

Analisis Proyeksi Graf Bipartit dan Sentralitas Jaringan untuk Deteksi Komunitas Otomatis pada Data Gejala-Penyakit

Zidane Uland Fakhry – 13525132

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jalan Ganesha 10 Bandung

E-mail: zidane.uland@gmail.com, 13525132@std.stei.itb.ac.id

Abstrak—Proses identifikasi penyakit saat ini sering kali menghadapi kendala akibat adanya kemiripan gejala antara satu penyakit dengan penyakit lainnya. Penelitian ini bertujuan untuk memetakan dan mengelompokkan hubungan gejala dan penyakit secara otomatis menggunakan pendekatan teori graf. Data kesehatan direpresentasikan ke dalam bentuk graf bipartit berbobot, yang kemudian ditransformasikan menggunakan metode proyeksi graf menjadi Jaringan Penyakit (*Disease Network*). Analisis struktur topologi jaringan dilakukan melalui penghitungan metrik sentralitas, meliputi *Degree Centrality*, *Betweenness Centrality*, dan *Closeness Centrality* untuk mengidentifikasi penyakit yang memiliki kemiripan gejala tertinggi dan bertindak sebagai jembatan klinis. Selanjutnya, Algoritma Louvain diterapkan untuk mendeteksi komunitas secara otomatis berdasarkan nilai modularitas graf. Hasil penelitian menunjukkan bahwa pendekatan berbasis graf ini mampu mengelompokkan penyakit secara otomatis ke dalam kluster-kluster spesifik yang sejalan dengan klasifikasi medis asli berdasarkan kemiripan gejalanya. Model ini diharapkan dapat dikembangkan lebih lanjut sebagai alternatif sistem pendukung keputusan klinis untuk membantu meminimalkan risiko salah diagnosis.

Kata Kunci— *Teori Graf, Graf Bipartit, Proyeksi Graf, Sentralitas Jaringan, Deteksi Komunitas, Algoritma Louvain.*

I. PENDAHULUAN

Kesehatan merupakan salah satu aspek penting dalam kehidupan manusia. Seiring berkembangnya zaman, tantangan di dunia medis semakin kompleks, salah satunya disebabkan oleh kemunculan berbagai jenis penyakit dengan karakteristik gejala yang serupa. Proses identifikasi atau diagnosis penyakit seringkali menghadapi kendala ketika beberapa penyakit yang berbeda memiliki kemiripan gejala. Baik pasien maupun tenaga medis terkadang kesulitan membedakan diagnosis awal jika hanya melihat gejala-gejala umum yang bermunculan secara bersamaan. Oleh karena itu, diperlukan suatu metode analisis data kesehatan yang mampu memetakan hubungan terstruktur antara gejala dan penyakit secara menyeluruh.

Dalam matematika diskrit, data relasional yang menghubungkan dua entitas berbeda seperti ini dapat dimodelkan secara matematis menggunakan teori graf [1]. Struktur data kesehatan yang terdiri dari himpunan gejala dan himpunan penyakit dapat direpresentasikan ke dalam bentuk graf bipartit (*bipartite graph*). Pada model awal, gejala dan penyakit bertindak sebagai simpul (*node*), sedangkan hubungan di antara keduanya direpresentasikan sebagai sisi (*edge*) [2]. Analisis dasar pada graf bipartit umumnya hanya menggunakan konsep derajat (*degree*) untuk melihat seberapa sering suatu gejala muncul atau seberapa banyak gejala yang dimiliki suatu penyakit. Namun, analisis *degree* ini belum mampu memperlihatkan gambaran umum mengenai kedekatan atau kemiripan karakteristik antar-penyakit itu sendiri.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan pendekatan analisis jaringan yang lebih mendalam melalui metode proyeksi graf (*one-mode projection*). Graf bipartit awal ditransformasikan menjadi sebuah Jaringan Penyakit (*Disease Network*), di mana setiap simpul hanya merepresentasikan penyakit, dan bobot (*weight*) dari sisi yang menghubungkannya dihitung berdasarkan jumlah irisan gejala yang sama [2]. Melalui graf proyeksi berbobot ini, analisis topologi jaringan yang lebih kompleks dapat dilakukan menggunakan metrik sentralitas seperti *Degree Centrality*, *Betweenness Centrality*, dan *Closeness Centrality* untuk menemukan penyakit-penyakit kunci yang berperan sebagai jembatan klinis [3].

