

# Análise de dados em instâncias do problema do corte máximo em grafos

Jailon W. B. Oliveira da Silva<sup>1</sup>, Carlos V. Dantas Araújo<sup>1</sup>, Pablo L. Braga Soares<sup>1</sup>

<sup>1</sup>Universidade Federal do Ceará – Campus Russas  
Avenida Felipe Santiago – Nº 411 62.900-624, Russas/CE, Brasil

{williambrunocc, carlosvinicio}@alu.ufc.br, pablo.soares@ufc.br

**Abstract.** *The maximum cut problem is one of the central problems in combinatorial optimization, with various practical applications. This work aims to conduct an exploratory analysis of graph instances for this problem using the Biq Mac Library. To this end, graph characteristics were extracted and analyzed using statistical and visualization techniques to identify patterns and anomalies. The results show significant correlations between some metrics and identify outliers that may influence the problem's solution. It is concluded that the analysis provides a solid understanding of the instances and makes the development of machine learning predictive models for the maximum cut problem feasible.*

**Resumo.** *O problema do corte máximo é um dos problemas centrais na otimização combinatória, com diversas aplicações práticas. Este trabalho tem como objetivo realizar uma análise exploratória de instâncias de grafos para esse problema, utilizando a Biq Mac Library. Para isso, características dos grafos foram extraídas e analisadas por meio de técnicas estatísticas e de visualização para identificar padrões e anomalias. Os resultados mostram correlações significativas entre algumas métricas e identificam outliers que podem influenciar a solução do problema. Conclui-se que a análise fornece uma base de entendimento sólida a cerca das instâncias e torna viável o desenvolvimento de modelos preditivos de aprendizado de máquina para o problema do corte máximo.*

## 1. Introdução

Segundo [Boros and Hammer 1991], o Problema do Corte Máximo é um dos problemas centrais da otimização combinatória. Ele consiste em encontrar um subconjunto  $S$  dos vértices que define um corte de peso máximo em um determinado grafo ponderado. Mais especificamente, dado um grafo  $G = (V, E)$ , onde  $V$  representa o conjunto de vértices e  $E$  o conjunto de arestas, um corte é o conjunto de arestas que possuem exatamente um ponto terminal em  $S$ . O objetivo é encontrar o subconjunto  $S$  dos vértices tal que a soma dos pesos das arestas no corte seja máxima.

A figura 1 (a) apresenta uma instância do problema do corte máximo, onde o grafo  $G = (V, E)$  possui  $|V| = 5$  vértices e  $|E| = 6$  arestas ponderadas. Os vértices são numerados de 1 a 5, e as arestas possuem pesos associados. Já a figura 1 (b) apresenta uma solução ótima para este problema. O conjunto de vértices considerado é  $S = \{1, 3, 5\}$ . O corte resultante, que é o conjunto de arestas que têm exatamente um de seus vértices em  $S$ , é dado por  $(1, 2), (1, 4), (2, 3), (2, 5), (3, 5), (4, 3)$ . Portanto, o peso total do corte é dado pelo somatório dos pesos das arestas de corte, resultando em 27.

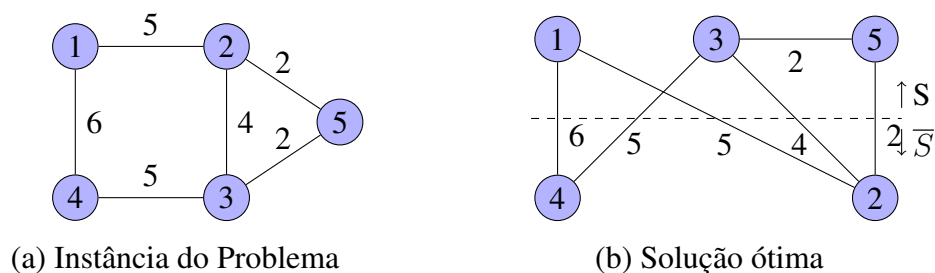


Figura 1. Instância do Problema do Corte Máximo

Dada a relevância prática do Problema do Corte Máximo, diversos algoritmos têm surgido na literatura, especialmente porque ele faz parte da classe de problemas de otimização combinatória, sendo que muitos problemas surgem diretamente em seu formato, enquanto outros podem ser reduzidos a ele. Essa versatilidade proporciona ao Problema do Corte Máximo, uma ampla gama de aplicações, abrangendo áreas como análise de redes sociais, agrupamento de dados, segmentação de imagens e física estatística [Agrawal et al. 2003, Otterbach et al. 2017, Sousa et al. 2013, Barahona et al. 1988].

