

Visualização exploratória de grafos de conhecimento: proposta de processo metodológico

Rafael Rocha¹, Gercina Ângela de Lima¹

¹Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Belo Horizonte, MG, Brasil

Abstract

Knowledge graphs are a data representation that maintains semantics through a structure with subject, predicate and object. With the advent of Big Data, greater storage brings several challenges to knowledge graphs. One of these challenges is extracting value from massive volume. In this context, interactive visual exploration is an effective approach. However, the problem is that, in the scientific literature, there are several methods. In this context, a methodological process is proposed with the help of Design Research Science (DSR) and the review of scientific literature.

Resumo

Os grafos de conhecimento são uma representação de dados que mantém a semântica por meio de uma estrutura com sujeito, predicado e objeto. Com o advento do *Big Data*, o maior armazenamento traz diversos desafios para os grafos de conhecimento. Um desses desafios é extrair valor de um volume massivo. Neste contexto, a exploração visual interativa é uma abordagem efetiva. No entanto, o problema é que, na literatura científica, há diversos métodos. Neste contexto, é proposto um processo metodológico com auxílio do *Design Research Science* (DSR) e da revisão da literatura científica.

Keywords

Knowledge graph, Big data, Visualization, Exploratory visualization

1. Introduction

Os estudos científicos sobre as representações em forma de grafo iniciaram na década de 1970, no entanto, o termo *grafo de conhecimento* ganhou maior projeção com a adoção da *Google* em seus produtos [1, 2]. Esta decisão foi seguida por outras empresas de tecnologia, que reconheceram o modelo de dados baseado em grafos como sendo mais vantajoso para determinadas aplicações.

Os grafos de conhecimento constituem-se de *nós*, que correspondem aos sujeitos e objetos, enquanto as *arestas* são os predicados [3]. O sujeito representa o conceito sobre uma declaração; o predicado explicita o tipo de relação e, por sua vez, o objeto é o valor da afirmação. Esta abordagem semântica é superior a outras abordagens, pois há mais flexibilidade na representação de coisas na realidade. Além disso, ela pode ter suporte, ou não, de uma ontologia que provê um metamodelo para desambiguação conceptual [4]. Por representar uma maior quantidade de dados, a superioridade semântica dos grafos de conhecimento contribui para colocar a engenharia de conhecimento em outro patamar.

Com o advento das tecnologias de *Big Data*, tornou-se possível armazenar dados em volumes não praticados anteriormente [5]. A maioria destas tecnologias possuem uma “semântica fraca”, ao passo que, os grafos de conhecimento possuem uma “semântica forte”, ou seja, já trazem embutidos em si suas partes intencional e extensional, isto amplia as vantagens em sua adoção[6]. Em outras palavras, a parte intencional é um modelo *ad hoc*, ou uma ontologia, e a parte extensional são os dados.

Entretanto, extrair valor de um volume massivo de dados é um problema recorrente apontado na literatura. Por outro lado, a exploração visual interativa permite ao usuário obter informação dinâmica e gradualmente [7]. Nessa perspectiva, os grafos de conhecimento contribuem para a integração em um número maior de dados ao conjunto, tornando a visualização mais viável. No entanto, o problema é que

Proceedings of the 17th Seminar on Ontology Research in Brazil (ONTOBRAS 2024) and 8th Doctoral and Masters Consortium on Ontologies (WTDO 2024), Vitória, Brazil, October 07-10, 2024.

✉ rafael-rocha@ufmg.br (R. Rocha); limagercina@gmail.com (G. d. Lima)

🆔 0000-0002-9719-1741 (R. Rocha); 0000-0003-0735-3856 (G. d. Lima)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

diversas técnicas de visualização exploratória de grafos de conhecimento são apresentadas de forma isolada na literatura, sem considerar adequadamente a jornada do usuário. Este estudo busca identificar e integrar os meios necessários para aperfeiçoar a ergonomia da visualização interativa desses grafos.

Neste contexto, este trabalho parte do pressuposto que integrar técnicas de visualização exploratória de grafos de conhecimento em um processo metodológico que considera a jornada do usuário resultará em uma ergonomia melhorada, facilitando a interação e a obtenção de *insights* pelos usuários. Assim busca-se responder: Quais são as técnicas e métodos mais eficazes para melhorar a ergonomia da visualização exploratória interativa de grafos de conhecimento, considerando as necessidades dos usuários?

