KG-PRE-view: Democratizing a TVCG Knowledge Graph through Visual Explorations

Category: Research

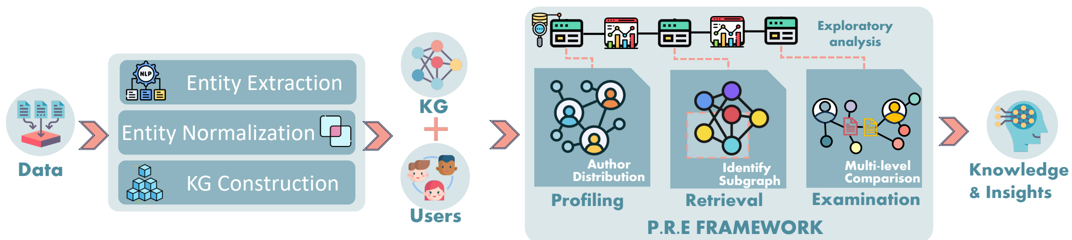


Figure 1: KG-PRE-view Overview: We first construct a TVCG knowledge graph and enable community access through visual exploratory tasks for enhanced decision-making.

Abstract

IEEE Transactions on Visualization and Computer Graphics (TVCG) publishes cutting-edge research in the fields of visualization, computer graphics, and virtual and augmented realities. Within the TVCG ecosystem, different stakeholders make decisions based on available information related to TVCG almost on a daily basis. The decisions involve various tasks such as the retrieval of research ideas and trends, the invitation of peer reviewers, and the selection of editorial board members, just to name a few. To make well-informed decisions in these contexts, a data-driven approach is necessary. However, the current IEEE digital library only provides access to individual papers. Transforming this wealth of data into valuable insights is a daunting task, requiring specialized expertise and effort in tasks such as data crawling, cleaning, analysis, and visualizations. To address the needs of the community in facilitating more efficient and transparent decision-making, we construct and publically release a TVCG knowledge graph (TVCG-KG). TVCG-KG is a structured representation of heterogeneous information, including the metadata of each publication such as *author, affiliation, title*, and semantic information such as *method, task, data*. Despite the widespread use of KGs in various downstream applications, a noticeable gap exists in the visualization literature regarding the full exploitation of the rich semantics embedded within KGs. While it might seem intuitive to just employ interactive graph-based visualization for KGs, we propose that knowledge discovery over KG is a series of visual exploratory tasks that can benefit from using multiple visualization techniques and designs. We conducted an evaluation of TVCG-KG quality and demonstrated its practical utility through several real-world cases. Our data and code are accessible via the following URL: `https://github.com/yasmineTYM/TVCG-KG.git`.

Index Terms: Visual Analytics for Knowledge Discovery—Knowledge Graph—Corpus Analysis—Text and Document data;

1 Introduction

IEEE Transactions on Visualization and Computer Graphics (TVCG) is a journal that publishes cutting-edge research on visualization, computer graphics, and virtual and augmented realities. Within the TVCG ecosystem, a variety of stakeholders routinely rely on available information related to TVCG in their decision-making processes. Researchers leverage digital publications and collaboration networks to identify research trends and seek potential collaborators. Committee members heavily depend on historical data to guide their decision-making, including nominating editors, identifying reviewers, and assigning papers to reviewers. Given these contexts and recognizing the potential of expanding data usability, there is a strong motivation to employ a data-driven approach to ensure well-informed decision-making.

The current IEEE digital library [1] provides access to digital versions of papers. However, using it directly in real-world scenarios poses several challenges. First, many exploratory questions related to TVCG require knowledge extracted from the content of papers. With existing databases, users need to synthesize, process, and analyze to transform information into insights. However, it is unrealistic to expect all users to possess the necessary engineering skills or have sufficient time to analyze. For example, questions like "What visual analytics approaches have been proposed for topic modeling?" necessitate further processing of the semantic content and identifying their connections. Second, consider a scenario where a knowledge base explicitly connects all knowledge snippets. In this situation, users do not need to deduce potential relationships through manual analysis. However, they may encounter difficulties when attempting to efficiently and interactively generate insights from the knowledge base. This difficulty arises from the heterogeneity of the data, the diverse needs of users, and the complexity of the tasks involved. Now consider the exploratory question, "Who is the best candidate to review a paper related to large language models and visual analytics?" Identifying relevant reviewers just by name is not enough; they also need to be profiled, analyzed, and compared to make an informed decision when selecting reviewers.

The first issue we tackle is the absence of a unified knowledge base constructed for the TVCG community. Recently, both general knowledge graphs like Wikipedia, DBPedia, and Google Knowledge Graph, as well as domain-specific knowledge graphs, have showcased their value in various downstream tasks. However, creating a TVCG knowledge graph from scratch is not trivial. Two critical aspects should be carefully considered: (1) **what** types of information should be incorporated in the KG to ensure sufficiency and efficiency in supporting real-world exploratory tasks? (2) **how** to extract, synthesize, and validate the multi-source information to guarantee the high quality and trustworthiness of the TVCG KG? To answer those two questions, we introduce our KG design rationale and the end-to-end construction pipeline, including *ontology definition, entity extraction, and normalization*. The key benefits of TVCG-KG over other relational datasets lie in its efficiency in connecting heterogeneous information. The KG can be easily queried via various querying languages, enabling a more diverse and comprehensive retrieval of information. We later briefly discuss the

[1] `https://www.computer.org/csdl/journal/tg`

semantics of graph queries and provide more detailed examples in the application-based evaluations of TVCG-KG.

Another key challenge we address is visual exploration over a knowledge graph. A recent interview with KG practitioners revealed a missing effort in the visualization literature for leveraging semantic-richness knowledge graphs [43]. While using node-link diagrams as a visual representation for knowledge graphs may seem intuitive [33], it often leads to problems like the hairball effects, reduced readability, and issues related to trustworthiness. In this paper, we aim to initiate a discussion on *whether KG visualization has to be a graph visualization.* Our answer is NO. We believe the visualization should focus on both the topological and the data instance aspects of the KG. While many existing efforts have been put into summarizing data patterns with proper visualizations for tabular and graph data, there is still a missing connection between data and graphs to facilitate knowledge graph explorations. To fill the gap, we introduce three visual exploratory tasks called PRE-view: profiling, retrieval, and examination, where each task is answered with data and visualization. Tasks can then be connected to form various exploration pipelines. Through this discussion, we hope to provide (1) a good understanding of how visualization can contribute to various KG tasks and (2) inspiration for novel visual designs to support specific needs in the KG community. Our contributions are outlined as follows:

- We construct a domain-specific knowledge graph, i.e., TVCG-KG, released as a public dataset for the community. This can benefit various downstream tasks such as question answering, knowledge discovery, and entity-based explorations.
- We introduce an end-to-end framework for domain-specific KG construction, utilizing the GPT-3.5 model. The approach can be easily adapted to similar efforts in other scientific domains.
- We propose a PRE-view approach to summarize exploratory tasks with corresponding visualization as answers. It enhances the data-driven decision-making process related to TVCG.

2 Related Work

2.1 Collections and analysis of VIS Publications

Collecting and publishing scholarly datasets can support various use cases in the research community. Similar efforts have been advocated in the visualization community. The pioneering dataset, VisPub [36], collects information on IEEE VIS publications from 1990-2022. This dataset promotes many work, such as topic analysis [37, 38], citation analysis [30, 91]. Another dataset, VitaLITy [56], mainly focuses on utilizing embeddings for paper retrieval in serendipitous discovery. It also contributes data from 38 popular data visualization publication venues. Besides text-based datasets, many datasets collect and categorize figures and tables from papers, such as VIS30K [11], VizNet [35], VisImages [16]. These image-based datasets enable the graphical content analysis, including color vision deficiencies [4], neural embedding for image retrieval [88].

While these datasets are valuable resources for conducting data analysis in visualization literature, there are several potential issues that TVCG-KG aims to address: (1) TVCG not only covers visualization but also extends to virtual reality and graphics. With the development of artificial intelligence, the boundaries of core technologies are becoming blurred. It is intriguing to explore the integration of multi-discipline research and see how they are connected. (2) Existing work focuses on the meta-data level, leaving the burden of semantic processing to users. In contrast, our TVCG-KG shifts such burdens into the KG construction process, thereby enhancing the efficiency of semantics-related explorations.

2.2 Construction of Knowledge Graph

The creation of KGs can be divided into two main streams: ontology- and non-ontology-based approaches. Non-ontology approaches extract entities and relationships from unstructured text, independent of pre-defined ontology [89]. On the other hand, ontology-based methods follow pre-defined rules to connect entities, which is the focus of our research. They are well-studied in the field of NLP. Mondal et al. [52] define an ontology for NLP-KG and propose an end-to-end framework including three distinct relation extractors to identify pre-defined relation types. Al-Khatib et al. [2] establish an argumentation KG and propose a supervised approach for relation detection. Similar efforts of KG construction have been applied to scientific literature. Chen et al. [10] present an ontology-based pipeline to build KGs from abstracts. However, they perform sentence classification first and then extract entities from each sentence, making it less suitable for scenarios where entities with different labels can co-occur in the same sentence. Tosi et al. [77] aim to structure knowledge in scientific literature, leveraging a Babelfy [54] to extract entities and map them to BabelNet [57], a knowledge graph built on WordNet [51]. Wise et al. [86] construct an ontology for COVID-19 from CORD19 and extract entities using SciSpacy.