Neste contexto, uma análise exploratória das instâncias do Problema do Corte Máximo presentes na Biq Mac Library, apresentados no trabalho de [Wiegele 2007], pode fornecer informações valiosas sobre as características estruturais dessas instâncias. Esta biblioteca contém uma coleção diversificada de instâncias que variam em tamanho, de 20 a 500 vértices, e em complexidade, com diferentes propriedades como número de triângulos, densidade, peso das arestas, comprimento médio do caminho entre vértices, centralidade e coeficiente de cluster. A análise dessas instâncias pode revelar padrões e tendências que influenciam a escolha e o desenvolvimento de algoritmos de solução.

Além disso, trabalhos anteriores também se concentraram na análise exploratória de dados, ressaltando a importância dessa abordagem para a compreensão de problemas complexos. Por exemplo, o estudo de [Jebb et al. 2016] demonstrou como a análise exploratória pode revelar padrões ocultos em grandes conjuntos de dados, trazendo *insights* fundamentais para a modelagem preditiva. Da mesma forma, [Ferreira 2022] aplicou técnicas exploratórias para analisar dados multivariados, destacando a identificação de anomalias e a estruturação de dados para facilitar a interpretação dos resultados. Outro exemplo é o trabalho de [Wongsuphasawat et al. 2019], que utilizou entrevistas com analistas de dados para entender os desafios e processos da análise exploratória, identificando correlações significativas e *outliers* que poderiam influenciar os resultados dos experimentos.

A diversidade presente nas instâncias do Biq Mac Library representa um desafio significativo para modelos de aprendizado de máquina, pois a variabilidade em complexidade e estrutura pode dificultar a capacidade de generalização dos modelos. Portanto, uma análise detalhada dessas instâncias é crucial para entender melhor suas características e para desenvolver algoritmos mais robustos e eficientes.

O restante deste trabalho está estruturado da seguinte forma: na Seção 2 são apresentados os conceitos e definições necessários de teoria dos grafos, incluindo características estruturais, conectividade, distâncias, centralidade e aglomeração. A Seção 3 descreve a criação e estruturação da base de dados das instâncias do problema do corte máximo.

A análise exploratória das instâncias é discutida na Seção 4, onde são apresentadas as distribuições das métricas, detecção de outliers e análise de centralidade. Finalmente, a Seção 5 conclui o trabalho, destacando os principais achados e sugestões para pesquisas futuras.

## 2. Conceitos, definições e medidas de grafos

Nesta seção, são apresentados alguns conceitos, definições e medidas obtidas dos grafos que serão analisadas neste estudo como características.

### 2.1. Estrutura de um grafo

- **Número de vértices:** A quantidade total de vértices em um grafo é uma medida fundamental que indica a quantidade de pontos de conexão no grafo [Diestel 2017];
- **Número de arestas e grau do vértice:** O número total de arestas em um grafo representa a quantidade de conexões entre os vértices, já o grau de um vértice é definido como o número de arestas conectadas a ele [Shrimali and Shah 2020];
- **Peso nas arestas:** O peso das arestas em um grafo é um valor atribuído a cada aresta que pode representar métricas relevantes como distância ou custo [Diestel 2017].

### 2.2. Conectividade e Componentes

- **Subgrafo:** Um subgrafo é um conjunto de vértices e arestas que pertence a um grafo maior [Diestel 2017];
- **Conexo:** Um grafo é considerado conexo se existe um caminho entre quaisquer dois vértices [Diestel 2017];
- **Número de Componentes Conectados:** O número de componentes conectados é a contagem dos subgrafos maximamente conexos dentro de um grafo [Shrimali and Shah 2020].

### 2.3. Distâncias e Caminhos

- **Diâmetro:** O diâmetro de um grafo é o maior valor de distância mínima entre quaisquer dois vértices em um grafo conexo [Shrimali and Shah 2020];
- **Excentricidade:** A excentricidade de um vértice é a maior distância mínima entre esse vértice e qualquer outro vértice no grafo [Diestel 2017];
- **Menores Caminhos:** Representa os comprimentos dos caminhos mais curtos entre todos os pares de vértices em um grafo conexo [Shrimali and Shah 2020];
- **Raio:** O raio de um grafo é a menor excentricidade entre todos os vértices [Diestel 2017].

### 2.4. Centralidade

- **Centralidade de Grau:** A centralidade de grau de um vértice é definida pelo número de arestas conectadas a ele [Latora et al. 2017];
- **Centralidade de Intermediação:** A centralidade de intermediação de um vértice é determinada pelo número de vezes que ele atua como intermediário ao longo dos caminhos mais curtos entre outros dois vértices [Latora et al. 2017].