2. Trabalhos relacionados

O processo exploratório visual permite ao usuário um caminho conciso para alcançar os objetivos nos conjuntos de dados em análise. Isto se materializa em passos a serem seguidos, mas que não limitam as ações do usuário em seu estudo. Neste contexto, a literatura científica traz trabalhos que apresentam propostas relacionadas a esta abordagem.

Em [8], os autores afirmam que os grafos de conhecimentos são representados por meio de nodos conectados para traduzir algumas métricas ou estéticas. Grafos esteticamente apresentados são mais fáceis de serem compreendidos, no entanto, os autores reiteram que algumas escolhas são subjetivas, enquanto outras são empiricamente testadas. Os autores propõem uma abordagem bem fundamentada em teorias cognitivas e guias de desenho, e culminam apresentando seis categorias de questões para orientar a visualização exploratória. Em sequência, estabelecem princípios cognitivos de *design* iniciais, além de disponibilizarem recursos interativos para adaptar a exploração.

Já em [9], a proposta é dividida em três partes: (i) sem interação e personalização: neste nível não são requeridos conhecimento de domínio e visão geral de alto nível; (ii) na interatividade média-alta requer-se um alto nível de necessidade de informação, além de uma visão geral de aspectos específicos; (iii) na última etapa, a alta interatividade e personalização, exige-se uma amostra detalhada ou uma consulta para respostas precisas. Ou seja, a heterogeneidade dos dados e as necessidades imprecisas dos usuários necessitam de uma abordagem adaptativa. Isto se resume, principalmente, em suporte a métodos exploratórios baseados em exemplos e similaridades, além de uma maior interatividade e personalização.

Ambos os trabalhos segmentam a forma como as interações na visualização devem ocorrer; no entanto, falta definir o processo ou a ordem que trará os melhores benefícios. Existem outros trabalhos relacionados, mas, para não extrapolar os requisitos de formatação do evento, optou-se por apresentar as duas produções consideradas mais relevantes para a proposta deste estudo.

3. Metodologia

Esta pesquisa utiliza o *Design Science Research* (DSR), um arcabouço metodológico paradigmático, para conduzir pesquisas [10]. Esta metodologia se diferencia de outros métodos por focar em artefatos de entrega que devem ser avaliados para atingir determinados objetivos. Além disso, pressupõe-se uma questão prática, que precisa ser abordada por meio da busca de uma solução para um problema. Para isso, há um processo interativo no qual cada etapa interage com outras para criar uma solução cíclica. Segundo [10], o ciclo é (i) investigação do problema, (ii) desenho da solução, (iii) validação do desenho, (iv) implementação da solução e (v) avaliação da implementação.

A etapa DSR (i) é definida na Seção 1, enquanto que, para a etapa (ii), utiliza-se a Revisão Sistemática da Literatura (RSL) científica para a identificação de abordagens e ferramentas de visualização. Para a pesquisa, foram selecionados os repositórios *Web of Science* (WoS) e Scopus. Os trabalhos foram recuperados por meio das strings dividida em dois campos de pesquisa: (i) a string “*visual exploration*” OR “*visuali?ation*”¹ aplicado ao campo título; (ii) a string “*knowledge graph*” OR “*causality graph*” OR

¹Foi utilizado interrogação para diferenciar *visualization* (inglês americano) e *visualisation* (inglês britânico)

“*semantic network*” OR “*semantic graph*” OR “*ontology*” aplicada a todos os campos. Os estudos seguem critérios de inclusão para elegibilidade à revisão; isto é, (i) publicação entre 2018 e 2022, (ii) disponível online, (iii) foco principal em KG, (iv) mostrar uma ferramenta para visualizar KG, (v) conter uma abordagem ou método para visualizar KG. Os critérios de exclusão, além do não cumprimento dos critérios de inclusão, são: (i) estudos duplicados (será utilizado apenas um); (ii) artigos curtos (iguais ou inferiores a seis páginas); (iii) e trabalhos redundantes (será considerada a versão completa); (iv) estudos não escritos em inglês, português ou espanhol. Foram recuperados 497 trabalhos, no total, ao se aplicar a busca nas plataformas. Na condução da revisão, foram selecionados 35 trabalhos contendo abordagens ou métodos para visualização exploratória (Seção 4.1); e 43 trabalhos com ferramentas (Seção 4.2).

