

PENGEMBANGAN ALGORITMA *MAXIMUM CLIQUE* *DYNAMIC-MACHINE LEARNING* (MCQD-ML) PADA JARINGAN INTERAKSI PROTEIN-PROTEIN

GILLAND FAUSTA PUTRA ACHYAR



**PROGRAM MAGISTER ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Pengembangan Algoritma *Maximum Clique Dynamic-Machine Learning* (MCQD-ML) pada Jaringan Interaksi Protein-Protein” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2026

Gilland Fausta Putra Achyar
M0503241036

RINGKASAN

GILLAND FAUSTA PUTRA ACHYAR. Pengembangan Algoritma *Maximum Clique Dynamic-Machine Learning* (MCQD-ML) pada Jaringan Interaksi Protein-Protein. Dibimbing oleh WISNU ANANTA KUSUMA, ANNISA dan HERU CAHYA RUSTAMAJI.

Maximum Clique Problem (MCP) merupakan salah satu permasalahan dalam teori graf yang tergolong NP-Hard dan banyak diaplikasikan dalam bioinformatika, termasuk analisis jaringan interaksi protein-protein (PPI). Salah satu pendekatan modern untuk menyelesaikan MCP adalah algoritma *Maximum Clique Dynamic-Machine Learning* (MCQD-ML), yang mengintegrasikan pembelajaran mesin untuk memprediksi parameter batas komputasi (T_{limit}) untuk mengefisienkan langkah pencarian klik. Namun, pada graf jaringan biologis (*Protein-Protein Interaction Network*/PPIN) yang bersifat *scale-free* dan kompleks, performa MCQD-ML sering kali belum optimal. Besarnya ruang pencarian dan banyaknya simpul yang tidak relevan (*noise*) dapat menghambat efisiensi model prediktif. Untuk mengatasi keterbatasan tersebut, penelitian ini bertujuan mengembangkan algoritma *Maximum Clique Dynamic-Machine Learning* (MCQD-ML) yang lebih efisien untuk identifikasi *maximum clique* pada jaringan PPI melalui dua strategi utama, yaitu pemangkasan graf adaptif menggunakan kombinasi *K-Core Decomposition* dan *Graph Attention Network* yang dioptimasi dengan *Particle Swarm Optimization* (GAT-PSO), serta prediksi dinamis parameter T_{limit} yang diformulasikan sebagai tugas klasifikasi multikelas menggunakan arsitektur *Graph Neural Network* (GNN) berbasis fitur sentralitas.

Eksperimen komputasi dilakukan terhadap 12.535 jaringan PPI untuk menguji efektivitas kerangka kerja yang diusulkan. Hasil pengujian menunjukkan bahwa tahap *k-core decomposition* dan *attention-based pruning* berbasis GAT-PSO berhasil mereduksi ruang pencarian secara signifikan, dengan tingkat eliminasi rata-rata sebesar 95,87% simpul dan 91,11% sisi graf awal. Ekstraksi empat metrik sentralitas mempertahankan struktur kluster biologis meskipun ruang pencarian telah direduksi secara signifikan. Pada tahap prediksi T_{limit} , model klasifikasi GIN terpilih berhasil memetakan karakteristik topologi graf sisa ke dalam tiga kelas komputasi secara stabil. Secara operasional, evaluasi sistem terintegrasi (*end-to-end*) pada himpunan graf irisan yang terisolasi secara ketat (*independent unseen test set*) berhasil mencatatkan rata-rata percepatan komputasi (*speed-up*) sebesar 72,19 kali lipat. Percepatan ini dicapai tanpa mengorbankan struktur biologi esensial, dibuktikan oleh nilai *Similarity Size* sebesar 93,20% dan *Jaccard Index* 0,8343. Lebih lanjut, analisis pengayaan fungsional lintas kingdom memvalidasi bahwa *maximum clique* yang ditemukan memiliki signifikansi biomedis. Hal ini terlihat dari kemampuannya dalam memetakan lokalisasi anatomi tumbuhan serta mengisolasi modul ribosom yang rentan menjadi target infeksi patogen. Hasil ini menunjukkan bahwa algoritma yang diusulkan tidak hanya meningkatkan efisiensi komputasi, tetapi juga mempertahankan relevansi biologis dari *maximum clique* yang diidentifikasi.

