Industry Chain Visual Exploration Based on Heterogeneous Graphs Analysis

Yi Wan, Xiaowen Zhang, Xiaoxu Chen, Jiacheng Tang, and Siming Chen

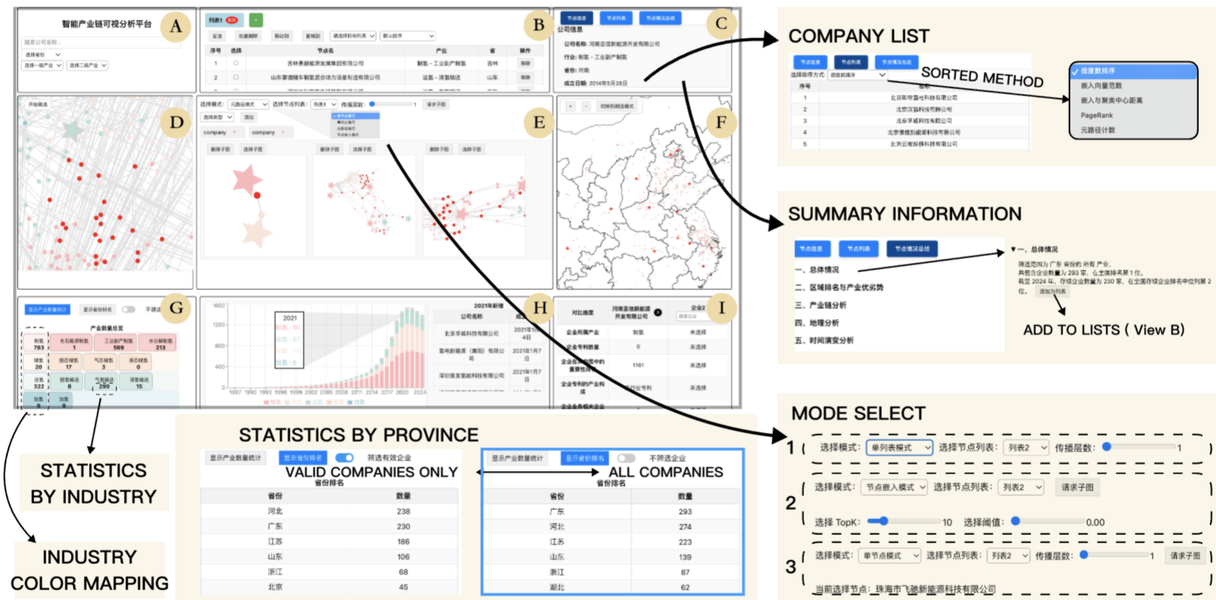

Fig. 1: The system interface of our system: (A)Company search and filtering, allowing users to filter firms by region and industry or search by name; (B) Company list management, enabling customized lists with batch operations and multi-dimensional sorting; (C) Node detail and selection, providing company profiles, sorted node lists, and summary templates; (D)Global force-directed graph view, visualizing company-industry relationships with interactive exploration; (E) Subgraph exploration, supporting multiple extraction strategies for comparative analysis; (F) Geographic map view, displaying company locations with interactive selection tools; (G)Industry and region statistics, presenting company distributions both by industry and by province, and showing industry color mapping in the whole system; (H)Time axis view, visualizing yearly company growth trends and supporting actions like hover and click for details;(I) Competitive comparison, enabling side-by-side firm analysis across multiple dimensions.

Abstract—The analysis of industry chains in emerging industries faces challenges due to the complexity of industrial structures and the limitations of traditional analytical methods. In this paper, we introduce a heterogeneous graph-based visual analytics approach for industry chain analysis. By modeling enterprises and industries with multi-type nodes and edges, we apply graph embedding and subgraph mining to identify key enterprises and their relationships. Additionally, we design an interactive visualization framework for multi-level and dynamic exploration. To validate our approach, we conduct a case study on the hydrogen energy industry, demonstrating the system's effectiveness in enhancing data coverage, visualization clarity, and analytical efficiency for strategic decision-making.

Index Terms—Industry Chain Analysis, Heterogeneous Graph Visualization, Interactive Visual Exploration

1 Introduction

Industry chains represent critical structures within modern economic systems, enabling efficient resource allocation, coordinated development, and sustained value creation across interconnected sectors. The resilience and structural integrity of industry chains significantly influence innovation capacity, economic growth, and competitiveness. Consequently, analyzing industry chain has emerged as a prominent research topic, drawing substantial attention from academia, industry, and policymakers.

Recently, rapid technological development and the emergence of new industries have dramatically reshaped global industrial landscapes. Small and medium-sized enterprises (SMEs), characterized by dynamic market activities, such as frequent entry and exit events and loosely connected business relationships, are becoming increasingly central. Traditional analytical methodologies, which are typically reliant on static expert surveys, structured reports, and manually curated datasets, struggle to meet the evolving demands of modern industry chain analysis due to limited scalability, time-consuming processes, and inadequate responsiveness to dynamic changes.

While recent studies have leveraged visualization and graph-based analytical techniques to facilitate industry analysis, most existing frame-

- *Yi Wan is with Fudan University. E-mail: 23210980075@m.fudan.edu.cn*
- *Xiaowen Zhang is with Fudan University.*
- *Xiaoxu Chen is with Zhejiang Zhelixin Credit Information Co., Ltd.*
- *Jiacheng Tang is with Zhejiang Zhelixin Credit Information Co., Ltd.*
- *Siming Chen is with Fudan University. E-mail: simingchen3@gmail.com*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

works operate predominantly at aggregated or macro levels, focusing largely on sectoral classifications or regional clusters instead of enterprises. Furthermore, these methods often depend heavily on highly structured datasets, reducing their effectiveness in industries characterized by sparse, heterogeneous, or unstructured data.

Accordingly, there remains a critical need for a framework that can integrate diverse information sources and dynamically reflect the structural evolution of industry chains. To satisfy the need, this paper proposes a scalable, fine-grained visual analytics framework leveraging heterogeneous graph structures to capture and dynamically represent industry chain evolution. Our approach integrates diverse data sources—such as business registration records and patent databases—to automatically derive enterprise and industry linkages, resulting in a robust heterogeneous graph representation. Through advanced methods such as heterogeneous graph embeddings, targeted subgraph extraction, and interactive visual exploration, our framework effectively identifies key enterprises, reveals complex industry structures, and tracks temporal changes in a scalable and intuitive manner.

We validate the efficacy and practicality of the proposed visual analytics framework through two detailed case studies within the hydrogen energy sector. These case studies illustrate the framework's capability from both governmental planning and enterprise operational perspectives, underscoring its utility and adaptability.

In summary, this paper contributes to the visual analytics community by:

- Integrating innovative graph visualization methods into the domain of industry chain analysis, thereby enhancing the analytical precision and interpretability of complex industry structures;
- Developing a systematic analytical workflow based on heterogeneous graphs, offering scalable, automated methodologies for the dynamic construction and exploration of intricate industry chains;
- Implementing an interactive visual analytics system that combines data processing, advanced graph analytics, and real-time visualization, providing government agencies, enterprises, and investors with a powerful decision-making tool capable of addressing contemporary analytical challenges.

