

Sheaf Heterogeneous Neural Networks

Luke Braithwaite



Read the paper: <https://arxiv.org/abs/2409.08036>



Source code available at: <https://github.com/AspieCoder1/mphil-acs-repo>

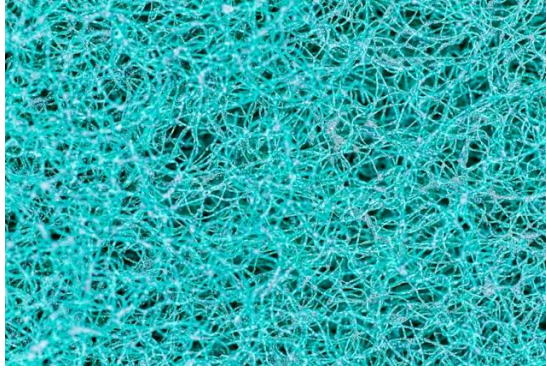
Background

- Relational data
- Heterogeneous graphs
- GNNs
- Heterogeneous GNNs

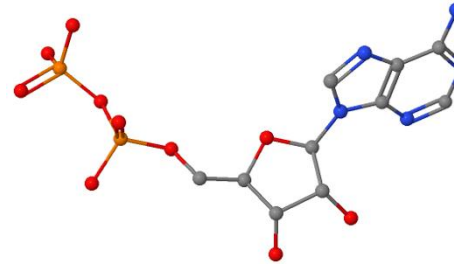
Relational data is everywhere



Proteins



Neuroscience



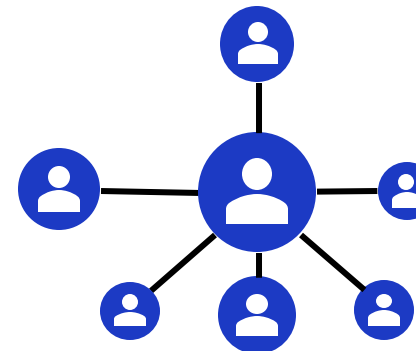
Chemistry



Transport



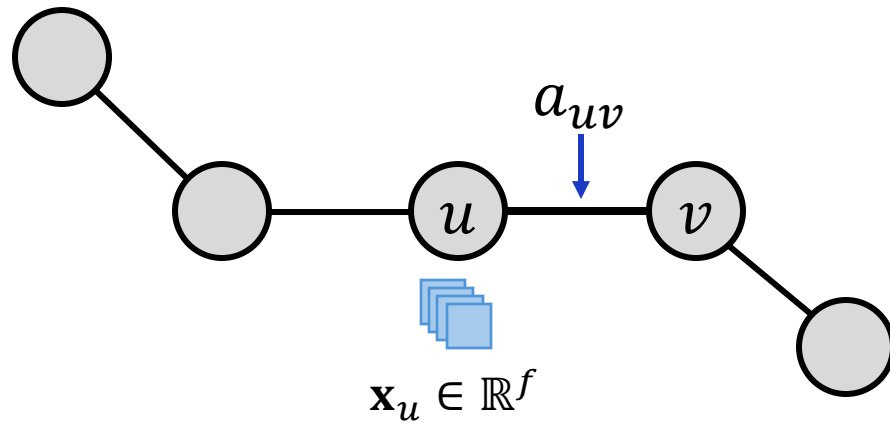
Robotics



Social networks

Graphs

A graph is a set of nodes connected by edges

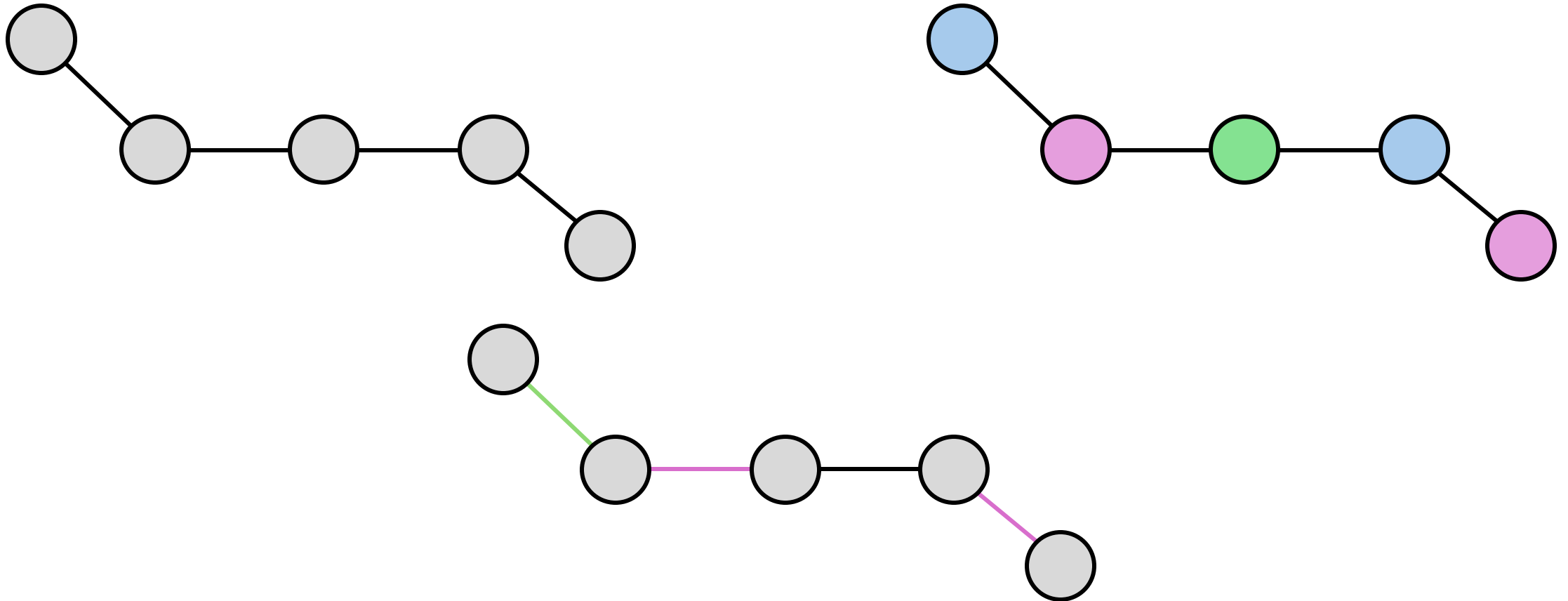


$$\mathcal{G} = (\mathbf{A}, \mathbf{X})$$

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

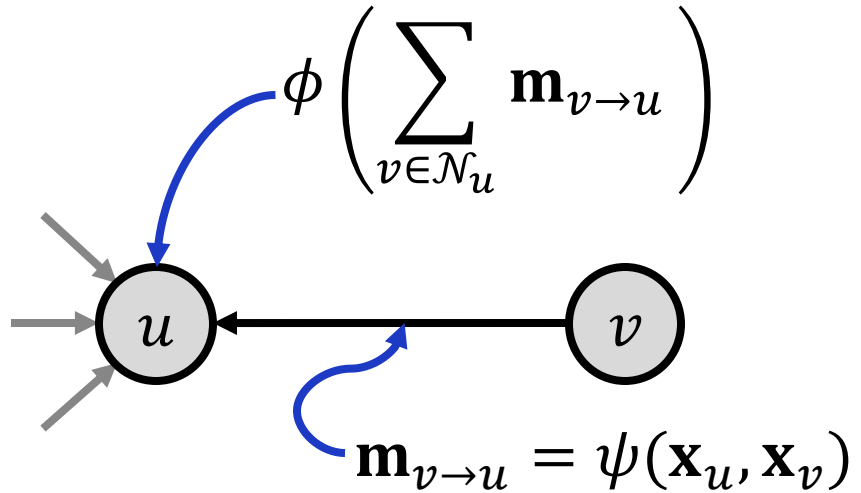
Heterogenous data

Heterogeneous data multiple node and edge types



Graph Neural Networks

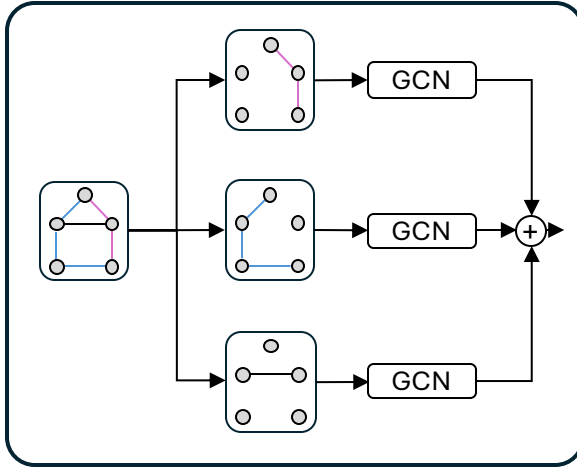
Node features are updated using local aggregation



