Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Technology Evolution-oriented Graph Representation Learning for Hidden Innovation Relationship Discovery in Power Grid Networks

Caihua Song

vxcrytf60652@outlook.com

Guangdong Power Grid Company (China)

ZiLan He

Guangdong Power Grid Company (China)

Zhaoxiong Guan

Guangdong Power Grid Company (China)

Kehuan Li

Guangdong Power Grid Company (China)

Article

Keywords:

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11313322/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Technology Evolution-oriented Graph Representation Learning for Hidden Innovation Relationship Discovery in Power Grid Networks

Caihua Song[1,*], He ZiLan[1], Guan Zhaoxiong[1], and Li Kehuan[1]

[1]Guangdong Power Grid Co., Ltd. Foshan Power Supply Bureau, Foshan, 510699, Guangdong, China
[*]vxcrytf60652@outlook.com

ABSTRACT

This study proposes a Power Grid Innovation Graph Representation Learning (PGI-GRL) framework for technology evolution-oriented hidden innovation relationship discovery in power grid networks. Existing approaches mainly rely on bibliometric analysis, text mining, or conventional graph learning methods, which are insufficient for modeling the evolutionary characteristics of heterogeneous innovation relationships and discovering latent technological connections. Scientific publications and patent documents are integrated to construct a heterogeneous innovation graph, where multiple innovation entities and relationship types are modeled to characterize different technological evolution patterns. Based on this graph, semantic features, structural dependencies, and heterogeneous relationship information are jointly learned through a technology evolution-oriented graph representation learning mechanism, enabling the framework to capture the evolutionary significance of different innovation interactions. A relationship prediction module is further developed to identify potential technological connections and latent innovation pathways. Experiments conducted on real-world power grid innovation data evaluate the proposed framework through link prediction, temporal relationship prediction, and hidden innovation discovery tasks. The results demonstrate that the proposed framework consistently outperforms representative graph embedding and graph neural network methods, showing superior capability in modeling complex innovation evolution patterns and discovering potential technology combinations. The proposed framework provides an effective data-driven approach for understanding technology evolution and supporting innovation relationship discovery in power grid networks.

1 Introduction

The rapid transformation of modern power systems has significantly accelerated technological innovation in the power grid domain[1]. With the increasing integration of renewable energy resources, the development of smart grid infrastructures, the adoption of digital twin technologies, the expansion of energy storage systems, and the application of artificial intelligence-based dispatching strategies, power grid innovation activities have become increasingly complex and interdisciplinary[2,3]. Unlike traditional technological development processes that mainly focus on isolated equipment improvement or individual system optimization, current power grid innovations are usually generated through the interaction of multiple technological domains, research organizations, industrial entities, and application scenarios[4]. As a result, innovation elements in the power grid domain exhibit heterogeneous characteristics, strong interdependencies, and complex relational structures[5]. Existing statistical and analytical approaches mainly concentrate on explicit associations[6], such as publication frequency, citation relationships, or keyword co-occurrence patterns, and therefore have limited capability to identify hidden technological connections and understand the evolutionary pathways of innovation relationships embedded in large-scale scientific and technological data.

Existing approaches for technological innovation analysis can be generally categorized into text mining methods, knowledge graph-based methods, and graph representation learning methods. Text mining approaches, including term frequency-inverse document frequency (TF-IDF), topic modeling, and transformer-based semantic representation methods such as BERT, have been widely applied to extract meaningful information from scientific documents[7]. These methods can effectively capture semantic characteristics and identify important technological concepts from textual resources. They mainly focus on textual similarity or keyword-level associations and lack the capability to model structural dependencies among different innovation elements. Knowledge graph-based approaches provide a more structured mechanism by representing entities and relationships in the form of interconnected networks, which has demonstrated effectiveness in scientific knowledge management and technology analysis. Nevertheless, existing knowledge graph construction methods often rely on manually predefined rules or limited relationship extraction strategies, making it difficult to represent diverse innovation interactions and adapt to complex technological evolution processes. More recently, graph representation learning methods have attracted increasing attention due to their ability to learn latent representations from graph structures. Methods such as Node2Vec, graph convolutional networks

(GCNs), graph attention networks (GATs), and GraphSAGE have achieved promising performance in various graph mining tasks. Conventional graph representation learning methods usually assume relatively homogeneous graph structures and cannot fully capture the heterogeneous entities, evolving temporal patterns, and multiple relationship types existing in power grid innovation networks[8]. Such methods typically model relationships as structural connections while overlooking the evolutionary meanings embedded in different innovation interactions. Addressing this limitation requires a framework that jointly models semantic information, structural characteristics, and heterogeneous innovation relationships to support technology evolution analysis and hidden innovation relationship discovery in the power grid domain.

To address the above challenges, this study proposes a Power Grid Innovation Graph Representation Learning Framework (PGI-GRL) for technology evolution-oriented hidden innovation relationship discovery. The proposed framework integrates scientific publications and patent documents to construct a heterogeneous innovation graph that represents multiple types of entities, including technologies, methods, devices, organizations, and applications, as well as diverse innovation relationships reflecting different technology evolution processes. Based on the constructed innovation graph, semantic and structural features of innovation elements are jointly initialized and incorporated into an adaptive relation-aware graph representation learning model. By dynamically considering different relationship types during graph information propagation, the proposed framework can effectively capture complex interactions among innovation elements. Furthermore, a relationship prediction module is developed to estimate potential connections between technological elements, enabling the discovery of hidden innovation relationships and potential technology evolution pathways that may provide insights into future power grid technology development[9]. The proposed PGI-GRL framework consists of four major stages, including innovation element extraction, heterogeneous innovation graph construction, graph representation learning, and innovation relationship prediction.

The main contributions of this study are summarized as follows.

1. A heterogeneous power grid innovation relationship network is constructed by integrating scientific publications and patent information, providing a unified graph-based representation framework for multiple types of innovation elements and their complex interactions, where different innovation relationships are modeled according to their roles in technology evolution.

2. An adaptive relation-aware graph representation learning model is proposed to learn technology evolution patterns by jointly modeling semantic information, topological structures, and heterogeneous innovation relationships of power grid innovation elements.

3. A graph learning-based potential innovation relationship prediction framework is developed to discover hidden innovation relationships and latent technology evolution pathways, providing data-driven support for innovation analysis and strategic technology planning in the power grid domain.

2 related works

2.1 Technology Innovation Mining in Power Grid Domain

The rapid development of modern power systems has significantly increased the complexity of technological innovation activities in the power grid domain. The integration of renewable energy resources, the deployment of smart grid infrastructures, the development of digital twin technologies, and the application of intelligent energy management strategies have promoted continuous technological evolution across multiple disciplines[10]. Unlike conventional power system innovations that primarily focus on individual components or isolated operational improvements, emerging power grid technologies increasingly rely on the interaction and integration of diverse knowledge domains, including power electronics, information communication, artificial intelligence, and data-driven optimization[11]. Consequently, technological innovation in the power grid domain has gradually evolved into a complex knowledge network involving multiple innovation entities, heterogeneous technologies, and interconnected research directions. Identifying the relationships among these innovation elements has become essential for understanding technological evolution, discovering emerging research opportunities, and supporting strategic decision-making in power system development.

Traditional approaches for technology innovation mining in the power grid domain mainly rely on bibliometric analysis and statistical methods[12]. By analyzing publication trends, citation patterns, and keyword distributions, these approaches have been widely used to identify research hotspots, evaluate technological development trends, and measure the influence of specific research areas. For example, citation analysis can reveal the evolution of influential studies and research communities, while keyword co-occurrence analysis can identify frequently studied technological concepts. These methods provide valuable insights into the overall landscape of power grid innovation. They mainly focus on explicit statistical associations and often treat technological elements as independent terms or aggregated indicators. As a result, they have limited capability to capture complex interactions among technologies and reveal hidden innovation relationships that emerge through cross-domain integration.

To overcome the limitations of statistical analysis, text mining-based approaches have been increasingly applied to extract semantic information from scientific literature and technical documents. Methods based on natural language processing, including term frequency-inverse document frequency (TF-IDF), topic modeling, word embedding, and transformer-based language models, have demonstrated strong capabilities in identifying technological concepts and representing semantic similarities among research topics. In the power grid domain, these approaches have been applied to analyze research trends, extract technical keywords, and classify scientific documents. Nevertheless, text-based methods mainly rely on textual information and semantic similarity, while the structural relationships among innovation elements are often overlooked. For example, two technologies may have limited textual similarity but may form meaningful innovation connections through application, improvement, or integration relationships. Existing text mining approaches remain insufficient for modeling the multi-dimensional and interconnected nature of power grid innovation networks[13].

With the increasing availability of large-scale scientific and technological data, technology innovation mining has gradually shifted from isolated information extraction toward relationship-oriented knowledge discovery. Recent studies have attempted to combine semantic analysis with structured representation methods to better characterize innovation processes. Existing approaches still face challenges in simultaneously modeling heterogeneous entities, multiple relationship types, and latent technological connections. These limitations indicate the need for advanced graph-based learning frameworks that can integrate semantic knowledge with relational structures and capture the evolutionary patterns embedded in complex power grid innovation relationships[14].