2 Related Works

2.1 Industry Analysis

Industry analysis is a fundamental aspect of macroeconomic research, encompassing various subfields such as performance evaluation, structural analysis, and policy assessment. Traditional methods typically rely on quantitative indicators and structured data, using static models to measure industry efficiency, concentration, or competitiveness [5, 10, 26]. While such approaches have laid the foundation for empirical industry studies, they often require domain expertise and lack flexibility in exploratory analysis.

Recent advances in network science have introduced graph-based models into industry chain analysis. In these models, enterprises are represented as nodes, and transactional, collaborative, or ownership relationships are modeled as edges. This abstraction supports tasks such as supplier recommendation and link prediction using graph neural networks or collaborative filtering techniques [17, 20, 28]. However, these methods primarily focus on algorithmic output and often provide limited interpretability for non-expert users.

To improve transparency and usability, visual analytics techniques have been adopted in industry-related systems. For instance, V4RIN [34] and RISeer [2] use map-based visualizations to show spatial distributions and evolutionary patterns of industrial clusters, combining interactive views with temporal or spatial encoding. Other systems, such as Hermes [19] and VIEA [33], explore economic flows across sectors and countries, integrating dashboards and flow-based network views to support exploration of industrial dynamics over time. These tools highlight the benefits of interactive, multi-view visualization in supporting exploratory analysis and decision-making.

Despite these advances, most existing systems focus on macro-level structures—such as regional aggregation or sector clustering—rather than enterprise-level relationships. Few studies explicitly model fine-grained, dynamic interactions between individual companies. One exception is the work of Arleo et al. [1], who constructed a weighted directed transaction network between firms based on multiple financial data sources and visualized it in a 3D geographic interface. However, such systems often rely on high-quality, proprietary data and are not designed to support analysis within specific industries using publicly accessible sources.

This study addresses these limitations by proposing a heterogeneous graph-based visual analytics framework focused on a specific industry. Built upon open enterprise data, our framework models inter-company relationships in a fine-grained manner and supports multi-scale visual exploration. In contrast to existing approaches, it enables users to analyze industry structures at the enterprise level, trace linkages within the supply chain, and interactively explore trends and key entities. This design aims to serve a broad range of users—including policymakers, business analysts, and investors—who may not have technical backgrounds in economics or graph theory but require interpretable, data-driven insights into complex industry networks.

2.2 Heterogeneous Graph Representation Learning

Heterogeneous graphs consist of multiple types of nodes and edges, and are widely used to model complex relational structures in domains such as social networks, citation graphs, e-commerce, and biological systems [27]. Compared with homogeneous graphs, heterogeneous graphs carry richer semantic information, which poses additional challenges for representation learning [25].

Early graph neural networks (GNNs) focused primarily on homogeneous graphs and often ignored node and edge type distinctions, limiting their ability to capture semantic heterogeneity. To address this, various heterogeneous graph representation methods have been proposed.

One major line of research employs attention-based mechanisms to model type-aware interactions. Models such as HGT [15], Simple-HGN [21], and HetSANN [14] extend the Transformer architecture to assign type-specific attention weights to different node and edge types, effectively capturing semantic dependencies while maintaining model scalability. Extensions like GAT [29] and SGAT [18] also use attention to capture both local and higher-order structures.

Another line of work leverages meta-paths—composite paths defined by node and edge types—to model semantic relations. Representative methods such as HAN [30] and MAGNN [9] aggregate information across meta-paths using hierarchical attention mechanisms. To reduce reliance on predefined meta-paths, models like HIN2Vec [8] and Meta-GNN [35] use random walks and path sampling to learn representations in a more flexible and adaptive manner.

In this study, we adopt the Heterogeneous Graph Transformer (HGT) due to its strong ability to model complex multi-type structures without requiring manual meta-path design. HGT captures both structural and semantic relationships through a unified attention framework, making it well-suited for modeling the diverse node and edge types in industry chain networks. Moreover, the concept of meta-paths also inspires our subgraph extraction strategy, enabling structural and semantic pattern-based exploration of specific industry segments.

2.3 Graph Visualization

Graph visualization is a key technique for improving the interpretability of network data. Among various methods, force-directed layouts are widely used in node-link diagrams, simulating physical forces to produce spatial arrangements of nodes and edges [11]. Variants such as Kamada-Kawai [16] and Fruchterman-Reingold [7] provide different optimization strategies for layout generation. However, as the size of the network increases, these layouts often suffer from edge clutter and reduced readability, even with GPU acceleration [6].

To address these challenges, researchers have proposed several strategies to improve visual clarity. Clustering-based abstraction can reduce visual complexity by aggregating nodes and visualizing inter-cluster links [4], although this may obscure intra-cluster relationships. Interactive techniques such as edge lensing and edge plucking allow users to locally disentangle overlapping edges [32], while edge bundling reduces clutter by routing similar edges together [13].

Beyond layout optimization, effective information retrieval remains a critical challenge in large-scale graph exploration. Subgraph querying enables users to focus on regions of interest. Techniques include feature-based neighborhood search [24] and similarity-based subgraph retrieval [23]. Filtering methods can simplify the graph by removing less relevant nodes or edges, though this risks altering the underlying topology [12, 31]. Navigation techniques such as topology-aware transitions [22] and visual guidance systems like VERTIGo [3] further support scalable graph exploration.

In this study, we adopt a force-directed layout to visualize the overall structure of the industry chain network, while selectively displaying only key node and edge types to minimize visual noise. Interactive features such as zooming, dragging, and semantic filtering are supported to enhance exploratory analysis. In addition, we incorporate flexible subgraph querying mechanisms to help users focus on specific segments, enabling both overview and detail-oriented exploration of complex industry structures.

3 METHODOLOGY

This section presents the overall visual analytics pipeline, including heterogeneous graph construction, subgraph extraction strategies, and interactive visual exploration. As shown in Fig. 2, the system is built around a heterogeneous graph model that integrates multi-source enterprise data. The resulting graph supports flexible subgraph extraction and interactive visualization, enabling detailed industry chain analysis at various levels of granularity.

3.1 Heterogeneous Graph Construction

Data Source. Our system uses publicly available enterprise data as input, including business registration information, patent records, and competitor relationships. In this study, the hydrogen energy industry—a strategic and emerging sector in China—was selected as the case domain. This industry features a large number of small and medium-sized enterprises (SMEs) and limited existing structured chain data, making it an ideal scenario for validating our framework. The dataset consists of 1,719 companies, 18,833 patents, and 7,965 competitor records obtained from a regional data platform.