## 2.5. Aglomeração e Clusterização

- **Número de Triângulos:** Um triângulo é um subgrafo de três vértices onde cada vértice está conectado a ambos os outros vértices [Diestel 2017];
- **Coefficiente de Aglomeração:** O coeficiente de aglomeração de um vértice é a proporção de triângulos que passam por esse vértice em relação ao número máximo possível de triângulos [Barabási 2016];
- **Transitividade:** A transitividade do grafo é a razão entre o número de triângulos e o número de trios conectados (três vértices que são conectados por pelo menos duas arestas) [Barabási 2016];
- **Densidade:** A densidade de um grafo é a razão entre o número de arestas e o número máximo possível de arestas [Shrimali and Shah 2020].

## 2.6. Visualização de dados

- **KDE (*Kernel Density Estimation*):** é um método para estimar a função de densidade de probabilidade dos dados de entrada, capturando suas características de distribuição de forma suave e contínua [Lin et al. 2020].
- **Box-Plot:** Método para demonstrar a distribuição de um conjunto de dados resumindo através de quartis e destacando potenciais outliers [Almeida and Martins 2023].

## 3. Base de dados

A abordagem central na construção da base de dados consiste em unir as instâncias uma abaixo da outra em um *dataframe*, organizando as características estruturais dos grafos em colunas, tais como número de vértices, número de arestas, número de triângulos, peso médio das arestas por vértices e densidade do grafo. Além dessas métricas globais, serão registrados os valores máximo, mínimo e médio observados para características específicas dos vértices, como centralidade e coeficiente de aglomeração. Essas medidas são cruciais para entender não apenas a estrutura geral dos grafos, mas também as propriedades individuais dos vértices, que podem influenciar decisões em análise de redes e otimização.

Para a criação da base de dados, foram utilizadas diversas bibliotecas e ferramentas da linguagem de programação *Python3*. A implementação dos grafos e a determinação das características de cada grafo foram realizadas por meio da biblioteca *NetworkX*. Para a criação e manipulação da base de dados, utilizou-se a biblioteca *Pandas*. Adicionalmente, as bibliotecas *Seaborn* e *Matplotlib.pyplot* deram suporte à análise exploratória dos grafos, permitindo a criação de visualizações gráficas. Toda a codificação feita para a análise exploratória foi conduzida em notebooks *Jupyter*, e o código da EDA e a base de dados completa está disponível no seguinte link: <https://shre.ink/Dt0A>.

## 4. Análise exploratória

Nesta seção, são apresentadas as análises feitas das características extraídas das instâncias dos grafos. A Seção 4.2 reúne gráficos de densidade para a distribuição de média da centralidade de grau, transitividade e densidade das instâncias. A Seção 4.3 aborda a detecção de *outliers* utilizando o método do intervalo interquartil (IQR), destacando a proporção de valores extremos para diferentes características dos grafos. A Seção 4.4

examina a distribuição da centralidade de intermediação, com gráficos que mostram a dispersão e a importância relativa dos vértices na conectividade dos grafos. Finalmente, a Seção 4.1 apresenta estatísticas gerais das instâncias, fornecendo uma visão abrangente das propriedades dos grafos analisados.

#### 4.1. Estatísticas gerais das instâncias

Nesta seção, são discutidas as estatísticas descritivas das propriedades das 326 instâncias de grafos analisadas, conforme apresentadas na Tabela 1. Essas estatísticas incluem medidas de tendência central, dispersão e extremos, proporcionando uma visão abrangente das características estruturais dos grafos.