2.2 Knowledge Graph and Heterogeneous Network Modeling

Knowledge graph-based approaches have emerged as an effective paradigm for representing complex relationships among heterogeneous information entities[15]. By organizing knowledge in the form of entities and relations, knowledge graphs provide a structured representation mechanism beyond conventional text-based analysis and enable the integration of diverse information sources[16]. In scientific and technological innovation analysis, knowledge graphs have been widely adopted to model relationships among concepts, research topics, organizations, patents, and technological resources. Compared with traditional statistical approaches that mainly capture explicit correlations, knowledge graphs can represent multi-dimensional interactions among innovation elements and provide a foundation for knowledge discovery and relationship analysis[17].

Existing studies have explored the construction of scientific knowledge graphs and patent knowledge graphs for understanding technological development patterns[18]. Scientific knowledge graphs typically integrate information from academic publications, including papers, authors, institutions, concepts, and citation relationships, to characterize the evolution of research fields and knowledge diffusion processes. Patent-oriented knowledge graphs mainly focus on technological entities, inventors, organizations, and patent classifications to analyze innovation activities and industrial technology development. These structured representations have demonstrated advantages in identifying technological clusters, analyzing collaboration patterns, and supporting technology management decisions. In the power grid domain, knowledge graph approaches have also been investigated for applications such as equipment knowledge management, fault diagnosis, energy system analysis, and technological trend identification.

Although knowledge graph-based methods provide a more expressive representation of innovation information, several limitations remain when applied to power grid technology innovation mining. Many existing approaches construct graphs based on limited relationship types, such as citation links or keyword associations, which cannot fully represent the diverse interactions among innovation elements. In practical power grid innovation processes, relationships among technologies may involve development, application, improvement, cooperation, and integration, and these heterogeneous relationships not only describe structural connections but also reflect different technology evolution processes. Conventional knowledge graph construction methods often rely on manually defined rules or predefined extraction strategies, which may reduce their adaptability when processing large-scale and evolving technological data. The rapid emergence of interdisciplinary technologies makes it increasingly difficult to manually specify all potential innovation relationships. Existing knowledge graph studies mainly focus on representing and querying existing knowledge, while they provide limited capability for predicting potential relationships among currently disconnected innovation elements.

To address these limitations, heterogeneous network modeling has attracted increasing attention because it enables the representation of multiple types of nodes and relations within a unified graph structure. Unlike homogeneous networks that assume identical node and edge types, heterogeneous networks can explicitly model complex interactions among different innovation entities, such as technologies, methods, devices, organizations, and applications. This characteristic makes heterogeneous network modeling particularly suitable for representing power grid innovation ecosystems, where technological advancement is often driven by cross-domain integration and multi-agent collaboration. Constructing a heterogeneous innovation network alone is insufficient for discovering latent technological connections, and effective learning mechanisms are required to extract meaningful representations from such complex graph structures.

Further research is required to develop technology evolution-oriented graph learning methods that can jointly exploit

semantic information, network topology, and heterogeneous relationship characteristics to discover latent innovation pathways in power grid networks. Such methods are expected to overcome the limitations of existing knowledge graph and heterogeneous network modeling approaches and enable more accurate discovery of hidden innovation relationships in the power grid domain.

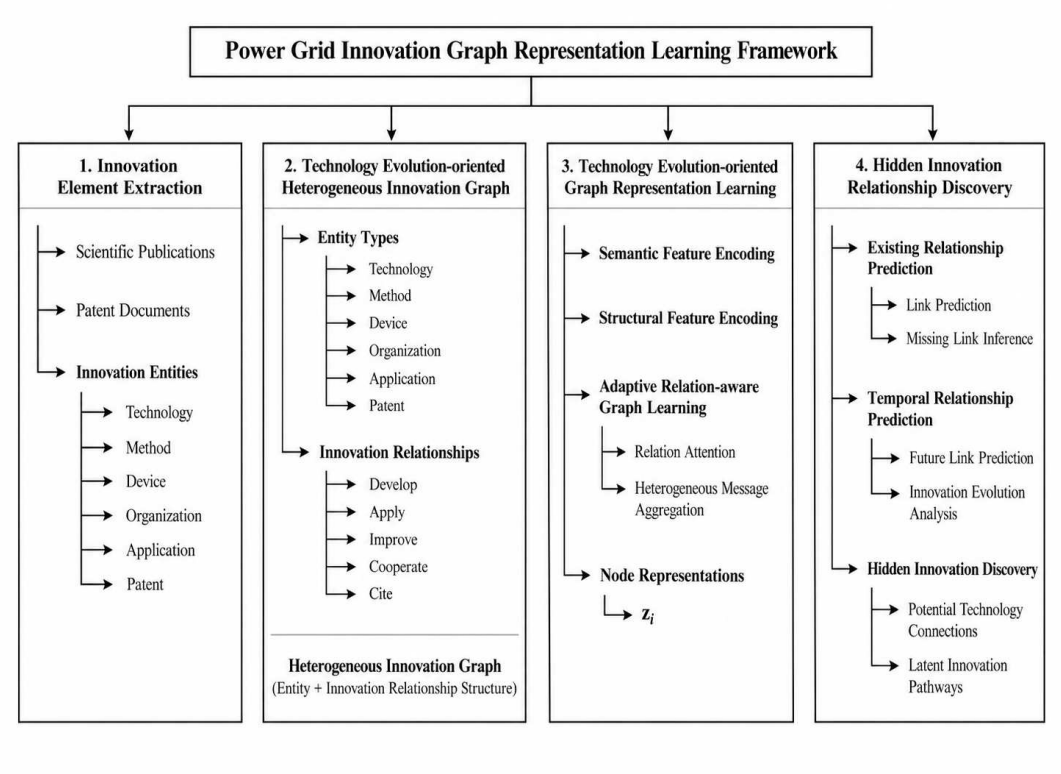


Figure 1. Overall framework of the proposed Power Grid Innovation Graph Representation Learning (PGI-GRL) framework, including innovation element extraction, technology evolution-oriented heterogeneous graph construction, adaptive relation-aware representation learning, and hidden innovation relationship discovery.

2.3 Graph Representation Learning for Innovation Relationship Discovery

Graph representation learning has recently become an important approach for extracting latent patterns from complex network structures[19]. Different from traditional statistical analysis methods that mainly rely on manually designed features, graph representation learning aims to automatically learn low-dimensional representations of nodes, edges, and subgraphs by exploiting topological structures and neighborhood information. These learned representations have demonstrated strong capabilities in various tasks, including node classification, link prediction, and network analysis[20]. Graph representation learning provides a promising solution for modeling technological innovation networks, where innovation elements and their relationships can be represented as interconnected graph structures reflecting technological evolution processes.

Early graph representation learning methods mainly focused on network embedding techniques that preserve structural proximity among nodes. Representative approaches such as DeepWalk, Node2Vec, and LINE learn node representations through random walk strategies or network neighborhood information. These methods have been successfully applied to social networks, recommendation systems, and knowledge discovery tasks. They generally assume that the underlying graph is homogeneous and that different edges represent similar types of interactions. Such assumptions limit their applicability to power grid innovation networks, where various relationships among technological elements, such as development, application, improvement, and cooperation, represent different technology evolution mechanisms and cannot be treated equally.

With the advancement of deep learning techniques, graph neural networks (GNNs) have been introduced to improve the capability of graph representation learning by aggregating information from neighboring nodes[21]. Graph convolutional networks (GCNs) extend convolution operations to graph structures and learn node representations by combining local neighborhood information. GraphSAGE further improves scalability by introducing neighborhood sampling strategies, enabling representation learning on large-scale graphs. Graph attention networks (GATs) incorporate attention mechanisms to assign different weights to neighboring nodes and enhance the adaptability of information aggregation. These methods have achieved promising

results in many graph mining tasks[22]; They are primarily designed for homogeneous graphs and often overlook the semantic differences among multiple node and edge types.

To address the limitations of homogeneous graph learning methods, heterogeneous graph neural networks have been developed to model networks containing multiple types of nodes and relations[23]. Approaches such as relational graph convolutional networks (R-GCN), heterogeneous attention networks (HAN), and heterogeneous graph transformers (HGT) introduce relation-specific or type-specific learning mechanisms to better capture complex graph structures[24]. These methods have shown advantages in applications involving knowledge graphs, recommendation systems, and scientific networks. Nevertheless, existing heterogeneous graph learning approaches still face challenges when applied to power grid innovation relationship discovery[25]. Many methods mainly focus on structural information while insufficiently incorporating the semantic characteristics of technological concepts extracted from scientific and patent documents. Existing models usually consider predefined relation types but lack adaptive mechanisms for learning how different innovation relationships contribute to technology evolution processes. Most existing studies emphasize representation learning or classification tasks, while the prediction of potential technological relationships among disconnected innovation elements remains insufficiently explored.

