Temporal GNN : a general review

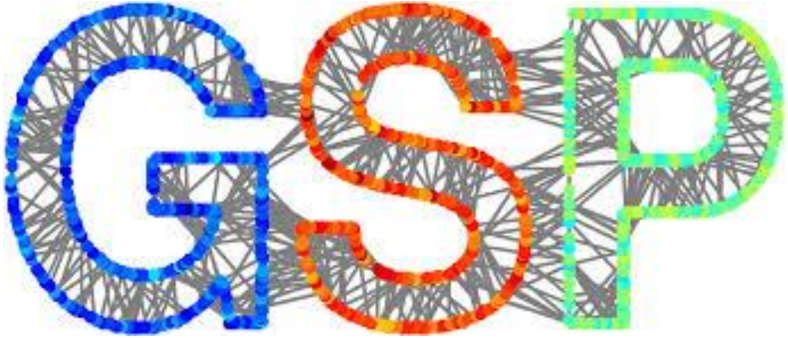
EE-626 Session

Who am I?

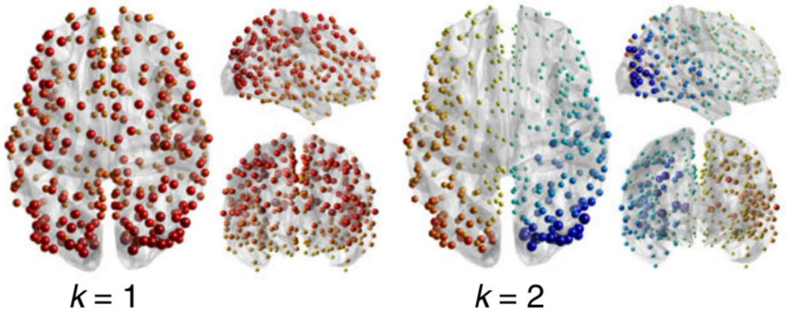
- My name is Michael Chan
- 2nd year PhD student in Neuro-X / Electrical Engineering @ Medical Imaging Processing Lab (MIPLab)
  - Bachelor in Mathematics
  - Master's in Data Science

What do I work on?

Method
(Graph Signal Processing)



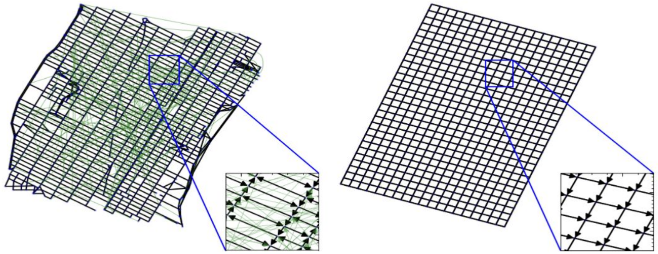
Application
(Structure Function)



What do I work on?

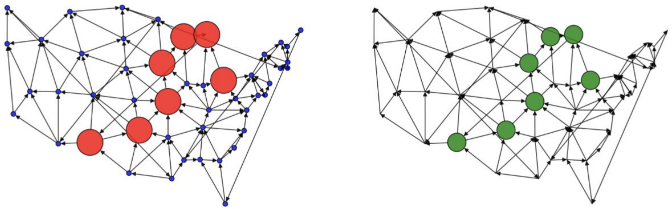
Methods : **directed** GSP

Graph Phase



(a) Midtown Manhattan graph (b) Regular 2D grid

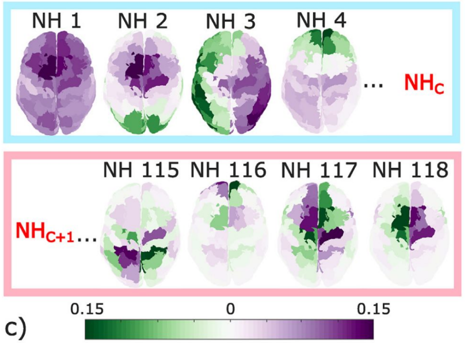
Graph Surrogates



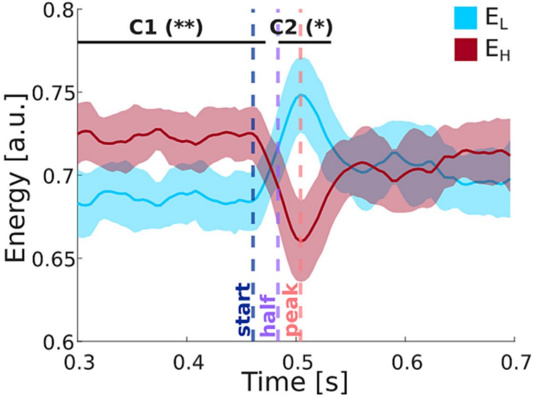
(d) Step Signal (e) Significant Nodes (Step Signal)

Application : **Epilepsy**

Brain eigenmodes



Interictal Spikes



Main paper

Graph Neural Networks for Temporal Graphs: State of the Art, Open Challenges, and Opportunities

Antonio Longa*
antonio.longa@unitn.it
University of Trento and Fondazione Bruno Kessler, Trento, Italy

Veronica Lachi*
veronica.lachi@student.unisi.it
University of Siena, Siena, Italy

Notation & Formalization

Definition 1 (Static Graph - SG) *A* Static Graph *is a tuple* $G = (V, E, X^V, X^E)$*, where* V *is the set of nodes,* $E \subseteq V \times V$ *is the set of edges, and* X^V, X^E *are* d_V*-dimensional node features and* d_E*-dimensional edge features.*

Definition 2 (Temporal Graph - TG) *A* Temporal Graph *is a tuple* $G_T = (V, E, V_T, E_T)$*, where* V *and* E *are, respectively, the set of all possible nodes and edges appearing in a graph at any time, while*

$$V_T := \{(v, x^v, t_s, t_e) : v \in V, x^v \in \mathbb{R}^{d_V}, t_s \leq t_e\},$$
$$E_T := \{(e, x^e, t_s, t_e) : e \in E, x^e \in \mathbb{R}^{d_E}, t_s \leq t_e\},$$

are the temporal nodes and edges, with time-dependent features and initial and final timestamps. A set of temporal graphs is denoted as $\mathcal{G}_T$.

Definition 3 (Discrete Time Temporal Graph - DTTG) *Let* $\Delta t > 0$ *be a fixed time-step and let* $t_1 < t_2 < \cdots < t_n$ *be timestamps with* $t_{k+1} = t_k + \Delta t$*. A* Discrete Time Temporal Graph G_{DT} *is a TG where for each* $(v, x^v, t_s, t_e) \in V_T$ *or* $(e, x^e, t_s, t_e) \in E_T$*, the timestamps* t_s, t_e *are taken from the set of fixed timestamps (i.e.,* $t_s, t_e \in \{t_1, t_2, \ldots, t_n\}$*, with* $t_s < t_e$*).*

Notation & Formalization

Definition 4 (Snapshot-based Temporal Graph - STG) *Let $t_1 < t_2 < \cdots < t_n$ be the ordered set of all timestamps t_s, t_e occurring in a TG G_T. Set*

$$V_i := \{(v, x^v) : (v, x^v, t_s, t_e) \in V_T, t_s \leq t_i \leq t_e\},$$
$$E_i := \{(e, x^e) : (e, x^e, t_s, t_e) \in E_T, t_s \leq t_i \leq t_e\},$$

and define the snapshots $G_i := (V_i, E_i)$, $i = 1, \ldots, n$. *Then a* Snapshot-based Temporal Graph *representation of G_T is the sequence*

$$\boxed{G_T^S := \{(G_i, t_i) : i = 1, \ldots, n\}}$$

of time-stamped static graphs.

Notation & Formalization

Definition 5 (Event-based Temporal Graph - ETG) *Let G_T be a TG, and let ε denote one of the following* events*:*

- Node insertion $\varepsilon_V^+ := (v, t)$*: the node v is added to G_T at time t, i.e., there exists $(v, x^v, t_s, t_e) \in V_T$ with $t_s = t$.*
- Node deletion $\varepsilon_V^- := (v, t)$*: the node v is removed from G_T at time t, i.e., there exists $(v, x^v, t_s, t_e) \in V_T$ with $t_e = t$.*
- Edge insertion $\varepsilon_E^+ := (e, t)$*: the edge e is added to G_T at time t, i.e., there exists $(e, x^e, t_s, t_e) \in E_T$ with $t_s = t$.*
- Edge deletion $\varepsilon_E^- := (e, t)$*: the edge e is removed from G_T at time t, i.e., there exists $(e, x^e, t_s, t_e) \in E_T$ with $t_e = t$.*

An Event-based Temporal Graph *representation of TG is a sequence of events*

$$\boxed{G_T^E := \{\varepsilon : \varepsilon \in \{\varepsilon_V^+, \varepsilon_V^-, \varepsilon_E^+, \varepsilon_E^-\}\}.}$$

Other specific modelling

- Sampling temporal network motifs
  - random walk to sample paths / subgraphs
  - supra-adjacency representation
    - DyANE Sato et al. 2019s
- Non-negative matrix factorization
- Statistical approaches to detect existence of edges / estimation
- Dynamic embedding projection based on previous snapshot
- Hypergraphs, multiplex graphs