2.3 Visualization of Knowledge Graph

As discussed in a recent interview with KG practitioners, there are many needs and opportunities for KG-based visualization research [43]. Nararatwong et al. [55] discussed several challenges associated with visualizing KGs due to their extensive and complex nature, while Gomez et al. [27] conducted a performance analysis of various visualization tasks for large-scale knowledge graphs.

Many existing efforts are on building visualization systems for KGs, with various focuses. Several visualization systems facilitate KG querying through visual query formulations (e.g., OptiqueVQS [73]) or graph-like queries (e.g., RDF Explorer [78], FedViz [21]). Another set of tools supports visualizing queried data from KGs [15, 85]. In addition, there are some systems designed and developed for domain-specific problems [59, 74], such as tax-related [61], spatiotemporal analysis of COVID-19 [39], dietary supplement analysis [32].

3 Notations

A knowledge graph, denoted as $\mathcal{K}$, stores facts as a graph representation. It captures entities as nodes, denoted as $\mathcal{E}$, such as *Alice*, *John*, *Microsoft* and *Google*. The relationships between entities are represented as links, denoted as $\mathcal{R}$, such as *is_member_of*. $\mathcal{K}$ comprises two essential components: an ontology $\mathcal{O}$ and a data model $\mathcal{M}$.

- The ontology, $\mathcal{O}$, defines the entity classes $\mathcal{C}$ and specifies how they are interconnected. Formally, $\mathcal{O} \in \mathcal{C} \times \mathcal{R} \times \mathcal{C}$. For example, $\mathcal{O}$ defines the <u>*author*</u>[2] class connected to <u>*affiliation*</u> class through the <u>*is_member_of*</u> relation.
- $\mathcal{M}$ consists of factual data instances that adhere to the rules defined in $\mathcal{O}$. Each fact is represented as a triplet of $\langle$*head, relation, tail*$\rangle$, denoted as $\mathcal{M} = \{(h, r, t) | h, t \in \mathcal{E}, r \in \mathcal{R}\}$. For instance, the data model $\mathcal{M}$ instances many factual triplets between <u>*affiliation*</u> and <u>*author*</u>, such as *(John, is_member_of, Microsoft) and (Alice, is_member_of, Google)*.

4 Requirement Analysis

In the TVCG ecosystem, various target users interact with the data as part of their daily academic routine, such as editors, reviewers, authors, and readers. To define the scope of our work, it is important to analyze users' real needs and identify tasks that can be better supported. However, given the diversity of stakeholders, these tasks can be multi-level and multi-dimensional. We collect tasks in three ways: (1) discussions with domain experts, (2) a literature review, and (3) the author's daily experience in academic research. We categorize those tasks and summarize them as follows:

[2] We use underline to indicate the class type and relation type defined in the ontology $\mathcal{O}$.

- *R1. Providing a comprehensive overview of the TVCG literature*. It is quite important in several practical scenarios. This overview task can be divided based on the *how* and *what* aspects of the overview process. As illustrated in Table 1, each specific task serves a valuable purpose for various stakeholders. For instance, the Editor-In-Chief (EIC) is responsible for monitoring the journal's content, and an overview can assist in tracking the evaluation of paper topics and the distribution of editors' research areas. Within the visualization literature community, researchers may seek collaborations by author profiling and gain inspiration from topic analysis.

Table 1: How and what to overview for TVCG literature.

What / How	By Papers	By Author	By Concept
Static	Document Clustering	Author Profiling	Topic Modeling
Dynamic	Distribution Change	Collaboration Analysis	Topic Evolution

- *R2. Retrieving heterogeneous information from a knowledge graph*. Once an overview of data is obtained, it is a key step to locate a subgraph that contains the desired information. This can be achieved by querying the knowledge graph, with further task divisions based on the target entity type:
  - *R2.1 Retrieve-by-Paper*: Identifying papers of interest is a common task for researchers to keep up with the advancements in their respective fields. This includes scenarios such as tracking the work of emerging scholars, exploring trending topics, or analyzing citation patterns.
  - *R2.2 Retrieve-by-Author*: Locating research scholars specializing in specific topics is another routine task. For instance, retrieving Associate Editors (AEs) and identifying their research areas ensures that the EIC can assign submissions properly and effectively.
  - *R2.3 Retrieve-by-Concept*: The essence of papers can be captured as building relationships between concepts. It is important for researchers to retrieve novel concepts to stay updated on evolving research trends and techniques. This has become increasingly important, given the rapid development of AI technology recently.
- *R3. Extracting more details based on the information of interest*. Once a subgraph containing desired information is identified, conducting an in-depth exploration becomes crucial. This exploration process may entail further information retrieval or summarization, and its execution can be categorized into three scopes:
  - *R3.1 Expanding an entity*: a target entity must retrieve all related information and perform entity summarization. For instance, when dealing with an author, various aspects may need to be explored comprehensively, such as publication and research focus.
  - *R3.2 Comparing several entities*: several targets facilitate comparison purposes. It can be done by performing expanding tasks separately first with further comparing.
  - *R3.3 Summarizing multiple entities*: a set of target entities may result in information overload, motivating proper visual summarization to present data patterns clearly and efficiently.

5 Method

As visualized in Fig. 1, the core of our framework is using knowledge graph $\mathcal{K}$ as a corpus representation. It offers several advantages which perfectly solve the deficiencies we mentioned before. First, $\mathcal{K}$ contains various types of entities, including metadata entity, such as *Author*, *Affiliation* and semantic entity extracted from paper content, such as *Method*, *Task*. In this way, it is efficient to perform semantic analysis and even more advanced analysis that requires both semantics and metadata. Second, $\mathcal{K}$ contains semantic relationships, providing contextual information about entities. It supports users to have a deeper understanding of data. Lastly, the graph nature of $\mathcal{K}$ offers flexibility and expressive queries. It is intuitive to perform pattern matching to identify target information from the graph.

As knowledge graph $\mathcal{K}$ as the backbone, our method answers the following two questions: (1) how to construct and query a TVCG knowledge graph? (2) how can it be used to solve real-world tasks?

5.1 TVCG Knowledge Graph

It is not trivial to build a domain-specific knowledge graph from scratch. To truly benefit the community, we first determine which information should be incorporated and then use the state-of-the-art model to extract the necessary information.

5.1.1 TVCG Data Preparation

We first prepare the most up-to-date TVCG publication dataset to build the knowledge graph, which contains TVCG papers from 1995 to August 2023.

Data Retrieval To prepare the dataset, we use a hierarchical retrieval strategy to retrieve TVCG papers from the Computer Society Digital Library. The strategy includes three levels of hierarchies: year→issue→papers. As TVCG grows, the number of issues increases over the years, so we first query the issues published every year. Then we query issue information regarding what papers are contained. Finally, we query paper details using the paper ID contained in each issue. We retrieved 5535 publications in total.

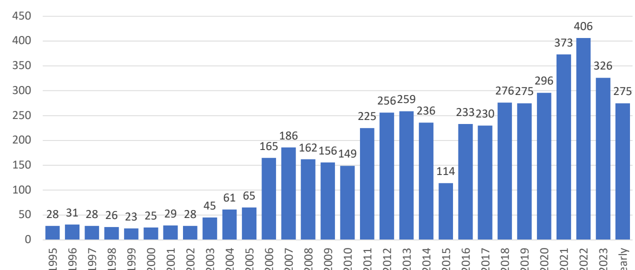


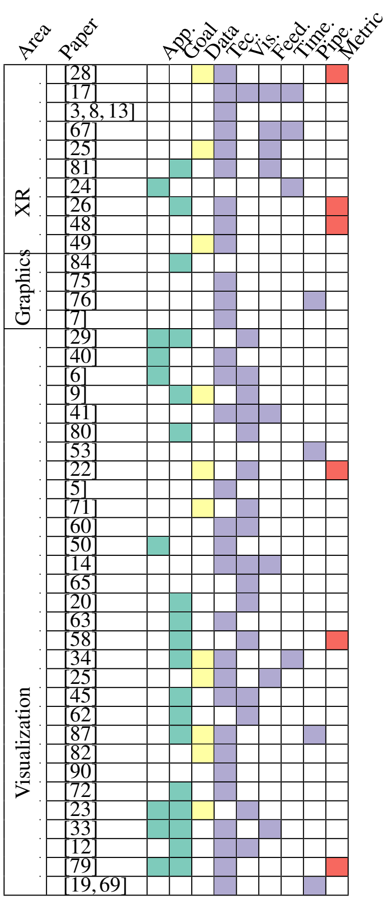
Figure 2: The number of papers published in TVCG from 1995 to 2023, including early access papers.

