Graph Machine Learning Project

Kaizer Rahaman — Sudheesh Kumar Potla

November 15, 2023

Abstract

Graph Representation Learning is among the latest topics of interest in Graph Machine Learning. Delving into unsupervised graph representation learning methods to capture informative node embeddings without relying on labeled data is challenging. There have been several significant associated baselines. However in our work we propose the use of Variational Auto-encoders(VAEs) for this objective. There has been less exploration about the use of VAE-based GNNs for generative tasks and unsupervised representation learning on graphs. We observe VAEs can help model the underlying graph structure more effectively. We are the first to use VAEs in the context of unsupervised learning for the task of Node Classification in literature.

1 Introduction

Graph Representation Learning is a field of machine learning that focuses on the development of techniques to learn meaningful representations of graph-structured data. Graphs are mathematical structures that consist of nodes and edges, where nodes represent entities, and edges represent relationships or connections between entities. Examples of graph-structured data include social networks, citation networks, biological networks, and knowledge graphs.

The goal of Graph Representation Learning is to encode nodes and edges in a way that captures the underlying structure and patterns within the graph. Traditional machine learning methods often struggle with graph-structured data because they typically assume that each data point is independent, whereas graphs inherently involve inter-dependencies and relational information.

We are interested to learn graph representations in an unsupervised manner. This is proposed keeping in mind the availability of limited and expensive labeled data in the real world. We point out the scarcity of using Variational Graph Auto-Encoders for this purpose. Although they have been used for the purpose of link predictions, VGAEs have not yet been proposed to solve the problem of Node Classification.

2 Related Work

2.1 Graph Representation Learning

Several conventional approaches to unsupervised graph representation learning also utilize the contrastive paradigm. Research in the field of unsupervised graph representation learning has primarily concentrated on the exploration of local contrastive patterns. This approach entails enforcing the similarity of embeddings for neighboring nodes. Nodes that appear in the same random walk are considered positive samples in this particular case. An illustration of this may be seen in the groundbreaking research conducted by DeepWalk, where the probabilities of node co-occurrence pairings are modeled through the utilization of noise-contrastive estimation. The effectiveness of random-walk-based methods has been demonstrated in their ability to factorize certain types of graph proximity, such as the transformation of the adjacent matrix. However, these methods tend to excessively prioritize the structural information contained in these graph proximities, and they also encounter significant challenges when applied to large-scale datasets. Moreover, these algorithms have been identified as prone to errors due to inadequate hyper-parameter tuning. In this study, the researchers implemented the **GRACE Algorithm** [ZXY+20] and achieved notable advancements in the task of Node Classification.

3 Deep Graph Contrastive Representation Learning with VAE

3.1 Ground Work

In unsupervised graph representation learning let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ denotes a unweighted and undirected graph, where $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represent the node set and edge set respectively. $\mathbf{X} \in \mathbb{R}^{\mathcal{N} \times \mathcal{F}}$ and $\mathbf{A} \in \{0,1\}^{\mathcal{N} \times \mathcal{N}}$ represent the feature matrix and adjacency matrix respectively, where $x_i \in \mathbb{R}^{\mathcal{F}}$ is the feature of v_i, and $\mathbf{A}_{ij} = 1$ iff $(v_i, v_j) \in \mathcal{E}$, (we assume diagonal elements set to 1, i.e. every node is connected to itself). There will be no given class information or labelled data of nodes in $\mathcal{G}$ during training. Main objective is to learn a Variational Graph Auto-Encoder: $f(\mathbf{X},\mathbf{A}) \in \mathbb{R}^{\mathcal{N} \times \mathcal{F}'}$ receiving the graph features and structure as input, that produces node embeddings in low dimensionality, i.e., $\mathcal{F}' \ll \mathcal{F}$. We denote $\mathbf{H} = f(\mathbf{X},\mathbf{A})$ as the learned representation of nodes, where is h_i is the embedding of node v_i.These representations can be used in node classification.

