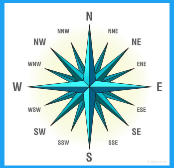
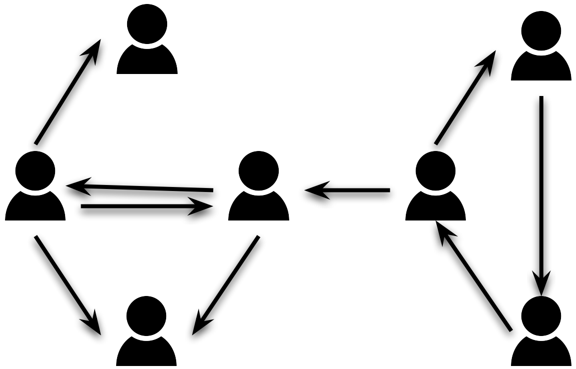
A Tale of Edge Directionality in Graph Neural Networks

Emanuele Rossi, Imperial College London
September 2023, Temporal Graph Reading Group



Graphs are Often Directed

Citation, social (interaction) and hyperlink networks among others



TGB Link Prediction Datasets

Scale	Name	Package	#Nodes	#Edges*	#Steps	Surprise	Metric
Small	tgbl-wiki-v2	0.1.2	9,227	157,474	152,757	0.108	MRR
Small	tgbl-review-v2	0.1.2	352,637	4,873,540	6,865	0.987	MRR
medium	tgbl-coin	0.1.2	638,486	22,809,486	1,295,720	0.120	MRR
large	tgbl-comment	0.1.2	994,790	44,314,507	30,998,030	0.823	MRR
large	tgbl-flight	0.1.2	18,143	67,169,570	1,385	0.024	MRR

Directed

The "Undirectedness" Assumption

Spectral GNNs [1] require an undirected graph to define convolution

$$\mathbf{y} = f_\theta(\mathbf{L})\mathbf{x}$$
$$= f_\theta(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top)\mathbf{x}$$
$$= \mathbf{U} f_\theta(\mathbf{\Lambda})\mathbf{U}^\top \mathbf{x}$$

Eigendecomposition requires a **symmetric Laplacian** → the graph has to be **undirected**

[1] M. Defferrard et al., "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering", NeurIPS 2016

The “Undirectedness” Assumption

Spatial Methods (MPNNs) also fail to deal with directionality

in- and out-neighbors treated equally

$$\mathbf{m}_i^{(k)} = \mathrm{AGG}^{(k)}\left(\{\{\mathbf{x}_j^{(k-1)} : (i,j) \in E\}\}\right)$$
$$\mathbf{x}_i^{(k)} = \mathrm{COM}^{(k)}\left(\mathbf{x}_i^{(k-1)}, \mathbf{m}_i^{(k)}\right)$$