Data Cleaning & Validation The publications retrieved from CSDL contain various types, including *Paper, Index, Editor's note, ERRATA, Reviewer List, and Covers*. To ensure a high-quality knowledge graph, we exclusively retain *Paper* type in the dataset. However, due to the inconsistencies in the original data source, type information is missing for some publications. Therefore, we use a keyword-matching approach to filter out other types of publications. To include all necessary keywords, we first manually check the number of papers for each issue every year as the ground truth. Then, we iteratively add and adjust keywords in a trial-and-error process. The iterative process continues until we confirm that the final results align with the ground truth. In the end, we filtered 4987 papers out of 5538 by removing other types of publications. To show the statistics of our final dataset, we present the number of papers in a bar chart as depicted in Fig. 2.

5.1.2 Ontology Design

While metadata are straightforward in describing the basic information of each paper, semantic information offers descriptive summaries of the paper's content. Designing an ontology is an important step to include both sources of information in the TVCG-KG. While the entity class of metadata can be directly derived from raw data, it is necessary to consider which dimension of semantic information should be incorporated as entity classes. There are four semantic dimensions that are widely used in entity extraction of academic papers [46,47]: *background, data, method, evaluation*. To validate and fine-tune these four dimensions based on TVCG papers, we conducted a survey of 47 survey papers selected from TVCG. For each paper, we record the semantic dimensions used for summarization, as listed in Table 2. From the survey, we observe that the four dimensions are aligned with TVCG papers. In addition, we notice there is a need for finer-grained dimensions. For example, *application* and *task/goal* are two detailed categorization of *background* information (●), which is adopted in our TVCG-KG ontology as shown in Fig. 3. During the survey, we noticed the *data* can be further divided into *input* and *output*, especially with most recent machine learning papers. However, the dimensions of *method* (●) vary a lot across sub-areas. To capture a more comprehensive list of *technique* entities, we decided not to differentiate different types of techniques. Regarding the *evaluation* aspect (●), we observe that it mainly focuses on the metric, as listed in Table 2. However, in practice, authors may also reference other methods as *baseline* model or evaluation *technique*. That is why we have incorporated these three dimensions in our ontology. The schematic design of our ontology $\mathcal{O}$ is illustrated in Fig. 3.

Table 2: A survey of survey papers from dimensions of **App**lication, **Goal**/task, **Data**, **Tec**hnique, **Vis**ualization, **Feed**back/interaction, **Time**/space/user case, **Pipe**line/component, **Metric**.



In addition to defining entity, the ontology $\mathcal{O}$ also specifies the interconnections between different entity types. One thing worth mentioning is that Paper entity connects to all semantic entities. This is because not every entity type can be extracted from paper content, given varying writing styles. Establishing these connections helps avoid the isolation of nodes that would be challenging to traverse and retrieve. If we focus on semantic entity alone, the descriptive connections among them is a good illustration of the main ideas

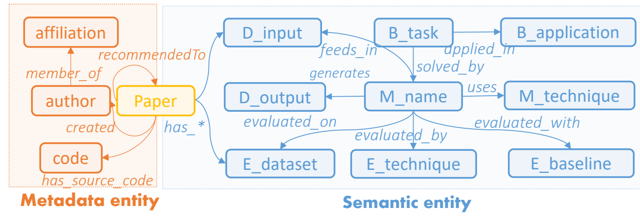


Figure 3: The schematic illustration of entities defined in the KG ontology. B, M, D, E are short for *background, method, data, evaluation*, respectively. *(To simplify the graph, we use semantic words to describe the relation and entity without using the real uri.)*

within each paper. To illustrate this point, we present an example of InSituNet [31], the best paper of IEEE Vis2019, in Fig. 4. It is clear to see the storyline of the paper: *InSituNet* is proposed to support the *parameter space exploration for ensemble simulations*. The method is evaluated on three simulations: *combustion, ocean, and cosmology*.

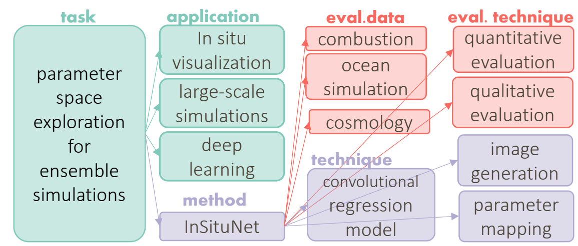


Figure 4: An example graph of the semantic entities for paper [31].

5.1.3 Knowledge Graph Construction

Following the ontology $\mathcal{O}$ poses a challenge when it comes to the automatic acquisition of semantic entities from text. Firstly, it requires the machine to have semantic comprehension capabilities for accurately extracting named entities. Second, due to the intrinsic nature of human language, one concept can be described in multiple ways. Consequently, finding a normalized version of these entities becomes necessary yet challenging. To solve the challenges, we propose an end-to-end pipeline for entity extraction and normalization using state-of-the-art models.

We exploit the power of Large Language Models (LLMs) to convert unstructured documents into structured entities. To achieve this, we engage in the prompt engineering process, where we carefully consider several key design elements in our final prompts: (1) **Instructions** describe the tasks to LLMs and provide specific guidelines to follow. Instead of framing the task as a general named entity extraction, we decided to tailor it to distill scientific concepts from publications. (2) **Data description** defines the desired output, i.e., entity classes as shown in Fig. 3. Note that they have a hierarchical structure. For example, *application* and *task* originate from *background* dimension. Deciding whether to include this structural information when instructing LLMs is another crucial design decision. (3) **Examples** can be incorporated into the prompts together with the instructions, which is known as the "show-and-tell" technique. We provide two examples in the prompt to efficiently leverage the power of LLMs. (4) **Model selection** is another important choice. In our experiment, we chose the *gpt-3.5-turbo* model, which was the most up-to-date choice at the time. We conducted an ablation study to validate these design choices, along with the complete prompt, which is available in our supplemental material.

To standardize entities, we utilize the Spotlight API to establish connections between the extracted entities and their corresponding

DBpedia resources [3]. This linking process facilitates normalization by connecting diverse descriptions with established concepts. It effectively prevents similar entities from becoming disconnected due to different language descriptions. For instance, both *interactive topic modeling* and *incremental hierarchical topic modeling* can be normalized to the concept of *topic model*. We also conduct entity normalization for *author* entities due to the presence of spelling errors and inconsistencies in the crawled data. To enhance the efficiency and precision of this normalization process, we implement a two-step approach involving both machine and human inspections. In the initial step, we identify candidate *author* entities based on a string-matching ratio exceeding the threshold of 0.8, which falls within the [0,1] range. A higher ratio signifies a stronger string match. Later, we manually check whether two names refer to the same person by spelling or missing initials. As a result, we removed 208 duplicate authors. Please refer to our git repository for more details regarding data cleaning.

5.1.4 Querying Knowledge Graph

The primary advantage of TVCG-KG over other relational scholar datasets is its effectiveness in acquiring heterogeneous information. Users can query it flexibly without the need for complex data wrangling. TVCG-KG can be stored in either RDF triplets or a Property Graph format. To query data from TVCG-KG, various query languages can be utilized, including SPARQL and Cypher. Their basic semantics can be abstracted to resemble a SQL query:

`SELECT` {*target*}, `WHERE` {*graph pattern*}, `FILTER` {*conditions*}

The {*target*} contains a list of variables or expressions (aggregation included) that are to be retrieved. The {*graph pattern*} specifies relevant graph patterns, while {*conditions*} is where to specify conditional expressions used to filter query results. For example, given the question: *"provide a list of papers published by Mystery Rivers."* The query abstraction can be described as follows:

`SELECT` {*paper*}, `WHERE` {*paper* $\xrightarrow{created}$ *author*},

`FILTER`{*author="Mystery Rivers"*}

The implementation of this abstraction into query languages varies based on different grammar. We prefer Cypher to SPARQL since it is more concise, intuitive, and readable. A corresponding Cypher query on TVCG-KG is shown as follows:

```
MATCH (p:Paper)-[:created]-(a:Author)
WHERE a.name="Mystery Rivers" RETURN p.title as title
```

It's easy to see that by matching a more intricate graph pattern, the advanced query can retrieve a wide range of information. More examples can be found in our task-driven evaluations.

5.2 Visual Explorations of Knowledge Graph

The general solution of visual exploration over KGs is the node-link diagrams (NLDs) [33] with graph-based interactions. While NLDs explicitly display the structural information, they have been criticized for lack of efficacy, scalability, and readability in the context of large KGs. To shed light on how visualization can help KG explorations, we discuss the inefficiency of using NLDs as KG visualizations and propose PRE-view as a solution. The key idea of PRE-view is that *KG visualization does not have to be graph visualization*. We will introduce why and how.

5.2.1 Why NLDs are not a good choice?

Without touching on the potential problems of using NLDs, we bring up a more fundamental question: is it sufficient to visualize the whole KG data only as NLDs? Unlike traditional graph data, given a KG, the connections among nodes strictly follow the ontology. In other words, we already know the KG structure from ontology, even without seeing data instances. For example, in the TVCG-KG, given an *author* entity, it must be connected to the *paper* entity by the ontology definition. What really matters is to retrieve those entities by using the relations instead of seeing the relations. Because, in the end, what users care about most is the data itself.