An effective graph representation learning framework for power grid innovation discovery should simultaneously consider semantic information, graph topology, and the evolutionary characteristics of heterogeneous innovation relationships[26]. By integrating these complementary information sources, such a framework can better capture the complex evolution patterns among innovation elements and identify hidden technological connections that are not explicitly observed in existing innovation networks.

2.4 Research Gap

Existing studies have made significant progress in analyzing technological innovation through statistical analysis, text mining, knowledge graph construction, and graph representation learning approaches. Bibliometric and text mining methods provide effective tools for extracting research trends and semantic information from large-scale scientific documents, while knowledge graph and heterogeneous network approaches improve the representation capability of complex innovation entities and relationships. Furthermore, graph representation learning methods have demonstrated strong potential in capturing latent structural patterns and predicting unknown relationships in complex networks. Despite these advances, existing approaches still face several challenges when applied to innovation relationship mining in the power grid domain.

Existing technology innovation analysis methods generally focus on either textual information or predefined structural relationships, while lacking a unified framework that integrates semantic characteristics and heterogeneous network structures. Scientific publications and patent documents contain rich technological knowledge, but textual similarity alone cannot fully reflect the functional relationships and evolutionary connections among innovation elements. Meanwhile, conventional knowledge graph approaches mainly rely on manually designed schemas and predefined relation extraction rules, which may limit their adaptability in rapidly evolving technology domains.

Existing graph representation learning methods have limited capability in modeling the diverse relationship characteristics of power grid innovation networks. Most traditional graph embedding methods and homogeneous graph neural networks assume that graph nodes and edges follow similar interaction patterns, which is inconsistent with practical innovation ecosystems where relationships such as technology development, technology application, technological improvement, and organizational cooperation represent different semantic meanings. Although heterogeneous graph learning methods have introduced relation-specific modeling strategies, they usually rely on fixed relation representations and insufficiently consider the dynamic importance of different relationships during information propagation.

Existing studies mainly focus on describing existing technological structures or performing conventional graph mining tasks, while the discovery of potential innovation relationships remains relatively underexplored[27]. For power grid technology development, identifying emerging connections among currently weakly related or disconnected innovation elements is particularly valuable for technology planning and strategic decision-making. Existing approaches rarely investigate how graph representation learning can be utilized to predict hidden innovation relationships by jointly considering semantic knowledge, structural dependencies, and heterogeneous relational information.

To address these research gaps, this study develops a technology evolution-oriented graph representation learning framework for mining hidden innovation relationships among power grid innovation elements. By constructing a heterogeneous innovation network from scientific publications and patent documents, integrating semantic and structural feature representations, and dynamically learning the importance of different relationship types, the proposed framework aims to provide a more effective solution for innovation relationship prediction and technology discovery in the power grid domain.

3 Methodology

3.1 Framework Overview

The objective of this study is to discover hidden innovation relationships among scientific and technological innovation elements in the power grid domain by learning informative representations from heterogeneous innovation networks. Considering that power grid innovation activities involve diverse entities, complex interactions, and heterogeneous relationships with different semantic meanings, a unified framework is required to integrate textual knowledge and graph structures for effective innovation relationship learning. This study proposes a Power Grid Innovation Graph Representation Learning Framework (PGI-GRL), which integrates heterogeneous graph construction, feature representation learning, adaptive relation-aware graph representation learning, and hidden innovation relationship discovery.

As illustrated in Fig. 1, the proposed framework consists of four sequential stages. Scientific publications and patent documents are collected from multiple public databases, and relevant innovation elements are extracted to construct a heterogeneous innovation graph containing different types of innovation entities and relationships. Semantic and structural features of innovation elements are initialized to provide informative representations for subsequent graph representation learning. An adaptive relation-aware graph representation learning model is developed to capture the dependencies among innovation elements by considering node characteristics, graph topology, and heterogeneous relationship types. Finally, the learned representations are utilized to predict potential relationships among innovation elements, enabling the discovery of hidden innovation relationships within the power grid innovation network.

The proposed PGI-GRL framework jointly exploits semantic knowledge and relational structures to generate comprehensive representations of innovation elements. By dynamically learning the contributions of heterogeneous relationships during graph propagation, the framework enables more effective modeling of complex innovation interactions and supports hidden innovation relationship discovery in the power grid domain.

3.2 Construction of Power Grid Innovation Heterogeneous Graph

The first step of the proposed PGI-GRL framework is to construct a heterogeneous innovation graph that can effectively represent the complex interactions among technological innovation elements in the power grid domain. Unlike conventional representations based on isolated textual information or simple statistical associations, the proposed graph integrates scientific publications and patent documents to capture both knowledge development and technology implementation processes. Scientific publications provide information related to theoretical exploration, methodological development, and emerging research directions, while patent documents reflect technological applications, engineering practices, and industrial innovation activities[28]. By combining these complementary information sources, a more comprehensive representation of power grid innovation activities can be established.

In this study, scientific publications and patent documents related to power grid technologies are collected from public academic and patent databases. The collected documents are selected within the period of 2018 to 2025 to capture recent technological developments associated with smart grids, renewable energy integration, energy storage systems, intelligent dispatching, and digital twin technologies. After removing duplicate and irrelevant records, the textual contents of the documents are processed through terminology normalization and entity alignment to reduce inconsistencies caused by different expressions of the same technological concepts. The processed documents are then utilized to extract innovation elements and construct the corresponding heterogeneous graph structure.

Considering the multi-dimensional characteristics of power grid innovation activities, six types of entities are defined in the heterogeneous innovation graph, including technologies, methods, devices, applications, organizations, and patents. The node set of the constructed graph is represented as

$$V = \{V_T, V_M, V_D, V_A, V_O, V_P\}, \tag{1}$$

where V_T, V_M, V_D, V_A, V_O, and V_P denote technology, method, device, application, organization, and patent entities, respectively. These heterogeneous entities describe different aspects of innovation activities and provide complementary information for innovation relationship modeling.

After identifying innovation entities, multiple relationship types are constructed to preserve the semantic differences among innovation interactions. The proposed heterogeneous innovation graph is defined as

$$G = (V, E, R), \tag{2}$$

where V represents the set of innovation entities, E denotes the relationship set among entities, and R represents different relationship types. Different from homogeneous graphs that represent interactions using a single connection type, the proposed

graph preserves multiple relationship categories to capture the semantic differences among innovation interactions. These heterogeneous relationships are not only structural connections but also reflect different stages and mechanisms of technology evolution. For example, develop relationships indicate technology generation, apply relationships represent technology deployment, and improve relationships reflect iterative evolution of existing solutions.The proposed graph considers five major relationship categories, including develop, apply, improve, cooperate, and cite. The develop relationship describes the creation or advancement of technological solutions, the apply relationship indicates the deployment of technologies in specific application scenarios, and the improve relationship represents optimization or enhancement among related technologies. The cooperate relationship captures collaborative innovation activities among organizations, while the cite relationship reflects knowledge diffusion through academic and patent references.

The constructed heterogeneous innovation graph provides a unified representation of diverse innovation elements and their interactions in the power grid domain. By preserving multiple entity types and relationship semantics, the graph structure enables subsequent learning models to distinguish different innovation interactions rather than treating all connections as identical. The constructed graph serves as the foundation for semantic and structural feature initialization and supports the adaptive relation-aware graph representation learning process for discovering hidden innovation relationships in the power grid domain.

3.3 Semantic and Structural Feature Initialization

The heterogeneous innovation graph constructed in Section 3.2 provides the structural basis for modeling relationships among power grid innovation elements. Graph topology alone cannot fully characterize the semantic meanings of different innovation entities, especially when technological concepts have similar structural positions but represent different knowledge contents[29]. This study initializes node representations by jointly incorporating semantic information extracted from scientific and technological documents and structural information derived from the heterogeneous innovation graph, enabling innovation elements to be represented from both knowledge and relational perspectives. The combination of semantic and structural features provides informative representations for subsequent adaptive relation-aware graph representation learning.

For each innovation element $v_i \in V$, the corresponding textual description is obtained from associated scientific publications, patent documents, and entity-related contexts. Since technological terms in the power grid domain often contain complex contextual meanings, conventional word-level representations are insufficient for capturing their semantic characteristics. A domain-independent transformer-based language representation model is employed to encode textual information. Given the textual description of an innovation element, the semantic representation is obtained as

$$h_i^s = Encoder(x_i), \tag{3}$$

where x_i denotes the textual information associated with node v_i, $Encoder(\cdot)$ represents the semantic encoding process, and h_i^s represents the semantic feature vector of the innovation element. The semantic representation captures contextual information among technological concepts and enables the model to distinguish innovation elements with different meanings.

Although semantic representations provide rich contextual information, they cannot explicitly describe the structural interaction patterns among innovation elements. Structural features are further extracted from the heterogeneous innovation graph. The structural representation reflects the topological characteristics of each node by considering its neighboring entities and relationship connections. For an innovation element v_i, its structural feature can be expressed as

$$h_i^g = GraphEncoder(v_i, G), \tag{4}$$

where G represents the heterogeneous innovation graph and $GraphEncoder(\cdot)$ denotes the structural feature extraction operation. The structural representation captures the connectivity patterns among innovation elements and provides complementary information to semantic features.

