

Comparação entre a detecção de comunidades e o algoritmo K-Means na descoberta do espectro do Transtorno do Déficit de Atenção e Hiperatividade (TDAH).

Marnes Adão Simão Cassule
*Programa de Pós-Graduação em
Engenharia Elétrica e Computação
Universidade Presbiteriana Mackenzie*
São Paulo, Brazil
marnesadao.cassule@mackenzista.com.br

Luiz Renato Rodrigues Carreiro
*Programa de Pós-Graduação em
Distúrbios do Desenvolvimento
Universidade Presbiteriana Mackenzie*
São Paulo, Brazil
luizrenato.carreiro@mackenzie.br

Leandro Augusto Silva
*Programa de Pós-Graduação em
Engenharia Elétrica e Computação
Universidade Presbiteriana Mackenzie*
São Paulo, Brazil
leandroaugusto.silva@mackenzie.br

Abstract—Transtorno do Déficit de Atenção/Hiperatividade (TDAH) é um transtorno do neurodesenvolvimento que afeta a atenção e o controle comportamental. Os sintomas são frequentemente tratados de forma individualizada, mas a heterogeneidade do transtorno sugere a necessidade de abordagens mais personalizadas. Este estudo utilizou uma rede complexa de pacientes com TDAH para investigar a possibilidade de diferentes perfis de prejuízo atencional. A análise da rede revelou a presença de dois grupos distintos: um grupo moderado e um grupo severo. O grupo moderado apresentou tempos de reação mais rápidos e maiores valores de centralidade do que o grupo severo. Essas descobertas sugerem que os dois grupos podem apresentar diferentes características atencionais, o que pode ter implicações para o diagnóstico e tratamento do TDAH. O estudo utilizou o algoritmo de clusterização KMeans para identificar os dois grupos. Os resultados do teste de hipótese forneceram suporte à hipótese alternativa, indicando que a divisão em dois grupos distintos com base na variabilidade atencional é estatisticamente significativa. Este estudo fornece evidências de que os pacientes com TDAH podem apresentar diferentes características atencionais. Essas descobertas podem levar ao desenvolvimento de abordagens de diagnóstico e tratamento mais personalizado para o TDAH.

Palavras-chave—déficit de atenção e hiperatividade/impulsividade, variabilidade atencional, detecção de comunidade, redes complexas, espectro do transtorno do déficit de atenção e hiperatividade, kmeans, cluster, elbow

I. INTRODUÇÃO

O TDAH (Transtorno de Déficit de Atenção/Hiperatividade) é um transtorno que se caracteriza pela presença constante de desatenção e/ou hiperatividade-impulsividade, o que pode interferir no funcionamento social, pessoal e no desenvolvimento da criança [7]. Para que o TDAH seja diagnosticado, é necessário que os sintomas sejam manifestados em mais de um ambiente, ou seja, o indivíduo deve apresentar sinais de TDAH na maioria dos ambientes que frequenta. Entre os sintomas que podem caracterizar a desatenção estão: constante distração em tarefas, faltas de persistência, dificuldade de manter o foco e desorganização, e não escutar quando alguém lhe dirige a palavra. Já para a hiperatividade, há comportamentos

excessivos na atividade motora em situações inadequadas (subir em coisas), incapacidade em brincar ou se envolver em atividades de lazer de uma forma calma e dificuldade em esperar a sua vez de falar em conversas. Na idade adulta, esses comportamentos podem se manifestar de forma diferente, como uma sensação de inquietude e pensamento acelerado. O transtorno tem início na infância e pode persistir até a idade adulta, e está presente em cerca de 5% das crianças e 2,5% dos adultos.

O TDAH se apresenta de várias formas e é multifatorial, não havendo um marcador biológico específico para o seu diagnóstico. Por isso, o processo de avaliação para diagnosticar o TDAH envolve uma avaliação criteriosa do paciente [11]. Para entender o TDAH, é necessário compreender como ocorre o desenvolvimento da atenção, nas suas diferentes formas, e como os déficits nas habilidades nesse processo podem levar aos problemas de atenção associados ao TDAH [12]. Portanto, a desatenção é uma característica determinante na avaliação das queixas de TDAH e pode prejudicar o paciente na realização de tarefas e levar a comportamentos impulsivos [7], [13], [14].

No entanto, devido às variações no nível de atenção, tratar todos os pacientes da mesma maneira pode não ser uma solução ideal [19]. Nesse contexto, a identificação de características dessa variação e a possibilidade de agrupar os pacientes se tornam desafios que podem ser abordados por meio de técnicas computacionais. Uma abordagem promissora para lidar com esse caso e identificar as variações de desatenção é o uso de Redes Complexas, especialmente a detecção de comunidades [1]–[3], [15] e o uso de algoritmos de clusterização como o K-means [32]. Essas duas técnicas têm sido aplicadas em diferentes domínios. O k-means tem sido utilizado em diversos campos, como detecção dinâmica de comunidades em redes sociais online [34], classificação de habilidades de estudantes [36], soluções de recuperação para nuvens híbridas [38], gerenciamento de conhecimento por

meio do agrupamento de documentos [39] e detecção robusta de comunidades em grafos [40]. Essas aplicações destacam a versatilidade do algoritmo k-means na identificação de clusters e comunidades em vários conjuntos de dados. Essas técnicas computacionais permitem analisar as interações entre os pacientes e identificar grupos ou subgrupos com características semelhantes.

Neste artigo, é realizado uma análise comparativa entre a técnica de detecção de comunidades em uma rede complexa, que representa 30 pacientes com TDAH, e a aplicação do algoritmo K-Means com o uso do método do cotovelo (Elbow Method) nos dados desses pacientes.

