

Visualização 3D para Bancos de Dados Orientados a Grafos: Um Estudo de Caso Utilizando Redes Sociais

Lucas C. Lemos¹, Célia G. Ralha¹

¹Departamento de Ciência da Computação - Instituto de Ciências Exatas
Universidade de Brasília (UnB), Brasília - DF - Brasil

lc.lemos@outlook.com, ghedini@unb.br

Abstract. *A known challenge of the graph-oriented database community includes aspects of visualization to enable quantitative and qualitative understanding of information. This work objective is to apply 3D visualization techniques to Neo4J online databases. The 3D visualization techniques are useful to minimize the data occlusion problem, enabling greater user interactivity. The tool with 3D visualization Graph2Vis is implemented with the Angular framework and the 3d-force-graph library. A case study using social networks formed by five graduate programs in Computing Studies associated with Brazilian universities illustrates the Graph2Vis usage.*

Resumo. *Um conhecido desafio da comunidade de banco de dados orientado a grafos inclui aspectos de visualização para viabilizar a compreensão quantitativa e qualitativa da informação. Desta forma, o objetivo deste trabalho é aplicar técnicas de visualização 3D para bancos de dados online Neo4J. Técnicas de visualização 3D têm demonstrado utilidade para minimizar o problema da oclusão de dados, além de viabilizar uma maior interatividade do usuário. A ferramenta Graph2Vis com visualização 3D foi implementado com o framework Angular e a biblioteca 3d-force-graph. Um estudo de caso utilizando redes sociais formada por cinco programas de pós-graduação em Computação vinculados a universidades brasileiras ilustra o uso do Graph2Vis.*

1. Introdução

O uso do banco de dados baseados em grafos Neo4j tem ganhado popularidade na comunidade de bancos NoSQL, ampliando a necessidade de visualização de grandes bases. O Neo4j possui uma área para visualização 2D dos dados de consulta denominado *Neo4j Browser* [Neo4j Inc. 2022]. Porém, [Lanum 2016] aponta que a oclusão influencia negativamente a experiência de visualização e análise de grafos, especialmente em contextos 2D, onde dois ou mais nós podem estar sobrepostos prejudicando análise e interpretação. Neste sentido, em consonância com o problema da oclusão existem outros aspectos que demandam soluções, como por exemplo, a viabilização de uma maior interatividade nas visualizações apresentadas pelo *Neo4j Browser*, sugerindo a aplicação de técnicas 3D.

Este trabalho tem como objetivo a aplicação de técnicas de visualização 3D para bancos de dados online Neo4J, minimizando o problema da oclusão de dados e aumentando a interatividade do usuário. Como prova de conceito foi desenvolvido a ferramenta Graph2Vis que aplica técnicas de visualização 3D usando o framework Angular e a biblioteca 3d-force-graph. Graph2Vis foi ilustrada com um estudo de caso de redes sociais formada por cinco programas de pós-graduação em Computação vinculados a universidades

brasileiras. Foram definidas questões investigativas para direcionar o escopo do trabalho visando ampliar o conhecimento a cerca da representação gráfica dos dados manipuladas nos bancos Neo4j, abrangendo aspectos de visualização e análise da informação:

1. Quais são as soluções teóricas e de ferramentas implementadas existentes na literatura para visualização 3D de bases de dados orientadas a grafo utilizando Neo4j?
2. Como desenvolver uma ferramenta de visualização 3D de banco de dados orientado a grafo utilizando Neo4j?
3. Qual estudo de caso pode ilustrar a ferramenta desenvolvida neste trabalho?

Como contribuição, cita-se a geração de visualizações 3D de grande volume de dados armazenados em Neo4j utilizando a ferramenta Graph2Vis de domínio de aplicação geral, mas ilustrada com um estudo de caso de redes sociais. Considera-se que a disponibilização do Graph2Vis à comunidade de banco de dados NoSQL é uma contribuição. O restante do artigo inclui na Seção 2, uma revisão da literatura, na Seção 3, a proposta de solução, na Seção 4, o estudo de caso, e na Seção 5, as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Para responder a primeira questão investigativa foi realizado uma revisão da literatura durante o mês de outubro de 2021 utilizando a plataforma Google Acadêmico, a qual inclui artigos de diversas bases, e.g., IEEE Xplore ([Tallat et al. 2019], [Müller et al. 2018]) e base Scopus ([García del Valle et al. 2021], [Yokoyama et al. 2021], [Carnaz et al. 2021], [Bukhari et al. 2021], [Zahoránszky-Kóhalmi et al. 2020]), somente citando alguns trabalhos recentes. As *strings* de busca *Visualization, Neo4j, Social Networks* foram utilizadas considerando o período de 2015 a 2021.

