IJCAI 2019 MACAO

Spatio-Temporal Attentive RNN for Node Classification in Temporal Attributed Graphs

Dongkuan Xu[1], Wei Cheng[2], Dongsheng Luo[1], Xiao Liu[1], Xiang Zhang[1]

[1]The Pennsylvania State University

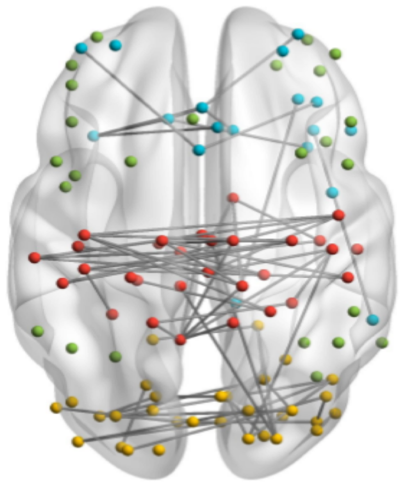
[2]NEC Laboratories America

PennState

NEC Laboratories America

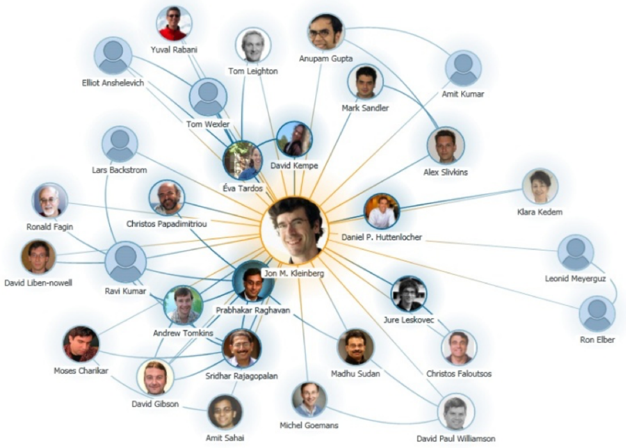
Relentless passion for innovation

Graph-Structure Data is Prevalent

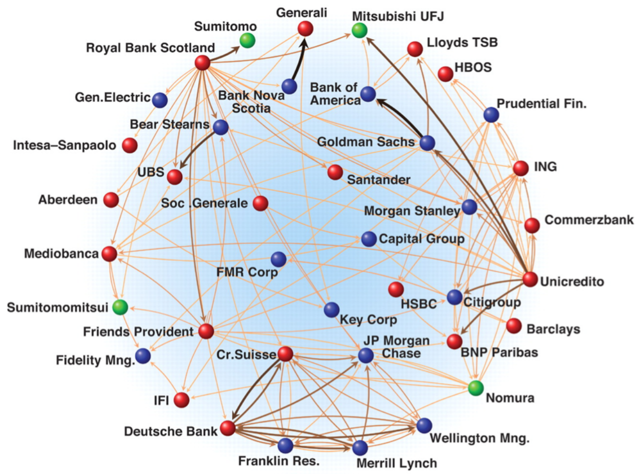


Brain Graph

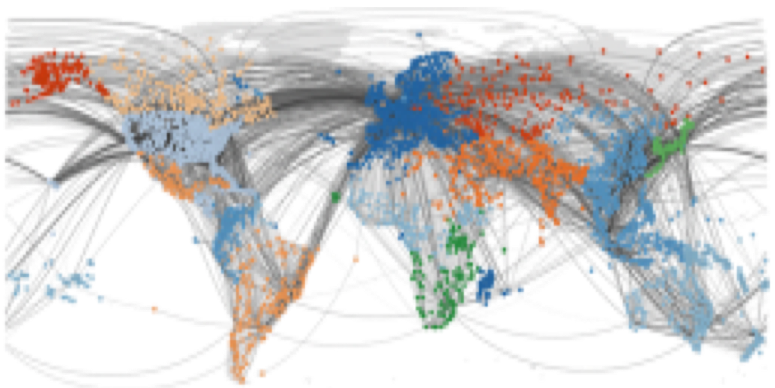
Social Graph



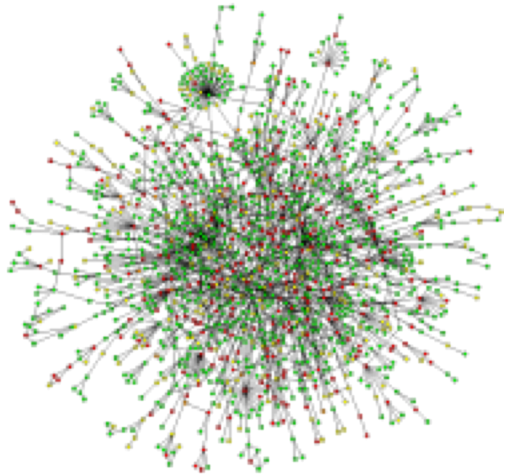
Co-author Graph



Financial Graph



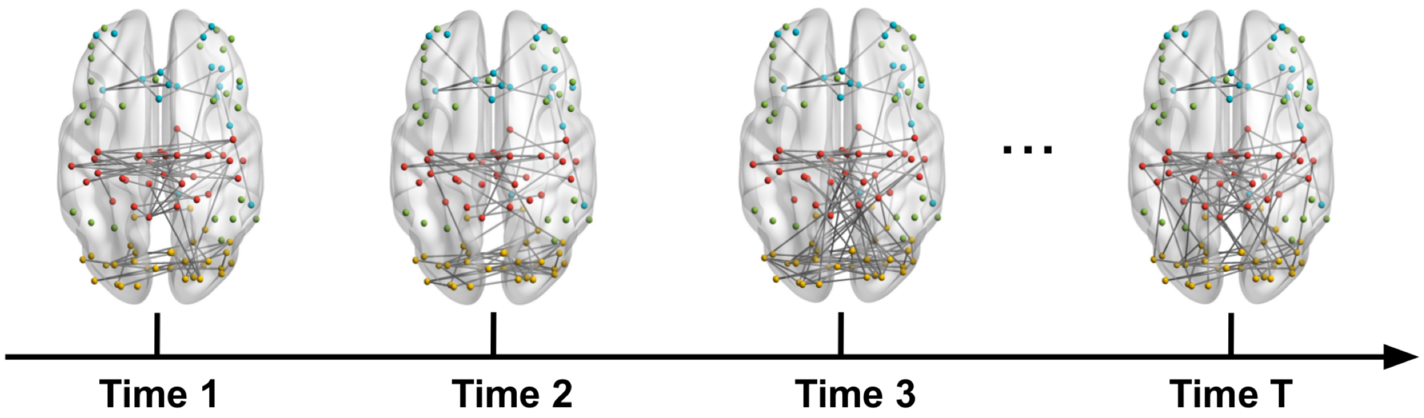
Traffic Graph



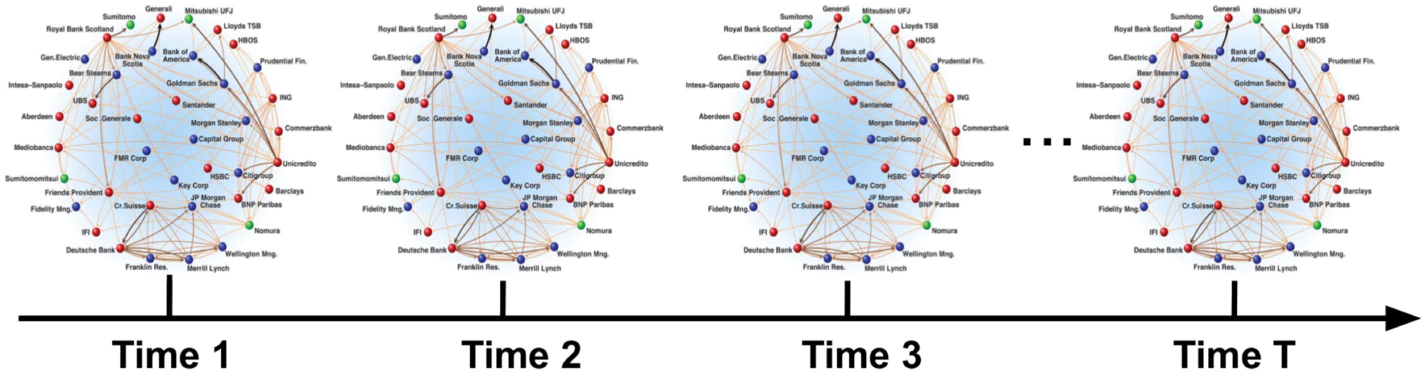
Protein-Protein-Interaction Graph

Temporal Graphs

- Graphs that vary over time
  - Graph topology + node attributes



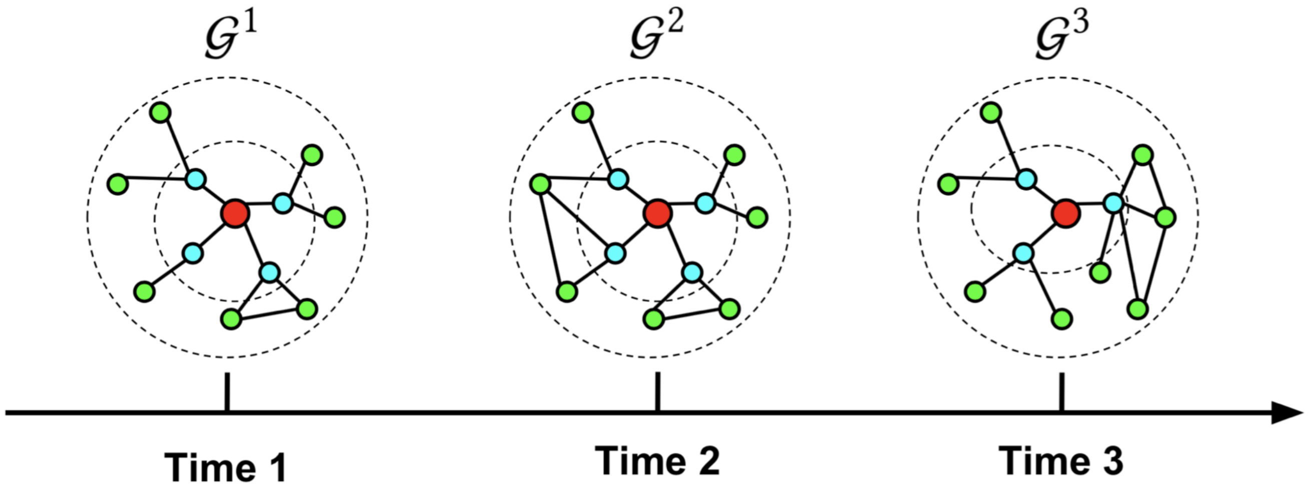
Temporal Brain Graph



Temporal Financial Graph

Problem Definition

- Node Classification in Temporal Attributed Graph



Goal: What Is The Category of The Red Node

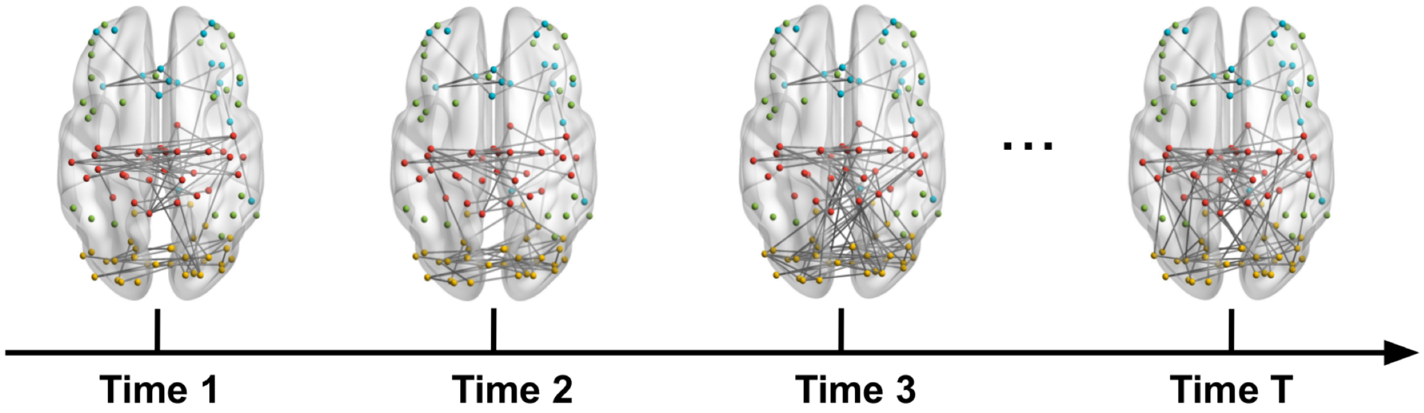
- Notations