A etapa (iii) da DSR será realizada com especialistas da área e alunos de pós-graduação. Nesse estágio, serão conduzidas entrevistas semi-estruturadas, tantas quantas forem necessárias, para a validação das propostas, com duração determinada. Em seguida, na etapa (iv) da DSR, ocorrerá a implementação da solução com o auxílio de especialistas da área, que atuarão como *alpha-testers*, desempenhando um papel crucial na detecção das principais inconsistências. Além disso, serão coletadas métricas de utilização para melhorias. Por fim, na etapa (v), será realizada a avaliação da solução.

Na seção seguinte, são apresentados os resultados preliminares.

4. Resultados preliminares

4.1. Abordagens

Após análise da literatura, foram selecionados 35 trabalhos de 497 recuperados que tratam das abordagens principais da pesquisa. Realizou-se uma análise nos 35 trabalhos utilizando-se seis categorias, que foram extraídas a partir dos termos mais enfatizados nos conteúdos dos artigos, conforme apresentados a seguir.

(i) *Função de expansão de nó* [11, 12, 13]. Este recurso permite aos usuários explorar o gráfico de conhecimento clicando duas vezes ou clicando no botão de opção de expansão de ícone ao redor do nó selecionado. O sistema exibe todos os nós conectados diretamente e seus relacionamentos, com um máximo de nós por consulta para evitar superlotação da tela;

(ii) *As informações do nó* ocorrem quando um nó é selecionado e a informação é exibida em algum lugar da tela [11, 12, 14, 13, 15, 16]. A função “Mostrar/Ocultar” reduz a carga cognitiva dos usuários para descrições mais longas.

(iii) Construção de uma *lista de conceitos* em forma de glossário [17, 18, 11, 14, 15] ou mesmo mostrando um diagrama de árvore [14, 19]. Essas abordagens ajudam os usuários a localizar rapidamente, proporcionando alta disponibilidade e eficácia da ferramenta.

(iv) *Cronologia* [14]. No domínio das ferramentas de visualização, a cronologia não é apenas uma sequência estática, mas uma estrutura dinâmica moldada pelo fazer e pelo refazer, uma ferramenta que transcende as restrições lineares, reconhecendo a natureza interativa do conhecimento humano. As ações iniciais envolvem coletar e organizar informações, mas o valor está na busca incessante por melhorias por meio do refazer. Este processo, que abrange a revisão e o refinamento, reflete *insights* e avanços em evolução. Neste contexto, a cronologia catalisa a melhoria contínua, capacitando os utilizadores a navegar em conhecimento responsivo, adaptável e em constante evolução.

(v) *As Métricas e estatísticas* [20, 14, 13, 21] nas ferramentas de visualização servem como base para extrair *insights* significativos dos dados. As métricas fornecem medidas quantificáveis, enquanto as técnicas estatísticas revelam padrões e tendências nas informações. Estas abordagens permitem que os utilizadores tomem decisões informadas, identifiquem correlações e avaliem o desempenho, aumentando a sua utilidade na análise científica e na tomada de decisões baseada em dados.

(vi) Usar um *mapa de calor* ou *tela de fundo* para ajudar os usuários a saber em qual *cluster* os nós estão localizados [22, 16]. A metáfora da imagem de fundo ajudou-os a compreender melhor o grafo.

Além dessas seis categorias, duas abordagens atípicas foram citadas: (i) [23] introdução de uma ferramenta de visualização de ontologias que é amigável e fácil de entender para especialistas e não especialistas em ontologias. Com a técnica de interação de busca “inteligente”, a ferramenta extrai e

exibe, de forma eficaz, uma visão ontológica contendo componentes relevantes para as necessidades do usuário. (ii) [12] apresentação de uma abordagem “tutorial”, na qual criam-se tutoriais interativos para ajudar os usuários a aprender rapidamente novos recursos. O tutorial “uma breve introdução” aparece primeiro e apresenta a barra de pesquisa e as caixas de informações do nó. Os usuários podem acessar novamente os tutoriais a qualquer momento usando um ícone.

Enfatiza-se, no entanto, que os resultados preliminares não foram esgotados, podendo ser ampliados em estudos futuros. A seção a seguir apresenta resumidamente os resultados sobre as ferramentas de visualização.