[2] J. Gilmer et al., “Neural Message Passing for Quantum Chemistry”, ICML 2017

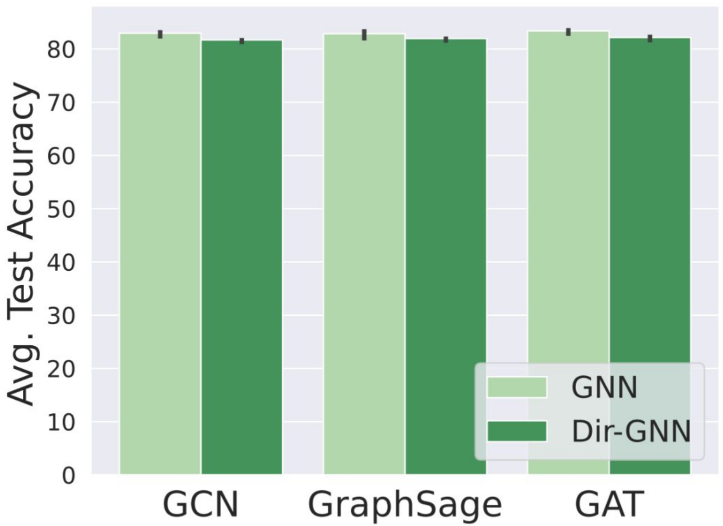
The “Undirectedness” Assumption

Making the graph undirected has become part of the standard preprocessing

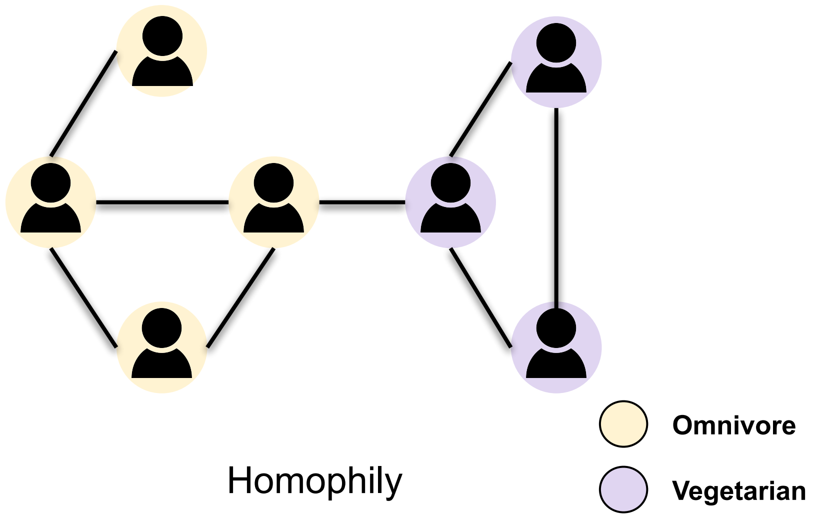
```python
def parse_npz(f):
    x = sp.csr_matrix((f['attr_data'], f['attr_indices'], f['attr_indptr']),
                      f['attr_shape']).todense()
    x = torch.from_numpy(x).to(torch.float)
    x[x > 0] = 1

    adj = sp.csr_matrix((f['adj_data'], f['adj_indices'], f['adj_indptr']),
                        f['adj_shape']).tocoo()
    row = torch.from_numpy(adj.row).to(torch.long)
    col = torch.from_numpy(adj.col).to(torch.long)
    edge_index = torch.stack([row, col], dim=0)
    edge_index, _ = remove_self_loops(edge_index)
    edge_index = to_undirected(edge_index, num_nodes=x.size(0))

    y = torch.from_numpy(f['labels']).to(torch.long)

    return Data(x=x, edge_index=edge_index, y=y)
```

The "Undirectedness" Assumption

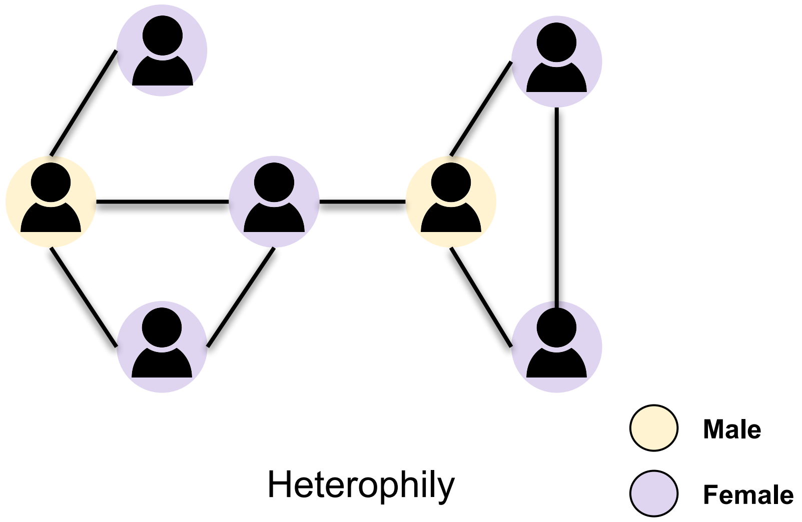
Undirected graphs perform equally well in common (homophilic) benchmarks



