

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Victor Barcelos Melazo

**Explorando redes complexas: a importância da
análise visual perante medidas estatísticas**

Uberlândia, Brasil

2021

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Victor Barcelos Melazo

**Explorando redes complexas: a importância da análise
visual perante medidas estatísticas**

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Orientador: Prof. Bruno Augusto Nassif Travençolo

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2021

Victor Barcelos Melazo

Explorando redes complexas: a importância da análise visual perante medidas estatísticas

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Trabalho aprovado. Uberlândia, Brasil, 01 de novembro de 2016:

**Prof. Bruno Augusto Nassif
Travençolo
Orientador**

Professor

Professor

Uberlândia, Brasil
2021

Resumo

O estudo de redes complexas apresenta diferentes estratégias de análise. Neste trabalho, a representação visual e o cálculo de medidas estatísticas de redes complexas distintas são abordados. O objetivo é explicitar a importância de combinar essas duas análises, uma vez que, padrões e comportamentos da rede podem não ser facilmente identificados somente com os resultados dos cálculos das medidas estatísticas. Cinco redes complexas diferentes foram utilizadas para exemplificar essa proposta. Suas representações visuais são apresentadas com o auxílio do software DyNetVis, e suas medidas estatísticas calculadas com a biblioteca NetworkX. Por fim, os fatores que indicam essa importância são colocados em destaque para demonstrar que essa combinação de estratégias pode ser mais explorada.

Palavras-chave: Redes complexas, representação visual, medidas estatísticas.

Lista de ilustrações

Figura 1 – Ilustração de Redes Complexas	10
Figura 2 – Ilustração de comunidades	10
Figura 3 – Exemplificação de Lista com ordenação diferente	12
Figura 4 – Ilustração de Matriz de Adjacência	13
Figura 5 – Ilustração da construção do Diagrama de Nós e Arestas de forma cumulativa durante o tempo	14
Figura 6 – Ilustração do diagrama de Nós e Arestas considerando as arestas ativas em cada instante de tempo	14
Figura 7 – Ilustração do Massive Sequence View	15
Figura 8 – Ilustração da propriedade Closeness	16
Figura 9 – Ilustração da propriedade <i>Betweenness Centrality</i>	17
Figura 10 – Ilustração das redes selecionadas para análise	22

Lista de tabelas

Tabela 1 – Dados sobre redes	21
Tabela 2 – Resultados das medidas estatísticas	22
Tabela 3 – Maiores valores da centralidade <i>Betweenness</i>	24

Sumário

1	INTRODUÇÃO	7
1.1	Objetivos	8
1.2	Organização da monografia	8
2	FUNDAMENTAÇÃO TEÓRICA E REVISÃO BIBLIOGRÁFICA	9
2.1	Redes Complexas	9
2.1.1	Comunidades	10
2.1.2	Redes Temporais	11
2.2	Visualização de Redes	11
2.2.1	Visualização Matricial	12
2.2.2	Diagrama de Nós e Arestas	13
2.2.3	<i>Massive Sequence View - MSV</i>	14
2.3	Medidas Estatísticas	15
2.3.1	Grau Médio	15
2.3.2	<i>Closeness Centrality</i>	15
2.3.3	<i>Betweenness Centrality</i>	16
2.3.4	Modularidade	17
2.4	Trabalhos correlatos	17
3	DESENVOLVIMENTO	20
3.1	Seleção de redes	20
3.2	Cálculo das medidas estatísticas	21
3.3	Comparação	23
4	CONCLUSÃO E TRABALHOS FUTUROS	25
4.1	Trabalhos Futuros	25
	REFERÊNCIAS	26

1 Introdução

Redes complexas são comumente utilizadas para modelar dados que representam interações entre elementos de um domínio. Tais elementos, que podem representar itens, como regiões, times de futebol ou hospitais de uma cidade, são representados por nós na rede e as interações são representadas por arestas que conectam os nós (ALBERT; BARABÁSI, 2002).

A rede temporal é definida pela inclusão do tempo na sua análise e representação. Essa característica é responsável por facilitar a percepção de determinados padrões temporais e anomalias, e também é responsável por diversos estudos sobre qual a melhor forma de representar redes que evoluem ao longo do tempo, as quais são comuns em cenários reais, como em redes sociais (LINHARES et al., 2019).

O estudo de redes é uma área de pesquisa extremamente ampla com diversas estratégias de análise. Neste contexto, as redes serão expostas a formas de visualização diferentes e análises estatísticas. Essas duas técnicas são escolhidas para que a comparação e avaliação de redes encontrem padrões e comportamentos importantes, os quais poderiam passar despercebidos.

O uso de medidas estatísticas é comum para análise de redes. Entretanto, fazer análises apenas matemáticas pode não ser suficiente, já que detalhes importantes podem passar despercebidos. Em muitas situações, resultados numéricos não são facilmente interpretados e padrões e comportamentos da rede podem não ser notados (GRANDJEAN; JACOMY, 2019).

As diferentes formas de visualização da informação aplicadas na análise dessas redes é benéfica neste contexto, pois nos disponibiliza ferramentas gráficas e até mesmo interativas capazes de incluir o usuário no processo de exploração dos dados. Isso torna mais fácil, rápido e confiável a procura e localização de padrões, anomalias, tendências e outros comportamentos da rede (WARE, 2013).

Os softwares de visualização capazes de gerar representações claras e objetivas ganharam grande relevância, visto que é possível mostrar as interações feitas em ordem cronológica, a quantidade de relações iniciadas entre nós, organizar os nós de inúmeras maneiras para melhorar a visualização do conjunto, agrupar em blocos atividades repetitivas, detectar comunidades, alterar a resolução de tempo estabelecida, entre outras possibilidades (LINHARES et al., 2019).

1.1 Objetivos

O objetivo deste trabalho é avaliar a complementariedade da análise de redes complexas de forma estatística e visual. Tem-se como hipótese que a análise visual como ferramenta de exploração e comparação de redes é relevante em cenários em que determinadas medidas estatísticas possuem valores semelhantes para redes com características diferentes. Os objetivos específicos deste trabalho são:

- Buscar por bases de redes de diferentes domínios e aplicar medidas estatísticas comumente utilizadas na análise de redes complexas;
- Avaliar se técnicas de visualização de redes complexas identificam padrões e comportamentos não observados pela análise estatística com as medidas avaliadas.

1.2 Organização da monografia

Adiante, a fundamentação teórica e a revisão bibliográfica serão apresentadas no Capítulo 2, parte do trabalho destinada a definição de conceitos relevantes para avaliação dos objetivos e a apresentação de trabalhos existentes na área de redes e visualização da informação. No Capítulo 3 a forma como este trabalho foi desenvolvido é documentada, as redes utilizadas para comparação são apresentadas, assim como suas representações visuais e os resultados dos cálculos de suas medidas estatísticas. Por fim, no Capítulo 4, o trabalho é concluído após uma ponderação sobre os dados obtidos e apresentados, levando em consideração os objetivos iniciais.

