UNIVERSITY OF CAMBRIDGE

Sheaves for Heterogeneous Data

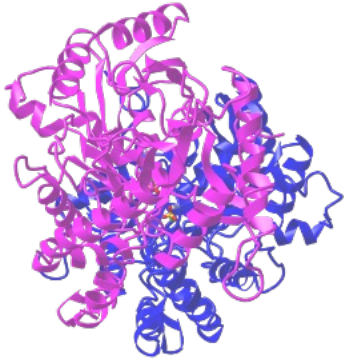
Luke Braithwaite

📖 **Read the paper:** https://arxiv.org/abs/2409.08036
🧑🏼‍💻 **Source code available at:** https://github.com/AspieCoder1/mphil-acs-repo

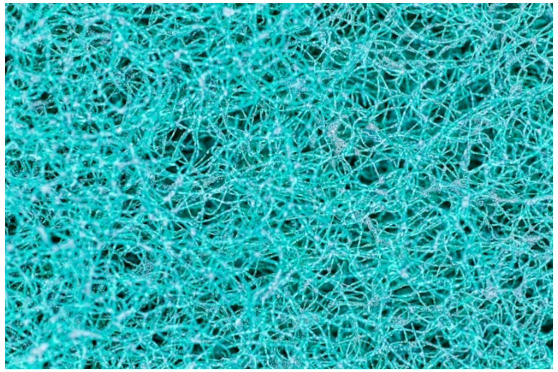
Background

- Relational data
- Heterogeneous graphs
- GNNs
- Heterogeneous GNNs

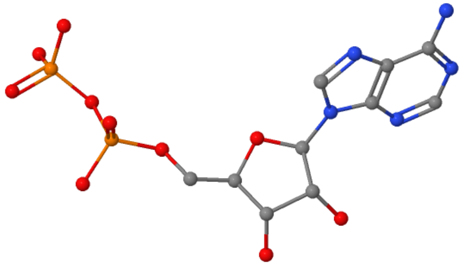
Relational data is everywhere



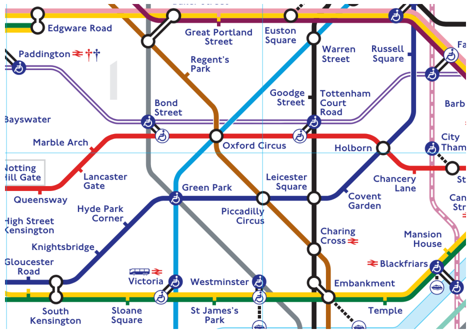
Proteins



Neuroscience

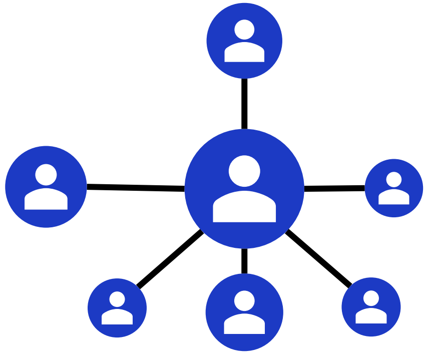


Chemistry



Transport

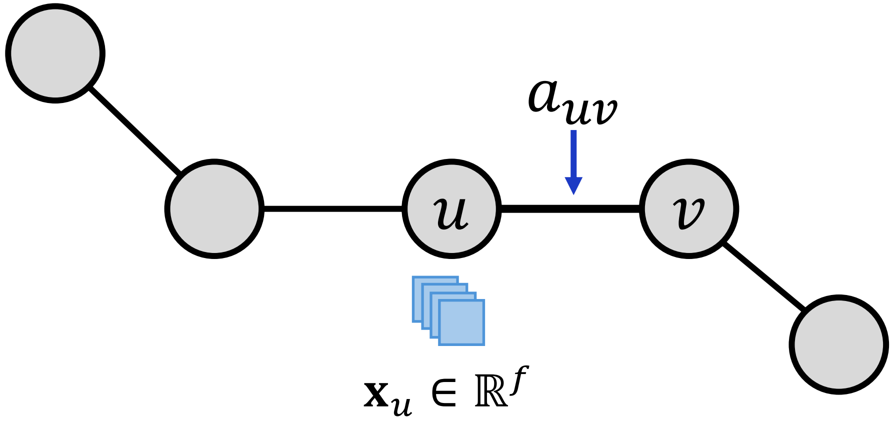
Robotics



Social networks

Graphs

A graph is a set of nodes connected by edges



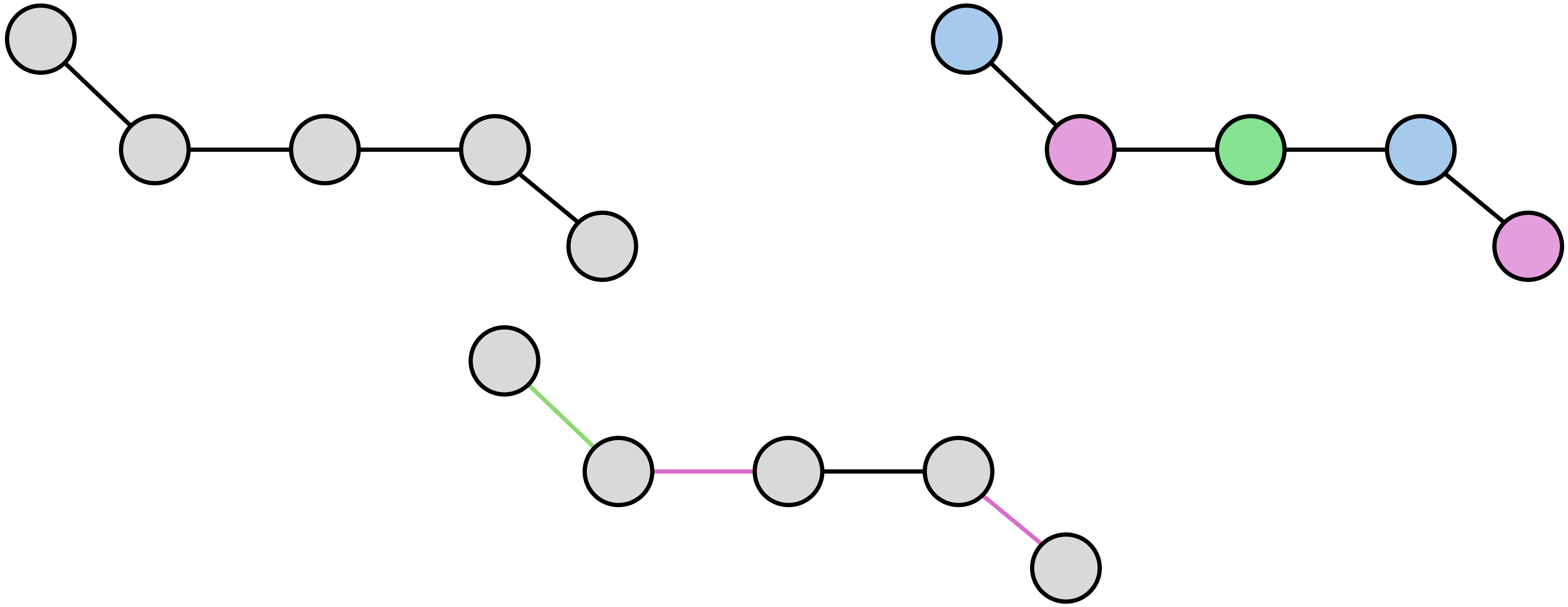
$$\mathcal{G} = (\mathbf{A}, \mathbf{X})$$