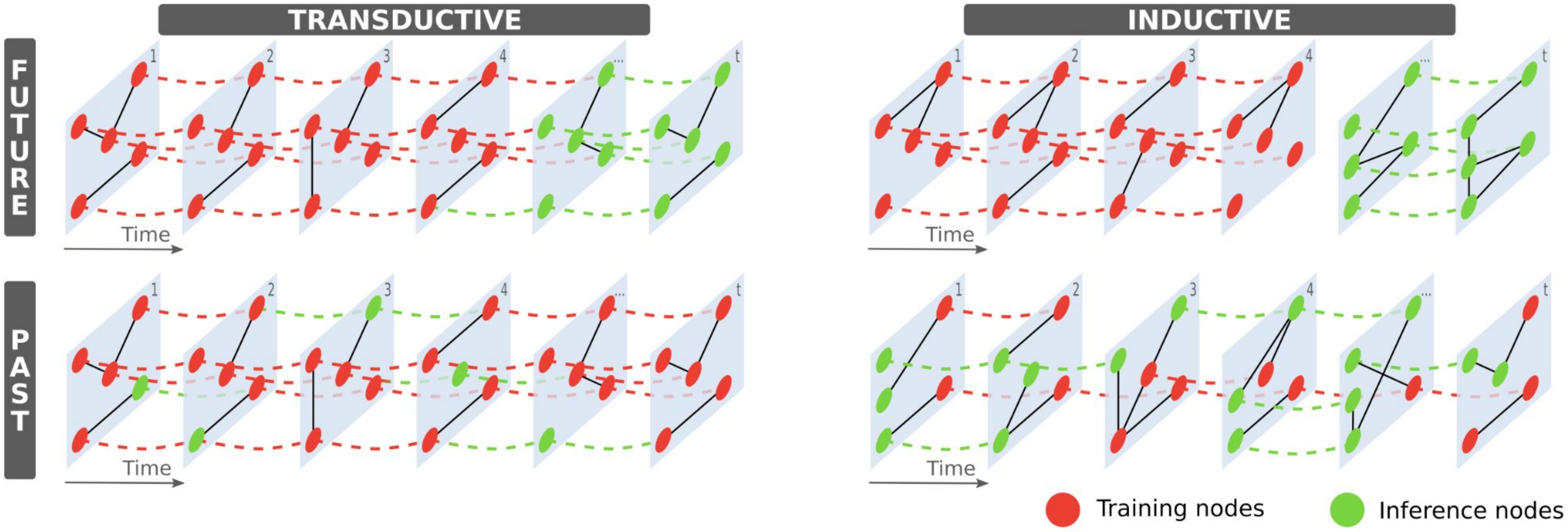
Learning in TGNN

Snapshots TGNN

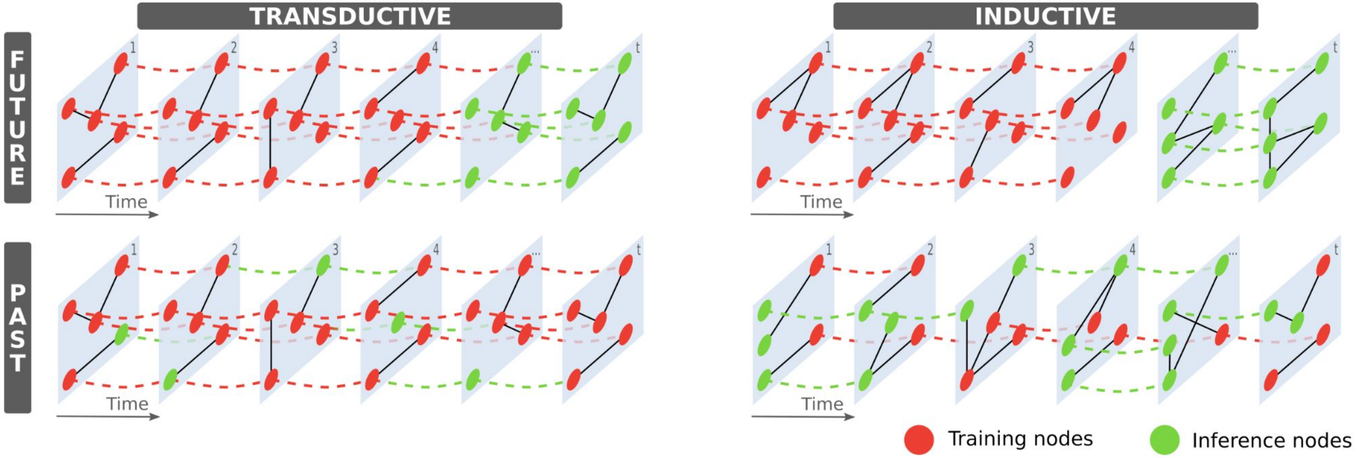
$$G_T^S := \{(G_i, t_i) : i = 1, \ldots, n\}$$

Event-based TGNN

$$G_T^E := \{\varepsilon : \varepsilon \in \{\varepsilon_V^+, \varepsilon_V^-, \varepsilon_E^+, \varepsilon_E^-\}\}.$$



Learning in TGNN



$$T_e^{all} := \max_{i=1,\dots,n} T_e(G_T^i),\ V^{all} := \cup_{i=1}^n V_i,\ E^{all} := \cup_{i=1}^n E_i,$$

- Transductive learning: *inference can only be performed on* $v \in V^{all}$, $e \in E^{all}$, *or* $G_T \subseteq_V G_T^i$ *with* $G_T^i \in \mathcal{G}_T$.
- Inductive learning: *inference can be performed also on* $v \notin V^{all}$, $e \notin E^{all}$, *or* $G_T \not\subseteq_V G_T^i$, *for all* $i = 1, \dots, n$.
- Past prediction: *inference is performed for* $t \leq T_e^{all}$.
- Future prediction: *inference is performed for* $t > T_e^{all}$.

Tasks in TGNN

- **Supervised Learning**
- Unsupervised Learning

Definition 7 $\boxed{\textbf{(Temporal Node Classification)}}$ *Given a TG $G_T = (V, E, V_T, E_T)$, the node classification task consists in learning the function*

$$f_{NC} : \boxed{V \times \mathbb{R}^+} \to \mathcal{C}$$

which maps each node to a class $C \in \mathcal{C}$, at a time $t \in \mathbb{R}^+$.

Definition 8 $\boxed{\textbf{(Temporal Edge Classification)}}$ *Given a TG $G_T = (V, E, V_T, E_T)$, the temporal edge classification task consists in learning a function*

$$f_{EC} : \boxed{E \times \mathbb{R}^+} \to \mathcal{C}$$

which assigns each edge to a class at a given time $t \in \mathbb{R}^+$.

Definition 9 $\boxed{\textbf{(Temporal Graph Classification)}}$ *Let $\mathcal{G}_T$ be a domain of TGs. The graph classification task requires to learn a function*

$$f_{GC} : \boxed{\mathcal{G}_T \times I^+} \to \mathcal{C}$$

that maps a temporal graph, restricted to a time interval $[t_s, t_e] \in I^+$, into a class.

Tasks in TGNN

- **Supervised Learning**
- Unsupervised Learning

Definition 7 (Temporal Node $\boxed{\textbf{Classification)}}$ *Given a TG $G_T = (V, E, V_T, E_T)$, the node classification task consists in learning the function*

$$f_{NC} : V \times \mathbb{R}^+ \to \mathcal{C}$$

which maps each node to a class $C \in \mathcal{C}$, at a time $t \in \mathbb{R}^+$.

- Regression counterpart
  - Temporal Link prediction
  - First time appearance of Edge

Tasks in TGNN

- Supervised Learning
- **Unsupervised Learning**

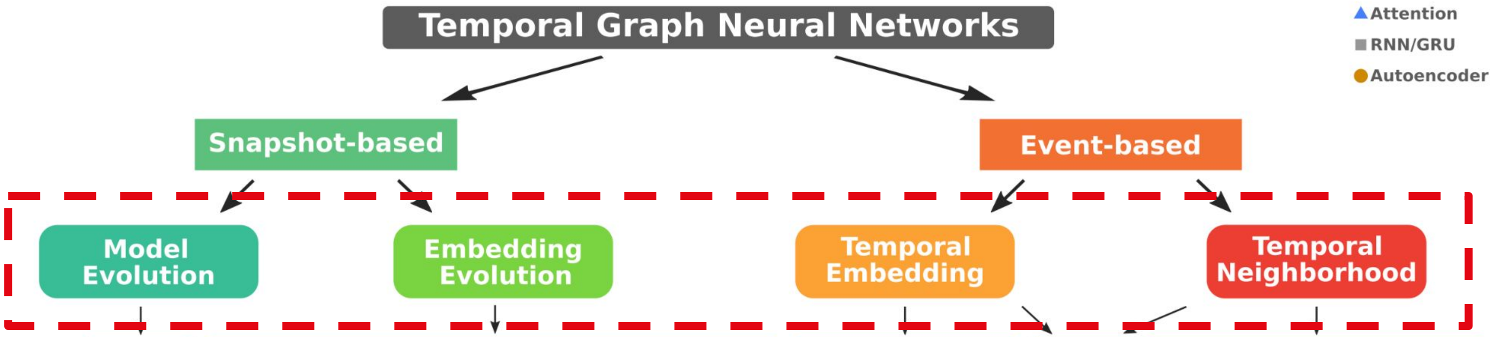
- Temporal Graph clustering (cluster per time sequence)
- Temporal Node clustering (cluster per node)
- Anomaly detection:

Definition 14 (Anomalous node detection) *Given a TG* $G_T = (V, E, V_T, E_T)$*, a TGNN and an application dependent threshold* $C \in (0,1)$*, a node* $v \in V$ *is detected as anomalous if*

$$\boxed{f_{DE}(\mathit{TGNN}(v, G_T; \theta)) < C,}$$

- Node anomaly (e.g virus starting point)
- Edge anomaly (e.g unlikely social interactions)
- Graph anomaly (e.g detecting mutagenic molecule)
- Vector embedding per time point