To obtain a comprehensive initial representation, semantic and structural features are combined through feature fusion. The initial representation of each innovation element is formulated as

$$h_i^0 = Concat(h_i^s, h_i^g), \tag{5}$$

where $Concat(\cdot)$ denotes the feature concatenation operation and h_i^0 represents the initial node representation used in subsequent graph representation learning. By integrating semantic characteristics and structural dependencies, the initialized node representations provide a comprehensive description of innovation elements for subsequent graph representation learning.

The obtained feature representations preserve both the semantic content and structural characteristics of power grid innovation elements. These representations are subsequently utilized in the adaptive relation-aware graph representation learning module, where heterogeneous relationship types are further considered to capture the dependencies among innovation elements.

3.4 Adaptive Relation-aware Graph Representation Learning

Although the semantic and structural features initialized in Section 3.3 provide informative representations of innovation elements, the heterogeneous innovation graph contains multiple relationship types with different semantic meanings and contributions to innovation interaction modeling. Conventional graph neural networks generally aggregate information from neighboring nodes without explicitly distinguishing relationship semantics, which may lead to inaccurate representation when applied to complex innovation networks[30]. For example, the relationship between a technology and its application scenario reflects functional deployment, whereas the relationship between two technologies may indicate improvement or integration. Treating these relationships as equivalent structural connections may weaken the ability of graph models to capture meaningful innovation interactions. Therefore, this study proposes an adaptive relation-aware graph representation learning mechanism to model heterogeneous innovation relationships and dynamically learn their contributions during graph information propagation[31]. The overall architecture of the proposed mechanism is illustrated in Fig. 2, where semantic and structural features are integrated with relation-specific attention and adaptive message aggregation to generate relation-aware node representations[32].

Given the heterogeneous innovation graph $G = (V, E, R)$, each node v_i is associated with an initial representation h_i^0 obtained from semantic and structural feature initialization. For each relationship type $r \in R$, the neighbor set of node v_i under relation r is denoted as N_i^r. The relation-specific message transmitted from neighboring node v_j to node v_i is formulated as

$$m_{ij}^r = W_r h_j^l, \tag{6}$$

where W_r represents the learnable transformation matrix associated with relation type r, and h_j^l denotes the representation of node v_j at the l-th graph learning layer. By introducing relation-specific transformation parameters, the model can preserve the semantic differences among various innovation relationships[33].

To further enhance the adaptability of relation modeling, an adaptive relation attention mechanism is designed to dynamically evaluate the contribution of different relationships during information aggregation[34]. Instead of assigning fixed importance to predefined relation types, the proposed mechanism determines relation weights according to the characteristics of connected innovation elements and their associated relationship types. The attention coefficient between node v_i and its neighbor v_j under relation r is calculated as

$$\alpha_{ij}^r = \frac{\exp(e_{ij}^r)}{\sum_{k \in N_i^r} \exp(e_{ik}^r)}, \tag{7}$$

where e_{ij}^r represents the attention score reflecting the relevance of the neighboring node v_j under relation type r. The attention score is computed by considering the representations of the connected nodes and their relationship embeddings, which enables the model to adaptively capture important innovation interactions.

Based on the relation-specific messages and adaptive attention coefficients, the representation of each innovation element is updated through multi-relation aggregation. The node representation at the next layer is calculated as

$$h_i^{l+1} = \sigma\left(\sum_{r \in R} \sum_{j \in N_i^r} \alpha_{ij}^r m_{ij}^r\right), \tag{8}$$

where $\sigma(\cdot)$ denotes the nonlinear activation function. Through this relation-aware message passing process, information from different types of innovation relationships can be selectively aggregated, allowing the model to capture complex heterogeneous interaction patterns within the power grid innovation network.

To obtain comprehensive innovation element representations, multiple graph learning layers are stacked to progressively capture higher-order dependencies among innovation entities. The final representation of each node is obtained as

$$z_i = h_i^L, \tag{9}$$

where L denotes the number of graph representation learning layers. The learned embedding z_i integrates semantic characteristics, graph topology, and heterogeneous relationship information, providing a comprehensive representation for subsequent hidden innovation relationship discovery.

The proposed adaptive relation-aware mechanism enables the model to distinguish heterogeneous innovation relationships and adjust information propagation according to graph context. This capability is particularly important for power grid innovation networks, where diverse innovation elements are connected through multiple relationship types with different semantic characteristics[35].

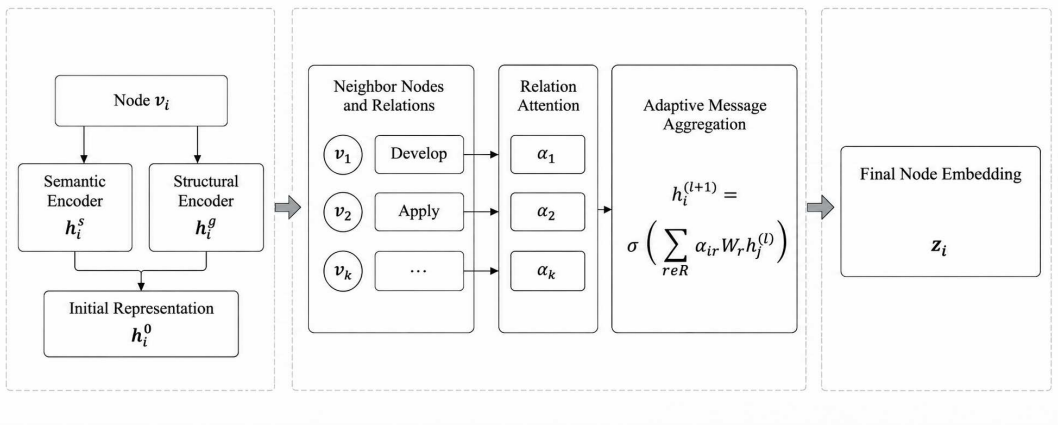


Figure 2. Illustration of the adaptive relation-aware graph representation learning mechanism, where semantic features, structural dependencies, and heterogeneous innovation relationships are jointly modeled to generate relation-aware node representations.

3.5 Hidden Innovation Relationship Discovery

The adaptive relation-aware graph representation learning module generates comprehensive embeddings for innovation elements by integrating semantic information, graph topology, and heterogeneous innovation relationship characteristics[36]. Based on these learned representations, this study further develops a hidden innovation relationship discovery mechanism to identify potential innovation relationships among innovation elements. The proposed discovery mechanism aims to identify latent relationships between currently weakly connected or disconnected innovation elements, thereby supporting the exploration of potential technological combinations in the power grid domain.

Given two innovation elements v_i and v_j, their learned representations obtained from the graph representation learning module are denoted as z_i and z_j, respectively. The potential relationship between these two elements is estimated by constructing a pairwise representation that captures their individual characteristics and interactive information. Specifically, the relationship representation is formulated as

$$z_{ij} = Concat(z_i, z_j, z_i \odot z_j), \tag{10}$$

where $Concat(\cdot)$ denotes the concatenation operation and $\odot$ represents element-wise multiplication. The concatenated representation preserves both independent node characteristics and latent interaction patterns, which provides informative features for hidden innovation relationship discovery.

A multilayer perceptron (MLP) classifier is employed to estimate the probability of a potential innovation relationship between two elements[37]. The prediction probability is calculated as

$$\hat{y}_{ij} = \sigma(W_p z_{ij} + b_p), \tag{11}$$

where W_p and b_p denote learnable parameters of the prediction layer, and $\sigma(\cdot)$ represents the sigmoid activation function. The output value $\hat{y}_{ij}$ indicates the likelihood that a potential relationship exists between innovation elements v_i and v_j.

During model optimization, binary cross-entropy loss is adopted as the training objective, which is defined as

$$L = -\sum_{(i,j)} [y_{ij} \log(\hat{y}_{ij}) + (1 - y_{ij}) \log(1 - \hat{y}_{ij})], \tag{12}$$

where y_{ij} represents the ground-truth relationship label and $\hat{y}_{ij}$ denotes the predicted probability. Positive samples correspond to observed relationships in the heterogeneous innovation graph, while negative samples are generated from unconnected innovation element pairs.

After training, the proposed discovery mechanism can estimate potential relationship scores for unconnected innovation elements and rank candidate technological connections according to their predicted probabilities. The high-scoring candidate pairs represent potential innovation relationships that may indicate emerging technology integration opportunities. By transforming graph representations into predictive innovation insights, the proposed framework extends beyond explicit relationship description and provides a data-driven approach for discovering hidden innovation relationships in power grid innovation networks.

4 Experimental Setup

4.1 Dataset Description

To evaluate the effectiveness of the proposed PGI-GRL framework, a power grid innovation dataset is constructed by integrating scientific publications and patent documents related to power system technologies. Scientific publications provide abundant information regarding theoretical research, methodological development, and emerging technological directions, while patent documents reflect practical applications, engineering implementations, and technology transfer activities. The combination of these two data sources enables a more comprehensive representation of innovation activities and provides sufficient heterogeneous relational information for innovation relationship representation learning.