Node Construction. Three types of nodes are defined: companies, patents, and industry categories. (1) *Industry nodes* are derived from the official technical classification published by the China National Intellectual Property Administration, which categorizes hydrogen energy into 5 first-level sub-sectors and 14 second-level sub-sectors (e.g., hydrogen production, storage, transportation, and usage). (2) *Company nodes* include registration details. Specifically, the company's survival status is determined by its registration date. And the company's registered address is converted into latitude and longitude data using Amap APIs to support subsequent spatial analysis. (3) *Patent nodes* include only non-design patents in an authorized state, which accurately reflect a company's innovation capability. These patents are deduplicated, resulting in a total of 11,727 unique patents.

Edge Construction. Among these three types of nodes, four types of semantic relationships (edges) are extracted:

Company–Industry edges are constructed by computing the semantic similarity between a company's business scope and the textual definitions of industry categories using the M3E embedding model. Cosine similarity is used to match each company to its most relevant industry.

Company–Company edges are created based on the Jaccard similarity of business descriptions, using tokenized word sets obtained through Jieba segmentation to establish links among companies with similar operations.

Company–Patent edges link companies to patents they hold, including cases with multiple assignees.

Patent–Industry edges are established using IPC codes mapped to industry categories based on official classification standards. A patent may link to multiple industries when mappings are ambiguous.

Graph Construction. With the node and link data structured from the input data, a graph is ready to construct. All edges are treated as undirected and bidirectional for model compatibility and network completeness, resulting in eight total edge types. The final heterogeneous graph includes three node types and four edge types, forming a rich semantic structure for downstream analysis.

Graph Embedding. To support similarity-based tasks and semantic subgraph extraction, we apply the Heterogeneous Graph Transformer (HGT) to learn node embeddings. HGT incorporates type-aware attention mechanisms for nodes and edges, enabling it to model the semantic dependencies in multi-relational, multi-typed networks. In each layer of the HGT [15], the process begins with message passing, where each node computes a type-weighted representation of itself. Next, mutual attention is applied to calculate the weighted importance of the types of connected nodes and the edges between them. These weighted features are then aggregated, followed by an additional aggregation of the weighted features from the neighboring nodes. This process results in the updated embedding for the node at that layer. This procedure is repeated across multiple layers, allowing the model to progressively refine the node embeddings by incorporating more complex interactions and higher-order dependencies from the graph structure.

3.2 Subgraph Extraction Strategies

Visualizing the entire graph may introduce visual clutter due to the number of nodes and edges. To address this, we provide three subgraph extraction strategies tailored to different user scenarios, allowing users to extract the subgraph of interest for detailed information.

k-hop Neighborhood Expansion. This method extracts a local subgraph centered around one or more target company nodes, expanding up to a user-defined number of hops while excluding high-degree industry nodes to maintain scalability. Related industry nodes are re-inserted post-expansion to retain semantic context. This approach is particularly useful for users who want to explore the local neighborhood of a target company, providing insights into the company's positioning. It can reveal how closely the target company is linked to others, whether there are connections with companies from different industries, and the company's patent holdings, among other details.

Meta-path Guided Extraction. For users with specific analytical goals, such as tracing upstream to downstream firms, this method defines meta-path patterns based on node types and attributes (e.g., industry level or region). The system first constructs an expanded neighborhood via k-hop expansion and then extracts valid paths matching the meta-path constraints. This enables flexible, pattern-based subgraph queries. This strategy is ideal for users who wish to identify and examine specific connection patterns in the network, such as following the flow from upstream to downstream firms in a supply chain or understanding the relationship between companies in different provinces to study inter-provincial collaboration.

Embedding Similarity-Based Extraction. When users lack a specific exploration pattern, this method recommends a semantically related subgraph based on node embeddings. Cosine similarity between a selected node and its peers is computed, and top-K similar nodes (above a defined threshold) are selected. Their neighbors are then included to enrich semantic context, resulting in a compact yet meaningful subgraph for further analysis. This method is especially useful when users want to expand their exploration from a few specific nodes to discover semantically related nodes, providing a way to explore interconnected nodes when there is no clear exploration pattern defined.

Sorting Strategies. To enhance the representation of extracted companies, six sorting strategies are implemented: insertion order, embedding norm, distance to cluster centroid, node degree, PageRank, and meta-path count(see Tab. 1). These methods facilitate a comprehensive exploration of company characteristics and importance among the extracted nodes. Spanning temporal, attribute-based, and structural dimensions, these sorting strategies enable deeper node filtering across diverse analytical contexts.

- Insertion Order: Nodes are sorted chronologically based on their addition to the list. If a node is added multiple times, it is ranked according to the most recent addition. This method is suitable for tracing exploratory paths and analyzing node inclusion sequences.