Taxonomy in TGNN



	EvolveGCN	VGRNN	DySAT	DynGESN	Roland	SSGNN	TGAT	NAT	TGL	APAN	DGNN	TGN
f_{NC}	FT	✗	✗	✗	✗	✗	FT,FI	✗	FT	FT	PT	FT
f_{EC}	FT	✗	✗	✗	✗	✗	✗	✗	✗	FT	✗	✗
f_{GC}	✗	✗	✗	PT	✗	✗	✗	✗	✗	✗	✗	✗
f_{LP}	FT	FI,FT,PT	PT,FT	✗	FT	✗	FT,FI	PI,FT	FT	FT	FT	FI,FT
f_{EP}	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Clustering	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
LDE	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Regression (node)	✗	✗	✗	FT	✗	FT	✗	✗	✗	✗	✗	✗
Anomaly detect.	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

Taxonomy in TGNN - Snapshot based

- **Model evolution**
  - GNN parameters evolve over time
  - e.g EvolveGNN Pareja et al. 2020, using RNN to update GCN
- **Embedding evolution**
  - RNN on embedding outputted by static GNN
  - DySAT : self attention for static node embedding for each time point, self attention per node to with past embeddings to generate new embeddings

Taxonomy in TGNN - Event based

- **Temporal embedding**
  - Explicit embedding for time - extending message passing considering time embedding computed from random fourier features
- **Temporal neighbourhood**
  - Learning node embedding by extending message passing when a neighbour of node is added or deleted

Challenges

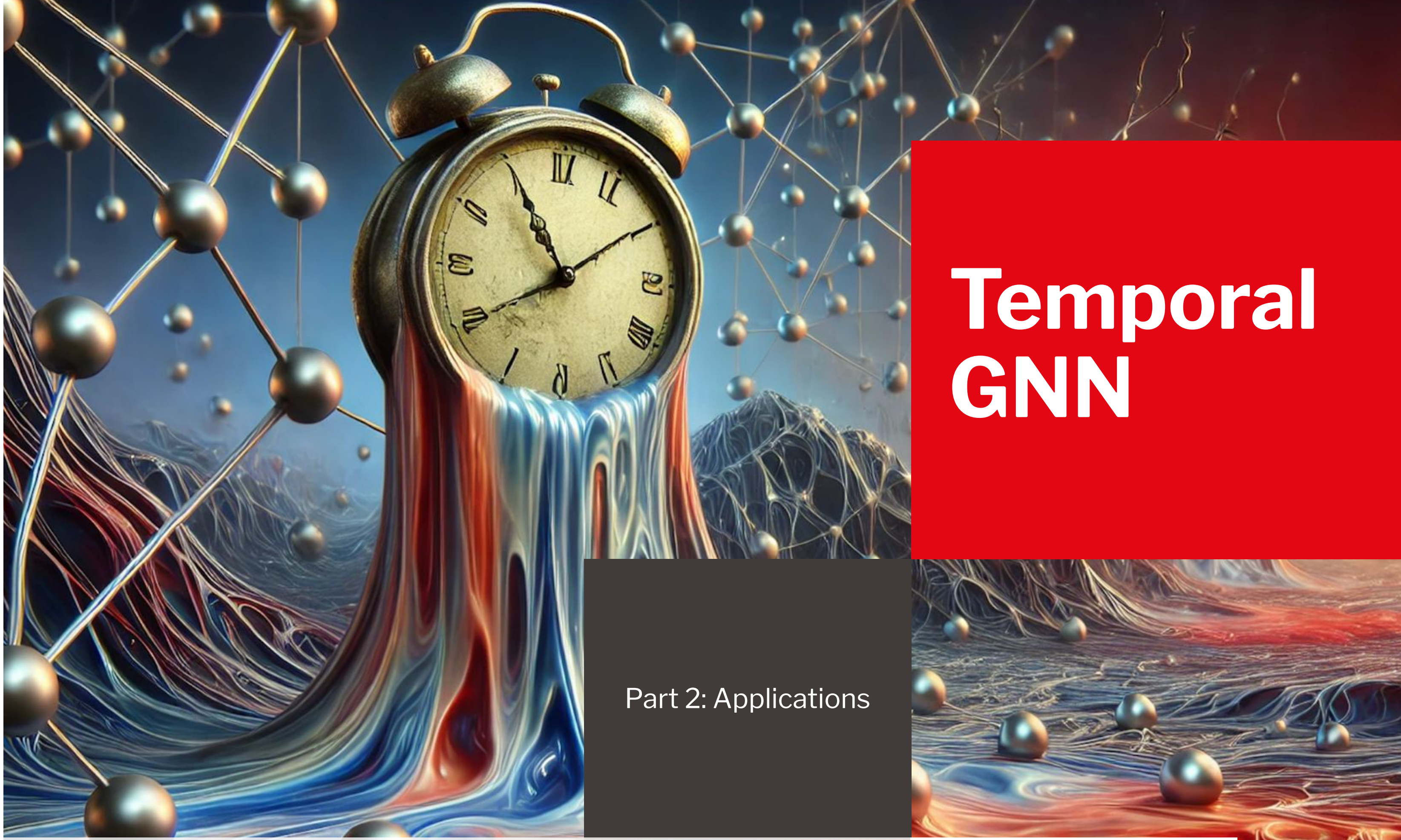
- Evaluation:
  - Standardized dataset / benchmark for performance evaluation
- Expressiveness:
  - Equivalent of Weisfeiler-Lehman isomorphism test in standard GNN - but for TGNN (learning power)
- Learnability:
  - Same problem of over-squashing / over-smoothing as in GNN
- Other future work: More physics driven modelling leveraging knowledge dynamical systems

Takeaway

- Organizes temporally related methods into axis of:
  - Taxonomy (STG / ETG)
  - Learning (Transductive / Inductive)
  - Practical way to handle temporality (model / embedding evolution etc…)
- Gives comprehensive and overall perspective on TGNN

Thanks!

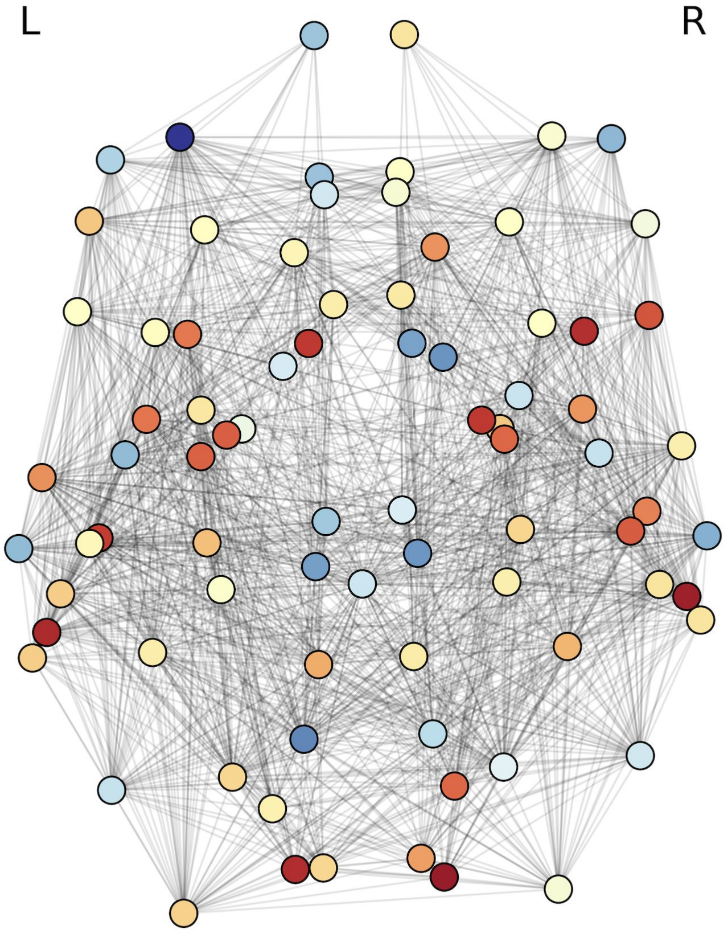
Temporal GNN



OpenAI. (2024). Image generated by ChatGPT using DALL·E, November 2024

Part 2: Applications

Alexandre Cionca



whoami

1st year PhD in EE

Medical Image Processing lab,
Prof. Dimitri Van De Ville

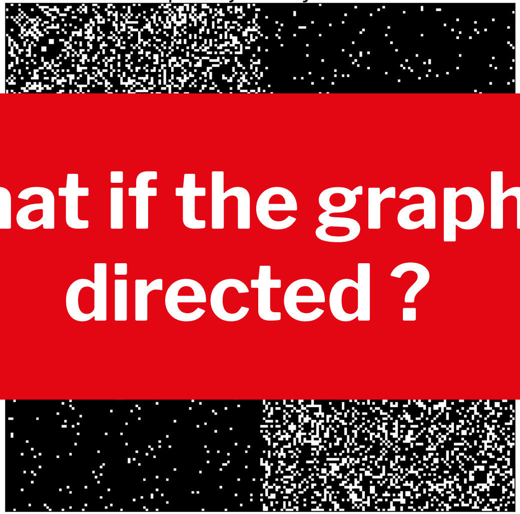
Detecting community in directed brain graphs

