



## **Processo para manipulação de grafos de conhecimento usando algoritmo de recomendação**

*Othon Brück*<sup>1</sup>

*Centro Universitário Academia, Juiz de Fora, MG*

*Evaldo de Oliveira da Silva*<sup>2</sup>

*Centro Universitário Academia, Juiz de Fora, MG*

Linha de Pesquisa: Engenharia de Software

### **RESUMO**

A partir da crescente geração e disponibilização de dados na sociedade atual, surge a necessidade de informações para tomadas de decisões ou para obtenção de conhecimento com base em grafos de conhecimento. Grafos de conhecimento agregam entidades (e seus atributos), relacionando-as por propriedade (predicados) formalizados por ontologias. Neste contexto, ontologias tornam-se importantes mecanismos de compartilhamento do conhecimento, pois organizam conceitos e como se relacionam para representar o conhecimento de um domínio. Técnicas de anotação semântica anotam *datasets* com o uso de ontologias com o objetivo de gerar grafos de conhecimentos enriquecidos com dados. Além disso, ontologias são definidas vistas como modelos onde as entidades formam o vocabulário que descreve os entes/indivíduos (instâncias) mais recorrentes do domínio. Entretanto, mesmo com seu poder de demonstração de seus relacionamentos, em grafos de conhecimento podem-se existir informações a mais que são capazes de ser extraídas. Com o objetivo de extrair mais conhecimento dos grafos, este trabalho tem o objetivo de utilizar o algoritmo *Apriori*, que abre caminho para recuperar e classificar dados a fim de apoiar a tomada de decisão em análise de dados, inclusive na área da saúde mental. Desta forma, este trabalho utiliza o *Apriori* para extrair informações de grafos de conhecimentos anotados com dados sobre tratamento da ansiedade. Espera-se com a execução do objetivo proposto, apresentar informações que possam melhorar a orientação dos tratamentos em ansiedade.

---

<sup>1</sup> Discente do Curso de Engenharia de Software do Centro Universitário Academia – UniAcademia. Endereço: Rua Aurora Tristão 595 ap 501 – Centro. Celular: (32) 98894-9694. E-mail: othon.bruck@hotmail.com.

<sup>2</sup> Docente do Curso de Engenharia de Software do Centro Universitário Academia. Orientador.

**Palavras-chave:** Machine Learning, Grafos de conhecimento, Ciência de Dados.

## **ABSTRACT**

*With the increasing generation and availability of data in today's society, the need for information to make decisions or to obtain knowledge based on knowledge graphs has been growing. Knowledge graphs aggregate entities (and their attributes), relating them by properties (predicates) formalized by ontologies. In this context, ontologies become important mechanisms for knowledge sharing, as they organize concepts and how they relate to each other to represent the knowledge of a domain. Semantic annotation techniques annotate datasets using ontologies with the goal of generating data-enriched knowledge graphs. Furthermore, ontologies are defined viewed as models where entities form the vocabulary that describes the most recurrent entities/individuals (instances) of the domain. However, even with its power to demonstrate its relationships, in knowledge graphs there may be more information that is able to be extracted. In order to extract more knowledge from graphs, this work aims to use the Apriori algorithm, which paves the way to retrieve and classify data to support decision making in data analysis, including in the area of mental health. Thus, this work uses Apriori to extract information from knowledge graphs annotated with data on anxiety treatment. It is expected with the execution of the proposed objective, to present information that can improve the targeting of treatments in anxiety.*

## **1 INTRODUÇÃO**

Nos últimos anos, com o avanço no desenvolvimento de novas ferramentas para capturar e analisar dados em diversas áreas de pesquisa e, como dito em Kambatla *et al* (2014), dados que passam do exabytes e continuam crescendo. Com esse rápido crescimento e a complexidade dos dados disponíveis, surge a necessidade de abordagens inovadoras e eficientes para lidar com essa quantidade massiva de informações. Nesse contexto, mostra que grafos de conhecimento têm ganhado destaque como uma poderosa estrutura para representar e explorar relações entre os dados (TIWARI, AL-WASADI e GAURAV, 2021).



Grafos representam a semântica dos dados impostas por ontologias, estruturados com o princípio de sujeito, predicado e objeto, onde o sujeito e o objeto representa a entidade que podem ser uma pessoa, sintoma, lugares entre outros, e o predicado a relação entre as entidades, permitindo a modelagem de relações complexas e o enriquecimento das informações por meio de anotações semânticas (Hogan *et al*, 2021).

Seguido dos grafos de conhecimento também existe o *machine learning* que é um subconjunto da inteligência artificial (IA), disciplina focada no desenvolvimento de algoritmos que permitem que os computadores aprendam e tenham um desempenho melhor sem necessariamente serem programados para cada tarefa. Tendo esse conceito em mente, podem ser utilizados em diferentes áreas como reconhecimento e previsão de padrões, classificação e recomendação, para auxiliar nas tomadas de decisões (JORDAN e MITCHELL, 2015).