**Tabela 1. Estatísticas descritivas das propriedades do grafo**

Características	mean	std	min	25%	50%	75%	max
Número de vértices	$1.5 \times 10^2$	$1.0 \times 10^2$	$2.0 \times 10^1$	$1.0 \times 10^2$	$1.0 \times 10^2$	$1.5 \times 10^2$	$5.0 \times 10^2$
Número de arestas	$5.8 \times 10^3$	$1.2 \times 10^4$	$1.1 \times 10^2$	$6.5 \times 10^2$	$3.1 \times 10^3$	$5.0 \times 10^3$	$1.2 \times 10^5$
Peso médio de aresta por vértice	$1.1 \times 10^3$	$7.1 \times 10^3$	$-3.5 \times 10^4$	$-1.1 \times 10^1$	$1.0 \times 10^0$	$4.8 \times 10^1$	$5.4 \times 10^4$
Número de componentes conectados	$1.0 \times 10^0$	$2.0 \times 10^{-1}$	$1.0 \times 10^0$	$1.0 \times 10^0$	$1.0 \times 10^0$	$1.0 \times 10^0$	$4.0 \times 10^0$
Diâmetro	$3.0 \times 10^0$	$2.5 \times 10^0$	$1.0 \times 10^0$	$2.0 \times 10^0$	$2.0 \times 10^0$	$3.0 \times 10^0$	$2.0 \times 10^1$
Excentricidade média	$2.7 \times 10^0$	$2.5 \times 10^0$	$1.0 \times 10^0$	$2.0 \times 10^0$	$2.0 \times 10^0$	$3.0 \times 10^0$	$2.0 \times 10^1$
Raio	$2.6 \times 10^0$	$2.5 \times 10^0$	$1.0 \times 10^0$	$2.0 \times 10^0$	$2.0 \times 10^0$	$3.0 \times 10^0$	$2.0 \times 10^1$
Densidade	$4.9 \times 10^{-1}$	$3.7 \times 10^{-1}$	$1.0 \times 10^{-2}$	$1.0 \times 10^{-1}$	$5.0 \times 10^{-1}$	$9.0 \times 10^{-1}$	$1.1 \times 10^0$
Média centralidade	$4.9 \times 10^{-1}$	$3.7 \times 10^{-1}$	$1.0 \times 10^{-2}$	$1.0 \times 10^{-1}$	$5.0 \times 10^{-1}$	$9.0 \times 10^{-1}$	$1.1 \times 10^0$
Máxima centralidade	$5.6 \times 10^{-1}$	$3.6 \times 10^{-1}$	$1.0 \times 10^{-2}$	$1.8 \times 10^{-1}$	$6.0 \times 10^{-1}$	$9.6 \times 10^{-1}$	$1.1 \times 10^0$
Mínima centralidade	$4.2 \times 10^{-1}$	$3.8 \times 10^{-1}$	$0.0 \times 10^0$	$4.4 \times 10^{-2}$	$3.3 \times 10^{-1}$	$8.2 \times 10^{-1}$	$1.1 \times 10^0$
Média centralidade intermediação	$3.8 \times 10^{-1}$	$1.9 \times 10^{-1}$	$5.1 \times 10^{-3}$	$3.9 \times 10^{-1}$	$4.9 \times 10^{-1}$	$5.0 \times 10^{-1}$	$5.0 \times 10^{-1}$
Máxima centralidade intermediação	$7.1 \times 10^{-1}$	$3.4 \times 10^{-1}$	$7.4 \times 10^{-3}$	$6.7 \times 10^{-1}$	$8.7 \times 10^{-1}$	$9.2 \times 10^{-1}$	$1.1 \times 10^0$
Mínima centralidade intermediação	$1.9 \times 10^{-2}$	$2.4 \times 10^{-2}$	$0.0 \times 10^0$	$0.0 \times 10^0$	$5.6 \times 10^{-3}$	$3.6 \times 10^{-2}$	$1.2 \times 10^{-1}$
Média coeficiente de aglomeração	$7.8 \times 10^{-2}$	$1.7 \times 10^{-1}$	$-4.5 \times 10^{-2}$	$-2.5 \times 10^{-4}$	$5.4 \times 10^{-5}$	$4.3 \times 10^{-3}$	$5.1 \times 10^{-1}$
Máximo coeficiente de aglomeração	$1.3 \times 10^{-1}$	$1.8 \times 10^{-1}$	$0.0 \times 10^0$	$8.6 \times 10^{-3}$	$2.5 \times 10^{-2}$	$1.6 \times 10^{-1}$	$5.8 \times 10^{-1}$
Mínimo coeficiente de aglomeração	$2.8 \times 10^{-2}$	$1.9 \times 10^{-1}$	$-7.2 \times 10^{-1}$	$-3.1 \times 10^{-2}$	$-1.0 \times 10^{-2}$	$0.0 \times 10^0$	$4.8 \times 10^{-1}$
Número de triângulos	$3.0 \times 10^5$	$1.4 \times 10^6$	$0.0 \times 10^0$	$1.9 \times 10^2$	$2.0 \times 10^4$	$1.2 \times 10^5$	$2.0 \times 10^7$
Transitividade	$4.8 \times 10^{-1}$	$3.7 \times 10^{-1}$	$0.0 \times 10^0$	$1.0 \times 10^{-1}$	$5.0 \times 10^{-1}$	$9.0 \times 10^{-1}$	$1.0 \times 10^0$

A Tabela 1 apresenta as estatísticas descritivas das propriedades dos grafos presentes em todas as 326 instâncias estudadas. As instâncias apresentam um número médio de vértices de 150, com um desvio padrão de 100. O valor mínimo observado é de 20 vértices, enquanto o máximo é de 500. Isso indica uma grande variabilidade no tamanho dos grafos, abrangendo desde grafos pequenos até grandes redes complexas. O número médio de arestas é de 5.800, com um desvio padrão considerável de 12.000, refletindo a diversidade estrutural das instâncias. O número mínimo de arestas é 110, e o máximo é 124.000, demonstrando a presença tanto de grafos esparsos quanto densamente conectados.