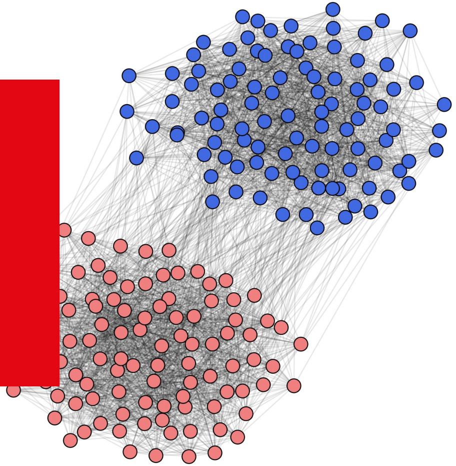
Community detection in graphs

Alexandre Cionca

Graph adjacency matrix

Graph adjacency matrix



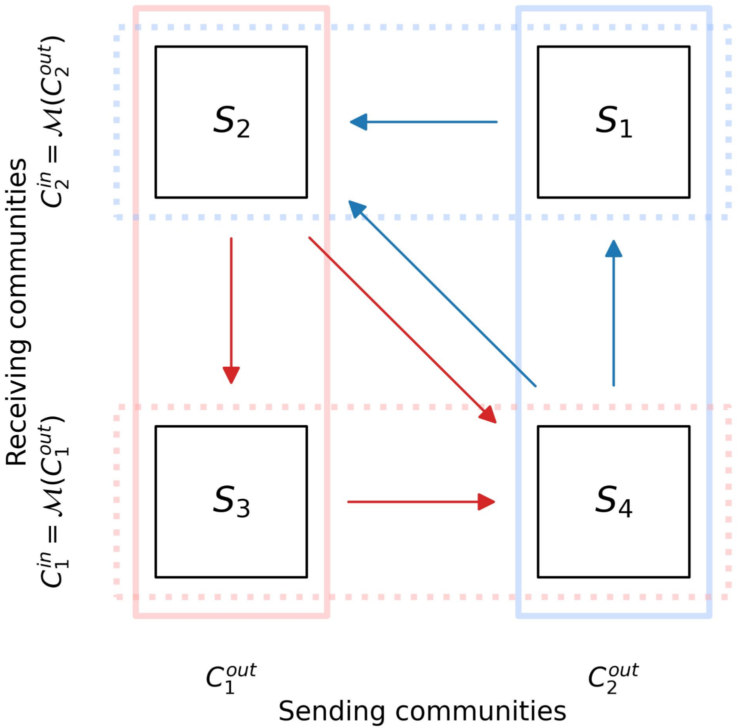


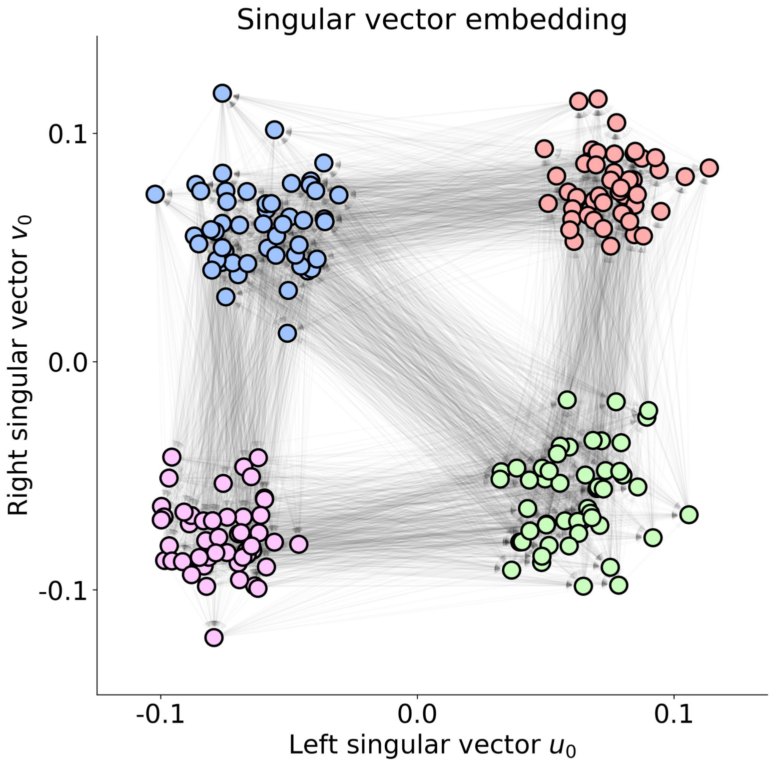
What if the graph is directed ?

Objective function → Optimize → Fine tune

Bimodularity

Consider sending and receiving communities



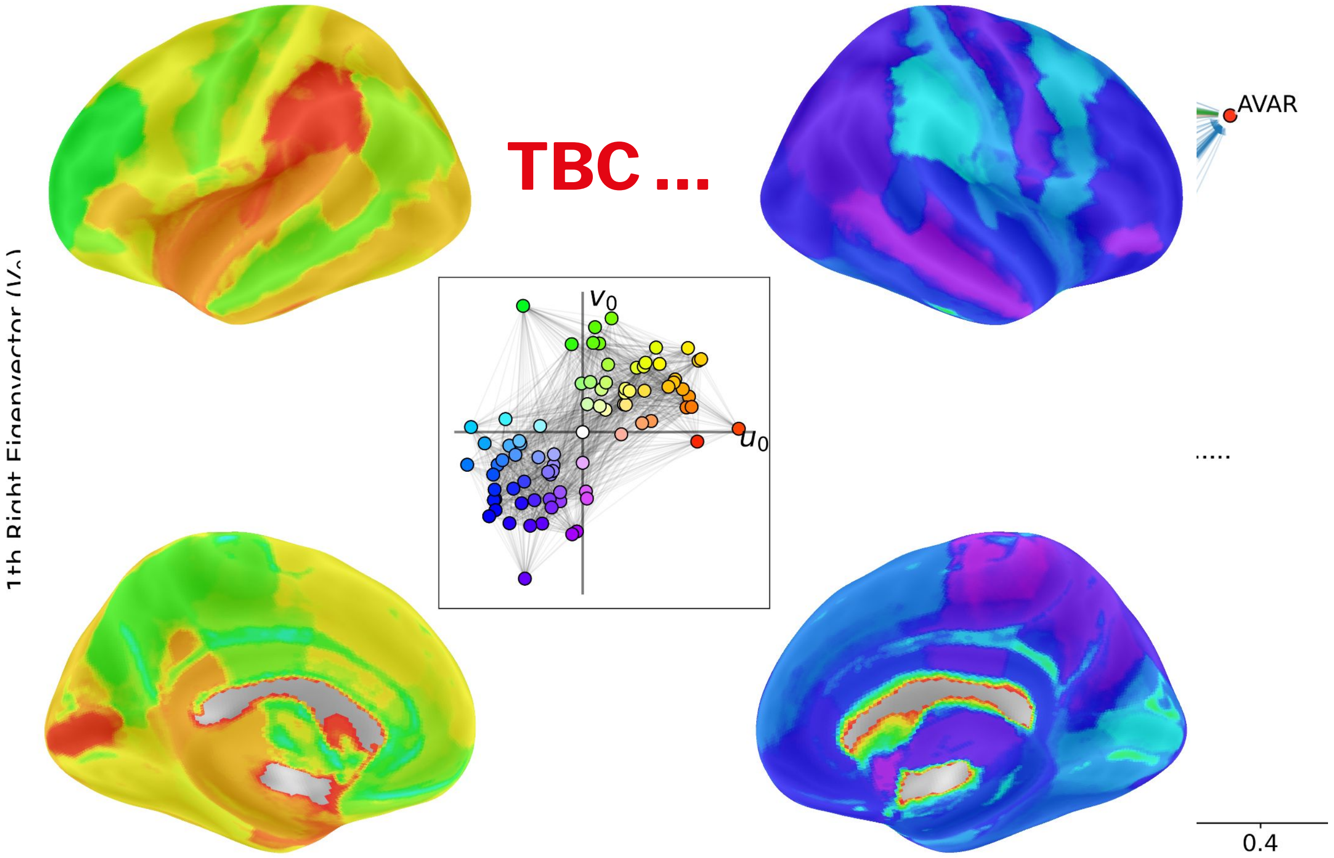


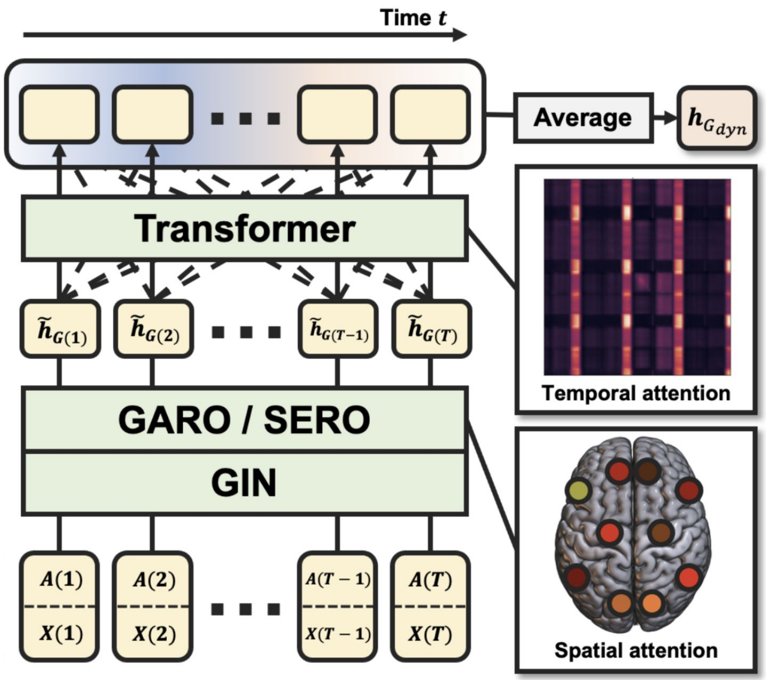
EPFL

Bimodularity

C. elegans directed network

TBC ...