While exploring KGs in NLD helps users gain new insights in an intuitive way, we argue that using NLDs for KGs may not be suitable for supporting decision-making, for several reasons [43]: (1) the hairball effect caused by a large number of entities and relations, resulting in difficulties in digesting meaningful information. (2) graph visualization primarily based on the underlying ontology makes it hard to identify intrinsic data patterns. (3) While graph-based operations supported by NLDs are useful, such as nodes expanding, deleting, and dragging, users can be overwhelmed by massive possible paths to explore.

5.2.2 Overview of PRE-view approach

Although NLDs only for KGs may not be a good choice, we still believe visualization is indispensable during KG exploration. However, the visualization should target both the properties of graphs, such as the topology and the nature of data, such as in tabular form that are extracted or computed from KGs. While existing work has investigated data patterns for tabular data [18, 70, 83] and graph data [42], the connections between them for knowledge graph exploration are still missing. To fill the gap, we introduce three visual exploratory tasks, called PRE-view, described as follows:

- ***Profiling***: presents the overall structure of the knowledge graph, including the overview of both ontology $\mathcal{O}$ and data model $\mathcal{M}$. It can help users to understand the knowledge graph or explore the overall data patterns.
- ***Retrieval***: helps users retrieve information of interest from the massive knowledge graph.
- ***Examination***: delves into the details of target entities, including examining one entity with its multi-level information, comparing two entities, and summarizing a set of entities.

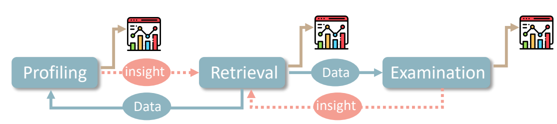


Figure 5: Interactions and communications between three tasks.

To satisfy the requirements we collected in Sect. 4, various exploratory pipelines can be built as a sequence of tasks, denoted as $P = \langle T_1, T_2, ..., T_k \rangle$. Therefore, it is important to introduce how each task can be connected to each other. In Fig. 5, we introduce two types of messages that can be exchanged among these tasks: *insight* and *data*. *Data* includes both raw data extracted from the knowledge graph and data computed from it through aggregation operations. *Retrieval* tasks exclusively generate data, which is then passed to *profiling* and *examination* tasks. On the other hand, *insight* refers to knowledge that results from human cognitive processes and motivates subsequent tasks. For example, the *profiling* task might uncover valuable insights, inspiring the next *retrieval* task.

5.2.3 Task Definition

To formally define each task, by extending a general task definition in KG exploration [44], we characterize a task T as a tuple of :

$$T = \langle \textit{type, goal, parameter, data, pattern} \rangle \tag{1}$$

Type: the type of task $\in$ { profiling, retrieval, and examination}.
Goal: one task can have various visualization goals that can be achieved with varying parameters and visualizations.

[3]`https://www.dbpedia-spotlight.org/api`

Table 3: We define three visual exploratory tasks, each with varying exploration goals according to its parameters. $\mathcal{M}$ is data model, $\mathcal{O}$ is the ontology. $\mathcal{C}$ is the classes of entities. f_{emb} is the embedding model for entities. f_{pro} is the dimensionality reduction algorithm, such as PCA and tSNE. Optional parameters are in square brackets, and required parameters are in angle brackets. G stands for raw graph data from KGs and T indicates tabular data that is derived from KGs.

Task	Visual Exploration Goal	Parameters	Data	Pattern
Profiling	**(G1)** To present the distribution of different entity classes	$\langle \mathcal{M}, \mathcal{O} \rangle, [\{c \vert c \in \mathcal{C}\}]$	T	Distribution/ Proportion
	(G2) To reveal structural relationships of entity classes.	$\langle \mathcal{O} \rangle, [\{c \vert c \in \mathcal{C}\}]$	G	Structure
	(G3) To explore clusters and inter-cluster understanding and comparison.	$\langle f_{emb}, f_{pro}, \mathcal{M} \rangle, [\{c \vert c \in \mathcal{C}\}]$	T	Clustering
	(G4) To discover the interested target entities among the overall distribution.	$\langle f_{emb}, f_{pro}, \mathcal{M}, \{e \vert e \in \mathcal{E}\} \rangle, [\{c \vert c \in \mathcal{C}\}]$	T	Clustering
Retrieval	**(G5)** To support exploratory analysis by flexibly building queries.	$\langle \mathcal{M}, Q \rangle,$	T	Multi\Structure
	(G6) To locate sub-graphs of interest and enable interactive browsing.	$\langle \mathcal{M}, Q \rangle,$	G	Structure
Examination	**(G7)** To identify multi-level details of several single entities.	$\langle \mathcal{M}, \{e \vert e \in \mathcal{E}\} \rangle$	G/T	Multi
	(G8) To compare entities by highlighting similarities and differences.	$\langle \mathcal{M}, \{e \vert e \in \mathcal{E}\} \rangle$	G/T	Multi
	(G9) To summarize a set of entities from multiple perspectives.	$\langle \mathcal{M}, \{e \vert e \in \mathcal{E}\} \rangle, [\{r \vert r \in \mathcal{R}\}]$	G/T	Multi

Parameters: the parameters for each task include ⟨required⟩ parameters and [optional] parameters as filtering conditions.
Data: The data is the output result based on the parameters and conditions. The data can be either sub-graph data (G) or tabular data (T) that go through aggregation operations.
Pattern: The data pattern of the resulting data determines which visualization chart is suitable. In addition to 10 data patterns identified from previous work [70] for tabular data, we added two more patterns: *clustering* tells the relationships of entities while *structure* is specific to show the structural information of graph data. *Multi* indicates all 12 data patterns.

The formal definition of three tasks is listed in Table 3. During the KG understanding stage, it is useful to show the basic statistics of the ontology [1], such as counting the number of entity/relation classes and their data instances **(G1)**, showing the ontology graph to reveal the structural relationships of entity classes **(G2)**. Target classes $\{c|c \in \mathcal{C}\}$ can be the optional parameters to filter statistical results. Besides ontology $\mathcal{O}$, it is also important to present the overall structure of the data instances, $\mathcal{M}$. However, users with various backgrounds can be interested in different global structures, such as semantic or topological relationships of all entities **(G3-G4)**. Therefore, we specify the embedding learning model as f_{emb} as a parameter to generate the desired entity embeddings and a projection algorithm f_{pro} for dimensionality reduction. In addition, users might have specific needs to identify target entities in an overview distribution, such as Fig. 8 (1). To achieve it, we also add a set of entities $\{e|e \in \mathcal{E}\}$ as optional parameters for highlighting and comparison purposes. For example, highlighting target papers in the paper distribution helps to identify similar papers.

While *profiling* provides a high-level overview of the KG, *retrieval* helps users zoom into a sub-graph for further exploration. For *retrieval* task, there are two types of visual exploration goals. One is efficiently displaying the patterns from tabular data computed from KG **(G5)**. Another is to enable interactive browsing for a sub-graph of interest **(G6)**.

Examination task delves into details of target entities. The goal varies based on the number of target entities. Examining one entity focuses on its own multi-level information, such as the distribution of its relations and its neighboring information **(G7)**. When two target entities are at hand, the goal is to compare them to identify their commonality and differences **(G8)**. With a number of entities, browsing each individual entity is less efficient than a visual summarization of all information. **(G8)**

6 Evaluation of TVCG-KG

Considering the diverse range of downstream tasks supported by KGs, multiple perspectives exist for assessing their quality [66]. In this study, we evaluate our TVCG-KG from three aspects: (1) an assessment based on its structure, (2) an evaluation of data quality, and (3) two usage scenarios for application/task-based evaluation.

6.1 Structure-based Statistical Assessment

In this section, we compute several metrics that reflect the structural statistics of TVCG-KG, including statistical properties of ontology, data model, and graph structure, as shown in Table 4.

Our schematic ontology depicted in Fig. 3, comprises 13 entity classes and 28 relationships among them. The full list of classes and relationships can be found in our supplemental material. To provide a clear perspective, we calculate the number of relations for each entity class and then compute the average. Additionally, We compute the number of entities and triplets to describe the TVCG-KG data models. Furthermore, it is worth noting that the TVCG-KG consists of a single connected component, signifying that all entities within it are interconnected without any instances of isolation.

Table 4: Structure-based assessment of TVCG-KG, including metrics of ontology, data model, and graph structure.

Ontology	# of entity class	13
	# of relations	28
	# of relations per class	4.54
Data Model	# of entities	81,033
	# of triplets	406,291
Graph Structure	Avg. in-degree	2.42
	Avg. out-degree	5.01
	# of weakly connected components	1

6.2 Data Quality Evaluation

In this section, we evaluate the quality of TVCG-KG by validating the relation consistency and interlinking ratio of TVCG-KG to other scientific KGs.