Como a quantidade de dados continua a crescer, os métodos analíticos antigos como demonstrados por Saranya. *et al* (2020) tornam-se cada vez mais demorados e ineficientes, para ter uma nova visão e melhor performance sobre os dados, o algoritmo *Apriori* (AL-MAOLEGI, Mohammed; ARKOK, Bassam, 2014) é um algoritmo de *machine learning* que desempenha um papel importante no auxílio desses problemas.

*Apriori* é um dos algoritmos de mineração mais famosos para encontrar associações de regras comuns em um conjunto de transações. Quando aplicados a conjuntos de dados clínicos, podem-se revelar padrões significativos que ajudam a entender melhor os fatores que levam a transtornos mentais e a tomar decisões clínicas informadas. De maneira geral, permite que esses dados sejam processados e interpretados de forma rápida e automatizada, abrindo caminho para descobertas e insights, como demonstrando em Dwyer, Falkai e Koutsouleris (2018) em suas abordagens utilizando de *machine learning* para a psicologia clínica e psiquiatria.

Organizar dados e gerar informações para instituições é uma parte vital para um negócio bem-sucedido como mostra em (LEE, In; SHIN, Yong Jae, 2020). Essa prática permite obter insights valiosos que impulsionam o crescimento e o desenvolvimento das empresas. Por meio da análise cuidadosa dos dados, é possível conhecer as tendências do público-alvo, identificar pontos de escassez e problemas que o afetam.



No entanto, deve-se ressaltar que, apesar da eficiência do grafo de conhecimento na construção de relacionamentos e estruturação de dados, nem todas as informações contidas no grafo são totalmente utilizadas em cada área. Dados valiosos geralmente são perdidos pela análise quantitativa tradicional, limitando os insights potenciais obtidos.

Para potencializar a geração de informações, será apresentado um processo para expandir as análises quantitativas utilizando os dados armazenados em grafos de conhecimento, com *machine learning* para auxiliar nas tomadas de decisões, visto que existem casos que comprovam a eficiência do uso dessa ferramenta no mundo real. Será demonstrado um caso de uso dos grafos na área da psicoterapia, onde tem-se os dados a partir da estratificação de riscos da saúde mental. Aplica-se o processo para prover informações que ajudam os profissionais em suas análises e avaliações.

No restante do artigo a Seção 2 descreve os conceitos da Terapia Cognitivo-Comportamental, grafos de conhecimento, plataformas de bancos de dados semânticos, RDF para acessar dados de bancos semânticos e algoritmos de recomendação. A Seção 3 apresenta o processo criado para manipular os grafos, usando as ferramentas apresentadas e gerando dados para análises. A Seção 4 mostra a aplicação do processo para recomendação de sintomas identificados a partir da Terapia Cognitivo-Comportamental. Ao final, na Seção 5, serão apresentadas as considerações finais.

## **2 REFERENCIAL TEÓRICO**

### **2.1 Terapia Cognitiva-Comportamental**

Técnicas são utilizadas por profissionais em saúde mental para descrever os casos e tratamentos psicoterápicos. A Terapia cognitivo-comportamental tem sido acolhida em diferentes campos que envolvam trabalhos terapêuticos, isso por apresentar um conjunto vasto de estudos e bases científicas, definindo-a como uma prática baseada em evidências em psicologia. Entre suas particularidades, considera-se o uso de protocolos de atendimentos e manuais que guiam a atuação prática a partir do desenho do caso clínico e possíveis diagnósticos. Ademais, tendo em vista que os transtornos de ansiedade se mostram cada vez mais presentes na vida da população, torna-se notória a necessidade crescente de estudo e tratamento destes no campo da Psicologia e de grande relevância para a pesquisa desse transtorno. É importante ressaltar que a



ansiedade se expressa em quatro proporções: através das emoções, do comportamento, dos pensamentos e do corpo. Sendo assim, o tratamento deve abranger todas essas dimensões por intermédio de intervenções emocionais, comportamentais, sociais e cognitivas. Dessa forma, a Terapia Cognitivo-Comportamental, abordagem escolhida para o presente trabalho, objetiva que a terapia seja diretiva, estruturada e orientada ao presente, com o intuito de que o paciente tenha autonomia a partir das intervenções feitas no processo terapêutico e atuando em todos esses âmbitos (MELO E LOURENÇO, 2020).

Para a montagem do protocolo a ser aplicado na pesquisa foi feito o levantamento bibliográfico na literatura sobre o que já se tinha em modelo de protocolo terapêutico em Terapia Cognitivo Comportamental (TCC) para o Transtorno de Ansiedade Generalizada (TAG). Diante disso, foi encontrado, algumas literaturas trazem sobre os protocolos de 16 sessões dentro da TCC. Como por exemplo: o estudo realizado com adultos mais velhos com TAG examinou a eficácia da TCC nos níveis de cortisol apresentados por esses indivíduos, comparando indivíduos que receberam 16 sessões de TCC combinada com o uso de medicação e indivíduos que receberam apenas medicação (REYES e FERMENN, 2017). Destaca-se que no próprio livro de Barlow (2016), o Manual Clínico dos transtornos psicológicos descreve que o tratamento passo a passo, o caso clínico de transtorno de ansiedade generalizada é descrito com 16 sessões.