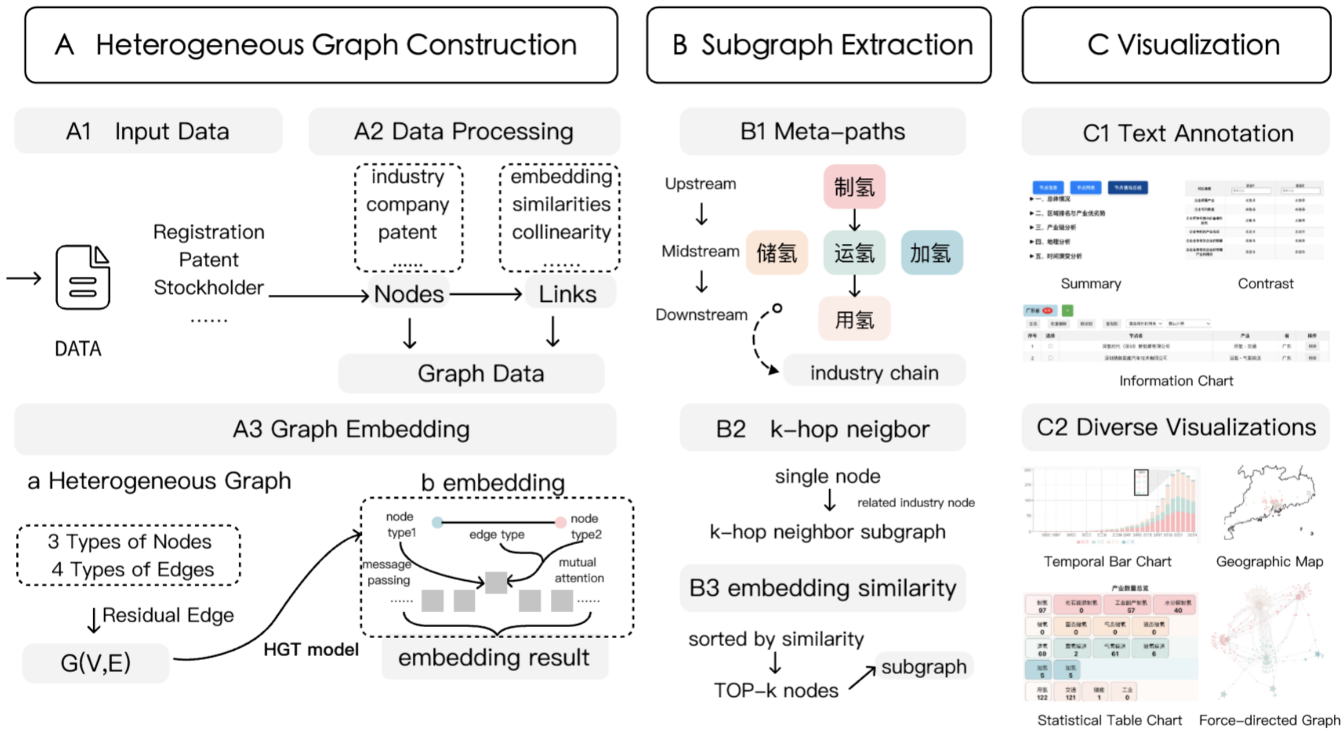

Fig. 2: The proposed pipeline processes multi-source enterprise data (A) by first (a1) integrating business registrations and company profiles to construct a heterogeneous graph. It then (a2) extracts semantic relationships to define nodes and edges, followed by (a3) encoding type-aware embeddings using Heterogeneous Graph Transformers (HGT). To facilitate graph exploration, (B) three subgraph extraction strategies are provided: (b1) meta-path guided extraction for industry chain analysis, (b2) k-hop neighborhood expansion, and (b3) similarity-based node ranking. Finally, (C) a visual analytics interface presents the extracted insights, combining (c1) interactive text annotations with (c2) coordinated multi-view visualizations for comprehensive industry chain analysis.

Table 1: Comparison of Six Ranking Methods.

Method	**Ranking Basis**	**Applicable Scenario**
Default Order	Node insertion time (ascending)	Exploration path tracing
Embedding Norm	Node embedding L_2 norm (descending)	Feature strength evaluation
Distance to Cluster Center	Euclidean distance to centroid (descending)	Anomaly detection, feature divergence analysis
Node Degree	Sum of in-degree and out-degree (descending)	Influence analysis
PageRank	PageRank score (descending)	Network influence analysis
Meta-path Count	Frequency of matching a given meta-path (descending)	Influence analysis in industrial supply chains

- Embedding Norm: Nodes are ranked by the L2 norm of their embedding vectors in descending order. Nodes with higher norms are considered to exhibit stronger feature intensity, aiding the identification of nodes with prominent characteristics in the embedding space.

- Distance to Cluster Centroid: Nodes are sorted by their Euclidean distance to the centroid of their assigned cluster, from largest to smallest. This approach helps identify outlier nodes or those deviating from cluster centers, allowing for focused analysis or exclusion of anomalous nodes.

- Node Degree: Nodes are ranked based on the sum of their in-degree and out-degree. Higher-degree nodes are more connected and likely to have greater relational significance within the network.

- PageRank: Based on the classic PageRank algorithm, nodes are assigned importance scores reflecting their global influence within the industrial graph. Sorting is performed in descending order of these scores.

- Meta-Path Count: Nodes are ranked by the frequency with which they participate in predefined meta-paths, specifically designed to reflect upstream-midstream-downstream industry relations, such as "hydrogen production company – storage/transport/refueling company – hydrogen utilization company." This method highlights a company's influence across the industrial value chain.

These multidimensional sorting mechanisms empower users to flexibly filter and analyze companies according to specific analytical goals by selecting appropriate methods based on their use cases and preferences, thereby enabling the efficient identification of both high- and low-ranking nodes and facilitating the targeted discovery of companies with particular attributes or strategic significance.

3.3 Visual Analytics Design

To support interactive analysis of complex subgraphs, we designed a multi-level visual analytics system with tightly integrated visual and textual representations.

Visualization Components. The system is primarily centered around graph visualization, complemented by various additional forms to better present the complex information and data relationships in the heterogeneous graph. In the design of graph visualization, the presentation of nodes and edges has been carefully optimized, with particular attention to shape and color of the nodes, to enhance the distinguishability of different node types and improve the overall readability of the graph. For example, industry nodes are represented by pentagons, company nodes by circles, and nodes representing different industry

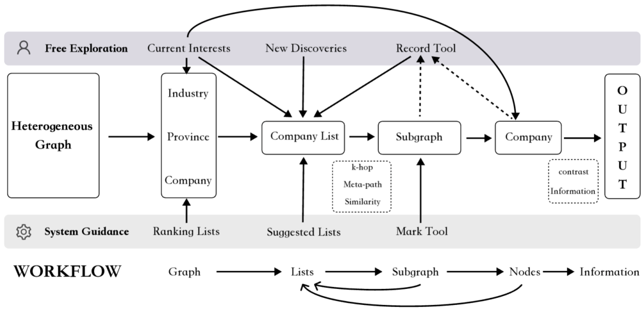

Fig. 3: the proposed interactive pipeline is to extract company lists from the global graph, generate subgraphs focused on the lists, discover key nodes within the subgraphs, delve into node information, and continuously update the list based on findings to obtain new subgraphs and key nodes, iterating progressively. Depending on the user, two distinct implementation paths are designed. In free exploration, the user starts from areas of current interest to make new discoveries and record them. In guided exploration, the user begins with ranked lists, receives suggested lists of focus, and uses the system's mark tools to identify important nodes within the subgraph.

levels are colored differently, consistent with the overall color mapping scheme.

Regarding node layout, a force-directed graph algorithm is employed, which simulates the repulsive forces between nodes and the attractive forces between edges. This iterative process adjusts the positions of nodes until a stable configuration is achieved, resulting in a clear and stable network structure [7]. This layout method effectively optimizes the spatial distribution of nodes, grouping similar nodes together while maintaining appropriate distances between distant nodes, thereby improving the visual effectiveness of the graph.

In addition to the graph layout, this system also designs several supplementary views to accomplish different analytical tasks. For parts of the data that are more complex or dense, textual annotations are used to help users quickly grasp key information. For instance, textual summaries are provided to describe the overall characteristics of the whole graph, comparisons of node attributes are displayed in tabular form, and key node information is listed to ensure users can quickly locate and explore specific nodes.

For data that can be visually presented, diverse visualizations are designed to enhance the user experience. For example, stacked bar charts are used to show the temporal changes in the number of companies in each industry, with the x-axis representing time and the y-axis representing quantity, while different colors of bars represent different industries. Geographical distribution maps are used to display the spatial distribution of companies, with company locations marked as points on the map, reflecting the concentration of companies in various regions. Industry statistical tables present categorization and quantity data for different industries, with color coding to distinguish between industry types and labels indicating the number of companies in each category. Additionally, association force-directed graphs visualize the closeness of relationships between nodes, where node distances reflect the strength of their relationships, and key nodes are highlighted using color and size for easy identification.

Interactive Exploration. To enable effective exploration of large-scale heterogeneous graphs, we designed an interactive analytical methodology that systematically integrates multiple visualization components. As illustrated in Fig. 3, our solution establishes a bidirectional exploration mechanism between textual and graphical representations, supporting progressive analytical refinement through three critical phases: (1) global overview establishment(graph), (2) local pattern identification(list and subgraph), and (3) individual node analysis.The system incorporates two complementary interaction paradigms tailored to different user expertise levels.