Kata kunci: *graph attention network*, jaringan interaksi protein-protein, *maximum clique problem*, *particle swarm optimization*, pembelajaran mesin



SUMMARY

GILLAND FAUSTA PUTRA ACHYAR. Development of Maximum Clique Dynamic-Machine Learning (MCQD-ML) Algorithm in Protein-Protein Interaction Networks. Supervised by WISNU ANANTA KUSUMA, ANNISA, and HERU CAHYA RUSTAMAJI.

The Maximum Clique Problem (MCP) is an NP-Hard problem in graph theory with extensive applications in bioinformatics, including the analysis of Protein-Protein Interaction (PPI) networks. A modern approach to solving the MCP is the Maximum Clique Dynamic-Machine Learning (MCQD-ML) algorithm, which integrates machine learning to predict the computational time limit parameter (T_{limit}) to improve the efficiency of clique searching. However, on biological network graphs (Protein-Protein Interaction Networks / PPINs), which are characteristically scale-free and complex, the performance of MCQD-ML is often suboptimal. The massive search space and the abundance of irrelevant nodes (noise) can hinder the efficiency of the predictive model. To address these limitations, this study aims to develop a more efficient Maximum Clique Dynamic-Machine Learning (MCQD-ML) algorithm for identifying maximum cliques in PPI networks through two main strategies, adaptive graph pruning using a combination of K - Core Decomposition and a Graph Attention Network optimized with Particle Swarm Optimization (GAT-PSO), as well as the dynamic prediction of the T_{limit} parameter, formulated as a multi-class classification task using a Graph Neural Network (GNN) architecture based on centrality features.

Computational experiments were conducted on 12,535 PPI networks to evaluate the effectiveness of the proposed framework. The experimental results demonstrate that the k-core decomposition and GAT-PSO-driven attention-based pruning stages successfully and significantly reduced the search space, achieving an average elimination rate of 95,87% of nodes and 91,11% of edges from the initial graphs. The extraction of four centrality metrics preserves the biological cluster structure even though the search space has been significantly reduced. In the T_{limit} prediction stage, the selected GIN classification model successfully and stably mapped the topological characteristics of the residual graphs into three computational time classes. Operationally, the end-to-end evaluation of the integrated system on an intersection set of strictly isolated graphs (independent unseen test set) achieved an average computational speed-up of 72,19 times. This acceleration was achieved without sacrificing essential biological structures, as evidenced by a Similarity Size of 93,20% and a Jaccard Index of 0,8343. Furthermore, cross-kingdom functional enrichment analysis validated that the discovered maximum cliques hold biomedical significance. This is evident from their ability to map plant anatomical localizations and isolate ribosomal modules that are highly susceptible to pathogenic infections. These results demonstrate that the proposed algorithm not only improves computational efficiency but also preserves the biological relevance of the identified maximum cliques.

Keywords: graph attention network, machine learning, maximum clique problem, particle swarm optimization, protein-protein interaction networks



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PENGEMBANGAN ALGORITMA *MAXIMUM CLIQUE DYNAMIC-MACHINE LEARNING* (MCQD-ML) PADA JARINGAN INTERAKSI PROTEIN-PROTEIN

GILLAND FAUSTA PUTRA ACHYAR

Tesis
sebagai salah satu syarat untuk melakukan penelitian dan
penulisan tesis pada
Program Studi Ilmu Komputer

**PROGRAM MAGISTER ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Tesis:
Dr. Toto Haryanto, S.Kom., M.Si.



Judul Tesis : Pengembangan Algoritma *Maximum Clique Dynamic-Machine Learning* (MCQD-ML) pada Jaringan Interaksi Protein-Protein
Nama : Gilland Fausta Putra Achyar
NIM : M0503241036

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Prof. Dr. Eng. Wisnu Ananta Kusuma, S.T., M.T.