Foi realizada uma análise detalhada das características de cada paciente em cada comunidade e cluster. O objetivo é compreender se determinadas comunidades estão mais associadas a sintomas leves ou mais agressivos, permitindo definir um espectro de TDAH. Essa abordagem auxiliará na identificação de padrões e variações na gravidade dos sintomas dentro das comunidades identificadas.

Ao explorar as características individuais e a estrutura das comunidades, espera-se obter insights valiosos sobre a heterogeneidade do TDAH. Esses insights podem contribuir para o desenvolvimento de intervenções e tratamentos mais personalizados e eficazes, adaptados às necessidades específicas de cada grupo de pacientes.

II. MATERIAIS E MÉTODOS

A. Base de Dados

O conjunto de dados utilizado nesta pesquisa consiste em informações sobre o desempenho de 30 crianças em um teste computadorizado (8 meninas e 22 meninos), com idades entre 6 e 12 anos, que foram diagnosticadas com TDAH. Conforme documentado no Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM-5) [7], o Transtorno de Déficit de Atenção e Hiperatividade (TDAH) é estimado em afetar aproximadamente 5% da população infantil, com uma prevalência de 2,5% na população adulta. Além disso, é notável uma disparidade de gênero associada a esta condição, revelando uma proporção de dois meninos afetados para cada menina durante a infância e uma proporção de 1,6 homens afetados para cada mulher na fase adulta. É importante ressaltar que essa discrepância de gênero está na raiz da razão pela qual o conjunto de dados em análise possui um número menor de meninas em comparação aos meninos. O teste foi projetado para medir a orientação atencional voluntária e automática baseada no tempo de reação entre a indicação do alvo até a ação de pressionamento da barra de espaço do teclado. O tempo é medido em milissegundos. Esses dados foram coletados com o objetivo de investigar diferenças no desempenho atencional entre os dois grupos de crianças.

Para a tarefa de orientação voluntária, os participantes foram apresentados a um ponto de fixação central na tela de um computador e duas caixas na lateral direita e esquerda da tela. Após um intervalo aleatório de tempo (800 à 1800 *ms*), uma seta apontando para a esquerda ou direita aparece ao lado do ponto de fixação. Depois de um período de tempo

específico (300 *ms* à 800 *ms*), um alvo é exibido em uma das duas caixas e permanece até que o participante pressione a barra de espaço de um teclado ou após 1500 *ms*. O alvo pode aparecer na posição indicada pela seta, considerada a condição válida em 70% das apresentações, ou na posição oposta indicada pela seta, considerada a condição inválida em 30% das apresentações. Os participantes são orientados a manter seu olhar no ponto de fixação e direcionar sua atenção para a posição indicada pela seta, respondendo o mais rápido possível, independentemente da posição do alvo.

Na tarefa de orientação automática, um ponto de fixação é apresentado no centro da tela juntamente com duas caixas, uma à direita e outra à esquerda. Após 700 *ms*, uma das duas caixas é iluminada (com 50% de probabilidade). Após um intervalo de 100 *ms* e 800 *ms*, um alvo (quadrado iluminado) é exibido dentro de uma das duas caixas e permanece assim até que o participante pressione a barra de espaço do teclado ou até que se atinja um tempo máximo de 1500 *ms*. Na condição ipsilateral, o alvo é exibido na mesma posição que o primeiro estímulo, e na condição contralateral, ele é exibido na posição oposta. Os participantes são instruídos a manter o olhar no ponto de fixação, ignorar o primeiro estímulo e responder ao alvo o mais rápido possível, independentemente da sua localização.

Com base nos resultados obtidos nos testes das atividades mencionadas, são gerados oito atributos que compõem o conjunto de dados. Os atributos são:

- **AutomContra100:** é um indicador em milissegundos utilizado para medir a velocidade do processamento da atenção em tarefa de orientação automática, indicando o tempo decorrido desde a exibição do alvo no ponto de fixação até o acionamento da barra de espaço em um intervalo de 100 *ms* na condição contralateral.
- **AutomContra800:** é uma medida em milissegundos usada para avaliar o tempo de reação em tarefas de orientação automática da atenção. Essa medida tem como objetivo mensurar o tempo transcorrido desde a exibição do alvo no ponto de fixação até o momento em que o participante aciona a barra de espaço em um intervalo de 800 *ms*, na condição contralateral.
- **AutomIpsol100:** mede a rapidez do processamento da atenção em tarefas de orientação automática em milissegundos, avaliando o tempo transcorrido desde a exibição do alvo no ponto de fixação até o acionamento da barra de espaço em um intervalo de 100 *ms* na condição ipsilateral.
- **AutomIpsol800:** mede o tempo de reação em tarefas de orientação automática da atenção em milissegundos, considerando o período de 800 *ms* na condição ipsilateral.
- **VoluntInvalida300:** é uma medida em milissegundos usada para avaliar o tempo de reação em uma tarefa de orientação voluntária de atenção. Ele mede o tempo desde a exibição do alvo no ponto de fixação até o acionamento da barra de espaço, em um intervalo de 300 *ms* após a indicação do alvo no lado incorreto.

