

Aplicações de Grafos com Ontologias e GNNs: Revisão de Casos nas Áreas de Saúde e Segurança

Allan Miller Silva Lima¹, João Vitor Mendes de Lira¹, José Thiago Torres da Silva¹

¹ Programa de Pós-Graduação em Engenharia da Computação
Universidade de Pernambuco (UPE) – Recife, PE – Brasil

amsl@ecomp.poli.br, jvml@ecomp.poli.br, jtts@ecomp.poli.br

Abstract. *This meta-paper describes the style to be used in articles and short papers for SBC conferences. For papers in English, you should add just an abstract while for the papers in Portuguese, we also ask for an abstract in Portuguese (“resumo”). In both cases, abstracts should not have more than 10 lines and must be in the first page of the paper.*

Resumo. *Este meta-artigo descreve o estilo a ser usado na confecção de artigos e resumos de artigos para publicação nos anais das conferências organizadas pela SBC. É solicitada a escrita de resumo e abstract apenas para os artigos escritos em português. Artigos em inglês deverão apresentar apenas abstract. Nos dois casos, o autor deve tomar cuidado para que o resumo (e o abstract) não ultrapassem 10 linhas cada, sendo que ambos devem estar na primeira página do artigo.*

1. Introdução

Nos últimos anos, as estruturas em grafo têm se destacado como uma poderosa ferramenta para representar e analisar dados complexos em diversas áreas. Sua capacidade de modelar relações semânticas, integrar diferentes tipos de informações e oferecer transparência nas análises as tornam especialmente úteis em domínios como saúde e segurança da informação, onde o volume e a complexidade dos dados exigem abordagens eficientes e interpretáveis.

Neste trabalho, exploramos como os grafos vêm sendo aplicados nesses campos, com foco em três estudos recentes: dois na área da saúde, que utilizam ontologias representadas como grafos RDF para organizar conhecimento, e um em segurança, que emprega redes neurais em grafos (GNNs) e técnicas de análise de sinais para detectar anomalias em logs de sistema. Ao examinar essas abordagens, buscamos entender não apenas os resultados alcançados, mas também as vantagens e desafios práticos da modelagem em grafo.

Além de apresentar os métodos e contribuições desses trabalhos, discutimos o papel dos grafos como uma ponte entre a representação semântica de conhecimento e as técnicas modernas de inteligência artificial. Por fim, refletimos sobre possíveis caminhos para pesquisas futuras que integrem essas duas perspectivas, destacando oportunidades para avanços tanto teóricos quanto aplicados.

2. Fundamentação Teórica

Nesta seção, apresentamos os fundamentos teóricos sobre grafos, sua estrutura e operações, bem como suas aplicações em contextos computacionais avançados. Embora tradicionalmente associados a redes sociais, mapas e estruturas de dados, os grafos vêm sendo cada vez mais explorados em áreas críticas como saúde e segurança digital. Particularmente, este trabalho se concentra em duas frentes de aplicação: (i) a utilização de grafos como base para representação de conhecimento em ontologias formais, com destaque para a área da saúde; e (ii) a modelagem de sequências de eventos e logs para detecção de anomalias em sistemas de segurança computacional. Ao longo da fundamentação, introduziremos os conceitos essenciais que serão retomados na análise dos trabalhos relacionados.

2.1. Grafos: Definição e Estrutura Básica

Grafos são estruturas matemáticas que nos ajudam a modelar relações entre objetos de maneira intuitiva e sistemática. Em termos simples, um grafo é composto por elementos (nós ou vértices) conectados por ligações (arestas), formando uma rede de dependências ou interações (Cormen et al. 2009). Essa estrutura é aplicável a uma ampla gama de domínios: em contextos clínicos, por exemplo, é possível representar um paciente, seus sintomas, diagnósticos e tratamentos como nós interligados por relações médicas; já em sistemas computacionais, eventos registrados em logs podem ser conectados para formar cadeias de execução que revelam o comportamento do sistema.