The scientific publication data are collected from IEEE Xplore and Web of Science databases, which contain extensive research outputs in the fields of power systems, smart grids, renewable energy integration, energy storage, and intelligent power management. Patent documents are collected from publicly available patent databases, including Google Patents and the China National Intellectual Property Administration. To focus on recent technological development trends, all documents are selected within the period from 2018 to 2025. A domain-oriented retrieval strategy is adopted using keywords related to smart grids, power system optimization, digital twin technologies, energy storage systems, and artificial intelligence-based grid management.The retrieval queries were constructed by combining core power grid terms with emerging technology-related terms to improve coverage of both fundamental research and practical applications. The search process considered titles, abstracts, keywords, and available metadata fields to identify relevant innovation-related documents. After removing duplicate records, incomplete documents, and irrelevant contents, the remaining documents were further screened according to their relevance to power grid innovation activities. Documents unrelated to power system technologies or lacking sufficient technical descriptions were excluded. The retained records were subsequently processed for entity extraction and relationship construction.The extracted innovation elements were normalized to ensure consistency among different expressions of the same technological concepts. Specifically, terminology normalization was performed to merge synonymous technical terms, while entity alignment was conducted to associate equivalent entities extracted from scientific publications and patent documents. The aligned entities were categorized into six types, including technologies, methods, devices, applications, organizations, and patents.

The heterogeneous relationships were constructed based on semantic associations identified from document contents, metadata, and citation information. Different relationship categories were defined according to their innovation meanings. Development relationships represent the generation or advancement of technological solutions, application relationships indicate the deployment of technologies in practical scenarios, improvement relationships describe optimization among related technologies, cooperation relationships reflect collaborative innovation activities, and citation relationships represent knowledge transmission among scientific and technological resources.

The constructed dataset contains approximately 12,000 scientific publications and 4,000 patent documents. The extracted documents are processed through terminology normalization and entity alignment to reduce inconsistencies caused by different expressions of the same technological concepts. Innovation entities are extracted from textual descriptions, metadata, and technical contexts, and entity matching is performed to merge semantically equivalent concepts. Based on the semantic associations among extracted entities, heterogeneous relationships are constructed to represent different innovation interactions, including development, application, improvement, cooperation, and citation relationships. Through innovation element extraction and relationship construction, the final heterogeneous innovation graph contains approximately 38,000 innovation entities and 150,000 relationships. The innovation entities include technologies, methods, devices, applications, organizations, and patents, while the relationships represent different innovation interactions with distinct semantic characteristics, including development, application, improvement, cooperation, and citation relationships. The statistics of the constructed heterogeneous innovation graph are summarized in Table 1.

The constructed dataset provides a comprehensive evaluation environment for innovation relationship discovery because it contains diverse entity types and heterogeneous interaction patterns. Compared with homogeneous citation networks or keyword co-occurrence graphs, the proposed dataset preserves richer semantic and structural information among power grid

Table 1. Statistics of the Power Grid Innovation Heterogeneous Graph

Category	Type	Number
Node	Technology	11,000
Node	Method	7,000
Node	Device	5,000
Node	Application	5,000
Node	Organization	4,000
Node	Patent	6,000
Edge	Develop	40,000
Edge	Apply	35,000
Edge	Improve	30,000
Edge	Cooperate	25,000
Edge	Cite	20,000

innovation elements. It is suitable for evaluating whether the proposed adaptive relation-aware graph representation learning framework can effectively capture heterogeneous innovation relationships and discover hidden innovation relationships.To improve the reliability of the constructed innovation graph, several quality control procedures were applied during dataset construction. Duplicate entities, ambiguous technological concepts, and inconsistent terminology were reviewed and corrected before graph generation. In addition, relationship construction rules were carefully designed to avoid introducing unsupported connections among innovation elements.

4.2 Experimental Tasks

To comprehensively evaluate the capability of the proposed PGI-GRL framework in discovering hidden innovation relationships, three experimental tasks are designed, including link prediction, temporal relationship prediction, and innovation discovery. These tasks evaluate the model from different perspectives, including relationship modeling capability, future relationship prediction ability, and practical technology discovery performance.

1) Link Prediction. The link prediction task is designed to evaluate whether the learned node representations can accurately identify potential relationships among innovation elements. Given two innovation entities in the heterogeneous innovation graph, the model predict the probability of a potential innovation relationship between them. The dataset is randomly divided into training, validation, and testing sets with a ratio of 80%, 10%, and 10%, respectively. Observed relationships are treated as positive samples, while unconnected entity pairs are randomly sampled as negative samples. This task directly evaluates the effectiveness of the proposed adaptive relation-aware graph representation learning mechanism.

2) Temporal Relationship Prediction. To evaluate the capability of discovering emerging innovation relationships, a temporal relationship prediction task is further conducted. Unlike conventional random data splitting, this task adopts a chronological evaluation strategy, where innovation relationships before 2023 are used for model training and relationships emerging during 2023–2025 are used for testing[38]. This setting simulates real-world technology development scenarios, where historical innovation information is utilized to predict future innovation relationships.

3) Innovation Discovery. The innovation discovery task is designed to demonstrate the practical application value of the proposed framework. Based on the predicted relationship probabilities, disconnected innovation element pairs are ranked according to their potential association scores. The top-ranked candidate relationships are analyzed as potential emerging technology combinations in the power grid domain. The top-ranked candidate relationships are further analyzed according to their predicted scores and technological relevance to demonstrate the practical applicability of the proposed framework.

4.3 Baselines

To evaluate the effectiveness of the proposed PGI-GRL framework, representative graph representation learning methods are selected as baselines. The compared methods cover three categories, including traditional network embedding approaches, homogeneous graph neural networks, and heterogeneous graph learning methods. These methods represent different stages of graph representation learning development and provide comprehensive comparisons regarding structural modeling and relationship learning capability.

The first category includes traditional network embedding methods, including DeepWalk, Node2Vec, and LINE. These approaches learn low-dimensional node representations by preserving graph proximity and structural characteristics. Although effective in capturing local network patterns, they generally ignore semantic differences among nodes and relationships.

The second category consists of homogeneous graph neural network methods, including GCN, GraphSAGE, and GAT. These methods improve representation learning by aggregating neighborhood information through graph convolution, sampling,

or attention mechanisms. They usually assume homogeneous node and edge structures and cannot explicitly model multiple innovation relationship types.

The third category includes heterogeneous graph learning methods, including R-GCN, HAN, and HGT. These approaches are designed to handle multiple node types and relationship types and therefore provide stronger comparisons for the proposed framework. Nevertheless, existing heterogeneous graph methods mainly rely on predefined relation representations and lack adaptive learning of heterogeneous relationship contributions. The proposed PGI-GRL framework is compared with these baselines under identical experimental settings to verify its capability in jointly modeling semantic information, structural dependencies, and heterogeneous innovation relationships.

4.4 Metrics

To comprehensively evaluate the performance of the proposed PGI-GRL framework, different evaluation metrics are adopted according to the characteristics of each experimental task. For relationship prediction tasks, including link prediction and temporal relationship prediction, the area under the receiver operating characteristic curve (AUC), average precision (AP), and F1-score are employed. AUC measures the overall ranking capability of the model in distinguishing positive and negative relationships, while AP evaluates the precision of predicted innovation relationships under different confidence thresholds. The F1-score provides a balanced evaluation by considering both precision and recall.

For the innovation discovery task, Precision@K is adopted to measure the proportion of correctly identified potential innovation relationships among the top-ranked candidates. For the innovation discovery task, Precision@K is adopted to evaluate the accuracy of the top-ranked candidate relationships generated by the proposed framework. These metrics jointly assess the predictive capability and practical applicability of the proposed framework in power grid hidden innovation relationship discovery.

4.5 Implementation Details

All experiments are implemented using the PyTorch framework with the PyTorch Geometric library for graph representation learning. The experiments are conducted on a workstation equipped with an NVIDIA GPU to accelerate model training. The proposed PGI-GRL framework and all baseline methods are evaluated under the same experimental environment to ensure fair comparisons[39].

For model training, the Adam optimizer is adopted with an initial learning rate of 0.001. The embedding dimension is set to 128, and the number of graph learning layers is set to two. Dropout is applied to reduce overfitting, with the dropout rate set to 0.3. The model is trained for a maximum of 200 epochs with early stopping based on validation performance. The parameters of all baseline models are selected according to their original implementations and optimized using the validation set.

To reduce the influence of random initialization, all experiments are independently repeated five times, and the average performance with standard deviation is reported. For relationship prediction tasks, the same training, validation, and testing splits are applied to all methods. For temporal relationship prediction, the chronological data partition described in Section 4.2 is maintained to evaluating the capability of discovering future innovation relationships.

5 Results and Discussion

5.1 Innovation Graph Characteristics

Before evaluating the prediction performance of the proposed PGI-GRL framework, the structural characteristics of the constructed power grid innovation heterogeneous graph are analyzed. The purpose of this analysis is to examine whether the proposed graph construction strategy can effectively represent the diversity and complexity of technological innovation relationships in the power grid domain. As described in Section 4.1, the constructed graph contains approximately 38,000 innovation entities and 150,000 heterogeneous relationships extracted from scientific publications and patent documents.