O modelo de visualização de grafos com três etapas proposto por [Chen et al. 2019] foi utilizado para classificar os 19 trabalhos selecionados (Tabela 1). A Etapa (i) de modelagem de relações inclui o pré-processamento de dados, modelagem de dados resultando na geração de grafos, e a mineração das relações nos grafos. A Etapa (ii) apresenta sete técnicas de visualização incluindo nós e links, matriz de adjacências, hipergrafo, diagrama de fluxo, grafo com informações geoespaciais, multi-atributos e preenchimento de espaços. A Etapa (iii) apresenta as funções de simplificação e interação utilizadas na visualização de grafos incluindo a clusterização de nós, filtragem do grafo, empacotamento de links, transformação baseada na topologia do grafo e limitação de dimensões. É possível utilizar abordagens híbridas com diversas técnicas, dependendo do objetivo de simplificação e interação definidos no modelo.

Nos trabalhos da literatura, as principais técnicas de visualização foram as de grafos, os nós e links, matriz de adjacências, hipergrafos, e preenchimento de espaços. Assim como as funções de clusterização de nós, filtragem do grafo e empacotamento de links (classificação de [Chen et al. 2019] na Tabela 1). Nota-se que a técnica de nós e links e a função de filtragem do grafo são as mais utilizadas. Considerando os domínios de conhecimento das ferramentas de visualização de dados em Neo4j, verifica-se que dos 19 trabalhos existe uma concentração na área de medicina (7), redes sociais e domínio geral (3), criminal (2), e a área de engenharia civil, recursos humanos, aviação e desenvolvimento de *software*. Este levantamento serviu para motivar a elaboração de uma solução de visualização 3D de domínio geral, visto que foi encontrado um único trabalho com essas características, mas aplicando técnicas 2D ([Rodrigues et al. 2015]).

Tabela 1. Trabalhos selecionados durante a revisão da literatura.

	Referência	Técnicas de Visualização	Funções Exploradas	Domínio de Aplicação
Técnicas e Funções de Visualização	[Bludau et al. 2021]	nós e links 2D	empacotamento de links	Geral
	[Kumar and Teo 2021]	nós e links 2D	filtragem do grafo	Eng. Civil
	[Tallat et al. 2019]	nós e links, hipergrafo 2D	filtragem do grafo	Medicina
	[Farooq et al. 2018]	nós e links, hipergrafo 2D	filtragem do grafo	Criminal
	[Rahman and Karim 2016]	nós e links, esférico, clusterizado 2D	filtragem do grafo, clusterização de nós	Redes Sociais
	[Summer et al. 2015]	nós e links 2D	filtragem do grafo	Geral
Ferramentas de Visualização	[Antweiler et al. 2021]	nós e links 2D	filtragem do grafo	Medicina
	[Yokoyama et al. 2021]	nós e links 2D	filtragem grafo, clusterização nós	Recursos Humanos
	[García del Valle et al. 2021]	nós e links 2D	filtragem do grafo	Medicina
	[Jo et al. 2021]	nós e links 2D	filtragem do grafo	Redes Sociais
	[Hassani-Pak et al. 2021]	nós e links 2D	filtragem do grafo	Medicina
	[Bukhari et al. 2021]	nós e links, matriz, preenchimento espaços 2D	filtragem do grafo	Medicina
	[Carnaz et al. 2021]	nós e links 2D	filtragem do grafo, empacotamento de links	Criminal
	[Zahoránszky-Kóhalmi et al. 2020]	nós e links 2D	filtragem grafo, pacotamento de links	Medicina
	[Müller et al. 2018]	hipergrafo, preenchimento de espaços 2D	filtragem do grafo	Desenvolvimento de <i>Software</i>
	[Hühne et al. 2018]	nós e links 3D	filtragem do grafo, clusterização de nós	Medicina
	[Kerzner et al. 2017]	nós e links, preenchimento de espaços 2D	filtragem do grafo	Aviação
	[Partl et al. 2016]	nós e links, preenchimento de espaços 2D	filtragem do grafo, clusterização de nós	Redes Sociais
	[Rodrigues et al. 2015]	nós e links 2D	filtragem do grafo, clusterização de nós	Geral