Application #1

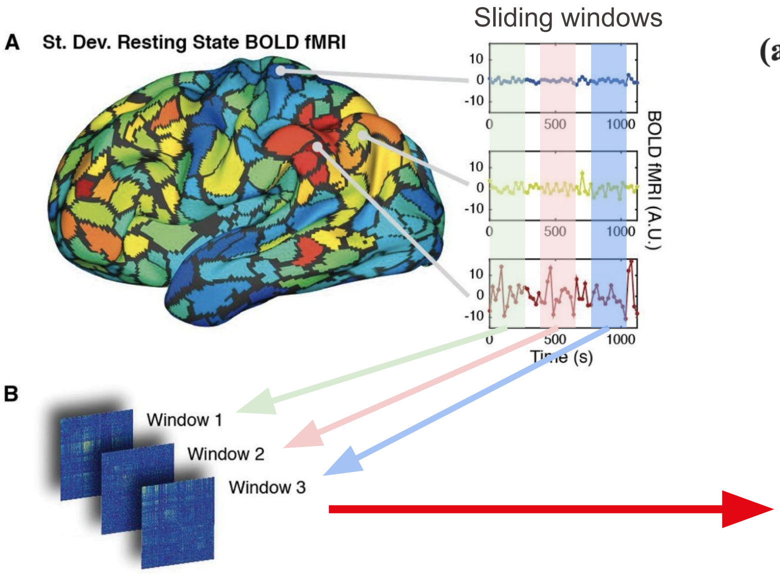
Learning Dynamic Graph Representation of Brain Connectome with Spatio-Temporal Attention

Kim, Byung-Hoon, Jong Chul Ye, and Jae-Jin Kim. "Learning dynamic graph representation of brain connectome with spatio-temporal attention." Advances in Neural Information Processing Systems 34 (2021): 4314-4327.

STAGIN

Spatio-Temporal Attention Graph Isomorphism Network

Dynamic brain connectivity



Gender and task predictions

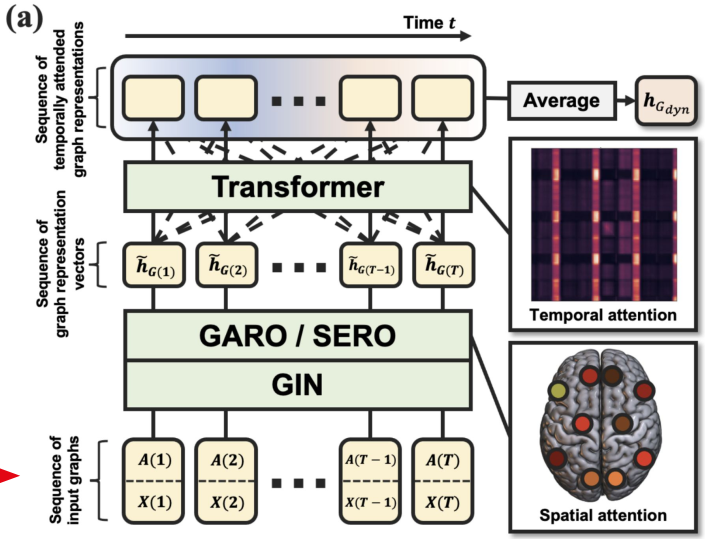


Fig.2 (edited) from Medaglia, John D., and Danielle S. Bassett. "*Network analyses and nervous system disorders*." arXiv preprint arXiv:1701.01101 (2017).

STAGIN

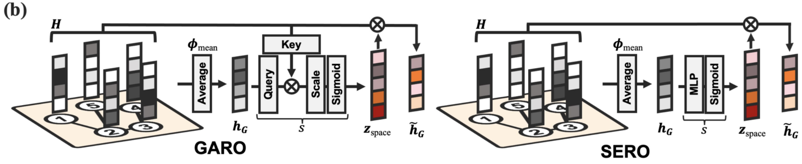
Spatio-Temporal Attention Graph Isomorphism Network

Graph Isomorphism Network

$$h_v^{(k)} = \texttt{MLP}^{(k)}\left(\left(1+\epsilon^{(k)}\right)\cdot h_v^{(k-1)} + \sum_{u\in\mathcal{N}(v)} h_u^{(k-1)}\right)$$

Attention-based readout (graph)

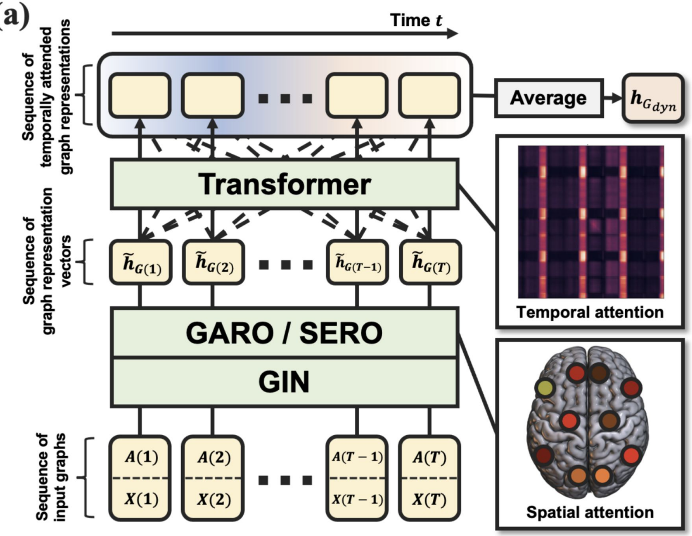
(b)



Single-head transformer (time)

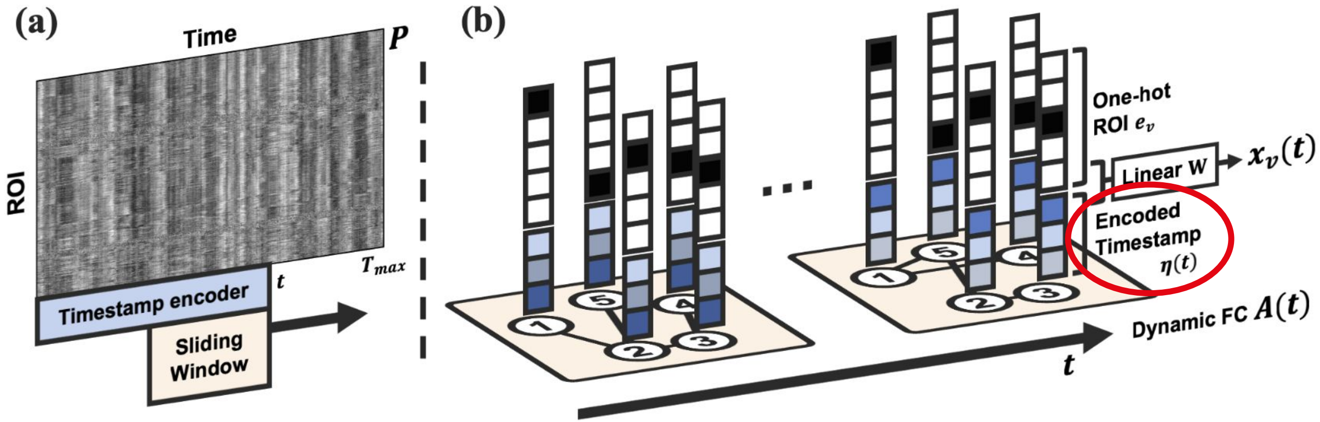
- Self-attention weight Z_{time}

(a)



Time-related features

Encoded timestamps



Learnable timestamp encoder

- Gated Recurrent Unit (GRU) with ROI timeseries as the input

EPFL

Prediction results

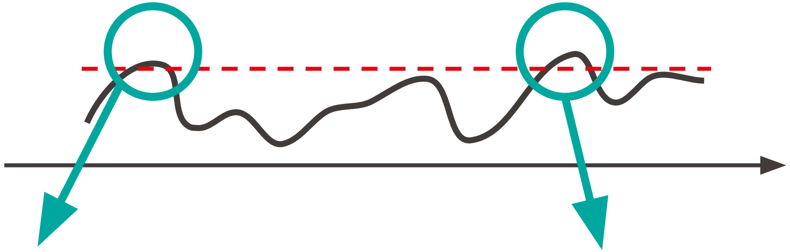
Alexandre Cionca

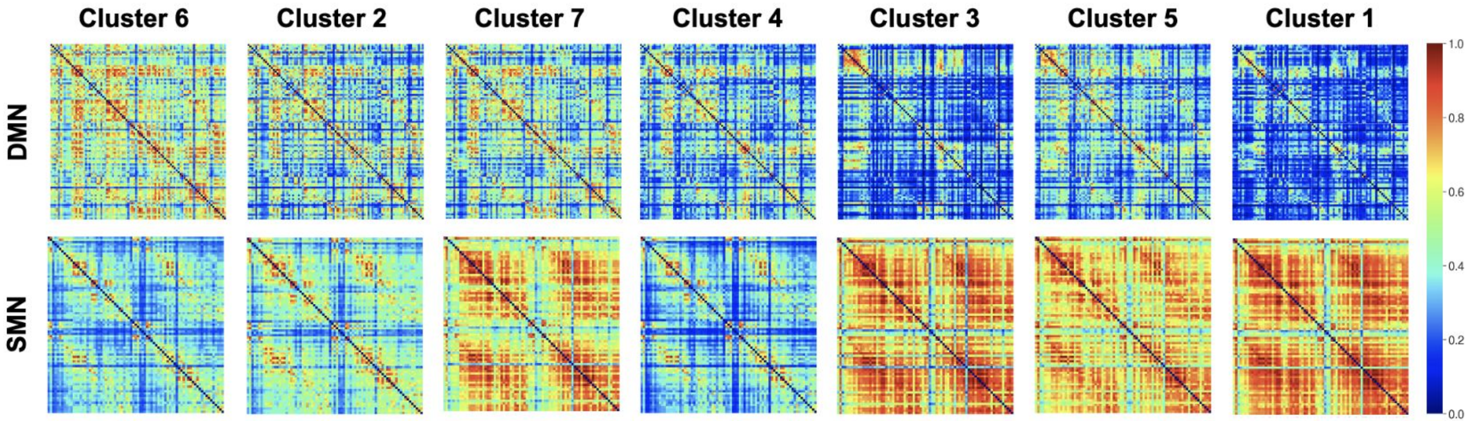
Table 1: Comparative study on HCP-Rest and HCP-Task dataset.