6.2.1 Consistency of Triplets

Another important measurement is how consistent the data in the knowledge graph are. We adopt the 10-fold strategy that evaluates the consistency of triplets in [86]. The idea is that we split the triplets into 10 separate folds. Then, we employ a Knowledge Graph Embedding (KGE) model trained on nine of these folds to predict the left-out fold. The underlying assumption is that if the triplets are consistent, the prediction performance should demonstrate stability with low variance across ten experiments.

In the experiment, we use TransE, one KGE model that learns the embedding for each triplet $\langle h, r, t \rangle$ such that $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$. Here h, r, t are head, relation, and tail defined in Sect. 3. To learn such embeddings, the model is trained to minimize the loss:

$$\mathcal{L} = [\sum_{(\mathbf{h},\mathbf{r},\mathbf{t}) \in \mathcal{M}} \sum_{(\mathbf{h}',\mathbf{r},\mathbf{t}')} \gamma + d(\mathbf{h} + \mathbf{r}, \mathbf{t}) - d(\mathbf{h}' + \mathbf{r}, \mathbf{t}')] \tag{2}$$

where $\gamma > 0$ is a margin hyperparameter. The $\langle \mathbf{h}', \mathbf{r}, \mathbf{t}' \rangle$ is the corrupted triplets with either the *head* or *tail* replaced by a random entity $\in \mathcal{E}$. Once the model is well-trained, it can score each $\langle h, r, t\rangle$ based on the plausibility of the relationship expressed in the triplet being true. For each test triplet, we corrupt the h and t separately as two corruption lists. Then, the model ranks the test triplet against the corruption lists. The Hits@K score is defined as the ratio of the number of test triplets that are ranked in top-K, formally as follows:

$$Hits@K = \frac{\sum_i^{|\mathcal{Q}|} 1 \text{ if } rank_{\langle h_i, r_i, t_i\rangle} \leq K}{|\mathcal{Q}|} \tag{3}$$

$\mathcal{Q}$ is the test set of triplets, and each $\langle h_i, r_i, t_i\rangle$ triplet belongs to $\in \mathcal{Q}$. The $Hit@K$ score is in the range of [0,1]. A higher score indicates better performance. We compute this score across 11 different K. For each K, we get 10 $Hit@K$ scores using the 10-fold strategy and present their statistical distribution using a box plot (see Fig. 6). The figure reveals that the variance in $Hit@K$ scores for each K is consistently low, indicating the stability of the model's performance. Since the model is trained using 9 out of 10 folds, this suggests a consistency of triplets within TVCG-KG. Consider the case $K = 6$, the model successfully ranks 60% of test triplets at the top 6 of a list containing 40,000 candidates. This highlights the quality of the triplets that teach the model to make accurate predictions.

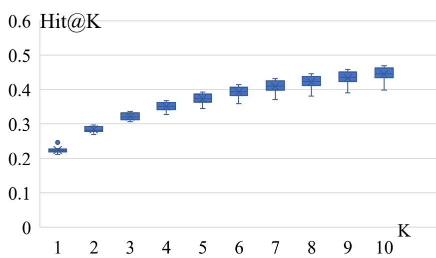


Figure 6: Statistical Distribution of 10 $Hit@K$ Scores for Each K Using a 10-fold Strategy.

6.2.2 Interlinking to External Knowledge Graphs

Interlinking assesses how much a KG can establish connections with other KGs [64]. KGs are preferred over other data formats due to their flexibility and scalability, enabling seamless integration and enrichment with new information. Therefore, the interlinking is also one important aspect of KG evaluation.

To evaluate the interlinking capabilities of TVCG-KG, we use Microsoft Academic Knowledge Graph (MAKG) as the target knowledge base for linkage. The reasons are three-folded: (1) The MAKG is one of the largest freely available scholars KG; (2) it contains over eight billion triplets with rich information; (3) it provides metadata for entities that are also contained in our TVCG-KG. We compare two important entity classes, *paper* and *author*. For each entity in TVCG-KG, we query the target entity from MAKG by matching the paper title or author name. It is achieved by calling the SPARQL endpoint of MAKG [4]. As a result, 89.07% of author entities and 73.89% of paper entities can be matched to the MAKG. On the one hand, this high matching ratio indicates the highly interconnected nature of TVCG-KG. On the other hand, the unmatched entities are mainly caused by the recent publications in TVCG. The continuous stream of research publications also highlights the importance of our end-to-end framework for the KG construction pipeline.

6.3 Task-based TVCG-KG Evaluation

One key advantage of TVCG-KG is that it directly captures the connections among various entities, reducing the time to collect and process data while inferring by the relations among the entities. In this section, we demonstrate the usefulness and effectiveness of the TVCG-KG data through several usage scenarios.

6.3.1 Implementation Details

We store TVCG-KG in Neo4j Aura, a cloud-based graph database service, and make it accessible for queries using the Cypher language. Our PRE-view demonstration is implemented entirely in JavaScript, including database queries, data processing, and visualization. The source code is available in our Git repository.

6.3.2 Author-driven analysis

Author-driven explorations are necessary in real-world decision-making, including tasks like locating potential collaborators and identifying suitable candidate reviewers. In this section, we demonstrate how TVCG-KG can facilitate author-driven explorations from multiple perspectives.

Author Profiling Analyzing TVCG authors can reveal intriguing insights, such as identifying research communities with strong collaboration and shared interests. To profile all TVCG authors, we utilize the TransE embeddings to generate an overview of author distributions. As introduced in Sect. 6.2.1, the TransE is trained on the triplets within TVCG-KG such that similar entities should be close in the embedding space. After projecting the high-dimensional embedding space to 2D space using t-SNE, neighboring entities indicate higher similarity. The scatterplot is shown in Fig. 7 (1). The presence of clusters in the plot validates that some collaborative groups among authors are well-captured in the TVCG-KG.

In this overview, authors can be highlighted and colored differently for various scenarios. To illustrate, we address a practical question from the Editor-in-Chief's (EIC) perspective: Do the current Associate Editors (AEs) on the editorial board have a comprehensive coverage of the TVCG topics?

To answer this question, we color authors based on their research areas within this overview. Initially, we collect information about AEs and categorize them into VIS, Graphics, and XR. We retrieve the author IDs of all the AEs and their collaborators using the query in Fig. 7 (1'). Authors are colored in blue (●), while AEs and their collaborators in the area of VIS (●), Graphics (●), and VR (●) are highlighted. In the figure, it is clear that the blue dots are nearly completely covered by highlighted dots. This suggests that AEs and their collaborators represent the TVCG area well.

Author Identification & Analysis We can filter and pinpoint target authors in the TVCG-KG in various ways. Furthermore, these methods can be interconnected to suit different usage scenarios. Here, we show an example of identifying top authors who have published papers on Virtual Reality, Augmented Reality, and Mixed Reality topics. To achieve it, an example query can be employed, as demonstrated in Fig. 7 (2'). Once we have the list of returned author names and their publication numbers, we visualize them in a bar chart, as depicted in Fig. 7 (2).

Author Examination & Summarization When comparing authors, users often need additional information in real-world scenarios to make informed decisions under various contexts. Our TVCG-KG offers multiple perspectives to summarize authors of interest. Following the previous example, when identifying the top 10 scholars in XR, it is essential to assess their research focus, activities, and impact before deciding to follow their work or collaborate with them. Our TVCG-KG simplified this process through automatic author profiling. Taking *Anatole Lecuyer* as an example, we can easily extract his collaboration network from TVCG-KG using a simple query (See Fig. 7 (3')). Visualizing this network in a node-link diagram (Fig. 7 (3)) reveals his active collaborations, including this closest collaborator, *Ferran Argelaguet*, another top 10 author in the field. We can also extract *Anatole Lecuyer*'s TVCG publications and present a timeline (Fig. 7 (4)). It shows increased activity after