- **VoluntInvalida800:** é uma medida em milissegundos usada para avaliar o tempo de reação em uma tarefa de orientação voluntária de atenção. Essa medida é obtida a partir do momento em que o alvo é apresentado no ponto de fixação até o momento em que a barra de espaço é acionada, em um intervalo de 800 ms após a indicação do alvo no lado incorreto
- **Voluntvalida300:** é uma medida em milissegundos usada para avaliar o tempo de reação em uma tarefa de orientação voluntária de atenção. Ele mede o tempo desde a exibição do alvo no ponto de fixação até o acionamento da barra de espaço, em um intervalo de 300 ms após a indicação do alvo no lado apropriado.
- **VoluntValida800:** é um valor em milissegundos(ms) que mede o tempo de reação desde o momento em que o alvo é exibido no ponto de fixação em uma tarefa de orientação voluntária de atenção, até o momento em que o usuário aciona a barra de espaço, em um intervalo de 800 ms entre a indicação do alvo e sua apresentação no lado indicado.

B. Métodos

Nesta subseção, serão apresentados detalhadamente os procedimentos e estratégias utilizadas para atingir o objetivo proposto neste trabalho. A metodologia geral esta representada de forma simplificada na figura 1 e foi elaborada com base em premissas científicas e em métodos validados pela literatura. Os passos e etapas envolvidos na coleta, análise e interpretação dos dados serão descritos, assim como as técnicas e ferramentas utilizadas. A execução desse processo envolve o uso das linguagens R e Python. No R, são executados os procedimentos de coleta e pré-processamento de dados, enquanto no Python, são conduzidas as análises do grafo. A etapa de análise de redes é realizada com o suporte da ferramenta Gephi¹.

Cabe ainda destacar que para fins de proposta deste trabalho, apenas o conjunto de dados de crianças com TDAH será analisado. Isto é feito para efeitos de contribuição do trabalho com o objetivo de descobrir um espectro no grafo.

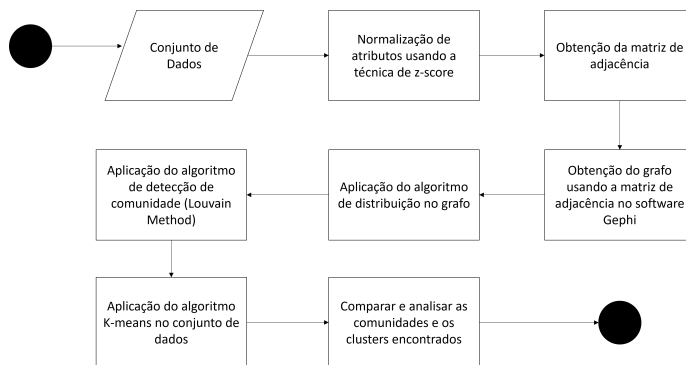


Fig. 1. Diagrama de Fluxo da Metodologia Utilizada na Pesquisa sobre Redes Complexas de Testes de Orientação da Atenção

¹<https://gephi.org/>

1) *Detectar Comunidades:* O processo (figura 1) inicia com a organização do conjunto de dados de tal forma que as crianças (sem nenhuma identificação pessoal) sejam organizadas em linhas e cada coluna os oito atributos em colunas, gerando assim um conjunto de dado com 30 objetos e 8 atributos.

O primeiro procedimento feito sobre o conjunto de dados é com a normalização dos atributos, usando o método z-score [29]. Este é um procedimento típico em análise de dados, mas sobretudo, na área da psicologia que tem como finalidade padronizar as diferentes escalas dos atributos por unidade de desvio padrão [28]. Este processo foi feito em R com o uso da função `scale()`, parte da biblioteca padrão da linguagem.

O segundo procedimento consiste em analisar a similaridade das crianças entre si. Assim, gera-se uma matriz em que se relacionada cada criança com toda a base, gerando uma matriz de similaridade de 30×30 para cada grupo (TDAH e controle). Cada elemento da matriz constitui a similaridade calculada usando a distância euclidiana, que é em R é feita com o uso da função `dist()`. Como o objetivo deste processo é gerar um grafo em que cada vértice é uma criança e a aresta o resultado da distância, a similaridade indica que os menores valores sejam os mais parecidos, mas em grafos se deseja expressa de forma inversa, ou seja, maior força de relacionamento. Desta forma, se faz necessário a seguinte operação na matriz de similaridade:

$$rel = \frac{1}{dist(x, y)} \quad (1)$$

sendo *rel* a relação entre duas crianças quaisquer *x* e *y* do conjunto de dados.

A partir da matriz de similaridade se faz a exportação dos dados para o formato desejado do Gephi usando a função `gephi_write_edges_from_df()` da biblioteca Gephi na linguagem R [26]. O arquivo resultante contém informações sobre a força de conexão entre os nós, que são representadas pelos pesos.

Com a matriz de adjacência importada no Gephi se pode fazer a manipulação e exploração da rede, bem como permite visualizar, espacializar, filtrar e exportar os resultados [26]. Com a importação não se permite fazer alguma inferência da rede, uma vez que a aparência inicial tem um aspecto caótico. Para tanto se decidiu pelo uso do algoritmo de distribuição de grafo denominado Force Atlas 2 [6]. O algoritmo utiliza uma abordagem de simulação física para espacializar uma rede, fazendo com que os nós se repilam mutuamente como partículas com cargas, enquanto as arestas os atraem como molas, resultando em uma representação gráfica mais clara e legível do grafo.

Com o grafo modelado, aplica-se as medidas de centralidades (intermediação, grau, proximidade) e posteriormente aplica-se a modularidade para a detecção das comunidades.

A medida de centralidade grau, atribui ao nó a importância de acordo a quantidade de ligações diretas que o vértice

estabelece com outros vértices do grafo [27]. Representado conforme a equação:

$$C_G(v) = \frac{d(v)}{n-1} \quad (2)$$

onde $C_G(v)$ é a centralidade de grau do vértice v , $d(v)$ é o grau do vértice v , e n é o número total de vértices no grafo. [18].

