Link Prediction on Multilayer Networks through Learning of Within-Layer and Across-Layer Node-Pair Structural Features and Node Embedding Similarity

Lorenzo Zangari
lorenzo.zangari@dimes.unical.it
DIMES, University of Calabria, Rende (CS), Italy

Domenico Mandaglio
d.mandaglio@dimes.unical.it
DIMES, University of Calabria, Rende (CS), Italy

Andrea Tagarelli
andrea.tagarelli@unical.it
DIMES, University of Calabria, Rende (CS), Italy

Overview

- We define **ML-Link**, a neural-network based learning framework for **link prediction on (attributed) multilayer networks.** Its purpose is to estimate the probability of edge existence within an arbitrary set of layers.
- We focus on augmenting multilayer Graph Neural Networks (GNNs) with **node-pair structural features** learned from both **within-** and **across-layer** information.
- We assess the significance of **ML-Link** on real-world and synthetic multilayer networks, conducting a comparative evaluation against MAGMA (Coscia et al., 2022), Pujari (Pujari et al., 2015), Jalili (Jalili et al., 2017), Hristova (Hristova et al., 2016), MAA (Aleta et al., 2020), MELL (Matsuno et al., 2018), CrossMNA (Chu et al., 2019), ML-GAT (Zangari et al., 2021), GATNE (Cen et al., 2019), Neo-GNN (Yun et al., 2021) and SEAL (Zhang et al., 2018).

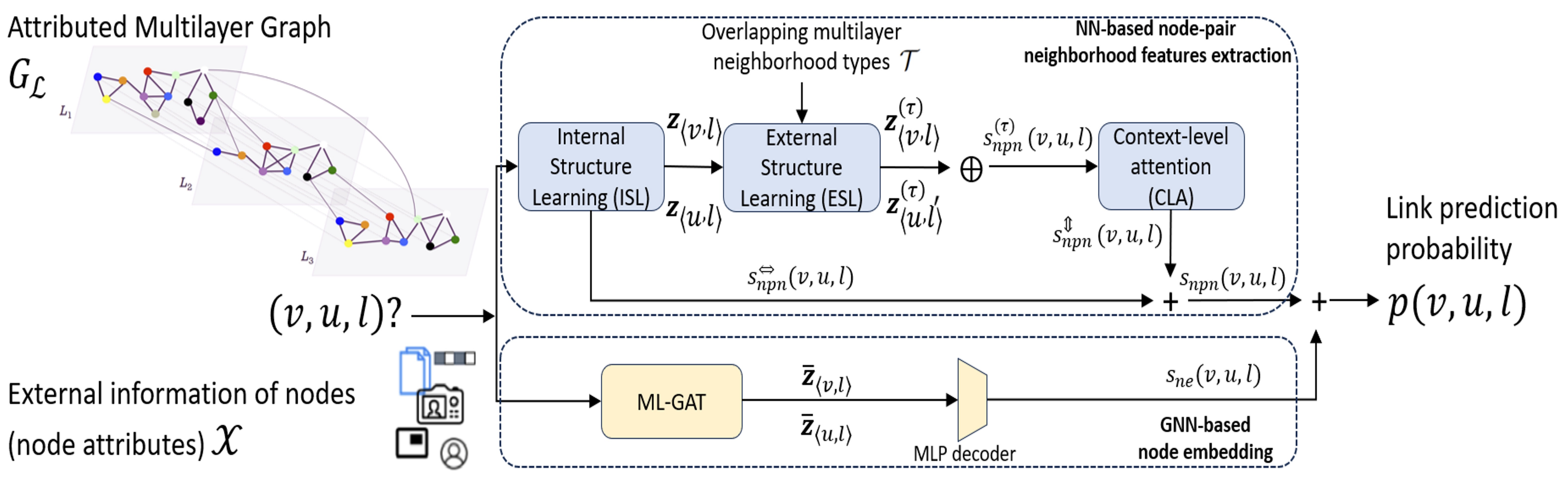
Motivation

- GNN representation learning design limits their ability to capture link-specific information, resulting in classic heuristics achieving comparable performance in link prediction tasks.
- We aim to inject node-pair-level (multilayer) structural features based on the shared neighborhoods of any two nodes into the learning process.
- Learning link structural information by considering the **overlapping neighborhood** between any pair of layers enable to generalize single-layer and multilayer link prediction heuristics.
- Exploiting different **overlapping multilayer neighbors** enable a holistic view of the multilayer neighborhood:
  - *Overlapping across-layer neighborhood* (OAN) considers the shared entitiy-neighbors across two layers.
  - *Multilayer Adamic-Adar neighborhood* (MAAN) considers triadic closure relations across two layers.