[4] https://makg.org/sparql-endpoint/

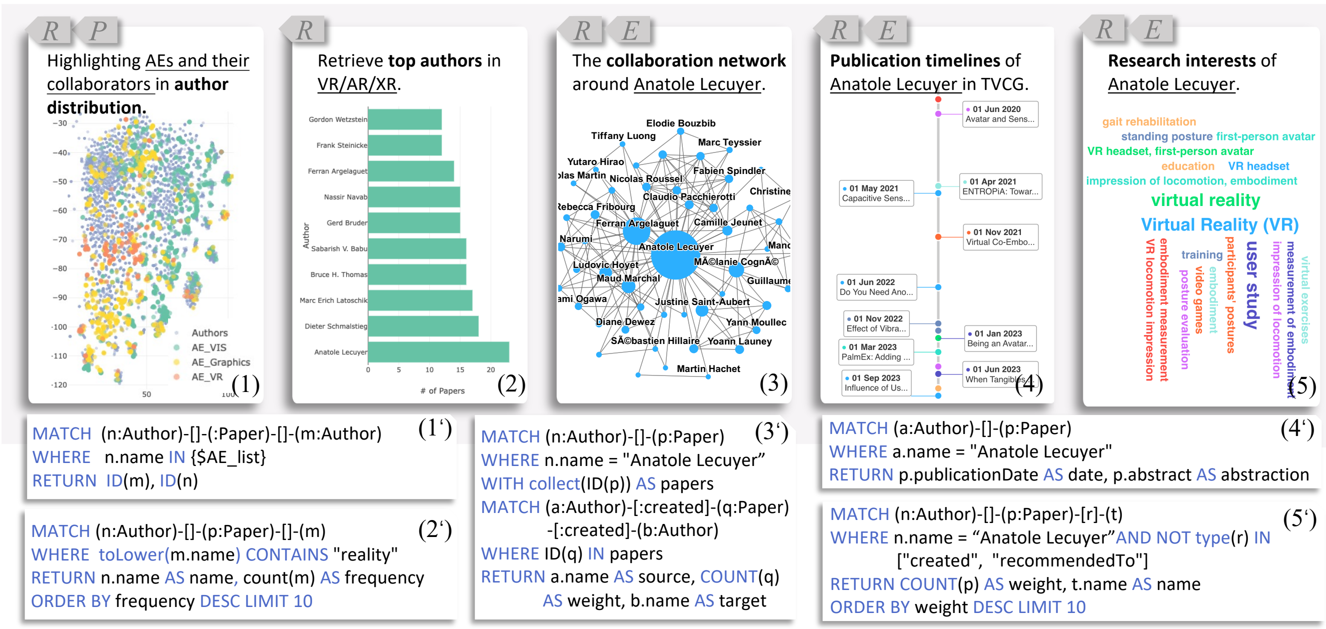


Figure 7: TVCG-KG supports various author-driven analysis tasks: (1) author profiling & overview; (2) author retrieval based on multiple conditions; automatic author examination and summarization from (3) collaboration network; (4) publication timeline, and (5) research interests. (1')-(5') are corresponding Cypher queries used to query TVCG-KG.

2021, suggesting he has become more active in TVCG recently. Additionally, we can traverse from his publications to semantic entities and create a word cloud to display his research interests (Fig. 7 (5)). Clearly, our TVCG-KG provides flexible and high-quality support for author-related tasks, including overviews, retrievals, and examinations.

6.3.3 Paper-driven analysis: Literature Review

Conducting a literature review is a fundamental task for researchers. It involves collecting and annotating relevant papers, and creating summaries. In this section, we illustrate how TVCG-KG can assist at each step through examples.

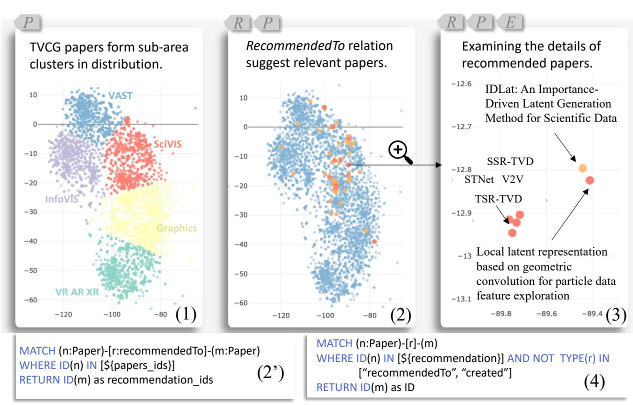


Figure 8: Profiling of TVCG Papers using TransE embeddings. (1) sub-area clusters; (2) recommending papers based on *recommendedTo* relation; (3) Zoomed in view from (2).

Profiling TVCG Paper Collections The initial step in our analysis is to profile the TVCG paper collection to build a basic understanding for further exploration. To accomplish this, we employ the TransE model to generate embeddings. Later, we applied the t-SNE algorithm for dimensionality reduction and the K-means algorithm for coloring purposes. We visualize the result in a scatterplot, as shown in Fig. 8 (1). We can manually check the paper titles within each cluster by hovering over them. Notably, we found that these clusters mainly align with sub-areas in TVCG, as labeled in the scatterplot. This finding implies that (1) the interrelationships among entities capture meaningful information, (2) the quality of these interrelationships is good such that even a basic model can effectively learn from them without confusion.

Paper Retrieval & Recommendation The TVCG-KG offers multiple ways to retrieve papers, such as keyword matching in titles or abstracts. After identifying several initial papers of interest, TVCG-KG enables paper recommendations through link traversal. The query for recommending through "recommendTo" relation is shown in Fig. 8 (2').

For illustration purposes, we selected a survey paper by Wang et al. [79] that has selected 19 papers from the TVCG. For a sanity check, we successfully retrieved all of them from the TVCG-KG by matching their paper titles. Furthermore, we manually verified the author lists, which showed a 100% alignment with the publications. Since the KG construction is accomplished through iterating each paper, these results prove the accuracy of our construction process. The cited papers are represented as red dots (●) in Fig. 8 (2-3), scattered around the *SciVIS* cluster, which also resonates with our earlier conclusion of the formed sub-areas.

By taking these cited papers as initial nodes, we traversed along the *RecommendTo* relation and arrived at a set of recommended *papers*, using the query in Fig. 8 (2'). The recommended papers are highlighted as orange dots (●) in Fig. 8 (2-3). From Fig. 8 (2), it is clear that the recommendations are close to the cited papers, scattered within the *SciVis* cluster. Zooming in on a specific area in Fig. 8 (2) reveals a more detailed view in Fig. 8 (3). To the left of Fig. 8 (3), we find four closely related publications, all focused on time-varying data analysis, suggesting the relationships between papers are well-captured in TVCG-KG such that models can successfully learn and map to distance. On the right, the two papers are related to the latent representation of scientific data. The recommended one in orange, i.e., IDLat [68], is a more recent paper not covered in this survey paper, but its topic aligns well with the survey paper. The finding indicates the usefulness of TVCG-KG in providing valuable recommendations and guidance when identifying related papers during the literature review.

Paper Labeling & Summarising During the TVCG-KG construction step, we pre-process the abstract of each paper and extract the important concepts as entities. This information is especially beneficial for labeling, categorizing, and summarizing papers. It can be easily accessed by querying the TVCG-KG without the need for data wrangling, as shown in Fig. 8 (4). To illustrate it, we retrieve all semantic entities connected to the cited papers in the same survey [79] and conduct two evaluation steps: one-to-one mapping for labeling and many-to-one mapping for summarization.

In the survey paper by Wang [79], the authors manually labeled the input data, output data, techniques, and tasks for each paper, which we consider as the ground truth. We then compare this ground truth with the entities queried from our TVCG-KG. The comparison of *input_data* is presented in Table 5. The table clearly illustrates a high alignment between the semantic entities with expert-labeled concepts, highlighted in green ( ). In other instances, our entities are more concrete, helping experts summarize general and broad concepts. For example, *cerebrovasculature data* can be linked to *3D volume data* in the context of medical imaging. In summary, the semantic entities within the TVCG-KG can significantly enhance summarization efficiency by providing essential information without the need to manually processing a large number of papers.

Table 5: The comparison of extracted entities from TVCG-KG with expert-labeled entities for papers cited in DL4SciVis [79].

TVCG-KG	Ground Truth
low-resolution volume sequences	two end volumes
low-resolution volume	low-resolution volume
low-resolution volumes	low-resolution two end volumes
fluid flow data	low-resolution flow map
single-view 2D medical images	3D/4D-CT projection or X-ray image
simulation parameters	simulation parameters
pairs of time steps of the source and target variable	source variable
collection of volume renderings	new viewpoint and transfer function
simulation and visualization parameters	ensemble simulation parameters
low resolution depth and normal field	low-resolution isosurface maps, optical flow
low-resolution input image	low-resolution image
spatiotemporal volumes	local spatiotemporal patch
volumetric data sets	intensity volume, opacity volume or transfer function
cerebrovasculature data	**3D volume patch**, multislice composited 2D MIP
brain volumes imaged using wide-field microscopy	batch of grayscale images
volumetric data	volume patch
large, unlabeled spatiotemporal scientific data	local spatiotemporal patches
collection of streamlines or stream surfaces generated from a flow field data set	streamline or stream surface
particle data	particle patch

The survey paper categorizes papers into five groups based on the designated *task*, which is taken as the ground truth here. To perform the comparison, we filtered the semantic entities by traversing the *has_task* relation of each paper and subsequently compared them with the ground truth. The goal is to evaluate whether the semantic entities contribute to the summarization of the group. Our findings indicate that the semantic entities provide enough context to summarize each group. Given the limited space for each group, we show the most closely related entities. Additional details and mappings can be found in our supplemental material.

