

PENERAPAN OVERLAPPING COMMUNITY DETECTION PADA HETEROGENEOUS NETWORK TERKAIT KANKER BERBASIS MACHINE LEARNING

RIDHO AL FATH NUSANTARA



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

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Penerapan Overlapping Community Detection pada Heterogeneous Network Terkait Kanker Berbasis Machine Learning” 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 skripsi ini.

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

Bogor, Juli 2026

Ridho Al Fath Nusantara
G6401221030

@aniprati IPB University

ABSTRAK

RIDHO AL FATH NUSANTARA. Penerapan *Overlapping Community Detection* pada *Heterogeneous Network* Terkait Kanker Berbasis *Machine Learning*. Dibimbing oleh WISNU ANANTA KUSUMA dan ANNISA.

Jaringan biologis merepresentasikan interaksi molekuler yang kompleks pada sistem kanker dan umumnya bersifat heterogen karena melibatkan berbagai jenis entitas biologis. Akibatnya, sebagian besar metode deteksi komunitas yang dirancang untuk jaringan homogen kurang optimal dalam menangkap hubungan lintas entitas. Penelitian ini bertujuan membangun representasi jaringan heterogen yang mengintegrasikan interaksi senyawa-protein dan protein-protein terkait kanker serta menerapkannya untuk mendeteksi komunitas yang saling tumpang tindih. Data interaksi biologis diperoleh dari Stanford Network Analysis Project dan data gen kanker dari Kyoto Encyclopedia of Genes and Genomes pada Human Cancer Pathway. Representasi jaringan dipelajari menggunakan embedding berbasis random walk, yaitu metapath2vec. Vektor hasil *embedding* kemudian digunakan untuk membangun jaringan homogen berbobot berdasarkan cosine similarity melalui pendekatan Top-K similarity. Jaringan hasil transformasi dianalisis menggunakan algoritma BigCLAM untuk mengidentifikasi komunitas tumpang tindih. Konfigurasi terbaik diperoleh pada $K=12$ komunitas dengan *overlapping modularity* 0,2366, *internal density* 0,4218, dan *conductance* 0,2578, menunjukkan komunitas yang kompak dan terpisah dengan baik. Sebanyak 33 gen teridentifikasi tumpang tindih antar komunitas, dengan PTEN dan CASP3 berperan sebagai hub multifungsional pada tiga komunitas. Pipeline yang dikembangkan juga berpotensi diadaptasi pada domain nonbiologis dengan struktur data serupa. Selain itu, pendekatan ini dapat mendukung penelitian lanjutan, seperti identifikasi biomarker dan penargetan obat pada kanker.

Kata kunci: BigCLAM, deteksi komunitas tumpang tindih, jaringan heterogen, kanker, metapath2vec

ABSTRACT

RIDHO AL FATH NUSANTARA. Application of *Overlapping Community Detection* on Cancer-Related *Heterogeneous Networks* Based on *Machine Learning*. Supervised by WISNU ANANTA KUSUMA and ANNISA.

Biological networks represent complex molecular interactions in cancer systems and are typically heterogeneous, involving diverse biological entities. As a result, community detection methods designed for homogeneous networks often fail to capture cross-entity relationships. This study aims to construct a heterogeneous network representation integrating compound-protein and cancer-related protein-protein interactions, and apply it to detect overlapping communities. Interaction data were obtained from the Stanford Network Analysis Project, and cancer gene data from the KEGG Human Cancer Pathway. The network representation was learned using the random walk-based embedding method metapath2vec. The



resulting embeddings were used to build a weighted homogeneous network based on cosine similarity via a Top-K similarity approach. This transformed network was then analyzed using the BigCLAM algorithm to identify overlapping communities. The optimal configuration was achieved at $K=12$ communities, with overlapping modularity of 0.2366, internal density of 0.4218, and conductance of 0.2578, indicating compact and well-separated communities. A total of 33 genes were identified as overlapping across communities, with PTEN and CASP3 acting as multifunctional hubs in three communities. The pipeline developed here has potential for adaptation to non-biological domains with similar data structures, and supports further research such as biomarker identification and drug target discovery in cancer.