3.2 Our Proposed Algorithm

In contrast to prior research that focuses on learning representations through the utilization of local-global links, our model integrates the Variational Graph Auto-Encoder and GRACE algorithm.In the GRACE framework, the process of learning embeddings involves the direct maximization of agreement at the node level between the embeddings.

In our model, at each iteration we generate two graph views from the original graph G, denoted as G_1, G_2, and denote node embeddings in the two generated views as $U = f(\tilde{X_1}, \tilde{A_1})$ and $V = f(\tilde{X_2}, \tilde{A_2})$. Considering contrastive approaches that rely on contrasting between node embeddings in different views, we perform two methods for graph corruption, removing edges for topology and masking features for node attributes.

We employ a contrastive objective (i.e., a discriminator) that distinguishes the embeddings of the same node in these two different views from other node embeddings For any node v_i, its embedding generated in one view, u_i, is treated as the anchor, the embedding of it generated in the other view, v_i, forms the positive sample, and embeddings of nodes other than v_i in the two views are naturally regarded as negative samples. Formally, we define the critic $\theta(u, v) = s(g(u), g(v))$, where s is the cosine similarity and g is a non-linear projection to enhance the expression power of the critic. The projection g is implemented with a two-layer multilayer perceptron (MLP).We define the pairwise objective for each positive pair (ui, vi) as:

$$\ell(\boldsymbol{u}_i, \boldsymbol{v}_i) = \log \frac{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}{\underbrace{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}_{\text{the positive pair}} + \underbrace{\sum_{k=1}^{N} \mathbb{1}_{[k \neq i]} e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_k)/\tau}}_{\text{inter-view negative pairs}} + \underbrace{\sum_{k=1}^{N} \mathbb{1}_{[k \neq i]} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_k)/\tau}}_{\text{intra-view negative pairs}}},$$

where $1_{[k \neq i]} \in \{0, 1\}$ is an indication function that equals to 1 iff $k \neq i$ In the decoder, we are performing the inner dot product Inner Dot Product $\left\langle {F_{1_0}}' \cdot {F_{1_1}}' \right\rangle$ and $\left\langle {F_{2_0}}' \cdot {F_{2_1}}' \right\rangle$. Calculates the inner product for all pairs of nodes in the graph, resulting in an adjacency matrix.

$$\mathcal{J}_{\text{contrastive}} = \frac{1}{2N} \sum_{i=1}^{N} l\left(u_i, v_i\right) + l\left(v_i, u_i\right)$$

$$\begin{aligned} \mathcal{J}_{\text{recon}}(z, \text{pos\_edge\_index}, \text{neg\_edge\_index}) = & -\frac{1}{|E_{\text{pos}}|} \sum_{(i,j) \in \text{pos\_edge\_index}} \log\left(\sigma\left(\mathbf{z}_i^T \cdot \mathbf{z}_j\right)\right) \\ & -\frac{1}{|E_{\text{neg}}|} \sum_{(i,j) \in \text{neg\_edge\_index}} \log\left(1 - \sigma\left(\mathbf{z}_i^T \cdot \mathbf{z}_j\right)\right) \end{aligned}$$

$$\text{Loss Function} \quad \mathcal{J}_{\text{total}} = \sum \left(\mathcal{J}_{\text{contrastive}}, \mathcal{J}_{\text{recon1}}, \mathcal{J}_{\text{recon2}}\right)$$