Homophily and Heterophily

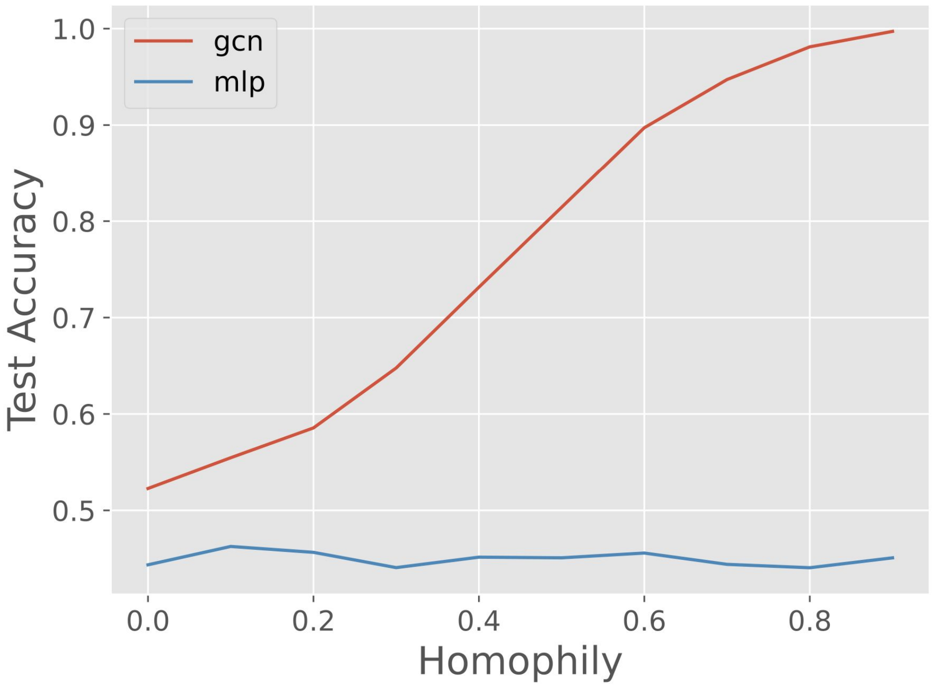


Homophily



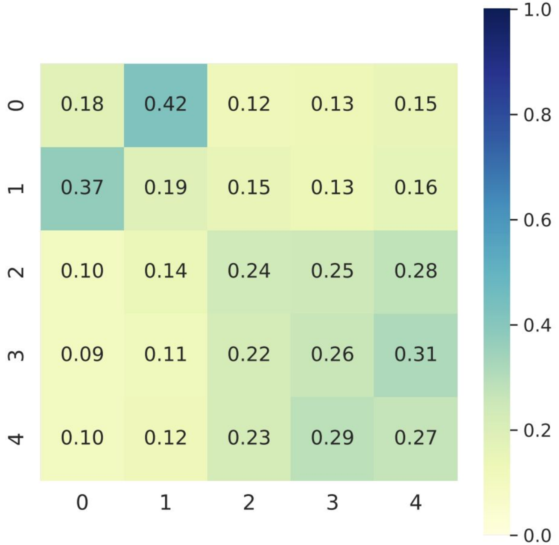
Heterophily

GNNs Struggle on Heterophilic Data



Measuring Homophily

Undirected Graphs



Class Compatibility Matrix

$$h = \frac{1}{|V|} \sum_{i \in V} \frac{\sum_{j \in \mathcal{N}(i)} I[y_i = y_j]}{d_i}$$