$\mathbb{R}^{n \times f}$ feature matrix

$n \times n$ adjacency matrix

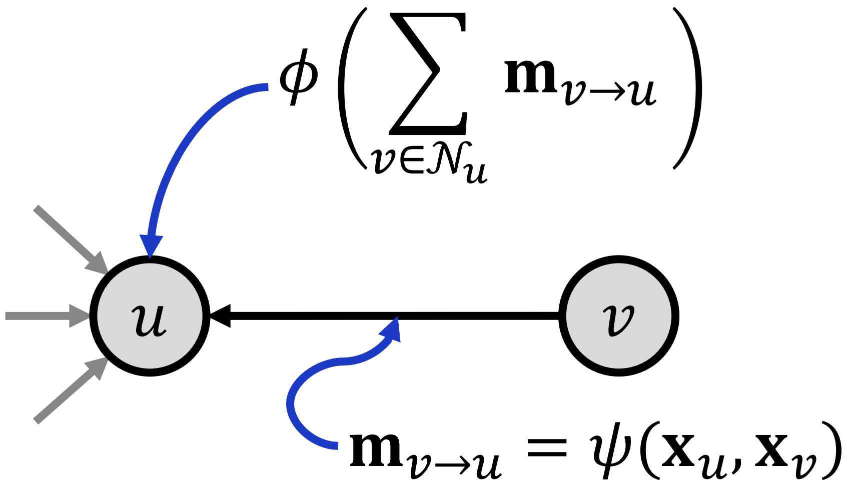
Heterogenous data

Heterogeneous data multiple node and edge types



Graph Neural Networks

Node features are updated using local aggregation

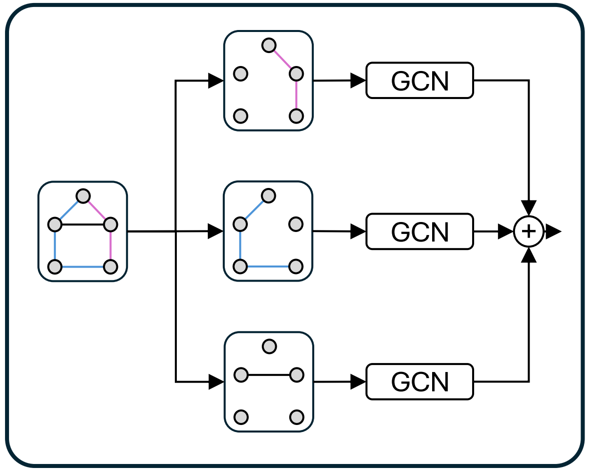


$$\mathbf{m}_u^{(l)} := \textsc{Agg}\left(\left\{\left(\boldsymbol{x}_u^{(l)}, \boldsymbol{x}_u^{(l)}\right) \middle| v \in \mathcal{V}\right\}\right)$$

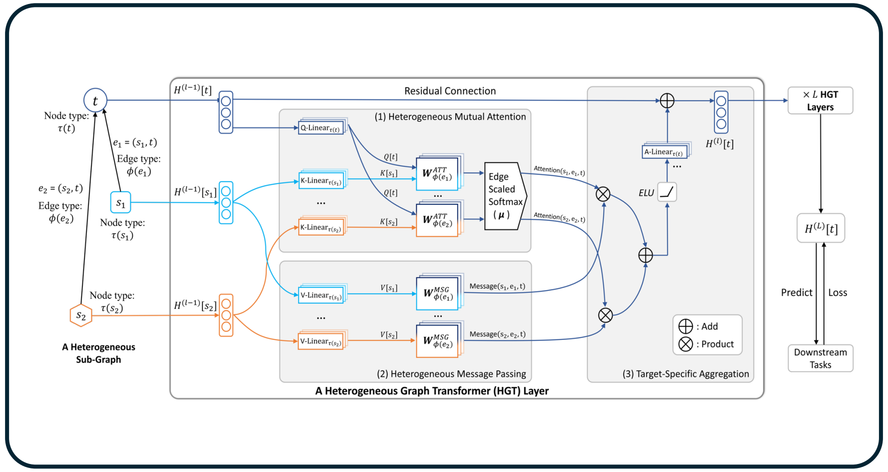
$$\mathbf{x}_u^{(l+1)} := \textsc{Upd}\left(\mathbf{x}_u^{(l)}, \mathbf{m}_u^{(l+1)}\right)$$

Heterogeneous Graph Neural Networks

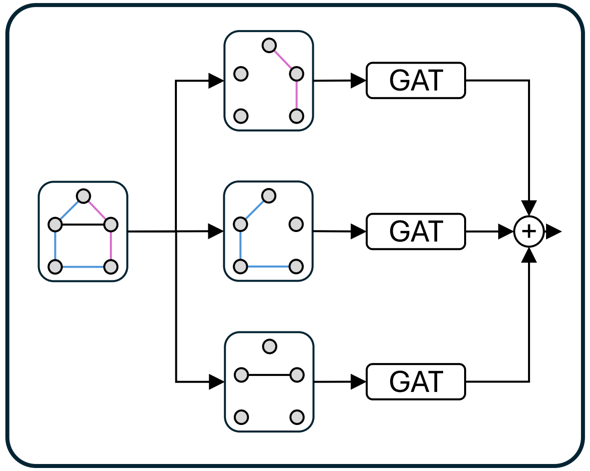
R-GCN[1]



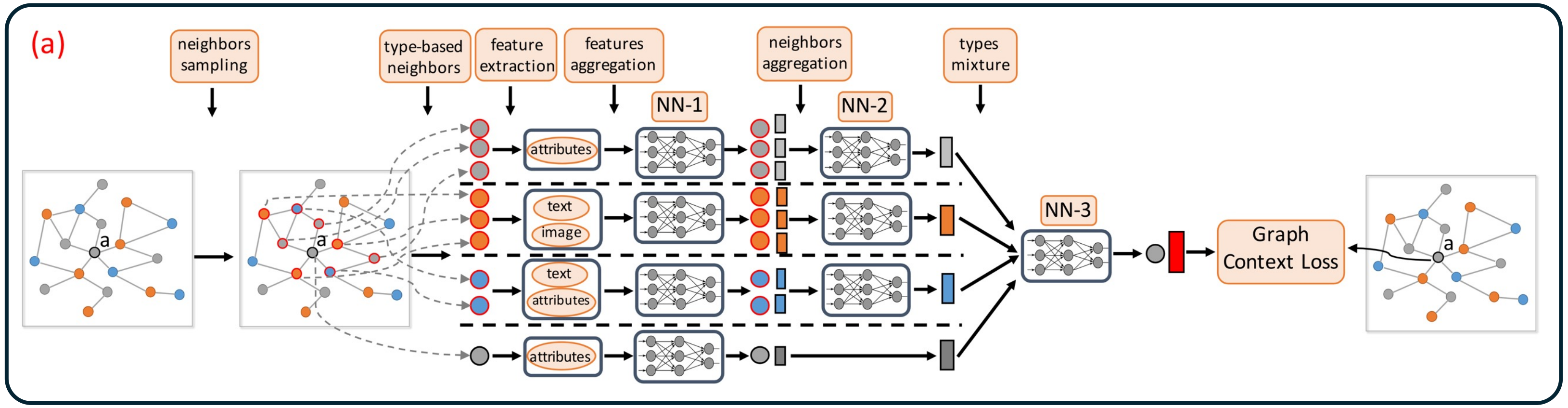
HGT[2]



HAN[3]



HetGNN[4]



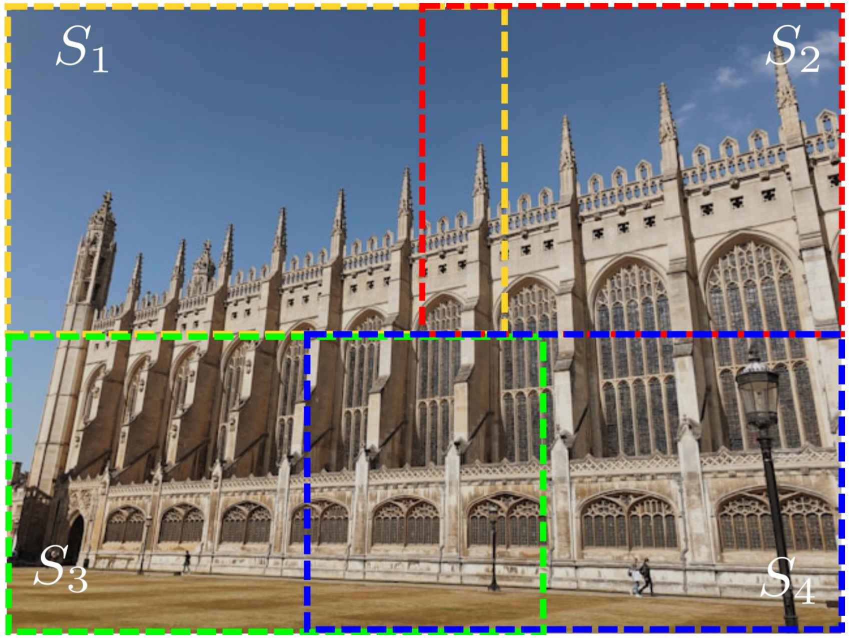
[1] Schlichtkrull et al., ‘Modeling Relational Data with Graph Convolutional Networks’, ESWC 2018.
[2] Hu et al., ‘Heterogeneous Graph Transformer’, WWW 2020.
[3] Wang et al., ‘Heterogeneous Attention Network’, WWW 2019.
[4] Zhang et al., ‘Heterogeneous Graph Neural Network’, KDD 2019.