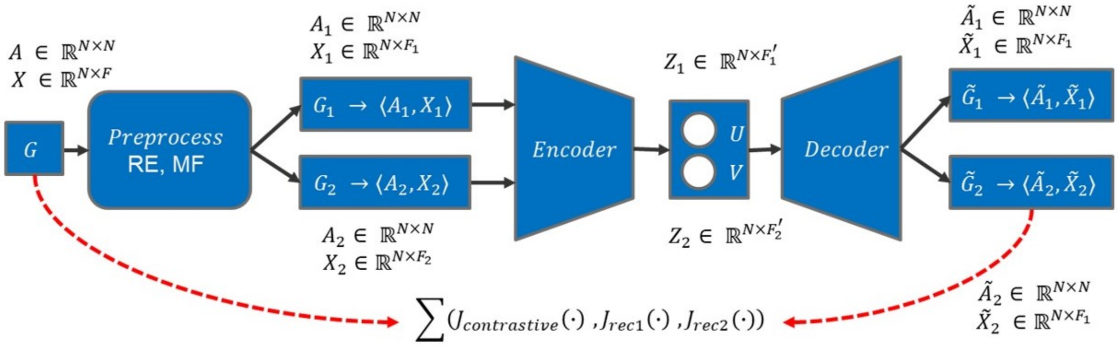


Figure 1: Our Model Architecture

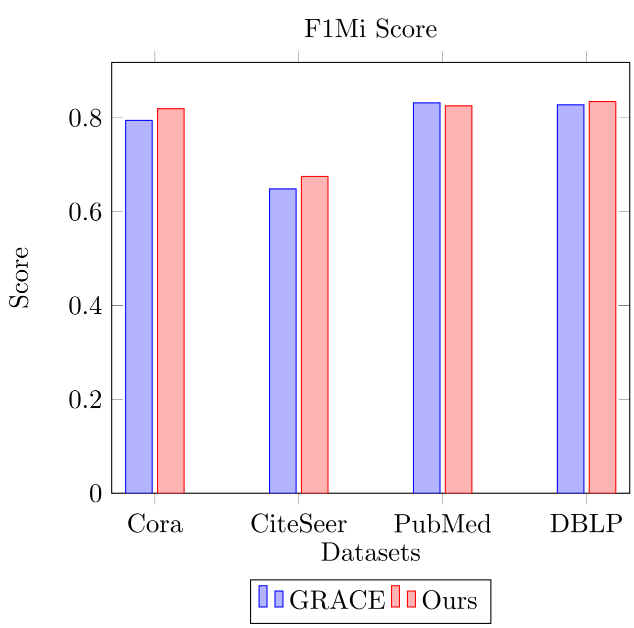
4 Results

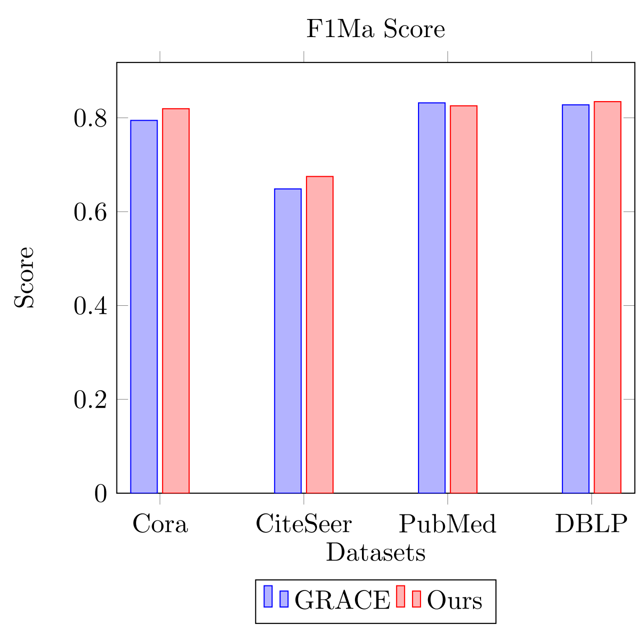
The F1Mi Score Comparison:

Dataset	Grace	Ours
Cora	0.7945	0.8194
CiteSeer	0.6485	0.6750
PubMed	0.8319	0.8256
DBLP	0.8277	0.8345

The F1Ma Score Comparison:

Dataset	Grace	Ours
Cora	0.7819	0.8039
CiteSeer	0.6108	0.6288
PubMed	0.8282	0.8188
DBLP	0.7814	0.7894





References

[ZXY+20] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep graph contrastive representation learning. *arXiv preprint arXiv:2006.04131*, 2020.