The *free exploration* mode is designed for expert users who can directly manipulate visual elements, such as selecting nodes in force-directed graphs or filtering regions in geospatial views. Users iteratively refine company lists, generate subgraphs, and analyze node details, dynamically adjusting their exploration focus based on emerging patterns. This approach allows for flexible, in-depth analysis tailored to specific research needs.

The *system guidance* mode assists general users by providing algorithmically ranked lists and annotated visual summaries. Users can start with industry rankings, regional distributions, or clustered companies, progressively refining their focus through structured prompts and visual guidance. This guided approach simplifies navigation through complex networks, making large-scale graph analysis more accessible to non-expert users.

In summary, our methodology combines scalable heterogeneous graph modeling, flexible subgraph extraction, and rich interactive visualization to support multi-level industry chain analysis. The system is designed to accommodate diverse user needs, from policy analysts and business managers to technical researchers, providing them with interpretable, actionable insights into complex industrial ecosystems.

4 System Design

This section presents the design of our visual analytics system for industry chain analysis. We first outline the key challenges in analyzing emerging industry ecosystems, followed by the corresponding design objectives. Then, we describe the structure and functions of the system components, highlighting how each view addresses specific analytical needs.

4.1 Design Challenges and Objectives

Analyzing industry chains in emerging sectors—such as hydrogen energy—poses unique difficulties due to the dominance of small and medium-sized enterprises (SMEs) and the lack of well-structured data. Our system addresses the following key challenges:

C1. Data processing and graph construction. Enterprise-related data is often fragmented, limited in coverage, and dominated by unstructured text. Constructing a meaningful graph structure from such data requires robust information extraction, semantic matching, and integration from heterogeneous sources.

C2. Large-scale graph visualization. Industry chains involve a large number of companies, each with numerous attributes such as patents and business domains. Visualizing both global structures and local relationships in a single coherent view remains challenging.

C3. Exploratory path selection. Users without technical expertise often struggle to determine starting points or exploration directions within complex heterogeneous graphs. Exploration frequently involves multiple steps, making it difficult to track history and maintain context.

Based on these challenges, we define the following system design objectives:

T1. Efficient data processing and graph modeling. The system must support automatic extraction of entities and relationships from diverse data sources (e.g., business registrations, patents), using natural language processing (NLP) and similarity computation to construct semantically meaningful heterogeneous graphs.

T2. Multi-scale visualization and interaction. To balance overview and detail, the system should allow users to explore both global industry structures and fine-grained company-level information. Flexible interactions and scalable layouts are required to support smooth transitions across scales.

T3. Guided exploration paths. The system should assist users in identifying starting points and reasonable expansion paths, especially in large graphs. This reduces exploration complexity and improves insight discovery efficiency.

T4. Exploration memory and record-keeping. To support multi-step analysis, the system should provide mechanisms for saving intermediate results and maintaining logical continuity across exploration stages.

4.2 System Overview and Architecture

The system is implemented as a web-based visual analytics platform, supporting the complete analytical pipeline in Fig. 2 from raw data

Table 2: Heterogeneous Graph Structure.

Node Type	Node Count	Edge Type	Edge Count
Industry	14	Industry-Company	1,719
Company	1,719	Company-Company	1,719
Patent	11,727	Company-Patent	5,494
		Patent-Industry	9,332
Total	**13,460**		**18,264**

acquisition to heterogeneous graph construction, subgraph extraction, and interactive visualization. Aligned with the interaction pipeline described in Fig. 3, the system ensures usability and effectively fulfills the design objectives.

The system design is illustrated in Fig. 1. The interface is built upon a graph model constructed from hydrogen industry data, consisting of three node types (industry, company, patent) and four edge types (see Tab. 2). The system provides nine coordinated views, enabling multi-level exploration from macro-level industry structures to individual companies.

Each view contributes to different aspects of analysis. Below we describe the design and functionality of selected core modules.

(A) Company Search and Filtering. The company search and filtering view offers three dimensions for selection: company, industry, and province. Users can define filtering criteria by selecting a province, primary industry, and secondary industry through dropdown menus. Selected criteria are displayed as tags next to the selection box, enabling users to clearly identify the current filter state and efficiently remove individual conditions as needed. The filtering system supports multi-selection. Once the filtering criteria are set, related views—including (C) the node detail and selection view, (D) the force-directed graph, (F) the map distribution view, (G) the industry and province distribution statistics view, and (H) the timeline view—are synchronously updated to reflect only the filtered companies and their associations. Statistical and temporal data are likewise recalculated based on the filtered results. This filtering mechanism enables more precise analysis of industry linkages and relevant information.

(B) Company List Management. Users can create multiple company lists for targeted analysis and precise node management. They can either add nodes individually or in bulk from other views or perform batch operations like moving items between lists. Furthermore, custom naming enables clear differentiation of list purposes. In addition, six sorting methods, including default sorting, embedding-based sorting, distance to cluster center, node degree, PageRank, and meta-path counting, help users comprehensively analyze company characteristics and significance within each list.

(C) Node Detail and Selection. This view consists of three components: Node Detail, Company List, and Summary Information. The Node Detail View displays comprehensive information about the most recently selected node.The Node List View presents a sorted list of nodes based on a specified ranking method, facilitating quick identification of target nodes. Its sorting options align with those in the (B) Company List Management View, except that the default sorting here is by node degree rather than insertion order. The Node Summary View provides multi-dimensional statistical insights based on the current province and industry filters, following a standardized template. Users can directly add the analyzed company list to (B) Company List Management View for further study. The summary covers five key aspects: (1) Overall Analysis for a macro-level overview, (2) Regional Ranking and Industry Strengths/Weaknesses to reveal competitive positioning, (3) Industry Chain Analysis for upstream and downstream relationships, (4) Geographical Analysis for spatial distribution patterns, and (5) Temporal Evolution Analysis to track industry dynamics. These combined insights help users efficiently identify key enterprise clusters and understand regional industry characteristics.

(D) Global Force-Directed Graph View. This view provides a comprehensive visualization of the relationships between filtered companies and industries based on physical simulation principles, where nodes are modeled as repelling charges and edges simulate attractive elastic forces. By carefully tuning edge lengths and repulsive parameters, the layout achieves a balance between visual clarity and structural compactness, effectively minimizing node overlap while preserving readability. Node attributes are distinguished through variations in shape and color. Users can interact with the graph by hovering to access detailed information or dragging nodes to refine local layouts for improved interpretability. This view enables users to observe clustering patterns and association strengths, identify potential core nodes or highly connected groups, and record selected nodes for further exploration. Overall, it offers a global perspective on the heterogeneous structure of company–industry relations, facilitating intuitive understanding of network distributions and revealing potential entry points for deeper analysis.

