

OGB-LSC Technical Report - Mycelia team

Chris Hawkes, Fengtong Xiao, and Mehdi Aghagolzadeh

AI Lab, The Trade Desk

November 2022

1 Introduction

2 Methods

2.1 Pre Processing

Input features for rGAT model is the concatenation of those:

- whitening: reduce 768 RoBERTa emb to 128 dimensions
- metapath2vec learned embedding
- metapath feature aggregation: author institute averaging from neighbor paper; assigning year info for author by getting the most frequent year from its connected paper neighbors; paper number of citations, number of authors
- PrevLabelDistributionEmb: input the previous year's label distribution as embedding and
- YearPosEmb: generate year positional embeddings
- node in and out degree information
- paper features from encoder based on GRACE contrastive loss
- paper features encoder based on BGRL contrastive loss

based on correlation plot on label distribution between different years, we observed there is a dramatic label distribution shift in 2008, so we only train on training samples that has year \neq 2009

2.1.1 Encoders Structure

RGAT + MaskLabelPred for unified propagation +

BGRL & Grace + shuffling for regularization, trained as above but with extra self-supervised objective added to the contrastive loss

2.2 Model Structure

$$MLP(roberta, meta2vec, degree, H1, H2, H3, metapaths)$$

Generate multiple embeddings for final finetune: fine tune structure: concat(non-metapath emb, max/min/sum_pooling(metapath_emb)) -> 2 layer NN

- paper degree
- metapath2vec learned embedding
- rGAT with BGRL on 80% validation
- rGAT with (Grace on full valid data)
- rGAT with (Grace on 80% of valid data + train data) x 2 (2 different random 80%)
- More embeddings from metapath-based random walk label feature generations (similar to 4th paper solution, but with 2 more metapaths: most recent 7 citations, most cited (papers that have been cited the most))

run 10 models, randomly dropout 10% on input features, randomly picking 8 metapath-based embeddings, bagging the results

3 Results and Conclusion

Train	Val	Test
a	b	?
a	b	?