



Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Relatório do Projeto

Parte 2

Nome do Integrante	RA
Cesar Hideki Imai	10402758
João Victor Dallapé Madeira	10400725
David Varão Lima Bentes	10402647
Pessoa	
André Franco Ranieri	10390470

Aplicação com busca semântica e representação de grafos

Definição do problema	2
Descrição textual	2
Transformer	2
Codificador	3
Busca vetorial	3
Componentes importados	5
Objetivos de Desenvolvimento Sustentável	5
Educação de qualidade	5
Indústria, inovação e infraestrutura	5
Testes de execução	6
Ler dados do arquivo	6
Gravar dados no arquivo grafo.txt	7
Inserir vértice	8
Inserir aresta	9
Remover vértice	10
Remover aresta	11
Exibir grafo	12
Exibir matriz	13
Apresentar a conexidade do grafo	14
Encerrar a aplicação	15
Buscar um índice pela palavra	16
Buscar uma palavra pelo índice	17
Fazer busca vetorial	18
Exibir grafo colorido	19
Exibir árvore parcial de custo mínimo	20
Apresentar caminho hamiltoniano	21
Apêndice	22





Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

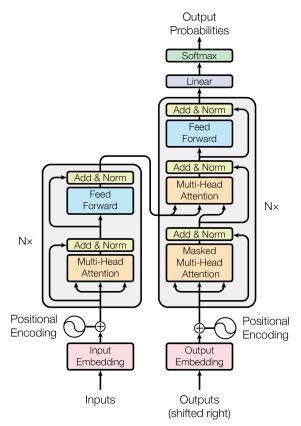
Definição do problema

O objetivo é utilizar a representação de grafos para uma melhor visualização acerca da busca semântica de palavras. A busca semântica é muito útil para consultas de dados não estruturados em que é necessário considerar o contexto e a relação entre diferentes itens. Exemplos de aplicações incluem a ferramenta de pesquisa do Google, *chatbots* como o ChatGPT e algoritmos de recomendação do YouTube.

Descrição textual

Transformer

A arquitetura *Transformer* foi um modelo de rede neural introduzida no artigo <u>Attention is All You Need</u>. Essa arquitetura contribuiu para avanços significativos no campo de processamento de linguagem natural (NLP), através da aplicação do mecanismo de auto-atenção e da codificação posicional. Também ela serviu de base para o desenvolvimento de grandes modelos de linguagem (LLMs), componentes fundamentais para a implementação de *chatbots*, como o ChatGPT e o Gemini.



Arguitetura Transformer (VASWANI et al., 2017, p. 3)





Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

O mecanismo de auto-atenção permite extrair a relação de uma palavra com outras palavras do mesmo texto ao ser processada. O que é importante para capturar dependências entre palavras, não importando a sua distância na sequência.

Já a técnica de codificação posicional é usada para inserir informações sobre a posição dos *tokens* dentro de uma sequência. O que foi essencial para permitir o processamento paralelo na arquitetura *Transformer*.

Codificador

Os codificadores são modelos de linguagem usados para:

- Tokenização É o processo de dividir um texto em unidades menores, chamadas de tokens. Esses tokens podem ser palavras, subpalavras, caracteres ou até mesmo frases, dependendo do contexto e da aplicação.
- **Embedding** É o processo de converter *tokens* em vetores numéricos multidimensionais. Esses vetores capturam informações semânticas e contextuais dos *tokens*. *Embeddings* permitem que modelos de aprendizado de máquina compreendam o significado dos *tokens*.
- Outras tarefas Como classificação de textos, detecção de entidades nomeadas (NER) e análise de tópicos.

O modelo de codificador BERTimbau é uma versão do modelo BERT (Bidirectional Encoder Representations from Transformers) que foi treinada com o *dataset* BrWaC (Brazilian Web as Corpus) entre outros dados em português. A pesquisa <u>BERTimbau: Pretrained BERT Models for Brazilian Portuguese</u> mostrou que a especialização do modelo em português contribuiu para um aumento na performance em comparação com a versão multilíngue mBERT.

Busca vetorial

A busca vetorial é uma técnica de recuperação de informações que utiliza *embeddings*. Ao invés de procurar correspondências exatas de palavras-chave, a busca vetorial se baseia em similaridade semântica

Principais métodos para medir similaridade:

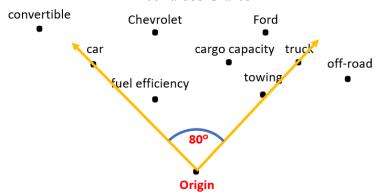
• **Similaridade por Cosseno -** Medida de similaridade que calcula o cosseno do ângulo entre dois vetores no espaço multidimensional.

$$\cos(heta) = rac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = rac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2} \sqrt{\sum\limits_{i=1}^{n} B_i^2}}$$





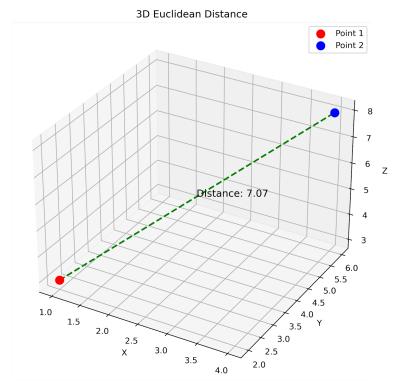
Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos



Fonte: https://en.wikipedia.org/wiki/File:CosineSimilarity.png

• **Distância Euclidiana** - Medida de similaridade que calcula a reta entre dois pontos em um espaço multidimensional.

$$d = \sqrt{\sum_{i=1}^{n} (b_i - a_i)^2}$$



Fonte: https://www.datacamp.com/pt/tutorial/euclidean-distance





Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Componentes importados

- **sentence-transformers** É um pacote que fornece ferramentas para usar modelos de *Transformers* para criar representações vetoriais (*embeddings*) e fazer buscas vetoriais.
- requests É um módulo para fazer requisições HTTP.
- pyvis É uma biblioteca para visualização de grafos interativos.
- time É uma biblioteca para acessar e manipular o tempo.
- ipywidgets É uma biblioteca para criar interfaces interativas no Jupyter Notebook.
- IPython.display É um módulo para exibir e controlar saídas no Jupyter Notebook.

Objetivos de Desenvolvimento Sustentável







































Educação de qualidade

O projeto contribui como ferramenta de estudo para *Large Language Models* (LLMs), modelos que processam linguagem natural e que são essenciais para sistemas de *chatbots*.

Indústria, inovação e infraestrutura

O projeto se propõe a facilitar as análises de relações semânticas com representações de grafos. O que pode auxiliar na análise de balanceamento de pesos de LLMs.

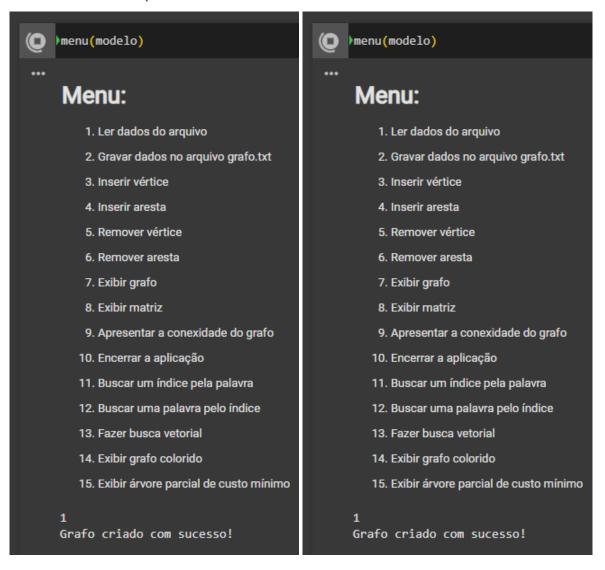




Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Testes de execução

Ler dados do arquivo

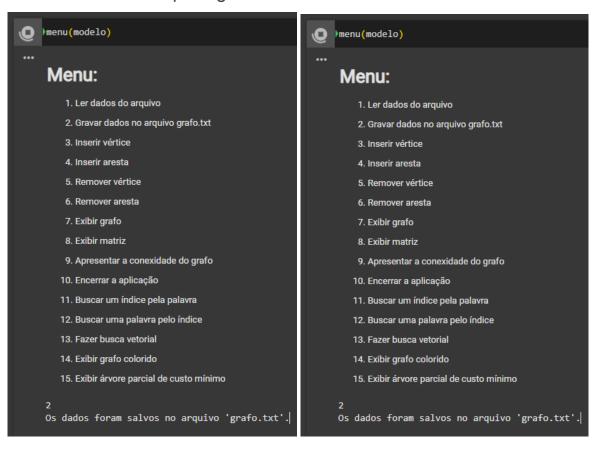






Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Gravar dados no arquivo grafo.txt

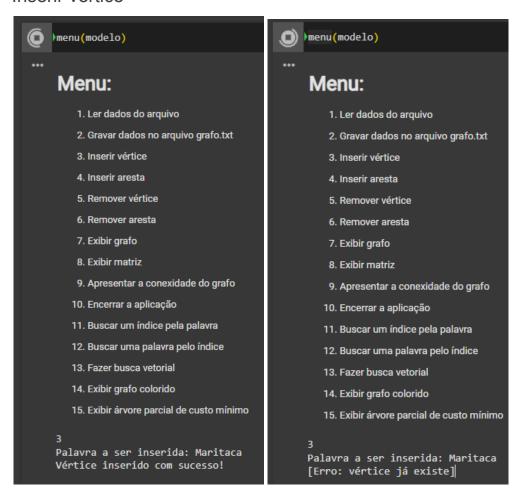






Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Inserir vértice

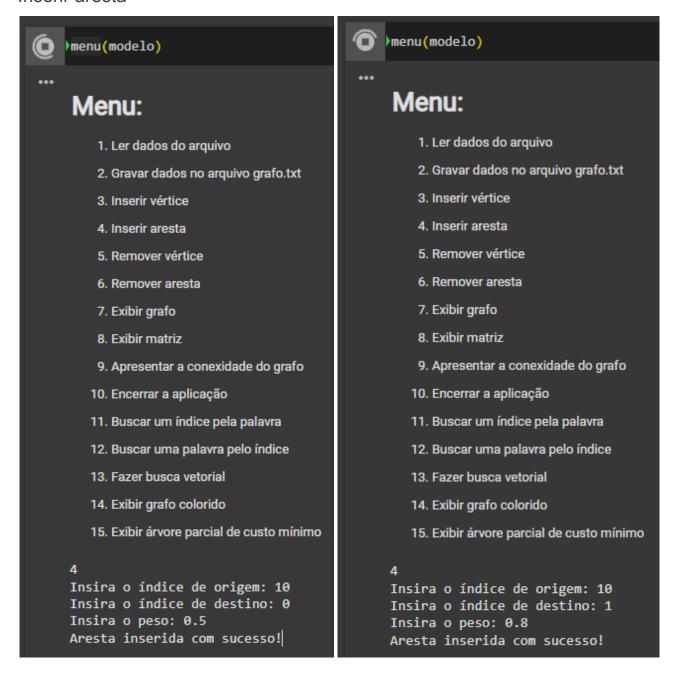






Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Inserir aresta

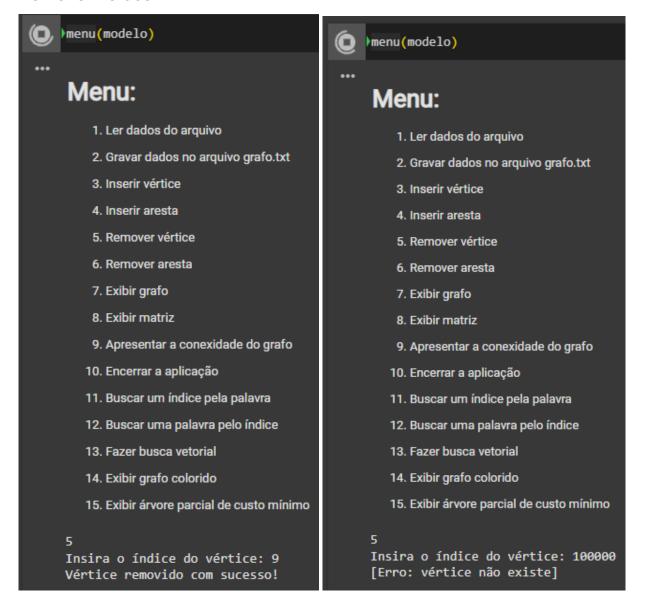






Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Remover vértice

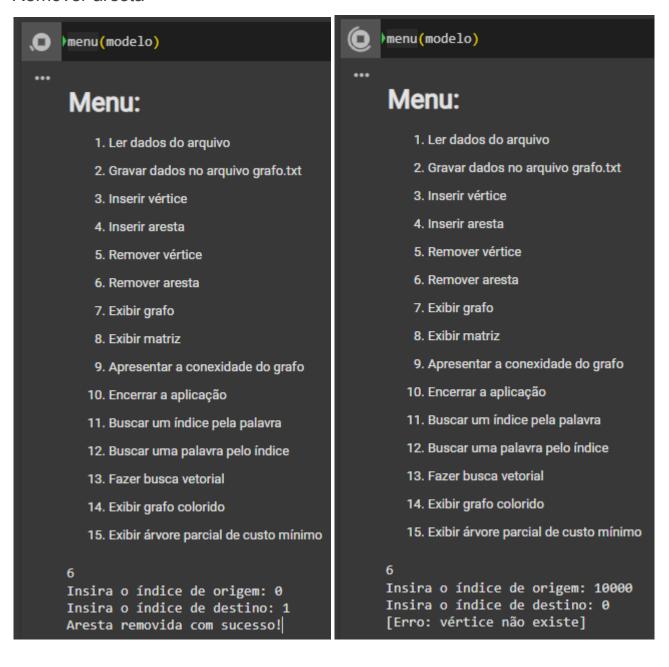






Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Remover aresta

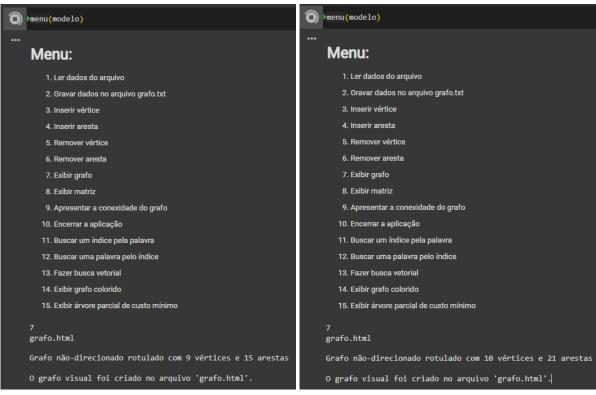


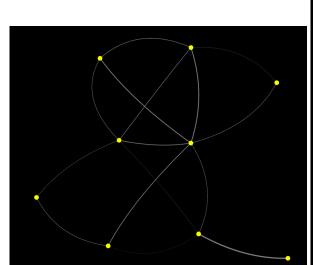


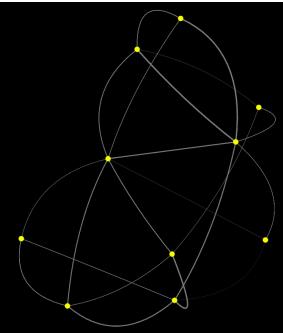


Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Exibir grafo





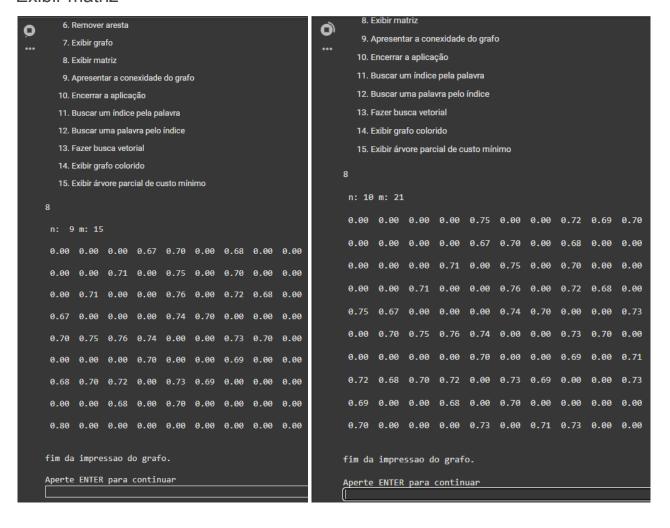






Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Exibir matriz

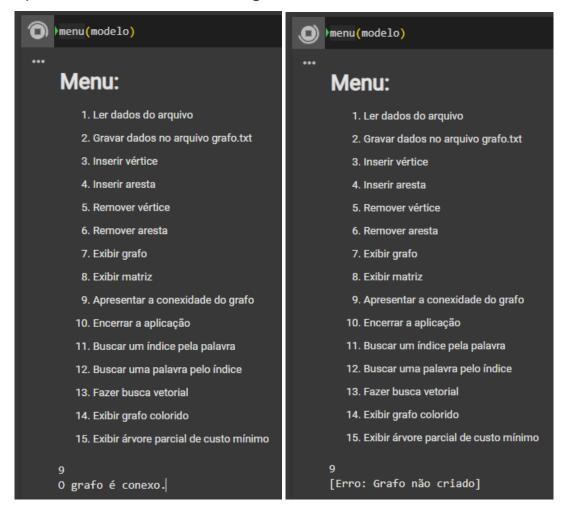






Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Apresentar a conexidade do grafo

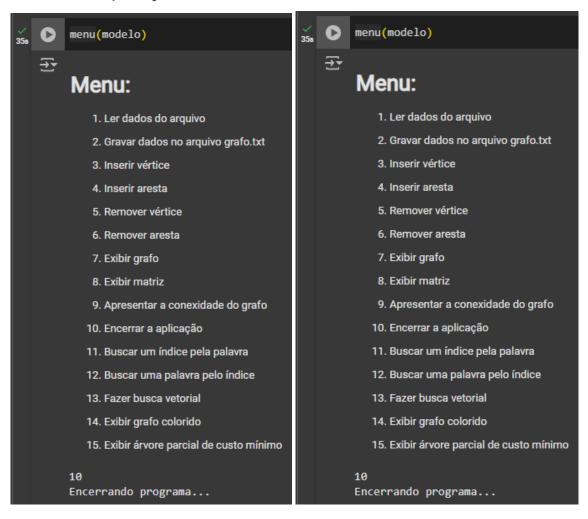






Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Encerrar a aplicação

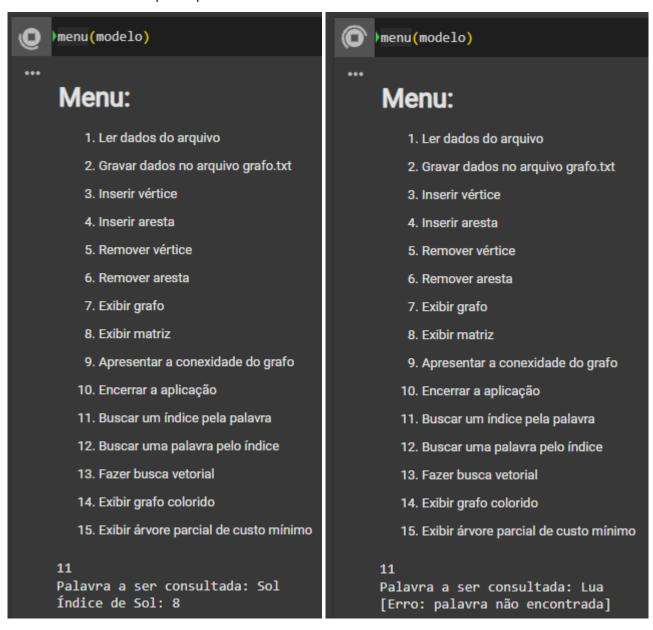






Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Buscar um índice pela palavra

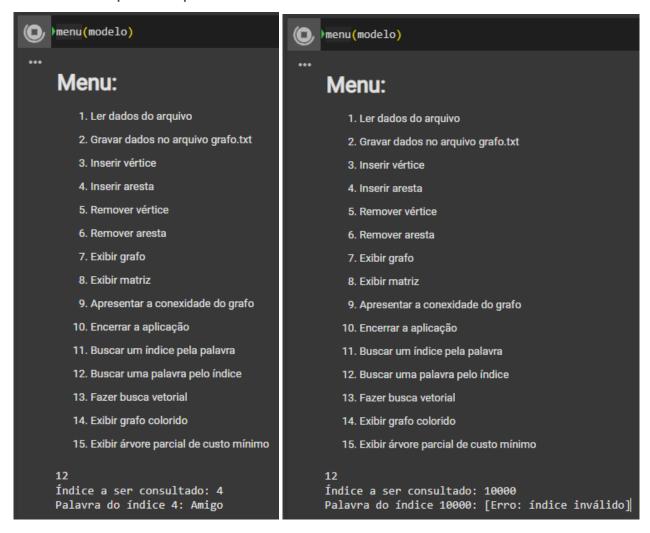






Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Buscar uma palavra pelo índice







Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Fazer busca vetorial



Michia.

- 1. Ler dados do arquivo
- 2. Gravar dados no arquivo grafo.txt
- 3. Inserir vértice
- 4. Inserir aresta
- 5. Remover vértice
- 6. Remover aresta
- 7. Exibir grafo
- 8. Exibir matriz
- 9. Apresentar a conexidade do grafo
- 10. Encerrar a aplicação
- 11. Buscar um índice pela palavra
- 12. Buscar uma palavra pelo índice
- 13. Fazer busca vetorial
- 14. Exibir grafo colorido
- 15. Exibir árvore parcial de custo mínimo

13

Insira uma palavra: Lua

Busca: Lua

Palavra: Sol Peso: 0.84

Palavra: Casa Peso: 0.65

Palavra: Escola Peso: 0.64

Aperte ENTER para continuar

()

Wichia.

- 1. Ler dados do arquivo
- 2. Gravar dados no arquivo grafo.txt
- 3. Inserir vértice
- 4. Inserir aresta
- 5. Remover vértice
- 6. Remover aresta
- 7. Exibir grafo
- 8. Exibir matriz
- 9. Apresentar a conexidade do grafo
- 10. Encerrar a aplicação
- 11. Buscar um índice pela palavra
- 12. Buscar uma palavra pelo índice
- 13. Fazer busca vetorial
- 14. Exibir grafo colorido
- 15. Exibir árvore parcial de custo mínimo

13

Insira uma palavra: Família

Busca: Família

Palavra: Casa Peso: 0.80

Palavra: Escola Peso: 0.77

Palavra: Montanha

Peso: 0.72

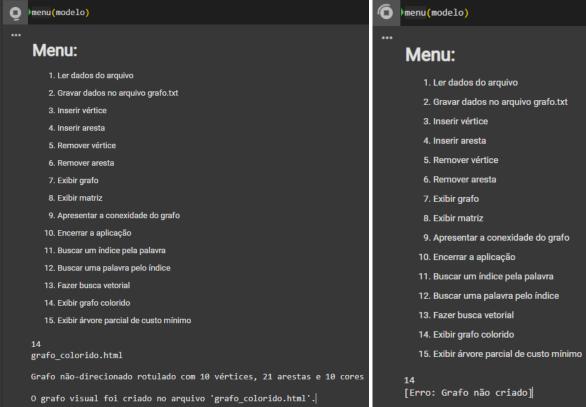
Aperte ENTER para continuar

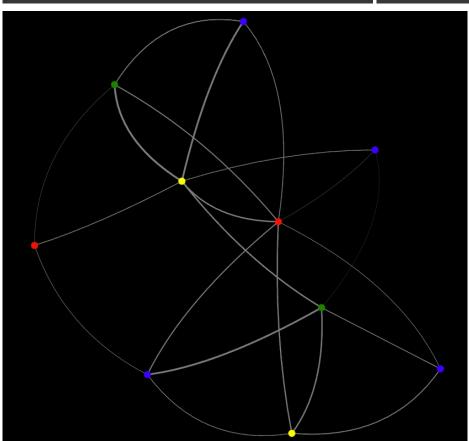




Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Exibir grafo colorido





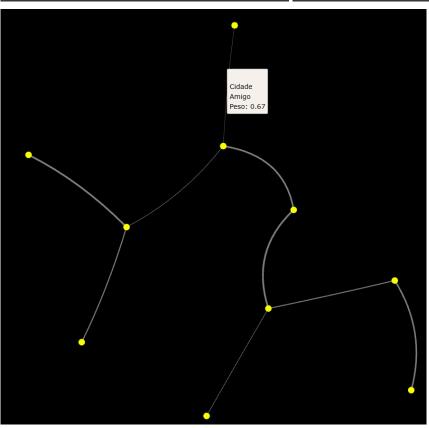




Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Exibir árvore parcial de custo mínimo



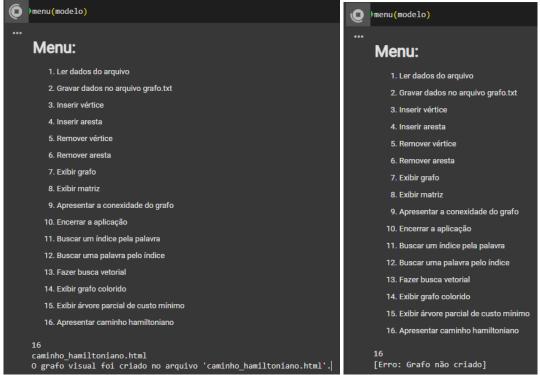


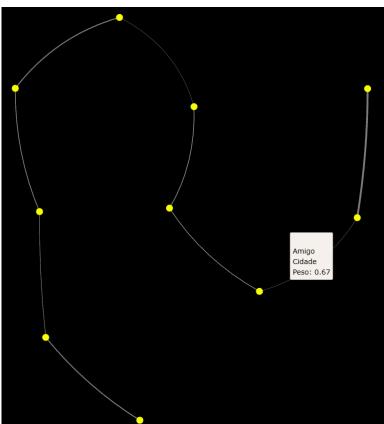




Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Apresentar caminho hamiltoniano









Prof. Dr. Ivan Carlos Alcântara de Oliveira Teoria dos Grafos

Apêndice

- Acesso ao projeto: https://github.com/Cehiim/TeoriaDosGrafos/tree/main/ProjetoFinal
- Vídeo do projeto: https://youtu.be/gusflpNNv3A