  - Temporal attributed graph $\mathbb{G} = (\mathcal{G}^1, \mathcal{G}^2, \cdots, \mathcal{G}^T)$, where $\mathcal{G}^t = (\mathcal{V}, \mathbf{A}^t, \mathbf{X}^t)$

    Node set $\mathcal{V}$, Adjacency matrix $\mathbf{A}^t \in \mathbb{R}^{N \times N}$, Node attributes $\mathbf{X}^t \in \mathbb{R}^{N \times d}$
- Definition
  - **Given: $\mathbb{G}$ and labels of a subset of nodes $\mathcal{V}_L$**
  - **Goal: to classify the nodes in subset $\mathcal{V}_U$, where $\mathcal{V} = \mathcal{V}_L \cup \mathcal{V}_U$**

Challenges

- Spatio-temporal contextual information
  - Topology and attributes
  - Temporal and spatial entangled

- Different factors influencing node representations
  - Time steps
  - Local neighbors



Dynamic Brain Graph

Existing Approaches

- **Static Graphs**
  - Random walk based: Deepwalk [1]
  - Convolutional neural network based: GCN [2]
  - Local sampling based: GraphSAGE [3]
  - Neural attention mechanism based: GAT [4]
- **Temporal Graphs**
  - Matrix factorization based: SLIDE [5]
  - Triad structure based: DynamicTriad [6]
  - Point process based: HTNE [7]
  - Neural network based: DynGEM [8]

- **STAR**

  Temporal graphs

  Handle attributes & topology

  Differentiate relative importance of different factors

1. Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In *SIGKDD*.
2. *Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.*
3. Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. In *NeurIPS.*
4. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. In *ICLR*.
5. *Li, J., Cheng, K., Wu, L., & Liu, H. (2018, February). Streaming link prediction on dynamic attributed networks. In SWDM.*
6. *Zhou, L., Yang, Y., Ren, X., Wu, F., & Zhuang, Y. (2018, April). Dynamic network embedding by modeling triadic closure process. In AAAI.*
7. *Zuo, Y., Liu, G., Lin, H., Guo, J., Hu, X., & Wu, J. (2018, July). Embedding temporal network via neighborhood formation. In SIGKDD.*
8. *Goyal, P., Kamra, N., He, X., & Liu, Y. (2018). Dyngem: Deep embedding method for dynamic graphs. arXiv preprint arXiv:1805.11273.*

Motivations

- Consider node attributes at different time steps
- Consider neighborhood at different time steps
- Differentiate importance of different time steps
- Differentiate importance of different neighbors