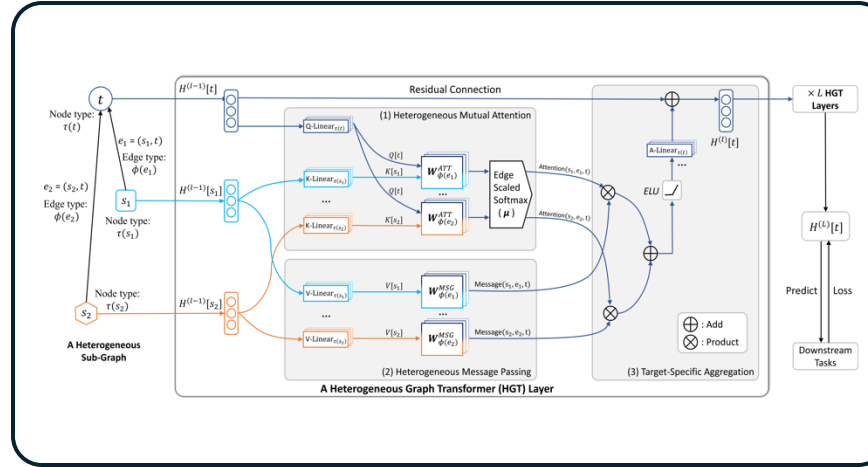
$$\mathbf{m}_u^{(l)} := \text{AGG} \left(\left\{ \left(\mathbf{x}_u^{(l)}, \mathbf{x}_u^{(l)} \right) \mid v \in \mathcal{V} \right\} \right)$$
$$\mathbf{x}_u^{(l+1)} := \text{UPD} \left(\mathbf{x}_u^{(l)}, \mathbf{m}_u^{(l+1)} \right)$$

Heterogeneous Graph Neural Networks

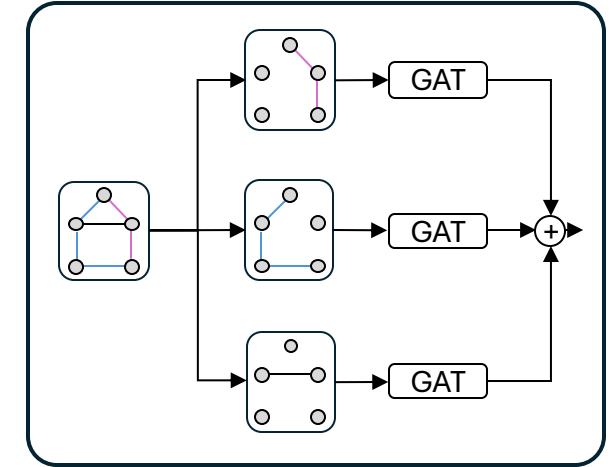
R-GCN^[1]



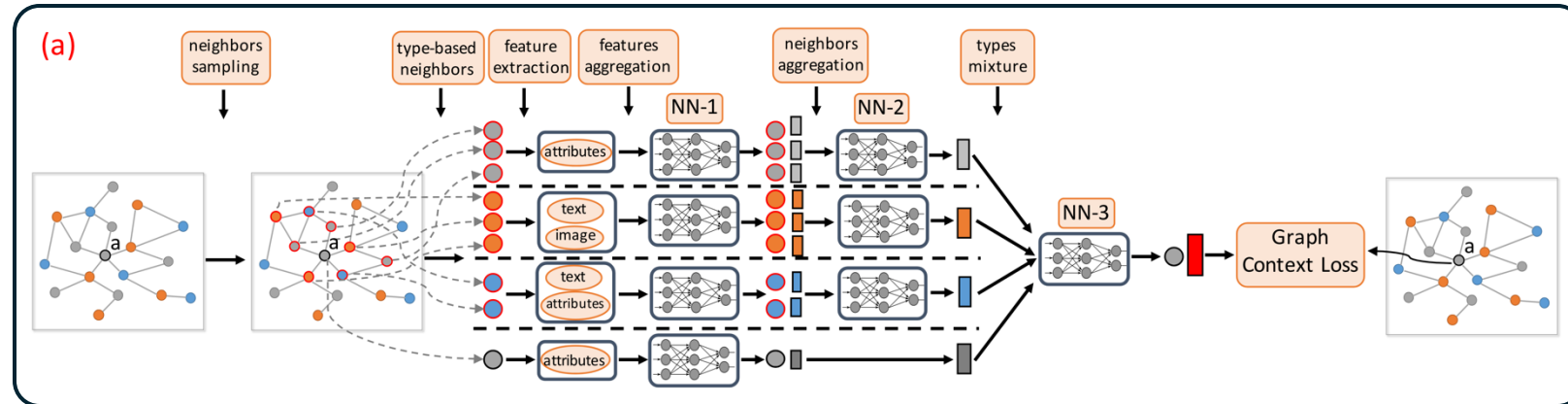
HGT^[2]



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.

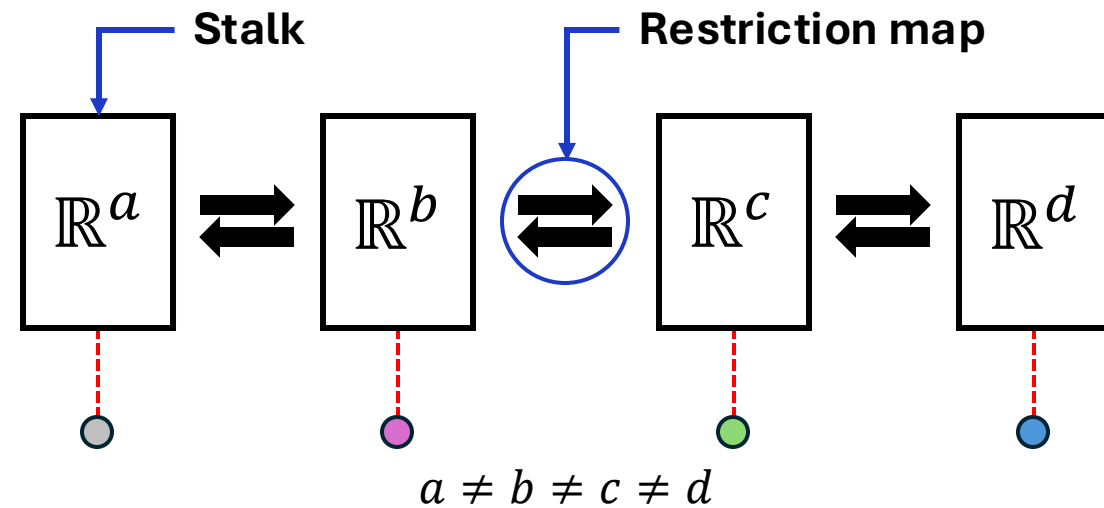
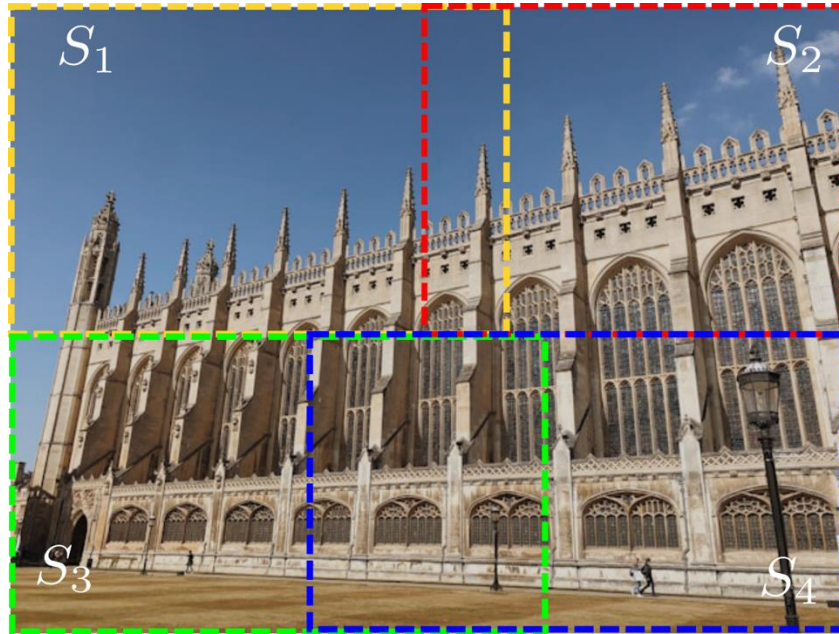
[4] Zhang et al., 'Heterogeneous Graph Neural Network', KDD 2019.