Keywords: BigCLAM, cancer, heterogeneous network, metapath2vec, overlapping community detection.

- 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 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.

PENERAPAN OVERLAPPING COMMUNITY DETECTION PADA HETEROGENEOUS NETWORK TERKAIT KANKER BERBASIS MACHINE LEARNING

RIDHO AL FATH NUSANTARA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

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



@Hak cipta milik IPB University

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 Skripsi:

1 Dr. Mushthofa, S.Kom, M.Sc

Judul Skripsi : Penerapan Overlapping Community Detection Pada
Heterogeneous Network Terkait Kanker Berbasis Machine Learning

Nama : Ridho Al Fath Nusantara

NIM : G6401221030

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Prof. Dr. Wisnu Ananta Kusuma, S.T, M.T.

19711110 200501 1 005



Pembimbing 2:

Dr. Eng. Annisa, S.Kom., M.Kom.

19790731 200501 2 002



Diketahui oleh

Ketua Program Sarjana Ilmu Komputer:

Dr. Sony Hartono Wijaya, S.Kom., M.Kom.

1981080 9200812 1 002



Tanggal Ujian:

21 Juli 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Judul yang dipilih dalam penelitian yang dilaksanakan sejak bulan November 2025 sampai dengan Juni 2026 ini ialah "Penerapan Overlapping Community Detection pada Heterogeneous Network Terkait Kanker Berbasis Machine Learning"

Dalam penyusunan karya ilmiah ini, penulis memperoleh banyak bantuan, bimbingan, dukungan, dan doa dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih yang sebesar-besarnya kepada kedua orang tua penulis, Bapak Suherman dan Ibu Maimunah, serta keluarga yang senantiasa memberikan doa, kasih sayang, dukungan, dan semangat kepada penulis. Ucapan terima kasih juga penulis sampaikan kepada Prof. Dr. Eng. Wisnu Ananta Kusuma, S.T., M.T. dan Dr. Eng. Annisa, S.Kom., M.Kom., selaku dosen pembimbing yang telah dengan penuh kesabaran memberikan bimbingan, arahan, saran, serta motivasi kepada penulis sejak awal hingga selesainya penelitian ini. Penulis juga mengucapkan terima kasih kepada Dr. Mushthofa, S.Kom., M.Sc. selaku dosen penguji yang telah memberikan masukan, kritik, dan saran yang sangat bermanfaat untuk penyempurnaan penelitian dan penulisan skripsi ini. Rasa terima kasih yang tulus juga penulis sampaikan kepada keluarga IKADA Bogor dan Asrama Putra Indramayu yang senantiasa memberikan dukungan selama proses penyusunan skripsi. Selain itu, penulis mengucapkan terima kasih kepada teman-teman Ilmu Komputer Angkatan 59 atas kebersamaan, bantuan, dan dukungan yang diberikan selama masa perkuliahan maupun penyusunan skripsi. Penulis juga menyampaikan apresiasi kepada semua pihak yang tidak dapat disebutkan satu per satu atas segala bantuan dan dukungan yang telah diberikan sehingga penelitian ini dapat diselesaikan