Model	HCP-Rest		HCP-Task	Type of FC	# Params
	Accuracy (%)	AUROC	Accuracy (%)		
STAGIN-SERO	**88.20** $\pm$ 1.33	**0.9296** $\pm$ 0.0187	**99.19** $\pm$ 0.20	Dynamic	1,209k
STAGIN-GARO	87.01 $\pm$ 3.00	0.9151 $\pm$ 0.0258	**99.02** $\pm$ 0.17	Dynamic	1,068k
ST-GCN [15]	76.95 $\pm$ 3.00	0.8545 $\pm$ 0.0316	98.92 $\pm$ 0.27	Dynamic	355k
MS-G3D [10]	79.16 $\pm$ 2.53	0.8912 $\pm$ 0.0329	-	Dynamic	3,045k
BAnD++ [36]	-	-	97.20 $\pm$ 0.57	None	2,010k
BAnD [36]	-	-	95.10 $\pm$ 0.62	None	2,010k
r-BAnD	-	-	98.90 $\pm$ 0.27	Dynamic	664k
GIN [23]	81.34 $\pm$ 2.40	0.8955 $\pm$ 0.0237	93.87 $\pm$ 0.66	Static	169k
GCN [24]	80.79 $\pm$ 2.00	0.8741 $\pm$ 0.0174	45.07 $\pm$ 1.63	Static	101k
GraphSAGE [31]	75.48 $\pm$ 1.97	0.8237 $\pm$ 0.0228	54.52 $\pm$ 0.97	Static	202k
ChebGCN [2]	77.76 $\pm$ 2.09	0.8582 $\pm$ 0.0233	73.06 $\pm$ 0.68	Static	704k

Clustering temporal attention

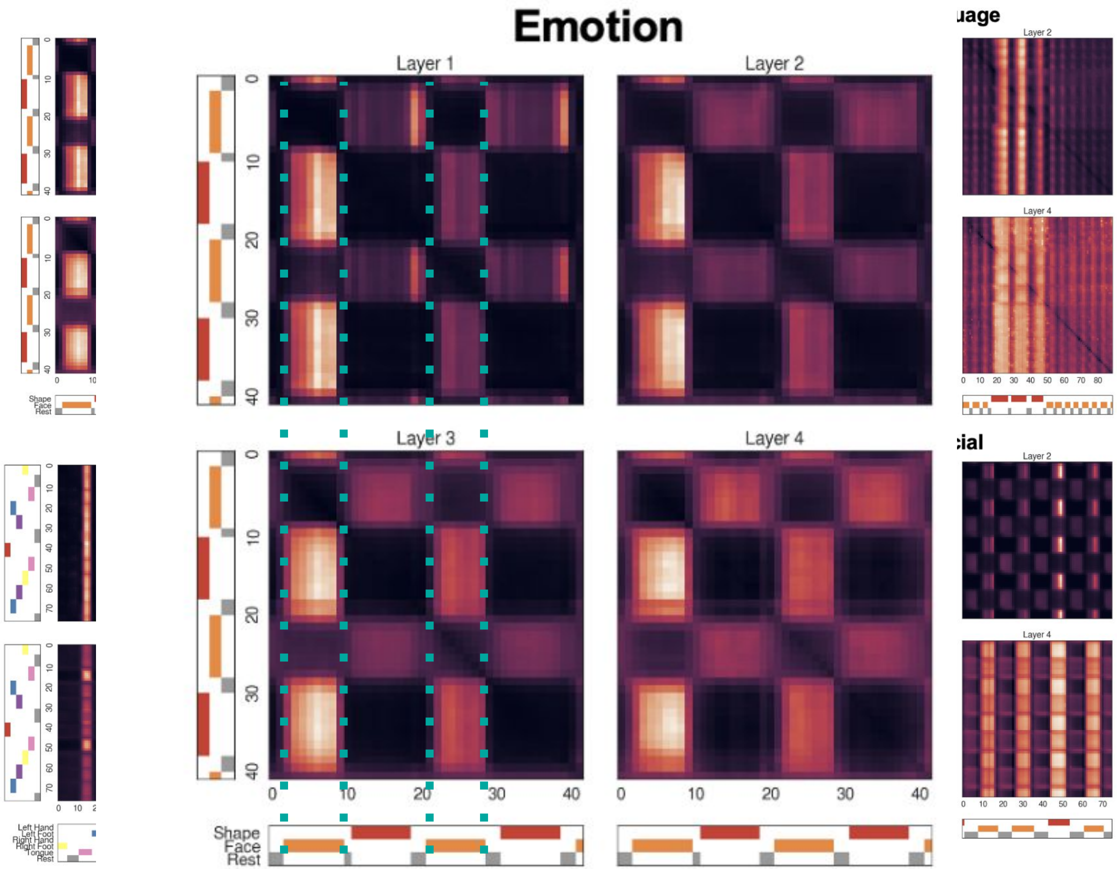
Temporal attention score per timepoint





	Cluster 6	Cluster 2	Cluster 7	Cluster 4	Cluster 3	Cluster 5	Cluster 1
Female	1.299	1.208	1.228	0.807	0.690	0.642	0.576
Male	0.927	0.950	0.967	1.036	1.079	1.041	1.128
F/M	**1.401**	**1.272**	**1.269**	**0.778**	**0.640**	**0.617**	**0.510**

Temporal attention (task)

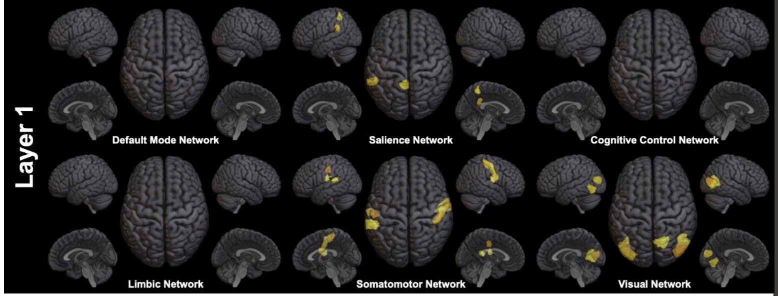


Ablation study

- Atlas Granularity
- Encoded timestamps

Table 3: Ablation study results.

Atlas	N	$\mathcal{L}_{\text{ortho}}$	z_{space}	Z_{time}	$\eta(t)$	Accuracy (%)	AUROC
Schaefer	400	✓	✓	✓	✓	88.20 ± 1.33	0.9296 ± 0.0187
		✗	✓	✓	✓	87.46 ± 3.56	0.9213 ± 0.0242
		✗	✗	✓	✓	86.55 ± 3.12	0.9260 ± 0.0216
		✗	✗	✗	✓	85.64 ± 2.47	0.9272 ± 0.0104
		✗	✗	✗	✗	82.34 ± 3.38	0.9005 ± 0.0256
AAL	116	✓	✓	✓	✓	85.36 ± 1.58	0.9216 ± 0.0116
Destrieux	150	✓	✓	✓	✓	85.73 ± 1.39	0.9235 ± 0.0126
Harvard-oxford	48	✓	✓	✓	✓	82.07 ± 1.11	0.9008 ± 0.0093



HUMAN BRAIN MAPPING

Influence Series

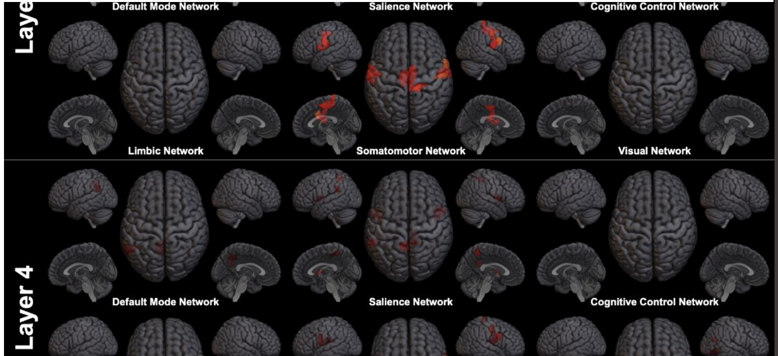
Research Article | Full Access

Functional connectivity predicts gender: Evidence for gender differences in resting brain connectivity

Chao Zhang, Chase C. Dougherty, Stefi A. Baum, Tonya White, Andrew M. Michael

First published: 10 January 2018 | https://doi.org/10.1002/hbm.23950 | Citations: 112

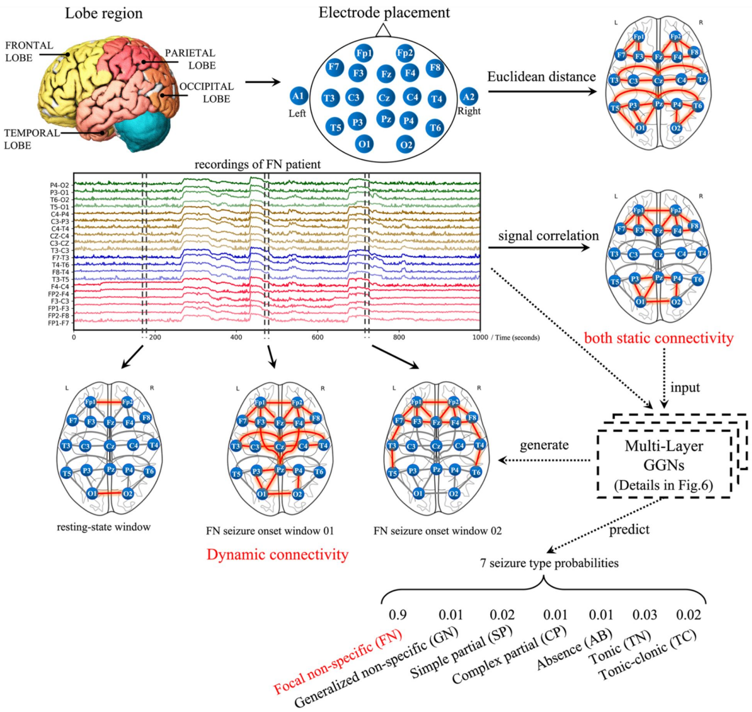
"... we achieved a **gender prediction accuracy of 87%** ..."



