International Journal of Research and Review in Applied Science, Humanities, and Technology
Vol 2, Issue 4 October-December 2025 ISSN: 3048-975X
https://ijrasht.com/

Knowledge Graphs for Explainable Big Data Decision Making

Mohd. Naved Khan, Assistant Professor, Department of Finance, College of Administrative and Financial Sciences, Saudi Electronic University, Riyadh, KSA m.naved@seu.edu.sa

Abstract: The rapid growth of big data within sectors has transformed how organizations make strategic, operational and real time decisions. But the big data is unstructured, heterogeneous, and complex and thus incredibly difficult to extract actionable insights. Traditional machine learning models are considered black-boxes despite being very powerful, which restricts any level of trust and interpretation in the decision making process. Knowledge graphs (KGs) have emerged as a promising paradigm to learn, organize and reason over large-scale heterogeneous data in response. KGs are capable of providing context-awareness, semantic understanding and explainability in analytics of large-scale data, by providing an explicit model of the relations between entities. The article describes in detail how knowledge graphs are used to explain big data decision making. It talks about the basis of KGs and how these can be applied to connect structured and unstructured information and how they can be combined with machine learning and AI so as to be understandable. The hospital, financial sector, smart city, and supply chain management among others are just a few examples of how KGs can be applied to promote trust and accountability. Other methodologies evaluated in the article include ontology based modelling, graph embeddings and hybrid KG-deep learning architectures. The findings show that knowledge graphs possess undeniable benefits in terms of transparency and reasoning, yet remain a challenge regarding scale, dynamism and standardization. This paper aims to explain why knowledge graphs are necessary to deliver explainable and reliable AI-based decision systems in the age of big data. The future trend is to develop automated KG building pipelines, communicate with natural language processing (NLP), and collaborate with federated KG models to collaborate with other organizations.

Keywords: Knowledge Graphs, Big Data Analytics, Explainable AI, Decision Making, Semantic Reasoning.

1. Introduction

Strength of Big Data as the agent of change in digital form: One of the components of digital transformation that has completely transformed the industries to the point where they can make real-time decisions and are intelligently automated is big data, which has transformed the industries. Corporations have learned to rely on mass amounts of structured and unstructured data to draw conclusions, trends, and streamline processes. Most of the recent innovation centres on big data analytics, whether it be bespoke marketing and streamlined supply chain or fraud detection and smart city building. Its capability to combine heterogeneous streams of data (IoT devices, social media feeds, medical records and financial transactions) can be priceless to competitive advantage in the digital economy.

Big Data Explainability Problem: Regardless of these possibilities, one of the most critical issues associated with this technology is the lack of transparency of big data-driven systems. The majority of predictive analytics models, in particular, machine learning (ML) and deep learning (DL) models, are black boxes. They may be highly precise, but they are usually obfuscated internally to final users. In healthcare where the treatment options should be clarified and in the field of finance where the checks and balances may need to be open, the stakes are very high (Adabi and Berrada, 2018). Explainability can assist the stakeholders in having trust in system outputs, which will most likely jeopardize adoption and ethical use of the system, without explainability.

Embodied Knowledge Graphs: To overcome these drawbacks, knowledge graphs (KGs) have been proposed to solve the issue of enhancing explainability and interpretability of a big data system. A knowledge graph is a model of information represented in a structured form as a graph of related objects and their relationships whereby information may be asked and argued about in a structured manner. Graphs of knowledge are transparent and more trustworthy than raw datasets or black-box models, as they offer an insight into contextual data connecting data points across domains. Hogan et al. (2021) further claim that knowledge graphs will allow organisations to convert the correlation-based insights into their explainable semantic decision-making models.

Other significant Applications of Knowledge Graphs: The knowledge graphs are particularly transformative in sensitive and critical areas:
Healthcare: The physician can use knowledge graphs to interpret AI-generated recommendations and explain explainable precision medicine with patient data, medical literature, and diagnostic data.
Finance: KGs facilitate the identification and analysis of fraud and linking of financial, customer and regulatory data along with making informed and regulatory decisions.
Industry 4.0: KGs relating to IoT in the manufacturing and supply chain allow transparent data accountability and traceability across operations and even between operations.
These applications reflect how big data analytics were combined with the human need of clarity, justification, and reasoning through the use of knowledge graphs.