Além disso, no trabalho de Melo e Lourenço (2020) para transtorno de ansiedade generalizada, utilizando a Terapia Cognitivo Comportamental e em um estudo de caso clínico, foram utilizadas técnicas como a psicoeducação, a reestruturação cognitiva e o relaxamento progressivo, as quais, ao longo das 14 sessões de tratamento, resultaram em uma significativa redução do quadro de ansiedade e no aprendizado de recursos para lidar com possíveis futuras situações ansiogênicas. O processo citado buscou proporcionar à paciente uma redução da ansiedade e um melhor entendimento de seu sistema cognitivo, através de intervenções psicoterapêuticas. Utilizou-se também a técnica de Registro de Pensamentos, questionamento socrático, reestruturação cognitiva, ensinamento de técnicas de respiração diafragmática e relaxamento progressivo que na avaliação final do tratamento que as queixas iniciais de ansiedade melhoraram de forma considerável. (MELO E LOURENÇO, 2020). É importante ressaltar



que mesmo que no estudo de Melo e Lourenço (2020) tenham usado 14 sessões. Trata-se de um caso com uma ansiedade relevante, retratando alta sintomatologia, bem como em virtude de se tratar de um tratamento com um número limitado de sessões trabalhadas, apesar dos protocolos recomendarem, em média, 16 sessões (BARLOW, 2016).

Em um outro estudo de Oliveira (2011), foi utilizado um protocolo com 12 sessões na intervenção com terapia cognitivo-comportamental em um relato de caso com sucesso em diminuição nos níveis de transtorno de ansiedade. Na primeira, houve a escuta e a proposta inicial de psicoterapia. Já na segunda, foi dado início ao seu tratamento, com a psicoeducação a respeito da TCC e do transtorno de ansiedade, sintomas, pensamento automático, erros cognitivos característicos do ansioso, tarefas de casa e responsabilização do paciente no processo psicoterápico. Explicada a conceituação cognitiva ao paciente, preenchendo um formulário próprio em que ele podia relatar uma situação ansiogênica, o pensamento automático emergente, a emoção no momento e o comportamento consequente, solicitou que ele repetisse o exercício, como tarefa de casa, relatando outros eventos ansiogênicos. E, então, nas sessões subsequentes, a explorar tais tarefas, a fim de proceder à reestruturação cognitiva a partir da identificação dos pensamentos automáticos, das emoções e das crenças disfuncionais do paciente.

Ademais, as quatro sessões subsequentes ao treino da conceituação cognitiva sempre eram iniciadas comentando a semana e, em seguida, treinando com o paciente respiração diafragmática e relaxamento progressivo, a fim de reduzir suas tensões musculares e sua ansiedade. E, por fim, as duas sessões subsequentes restringiram-se à avaliação da resolução do problema escolhida pelo paciente. Foram necessárias apenas 12 semanas para esse processo, com bons resultados, uma vez que o objetivo principal foi alcançado. (OLIVEIRA, 2011).

Com a técnica TCC, serão gerados os dados informativos sobre os pacientes como por exemplo sintomas. Utiliza-se esses dados na execução do processo para permitir o entendimento das associações entre os elementos para tratamento em saúde mental.





## 2.2 Grafos de conhecimento (GC)

Em HOGAN *et al* (2021) GCs são uma representação estruturada de informações e conhecimento, organizados na forma de um grafo. Composto por uma estrutura de nós (ou vértices) representando entidades e os relacionamentos entre essas entidades, representados por arestas que conectam os nós. Esses gráficos são usados para modelar e representar o conhecimento sobre um determinado campo de forma mais estruturada e relevante.

Outro estudo, de Parmani (2021), destaca o papel dos grafos na análise de redes sociais. Fornecendo uma estrutura flexível para representar os diversos elementos presentes em uma rede social, como usuários, interações, interesses e comunidades. Com isso é capaz de identificar os padrões, a detecção de comunidades e a análise de influenciadores, auxiliando para um entendimento das interações sociais e permitindo visualização das diversas relações dentro da rede.

Além disso, os GCs são essenciais para a análise de dados e a inteligência artificial em geral. Eles permitem a estruturação e a organização dos dados de forma mais eficiente, facilitando a extração de *insights* valiosos como demonstrados em Parmani (2021). Os algoritmos de *machine learning* e mineração de dados podem explorar as conexões no grafo para identificar padrões, tendências e correlações ocultas, melhorando a tomada de decisões e a previsão de eventos futuros.

O GC vai ser responsável por armazenar os dados que serão utilizados no processo, aproveitando da sua estrutura flexível para entender interações e as entidades presentes.

## 2.3 Plataformas de bancos de dados semânticos

Na área de tecnologia, é importante guardar dados, a fim de realizar pesquisas sobre os dados e gerar informação útil para o negócio, por isso existem as plataformas de bancos de dados semânticos, que ajudam a fazer esse armazenamento. Com eles é possível gerenciar os dados projetados para criar, atualizar, deletar e ler informações de maneira semântica, dentro do artigo de Gutierrez, Hurtado e Mendelzon (2004), conta sobre sua estrutura e a necessidade de seguir os fundamentos.