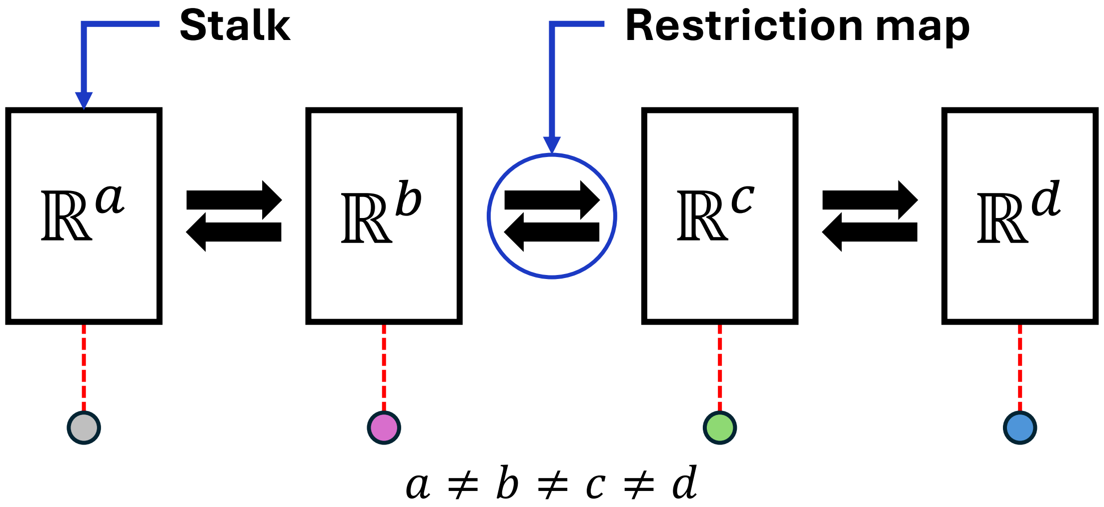
Sheaves for heterogeneous data

- Cellular sheaves
- Neural Sheaf Diffusion
- Sheaves model heterogeneity

Motivating sheaves

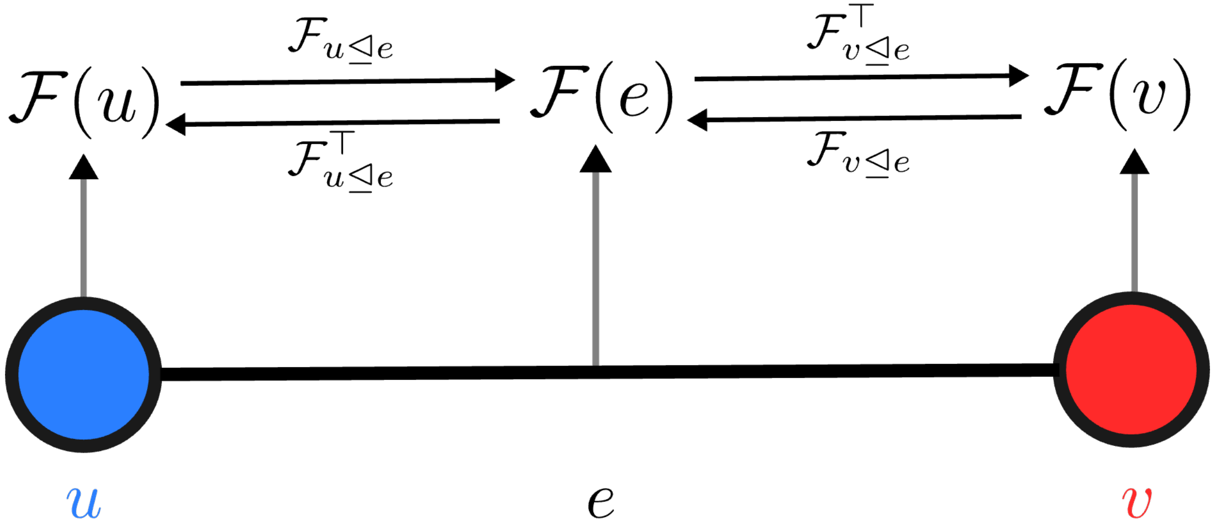
Local data assignment → consistent global representation





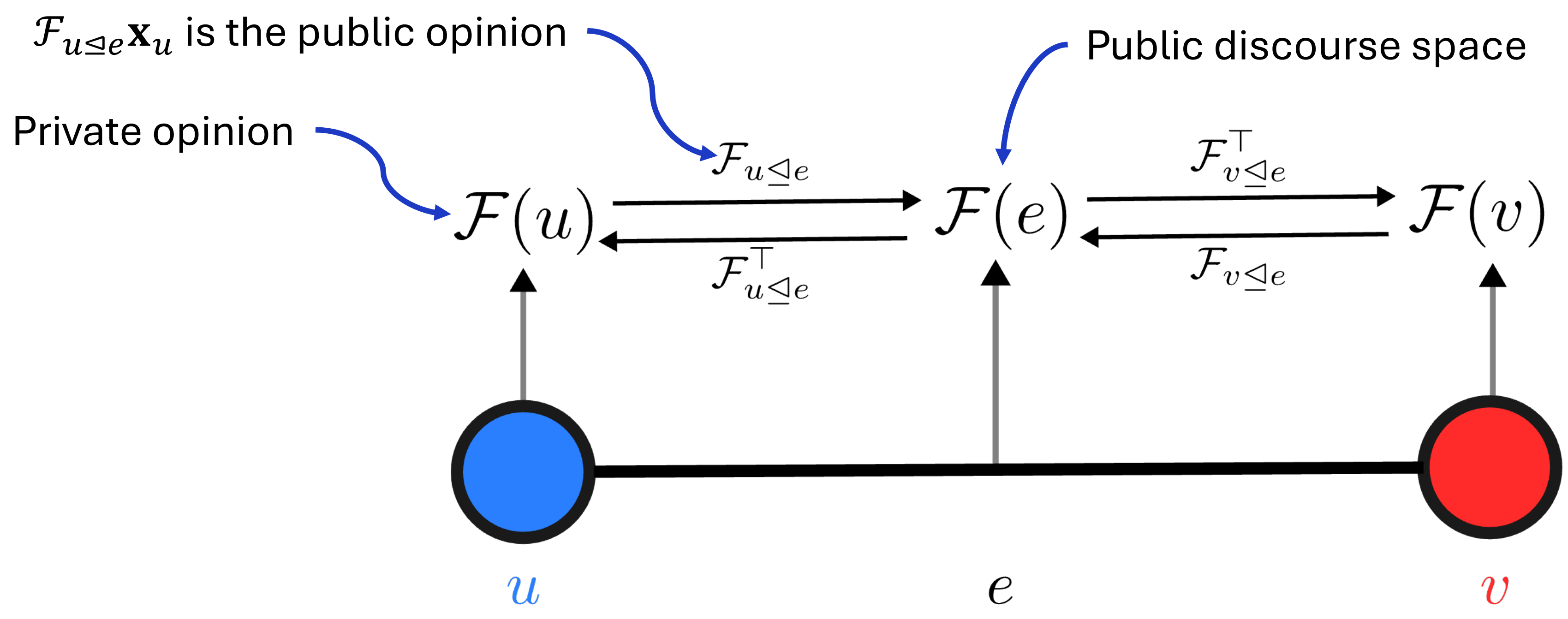
Cellular sheaves

- **Node stalks** $\mathcal{F}(u)$ attached to each node
- **Edge stalks** $\mathcal{F}(e)$ attached to each edge
- **Restriction map** $\mathcal{F}_{u \trianglelefteq e}$ for each node-edge incidence pair



So what is a sheaf?