Pembimbing 2:
Dr. Eng. Annisa, S.Kom., M.Kom.

Pembimbing 3:
Dr. Heru Cahya Rustamaji, S.Si., M.T.



Diketahui oleh

Ketua Program Studi:
Dr. Toto Haryanto, S.Kom., M.Si.
NIP 198211172014041001

Dekan Sekolah Sains Data, Matematika dan Informatika:
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.
NIP 196607021993021001

Tanggal Ujian:
27 Juli 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur tak terhingga penulis panjatkan kepada Allah Subhānahu wa Ta‘ālā atas segala karunia, kasih sayang, pertolongan, dan hidayah-Nya sehingga karya ilmiah ini berhasil diselesaikan. Sungguh, semua karena Allah, dan hanya Allah tempat terbaik dalam meminta dan memohon pertolongan. Tema penelitian yang dilaksanakan sejak bulan Agustus 2025 sampai bulan Juli 2026 ini berkaitan dengan bioinformatika, dengan judul “Pengembangan Algoritma *Maximum Clique Dynamic-Machine Learning* (MCQD-ML) pada Jaringan Interaksi Protein-Protein”. Ucapan terima kasih secara khusus penulis haturkan kepada:

1. Bapak Prof. Dr. Eng. Wisnu Ananta Kusuma, S.T., M.T., Ibu Dr. Eng. Annisa, S.Kom., M.Kom., dan Bapak Dr. Heru Cahya Rustamaji, S.Si., M.T. selaku komisi pembimbing yang telah meluangkan waktu untuk membimbing, mengarahkan, berdiskusi, serta memberikan tunjuk ajar yang sangat berharga kepada Penulis dari awal hingga selesainya Tesis ini.
2. Bapak Dr. Toto Haryanto, S.Kom., M.Si. selaku Ketua Program Studi Magister Ilmu Komputer sekaligus penguji ujian Tesis yang telah memberikan saran dan masukan membangun untuk penyempurnaan karya tulis ini, dan Bapak Dr. Aziz Kustiyo S.Si., M.Kom. selaku Sekretaris Program Studi, beserta jajaran staf administrasi program studi yang telah banyak memfasilitasi Penulis.
3. Lembaga Pengelola Dana Pendidikan (LPDP) Kementerian Keuangan Republik Indonesia yang telah memberikan dukungan beasiswa penuh dan pendanaan selama masa studi, sehingga Penulis dapat berfokus menyelesaikan pendidikan magister ini dengan baik dan lancar.
4. Keluarga tercinta Penulis; Ayahanda Dani Achyar Kidi dan Ibunda Atikah Achyar, Reffy Rizkia Putri Achyar beserta Hari Sandi, Nazwa Queenza Putri Achyar, serta keponakan tersayang, Anaking Mahasadu Shankara, yang senantiasa mencurahkan kasih sayang luar biasa, dukungan moral maupun materil tanpa henti, serta doa yang tidak pernah putus mengiringi setiap langkah perjuangan Penulis.
5. Sahabat seperjuangan Penulis, teman-teman Magister Ilmu Komputer angkatan 2023-2024 yang selalu menjadi tempat bertukar pikiran dan saling menyemangati.
6. Seluruh pihak lainnya yang tidak dapat Penulis sebutkan satu per satu, yang telah memberikan semangat, bantuan, dan dukungan berarti selama Penulis menempuh pendidikan magister ini.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2026

