NetSciX'24

Deep Community Detection with Spatio-Temporal Graph Neural Networks

Nelson Reis[1,2] nelson.reis@phd.unipi.it

Emanuele Carlini[1] emanuele.carlini@isti.cnr.it

Salvatore Trani[1] salvatore.trani@isti.cnr.it

[1] *ISTI-CNR, Pisa, Italy* [2] *University of Pisa, Italy*

Context

Identifying mesoscale patterns in large graphs allows the study of increasingly complex networks, further the understanding of their dynamics. However, even though the task of community detection has been a thriving research topic for decades, most of the introduced methods focuses on the issue of static graphs, while most real-world networks are instead **temporally evolving**.

The task of detecting mesoscale structures on dynamic attributed graphs (**Fig. 1**) is therefore still an underexplored and compelling topic of research, which offers a promising opportunity to improve on the state of the art of community detection. Our goal is to tackle it by introducing a novel task-oriented **Graph Neural Network** (**GNN**) for clustering spatio-temporal signals.

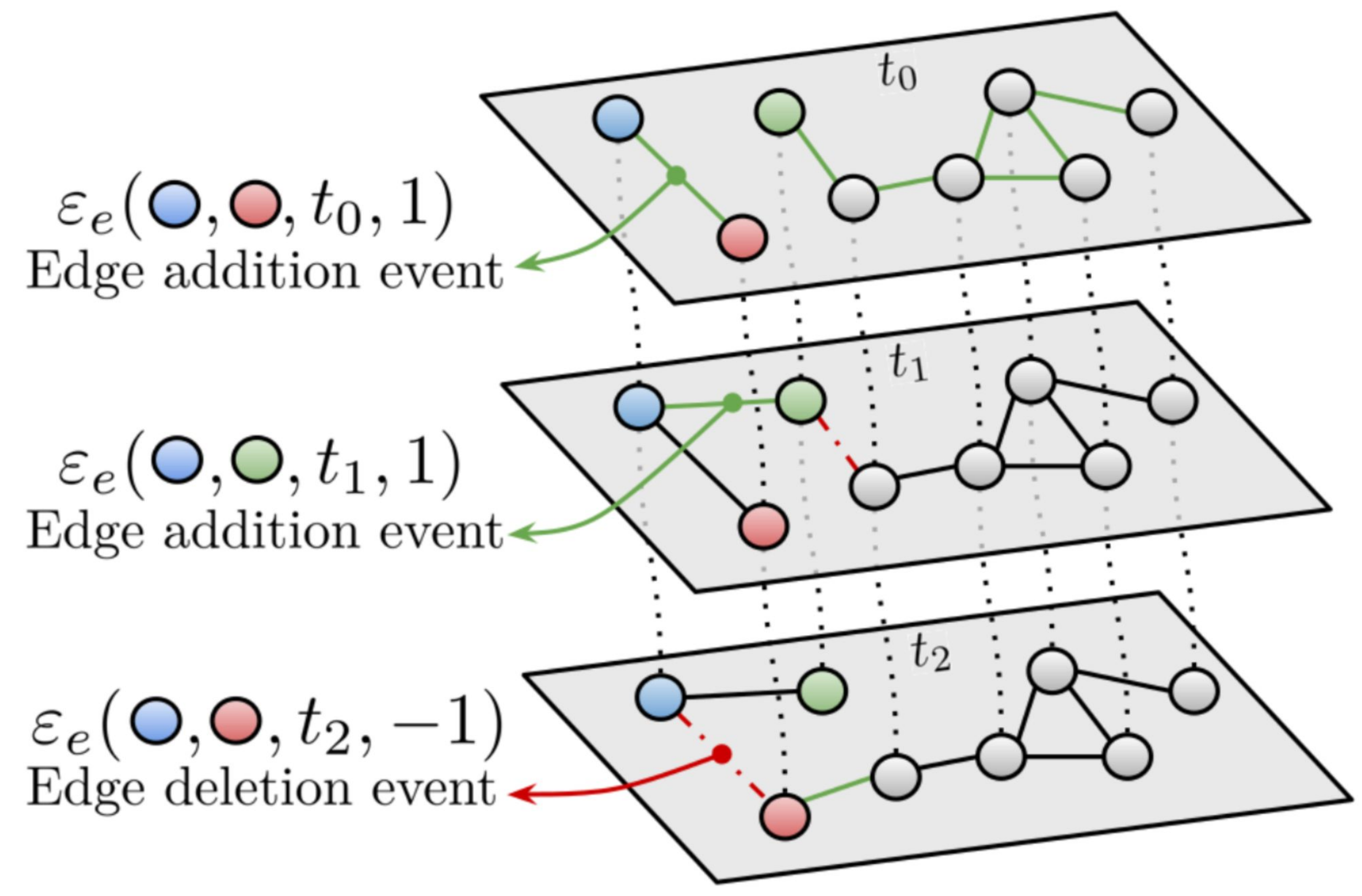

Fig. 1. A temporal graph depicted in three snapshots. Dotted lines represent edge couplings among temporal nodes.