2 Fundamentação Teórica e Revisão Bibliográfica

Neste capítulo conceitos importantes para melhor entendimento do trabalho serão apresentados. Na Seção 2.1 a definição de redes e suas principais características serão introduzidas, para que na Seção 2.2 as formas mais comuns de visualização de redes sejam apresentadas e definidas. A Seção 2.3 será sobre as medidas estatísticas comumente utilizadas na análise de redes e suas descrições e a Seção 2.4 trará os trabalhos correlatos da área de visualização, comparação e análise de redes, os quais representam a importância do desenvolvimento deste trabalho.

2.1 Redes Complexas

De forma simplificada, uma rede complexa é um conjunto de informações que registra a interação entre entidades diferentes (ALBERT; BARABÁSI, 2002). Ao interpretar essa base de dados podemos reconhecer padrões de comportamento, fenômenos relevantes e características desconhecidas, por exemplo.

As redes podem ser aplicadas para representar sistemas do mundo real em qualquer situação em que exista uma relação ou interação entre dois ou mais objetos distintos. Por isso, esses objetos podem ilustrar, por exemplo, pessoas, animais, carros, computadores e restaurantes. O objetivo dessa representação é ser capaz de analisar um domínio específico de dados e gerar conhecimentos a partir dele.

Uma rede complexa é definida como um grafo $G(V, E)$, sendo V o conjunto de vértices (ou nós), que representam os objetos do domínio sendo estudado, e E o conjunto de tuplas (nx, ny) , com $nx \in V$ e $ny \in V$ (ZAIDI; MUELDER; SALLABERRY, 2014). Cada tupla refere-se a uma aresta, isto é, a uma interação existente entre dois objetos. Aplicações de redes complexas podem ser encontradas, por exemplo, na malha rodoviária brasileira, em que as cidades são tratadas como os nós do grafo e as rodovias as arestas, ou nas citações entre artigos científicos, em que uma aresta entre dois nós (artigos) representa uma citação.

As arestas em uma rede complexa podem ser direcionadas ou não-direcionadas, conforme ilustra a Figura 1. Em uma rede de contato definida pela proximidade entre pessoas, por exemplo, temos que, se A está próximo a B então B também está próximo a A (rede não-direcionada). Por outro lado, em uma rede de citações, A citar B não quer dizer, necessariamente, que B também citou A (rede direcionada). Em redes complexas

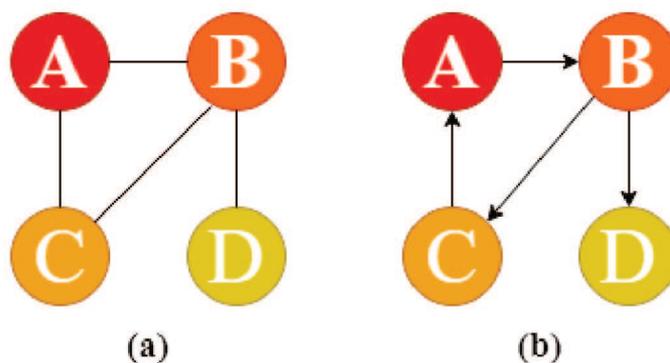


Figura 1 – Representação visual de uma rede complexa. (a) Rede não-direcionada. (b) Rede direcionada.

do mundo real, importantes propriedades também podem ser percebidas. É o caso, por exemplo, da formação de comunidades, assunto que será melhor explorado a seguir.

2.1.1 Comunidades

A análise de redes complexas pode ser uma tarefa difícil devido a quantidade de informações armazenadas. Em redes com muitos nós ou de muitas arestas, certas características e padrões certamente passam despercebidos. Por isso, a detecção de comunidades pode ser vantajosa, pois ajuda a manifestar os padrões despercebidos.

As comunidades são detectadas levando em consideração características semelhantes entre nós, agrupando-os para facilitar a análise das redes complexas, como ilustrado na Figura 2. Nessa imagem, os nós semelhantes são visualmente agrupados e recebem uma cor para caracterizar eles. O processo de detecção das comunidades pode considerar diferentes fatores, por exemplo, agrupar nós que mais se comunicam ou então agrupar nós que compartilham propriedades comuns, como cidades de um mesmo estado ou artigos de uma mesma área de conhecimento (FORTUNATO, 2010).

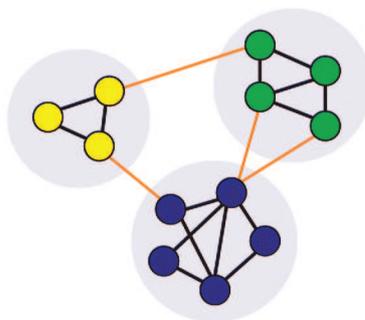


Figura 2 – Ilustração de comunidades em uma rede. Figura retirada de (LINHARES et al., 2020).

A detecção de comunidades em redes complexas é feita ao escolher algoritmos para

executar essa tarefa, visto que é possível criar diferentes métodos de seleção e caracterização de comunidades. Assim, diferentes algoritmos detectam diferentes comunidades dentro de uma mesma rede complexa. Portanto, a eficiência desses algoritmos precisa ser avaliada de acordo com a rede em análise.

Algoritmos como o Louvain (BLONDEL et al., 2008) e o Infomap (ROSVALL; BERGSTROM, 2008) são exemplos de métodos eficientes recomendados para detectar comunidades (LINHARES et al., 2020).

2.1.2 Redes Temporais

As redes temporais compreendem um tipo de rede complexa que tem como principal característica a introdução do tempo como fator complementar para a indicação de interações feitas entre objetos distintos. Em outras palavras, uma aresta conectando dois nós A e B que existe no instante de tempo 0 pode ou não persistir no instante de tempo 1.

Uma rede temporal é definida como um grafo $G(V, E)$, sendo V o conjunto de vértices (ou nós), que representam os objetos do domínio sendo estudado, e E o conjunto de tuplas (nx, ny, tk) , com $nx, ny \in V$ e tk representando o instante de tempo em que a interação entre nx e ny existe (LINHARES et al., 2019). Aplicações de redes temporais podem ser encontradas, por exemplo, em redes de transporte aéreo (ROCHA, 2009) e na comunicação entre pessoas em ambientes controlados (LINHARES et al., 2017), entre outras.

2.2 Visualização de Redes

A visualização de redes existe dentro da área de pesquisa Visualização da Informação (VI), a qual busca e desenvolve métodos para representações visuais. A forma como a informação é representada tem forte impacto na compreensão e interpretação de domínios abstratos diversos, já que facilita a análise de grandes bases de dados, o reconhecimento de padrões, o descobrimento de problemas com o tipo de dados coletados para estudo, entre outros (WARE, 2013).

Visualização de Redes é um campo de estudo extenso e em crescimento, em função da sua vasta área de aplicação e da quantidade de informações relevantes que são deduzidas a partir de sua aplicação eficaz. Assim como existem diferentes tipos de redes, existem diferentes tipos de representação de redes, as quais se encaixam melhor com determinadas características de cada domínio (PONCIANO et al., 2021).