Gilland Fausta Putra Achyar



DAFTAR ISI

DAFTAR TABEL	ix
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	x
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	4
1.4 Manfaat	4
1.5 Ruang Lingkup	4
TINJAUAN PUSTAKA	5
2.1 Teori Graf	5
2.2 Jaringan Interaksi Protein-Protein	5
2.3 <i>Maximum Clique Problem</i>	6
2.4 Algoritma <i>Maximum Clique Dynamic</i> (MCQD)	7
2.5 Teknik <i>Pruning</i> pada Graf	8
2.6 Algoritma <i>K-Means Clustering</i>	9
2.7 <i>Graph Neural Networks</i> (GNN)	9
2.8 <i>Particle Swarm Optimization</i> (PSO)	12
METODE	13
3.1 Tahapan Penelitian	13
3.2 Praproses Data	13
3.3 Pemodelan <i>Pruning</i> Graf	15
3.4 Pemodelan Prediksi <i>Tlimit</i>	19
3.5 Evaluasi Efisiensi dan Akurasi	24
3.6 <i>Enrichment Analysis</i>	28
3.7 Peralatan Penelitian	28
HASIL DAN PEMBAHASAN	29
4.1 Validasi Algoritma MCQD pada <i>Dataset Benchmark</i>	29
4.2 Hasil Akuisisi dan Karakteristik <i>Dataset PPI</i>	30
4.3 Hasil <i>K-Core Decomposition</i>	31
4.4 Hasil <i>Attention-based Pruning</i>	33
4.5 Pemodelan Prediksi <i>Tlimit</i>	42
4.6 Evaluasi Efisiensi dan Akurasi	55
4.7 Validasi Fungsional Modul Protein Berbasis <i>Enrichment Analysis</i>	58
SIMPULAN DAN SARAN	65
5.1 Simpulan	65
5.2 Saran	65
DAFTAR PUSTAKA	66
LAMPIRAN	73
RIWAYAT HIDUP	77



DAFTAR TABEL

1	Pendekatan dalam penyelesaian <i>maximum clique problem</i>	7
2	Hasil pengujian MCQD pada <i>dataset benchmark</i>	29
3	Karakteristik jaringan interaksi protein-protein	30
4	Hasil reduksi simpul dan sisi pada tahap <i>k-core decomposition</i>	32
5	Matriks fitur topologi dan label pada sampel graf 1001240	34
6	Distribusi pembagian <i>dataset</i> berdasarkan topologi jaringan	35
7	Ringkasan performa terbaik pada pelatihan <i>baseline</i> GAT	36
8	Batasan ruang pencarian <i>hyperparameter</i> PSO pada model GAT	37
9	Konfigurasi <i>hyperparameter</i> optimal hasil pencarian PSO	38
10	Perbandingan performa model GAT dengan optimasi PSO	38
11	Hasil pemangkasan pada sampel jaringan PPI	39
12	Evaluasi preservasi topologi dan akurasi <i>maximum clique</i>	40
13	Hasil evaluasi studi ablasi pemangkasan jaringan	41
14	Identifikasi Tlimit optimal pada sampel jaringan pasca-pemangkasan	43
15	Matriks fitur <i>node-level</i> pada graf 100053	44
16	Distribusi stratifikasi <i>subset dataset</i> GAT	49
17	Strategi pembobotan penalti pada ketimpangan kelas target	50
18	Perbandingan kinerja arsitektur GNN <i>baseline</i>	50
19	Evaluasi performa integrasi LGGAN terhadap arsitektur GNN	51
20	Batasan ruang pencarian <i>Hyperparameter</i> PSO pada model GIN	53
21	Komparasi performa GIN <i>baseline</i> dan GIN-PSO	53
22	Ringkasan hasil pengujian terintegrasi MCQD-ML	55
23	Rangkuman rata-rata kinerja sistem terintegrasi MCQD-ML	56
24	Analisis multiplisitas dan <i>overlapping</i> klik maksimum alternatif	56
25	Ringkasan analisis pengayaan fungsional pada jaringan <i>S. pombe</i>	59