Lebih jauh lagi, struktur jaringan penyakit yang terbentuk dapat dimanfaatkan untuk melakukan pengelompokkan secara otomatis melalui teknik deteksi komunitas (*community detection*) menggunakan Algoritma Louvain [4]. Algoritma ini bekerja dengan cara mengoptimalkan nilai modularitas graf untuk mengelompokkan penyakit-penyakit ke dalam kluster tertentu berdasarkan kemiripan topologi gejalanya.

Melalui penerapan konsep-konsep ini, penelitian ini bertujuan untuk memodelkan data relasi gejala-penyakit dari graf bipartit menjadi graf proyeksi Jaringan Penyakit yang berbobot. Selain itu, penelitian ini dilakukan untuk

mengidentifikasi karakteristik topologi jaringan menggunakan metrik sentralitas kompleks guna menemukan penyakit yang memiliki tingkat kemiripan gejala tertinggi, serta menganalisis kinerja Algoritma Louvain dalam melakukan deteksi komunitas secara otomatis. Dengan pendekatan ini, hasil penelitian diharapkan dapat memberikan sudut pandang baru dalam analisis data medis terstruktur serta berpotensi menjadi dasar pengembangan sistem pendukung keputusan klinis untuk meminimalkan risiko kesalahan diagnosis.

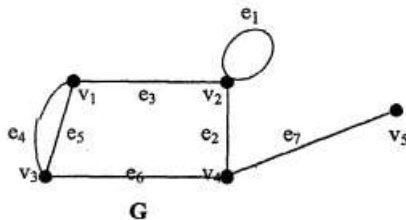
II. LANDASAN TEORI

A. Teori Graf dan Graf Bipartit

Graf didefinisikan sebagai pasangan himpunan $(G = (V, E))$, dimana (V) merupakan himpunan titik atau simpul (*vertex/node*) dan (E) merupakan himpunan sisi (*edge*) yang menghubungkan pasangan titik tertentu. Dalam representasi graf, titik digunakan untuk merepresentasikan objek, sedangkan sisi digunakan untuk menunjukkan hubungan antar objek.

Graf dapat direpresentasikan dalam bentuk diagram yang terdiri atas kumpulan simpul dan sisi. Representasi graf banyak digunakan dalam berbagai bidang, seperti jaringan komputer, media sosial, sistem rekomendasi, serta analisis data kesehatan. Pada penelitian ini, teori graf digunakan untuk merepresentasikan hubungan antara gejala dan penyakit dalam bentuk simpul dan sisi sehingga hubungan antar keduanya dapat dianalisis secara lebih terstruktur.

Gambar berikut menunjukkan contoh ilustrasi diagram graf yang terdiri atas simpul dan sisi.



Gambar 1. Contoh ilustrasi diagram graf yang terdiri atas simpul dan sisi, diambil dari [1].

Graf bipartit merupakan graf khusus di mana himpunan simpulnya dapat dibagi menjadi dua kelompok berbeda, sehingga setiap sisi hanya menghubungkan simpul dari kelompok yang berbeda. Dengan kata lain, tidak terdapat hubungan antar simpul dalam kelompok yang sama. Secara umum, graf bipartit dapat dinyatakan sebagai:

$$G = (X, Y, E)$$

dengan:

- X dan Y sebagai dua himpunan simpul yang berbeda,
- E sebagai himpunan sisi yang menghubungkan simpul pada kedua himpunan tersebut.

Pada penelitian ini, graf bipartit digunakan untuk merepresentasikan hubungan antara gejala dan penyakit. Gejala dan penyakit direpresentasikan sebagai dua kelompok simpul yang berbeda, sedangkan hubungan antara gejala dan penyakit direpresentasikan sebagai sisi. Dalam representasi tersebut, simpul gejala hanya terhubung dengan simpul penyakit dan tidak memiliki hubungan dengan simpul gejala lainnya. Dengan representasi tersebut, hubungan antar gejala dan penyakit dapat dianalisis secara lebih terstruktur untuk mengidentifikasi gejala yang paling dominan.

B. Proyeksi Graf (One-Mode Projection)

Graf bipartit memberikan representasi relasi antar dua entitas yang berbeda, namun kurang optimal untuk menganalisis hubungan internal di dalam satu entitas saja. Untuk melihat bagaimana keterkaitan antar-penyakit berdasarkan gejala yang tumpang tindih, dilakukan proyeksi graf satu mode (*one-mode projection*).