The distribution of different types of innovation entities is presented in Table 1 and Fig. 3(a). The statistical results show that technology nodes constitute the largest proportion among all innovation entities, indicating that technological concepts represent the core elements of power grid innovation activities. Method and device entities also account for considerable proportions, reflecting the increasing integration of algorithmic approaches and engineering implementations in modern power systems. In addition, application, organization, and patent entities provide complementary information regarding practical deployment scenarios, innovation participants, and intellectual property development. The coexistence of these heterogeneous entities demonstrates that power grid innovation is not driven by isolated technological components but emerges through interactions among knowledge, methods, applications, and organizations.

The relationship distribution further illustrates the heterogeneous characteristics of the constructed innovation graph, as shown in Fig. 3(b). Different relationship types, including develop, apply, improve, cooperate, and cite, represent distinct innovation interaction patterns. Among them, development and application relationships occupy a relatively large proportion, suggesting that technological advancement in the power grid domain is strongly associated with the development of technical

solutions and their practical deployment. Improvement and cooperation relationships further reveal the iterative optimization process and collaborative characteristics of innovation activities, while citation relationships reflect knowledge diffusion across scientific and technological resources.

Based on the entity and relationship distributions shown in Fig. 3, the constructed heterogeneous graph exhibits several important characteristics for innovation relationship mining. The coexistence of multiple entity and relationship types provides richer information than conventional homogeneous citation networks or keyword co-occurrence graphs. The diverse connection patterns indicate that innovation relationships cannot be accurately modeled by treating all interactions equally. Different relationship types carry distinct semantic meanings and may contribute differently to technological evolution. The integration of scientific publications and patent information enables the graph to capture both knowledge generation and technology application processes.

These characteristics provide a suitable foundation for the subsequent adaptive relation-aware graph representation learning model. By preserving heterogeneous entities and relationship semantics, the constructed innovation graph allows the proposed framework to learn more informative representations and discover potential technological connections that are difficult to identify through traditional statistical or text-based approaches.

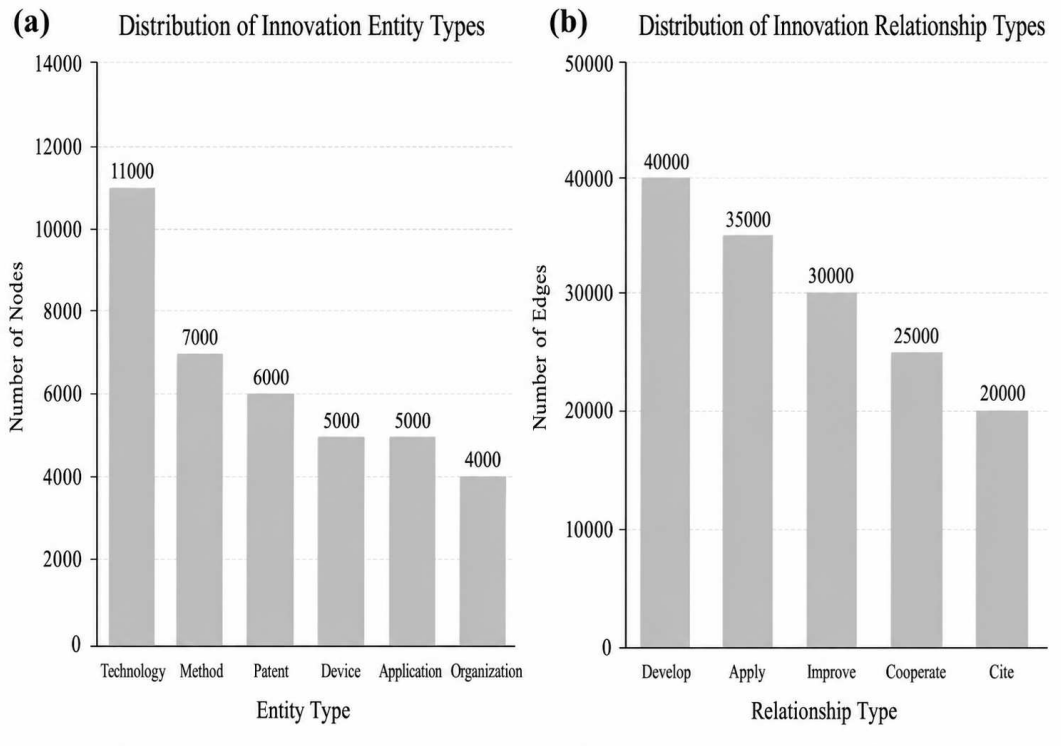


Figure 3. Distribution characteristics of the constructed power grid innovation heterogeneous graph, including entity types and relationship types.

5.2 Relationship Prediction Performance

The performance comparison of different methods on innovation relationship prediction is presented in Table 2. The results demonstrate that graph-based representation learning methods generally outperform traditional network embedding approaches, indicating that modeling neighborhood dependencies is beneficial for capturing complex interactions among power grid innovation elements. Specifically, DeepWalk, Node2Vec, and LINE achieve relatively limited performance, with AUC values ranging from 0.804 to 0.826. These methods mainly preserve local structural proximity through random walk-based representation learning while ignoring semantic characteristics and heterogeneous relationship information contained in scientific publications and patent documents, which limits their capability for complex innovation relationship modeling.

Compared with traditional embedding methods, GCN, GraphSAGE, and GAT achieve consistent improvements, obtaining AUC values of 0.861, 0.873, and 0.881, respectively. These improvements indicate that neighborhood aggregation mechanisms can effectively enhance innovation element representations by incorporating structural dependencies from connected entities. However, these homogeneous graph neural networks still treat different relationship types in a similar manner and cannot

explicitly distinguish the semantic differences among develop, apply, improve, cooperate, and cite relationships in the heterogeneous innovation graph.

The heterogeneous graph learning methods further improve prediction performance. R-GCN, HAN, and HGT achieve AUC values of 0.901, 0.913, and 0.924, respectively, demonstrating the effectiveness of explicitly modeling heterogeneous entities and relationship types in power grid innovation networks. The improvement over homogeneous graph learning methods indicates that relationship semantics provide important information for identifying potential innovation connections. Nevertheless, these approaches mainly rely on predefined relation representations and lack adaptive mechanisms for dynamically evaluating the importance of different innovation interactions.

The proposed PGI-GRL achieves the best performance with an AUC of 0.942 ± 0.003, an AP of 0.931 ± 0.003, and an F1-score of 0.896 ± 0.004. Compared with the strongest baseline HGT, PGI-GRL improves AUC from 0.924 to 0.942, corresponding to an increase of 1.8 percentage points. The AP and F1-score are improved from 0.912 to 0.931 and from 0.879 to 0.896, respectively. Meanwhile, the relatively small standard deviations indicate that PGI-GRL maintains stable performance across repeated experiments. The improvement can be attributed to the proposed adaptive relation-aware graph representation learning mechanism, which jointly integrates semantic features extracted from technological documents, structural dependencies from heterogeneous graphs, and dynamically learned relation importance during message propagation. By selectively aggregating information from different innovation relationships, PGI-GRL generates more discriminative representations for identifying potential innovation relationships, thereby enabling more effective hidden innovation relationship discovery in complex power grid innovation networks.

Table 2. Performance Comparison of Different Methods on Innovation Relationship Prediction.

Method	AUC	AP	F1-score
DeepWalk	0.812±0.008	0.798±0.009	0.761±0.010
Node2Vec	0.826±0.007	0.811±0.008	0.774±0.009
LINE	0.804±0.009	0.790±0.010	0.752±0.011
GCN	0.861±0.006	0.846±0.007	0.812±0.008
GraphSAGE	0.873±0.006	0.858±0.006	0.824±0.007
GAT	0.881±0.005	0.867±0.006	0.832±0.007
R-GCN	0.901±0.005	0.887±0.006	0.851±0.006
HAN	0.913±0.004	0.899±0.005	0.864±0.006
HGT	0.924±0.004	0.912±0.004	0.879±0.005
PGI-GRL	0.942±0.003	0.931±0.003	0.896±0.004

5.3 Ablation Study

The effectiveness of different components in the proposed PGI-GRL framework is further investigated through ablation experiments, and the results are presented in Table 3. The complete PGI-GRL model achieves the highest performance with an AUC of 0.942 ± 0.003, an AP of 0.931 ± 0.003, and an F1-score of 0.896 ± 0.004, confirming the effectiveness of jointly modeling semantic information, structural dependencies, and heterogeneous relationship characteristics for innovation relationship prediction. The relatively small standard deviations also indicate that the complete framework maintains stable optimization behavior across repeated experiments.

Removing semantic features decreases the AUC from 0.942 to 0.929, accompanied by reductions in AP from 0.931 to 0.916 and F1-score from 0.896 to 0.882. This degradation indicates that textual information extracted from scientific publications and patent documents provides complementary knowledge for distinguishing innovation elements with different technological meanings. Although graph structures can preserve relational patterns, semantic representations remain important for capturing conceptual differences among technologies, methods, and applications. The increased variation compared with the complete model (± 0.005 for AUC) further suggests that semantic information contributes to more robust representation learning.

