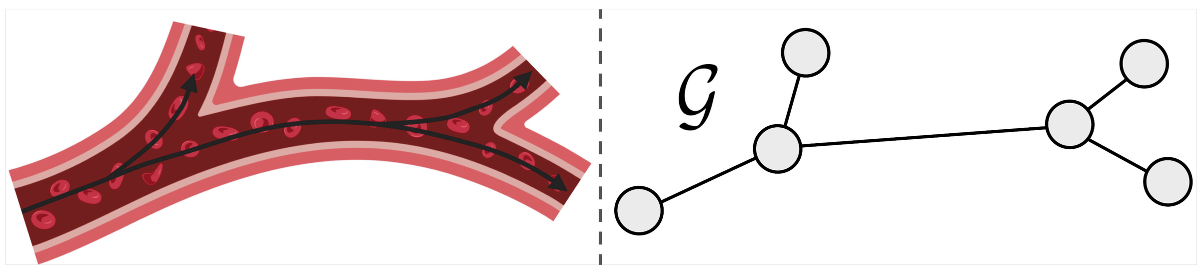
Link Prediction for Flow-Driven Spatial Networks

JAN 4-8 WACV 2024 WAIKOLOA HAWAII

Bastian Wittmann, University of Zurich
Johannes C. Paetzold, Technical University of Munich
Chinmay Prabhakar, University of Zurich
Daniel Rueckert, Technical University of Munich
Bjoern Menze, University of Zurich

Motivation:

Networks (or graphs) can serve as efficient representations of real-world, ultra-complex systems and can be further classified into different categories. A prominent category is represented by undirected networks embedded in a Euclidean space constrained by geometry, called spatial networks. In this work, we are focusing on **spatial networks, where a form of physical exchange or flow can be described characteristic functional properties of the underlying physical system** (eg, whole-brain vessel graphs, road networks, or global exchange networks). We refer to such networks as **flow-driven spatial networks**.



Flow-driven spatial network (right) representing vasculature (left).

The graph generation process of flow-driven spatial networks is, however, typically represented by a **multi-stage and imperfect process** (segmentation, skeletonization, and subsequent graph pruning), which **introduces noise and artifacts** to the graph representation. In this context, we introduce a novel link prediction algorithm tailored to flow-driven spatial networks to **tackle under-/over-connectivity**, a prominent artifact that limits the application of, for example, whole-brain vessel graphs for subsequent medically relevant downstream tasks, such as the diagnosis, treatment, and analysis of neurovascular brain disorders (eg, aneurysms or strokes).

Rationale & Contributions:

In this work, we aim to **model simplified physical flow in spatial networks** to introduce a **strong inductive bias** for the link prediction task in flow-driven spatial networks.

Our framework relies on the assumption that flow, as an **unobserved functional property**, should be reflected in the network's **structural properties** (such as bifurcation angles or length of edges).

We summarize our core contributions as follows:

1) We propose an **attentive, neighborhood-aware message-passing layer**, called GAV layer, which updates vector embeddings, mimicking the (change in) direction and magnitude of physical flow in spatial networks.

2) We introduce a **readout module** that aggregates vector embeddings in a physically plausible way and thus facilitates the interpretability of results.

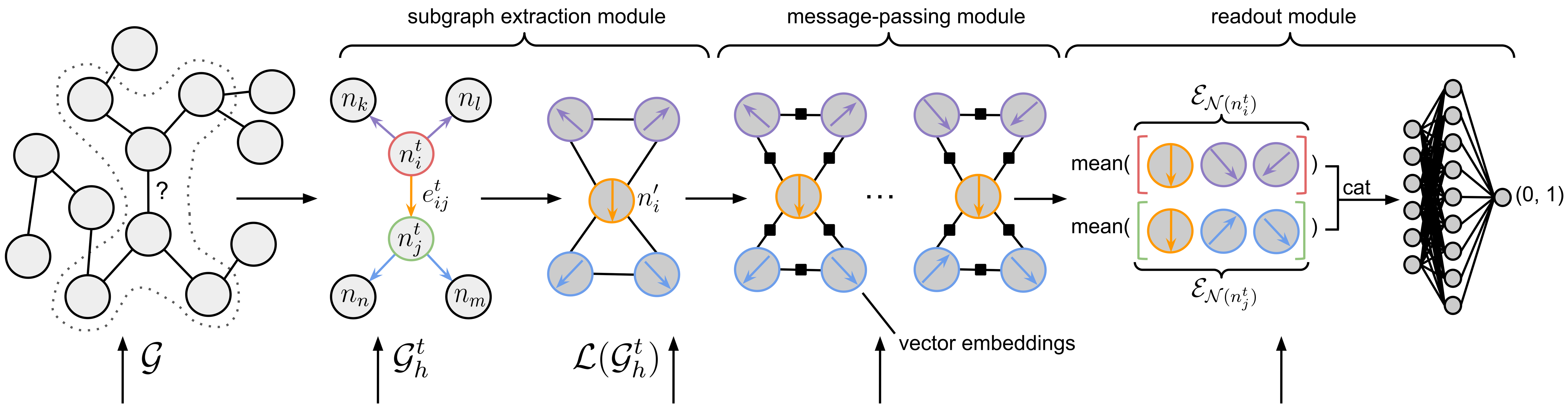
3) We demonstrate **state-of-the-art performance** across all metrics in extensive experiments on eight flow-driven spatial networks, including the **Open Graph Benchmark's ogbl-vessel benchmark (98.38 vs. 87.98 AUC)**.