A centralidade de intermediação baseia-se no fluxo de informação através dos nós, por isso, é também chamada centralidade do fluxo. Em seu cálculo assume-se que a informação flui sempre através dos caminhos mais curtos. A medida de intermediação calcula o número de vezes que um vértice atua como ponte ao longo do caminho mais curto entre dois outros vértices [17].

Para calcular a intermediação de um vértice v_k no grafo G com n vértices, é necessário somar todas as intermediações parciais de v_k em G . O valor encontrado será denotado por $C_B(v_k)$ e mede-se a centralidade de intermediação de v_k em G , da seguinte maneira:

$$C_B(v_k) = \sum_{s \neq v_k \neq t} \frac{\sigma_{st}(v_k)}{\sigma_{st}} \quad (3)$$

onde $C_B(v_k)$ é a centralidade de intermediação do vértice v_k , σ_{st} é o número de caminhos mais curtos entre os vértices s e t , e $\sigma_{st}(v_k)$ é o número de caminhos mais curtos entre s e t que passam pelo vértice v_k .

A centralidade de proximidade indica o quão próximo um vértice se localiza de outro vértice [27]. Dado que um vértice v_k pertence a um grafo G , a centralidade de proximidade desse vértice é calculada fazendo o inverso da soma das menores distâncias de v_k para todos os outros vértices pertencentes a G .

$$C_P(v) = \frac{1}{\sum_{u \neq v} d(u, v)} \quad (4)$$

onde $C_P(v)$ é a centralidade de proximidade do vértice v e $d(u, v)$ é a distância entre os vértices u e v .

Após o cálculo das medidas de centralidade, aplica-se o algoritmo de modularidade [16], que avalia a conectividade de partições de um grafo segundo a configuração (grau, betweenness, intermediação) e permite a detecção de comunidade [15]. O algoritmo detecção de comunidade do gephi é implementado com o método Louvain [10], o qual é um algoritmo heurístico baseado na modularidade. O método Louvain descobre automaticamente comunidades caracterizadas por vértices com maior interação entre si, em relação aos vértices que estão fora da comunidade.

O algoritmo Louvain [10] é iterativo e cada iteração melhora a modularidade global do grafo. A modularidade final obtida pelo algoritmo Louvain representa a divisão do grafo em comunidades que maximizam a estrutura interna e minimizam as conexões entre as comunidades. É um método eficiente para detectar comunidades em grafos. Ele opera otimizando a modularidade do grafo, movendo vértices entre comunidades

de forma iterativa. O algoritmo é capaz de encontrar uma partição de comunidades de alta qualidade em grandes grafos, tornando-o uma ferramenta valiosa para análise de redes complexas em várias áreas de estudo.

Algorithm 1 Algoritmo de Louvain

```

0: Inicialize cada vértice como uma comunidade separada
0: repeat
0:   repeat
0:     for cada vértice  $v$  do
0:       Calcule o ganho de modularidade  $\Delta Q$  movendo  $v$ 
para cada comunidade vizinha  $C$ 
0:       Defina  $C_{\max}$  como a comunidade que maximiza o
 $\Delta Q$ 
0:       Atualize a comunidade de  $v$  para  $C_{\max}$  e armazene
o  $\Delta Q$  correspondente
0:   until não houver mais melhorias na modularidade
0:   Mesclar comunidades com os mesmos rótulos
0:   Reatribuir novos rótulos às comunidades mescladas
0: until não houver mais melhorias na modularidade =0

```

2) *K-Means E Método Elbow*: O algoritmo K-Means [32], [33] é um método de agrupamento que busca minimizar a função objetivo de distorção, dada por:

$$J = \sum_i \sum_j \|x_i - \mu_j\|^2 \quad (5)$$

onde x é um ponto de dados, μ_j é o centroide do j -ésimo cluster e $\|\cdot\|$ representa a norma Euclidiana. O algoritmo visa encontrar a atribuição ótima dos pontos aos clusters e os centroides correspondentes que minimizam a função objetivo. O processo de execução do K-Means é realizado em etapas iterativas:

- 1) Inicialização: Selecionam-se K centroides iniciais, geralmente aleatoriamente ou usando algum método heurístico.
- 2) Atribuição: Calcula-se a distância entre cada ponto de dados e os centroides. Cada ponto é atribuído ao centroide mais próximo, formando clusters iniciais.
- 3) Atualização: Recalculam-se os centroides para cada cluster com base na média dos pontos atribuídos a ele.
- 4) Iteração: Repetem-se as etapas 2 e 3 até que ocorra a convergência, ou seja, os centroides não se movam significativamente ou o número máximo de iterações seja atingido.

O resultado final do algoritmo é uma configuração de clusters otimizada, onde cada ponto de dados é atribuído a um cluster específico. Essa configuração é determinada pelos centroides que representam o centro de cada cluster. O algoritmo K-Means é caracterizado por sua simplicidade e eficiência computacional, tornando-o amplamente utilizado em diversas áreas de pesquisa [35], [37]. No entanto, ele pode estar sujeito a problemas de inicialização inadequada, que podem levar a soluções subótimas ou presas em mínimos locais da função objetivo. Um desses problemas é a definição do número de

clusters, muitas vezes por suposição e também pela suposição inicial do centróide do cluster. Para impedir esse problema, é necessário aplicar um método de otimização, sendo um dos métodos usados o Elbow method [35], [37]. Este método é um método visual para testar a consistência do melhor número de clusters, comparando a diferença da soma dos erros quadrados (SSE) de cada cluster. A diferença mais extrema que forma o ângulo do cotovelo mostra o melhor número de clusters.