Sheaves for heterogeneous data

- Cellular sheaves
- Neural Sheaf Diffusion
- Sheaves model heterogeneity

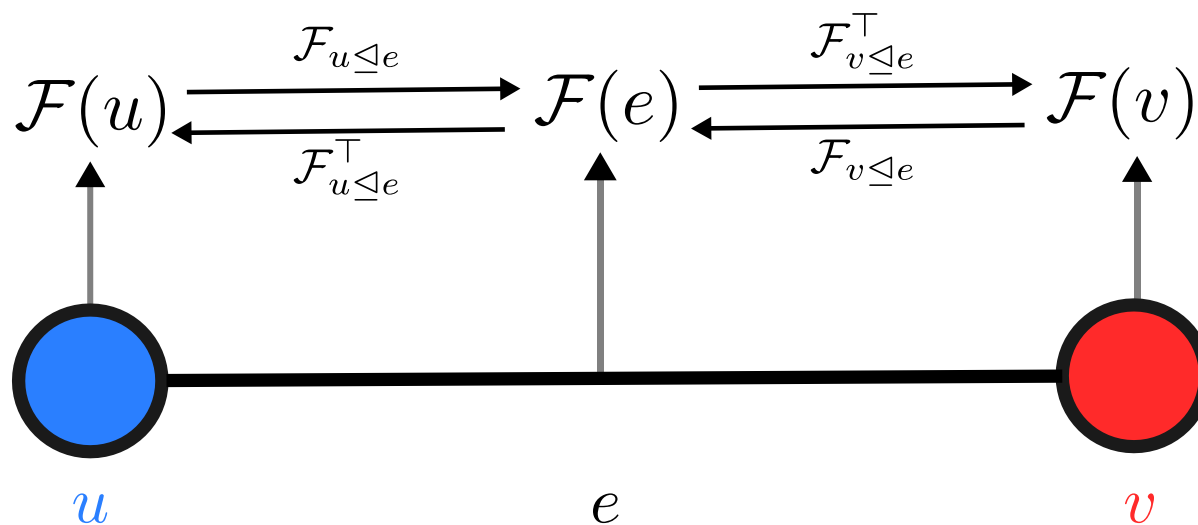
Motivating sheaves

Local data assignment \rightarrow consistent global representation



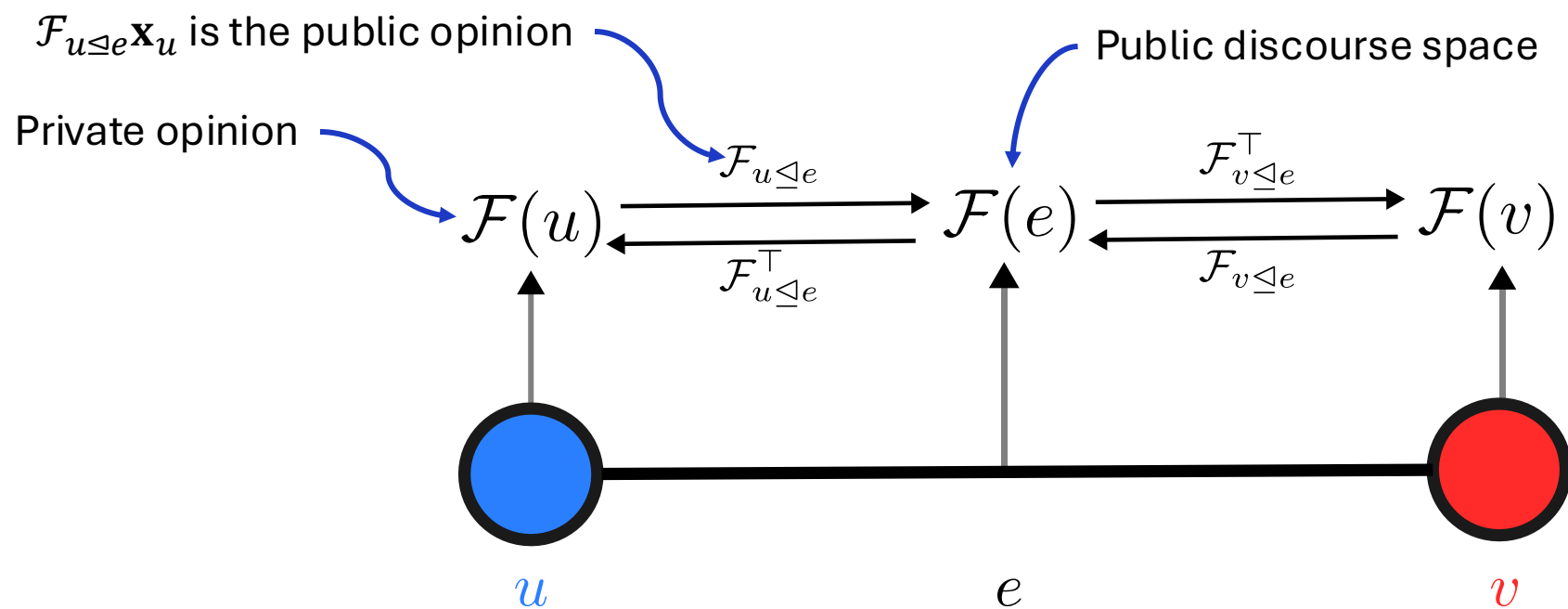
Cellular sheaves

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



So what is a sheaf?

