

MatheusPettersBevilaqua

2024-07-03 - 20:00 -> 20:30 -> 30 minutos => 20:39

Por curiosidade ... Veio primeiro o interesse por Grafos Redes Complexas ou Medicamentos
Quanto CIDs não levaram ao UNII?

Web scrapping

ANÁLISE E ASSOCIAÇÃO DE MEDICAMENTOS PLANTAS FITOTERÁPICAS ATRAVÉS DE REDES COMPLEXAS

Matheus Petters Bevilaqua, Aurélio Faustino Hoppe – Orientador

Curso de Bacharel em Ciência da Computação
Departamento de Sistemas e Computação
Universidade Regional de Blumenau (FURB) – Blumenau, SC – Brasil
mbevilaqua@furb.br, aureliof@furb.br

Resumo: Este trabalho apresenta o desenvolvimento de um modelo de redes complexas para associação de medicamentos fitoterápicos a medicamentos químicos, visando estabelecer relações de equivalência. Para tanto, foram utilizados dados do DrugBank e do IMPPAT, criando-se duas redes complexas: uma bipartida, com dados de plantas fitoterápicas e compostos utilizados em fármacos, e outra tripartida, com a inclusão de um terceiro grupo - partes de plantas - que intermedeia a relação entre fitoterápicos e químicos. Ambas as redes foram processadas pelo algoritmo de Louvain para formar comunidades e permitir a análise dos grupos, sendo validadas pelo método de Within-inter-cluster e pelo coeficiente de Jaccard. Observou-se, por exemplo, que plantas de um mesmo gênero tendem a estar relacionadas a uma mesma família de medicamentos. Concluiu-se que a abordagem por redes complexas e detecção de comunidades mostrou-se promissora para revelar padrões ocultos e orientar a busca por novas soluções nas áreas da fitoterapia e farmacologia, fornecendo informações importantes sobre o relacionamento entre plantas e compostos químicos.

Palavras-chave: Medicamentos. Fitoterápicos. Remédios químicos. Associação. Redes complexas. Detecção de comunidades. Web scrapping.

1 INTRODUÇÃO

Os erros de medicação representam um desafio global com impactos alarmantes, gerando custos anuais de 42 bilhões de dólares e resultando em pelo menos uma morte por dia (Organização Pan-Americana da Saúde - OPAS, 2021). No Brasil, a situação também é preocupante, no qual o Ministério da Saúde (2020?) estima que mais da metade dos medicamentos são prescritos, dispensados ou vendidos inadequadamente, levando ao uso incorreto por parte de metade dos pacientes. Diante dessa problemática, segundo o Ministério da Saúde (2012, p. 13), a fitoterapia emerge como uma alternativa terapêutica complementar promissora, utilizando plantas ou seus derivados como constituintes ativos em medicamentos.

De acordo com a Organização Mundial da Saúde (2015), a fitoterapia encontra respaldo tanto no conhecimento popular tradicional quanto em crescentes evidências científicas, sendo utilizada por 60% da população mundial. No Brasil, a prática é regulamentada pela Agência Nacional de Vigilância Sanitária (Anvisa) e, como destacado por Santos *et al.* (2011), oferece a oportunidade de tratar o organismo de forma eficaz com medicamentos naturais.

No Brasil, o uso de fitoterápicos começou a ser adotado nas últimas décadas pelo Ministério da Saúde por meio de políticas como a Política Nacional de Plantas Medicinais e Fitoterápicos (Ministério da Saúde, 2006), a Política Nacional de Práticas Integrativas e Complementares no SUS (Ministério da Saúde, 2015) e o caderno de Práticas Integrativas e Complementares (Ministério da Saúde, 2012), que buscam integrar práticas da medicina tradicional à medicina convencional. Tais políticas, segundo o Ministério da Saúde (2015), visam garantir o acesso seguro e eficaz a esses recursos naturais, contribuindo para a diversificação terapêutica e o fortalecimento da produção de fitoterápicos no país.

Segundo Huang *et al.* (2013), aplicando o conhecimento e uso da fitoterapia ao problema da associação medicamentosa, observa-se que é possível utilizar medicamentos fitoterápicos para complementar ou até substituir um medicamento sintético no tratamento de algum sintoma ou enfermidade. Para isso, segundo os autores, é necessário conhecer e conseguir prever os possíveis sintomas e reações que um medicamento pode causar em um dado indivíduo, para então decidir se e com qual medicamento fitoterápico o paciente pode ser tratado, ou ainda se algum de seus medicamentos pode ser substituído por um fitoterápico. Nesse sentido, a farmacologia de redes surge como uma ferramenta poderosa para mapear e analisar as complexas interações entre moléculas, identificando potenciais sinergias e antagonismos entre componentes fitoterápicos e medicamentos convencionais.

Além disso, Huang *et al.* (2013) também apontam para a necessidade da criação de diretrizes claras para a prescrição e o uso racional de fitoterápicos, respaldadas em evidências científicas robustas. A formação continuada dos profissionais de saúde, capacitando-os para a prescrição segura e eficaz de plantas medicinais, também é crucial para garantir o sucesso dessa abordagem terapêutica integrada. A educação em saúde voltada para a população, por sua vez, é

fundamental para promover o uso consciente e responsável dos fitoterápicos, desmistificando crenças e incentivando a busca por orientação profissional qualificada.

Nesse cenário, de acordo com Jia *et al.* (2017), o uso de redes complexas parece promissor na identificação das correlações presentes entre sintéticos e fitoterápicos, e na identificação de quais podem servir para tratar os sintomas no lugar de um medicamento químico. Ainda segundo os autores, o uso de redes complexas pode ser utilizado para analisar e estudar relações entre objetos de diversas classes, como domínios biológicos e redes sociais. Para Barabasi *et al.* (2002), o uso de redes complexas é cada vez mais recorrente para modelar sistemas reais e artificiais, pois essas redes podem representar desde a análise de reações químicas até a dinâmica dos relacionamentos que permeiam a sociedade, como, por exemplo, as redes sociais e a internet. Além disso, a abordagem de redes complexas é utilizada para representação de conhecimento, tratamento de dados e modelagem de sistemas complexos, ou seja, sistemas formados por muitas partes inter-relacionadas, possivelmente de forma não linear, que apresentam comportamentos emergentes e em multiescolhas (BARABASI, 2002).

Diante do exposto, este trabalho propõe um modelo computacional baseado em redes complexas para investigar a associação entre plantas, compostos fitoterápicos e medicamentos farmacológicos. O objetivo central da pesquisa é o desenvolvimento de um sistema capaz de correlacionar medicamentos fitoterápicos a seus potenciais equivalentes sintéticos. Para tanto, a pesquisa se desdobra nos seguintes objetivos específicos: (i) verificar a viabilidade do uso de redes complexas para identificar a correlação entre um fármaco fitoterápico e um equivalente sintético; (ii) identificar as propriedades topológicas da rede, analisando indicadores como grau nodal, intermediação, agrupamento, modularidade, etc. para entender a estrutura da rede e como os diferentes elementos (plantas, compostos, medicamentos) se conectam; (iii) mapear e caracterizar os principais clusters (agrupamentos) formados na rede, observando se eles agrupam elementos com características similares (por exemplo, mesma classe terapêutica, entre outras).

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo está subdividido em quatro seções. A seção 2.1 discorre sobre fitoterapia. A seção 2.2 apresenta conceitos fundamentais de grafos, enquanto a seção 2.3 aborda redes complexas e sua capacidade de representar a situação estudada. Por fim, a seção 2.4 apresenta os trabalhos correlatos identificados durante a revisão bibliográfica, contextualizando a pesquisa atual em relação ao estado da arte.

2.1 FITOTERAPIA

Segundo o Ministério da Saúde (2012, p. 13), fitoterapia é uma abordagem terapêutica que faz uso de medicamentos cujos componentes ativos são provenientes de plantas ou substâncias derivadas de vegetais. Além disso, a prática da fitoterapia, assim como outras práticas integrativas à medicina é antiga e foi observada em diversas culturas ao redor do mundo. No Brasil, esta prática foi incentivada pelo governo, sendo regulamentada pela Agência Nacional de Vigilância Sanitária (ANVISA). Ainda de acordo com o Ministério da Saúde (2012, p. 13), as raízes da fitoterapia são antigas e a prática foi observada no começo de civilizações anciãs no continente asiático, em particular na Índia e na China. Nessas civilizações, há registros escritos do uso da técnica, indicando que conforme essas culturas se desenvolviam, o uso de plantas medicinais já era uma prática explorada e utilizada para tratar de enfermidades e sintomas diversos.

O Ministério da Saúde (2012, p. 13) destaca ainda que enquanto medicamentos ocidentais são utilizados para curar doenças graves ou agudas, a fitoterapia é usada especialmente em casos de doenças crônicas. Ainda de acordo com o órgão público, a disseminação da medicina tradicional chinesa através do mundo contribuiu para a popularização da prática em diversos países. Por isso, para a Organização Mundial da Saúde, o uso de plantas medicinais é o recurso de medicina tradicional mais utilizado ao redor do mundo, sendo extensivamente incorporado nos serviços de saúde pública da maioria dos países, sobretudo na atenção primária a saúde (WORLD HEALTH ORGANIZATION, 2015). No Brasil, de acordo com o Ministério da Saúde (2012, p. 9), o uso de fitoterápicos e derivados de plantas vem sendo usados pela população brasileira há muito tempo, sendo um dos principais recursos da medicina tradicional e da medicina complementar e alternativa. O Ministério da Saúde (2006, p. 9) também destaca que o Brasil é o país com maior biodiversidade no planeta e que detém um grande conhecimento tradicional associado ao uso de medicamentos naturais, tendo grande potencial para pesquisas e desenvolvimento de tecnologias acerca da terapêutica fitoterápica.

No contexto regulatório, segundo Oshiro *et al.* (2016), muitos países possuem diretrizes específicas para o registro e uso de medicamentos fitoterápicos. De fato, a caracterização e padronização das plantas utilizadas, a identificação das substâncias ativas, e a garantia de qualidade, segurança e eficácia são pontos cruciais nesse processo. Diferentemente dos medicamentos sintéticos, os fitoterápicos frequentemente envolvem uma abordagem mais holística, explorando os benefícios terapêuticos de compostos naturais presentes nas plantas. A utilização de plantas medicinais, fundamentada em saberes populares e na convicção de que, por serem de origem natural, não provocam efeitos indesejados, resultou em pouca avaliação por meio de estudos pré-clínicos e clínicos de plantas medicinais.