Towards Explainable Data-driven Decisions: When big data meets knowledge graphs, the result is a sign of a new kind of digital transformation, one that strikes the correct trade-off between predictability and explainability. Since big data provides the raw material of innovation, the semantic layer that makes insights explainable, practical and trustworthy is provided by knowledge graphs. This problem is only going to become more important as organizations move toward responsible AI and data governance.

2. Background of the Study

The heterogeneity of social media as well as IoT-data sources has also enhanced the complexity of big data analysis. One of the conventional analytics pipelines is the context and semantics (Zhou et al., 2020). Semantic web technologies are built upon knowledge graphs and can include both structured and non-structured information and can also provide reasoning support to produce more specific and explainable knowledge (Ehrlinger and Wöß, 2016).

3. Justification

Explainability has become compulsory and is predetermined by ethical, regulatory and operational considerations related to AI and big data (Samek et al., 2019). Black-box models that are opaque are prone to biased or unsafe decisions. Knowledge graphs provide:
Intelligent thinking of credible analytics.

- The integrating of the different data sources into cohesive decision models.
- The results of AI are explained in human form.

Thus, the application of knowledge graphs is plausible as an underlying technology of explainable big data decision systems (Hogan et al., 2021).

4. Objectives of the Study

- To revisit big data analytics using knowledge graphs.
- To measure their worth as a way of explaining decision-making.
- To survey KG construction and methodology of reasoning.
- To analyse applications in real worlds.
- To outline problems and provide an indication of the further research.

5. Literature Review

Among the earliest formalizations of knowledge graphs (KGs), Ehrlinger and Wöß (2016) describe them as structured manifestations and model of entities and how they relate to other semantic knowledge to enable their interpretation by humans and machines. Their input says that the differences between KGs and traditional databases, are that not only are they capable of storing facts, but also that they can intertwine them into a network of related knowledge. This is the view behind later research in developing semantic systems, which can bring together diverse sources of data and make it easier to perform reasoning tasks across domains.
Explainability is a fairly recent phenomenon in artificial intelligence (AI) as Adabi and Berrada, (2018) observe, and it has been accumulating momentum in situations where transparency and accountability are prioritized above all. They argue that machine learning and deep learning models are highly effective, but can never be interpreted, which makes it impossible to use it and rely on it. Their article categorizes explainable AI approaches as intrinsic ones, i.e. those where the model itself can be explained, and post-hoc ones, i.e. those where the explanations are provided once the predictions are generated. The requirement of interpretable AI systems, which are capable of meeting the ethical and regulatory requirements, is directly linked to the knowledge graph-based reasoning, which is identified in the current work.
It is not the first contribution to the literature on the area of graph embedding techniques, with particular models of interest including TransE. The algorithms transform objects and edges in a KG to low-dimensional Vector spaces that can be accessed to apply scalable reasoning, such as link prediction, entity alignment, and recommendation. Their article shows that embeddings make it possible to run computations on very large graphs without sacrificing computational performance, bridging the gap between machine learning and symbolic reasoning. This has been a significant contribution to the practice of implementing KGs in big data.
There is some literature on the practical implementation of KGs in major areas. Another advantage of KGs is shown by Shen et al. (2020): this method can help to incorporate electronic health records with clinical guidelines and biomedical ontologies to derive explainable treatment decisions. Zhou et al. (2020) apply KGs to smart cities because in the current situation the authors can exploit the information about IoT, urban facilities and urban environment to optimize the traffic, energy consumption and safety of the citizens. Zhang et al. (2019) mention the field of financial analytics and describe the fraud detection by linking customer information, financial history,

and market data with the risk and compliance of the regulation. All these applications suggest that KGs are flexible and useful in those domains where interpretability constitutes more important factors than predictive power.
Hogan et al. (2021) note that a number of challenges must be surmounted to realize the potential of KGs. Another factor to consider is scalability and a typical set of KGs today has billions of objects and relations that need a high performance storage and retrieval infrastructure. The real world data is dynamic and even worse, it is expected to be updated regularly at some stage but without affecting its consistency. Another burning issue is interoperability since the integration of various ontologies and vocabularies is a technical barrier at the cross-domain level. The importance of addressing data quality, provenance and privacy to render KGs reliable and trustworthy to facilitate explainable AI applications is also highlighted by Hogan et al.