Summary

- + Interpretation of temporal attention !
- - Accuracy is setting dependent
- - Gender and task prediction is an "easy" task

What do you think ?



Application #2

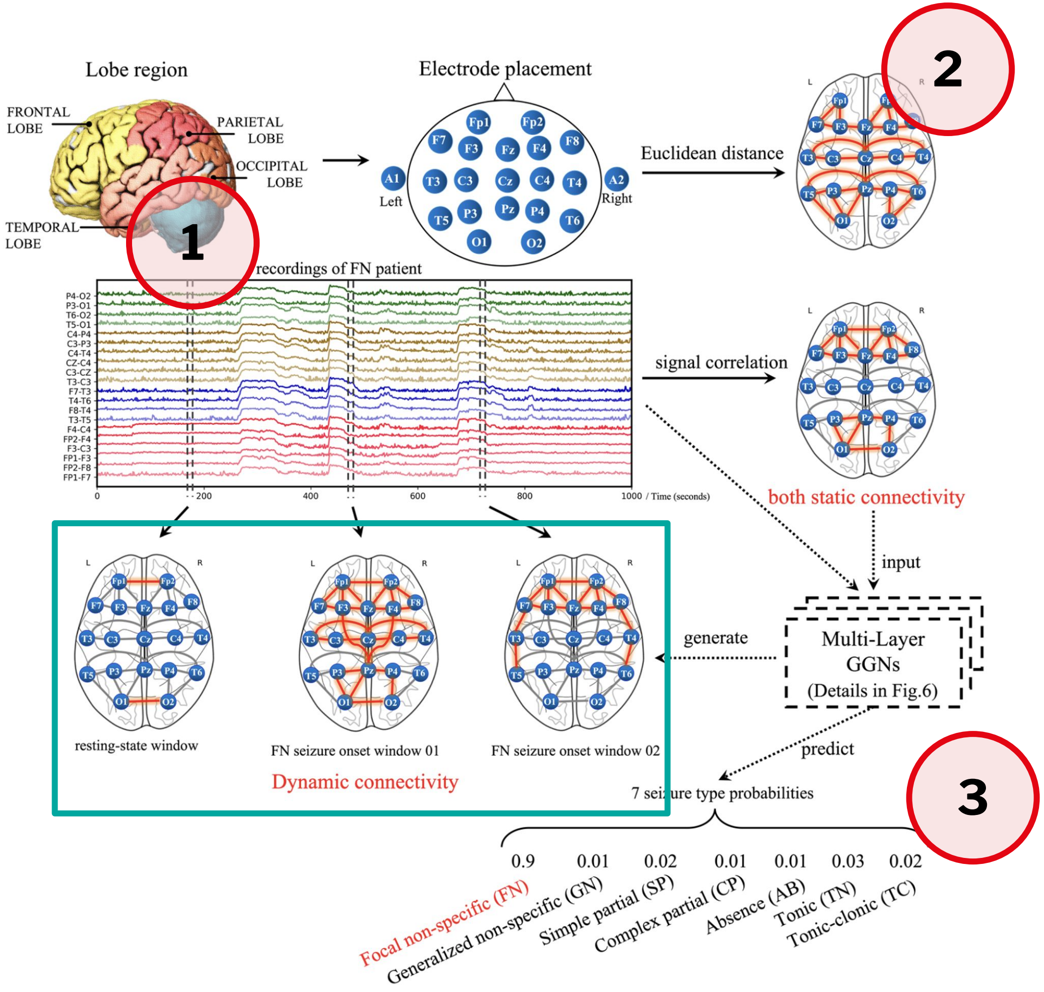
Graph-generative neural network for EEG-based epileptic seizure detection via discovery of dynamic brain functional connectivity

Li, Zhengdao, et al. "Graph-generative neural network for EEG-based epileptic seizure detection via discovery of dynamic brain functional connectivity." Scientific reports 12.1 (2022): 18998.

Scope

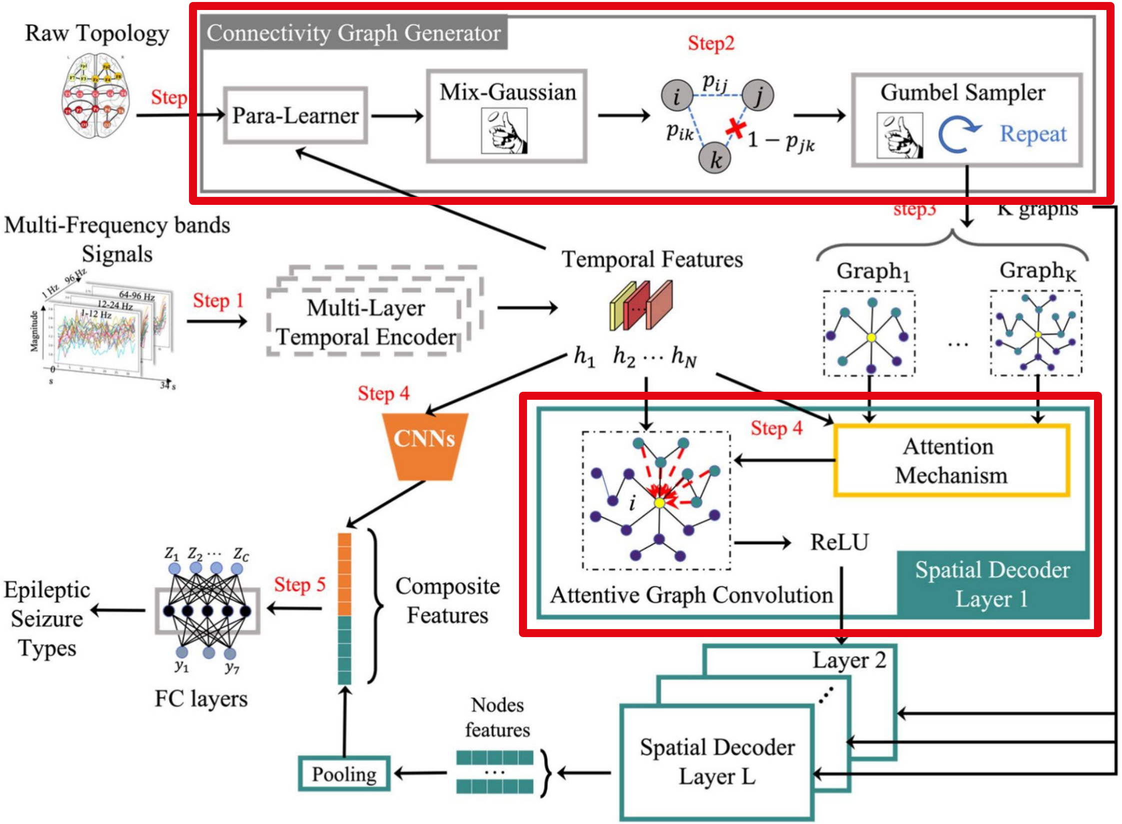
1. EEG acquisition
2. Electrodes graph
3. 7 types of epileptic seizures

- Generate seizure specific graphs (GGN)



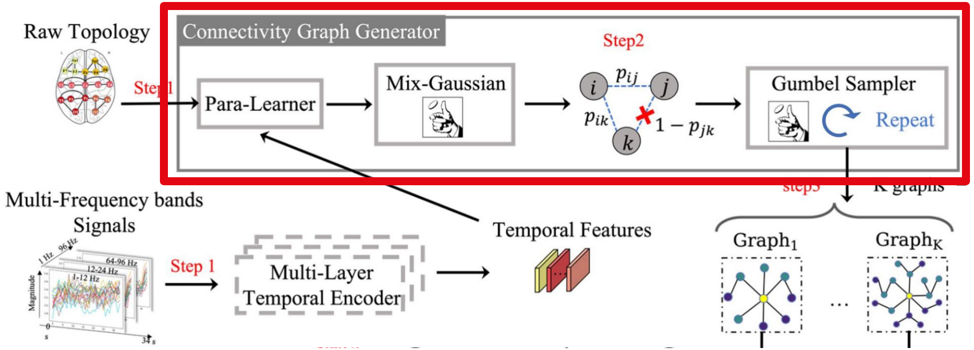
GGN

Graph Generator Neural network



CGN

Connectivity Graph Generator

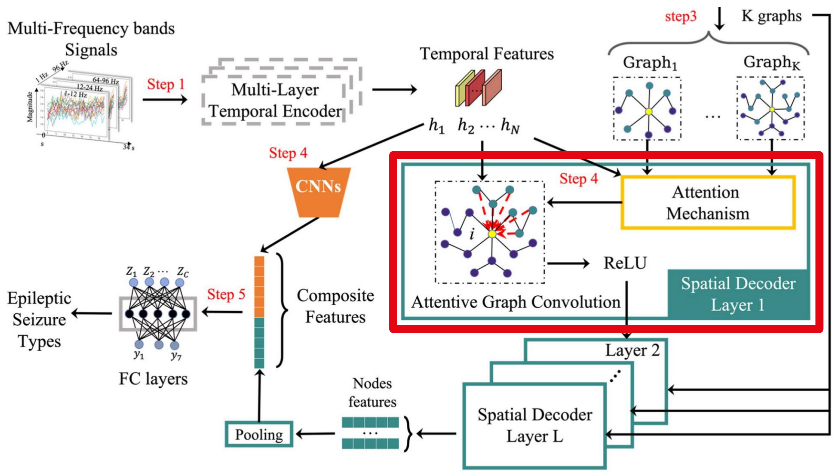


Para-Learner:
- 3 Message passing GNNs
- Raw topology & temporal features as inputs

