NeuroPath: A Neural Pathway Transformer for Joining the Dots of Human Connectomes

Department of Computer Science

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Ziquan Wei, Tingting Dan, Jiaqi Ding, Guorong Wu

on NeurIPS 2024

Advanced Computational Medicine Laboratory

ACM Lab

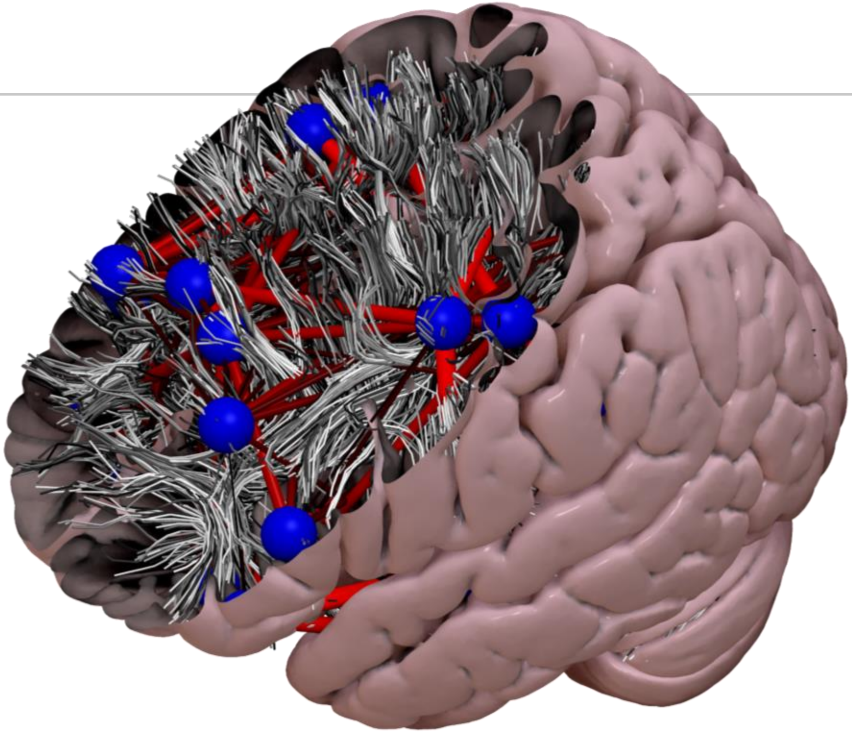
Outline

- Background
- Research problems
- Motivations
- Methods
- Experimental designs
- Results
- Conclusion

Background: Network neuroscience

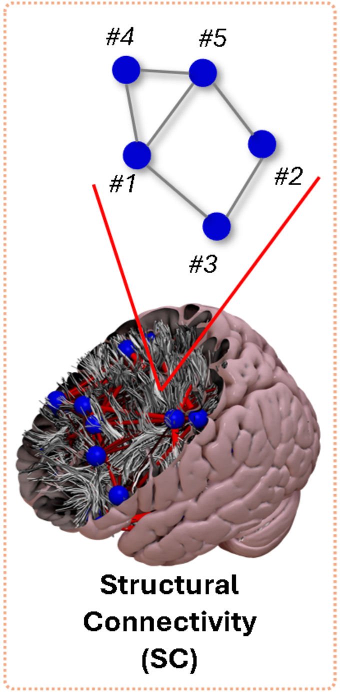
Predict the brain state given human brain networks

1. MRI is a non-invasive 3D imaging
2. Pre-defined brain atlas partition 3D images into regions
3. Regions refer to nodes of the graph along with regional signals.

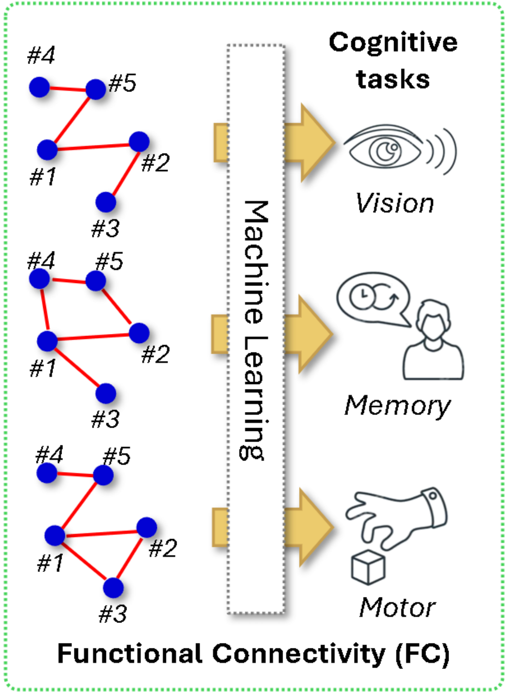


Gray matter surface & human connectome

Background: Two types of graph



Stable across scanning sessions



Depends on subject status, e.g., cognitive tasks

Gray links are SC

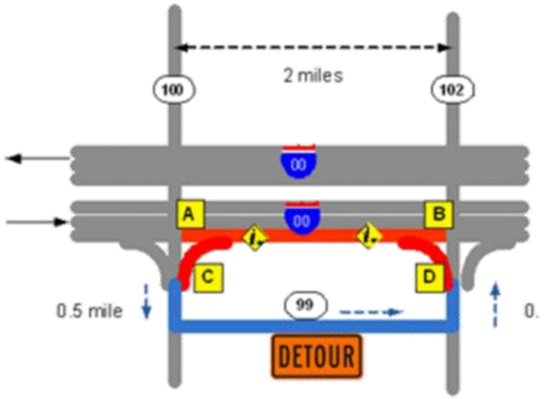
- **DWI** shows the orientation of white matter fibers
- Relatively static

Red links are FC

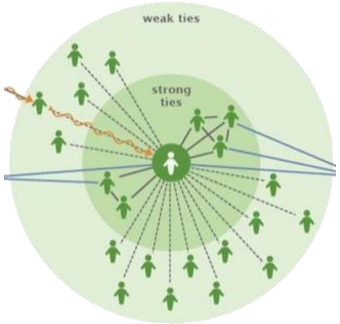
- **fMRI** measures blood-oxygen level dependent (BOLD)
- Region-wise Pearson correlation indicates the edge.
- Dynamic

Problem 1: Inter-subject variations

- Everyone has a different SC and FC, but has the similar function
    - Structure and functions are changing each other



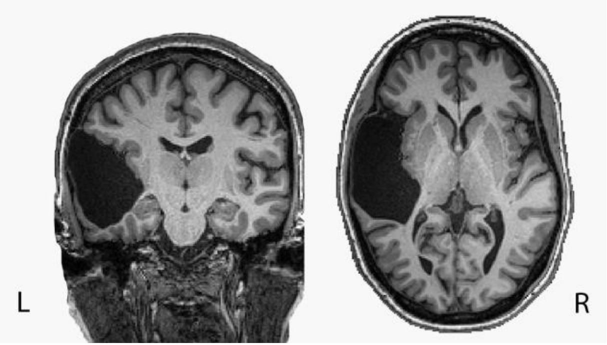
Route planning



Social connecting via a common friend

Problem 1: Inter-subject variations

- Everyone has a different SC and FC, but has the similar function
  - Structure and functions are changing each other