4.2. Ferramentas

Após análise da literatura, foram selecionados 42 trabalhos dos 497 recuperados que citam ferramentas de visualização. Os artigos que contêm ferramentas de visualização foram inicialmente selecionados a partir da verificação da existência, ou não, de link para alguma ferramenta ou para o *host* do código-fonte. Se encontrado o *link* no documento, a ferramenta é, também, classificada e marcada como ativa.

Como resultado, a pesquisa, em desenvolvimento, identificou 36 ferramentas de visualização. No entanto, 13 ferramentas não possuíam *link*, enquanto 3 demonstraram estar descontinuadas. Desta forma, apenas 20 ferramentas foram consideradas ativas e continuarão a ser analisadas neste estudo.

Atualmente, essa pesquisa está sendo realizada com foco dedicado à análise dos resultados obtidos. À medida que a investigação se aprofunda em suas complexidades, com um exame meticuloso dos dados, os resultados podem atingir os esforços atuais e o objetivo proposto. Já em relação às ferramentas, apesar de não terem obtido grande avanço na classificação, já foram encontradas em quantidade suficiente para o prosseguimento da pesquisa. Esta combinação dinâmica de investigação em progresso e do avanço a partir das etapas definidas da nossa metodologia destaca a natureza abrangente do trabalho atual.

4.3. Discussão

O objetivo da pesquisa é uma proposta de processo metodológico para auxiliar na visualização exploratória, recorrendo à revisão da literatura científica, que trouxe insumos suficientes para prosseguimento das fases seguintes do DSR.

Na literatura científica foram identificadas 8 abordagens que auxiliam na exploração dos grafos de conhecimento, além de 20 ferramentas que têm pré-requisitos para avançar na investigação. Estes achados, juntamente com os trabalhos relacionados, auxiliam na resposta para a pergunta apresentada na introdução.

Os resultados alcançados até o momento trazem robustez para a pesquisa, no entanto, estão em progresso.

5. Considerações finais

Esta pesquisa iniciou identificando uma dificuldade na exploração visual interativa de grafos de conhecimento, por não haver processo metodológico consolidado, a obtenção de informação pelo usuário é prejudicada. Sabe-se que os grafos de conhecimento tem como pressuposto modelar a forma como os humanos organizam e processam o conhecimento. Nesse contexto, as contribuições advindas da ciências cognitivas, podem auxiliar no processo de compreensão da rede de associações que ocorre no cérebro humano durante os processos cognitivos tanto na representação dessa estrutura quanto no uso de informações de maneira eficiente, facilitando a recuperação.

Neste contexto, esta pesquisa tem como objetivo analisar a fundamentação teórica da exploração visual interativa dos grafos de conhecimento. Assim, propor um processo metodológico compilando técnicas e abordagens da literatura científica. Parte-se do pressuposto que esta integração trará melhores resultados aos usuários.

A metodologia DSR possui uma dinâmica interativa de construção, resultando em entregas pragmáticas mais eficientes. A pesquisa em andamento trouxe diversos métodos e ferramentas para exploração

visual que contribuirá para a primeira versão da proposta. Os próximos passos serão a construção do primeiro esboço e rodadas de entrevistas semi-estruturadas.

Os resultados desta pesquisa contribuirão para diversas áreas que utilizem os grafos de conhecimento, além disso inova ao aplicar a DSR para checar no constructo final. Assim, expandindo as abordagens já existentes resolvendo o problema da visualização exploratória interativa.