(E) Subgraph Exploration. This view enables in-depth analysis of companies listed in the (B) Company List Management View by generating subgraphs from the overall heterogeneous network, revealing association patterns with specific node types and information such as patent ownership. This view consists of two components: parameter configuration and subgraph visualization. Users can select from four modes—Single Node, Single List, Meta-Path, and Node Embedding—each requiring distinct parameter inputs. In Single Node mode, the most recently added node in the selected list is used as the focal node, and a user-defined number of k-hop neighbors are included in the subgraph. In Single List mode, all nodes in the selected list and their k-hop neighbors are visualized. The Meta-Path mode extracts paths and related nodes that conform to user-defined meta-path patterns within an expanded subgraph generated from the selected list and a specified propagation depth. Meta-paths are configured by combining node types—industry, company, and patent—with optional company-level attributes such as primary and secondary industry classification or province. For example, "company-patent" denotes paths between company and patent nodes. In Node Embedding mode, the system identifies a fixed number (TopK) of nodes most similar in embedding space to each selected node, subject to a similarity threshold between 0 and 1. After setting the mode and relevant parameters, users can generate the subgraph by clicking "Request Subgraph." The rendering style follows the layout principles of the global force-directed graph (D). Nodes highlighted in red represent those in the currently selected company list from view (B), allowing users to observe how companies from different lists are distributed within the subgraph. Additionally, the interface supports subgraph deletion and parameter adjustments; users can delete a subgraph with the “Delete Subgraph” button or update its configuration by selecting it and re-submitting revised parameters.

(F) Geographic Map View. Company nodes are displayed based on geolocation in this view. By default, the map is zoomed to display the entire country, enabling users to quickly identify the nationwide distribution of industries. Users can adjust the view through zooming and panning operations to focus on specific provinces or regions. This view supports interactive functionalities similar to Global Force-Directed Graph View. Nodes are color-coded to represent companies from different primary industries. Users can hover over nodes to view detailed information or select nodes via clicking or lasso selection. To more clearly illustrate the geographic distribution of companies across the country, node sizes are set relatively large, which may cause overlapping among nearby nodes. Users can leverage the filtering functionality to select densely clustered regions and examine specific company distributions in the node list panel.

(G) Industry and Region Statistics. The view consists of two modules: Statistics by Industry and Statistics by Province. The Statistics by Industry module presents the distribution of both primary and secondary industries based on the currently applied provincial filter. Each industry block is color-coded, with numerical labels indicating the number of companies within each industry. Users can intuitively observe the industry-color mapping from the blocks. Additionally, filters can be applied directly by clicking on specific industry blocks The Statistics by Province module displays a ranking of provinces by company count, based on the currently selected industry filter. Users can also apply a province-level filter by clicking on a corresponding row in the ranking list. Furthermore, this view includes a toggle switch

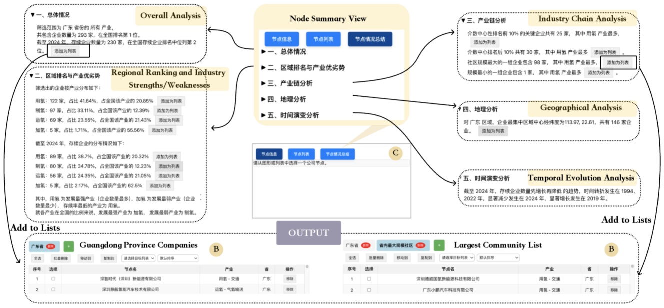

Fig. 4: node summary view covers information from five key aspects: (1) Overall Analysis, (2) Regional Ranking and Industry Strengths/Weaknesses, (3) Industry Chain Analysis, (4) Geographical Analysis, and (5) Temporal Evolution Analysis, where we identify two key lists: Guangdong Province Companies and Largest Community List.

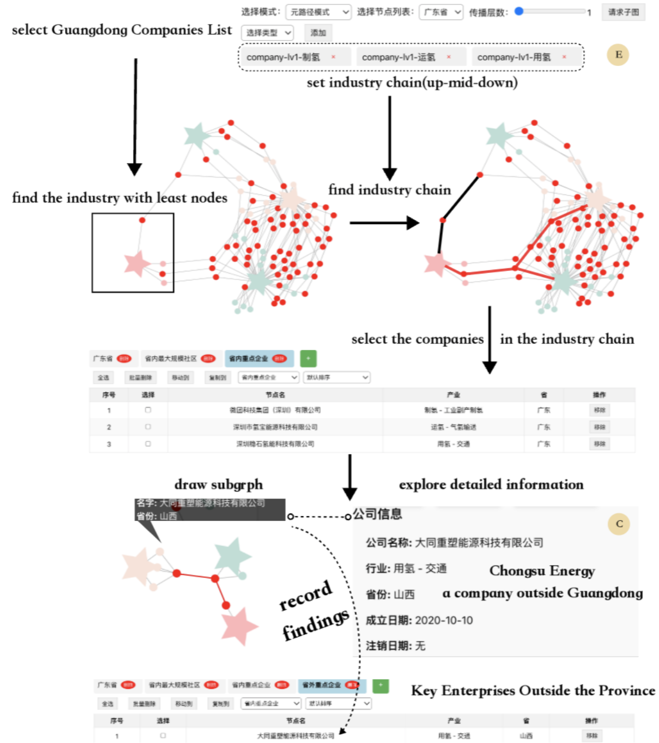

Fig. 5: For the meta-path mining method, we select the Guangdong Province Companies List as the target list and use a meta-path as the extraction strategy. By setting the meta-path as the industry chain, we find a complete industry chain in the subgraph. After recording the companies in the industry chain, we draw a new subgraph and identify an important company outside the province, which is later recorded in the key enterprises outside the province list.

labeled "Filter by Active Companies". By default, all companies are included in the statistics. When the switch is enabled, only currently operating (active) companies are counted, allowing users to focus on the current landscape of industry distribution.

(H) Time Axis View. A stacked bar chart displays the number of companies per year across industries. Clicking a year reveals newly registered firms for that year, supporting temporal exploration.

(I) Competitive Comparison. This view enables side-by-side analysis of two firms based on multiple dimensions, such as industry relevance, patent portfolios, graph centrality, and related firms. It helps users assess relative competitiveness.

5 Case Study

To evaluate the effectiveness and practicality of our system, we conducted two representative case studies involving real-world analysis tasks from distinct user roles: (1) a government policymaker aiming to optimize regional hydrogen industry strategies, and (2) an enterprise user seeking to understand its position within the industrial value chain. These cases illustrate how our system supports both top-down and bottom-up exploration paths, combining interactive visualizations, subgraph extraction, and comparative analysis.

5.1 Case 1: Supporting Regional Hydrogen Policy and Enterprise Introduction

Objective. In this case, the user acts as a regional government planner from Guangdong Province—one of China's leading regions in hydrogen energy development. The goal is to assess the province's hydrogen industry layout, identify core enterprises for strategic support, and discover promising external firms for collaboration or investment to improve weak links in the value chain.

Workflow. Based on user exploration needs, this case follows an analytical path from macro to micro and from overall to specific levels. First, a general exploration is conducted to form a list of key enterprises within the province. Next, the key enterprise list is further refined, and finally, based on this provincial key enterprise list, a list of key enterprises from outside the province is derived. This complete process provides an important reference for the government in formulating domestic industrial policies and attracting foreign investments.