3. Proposta

A segunda questão investigativa aborda o desenvolvimento de uma ferramenta para visualização 3D de banco de dados Neo4j com domínio de aplicação geral. A ferramenta *open-source* foi denominada Graph2Vis (*Graph to Visualization*),¹ potencializando a capacidade de visualização dos grafos em formato de nós e *links*.²

Ciclo de Processamento do Graph2Vis – a interface Web é o componente principal que realiza a interação do usuário nas três etapas conforme apresentado na Figura 1(a), conexão ao banco de dados Neo4j (login/senha), consulta e parâmetros de visualização (Figura 1(b)), e visualização dos grafos em formato de nós e *links* 3D.

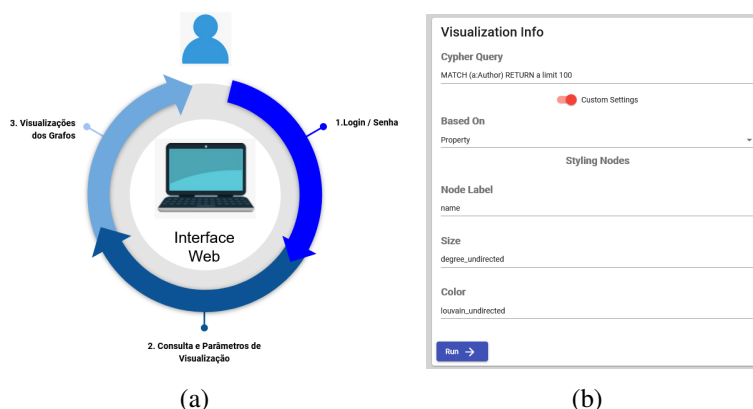


Figura 1. Ciclo de Processamento e Interface do Graph2Vis.

Arquitetura – para a implementação do Graph2Vis foi utilizado o *Model, View, Control-*

¹Código e documentação disponíveis em <https://gitlab.com/InfoKnow/graph2viz>.

²Graph2Vis está disponível para uso em <https://graph2viz.herokuapp.com/>.

ler (MVC) [Gamma et al. 1994]. Conforme apresentado na Figura 2, a camada controladora recebe e envia a requisição do usuário (1 e 2). A camada modelo recebe a requisição do controle (3) e realiza a consulta ao banco Neo4j (4). A camada controle recebe o resultado da consulta (5) e atualiza a camada visão com grafos 3D (6 e 7).

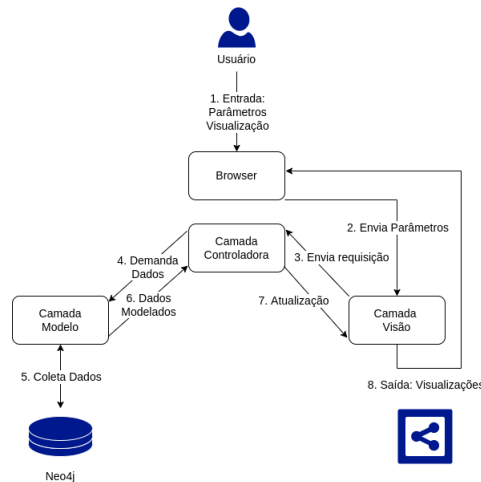


Figura 2. Arquitetura do Graph2Vis.

4. Estudo de caso

A terceira questão investigativa utilizou uma rede social formada por 194 pesquisadores vinculados a cinco programas de pós-graduação na área de Ciência da Computação de universidades públicas brasileiras: Universidade de Brasília (UnB), Universidade de São Paulo (USP), Universidade Federal do Amazonas (UFAM), Universidade Federal do Rio Grande do Norte (UFRN) e Universidade Federal de Minas Gerais (UFMG). Essa rede foi desenvolvida no projeto *SCI-synergy* (<http://165.227.113.212>) do Grupo de Pesquisa do CNPq *InfoKnow - Computer Systems for Information and Knowledge Treatment*. A rede ilustra relações de coautoria em publicações científicas extraídas do *Digital Bibliography and Library Project* (DBLP) com a construção de uma base no Neo4j. Para potencializar a visualização dos nós do grafo foram computadas duas medidas de centralidade executadas utilizando rotinas na linguagem CYPHER, e a contagem de relação de coautorias de cada autor *COAUTHOR_WEIGHT*.