Tendo em vista a importância dessa regulamentação, de acordo com Oshiro *et al.* (2016), o governo brasileiro publicou legislações específicas sobre o registro de fitoterápicos no Brasil. Como evidenciado por Oshiro *et al.* (2016), a normatização do registro de fitoterápicos é um tema antigo, sendo abordado pela primeira vez através da Portaria nº 22, de 30 de outubro de 1967, exigindo, entre outras coisas, informações sobre a identificação botânica, caracterização da droga vegetal, características físico-químicas, fitoquímicas e fundamentação para o emprego terapêutico. Ainda segundo Oshiro *et al.* (2016), desde a publicação dessa portaria, várias atualizações a essa legislação foram feitas, sendo a mais recente a Resolução nº 26 de 13 de maio de 2014. Nessas atualizações, várias melhorias para comprovação de segurança e eficácia foram anexadas, como a apresentação de estudos de toxicologia e farmacologia pré-clínica e clínica, ou a comprovação bibliográfica de uso seguro por um período de pelo menos 20 anos. A legislação de 2014 ainda diferencia os medicamentos fitoterápicos de produtos tradicionais fitoterápicos, sendo estes os com eficácia comprovada por literatura técnico-científica e aqueles os comprovados por estudos científicos (OSHIRO *et al.*, 2016). A legislação brasileira, ao diferenciar medicamentos fitoterápicos de produtos tradicionais, reforça a necessidade de um sistema robusto para identificação e classificação de substâncias, capaz de garantir a precisão na composição e segurança dos produtos.

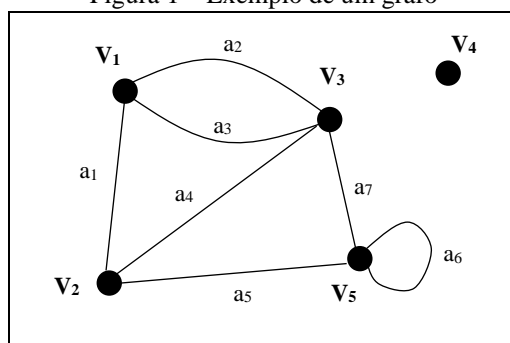
Nesse contexto de mapeamento detalhado, a identificação precisa das substâncias presentes nos fitoterápicos é crucial. Em 2006, a agência americana Food and Drug Administration (FDA) criou o Unique Ingredient Identifier (UNII), um identificador único e alfanumérico ligado a uma substância específica (FOOD AND DRUG ADMINISTRATION, 2024?). De acordo com a instituição, esse identificador serve para gestão das informações ligadas a uma substância, como a sua estrutura molecular e propriedades descritivas, trazendo mais precisão na identificação de substâncias presentes em alimentos ou remédios, por exemplo. O identificador UNII mostra-se como uma ferramenta poderosa para o mapeamento de fitoterápicos, facilitando a identificação de substâncias comuns a plantas medicinais e medicamentos sintéticos, e contribuindo para a pesquisa e desenvolvimento de novos produtos.

2.2 TEORIA DOS GRAFOS

A teoria dos grafos, segundo Melo (2014), tem sua origem no século XVIII, que para a história da matemática, é considerada uma origem recente. O autor ainda comenta que no século XX, a teoria dos grafos já estava desenvolvida, apresentando a sua importância devido as ligações e aplicações com outras ciências. Melo (2014) complementa que mesmo o termo grafo sendo utilizado pela primeira vez por J. J. Sylvester em 1877, a resolução de Euler do Problema das Pontes de Königsberg, publicada em 1736, é referida como primeira publicação da teoria.

A teoria dos grafos, segundo Feofiloff, Kohayakama e Wakabayashi (2011, p. 5) “[...] estuda objetos combinatórios – grafos – que são um bom modelo para muitos problemas para vários ramos da matemática, da informática, da engenharia e da indústria”. Rabuske (1992) comenta que a teoria dos grafos está relacionada com muitos ramos da matemática, como teoria dos grupos, matrizes, probabilidade, análise numérica, topologia e combinatória. A Figura 1 mostra a representação gráfica de um grafo.

Figura 1 – Exemplo de um grafo



Fonte: adaptado de Rabuske (1992).

Na Figura 1 pode ser observado um grafo, que segundo definição de Costa (2011, p. 17) “um grafo (finito) G é formado por um par $(V(G), A(G))$ onde $V(G)$ é um conjunto finito não vazio e $A(G)$ uma família de pares não ordenados de elementos, não necessariamente distintos, de $V(G)$ ”. Para o grafo da Figura 1, têm-se o seguinte conjunto de vértices, $V(G) = \{V_1, V_2, V_3, V_4, V_5\}$ e arestas, $A(G) = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7\}$.

Pinto (2018) descreve que um grafo pode ser categorizado como direcional, no qual as arestas têm um sentido, com um vértice de origem e um vértice de destino, ou não direcional, onde o sentido da aresta não importa. Carvalho (2020) destaca que a característica principal de um vértice, é o seu grau, que representa o número de ligações que o vértice possui e, a partir dele, é possível determinar o grau médio da rede e, consecutivamente, identificar os nós *Hub*. Segundo Rabuske (1992, p. 17), a terminologia utilizada na teoria dos grafos compreende as definições dos elementos, características e propriedades dos grafos. Os elementos de um grafo são:

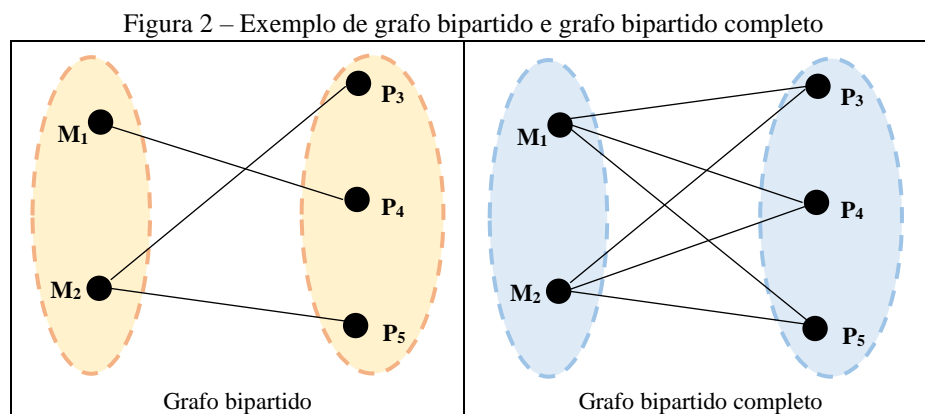
- a) laços: são arestas que as extremidades estão ligadas ao mesmo vértice;
- b) vértices adjacentes: quando dois vértices estão ligados pela mesma aresta;
- c) arestas paralelas: São duas arestas com as mesmas extremidades, ou seja, estão ligadas aos mesmos vértices;
- d) vértice isolado: É um vértice que não é adjacente a nenhum outro vértice;
- e) grau de um vértice: número de arestas que se ligam a um vértice.

A Figura 1 ilustra os seguintes exemplos: As arestas a_2 e a_3 são paralelas porque incidem nos mesmos vértices; o vértice V_4 é um vértice isolado visto que não se liga a nenhum outro; o vértice V_5 tem um laço, que se caracteriza por ser a origem e o destino de uma aresta; o vértice V_2 é adjacente aos vértices V_1 , V_3 e V_5 , o vértice V_5 é adjacente aos vértices V_1 e V_3 e a si mesmo, o vértice V_1 é adjacente aos vértices V_2 e V_3 ; e os vértices V_1 e V_2 tem grau 3, o vértice V_4 tem grau 0, o vértice V_3 e V_5 tem grau 4.

Nesse contexto, surge também a relevância da análise de subgrafos. Um subgrafo consiste em um subconjunto de vértices e arestas de um grafo maior (SCIENCE DIRECT, 2021?). Formalmente, dado um grafo $G=(V,E)$, onde V é o conjunto de vértices e E é o conjunto de arestas, um subgrafo $G'=(V',E')$ é tal que $V' \subseteq V$ e $E' \subseteq E$, com as arestas em E' conectando apenas os vértices em V' . No grafo da Figura 1, um possível subgrafo seria a relação entre V_1 e V_2 , no qual o conjunto de vértices $V' = \{V_1, V_2\}$ e o conjunto de arestas $E' = \{a_1\}$. Contudo, subgrafos são utilizados para analisar propriedades locais de grafos complexos, permitindo uma investigação detalhada das estruturas internas e características específicas.

Pavlopoulos *et al.* (2018) destacam que os grafos podem ser categorizados em classes, dentre as quais se destacam o grafo completo, o regular e o bipartido. Um grafo completo é definido como um grafo em que, para qualquer par de vértices distintos, existe uma aresta conectando-os diretamente. Formalmente, um grafo $G = (V, E)$ é completo se, e somente se, para todo $u, v \in V$, com $u \neq v$, temos $\{u, v\} \in E$. Em outras palavras, a relação de adjacência entre os vértices é máxima. Por outro lado, um grafo regular é caracterizado por uma distribuição uniforme de arestas incidentes em cada vértice. Formalmente, um grafo G é dito regular de grau k se, e somente se, todo vértice $v \in V$ possui grau k , ou seja, $d(v) = k$. Isso significa que cada vértice está conectado a exatamente k outros vértices.

Um grafo bipartido é um grafo em que os vértices podem ser divididos em dois conjuntos e nenhuma aresta do grafo se liga a dois vértices do mesmo conjunto, ou seja, cada aresta liga vértices de conjuntos diferentes. Quando um grafo bipartido possui os vértices de um conjunto ligados a todos os vértices do outro conjunto, ele é chamado de grafo bipartido completo (Pavlopoulos *et al.*, 2018). A Figura 2 mostra um exemplo de grafo bipartido e grafo bipartido completo.



Fonte: adaptado de Pavlopoulos *et al.* (2018).

Na Figura 2 na **esquerda** é mostrado um grafo bipartido com dois conjuntos de vértices que representam plantas e medicamentos, o conjunto de plantas são rotulados com nomes (P_3 , P_4 e P_5), e o conjunto de medicamentos tem vértices rotulados com nomes (M_1 e M_2). No grafo a direita da Figura 2 foram acrescentadas ligações ao grafo para que todos os vértices do conjunto plantas tenham ligações com todos os vértices do conjunto medicamentos e vice-versa, dessa forma o grafo passou a ser um grafo bipartido completo. Em outras palavras, em um grafo bipartidos todas as conexões ocorrem entre vértices de subconjuntos diferentes, tornando-os particularmente úteis para modelar relações entre dois tipos distintos de entidades, como plantas medicinais e compostos químicos. Este tipo de grafo é amplamente utilizado em diversas áreas da ciência, incluindo biologia, ciência da computação e estudos de redes complexas, devido à sua capacidade de representar de maneira clara e eficiente interações bipartidas complexas. A estrutura dos grafos bipartidos permite uma análise detalhada das conexões entre os dois tipos de nós, oferecendo insights sobre a organização e a dinâmica das redes que eles representam (Pavlopoulos *et al.*, 2018).