Macro-Level Analysis of Industrial Strengths and Weaknesses. First, users can overview the data in two ways. On the one hand, by filtering by province, users can analyze visualization results across different views to identify company association patterns, time-series trends, geographic distribution patterns, and overall quantity trends. On the other hand, users can also utilize the (C) node details and selection view to examine insights from the node summary view.

Through the summary view(see Fig. 4), users find that although Guangdong Province has a relatively high number of hydrogen production enterprises (97), these companies account for only 12.39% of the national hydrogen production industry. This indicates Guangdong's relative weakness in hydrogen production. Additionally, the summary view clarifies which lists require further attention—specifically, the list of existing enterprises in Guangdong and the largest community enterprise list. Users can add these lists to their focus list using shortcut buttons in the view. These lists help further explore the core strengths, weaknesses, key nodes, and development potential of Guangdong's hydrogen energy industry, providing a scientific basis for subsequent industrial analysis and investment attraction efforts.

Identifying Representative Companies within the Province.A complete industrial chain is essential for understanding the industry comprehensively. Users explore it by either analyzing all provincial enterprises through meta-path mining or identifying key enterprises and mapping their upstream and downstream relationships.

For meta-path mining method, by setting the meta-path "hydrogen production - hydrogen transportation - hydrogen utilization," only one complete industrial chain in Guangdong is identified, highlighting weak enterprise synergy(see Fig. 5). Adding three core enterprises including WenShi Hydrogen to a new list and extracting related firms via the "company-to-company" meta-path reveals that key external enterprises play vital roles in linking Guangdong's hydrogen energy supply chain. These are included in the "Key Enterprises Outside the Province" list to explore inter-provincial cooperation.

Further analysis starting from key enterprises uses the largest community network, identified via the Louvain algorithm, with the top 20 enterprises ranked by PageRank as exploration starting points. Subgraph analysis shows that most connections remain industry-specific, with only one company crossing industries. Adding four of its connected enterprises reveals its central network role and extensive patent holdings, proving its significance.

Through filtering, analysis, and validation, a final list of seven key enterprises emerges. Users can adjust selection criteria to generate targeted lists supporting industrial policy and corporate strategy.

Exploring Investment Attraction Strategies. When exploring enterprises outside the province, users need to comprehensively consider industrial, geographical, and associative factors to formulate diversified cooperation strategies.

Starting with weak industries, users choose to analyze the hydrogen production sector. After filtering, they find that Hebei Province has a significant advantage in this field. They further select enterprises in Hebei that are closely related to hydrogen production and extract a subgraph containing these enterprises along with existing enterprises in Guangdong. Through meta-path analysis, they identify connections between "Guangdong enterprises - Hebei enterprises" and incorporate Hebei enterprises that are already linked to Guangdong firms into the "Key Enterprises Outside the Province" list as potential cooperation partners.

Additionally, users can identify external enterprises by analyzing subgraphs related to the seven key provincial enterprises and find that a Shanxi company, Chongsu Energy, plays a crucial role in the subgraph. Also, users can adopt a similar approach to weak industries by constructing a subgraph between Guangdong and neighboring provinces to uncover key associations and potential partnerships. Finally, users can input these enterprises and their linked Guangdong counterparts into the competitive comparison view for a multidimensional analysis to assess their value for introduction.

Outcome and Validation. In summary, the system produces:(1) a list of seven Guangdong-based firms covering upstream, midstream, and downstream sectors, including WenShi Hydrogen and Yaqing Technology, and (2) a curated list of external firms from Hebei, Shanxi, and others, matched by innovation relevance or structural linkage.

Cross-referencing public sources validates findings: (1) WenShi is backed by major SOEs and has 30+ patents , (2) Chongsu Energy is actively engaged in municipal hydrogen projects in Guangdong. Thus, the system effectively supports planning-level decisions for industrial development and cross-regional cooperation.

5.2 Case 2: Enterprise-Level Industrial Positioning

To support enterprise-level strategic analysis, the system was used to investigate the industrial positioning of Dongfang Electric Group, a major enterprise in China's hydrogen energy sector. The objective was to assess the company's network role, innovation capability, and potential for external collaboration.

First, the user searched and identified the company using the keyword search module. The company ranked second in node degree, indicating relatively high connectivity. However, in the global force-directed layout, it appeared structurally peripheral and primarily connected to downstream (use-phase) firms, with limited interaction with upstream and midstream entities.

Next, a k-hop subgraph was extracted to explore local connectivity. The result showed that Dongfang Electric had only one direct peer-level link—Liaoning Hongtu Power Technology. Further subgraph analysis revealed that this peer exhibited broader relational ties across the network, whereas Dongfang Electric maintained few external connections(see Fig. 6).

To understand this asymmetry, the user employed the visual comparison tool. Results showed that Dongfang Electric had a significantly stronger innovation profile, including a larger number of patents and deeper technical specialization, while the peer firm had limited innovation capacity. This suggests that Dongfang Electric's limited reliance on external connections stems not from marginality, but from its internal ability to operate across multiple segments of the value chain.

Finally, external validation confirmed that Dongfang Electric has upstream-to-downstream capabilities, including fuel cell production and hydrogen storage technologies. This aligns with the system's findings and highlights its effectiveness in identifying structural roles, assessing innovation competitiveness, and supporting collaboration decisions.

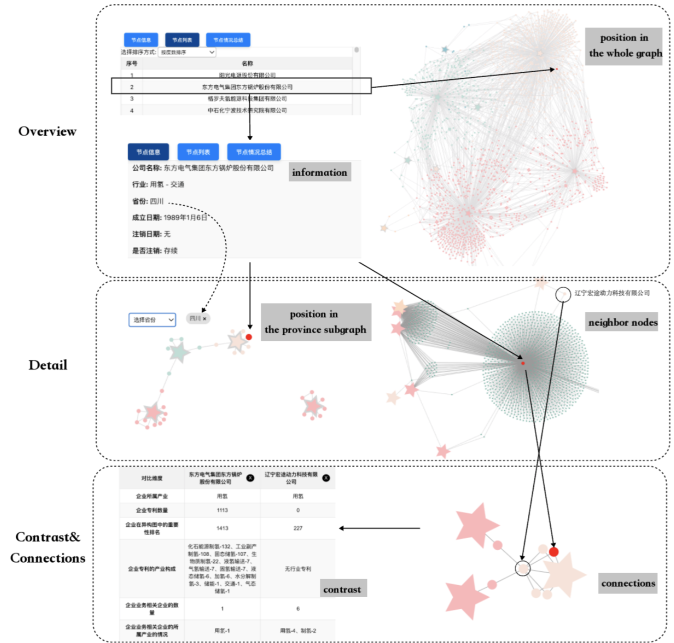

Fig. 6: An overview comprehension is acquired from the whole graph. After that, more detailed information is shown in the subgraph, both in the province subgraph and in the neighboring subgraph, where we find the most related node. For the node we find, we contrast it with the target node and explore the connections between them for a better understanding.