C. Resultados

Como mencionado anteriormente, no presente estudo foi gerada uma rede composta por 30 elementos representando casos positivos de TDAH. A representação gráfica dessa rede pode ser observada na Figura 2. É importante ressaltar que as cores utilizadas na figura são meramente ilustrativas e não possuem associação com informações além das análises realizadas em cada procedimento experimental.

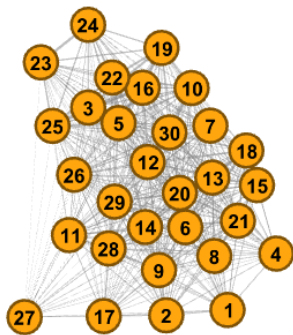


Fig. 2. Rede com elementos (pacientes) com TDAH.

Ao aplicar a medida de centralidade grau na rede, cujos resultados são ilustrados na Figura 3, foi constatado que 80% dos nós na rede de pacientes com TDAH estão conectados aos outros 29 nós. Por outro lado, os restantes 20% dos nós, correspondentes a 6 nós específicos (6, 14, 15, 19, 25, 29), estão conectados apenas a 28 nós, conforme representado na Figura 3. A alta conectividade observada entre os 80% dos nós da rede sugere uma propensão a interações sociais intensas ou uma maior disseminação de informações dentro desse grupo. Além disso, essa conectividade também pode indicar que os pacientes que se conectam na rede compartilham características semelhantes em relação às atividades de atenção voluntária e automáticas.

Por outro lado, o grupo de pacientes com TDAH que apresenta um grau ligeiramente menor de conexões sugere uma menor interconectividade dentro desse subgrupo. Essa observação pode ser indicativa de uma menor propensão a interações sociais intensas ou a uma disseminação de informações menos abrangente. É importante ressaltar que essas conclusões são baseadas nos resultados da medida de centralidade grau e podem fornecer insights relevantes sobre a estrutura da rede e a dinâmica de conexões entre os pacientes com TDAH analisados.

Com relação à medida de intermediação na rede, foi observado que 80% dos nós com o maior grau apresentam uma frequência de 0,10 em aparecer nos caminhos mais curtos entre os vértices do grafo. O que é consistente com os resultados da medida de grau, que também identificou esses nós como os mais conectados. Por outro lado, os 20% restantes (6, 14, 15, 19, 25, 29) possuem uma taxa de 0,07. Essa diferença indica que os nós destacados pela medida de intermediação desempenham um papel significativo na intermediação de informações na rede. É provável que eles sejam responsáveis por conectar diferentes partes da rede e facilitar a comunicação entre os nós.

Além disso, a medida de proximidade atribui importância semelhante aos nós evidenciados pelas medidas de grau e intermediação. Esses nós apresentam a menor distância entre os vértices do grafo em relação aos outros 20% dos nós. Isso sugere que esses nós têm maior proximidade e são potencialmente cruciais para a conectividade geral da rede. Eles podem ajudar a manter a rede estável e evitar que ela se fragmente.

Essas descobertas ressaltam a relevância dos nós com alta intermediação e proximidade na estrutura e dinâmica da rede. Eles desempenham um papel fundamental na transferência eficiente de informações e na manutenção da conectividade entre os vértices do grafo.

Destas análises pode-se concluir haver algum tipo de variação na rede descobertas em forma de comunidade. Por isto, como resultado da aplicação do algoritmo de modularidade foi possível detectar a existência de duas comunidades claramente distintas no grafo. Essa semelhança é importante porque ela sugere que as técnicas estão identificando os mesmos grupos de pacientes com TDAH.

As comunidades detectadas serão classificadas em grupo Moderado e Severo, conforme ilustrado na figura 4. A atribuição se dará por conseguinte em análise aos atributos que compõem o conjunto de dados, portanto, que representam cada nó.

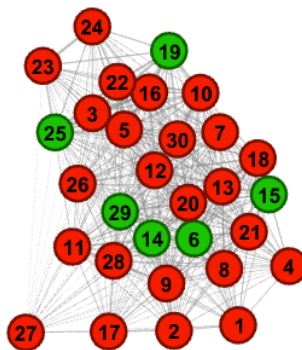


Fig. 3. Resultados de centralidade.

O algoritmo K-Means foi aplicado ao conjunto de dados contendo resultados de testes de atenção em crianças com diagnóstico positivo de TDAH (Transtorno do Déficit de Atenção com Hiperatividade). Inicialmente, o método do co-

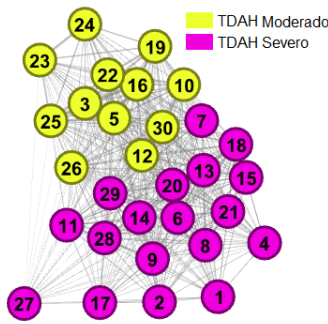


Fig. 4. Comunidades descobertas.

tovelo (Elbow method) foi utilizado para determinar o número ideal de clusters. Com base na análise visual do gráfico da Figura 5, o método indicou a presença de 2 clusters como o número ótimo.

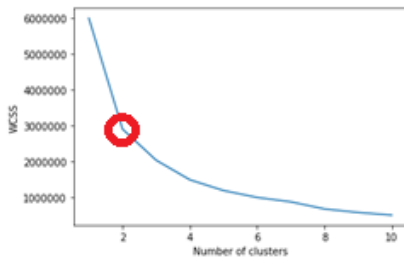


Fig. 5. Número de Cluster Ideal

A quantidade de comunidades detectadas corresponde ao número de clusters determinado pelo método Elbow ao aplicar o algoritmo nos dados utilizados nos experimentos. Além disso, a quantidade de elementos em cada cluster é igual à quantidade de elementos em cada comunidade detectada na rede complexa, conforme apresentado na tabela II-C.