Learning on graphs

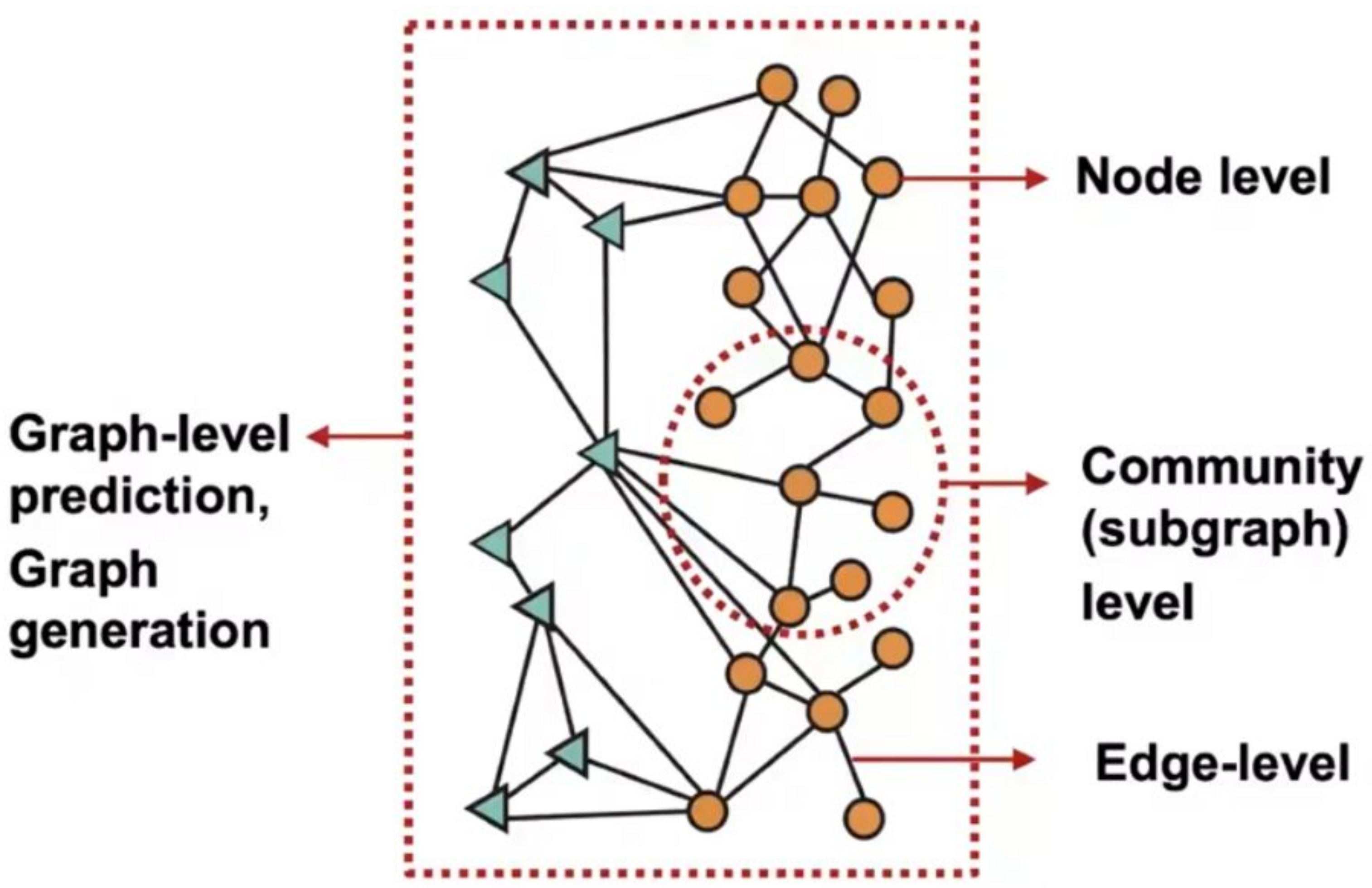

Fig. 2. GNN task levels. Leskovec et al., 2021.

GNNs are models that perform feature representation learning on data represented in non-euclidean domains, such as network graphs. They rely on a message-passing paradigm with permutation invariant or permutation equivariant functions to aggregate the neighborhood' information of nodes and/or edges.

Their main strength lies in considering both the **topological structure** of a graph and the **attribute features** of its high-order (*n*-hop) neighbors, which purported them to the state of the art in many node-, edge-, and (sub)graph-level prediction tasks (**Fig. 2**).

Meanwhile, (Spatio-)Temporal GNNs integrate mechanisms such as **Gated Recorrent Units** and **Long-Short Term Memory**, extending such a strength to learning on dynamic graph signals.

Data augmentation

Our preliminary tests revealed that simply **introducing the temporal dimension through data augmentation techniques on a structural level** slightly improved prediction on an existing time-agnostic GNN model trained on the PubMed dataset.

	ACC	NMI	ARI	F1
PubMed	0.577	0.190	0.169	0.577
PubMedTemporal (t=2, b=edge)	0.584	0.194	0.175	0.585
PubMedTemporal (t=2, b=node)	0.584	0.195	0.175	0.585
PubMedTemporal (t=3, b=edge)	0.591	0.198	0.178	0.593
PubMedTemporal (t=3, b=node)	0.591	0.198	0.178	0.593
PubMedTemporal (t=5, b=node)	0.592	0.204	0.184	0.595
PubMedTemporal (t=5, b=edge)	0.592	0.204	0.184	0.595
PubMedTemporal (t=4, b=node)	0.593	0.203	0.184	0.595
