OVERSAMPLING IN HETEROGENEOUS GRAPHS USING SMOTE

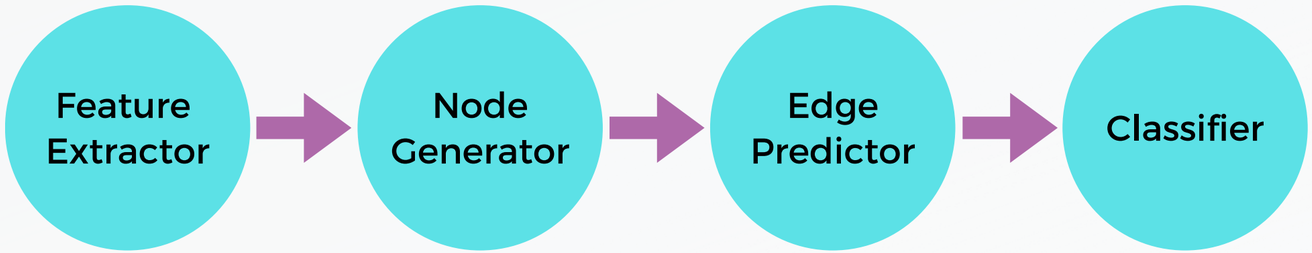
PROJECT PROPOSAL BY: ADHILSHA AND DEEPENDRA

> Objectives

- To address the challenge of imbalanced data in heterogeneous graph (HG) learning to improve the performance of downstream tasks.
- To apply SMOTE on HGs using a novel oversampling strategy.
- To develop a generalized domain independent oversampling technique.

homogeneous

heterogeneous



Related Papers

- HetGNN: Zhang et al. "Heterogeneous graph neural network." SIGKDD 2019.}
- MagNN: Fu et al. "Magnn: Metapath aggregated GNN for heterogeneous graph embedding." The Web Conf. 2020.}
- HAN: Wang et al. "Heterogeneous graph attention network." WWW Conf. 2019.}
- Review: Shi. "Heterogeneous Graph Neural Networks." GNN Book, 2022.}

Datasets

- **Bibliographic-Based:** DBLP, ACM, and AMiner.
- **Business/Review-Based:** Amazon, Yelp Challenge.
- **Movie-Based:** IMDB

Expected Results

We expect the implementation of the specialized oversampling strategy to be completed by Midway for us to continue the generalization part from beyond that.

Baselines to Implement

- The selected baseline model will serve as the foundation for our experiments. We will evaluate its performance on a chosen dataset and apply our specialized oversampling strategy to it.
- **Meta Path-Based Baseline Models:** HAN, MAGNN and other relevant models.
- **Deep-Based Baseline Models:** HetGNN.

Work Division

- **Adhil:** Literature Review, documentation, data preprocessing, edge generator and downstream task.
- **Deependra:** Literature Review, documentation, baseline model experimentation, Node generation,

Midway Plans

- Extensive Literature review on current state-of-the-art work on deep models for HGs.
- Shortlisting a specific dataset and generating its imbalanced subset.
- Shortlisting baseline feature extractor deep model and implementing it on the dataset.
- Deciding on the oversampling strategy and implementing it for the "specific" model and dataset.