A média do peso das arestas por vértice é 1.100, com um desvio padrão de 7.100. A grande variação, com valores negativos e um máximo de 54.000, sugere uma diversidade significativa nas métricas de peso atribuídas às arestas. As instâncias apresentam um número médio de componentes conectados de 1, com um desvio padrão de 0.2, indicando que a maioria dos grafos é altamente conectada, com poucos subgrafos desconectados. O diâmetro dos grafos, apresenta presença de um valor incomum, onde a média é 3, isto é, embora a maioria dos grafos tenha um diâmetro de 2 a 5, alguns têm diâmetros muito maiores, como 20, o que é fora do comum em comparação com a maioria dos casos analisados.

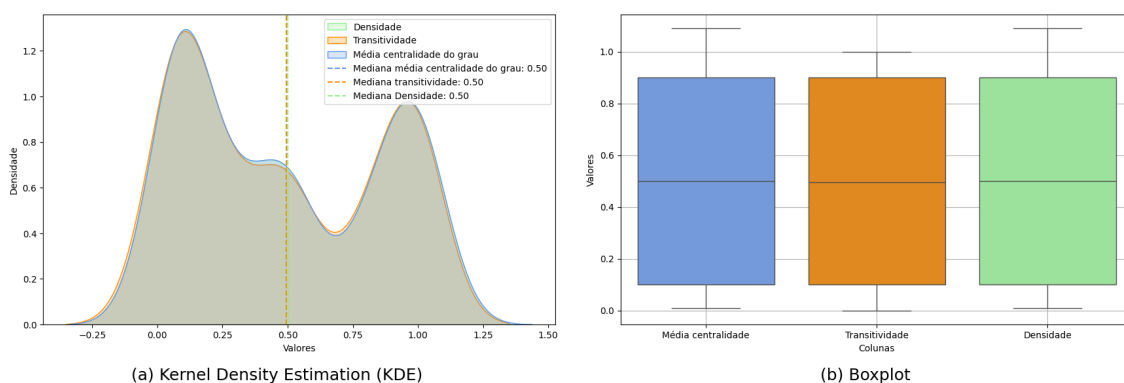
A excentricidade média é 2.7, com um desvio padrão de 2.5, variando de 1 a 20. Esta métrica reflete a variabilidade nas distâncias máximas de qualquer vértice para os demais vértices no grafo. O raio médio também é 2.6, sugerindo a presença de grafos

tanto compactos quanto mais espalhados. A densidade média dos grafos é 0.49, com um desvio padrão de 0.37. Os valores variam de 0.01 a 1.1, sugerindo uma ampla gama de densidades, desde grafos muito esparsos até grafos completos. Note que a densidade de 1.1 sugere a presença de laços em um grafo, em pelo menos uma instância das 326. Além disso, a centralidade média segue a mesma tendência da densidade média, indicando a variação na importância relativa dos vértices dentro das redes.

O número médio de triângulos é 300.000, com um desvio padrão de 1.400.000, variando de 0 a 20.000.000, indicando uma ampla variação na presença de subgrafos que formam triângulos. Essas estatísticas descritivas fornecem uma visão abrangente das propriedades estruturais dos grafos analisados e destacam a diversidade e complexidade das instâncias do problema do corte máximo.

## 4.2. Transitividade, centralidade de grau e densidade

As colunas Densidade, Transitividade e Média da centralidade do grau, conforme a Tabela 1, apesar de representarem características distintas, possuem estatísticas básicas muito semelhantes. Por isso, foram selecionadas para análise conjunta, com o objetivo de compreender como essas características dos grafos influenciam não apenas de forma individual, mas também em conjunto.



**Figura 2. Distribuição de Média da Centralidade de Grau, Transitividade e Densidade**

A Figura 2 (a) revela que a média das centralidades de grau dos vértices e a transitividade exibem um KDE extremamente similar, com a densidade do grafo também mostrando uma distribuição idêntica entre essas métricas. A Figura 2 (b) mostra que as estatísticas básicas, como primeiro, segundo e terceiro quartil, são também bastante semelhantes, com a densidade do grafo mantendo uma consistência equivalente. Essa consistência sugere uma correlação estrutural significativa, sendo confirmada pela correlação de Pearson, que apresentou uma pontuação máxima de 1.0. Isso indica que instâncias do problema do corte máximo com maior transitividade tendem a apresentar uma média mais alta das centralidades de grau de seus vértices, assim como uma densidade similar.