Jika terdapat graf bipartit $G = (V_1 \cup V_2, E)$, maka proyeksi graf terhadap himpunan V_2 (penyakit) akan menghasilkan graf baru $G' = (V_2, E')$. Pada graf proyeksi ini, dua simpul $u, v \in V_2$ akan terhubung oleh sebuah sisi jika dan hanya jika kedua simpul tersebut bertetangga dengan setidaknya satu simpul yang sama di V_1 .

Pada penelitian ini, proyeksi yang digunakan adalah graf berbobot (*weighted graph*). Bobot (w) pada sisi yang menghubungkan simpul penyakit u dan v ditentukan oleh jumlah irisan (*intersection*) simpul gejala yang sama pada graf bipartit awal, yang dapat dirumuskan sebagai:

$$w(u, v) = |N(u) \cap N(v)|$$

dimana $N(u)$ dan $N(v)$ adalah himpunan simpul tetangga (gejala) dari masing-masing penyakit pada graf bipartit awal.

C. Metrik Sentralitas Jaringan (Network Centrality)

Sentralitas adalah metrik struktural yang digunakan untuk menentukan tingkat kepentingan atau posisi strategis suatu simpul di dalam jaringan topologi graf. Terdapat tiga metrik sentralitas yang digunakan dalam analisis jaringan penyakit ini:

- *Degree Centrality*: metrik paling mendasar yang mengukur jumlah sisi yang terhubung secara langsung pada suatu simpul. Pada graf proyeksi berbobot, *degree centrality* yang dinormalisasi untuk simpul i didefinisikan sebagai:

$$C_D(i) = \frac{\sum_j w(i, j)}{|V| - 1}$$

di mana $\sum_j w(i, j)$ melambangkan bobot sisi antara simpul i dan j , sedangkan $|V|$ adalah total simpul dalam graf proyeksi.

- *Betweenness Centrality*: mengukur seberapa sering suatu simpul dilewati oleh lintasan terpendek (*shortest path*) yang menghubungkan seluruh pasangan simpul lain di dalam graf. Simpul dengan *betweenness* tinggi bertindak sebagai jembatan (*bridge*) informasi. Rumus matematisnya adalah:

$$C_B(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

di mana σ_{st} adalah total lintasan terpendek dari simpul s ke simpul t , dan $\sigma_{st}(i)$ adalah jumlah lintasan terpendek dari s ke t yang melewati simpul i .

- *Closeness Centrality*: Mengukur seberapa dekat suatu simpul dengan seluruh simpul lainnya berdasarkan rata-rata jarak lintasan terpendek. Nilai yang tinggi menunjukkan bahwa simpul tersebut berada di pusat jaringan topologi. Rumusnya adalah:

$$C_C(i) = \frac{|V| - 1}{\sum_{j \neq i} d(i, j)}$$

di mana $d(i, j)$ menyatakan jarak lintasan terpendek antara simpul i dan j .

D. Deteksi Komunitas dan Algoritma Louvain

Deteksi komunitas (*community detection*) dalam teori jejaring bertujuan untuk memecah graf ke dalam beberapa kelompok (*cluster* atau *community*). Sebuah komunitas ideal dalam graf ditandai dengan struktur simpul di dalam kelompok yang memiliki tingkat kepadatan hubungan (*edge*) yang sangat tinggi, namun memiliki hubungan yang renggang dengan simpul di luar kelompok tersebut.

Algoritma Louvain adalah metode praktis berbasis optimasi nilai modularitas (Q) untuk mengekstraksi struktur komunitas secara cepat pada graf berukuran besar. Modularitas merupakan nilai skalar antara -1 hingga 1 yang mengukur kekuatan relatif dari pembagian komunitas dibandingkan dengan graf acak. Secara matematis, formulasi modularitas didefinisikan sebagai:

$$Q = \frac{1}{2m} \sum_{i,j} \left[w(i, j) - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

di mana $w(i, j)$ adalah bobot sisi antara simpul i dan j , k_i dan k_j adalah jumlah bobot sisi yang terhubung ke simpul masing-masing, m adalah total bobot seluruh sisi pada graf, dan $\delta(c_i, c_j)$ adalah fungsi delta Kronecker yang bernilai 1 jika

simpul i dan j berada di komunitas yang sama, serta bernilai 0 untuk kondisi sebaliknya.