A capacidade de interpretação humana pode ser muito limitada caso os dados apresentados não estejam organizados de forma intuitiva, principalmente, quando redes

grandes são manipuladas. A Figura 3(a) mostra uma lista de interações existentes entre nós de uma rede complexa, e é possível perceber que mesmo em uma lista pequena o comportamento entre as letras do alfabeto é difícil de ser notado, ou seja, formas de visualização eficientes são imprescindíveis para o desenvolvimento dessa área de estudo. A Figura 3(b) mostra uma ordenação alternativa, na qual os nós que aparecem mais de uma vez são agrupados o que facilita a visualização de quantas arestas incidem sobre cada nó.

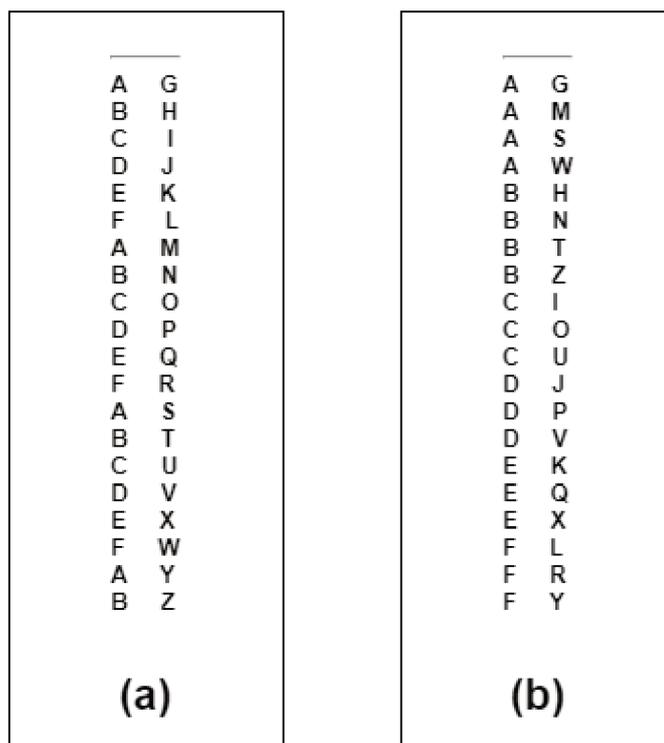


Figura 3 – Exemplificação de listas para a análise de redes complexas. Em (a) os nós foram colocados na lista de acordo com sua ordem de criação e em (b) temos uma lista ordenada de forma que os nós mais recorrentes são agrupados.

Embora manipulações na lista possam otimizar a análise um pouco, a análise ainda pode ser bastante trabalhosa, especialmente no caso de redes grandes em termos de quantidade de nós e arestas. Diferentes representações visuais podem então ser adotadas visando um estudo mais eficaz da rede. A seguir, três estratégias serão representadas.

2.2.1 Visualização Matricial

A visualização matricial utiliza matrizes de adjacência para representar as redes complexas, as quais possuem propriedades importantes para análise e comparação, por exemplo, a possibilidade de visualizar todos os nós e suas incidências de uma só vez. Desse modo, as matrizes de adjacência são quadradas e, quando as redes são não direcionadas, simétricas. Logo, com a existência de uma ligação entre dois nós as células que a representa são coloridas, como na Figura 4 em que tanto a interação AB quanto a BA estão coloridas.

	A	B	C	D
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figura 4 – Ilustração de uma matriz de adjacência.

A visualização matricial é comumente utilizada para estimar a densidade de uma rede, ou seja, conhecer o número de arestas com relação ao número de nós. Também é usada para comparar redes diferentes e para reconhecer comunidades (BEHRISCH et al., 2016).

É possível utilizar animações na representação das matrizes para analisar redes temporais, de forma que cada frame produzido pela animação mostra um instante de tempo colorindo apenas as arestas ativas no momento. Entretanto, a utilização da animação pode dificultar a visualização da rede temporal caso ela disponha os frames dos instantes de tempo de forma rápida, ou caso a rede seja muito grande e a identificação e ordenação dos nós seja comprometida (BEHRISCH et al., 2016; LINHARES et al., 2017).

2.2.2 Diagrama de Nós e Arestas

O Diagrama de Nós e Arestas é uma forma de visualização de redes. Nessa visualização, como na Figura 1(a), os nós são representados por círculos e as arestas, que os conectam, por linhas retas.

A identificação de padrões visuais no diagrama é uma das grandes utilidades que essa forma de representação possui, em função de todos os nós e arestas estarem dispostos na tela. É possível gerar visualizações com informações sobre nós, por exemplo, quais deles interagem mais ou menos rede, ou então quais nós interagem mais entre si.

O processo de animação da visualização, utilizando diagramas de nós e arestas, para representar redes temporais também pode ser aplicado aqui, porém com os mesmos obstáculos que a animação matricial. Nesse caso, pode ocorrer a sobreposição de arestas, já que o posicionamento dos nós e das arestas permitem essa situação, o que pode esconder importantes padrões e comportamentos para a análise da rede (LINHARES et al., 2021).

O processo de animação da rede temporal pode adotar estratégias para diminuir o impacto da sobreposição das arestas, como vemos na Figura 5 e na Figura 6, nas quais a forma como a rede é apresentada ao longo dos instantes de tempo é diferente. Na Figura 5 a medida que o tempo passa as arestas dos instantes anteriores permanecem ativas no

diagrama, já na Figura 6 esse acúmulo não acontece, somente as arestas ativas no instante de tempo especificado aparecem no diagrama (PONCIANO et al., 2021)

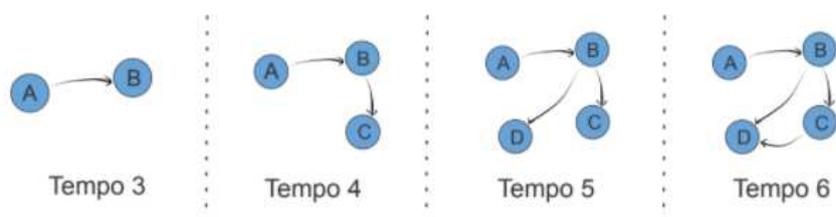


Figura 5 – Ilustração da construção do diagrama de nós e arestas de forma cumulativa durante o tempo. Figura retirada de (LINHARES, 2020)

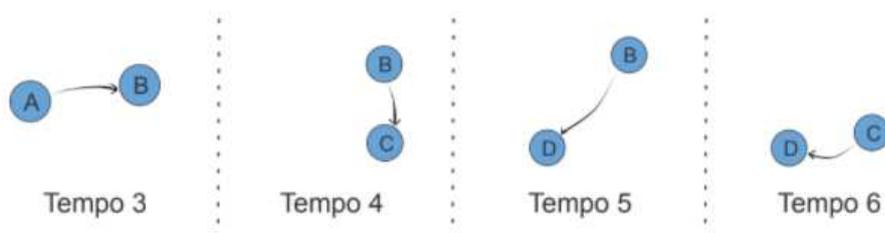


Figura 6 – Ilustração do diagrama de nós e arestas considerando as arestas ativas em cada instante de tempo. Figura retirada de (LINHARES, 2020)

Outras estratégias de representação no Diagrama de Nós e Arestas também podem ser adotadas para melhorar a visualização tanto de redes temporais animadas quanto de redes complexas. A coloração dos nós de acordo com características semelhantes ou de acordo com o número de interações feitas (PONCIANO et al., 2021), e a reordenação espacial dos nós são exemplos dessas estratégias que melhoram a visualização e a identificação de padrões de uma rede.