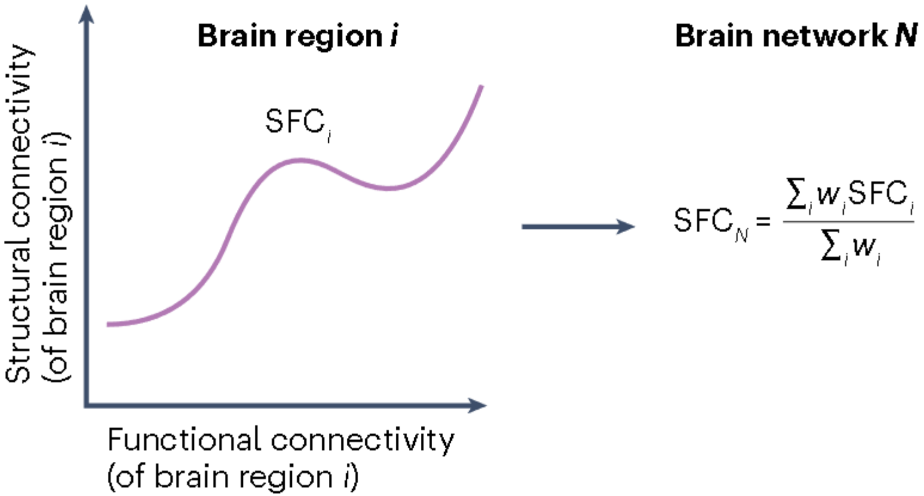


Brain resilience for white-matter lesions

How to couple structure and function together?

Previous works

- Univariate structure-function coupling
  - Third nodes are not involved



[1] Fotiadis, Panagiotis, et al. "Structure–function coupling in macroscale human brain networks." *Nature Reviews Neuroscience* (2024): 1-17.

Problem 2

- Graph topology representation learning
  - Finding all high-order structures is a NP-problem
  - NP-problem is not desirable in network neuroscience, since FC degree is high.

	Type	Time
PathNN [40]	All simple paths	5.23s ($H = 4$), 650s ($H = 5$)
Graphormer [37]	Shortest distance	270ms ($H = 7$)
	None	-

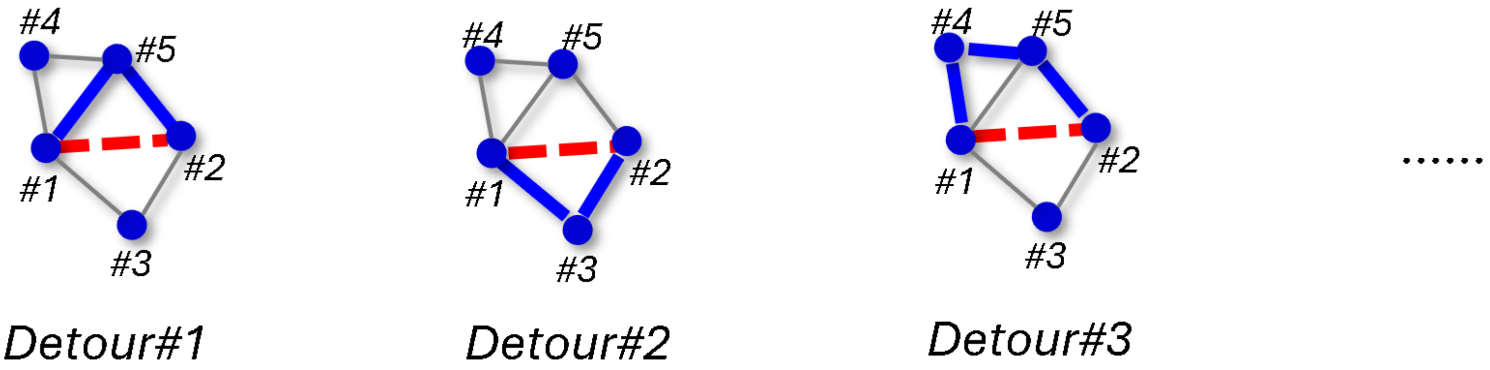
How can we train machine learning on neural pathways?

Previous works

- Methods dedicated on brain networks did not consider high-order structures
  - BNT, BolT, BrainGNN, BrainCNN…
- High-order graph neural networks need to find high-order structures in advance
  - GSN, PathNN…

Motivation 1

A multivariate structure-function relationship named topological **<u>detour</u>** is one of the biases causing inter-subject variations in brain network.



Motivation 2

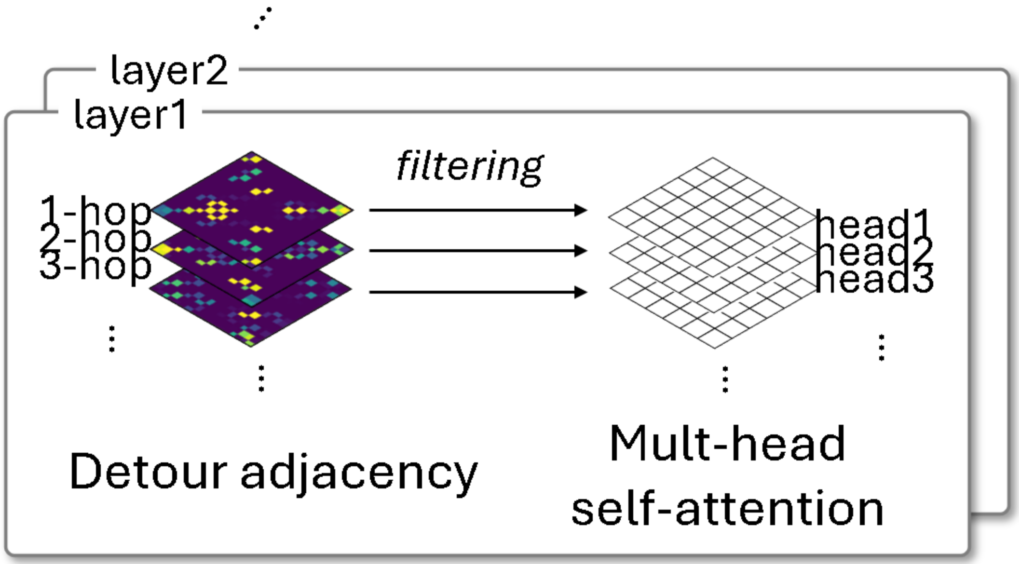
Transformer self-attention matrix can be treated as a graph adjacency matrix. Then it can be reduced to path adjacency

- Transformer:

$$\mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

- Graphormer:

$$A_{ij} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b_{\phi(v_i, v_j)}$$



