



Universidade Federal do Piauí - UFPI

Centro de Ciências da Natureza

Departamento de Computação

Teresina - PI

Aplicação do problema do clique na filogenética de bactérias

Participantes:

Kauã Marques do Nascimento

Lucas Herlon Santos Moreira da Cunha

Davi do Nascimento Santos

Introdução

Definição do Problema

O Problema do Clique é um problema em Teoria dos Grafos, onde temos um grafo não direcionado e queremos encontrar um grupo de vértices (um "clique") onde cada par de vértices está conectado por uma aresta. Um clique maximal é um clique para o qual não se pode adicionar mais vértices sem que este perca a característica de clique. O problema consiste na seguinte pergunta: Dado um grafo não direcionado e um número k . Existe um grupo de k vértices no grafo onde cada par de vértices está conectado por uma aresta?

Exemplo: Se temos um grafo com vértices $\{A, B, C, D\}$ e arestas $\{(A, B), (A, C), (B, C), (B, D)\}$, e $k = 3$, queremos saber se há um grupo de 3 vértices onde cada par está conectado por uma aresta. Neste exemplo, $\{A, B, C\}$ é um clique.

O Problema do Clique é classificado como NP-completo, uma categoria na teoria da complexidade computacional que indica alta dificuldade na busca de soluções eficientes (rápidas). A complexidade NP-completo implica que encontrar uma solução ótima em tempo polinomial para todos os casos conhecidos é altamente improvável. Diante dessa complexidade, recorre-se a heurísticas, abordagens que buscam soluções subótimas em um tempo razoável, embora sem garantia de precisão. O Problema do Clique é reconhecido como um dos desafios mais intratáveis na ciência da computação, refletindo a busca contínua por estratégias mais eficazes, apesar da ausência de uma solução ideal universalmente aplicável.

No escopo deste trabalho estamos interessados em procurar em um grafo os cliques maximais que o compõem, uma vez que essa estrutura pode ser usada para modelar problemas em que precisamos agrupar indivíduos em uma mesma categoria. Aqui iremos aplicar o problema do clique na biologia, em especial na filogenética.

Aplicação prática proposta

A aplicação prática está na identificação e classificação de grupos de bactérias com base em características evolutivas específicas. Para alcançar essa classificação, utiliza-se o problema do clique como uma ferramenta de modelagem do relacionamento entre as bactérias. O grafo é construído onde os vértices representam as bactérias e as arestas são estabelecidas com base na quantidade de características evolutivas que as bactérias compartilham. Por exemplo, se duas bactérias compartilham 5 características evolutivas em comum, elas estarão ligadas por uma aresta, indicando um certo grau de parentesco. Uma vez que o grafo estiver modelado, analisaremos os cliques maximais para determinar a quantidade de grupos formados, onde cada clique maximal representa um grupo de bactérias com características evolutivas semelhantes. A contagem dos vértices em cada clique maximal nos permitirá catalogar as bactérias que compartilham características evolutivas parecidas, possibilitando a classificação das mesmas.

Referências

Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP). Complexidade Computacional: O que são problemas NP-completos? Disponível em: https://www.ime.usp.br/~pf/analise_de_algoritmos/aulas/NPcompleto.html. Acesso em: 16/10/2023.

Holland, Barbara & Spencer, Hamish & Worthy, Trevor & Kennedy, Martyn. (2010). Identifying Cliques of Convergent Characters: Concerted Evolution in the Cormorants and Shags. *Systematic biology*. 59. 433-45. 10.1093/sysbio/syq023.

Bhar A, Gierse LC, Meene A, Wang H, Karte C, Schwaiger T, Schröder C, Mettenleiter TC, Urich T, Riedel K and Kaderali L (2022) Application of a maximal-clique based community detection algorithm to gut microbiome data reveals driver microbes during influenza A virus infection. *Front. Microbiol.* 13:979320. doi: 10.3389/fmicb.2022.979320