Ao contrário dos bancos de dados tradicionais, que geralmente armazenam dados em tabelas com estruturas fixas como colunas ou linhas. Os bancos de dados semânticos utilizam modelos de dados flexíveis capazes de representar relacionamentos complexos entre entidades e atributo, facilitando a descoberta de informações relevantes e a realização de análises mais avançadas (GUTIERREZ, HURTADO E MENDELZON, 2004).

Neste trabalho foi utilizado o Blazegraph<sup>3</sup>, que é um banco de dados de grafos de código aberto que funciona como um servidor autossuficiente e promete uma alta performance, permitindo suportar consultas SPARQL.

O SPARQL<sup>4</sup> (*Standard Query Language and Protocol for Linked Open Data*) em Pérez, Arenas e Gutierrez, (2009) é uma linguagem de consulta para o *Resource Description Framework* (RDF), utilizada para expressar consultas em diversas fontes de dados, seja diretamente em RDF ou por meio de um middleware. Com sua estrutura, pode-se realizar consultas indicando os relacionamentos desejados e obter os dados necessários (PEREZ, ARENAS e GUTIERREZ, 2009).

RDF é um padrão para intercâmbio de dados usado para representar dados altamente interconectados. Cada instrução RDF é uma estrutura de três partes que consiste em recursos onde cada recurso é identificado por um URI (*Universal Resource Identifier*). A representação de dados em RDF permite que as informações sejam facilmente identificadas, desambiguadas e interconectadas por sistemas de inteligência artificial (ONTOTEXT, 2023).

#### 2.4 RDF para acessar bancos de dados semânticos

O RDF segundo o livro de Powers (2003) é uma estrutura geral para representar dados na web e é usado para integrar informações de várias fontes, permitindo a padronização de dados com base em relações conhecidas como triplas. Essas triplas consistem em sujeito, predicado e objeto. Com o RDF, é possível expressar relações entre recursos, como documentos, pessoas e objetos abstratos.

---

<sup>3</sup> BlazeGraph. Disponível em: <https://blazegraph.com>

<sup>4</sup> SPARQL. Disponível em: <https://www.w3.org/TR/rdf-sparql-query/>





Dentro das bibliotecas que se utiliza durante o trabalho, a mais importante para acessar o banco de dados semântico Blazegraph foi o Pymantic<sup>5</sup>. Trata-se de uma biblioteca de fácil uso que trabalha com dados RDF.

Por meio do RDF e da biblioteca Pymantic, permite que se conecte ao banco de dados semântico e criar consultas SPARQL dentro do Python.

## 2.5 Algoritmos de Recomendação (AR)

Os ARs desempenham um papel fundamental no mundo atual como demonstra em Cremonesi, Koren e Turrin, (2010), em que sendo constantemente bombardeados por uma quantidade imensa de informações e opções. Elas são responsáveis por filtrar e direcionar conteúdos personalizados, produtos ou serviços que melhor se adequem às preferências e necessidades individuais dos usuários.

Um exemplo notável de AR é o algoritmo *Apriori*. Criado por Rakesh Agrawal e Ramakrishnan Srikant em 1994, com intuito de principalmente para detectar itens ou elementos mais frequentes dentro de transações de bancos de dados e criar regras de associações entre os elementos, podendo ser utilizado em sistemas de recomendações como filmes, compras de produtos, propagandas, investimento e entre outros.

O algoritmo *Apriori* como dito em (AL-MAOLEGI, Mohammed; ARKOK, Bassam, 2014) funciona identificando conjuntos de itens que ocorrem frequentemente juntos em um conjunto de dados. Esses conjuntos de itens são chamados de "*itemsets* frequentes". Em seguida, com base nesses *itemsets* frequentes, o algoritmo gera regras de associação, que são usadas para fazer recomendações.

O processo do algoritmo *Apriori* pode ser dividido em algumas etapas principais. Primeiro, identifica todos os *itemsets* de tamanho 1 (ou seja, conjuntos de um único item) que ocorrem com frequência suficiente no conjunto de dados. Em seguida, combina esses *itemsets* frequentes para formar *itemsets* de tamanho 2 e verifica novamente sua frequência de ocorrência. Esse processo é repetido para gerar *itemsets* de tamanhos maiores, até que não seja possível encontrar mais *itemsets* frequentes.

---

<sup>5</sup> Pymantic. Disponível em: <https://github.com/norcalrdf/pymantic>



Dentro do algoritmo *Apriori*, existem conceitos importantes, como o Suporte, que representa a porcentagem de transações contendo um determinado item. No caso, seria a quantidade de transações que contêm o elemento X dividida pela quantidade total de elementos no conjunto de dados. Isso permite determinar a popularidade do elemento. Outro conceito relevante é a Confiança, que indica a veracidade de uma regra. Por exemplo, se um usuário compra o elemento X, qual é a probabilidade dele comprar o elemento Y? Para descobrir a probabilidade dessa regra, é realizado o cálculo das transações de X e Y dividido pelas transações de X, obtendo assim uma porcentagem que indica a probabilidade de quem adquire o elemento X também adquirir o elemento Y. Por último, existe o *Lift*, que é uma medida da probabilidade de um elemento ser comprado, levando em consideração sua popularidade. Nesse caso, a função seria calculada como  $\text{Confiança}(X \rightarrow Y)$  dividido pelo  $\text{Suporte}(Y)$ . Quando o valor do *Lift* é inferior a 1, é bastante provável que não exista uma relação clara entre os elementos.