PubMedTemporal (t=4, b=edge)	0.593	0.203	0.184	0.595
PubMedTemporal (t=6, b=edge)	0.598	0.206	0.186	0.601
PubMedTemporal (t=6, b=node)	**0.598**	**0.206**	**0.186**	**0.601**

Tab. 1. Results obtained from an average of 20 runs using a 2-layer GAT (Veličković et al., 2018) with a self-supervised DAEGC (Wang et al., 2019) autoencoder, with one epoch for pretrain and 500 epochs for train.

We first converted the originally static network composed of 19 171 nodes (papers) and 44 335 edges (citations) to distinct **snapshot-based** temporal graphs, balancing (***b***) the number of nodes or edges in each time slice (***t***). Afterwards, the graph was flattened and edge couplings connecting temporal nodes were added, resulting in a a unified temporal graph that encodes the temporal information in the topology of a statically defined graph.

In spite of the fixed attribute feature sets of nodes over time, the results (**Tab. 1**) show **minor but consistent improvements** in the metrics used for evaluating the unified temporal graph configurations, with the same seeds and set of hyperparameters.

Our next steps are in the direction of implementing and experimenting with distinct model architectures for **learning on dynamic graphs**, as well as comparing against time-agnostic models for clustering with static and augmented graph data sets.

Consiglio Nazionale delle Ricerche

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"

HPC lab

Università di Pisa

Finanziato dall'Unione europea NextGenerationEU