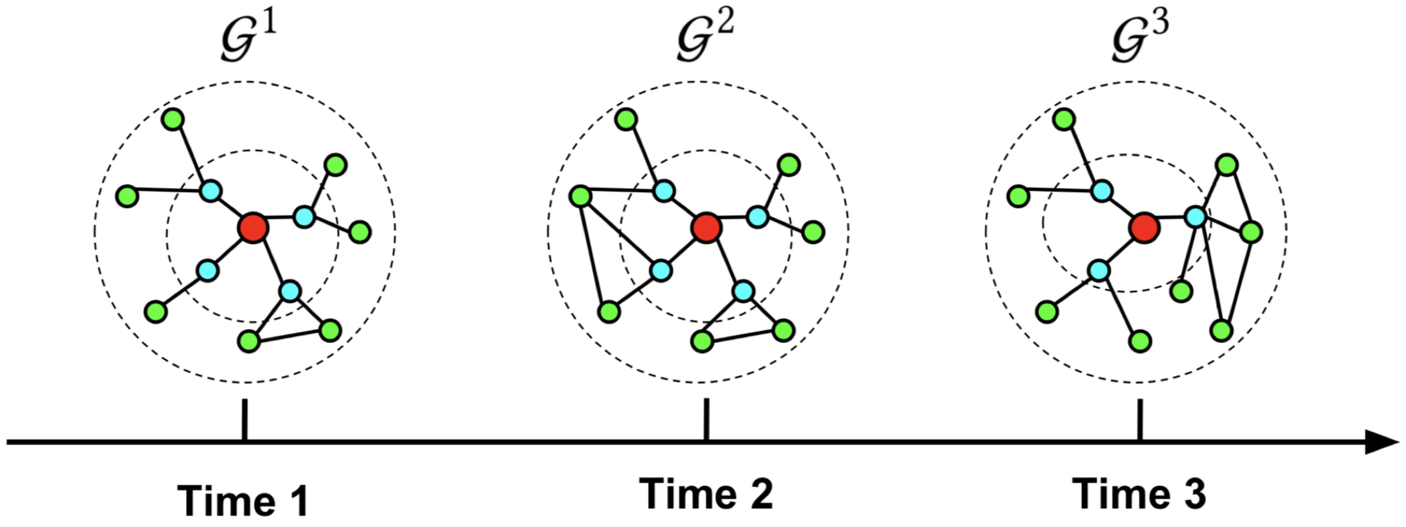
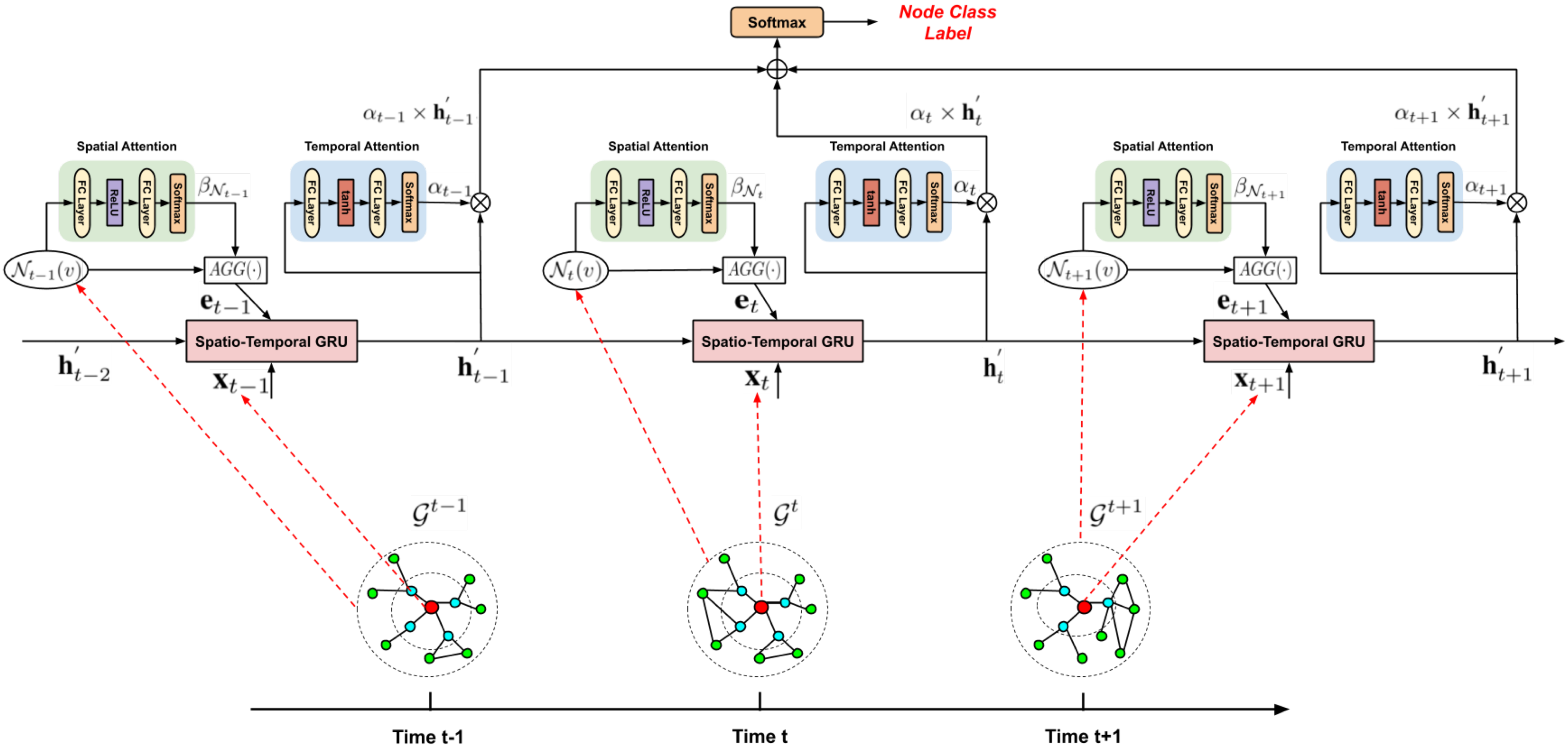


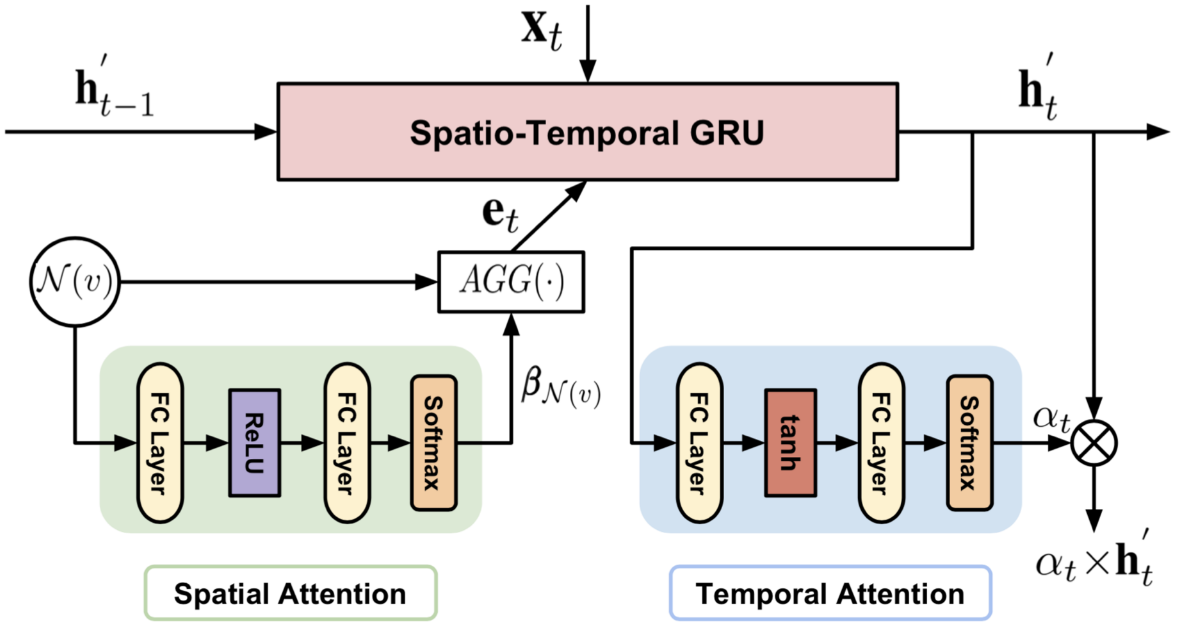
Diagram of temporal graph with 3 snapshots