Methods: Define the detour pathway

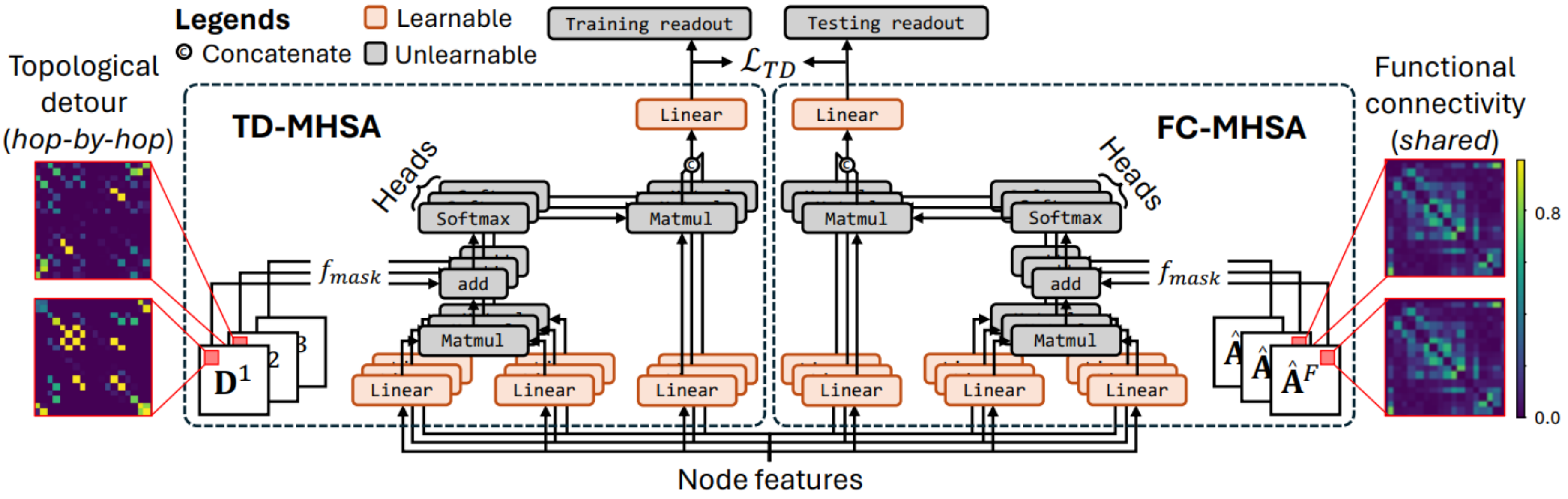
Detour adjacency matrix $$\mathbf{D}^h := \Big(\underbrace{(\hat{\mathbf{A}}^{\mathbf{S}})^h > 0}_{\text{Paths in SC graph}}\Big) \cdot \Big(\hat{\mathbf{A}}^{\mathbf{F}}\Big)$$

Edges in FC graph

Instead of knowing all paths, **D** indicates if there are h-long detours connecting a pair of node.

Methods: NeuroPath

Twin branch design of multi-head self-attention (MHSA)



Theoretical analysis: Pathway Representation

Get the weights of detours to one FC edge.

Fact 3.1. *The top pathway representations are obtained by* $\arg\max_{j,h}\left(\frac{1}{h}\sum_{j\in\mathbf{p}}\mathbf{S}_{ij}\bar{\gamma}_h\bar{\mathbf{W}}_{i\sim j}\right)$, *where* $\mathbf{S}$ *denotes the softmax of self-attention,* $\mathbf{p}\subset\mathbf{P}_i^H$ *is a set of node index of a path and* $\mathbf{P}_i^H$ *is the node collection of neural pathways within* H*-hop starting at* i*-th node.*

When interpret the results:

via Depth-First Search algorithm

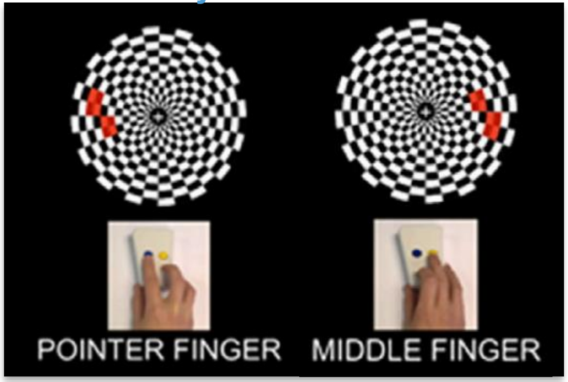
Statistical analysis: Detour degree

Paired *t*-test on how different $\boxed{\text{detour degree}}$ between task and resting

$\sum \mathbf{D}^{h}_{i,\cdot}$

Paired

Subject#1 Task | Subject#1 Resting | Subject#2, …, Subject#732

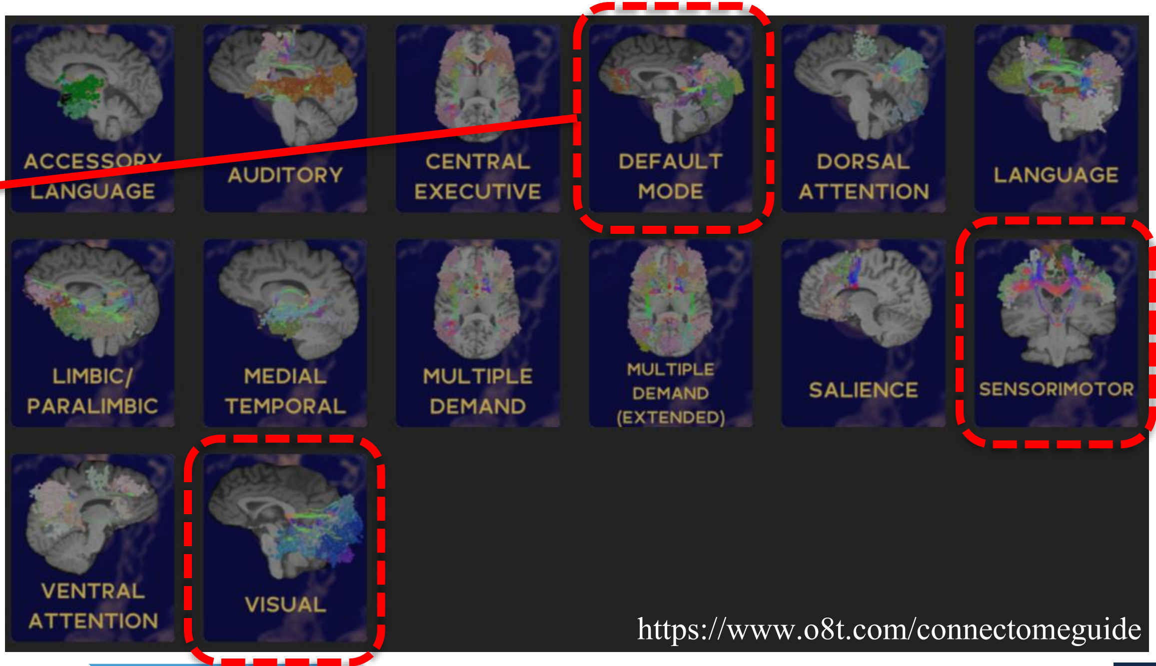


[2] Susan et. al., The lifespan human connectome project in aging: an overview. *Neuroimage*, 185:335–348, 2019.