DAFTAR GAMBAR

1	Ilustrasi graf untuk penentuan <i>maximum clique</i>	1
2	Ilustrasi struktur graf (a) Graf tidak berarah (b) Graf berarah	5
3	Representasi graf dengan matriks ketetanggaan	5
4	Representasi struktur modul dalam graf interaksi protein-protein	6
5	Ilustrasi <i>maximum clique</i> dalam graf	6
6	Algoritma <i>MaxCliqueDyn</i> (MCQD)	8
7	Proses reduksi graf awal hingga penghapusan simpul dengan keterhubungan rendah	8
8	<i>Message passing</i> dalam memperbarui simpul	9
9	Skema arsitektur propagasi <i>multi-layer</i> GCN	10
10	Mekanisme atensi dan <i>multi-head attention</i>	10
11	Daya ekspresif antara fungsi agregasi <i>sum</i> , <i>mean</i> , dan <i>max</i>	11
12	Ilustrasi grafis dari algoritma PSO	12
13	Tahapan penelitian	13
14	Alur praproses data interaksi protein dari STRING	15
15	Tahapan <i>pruning</i> untuk penyederhanaan struktur graf	15

16	Proses pemangkasan 2-core pada graf	16
17	Pemangkasan <i>confidence score</i> pada ambang batas 0,5	19
18	Tahapan pembangunan model prediktif <i>Tlimit</i> berbasis GAT-PSO	19
19	Distribusi densitas dan rata-rata derajat jaringan PPI	30
20	Distribusi nilai ambang batas (k) pada tahap pemangkasan	32
21	Distribusi nilai kelas fitur sentralitas simpul jaringan. (A) <i>Degree centrality</i> . (B) <i>Closeness centrality</i> . (C) <i>Eigenvector centrality</i> . (D) <i>Betweenness centrality</i> .	34
22	Kurva konvergensi PSO pada pencarian <i>hyperparameter</i>	37
23	Perbandingan antara model berbasis fitur statistik dengan fitur sentralitas. (A) Kurva ROC. (B) Kurva <i>Precision-Recall</i> .	41
24	Dinamika waktu eksekusi MCQD terhadap variasi parameter <i>Tlimit</i> . (A) Graf 100053. (B) Graf 1000565. (C) Graf 1001240.	43
25	Distribusi densitas pada graf pasca-pemangkasan	45
26	Distribusi <i>Best Tlimit</i> dan Log10 pada <i>dataset</i> PPI	46
27	Analisis <i>elbow method</i> pada penurunan WCSS	48
28	Pemetaan kelas <i>Tlimit</i> berdasarkan <i>K-Means</i> 1D	48
29	Evaluasi kinerja model dengan dan tanpa augmentasi data LGGAN. (A) <i>Macro-Average ROC</i> . (B) <i>Macro-Average Precision-Recall</i> .	52
30	Konvergensi Partikel PSO dalam Optimasi <i>Hyperparameter</i>	54
31	Perbandingan kinerja klasifikasi antara model GIN dan GIN-PSO. (A) <i>Macro-Average ROC</i> . (B) <i>Macro-Average Precision-Recall</i> .	54
32	(A) Peringkat Signifikansi Ontologi Fungsional dan (B) Jejaring Konektivitas Antar-Istilah Pengayaan (<i>Enrichment Network</i>) pada <i>S. pombe</i>	59
33	(A) Peringkat Signifikansi Ontologi Fungsional dan (B) Jejaring Konektivitas Antar-Istilah Pengayaan (<i>Enrichment Network</i>) pada <i>R. norvegicus</i>	61
34	(A) Peringkat Signifikansi Ontologi Fungsional dan (B) Jejaring Konektivitas Antar-Istilah Pengayaan (<i>Enrichment Network</i>) pada <i>A. thaliana</i>	62
35	(A) Peringkat Signifikansi Ontologi Fungsional dan (B) Jejaring Konektivitas Antar-Istilah Pengayaan (<i>Enrichment Network</i>) pada <i>Homo sapiens</i>	63

DAFTAR LAMPIRAN

1	Hasil Analisis Pengayaan Fungsional pada Jaringan <i>Rattus norvegicus</i>	74
2	Hasil Analisis Pengayaan Fungsional pada Jaringan <i>Arabidopsis thaliana</i>	75
3	Hasil Analisis Pengayaan Fungsional pada Jaringan <i>Homo sapiens</i>	76