1. *data generation*: time-varying data generation, spatiotemporal super-resolution, medical image reconstruction.
2. *vis generation*: volume rendering, upsampling, shading, parameter space exploration.
3. *prediction*: volumetric ambient occlusion prediction, complex behavior detection.
4. *object detection& segmentation*: vessel segmentation, image composition, segmentation.
5. *feature learning&extraction*: tracking, latent representation, feature extraction, interactive example-based queries

7 Discussion And Future Work

Compared to other data structures like traditional graphs or relational databases, knowledge graphs are attractive due to their high schema flexibility, easy data integration, and rich semantic encodings. We have demonstrated our TVCG-KG can be queried flexibly to support decision-making processes. However, we also identified several aspects that can be further improved in the future:

First, the abstract contains important yet somewhat limited information. The semantic entities can be further enriched by processing the full-text of papers. Second, due to the flexibility of the knowledge graph, TVCG-KG can be further extended by integrating text-based and image-based databases in the visualization field. Such a multi-modal approach could stimulate the exploration of various cross-modal tasks, such as improving representation, retrieval, and recommendation processes. These tasks, often requiring diverse data sources, cannot be effectively supported by any single database alone. Third, administrative-related information can be added and only visible to those with proper access. Information such as reviewer periods, submission dates, and reviewer identities can be integrated seamlessly. This would simplify administrative processes and enhance transparency.

8 Conclusion

This paper proposes KG-PRE-view to democratize a TVCG knowledge graph through visual explorations. We first construct a TVCG-based knowledge graph, improving the efficiency of the decision-making process by querying heterogeneous data from KG. Given the massive amount of information captured in TVCG-KG, we propose a PRE-view approach to incorporate visualization into the KG exploration pipelines. By applying PRE-view to TVCG-KG, we perform task-driven evaluations through multiple usage scenarios. In addition, we also evaluate the quality of TVCG-KG from several other aspects.

References

[1] Z. Abedjan, T. Grütze, A. Jentzsch, and F. Naumann. Profiling and mining rdf data with prolod++. In *2014 IEEE 30th International Conference on Data Engineering*, pp. 1198–1201. IEEE, 2014.

[2] K. Al-Khatib, Y. Hou, et al. End-to-end argumentation knowledge graph construction. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, pp. 7367–7374, 2020.

[3] M. Al Zayer, P. MacNeilage, and E. Folmer. Virtual locomotion: a survey. *IEEE transactions on visualization and computer graphics*, 26(6):2315–2334, 2018.

[4] K. Angerbauer, N. Rodrigues, R. Cutura, S. Öney, N. Pathmanathan, C. Morariu, D. Weiskopf, and M. Sedlmair. Accessibility for color vision deficiencies: Challenges and findings of a large scale study on paper figures. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–23, 2022.

[5] H. Bhatia, G. Norgard, V. Pascucci, and P.-T. Bremer. The helmholtz-hodge decomposition—a survey. *IEEE Transactions on visualization and computer graphics*, 19(8):1386–1404, 2012.

[6] N. Bressa, H. Korsgaard, A. Tabard, S. Houben, and J. Vermeulen. What's the situation with situated visualization? a survey and perspectives on situatedness. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):107–117, 2021.

[7] J. L. Cárdenas-Donoso, C. J. Ogayar, F. R. Feito, and J. M. Jurado. Modeling of the 3d tree skeleton using real-world data: a survey. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[8] P. Caserman et al. A survey of full-body motion reconstruction in immersive virtual reality applications. *IEEE transactions on visualization and computer graphics*, 26(10):3089–3108, 2019.

[9] P. Caserta and O. Zendra. Visualization of the static aspects of software: A survey. *IEEE transactions on visualization and computer graphics*, 17(7):913–933, 2010.

[10] H. Chen et al. An automatic literature knowledge graph and reasoning network modeling framework based on ontology and natural language processing. *Advanced Engineering Informatics*, 42:100959, 2019.

[11] J. Chen, M. Ling, R. Li, P. Isenberg, T. Isenberg, M. Sedlmair, T. Möller, R. S. Laramee, H.-W. Shen, K. Wünsche, et al. Vis30k: A collection of figures and tables from ieee visualization conference publications. *IEEE Transactions on Visualization and Computer Graphics*, 27(9):3826–3833, 2021.

[12] Q. Chen, S. Cao, J. Wang, and N. Cao. How does automation shape the process of narrative visualization: A survey of tools. *IEEE Transactions on Visualization and Computer Graphics*, 2023.

[13] D. Cohen-Or, Y. L. Chrysanthou, C. T. Silva, and F. Durand. A survey of visibility for walkthrough applications. *IEEE Transactions on Visualization and Computer Graphics*, 9(3):412–431, 2003.

[14] M. F. De Oliveira and H. Levkowitz. From visual data exploration to visual data mining: A survey. *IEEE transactions on visualization and computer graphics*, 9(3):378–394, 2003.

[15] M. E. Deagen et al. Fair and interactive data graphics from a scientific knowledge graph. *Scientific Data*, 9(1):1–11, 2022.

[16] D. Deng, Y. Wu, X. Shu, J. Wu, S. Fu, W. Cui, and Y. Wu. Visimages: A fine-grained expert-annotated visualization dataset. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[17] F. Diller, G. Scheuermann, and A. Wiebel. Visual cue based corrective feedback for motor skill training in mixed reality: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[18] R. Ding, S. Han, Y. Xu, H. Zhang, and D. Zhang. Quickinsights: Quick and automatic discovery of insights from multi-dimensional data. In *Proceedings of the 2019 International Conference on Management of Data*, pp. 317–332, 2019.

[19] V. Domova and K. Vrotsou. A model for types and levels of automation in visual analytics: a survey, a taxonomy, and examples. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[20] G. M. Draper, Y. Livnat, and R. F. Riesenfeld. A survey of radial methods for information visualization. *IEEE transactions on visualization and computer graphics*, 15(5):759–776, 2009.

[21] S. S. e Zainab, M. Saleem, Q. Mehmood, D. Zehra, S. Decker, and A. Hasnain. Fedviz: A visual interface for sparql queries formulation and execution. In *VOILA@ ISWC*, p. 49, 2015.

[22] M. Espadoto et al. Toward a quantitative survey of dimension reduction techniques. *IEEE transactions on visualization and computer graphics*, 27(3):2153–2173, 2019.

[23] P. Federico et al. A survey on visual approaches for analyzing scientific literature and patents. *IEEE transactions on visualization and computer graphics*, 23(9):2179–2198, 2016.

[24] C. G. Fidalgo, Y. Yan, H. Cho, M. Sousa, D. Lindlbauer, and J. Jorge. A survey on remote assistance and training in mixed reality environments. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2291–2303, 2023.

[25] A. Fonnet et al. Survey of immersive analytics. *IEEE transactions on visualization and computer graphics*, 27(3):2101–2122, 2019.

[26] A. Genay, A. Lécuyer, and M. Hachet. Being an avatar "for real": a survey on virtual embodiment in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):5071–5090, 2021.

[27] J. Gómez-Romero, M. Molina-Solana, A. Oehmichen, and Y. Guo. Visualizing large knowledge graphs: A performance analysis. *Future Generation Computer Systems*, 89:224–238, 2018.

[28] J. Grubert et al. A survey of calibration methods for optical see-through head-mounted displays. *IEEE transactions on visualization and computer graphics*, 24(9):2649–2662, 2017.

[29] Y. Guo, S. Guo, Z. Jin, S. Kaul, D. Gotz, and N. Cao. Survey on visual analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):5091–5112, 2021.

[30] H. Hao, Y. Cui, Z. Wang, and Y.-S. Kim. Thirty-two years of ieee vis: Authors, fields of study and citations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1016–1025, 2022.

[31] W. He, J. Wang, H. Guo, K.-C. Wang, H.-W. Shen, M. Raj, Y. S. Nashed, and T. Peterka. Insitunet: Deep image synthesis for parameter space exploration of ensemble simulations. *IEEE transactions on visualization and computer graphics*, 26(1):23–33, 2019.

[32] X. He, R. Zhang, R. Rizvi, et al. Aloha: developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design. *BMC medical informatics and decision making*, 19(4):1–18, 2019.

[33] I. Herman, G. Melançon, and M. S. Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on visualization and computer graphics*, 6(1):24–43, 2000.

[34] F. Hohman et al. Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE transactions on visualization and computer graphics*, 25(8):2674–2693, 2018.

[35] K. Hu, S. Gaikwad, M. Hulsebos, M. A. Bakker, E. Zgraggen, C. Hidalgo, T. Kraska, G. Li, A. Satyanarayan, and Ç. Demiralp. Viznet: Towards a large-scale visualization learning and benchmarking repository. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–12, 2019.

[36] P. Isenberg, F. Heimerl, S. Koch, T. Isenberg, P. Xu, C. Stolper, M. Sedlmair, J. Chen, T. Möller, and J. Stasko. vispubdata.org: A Metadata Collection about IEEE Visualization (VIS) Publications. *IEEE Transactions on Visualization and Computer Graphics*, 23, 2017. To appear. doi: 10.1109/TVCG.2016.2615308

[37] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller. Visualization as seen through its research paper keywords. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):771–780, 2016.

[38] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller. Visualization as seen through its research paper keywords. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):771–780, 2017. doi: 10.1109/TVCG.2016.2598827

[39] B. Jiang, X. You, K. Li, T. Li, X. Zhou, and L. Tan. Interactive analysis of epidemic situations based on a spatiotemporal information knowledge graph of covid-19. *IEEE Access*, 10:46782–46795, 2022. doi: 10.1109/ACCESS.2020.3033997

[40] M. W. Jones, J. A. Baerentzen, and M. Sramek. 3d distance fields: A survey of techniques and applications. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):581–599, 2006.