Statistical analysis: Detour degree

Hypothesis: Degree is differentiated by main networks

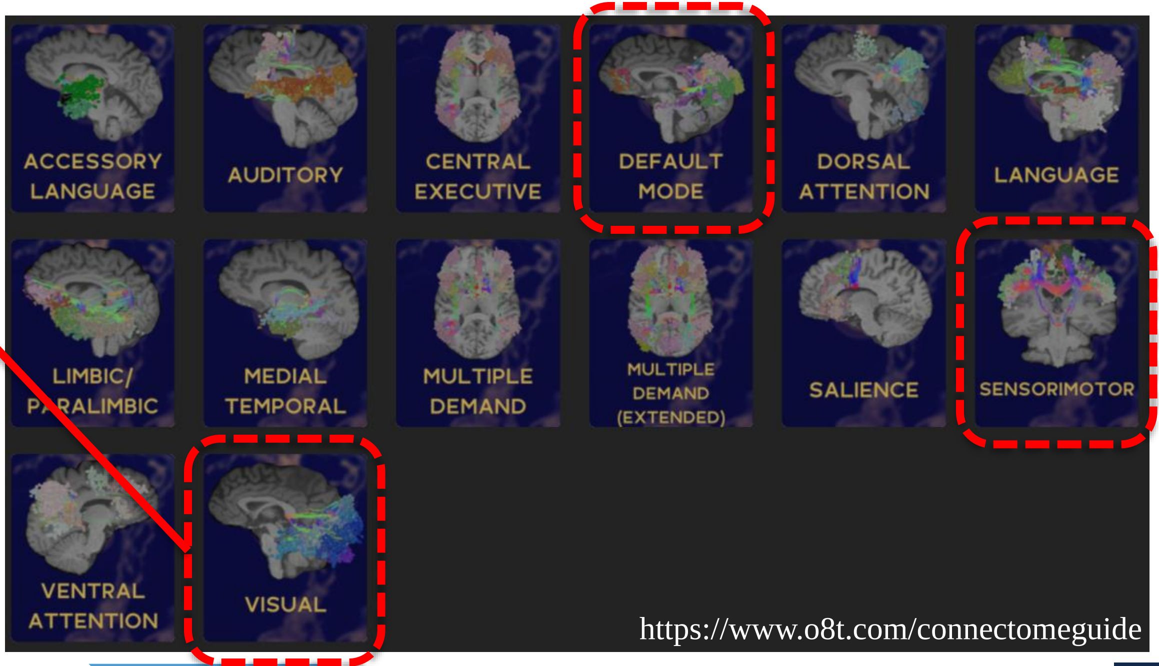
The **default mode network** is a distributed network of brain regions most active and connected during **rest**



Statistical analysis: Detour degree

Hypothesis: Node degree is differentiated by main networks

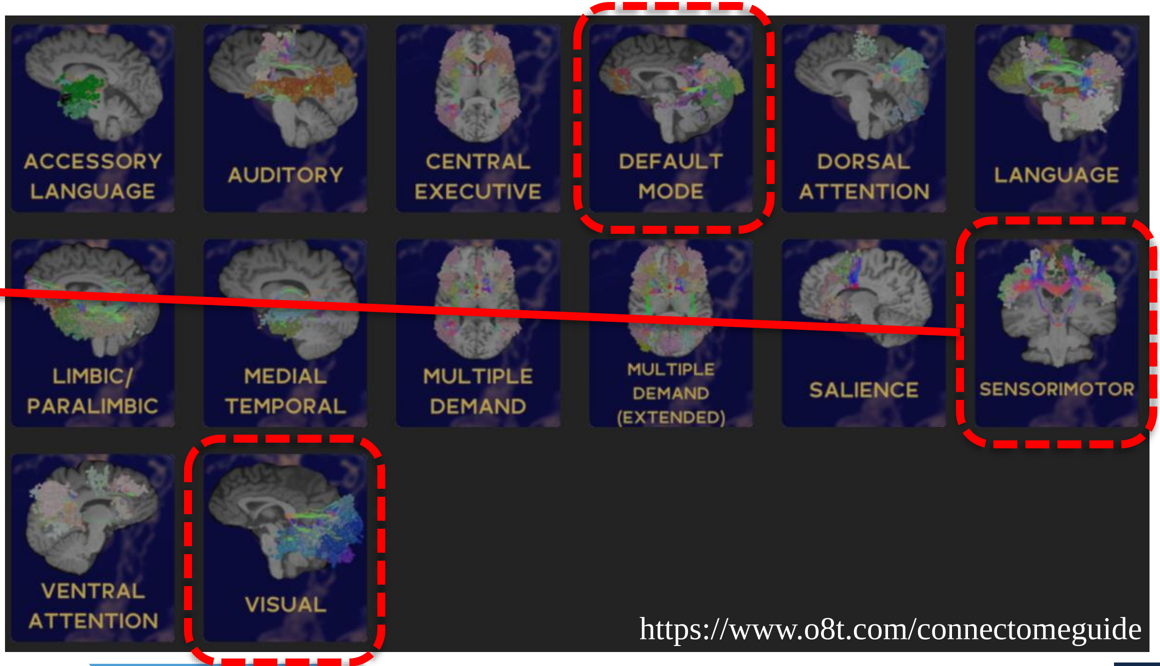
Visual network is consisted of the **primary visual areas**



Statistical analysis: Detour degree

Hypothesis: Node degree is differentiated by main networks

The precentral gyrus contained architecture has been known to be involved in **motor function** has been known since electrical stimulation experiments in dogs in the last 19th century.

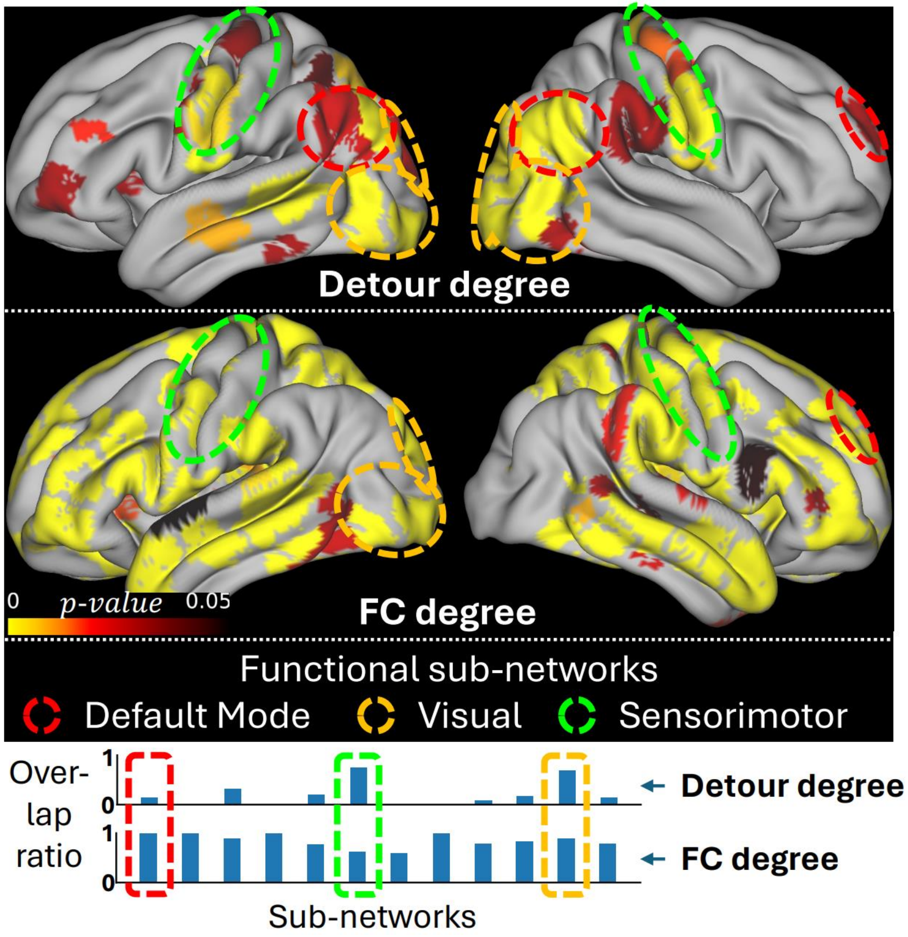


Methods

Observation of p-values

Test on HCP-Aging (n=732) dataset

- **Detour degree** is significantly different only in DMN, VN, SMN
- **FC degree** is significantly different in most of regions .