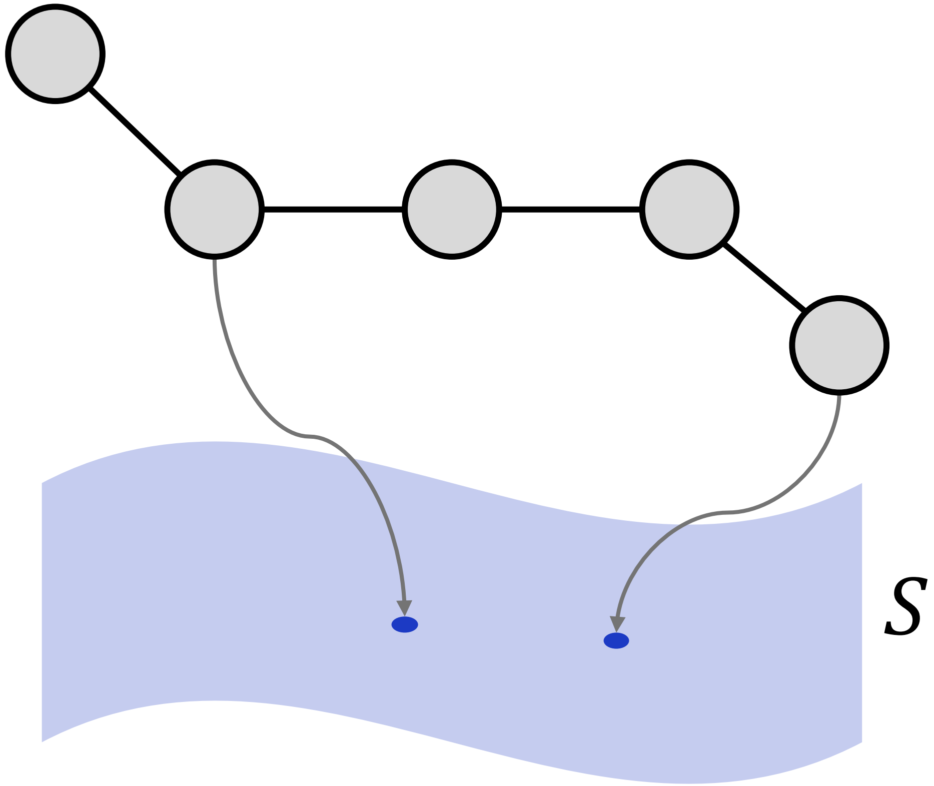
Opinion dynamics[1] provides a nice perspective

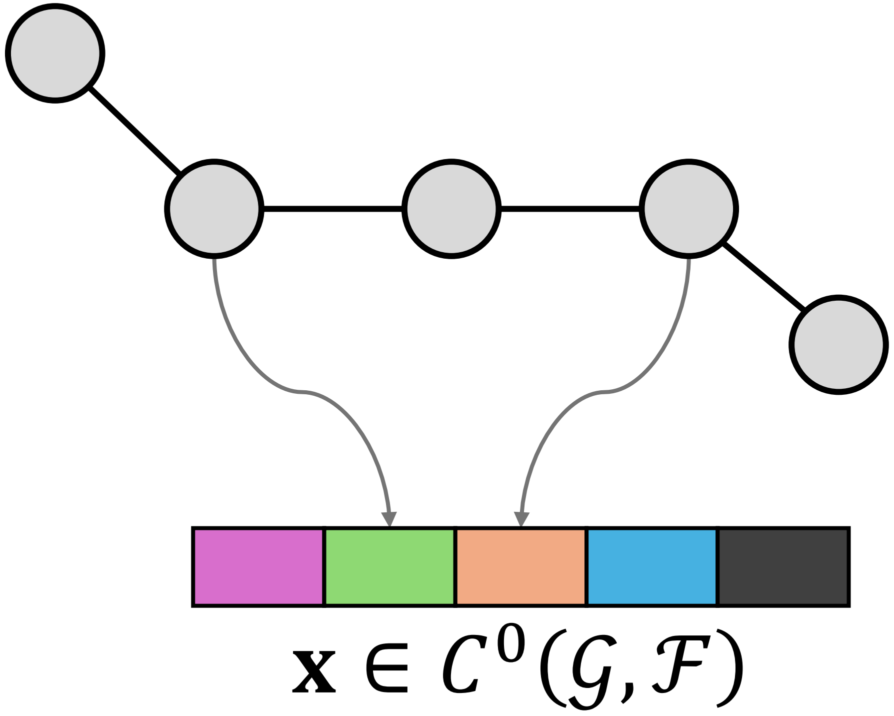


[1] Hansen and Ghirst, 'Opinion Dynamics on Sheaf Discourses', 2020, arXiv:2005.12798 [math.DS]

Why sheaves?

The underlying topology models the heterogeneity

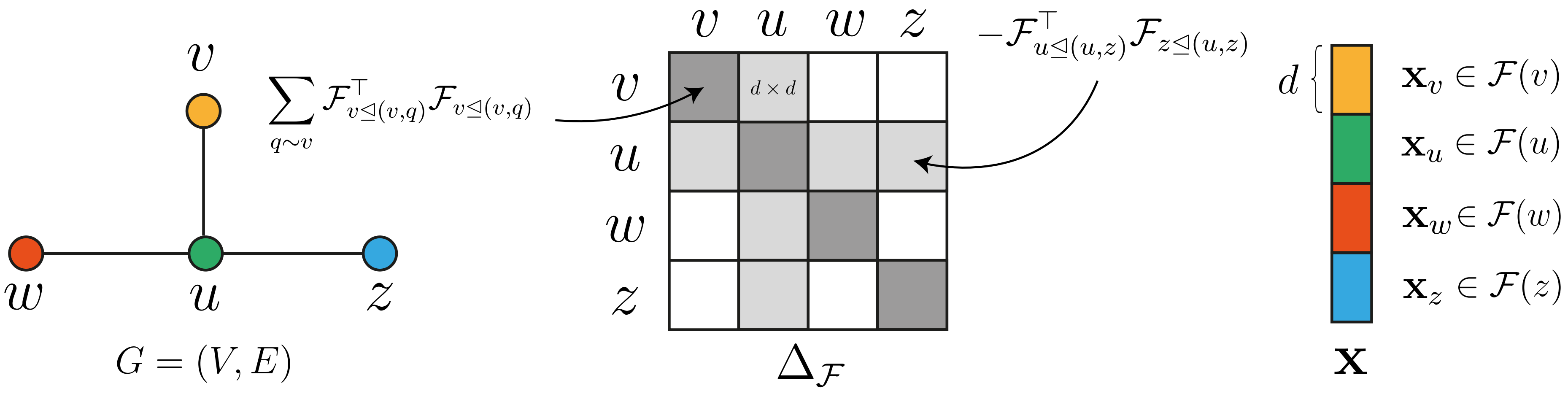




*Here $C^0(\mathcal{G},\ \mathcal{F}) = \bigoplus_{u \in \mathcal{V}} \mathcal{F}(u)$, or the block matrix formed by stacking each node stalk representation.

Neural Sheaf Diffusion[1]

Attaches a sheaf to a Graph Convolutional Network



$$\mathbf{Y} = \sigma\big((\mathbf{I}_{nd} - \Delta_{\mathcal{F}})(\mathbf{I}_n \otimes \mathbf{W}_1)\mathbf{X}\mathbf{W}_2\big)$$
$$\mathcal{F}_{u \trianglelefteq e} = \mathrm{MLP}\big(\mathbf{x}_u \| \mathbf{x}_v\big)$$

[1] Bodnar et al., 'Neural Sheaf Diffusion: A Topological Perspective on Heterophily and Oversmoothing in GNNs', NeurIPS 2022.