2.3 REDES COMPLEXAS

Segundo Nahal (2014), as redes complexas são estruturas topológicas que não seguem um padrão regular, são baseadas em grafos, formadas por elementos (vértices) e conexões (arestas). De modo simples, as redes complexas são representações de um domínio e da relação entre seus elementos através de ligações. Neste contexto, nota-se a importância de conhecimentos de teoria dos grafos para representar e entender as relações entre os elementos que constituem a rede. Mata (2020) ainda expande o conceito, afirmando que redes complexas são denominadas dessa forma pois o comportamento da rede como um todo não pode ser definido pelos seus componentes individuais.

Nahal (2014) ressalta que uma característica importante das redes complexas é a sua natureza dinâmica. A partir disso, entendendo que os objetos e relacionamentos dentro da rede são dinâmicos, torna-se necessário analisar o comportamento da rede não só em um instante, mas em uma janela temporal. A observação da rede através de um período de tempo possibilita observar o comportamento e o movimento das relações e, por consequência, da rede como um todo.

Além disso, segundo Ribeiro (2017), a aplicação de técnicas de análise de comunidades em redes complexas revela agrupamentos naturais de nós que compartilham propriedades ou funções semelhantes. A identificação dessas comunidades contribui para uma compreensão mais refinada das estruturas subjacentes e dos padrões de interconexão na rede. A análise de comunidades não apenas fornece informações sobre a modularidade da rede, mas também destaca a importância de determinados grupos de elementos em relação ao funcionamento global. Ribeiro (2017) ainda destaca que essa abordagem é particularmente valiosa na identificação de subconjuntos de interações em sistemas complexos, como redes biológicas, no qual grupos de proteínas podem formar módulos funcionais que desempenham papéis específicos e coordenados no contexto mais amplo do organismo.

Nesse contexto, dentre os algoritmos de detecção de comunidades tem-se o algoritmo de Louvain. O método Louvain é um algoritmo heurístico que se baseia na otimização da modularidade Q (BLONDEL *et al.*, 2008). Ele pode ser visto como um método de agrupamento aglomerativo, onde a entrada do algoritmo é uma rede com n vértices. O algoritmo é dividido em duas fases distintas: (i) a otimização da modularidade e (ii) a agregação da comunidade. Para a otimização da modularidade o algoritmo utiliza a fórmula apresentada na Equação 1, que ordena todos os vértices da rede aleatoriamente, e percorre os vértices trocando-os de comunidades até obter um aumento da modularidade.

$$\Delta M = \left[\frac{\sum_{in} + 2k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (1)$$

Na Equação 1, \sum_{in} representa a soma dos pesos das arestas que fazem parte da comunidade, \sum_{tot} a soma dos pesos de todas as arestas que têm vínculo com algum vértice da comunidade, k_i representa a soma dos pesos das arestas vinculadas ao vértice i , k_i e in são a soma dos pesos das arestas do vértice i para com os outros vértices da comunidade, e com m representando a soma dos pesos de todas as arestas da comunidade. No processo de agregação da comunidade todos os vértices pertencentes a mesma comunidade são agrupados em um único vértice, suas arestas também são agrupadas, com as arestas internas transformadas em uma aresta em loop (RITA, 2020). No Quadro 1 é apresentado o pseudocódigo do algoritmo de detecção de comunidade Louvain, sendo possível observar cada fase do algoritmo.

Quadro 1 – Pseudocódigo do algoritmo Louvain

```
enquanto houver mudança da modularidade faça
  enquanto mudar vértices de comunidade melhora a modularidade faça
    cria uma comunidade para cada vértice  $v \in V(G)$ ;
    para cada  $v_i \in V(G)$  faça
      para cada  $v_j$  vizinho de  $v_i$  faça
        avalia o ganho de modularidade obtido ao mover o vértice  $v_i$  para a comunidade de  $v_j$ 
      fim para cada
      se o ganho máximo é positivo então
         $v_i$  é movido para a comunidade do vizinho onde o ganho foi máximo;
      senão
         $v_i$  permanece em sua comunidade;
      fim se
    fim para cada
  fim enquanto
  cria  $G_1$ , onde cada comunidade é um vértice;
  para cada  $v_i, v_j \in V(G_1)$  faça
     $e_{ij}$  recebe como peso a soma dos pesos dos links entre as duas comunidades originais;
  fim para cada
   $G \leftarrow G_1$ , retorna árvore hierárquica do grafo  $G$  indicando as comunidades detectadas;
```

Fonte: Aires e Nakamura (2017).

Ao subdividir o grafo em comunidades, o algoritmo de Louvain revela informações cruciais sobre a organização e o funcionamento da rede. A identificação de comunidades permite analisar a dinâmica de interação entre os nós, revelando influenciadores, mapeando o fluxo de informações e proporcionando uma compreensão profunda de como elementos se propagam. Essa técnica também possibilita a detecção de grupos com funções similares, inferindo papéis e características comuns entre os nós, com aplicações que vão da biologia de sistemas à ciência da informação, auxiliando na segmentação de usuários em redes sociais ou agrupamento de documentos por áreas de pesquisa.

Segundo Valverde-Rebaza e Lopes (2012) a partir da detecção das comunidades, pode-se realizar previsões para encontrar os vínculos faltantes ou futuros. No qual, a partir de um par de vértices desconexos o algoritmo deve mensurar a probabilidade de vínculo futuro entre eles. Nesse contexto, têm-se o método de Within-inter-cluster e o algoritmo de Jaccard Coefficient.

De acordo Valverde-Rebaza e Lopes (2012), o método Within-inter-cluster, necessita de uma execução prévia de algum método de detecção de comunidade na rede. Para realizar a estimativa de probabilidade de vínculo sobre dois vértices desconexos contidos na mesma comunidade (Within cluster), é feita a divisão do número total de vizinhos entre os vértices com o número de vizinhos em comum pertencentes a mesma comunidade, conforme determinado pela Equação 2.

$$P(\Lambda_{x,y}|x^{c_\alpha}, y^{c_\alpha}) = \frac{|\Lambda_{x,y}^W|}{|\Lambda_{x,y}|} \quad (2)$$

Na estimativa de probabilidade de vínculo entre vértices desconexos de diferentes comunidades (Inter cluster), a divisão é executada com a quantidade de vizinhos em comum, que contenham vínculos em mais de uma comunidade, pela quantidade total de vizinhos em comum, determinado pela Equação 3.

$$P(\Lambda_{x,y}|x^{c_\alpha}, y^{c_\beta}) = \frac{|\Lambda_{x,y}^{IC}|}{|\Lambda_{x,y}|} \quad (3)$$

O algoritmo do coeficiente de Jaccard, determina a probabilidade de um vértice aleatório pertencente a união dos conjuntos de vizinhos dos dois vértices estar na interseção deles (BRAGA, 2017), sendo representados pela fórmula apresentada na Equação 4.

$$JC(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \quad (4)$$

O coeficiente de Jaccard também resolve uma deficiência da métrica Vizinhos em Comum, em alguns casos a quantidade de vizinhos em comum pode ser grande, porém não muito significativo em relação as similaridades (CAVALCANTE, 2020).

2.4 TRABALHOS CORRELATOS

Nesta seção, são apresentados estudos que possuem objetivos e funcionalidades análogos aos do presente trabalho. O Quadro 2 aborda a detecção da interação entre medicamentos em uma aplicação de redes complexas (WESTARB, 2021). No Quadro 3 é apresentado um estudo sobre a representação de interações bioquímicas na forma de um grafo efetivo (GATES *et al.*, 2021). Por fim, no Quadro 4 aborda um modelo para correlação de sintomas em pacientes com esteatose hepática, comparando diagnósticos da medicina moderna com diagnósticos da medicina tradicional chinesa através de redes complexas (WANG *et al.* 2019).

Quadro 2 – Aplicação de redes complexas para verificação de interação medicamentosa

Referência	Westrab (2021)
Objetivos	Desenvolver uma aplicação que utilize redes complexas para identificar e analisar interações medicamentosas
Principais funcionalidades	Um modelo que utiliza de uma rede complexa para relacionar as interações entre diferentes fármacos.
Ferramentas de desenvolvimento	O autor desenvolveu o modelo através de dados de fármacos obtidos do DrugBank e utilizou-se da biblioteca NetworkX para confecção das redes de proteínas alvo, proteínas carreadoras, proteínas de transporte e de enzimas. Os algoritmos Louvain e Girvan-Newman foram utilizados para subdividir a rede em comunidades que compartilhassem de características similares.
Resultados e conclusões	O autor obteve resultados satisfatórios com a identificação de 371 interações validadas pelo algoritmo Girvan-Newman e 58 pelo algoritmo Louvain.

Fonte: elaborado pelo autor.

Quadro 3 – The effective graph reveals redundancy, canalization, and control pathways in biochemical regulation

Referência	Gates <i>et al.</i> (2021)
Objetivos	Desenvolver modelo de rede complexa que fosse similar à modelos biológicos reais para estudar o comportamento desses modelos.
Principais funcionalidades	Foi desenvolvido um modelo que consegue analisar a rede de modo a encontrar os caminhos e variáveis mais importantes na escala geral, que ditassem o comportamento da rede.
Ferramentas de desenvolvimento	O autor desenvolveu o modelo a partir de um grafo direcionado onde os nós eram autômatos booleanos que proporcionava uma visão probabilística das dinâmicas da rede. Em seguida foi calculada a redundância das entradas dos autômatos através do algoritmo de Quine-McCluskey e de uma segunda equação auxiliar que indicava a conectividade da rede. Por fim, utilizou-se o coeficiente de Gini para analisar a inequidade das interações e os resultados foram comparados com modelos reais e aleatórios.
Resultados e conclusões	Os autores obtiveram bons resultados indicando que o grafo efetivo é uma ferramenta poderosa para estudar sistemas biológicos complexos, permitindo uma caracterização probabilística das interações causais e dinâmicas em forma de grafo, nas quais poucas variáveis são mais importantes e ditam o comportamento geral da rede.

Fonte: elaborado pelo autor.