Algoritma Louvain bekerja melalui dua tahap yang diulang secara iteratif:

1) *Tahap Optimasi Modularitas Lokal*: Setiap simpul diposisikan ke dalam komunitasnya sendiri secara modular. Secara bergiliran, simpul i dipindahkan ke komunitas tetangganya hanya jika pemindahan tersebut menghasilkan peningkatan nilai modularitas (ΔQ) terbesar yang positif. Proses ini berhenti jika tidak ada lagi perpindahan simpul yang dapat meningkatkan nilai modularitas.

2) *Tahap Agregasi Jaringan*: Simpul-simpul yang berada di dalam komunitas yang sama digabungkan (*coarse-grained*) menjadi satu simpul meta (*meta-node*) baru. Sisi antar simpul meta dihitung berdasarkan akumulasi bobot sisi asli antar-komunitas. Setelah graf baru yang lebih ringkas terbentuk, algoritma akan kembali menjalankan tahap pertama hingga nilai modularitas global mencapai titik maksimum dan konvergen.

III. METODOLOGI PENELITIAN

Metodologi pengerjaan dalam penelitian ini dirancang secara sistematis melalui lima tahapan utama, dimulai dari pengumpulan data hingga visualisasi hasil komunitas. Seluruh tahapan eksperimen ini diimplementasikan menggunakan bahasa pemrograman Python 3.12 dengan memanfaatkan pustaka utama seperti *Pandas* untuk manipulasi data, *NetworkX* untuk pemodelan topologi graf, serta Python-Louvain untuk eksekusi algoritma deteksi komunitas.

A. Pengumpulan Data dan Pra-pemrosesan Data

Penelitian ini menggunakan dataset kesehatan berupa tabel biner relasi gejala dan penyakit yang diperoleh dari platform Kaggle dengan judul *Disease Prediction Using Machine Learning* [3]. Dataset tersedia dalam format CSV dan terdiri dari ratusan baris data pasien yang mencakup 41 jenis penyakit (*prognosis*) unik dan 132 indikator gejala (*symptoms*).

TABLE I. TABEL RINGKASAN KARAKTERISTIK DATASET AWAL

Parameter	Kuantitas/Keterangan
Sumber Data	Kaggle Repository (Secondary Data)
Total Baris Data	4920 Baris
Jumlah Node Penyakit	41 Jenis Penyakit
Jumlah Node Gejala	132 Indikator Gejala
Format Nilai Relasi	Biner (0: Tidak ada, 1: Memiliki Gejala)

Sebelum memodelkan graf, tahap pra-pemrosesan (*preprocessing*) dilakukan untuk membersihkan struktur data. Kolom kosong atau kolom tidak valid yang tidak memiliki nama akibat kesalahan *export* (seperti kolom berformat *Unnamed*) dideteksi dan dieliminasi menggunakan pustaka *Pandas*. Selanjutnya, karena dataset asli berisi banyak baris rekaman

medis untuk penyakit yang sama, dilakukan proses agregasi data menggunakan fungsi *grouping* berdasarkan nilai maksimum (*groupby().max()*). Proses agregasi ini krusial untuk menghasilkan satu baris relasi unik dan biner (0 atau 1) untuk setiap jenis penyakit tanpa adanya duplikasi data klinis.

B. Representasi Graf Bipartit Awal

Setelah data siap dan bersih, langkah selanjutnya adalah memodelkan data tersebut ke dalam struktur graf. Karena dataset ini menghubungkan dua objek yang berbeda—yaitu gejala dan penyakit—maka model matematika yang paling tepat adalah menggunakan Graf Bipartit.

Proses pembuatan graf ini dilakukan secara terprogram menggunakan bantuan pustaka *NetworkX*. Langkah pertama yang dilakukan adalah membuat sebuah objek graf kosong. Kemudian, sistem akan mendaftarkan dua kelompok simpul (*node*) secara terpisah. Kelompok pertama adalah simpul untuk gejala yang berjumlah 132 simpul, dan kelompok kedua adalah simpul untuk penyakit yang berjumlah 41 simpul. Untuk membedakan kedua kelompok ini di dalam program, simpul gejala diberi tanda berupa atribut *bipartite=0*, sedangkan simpul penyakit diberi tanda *bipartite=1*.