NSD performs well on benchmarks

NSD is smaller than R-GCN with similar performance

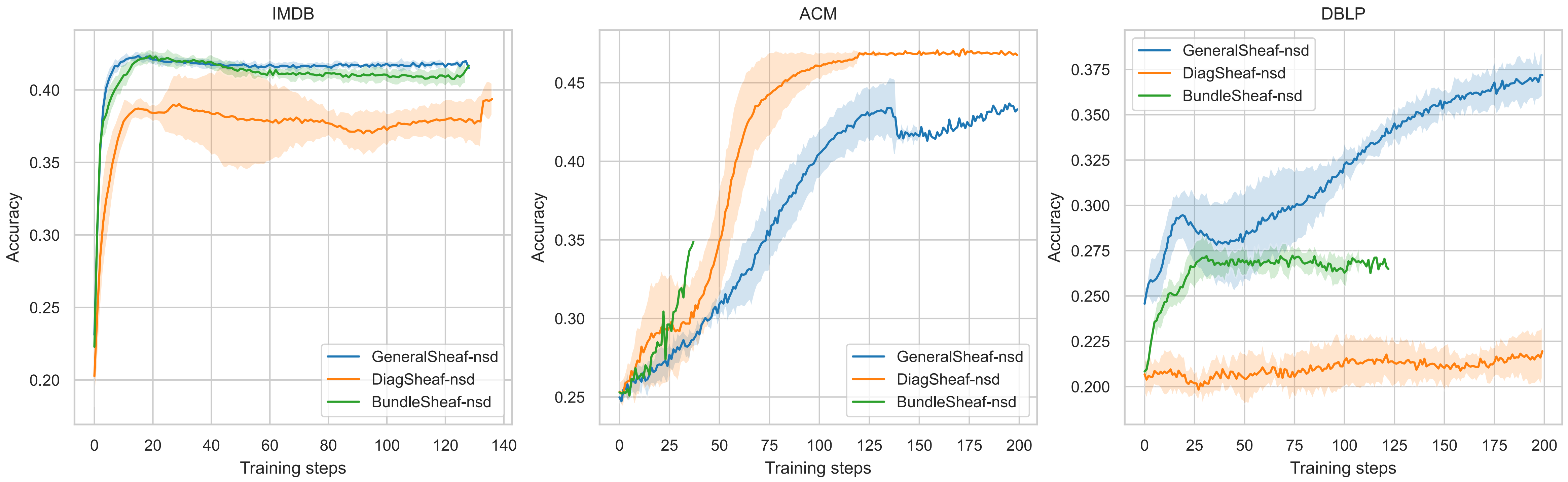
	ACM		DBLP		IMDB	
	Macro F1	Micro F1	Macro F1	Micro F1	Macro F1	Micro F1
GAT	75.80 ± 10.69	77.91 ± 8.66	95.47 ± 0.44	95.70 ± 0.42	84.12 ± 0.96	85.31 ± 0.92
GCN	89.09 ± 3.66	89.14 ± 3.60	**96.31 ± 0.73**	**96.57 ± 0.63**	82.41 ± 1.15	83.99 ± 0.92
HAN	86.95 ± 6.19	86.64 ± 6.43	94.74 ± 0.81	95.01 ± 0.73	13.53 ± 0.24	38.70 ± 1.13
R-GCN	**95.81 ± 0.39**	**95.75 ± 0.39**	**96.79 ± 0.39**	**97.01 ± 0.34**	**88.16 ± 0.67**	**89.08 ± 0.63**
HGT	**93.24 ± 3.19**	**93.30 ± 2.91**	93.91 ± 1.08	94.26 ± 1.09	**87.74 ± 0.76**	**88.45 ± 0.71**
Sheaf-NSD	**94.97 ± 0.41**	**94.94 ± 0.42**	**96.69 ± 0.82**	**96.89 ± 0.79**	**86.70 ± 0.90**	**87.50 ± 0.78**

Sheaf-NSD 111x smaller than R-GCN

	LastFM		MovieLens	
	AUPR	AUROC	AUPR	AUROC
GAT	62.88 ± 0.18	50.69 ± 0.63	97.06 ± 0.24	97.47 ± 0.21
GCN	**96.84 ± 0.10**	**96.42 ± 0.08**	**99.57 ± 0.03**	**99.51 ± 0.03**
HAN	82.48 ± 3.86	78.47 ± 3.04	63.49 ± 0.14	52.06 ± 0.27
R-GCN	**96.86 ± 0.07**	**96.97 ± 0.05**	**99.06 ± 0.05**	**99.13 ± 0.04**
HGT	–	–	–	–
Sheaf-NSD	**97.16 ± 0.19**	**96.58 ± 0.18**	**99.66 ± 0.04**	**99.57 ± 0.03**

Sheaf-NSD 209x smaller than R-GCN

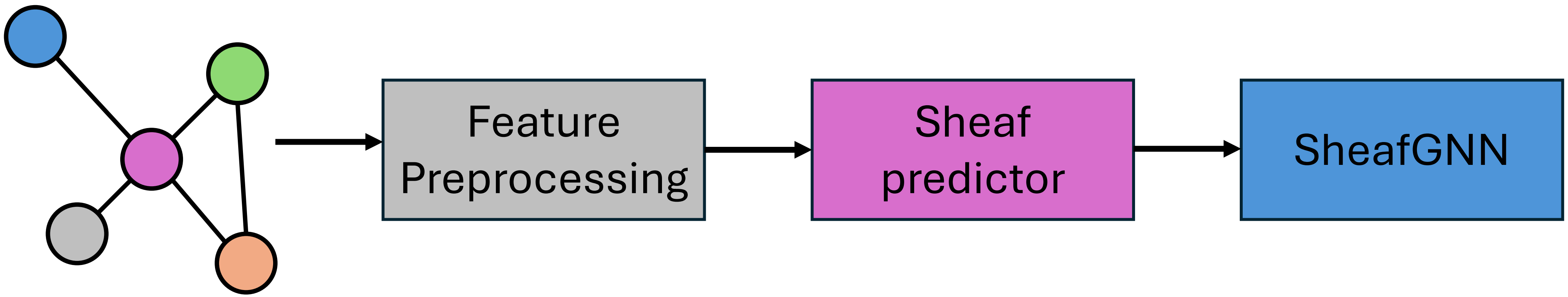
Sheaves implicitly learn types



HetSheaf

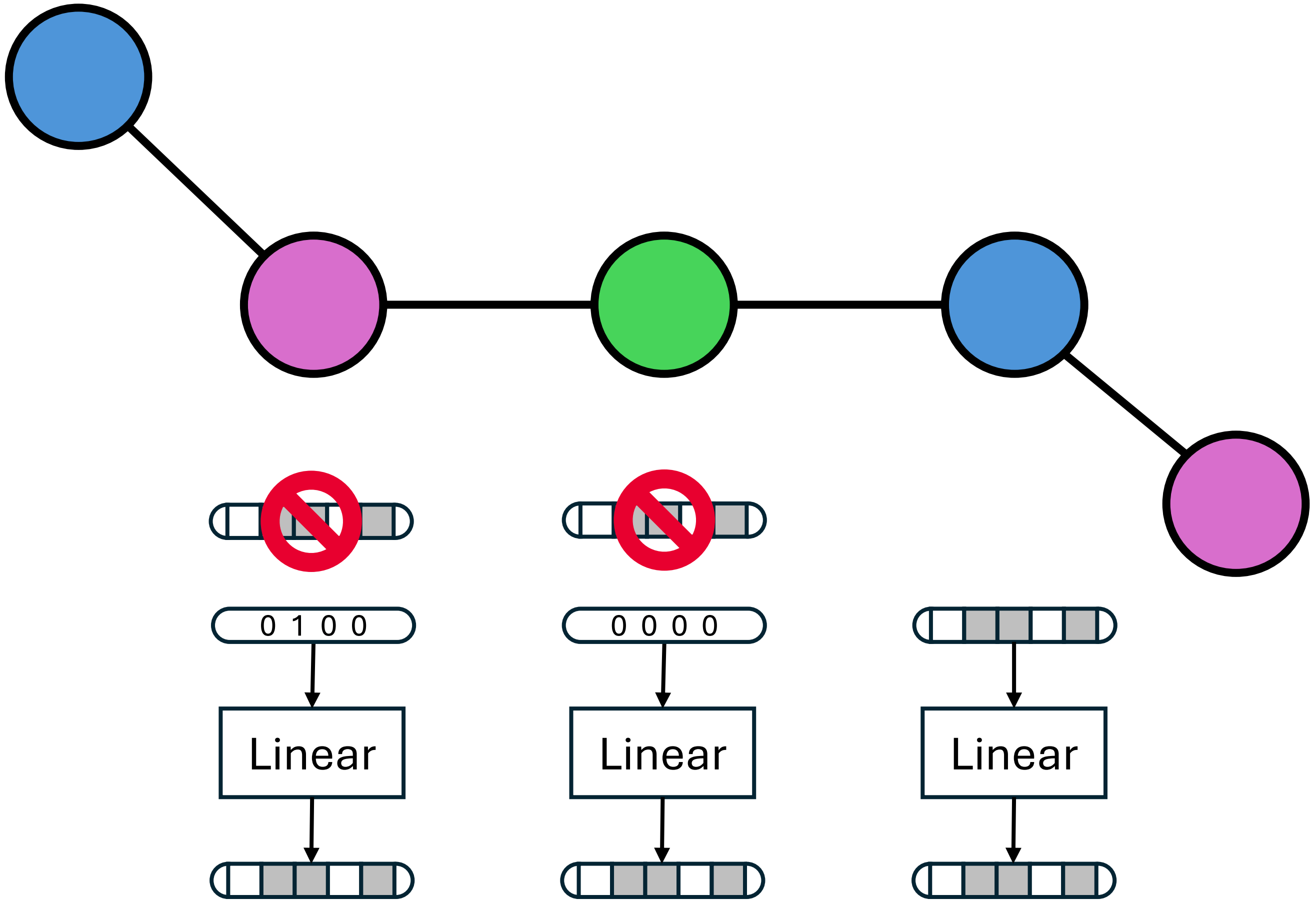
A general framework for heterogeneous sheaf neural networks