References

- [1] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A.-C. N. Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, Knowledge graphs, *ACM Comput. Surv.* 54 (2021).
- [2] J. P. McCusker, J. Erickson, K. Chastain, S. Rashid, R. Weerawarana, D. McGuinness, What is a knowledge graph, *Semantic Web Journal* (2018) 2–14.
- [3] S. Shin, X. Jin, J. Jung, K.-H. Lee, Predicate constraints based question answering over knowledge graph, *Information Processing & Management* 56 (2019) 445–462.
- [4] M. B. Almeida, Revisiting ontologies: A necessary clarification, *Journal of the American Society for Information Science and Technology* 64 (2013) 1682–1693.
- [5] G. Chen, J. Hou, C. Liu, K. Hu, J. Wang, Visualization analysis of cross research between big data and construction industry based on knowledge graph, *Buildings* 12 (2022) 1812.
- [6] K. Wang, G. Qi, J. Chen, Y. Huang, T. Wu, Embedding ontologies via incorporating extensional and intensional knowledge, 2024.
- [7] D. J. Janvrin, R. L. Raschke, W. N. Dilla, Making sense of complex data using interactive data visualization, *Journal of Accounting Education* 32 (2014) 31–48.
- [8] W. Huang, J. Luo, T. Bednarz, H. Duh, Making graph visualization a user-centered process, *Journal of Visual Languages & Computing* 48 (2018) 1–8.
- [9] M. Lissandrini, T. B. Pedersen, K. Hose, D. Mottin, Knowledge graph exploration: where are we and where are we going?, *ACM SIGWEB Newsletter* (2020) 1–8.
- [10] R. Wieringa, Design science as nested problem solving, in: *Proceedings of the 4th international conference on design science research in information systems and technology*, ACM, 2009, p. 8.
- [11] J. Demelo, K. Sedig, Forming cognitive maps of ontologies using interactive visualizations, *Multimodal Technologies and Interaction* 5 (2021).
- [12] X. He, R. Zhang, R. Rizvi, J. Vasilakes, X. Yang, Y. Guo, Z. He, M. Prospero, J. Huo, J. Alpert, J. Bian, Aloha: developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design, *BMC Medical Informatics and Decision Making* 19 (2019) 150.
- [13] M. Nečaský, Štěpán Stenclák, Interactive and iterative visual exploration of knowledge graphs based on shareable and reusable visual configurations, *Journal of Web Semantics* 73 (2022) 100713.
- [14] A. Kleebaum, B. Paech, J. O. Johanssen, B. Bruegge, Continuous rationale visualization, in: *2021 Working Conference on Software Visualization (VISSOFT)*, 2021, pp. 33–43.
- [15] J. Peng, D. Xu, R. Lee, S. Xu, Y. Zhou, K. Wang, Expediting knowledge acquisition by a web framework for knowledge graph exploration and visualization (kgev): case studies on covid-19 and human phenotype ontology, *BMC Medical Informatics and Decision Making* 22 (2022) 147.
- [16] V. Wiens, S. Lohmann, S. Auer, Gizmo - a customizable representation model for graph-based visualizations of ontologies., in: *Proceedings of the 10th International Conference on Knowledge Capture - K-CAP '19*, ACM Press, 2019.
- [17] S. Alter, Extending a work system metamodel using a knowledge graph to support IS visualization and development, in: S. Link, I. Reinhartz-Berger, J. Zdravkovic, D. Bork, S. Srinivasa (Eds.), *Proceedings of the ER Forum and PhD Symposium 2022 co-located with 41st International Conference on Conceptual Modeling (ER 2022)*, Virtual Event, Hyderabad, India, October 17, 2022, volume 3211 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022.
- [18] T. Basyuk, A. Vasyliuk, Approach to a subject area ontology visualization system creating, in:

- N. Sharonova, V. Lytvyn, O. Cherednichenko, Y. Kupriianov, O. Kanishcheva, T. Hamon, N. Grabar, V. Vysotska, A. Kowalska-Styczen, I. Jonek-Kowalska (Eds.), Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021). Volume I: Main Conference, Lviv, Ukraine, April 22-23, 2021, volume 2870 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 528–540.
- [19] S. Sheng, P. Zhou, X. Wu, Cpv: A tree structure information extraction and visualization tool for big knowledge graph, in: 2019 IEEE International Conference on Big Knowledge (ICBK), 2019, pp. 221–228.
- [20] F. Desimoni, N. Bikakis, L. Po, G. Papastefanatos, A comparative study of state-of-the-art linked data visualization tools, in: *CEUR WORKSHOP PROCEEDINGS*, volume 2778, CEUR-WS, 2020, pp. 1–13.
- [21] J. Zhu, Q. Zhao, E. Katsevich, C. Sabatti, Exploratory gene ontology analysis with interactive visualization, *Scientific Reports* 9 (2019) 7793.
- [22] H. Ha, H. Han, S. S. Mun, S. Bae, J. Lee, K. Lee, An Improved Study of Multilevel Semantic Network Visualization for Analyzing Sentiment Word of Movie Review Data, *Applied Sciences* 9 (2019) 2419.
- [23] M. Neji, F. Ghorbel, B. Gargouri, A smart search-based ontology visualization tool using sparql patterns, in: C. Douligeris, D. Karagiannis, D. Apostolou (Eds.), *Knowledge Science, Engineering and Management*, Springer International Publishing, Cham, 2019, pp. 33–44.