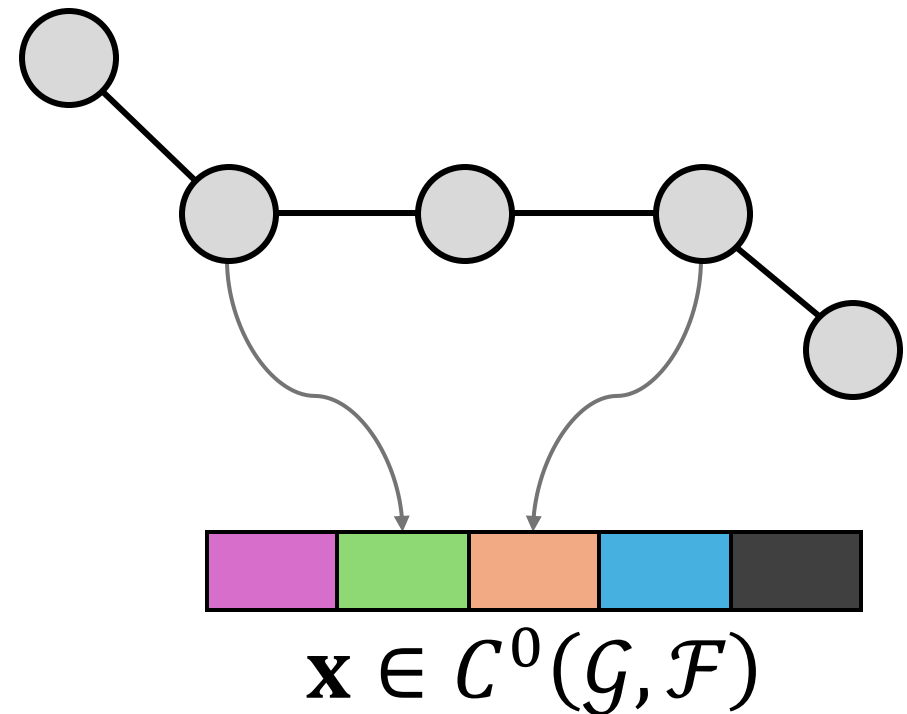
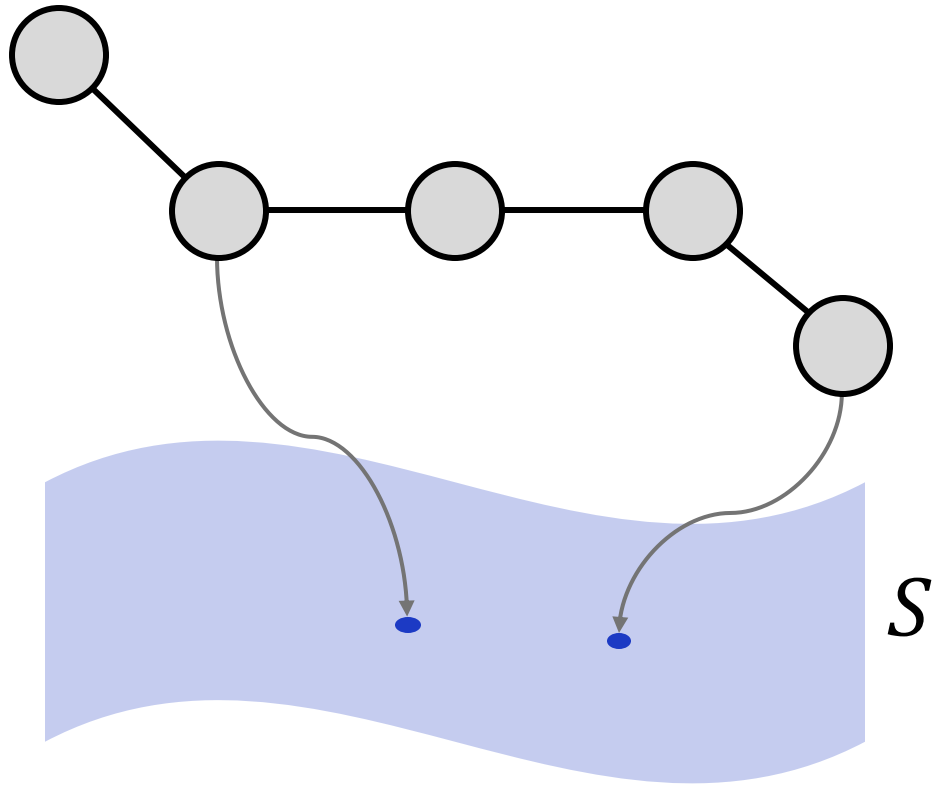
Opinion dynamics^[1] provides a nice perspective



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

Why sheaves?

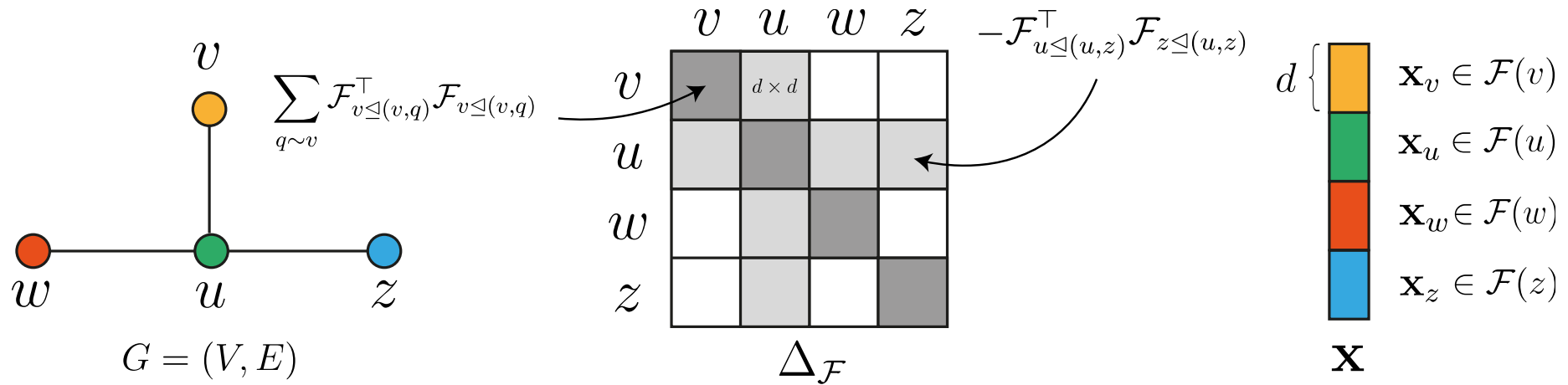
The underlying topology models the heterogeneity



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

Neural Sheaf Diffusion^[1]

Attaches a sheaf to a **Graph Convolutional Network**



$$\mathbf{Y} = \sigma((\mathbf{I}_{nd} - \Delta_{\mathcal{F}})(\mathbf{I}_n \otimes \mathbf{W}_1)\mathbf{X}\mathbf{W}_2)$$