Quadro 4 – Bridging fatty liver disease and traditional chinese medicine: A complex network approach

Referência	Wang <i>et al.</i> (2019)
Objetivos	Relacionar através de redes complexas a medicina tradicional chinesa (MTC) e a medicina moderna (MM) no contexto de esteatose hepática (HE).
Principais funcionalidades	Os autores compararam os sintomas da MTC e MM, criando redes com os sintomas observados nas análises físicas dos pacientes, simplificando os registros a uma variável binária.
Ferramentas de desenvolvimento	Os autores construíram duas redes tendo como nós os sintomas obtidos das análises da MTC e MM, onde uma rede era voltada à pacientes com EH e outra à pacientes saudáveis de controle. As métricas usadas para caracterizar e comparar as redes Gp e Gc foram: grau médio, força, tamanho médio dos caminhos, coeficiente de <i>clustering</i> e assortatividade.
Resultados e conclusões	Os autores contataram que sete sintomas da MM eram ligados a 18 sintomas diferentes da MTC o que indica que são sintomas cruciais para refletir o estado de saúde do paciente, embora ainda faltassem conexões diretas entre os dois grupos de sintomas. Ainda assim, analisando a topologia e métricas da rede os autores concluíram que embora os resultados da MTC fossem mais baixos que os da MM, a MTC ainda se mostrava razoavelmente confiável.

Fonte: elaborado pelo autor.

3 DESCRIÇÃO DO MODELO

Este capítulo aborda os aspectos fundamentais relacionados à criação do modelo computacional para correlação entre compostos fitoterápicos e medicamentos químicos, valendo-se de redes complexas para a modelagem da base e algoritmos de comunidades para análise das relações encontradas. A partir deste objetivo, foram definidos os seguintes Requisitos Funcionais (RF), no Quadro 5, e Requisitos Não Funcionais (RNF), no Quadro 6, da aplicação desenvolvida.

Quadro 5 – Requisitos funcionais do modelo

Requisitos Funcionais
RF01 – O modelo deve relacionar os fármacos e fitoquímicos através de suas substâncias
RF02 – O modelo deve ilustrar a forma e dinâmica da rede
RF03 – O modelo deve indicar medidas de densidade, grau médio, coesão e estruturas de comunidades para possibilitar análise e estabelecer relações sobre a rede

Fonte: elaborado pelo autor.

Quadro 6 – Requisitos não funcionais do modelo

Requisitos Não Funcionais
RNF01 – O modelo deve relacionar os fármacos e medicamentos fitoterápicos através de redes complexas
RNF02 – O modelo deve ser desenvolvido em Python
RNF03 – A base de dados utilizada pelo modelo deve ser estruturada em um arquivo JSON
RNF04 – A base deve ser modelada de modo a facilitar a leitura e formação da rede complexa

Fonte: elaborado pelo autor.

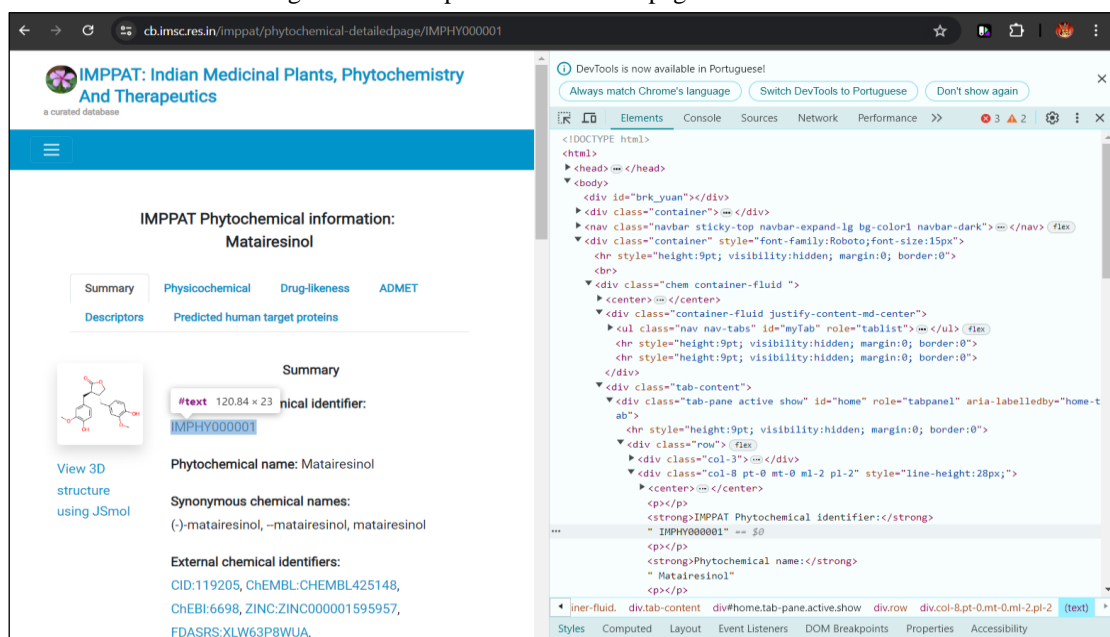
A execução deste trabalho baseou-se na análise de comunidades sobre redes complexas, buscando associar plantas fitoterápicas e fármacos químicos. Para tal, definiu-se um processo em cinco etapas, sendo elas: (i) obtenção dos dados, (ii) filtragem e processamento das informações dos compostos químicos e das partes de plantas, (iii) modelagem da rede, (iv) detecção das comunidades e, (v) a análise das associações, conforme pode ser visualizado na Figura 3.

O diagrama ilustra o processo de análise de redes sociais em cinco etapas sequenciais:

- Obtenção dos dados:** Representado por um ícone de um banco de dados (cilindro).
- Limpeza e organização dos dados:** Representado por ícones de engrenagens e um funil.
- Modelagem da rede:** Representado por um ícone de uma rede simples com dois nós e uma aresta.
- Deteção de comunidades:** Representado por um ícone de uma rede com múltiplos nós e arestas, formando uma estrutura de comunidade.
- Análise das associações:** Representado por um ícone de uma rede com um nó central e múltiplos nós periféricos, com um símbolo de barras de gráfico no centro.

Os dados dos medicamentos utilizados para montagem das redes foram obtidos através do DrugBank, disponíveis em (<https://go.drugbank.com/>), e os dados de plantas por sua vez foram obtidos através do banco de dados Indian Medicinal Plants, Phytochemistry And Therapeutics (IMPPAT), em (<https://cb.imsc.res.in/imppat/>). Para a seleção da base de dados utilizada na obtenção de informações sobre as plantas, foram pesquisadas diversas bases de diferentes regiões do mundo, cada qual com informações específicas. O IMPPAT foi selecionado por conter informações sobre partes específicas das plantas relacionadas a diferentes substâncias, além da possibilidade de integrar seus dados com as informações disponíveis em outras fontes.

Figura 4 – Exemplo de estrutura da página do IMPPAT



8

Figura 5 – Exemplo de página do IMPPAT ressaltando o código do composto na URL e o campo CID procurado.

Fonte: elaborado pelo autor.

O link associado ao CID direciona para a página do PubChem correspondente ao composto em questão. Nessa página, encontram-se informações sobre a substância, como estrutura química, identificadores de outras plataformas, propriedades químicas, entre outras. Dentre as informações disponíveis, obteve-se o Unique Ingredient Identifier (UNII), destacado na Figura 6. O UNII é um dos identificadores presentes na base de dados do DrugBank e foi utilizado para diferenciar as substâncias.

Figura 6 – Exemplo de página do Pubchem, na seção onde foi encontrado o identificador UNII

Fonte: elaborado pelo autor.

Nesta etapa, optou-se pela biblioteca Selenium para a obtenção de dados da página do PubChem. A escolha justifica-se pelo carregamento dinâmico do conteúdo da página, o que inviabilizava a utilização exclusiva da biblioteca BeautifulSoup. A Figura 7 exibe um trecho do código em que o webdriver, configurado para uso da biblioteca Selenium, realiza uma requisição à página. O objetivo é buscar o texto "UNII" dentro de um caminho específico na árvore de tags da seção com o ID "UNII", utilizando os métodos `get` e `find_element`.

Figura 7 – Extrato do código que exibe o uso da biblioteca Selenium

```
driver.get(unii_link) # Ex: https://pubchem.ncbi.nlm.nih.gov/compound/119205
time.sleep(1)
unii_elem = driver.find_element(By.CSS_SELECTOR, "section[id='UNII']>div[class='py-3']>div[class='break-words']>a")
print(unii_elem.text)
```

Fonte: elaborado pelo autor.

Nesta etapa do *web scrapping*, foram coletados os nomes dos compostos, juntamente com seus respectivos códigos IMPHY e UNII. É importante destacar que muitos compostos presentes no IMPPAT não possuíam identificadores UNII associados. Portanto, foram considerados apenas os casos em que havia correspondência entre o fitoquímico e um

UNII na base de dados do PubChem. Os registros obtidos foram armazenados em um arquivo JSON, contendo uma lista de objetos com as informações coletadas, conforme ilustrado na Figura 8.

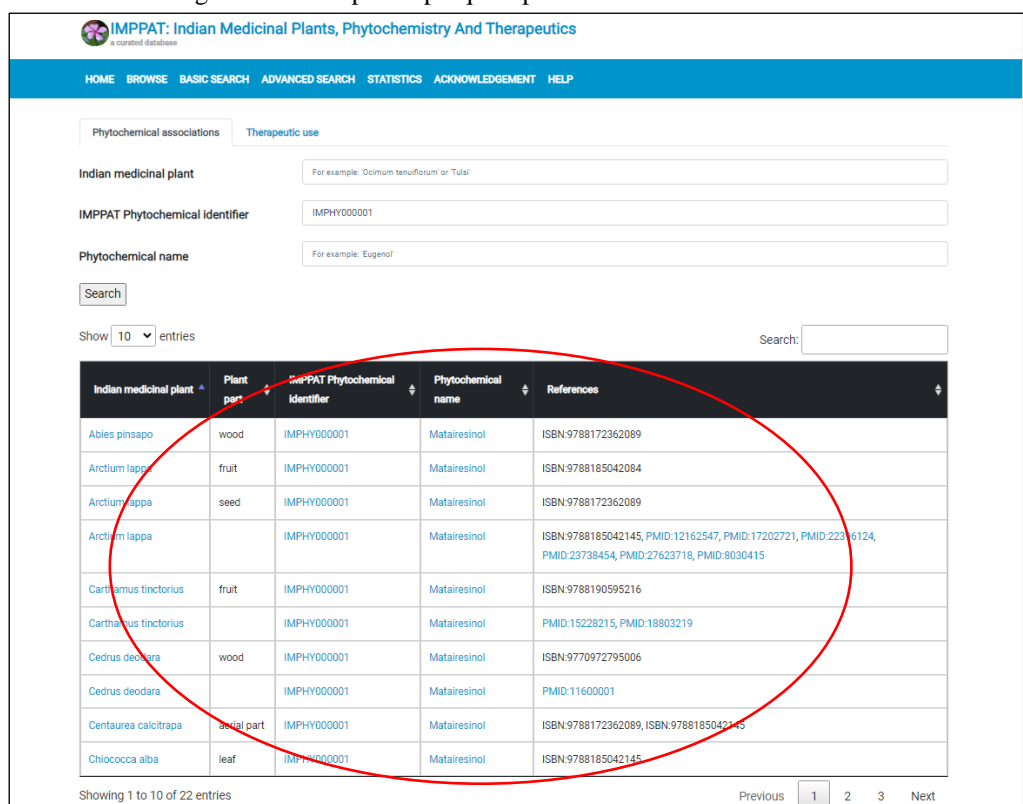
Figura 8 – Exemplo de resultado da primeira rodada de *web scrapping*

```
{
  "IMPPAT Link": "https://cb.imsc.res.in/imppat/phytochemical-detailedpage/IMPHY000001",
  "IMPPAT Identifier": "IMPHY000001",
  "Phytochemical Name": "Matairesinol",
  "Synon. Chemical Names": "(-)-matairesinol, --matairesinol, matairesinol",
  "CID": "119205",
  "UNII Link": "https://pubchem.ncbi.nlm.nih.gov/compound/119205",
  "UNII": "XLW63P8WUA"
},
```

Fonte: elaborado pelo autor.