Keywords: geometric deep learning, link prediction, spatial networks, graph neural networks, vessel graphs

GitHub

ogbl-vessel leaderboard

The Graph Attentive Vectors Link Prediction Framework (GAV):



flow-driven spatial network (eg, representative of vasculature)

extract **subgraph** to reduce computational complexity & **initialize flow in vessels as vector embeddings**

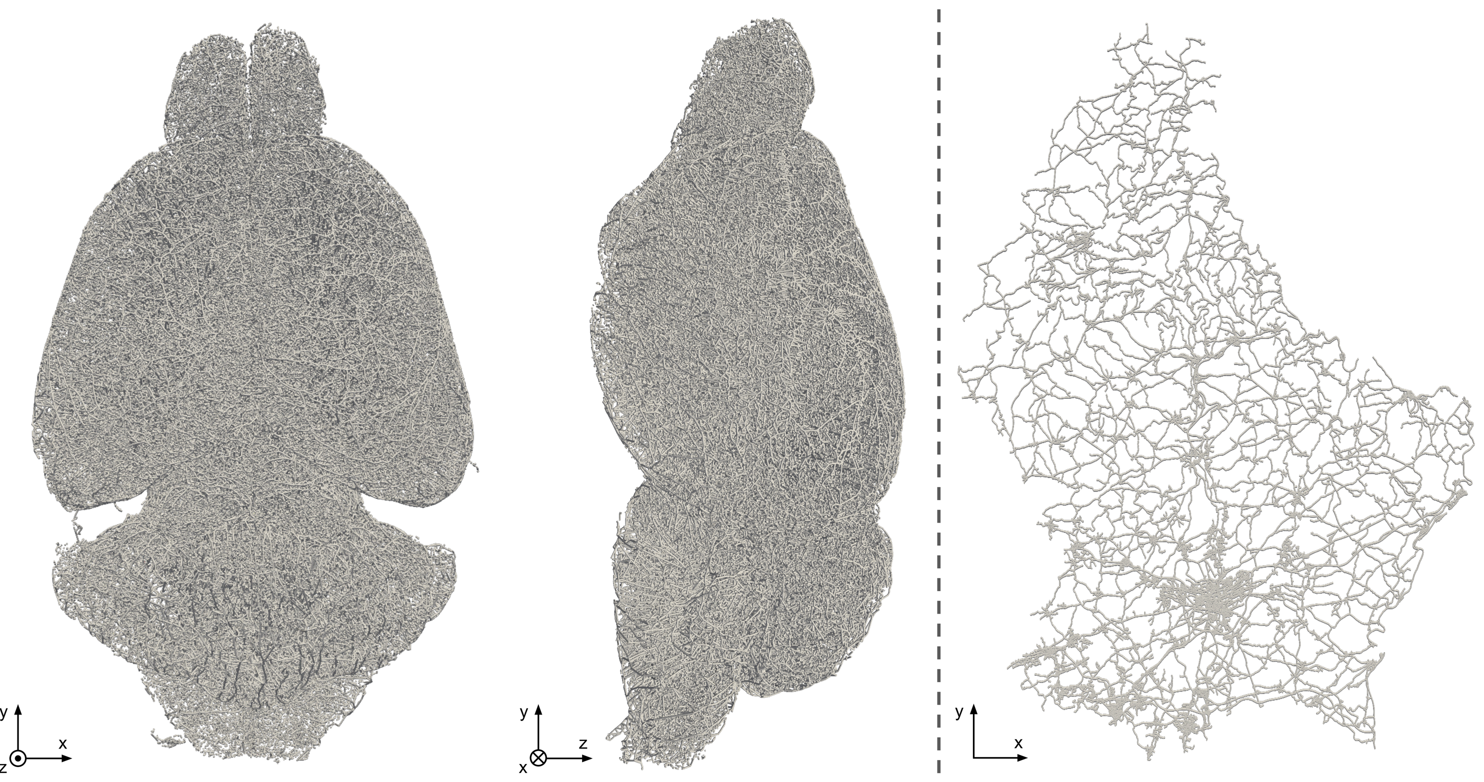
transform subgraph to a **line graph** to embed vector embeddings in nodes for message-passing (MP)

update vector embeddings mimicking simplified physical flow in spatial networks via MP

aggregate vector embeddings around target nodes to predict whether they should be connected or not