Node Homophily

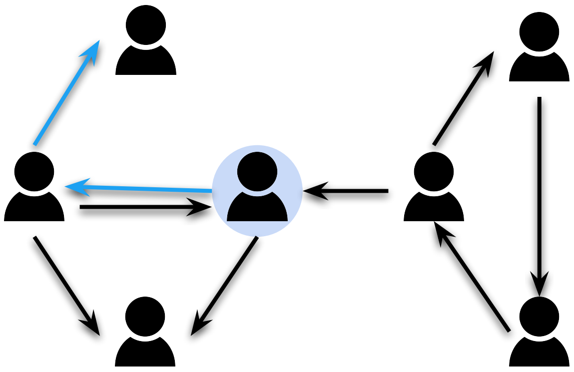
Measuring Homophily

Weighted directed graphs

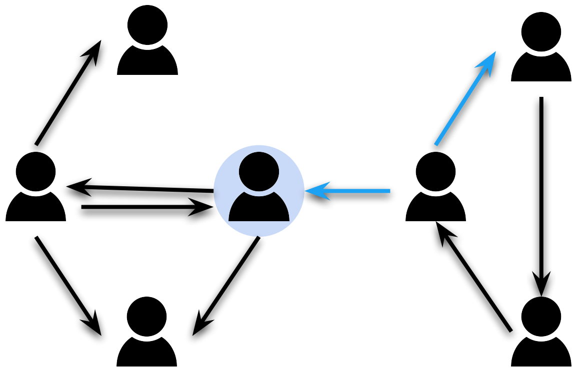
$$h(\mathbf{S}) = \frac{1}{|V|} \sum_{i \in V} \frac{\sum_{j \in \mathcal{N}(i)} s_{ij} I[y_i = y_j]}{\sum_{j \in \mathcal{N}(i)} s_{ij}}$$

Directed 2-hops

There are four different 2-hops for directed graphs



$\mathbf{A}^2$



$\mathbf{A}^\top \mathbf{A}$

Effective Homophily

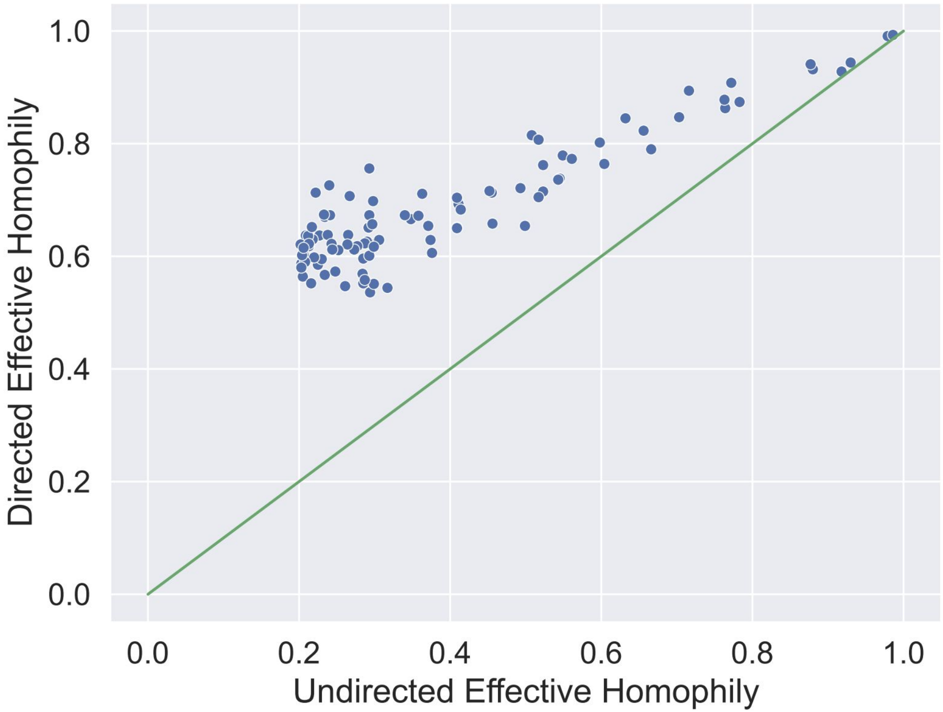
Going beyond the immediate neighbors

$$h^{(\mathrm{eff})} = \max_{k \geq 1} \max_{\mathbf{C} \in \mathcal{B}^k} h(\mathbf{C})$$

Higher-order hops

Directionality Enhances Effective Homophily

Synthetic graphs



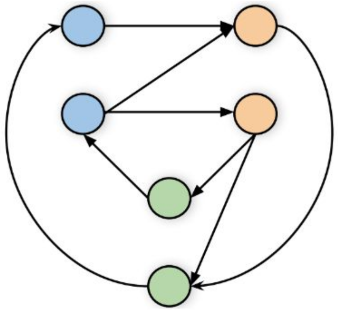
Directionality Enhances Effective Homophily