Formalmente, um grafo é definido como um par $G = (V, E)$, onde:

- **V** é o conjunto de vértices (os elementos individuais, como pacientes, condições clínicas ou eventos de sistema);
- **E** é o conjunto de arestas (as conexões ou relações entre os vértices, como “apresenta sintoma”, “executa processo” ou “causa evento”).

A Figura 1 ilustra um grafo simples, no qual os círculos representam os vértices e as linhas que os conectam representam as arestas. Mesmo com essa estrutura elementar, é possível capturar relações complexas e úteis para tarefas como raciocínio automático, inferência semântica ou detecção de anomalias.

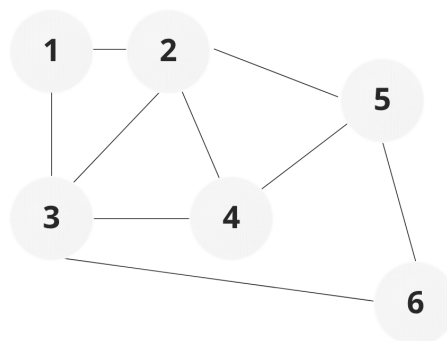


Figura 1. Representação de um grafo

2.2. Classificações de Grafos

Os grafos podem ser classificados conforme diferentes características:

- **Direcionamento:** em grafos direcionados, as arestas possuem orientação, representando relações assimétricas entre vértices. Já em grafos não direcionados, as conexões são bidirecionais.
- **Ponderação:** grafos ponderados atribuem pesos numéricos às arestas, geralmente utilizados para expressar custo, distância ou intensidade de relação. Em contrapartida, grafos não ponderados consideram todas as conexões equivalentes.
- **Densidade:** a densidade do grafo refere-se à proporção entre o número de arestas existentes e o número máximo possível de arestas. Essa métrica é fundamental para avaliar a complexidade da estrutura relacional.

2.3. Principais Operações sobre Grafos

Diversas operações clássicas sobre grafos são fundamentais para aplicações práticas nas áreas de saúde e segurança. Entre as mais relevantes, destacam-se:

- **Buscas em profundidade (DFS) e em largura (BFS)** — utilizadas, por exemplo, para explorar cadeias de eventos em logs ou propagar relações entre conceitos em uma ontologia clínica;
- **Caminhos mínimos** — úteis para identificar trajetórias diagnósticas mais prováveis ou cadeias críticas em sistemas;
- **Centralidade de nós** — importante para avaliar quais sintomas ou eventos são mais influentes em uma rede semântica ou de segurança.

Essas operações servem como base para algoritmos mais complexos, como inferência em grafos de conhecimento, detecção de anomalias por GNNs e raciocínio automatizado em sistemas baseados em ontologias.

2.4. Aspectos Computacionais

Na prática, a forma como os grafos são representados impacta diretamente sua aplicação. Estruturas como listas de adjacência são preferidas quando se trabalha com dados esparsos, como relações clínicas mapeadas em ontologias médicas. Já matrizes de adjacência podem ser mais eficientes quando se deseja verificar rapidamente a conexão entre eventos, como em fluxos de logs em sistemas de segurança.

Algoritmos de busca e propagação de informação, como BFS e DFS, têm complexidade linear ($O(V + E)$) e são fundamentais tanto para percorrer triplas RDF em consultas semânticas, quanto para extrair subgrafos em técnicas de detecção de anomalias. Em modelos baseados em aprendizado, como GNNs, a representação do grafo influencia diretamente o desempenho computacional, especialmente em tarefas que envolvem volumes massivos de dados ou exigem atualização contínua dos vértices.

Assim, aspectos como escalabilidade, eficiência de consulta e custo computacional tornam-se cruciais ao aplicar grafos em domínios sensíveis como saúde e segurança da informação.