$$\mathcal{F}_{u \leq e} = \text{MLP}(\mathbf{x}_u \parallel \mathbf{x}_v)$$

NSD performs well on benchmarks

NSD is smaller than R-GCN with similar performance

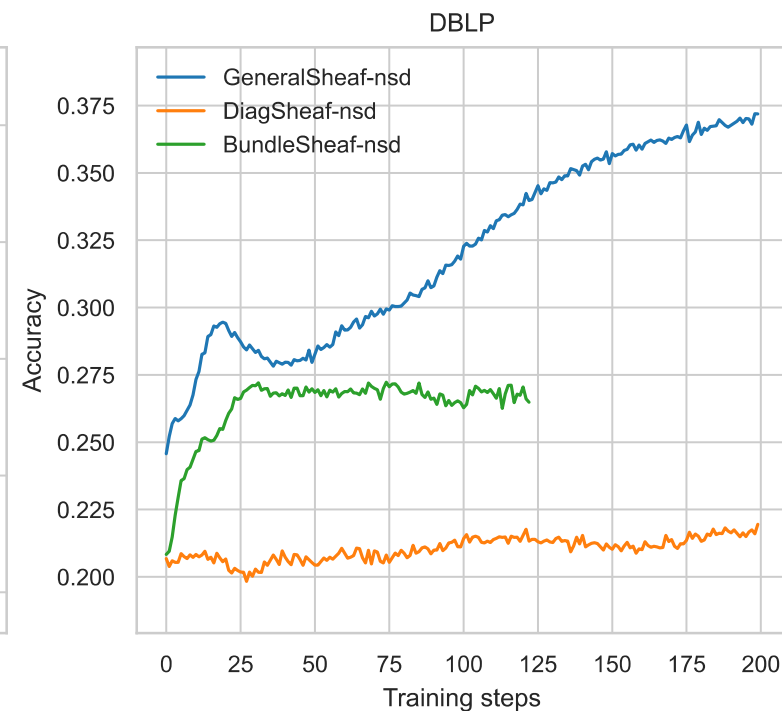
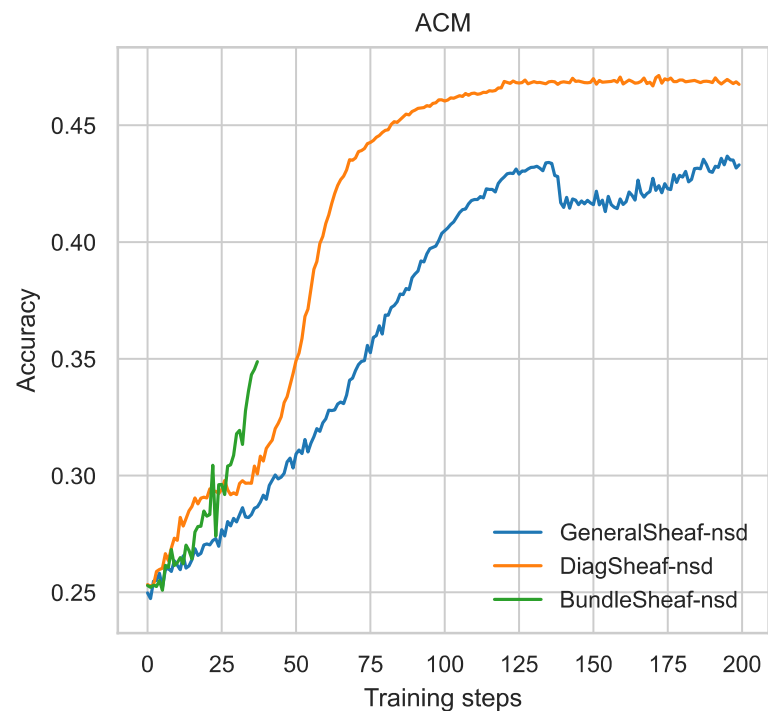
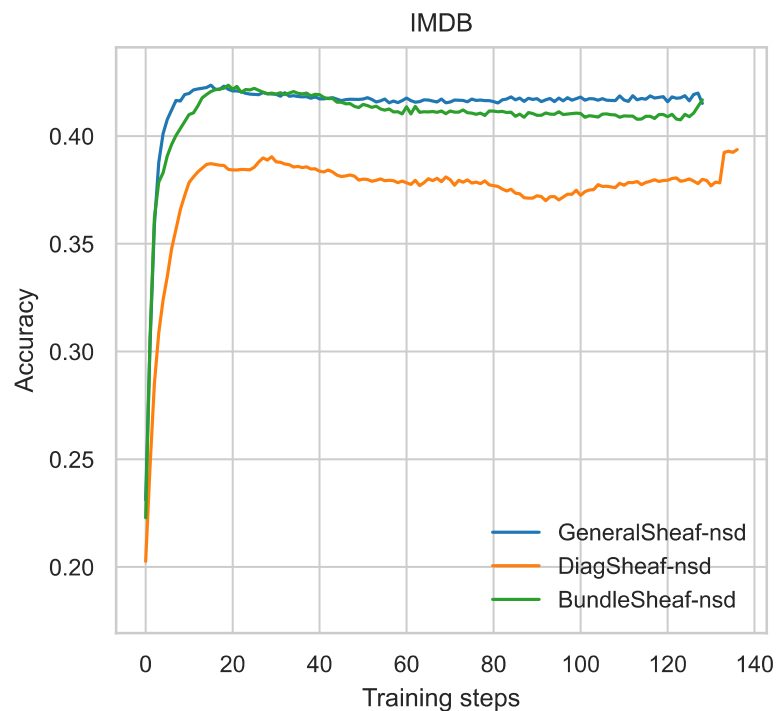
	ACM		DBLP		IMDB	
	Macro F1	Micro F1	Macro F1	Micro F1	Macro F1	Micro F1
GAT	75.80 ± 10.69	77.91 ± 8.66	95.47 ± 0.44	95.70 ± 0.42	84.12 ± 0.96	85.31 ± 0.92
GCN	89.09 ± 3.66	89.14 ± 3.60	96.31 ± 0.73	96.57 ± 0.63	82.41 ± 1.15	83.99 ± 0.92
HAN	86.95 ± 6.19	86.64 ± 6.43	94.74 ± 0.81	95.01 ± 0.73	13.53 ± 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