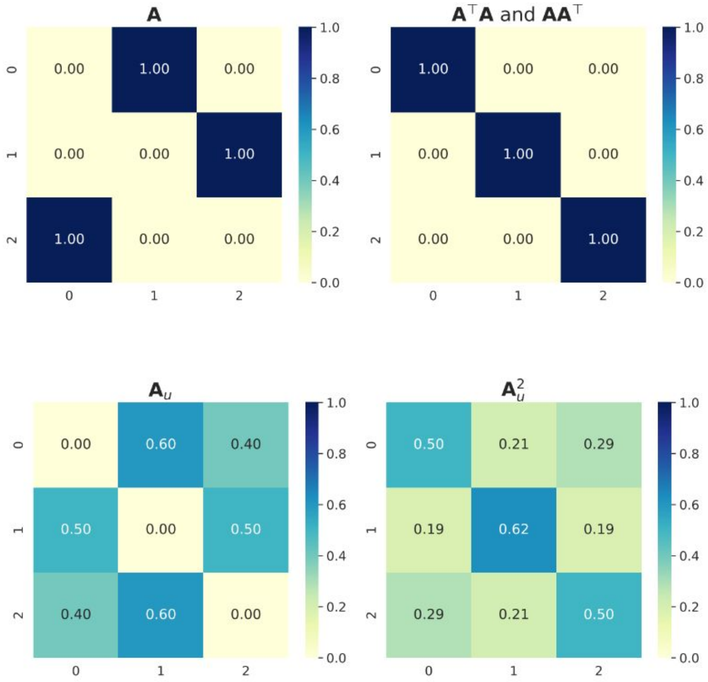
Real-world datasets

		$\mathbf{A}_u$	$\mathbf{A}_u^2$	$h_u^{(\text{eff})}$	$\mathbf{A}$	$\mathbf{A}^\top$	$\mathbf{A}^\top\mathbf{A}$	$\mathbf{A}\mathbf{A}^\top$	$h_d^{(\text{eff})}$	$h_{\text{gain}}^{(\text{eff})}$
Homophilic	CITESEER-FULL	0.958	0.951	0.958	0.954	0.959	0.971	0.951	0.971	1.36%
	CORA-ML	0.810	0.767	0.810	0.808	0.833	0.803	0.779	0.833	2.84%
	OGBN-ARXIV	0.635	0.548	0.635	0.632	0.675	0.658	0.556	0.675	6.3%
Heterophilic	CHAMELEON	0.248	0.331	0.331	0.249	0.274	0.383	0.335	0.383	15.71%
	SQUIRREL	0.218	0.252	0.252	0.219	0.210	0.257	0.258	0.258	2.38%
	ARXIV-YEAR	0.289	0.397	0.397	0.310	0.403	0.487	0.431	0.487	22.67%
	SNAP-PATENTS	0.221	0.372	0.372	0.266	0.271	0.478	0.522	0.522	40.32%
	ROMAN-EMPIRE	0.046	0.365	0.365	0.045	0.042	0.535	0.609	0.609	66.85%

Directionality Enhances Effective Homophily

An intuitive example





Dir-GNN

Aggregate from both in- and out-neighbors, but separately

$$\mathbf{m}_{i,\leftarrow}^{(k)} = \mathrm{AGG}_{\leftarrow}^{(k)}\left(\{\{\mathbf{x}_j^{(k-1)} : (j,i) \in E\}\}\right)$$
$$\mathbf{m}_{i,\rightarrow}^{(k)} = \mathrm{AGG}_{\rightarrow}^{(k)}\left(\{\{\mathbf{x}_j^{(k-1)} : (i,j) \in E\}\}\right)$$
$$\mathbf{x}_i^{(k)} = \mathrm{COM}^{(k)}\left(\mathbf{x}_i^{(k-1)}, \mathbf{m}_{i,\leftarrow}^{(k)}, \mathbf{m}_{i,\rightarrow}^{(k)}\right)$$