O algoritmo *Apriori* no processo tem como dever identificar relações entre os elementos presentes inseridos no algoritmo e assim gerar regras de associações, com as informações sobre quais entidades têm a associação assim como seu "Suporte", "Confiança" e "Lift", que podem ser analisadas e consultadas por profissionais da área e auxiliar em suas tomadas de decisões.

### **3 DEFINIÇÃO DE UM PROCESSO PARA MANIPULAÇÃO DE GC USANDO ALGORITMO DE RECOMENDAÇÃO**

Será demonstrado o processo "AprioriGraph"<sup>6</sup> que consiste em utilizar as ferramentas levantadas durante este trabalho. Para realizar o processo, será realizado um levantamento do que precisa ser analisado. Percorrendo pelo banco de dados semântico que estejam utilizando, com uma consulta que atenda aos critérios. Com a ajuda de uma biblioteca para se conectar ao banco de dados semânticos.

Após a consulta, é necessário fazer uma limpeza dos dados obtidos, já que em muitos casos, eles não estão representados da maneira que necessita ou apresentam

---

<sup>6</sup> AprioriGraph. Disponível em: <https://github.com/OthonBruck/apriorigraph>



uma estrutura indesejada. Nesses casos, deve ser feita uma limpeza das informações para estruturá-las da forma necessária, a fim de encaixá-las no algoritmo *Apriori*.

Nessa etapa, com os dados estruturados da forma que pode ser utilizado no algoritmo *Apriori*. Com o algoritmo, obter as relações mais significativas entre os elementos, permitindo determinar tendências entre eles. Dando como exemplo, se na base de dados existem muitas relações em que os pacientes apresentam os determinados sintomas A, B e C, e em uma nova consulta um paciente é atendido e são identificados apenas os sintomas A e C, o algoritmo indicaria ao profissional a possibilidade de o paciente também ter ou desenvolver o sintoma B.

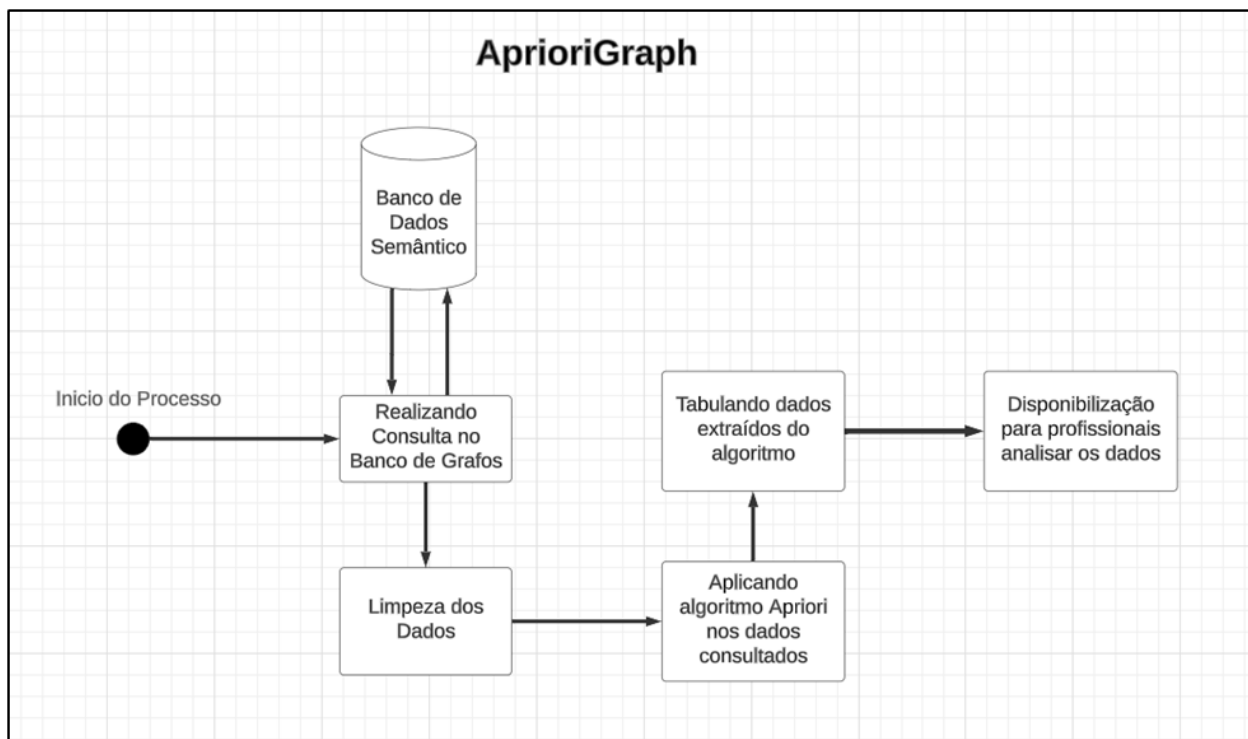
Por fim, realiza-se a tabulação dos dados extraídos dos grafos e transformados pelo algoritmo, a fim de facilitar a visualização dos dados. No caso da continuação do exemplo dado no *Apriori*, terão cinco colunas com as informações extraídas pelo algoritmo podem ser elas “Sintomas que o paciente possui”, “Sintomas que o paciente ainda pode ter/desenvolver”, “Suporte”, “Confiança” e “Lift”.