2.2.3 Massive Sequence View - MSV

A visualização de redes, estritamente, temporais baseada no *Massive Sequence View* (MSV) utiliza de artefatos visuais ilustrados na Figura 7, como a utilização do eixo Y para representar os nós, o eixo X para as unidades de tempo com a inserção das arestas de forma vertical (ELZEN et al., 2014). Entretanto, com a escalabilidade das redes certos problemas, como a sobreposição das arestas ou a quantidade de nós, torna a capacidade de extrair informações um grande obstáculo. Em função disso, propostas existentes na literatura buscam resolver esses problemas na visualização (LINHARES et al., 2019).

A identificação de comunidades pode ser muito útil em redes grandes, porém a visualização com o fator tempo em destaque exige outras formas de organização visual.

$$ClosenessCentrality(u) = \frac{1}{\sum_{v \in V} SP(u, v)} \quad (2.1)$$

Closeness Centrality é uma ótima medida para detectar nós capazes de espalhar informações de maneira mais eficiente pela rede. O valor do cálculo efetuado pode variar de 0 até 1, de forma que, caso o valor esteja próximo de 0, o nó em avaliação se encontra distante dos demais e, caso o valor esteja próximo de 1, significa que o nó em avaliação é mais próximo dos demais nós da rede. Logo, um valor baixo indica que mais arestas deverão ser utilizadas para ir de um nó a outro da rede e um valor mais alto implica que os caminhos entre u e qualquer outro nó sejam consideravelmente menores (GOLDSTEIN; VITEVITCH, 2017).

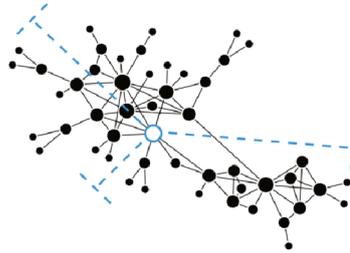


Figura 8 – Ilustração da propriedade *closeness* em uma rede. Figura retirada de (GRANDJEAN; JACOMY, 2019).

Na Figura 8 pode-se visualizar que o nó destacado se encontra em uma posição crucial dentro da rede por estar entre dois grupos de nós mais conectados entre si. Em função disso, o valor de *Closeness Centrality* do nó destacado seria elevado, já que se comunica com a maioria dos nós da rede com a utilização de um número menor de arestas.

2.3.3 *Betweenness Centrality*

A medida estatística *Betweenness Centrality* é definida pela contabilização da quantidade de vezes que um nó u aparece nos menores caminhos entre todo par possível de nós da rede. Na equação 2.2, SP_{sv} é o número de menores caminhos entre os nós s e v e $SP_{sv}(u)$ é o número de menores caminhos entre s e v que passam por u (FREEMAN, 1977).

$$BetweennessCentrality(u) = \sum_{s \neq u \neq v \in V} \frac{SP_{sv}(u)}{SP_{sv}} \quad (2.2)$$

Betweenness Centrality é recomendada para identificar nós que conectam comunidades. O valor do cálculo efetuado pode variar de 0 até 1 e como a medida estatística é calculada para cada nó, quanto maior seu valor maior a probabilidade do nó em análise agir como ponte entre diferentes partes da rede (GOLBECK, 2015).

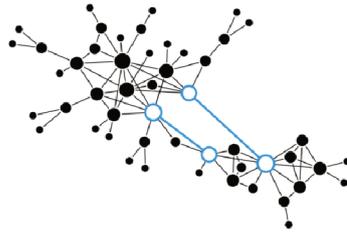


Figura 9 – Ilustração da propriedade *Betweenness Centrality* em um grafo. Figura retirada de (GRANDJEAN; JACOMY, 2019).

Na Figura 9 é possível analisar as pontes criadas entre os dois pares de nós destacados, as quais certamente possuem grande influência no valor da medida *Betweenness Centrality* de cada um desses nós, de maneira que seus valores são maiores que os demais.

2.3.4 Modularidade

A modularidade é uma medida usada, principalmente, na análise de comunidades dentro de uma rede. Ela avalia o nível de envolvimento dos nós dentro dessas partições, o qual pode variar de -1 até 1, indicando o pior e o melhor nível de envolvimento respectivamente.

$$\text{Modularidade} = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \quad (2.3)$$

Na Equação 2.3 temos que m é o número de arestas na rede e A_{ij} é o valor da matriz de adjacência A , a qual representa a existência de uma aresta entre os nós i e j com 1 caso exista e 0 caso não exista. A variável k_i é o número de arestas que incidem em i e $\delta(c_i, c_j) = 1$ se i e j são da mesma comunidade ou $\delta(c_i, c_j) = 0$ caso contrário. Valores encontrados entre 0,3 e 0,7 geralmente indicam uma boa divisão de comunidades (LINHARES et al., 2020).

2.4 Trabalhos correlatos

Esta seção é dedicada para discussão sobre trabalhos existentes na área de análise de redes e visualização da informação, assim como os pontos que inspiraram a construção deste trabalho.

De acordo com (BHADRA et al., 2009), o estudo de redes tem ganhado ainda mais importância por ser uma ferramenta fundamental no estudo de dados. As abordagens atuais já são capazes de quantificar diversos aspectos de uma rede, já possuindo técnicas de identificação de nós importantes e sendo capazes de comparar a topologia de redes

distintas. No artigo os autores comparam o comportamento de vespas com o de crianças em suas salas de aula e chegaram a conclusão, com o uso de medidas estatísticas básicas, que as redes formadas por esses dois grupos são diferentes. No artigo as redes são representadas de forma simples por diagramas de nós e arestas e a análise por meio da visualização é pouco explorada.

Em (RIDER et al., 2014) os autores enfatizam a importância das propriedades escolhidas para construção de redes e como suas interações serão tratadas. Redes de componentes biológicos, principalmente, expressões genéticas são as mais abordada pelo estudo. O peso das arestas é um exemplo de propriedade abordada no artigo. Neste caso a comparação da mesma rede, construída de diferentes formas, é feita para encontrar a forma mais vantajosa e identificar particularidades nos dados da rede. O artigo (VACCA, 2020) também chega a mesma conclusão comparando algumas redes de forma individual mas representadas de formas diferentes. Em ambos os artigos encontramos a utilização de medidas estatísticas, como *Closeness Centrality*, *Betweenness Centrality* e Modularidade, as quais são utilizadas para buscar a melhor forma de comparar redes.

Em (LINHARES et al., 2020) o uso combinado de medidas estatísticas, como modularidade, e um método de visualização de redes baseado em matriz é utilizado para encontrar a forma de detecção de comunidades mais apropriada em cenários específicos. Os algoritmos de detecção de comunidades utilizados foram o Louvain e o Infomap, e as comparações foram feitas sempre sobre uma mesma rede. Logo, apesar de apresentar um processo de avaliação de redes completo, o objetivo é identificar diferenças nos métodos de detecção de comunidades propostos e encontrar o melhor para um determinado cenário.