1. *degree* - representa o número de relações que incidem em determinado nó do grafo, indicando o grau de conectividade direta entre autores e coautores, determinando o tamanho de cada nó na rede. A consulta CYPHER utilizada é:

```

MATCH (a:Author)-[r:COAUTHOR]-(b:Author)
WITH a,b,count(r) as dg
SET a.degree_undirected = dg

```

2. *louvain* - representa qual comunidade na rede determinado nó pertence, indicando grupos de nós que tem grande probabilidade de compartilhar propriedades ou papéis semelhantes, sendo representado por uma respectiva cor. A consulta CYPHER para a comunidade de cada autor é:

```

CALL algo.louvain(
MATCH (a:Author) where (a)-[:COAUTHOR]->() RETURN id(a) as id,
MATCH (a)-[:COAUTHOR]->()<-[:COAUTHOR]-(b) RETURN id(a) as source,
id(b) AS target, count(*) as 'louvain_undirected', {graph:'neo4j',iterations:1})

```

3. *COAUTHOR_WEIGHT* - representa o total de coautorias de cada autor proporcionando uma visualização mais coesa, a qual resume a quantidade de relações de coautoria na apresentação dos grafos. A rotina CYPHER utilizada é:

```
MATCH (a:Author)-[r:COAUTHOR]->(b:Author) WITH a, b, count(r) as W
CREATE (a)-[O:COAUTHOR_WEIGHT{weight:W}]->(b)
```

Considerando os 31 pesquisadores vinculados a UnB, a rede científica totaliza 1.621 autores e 3.619 relações de coautoria. A Figura 3(a) ilustra a rede de coautoria de um pesquisador com informações detalhadas representado pelo nó central, com cor para coautor e tamanho diferente de nó para o grau de colaboração. A Figura 3(b) apresenta informações da relação entre autores com o peso da relação (*COAUTHOR_WEIGHT*). Na Figura 3(c) é possível identificar *clusters* de autores através de cores e tamanhos variados de nós, bem como nós intermediários entre os *clusters*. A Figura 3(d) expande a distância entre os nós do grafo gerando uma visualização menos densa. Pode-se aplicar diferentes níveis de *zoom* e rotação aos grafos viabilizando maior interação do usuário. A Figura 4 apresenta a rede de coautoria de quatro programas de pós-graduação, 4(a) USP, 4(b) UFAM, 4(c) UFRN e 4(d) UFMG (maior rede de coautoria).

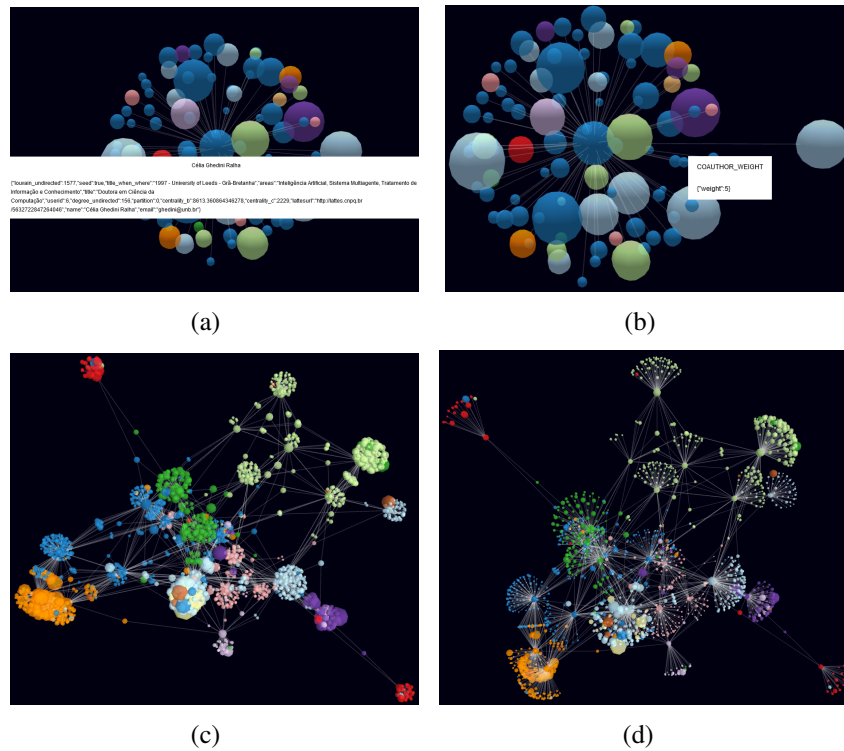


Figura 3. Visualização da rede de coautoria da UnB.

