

ABSTRAK

Penyakit Tuberkulosis (TB) melibatkan interaksi protein yang sangat kompleks dalam mekanisme patogenesisnya. Untuk memahami mekanisme penyakit ini secara mendalam, diperlukan pendekatan komputasi graf yang mampu mengekstraksi jaringan interaksi secara efisien. Namun, besarnya skala dan kepadatan data interaksi protein (*hairball network*) menuntut adanya algoritma pencarian yang mampu bekerja secara cepat dan akurat untuk menemukan pusat interaksi utama tanpa mengorbankan performa komputasi. Keterbatasan tersebut menunjukkan perlunya sistem komputasi yang tidak hanya tangguh secara matematis, tetapi juga mampu menghasilkan data biologis yang relevan. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi efisiensi waktu komputasi algoritma CliSAT pada jaringan interaksi protein dengan tingkat kepadatan yang bervariasi, serta menganalisis korelasi antara kelompok protein yang teridentifikasi sebagai *maximum clique* dengan peran fungsionalnya.

Metode yang diusulkan menggunakan pendekatan pemodelan graf dan optimasi algoritma CliSAT untuk menyelesaikan *Maximum Clique Problem*. Proses diawali dengan penyaringan interaksi protein berdasarkan ambang batas *confidence score* untuk mengeliminasi probabilitas interaksi yang lemah. Proses ini mempartisi jaringan ke dalam tiga skenario tingkat kepadatan, yaitu *Medium* (0.4), *High* (0.7), dan *Highest* (0.9). Struktur topologi graf yang dihasilkan kemudian ditransformasikan ke dalam representasi baku berformat DIMACS (.clq) agar matriks relasionalnya dapat dibaca dan dieksekusi oleh mesin algoritma CliSAT. Evaluasi dilakukan dengan mengukur efisiensi waktu eksekusi komputasi pada ketiga skenario, serta melakukan analisis pengayaan biologis untuk memastikan peran fungsional dari kelompok protein target yang berhasil diekstraksi.

Hasil penelitian menunjukkan bahwa algoritma CliSAT sangat efisien dalam mengatasi kepadatan jaringan, dengan waktu eksekusi pencarian *maximum clique* sebesar 0.01316 detik untuk 23 protein (*Medium*), 0.00401 detik untuk 15 protein (*High*), dan 0.00332 detik untuk menyisakan 9 protein inti (*Highest*). Analisis lanjutan memastikan bahwa kelompok protein yang diekstraksi memiliki korelasi fungsional yang sangat kuat dengan sistem pertahanan tubuh inang, khususnya pada mekanisme modifikasi kromatin, aktivasi limfosit, dan persinyalan sitokin. Kesimpulannya, algoritma CliSAT terbukti tangguh dalam mereduksi kompleksitas jaringan secara cepat, dan kelompok protein yang dihasilkan dapat dipastikan memiliki peran fungsional yang relevan sebagai data rujukan terpercaya untuk penelitian medis lanjutan.

Kata Kunci: Tuberkulosis, Jaringan Interaksi Protein, *Maximum Clique*, CliSAT, Bioinformatika

ABSTRACT

Tuberculosis (TB) disease involves highly complex protein interactions in its pathogenesis mechanism. To thoroughly understand this disease mechanism, a graph computational approach capable of efficiently extracting interaction networks is required. However, the massive scale and density of protein interaction data (hairball network) demand a search algorithm capable of working quickly and accurately to find the main interaction centers without sacrificing computational performance. These limitations highlight the need for a computational system that is not only mathematically robust but also capable of generating relevant biological data. Therefore, this study aims to evaluate the computational time efficiency of the CliSAT algorithm on protein interaction networks with varying density levels, as well as to analyze the correlation between the protein groups identified as the maximum clique and their functional roles.

The proposed method uses a graph modeling approach and CliSAT algorithm optimization to solve the Maximum Clique Problem. The process begins with filtering protein interactions based on a confidence score threshold to eliminate weak interaction probabilities. This process partitions the network into three density level scenarios, namely Medium (0.4), High (0.7), and Highest (0.9). The resulting graph topology structure is then transformed into a standard representation in DIMACS format (.clq) so that its relational matrix can be read and executed by the CliSAT algorithm engine. Evaluation is carried out by measuring the computational execution time efficiency across the three scenarios, as well as performing biological enrichment analysis to ensure the functional role of the successfully extracted target protein groups.

The results showed that the CliSAT algorithm is highly efficient in overcoming network density, with a maximum clique search execution time of 0.01316 second for 23 proteins (Medium), 0.00401 second for 15 proteins (High), and 0.00332 second leaving 9 core proteins (Highest). Further analysis confirmed that the extracted protein groups have a very strong functional correlation with the host's immune defense system, particularly in the mechanisms of chromatin modification, lymphocyte activation, and cytokine signaling. In conclusion, the CliSAT algorithm is proven to be robust in rapidly reducing network complexity, and the resulting protein groups are confirmed to have relevant functional roles as reliable reference data for further medical research.

Keywords: *Tuberculosis, Protein-Protein Interaction, Maximum Clique, CliSAT, Bioinformatics*