Setelah semua simpul terdaftar, langkah berikutnya adalah membuat hubungan atau sisi (*edge*) yang menghubungkan simpul gejala dengan simpul penyakit. Proses ini dilakukan dengan cara membaca ulang tabel data biner yang sudah dibersihkan tadi. Program akan mengecek setiap baris penyakit dan kolom gejala satu per satu. Jika pada tabel data ditemukan angka 1, itu artinya gejala tersebut dialami oleh penyakit yang bersangkutan. Berdasarkan angka 1 tersebut, program akan menarik sebuah garis lurus (sisi) tanpa arah yang menghubungkan simpul gejala dengan simpul penyakitnya. Jika angka pada tabel adalah 0, maka program tidak akan membuat garis apa pun.

C. Proyeksi Graf Menjadi Jaringan Penyakit Berbobot

Model graf bipartit pada tahap sebelumnya sangat baik untuk melihat hubungan langsung antara gejala dan penyakit. Namun, model tersebut kurang efektif jika kita ingin menganalisis hubungan kemiripan antara sesama penyakit. Oleh karena itu, pada tahap ketiga ini, graf bipartit tersebut diubah bentuknya melalui metode proyeksi graf (*one-mode projection*).

Proses proyeksi ini dilakukan menggunakan fungsi khusus dari *NetworkX* bernama *weighted_projected_graph*. Fokus utama dari tahap ini adalah mengisolasi simpul penyakit saja dan mengeliminasi simpul gejala dari struktur jaringan. Hasil dari proyeksi ini adalah sebuah graf baru yang isinya hanya berupa simpul-simpul penyakit, yang dinamakan Jaringan Penyakit (*Disease Network*).

Di dalam graf baru ini, dua buah simpul penyakit akan otomatis terhubung oleh sebuah garis jika dan hanya jika kedua penyakit tersebut memiliki minimal satu gejala yang sama pada graf bipartit awal. Hubungan antar-penyakit ini bersifat berbobot (*weighted*). Bobot pada garis tersebut dihitung berdasarkan seberapa banyak jumlah gejala yang beririsan atau sama di antara kedua penyakit tersebut. Sebagai contoh, jika Penyakit A dan Penyakit B sama-sama memiliki 5 gejala yang persis sama,

maka garis yang menghubungkan simpul Penyakit A dan Penyakit B akan memiliki nilai bobot sebesar 5. Semakin besar nilai bobotnya, berarti kedua penyakit tersebut memiliki kemiripan klinis yang semakin tinggi.

D. Penghitungan Metrik Sentralitas Jaringan

Setelah Jaringan Penyakit yang berbobot berhasil terbentuk, tahap keempat adalah melakukan analisis struktur jaringannya. Analisis ini bertujuan untuk melihat posisi dan tingkat pentingnya suatu penyakit di dalam keseluruhan jaringan berdasarkan bentuk hubungannya. Untuk melakukan hal ini, dihitunglah tiga jenis nilai sentralitas (*centrality*) pada setiap simpul penyakit:

1) *Degree Centrality*: Perhitungan ini digunakan untuk melihat seberapa banyak total hubungan berbobot yang dimiliki oleh suatu penyakit ke seluruh jaringan. Penyakit dengan nilai degree yang tinggi mengindikasikan bahwa penyakit tersebut berbagi banyak gejala umum dengan banyak penyakit lainnya.

2) *Betweenness Centrality*: Perhitungan ini bertujuan untuk mencari penyakit yang sering menjadi "jembatan" atau penghubung lintasan terpendek di antara kelompok penyakit lain. Nilai ini sangat penting karena penyakit dengan betweenness tinggi bisa menjadi titik kritis yang menghubungkan dua kelompok penyakit yang sebenarnya berbeda jenis.

3) *Closeness Centrality*: Perhitungan ini digunakan untuk mengukur seberapa dekat jarak suatu penyakit dengan seluruh penyakit lain di dalam jaringan. Penyakit yang memiliki nilai closeness tinggi berarti posisinya berada di tengah-tengah (pusat) struktur jaringan dan memiliki akses tercepat ke simpul lainnya.