Cluster #	Cluster Size
0	18
1	12

TABELA I

QUANTIDADE DE ELEMENTOS EM CADA CLUSTER

Os resultados obtidos por meio da aplicação do algoritmo K-means e do método Elbow revelaram uma correspondência significativa entre os elementos pertencentes a cada comunidade detectada e os clusters identificados através da aplicação do algoritmo K-means e do método Elbow. Essa consistência nas atribuições dos elementos a diferentes grupos é uma evidência robusta da concordância entre os métodos utilizados para análise de comunidades em redes complexas.

Na figura 6, é possível observar que os dois grupos possuem características distintas. Os pacientes com TDAH severo apresentam, em geral, tempos de reação mais longos nas métricas AutomContra100 e AutomContra800.

Foi realizado o teste T de Student [41] para duas amostras independentes para comparar as médias dos grupos encontrados. Os resultados do teste mostraram que há uma diferença

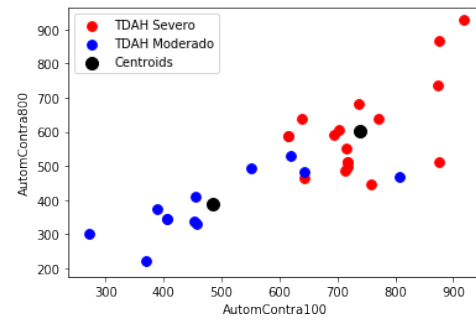


Fig. 6. Gráfico de dispersão dos clusters.

significativa entre as médias dos dois grupos para todos os atributos. Os valores t e os p-valores para cada atributo são mostrados na tabela IV.

A similaridade entre as atribuições dos elementos em cada comunidade e cluster reforça a validade dessas técnicas no contexto da análise de redes complexas, pois sugere que elas são capazes de identificar estruturas intrínsecas e padrões existentes nos dados.

É importante ressaltar que essa correspondência entre as comunidades detectadas pelo método Louvain e os clusters obtidos pelo K-means e pelo método Elbow fornece um forte suporte para a interpretação dos resultados, pois sugere que as técnicas estão identificando os mesmos grupos de pacientes com TDAH. Essa convergência metodológica contribui para a compreensão mais aprofundada dos padrões de relacionamento entre os elementos e oferece insights valiosos para a caracterização e interpretação das comunidades em redes complexas.

Cluster #	Elementos
0	1, 2, 4, 6, 7, 8, 9, 11, 13, 14, 15, 17, 18, 20, 21, 27, 28, 19
1	3, 5, 10, 12, 16, 19, 22, 23, 24, 25, 26, 30

TABELA II
ELEMENTO NOS SEUS CLUSTERS

Com base na análise exploratória dos dados de teste de atenção, foi constatado que os dois grupos detectados nas duas técnicas, o grupo TDAH Moderado e o grupo TDAH Severo, apresentaram diferenças significativas em relação aos atributos de atenção. O grupo TDAH Moderado obteve tempos de reação significativamente menores em todas as métricas de atenção, em comparação com o grupo TDAH Severo. Em contraste, o grupo TDAH Severo apresentou um tempo de processamento da atenção mais prolongado em todas as atividades testadas.

III. DISCUSSÃO E CONCLUSÃO

O presente estudo comparou duas abordagens para identificar variações de valores provenientes de testes realizados em pacientes com TDAH: a técnica de detecção de comunidades em redes complexas e a aplicação do algoritmo K-means com o método do cotovelo (Elbow method). O objetivo do estudo foi investigar a possibilidade de definir perfis de prejuízo aten-

Atributo	TDAH Moderado	TDAH Severo
AutomContra100	272 - 807	615 - 915,50
AutomContra800	220 - 529,80	445,50 - 927,50
AutomIpsoa100	343,50 - 622,95	545,25 - 910
AutomIpsoa800	224,50 - 494,15	404 - 717,70
VoluntInvalida300	287,50 - 630,50	509 - 1192
VoluntInvalida800	255,50 - 653	471 - 1226
Volunvalida300	270 - 495,50	457 - 857,50
VoluntValida800	232 - 456,50	384,50 - 757

TABELA III

VALORES MÍNIMOS E MÁXIMOS DOS ATRIBUTOS DAS COMUNIDADES IDENTIFICADAS: UMA ANÁLISE COMPARATIVA. MÉTRICAS EXPRESSAS EM MILISSEGUNDOS

TABELA IV
RESULTADOS DOS TESTES

Atributo	t	p
AutomContra100	0.056478	4.741041×10^{-6}
AutomContra800	5.115841	2.020524×10^{-5}
AutomIpsoa100	6.321734	7.734352×10^{-7}
AutomIpsoa800	6.905452	1.662291×10^{-7}
VoluntInvalida300	5.847432	2.761150×10^{-6}
VoluntInvalida800	4.124490	3.006908×10^{-4}
Volvalida300	5.521771	6.676710×10^{-6}
VoluntValida800	5.321319	1.152786×10^{-5}

cional, contribuindo assim para a caracterização do espectro do TDAH.

Os resultados revelaram que as duas abordagens foram capazes de identificar dois grupos distintos de pacientes com TDAH: um grupo moderado e um grupo severo. Esses resultados sugerem que pacientes com TDAH podem apresentar diferenças comportamentais em relação à capacidade de processamento da atenção. Essas descobertas são de grande importância para compreender a variação interindividual no desempenho atencional e podem contribuir para o desenvolvimento de abordagens de tratamento personalizadas para pacientes com TDAH.

