

Node Feature Extraction by Self-Supervised Multi-scale Neighborhood Prediction

Eli Chien¹ Wei-Cheng Chang² Cho-Jui Hsieh³ Hsiang-Fu Yu²
Jiong Zhang² Olgica Milenkovic¹ Inderjit S. Dhillon²

University of Illinois Urbana-Champaign, Illinois, USA¹

Amazon Search, USA²

University of California, Los Angeles, USA³



ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

amazon | science

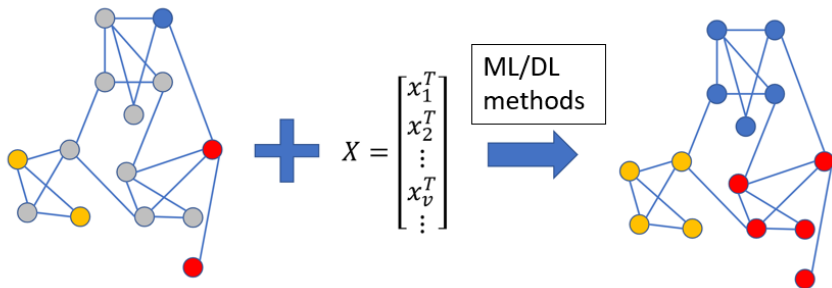
UCLA

ICLR 2022

<https://github.com/amzn/pecos/tree/mainline/examples/giant-xrt>

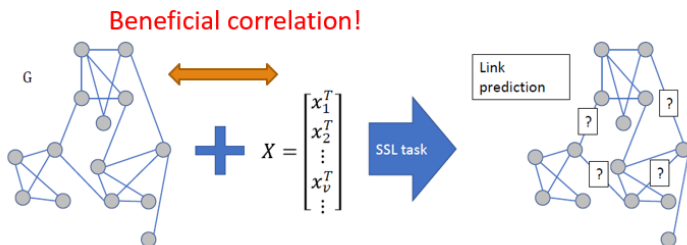
Machine learning on graphs

- Machine learning on graphs, **especially graph neural networks (GNNs)**, has achieved great success recently.
- Although there are various tasks (i.e. graph classification), we focus on **node classification** in this work.
- Standard GNN pipeline assumes the input is a graph paired with **numerical** node features.



Self-supervised learning (SSL) in GNNs

- Training large scale GNNs usually requires numerous task-specific labeled data, which can be expensive in practice.
- Recently, researchers proposed to alleviate this issue with SLL.
- The idea is to design SSL tasks using only X and G **without task-specific labels**.
- By solving SSL tasks, we hope to either get a more informative node embedding or a well-initialized GNN.
- **Key intuition: X and G are both correlated to downstream tasks.**

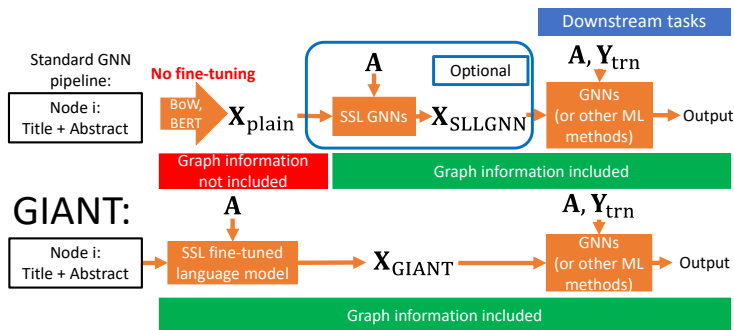


Issue of standard GNN pipeline

- A critical yet neglected problem: How to obtain the numerical node features from raw data, such as text, images and audio signals?
- Standard approach is to use graph agnostic methods: Bag-of-Words (BoW), pretrained BERT ... etc.
- Clearly, using graph agnostic methods is suboptimal, as the beneficial correlation between graph topology and raw features is ignored.

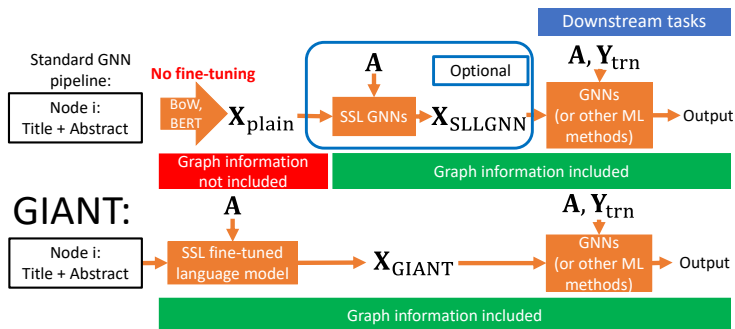
GIANT framework

- We propose Graph Information Aided Node feature Extraction (GIANT) framework to resolve the aforementioned issue.