Separate aggregation
of in- and out-neighbors

From GCN to Dir-GCN

A general framework which can be used to extend any MPNN to directed graphs

$$\mathbf{X}^{(k)} = \sigma\left(\mathbf{A}_u \mathbf{X}^{(k-1)} \mathbf{W}^{(k)}\right)$$

$$\tilde{\mathbf{A}}_u = \mathbf{D}_u^{-1/2} \mathbf{A}_u \mathbf{D}_u^{-1/2}$$

$$\mathbf{X}^{(k)} = \sigma\left(\mathbf{A}_{\rightarrow} \mathbf{X}^{(k-1)} \mathbf{W}_{\rightarrow}^{(k)} + \mathbf{A}_{\rightarrow}^{\top} \mathbf{X}^{(k-1)} \mathbf{W}_{\leftarrow}^{(k)}\right)$$

$$\mathbf{A}_{\rightarrow} = \mathbf{D}_{\rightarrow}^{-1/2} \mathbf{A} \mathbf{D}_{\leftarrow}^{-1/2}$$

Dir-GNN Leads to More Homophilic Aggregations

It treats different 2-hops differently

$$\mathbf{X}^{(2)} = \mathbf{A}^2_{\rightarrow}\mathbf{X}^{(0)}\mathbf{W}^{(1)}_{\rightarrow}\mathbf{W}^{(2)}_{\rightarrow} + (\mathbf{A}^{\top}_{\rightarrow})^2\mathbf{X}^{(0)}\mathbf{W}^{(1)}_{\leftarrow}\mathbf{W}^{(2)}_{\leftarrow}$$
$$+ \boxed{\mathbf{A}_{\rightarrow}\mathbf{A}^{\top}_{\rightarrow}}\mathbf{X}^{(0)}\mathbf{W}^{(1)}_{\leftarrow}\mathbf{W}^{(2)}_{\rightarrow} + \boxed{\mathbf{A}^{\top}_{\rightarrow}\mathbf{A}_{\rightarrow}}\mathbf{X}^{(0)}\mathbf{W}^{(1)}_{\rightarrow}\mathbf{W}^{(2)}_{\leftarrow}$$

Expressivity Analysis

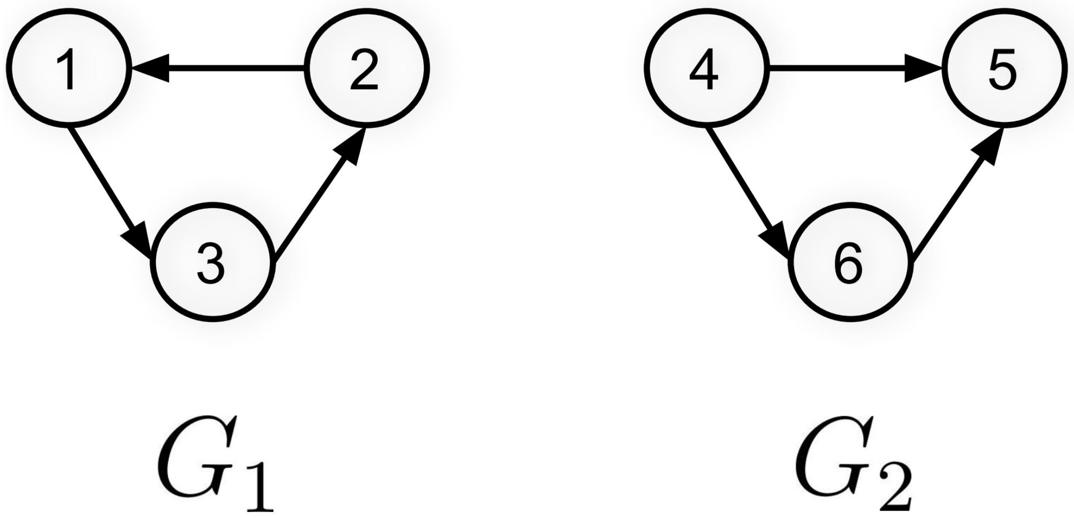
Dir-GNN is strictly more expressive than MPNNs

Theorem 4.1 (Informal). *Dir-GNN is as expressive as D-WL if* $\mathrm{AGG}_{\rightarrow}^{(k)}$, $\mathrm{AGG}_{\leftarrow}^{(k)}$, *and* $\mathrm{COM}^{(k)}$ *are injective for all* k.