Spatio-temporal attentive recurrent neural network model (STAR)

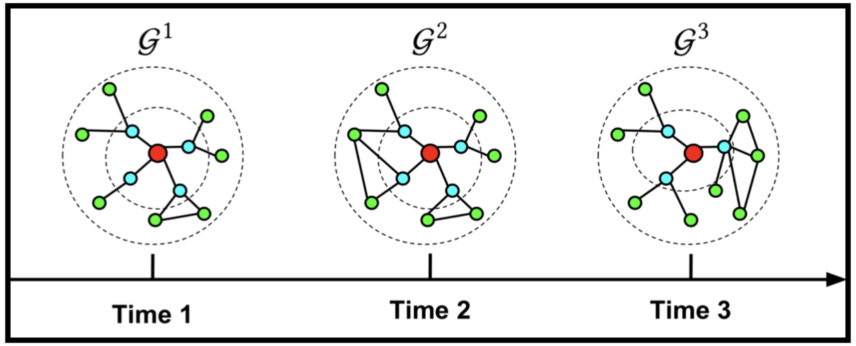


Temporal Attributed Graphs with 3 Snapshots

Architecture of STAR at each time step

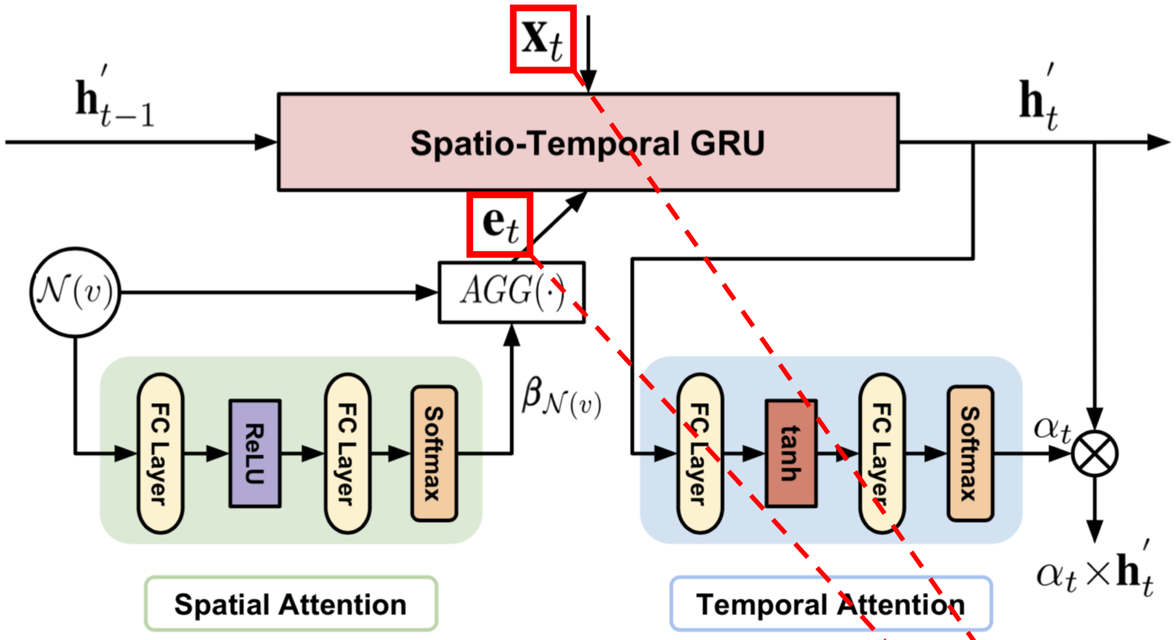


- Components:
  - Spatio-temporal gated recurrent unit
  - Spatial attention module
  - Temporal attention module

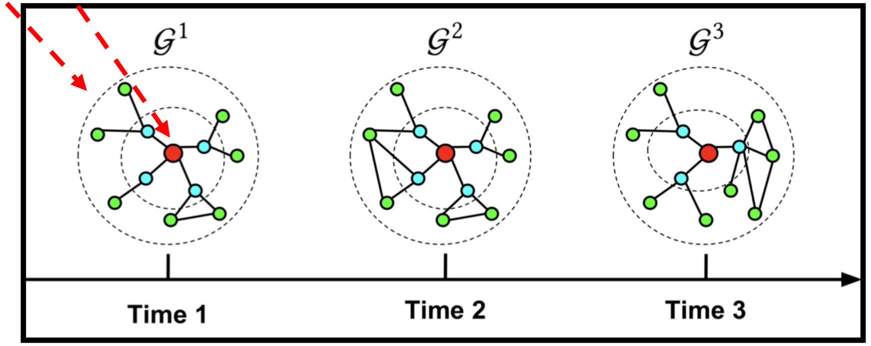


Temporal graph with 3 snapshots

Architecture of STAR at each time step



- Components:
  - Spatio-temporal gated recurrent unit
  - Spatial attention module
  - Temporal attention module