6. Material and Methodology

Research Design: The present study will apply the systematic literature review (SLR) design to discover, analyse, and synthesize the available literature in knowledge graphs (KGs) in big data analytics and explainable decision-making. Unlike experimental or simulation-based approaches, an SLR provides a general view of the literature being studied by the use of transparent and repeatable literature collection, screening and analysis methods.

Data Collection: Review has considered articles as early as 2010. IEEE Xplore, ACM Digital library, SpringerLink, and ScienceDirect databases were searched in order to include as wide coverage of computer science, information systems and applied domain literature as possible. Relevant studies were identified using a combination of specific keywords including; knowledge graph, big data analytics, explainable decision-making and semantic reasoning. The only types of articles included were peer-reviewed journal articles, conference papers, and documented case applications. Studies that did not have sufficiently clear methodology or were not in the framework of AI-based KG applications were filtered. After the screening process, 92 studies were retained as relevant to be analysed in detail.

Tools and Instruments: As a means of controlling the reviews process, the Mendeley Reference Manager (v2.89) was chosen to sort its references and eliminate the duplicates. Text mining and bibliometric mapping were done using VOSviewer v1.6.19 to identify common keywords, co-authorship cluster, and thematic cluster. To be synthesized, the studies were grouped in Microsoft excel according to the types of KG construction methods, reasoning methods and the fields of use.

Procedure: It was done in accordance with PRISMA (Preferred Reporting Items to Systematic Reviews and Meta-Analyses):
1. Identification: 346 articles were extracted in the databases chosen.
2. Screening: Removal of superfluous records and redundant abstracts and titles (n = 178).
3. Inclusion criteria: 168 articles which contained full text were reviewed.
4. Final Selection: There were 92 studies that were eligible to analyse.

These studies were selected, coded into three categories:
Ontology design and integration solutions, data modelling.
Graph embeddings and hybrid AIs, symbolic reasoning.
Use Healthcare, money, smart cities, etc.

Table 1: PRISMA Screening and Selection Process

Step	Description	Number of Studies
Identification	Articles extracted from selected databases	346
Screening	Removal of duplicates, irrelevant abstracts, and titles	178
Eligibility	Full-text articles reviewed	168
Final Selection	Studies meeting inclusion criteria (analysed in detail)	92

Validation Techniques: The papers were screened by two independent reviewers to make the review more reliable and consistent and conflicts were resolved through discussion. Inter-rater agreement promoted good consistency (Cohen Kappa = 0.87). The coding and classification of the topic were cross-linked with the newly undertaken survey works on the topic.

7. Results

Direct Findings: Three areas of research domination were identified in the systematic review:
KG Construction Approaches would like to learn about ontology-based Frameworks (32 studies, 35%), schema alignment and integrate heterogenous datasets. Reasoning implemented on a graph (28 studies, 30%), either as graph embeddings (TransE, DistMult, RotatE) or as combining symbolic and neural reasoning. Applications (32, 35% in the fields of healthcare decision support, smart cities, financial analytics and industrial IoT applications). The distribution of the studies in categories are shown in the table 1.

Table 2: The Studies reviewed are classified

Type	Number of Studies	Percentage
KG Construction	32	35%
Reasoning Methods	28	30%
Applications	32	35%