Theorem 4.2 (Informal). *Dir-GNN is strictly more expressive than both MPNN-U and MPNN-D.*

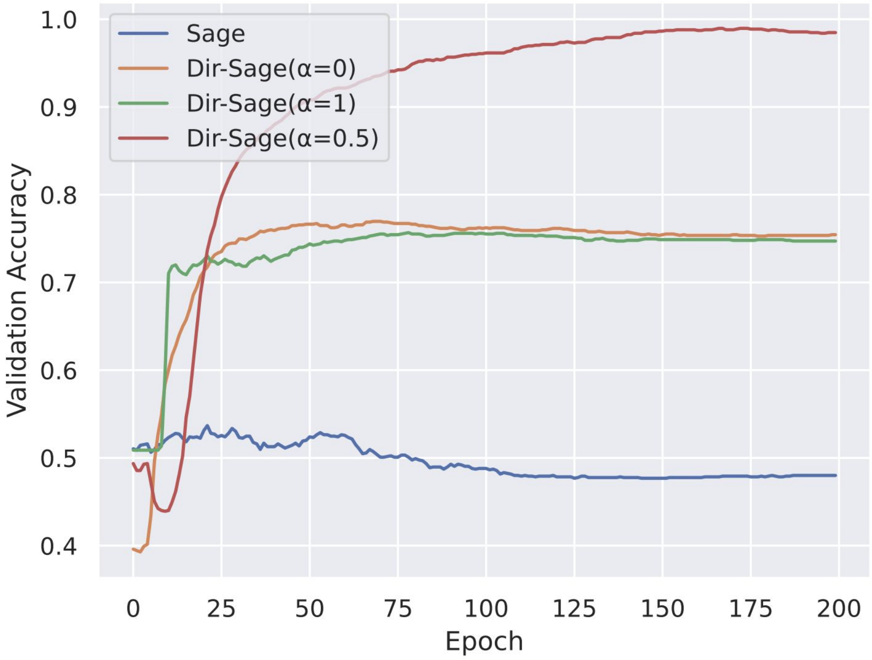
Dir-GNN $\sqsubseteq$ MPNN-U

MPNN-U fails to distinguish the two graphs below



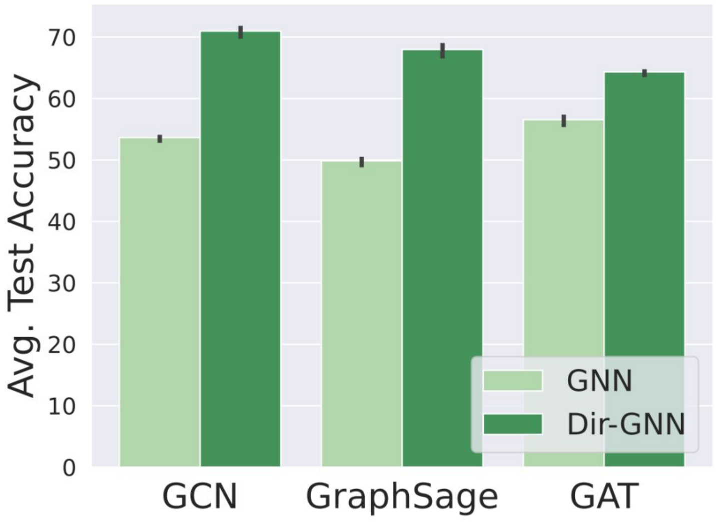
Empirical Results

Synthetic task where the label of a node depends both on in- and out-neighbors

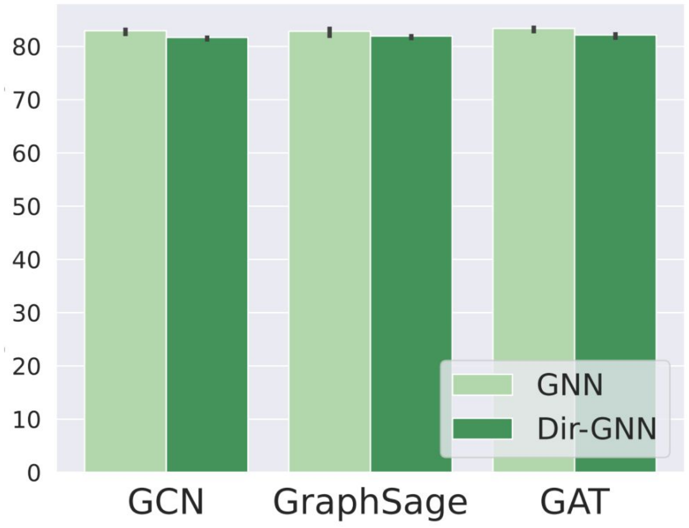


Empirical Results

Directionality leads to significant improvement on heterophilic datasets



(a) Heterophilic Graphs



(b) Homophilic Graphs

Empirical Results

Dir-GNN achieves state-of-the-art results on five heterophilic benchmarks

	Squirrel	Chameleon	Arxiv-year	Snap-patents	Roman-Empire
MLP	28.77 ± 1.56	46.21 ± 2.99	36.70 ± 0.21	31.34 ± 0.05	64.94 ± 0.62
GCN	53.43 ± 2.01	64.82 ± 2.24	46.02 ± 0.26	51.02 ± 0.06	73.69 ± 0.74
H_2GCN	37.90 ± 2.02	59.39 ± 1.98	49.09 ± 0.10	OOM	60.11 ± 0.52
GPR-GNN	54.35 ± 0.87	62.85 ± 2.90	45.07 ± 0.21	40.19 ± 0.03	64.85 ± 0.27
LINKX	61.81 ± 1.80	68.42 ± 1.38	56.00 ± 0.17	61.95 ± 0.12	37.55 ± 0.36
FSGNN	74.10 ± 1.89	78.27 ± 1.28	50.47 ± 0.21	65.07 ± 0.03	79.92 ± 0.56
ACM-GCN	67.40 ± 2.21	74.76 ± 2.20	47.37 ± 0.59	55.14 ± 0.16	69.66 ± 0.62
GloGNN	57.88 ± 1.76	71.21 ± 1.84	54.79 ± 0.25	62.09 ± 0.27	59.63 ± 0.69
Grad. Gating	64.26 ± 2.38	71.40 ± 2.38	63.30 ± 1.84	69.50 ± 0.39	82.16 ± 0.78