Trabalhos futuros podem enriquecer nossa compreensão do TDAH com base nos resultados deste estudo. A inclusão de um grupo controle de pacientes sem TDAH permitiria validar as abordagens utilizadas e verificar a representatividade dos atributos considerados. Além disso, uma análise mais aprofundada das medidas de centralidade, intermediação e grau poderia identificar indicadores mais precisos de prejuízo atencional. Uma comparação das redes de orientação da atenção antes e após tratamento em indivíduos com TDAH poderia avaliar possíveis mudanças na organização da rede e sua relação com a melhoria dos sintomas.

Essas investigações adicionais certamente contribuirão para o avanço do conhecimento sobre o TDAH e para aprimorar a compreensão dos mecanismos subjacentes ao transtorno.

AGRADECIMENTO

Profunda gratidão à Universidade Presbiteriana Mackenzie e à coordenação do programa de pós-graduação em Engenharia Elétrica e Computação pelo inestimável suporte oferecido durante todo o processo de desenvolvimento desta pesquisa.

REFERENCES

- [1] Sah, P.; Singh, L.O.; Clauzet, A.; Bansal, S. Exploring community structure in biological networks with random graphs. *BMC Bioinformatics* **2014**, *15*(1), 220. <https://doi.org/10.1186/1471-2105-15-220>.
- [2] Yassine, S.; Kadry, S.; Sicilia, M.-A. Application of community detection algorithms on learning networks. The case of Khan Academy repository. *Computer Applications in Engineering Education* **2020**, *29*(2), 411–424. <https://doi.org/10.1002/cae.22212>.
- [3] Betzel, R. Community detection in network neuroscience. *MDPI Complexity* **2020**, *2020*(1), 1–15. <https://doi.org/10.3390/complexity5010001>.
- [4] Boulic, R.; Renault, O. 3D Hierarchies for Animation. In *New Trends in Animation and Visualization*; Magnenat-Thalmann, N., Thalmann, D., Eds.; John Wiley Sons Ltd.: 1991; pp. 1–10.
- [5] Smith, A.; Jones, B. On the Complexity of Computing. In *Advances in Computer Science*; Smith-Jones, A.B., Ed.; Publishing Press: 1999; pp. 555–566.
- [6] Jacomy, M.; Venturini, T.; Heymann, S.; Bastian, M. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS One* **2014**, *9*(6), e98679. <https://doi.org/10.1371/journal.pone.0098679>.
- [7] American Psychiatric Association. *Manual diagnóstico e estatístico de transtornos mentais: DSM-5*, 5th ed.; Autor: Washington, DC, USA, 2013.
- [8] Piaget, J. *A linguagem e o pensamento da criança*, 7th ed.; Martins Fontes: São Paulo, SP, Brazil, 1999.
- [9] Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*(12), 7821–7826. <https://doi.org/10.1073/pnas.122653799>.
- [10] Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.: Theory Exp.* **2008**, *2008*(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/p10008>.
- [11] Al-Ansari, A. Measurement of Impairment among Children with Attention Deficit Hyperactivity Disorder as Part of Evaluating Treatment Outcome. *Sultan Qaboos Univ. Med. J.* **2013**, *13*, 296–300. <https://doi.org/10.12816/0003237>.
- [12] Leite, H.A.; Tuleski, S.C. Psicologia Histórico-Cultural e desenvolvimento da atenção voluntária: novo entendimento para o TDAH. *Rev. FapUNIFESP* **2011**, *15*(1), 111–119. <https://doi.org/10.1590/s1413-85572011000100012>.
- [13] Araujo, R.R.; Carreiro, L.R.R. Orientação voluntária e automática da atenção e indicadores de desatenção e hiperatividade em adultos. *Avaliação Psicológica* **2009**, *8*, 325–336. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712009000300005&nrm=isoURL.
- [14] Cuvillier, A. *ABC de Psicologia*, 1935; Companhia Editorial Nacional: São Paulo, Brazil.
- [15] Newman, M.E.J.; Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **2004**, *69*(2), 026113. <https://doi.org/10.1103/physreve.69.026113>.
- [16] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **2002**, *12*(99), 7821–7826. <https://doi.org/10.1073/pnas.122653799>.
- [17] Newman, M.E.J. A measure of betweenness centrality based on random walks. *Soc. Networks* **2005**, *27*(1), 39–54. <https://doi.org/10.1016/j.socnet.2004.11.009>.
- [18] Newman, M. *Networks: An Introduction*; Oxford University Press, Inc.: USA, 2010. <https://doi.org/10.5555/1809753>.
- [19] Mariani, M.; Rocha Reis Lellis, V.; Novaes, R.; Schwartzman, J.; Triguero Veloz Teixeira, M.C.; Carreiro, L.R. Voluntary and Automatic Orienting of Attention in Children with Attention Deficit Hyperactivity Disorder. *Arch. Clin. Psychiatry* **2021**, *48*, <https://doi.org/10.15761/0101-60830000000298>.
- [20] Dalgalarondo, P. *Psicopatologia e Semiologia dos Transtornos Mentais*; Artmed Editora, 2018. <https://books.google.com.br/books?id=8R5vDwAAQBAJURL>.
- [21] Mendes, Y.; Almeida, E.; Cabral, R.; Queiroz, F. Uso de Redes Complexas para Classificação de Doenças Pulmonares Intersticiais em Imagens de Tomografia Computadorizada. In *Proceedings of the Anais Principais do Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2020)*; Sociedade Brasileira de Computação - SBC: Porto Alegre, RS, Brasil, 2020; <https://doi.org/10.5753/sbcas.2020.11528>.