O estudo de diversas redes simultaneamente é abordado estatisticamente em (BOX-STEF.; CHRIS., 2015). Nele, a forma como organizações associativas entre indústrias e empresas interagem é apresentada e a como essa relação influencia por meio do espaço, tempo, tamanho e problemas, outras redes. O artigo traz como exemplo restrições de políticas organizacionais norte americanas de acordo com a área em que elas atuam e de acordo com a necessidade da coalizão feita entre organizações associativas que possuem interesses em comum.

Em (Miller et al., 2015) e (KOMOLAFE et al., 2019) processos para detecção de nós com comportamento anômalo dentro da rede são propostos para análise estatística de uma rede complexa.

Portanto, existem artigos na área de pesquisa de visualização da informação que exploram a comparação de redes ou de diferentes formas de representação visual para uma mesma rede. As pesquisas feitas em (BHADRA et al., 2009) e em (FAUST; SKVORETZ, 2002) debatem sobre a comparação de redes e chegam a conclusão de que por meio de certas metodologias conseguimos verificar a existência ou não de similaridades entre redes com componentes distintos, mas a influência da construção da representação visual das

redes pode ser ainda mais pesquisada.

3 Desenvolvimento

Neste capítulo a metodologia utilizada para o desenvolvimento do trabalho será apresentada e aplicada sobre redes selecionadas, as quais apresentarão diferenças na comparação de suas medidas estatísticas e de suas representações visuais. Para isso, somente redes complexas serão utilizadas e a forma de visualização escolhida foi de diagrama de nós e arestas.

A escolha de somente um tipo de rede e de visualização foi feita para manter a objetividade sobre uma avaliação entre duas estratégias de interpretação e estudo de redes que são bem amplos.

Na Seção 3.1 serão apresentados os meios pelos quais as bases de dados referentes as redes complexas foram encontrados, o software utilizado para representação visual e as definições das redes selecionadas para comparação. Na Seção 3.2 os resultados dos cálculos das medidas estatísticas de cada rede são introduzidos, bem como a biblioteca utilizada para obter esses dados. Na Seção 3.3 a interpretação das informações disponíveis é feita ao comparar as visualizações e as medidas estatísticas.

3.1 Seleção de redes

Existem diversos *websites* e repositórios disponíveis na internet que oferecem, de forma gratuita, bases de dados referentes a redes. Como o principal objetivo do trabalho é encontrar redes complexas com representações visuais diferentes e medidas estatísticas semelhantes, inicia-se a busca de exemplares para demonstrar a importância da representação visual na análise de redes.

A seleção das redes para análise partiu, primeiramente, pela discrepância entre a organização global dos nós da rede, ou seja, com o auxílio de um software capaz de representar as redes levando em consideração suas propriedades, foram selecionadas redes com padrões visuais diferentes. Para isso, o software DyNetVis (LINHARES et al., 2017), capaz de representar redes como diagramas de nós e arestas, matriz e *massive sequence view*, foi escolhido para visualizar a estrutura das redes complexas.

Neste trabalho, o software DyNetVis irá representar as estruturas das redes complexas com diagramas de nós e arestas, e a estratégia utilizada para posicionar os nós é direcionada pela força (LINHARES et al., 2017). Isso significa que os nós conectados entre si buscam se aproximar e se agrupar, o que pode tornar as características da rede esteticamente mais fáceis de serem descobertas.

Para o desenvolvimento deste trabalho foram selecionadas 5 redes complexas para

Tabela 1 – Dados sobre redes

	(a)	(b)	(c)	(d)	(e)
Nº de Nós	2,361	10,351	2,017	4,941	6.474
Nº de Arestas	7,181	20,934	2,929	6,593	12.571

demonstração, são elas:

1. Rede de interação entre proteínas durante o brotamento da levedura, em que cada nó representa uma proteína e cada aresta uma interação entre duas proteínas (BU et al., 2003).
2. Rede de colaboração entre autores para resolução de artigos sobre NCSTRL (*Network Computer Science Technical Reference Library*), em que cada nó representa um cientista e cada aresta é colocada quando houve uma coautoria entre eles (LATORA; NICOSIA; RUSSO, 2017).
3. Rede de interação entre proteínas de células de uma levedura madura, em que cada nó representa uma proteína e são conectadas caso elas interajam fisicamente dentro da célula (WATTS; STROGATZ, 1998).
4. Rede representativa da planta de energia dos Estados Unidos, na qual cada aresta representa a ligação via cabo entre usinas de energia, transformadores ou consumidores presentes na rede (WATTS; STROGATZ, 1998).
5. Rede representativa da internet no nível de sistemas autônomos, na qual cada nó representa um sistema autônomo e a existência de uma aresta significa que eles se comunicam (LATORA; NICOSIA; RUSSO, 2017).

Na Tabela 1, o número de nós e o número de arestas de cada uma das cinco redes são apresentados para também serem levados em consideração na análise dos resultados das medidas estatísticas.

Na Figura 10 é possível observar a discrepância da organização global dos nós em cada rede mesmo com a utilização do mesmo método de posicionamento dos nós na construção da representação visual.

3.2 Cálculo das medidas estatísticas

As medidas estatísticas foram calculadas utilizando o pacote NetworkX (HAGBERG; SWART; CHULT, 2008), desenvolvido na linguagem de programação python para criação, manipulação, e estudo de estruturas, dinâmicas e funções de redes complexas. A aplicação do pacote NetworkX em (LADD et al., 2017) sobre uma sociedade de amigos

