

Project Proposal: Implementing Graph Attention Networks

Valle, Ruy & Soni, Kushagra
{rrv4, soni14}@illinois.edu

Group ID: 179

Paper ID: 7

Presentation link: [TBU](#)

Code link: https://github.com/kushagrasoni/CS598_DLH_GAT_Implementation

1 Introduction

Graph Attention Networks (GATs) [1] are a type of neural network architecture designed for processing graph-structured data. They aim to overcome the limitations of previous architectures by being able to operate on arbitrarily structured graphs in a parallelizable manner. GATs use a masked shared self-attention mechanism to assign weights to each node's neighbors and combine their features, resulting in a new set of features for the node in question. GATs have shown promising results in various graph-based tasks, including node classification and link prediction.

2 Background and Problem Statement

Before Graph Attention Networks (GATs), several neural network architectures were proposed to process graph-structured data, including Graph Convolutional Networks (GCNs), GraphSAGE, and DeepWalk. However, these architectures had one or more of the following limitations:

- Inability to operate on arbitrarily structured graphs: GCNs, for example, are limited to processing homogeneous graphs where all nodes have the same features.
- Need to sample from input graphs: some architectures, like GraphSAGE, require sampling from input graphs to create mini-batches, which can lead to information loss.
- Learning separate weight matrices for different node degrees: GCNs use different weight matrices for nodes with different degrees, making the model less scalable.
- Inability to parallelize across nodes: traditional GCNs perform summation operations over a node's neighbors, which cannot be parallelized.

Overall, prior architectures faced difficulties in handling the complexity and heterogeneity of real-world graphs, which motivated the development of Graph Attention Networks.

GATs aim to be able to operate on arbitrarily structured graphs in a manner that is parallelizable across nodes in the graph, thus having none of the limitations mentioned previously. GATs use a masked shared self-attention mechanism. The mask ensures that, for a given node, only features from first-degree neighbours are taken into consideration. The self-attention mechanism allows the model to assign arbitrary weights to each of a given node's neighbours, which then allows the neighbours' features to be combined, which results in a new set of features for the node in question. The paper only mentions using the resulting features for classification tasks, but it should also be possible to use these features for regression tasks.

3 Objectives

Our goal is to reproduce the results reported in the original paper, which used the Cora, Citeseer, Pubmed, and Protein-protein interaction datasets. We will use the same architectural details as in the paper, aiming for

- Classification accuracies of approximately 83.0%, 72.5%, and 79% in the Cora, Citeseer, and Pubmed datasets, respectively,
- Micro-averaged F1 score of approximately 0.973 in the protein-protein interaction dataset.

In all four cases, the results found were as good or better as the results found in previous studies. We aim to test the hypothesis that we will also get better results than previous studies and the hypothesis will be similar to those in the paper. Furthermore, we aim to conduct a few ablation studies and architectural variations, such as:

- Replacing the self-attention mechanism with other metrics, such as constants, random weights, and Pearson and Spearman correlation coefficients.
- Replacing the shared weight matrix W with the encoding phase of an autoencoder, by first training an autoencoder on the data.

4 Methodology

We will use the publicly available data from the original paper and Google Colab for training and testing, which should provide sufficient computational resources given the relatively small dataset sizes (the largest being 15.5 MB when compressed).

To achieve our goals, we plan to modify the official GAT implementation [2] to suit our needs and make it possible to run the model variants described above.

Through this project, we hope to deepen our understanding of GATs and contribute to the ongoing research in this field.

5 Expected Results and Outcomes

We hypothesize that our results will be as good or better than those found in previous studies.

6 Limitations and Challenges

We would like to mention that most of the datasets used are quite small and that computational resources should not be an issue. We would also identify any potential limitations or challenges that may arise during the project

References

- [1] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations (ICLR)*, 2018.
- [2] Guillem Cucurull Petar Veličković. Gat. <https://github.com/PetarV-/GAT>, 2018.