

Ciência da Computação / Sistemas de Informação

## **Ferramenta para compressão de grafos de dados de proteínas**

Hélio Henrique Medeiros Silva - 9º módulo de ciência da computação, UFLA, iniciação científica voluntária

Vinícius dos Santos Dias - Orientador DCC, UFLA - Orientador(a)

### **Resumo**

Armazenar dados de proteínas em formato de grafo demanda grande quantidade de memória, dado que essas estruturas podem conter de centenas a milhares de átomos. Este trabalho teve como objetivo reduzir o espaço necessário para esse armazenamento. Para isso, foi desenvolvida uma função que recebe como entrada uma lista de grafos de proteínas, modelados com a biblioteca Graphein em Python, e retorna um único grafo contendo todos os vértices e arestas, sem repetições. O fator de compressão (relação entre o tamanho dos grafos originais e o do grafo comprimido) foi avaliado em experimentos realizados com datasets (listas de proteínas) disponibilizados pelo Graphein, variando de 43 MB a mais de 6 GB após modelados como grafos. Os resultados mostraram taxas de compressão entre 1,4x (40%) e 4,78x (378%), com média de 2,3x (130%). A compressão mostrou ser mais efetiva em grafos maiores, atingindo seus valores mais altos (4.78x) para os maiores grafos (com mais de 6GB). Conclui-se que o método é eficiente em todos os casos, destacando-se pela maior utilidade em cenários de grande volume de dados.

Palavras-Chave: Grafos, Compressão de dados, Dados biológicos.

Instituição de Fomento: N/A

Link do pitch: <https://www.youtube.com/watch?v=Rc3ZuA82Tw8>