Por último, o gráfico KDE evidencia dois picos de densidade, sugerindo a presença de duas regiões de alta densidade de dados, indicando duas populações distintas. Além disso, a distribuição similar dessas métricas pode influenciar diretamente estratégias de otimização. Por exemplo, identificar regiões com vértices de alta centralidade de grau em grafos altamente transitivos e densos pode facilitar a identificação de cortes que maximizem a separação entre os dois *clusters* de vértices.

### 4.3. Detecção de outliers

A Figura 3 mostra as porcentagens de *outliers* nas colunas da base de dados. Nota-se que certas propriedades dos grafos não seguem um padrão normal, conforme definido pelo Intervalo Interquartil (IQR). Cada porcentagem indica a proporção de valores considerados extremos em relação ao restante da distribuição de cada coluna [Moore et al. 2009]. Por exemplo, as porcentagens das colunas máximo, mínimo e médio do coeficiente de aglomeração sugerem que uma parte significativa dos valores dessa propriedade do grafo está fora do intervalo intermediário entre o primeiro e o terceiro quartil da distribuição dos dados.

As porcentagens dessas três colunas indicam que 31,90%, 24,85% e 15,03% das 326 instâncias da base de dados apresentam valores considerados *outliers* em relação aos atributos específicos mencionados, refletindo uma alta variação substancial e características únicas nos dados analisados. Essa alta variação é ilustrada pelos gráficos de KDE e *boxplot* dessas três colunas, presentes na Figura 4, que evidenciam a distribuição desigual e a presença de valores extremos. Essas observações podem ser cruciais para estratégias de clusterização mais precisas e para a identificação de grupos distintos dentro do conjunto de dados.

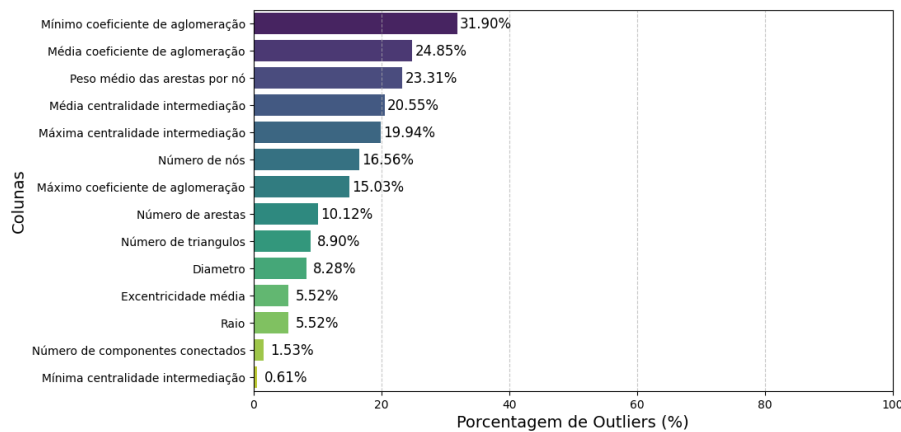


Figura 3. Visualização da Percentagem de Outliers com Método IQR

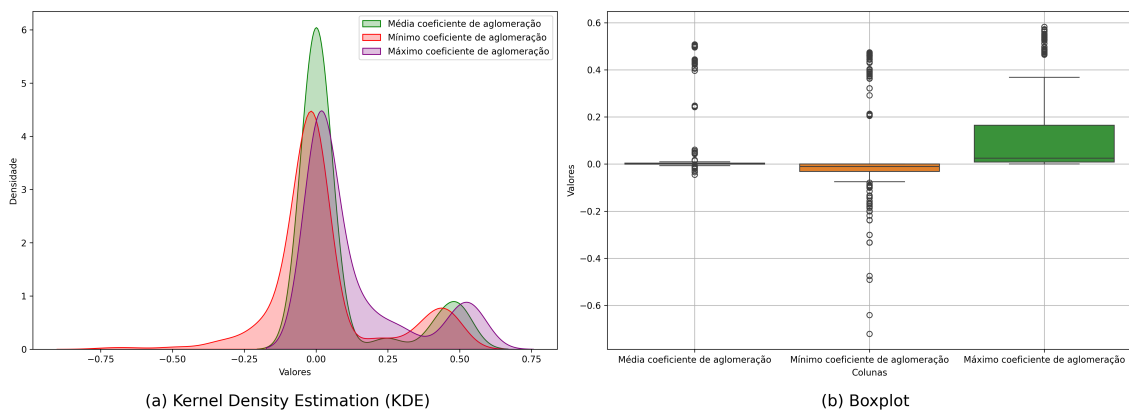


Figura 4. Análise do Coeficiente de Aglomeração

#### 4.4. Distribuição de Centralidade de Intermediação

A Figura 5 fornece uma visão sobre a centralidade de intermediação das instâncias, destacando a distribuição dos valores mínimos, médios e máximos dessas métricas. Análisar essa distribuição é crucial para encontrar detalhes importantes da estrutura e conectividade dos grafos. Observando a mínima centralidade, é perceptível que os valores são altamente concentrados em torno de zero, esse agrupamento indica que muitos vértices possuem uma influência quase nula na conectividade geral dos grafos.