DiGCN	37.74 ± 1.54	52.24 ± 3.65	OOM	OOM	52.71 ± 0.32
MagNet	39.01 ± 1.93	58.22 ± 2.87	60.29 ± 0.27	OOM	88.07 ± 0.27
Dir-GNN	$\mathbf{75.31 \pm 1.92}$	$\mathbf{79.71 \pm 1.26}$	$\mathbf{64.08 \pm 0.26}$	$\mathbf{73.95 \pm 0.05}$	$\mathbf{91.23 \pm 0.32}$

Dir-GNN for Temporal Graphs

TGN [3] uses direction in message function, but discards it for the graph aggregation

Message Function

$$\mathbf{m}_u(t) = \mathrm{msg_s}\left(\mathbf{s}_u(t^-), \mathbf{s}_v(t^-), t, \mathbf{e}\right),$$
$$\mathbf{m}_v(t) = \mathrm{msg_d}\left(\mathbf{s}_v(t^-), \mathbf{s}_u(t^-), t, \mathbf{e}\right)$$

Direction-aware

Graph Aggregation

$$\mathbf{Z}(t) = \mathrm{GNN}(\mathbf{G}(t), \mathbf{E}(t), \mathbf{X}(t), \mathbf{S}(t))$$

Direction-unaware

[3] Rossi et al., “Temporal Graph Networks For Deep Learning On Dynamic Graphs”, ICML 2020 GRL Workshop;

Conclusion

Dir-GNN achieves state-of-the-art results on five heterophilic benchmarks

- Edge **directionality** has largely been **ignored** in GNNs
- Preserving directionality can make heterophilic datasets more **homophilic**
- We introduce **Dir-GNN**, a general framework for learning on directed graphs
- Dir-GNN is **more expressive** than MPNNs on directed graphs
- Dir-GNN leads to **large improvements** on **heterophilic** datasets
- Many temporal datasets are directed!

Questions?

@emaros96
www.emanuelerossi.co.uk