6 Conclusion and Future Work

Effective analysis of industrial chains is crucial for strategic planning, yet current analytical tools often lack the capacity to extract critical information and support multi-dimensional exploration from complex datasets. Limitations such as insufficient visual clarity, constrained interaction mechanisms, and poor decision-support capabilities restrict the practical utility of existing methods.

To address these challenges, this paper presents a multi-level visual analytics system for dynamic industrial chain analysis. Leveraging heterogeneous graph modeling, graph embedding, and interactive visualization techniques, the system enables users to identify core enterprises, uncover structural relationships, and support informed decision-making.

Despite the promising results, several limitations remain. First, while the system currently supports graph construction using public enterprise data, it still relies on user-provided datasets. Future versions will integrate preloaded datasets aligned with user-selected industries while retaining support for user-uploaded data to enrich the analysis.

Second, the current subgraph extraction strategies are primarily based on static attributes and lack explicit modeling of temporal dynamics. This limits the system's capacity to capture the evolution of industrial ecosystems over time. Incorporating time-aware analysis will improve the system's ability to model industry cycles, market volatility, and supply chain resilience.

Third, the system's flexibility and customization capabilities can be further enhanced. Future iterations will focus on improving user-defined configurations, enhancing readability, and deepening analytical insights.

Ongoing work will pursue three directions: (1) expanding data sources and establishing a comprehensive knowledge base to support plug-and-play industrial analysis across domains; (2) incorporating time-series modeling to reveal dynamic transitions in industrial networks; and (3) conducting user studies to refine interactive visualization features for more intelligent and personalized decision support in policy-making, enterprise strategy, and supply chain optimization.

References

[1] A. Arleo, C. Tsiganos, C. Jia, et al. Sabrina: Modeling and visualization of financial data over time with incremental domain knowledge. In *IEEE*, 2019. 2

[2] L. Chen, Y. Ouyang, H. Zhang, et al. Riseer: Inspecting the status and dynamics of regional industrial structure via visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1070–1080, 2022. 2

[3] E. Cuenca, A. Sallaberry, D. Ienco, et al. Vertigo: A visual platform for querying and exploring large multilayer networks. *IEEE Transactions on Visualization and Computer Graphics*, 28(3):1634–1647, 2021. 3

[4] P. Eades, Q.-W. Feng, and X. Lin. Straight-line drawing algorithms for hierarchical graphs and clustered graphs. In *Springer*, 1997. 2

[5] G. Ellison and E. L. Glaeser. Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105(5):889–927, 1997. 2

[6] Y. Frishman and A. Tal. Multi-level graph layout on the gpu. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1310–1319, 2007. 2

[7] T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, 1991. 2, 5

[8] T.-Y. Fu, W.-C. Lee, and Z. Lei. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In *IEEE*, 2017. 2

[9] X. Fu, J. Zhang, Z. Meng, et al. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. In *IEEE*, 2020. 2

[10] R. Färe, S. Grosskopf, and M. Norris. Productivity growth, technical progress, and efficiency change in industrialized countries. *The American Economic Review*, pp. 66–83, 1994. 2

[11] E. R. Gansner, E. Koutsofios, S. C. North, et al. A technique for drawing directed graphs. *IEEE Transactions on Software Engineering*, 19(3):214–230, 1993. 2

[12] D. Hennessey, D. Brooks, A. Fridman, et al. A simplification algorithm for visualizing the structure of complex graphs. In *IEEE*, 2008. 3

[13] D. Holten and J. J. Van Wijk. Force-directed edge bundling for graph visualization. In *Wiley Online Library*, 2009. 3

[14] H. Hong, H. Guo, Y. Lin, et al. An attention-based graph neural network for heterogeneous structural learning. In *IEEE*, 2020. 2

[15] Z. Hu, Y. Dong, K. Wang, et al. Heterogeneous graph transformer. In *IEEE*, 2020. 2, 3

[16] T. Kamada and S. Kawai. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7–15, 1989. 2

[17] E. E. Kosasih and A. Brintrup. A machine learning approach for predicting hidden links in supply chain with graph neural networks. *International Journal of Production Research*, 60(17):5380–5393, 2022. 2

[18] S. H. Lee, F. Ji, and W. P. Tay. Sgat: Simplicial graph attention network. *arXiv preprint arXiv:2207.11761*, 2022. 2

[19] R. A. Leite, A. Arleo, J. Sorger, et al. Hermes: Guidance-enriched visual analytics for economic network exploration. *Visual Informatics*, 4(4):11–22, 2020. 2

[20] Y. Li, Z. Zhu, L. Chen, et al. Exploring large-scale financial knowledge graph for smes supply chain mining. *IEEE Transactions on Knowledge and Data Engineering*, 2023. 2

[21] Q. Lv, M. Ding, Q. Liu, et al. Are we really making much progress? revisiting, benchmarking and refining heterogeneous graph neural networks. In *IEEE*, 2021. 2

[22] T. Moscovich, F. Chevalier, N. Henry, et al. Topology-aware navigation in large networks. In *IEEE*, 2009. 3

[23] R. Pienta, F. Hohman, A. Endert, et al. Vigor: Interactive visual exploration of graph query results. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):215–225, 2017. 3

[24] R. Pienta, M. Kahng, Z. Lin, et al. Facets: Adaptive local exploration of large graphs. In *SIAM*, 2017. 3

[25] Y. Ren, B. Liu, C. Huang, et al. Heterogeneous deep graph infomax. *arXiv preprint arXiv:1911.08538*, 2019. 2

[26] S. A. Rhoades. The herfindahl-hirschman index. *Fed Res Bull*, 79:188, 1993. 2

[27] C. Shi, Y. Li, J. Zhang, et al. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering*, 29(1):17–37, 2016. 2

[28] Y. Tu, W. Li, X. Song, et al. Using graph neural network to conduct supplier recommendation based on large-scale supply chain. *International Journal of Production Research*, pp. 1–14, 2024. 2

[29] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. *stat*, 1050(20):10–48550, 2017. 2

[30] X. Wang, H. Ji, C. Shi, et al. Heterogeneous graph attention network. In *IEEE*, 2019. 2

[31] Y. Wang, Q. Shen, D. Archambault, et al. Ambiguityvis: Visualization of ambiguity in graph layouts. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):359–368, 2015. 3

[32] N. Wong and S. Carpendale. Supporting interactive graph exploration using edge plucking. In *SPIE*, 2007. 3

[33] Z. Wu, S. Liu, Y. Zhou, et al. Viea: A visualization system for industrial economics analysis based on trade data. *IEEE Transactions on Computational Social System*, 2023. 2

[34] W. Xiong, C. Yu, C. Shi, et al. V4rin: visual analysis of regional industry network with domain knowledge. *Visual Computing for Industry, Biomedicine, and Art*, 7(1):11, 2024. 2

[35] F. Zhou, C. Cao, K. Zhang, et al. Meta-gnn: On few-shot node classification in graph meta-learning. In *IEEE*, 2019. 2