A segunda etapa de *web scrapping* teve como objetivo buscar no IMPPAT as partes das plantas associadas a cada um dos compostos químicos. Para isso, utilizou-se a página de busca do IMPPAT, pesquisando pelos códigos IMPHY obtidos na primeira etapa. Os dados, apresentados em formato de tabela (Figura 9), forneceram o nome da planta e a parte da planta que continha a substância em questão. É importante ressaltar que, em alguns casos, a parte da planta não estava especificada, havendo apenas o registro da associação entre a planta e a substância. Nesses casos, o registro foi salvo com a informação da parte da planta em branco. Assim como na etapa anterior, os resultados foram armazenados em uma lista no formato JSON, conforme ilustrado na Figura 10.

Figura 9 – Exemplo de pesquisa por IMPHY na base do IMPPAT



Indian medicinal plant	Plant part	IMPPAT Phytochemical Identifier	Phytochemical name	References
Abies pinsapo	wood	IMPHY000001	Matairesinol	ISBN:9788172362089
Arctium lappa	fruit	IMPHY000001	Matairesinol	ISBN:9788185042084
Arctium lappa	seed	IMPHY000001	Matairesinol	ISBN:9788172362089
Arctium lappa		IMPHY000001	Matairesinol	ISBN:9788185042145, PMID:12162547, PMID:17202721, PMID:2236124, PMID:23738454, PMID:27623718, PMID:8030415
Carthamus tinctorius	fruit	IMPHY000001	Matairesinol	ISBN:9788190595216
Carthamus tinctorius		IMPHY000001	Matairesinol	PMID:15228215, PMID:18803219
Cedrus deodara	wood	IMPHY000001	Matairesinol	ISBN:9770972795006
Cedrus deodara		IMPHY000001	Matairesinol	PMID:11600001
Centaurea calcitrapa	seed part	IMPHY000001	Matairesinol	ISBN:9788172362089, ISBN:9788185042145
Chiococca alba	leaf	IMPHY000001	Matairesinol	ISBN:9788185042145

Fonte: elaborado pelo autor.

Figura 10 – Exemplo de resultado da segunda etapa de *web scrapping*

```
{
  "Plant name": "Abies pinsapo",
  "Plant part": "wood",
  "IMPPAT ID": "IMPHY000001"
},
```

Fonte: elaborado pelo autor.

Os dados do DrugBank foram obtidos diretamente do site da empresa por meio de uma licença acadêmica. Os dados estavam disponíveis em um arquivo XML, contendo informações detalhadas sobre uma vasta gama de medicamentos, incluindo nome, informações de venda e distribuição (como nome comercial e laboratório),

identificadores, referências acadêmicas, interações com outros medicamentos da base, entre outras. Para este projeto, extraiu-se de cada medicamento o nome e o UNII.

Após a coleta dos dados de todas as bases, procedeu-se à união e tabulação. Os dados foram então reestruturados e filtrados para utilizar apenas aqueles relacionados a plantas associadas a compostos fitoterápicos que possuíam equivalência nos dados de compostos químicos, e vice-versa. No total, 583 substâncias foram identificadas, relacionadas a 2256 plantas diferentes. Essa combinação de dados possibilitou a construção da rede, relacionando os compostos presentes nas partes das plantas com os compostos encontrados em medicamentos por meio do UNII. A estrutura final da base de dados foi armazenada em um arquivo JSON, com os dados organizados por planta. Cada entrada no arquivo contém uma lista das partes da planta encontradas e a respectiva substância presente em cada parte, conforme exemplificado na Figura 11.

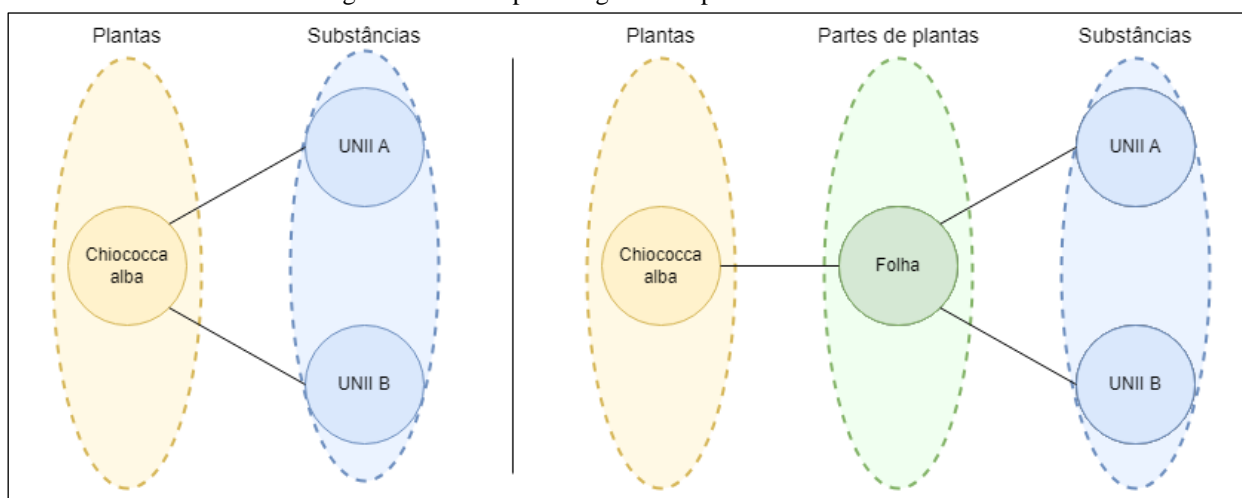
Figura 11 – Exemplo de entrada da tabulação da base de dados

```
{
  "Plant": "Chiococca alba",
  "Parts": [
    {
      "Part": "leaf",
      "Substance": {
        "Drug": "Matairesinol",
        "UNII": "XLW63P8WUA",
        "IMPHY": "IMPHY000001"
      }
    },
    {
      "Part": "leaf",
      "Substance": {
        "Drug": "Mannitol",
        "UNII": "3OWL53L36A",
        "IMPHY": "IMPHY011729"
      }
    }
  ]
},
```

Fonte: elaborado pelo autor.

A partir dos dados obtidos, foram criadas duas redes. A primeira consiste em um grafo bipartido com um grupo de plantas e outro de compostos químicos (à esquerda na Figura 12). A segunda rede relaciona três grandes grupos: um de plantas, um de partes de plantas e outro para os compostos (à direita na Figura 12). Ambos os modelos podem ser visualizados no diagrama da Figura 12.

Figura 12 – Exemplos de grafos da planta *Chiococca alba*

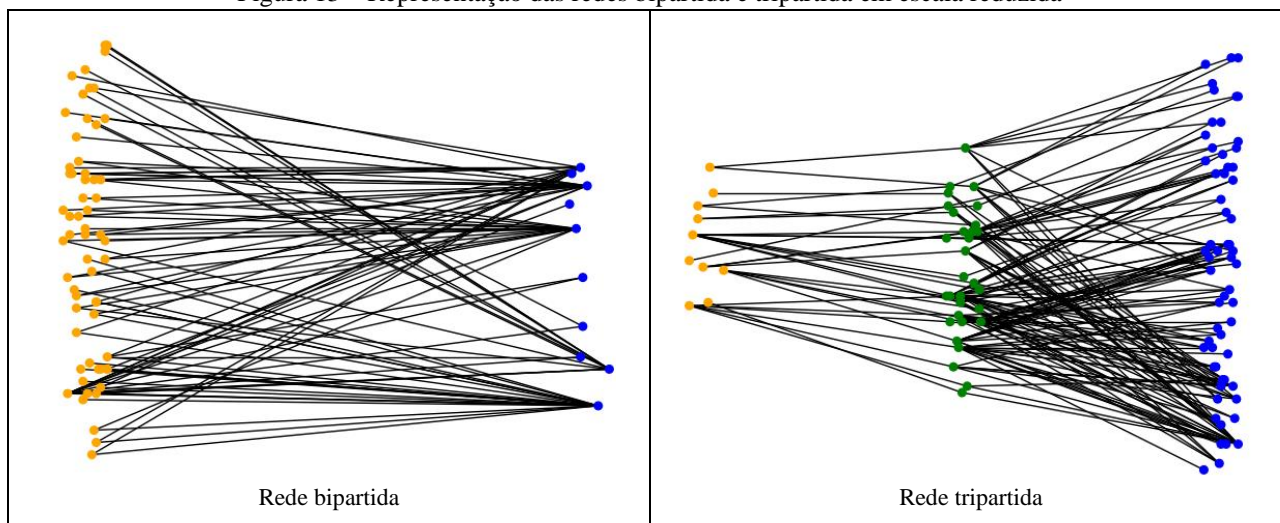


Fonte: elaborado pelo autor.

Para a elaboração das redes, utilizou-se a biblioteca NetworkX na linguagem Python. As redes foram construídas utilizando os dados coletados na primeira etapa do projeto, seguindo o modelo descrito na Figura 12. A rede bipartida foi composta por um total de 2839 nós, dos quais 2256 representavam plantas e 583 representavam substâncias. A rede que incluía a camada adicional para as partes de plantas contava com um total de 8040 nós, somando-se os nós da rede anterior a 5201 novos nós representando partes de plantas. As redes bipartida e tripartida podem ser visualizadas na Figura 13,

porém em uma escala de quantidade de nós menor para facilitar a visualização. Na Figura 13, os nós laranja representam os compostos químicos enquanto os azuis representam as plantas. No caso da rede tripartida existem também os nós verdes, que simbolizam as partes de planta.