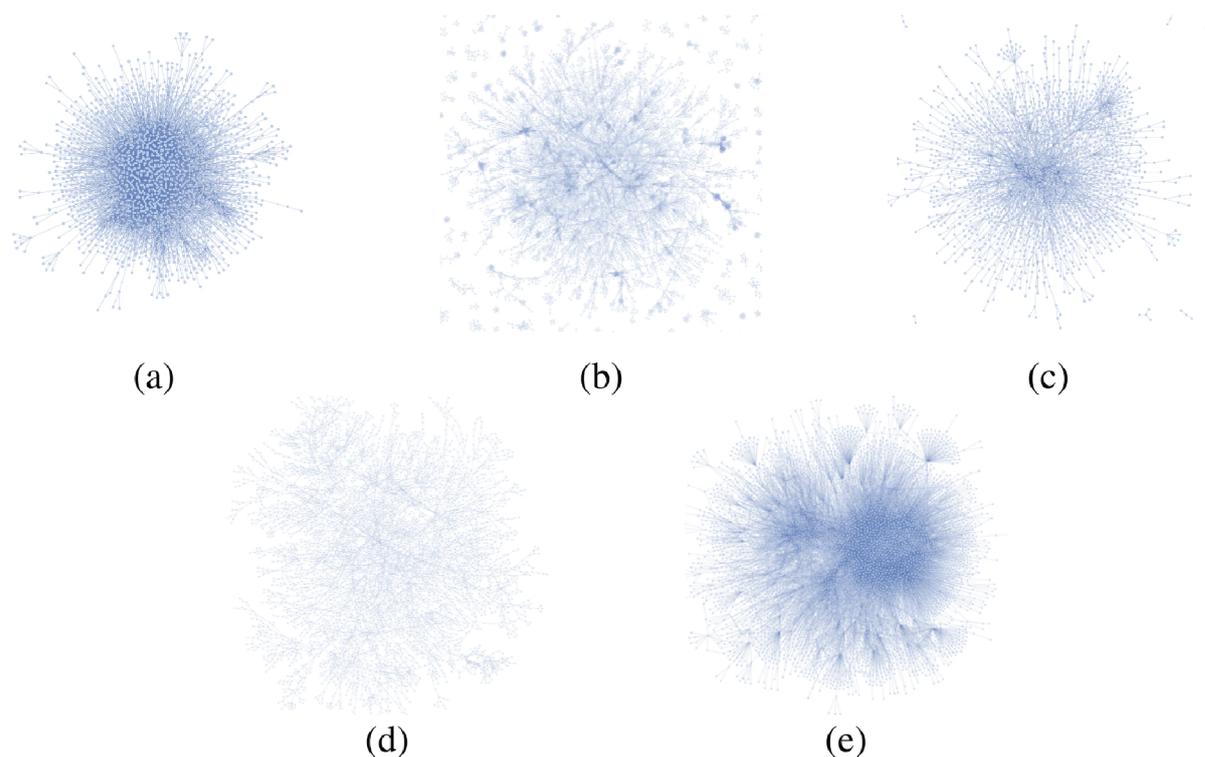


Figura 10 – Representação visual das redes. (a) Interação entre proteínas durante o brotamento da levedura. (b) Colaboração entre autores. (c) Interação entre proteínas de células de levedura. (d) Planta de energia nos Estados Unidos. (e) Internet no nível de sistemas autônomos. Figuras geradas utilizando o software DyNetVis

Tabela 2 – Resultados das medidas estatísticas

	(a)	(b)	(c)	(d)	(e)
Grau Médio	6,0830	4,0448	2,9043	2,6687	3,8835
Média <i>Closeness</i>	0,2699	0,0756	0,1938	0,0670	0,3410
Média <i>Betweenness</i>	0,0086	0,0333	0,0114	0,0300	0,0040
Modularidadeq	0,6026	0,95	0,7696	0,9338	0,6019

que se comunicavam na Europa, no século XVII foi usada como base para o desenvolvimento do código capaz de gerar relatórios sobre redes complexas envolvendo o cálculo de suas medidas estatísticas.

As funções prontas já disponibilizadas no pacote NetworkX para calcular, de forma trivial, o grau médio, as medidas de centralidade *Closeness* e *Betweenness* para cada nó, e a modularidade da rede, permite processar as cinco redes para comparar seus resultados e buscar comportamentos semelhantes em visualizações discrepantes.

Na Tabela 2 são apresentados os resultados dos cálculos das medidas estatísticas para cada rede e, com relação aos valores apresentados para as centralidades *Closeness* e *Betweenness*, os resultados foram ordenados e a média de 10% dos maiores resultados de cada rede foi calculada para comparação e análise.

3.3 Comparação

As redes apresentadas e definidas nas sessões anteriores apresentam semelhanças relacionadas as medidas estatísticas interessantes, as quais serão explicitadas a seguir.

Na Figura 10 nota-se que todas as cinco redes possuem um número grande de nós e arestas, porém cada uma possui peculiaridades que as tornam diferentes das demais, por exemplo, a visualização das redes (a), (c) e (e) são parecidas mas é visível que a concentração de nós no centro da rede (a) é maior do que na rede (c), como as duas representam interações entre proteínas de uma célula é possível interpretar que durante a fase de brotamento da levedura a atividade proteica celular é maior, já que a rede (a) tem o dobro de arestas com relação a rede (c).

Na rede (e), apesar da visualização com as redes (a) e (c) se assemelharem por uma concentração de nós, o maior núcleo da rede não está posicionado no centro da imagem, uma vez que, existem outros aglomerados de nós que também ocupam parte do espaço e esse espalhamento de nós no layout serve como uma pista visual para a existência de comunidades dentro da rede. Neste caso, como cada nó representa um sistema autônomo essas sugestões de comunidades pode indicar maior proximidade física ou então lógica entre os nós.

Ainda sobre Figura 10, as redes (b) e (d) possuem estruturas visuais explicitamente diferentes das demais, as duas são mais esparsas, ou seja, não encontramos grandes concentrações de nós, mas ainda assim se diferem entre si. A rede (b), além de ser uma rede que possui componentes conexos, que são grupos de nós conectados entre si mas sem nenhuma conexão com os demais nós da rede, apresenta pequenas concentrações de nós, isso pode representar cientistas de um mesmo laboratório ou universidade que naturalmente publicavam mais artigos juntos e no caso dos componentes conexos, eles podem representar pequenos grupos de pesquisa ou cientistas que não tomam essa linha de pesquisa como principal.

A rede (d), referente a distribuição geográfica de usinas elétricas, transformadores e consumidores não apresenta nenhuma concentração de nós e é uma rede conexa, ou seja, toda a rede está, necessariamente, interligada por cabos de energia.

O grau médio de cada rede especificado na Tabela 2 pode revelar se uma rede é densa ou não. Entretanto, ao analisar a rede (b), com grau médio igual a 4,0448, e a rede (e), com grau médio igual a 3,8835, na Figura 10, percebemos que esse valor parecido não indica a presença de aglomerações de nós na representação visual para todos os casos. O mesmo acontece entre as redes (c) e (d), com valores de grau médio iguais a 2,9043 e 2,6687, respectivamente, e estruturas visuais diferentes.

Os valores de centralidade *Closeness* apresentadas na Tabela 2 mostram que essa é uma medida coerente neste contexto, de forma que, pode refletir bem como a representação

Tabela 3 – Maiores valores da centralidade *Betweenness*

	(a)	(b)	(c)	(d)	(e)
Máx. <i>Betweenness</i>	0,0377	0,03521	0,1816	0,2883	0,3632
	0,0373	0,0329	0,1533	0,2817	0,1650
	0,0371	0,02893	0,1101	0,2797	0,1535

visual de uma rede irá se parecer. Na Figura 10 as redes (a), (c) e (e), com valores de centralidade *Closeness* iguais a 0,2699, 0,1938 e 0,3410, respectivamente, realmente apresentam representações visuais mais semelhantes. O mesmo acontece com as redes (b) e (d), com valores de centralidade *Closeness* iguais a 0,0756 e 0,0670, respectivamente, e estruturas visuais similares.

A análise da centralidade *Betweenness* é feita de acordo com os primeiros maiores valores encontrados dentro de cada rede, uma vez que, a medida encontra nós que funcionam como ponte entre comunidades, então a média feita com 10% dos maiores valores pode esconder nós que exercem essa função. Na Tabela 3 são expostos os três maiores valores da centralidade *Betweenness* encontrados em cada rede e ao compará-los com os valores médios de *Betweenness Centrality* presentes na Tabela 2 conclui-se, principalmente, que para as redes (a), (d) e (e) essa análise é mais vantajosa.