Mix-Gaussian:
- Generate edge probabilistic distribution

Gumbel Sampler:
- Generate graph edges based on the probability

Spatial decoder



Attention mechanism

- Refines graphs to generate local and global spatial representations
- Learns the best neighboring range for different seizure types

Attentive graph convolution

- Update node representation

EPFL

Good predictive power

a — CNN

The Ground Truth \ Predictions	FN	GN	CP	AB	TN	SP	TC
FN	0.91	0.03	0.06	0	0	0	0
GN	0.38	0.54	0.09	0	0	0	0
CP	0.64	0.04	0.32	0	0	0	0
AB	0	0.18	0	0.82	0	0	0
TN	0.75	0	0.25	0	0	0	0
SP	0.12	0	0	0	0	0.88	0
TC	0.12	0.69	0.12	0	0	0.06	0

c — Transformer

The Ground Truth \ Predictions	FN	GN	CP	AB	TN	SP	TC
FN	0.92	0.03	0.05	0	0	0	0
GN	0.29	0.64	0.05	0.01	0	0	0.02
CP	0.45	0.03	0.51	0	0.01	0	0
AB	0	0	0	1	0	0	0
TN	0.15	0	0	0	0.85	0	0
SP	0.18	0	0	0	0	0.82	0
TC	0	0.25	0.06	0.06	0	0	0.62

b — GNN

The Ground Truth \ Predictions	FN	GN	CP	AB	TN	SP	TC
FN	0.91	0.06	0.03	0	0	0	0
GN	0.24	0.72	0.05	0	0	0	0
CP	0.59	0.07	0.34	0	0	0	0
AB	0	0	0	1	0	0	0
TN	0.4	0.6	0	0	0	0	0
SP	0.12	0	0	0	0	0.88	0
TC	0	0.44	0.25	0.12	0	0	0.19

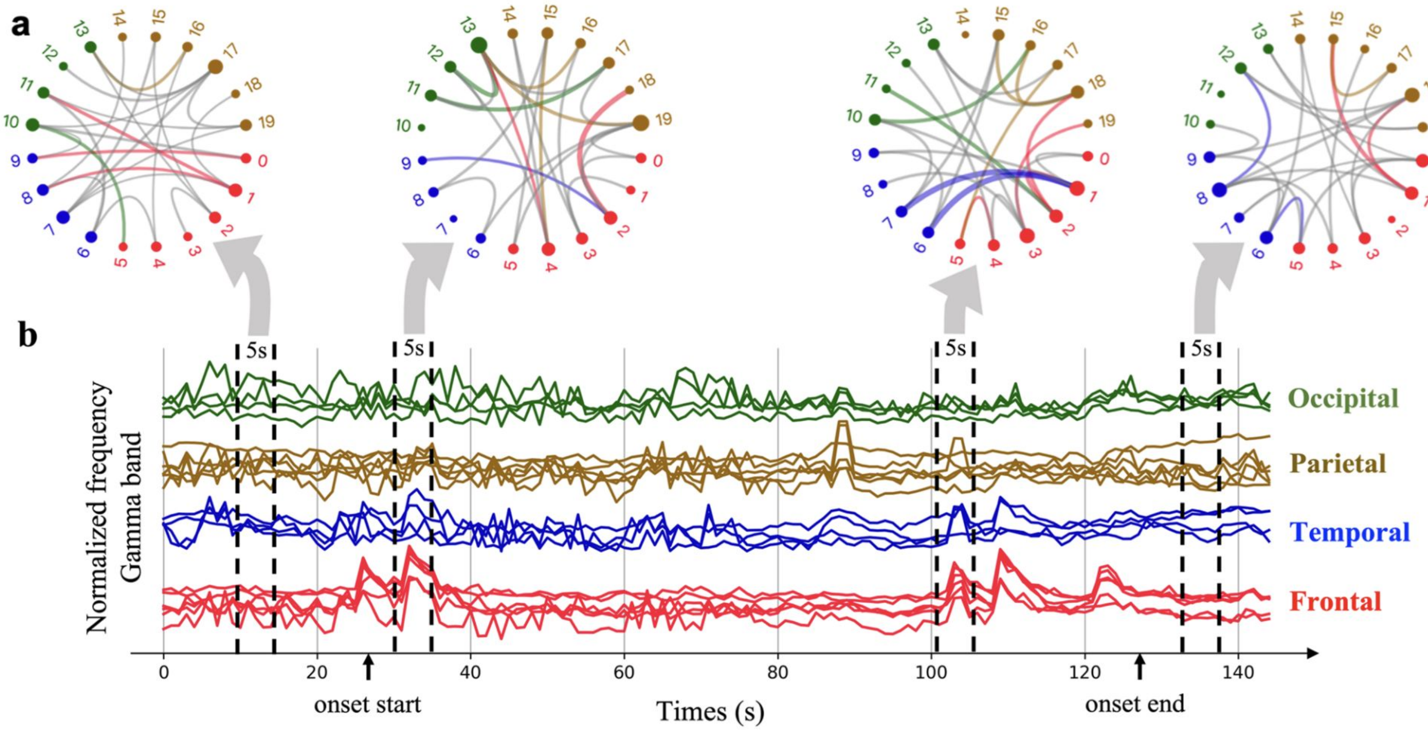
d — GGN

The Ground Truth \ Predictions	FN	GN	CP	AB	TN	SP	TC
FN	0.94	0.03	0.03	0	0	0	0
GN	0.13	0.82	0.04	0	0	0	0.01
CP	0.07	0.03	0.88	0	0	0.01	0.01
AB	0	0	0	1	0	0	0
TN	0	0	0	0	1	0	0
SP	0.06	0.06	0	0	0	0.88	0
TC	0.06	0.06	0	0	0	0	0.88

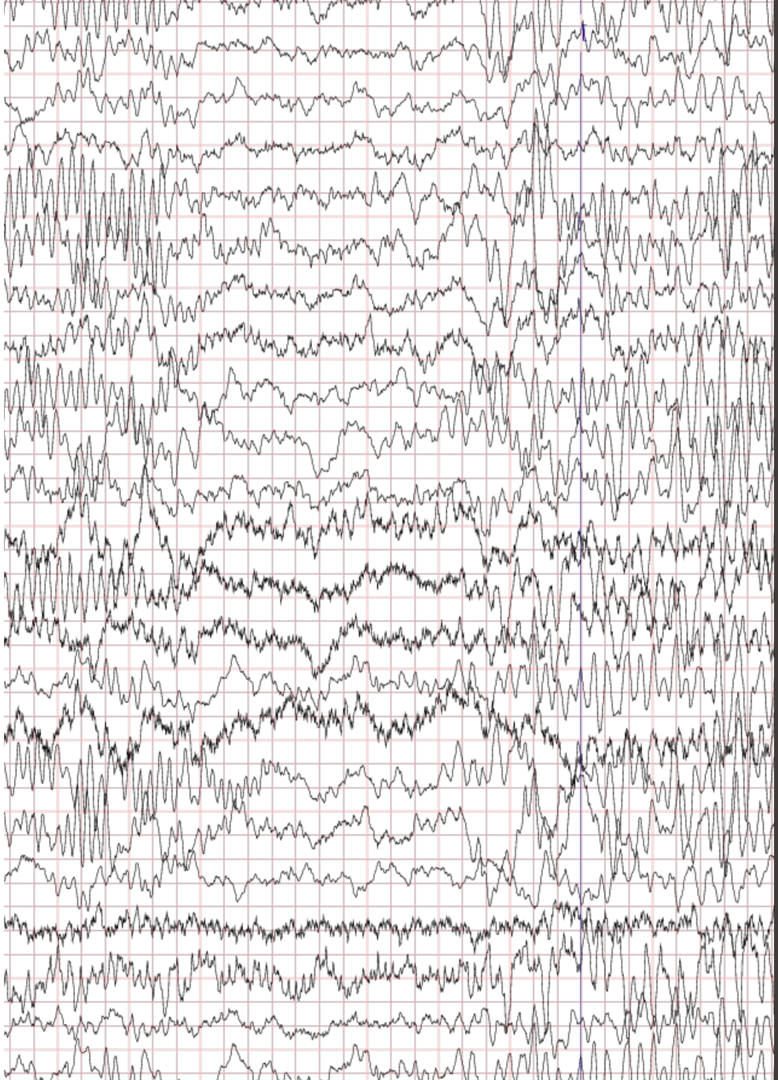
0 0.5 1

Spatio-temporal connectivity

Focal non-specific (FN) onset connectivity (1 patient)



Alexandre Cionca



Summary

+ Generation of (meaningful) graphs!
+ Elegant structure

- Not yet suited for clinical application

What do you think ?