Temporal graph with 3 snapshots

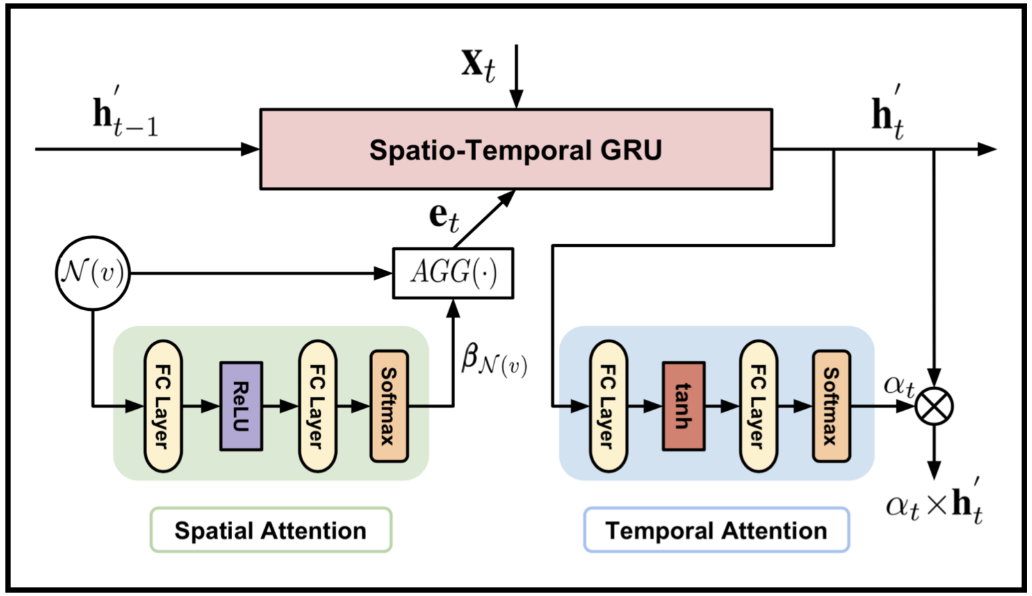
STAR

- Objective function

$$J = L_{ce} + \lambda_1 P_{att} + \lambda_2 P_{nn}$$

- Spatio-temporal gated recurrent unit

$$\begin{aligned}
\mathbf{z}_t^{'} &= \sigma(\mathbf{W}_z^{'}[\mathbf{x}_t \oplus \mathbf{h}_{t-1}^{'} \oplus \mathbf{e}_t] + \mathbf{b}_z^{'}),\\
\mathbf{r}_t^{'} &= \sigma(\mathbf{W}_r^{'}[\mathbf{x}_t \oplus \mathbf{h}_{t-1}^{'} \oplus \mathbf{e}_t] + \mathbf{b}_r^{'}),\\
\mathbf{s}_t^{'} &= \sigma(\mathbf{W}_s^{'}[\mathbf{x}_t \oplus \mathbf{h}_{t-1}^{'} \oplus \mathbf{e}_t] + \mathbf{b}_s^{'}),\\
\tilde{\mathbf{h}}_t^{'} &= \tanh(\mathbf{W}_h^{'}[\mathbf{x}_t \oplus (\mathbf{r}_t^{'} \odot \mathbf{h}_{t-1}^{'}) \oplus (\mathbf{s}_t^{'} \odot \mathbf{e}_t)] + \mathbf{b}_h^{'}),\\
\mathbf{h}_t^{'} &= (\mathbf{1} - \mathbf{z}_t^{'}) \odot \mathbf{h}_{t-1}^{'} + \mathbf{z}_t^{'} \odot \tilde{\mathbf{h}}_t^{'},
\end{aligned}$$



Overall architecture of STAR

- Spatial attention module

$$\beta_u^k = \frac{\exp\{F(\mathbf{w}_k^\top[\mathbf{V}_k \mathbf{g}_{t(u)}^k \oplus \mathbf{V}_k \mathbf{g}_{t(v)}^k])\}}{\sum_{v' \in \mathcal{N}(v)} \exp\{F(\mathbf{w}_k^\top[\mathbf{V}_k \mathbf{g}_{t(v')}^k \oplus \mathbf{V}_k \mathbf{g}_{t(v)}^k])\}}$$

- Temporal attention module

$$\alpha_t = \frac{\exp\{\tilde{\mathbf{w}}^\top \tanh(\tilde{\mathbf{V}} \mathbf{h}_t^{'})\}}{\sum_{i=1}^T \exp\{\tilde{\mathbf{w}}^\top \tanh(\tilde{\mathbf{V}} \mathbf{h}_i^{'})\}}$$

