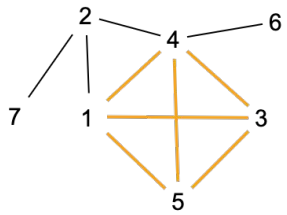
Max Clique Problem: Applications and Approximation

Aaron Banse, Brendan Banfield, Sophie Quinn, Kellan Knop

Max Clique Problem

Given a graph, what is the size of the largest complete subgraph, or **clique**?



Example: maximum clique is of size 4.

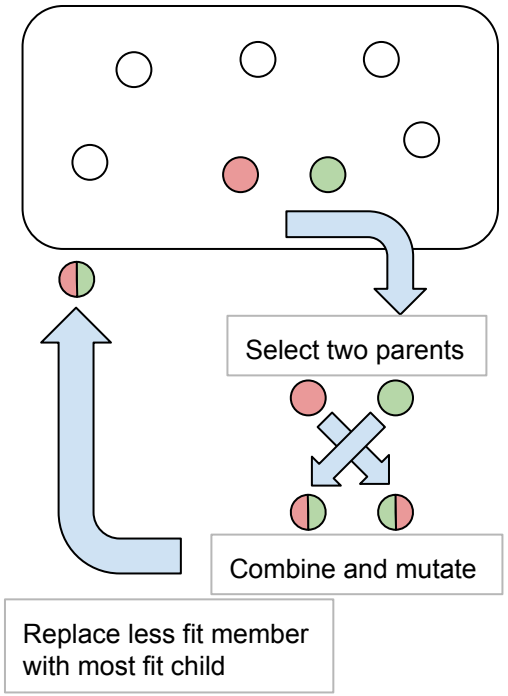
NP-Hard problem!

Protein Structural Similarity

A notable application of the max clique problem is efficiently finding protein structural similarity. Given two protein graphs, we can construct a product graph encoding local similarity between the two proteins.

Hence, a clique in this graph represents an area of matching structure.
The size of the max clique, then, is a good representative of the two proteins structural similarity.

Genetic Approximation Algorithm

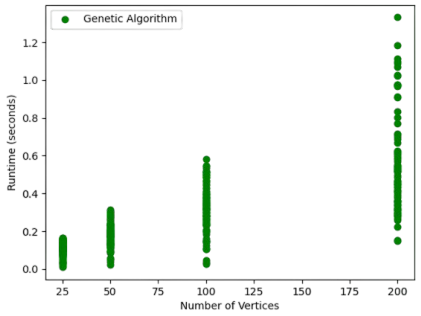


Each member of the population is a clique, meaning its fitness is simply the size of the clique. We progressively improve the population until we cannot anymore, at which point we return the largest clique in the population.

Results

We tested several other algorithms alongside the genetic algorithm on a series of randomized graphs and protein product graphs. The genetic algorithm performed exceptionally well, finding the correct max clique **92.6% of the time** for the randomized graphs, and was **fully accurate** for the protein product graphs.

Unlike exact solutions to the max clique problem, this algorithm does not exhibit exponential runtime, making it efficient even for larger graphs, as seen below.



References

Huang, B. (2019). *Finding Maximum Clique with a Genetic Algorithm*.

Depolli, M., Konc, J., & Janežič, D. (2013). *Exact Parallel Maximum Clique Algorithm for General and Protein Graphs. Journal of Chemical Information and Modeling*, **53**(9), 2217-2228.