HetSheaf pipeline



Feature preprocessing

Linear layers used to project features to same dimensionality



Heterogeneous sheaf predictors

node features in $\mathcal{F}(u)$

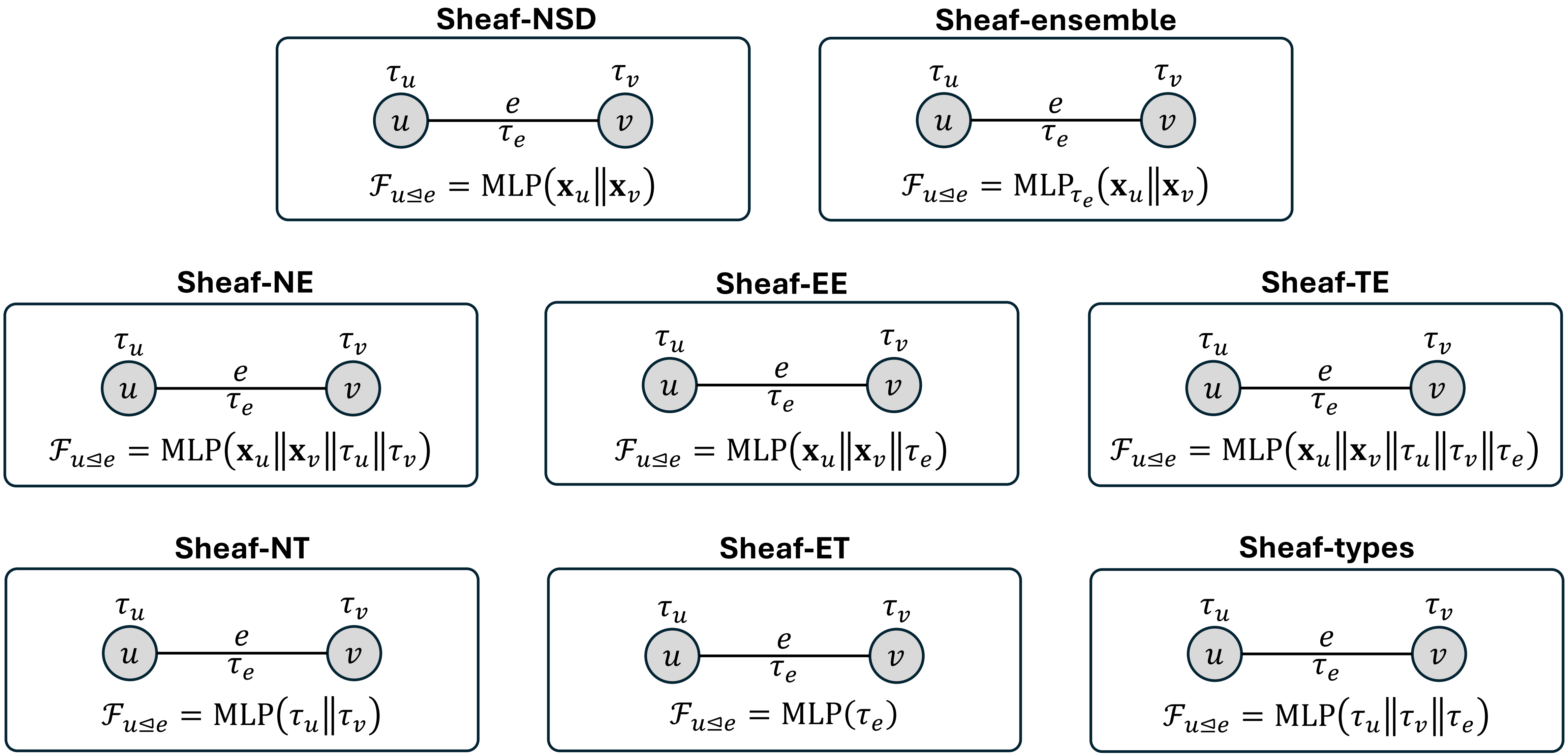
type of node u

type of edge e

$$\mathcal{F}_{u \trianglelefteq (u,v)} = \mathbf{\Phi}\left(\mathbf{x}_u, \mathbf{x}_v, \phi(u), \phi(v), \psi(e)\right)$$

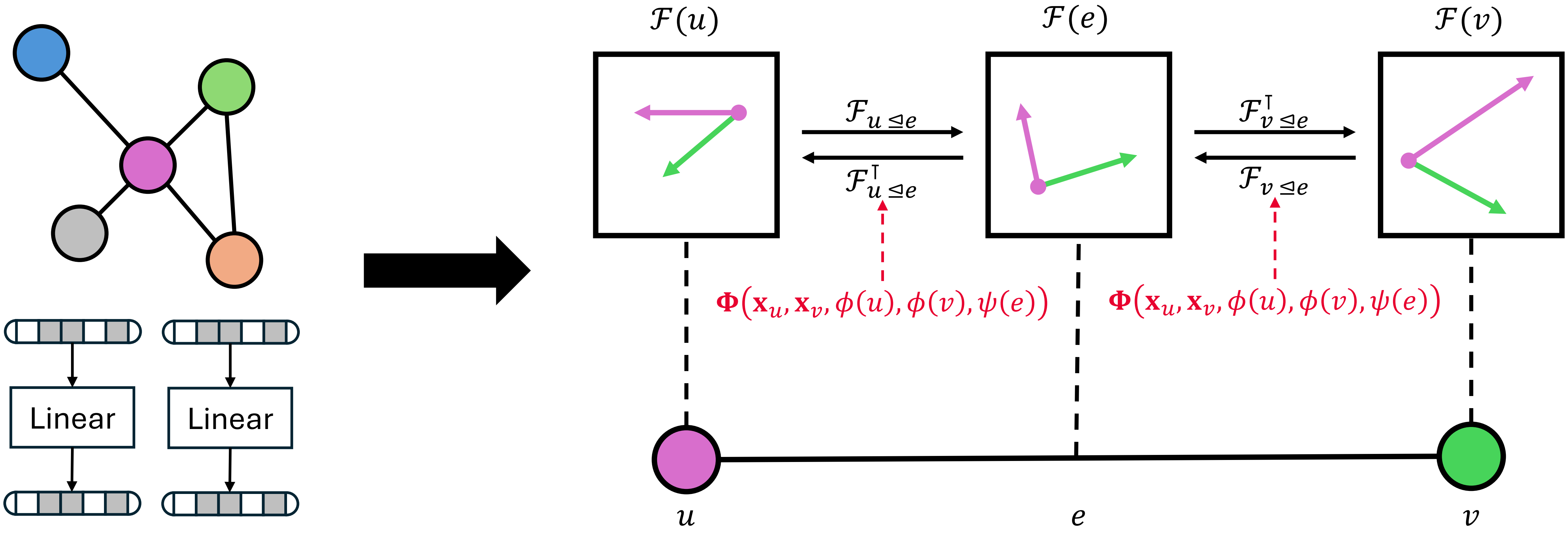
node features in $\mathcal{F}(v)$

type of node v



* Each type is assumed to be a one-hot encoded vector, $\tau_e \coloneqq \mathbf{e}_{\psi(e)}$ for $e \in \mathcal{E}$ and $\tau_u \coloneqq \mathbf{e}_{\phi(u)}$ for $u \in \mathcal{V}$.