5. Conclusão

Aspectos de visualização de dados em forma de grafos ainda apresenta desafios para viabilizar a compreensão quantitativa e qualitativa da informação. Neste sentido, este trabalho aplicou técnicas de visualização 3D para bancos de dados online Neo4j apresentando uma ferramenta denominada Graph2Vis. Foram definidas três questões investigativas (Seção 1), as quais foram respondidas ao longo do manuscrito (Seções 2, 3 e 4). Como trabalhos futuros, vislumbra-se novas funcionalidades para o Graph2Vis, tais como a integração de outras bibliotecas de visualização, e.g., *Vis*, *Neovis*, *Popoto*, *Sigma*, *Vivagraph*, sendo que cada biblioteca requer um modelo específico de dados de entrada para geração dos grafos.

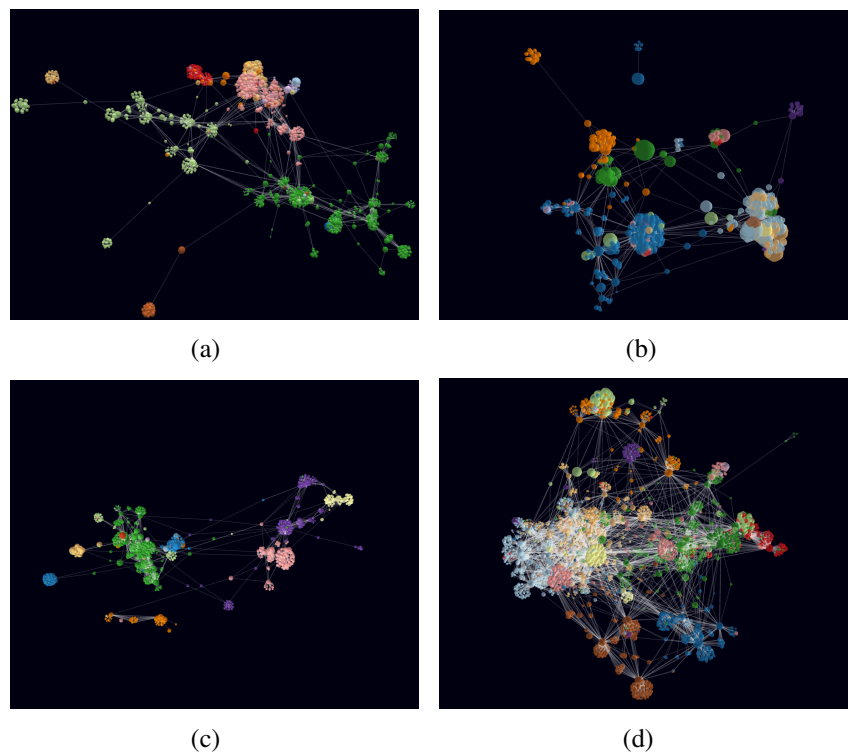


Figura 4. Visualização das redes de coautoria da USP (a), UFAM (b), UFRN (c) e UFMG (d).

Referências

- Antweiler, D., Sessler, D., Ginzel, S., and Kohlhammer, J. (2021). Towards the Detection and Visual Analysis of COVID-19 Infection Clusters. In Vrotsou, K. and Bernard, J., editors, *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Ass.
- Bludau, M.-J., Dörk, M., and Tominski, C. (2021). Unfolding Edges for Exploring Multivariate Edge Attributes in Graphs. In Byška, J., Jänicke, S., and Schmidt, J., editors, *EuroVis 2021 - Posters*. The Eurographics Association.
- Bukhari, S. A. C., Pawar, S., Mandell, J., Kleinstein, S. H., and Cheung, K.-H. (2021). LinkedImm: a linked data graph database for integrating immunological data. *BMC bioinformatics*, 22(S9):105–105.
- Carnaz, G., Nogueira, V. B., and Antunes, M. (2021). A graph database representation of portuguese criminal-related documents. *Informatics*, 8(2).
- Chen, Y., Guan, Z., Zhang, R., Du, X., and Wang, Y. (2019). A survey on visualization approaches for exploring association relationships in graph data. *J. Visualization*, 22.
- Farooq, A., Joyia, G. J., Uzair, M., and Akram, U. (2018). Detection of influential nodes using social networks analysis based on network metrics. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*.
- Gamma, E., Helm, R., and Johnson, R. E. (1994). *Design Patterns. Elements of Reusable Object-Oriented Software*. Addison-Wesley Longman, Amsterdam, 1st edition.