Na Tabela 3 pode-se perceber uma semelhança entre os valores das redes (a) e (b), as quais apresentam visualizações bem diferentes na Figura 10, e também entre as redes (d) e (e), que são visualmente bem diferentes mas a média dos três maiores valores da centralidade *Betweenness*, 0,2832 e 0,2272 respectivamente, demonstra que elas possuem nós que trabalhariam como pontes de forma similar.

A última linha da Tabela 2 introduz o valor da medida estatística Modularidade de cada rede. O resultado dos cálculos referentes as redes (a) e (e) chamam bastante atenção pois, além dos valores serem praticamente iguais e estarem dentro do intervalo $[0.3, 0.7]$, que indica um bom envolvimento dos nós dentro de suas respectivas comunidades, a representação visual das duas na Figura 10 apresentam certas diferenças, por exemplo, a rede (e) é mais descentralizada e possui limiares entre suas comunidades mais aparentes.

4 Conclusão e Trabalhos Futuros

A análise de redes é uma área de pesquisa grande e que pode ser aplicada a diferentes domínios do conhecimento. Neste trabalho duas linhas de estudo importantes foram apresentadas: o cálculo de medidas estatísticas e a utilização de ferramentas capazes de criar representações visuais de redes. Desta maneira, os objetivos deste trabalho foram alcançados, já que foi possível fazer uma comparação entre a análise visual e a estatística. No caso, foram encontrados resultados numéricos semelhantes para redes com comportamentos visuais distintos.

Este trabalho aplica as duas estratégias de análise em redes complexas que foram captadas em diferentes cenários, como em interações proteicas de vegetais, em plantas de energia e em citações entre cientistas, com o objetivo de achar medidas estatísticas semelhantes até mesmo em bases de dados que representam domínios completamente diferentes.

Portanto, com a aplicação da visualização de redes complexas foram observados padrões e comportamentos que poderiam passar despercebidos caso somente a análise estatística fosse feita. Logo, constata-se a importância de analisar uma rede de formas diferentes, para que os estudos se complementem e a base de dados seja compreendida de forma completa.

4.1 Trabalhos Futuros

O presente trabalho possui limitações quanto a profundidade da análise e quantidade de redes analisadas. No entanto, os indícios apresentados aqui geram demandas para novas análises.

Este trabalho pode se estender ainda sobre as semelhanças encontradas entre medidas estatísticas e representações visuais de redes temporais, as quais foram apresentadas e que adicionam níveis de dificuldades ainda maiores em sua análise, principalmente, pela adição do fator tempo e pelo tamanho que essas bases de dados podem atingir. Além disso, ainda existem outras medidas de redes complexas que podem ser levadas em consideração na análise de redes e que poderiam, por exemplo, explicitar melhor que duas redes possuem representações visuais diferentes.

Referências

- ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. **Rev. Mod. Phys.**, American Physical Society, v. 74, p. 47–97, Jan 2002. Disponível em: <<https://doi.org/10.1103/RevModPhys.74.47>>. Citado 2 vezes nas páginas 7 e 9.
- BEHRISCH, M.; BACH, B.; RICHE, N. H.; SCHRECK, T.; FEKETE, J.-D. Matrix reordering methods for table and network visualization. **Computer Graphics Forum**, v. 35, n. 3, p. 693–716, 2016. Disponível em: <<https://doi.org/10.1111/cgf.12935>>. Citado na página 13.
- BHADRA, A.; JORDÁN, F.; SUMANA, A.; DESHPANDE, S.; GADAGKAR, R. A comparative social network analysis of wasp colonies and classrooms: Linking network structure to functioning. **Ecological Complexity**, v. 6, p. 48–55, 03 2009. Disponível em: <<https://doi.org/10.1016/j.ecocom.2008.10.004>>. Citado 2 vezes nas páginas 17 e 18.
- BLONDEL, V. D.; GUILLAUME, J.-L.; LAMBIOTTE, R.; LEFEBVRE, E. Fast unfolding of communities in large networks. **Journal of Statistical Mechanics: Theory and Experiment**, IOP Publishing, v. 2008, n. 10, p. P10008, oct 2008. Disponível em: <<https://doi.org/10.1088/1742-5468/2008/10/P10008>>. Citado na página 11.
- BOX-STEF., J. M.; CHRIS., D. P. Comparing membership interest group networks across space and time, size, issue and industry. **Network Science**, Cambridge University Press, v. 3, n. 1, p. 78–97, 2015. Disponível em: <<https://doi.org/10.1017/nws.2015.6>>. Citado na página 18.
- BU, D.; ZHAO, Y.; CAI, L.; XUE, H.; ZHU, X.; LU, H.; ZHANG, J.; SUN, S.; LING, L.; ZHANG, N.; LI, G.-J.; CHEN, R. Topological structure analysis of the protein-protein interaction network in budding yeast. **Nucleic acids research**, v. 31, p. 2443–50, 06 2003. Disponível em: <<https://doi.org/10.1093/nar/gkg340>>. Citado na página 21.
- ELZEN, S.; HOLTEN, D.; BLAAS, J.; WIJK, J. J. van. Dynamic network visualization with extended massive sequence views. **IEEE Transactions on Visualization & Computer Graphics**, IEEE Computer Society, Los Alamitos, CA, USA, v. 20, n. 08, p. 1087–1099, aug 2014. ISSN 1941-0506. Disponível em: <<https://doi.org/10.1109/TVCG.2013.263>>. Citado 2 vezes nas páginas 14 e 15.
- FAUST, K.; SKVORETZ, J. Comparing networks across space and time, size and species. **Sociological Methodology**, v. 32, n. 1, p. 267–299, 2002. Disponível em: <<https://doi.org/10.1111/1467-9531.00118>>. Citado na página 18.
- FORTUNATO, S. Community detection in graphs. **Physics Reports**, v. 486, n. 3, p. 75 – 174, 2010. ISSN 0370-1573. Disponível em: <<https://doi.org/10.1016/j.physrep.2009.11.002>>. Citado na página 10.
- FREEMAN, L. C. A set of measures of centrality based on betweenness. **Sociometry**, [American Sociological Association, Sage Publications, Inc.], v. 40, n. 1, p. 35–41, 1977.

ISSN 00380431. Disponível em: <<https://doi.org/10.2307/3033543>>. Citado na página 16.

GOLBECK, J. Chapter 21 - analyzing networks. In: GOLBECK, J. (Ed.). **Introduction to Social Media Investigation**. Boston: Syngress, 2015. p. 221–235. ISBN 978-0-12-801656-5. Disponível em: <<https://doi.org/10.1016/B978-0-12-801656-5.00021-4>>. Citado na página 16.

GOLDSTEIN, R.; VITEVITCH, M. S. The influence of closeness centrality on lexical processing. **Frontiers in Psychology**, v. 8, p. 1683, 2017. ISSN 1664-1078. Disponível em: <<https://doi.org/10.3389/fpsyg.2017.01683>>. Citado na página 16.