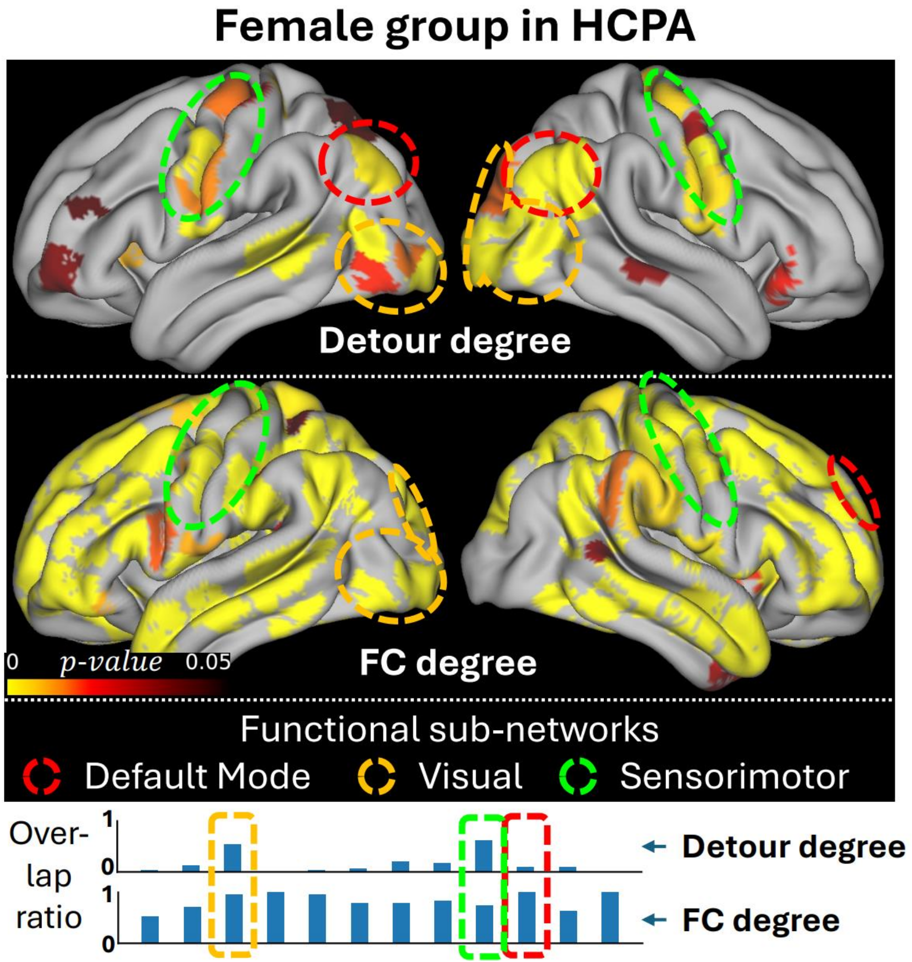
- FC threshold is 0.5
- SC threshold is 0.1
- H=6

Methods

Observation of *p*-values

Test on female group

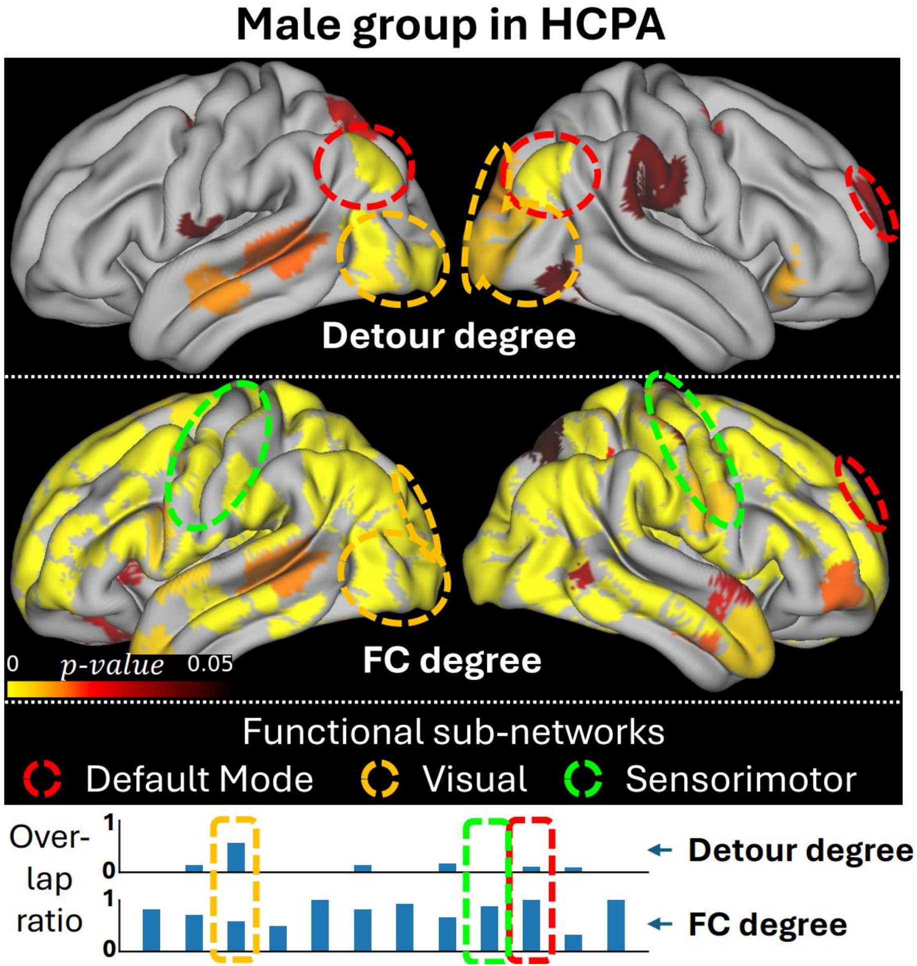
Observation is the same



Observation of p-values

Test on male group

Observation is partially the same



Experimental designs

Task/rest classification

AD/Normal classification

	HCPA		UKB		ADNI		OASIS	
	static	dynamic	static	dynamic	static	dynamic	static	dynamic
$\|G\|$	4,863	18,306	5,890	22,600	138	294	402	1,678
$\|C\|$	4	4	2	2	2	2	2	2
$avg(D)$	6.53	13.52	12.85	36.07	44.44	43.89	56.47	59.36

- **FC graph construction:**
  1. Entire session
  2. 100 timepoints
- **Node feature:**
  a. Correlation
  b. BOLD signal

Experimental designs

Comparing methods

Venue	Method	Category
	MLP	baselines
	GCN	baselines
2021, MIA	BrainGNN	Methods dedicated on brain networks
2022, NeurIPS	BNT	Methods dedicated on brain networks
2023, MIA	BolT	Methods dedicated on brain networks
2021, NeurIPS	Graphormer	General graph transformers
2023, ICLR	NAGphormer	General graph transformers
	NeuroPath	

Experiment 1

Task/rest?

