# ABSTRAK

Kanker payudara merupakan penyakit dengan kompleksitas molekuler tinggi karena melibatkan interaksi banyak protein dalam berbagai jalur biologis. Analisis jaringan interaksi protein memungkinkan pemahaman hubungan antar protein secara sistematis. Namun, jaringan biologis bersifat kompleks dan memungkinkan satu protein berada pada lebih dari satu kelompok fungsi, sehingga diperlukan metode deteksi komunitas yang mampu mengidentifikasi struktur overlapping. Penelitian ini bertujuan untuk menerapkan algoritma Local Greedy Extended Dynamic Overlapping Community Detection (GLOD) pada jaringan interaksi protein kanker payudara serta mengevaluasi kualitas komunitas yang terbentuk secara struktural dan fungsional.

Data protein diperoleh melalui API UniProt menggunakan kata kunci breast cancer dan menghasilkan 2.690 entri. Setelah seleksi atribut Gene Symbol dan proses deduplikasi, diperoleh 2.010 protein unik. Jaringan interaksi dibangun menggunakan STRING DB dengan confidence score 0,900 pada organisme Homo sapiens, menghasilkan 1.823 simpul dan 2.646 sisi. Setelah ekstraksi giant component, jaringan akhir terdiri dari 864 simpul dan 2.567 sisi. Algoritma GLOD diterapkan dengan parameter alpha 0,75 dan merging threshold 0,2 untuk mendeteksi komunitas overlapping. Sistem dikembangkan menggunakan Python dengan framework Django dan diuji menggunakan metode whitebox testing untuk memastikan setiap modul berjalan sesuai logika perancangan.

Hasil penelitian menunjukkan bahwa algoritma GLOD berhasil mengidentifikasi 16 komunitas overlapping dengan nilai rata-rata Normalized Node Cut (NNC) sebesar 0,3691. Berdasarkan distribusi kuartil, 4 komunitas termasuk kategori baik, 8 kategori cukup, dan 4 kategori rendah. Secara fungsional, enrichment analysis menunjukkan pengayaan signifikan pada beberapa jalur utama, antara lain Jalur dalam kanker pada Komunitas 1 (54,76%, -29,03), Kanker kolorektal pada Komunitas 2 (66,67%, -26,57), Proteoglikan dalam kanker pada Komunitas 6 (64,29%, -16,33), Proses siklus sel mitotik pada Komunitas 3 (66,67%, -28,23), serta Remodeling kromatin pada Komunitas 10 dan 14 (89,47%, -25,46). Nilai Log10(P) yang negatif menunjukkan signifikansi statistik yang kuat. Seluruh pengujian sistem menunjukkan hasil berhasil tanpa kesalahan fungsional. Penelitian ini menunjukkan bahwa pendekatan deteksi komunitas overlapping berbasis GLOD mampu menghasilkan struktur modular yang konsisten secara topologi dan relevan secara fungsional pada jaringan interaksi protein kanker payudara.

**Kata Kunci:** kanker payudara, GLOD, deteksi komunitas overlap, jaringan interaksi protein, enrichment analysis

# ABSTRACT

*Breast cancer is a complex disease that involves many proteins interacting in different biological pathways. Studying protein interaction networks helps to understand how these proteins are connected and work together. However, biological networks are complex, and one protein can belong to more than one functional group. Therefore, an overlapping community detection method is needed. This research aims to apply the Local Greedy Extended Dynamic Overlapping Community Detection (GLOD) algorithm to a breast cancer protein interaction network and to evaluate the structural and functional quality of the detected communities.*

*Protein data were collected from the UniProt API using the keyword breast cancer, resulting in 2,690 protein entries. After selecting the Gene Symbol attribute and removing duplicate data, 2,010 unique proteins were obtained. The protein interaction network was built using STRING DB with a confidence score of 0.900 for Homo sapiens, producing 1,823 nodes and 2,646 edges. After extracting the giant component, the final network contained 864 nodes and 2,567 edges. The GLOD algorithm was applied using alpha = 0.75 and merging threshold = 0.2 to detect overlapping communities. The system was developed using Python and the Django framework. System testing was conducted using the whitebox method to ensure that all modules worked according to the design.*

*The results show that the GLOD algorithm successfully identified 16 overlapping communities with an average Normalized Node Cut (NNC) value of 0.3691. Based on quartile distribution, 4 communities were categorized as good, 8 as moderate, and 4 as low quality. Functional enrichment analysis showed significant enrichment in several main pathways, such as Pathways in cancer in Community 1 (54.76%, -29.03), Colorectal cancer in Community 2 (66.67%, -26.57), Proteoglycans in cancer in Community 6 (64.29%, -16.33), Mitotic cell cycle process in Community 3 (66.67%, -28.23), and Chromatin remodeling in Communities 10 and 14 (89.47%, -25.46). The negative Log10(P) values indicate strong statistical significance. All system testing results were successful. This study shows that the GLOD-based overlapping community detection approach can identify modular and functionally consistent structures in a breast cancer protein interaction network.*

**Keywords:** *breast cancer, GLOD, overlapping community detection, protein interaction network, enrichment analysis*