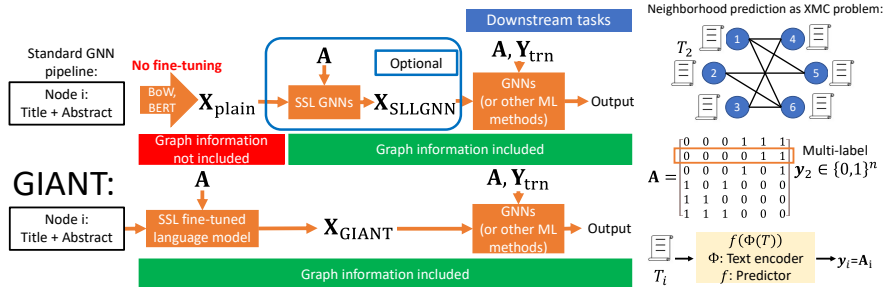
GIANT framework

- Raw feature: raw text.
- We design a new SSL task, **neighborhood prediction**, which can work for both **homophily** and **heterophily** cases.



GIANT framework with XR-Transformer

- We show that **neighborhood prediction** is an **eXtreme Multilabel Classification (XMC)** problem¹.
- Thus, we can leverage the state-of-the-art method, **XR-Transformer (XRT)**², to solve it.

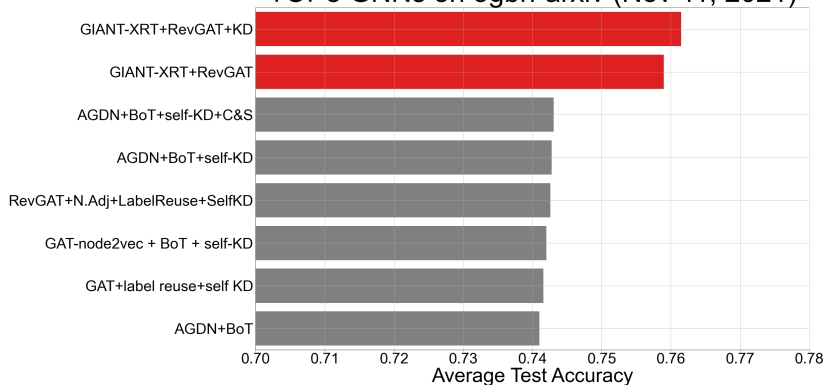


¹Taming pre-trained transformers for extreme multi-label text classification. Chang et al. KDD 2020.

²Fast multi-resolution transformer fine-tuning for extreme multi-label text classification. Zhang et al. NeurIPS 2021.

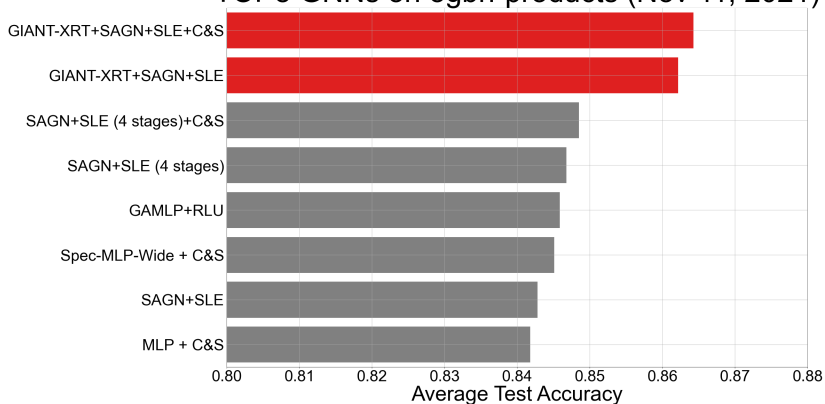
Results on OGB-arxiv

TOP8 GNNs on ogbn-arxiv (Nov 11, 2021)



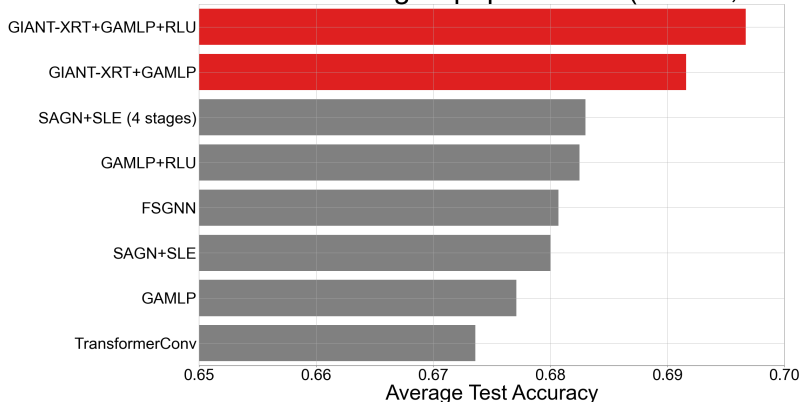
Results on OGB-products

TOP8 GNNs on ogbn-products (Nov 11, 2021)



Results on OGB-papers100M

TOP8 GNNs on ogbn-papers100M (Nov 11, 2021)



More results

- We have showed that standard **Link Prediction** methodology is not a suitable SSL task for **heterophily graphs**, while our **neighborhood prediction** is.
- By leveraging the graph learning analysis, we are able to prove the benefit of components in XRT.
- We conduct extensive experiments and ablation study to show the effectiveness of GIANT + XRT.
- Please check our paper and poster for more details.

Thanks for your attention!

Please also check our work on hypergraph neural networks³!

³You are AllSet: A Multiset Learning Framework for Hypergraph Neural Networks,
Chien et al. ICLR 2022