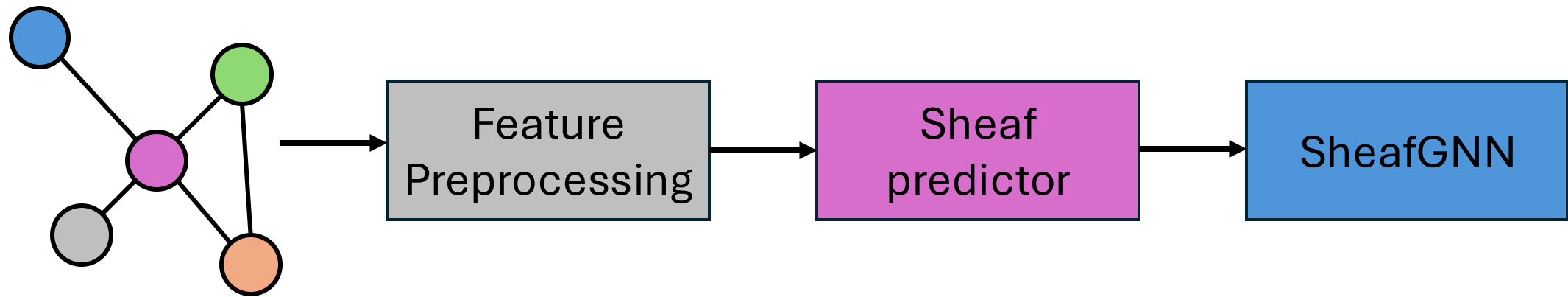
Sheaves implicitly learn types



HETSHEAF

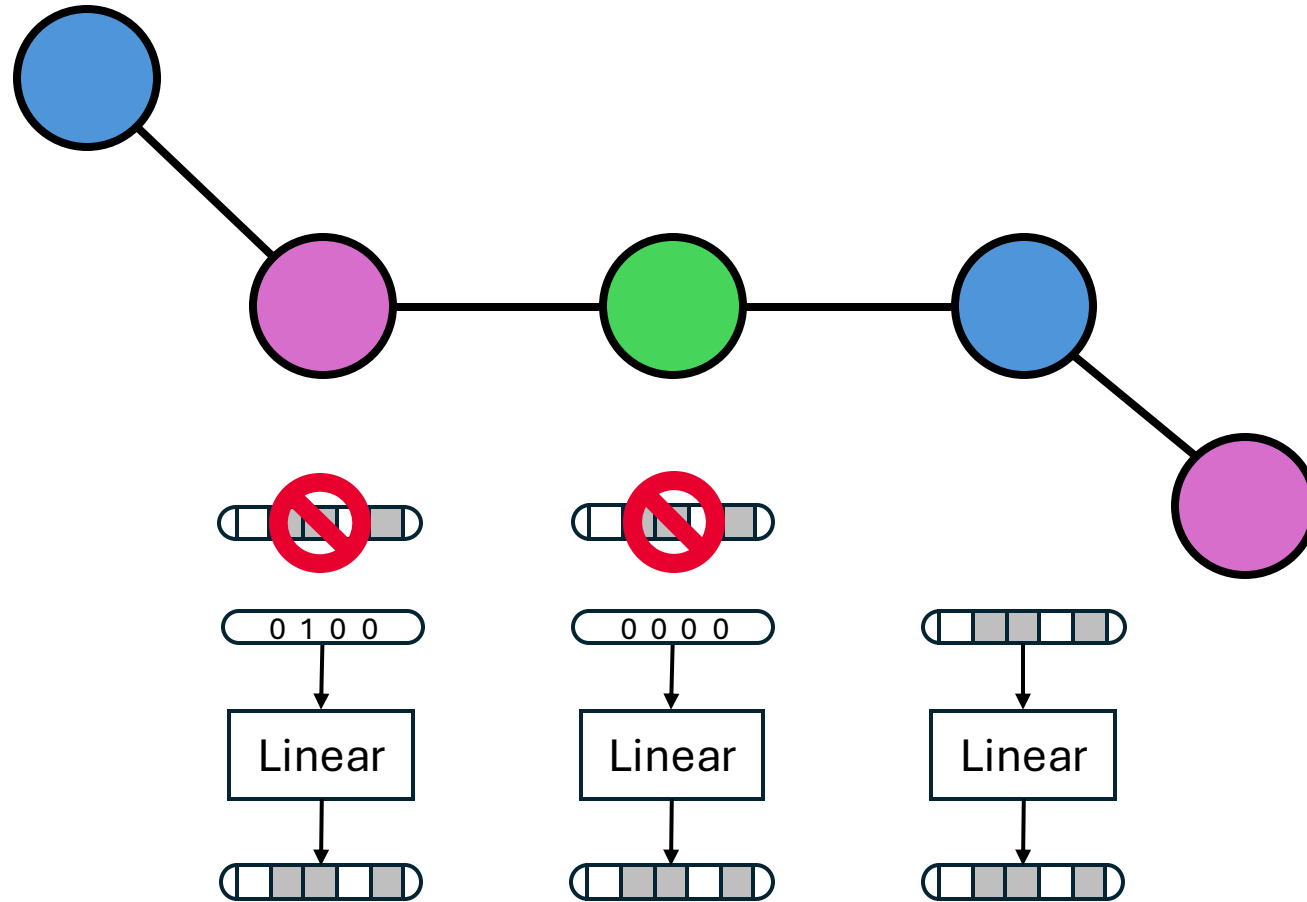
A general framework for heterogeneous sheaf neural networks

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

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

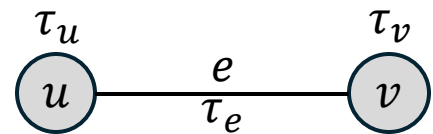
node features in $\mathcal{F}(v)$

type of node v

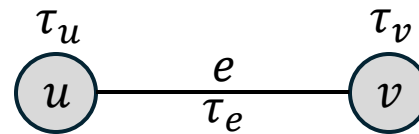
The diagram illustrates the components of the heterogeneous sheaf predictor formula. Blue arrows point from descriptive labels to the corresponding arguments in the function Φ :

- An arrow from "node features in $\mathcal{F}(u)$ " points to \mathbf{x}_u .
- An arrow from "type of node u " points to $\phi(u)$.
- An arrow from "type of node v " points to $\phi(v)$.
- An arrow from "type of edge e " points to $\psi(e)$.

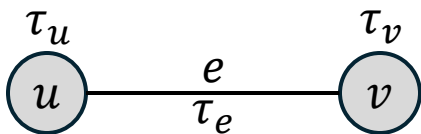
The label "node features in $\mathcal{F}(v)$ " is positioned below the formula but has no arrow pointing to a specific argument.

Sheaf-NSD

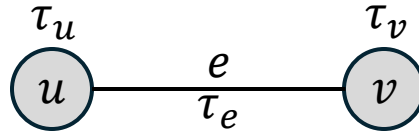
$$\mathcal{F}_{u \leq e} = \text{MLP}(\mathbf{x}_u \| \mathbf{x}_v)$$

Sheaf-ensemble

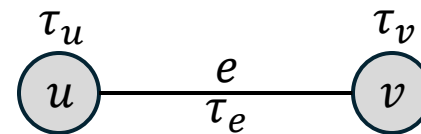
$$\mathcal{F}_{u \leq e} = \text{MLP}_{\tau_e}(\mathbf{x}_u \| \mathbf{x}_v)$$

Sheaf-NE

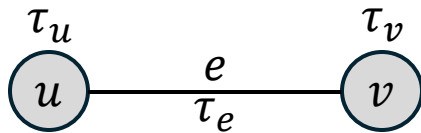
$$\mathcal{F}_{u \leq e} = \text{MLP}(\mathbf{x}_u \| \mathbf{x}_v \| \tau_u \| \tau_v)$$

Sheaf-EE

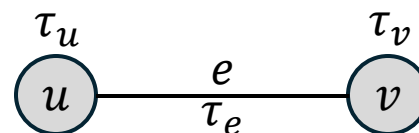
$$\mathcal{F}_{u \leq e} = \text{MLP}(\mathbf{x}_u \| \mathbf{x}_v \| \tau_e)$$

Sheaf-TE

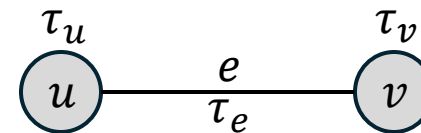
$$\mathcal{F}_{u \leq e} = \text{MLP}(\mathbf{x}_u \| \mathbf{x}_v \| \tau_u \| \tau_v \| \tau_e)$$

Sheaf-NT

$$\mathcal{F}_{u \leq e} = \text{MLP}(\tau_u \| \tau_v)$$

Sheaf-ET

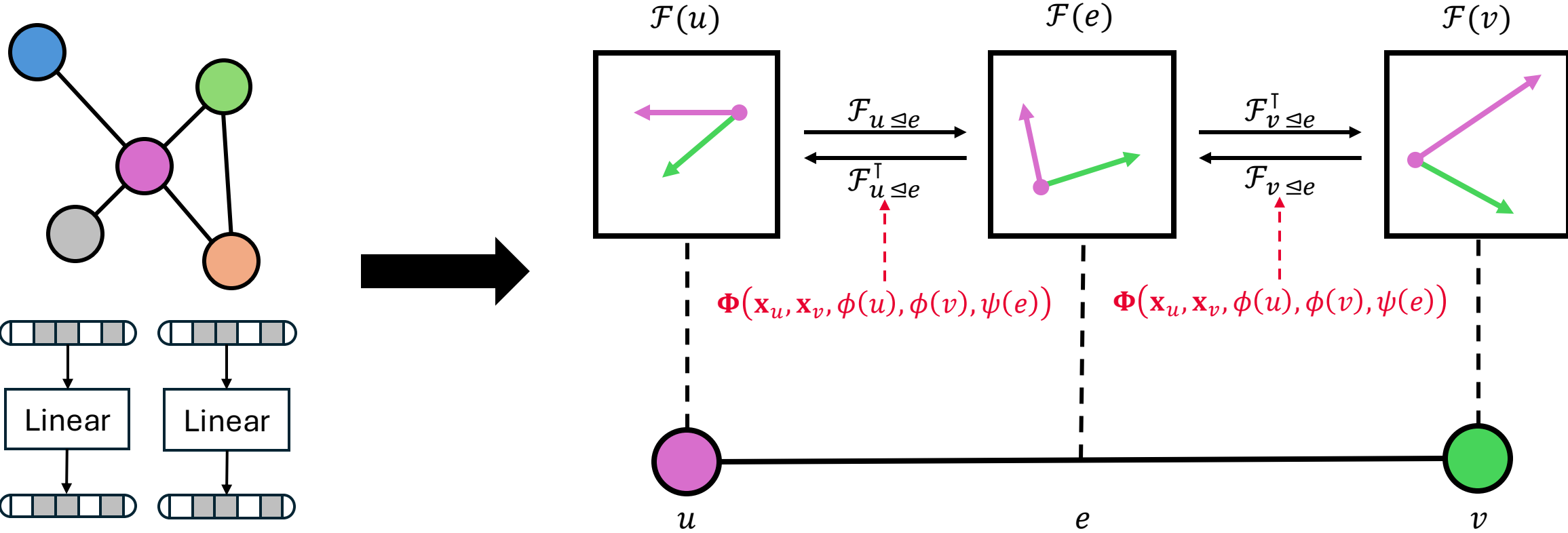
$$\mathcal{F}_{u \leq e} = \text{MLP}(\tau_e)$$