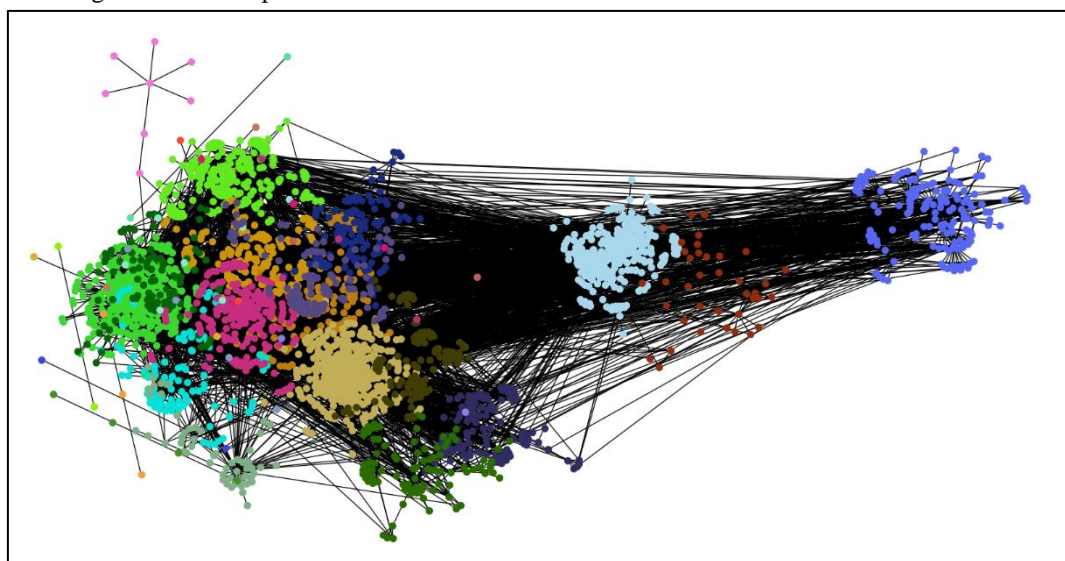
Figura 13 – Representação das redes bipartida e tripartida em escala reduzida



Fonte: elaborado pelo autor.

Com as redes modeladas, utilizou-se o algoritmo de Louvain para subdividir a rede em comunidades.

Figura 14 – Exemplo de divisão em comunidades na rede através do método de Louvain



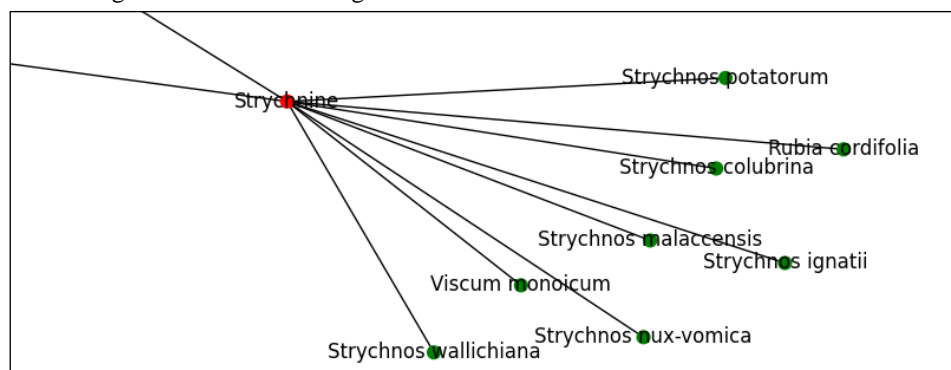
Fonte: elaborado pelo autor.

A análise da estrutura das comunidades geradas pelos algoritmos revelou que na maioria dos casos, as comunidades são organizadas de plantas que pertencem a um mesmo gênero, mas de espécies diferentes. Nesses casos, é comum que várias plantas do mesmo gênero compartilhem de uma mesma substância ou grupo de substâncias. Isso indica que nesses casos mesmo plantas de espécies diferentes dentro de um mesmo gênero podem ser usadas para obtenção de uma mesma substância na fabricação de um medicamento, dada a sua similaridade, sugerindo que a produção de compostos bioativos específicos com potencial medicinal pode ser uma característica conservada dentro de certos gêneros de plantas.

Um exemplo disso pode ser ilustrado na Figura 15 e na Figura 16, através de um subgrafo de uma das comunidades geradas na execução do algoritmo Louvain. Nesse caso, a comunidade é formada por várias plantas de gêneros diferentes que se conectam a um grupo específico de substâncias. Na imagem, é possível notar que várias plantas do gênero *Solanum* e *Strychnos* estão mais fortemente ligadas a uma substância que outras plantas, como é o caso de *Solamargine* e *Solasonine* no gênero *Solanum* e *Strychnine* no gênero *Strychnos*.

[illegible]

Figura 16 – Parte do subgrafo de uma comunidade do método de Louvain.



Essa similaridade intragenérica é especialmente importante do ponto de vista prático. A identificação de uma substância bioativa de interesse medicinal em uma espécie pode indicar que outras espécies do mesmo gênero também a produzem, diversificando as fontes de matéria-prima e abrindo caminho para a seleção de espécies com maior rendimento, cultivo mais fácil ou menor risco de extinção.

A presença de comunidades atreladas por plantas ou substâncias que conectam vários grupos reflete a diversidade e a interconexão das informações biológicas e químicas presentes nesses grupos, indicando que esses grupos podem ter sido de alguma forma relacionados na medicina tradicional indiana, seja por similaridade biológica das plantas ou por usos terapêuticos similares. Existem também os casos de comunidades isoladas, formadas apenas por uma planta e a substância a qual ela se liga. Um exemplo disso pode ser observado em uma comunidade formada apenas pela planta *Faradaya splendida* e pelo composto *Carbromal*. A relação de comunidade exclusiva entre esses dois compostos pode significar que essa é uma associação específica entre esses dois elementos, incomum a outras plantas da rede. Isso é sustentado pelo fato de que a substância só aparece ligada a essa planta em ambas as redes, e o contrário também é verdadeiro. Essa ligação única entre a planta e a substância pode indicar também uma associação terapêutica específica. De acordo com a própria base do DrugBank, a substância é classificada como um hipnótico e sedativo, sendo um psicléptico com efeito calmante, que pode revelar seus possíveis usos como fitoterápico. Embora não se tenha encontrado referências diretas ao uso como fitoterápico, essa planta já foi usada pela população aborígene para paralisar peixes a fim

de coletá-los (FREEMAN, 2017). Essa ligação fortalece o argumento de forte relacionamento entre a relação dessa substância com a planta para uso fitoterápico.

A partir da execução do algoritmo de Louvain, gerou-se também algumas métricas. As métricas analisadas em cada rede foram: densidade, grau médio dos nós e coesão. As médias das métricas alcançadas em cada algoritmo podem ser visualizadas no Quadro 7, juntamente com a quantidade de comunidades encontradas.

Quadro 7 – Métricas obtidas com a execução dos algoritmos

Rede	Densidade	Grau médio	Coesão	Comunidades
Bipartida	0.374068	2.401099	0.374068	32
Tripartida	0.191643	2.545486	0.191643	49

Fonte: elaborado pelo autor.

A aplicação do algoritmo de Louvain, nas redes bipartidas e tripartidas construídas a partir dos dados de plantas, compostos químicos e suas relações, permitiu a identificação de comunidades e a análise de métricas importantes para a compreensão da estrutura e organização dessas redes. O Quadro 7 apresenta as médias das métricas de densidade, grau médio dos nós e coesão para cada rede, juntamente com a quantidade de comunidades encontradas.

Na rede bipartida, a densidade de 0.374068 indica que aproximadamente 37% das conexões possíveis entre plantas e compostos químicos realmente existem. Essa densidade relativamente alta sugere uma interação significativa entre esses dois grupos, com diversas plantas compartilhando compostos químicos em comum. O grau médio dos nós de 2.401099 indica que, em média, cada nó na rede (planta ou composto) está conectado a cerca de 2 a 3 outros nós. A coesão, que coincide com a densidade nesse caso, reforça a ideia de uma estrutura relativamente conectada na rede bipartida, com 32 comunidades distintas formadas.

Na rede tripartida, a inclusão das partes das plantas levou a uma densidade menor (0.191643), o que era esperado devido à maior complexidade da rede. Essa densidade indica que cerca de 19% das conexões possíveis entre plantas, partes das plantas e compostos químicos estão presentes. O grau médio dos nós, ligeiramente superior ao da rede bipartida (2.545486), sugere que a inclusão das partes das plantas aumentou a conectividade média da rede. A coesão, também coincidindo com a densidade, aponta para uma estrutura menos coesa em comparação à rede bipartida, o que é natural considerando a maior quantidade de nós e a especificidade das relações. Nessa rede, foram identificadas 49 comunidades.

A análise comparativa das métricas entre as redes bipartida e tripartida demonstra que a inclusão das partes das plantas, apesar de diminuir a densidade geral, aumenta a complexidade e a conectividade da rede, resultando em um número maior de comunidades. Além disso, também se buscou também gerar métricas sobre as ligações dentro dos clusters e entre diferentes clusters, além do coeficiente de Jaccard das redes bipartida e tripartida. Os resultados podem ser observados no Quadro 8.

Quadro 8 – Ligações internas e entre comunidades e coeficiente de Jaccard

Rede	Ligações dentro das comunidades	Ligações entre comunidades	Coeficiente de Jaccard
Bipartida	6188	4279	0.111070
Tripartida	14543	5297	0.130837

Fonte: elaborado pelo autor.

A análise das redes bipartida e tripartida revela que a rede tripartida apresenta uma quantidade significativamente maior de arestas dentro dos clusters, indicando uma maior densidade de conexões internas entre os nós pertencentes às mesmas comunidades, o que é esperado dada a camada adicional de partes de plantas presentes nesse contexto. Além disso, embora a rede tripartida tenha também mais arestas entre clusters em comparação à rede bipartida, a diferença é menos acentuada do que a observada nas arestas internas. Esse padrão indica que a estrutura da rede tripartida é mais modular, com uma predominância de conexões internas indicando que as comunidades na rede tripartida são mais bem definidas e possuem uma maior interação interna. Não obstante a isso, a menor diferença proporcional entre as arestas dentro e entre clusters na rede bipartida pode sugerir uma conectividade mais dispersa e menos modular. Isso pode significar que essa é uma rede onde as comunidades são menos isoladas e mais interconectadas se comparadas com a tripartida. Por sua vez, o coeficiente de Jaccard mede a similaridade entre os conjuntos de vizinhos dos nós conectados, indicando a probabilidade de dois nós conectados compartilharem vizinhos comuns. Um coeficiente mais alto na rede tripartida implica que há mais sobreposição nos conjuntos de vizinhos, indicando uma maior coesão local dentro das comunidades.