- [22] Kido, G.S.; Igawa, R.A.; Barbon Jr., S. Topic Modeling based on Louvain method in Online Social Networks. In Proceedings of the Anais do Simpósio Brasileiro de Sistemas de Informação (SBSI); Sociedade Brasileira de Computação: Porto Alegre, RS, Brasil, 2016; <https://doi.org/10.5753/sbsi.2016.5982doi>.
- [23] Almeida, T.; Nakamura, F.; Nakamura, E. Uma Abordagem Baseada em Redes Complexas para Análise de Depoimentos Legais. In Proceedings of the Anais do XXXVI Concurso de Trabalhos de Iniciação Científica da SBC; Sociedade Brasileira de Computação - SBC: Porto Alegre, RS, Brasil, 2017; <https://sol.sbc.org.br/index.php/ctic/article/view/3240URL>.
- [24] Caminha, C.; Furtado, V.; Pinheiro, V.; Ponte, C. Detecção de comunidades em redes complexas para identificar gargalos e desperdício de recursos em sistemas de ônibus. *Anais do VI Brazilian Workshop on Social Network Analysis and Mining 2017, São Paulo*, ISSN 2595-6094, DOI: 10.5753/brasnam.2017.3262, URL: <https://sol.sbc.org.br/index.php/brasnam/article/view/3262>.
- [25] Freeman, L.C. Centrality in social networks conceptual clarification. *Social Networks* **1978**, *1*(3), 215-239. ISSN 0378-8733. DOI: [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7). URL: <https://www.sciencedirect.com/science/article/pii/0378873378900217>.
- [26] Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media*, 2009. Disponível em: <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>.
- [27] Freeman, L.C. Centrality in social networks conceptual clarification. *Social Networks*, **1**(3), 215-239, 1978. ISSN 0378-8733. Disponível em: <https://www.sciencedirect.com/science/article/pii/0378873378900217>.
- [28] Lloyd-Smith, P. Kuhn-Tucker and Multiple Discrete-Continuous Extreme Value Model Estimation and Simulation in R: The rmdcev Package. *The R Journal*, **12**(2), 251-265, 2021. DOI: 10.32614/RJ-2021-015. Disponível em: <https://doi.org/10.32614/RJ-2021-015>.
- [29] Curtis, A.; Smith, T.; Ziganshin, B.; Elefteriades, J. The Mystery of the Z-Score. *AORTA*, **04**(04), 124-130, 2016. DOI: 10.12945/j.aorta.2016.16.014. Disponível em: <https://doi.org/10.12945/j.aorta.2016.16.014>.
- [30] Jones, P.J.; Mair, P.; McNally, R.J. Visualizing Psychological Networks: A Tutorial in R. *Frontiers in Psychology*, **9**, 2018. DOI: 10.3389/fpsyg.2018.01742. Disponível em: <https://doi.org/10.3389/fpsyg.2018.01742>.
- [31] Sato, J.R.; Vidal, M.C.; Santos, S.d.S.; Massirer, K.B.; Fujita, A. Complex network measures in autism spectrum disorders. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **15**(2), 581-587, 2018. Disponível em: <https://ieeexplore.ieee.org/document/8270619>.
- [32] J. B. MacQueen, "Some Methods for Classification and Analysis of MultiVariate Observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. Le Cam and J. Neyman (eds.), vol. 1, pp. 281-297, University of California Press, 1967.
- [33] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129-137, Mar. 1982. doi: 10.1109/tit.1982.1056489.
- [34] Xu, X., Lv, W., Zhang, B., ShuaiPeng, Z., Wei, W., Li, Y. (2021). A Novel Emerging Topic Identification and Evolution Discovery Method On Time-evolving And Heterogeneous Online Social Networks. *Complexity*, (2021), 1-14. <https://doi.org/10.1155/2021/8859225>
- [35] Ezequiel López-Rubio and Esteban J. Palomo and Francisco Ortega-Zamorano.(2018). Unsupervised learning by cluster quality optimization. *Information Sciences*, 436-437, 31-35. <https://doi.org/10.1016/j.ins.2018.01.007>
- [36] Nurhidayati, N. M., Khasanah, N. N. (2021). Penggunaan Metode K-means Cluster Untuk Mengklasifikasikan Kemampuan 4c Mahasiswa. *JIMT*, **2**(18), 160-169. <https://doi.org/10.22487/2540766x.2021.v18.i2.15615>
- [37] Umargono, E.; Suseno, J. and S. K., V. (2020). K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median. In Proceedings of the International Conferences on Information System and Technology - CONRIST; ISBN 978-989-758-453-4, SciTePress, pages 234-240. DOI: 10.5220/0009908402340240
- [38] V. V. B., Chen, J. I., Ahmad, B. H. B., Khan, Z. F. (2022). Guest Editorial. *IJIUS*, **1**(10), 1-2. <https://doi.org/10.1108/ijius-01-2022-116>
- [39] Arnarsson, Í. Ö., Frost, O., Gustavsson, E., Jirstrand, M., Malmqvist, J. (2021). Natural Language Processing Methods For Knowledge Management—applying Document Clustering For Fast Search and Grouping Of Engineering Documents. *Concurrent Engineering*, **2**(29), 142-152. <https://doi.org/10.1177/1063293x20982973>
- [40] Al-sharoha, E., Ababneh, B. M., Al-khassaweneh, M. (2021). Robust Community Detection In Graphs. *IEEE Access*, (9), 118757-118770. <https://doi.org/10.1109/access.2021.3105692>
- [41] Student. (1908). The probable error of a mean. *Biometrika*, 1-25. JSTOR.