Red: 1^{st}

Blue: 2^{nd}

Orange: 3^{rd}

	HCPA CORR		HCPA BOLD		UKB CORR		UKB BOLD	
	static	dynamic	static	dynamic	static	dynamic	static	dynamic
• Accuracy								
MLP	$96.01_{\pm 0.50}$	$92.57_{\pm 0.32}$	$93.42_{\pm 0.58}$	$83.64_{\pm 1.33}$	$99.00_{\pm 0.15}$	$97.70_{\pm 0.32}$	$99.05_{\pm 0.48}$	$96.42_{\pm 0.60}$
GCN	$95.85_{\pm 0.93}$	$91.95_{\pm 0.45}$	$92.94_{\pm 0.58}$	$84.60_{\pm 0.45}$	$99.00_{\pm 0.22}$	$97.54_{\pm 0.24}$	$99.31_{\pm 0.33}$	$93.39_{\pm 0.71}$
BrainGNN	$90.85_{\pm 1.35}$	$86.06_{\pm 2.64}$	$89.38_{\pm 2.88}$	$72.62_{\pm 3.33}$	$97.54_{\pm 0.52}$	$95.32_{\pm 1.68}$	$90.33_{\pm 2.72}$	$86.11_{\pm 4.04}$
BNT	$97.92_{\pm 0.65}$	$94.18_{\pm 0.35}$	$92.57_{\pm 1.19}$	$86.55_{\pm 0.37}$	$98.71_{\pm 0.34}$	$97.15_{\pm 0.49}$	$98.64_{\pm 0.18}$	$95.98_{\pm 0.44}$
BolT	$96.40_{\pm 0.41}$	$91.68_{\pm 0.38}$	$95.78_{\pm 0.55}$	$91.92_{\pm 0.69}$	$99.13_{\pm 0.33}$	$97.61_{\pm 0.23}$	$99.29_{\pm 0.26}$	$98.22_{\pm 0.31}$
Graphormer	$78.80_{\pm 5.89}$	$78.73_{\pm 1.91}$	$59.63_{\pm 6.07}$	$65.01_{\pm 3.84}$	$92.76_{\pm 10.05}$	$81.98_{\pm 9.83}$	$86.82_{\pm 12.42}$	$55.56_{\pm 21.08}$
NAGphormer	$93.67_{\pm 0.96}$	$90.73_{\pm 0.64}$	$94.76_{\pm 1.15}$	$82.02_{\pm 1.77}$	$98.79_{\pm 0.35}$	$96.83_{\pm 0.36}$	$99.22_{\pm 0.36}$	$92.90_{\pm 0.69}$
NeuroPath	$96.69_{\pm 0.54}$	$92.76_{\pm 0.52}$	$95.03_{\pm 1.93}$	$87.54_{\pm 0.77}$	$99.22_{\pm 0.24}$	$97.77_{\pm 0.21}$	$99.59_{\pm 0.21}$	$94.12_{\pm 0.75}$
• F1 score								
MLP	$96.01_{\pm 0.49}$	$92.52_{\pm 0.35}$	$93.42_{\pm 0.58}$	$82.86_{\pm 1.68}$	$99.00_{\pm 0.15}$	$97.69_{\pm 0.32}$	$99.05_{\pm 0.49}$	$96.42_{\pm 0.60}$
GCN	$95.85_{\pm 0.95}$	$91.90_{\pm 0.41}$	$92.98_{\pm 0.60}$	$83.95_{\pm 0.37}$	$99.00_{\pm 0.22}$	$97.53_{\pm 0.24}$	$99.31_{\pm 0.33}$	$93.36_{\pm 0.71}$
BrainGNN	$90.92_{\pm 1.41}$	$85.43_{\pm 3.37}$	$89.38_{\pm 2.92}$	$64.40_{\pm 6.67}$	$97.54_{\pm 0.52}$	$95.30_{\pm 1.71}$	$90.35_{\pm 2.70}$	$86.09_{\pm 4.18}$
BNT	$97.92_{\pm 0.66}$	$94.16_{\pm 0.35}$	$92.57_{\pm 1.22}$	$86.45_{\pm 0.40}$	$98.71_{\pm 0.34}$	$97.15_{\pm 0.49}$	$98.64_{\pm 0.18}$	$95.97_{\pm 0.43}$
BolT	$96.38_{\pm 0.43}$	$91.66_{\pm 0.39}$	$95.78_{\pm 0.57}$	$91.76_{\pm 0.78}$	$99.13_{\pm 0.34}$	$97.60_{\pm 0.23}$	$99.29_{\pm 0.26}$	$98.22_{\pm 0.31}$
Graphormer	$77.29_{\pm 7.15}$	$75.26_{\pm 2.66}$	$53.05_{\pm 5.81}$	$57.04_{\pm 1.46}$	$92.67_{\pm 10.25}$	$80.19_{\pm 11.35}$	$86.54_{\pm 13.03}$	$50.12_{\pm 27.58}$
NAGphormer	$93.69_{\pm 0.95}$	$90.64_{\pm 0.68}$	$94.76_{\pm 1.16}$	$81.06_{\pm 2.03}$	$98.79_{\pm 0.35}$	$96.82_{\pm 0.35}$	$99.22_{\pm 0.36}$	$92.88_{\pm 0.68}$
NeuroPath	$96.70_{\pm 0.54}$	$92.72_{\pm 0.54}$	$95.09_{\pm 1.86}$	$87.03_{\pm 0.95}$	$99.22_{\pm 0.24}$	$97.77_{\pm 0.21}$	$99.59_{\pm 0.21}$	$94.11_{\pm 0.75}$

Experiment 1

AD/Normal ?