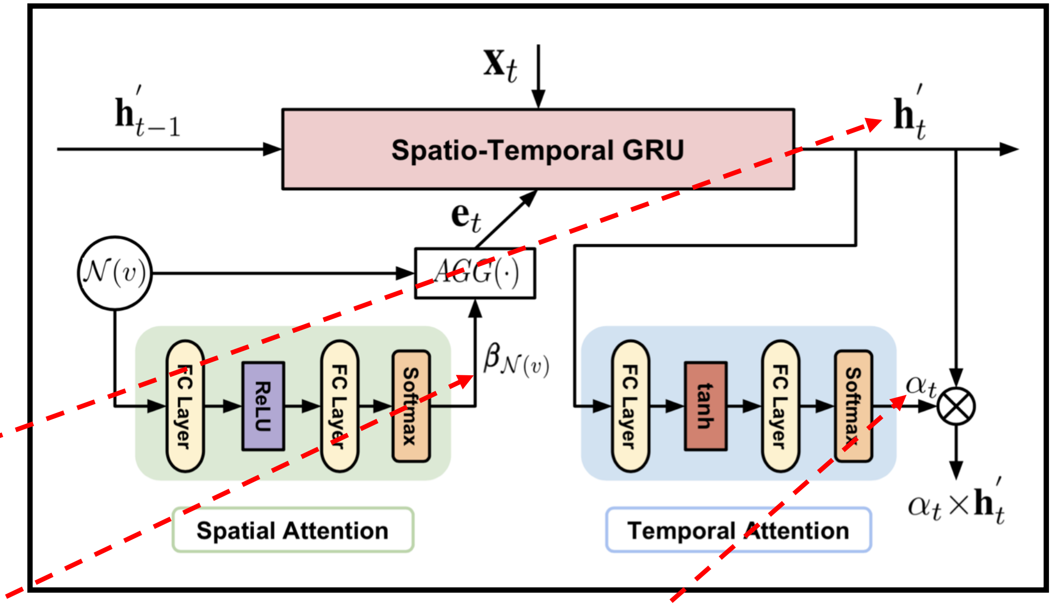
STAR

- Objective function

$$J = L_{ce} + \lambda_1 P_{att} + \lambda_2 P_{nn}$$

- Spatio-temporal gated recurrent unit

$$\begin{aligned}
\mathbf{z}_t^{'} &= \sigma(\mathbf{W}_z^{'}[\mathbf{x}_t \oplus \mathbf{h}_{t-1}^{'} \oplus \mathbf{e}_t] + \mathbf{b}_z^{'}),\\
\mathbf{r}_t^{'} &= \sigma(\mathbf{W}_r^{'}[\mathbf{x}_t \oplus \mathbf{h}_{t-1}^{'} \oplus \mathbf{e}_t] + \mathbf{b}_r^{'}),\\
\mathbf{s}_t^{'} &= \sigma(\mathbf{W}_s^{'}[\mathbf{x}_t \oplus \mathbf{h}_{t-1}^{'} \oplus \mathbf{e}_t] + \mathbf{b}_s^{'}),\\
\tilde{\mathbf{h}}_t^{'} &= \tanh(\mathbf{W}_h^{'}[\mathbf{x}_t \oplus (\mathbf{r}_t^{'} \odot \mathbf{h}_{t-1}^{'}) \oplus (\mathbf{s}_t^{'} \odot \mathbf{e}_t)] + \mathbf{b}_h^{'}),\\
\boxed{\mathbf{h}_t^{'}} &= (\mathbf{1} - \mathbf{z}_t^{'}) \odot \mathbf{h}_{t-1}^{'} + \mathbf{z}_t^{'} \odot \tilde{\mathbf{h}}_t^{'},
\end{aligned}$$



Overall architecture of STAR

- Spatial attention module

$$\boxed{\beta_u^k} = \frac{\exp\{F(\mathbf{w}_k^{\top}[\mathbf{V}_k \mathbf{g}_{t(u)}^k \oplus \mathbf{V}_k \mathbf{g}_{t(v)}^k])\}}{\sum_{v' \in \mathcal{N}(v)} \exp\{F(\mathbf{w}_k^{\top}[\mathbf{V}_k \mathbf{g}_{t(v')}^k \oplus \mathbf{V}_k \mathbf{g}_{t(v)}^k])\}}$$

- Temporal attention module

$$\boxed{\alpha_t} = \frac{\exp\{\tilde{\mathbf{w}}^{\top} \tanh(\tilde{\mathbf{V}} \mathbf{h}_t^{'})\}}{\sum_{i=1}^{T} \exp\{\tilde{\mathbf{w}}^{\top} \tanh(\tilde{\mathbf{V}} \mathbf{h}_i^{'})\}}$$

Baseline Methods & Datasets

Comparison of baseline methods

	Method	Models Temporal	Models Neighborhood	Handles Attribute	Applies Attention
For static graphs	DeepWalk	×	✓	×	×
	node2vec	×	✓	×	×
	GCN	×	✓	✓	×
	GraphSAGE	×	✓	✓	×
For temporal graphs	LSTM	✓	×	✓	×
	GRU	✓	×	✓	×
	DynGEM	✓	✓	×	×
	DynAERNN	✓	✓	×	×
	STAR	✓	✓	✓	✓

Description of the datasets

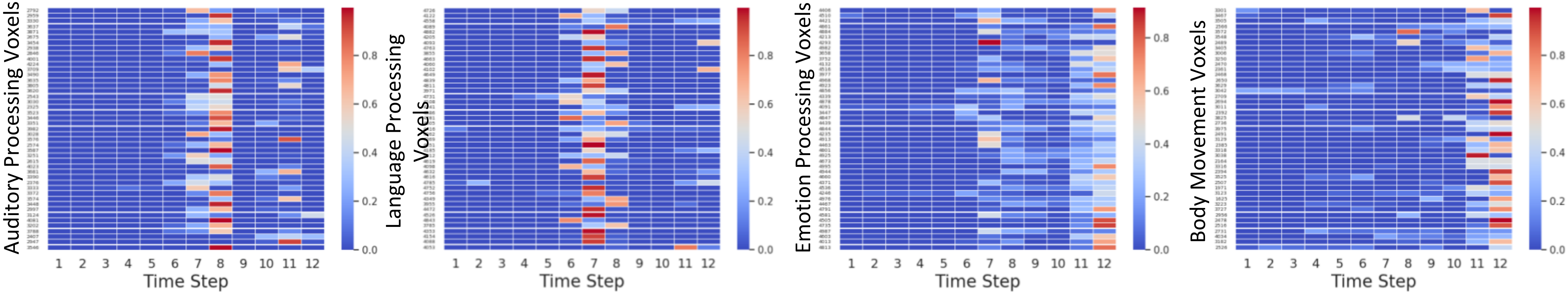
Dataset	# Nodes	# Edges	# Attributes	# Time Steps	# Classes
Brain	5000	1955488	20	12	10
Reddit	8291	264050	20	10	4
DBLP-5	6606	42815	100	10	5
DBLP-3	4257	23540	100	10	3