Analisando-se o tamanho e formação das comunidades obteve-se também novas perspectivas quanto as dinâmicas dos elementos na rede. Nas comunidades de tamanho médio, entre cerca de 90 e 200 nós, observou-se que normalmente a comunidade é fortemente ligada a alguns nós centrais que servem como “hubs” no centro da comunidade, conectando os outros nós. Em uma das comunidades, por exemplo, observou-se que os compostos *Cianidanol* e *Epicatechin* estavam no centro da comunidade, se ligando a diversas plantas que por sua vez ligavam a novos compostos criando seus próprios subgrupos na comunidade. Ainda assim, nota-se que ambos os medicamentos apresentam alvos similares na metabolização, atuando no intestino. Isso mostra que para essas comunidades é normal que os agrupamentos

sejam entre plantas e compostos que atuam em uma região específica do corpo, ou que possam ser utilizadas em um tratamento de um grupo de enfermidades similar.

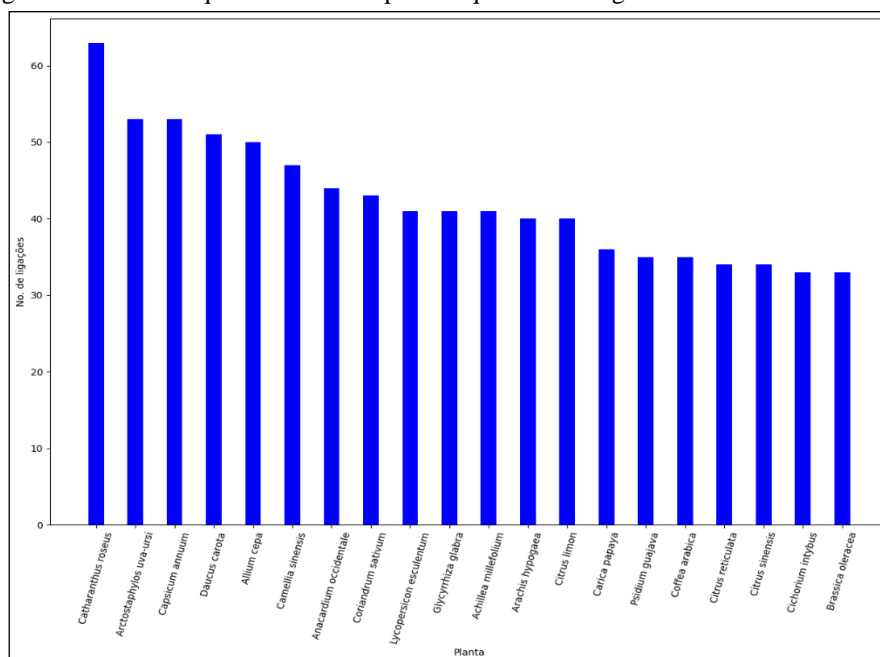
Em comunidades grandes, notou-se que esse comportamento ainda se sustenta. A grande quantidade de nós fortemente conectados indica a presença de áreas de convergência terapêutica, onde uma grande quantidade de plantas e compostos podem ser utilizados para o tratamento de um mesmo grupo de sintomas relacionados. Nessas mesmas comunidades, porém, também em virtude do seu tamanho massivo, podem existir casos no qual os nós de ramificações diferentes possam ser usados para práticas diferentes, embora ainda estejam ligados a um mesmo grupo de elementos mais fortemente conectados.

Em comunidades pequenas, com pouquíssimos nós, e em comunidades isoladas, a análise revela um comportamento mais peculiar. A presença de poucas ou somente uma planta relacionada ao composto em uma comunidade isolada pode significar que a planta é rara ou de difícil obtenção, ou que o uso terapêutico da substância seja mais específico. Embora alguns dos medicamentos nessas redes atuem em áreas que outros medicamentos na rede também atuem, existem casos mais específicos, como o do psicoléptico *Carbromal* apresentado anteriormente.

Com relação as famílias de plantas que formam as redes, não se observaram casos em que um gênero específico dominasse uma comunidade toda, embora isso aconteça em alguns subgrupos. Um exemplo disso pode ser visto no caso dos gêneros *Solanum* e *Strychnos* apresentados previamente, onde embora as plantas estejam fortemente ligadas a um composto específico, sua dinâmica ainda seguia a do resto da comunidade, sem muita expressão na dinâmica geral exceto pela concentração em volta de um composto específico.

Além da dinâmica de comunidades, outras análises foram realizadas na rede. A primeira buscou identificar a planta com maior potencial para a produção de medicamentos, utilizando a rede bipartida. A lógica empregada considerou que cada conexão entre uma planta e uma substância diferente representava um fármaco potencialmente produzido a partir daquela planta. A análise da quantidade de arestas de cada planta na rede revelou que a *Catharanthus roseus* seria a mais promissora para a produção de medicamentos, como ilustra a Figura 17.

Figura 17 – Gráfico que mostra as 20 plantas que mais se ligam a substâncias diferentes.



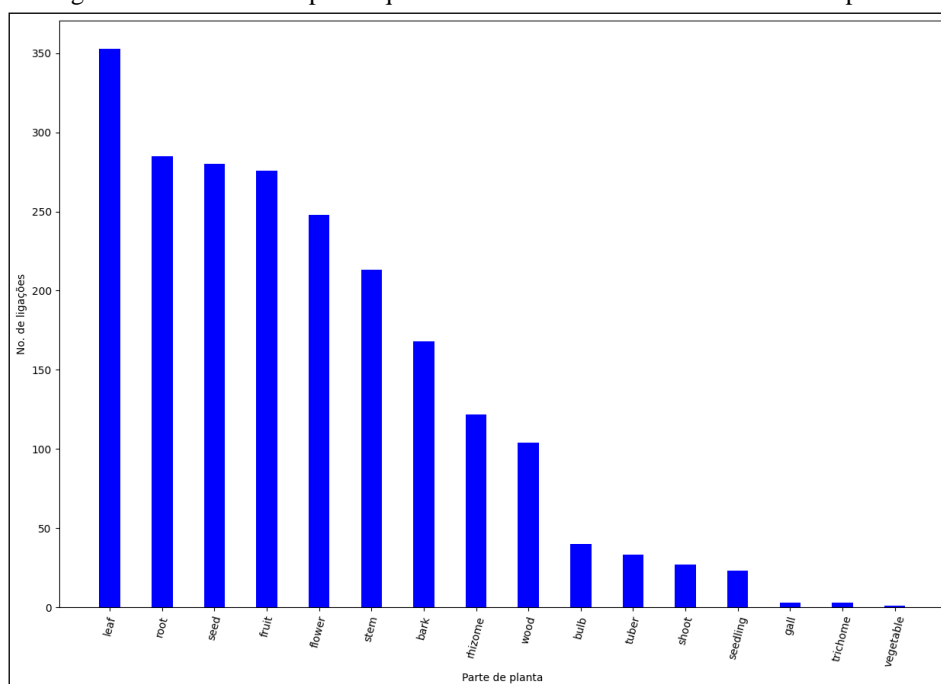
Fonte: elaborado pelo autor.

A partir da Figura 17, pode-se observar que o grande número de ligações dessas plantas sugere que elas possam ser plantas promissoras para a utilização fitoterápica, podendo usar de várias substâncias com princípios similares para o tratamento de algum sintoma, ou até para tratamentos diferentes, dependendo do contexto de uso de cada substância que constituem suas partes. Além disso, nota-se que todas as plantas citadas também são fortes candidatas para produção de medicamentos, seja como fonte de várias substâncias ou como uma fonte com alta concentração de uma substância específica em suas partes. Ainda assim, ressalta-se que o número de ligações na rede não necessariamente dita a eficácia de uso dessas plantas, dado que isso também pode ser influenciado por outros fatores como facilidade de obtenção e bioatividade.

Seguindo a mesma lógica, buscou-se identificar qual parte da planta era a fonte mais comum de substâncias bioativas. Para isso, analisou-se no gráfico quais partes se conectavam a diferentes compostos, agrupando-as por tipo,

como "flor", por exemplo. Os resultados revelaram que as folhas são as fontes mais comuns, seguidas pelas raízes e sementes. Casos em que a parte da planta não estava especificada nos registros foram desconsiderados. A Figura 18 apresenta o gráfico gerado por essa análise.

Figura 18 – Gráfico das partes que mais se relacionam com diferentes compostos



Fonte: elaborado pelo autor.

Esses resultados sugerem que a fonte primária de extração de substâncias com princípios ativos seja para uso fitoterápico ou extração para a indústria farmacêutica também venha das folhas das plantas, considerando o escopo das plantas definidas na base. Ainda assim, novamente é importante se fazer a ressalva de que dados como a bioatividade, concentração da substância na planta e facilidade de acesso ou cultivo das plantas não estão sendo considerados nesse caso.

4 CONCLUSÕES

O uso de fitoterápicos como tratamento representa uma alternativa importante no combate a doenças, incentivada pelo Ministério da Saúde por oferecer uma abordagem mais natural em comparação aos medicamentos sintéticos. Apesar do incentivo, a prática ainda não está amplamente difundida entre profissionais de saúde, que geralmente optam por fármacos convencionais. A partir deste contexto, este trabalho buscou correlacionar plantas medicinais a medicamentos por meio da análise de redes complexas. Foram utilizados dados provenientes da DrugBank, uma plataforma com vasta base de dados sobre medicamentos, e do IMPPAT, que contém informações sobre plantas medicinais tradicionais da fitoterapia indiana.

Para o desenvolvimento deste trabalho optou-se pela linguagem Python, utilizando a biblioteca NetworkX para modelagem e análise das redes. O algoritmo de Louvain, crucial para a detecção de comunidades, apresentou resultados satisfatórios, identificando 32 e 49 comunidades nas redes bipartida e tripartida, respectivamente. Observou-se que as comunidades se estruturavam, majoritariamente, em torno de uma substância ou grupo de substâncias associadas a plantas do mesmo gênero, formando "famílias" de plantas ou grupos de "famílias". Em outros casos, também se percebeu que comunidades isoladas podem evidenciar fortes ligações entre a planta e o composto associado, aumentando a confiabilidade das relações.