The removal of structural features results in a further performance decline, reducing the AUC to 0.924 ± 0.005. This demonstrates that topology information plays an important role in modeling innovation dependencies, as relationships among technological entities are not solely determined by semantic similarity but also by their positions and interactions within the heterogeneous innovation graph. Without structural dependencies, the model loses important relational patterns among connected innovation elements.

The largest performance degradation is observed when relation-aware learning is removed, where AUC decreases to 0.911 ± 0.007, AP decreases to 0.897 ± 0.007, and F1-score decreases to 0.858 ± 0.008. This result verifies that explicitly modeling heterogeneous relationship types is critical for power grid innovation networks. Without relation-aware aggregation,

different interactions such as development, application, improvement, cooperation, and citation are treated similarly, which weakens the capability of the model to capture complex innovation evolution patterns.

Replacing the adaptive mechanism with fixed relation weighting reduces AUC from 0.942 to 0.917 ± 0.006. Although relation information is still preserved, the absence of dynamically learned relation importance limits the model's ability to adjust information propagation according to different graph contexts. These results demonstrate that the adaptive relation-aware mechanism provides essential improvements by enabling more flexible and discriminative representation learning for heterogeneous innovation relationship discovery.

Table 3. Ablation Study of the Proposed PGI-GRL Framework.

Variant	AUC	AP	F1-score
Full PGI-GRL	0.942±0.003	0.931±0.003	0.896±0.004
w/o Semantic	0.929±0.005	0.916±0.005	0.882±0.006
w/o Structural	0.924±0.005	0.910±0.006	0.876±0.006
w/o Relation-aware	0.911±0.007	0.897±0.007	0.858±0.008
w/o Adaptive	0.917±0.006	0.903±0.006	0.865±0.007

5.4 Temporal Innovation Prediction

5.5 Temporal Innovation Prediction

The temporal innovation relationship prediction results are presented in Table 4. Compared with the conventional link prediction setting, all methods exhibit performance degradation, demonstrating that predicting future innovation relationships is more challenging than reconstructing existing connections. This is mainly because temporal prediction requires models to capture evolving innovation interaction patterns rather than only relying on observed structural relationships.

Traditional network embedding approaches, including DeepWalk, Node2Vec, and LINE, achieve relatively limited performance, with AUC values of 0.764 ± 0.010, 0.781 ± 0.009, and 0.752 ± 0.011, respectively. These methods mainly preserve historical topological proximity and lack the capability to incorporate semantic evolution and heterogeneous interaction patterns, making them less effective for identifying emerging innovation relationships.

Graph neural network methods obtain improved results by leveraging neighborhood information. GCN, GraphSAGE, and GAT achieve AUC values of 0.823 ± 0.008, 0.837 ± 0.007, and 0.846 ± 0.007, respectively. These improvements indicate that graph-based message aggregation can provide more informative representations for temporal relationship prediction. However, these homogeneous graph learning methods still have limitations because they cannot explicitly distinguish the varying contributions of different innovation relationship types during technological evolution.

Heterogeneous graph learning methods further improve temporal prediction performance. R-GCN, HAN, and HGT achieve AUC values of 0.872 ± 0.006, 0.884 ± 0.006, and 0.897 ± 0.005, respectively. These results demonstrate that explicitly modeling multiple entity types and relationship categories contributes to capturing future innovation patterns. Nevertheless, these approaches mainly rely on predefined relation representations and lack adaptive mechanisms for dynamically adjusting the importance of different innovation interactions.

The proposed PGI-GRL achieves the best temporal prediction performance with an AUC of 0.918 ± 0.004, an AP of 0.906 ± 0.005, and a Recall@50 of 0.793 ± 0.006. Compared with the strongest baseline HGT, PGI-GRL improves AUC from 0.897 to 0.918, AP from 0.883 to 0.906, and Recall@50 from 0.768 to 0.793, corresponding to improvements of 2.1, 2.3, and 2.5 percentage points, respectively. The relatively small standard deviations indicate that the proposed framework maintains stable prediction capability under temporal evolution settings. The superior performance can be attributed to the adaptive relation-aware graph representation learning mechanism, which dynamically adjusts the contribution of different innovation relationships while jointly exploiting semantic information and structural dependencies. This design enables PGI-GRL to better capture evolving innovation interaction patterns and identify potential future technological connections from historical power grid innovation networks.

5.6 Hidden Innovation Relationship Discovery

Beyond quantitative relationship prediction evaluation, the practical capability of the proposed PGI-GRL framework is further investigated through hidden innovation relationship discovery. The objective of this experiment is to identify potential innovation relationships that are not explicitly observed in the existing innovation graph but may represent meaningful potential technology integration opportunities. Based on the predicted relationship scores generated by PGI-GRL, disconnected innovation element pairs are ranked according to their potential association probabilities. The top-ranked hidden innovation relationships are presented in Table 5. These candidate pairs achieve high prediction scores despite not being directly connected in the original graph, demonstrating the capability of the proposed framework to identify latent innovation associations.

Table 4. Performance Comparison on Temporal Innovation Relationship Prediction.

Method	AUC	AP	Recall@50
DeepWalk	0.764±0.010	0.741±0.011	0.612±0.012
Node2Vec	0.781±0.009	0.758±0.010	0.628±0.011
LINE	0.752±0.011	0.733±0.012	0.601±0.013
GCN	0.823±0.008	0.801±0.009	0.681±0.010
GraphSAGE	0.837±0.007	0.816±0.008	0.697±0.009
GAT	0.846±0.007	0.829±0.008	0.709±0.009
R-GCN	0.872±0.006	0.854±0.007	0.736±0.008
HAN	0.884±0.006	0.869±0.006	0.751±0.007
HGT	0.897±0.005	0.883±0.006	0.768±0.007
PGI-GRL	0.918±0.004	0.906±0.005	0.793±0.006

Table 5. Top-ranked Hidden Innovation Relationships Discovered by PGI-GRL.

Rank	Node 1	Node 2	Score
1	Digital Twin	AI Grid Optimization	0.931
2	Energy Storage	Virtual Power Plant	0.918
3	Renewable Forecasting	Intelligent Dispatching	0.904
4	Large Language Model	Grid Knowledge Management	0.887
5	Edge Computing	Smart Grid Monitoring	0.873

The discovered relationships reveal several potential technology integration patterns in the power grid domain. For example, the predicted associations between Digital Twin and AI Grid Optimization, Energy Storage and Virtual Power Plant, and Renewable Forecasting and Intelligent Dispatching reflect emerging trends in intelligent operation, distributed energy coordination, and data-driven power system management. In addition, the identified relationships involving Large Language Models and Grid Knowledge Management as well as Edge Computing and Smart Grid Monitoring indicate potential opportunities for integrating emerging computing technologies with power system applications. These results suggest that PGI-GRL can capture cross-domain innovation relationships that may not be easily identified through conventional text similarity or citation-based analysis.To provide an intuitive illustration of these discovered relationships, the top-ranked hidden innovation relationships are further visualized in Fig. 4.

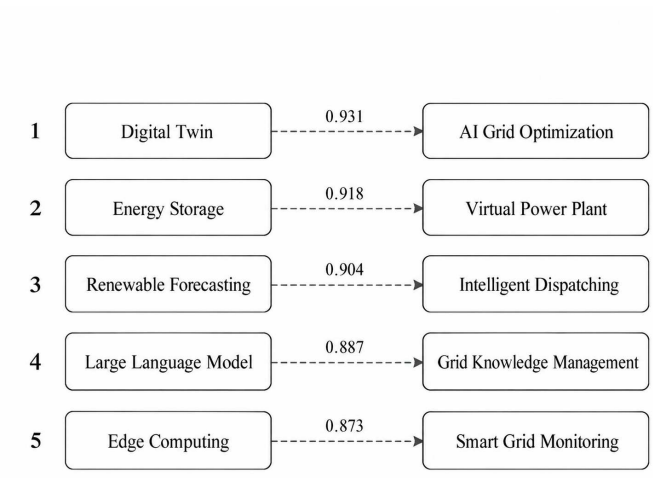


Figure 4. Visualization of top-ranked candidate innovation relationships predicted by the proposed PGI-GRL framework for hidden innovation relationship discovery.

The hidden innovation relationship discovery results demonstrate that the proposed framework extends innovation analysis from existing relationship modeling toward hidden innovation relationship discovery and potential technology exploration. By jointly learning semantic information, structural dependencies, and heterogeneous relationship characteristics, PGI-GRL provides a data-driven approach for identifying emerging innovation combinations and supporting strategic technology planning in the power grid domain.