Semua hasil perhitungan dari ketiga metrik sentralitas ini kemudian dikumpulkan dan dimasukkan ke dalam sebuah tabel (*DataFrame Pandas*) agar nantinya bisa diurutkan dari nilai yang paling tinggi hingga terendah untuk dianalisis lebih lanjut.

E. Penghitungan Metrik Sentralitas Jaringan

Tahap terakhir dalam metodologi ini adalah melakukan pengelompokan penyakit secara otomatis menggunakan teknik deteksi komunitas (*community detection*) dengan Algoritma Louvain. Tujuan utama dari tahap ini adalah mengelompokkan penyakit-penyakit yang memiliki karakteristik gejala mirip ke dalam satu kelompok (komunitas) yang sama tanpa perlu dikelompokkan secara manual oleh manusia.

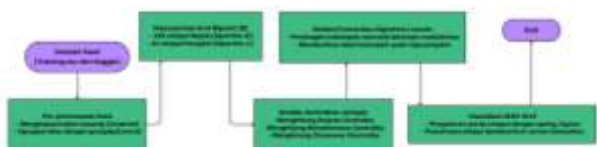
Proses pengelompokan ini dijalankan lewat modul program *community_louvain.best_partition* dengan memasukkan objek Jaringan Penyakit berbobot yang sudah dibuat sebelumnya. Algoritma Louvain akan bekerja secara berulang-ulang dalam dua tahapan utama. Pertama, algoritma akan mencoba memindahkan satu simpul penyakit ke kelompok tetangganya dan menghitung apakah pemindahan tersebut membuat pembagian kelompok menjadi lebih bagus atau tidak (diukur lewat nilai matematika bernama modularitas). Kedua, setelah kelompok-kelompok kecil yang rapi terbentuk, algoritma akan menyatukan kelompok tersebut menjadi satu simpul besar, lalu

mengulangi proses pencarian kelompok lagi sampai pembagian kelompoknya sudah maksimal dan tidak bisa ditingkatkan lagi.

Setelah proses perhitungan selesai, setiap penyakit akan mendapatkan label nomor kelompoknya masing-masing. Untuk mempermudah dalam membaca hasil pengelompokan tersebut, sistem akan membuat gambar visualisasi graf menggunakan pustaka *Matplotlib*. Posisi simpul-simpul penyakit akan diatur secara otomatis menggunakan visualisasi dinamis bernama *spring_layout* (gaya pegas), di mana penyakit yang gejalanya mirip akan otomatis saling mendekat di dalam gambar. Terakhir, setiap simpul penyakit akan diberi warna yang berbeda-beda secara otomatis berdasarkan nomor komunitas hasil perhitungan Algoritma Louvain, sehingga kelompok penyakit yang terbentuk dapat langsung terlihat dengan jelas secara visual.

Seluruh rangkaian tahapan pengerjaan sistem mulai dari pemrosesan data mentah hingga tahap visualisasi akhir ini dirangkum ke dalam diagram alir pada Gambar 2. berikut.

Alur Kerja Pengerjaan Sistem



Gambar 2. Alur kerja pengerjaan sistem.

IV. HASIL DAN PEMBAHASAN

A. Hasil Pembentukan Graf dan Proyeksi Graf

Eksperimen dimulai dengan memasukkan data dari *file Training.csv*. Setelah melewati tahap pembersihan data dari kolom kosong dan penggabungan baris duplikat, program berhasil membangun Graf Bipartit awal yang utuh. Graf bipartit ini memiliki total 173 simpul (terdiri dari 132 simpul gejala dan 41 simpul penyakit) yang saling dihubungkan oleh sisi-sisi biner.

Graf bipartit tersebut kemudian ditransformasikan menggunakan metode proyeksi satu mode menjadi Jaringan Penyakit. Dari hasil proyeksi, terbentuk sebuah graf baru yang hanya berisi 41 simpul penyakit. Seluruh simpul penyakit ini saling terhubung satu sama lain jika mereka berbagi minimal satu gejala yang sama.

Berdasarkan hasil pengurutan bobot garis (*edge weight*) tertinggi pada graf proyeksi, ditemukan kombinasi pasangan penyakit yang memiliki tingkat tumpukan kombinasi gejala paling parah di dalam dataset. Pasien dengan gejala-gejala ini sangat rawan menerima kesalahan diagnosis awal karena kemiripan klinisnya yang sangat tinggi. Tiga pasangan penyakit dengan bobot irisan gejala tertinggi ditunjukkan pada TABLE II berikut.