- García del Valle, E. P., Lagunes García, G., Prieto Santamaría, L., Zanin, M., Menasalvas Ruiz, E., and Rodríguez-González, A. (2021). DisMaNET: A network-based tool to cross map disease vocabularies. *Comp. Methods & Prog. Biomedicine*, 207:106233.
- Hassani-Pak, K., Singh, A., Brandizi, M., Hearnshaw, J., Parsons, J. D., Amberkar, S., Phillips, A. L., Doonan, J. H., and Rawlings, C. (2021). KnetMiner: a comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. *Plant Biotechnology Journal*, 19(8):1670–1678.
- Hühne, R., Kessler, V., Fürstberger, A., Kühlwein, S., Platzer, M., Sühnel, J., Lausser, L., and Kestler, H. (2018). 3D Network exploration and visualisation for lifespan data. *BMC Bioinformatics*, 19(390).
- Jo, S., Park, B., Lee, S., and Kim, J. (2021). OLGAVis: On-Line Graph Analysis and Visualization for Bibliographic Information Network. *Applied Sciences*, 11(9).
- Kerzner, E., Lex, A., Sigulinsky, C. L., Urness, T., Jones, B. W., Marc, R. E., and Meyer, M. (2017). Graffinity: Visualizing connectivity in large graphs. *Computer Graphics Forum (EuroVis)*, 36(3):251–260.
- Kumar, V. and Teo, E. A. L. E. (2021). Exploring the application of property graph model in visualizing cobie data. *Journal of facilities management*, 19(4):500–526.
- Lanum, C. (2016). *Visualizing Graph Data*. Manning Publications, 1 edition.
- Müller, R., Mahler, D., Hunger, M., Nerche, J., and Harrer, M. (2018). Towards an open source stack to create a unified data source for software analysis and visualization. In *2018 IEEE Working Conference on Software Visualization (VISSOFT)*, pages 107–111.
- Neo4j Inc. (2022). Neo4j Browser User Interface Guide. <https://neo4j.com/developer/neo4j-browser/>. Acessado em: 12 de maio de 2022.
- Partl, C., Gratzl, S., Streit, M., Wassermann, A. M., Pfister, H., Schmalstieg, D., and Lex, A. (2016). Pathfinder: Visual analysis of paths in graphs. *Computer Graphics Forum (EuroVis '16)*, 35(3):71–80.
- Rahman, M. and Karim, R. (2016). Comparative study of different methods of social network analysis and visualization. In *Int. Conf. Networking Systems & Security*.
- Rodrigues, J. F., Tong, H., Traina, A. J. M., Faloutsos, C., and Leskovec, J. (2015). Gmine: A system for scalable, interactive graph visualization and mining. In *Proc. 32nd Int. Conf. on Very Large Data Bases*, page 1195–1198.
- Summer, G., Kelder, T., Ono, K., Radonjic, M., Heymans, S., and Demchak, B. (2015). cyNeo4j: connecting Neo4j and Cytoscape. *Bioinformatics*, 31(23):3868–3869.
- Tallat, R., Amir Latif, R. M., Ali, G., Zaheer, A. N., Farhan, M., and Aslam Shah, S. U. (2019). Visualization and analytics of biological data by using different tools and techniques. In *IEEE 16th Int. Bhurban Conf. Applied Sci. & Tech.*, pages 291–303.
- Yokoyama, T. T., Okada, M., and Taniguchi, T. (2021). Panacea: Visual exploration system for analyzing trends in annual recruitment using time-varying graphs. *PLOS ONE*, 16(3):1–22.
- Zahoránszky-Kóhalmi, G., Sheils, T., and Oprea, T. (2020). SmartGraph: a Network Pharmacology Investigation Platform. *Journal of Cheminformatics*, 12(5).