A média da centralidade apresenta uma distribuição mais espalhada, com um pico próximo a 0.5, indicando uma diversidade maior na importância dos vértices em termos de conexão. Com uma média de 0.49, há uma quantidade significativa de vértices com influência moderada, detalhe crucial para a manutenção da conectividade em várias instâncias. Em contraste, a máxima centralidade de intermediação mostra uma distribuição ampla e menos concentrada, com um pico em torno de 0.87. Isso mostra a presença de vértices com alta centralidade de intermediação, que são essenciais para a conectividade da rede, são os chamados “vértices-chave”.

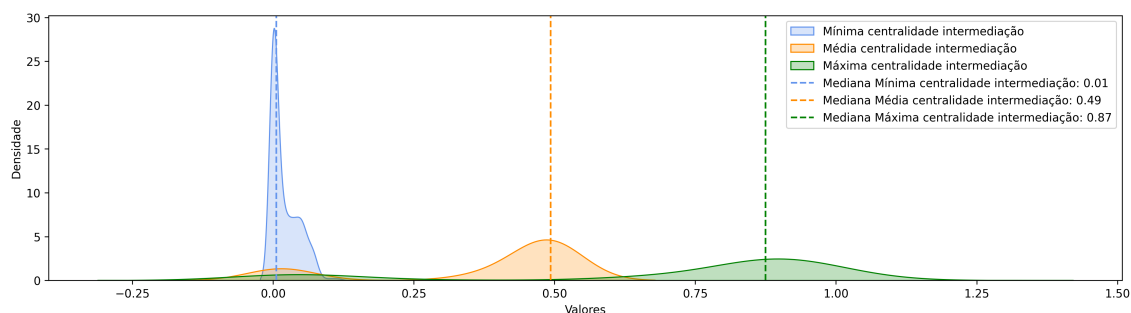
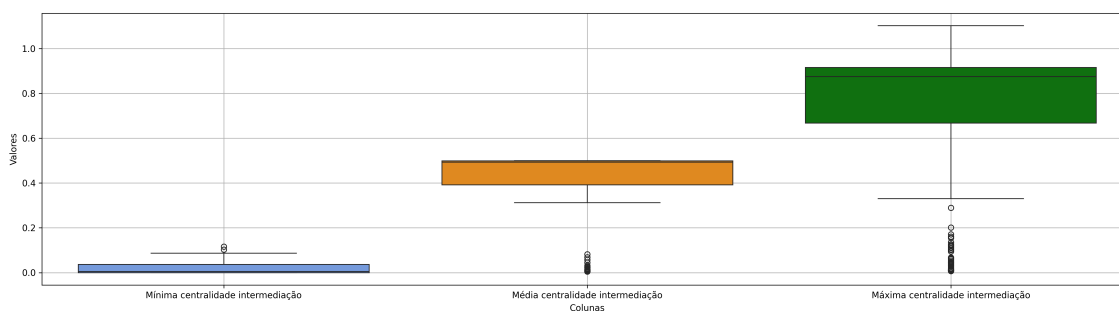


Figura 5. Centralidade de Intermediação

A Figura 6 representa um *boxplot* que fornece uma visão complementar para os valores de centralidade de intermediação. O *boxplot* da mínima centralidade de intermediação mostra baixa variação, com a maioria dos valores próximos de zero, embora haja alguns *outliers* que indicam a presença de vértices com centralidade um pouco superior, mas ainda baixa. Esses vértices, apesar de não serem críticos, têm um impacto maior do que a maioria. Em contraste, o *boxplot* da média e máxima de centralidade mostra variações e *outliers* consideravelmente maiores, confirmando a presença de vértices extremamente importantes, como citado anteriormente.

Em termos práticos, a diversidade na centralidade de intermediação pode ser percebida pela considerável variação de distribuições, o que indica uma estrutura complexa nos grafos, com exemplos bem claros de vértices que desempenham papéis importantes na conectividade e outros que não possuem tanta relevância em termos de conexão. Ao identificar e reforçar vértices com alta centralidade de intermediação, é possível reforçar a robustez dos grafos, levando em conta que esses vértices, por terem bastante influência para a conectividade do grafo, se protegidos ou otimizados, podem prevenir fragmentações graves em casos de perda de dados ou falhas.