Com as informações disponibilizadas, o profissional da área poderá analisar e ou ser auxiliado em suas tomadas de decisões com base nos dados gerados pelo algoritmo *Apriori*. Para ilustrar o processo mencionado foi criado um diagrama com as etapas do AprioriGraph no LucidChart<sup>7</sup>.

---

<sup>7</sup> LucidChart. Disponível em: <https://www.lucidchart.com>

Figura 1 – Diagrama do processo AprioriGraph.



Fonte: Imagem produzida pelo próprio autor.

#### 4 APLICAÇÃO DO APRIORIGRAPH PARA RECOMENDAÇÃO DE SINTOMAS IDENTIFICADOS A PARTIR DA TCC

Esta seção apresenta a aplicação do AprioriGraph. O grafo utilizado foi extraído da base de dados do projeto de pesquisa intitulado “Processo sistemático fundamentado em modelagem ontológica para representar o conhecimento em análise quali-quantitativa”. Esse projeto de pesquisa é realizado no Centro Universitário UniAcademia, na Clínica de Psicologia da mesma instituição, e tem a participação de alunos e professores dos cursos de graduação em Psicologia e Engenharia de Software. O grafo possui dados sintéticos extraídos de artigos científicos da área de psicologia, e já analisados pelos integrantes do projeto de pesquisa. O script *sdd2rdf* ofereceu suporte para geração do grafo aqui utilizado.

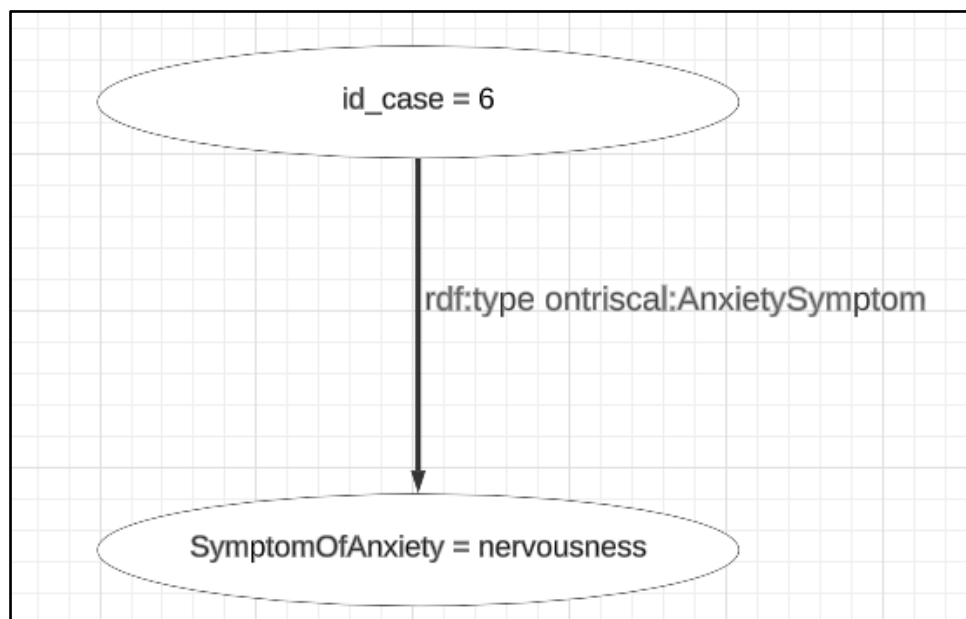
Para iniciar o AprioriGraph, coleta-se os dados que deseja analisar. Neste caso em específico, objetiva-se a análise de sintomas de ansiedade em pacientes. Percorrendo por todo o banco de dados semântico utilizando o Blazegraph, realizando



uma consulta SPARQL que atenda aos critérios. Os dados sintéticos dados possuem informações sobre os sintomas de 170 pacientes.

Abaixo demonstra uma figura demonstrativa dos dados em grafo que se utiliza a entidade chamada “id\_case” que representa o paciente, outra entidade chamada “SymptomOfAnxiety” para representar o sintoma e a seta que liga ambos que é o relacionamento entre eles.

**Figura 2** – Demonstração exemplo de um grafo.



Fonte: Imagem produzida pelo próprio autor.