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

Bogor, Juli 2026

Ridho Al Fath Nusantara



DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	x
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Ruang Lingkup	3
TINJAUAN PUSTAKA	4
2.1 Jaringan Biologis dan Kompleksitas Sistem Kanker	4
2.2 Jaringan Heterogen dalam Analisis Biologis	4
2.3 Deteksi Komunitas Tumpang Tindih pada Jaringan Biologis	4
2.4 Representasi Jaringan Heterogen Berbasis <i>Machine Learning</i>	6
2.5 Evaluasi Kualitas Komunitas pada Jaringan Biologis Heterogen	7
METODE	9
3.1 Data Penelitian	9
3.2 Tahapan Penelitian	9
3.2.1 Pengumpulan Data	10
3.2.2 Praproses Data	10
3.2.3 Konstruksi Jaringan Heterogen	10
3.2.4 Pembelajaran Representasi Node Jaringan Heterogen	11
3.2.5 Konstruksi Jaringan Homogen Berbobot	12
3.2.6 Deteksi dan Evaluasi Komunitas Tumpang Tindih	13
3.2.6.1 Eksperimen variasi jumlah komunitas	13
3.2.6.2 Analisis sensitivitas <i>hyperparameter</i>	15
3.2.6.3 Deteksi komunitas pada konfigurasi optimal	15
3.2.6.4 Analisis karakteristik biologis komunitas	16
3.2.6.5 Rekonstruksi graf heterogen pada komunitas terdeteksi	17
3.2.7 Implementasi Website	17
3.3 Lingkungan Pengembangan	17
HASIL DAN PEMBAHASAN	19
4.1 Pengumpulan Data	19
4.2 Praproses Data	20
4.3 Konstruksi Jaringan Heterogen	21
4.4 Pembelajaran Representasi Node Jaringan Heterogen	22
4.5 Konstruksi Jaringan Homogen Berbobot	23
4.6 Deteksi dan Evaluasi Komunitas Tumpang Tindih	26
4.6.1 Eksperimen Variasi Jumlah Komunitas	26
4.6.2 Analisis Sensitivitas <i>Hyperparameter</i>	28
4.6.3 Deteksi Komunitas pada Konfigurasi Optimal	31



4.6.4	Analisis Karakteristik Biologis Komunitas	34
4.6.5	Rekonstruksi Graf Heterogen pada Komunitas Terdeteksi	39
4.7	Implementasi Website	46
V	SIMPULAN DAN SARAN	52
5.1	Simpulan	52
5.2	Saran	52
	DAFTAR PUSTAKA	54
	LAMPIRAN	57
	RIWAYAT HIDUP	122

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.



DAFTAR TABEL

1	Jenis metapath	11
2	Data PPI dari SNAP	19
3	Data CPI dari SNAP	19
4	Data <i>gene symbol</i> dari KEGG	19
5	Hasil mapping <i>gene symbol</i> ke entrez ID	20
6	Senyawa dengan target protein terbanyak	21
7	Node tetangga terdekat dari node protein referensi (Protein TFG)	23
8	Node tetangga terdekat dari node senyawa referensi (Senyawa Imatinib)	23
9	Protein dengan hubungan terkuat	24
10	Hasil evaluasi variasi k dengan konfigurasi <i>hyperparameter default</i>	26
11	Sensitivitas top-k similarity pada K=11	28
12	Sensitivitas vector size pada K=11	28
13	Sensitivitas epoch pada K=11	29
14	Sensitivitas walk length pada K=11	29
15	<i>Coefficient of variation</i> (CV) <i>hyperparameter</i> dan ranking stabilitas	29
16	Hasil optimasi K dengan konfigurasi <i>hyperparameter</i> optimal	30
17	Distribusi anggota dan gen overlapping di setiap komunitas pada konfigurasi optimal (K=12)	32
18	Distribusi gen overlapping antar komunitas pada K=12	32
19	Top-3 Enriched pathway per komunitas berdasarkan KEGG pathways	35
20	Top-3 Enriched Go biological process per Komunitas berdasarkan gene	36
21	Top-3 Enriched pathway per komunitas berdasarkan reactome pathways	37
22	Distribusi label eksplisit kanker terhadap sub-pathway/proses pada	38
23	Distribusi senyawa per komunitas (K=12)	39
24	Afiliasi protein target senyawa pada komunitas K1	43

DAFTAR GAMBAR

1	Tahapan penelitian	9
2	Topologi graf heterogen	22
3	Topologi graf homogen berbobot	25
4	Tren <i>overlapping modularity</i> terhadap variasi jumlah komunitas pada konfigurasi default	28
5	Deteksi komunitas overlapping pada konfigurasi optimal (K=12)	31
6	Distribusi subgraf komunitas overlapping (K=12)	34
7	Komunitas 1 dengan overlay node senyawa (K=12)	42
8	Tampilan website	50

DAFTAR LAMPIRAN

Data konversi <i>gene symbol</i> ke entrez ID	58
---	----

2	Keanggotaan gen pada tiap komunitas	66
3	Data konversi pubchem CID ke nama senyawa	75
4	Afiliasi protein target senyawa pada seluruh komunitas	78

@Hak cipta milik IPB University

IPB University





@Hak cipta milik IPB University

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.