[41] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE transactions on visualization and computer graphics*, 19(3):495–513, 2012.

[42] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pp. 1–5, 2006.

[43] H. Li, G. Appleby, C. D. Brumar, R. Chang, and A. Suh. Characterizing the users, challenges, and visualization needs of knowledge graphs in practice. *arXiv preprint arXiv:2304.01311*, 2023.

[44] M. Lissandrini et al. Knowledge graph exploration systems: are we lost. In *CIDR*, vol. 22, pp. 10–13, 2022.

[45] S. Liu, X. Wang, C. Collins, W. Dou, F. Ouyang, M. El-Assady, L. Jiang, and D. A. Keim. Bridging text visualization and mining: A task-driven survey. *IEEE transactions on visualization and computer graphics*, 25(7):2482–2504, 2018.

[46] Y. Luan, L. He, M. Ostendorf, and H. Hajishirzi. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. *arXiv preprint arXiv:1808.09602*, 2018.

[47] Y. Luan, D. Wadden, L. He, A. Shah, M. Ostendorf, and H. Hajishirzi. A general framework for information extraction using dynamic span graphs. *arXiv preprint arXiv:1904.03296*, 2019.

[48] T. Luong, A. Lecuyer, N. Martin, and F. Argelaguet. A survey on affective and cognitive vr. *IEEE transactions on visualization and computer graphics*, 28(12):5154–5171, 2021.

[49] E. Marchand, H. Uchiyama, and F. Spindler. Pose estimation for augmented reality: a hands-on survey. *IEEE transactions on visualization and computer graphics*, 22(12):2633–2651, 2015.

[50] A. M. McNutt. No grammar to rule them all: A survey of json-style dsls for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 1912.

[51] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[52] I. Mondal, Y. Hou, and C. Jochim. End-to-end nlp knowledge graph construction. *arXiv preprint arXiv:2106.01167*, 2021.

[53] K. Moreland. A survey of visualization pipelines. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):367–378, 2012.

[54] A. Moro, A. Raganato, and R. Navigli. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics*, 2:231–244, 2014.

[55] R. Nararatwong, N. Kertkeidkachorn, and R. Ichise. Knowledge graph visualization: Challenges, framework, and implementation. In *2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pp. 174–178, 2020. doi: 10.1109/AIKE48582.2020.00034

[56] A. Narechania, A. Karduni, R. Wesslen, and E. Wall. Vitality: Promoting serendipitous discovery of academic literature with transformers & visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):486–496, 2021.

[57] R. Navigli and S. P. Ponzetto. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial intelligence*, 193:217–250, 2012.

[58] A. Pandey, U. H. Syeda, C. Shah, J. A. Guerra-Gomez, and M. A. Borkin. A state-of-the-art survey of tasks for tree design and evaluation with a curated task dataset. *IEEE Transactions on Visualization and Computer Graphics*, 28(10):3563–3584, 2021.

[59] M. A. Pellegrino et al. Move cultural heritage knowledge graphs in everyone's pocket. *Semantic Web*, (Preprint):1–37, 2020.

[60] B. Preim et al. Survey of the visual exploration and analysis of perfusion data. *IEEE transactions on visualization and computer graphics*, 15(2):205–220, 2008.

[61] Y. Qiu et al. Tax-kg: Taxation big data visualization system for knowledge graph. In *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*, pp. 425–429. IEEE, 2020.

[62] G. J. Quadri et al. A survey of perception-based visualization studies by task. *IEEE transactions on visualization and computer graphics*, 2021.

[63] M. Rautenhaus et al. Visualization in meteorology—a survey of techniques and tools for data analysis tasks. *IEEE Transactions on Visualization and Computer Graphics*, 24(12):3268–3296, 2017.

[64] D. Ringler and H. Paulheim. One knowledge graph to rule them all? analyzing the differences between dbpedia, yago, wikidata & co. In *KI 2017: Advances in Artificial Intelligence: 40th Annual German Conference on AI, Dortmund, Germany, September 25–29, 2017, Proceedings 40*, pp. 366–372. Springer, 2017.

[65] H.-J. Schulz, S. Hadlak, and H. Schumann. The design space of implicit hierarchy visualization: A survey. *IEEE transactions on visualization and computer graphics*, 17(4):393–411, 2010.

[66] S. Seo, H. Cheon, H. Kim, and D. Hyun. Structural quality metrics to evaluate knowledge graph quality.

[67] M. Sereno, X. Wang, L. Besançon, M. J. Mcguffin, and T. Isenberg. Collaborative work in augmented reality: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2530–2549, 2020.

[68] J. Shen, H. Li, J. Xu, A. Biswas, and H.-W. Shen. Idlat: An importance-driven latent generation method for scientific data. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):679–689, 2022.

[69] L. Shen et al. Towards natural language interfaces for data visualization: A survey. *IEEE transactions on visualization and computer graphics*, 2022.

[70] D. Shi, X. Xu, F. Sun, Y. Shi, and N. Cao. Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):453–463, 2020.

[71] H. Shiravi, A. Shiravi, and A. A. Ghorbani. A survey of visualization systems for network security. *IEEE Transactions on visualization and computer graphics*, 18(8):1313–1329, 2011.

[72] M. Šik and J. Křivánek. Survey of markov chain monte carlo methods in light transport simulation. *IEEE transactions on visualization and computer graphics*, 26(4):1821–1840, 2018.

[73] A. Soylu, E. Kharlamov, et al. Optiquevqs: A visual query system over ontologies for industry. *Semantic Web*, 9(5):627–660, 2018.

[74] Y. Sun, J. Yu, and M. Sarwat. Demonstrating spindra: A geographic knowledge graph management system. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pp. 2044–2047, 2019. doi: 10.1109/ICDE.2019.00235

[75] G. K. Tam, Z.-Q. Cheng, Y.-K. Lai, F. C. Langbein, Y. Liu, D. Marshall, R. R. Martin, X.-F. Sun, and P. L. Rosin. Registration of 3d point clouds and meshes: A survey from rigid to nonrigid. *IEEE transactions on visualization and computer graphics*, 19(7):1199–1217, 2012.

[76] X. Tian et al. A survey of smooth vector graphics: Recent advances in representation, creation, rasterization and image vectorization. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[77] M. D. L. Tosi and J. C. Dos Reis. Scikgraph: A knowledge graph approach to structure a scientific field. *Journal of Informetrics*, 15(1):101109, 2021.

[78] H. Vargas et al. Rdf explorer: A visual sparql query builder. In *The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part I 18*, pp. 647–663. Springer, 2019.

[79] C. Wang and J. Han. Dl4scivis: A state-of-the-art survey on deep learning for scientific visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[80] J. Wang, S. Hazarika, C. Li, and H.-W. Shen. Visualization and visual analysis of ensemble data: A survey. *IEEE transactions on visualization and computer graphics*, 25(9):2853–2872, 2018.

[81] L. Wang, M. Huang, R. Yang, H.-N. Liang, J. Han, and Y. Sun. Survey of movement reproduction in immersive virtual rehabilitation. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[82] Q. Wang, Z. Chen, Y. Wang, and H. Qu. A survey on ml4vis: Applying machine learning advances to data visualization. *IEEE transactions on visualization and computer graphics*, 28(12):5134–5153, 2021.

[83] Y. Wang et al. Datashot: Automatic generation of fact sheets from tabular data. *IEEE transactions on visualization and computer graphics*, 26(1):895–905, 2019.

[84] K. Ward et al. A survey on hair modeling: Styling, simulation, and rendering. *IEEE transactions on visualization and computer graphics*, 13(2):213–234, 2007.

[85] J. Wei et al. Vision-kg: topic-centric visualization system for summarizing knowledge graph. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pp. 857–860, 2020.

[86] C. Wise, V. N. Ioannidis, et al. Covid-19 knowledge graph: accelerating information retrieval and discovery for scientific literature. *arXiv preprint arXiv:2007.12731*, 2020.

[87] A. Wu, Y. Wang, X. Shu, D. Moritz, W. Cui, H. Zhang, D. Zhang, and H. Qu. Ai4vis: Survey on artificial intelligence approaches for data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[88] Y. Ye, R. Huang, and W. Zeng. Visatlas: An image-based exploration and query system for large visualization collections via neural image embedding. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[89] W. Yun, X. Zhang, Z. Li, H. Liu, and M. Han. Knowledge modeling: A survey of processes and techniques. *International Journal of Intelligent Systems*, 36(4):1686–1720, 2021.

[90] L. Zhou and C. D. Hansen. A survey of colormaps in visualization. *IEEE transactions on visualization and computer graphics*, 22(8):2051–2069, 2015.

[91] Z. Zhou, C. Shi, M. Hu, and Y. Liu. Visual ranking of academic influence via paper citation. *Journal of Visual Languages & Computing*, 48:134–143, 2018.