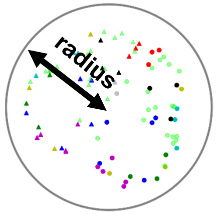


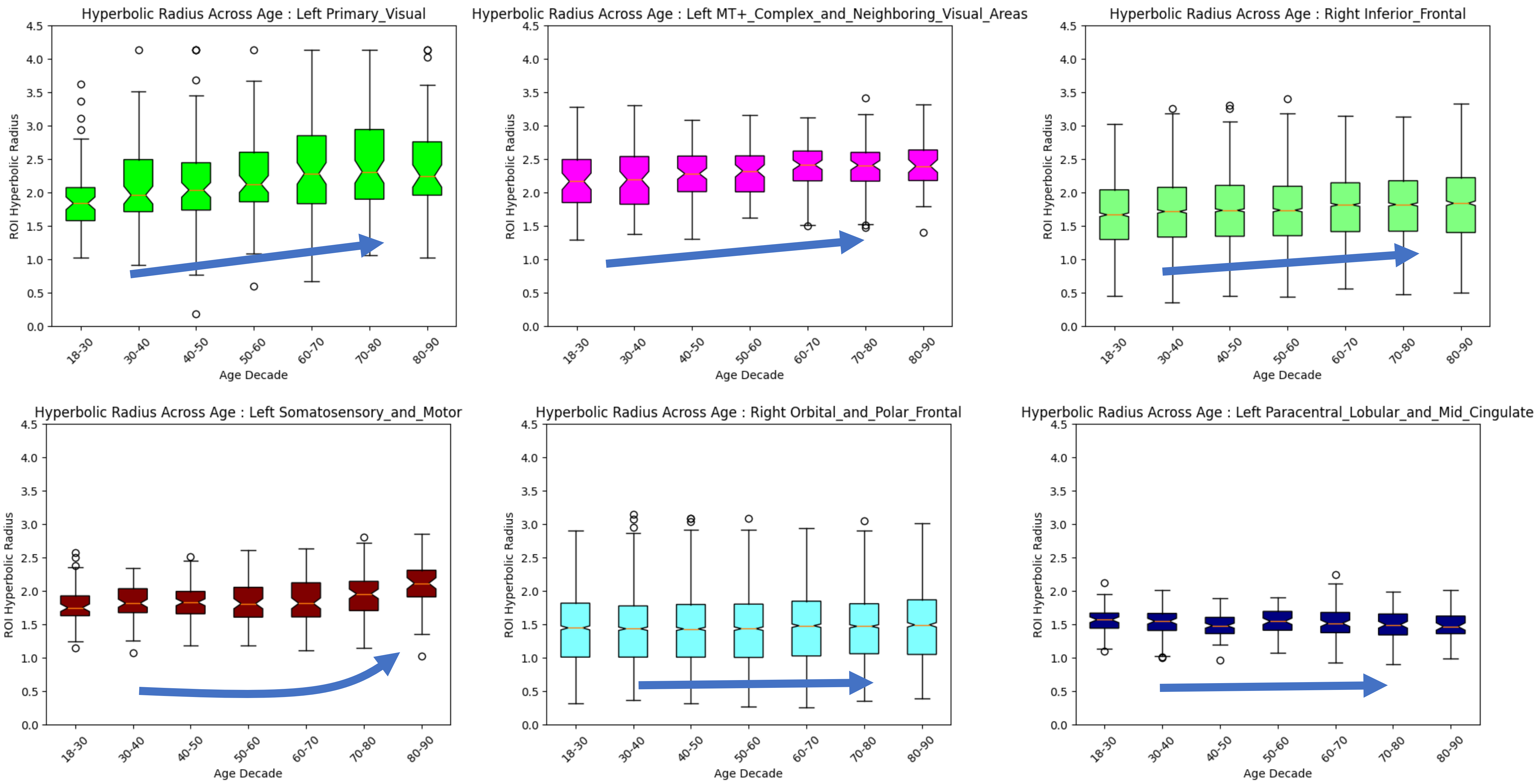

**PV**: Primary Visual
**MT_CNVS**: MT+ Complex and Neighboring Visual Areas
**DSV**: Dorsal Stream Visual
**EV**: Early Visual
**VSV**: Ventral Stream Visual
**SM** :Somatosensory and Motor
**Pre**: Premotor
**PC**: Posterior Cingulate

**EA**: Early Auditory
**TPOJ**: Temporo-Parieto Occipital Junction
**DP**: Dorsolateral Prefrontal
**SP**: Superior Parietal
**PLMC**: Paracentral Lobular and Mid Cingulate
**ACMP**: Anterior Cingulate and Medial Prefrontal
**OPF**: Orbital and Polar Frontal
**IF**: Inferior Frontal

**PO**: Posterior Opercular
**IFO**: Insular and Frontal Opercular
**AA**: Auditory Association
**IP**: Inferior Parietal
**MT**: Medial Temporal
**LT**: Lateral Temporal

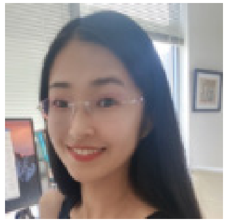
# Hierarchy of subnetworks across age



# Acknowledgements

## MIT Team


Mengjia Xu


Cole Baker


Hugo Ramirez


Wiley Smith

## Madrid / Complutense Univ. Team


Fernando Maestu


Isabel Suarez Mendez

## Trento Team


Davide Tabarelli

## Paris / IM2A Team


Katia Andrade


Thomas Guieysse

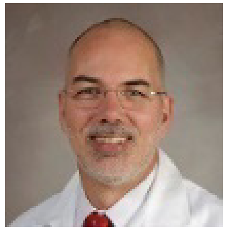

Sol Razafimahatratra

## USC Team


Takfarinas Medani

## UTHealth Team


John Mosher


Michael Funke

## MGH Team


Quanzheng Li

NIH National Institute on Aging

J-Clinic
ABDUL LATIF JAMEEL CLINIC FOR MACHINE LEARNING IN HEALTH

MISTI MIT Global Experiences