Sheaf-types

$$\mathcal{F}_{u \leq e} = \text{MLP}(\tau_u \| \tau_v \| \tau_e)$$

* Each type is assumed to be a one-hot encoded vector, $\tau_e := \mathbf{e}_{\psi(e)}$ for $e \in \mathcal{E}$ and $\tau_u := \mathbf{e}_{\phi(u)}$ 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 \pm 107.0	77.91 \pm 8.66	95.47 \pm 0.44	95.70 \pm 0.42	84.12 \pm 0.96	85.31 \pm 0.92
GCN [47]	89.09 \pm 3.66	89.14 \pm 3.60	96.31 \pm 0.73	96.57 \pm 0.63	82.41 \pm 1.15	83.99 \pm 0.92
HAN [74]	86.95 \pm 6.19	86.64 \pm 6.43	94.74 \pm 0.81	95.01 \pm 0.73	13.53 \pm 0.24	38.70 \pm 1.13
RGCN [62]	95.81 \pm 0.39	95.75 \pm 0.39	96.79 \pm 0.39	97.01 \pm 0.34	88.16 \pm 0.67	89.08 \pm 0.63
HGT [41]	93.24 \pm 3.19	93.30 \pm 2.91	93.91 \pm 1.08	94.26 \pm 1.09	87.74 \pm 0.76	88.45 \pm 0.71
O(d)-nsd [7]	94.64 \pm 1.02	94.59 \pm 1.03	96.32 \pm 0.46	96.55 \pm 0.42	86.35 \pm 1.29	87.20 \pm 1.07
Diag-nsd [7]	94.42 \pm 0.51	94.42 \pm 0.48	95.25 \pm 0.70	95.52 \pm 0.67	86.36 \pm 0.94	87.26 \pm 0.78
Gen-nsd [7]	94.97 \pm 0.41	94.94 \pm 0.42	96.69 \pm 0.82	96.89 \pm 0.79	86.70 \pm 0.90	87.50 \pm 0.78
Sheaf-TE (ours)	96.11 \pm 0.49	96.09 \pm 0.51	97.93 \pm 0.36	98.08 \pm 0.31	86.85 \pm 0.81	87.67 \pm 0.80
Sheaf-ensemble (ours)	96.16 \pm 0.52	96.12 \pm 0.54	97.46 \pm 0.64	97.62 \pm 0.60	86.92 \pm 1.10	87.79 \pm 0.95
Sheaf-NE (ours)	96.13 \pm 0.39	96.09 \pm 0.38	97.68 \pm 0.55	97.83 \pm 0.51	86.87 \pm 1.01	87.73 \pm 0.81
Sheaf-EE (ours)	96.39 \pm 0.37	96.35 \pm 0.36	97.57 \pm 0.69	97.73 \pm 0.62	87.12 \pm 0.75	87.88 \pm 0.67
Sheaf-NT (ours)	96.12 \pm 0.36	96.12 \pm 0.32	97.88 \pm 0.47	98.04 \pm 0.43	86.92 \pm 0.95	87.76 \pm 0.85
Sheaf-ET (ours)	95.84 \pm 0.65	95.82 \pm 0.65	97.69 \pm 0.47	97.83 \pm 0.47	86.12 \pm 0.82	87.05 \pm 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.

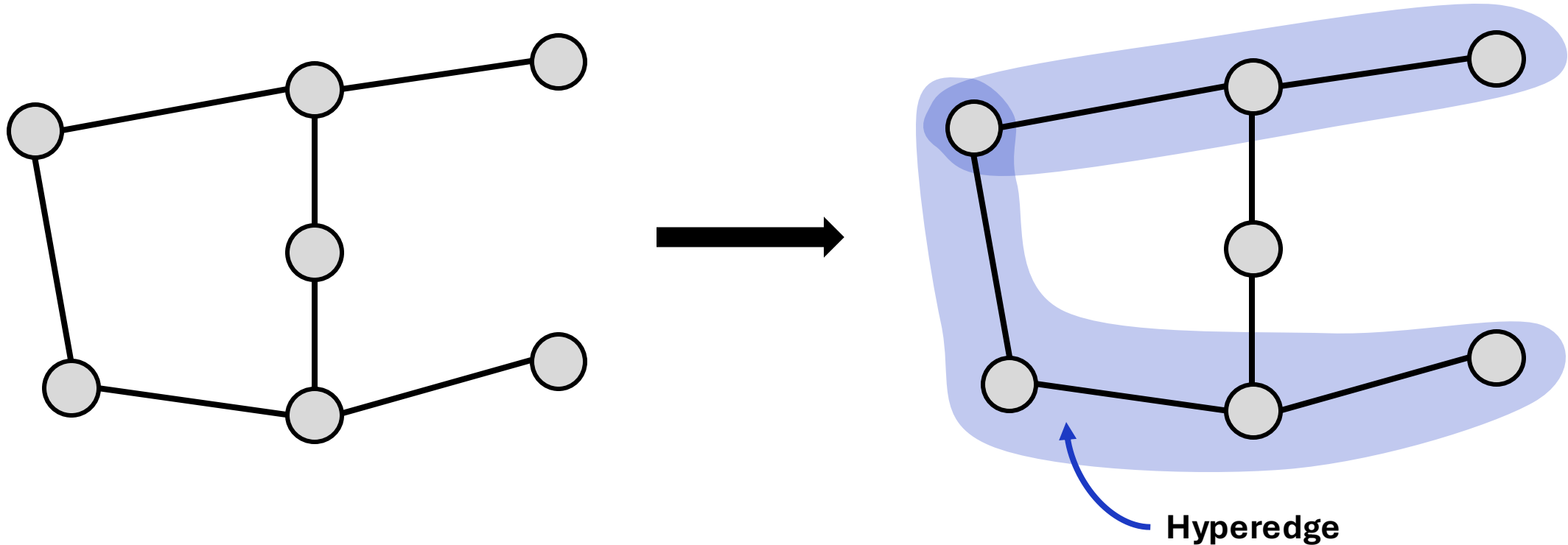
	LastFM		MovieLens	
	AUPR	AUROC	AUPR	AUROC
GAT	62.88 \pm 0.18	50.69 \pm 0.63	97.06 \pm 0.24	97.47 \pm 0.21
GCN	96.84 \pm 0.10	96.42 \pm 0.08	99.57 \pm 0.03	99.51 \pm 0.03
HAN	82.48 \pm 3.86	78.47 \pm 3.04	63.49 \pm 0.14	52.06 \pm 0.27
R-GCN	96.86 \pm 0.07	96.97 \pm 0.05	99.06 \pm 0.05	99.13 \pm 0.04
HGT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