Final architecture



Type information improves performance

The sheaf learners achieve SOTA or competitive results

Table 5.1: **Performance on heterogeneous node classification.** Results for the SheafGNN architectures and baselines from the literature are shown. The average macro and micro F1 score and standard deviation after 10 runs. The top three models are coloured by **First**, **Second** and **Third**.

	ACM		DBLP		IMDB	
	Macro F1	Micro F1	Macro F1	Micro F1	Macro F1	Micro F1
GAT [69]	75.8 ± 107.0	77.91 ± 8.66	95.47 ± 0.44	95.70 ± 0.42	84.12 ± 0.96	85.31 ± 0.92
GCN [47]	89.09 ± 3.66	89.14 ± 3.60	96.31 ± 0.73	96.57 ± 0.63	82.41 ± 1.15	83.99 ± 0.92
HAN [74]	86.95 ± 6.19	86.64 ± 6.43	94.74 ± 0.81	95.01 ± 0.73	13.53 ± 0.24	38.70 ± 1.13
RGCN [62]	95.81 ± 0.39	95.75 ± 0.39	96.79 ± 0.39	97.01 ± 0.34	**88.16 ± 0.67**	**89.08 ± 0.63**
HGT [41]	93.24 ± 3.19	93.30 ± 2.91	93.91 ± 1.08	94.26 ± 1.09	**87.74 ± 0.76**	**88.45 ± 0.71**
O(d)-nsd [7]	94.64 ± 1.02	94.59 ± 1.03	96.32 ± 0.46	96.55 ± 0.42	86.35 ± 1.29	87.20 ± 1.07
Diag-nsd [7]	94.42 ± 0.51	94.42 ± 0.48	95.25 ± 0.70	95.52 ± 0.67	86.36 ± 0.94	87.26 ± 0.78
Gen-nsd [7]	94.97 ± 0.41	94.94 ± 0.42	96.69 ± 0.82	96.89 ± 0.79	86.70 ± 0.90	87.50 ± 0.78
Sheaf-TE (ours)	96.11 ± 0.49	96.09 ± 0.51	**97.93 ± 0.36**	**98.08 ± 0.31**	86.85 ± 0.81	87.67 ± 0.80
Sheaf-ensemble (ours)	**96.16 ± 0.52**	**96.12 ± 0.54**	97.46 ± 0.64	97.62 ± 0.60	86.92 ± 1.10	87.79 ± 0.95
Sheaf-NE (ours)	**96.13 ± 0.39**	96.09 ± 0.38	97.68 ± 0.55	97.83 ± 0.51	86.87 ± 1.01	87.73 ± 0.81
Sheaf-EE (ours)	**96.39 ± 0.37**	**96.35 ± 0.36**	97.57 ± 0.69	97.73 ± 0.62	**87.12 ± 0.75**	**87.88 ± 0.67**
Sheaf-NT (ours)	96.12 ± 0.36	**96.12 ± 0.32**	**97.88 ± 0.47**	**98.04 ± 0.43**	86.92 ± 0.95	87.76 ± 0.85
Sheaf-ET (ours)	95.84 ± 0.65	95.82 ± 0.65	**97.69 ± 0.47**	**97.83 ± 0.47**	86.12 ± 0.82	87.05 ± 0.69

Table 5.3: **Performance on heterogeneous link prediction benchmarks.** Results for the three base SheafGNN architectures and baselines from the literature are shown. The table shows the average and standard deviation of the binary AUROC and AUPR scores after 10 runs with the top three models, coloured **First**, **Second** and **Third**. The runs labelled '–' were caused by an out-of-memory error of the GPU.

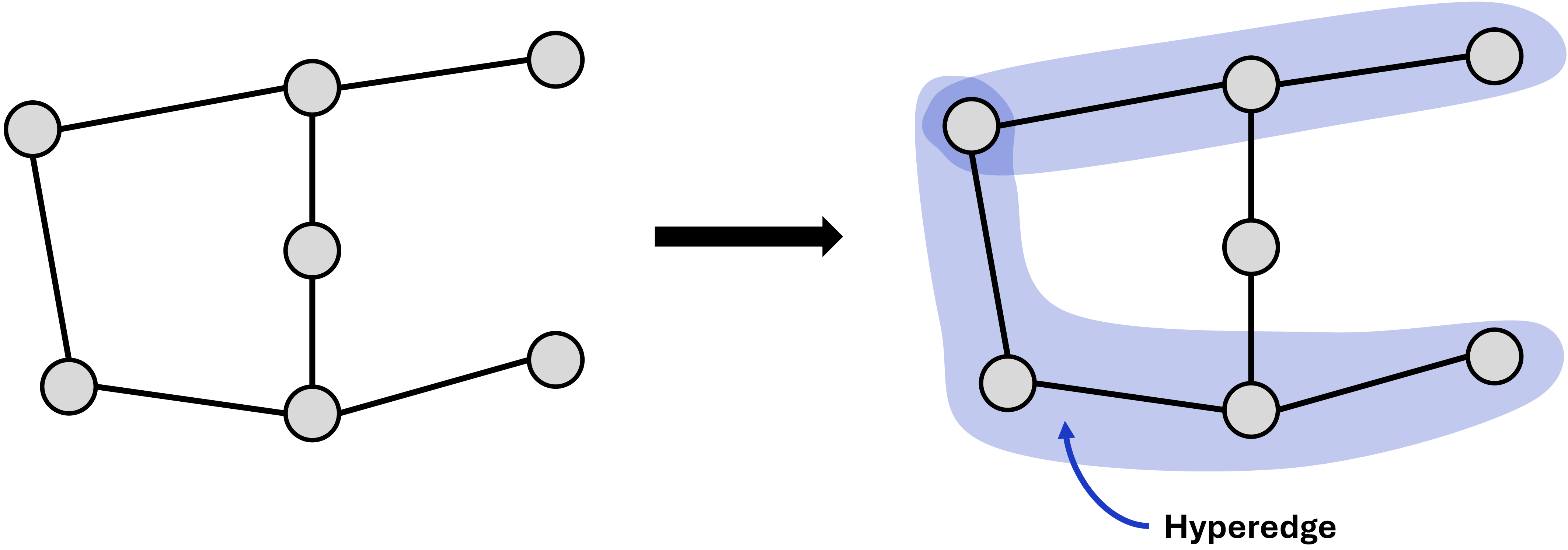
	LastFM		MovieLens	
	AUPR	AUROC	AUPR	AUROC
GAT	62.88 ± 0.18	50.69 ± 0.63	97.06 ± 0.24	97.47 ± 0.21
GCN	96.84 ± 0.10	96.42 ± 0.08	99.57 ± 0.03	99.51 ± 0.03
HAN	82.48 ± 3.86	78.47 ± 3.04	63.49 ± 0.14	52.06 ± 0.27
R-GCN	96.86 ± 0.07	96.97 ± 0.05	99.06 ± 0.05	99.13 ± 0.04
HGT	–	–	–	–
Sheaf-nsd	97.16 ± 0.19	96.58 ± 0.18	**99.66 ± 0.04**	**99.57 ± 0.03**
Sheaf-TE (ours)	97.71 ± 0.52	97.23 ± 0.63	99.65 ± 0.03	99.57 ± 0.04
Sheaf-ensemble (ours)	**98.21 ± 0.15**	**97.71 ± 0.18**	**99.68 ± 0.04**	**99.59 ± 0.04**
Sheaf-NE (ours)	**97.90 ± 0.68**	**97.51 ± 0.51**	99.66 ± 0.04	99.57 ± 0.04
Sheaf-EE (ours)	97.51 ± 0.44	96.91 ± 0.52	**99.67 ± 0.05**	**99.57 ± 0.05**
Sheaf-NT (ours)	**98.24 ± 0.13**	**97.80 ± 0.18**	99.61 ± 0.03	99.52 ± 0.03
Sheaf-ET (ours)	97.84 ± 0.32	97.260 ± 0.003	99.64 ± 0.03	99.54 ± 0.03