6 Conclusion

This study proposed a Power Grid Innovation Graph Representation Learning (PGI-GRL) framework for discovering hidden innovation relationships among scientific and technological innovation elements in the power grid domain. By integrating scientific publications and patent documents, the proposed framework constructs a heterogeneous innovation graph and jointly models semantic features, structural dependencies, and heterogeneous innovation relationships through an adaptive relation-aware graph representation learning mechanism. Extensive experiments on relationship prediction, temporal relationship prediction, and hidden innovation relationship discovery tasks demonstrate that PGI-GRL achieves consistent improvements over representative graph embedding and graph neural network methods, indicating its effectiveness in capturing complex innovation interactions and identifying potential innovation relationships. The proposed framework provides a data-driven approach for discovering hidden innovation relationships and supporting technology evolution analysis and strategic innovation planning in the power grid domain. Future work will focus on incorporating larger-scale and continuously updated innovation data and enhancing model interpretability through the integration of domain expert knowledge.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability Statement

The processed heterogeneous innovation graph dataset generated in this study is available at: https://snippets.cacher.io/snippet/a87739a78a709a48f5ef The dataset includes innovation entity records, heterogeneous relationship information, and data processing validation files used for constructing the power grid innovation graph. The original scientific publications and patent documents were obtained from publicly available databases, including IEEE Xplore, Web of Science, Google Patents, and the China National Intellectual Property Administration (CNIPA), and are subject to the access policies of the respective data providers.

Author Contributions

CS conceived the study, designed the methodology, supervised the research process, and was responsible for manuscript preparation and revision. HZ contributed to data processing, experimental analysis, visualization, and manuscript drafting. GZ contributed to data acquisition, resource preparation, and experimental validation. LK contributed to literature investigation, result analysis, and manuscript refinement. All authors reviewed and approved the final manuscript.

Funding

No Funding

References

1. Purkait, P., Basu, M. & Nath, S. R. Renewable energy integration to electric power grid: Opportunities, challenges, and solutions. *Challenges opportunities distributed renewable power* 37–100 (2024).
2. Balamurugan, M., Narayanan, K., Raghu, N., Arjun Kumar, G. & Trupti, V. Role of artificial intelligence in smart grid–a mini review. *Front. Artif. Intell.* **8**, 1551661 (2025).
3. Ismail, F. B., Al-Faiz, H., Hasini, H., Al-Bazi, A. & Kazem, H. A. A comprehensive review of the dynamic applications of the digital twin technology across diverse energy sectors. *Energy Strateg. Rev.* **52**, 101334 (2024).
4. Pandey, U., Pathak, A., Kumar, A. & Mondal, S. Applications of artificial intelligence in power system operation, control and planning: a review. *Clean Energy* **7**, 1199–1218 (2023).
5. Choi, J., Lee, C. & Yoon, J. Exploring a technology ecology for technology opportunity discovery: A link prediction approach using heterogeneous knowledge graphs. *Technol. Forecast. Soc. Chang.* **186**, 122161 (2023).
6. Arun, M. *et al.* Integration of energy storage systems and grid modernization for reliable urban power management toward future energy sustainability. *J. energy storage* **114**, 115830 (2025).
7. Verbeek, A. *et al.* Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics* **54**, 399–420 (2002).

8. Zhou, X., Wang, J., Wang, X. & Chen, S. Optimal dispatch of integrated energy system based on deep reinforcement learning. *Energy Reports* **9**, 373–378 (2023).

9. Fawzy, S. M., M. Ali, M. K. & Allam, N. K. Artificial intelligence-driven materials design for next-generation sustainable energy technologies. *ACS Sustain. Chem. & Eng.* **14**, 4745–4761 (2026).

10. Mchirgui, N., Quadar, N., Kraiem, H. & Lakhssassi, A. The applications and challenges of digital twin technology in smart grids: A comprehensive review. *Appl. Sci.* **14**, 10933 (2024).

11. Hu, Y. *et al.* Construction and application of conditional event-based knowledge graphs for electric power standards. *Expert. Syst. with Appl.* 129291 (2025).

12. Cheng, L. *et al.* Real-time dispatch via expert knowledge driven deep reinforcement learning. *CSEE J. Power Energy Syst.* (2023).

13. Ren, H., Zhang, L. & Wang, Q. A general methodology for technology opportunity discovery based on opportunity evaluation and optimization. *IEEE Transactions on Eng. Manag.* **71**, 6725–6740 (2023).

14. Afzal, M. A., Rahman, S. U. & Aslam, M. T. Mapping the nexus of sustainability, innovation, and renewable energy: a bibliometric analysis of green technology research. *The Critical Rev. Soc. Sci. Stud.* **3**, 1176–1200 (2025).

15. Rajabi, E. & Etminani, K. Knowledge-graph-based explainable ai: A systematic review. *J. information science* **50**, 1019–1029 (2024).

16. Li, K. *et al.* Knowledge graph-based multimodal neural networks for smart-grid defect detection. *Eng. Reports* **6**, e12824 (2024).

17. Khaleel, M. *et al.* Impact of smart grid technologies on sustainable urban development. *Int. J. Electr. Eng. Sustain.* **2**, 62–82 (2024).

18. Lee, J., Park, S. & Lee, J. A study on the monitoring of technology innovation through patent analysis. *Humanit. Soc. Sci. Commun.* **13**, 70 (2025).

19. Qin, T., Su, J. & Wu, C. Hegformer: representation learning on heterogeneous graph neural network with efficient transformer. In *2024 IEEE International Conference on Knowledge Graph (ICKG)*, 266–273 (IEEE, 2024).

20. Hu, Z., Dong, Y., Wang, K. & Sun, Y. Heterogeneous graph transformer. In *Proceedings of the web conference 2020*, 2704–2710 (2020).

21. Ye, Z., Kumar, Y. J., Sing, G. O., Song, F. & Wang, J. A comprehensive survey of graph neural networks for knowledge graphs. *Ieee Access* **10**, 75729–75741 (2022).

22. Chen, C. *et al.* A survey on graph neural networks and graph transformers in computer vision: A task-oriented perspective. *IEEE Transactions on Pattern Analysis Mach. Intell.* **46**, 10297–10318 (2024).

23. Dwivedi, V. P. *et al.* Relational graph transformer. In *International Conference on Learning Representations*, vol. 2026, 154348–154369 (2026).

24. Hu, J. *et al.* Heterogeneous graph neural networks for link prediction in biomedical networks. *Bioinforma. advances* **5**, vbaf187 (2025).

25. Zheng, X. *et al.* Graph neural networks for graphs with heterophily: A survey. *IEEE Transactions on Knowl. Data Eng.* (2026).

26. Sun, C. *et al.* A survey on gnn-based link prediction: Techniques, applications, and challenges. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **16**, e70093 (2026).

27. Zheng, Y., Yi, L. & Wei, Z. A survey of dynamic graph neural networks. *Front. Comput. Sci.* **19**, 196323 (2025).

28. Li, G., Siddharth, L. & Luo, J. Embedding knowledge graph of patent metadata to measure knowledge proximity. *J. Assoc. for Inf. Sci. Technol.* **74**, 476–490 (2023).

29. Ju, W. *et al.* Towards graph contrastive learning: A survey and beyond. *arXiv preprint arXiv:2405.11868* (2024).

30. Yang, Z. *et al.* Hop-based heterogeneous graph transformer. In *ECAI 2024: 27th European Conference on Artificial Intelligence, 19–24 October 2024, Santiago de Compostela, Spain–Including 13th Conference on Prestigious Applications of Intelligent Systems (PAIS 2024)*, 2346–2353 (SAGE Publications Pvt. Ltd 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP, 2024).

31. Thomas, J. M., Moallemy-Oureh, A., Beddar-Wiesing, S. & Holzhüter, C. Graph neural networks designed for different graph types: A survey. *arXiv preprint arXiv:2204.03080* (2022).

32. Huo, G. *et al.* Hierarchical spatio–temporal graph convolutional networks and transformer network for traffic flow forecasting. *IEEE Transactions on Intell. Transp. Syst.* **24**, 3855–3867 (2023).

33. Fang, Y. *et al.* Relation-aware graph convolutional networks for multi-relational network alignment. *ACM Transactions on Intell. Syst. Technol.* **14**, 1–23 (2023).

34. Sun, Y., Wang, Y., Chen, Y. & Zhao, J. Hierarchical semantic-aware heterogeneous graph transformer model. In *2026 International Conference on Digital Analysis and Processing, Intelligent Computation (DAPIC)*, 32–36 (The Academy of Engineering and Education, 2026).

35. Yu, L. *et al.* Heterogeneous graph representation learning with relation awareness. *IEEE Transactions on Knowl. Data Eng.* **35**, 5935–5947 (2022).

36. Li, X., Tian, Y. & Ji, S. Semantic-and relation-based graph neural network for knowledge graph completion: X li et al. *Appl. Intell.* **54**, 6085–6107 (2024).

37. Wang, Y. *et al.* Efficient link prediction via gnn layers induced by negative sampling (2024).

38. Liu, H. *et al.* Todynet: temporal dynamic graph neural network for multivariate time series classification. *Inf. Sci.* **677**, 120914 (2024).

39. Zheng, C. *et al.* Bytegnn: efficient graph neural network training at large scale. *Proc. VLDB Endow.* **15**, 1228–1242 (2022).