International Journal of Data Warehousing and Mining
Volume 21 • Issue 1 • January-December 2025

A Heterogeneous Network Text Attribute Fusion Method Based on Multi-Level Semantic Relation Contrastive Learning

Wei Zhang
https://orcid.org/0009-0000-9306-8116
Qinghai Normal University, China

Zhonglin Ye
https://orcid.org/0000-0002-2429-3325
Qinghai Normal University, China

ABSTRACT

Contrastive learning enables models to learn graph structural information through self-supervised learning in the absence of labels. However, real-world networks often contain both graph structural information and incomplete node attribute information. Based on this, this paper proposes a heterogeneous network text attribute fusion method based on multi-layer semantic relation contrastive learning. Firstly, the heterogeneous network components are reconstructed using semantic and thematic attribute acquisition methods at different levels, obtaining semantic representations of text attributes at various levels of abstraction. Then, the contrastive learning component of the heterogeneous network is employed to maximize the correlation between different views of the heterogeneous network, allowing the two heterogeneous networks to align in this space. This alignment helps to uncover the latent connections between text attribute features across different views, thereby achieving the fusion of information between views.

KEYWORDS

Network Representation Learning, Knowledge Modeling, Comparative Learning, Node Attribute, Joint Learning

INTRODUCTION

Complex networks serve as abstract models for understanding complex systems, where the formation of communities through node and edge interconnections reflects real-world community structures that reciprocally shape network connectivity. Heterogeneous networks, involving different types of nodes and edges, require consideration of the complex relationships between these types, which makes their analysis and study in graph analysis and mining more intricate. In the real world, most networks are heterogeneous, such as social networks (Hamilton et al., 2017; Wang et al., 2016), citation networks (Atwood et al., 2016; Kipf et al., 2016), and traffic networks (Liang et al., 2024; Zhang et al., 2018), where the textual data and community structure information of the nodes contain rich knowledge and relationships. Therefore, effectively leveraging this information-rich heterogeneous data, which contains text and community attributes, has become a hotspot in current research on network representation learning.

DOI: 10.4018/IJDWM.378680

Although graph neural networks (GNNs) were initially designed for tasks in homogeneous networks, such as graph classification (Duvenaud et al., 2015; Lee et al., 2019; Ying et al., 2018), link prediction (Schlichtkrull et al., 2018; Vercheval et al., 2020; Zhang et al., 2018), and node classification (Kipf et al., 2016; Wang et al., 2019), some outstanding GNN methods in recent years have been able to handle heterogeneous networks containing multiple types of nodes and edges. There are some methods to encode the structural information, such as the heterogeneous attention network (HAN; Wang et al., 2019), the metapath aggregated GNN for heterogeneous graph embedding (Fu et al., 2020), community attribute variational autoencoders (Shi et al., 2019), and network representation learning with community awareness and its applications in brain networks (Shi et al., 2022). These algorithms demonstrate that incorporating heterogeneous networks and community information into neural networks is beneficial for learning more features in complex networks. There are also methods that combine node structure and attributes, such as multimodal deep network embedding with integrated structure and attribute information (Zheng et al., 2020).

In network representation learning, a contrastive learning approach is employed, where the core idea is to maximize the similarity between positive samples (similar samples) and minimize the similarity between negative samples (dissimilar samples). By designing self-supervised tasks, contrastive learning enables models to learn meaningful representations, such as rich structural and semantic information from data in the absence of labels, achieving success in various fields, such as image processing and natural language processing.

Therefore, when effectively integrating multiple heterogeneous information sources from nonlinear relationships and obtaining a large amount of labeled data, contrastive learning does not rely on task-specific labels, it becomes more adaptable. This paper proposes a novel text attribute fusion method, namely the method of text attribute fusion in heterogeneous networks under deep semantic contrastive learning. Specifically, this is a joint learning approach consisting of two fundamental components: a reconstructing heterogeneous network component and a heterogeneous network contrastive learning component.

The reconstructing heterogeneous network component employs different semantic acquisition methods at various levels to reconstruct the heterogeneous network, aiming to obtain semantic representations at different abstraction levels in text attributes. The contrastive learning component of the heterogeneous network aligns the two heterogeneous networks in the representation space by maximizing the correlation between heterogeneous network views. This allows for the exploration of potential relationships between text attribute features in different views.

The integration of textual attributes and community structures in heterogeneous networks faces three core challenges, The first challenge is in regards to a semantic gap between low-level lexical features (e.g., word frequency) and high-level thematic features in texts, where direct fusion may lead to feature space misalignment. The second challenge is in regards to the macroscopic topological relationships reflected by community structures often exhibit inconsistencies with the microscopic semantics expressed in textual content. The third challenge is that existing methods typically process text embedding and community detection separately, neglecting the explanatory role of textual semantics in community formation. This paper addresses these cross-level representation alignment issues through joint modeling of multi-level textual semantics (word/topic) and multi-scale network structures (node/community).

The main contributions of this paper are as follows:

- This study proposes a novel heterogeneous GNN (HGNN) framework that achieves multi-level semantic representation learning by integrating network reconstruction and contrastive learning components. The framework employs a dynamic network reconstruction module to extract multi-level abstract features from textual attributes, and designs a cross-view contrastive learning mechanism to achieve coordinated optimization of structure and attributes in a shared