Experiment Results

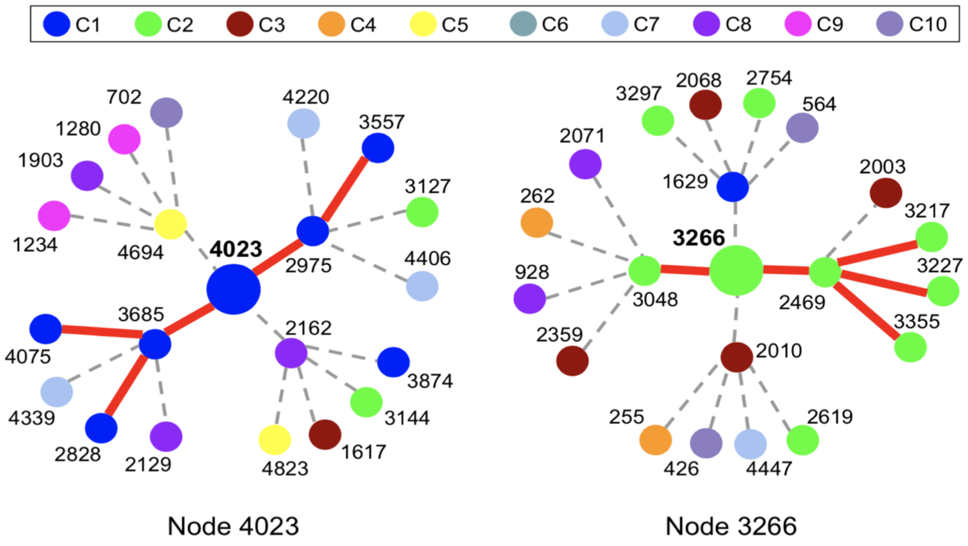
Node classification comparison (%)

	Method	Brain			DBLP-3			DBLP-5			Reddit		
		ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
For static graphs	DeepWalk	71.4	97.2	70.2	49.7	60.1	50.5	35.4	61.0	26.9	47.5	71.9	46.8
	node2vec	71.0	96.8	70.6	51.6	63.0	51.6	36.9	64.2	27.2	48.0	72.2	47.9
	GCN	65.0	86.7	60.1	47.4	90.4	51.5	33.7	50.0	28.9	23.9	50.0	17.3
	GraphSAGE	69.4	96.7	74.1	71.8	87.0	70.8	71.0	90.7	69.7	42.5	66.8	42.5
For temporal graphs	LSTM	83.6	98.6	84.6	81.9	92.5	81.7	74.1	91.4	74.1	40.2	66.5	40.6
	GRU	81.6	98.6	82.2	82.5	93.7	83.2	75.6	91.5	75.2	42.1	67.2	41.9
	DynGEM	71.0	97.2	70.2	52.3	59.0	52.8	31.6	54.6	9.9	39.9	66.2	41.5
	DynAERNN	46.6	89.0	47.0	50.2	53.5	50.3	36.8	55.9	16.0	28.9	53.6	18.6
Variants	STAR-NH	84.7	98.4	86.1	83.1	94.4	83.5	76.6	92.2	75.9	42.3	67.1	42.1
	STAR-TA	81.3	93.5	81.7	78.2	86.6	78.3	74.5	91.7	74.7	46.1	71.3	46.2
	STAR-SA	79.5	90.2	79.9	78.3	86.5	79.6	72.1	88.5	72.6	44.6	68.0	44.4
	STAR	**89.2**	**99.2**	**90.0**	**86.2**	**97.1**	**86.7**	**80.3**	**95.5**	**80.7**	**50.8**	**75.0**	**51.1**

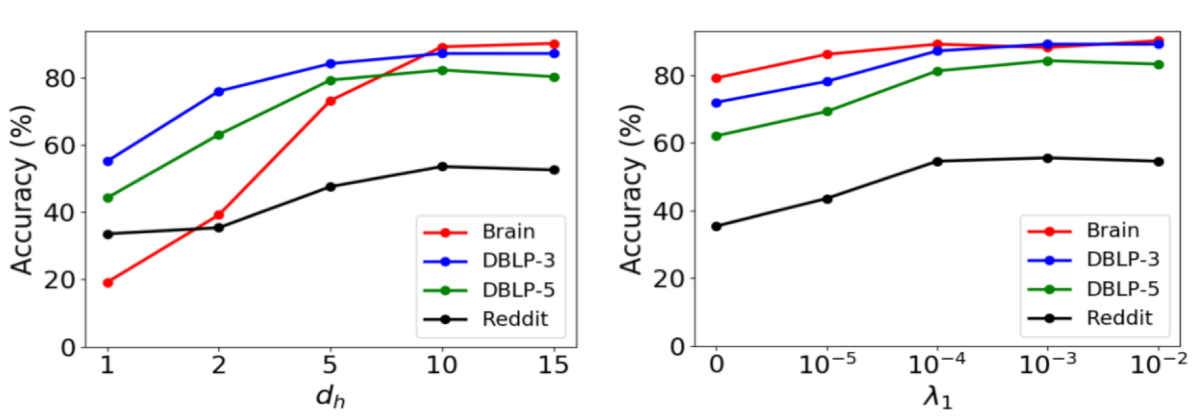
Experiment Results



Attention value of different time steps of four categories of voxels



Attention value of different neighbors of two nodes



(a) Varying d_h (b) Varying λ_1

Parameter sensitive analysis

Thanks!

Q & A