Future work

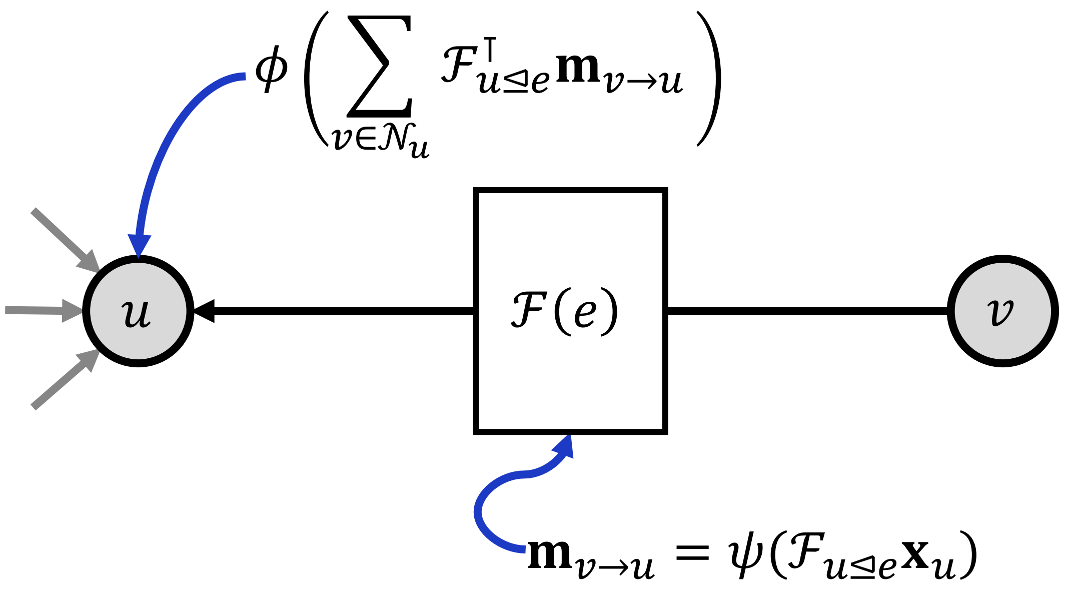
- Lifting to hypergraphs
- Generalised sheaf message passing
- Topological sheaves

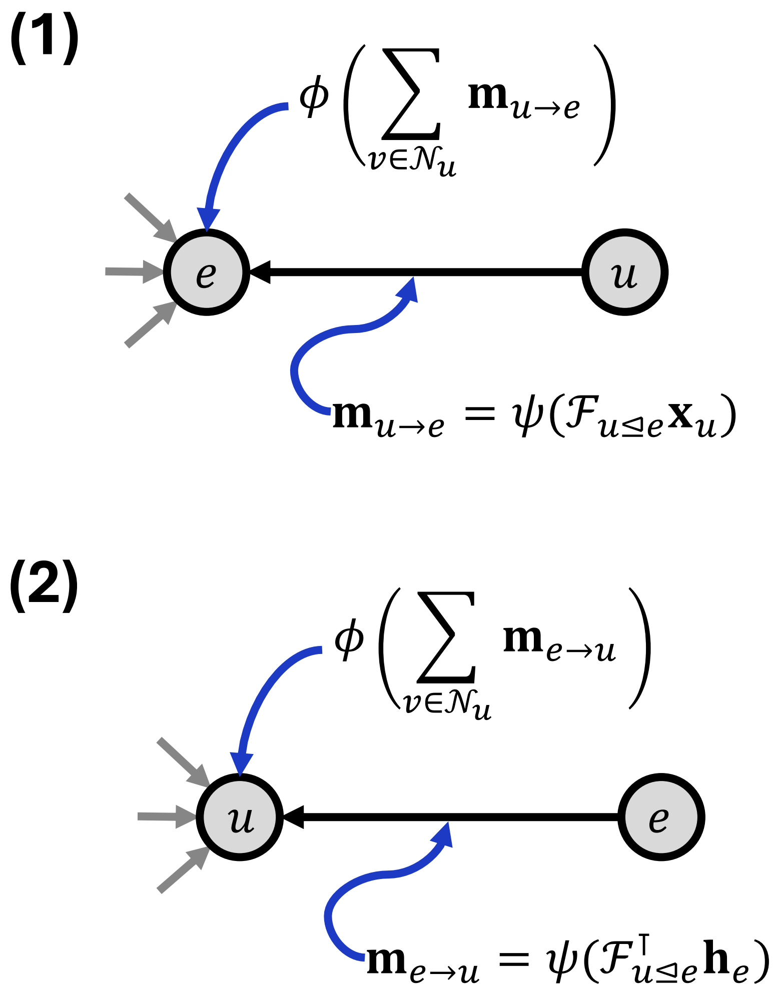
Accounting for higher order interactions

Hypergraphs connect an arbitrary set of nodes

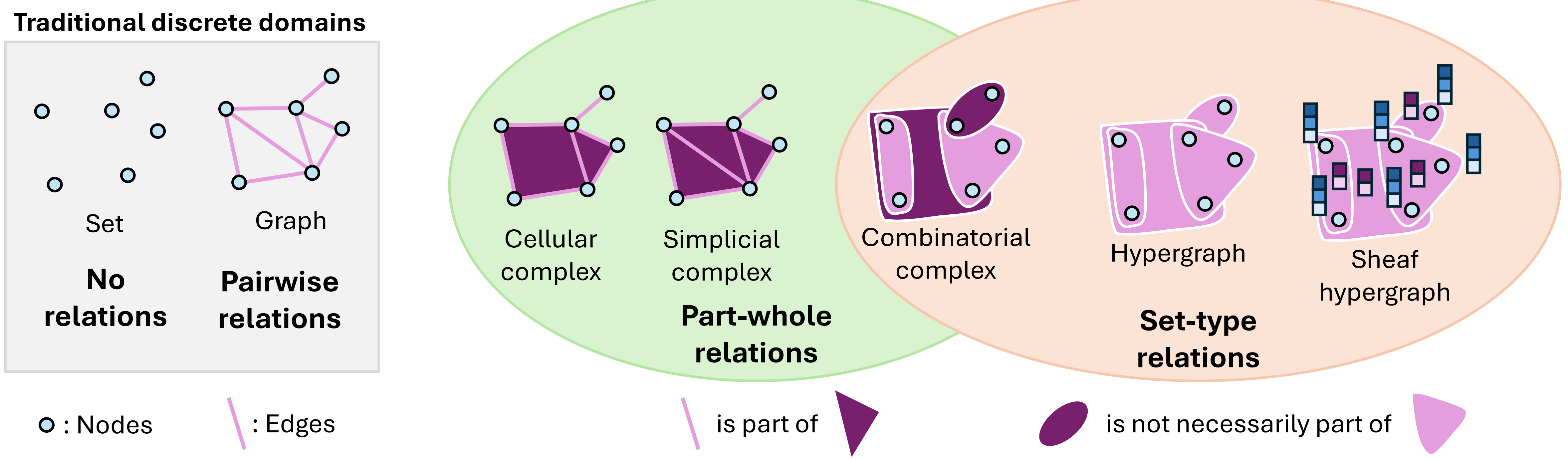


Generalised sheaf message passing





Sheaf Topological Neural Networks



*Adapted from Papillon et al., 'Architectures of Topological Deep Learning: A Survey on Topological Neural Networks', 2023, arXiv:2304.10031 [cs.LG]

Summary

- Sheaves provide a natural way to model heterogeneity
- Sheaf predictors may be parameterised to include type information
- Type information improves model performance
- These results are competitive or SOTA across all benchmarks
- We can define more general sheaf message passing approaches