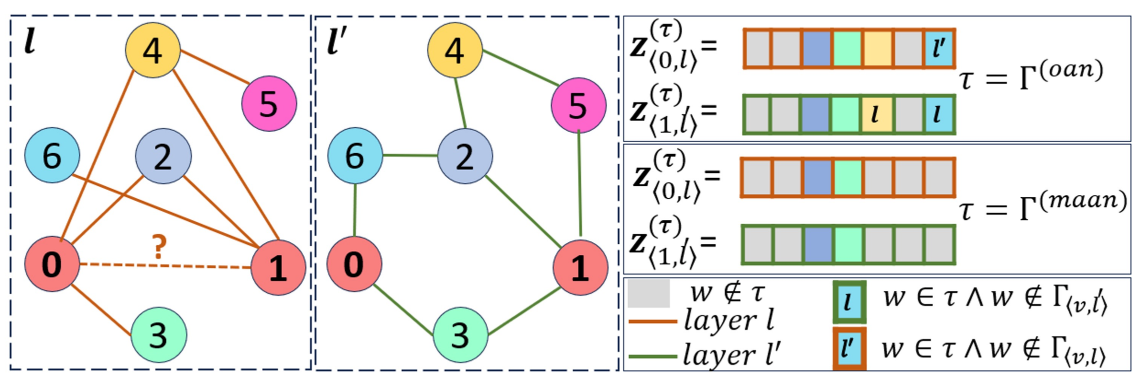
ML-Link

ML-Link is an end-to-end framework for link prediction based on two components:

1. **NN-based node pair neighborhood features extraction (NN-NPN)** learns node-pair structural information by leveraging the **ISL**, **ESL** and **CLA** modules.
2. **GNN-based node embedding (GNN-NE)** learns node representation based on external available node information through a multilayer GNN module.



- **ISL** captures within-layer topological information between the target nodes.
- **ESL** captures multilayer information between the entities related to the target nodes, through the context-aware vectors $z^{(\tau)}$, where τ is a type of overlapping multilayer neighborhood (e.g., MAAN).



- **CLA** weights the importance of each type of overlapping multilayer neighborhood through an attention mechanism.

Results

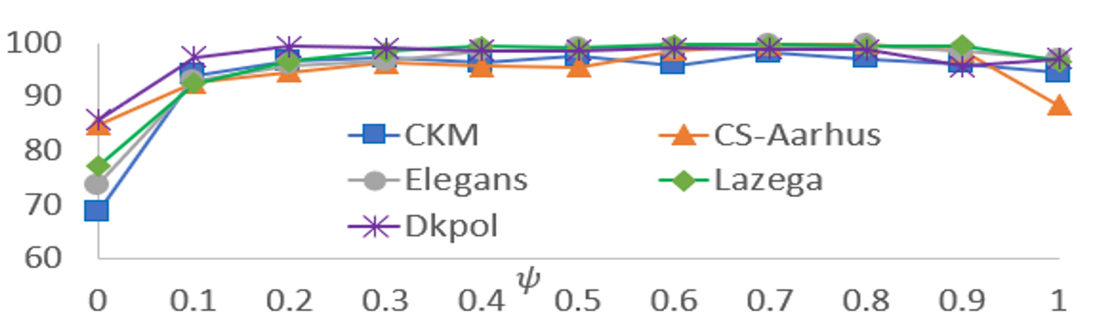
Comparative evalution: ML-Link outperforms 11 competing methods and 6 heuristics (ensemble) in terms of AUC and AP.