GRANDJEAN, M.; JACOMY, M. Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics. In: **Digital Humanities**. Utrecht, Netherlands: [s.n.], 2019. Disponível em: <<https://halshs.archives-ouvertes.fr/halshs-02179024>>. Citado 3 vezes nas páginas 7, 16 e 17.

HAGBERG, A.; SWART, P.; CHULT, D. S. Exploring network structure, dynamics, and function using networkx. 1 2008. Disponível em: <<https://www.osti.gov/biblio/960616>>. Citado na página 21.

KOMOLAFE, T.; QUEVEDO, A. V.; SENGUPTA, S.; WOODALL, W. H. Statistical evaluation of spectral methods for anomaly detection in static networks. **Network Science**, Cambridge University Press, v. 7, n. 3, p. 319–352, 2019. Disponível em: <<https://doi.org/10.1017/nws.2019.14>>. Citado na página 18.

LADD, J.; OTIS, J.; WARREN, C.; WEINGART, S. Exploring and analyzing network data with python. **The Programming Historian**, 08 2017. Disponível em: <<https://doi.org/10.46430/phen0064>>. Citado na página 21.

LATORA, V.; NICOSIA, V.; RUSSO, G. **Complex Networks: Principles, Methods and Applications**. 1st. ed. USA: Cambridge University Press, 2017. ISBN 1107103185. Disponível em: <<https://doi.org/10.1017/9781316216002>>. Citado na página 21.

LINHARES, C.; PONCIANO, J.; PEREIRA, F.; ROCHA, L.; PAIVA, J. G.; TRAVENÇOLO, B. Visual analysis for evaluation of community detection algorithms. **Multimedia Tools and Applications**, v. 79, 07 2020. Disponível em: <<https://doi.org/10.1007/s11042-020-08700-4>>. Citado 4 vezes nas páginas 10, 11, 17 e 18.

LINHARES, C.; PONCIANO, J.; ROCHA, L.; PAIVA, J. G.; TRAVENÇOLO, B. Análise temporal de uma rede de contato hospitalar utilizando técnicas de visualização de informação. In: **Anais do XVII Workshop de Informática Médica**. Porto Alegre, RS, Brasil: SBC, 2017. ISSN 0000-0000. Disponível em: <<https://doi.org/10.5753/sbcas.2017.3696>>. Citado na página 11.

LINHARES, C. D. G. Técnicas avançadas para análise visual de redes temporais. 2020. Disponível em: <<http://doi.org/10.14393/ufu.te.2020.375>>. Citado na página 14.

LINHARES, C. D. G.; PONCIANO, J. R.; PAIVA, J. G. S.; TRAVENÇOLO, B. A. N.; ROCHA, L. E. C. Visualisation of structure and processes on temporal networks. In: _____. **Temporal Network Theory**. Cham: Springer International Publishing, 2019. p. 83–105. ISBN 978-3-030-23495-9. Disponível em: <https://doi.org/10.1007/978-3-030-23495-9_5>. Citado 3 vezes nas páginas 7, 11 e 14.

LINHARES, C. D. G.; PONCIANO, J. R.; PAIVA, J. G. S.; TRAVENÇOLO, B. A. N.; ROCHA, L. E. C. A comparative analysis for visualizing the temporal evolution of contact networks: a user study. **Journal of Visualization**, Springer Science and Business Media LLC, jun 2021. Disponível em: <<https://doi.org/10.1007/s12650-021-00759-x>>. Citado na página 13.

LINHARES, C. D. G.; TRAVENÇOLO, B. A. N.; PAIVA, J. G. S.; ROCHA, L. E. C. Dynetvis: A system for visualization of dynamic networks. In: **Proceedings of the Symposium on Applied Computing**. New York, NY, USA: ACM, 2017. (SAC '17), p. 187–194. ISBN 978-1-4503-4486-9. Disponível em: <<http://doi.acm.org/10.1145/3019612.3019686>>. Citado 2 vezes nas páginas 13 e 20.

Miller, B. A.; Beard, M. S.; Wolfe, P. J.; Bliss, N. T. A spectral framework for anomalous subgraph detection. **IEEE Transactions on Signal Processing**, v. 63, n. 16, p. 4191–4206, 2015. Disponível em: <<https://doi.org/10.1109/TSP.2015.2437841>>. Citado na página 18.

PONCIANO, J. R.; LINHARES, C. D. G.; ROCHA, L. E. C.; FARIA, E. R.; TRAVENÇOLO, B. A. N. A streaming edge sampling method for network visualization. **Knowledge and Information Systems**, Springer Science and Business Media LLC, v. 63, n. 7, p. 1717–1743, apr 2021. Disponível em: <<https://doi.org/10.1007/s10115-021-01571-7>>. Citado 2 vezes nas páginas 11 e 14.

RIDER, A. K.; MILENKOVIĆ, T.; SIWO, G. H.; PINAPATI, R. S.; EMRICH, S. J.; FERDIG, M. T.; CHAWLA, N. V. Networks' characteristics are important for systems biology. **Network Science**, Cambridge University Press, v. 2, n. 2, p. 139–161, 2014. Disponível em: <<https://doi.org/10.1017/nws.2014.13>>. Citado na página 18.

ROCHA, L. E. C. da. Structural evolution of the brazilian airport network. **Journal of Statistical Mechanics: Theory and Experiment**, IOP Publishing, v. 2009, n. 04, p. P04020, apr 2009. Disponível em: <<https://doi.org/10.1088/1742-5468/2009/04/P04020>>. Citado na página 11.

ROSVALL, M.; BERGSTROM, C. T. Maps of random walks on complex networks reveal community structure. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 105, n. 4, p. 1118–1123, 2008. ISSN 0027-8424. Disponível em: <<https://doi.org/10.1073/pnas.0706851105>>. Citado na página 11.

SABIDUSSI, G. The centrality index of a graph. **Psychometrika**, v. 31, n. 4, p. 581–603, 1966. Disponível em: <<https://doi.org/10.1007/BF02289527>>. Citado na página 15.

VACCA, R. Structure in personal networks: Constructing and comparing typologies. **Network Science**, Cambridge University Press, v. 8, n. 2, p. 142–167, 2020. Disponível em: <<https://doi.org/10.1017/nws.2019.29>>. Citado na página 18.

WARE, C. **Information Visualization (Third Edition)**. Third edition. Boston: Morgan Kaufmann, 2013. (Interactive Technologies). ISBN 978-0-12-381464-7. Disponível em: <<https://doi.org/10.1016/C2009-0-62432-6>>. Citado 2 vezes nas páginas 7 e 11.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of 'small-world' networks. **Nature**, v. 393, n. 6684, p. 440–442, 1998. Disponível em: <<https://doi.org/10.1038/30918>>. Citado na página 21.

ZAIDI, F.; MUELDER, C.; SALLABERRY, A. Analysis and visualization of dynamic networks. **CoRR**, abs/1409.5034, 2014. Disponível em: <https://doi.org/10.1007/978-1-4614-6170-8_382>. Citado na página 9.