Future work

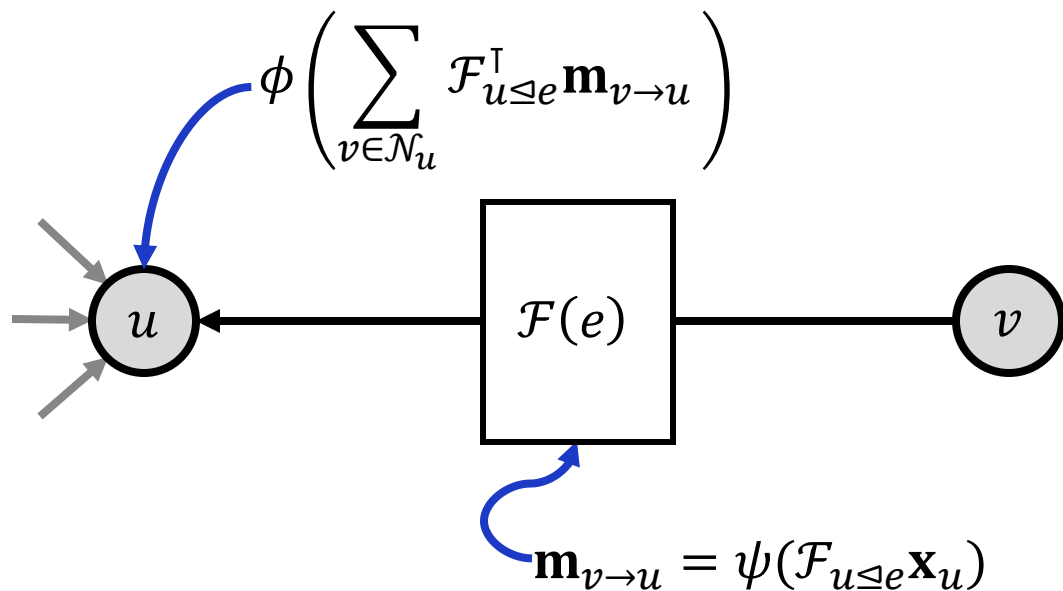
- Lifting to hypergraphs
- Generalised sheaf message passing
- Topological sheaves

Accounting for higher order interactions

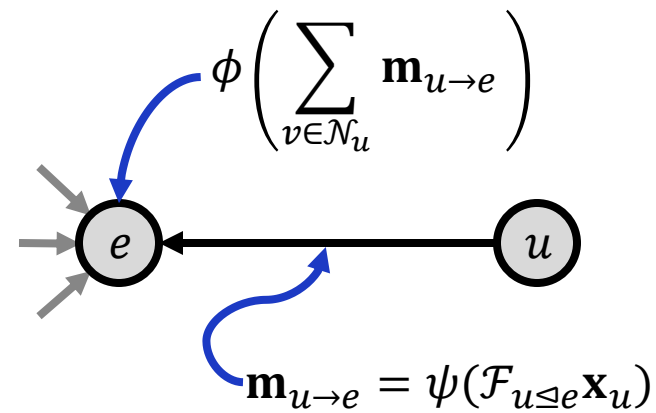
Hypergraphs connect an **arbitrary set** of nodes



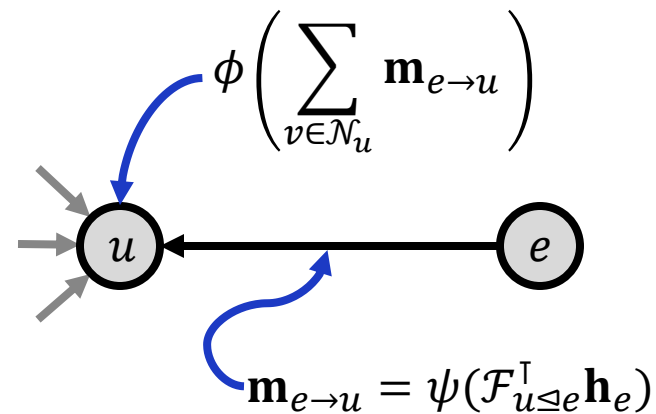
Generalised sheaf message passing



(1)

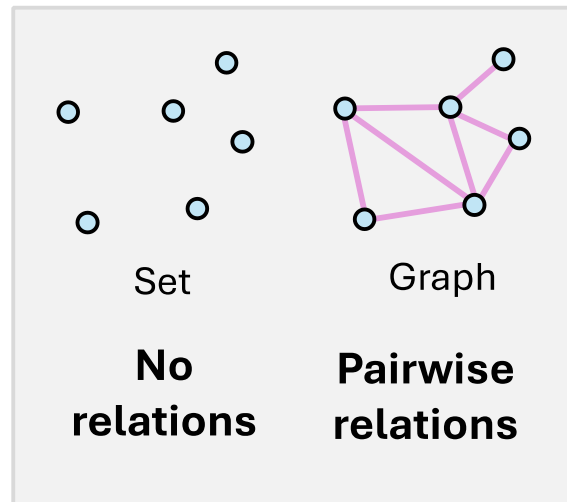


(2)



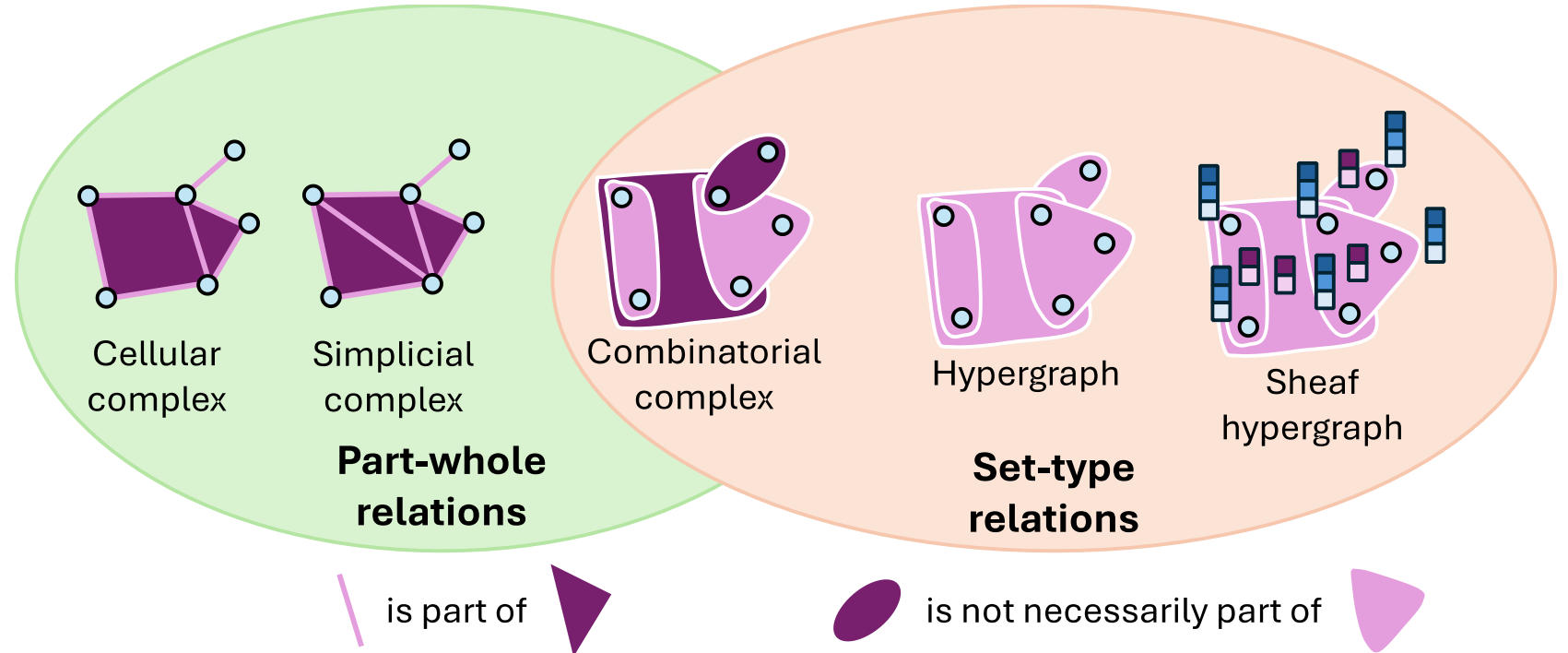
Sheaf Topological Neural Networks

Traditional discrete domains



○ : Nodes

— : Edges



— is part of

is not necessarily part of

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