Method	Cs-Aarhus	CKM	Elegans	Lazega	DkPol	ArXiv
ML-Link	**97.208** **97.348**	**99.269** **99.268**	**99.646** **99.645**	**99.557** **99.579**	**99.552** **99.515**	**99.342** **99.470**
Ensemble	89.831 89.520	73.528 72.906	80.322 79.759	81.860 80.398	92.124 92.423	99.171 99.293
MAGMA	85.606 80.619	92.341 89.659	96.176 96.335	82.188 79.036	90.749 89.632	96.238 96.114
Pujari	83.218 75.559	69.225 74.774	77.017 76.763	64.564 58.747	79.241 71.735	OOT OOT
Jalili	80.717 76.270	79.730 70.188	67.987 65.248	59.801 55.223	73.408 72.701	OOT OOT
Hristova	79.766 60.176	71.803 61.44	56.198 54.097	55.054 53.626	62.586 53.295	OOT OOT
MAA	92.083 91.611	85.151 86.692	86.025 84.422	79.682 78.260	90.719 89.438	OOT OOT
MELL	73.641 77.517	68.357 77.521	82.093 88.644	64.262 70.328	45.918 48.570	OOM OOM
CrossMNA	78.589 75.457	88.317 87.859	88.389 87.203	74.54 69.68	68.371 61.268	98.318 98.426
ML-GAT	89.432 88.754	88.517 86.751	96.307 95.236	72.623 69.047	85.382 84.015	82.635 76.617
GATNE	85.096 84.459	90.033 88.445	88.389 87.203	78.352 75.231	75.579 73.317	98.914 99.187
Neo-GNN	83.370 82.986	89.094 87.591	82.793 82.405	78.956 78.428	81.084 81.983	92.176 93.847
SEAL	81.986 82.316	83.898 83.651	87.979 86.517	81.429 80.140	95.004 94.684	98.823 98.816

Ablation analysis: all architectural components of ML-Link are effective.

Method	Cs-Aarhus	CKM	Elegans	Lazega	DkPol
GNN-NE	89.432 88.754	88.517 86.751	96.307 95.236	72.603 69.047	85.382 84.015
ISL	84.622 84.848	62.450 68.512	75.158 73.560	78.905 77.035	85.924 85.699
ISL w/ GNN-NE	91.16 91.07	89.341 88.645	96.253 95.377	80.577 79.867	90.398 92.031
ISL w/ ESL ($\Gamma^{(oan)}$)	90.410 89.946	72.301 76.821	82.172 79.523	78.294 77.455	85.531 82.478
ISL w/ ESL ($\Gamma^{(maan)}$)	89.175 88.769	70.723 75.393	79.369 77.826	80.527 78.456	87.915 86.236
ISL w/ ESL	90.284 89.566	73.191 77.709	82.693 80.294	81.254 79.200	87.521 86.228
NN-NPN	95.927 95.481	98.561 97.576	98.828 98.865	99.023 99.104	98.036 98.560
ML-Link	**97.208** **97.348**	**99.269** **99.268**	**99.646** **99.645**	**99.557** **99.579**	**99.552** **99.515**

Sensitivity analysis: ML-Link is robust w.r.t. its main hyper-parameter ψ, which weights the importance of ISL vs ESL and CLA modules.



Efficiency analysis: ML-Link tends to be faster than the strongest competing method on two sets of syntethic networks.

	$\beta = 0.1$			$\beta = 0.5$		
	ML-Link		MAGMA	ML-Link		MAGMA
$\lvert V_{\mathcal{L}} \rvert$	GPU	CPU	CPU	GPU	CPU	CPU
1500	0.013	0.095	0.027	0.023	0.167	0.043
3000	0.020	0.403	0.154	0.034	0.718	0.453
4500	0.030	1.193	0.670	0.059	2.271	1.595
6000	0.060	3.792	5.173	0.120	7.280	5.636
7500	0.102	7.938	8.489	0.211	15.589	12.478
9000	0.179	17.627	20.339	0.397	33.054	29.131
10500	0.339	31.074	34.781	0.740	60.217	54.932
12000	0.641	62.395	77.394	1.322	117.877	148.338
13500	1.366	88.707	127.924	2.659	184.375	235.066

Scan the QR code to access the ML-Link GitHub repository

UNIVERSITÀ DELLA CALABRIA
DIPARTIMENTO DI INGEGNERIA INFORMATICA, MODELLISTICA, ELETTRONICA E SISTEMISTICA
DIMES
AI & Data Science Lab
Future Artificial Intelligence Research