BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Ethical and Privacy Issues in Knowledge Graphs

ASSIGNMENT 1 PART-B
LITERATURE SURVEY

Group Number: 26
- DIVYANSHU SINGH (2022AC05011)
- SOURAV PANI (2022AC05206)
- SWARAJYA RANJAN BHANJA (2022AC05599)
- BALAJI T (2022AC05612)

Abstract

Knowledge graphs (KGs) are powerful tools for organizing and analyzing data, but they raise significant ethical and privacy concerns. This survey explores current research on these issues, focusing on bias, fairness, transparency, accountability, consent, data leakage, ownership, and re-identification risks. It reviews technical solutions like fairness-aware algorithms, explainable AI, differential privacy, and federated learning, as well as ethical frameworks and interdisciplinary approaches. Addressing these challenges is crucial for responsibly developing and deploying knowledge graphs, ensuring their benefits without compromising individual rights and privacy.

Introduction

Knowledge graphs (KGs) are structured representations of knowledge that integrate information from diverse sources into a graph format, where entities are nodes and relationships are edges. These graphs are widely used in applications such as search engines, recommendation systems, and natural language processing. However, the use of knowledge graphs raises significant ethical and privacy concerns. This literature survey aims to explore these issues and the current state of research in this area.

Knowledge graphs have become a crucial technology in organizing and utilizing vast amounts of data. They enable the creation of interconnected data points that can be easily queried and analyzed. Google, for instance, uses a knowledge graph to enhance search results by providing contextually relevant information.

Ethical Issues in Knowledge Graphs

a. Bias and Fairness

One major ethical concern is bias in knowledge graphs. Since KGs are often built from human-generated data, they can inherit and even amplify existing biases.

- **Source of Bias**: Data sources may contain inherent biases based on the demographics of the data creators.
- **Impact**: Biased knowledge graphs can lead to unfair outcomes in applications like job recommendation systems or automated decision-making.
- **Mitigation**: Research suggests using fairness-aware algorithms and auditing mechanisms to detect and mitigate biases.

b. Transparency and Accountability

Transparency in how knowledge graphs are constructed and used is critical.

- **Challenges**: Proprietary algorithms and data sources can make it difficult to understand how decisions are made.
- **Solutions**: Efforts are being made to develop explainable AI (XAI) techniques that can provide insights into the decision-making process of systems using knowledge graphs.

c. Consent and Autonomy

The inclusion of personal data in knowledge graphs raises concerns about user consent and autonomy.

- **Issues**: Individuals may not be aware that their data is being used or may not have consented to its use in certain ways.
- **Research**: Studies advocate for better data governance frameworks and mechanisms for obtaining and managing user consent.

Privacy Issues in Knowledge Graphs

a. Data Leakage

Knowledge graphs can inadvertently expose sensitive information.

- **Risks**: Linking different data sources can lead to unintended inferences about individuals.
- **Techniques**: Differential privacy and anonymization techniques are being researched to protect sensitive data in knowledge graphs.

b. Data Ownership and Control

The question of who owns and controls the data in knowledge graphs is complex.

- **Ownership**: Individuals might not have control over how their data is used once it is integrated into a knowledge graph.
- **Solutions**: Proposals include data provenance tracking and giving individuals more control over their data through decentralized knowledge graph frameworks.

c. Re-identification Risks

Even anonymized data can sometimes be re-identified when integrated into a knowledge graph.

- **Challenges**: Combining data from multiple sources increases the risk of re-identification.
- **Approaches**: Research focuses on developing robust anonymization methods and monitoring systems to detect and prevent re-identification.

Current Research Trends in Ethical and Privacy Issues in Knowledge Graphs

a. Privacy-Preserving Techniques

- **Homomorphic Encryption**: Homomorphic encryption allows computations to be performed on encrypted data without needing to decrypt it first. This ensures data privacy is maintained throughout the processing pipeline.

  Example: Research by Acar et al. (2018) demonstrates the application of homomorphic encryption in securing cloud-based machine learning systems, which can be extended to protect sensitive data in knowledge graphs

- **Federated Learning**: Federated learning enables the training of machine learning models across multiple decentralized devices or servers holding local data samples, without exchanging them.

  Example: Google's Gboard, a virtual keyboard, utilizes federated learning to improve next-word prediction models without compromising user privacy.

b. Ethical Frameworks and Guidelines

- **Development of Guidelines**: There is an increasing emphasis on developing comprehensive ethical guidelines for the construction and use of knowledge graphs.

  Example: The European Commission's guidelines on trustworthy AI include principles relevant to knowledge graphs, such as transparency, accountability, and privacy.

- **Case Studies**: Analyzing real-world applications to identify best practices and potential pitfalls is crucial in developing robust ethical frameworks.

  Example: Studies on the use of knowledge graphs in healthcare highlight the importance of consent and transparency, as these applications involve sensitive personal data.

c. Interdisciplinary Approaches

- **Collaboration**: Combining insights from computer science, law, and ethics to address the multifaceted issues related to knowledge graphs.
- **Workshops and Conferences**: Dedicated forums for discussing the ethical and privacy implications of knowledge graphs.

d. Bias Detection and Mitigation

- **Auditing Mechanisms**: Developing tools and techniques to audit knowledge graphs for bias and ensure fairness.

  Example: The Linked Data for Fairness project aims to create tools for detecting and mitigating bias in knowledge graphs.

- **Fairness-Aware Algorithms**: Research focuses on developing algorithms that can identify and correct biases in the data used to build knowledge graphs.

  Example: The Fairness Constraints framework by Biega et al. (2018) applies constraints during the construction of knowledge graphs to ensure fair representation.

Conclusion

The use of knowledge graphs presents significant ethical and privacy challenges. Addressing these issues requires a multi-faceted approach, incorporating technical solutions, ethical guidelines, and interdisciplinary collaboration. Ongoing research is crucial to develop methods that ensure the responsible use of knowledge graphs, balancing their powerful capabilities with the need to protect individuals' rights and privacy.

References

- ***Knowledge Graphs: A Survey***

  - Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Decker, S., Gutiérrez, C., ... & Zimmermann, A. (2021). Knowledge Graphs. *ACM Computing Surveys (CSUR)*, 54(4), 1-37.
  - Link to paper

- ***Fairness and Bias in Knowledge Graphs***

  - Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys (CSUR)*, 54(6), 1-35.
  - Link to paper

- ***Privacy-Preserving Techniques in Knowledge Graphs***

  - N Sharma and R Bhatt 2022, Privacy preserving knowledge graph for healthcare applications
  - Link to Paper