Red: 1st

Blue: 2nd

Orange: 3rd

	ADNI CORR		ADNI BOLD		OASIS CORR		OASIS BOLD	
	static	dynamic	static	dynamic	static	dynamic	static	dynamic
• Accuracy								
MLP	$79.26_{\pm 10.34}$	$82.68_{\pm 5.71}$	$80.67_{\pm 7.26}$	$82.93_{\pm 6.35}$	$89.28_{\pm 3.58}$	$89.32_{\pm 3.18}$	$88.99_{\pm 3.52}$	$89.02_{\pm 3.25}$
GCN	$84.22_{\pm 6.92}$	$83.30_{\pm 6.30}$	$80.67_{\pm 7.26}$	$83.53_{\pm 5.39}$	$88.80_{\pm 2.88}$	$88.30_{\pm 3.54}$	$88.27_{\pm 4.87}$	$88.49_{\pm 3.16}$
BrainGNN	$82.07_{\pm 6.86}$	$83.30_{\pm 5.42}$	$82.07_{\pm 6.86}$	$83.42_{\pm 6.05}$	$89.29_{\pm 4.75}$	$89.65_{\pm 3.31}$	$87.76_{\pm 4.64}$	$89.27_{\pm 3.36}$
BNT	$82.81_{\pm 6.47}$	$83.30_{\pm 6.30}$	$82.67_{\pm 4.40}$	$84.33_{\pm 6.99}$	$89.02_{\pm 3.48}$	$89.98_{\pm 2.75}$	$88.75_{\pm 4.36}$	$89.57_{\pm 3.02}$
BolT	$82.00_{\pm 3.51}$	$80.34_{\pm 2.82}$	$81.41_{\pm 7.08}$	$80.80_{\pm 7.53}$	$88.30_{\pm 3.77}$	$88.97_{\pm 3.04}$	$87.54_{\pm 4.62}$	$88.50_{\pm 3.35}$
Graphormer	$82.74_{\pm 5.89}$	$83.28_{\pm 5.80}$	$83.48_{\pm 5.31}$	$81.28_{\pm 6.58}$	$88.55_{\pm 4.22}$	$88.57_{\pm 3.18}$	$87.49_{\pm 5.19}$	$88.98_{\pm 3.26}$
NAGphormer	$82.74_{\pm 5.89}$	$82.79_{\pm 5.82}$	$81.33_{\pm 6.09}$	$82.17_{\pm 5.73}$	$89.53_{\pm 3.33}$	$88.64_{\pm 3.85}$	$89.02_{\pm 3.48}$	$89.21_{\pm 3.44}$
NeuroPath	$85.56_{\pm 4.97}$	$83.82_{\pm 3.94}$	$83.48_{\pm 5.31}$	$83.68_{\pm 5.64}$	$90.01_{\pm 3.42}$	$89.49_{\pm 3.33}$	$89.02_{\pm 3.48}$	$89.21_{\pm 3.44}$
• F1 score								
MLP	$74.72_{\pm 8.67}$	$77.98_{\pm 5.79}$	$74.96_{\pm 9.17}$	$76.75_{\pm 7.08}$	$87.05_{\pm 5.00}$	$86.74_{\pm 4.81}$	$85.27_{\pm 4.82}$	$85.02_{\pm 5.03}$
GCN	$78.53_{\pm 9.76}$	$76.95_{\pm 8.17}$	$76.19_{\pm 8.50}$	$77.87_{\pm 5.66}$	$84.75_{\pm 5.56}$	$85.71_{\pm 4.10}$	$85.56_{\pm 5.55}$	$85.86_{\pm 3.97}$
BrainGNN	$76.57_{\pm 10.01}$	$79.14_{\pm 8.02}$	$75.11_{\pm 9.69}$	$78.82_{\pm 6.96}$	$86.07_{\pm 5.71}$	$85.12_{\pm 4.90}$	$84.94_{\pm 5.22}$	$84.50_{\pm 5.00}$
BNT	$79.68_{\pm 6.15}$	$78.71_{\pm 6.67}$	$80.16_{\pm 8.01}$	$80.50_{\pm 8.40}$	$86.07_{\pm 3.19}$	$86.73_{\pm 3.57}$	$85.32_{\pm 4.85}$	$85.67_{\pm 4.04}$
BolT	$79.64_{\pm 4.33}$	$76.89_{\pm 7.75}$	$76.68_{\pm 8.77}$	$77.92_{\pm 8.62}$	$85.49_{\pm 3.85}$	$84.91_{\pm 4.76}$	$84.91_{\pm 4.76}$	$84.70_{\pm 4.39}$
Graphormer	$78.14_{\pm 6.03}$	$78.29_{\pm 5.22}$	$77.78_{\pm 5.51}$	$76.74_{\pm 6.95}$	$84.77_{\pm 5.24}$	$85.67_{\pm 3.72}$	$85.44_{\pm 4.73}$	$84.40_{\pm 4.77}$
NAGphormer	$76.57_{\pm 6.67}$	$76.46_{\pm 6.93}$	$75.40_{\pm 8.58}$	$77.80_{\pm 7.01}$	$85.61_{\pm 4.79}$	$84.76_{\pm 4.58}$	$83.87_{\pm 5.02}$	$84.48_{\pm 4.58}$
NeuroPath	$83.29_{\pm 4.45}$	$79.93_{\pm 5.83}$	$77.35_{\pm 7.35}$	$78.05_{\pm 5.88}$	$86.37_{\pm 5.03}$	$86.26_{\pm 4.48}$	$87.02_{\pm 3.77}$	$85.01_{\pm 4.48}$

Results

Experiment 1

Average ranking

	MLP	GCN	BrainGNN	BNT	BolT	Graphormer	NAGphormer	*NeuroPath*
HCPA	4.0	4.5	7.0	2.8	<u>2.5</u>	8.0	5.3	**2.0**
UKB	3.0	3.75	7.0	5.0	<u>2.3</u>	8.0	5.3	**1.8**
ADNI	6.9	4.4	4.1	<u>2.1</u>	5.8	4.6	5.8	**1.6**
OASIS	3.3	5.3	4.4	<u>2.8</u>	6.5	6.4	5.0	**2.5**

NeuroPath is the best in the average performance

Experiment 2

Zero-shot learning between four datasets using BOLD as node feature

	OASIS→ADNI		ADNI→OASIS	
	static	dynamic	static	dynamic
Graphormer	$77.63_{\pm 2.89}$	$77.24_{\pm 7.34}$	$79.69_{\pm 7.71}$	$\mathbf{83.55}_{\pm 6.90}$
NAGphormer	$73.11_{\pm 5.90}$	$78.09_{\pm 7.24}$	$69.63_{\pm 10.99}$	$78.09_{\pm 7.08}$
NeuroPath	$\mathbf{79.78}_{\pm 3.53}$	$\mathbf{81.57}_{\pm 7.24}$	$\mathbf{80.03}_{\pm 8.50}$	$\underline{79.65_{\pm 6.35}}$
	HCPA→UKB		UKB→HCPA	
	static	dynamic	static	dynamic
Graphormer	$39.09_{\pm 28.14}$	$50.97_{\pm 4.01}$	$57.78_{\pm 14.50}$	$64.36_{\pm 7.90}$
NAGphormer	$74.49_{\pm 4.01}$	$70.17_{\pm 1.31}$	$89.77_{\pm 0.94}$	$73.44_{\pm 0.70}$
NeuroPath	$\mathbf{91.29}_{\pm 2.10}$	$\mathbf{72.08}_{\pm 2.15}$	$\mathbf{90.61}_{\pm 3.65}$	$\mathbf{75.62}_{\pm 2.98}$

Experiment 3: Ablation studies

Twin branch is better than single branch

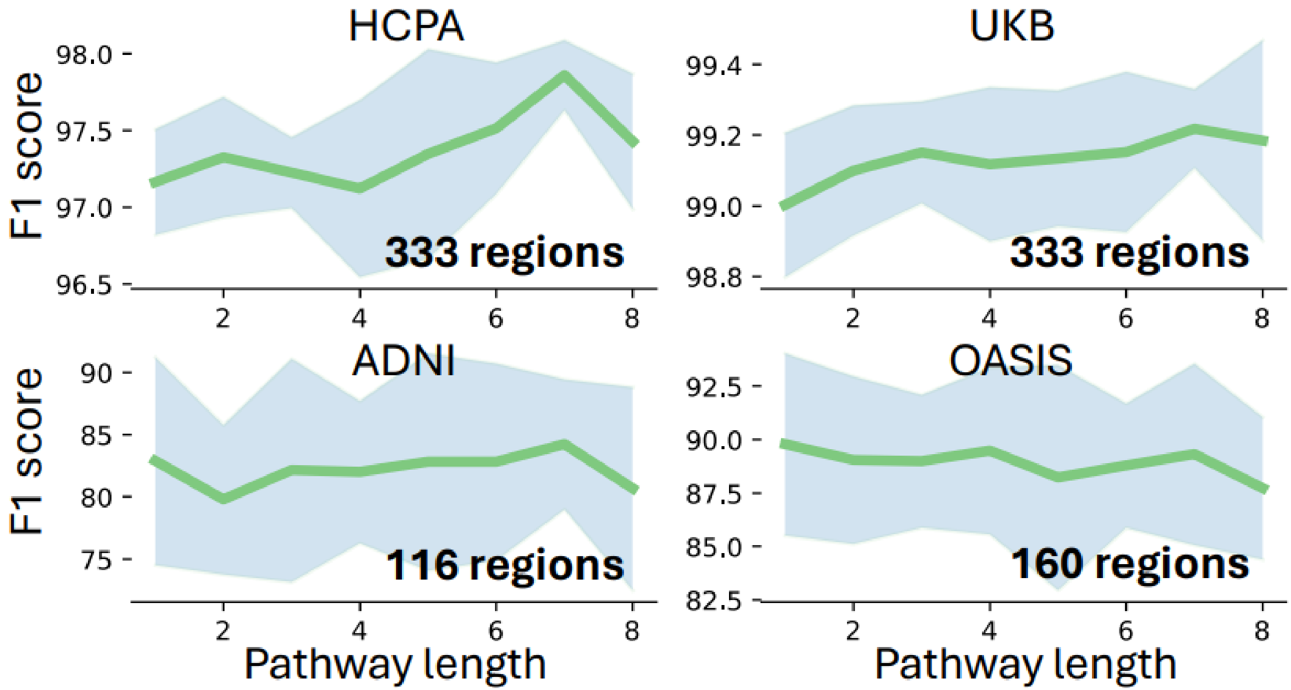
	ADNI		OASIS		HCPA		UKB	
	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
None	$82.42_{\pm 5.98}$	$78.65_{\pm 7.37}$	$88.52_{\pm 3.48}$	$86.19_{\pm 3.81}$	$97.53_{\pm 0.50}$	$97.53_{\pm 0.51}$	$99.53_{\pm 0.22}$	$99.53_{\pm 0.22}$
w/ TD-MHSA	$82.74_{\pm 7.88}$	$77.51_{\pm 9.39}$	$89.05_{\pm 3.99}$	$86.11_{\pm 4.32}$	$97.33_{\pm 0.44}$	$97.34_{\pm 0.43}$	$99.10_{\pm 0.13}$	$99.10_{\pm 0.13}$
w/ FC-MHSA	$81.93_{\pm 3.25}$	$80.97_{\pm 4.20}$	$89.31_{\pm 4.36}$	**$86.58_{\pm 5.87}$**	$97.72_{\pm 0.34}$	$97.72_{\pm 0.34}$	$99.25_{\pm 0.18}$	$99.25_{\pm 0.18}$
w/ both	**$85.56_{\pm 4.97}$**	**$83.29_{\pm 4.45}$**	**$90.01_{\pm 3.42}$**	$\underline{86.37_{\pm 5.03}}$	**$98.23_{\pm 0.45}$**	**$98.23_{\pm 0.45}$**	**$99.59_{\pm 0.21}$**	**$99.59_{\pm 0.21}$**