Datasets/Flow-Driven Spatial Networks:

We conduct experiments on eight flow-driven spatial networks, including the **ogbl-vessel benchmark** and seven additionally sourced datasets representing murine **whole-brain vessel graphs** acquired via different imaging techniques and **road networks** of four European countries.



Visualization of a whole-brain vessel graph and a road network.

The table below summarizes properties of the utilized datasets:

Dataset Name	# Nodes	# Edges	Node Features	Description
ogbl-vessel	3,538,495	5,345,897	x-, y-, z-coordinates	BALB/c mouse strain[1]
c57-tc-vessel	3,820,133	5,614,677	x-, y-, z-coordinates	C57BL/6 mouse strain[1]
cd1-tc-vessel	3,645,963	5,791,309	x-, y-, z-coordinates	CD-1 mouse strain[1]
c57-cc-vessel	6,650,580	9,054,100	x-, y-, z-coordinates	C57BL/6 mouse strain[2]
belgium-road	1,441,295	1,549,970	x-, y-coordinates	Belgium
italy-road	6,686,493	7,013,978	x-, y-coordinates	Italy
netherlands-road	2,216,688	2,441,238	x-, y-coordinates	Netherlands
luxembourg-road	114,599	119,666	x-, y-coordinates	Luxembourg

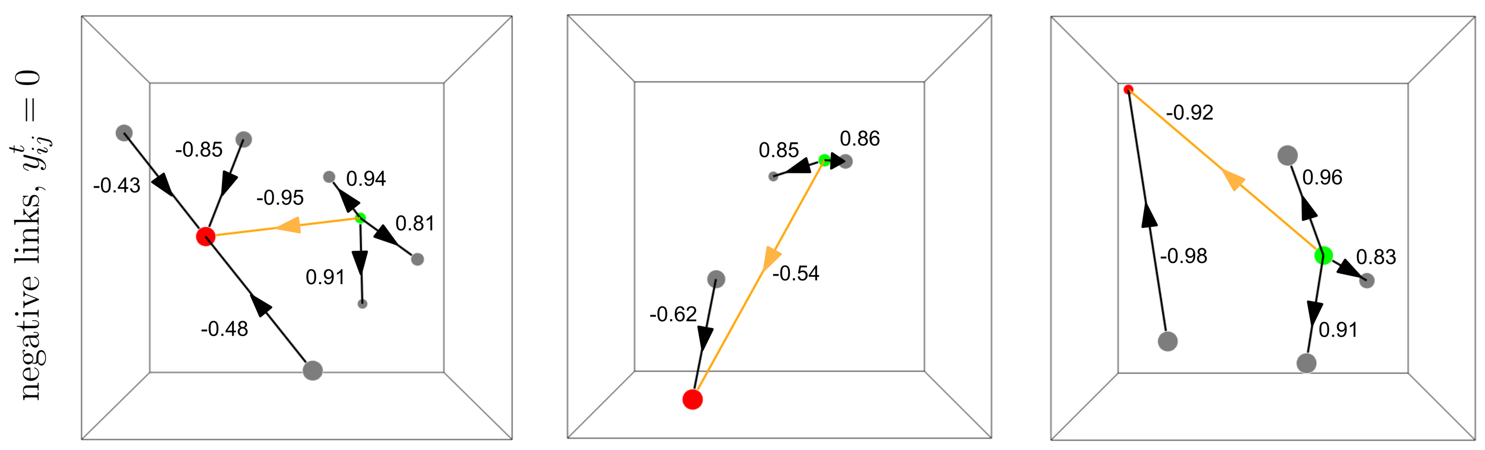
[1] tissue clearing (tc) and light-sheet microscopy imaging [2] corrosion casting (cc) and SRμCT imaging

Quantitative Link Prediction Results:

GAV achieves state-of-the-art results across all metrics and datasets:

Dataset	Model	# Params	AUC (%)	Hits@100 (%)	Hits@50 (%)
ogbl-vessel	GCN	396,289	43.53 ± 9.61	-	-
	MLP	1,037,577	47.94 ± 1.33	-	-
	Adamic-Adar	0	48.49 ± 0.00	-	-
	GraphSAGE	396,289	49.89 ± 6.78	-	-
	SAGE+JKNet	273	50.01 ± 0.07	-	-
	SGC	897	50.09 ± 0.11	-	-
	LRGA	26,577	54.15 ± 4.37	-	-
	SEAL	172,610	80.50 ± 0.21	-	-
	S3GRL (PoS$^+$)	2,382,849	80.56 ± 0.06	-	-
	SUREL+	56,353	84.96 ± 0.68	-	-
	SIEG	752,716	87.98 ± 1.00	-	-
	SEAL+EdgeConv (ours)	49,346	97.53 ± 0.32	16.09 ± 10.48	9.37 ± 6.18
	GAV (ours)	8,194	98.38 ± 0.02	34.77 ± 0.94	28.02 ± 1.58
c57-tc-vessel	SEAL	43,010	78.21	0.12	0.06
	SEAL+EdgeConv (ours)	49,346	97.23	16.71	10.39
	GAV (ours)	8,194	98.24	33.26	26.89
cd1-tc-vessel	SEAL	43,010	83.60	0.27	0.16
	SEAL+EdgeConv (ours)	49,346	97.91	17.05	11.57
	GAV (ours)	8,194	98.72	35.82	27.25
c57-cc-vessel	SEAL	43,010	83.75	0.65	0.44
	SEAL+EdgeConv (ours)	49,346	97.49	7.21	3.35
	GAV (ours)	8,194	97.99	18.90	14.58
belgium-road	SEAL	43,010	86.73	1.25	0.68
	SEAL+EdgeConv (ours)	49,346	96.98	0.55	0.55
	GAV (ours)	8,194	99.29	47.44	38.60
italy-road	SEAL	43,010	90.07	0.32	0.16
	SEAL+EdgeConv (ours)	49,346	90.24	0.26	0.17
	GAV (ours)	8,194	99.41	28.49	20.08
netherlands-road	SEAL	43,010	84.19	0.00	0.00
	SEAL+EdgeConv (ours)	49,346	96.06	3.91	2.20
	GAV (ours)	8,194	99.44	37.55	26.97
luxembourg-road	SEAL	43,010	89.79	11.39	6.15
	SEAL+EdgeConv (ours)	49,346	97.53	59.79	39.15
	GAV (ours)	8,194	99.31	85.88	76.84

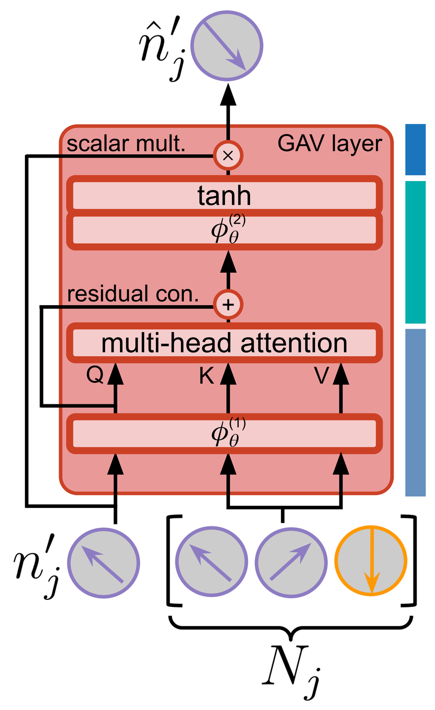
Interpretability of Results:



Visualization of updated vector embeddings representative of flow.

We find that GAV modifies vector embeddings in subgraphs extracted around **negative, implausible target links** to establish **learned sink/source flow** between target nodes, which stands in drastic contrast to the behavior of flow in spatial networks.

Message-Passing (MP) in the GAV Layer:



3) scalar multiplication with original vector embedding

$\hat{n}'_j = s_j \cdot n'_j$

2) predict scalar used to update vector embedding

$s_j = \tanh(\phi_\theta^{(2)}(\tilde{n}'_j + \phi_\theta^{(1)}(n'_j)))$

1) intermediate, neighborhood aware representation

$\tilde{n}'_j = \mathrm{MultiHeadAttn}(\phi_\theta^{(1)}(n'_j),\ \phi_\theta^{(1)}(N_j),\ \phi_\theta^{(1)}(N_j))$

Scalar multiplication via learned scalar $s_i \in (-1, 1)$ mimics (change in) direction and magnitude of flow.