**Figura 6. Boxplot da Centralidade de Intermediação**

## 5. Conclusão

A análise exploratória das instâncias do problema do corte máximo trouxe percepções significativas a cerca das características estruturais dos grafos. Foi observado que existe uma correlação substancial entre a transitividade, a centralidade de grau e a densidade, indicando que instâncias com maior transitividade tendem a ter vértices com maior centralidade de grau e uma densidade similar. Essas correlações podem ser exploradas para otimizar a identificação de cortes que maximizem a separação entre *clusters* de vértices.

A detecção de *outliers* destacou a presença de muitas variações nas características dos grafos, como o coeficiente de aglomeração e a centralidade de intermediação. A análise mostrou que certos vértices possuem uma influência desproporcional na conectividade dos grafos, sugerindo que a proteção ou otimização desses “vértices-chave” pode aumentar a robustez dos grafos contra falhas.

Além disso, a análise estatística das instâncias forneceu uma visão geral das propriedades estruturais dos grafos, com ênfase na diversidade e complexidade das instâncias do problema do corte máximo. A presença de valores extremos em várias métricas reforça a necessidade conhecida de algoritmos robustos, capazes de lidar com essas variações e de aproveitar essas características para melhorar o desempenho das soluções.

Para trabalhos futuros, é esperada a utilização das características estruturais identificadas nesta análise para construir modelos preditivos de aprendizado de máquina voltados para o problema do corte máximo. A expansão da análise para incluir outras bibliotecas de instâncias também pode fornecer uma visão mais abrangente e contribuir para a generalização dos modelos desenvolvidos. A utilização das informações obtidas nas análises pode potencialmente melhorar a eficácia dos algoritmos de otimização utilizados para resolver o problema do corte máximo em diferentes contextos.

## Referências

- Agrawal, R., Rajagopalan, S., Srikant, R., and Xu, Y. (2003). Mining newsgroups using networks arising from social behavior. In *Proceedings of the 12th international conference on World Wide Web*, pages 529–535.
- Almeida, A. N. and Martins, S. D. (2023). Evolução da qualidade dos planos de mitigação e monitoramento nos estudos de impacto ambiental. *Revista Gestão & Sustentabilidade Ambiental*, 12(1):e12355–e12355.
- Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.

- Barahona, F., Grötschel, M., Jünger, M., and Reinelt, G. (1988). An application of combinatorial optimization to statistical physics and circuit layout design. *Operations Research*, 36(3):493–513.
- Boros, E. and Hammer, P. L. (1991). The max-cut problem and quadratic 0–1 optimization; polyhedral aspects, relaxations and bounds. *Annals of Operations Research*, 33(3):151–180.
- Diestel, R. (2017). *Graph Theory*. Springer.
- Ferreira, M. M. C. (2022). Quimiometria iii-revisitando a análise exploratória dos dados multivariados. *Química Nova*, 45(10):1251–1264.
- Jebb, A. T., Parrigon, S., Woo, S. E., and Loving, J. L. (2016). Exploratory data analysis and visualization: A way to extract knowledge from complex data. *Journal of Computational Science*, 18:3–15.
- Latora, V., Nicosia, V., and Russo, G. (2017). *Complex Networks: Principles, Methods and Applications*. Cambridge University Press, Cambridge.
- Lin, M., Xu, W., Lin, Z., and Chen, R. (2020). Determine owa operator weights using kernel density estimation. *Economic research-Ekonomska istraživanja*, 33(1):1441–1464.
- Moore, D. S., McCabe, G. P., and Craig, B. A. (2009). *Introduction to the Practice of Statistics*, volume 4. WH Freeman New York.
- Otterbach, J. S., Manenti, R., Alidoust, N., Bestwick, A., Block, M., Bloom, B., Caldwell, S., Didier, N., Fried, E. S., Hong, S., et al. (2017). Unsupervised machine learning on a hybrid quantum computer. *arXiv preprint arXiv:1712.05771*.
- Shrimali, N. P. and Shah, N. H. (2020). *Recent Advancements in Graph Theory*. CRC Press, Boca Raton.
- Sousa, S., Haxhimusa, Y., and Kropatsch, W. G. (2013). Estimation of distribution algorithm for the max-cut problem. In *Graph-Based Representations in Pattern Recognition: 9th IAPR-TC-15 International Workshop, GbRPR 2013, Vienna, Austria, May 15-17, 2013. Proceedings 9*, pages 244–253. Springer.
- Wiegele, A. (2007). Biq mac library—a collection of max-cut and quadratic 0-1 programming instances of medium size. *Preprint*, 51.
- Wongsuphasawat, K., Liu, Y., and Heer, J. (2019). Goals, process, and challenges of exploratory data analysis: An interview study. *arXiv preprint arXiv:1911.00568*.