- Graph construction: Entire session
- Node feature: BOLD signal

Experiment 3: Ablation studies

Hyperparameter H controls how long a detour pathway is modeled

Finer regions have longer detour pathways



- Graph construction: Entire session
- Node feature: BOLD signal

Results

Experiment 3: Ablation studies

Deeper graph neural networks can be unstable

Layer #	HCPA 4	HCPA 8	HCPA 16	Rank	UKB 4	UKB 8	UKB 16	Rank
BNT	91.81	93.41	93.28	3.67	88.63	96.32	97.45	3.00
BolT	97.01	97.81	88.23	2.33	81.36	89.20	89.84	4.00
Graphormer	64.08	47.01	50.84	5.00	43.42	43.44	59.46	5.00
NAGphormer	96.89	97.26	97.22	2.33	99.24	98.95	99.20	2.00
NeuroPath	97.76	97.72	96.60	**1.67**	99.59	99.61	99.44	**1.00**
	ADNI			Rank	OASIS			Rank
BNT	76.39	75.91	77.28	3.67	85.32	85.96	85.21	3.33
BolT	75.93	78.67	78.23	2.67	85.30	84.55	85.55	3.67
Graphormer	78.58	74.12	74.12	4.00	84.45	83.87	83.87	5.00
NAGphormer	75.86	77.15	78.44	3.00	86.05	86.49	85.78	1.67
NeuroPath	78.93	78.42	78.32	**1.67**	86.16	86.77	85.78	**1.00**

Ours is the most stable model in comparison

- Graph construction: Entire session
- Node feature: BOLD signal

Results

Experiment 3: Ablation studies

Higher threshold ➡️ smaller degree

Lower threshold ➡️ bigger degree

Ours is the most stable model in comparison

	HCPA			Rank	UKB			Rank
FC threshold	0.3	0.5	0.7		0.3	0.5	0.7	
BNT	95.73	92.57	84.51	4.00	76.41	98.64	94.46	4.33
BolT	87.02	95.78	94.68	3.00	86.98	99.29	87.04	3.67
Graphormer	90.41	53.05	88.43	4.33	97.76	86.54	96.73	3.67
NAGphormer	96.08	94.76	96.85	2.33	97.80	99.22	98.78	2.33
NeuroPath	97.57	95.09	97.32	**1.33**	99.27	99.59	99.15	**1.00**
	ADNI			Rank	OASIS			Rank
BNT	77.74	80.16	77.92	**1.67**	85.14	85.32	86.05	3.67
BolT	74.33	76.68	76.53	4.00	84.98	84.91	84.67	4.67
Graphormer	75.82	77.78	75.17	3.33	86.23	85.44	87.15	2.00
NAGphormer	72.55	75.40	77.29	4.67	86.32	83.87	85.78	3.67
NeuroPath	78.36	77.35	79.49	**1.67**	86.59	87.02	86.13	**1.33**

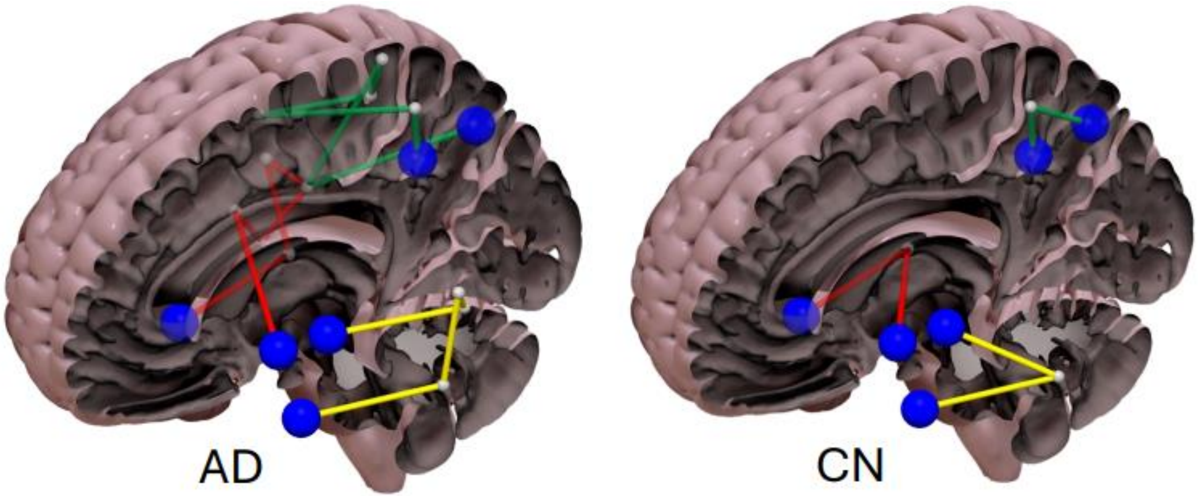
- Graph construction: Entire session
- Node feature: BOLD signal

Results

Experiment 4: Case study

Run DFS algorithm on three FC links that are significant in subcortical, entorhinal cortex, occipital lobe, and parietal lobe

AD prediction uses longer detours than the Normal



Conclusions

1. Detour in SC is one of inter-subject variations in brain networks with **statistical evidence**.
2. Path representation learning can be implemented without searching them in advance by reducing/filtering self-attention with **theoretical supports**.
3. NeuroPath has the best and stable performance in **real-world applications** including zero-shot learning.

Conclusions

Future directions

1. **Theory**: Is there a specific reduction of self-attention softmax theoretically refers to all nodes of all simple paths?
2. **Statistics**: How edge weights of FC and SC related to detour adjacency?
3. **Real-world application**: Other than Logistic/non-linear regression, how to apply to generative or unsupervised tasks?

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL