UNIVERSIDADE FEDERAL
DE SANTA CATARINA

Universidade Federal de Santa Catarina

Centro de Blumenau
Departamento de Matemática

PIBIC
Relatório Final

Citar o nome do meu projeto de pesquisa ao qual este projeto está vinculado

~~DMDGP: Um Problema Real~~

Não gosto desse nome
MUDE.

Guilherme Philippi, ~~orientado por Felipe Delfini Caetano Fidalgo~~
g.philippi@grad.ufsc.br, ~~felipe.fidalgo@ufsc.br~~

Orientador: ___ ___ (email)

16 de agosto de 2019

Sumário

1 Introdução 3

2 Grafos: Aspectos Gerais 4

2.1 Classificações 5

3 Um Passeio pela Bioquímica 6

3.1 Carbono 6

3.2 Classificação Macromolecular 7

3.3 Configuração Molecular 8

3.4 Aminoácidos 9

3.5 Estrutura das Proteínas 11

3.6 *Worldwide Protein Data Bank* 12

3.6.1 *Sotware PDBReader* 15

4 *Molecular Distance Geometry Problem* 18

4.1 Geometria de Distâncias 18

4.2 Representações de Átomos em Coordenadas 18

4.3 Ressonância Magnética Núclear 21

4.4 Modelagem Matemática 21

4.4.1 MDGP: Uma definição formal 22

4.5 Modelagem Computacional 23

4.6 Estudando o Conjunto Solução de um MDGP 25

4.7 Ordenação Conveniente dos Vértices 27

4.8 *Discretizable Molecular Distance Geometry Problem* 29

5 *Branch-and-Prune* 31

5.1 O Algorítimo 31

5.2 Estrutura Algorítmica 35

5.3 Simulações Computacionais 36

5.3.1 Representando distâncias com matrizes 36

5.3.2 Medida de distância entre resultados 36

5.3.3 Experimentos 36

6 Considerações Finais 37

Referências 38

Apêndice A - Lei dos Cossenos e Ângulos Entre dois Vetores no $\mathbb{R}^3$ 40

Apêndice B - Matrizes como Transformações Lineares e Sobre B_i 42

Apêndice C - Vinte Aminoácidos Naturais 50

Abstract

applied to proteins

study

In this work we ~~presented~~ the ~~basic principles of the~~ Discretizable Molecular Distance Geometry Problem (DMDGP), as well as the necessary tools for its introduction, going from graph theory to ~~the~~ biomolecular structure ~~of the primordial structures of study~~. We also ~~present~~ some recent results on the ordering that makes up the problem. The text concludes with a study of the main algorithm described to solve the problem and a brief section of computer simulations.
Keywords: DMDGP, Distance geometry, Optimization.

composes

of a protein graph

deal with

Resumo

Neste trabalho foram apresentados os princípios básicos sobre o Discretizable Molecular Distance Geometry Problem (DMDGP), bem como as ferramentas necessárias para sua introdução, passando da teoria de grafos à estrutura biomolecular das estruturas primordiais de estudo. Também apresentamos alguns resultados recentes sobre a ordenação que compõe o problema. O texto se encerra com um estudo sobre o principal algoritmo descrito para solucionar o problema e uma brevê seção de simulações computacionais.
Palavras-chave: DMDGP, Geometria de Distâncias, Otimização.

~~Estado~~ ~~São dadas~~

1 Introdução

entre Referência.

Existe uma relação muito forte ~~com~~ a forma geométrica das moléculas orgânicas e suas funções em organismos vivos. Pode-se fazer uma analogia destes organismos com um grande quebra-cabeça, cheio de peças tão variadas quanto se queira, onde cada peça tem um local específico ~~no grande quebra-cabeça~~, de forma que sua ~~posição~~ e função é dado justamente pela forma geométrica ~~de cada peça~~. Imagine este grande quebra-cabeça de forma tridimensional ao invés de plana, com peças em formatos diversos e escalafobéticos, logo, você não estará tão longe de entender como funciona um organismo vivo.

Retirar

Outrora, em pesquisas sobre a molécula de DNA (ácido desoxirribonucleico), descobriu-se que essa era parte fundamental da produção de um dos pilares para a vida: a proteína. ~~Esta será o alvo principal deste estudo~~ e possui uma gama extensa de funções em nosso organismo ~~— como se fossem as tais peças do quebra-cabeça da analogia acima~~. Podemos dizer que somos feitos de proteínas. Organismos vivos lutam constantemente contra a desorganização intrínseca do universo e as proteínas tem papel importantíssimo nessa luta. São elas as estruturas que utilizamos para nos organizar, gerando informação, ao possibilitarem um mecanismo procedural natural para a vida, como com o seu papel no transporte de oxigênio (hemoglobina), na proteção do corpo contra organismos patogênicos (imunoglobulina), com a catalização de reações químicas (apoenzima), além de outras inúmeras funções primordiais no nosso organismo [1]. ~~Perceba a importância fundamental em estudar a estrutura geométrica de cada uma dessas proteínas e sua relação com suas funções.~~

O que é isso?

Por conta dessa motivação tem-se esforços como o de Kurt Wüthrich, que propôs que se utilizasse experimentos de *Ressonância Magnética Nuclear* (RMN) para calcular a estrutura tridimensional de uma molécula de proteína, ganhando o premio Nobel da Química em 2002 [2]. ~~Porém, a partir dessa estratégia tivemos novos problemas.~~ A RMN não tem como resultado a estrutura tridimensional de uma proteína, mas sim distâncias entre átomos relativamente próximos que compõem a proteína. Para poder calcular a estrutura de uma proteína a partir dessas distâncias, surgira um novo problema, conhecido na literatura como *Problema de Geometria de Distâncias Moleculares*, ou simplesmente, MDGP [3].

com imprecisões, o que é um inconveniente

juntamente com as informações provenientes de outras fontes físicas e químicas

~~Ao longo desse texto estudaremos este problema, suas variações, seus problemas relacionados, as ferramentas que nos facilitarão a resolve-los, suas soluções e as complexidades relacionadas a estas soluções.~~

Partindo para uma introdução a ferramenta matemática usada para definir o nosso problema, o capitulo 2 da aspectos essenciais para este texto sobre a teoria de grafos, donde segue-se para uma contextualização do problema do ponto de vista biomolecular, explorando cada detalhe da estrutura das proteínas no capítulo 3. Após, segue-se para a seção principal do texto, o capitulo 4, que se encerra com a definição formal do problema, nos permitindo tratar do algorítimo que é responsável por sua solução, introduzido no capítulo 5.

O texto se encerra com os três apêndices: A, B e C. O primeiro trata de uma brevê explicação sobre a lei dos cossenos, que nos é essencial; O apêndice B constrói, junto ao leitor, uma linha de raciocínio para definir o conjunto de operações que compõe a matriz B_i; Por último, no apêndice C, temos uma representação química das estruturas básicas que compõe as proteínas.

Um conjunto de referências podem ser encontrados no fim do documento.

Resumir

seja objetivo

Mencionar brevemente a discretização do espaço de buscas.

Nosso foco não é ele, mas o DMDGP.

e aplicado em diversas áreas para além da Matemática.

um breve resumo

2 Grafos: Aspectos Gerais

Esta seção tem como objetivo apresentar a *teoria de grafos*, ~~um~~ tema muito estudado por diversos matemáticos~~, o que dá uma certa sensação de importância histórica na matéria~~.

sensação não é Ciência!

Reescrever

Pode-se dizer que em 1736 é que a teoria teve início, com base no artigo publicado por Leonhard Euler, sobre as 7 pontes de Königsberg [4] [5]. Esse é o problema que normalmente introduz quem está começando a trabalhar com grafos — se trata do desafio de ligar todos os pontos de um desenho sem tirar o lápis do papel e sem passar duas vézes no mesmo ponto. Diz a lenda que esta era uma dúvida rotineira dos moradores de Koenigsberg, que tentavam constantemente fazer o trecho a pé motivados a resolverem o desafio — algo que, tanto pelo tamanho demasiadamente grande do percurso quanto pela falta de fé de que esse problema realmente motivaria alguém não ligado a matemática, pessoalmente, acho que alguém inventou essa última parte.

estória

Ciência não é opinião

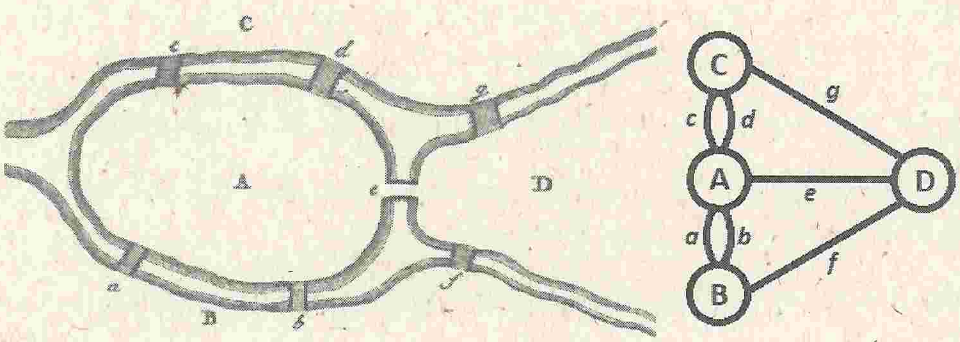

Figura 1: Ilustração original do problema [4] e sua representação em Grafos.

Isso pode ser descrito com mais detalhes para intro, aqui o formalismo

Muitas situações reais podem ser convenientemente representadas por diagramas contendo um conjunto de pontos e linhas ligando pares desses pontos — ~~que é a~~ definição ~~empírica~~ de um grafo. Por exemplo, esses pontos podem ser pessoas, onde as linhas representam pares de amigos.

Existe uma íntima relação entre Grafos e algoritmos. De fato, podemos inclusive representar um algoritmo por um grafo [6]. Na verdade, a definição de grafos é tão abrangente que podemos ver suas ligações com diversas áreas do conhecimento.

> **Grafo:** Um grafo G é uma tripla ordenada da forma $(V(G), E(G), \psi_G)$, composto por um conjunto de *vértices* $V(G)$, de arestas $E(G)$ e uma *função de incidência* ψ_G que, por sua vez, associa a cada aresta de $V(G)$ um par não ordenado de vértices (nem sempre distintos) de $E(G)$. Costumamos dizer que as arestas ligam os vértices.

TROCAR DE POSIÇÃO

Pede-se uma pequena pausa para a licença poética: Acreditamos que uma das mais belas características da representação dos grafos é que o conjunto de vértices V aceita qualquer elemento, ou seja, pode ser tanto um prédio, quanto um átomo. Daí vem o grande poder de abstração da representação de grafos.

Poesia não é Ciência

~~Essa abstração na representação de grafos é tamanha que nos motiva a usar um~~ paradigma diferente para apresenta-los. Na literatura, grafos sempre são usados em paralelo com desenhos de diagramas que tentam demonstrar, geometricamente, a que está associado os conjuntos que os definem. ~~Porém, essa representação é enganosa: Os vértices simplesmente não tem uma posição bem definida no espaço.~~

No que se segue, definiremos algumas características elementares que serão utilizadas durante esse texto. Para um estudo mais completo obre essa teoria ~~(a qual recomendamos profundamente), favor considerar~~ [5] e [7].

VIDE.

2.1 Classificações

Existem duas definições de extrema importância para o nosso problema molecular (tema central desse texto): o conceito de grafo completo e o de estruturas k-cliques. Porém, seremos obrigado a definir alguns outros conceitos prévios, como segue.

Laço: Uma aresta $\{e_i, e_j\} \in E$ tal que $i = j$.

Também, caso existam duas arestas iguais ($\{e_i, e_j\}$ e $\{e_j, e_i\}$, por exemplo), com as mesmas extremidades são ditos *paralelos.*

Grafo simples: Um grafo que não possui laços ou arestas paralelas.

Grafo Completo: Um grafo simples onde quaisquer dois vértices são adjacentes é denominado *grafo completo.*

Outro conceito que nos será de grande utilidade é o de subgrafo.

Subgrafo: É um grafo resultante de um subconjunto de vértices e outro subconjunto de arestas de outro grafo. Isto é, seja $G = (V, E), G' = (V', E')$ é dito um subgrafo de G se (V', E') é um grafo, $V' \subseteq V$ e $E' \subseteq E$.

E, finalmente

k-**Clique**: é um subgrafo G' com k vértices tal que G' é completo.

Em especial, também podemos interpretar as arestas como *caminhos* e, se o fizermos, podemos pensar em alguma forma de métrica para esses caminhos. Esse pensamento da origem aos grafos ditos *ponderados.*

Grafo Ponderado: É um grafo que possui uma função $d(E) \to \mathbb{R}$ associada, isto é, o grafo que possui valores numéricos atribuídos as suas arestas.

Para nós essa última definição terá uma maior importância, visto que nosso problema é localizar um conjunto de pontos dado um conjunto de distâncias entre eles [1]. Perceba que se trata do problema de desenhar o diagrama que representa esse grafo ponderado, completo [8], respeitando as suas distâncias no papel.

FORA DE CONT

QUE ISSO?

Ah, é!!
O leitor não foi informado disso previamente e estará boiando...