Depois da consulta, executa-se a limpeza dos dados, pois eles não estão apresentados de uma forma que possa ser utilizada no algoritmo *Apriori* que está sendo utilizado. Portanto, é necessário realizar a limpeza e organização dos dados de acordo com a estrutura desejada, que consiste em listar os sintomas dos pacientes em linhas dentro de um arquivo CSV. A Figura 3 demonstra como ficam os dados após a limpeza, considerando que as linhas são cada paciente e as colunas os sintomas que foram registrados.

**Figura 3** – Dados representativos dos sintomas de ansiedade de cada paciente em CSV.

```

anxiety.csv
27 feeling terrified and scared
28 headaches,family pressure
29 accelerated heart,hot flashes,muscle tension
30 difficulty breathing,sweating,constant fear
31 feeling of suffocation,burning face feeling
32 muscle tension,headaches
33 inability to relax,feeling terrified and scared,family pressure
34 accelerated heart,difficulties initiating sleep polyphasic and non-restorative sleep
35 hot flashes,constant fear,nervousness
36 feeling of suffocation,sweating,headaches
37 burning face feeling,muscle tension,family pressure
38 difficulty breathing,feeling terrified and scared
39 accelerated heart,feeling of suffocation,difficulty breathing
40 hot flashes,sweat,muscle tension
41 inability to relax,constant fear,nervousness
42 feeling terrified and scared,difficulties initiating sleep polyphasic and non-restorative sleep,headaches
43 family pressure
44 feeling of suffocation,sweat
45 burning face feeling,muscle tension,constant fear
46 difficulties initiating sleep polyphasic and non-restorative sleep,headaches
47 accelerated heart,hot flashes,family pressure
48 difficulty breathing,muscle tension
49 accelerated heart,difficulty breathing,hot flashes,sweat,muscle tension,constant fear
50 feeling of suffocation,hot flashes,sweat, burning face feeling,inability to relax,constant fear

```

Fonte: Imagem produzida pelo próprio autor.

Nesta etapa, a partir dos dados estruturados de acordo com a necessidade de pesquisa utiliza-se o algoritmo *Apriori*. Esse algoritmo permite identificar as relações mais relevantes entre os sintomas de ansiedade, possibilitando a determinação de tendências entre eles. Ao utilizar os dados e o algoritmo, obter-se essas informações em *log* e já é capaz identificar as recomendações como mostra na Figura 4, divididos pelas colunas, “Sintoma presente” representa caso o paciente tenha esse sintoma, “Sintoma que pode ocorrer/desenvolver” o sintoma que tem como probabilidade de ter/desenvolver, “Suporte” demonstrando a frequência dentro do *dataset*, “Confiança” que representa o probabilidade da regra acontecer e “*Lift*” representando se a regra dentro do *dataset* é confiável.

**Figura 4** – Log das informações geradas após a passagem do algoritmo *Apriori*.

	Sintoma presente	Sintoma que pode ocorrer/desenvolver	Suporte	Confiança	Lift
10	hot flashes	sweat	0.141176	0.352941	2.000000
11	inability to relax	nervousness	0.029412	0.277778	1.967593
9	feeling of suffocation	sweat	0.111765	0.339286	1.922619
7	difficulty breathing	sweat	0.076471	0.333333	1.888889
2	accelerated heart	sweating	0.176471	0.454545	1.717172

Fonte: Imagem produzida pelo próprio autor.



Por fim, é necessário utilizar a tabulação dos dados extraídos dos grafos e processados pelo algoritmo, o que facilita a visualização das informações, neste caso foi feito um *script*, para criar um Excel com as informações. Esses dados são apresentados em três colunas: "Sintoma presente", "Sintoma que pode ocorrer/desenvolver" e "Suporte", incluindo "Confiança" e "Lift". Na Figura 5 mostra após a tabulação dos dados dentro de um Excel.

**Figura 5** – Dados representativos das associações realizadas pelo algoritmo *Apriori* em Excel.

1	Sintoma presente	Sintoma que p	Suporte	Confiança	Lift
2	hot flashes	sweat	0.14	0.35	2
3	inability to relax	nervousness	0.03	0.28	1.97
4	feeling of suffocation	sweat	0.11	0.34	1.92
5	difficulty breathing	sweat	0.08	0.33	1.89
6	accelerated heart	sweating	0.18	0.45	1.72

Fonte: Imagem produzida pelo próprio autor.

Com isso, inicia-se a análise e verificação das tendências de sintomas dentro do conjunto de dados. Nesse caso, filtrando a tabela do maior para o menor em relação ao Lift, que é uma medida utilizada para determinar a existência de uma relação entre as regras. Por exemplo, observa-se que o sintoma "sweat" está relacionado a uma probabilidade de 35% de ocorrer caso haja o sintoma "hot flashes", 34% de chance em caso de "feeling of suffocation" e 33% de chance em caso de "difficulty breathing" estejam presentes, com base na confiança.

Baseando-se nas informações geradas, podem auxiliar profissionais nas tomadas de decisões da área de psicoterapia sobre os pacientes que estão lidando, além de também possibilitar a descobertas de relações entre os sintomas dos seus pacientes.

## 5 CONSIDERAÇÕES FINAIS

Considerando o conteúdo deste trabalho explorando a aplicação do AR em GCs com dados da psicoterapia, torna-se possível observar o potencial dessas técnicas para auxiliar tanto os profissionais da área quanto os pacientes, proporcionando um apoio personalizado e eficaz no processo terapêutico.