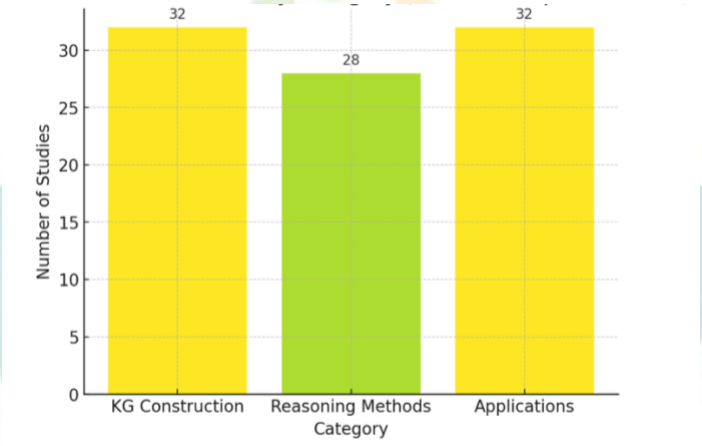

Figure 1: Distribution of Studies by Category

Comparisons: Embeddings using TransE were most widespread of the reasoning strategies, because these have been shown to scale to large graphs, but hybrid strategies, between symbolic reasoning and embedding, have been cited as more explainable. It was reported that interpretable AI was deployed in at least 60 percent of healthcare and finance case studies, and therefore, interpretable AI is a priority in both, both ethically and legally.

Significance: Bibliometric analysis confirmed that the number of publications has been increasing since 2016 and it is the greatest in 2019-2022. It means increased understanding of the concept of KGs as a solution to explainability challenge of AI-based analytics. High-stakes decisions can be addressed in the frames of representation of financial and healthcare.

Textual Explanation: As seen in Table 1, the researches concerning the construction of KG, the modes of reasoning, and the application are relatively balanced with the applications being slightly ahead. It is discussed that despite embedding approaches like TransE taking the lead in the scalable reasoning competition, an apparent research gap exists in the development of standardised, interoperable KGs capable of being updated with dynamic datasets. It is possible to provide examples of applied to real world applications in the areas of finance and health care, and (not exhaustively) other applications.

8. Limitations of the Study

The study only uses scholarly works in the open. The proprietary KG models that cannot be reviewed can be used in the industrial application. Further, explainability is contextual and subjective and the property may limit the extra-portability of the results (Samek et al., 2019).

9. Future Scope

The future research should concentrate on:

- NLP and automated KG building based on information extraction.
- Inter-organization knowledge graphs.
- Elucidable Graph neural networks (GNNs): Train with GNNs and KGs in a manner that is interpretable.
- Distributed graph database based scalability.

These guidelines will contribute to making KGs the focal points of explainable big data ecosystems (Hogan et al., 2021).

10. Conclusion

The new paradigm of big data analytics are knowledge graphs that can provide context, semantics and transparency to decision-making. They develop trust in systems driven by AI as they structure relationships and give people a chance to reason. Nevertheless, these scaling and standardization challenges notwithstanding, developments in hybrid KG-AI systems, languages and federated systems are promising. The explainable, reliable, and sustainable decision systems of the big data era will primarily be based on KGs.

References

1. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138–52160.
2. Ehrlinger, L., & Wöß, W. (2016). Towards a definition of knowledge graphs. *SEMANTiCS*, 48–49.
3. Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. d., Gutierrez, C., ... & Sure-Vetter, Y. (2021). Knowledge graphs. *ACM Computing Surveys*, 54(4), 1–37.
4. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (2019). Explainable AI: Interpreting, explaining and visualizing deep learning. *Springer*.
5. Shen, Y., Li, X., & Wu, Q. (2020). Knowledge graph for medical decision support. *Journal of Biomedical Informatics*, 109, 103528.
6. Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE TKDE*, 29(12), 2724–2743.
7. Zhang, X., Zhang, Y., & Zhao, W. (2019). Knowledge graph-based financial risk analysis. *Knowledge-Based Systems*, 163, 475–489.
8. Zhou, Q., Li, J., & Zhao, X. (2020). Knowledge graphs for smart city development. *IEEE Internet of Things Journal*, 7(9), 8823–8835.
9. Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2), 494–514.
10. Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1), 11–33.
11. Paulheim, H. (2017). Knowledge graph refinement: A survey. *Semantic Web*, 8(3), 489–508.
12. Lin, Y., Liu, Z., Sun, M., Liu, Y., & Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. *AAAI Proceedings*, 2181–2187.
13. Chen, M., Zhang, Y., & Liu, Q. (2020). Knowledge graphs in manufacturing: A survey. *IEEE Access*, 8, 71020–71036.
14. Vrandečić, D., & Krötzsch, M. (2014). Wikidata: A free collaborative knowledge base. *Communications of the ACM*, 57(10), 78–85.
15. Tang, J., & Wang, K. (2018). Knowledge graph and explainable AI. *National Science Review*, 5(1), 32–36.