Em relação aos trabalhos correlatos, este se diferencia do de Gates *et al.* (2021) em proposta e escopo. Apesar de ambos empregarem a análise de redes complexas, o contexto de Gates *et al.* (2021) difere significativamente ao abordar domínios biológicos, além de utilizar metodologia distinta. A pesquisa se assemelha mais à de Wang *et al.* (2019), que também utiliza redes complexas, mas com o objetivo de correlacionar sintomas da medicina tradicional chinesa aos da medicina moderna, adotando metodologia diferente. Entretanto, o trabalho que apresenta maior similaridade é o de Westarb (2021), pois ambos realizam análises de comunidades em seus respectivos grafos.

Por fim, pode-se concluir que a proposta de conectar plantas e medicamentos por meio do UNII foi realizada com sucesso. A partir dessa conexão, a investigação aprofundou-se na análise da estrutura e do comportamento da rede complexa formada, revelando importantes detalhes sobre seus elementos e comunidades. A metodologia empregada, combinando análise de redes complexas e estudo de comunidades, demonstrou grande potencial para descobertas e

inovações nas áreas de fitoterapia e farmacologia, expondo padrões específicos e abrindo novas perspectivas de pesquisa. Os resultados obtidos, além de robustos, fornecem informações relevantes sobre a relação entre plantas e compostos químicos.

Como sugestões para trabalhos futuros, propõe-se: (i) utilizar outros identificadores para ampliar as associações entre medicamentos e plantas medicinais; (ii) incluir bases de dados adicionais, expandindo o escopo para além da fitoterapia indiana e abrangendo outras regiões e suas plantas típicas; (iii) incluir relacionamento com alguma base de dados de plantas que possibilitem obter informações sobre o uso das plantas em medicinas tradicionais, a fim de aprofundar as análises entre o efeito da substância no organismo e o uso na medicina tradicional; (iv) desenvolver uma interface gráfica para facilitar a visualização e análise da rede.

REFERÊNCIAS

- AIRES, Victoria P. S.; NAKAMURA, Fabiola G. Detecção de Comunidades em Redes Sociais: Relacionando o Método Louvain a Medidas de Centralidade. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 37., 2017, São Paulo. Anais do XXXVII Congresso da Sociedade Brasileira de Computação. Porto Alegre: Sbc, 2017. p. 2442-2451. Disponível em: <<https://sol.sbc.org.br/index.php/ctic/article/view/3236/3198>>. Acesso em 21 jun. 2024.
- BARABASI, Albert-Laszlo. *et al.* Evolution of the social network of scientific collaborations. **Physica A: Statistical mechanics and its applications**, Elsevier, v. 311, n. 3, p. 590–614, 2002.
- BLONDEL, Vincent D. *et al.* **Fast unfolding of communities in large networks**. Journal Of Statistical Mechanics: Theory And Experiment: Iop Publishing Ltd, v. 2008, out. 2008.
- BRAGA, Camila Souto da Cunha Brendel. **Predição de links em uma rede heterogênea baseada em dados geolocalizados**. 2017. 38 f. TCC (Graduação) - Curso de Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco, Recife, 2017. Disponível em: <https://www.cin.ufpe.br/~tg/2017-2/cscbb-tg>. Acesso em: 22 jun. 2024.
- CAVALCANTE, Argus Antonio Barbosa. **Predição de links em redes sociais: resgatando informações contextuais e topológicas no histórico de evolução do grafo**. 2020. 92 f. Dissertação (Mestrado) - Curso de Sistemas e Computação, Instituto Militar de Engenharia, Rio de Janeiro, 2020. Disponível em: https://bdex.eb.mil.br/jspui/bitstream/123456789/9042/1/dissertacao_argus_cavalcante_vfinal.pdf. Acesso em: 24 jun. 2024.
- CARVALHO, Marco Antonio M. **BCC204 - Teoria dos grafos**. Ouro Preto, [2020]. Disponível em: <http://www.decom.ufop.br/marco/site_media/uploads/bcc204/04_aula_04.pdf>. Acesso em: 15 jun. 2024.
- COSTA, Polyana P. **Teoria de grafos e suas aplicações**. 2011. 77 f. Dissertação (Mestrado Profissional em Matemática Universitária) – Curso de Pós-Graduação em Profissional em Matemática, Instituto de Geociências e Ciências Exatas da Universidade Estadual Paulista, Rio Claro.
- FEOFILOFF, Paulo; KOHAYAKAWA, Yoshirau; WAKABAYASHI, Yoshiko. **Uma Introdução Sucinta à Teoria dos Grafos**. São Paulo, [2011]. Disponível em: <<https://www.ime.usp.br/~pf/teoriadosgrafos/texto/TeoriaDosGrafos.pdf>>. Acesso em: 17 jun. 2024.
- FOOD AND DRUG ADMINISTRATION. **FDA's Global Substance Registration System**. [S.l.], [2024?]. Disponível em: <<https://www.fda.gov/industry/fda-data-standards-advisory-board/fdas-global-substance-registration-system>>. Acesso em: 18 junho 2024.
- FREEMAN, Jenny. **Faradaya splendida**. South Brisbane, [2017]. Disponível em: <https://www.slq.qld.gov.au/blog/faradaya-splendida>. Acesso em: 22 jun. 2024.
- GATES, Alexander J. *et al.* The effective graph reveals redundancy, canalization, and control pathways in biochemical regulation and signaling. **Proceedings of the national academy of sciences**, v. 118, n. 12, 2021.
- HUANG, Jialiang *et al.* **Systematic Prediction of Pharmacodynamic Drug-Drug Interactions through Protein-Protein-Interaction Network**. Plos Computational Biology, [S.l.], v. 9, n. 3, p. 1-9, 21 mar. 2013. Public Library of Science (PLOS). Disponível em: <<http://dx.doi.org/10.1371/journal.pcbi.1002998>>. Acesso em: 18 junho 2024.
- JIA, Caiyan *et al.* Node attribute-enhanced community detection in complex networks. **Scientific Reports**, v. 7, 2017.
- MATA, Angélica S. da. Complex Networks: a Mini-review. **Brazilian Journal of Physics**. v. 50, n. 5, p. 658–672, 2020.
- MELO, Gildson Soares de. **Introdução à Teoria dos Grafos**. 2014. 35 f. Dissertação (Mestrado Profissional em Matemática) - Universidade Federal do Paráíba, João Pessoa, 2014.

MINISTÉRIO DA SAÚDE. **Política nacional de plantas medicinais e fitoterápicos**. 1. ed. Brasília, 2006. 60 p. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/politica_nacional_fitoterapicos.pdf. Acesso em: 21 jun. 2024.

_____. **Política nacional de práticas integrativas e complementares no SUS**. 2. ed. Brasília, 2015. 95 p. Disponível em:

https://bvsms.saude.gov.br/bvs/publicacoes/politica_nacional_praticas_integrativas_complementares_2ed.pdf. Acesso em: 28 set. 2023.

_____. **Práticas integrativas e complementares: Plantas medicinais e fitoterapia na atenção básica**. 1. ed. Brasília, 2012. 154 p. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/praticas_integrativas_complementares_plantas_medicinais_cab31.pdf. Acesso em: 25 set. 2023.

_____. **Uso racional de medicamentos**. [S.l.], [2020?]. Disponível em: <<https://www.gov.br/saude/pt-br/composicao/sectics/daf/uso-racional-de-medicamentos>>. Acesso em: 18 junho 2024.

NAHAL, Jessica. **Introdução a redes complexas**. [2014]. Disponível em: <https://blog.dp6.com.br/introdução-a-redes-complexas-df73b623d67f>. Acesso em: 18 junho 2024.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. **OPAS/OMS apresenta iniciativa para reduzir erros de medicação**. [S.l.], [2021]. Disponível em: <<https://www.paho.org/pt/noticias/18-8-2021-opasoms-apresenta-iniciativa-para-reduzir-erros-medicacao>>. Acesso em: 18 junho 2024.

OSHIRO, Mariana C. *et al.* A evolução do registro e prescrição de fitoterápicos no Brasil sob a perspectiva legal e sanitária. **Revista Visa em Debate: Ciência, sociedade e tecnologia**. v. 4, n. 4, p. 116–122, 2016.

PAVLOPOULOS, Georgios A. *et al.* Bipartite graphs in systems biology and medicine: a survey of methods and applications. **GigaScience**, v. 7, n. 4, 2018.

PINTO, Eduardo R. **Estudo da dinâmica de epidemias em redes complexas**. 2018. 89f. Dissertação (Mestrado em Biometria) - Curso de Pós-Graduação em Biometria, Universidade Estadual Paulista, Botucatu.

RABUSKE, Márcia A. **Introdução à teoria dos grafos**. 1 ed. Florianópolis: Ed. da UFSC, 1992.

RIBEIRO, Larissa F. **Redes sem Escala Típica: Visão Geral, Modelos Alternativos e Técnicas Computacionais**. 2017. 93f. Dissertação (Mestrado em Física) - Curso de Pós-Graduação em Física, Universidade Federal do Rio Grande do Norte, Natal.

RITA, Luís. **Louvain Algorithm**. [2020]. Disponível em: <https://towardsdatascience.com/louvain-algorithm-93fde589f58c>. Acessado em: 22 jun. 2024.

SANTOS, R. L. *et al.* Análise sobre a fitoterapia como prática integrativa no Sistema Único de Saúde. **Revista Brasileira de Plantas Medicinais**, Maringá, v. 13, n. 4, p. 486-491, 2011.

SCIENCE DIRECT. **Subgraphs**. [S.l.], [2021?]. Disponível em: <https://www.sciencedirect.com/topics/computer-science/subgraphs>. Acesso em: 23 jun. 2024.

VALVERDE-REBAZA, Jorge Carlos; LOPES, Alneu de Andrade. **Link Prediction in Complex Networks Based on Cluster Information**. 2012. 101 f. TCC (Graduação) - Curso de Ciências da Computação, Ciências da Computação, Universidade de São Paulo, São Carlos, 2012.

WANG, Wenjing *et al.* Bridging fatty liver disease and traditional chinese medicine: A complex network approach. **2019 IEEE International Symposium on Circuits and Systems (ISCAS)**, Sapporo, Japão, p. 1-5 2019.

WESTARB, Gustavo. **Aplicação de redes complexas para verificação de interação medicamentosa**. 2021. 16 f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Centro de Ciências Exatas e Naturais, Universidade Regional de Blumenau, Blumenau.

WORLD HEALTH ORGANIZATION. **Biodiversity and Health**. [S.l.], [2015]. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/biodiversity-and-health>>. Acesso em: 23 jun. 2024.