Ao utilizar os dados da psicoterapia em um GC, é possível representar as interações entre os diferentes elementos envolvidos no processo terapêutico, como os sintomas de ansiedade. Essa representação com os dados coletados ao longo das sessões, permite que os ARs identifiquem padrões e relações entre os elementos, fornecendo recomendações personalizadas para os pacientes e profissionais terapeutas.

Entretanto, mesmo com os benefícios visíveis neste trabalho, deve-se enfatizar que ainda há possibilidade de melhorar as recomendações no contexto psicoterapêutico, para isso é necessário ter estudos futuros que leve em consideração não apenas os sintomas apresentados pelo paciente, mas também seu *background* e história pessoal. Já que compreender as experiências passadas, traumas, padrões de comportamento e características individuais de cada paciente que é uma das técnicas que se utilizam dentro da terapia pode fornecer informações valiosas para um tratamento mais eficaz.

## REFERÊNCIAS

AL-MAOLEGI, Mohammed; ARKOK, Bassam. An improved Apriori algorithm for association rules. **arXiv preprint arXiv:1403.3948**, 2014.

CHEN, Xiaojun; JIA, Shengbin; XIANG, Yang. A review: Knowledge reasoning over knowledge graph. **Expert Systems with Applications**, v. 141, p. 112948, 2020.

CREMONESI, Paolo; KOREN, Yehuda; TURRIN, Roberto. Performance of recommender algorithms on top-n recommendation tasks. In: **Proceedings of the fourth ACM conference on Recommender systems**. 2010. p. 39-46.

DWYER, Dominic B.; FALKAI, Peter; KOUTSOULERIS, Nikolaos. Machine learning approaches for clinical psychology and psychiatry. **Annual review of clinical psychology**, v. 14, p. 91-118, 2018.



GUTIERREZ, Claudio; HURTADO, Carlos; MENDELZON, Alberto O. Foundations of semantic web databases. In: **Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems**. 2004. p. 95-106.

HOGAN, Aidan *et al.* Knowledge graphs. **ACM Computing Surveys (CSUR)**, v. 54, n. 4, p. 1-37, 2021.

INOKUCHI, Akihiro; WASHIO, Takashi; MOTODA, Hiroshi. An apriori-based algorithm for mining frequent substructures from graph data. In: **Principles of Data Mining and Knowledge Discovery: 4th European Conference, PKDD 2000 Lyon, France, September 13–16, 2000 Proceedings 4**. Springer Berlin Heidelberg, 2000. p. 13-23.

LEE, In; SHIN, Yong Jae. Machine learning for enterprises: Applications, algorithm selection, and challenges. **Business Horizons**, v. 63, n. 2, p. 157-170, 2020.

JORDAN, Michael I.; MITCHELL, Tom M. Machine learning: Trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255-260, 2015.

KAMBATLA, Karthik *et al.* Trends in big data analytics. **Journal of parallel and distributed computing**, v. 74, n. 7, p. 2561-2573, 2014.

LINDSAY, Grace W. Attention in psychology, neuroscience, and machine learning. **Frontiers in computational neuroscience**, v. 14, p. 29, 2020.

ONTOTEXT. **What is RDF?**. Disponível em: <https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf/> . Acesso em : 30 de jun de 2023.

PARNAMI, Archit *et al.* Transformation of Node to Knowledge Graph Embeddings for Faster Link Prediction in Social Networks. **arXiv preprint arXiv:2111.09308**, 2021.

PÉREZ, Jorge; ARENAS, Marcelo; GUTIERREZ, Claudio. Semantics and complexity of SPARQL. **ACM Transactions on Database Systems (TODS)**, v. 34, n. 3, p. 1-45, 2009.

POWERS, Shelley. **Practical RDF: solving problems with the resource description framework**. " O'Reilly Media, Inc.", 2003.

PUJARA, Jay *et al.* Knowledge graph identification. In: **The Semantic Web–ISWC 2013: 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I 12**. Springer Berlin Heidelberg, 2013. p. 542-557.



SARANYA, T. *et al.* Performance analysis of machine learning algorithms in intrusion detection system: A review. **Procedia Computer Science**, v. 171, p. 1251-1260, 2020.

SMITH, Laurence D. *et al.* Constructing knowledge: The role of graphs and tables in hard and soft psychology. **American Psychologist**, v. 57, n. 10, p. 749, 2002.

TIWARI, Sanju; AL-ASWADI, Fatima N.; GAURAV, Devottam. Recent trends in knowledge graphs: theory and practice. **Soft Computing**, v. 25, p. 8337-8355, 2021.

WANG, Zhen *et al.* Knowledge graph embedding by translating on hyperplanes. In: **Proceedings of the AAAI conference on artificial intelligence**. 2014.

YARKONI, Tal; WESTFALL, Jacob. Choosing prediction over explanation in psychology: Lessons from machine learning. **Perspectives on Psychological Science**, v. 12, n. 6, p. 1100-1122, 2017.