3. Trabalhos Relacionados

Nesta seção, são apresentados e analisados três trabalhos que ilustram diferentes aplicações da teoria dos grafos em contextos reais. Os dois primeiros artigos estão inseridos na área da saúde e exploram o uso combinado de grafos RDF e ontologias biomédicas para estruturar e representar informações clínicas de forma semântica. Já o terceiro artigo, voltado à segurança cibernética, mostra como grafos de proveniência podem ser utilizados para modelar relações entre eventos de sistema e detectar anomalias com alta precisão. Em comum, os três estudos adotam os grafos como elemento central de representação, seja para organizar conhecimento médico ou para capturar padrões complexos em dados de logs, destacando a versatilidade da estrutura gráfica em domínios distintos.

3.1. LLMs para Construção Inteligente de Grafos RDF na Saúde

O artigo “*Large Language Models for Intelligent RDF Knowledge Graph Construction*”, publicado em 2025 (Mavridis et al. 2025), investiga como modelos de linguagem de larga escala (LLMs), como o GPT-4o, podem ser empregados para automatizar a construção de grafos de conhecimento na área médica. A proposta central do estudo é utilizar LLMs para mapear termos clínicos extraídos de textos livres em conceitos padronizados da ontologia SNOMED CT, a partir da geração de triplas RDF que estruturam o conhecimento em forma de grafo. Esse tipo de estrutura tem ganhado relevância justamente pela sua capacidade de representar, de maneira explícita, relações semânticas complexas entre entidades clínicas, como diagnósticos, sintomas e tratamentos. Dessa forma, o artigo insere-se diretamente no escopo de aplicação de grafos como suporte à representação formal e inferência automática em domínios sensíveis como o da saúde. f

O problema de pesquisa enfrentado diz respeito à dificuldade histórica de transformar textos médicos não estruturados — como prontuários eletrônicos ou relatórios clínicos — em dados formais, padronizados e semanticamente integráveis. Enquanto abordagens tradicionais exigem curadoria manual ou sistemas baseados em regras fixas, o artigo propõe que os LLMs sejam capazes de compreender o conteúdo textual e converter as informações extraídas em estruturas formais baseadas em grafos. Com isso, o trabalho busca responder à pergunta de como é possível automatizar, com precisão, a geração de grafos semânticos que reflitam o conteúdo de registros médicos reais.

A metodologia consiste em utilizar o GPT-4o para gerar triplas RDF a partir de fragmentos de texto clínico, onde cada conceito identificado se transforma em um nó do grafo e as relações entre eles em arestas dirigidas. Ferramentas auxiliares, como o BioBERT, foram utilizadas para refinar o reconhecimento de entidades biomédicas, enquanto mecanismos de verificação semântica, como o ChromaDB, ajudaram a validar o mapeamento com a ontologia alvo. Ao final, o sistema é capaz de construir um grafo de conhecimento médico onde, por exemplo, termos como “hiperglicemia”, “nível elevado de glicose” e “diabetes tipo 2” são interligados por relações semânticas precisas, permitindo consultas estruturadas e inferência de informações clínicas relevantes.

Os resultados obtidos foram bastante expressivos. O modelo atingiu um F1-score de 96,26%, superando soluções anteriores tanto em precisão quanto em cobertura conceitual. Além disso, o grafo resultante apresentou alta coesão semântica, possibilitando inferências como a identificação de relações causais entre sintomas e diagnósticos, e a extração de planos terapêuticos comuns a grupos de pacientes com características seme-

lhantes. Ainda assim, o estudo reconhece limitações: o desempenho do sistema depende fortemente da qualidade dos dados de entrada e da cobertura da ontologia utilizada; além disso, a validação ocorreu em cenários laboratoriais, o que levanta questões sobre sua aplicação em ambientes clínicos reais.

Dessa forma, O artigo mostra como grafos RDF podem ser usados na área da saúde para estruturar informações clínicas a partir de textos livres. Ao empregar LLMs no mapeamento de termos médicos para ontologias formais, o estudo contribui para a padronização semântica dos dados. Embora não explore inferências diretamente, destaca o papel dos grafos na organização e integração de informações clínicas em sistemas computacionais.

3.2. Ontologias e Grafos na Estratificação de Riscos em Saúde Mental

O segundo artigo analisado, intitulado “*Produção de Grafos de Conhecimento em Estratificação de Riscos em Saúde Mental a partir de Dicionários de Dados Semânticos*” (Silva and Bax 2023), propõe uma abordagem ontológica para organizar dados clínicos coletados em processos de avaliação de saúde mental. O estudo parte do desafio de lidar com informações provenientes de questionários mistos — com perguntas abertas e fechadas — que geram dados qualitativos e quantitativos heterogêneos. A proposta consiste em aplicar ontologias para desambiguar os termos clínicos e formalizar semanticamente os dados, gerando, como resultado, grafos de conhecimento expressos em RDF. Ao longo do artigo, os autores descrevem o processo de transformação dos dados brutos em estruturas formais que podem ser exploradas computacionalmente, especialmente na organização do planejamento terapêutico individualizado.

O problema de pesquisa gira em torno da dificuldade de integrar e analisar dados clínicos heterogêneos, comuns em serviços de atenção psicossocial. Enquanto dados estruturados, como pontuações de sintomas, são relativamente fáceis de tratar, dados não estruturados — como respostas livres — carecem de um processo de codificação e representação mais sistemático. A solução apresentada baseia-se no uso de uma ontologia específica para estratificação de riscos em saúde mental (OERSM), e na técnica conhecida como *Semantic Data Dictionary* (SDD), que permite mapear os dados para conceitos ontológicos e gerar automaticamente grafos RDF que preservam a semântica das informações clínicas.

A metodologia é aplicada a um estudo de caso envolvendo dados simulados de pacientes avaliados por uma equipe multiprofissional de um hospital psiquiátrico. Cada entrada no dataset é enriquecida semanticamente com o uso de dicionários ontológicos e metadados, e o script *sdd2rdf* converte esses dados em triplas RDF que compõem o grafo de conhecimento. No grafo resultante, cada paciente, sintoma, pontuação e risco associado torna-se uma entidade identificável e interligada. A estrutura gráfica permite consultas do tipo: “Quais sintomas estão mais associados a pacientes com risco alto?” ou “Quais intervenções estão previstas no plano terapêutico para determinado nível de risco?”. Os grafos, assim, não apenas representam os dados, mas também facilitam a navegação semântica e a extração de informações relevantes ao cuidado clínico.

O artigo não busca aplicar raciocínio automático ou modelos de inferência complexos, mas sim estabelecer uma base ontológica sólida para a organização dos dados em saúde mental. Essa etapa de estruturação é essencial para permitir que, futuramente,

os dados possam ser reutilizados em sistemas inteligentes de apoio à decisão. Ao utilizar os grafos como forma de representação intermediária, o trabalho permite integrar informações que antes estavam dispersas, facilitando análises mais robustas e sistemáticas da situação clínica dos pacientes.

Este estudo evidencia, portanto, a relevância dos grafos no contexto da saúde mental como ferramenta para estruturar, organizar e padronizar informações clínicas de forma semântica. Ao combinar a modelagem ontológica com a geração automática de grafos RDF, o artigo demonstra como é possível transformar dados fragmentados em representações formais e integráveis. Ainda que o foco não esteja na análise automatizada dos grafos, fica claro que sua presença é central para viabilizar tanto a interoperabilidade entre sistemas quanto o aprimoramento dos processos de estratificação e planejamento terapêutico.

3.3. LogSHIELD: Grafos e Detecção de Anomalias em Segurança Cibernética

O artigo de Roy and Chen 2024 apresenta o LogSHIELD, um framework de detecção de anomalias em tempo real baseado na construção e análise de grafos de proveniência extraídos de logs de sistemas. Em ambientes de segurança cibernética, onde a rastreabilidade de ações e a detecção precoce de atividades maliciosas são cruciais, a estruturação dos dados de sistema como grafos permite capturar relações causais e sequenciais entre eventos, agregando contexto semântico às análises automatizadas.

O estudo propõe duas abordagens complementares: uma baseada em *Graph Neural Networks* (GNN), que utiliza redes neurais para capturar representações semânticas dos nós e vizinhanças nos grafos; e outra baseada em *Frequency Domain Analysis* (FDA), que transforma os padrões de conectividade dos grafos em sinais e os analisa como séries temporais para identificar desvios anômalos. Ambas exploram o grafo como estrutura central para representar e inferir padrões complexos de comportamento, o que reforça a relevância dos grafos como base para sistemas inteligentes de detecção.

Ao avaliar grandes volumes de dados de logs, o LogSHIELD demonstrou capacidade de identificar ataques sofisticados, como *ransomwares* e *rootkits*, com alto desempenho de detecção (AUC de até 98%). O uso de técnicas como *Random Walk with Restart* para amostragem de vizinhanças e *clustering* estatístico evidencia como os grafos permitem não apenas representar eventos, mas também extrair padrões de forma eficiente e escalável. O modelo baseado em FDA, por exemplo, se destaca por seu baixo tempo de resposta, sugerindo uma alternativa viável para cenários em que a latência é crítica.

Este artigo exemplifica como grafos estruturados a partir de dados brutos podem ser aplicados com sucesso na área de segurança da informação, oferecendo não apenas melhor desempenho na detecção, mas também interpretabilidade e modularidade na construção de soluções. A lógica por trás da detecção — baseada em conexões, padrões locais e estruturas relacionais — reforça a utilidade dos grafos como modelo base para sistemas inteligentes em contextos que exigem análise contextual de grandes volumes de dados.

4. Discussão e Conclusões

Os três artigos analisados mostram que os grafos podem ser aplicados de formas bem diferentes, dependendo do contexto e do objetivo. Nos dois primeiros casos, voltados à área

da saúde, os grafos são usados junto com ontologias para estruturar informações clínicas de forma mais clara e padronizada. Já no terceiro, voltado à segurança cibernética, os grafos servem para representar relações entre eventos e detectar comportamentos suspeitos com base nesses padrões.

Mesmo com abordagens distintas, todos os trabalhos reforçam a ideia de que os grafos são uma ferramenta versátil, capaz de lidar com dados complexos e variados. Além das aplicações mais tradicionais, como redes sociais ou rotas, os artigos mostram que é possível usar grafos para representar conhecimento, organizar informações clínicas, ou até detectar ataques em sistemas. Isso abre espaço para pensar em novas aplicações que combinem diferentes estratégias, como unir a estrutura semântica das ontologias com métodos de análise de comportamento. Com isso, os grafos se mostram uma boa base para soluções em áreas diversas e com demandas bem específicas.

Referências

- [Cormen et al. 2009] Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to Algorithms*. MIT Press.
- [Mavridis et al. 2025] Mavridis, A., Tegos, S., Anastasiou, C., Papoutsoglou, M., and Meditskos, G. (2025). Large language models for intelligent rdf knowledge graph construction: results from medical ontology mapping. *Frontiers in Artificial Intelligence*, Volume 8 - 2025.
- [Roy and Chen 2024] Roy, K. C. and Chen, Q. (2024). Logshield: A graph-based real-time anomaly detection framework using frequency analysis.
- [Silva and Bax 2023] Silva, E. O. and Bax, M. P. (2023). Produção de grafos de conhecimento e estratificação de riscos em saúde mental a partir de dicionários de dados semânticos. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 16.