TABLE II. TOP 3 GEJALA DAN PENYAKIT BERDASARKAN NILAI DERAJAT PADA GRAF BIPARTIT

Kategori	Nama Simpul (Node)	Nilai Derajat (Degree)	Analisis Kontekstual
Top Gejala	vomiting (Muntah)	17	Menunjukkan gejala yang paling tidak spesifik dan muncul di hampir sebagian besar rumpun penyakit berbeda dalam dataset.
	fatigue (Kelelahan)	17	
	high fever (Demam Tinggi)	12	
Top Penyakit	Common Cold (Flu Biasa)	17	Menunjukkan bahwa penyakit ini memiliki kompleksitas klinis tertinggi karena membutuhkan kombinasi 17 gejala berbeda untuk terdiagnosis.
	Tuberculosis (TBC)	16	
	Dengue (Demam Berdarah)	14	

B. Analisis Sentralitas Jaringan Penyakit

Analisis struktural dilakukan untuk mencari tahu penyakit mana saja yang memiliki posisi paling strategis atau menjadi pusat penyebaran informasi gejala di dalam jaringan. Hasil perhitungan tiga metrik sentralitas (*Degree*, *Betweenness*, dan *Closeness*) untuk lima penyakit teratas dirangkum dalam TABLE III.

TABLE III. NILAI TOP 5 SENTRALITAS SIMPUL JARINGAN PENYAKIT

Penyakit	Degree Centrality	Betweenness Centrality	Closeness Centrality
Dengue	0.925	0.097499	0.930233
GERD	0.550	0.064413	0.689655
Typhoid	0.800	0.063215	0.816327
Drug Reaction	0.275	0.061244	0.563380
Impetigo	0.375	0.052696	0.606061

Berdasarkan data pada TABLE III, karakteristik topologi dari Jaringan Penyakit yang terbentuk dapat dianalisis secara mendalam sebagai berikut:

1) *Analisis Degree Centrality*: Penyakit *Dengue* mencatatkan nilai *Degree Centrality* tertinggi, yaitu sebesar 0.925, disusul oleh *Typhoid* sebesar 0.800. Nilai derajat yang sangat tinggi pada graf proyeksi berbobot ini mengindikasikan bahwa *Dengue* dan *Typhoid* berbagi banyak sekali gejala klinis yang sama dengan mayoritas penyakit lain di dalam jaringan.

Secara medis, hal ini merepresentasikan sifat penyakit infeksi sistemik yang manifestasi klinis awalnya didominasi oleh gejala-gejala umum (seperti demam tinggi, kelelahan, dan sakit kepala). Akibatnya, simpul-simpul ini memiliki konektivitas yang sangat padat ke seluruh area graf.

2) *Analisis Betweenness Centrality*: Metrik ini menjadi poin paling krusial dalam mendeteksi penyakit yang bertindak sebagai jembatan antar-kelompok (*clinical bridge*). *Dengue* memimpin dengan nilai 0.097499, yang berarti selain memiliki banyak gejala umum, penyakit ini menjadi rute lintasan terpendek utama yang menghubungkan berbagai kluster penyakit berbeda. Namun, fenomena topologi yang jauh lebih menarik terlihat pada penyakit seperti *GERD* (0.064413), *Drug Reaction* (0.061244), dan *Impetigo* (0.052696). Ketiga penyakit ini memiliki nilai *Degree Centrality* yang relatif rendah (misalnya *Drug Reaction* hanya 0.275), namun berhasil masuk ke dalam jajaran Top 5 *Betweenness Centrality*. Secara matematis dan klinis, ini membuktikan bahwa meskipun penyakit-penyakit tersebut tidak memiliki banyak gejala umum, mereka memiliki kombinasi gejala spesifik yang unik sehingga mampu bertindak sebagai sekering atau jembatan penghubung antara satu rumpun spesialisasi medis dengan rumpun medis lainnya. Jika terjadi galat diagnosis pada simpul jembatan ini, prediksi klinis dapat melompat ke kluster penyakit yang sama sekali keliru.

3) *Analisis Closeness Centrality*: Selaras dengan metrik derajat, *Dengue* dan *Typhoid* menempati posisi teratas dengan nilai *Closeness Centrality* masing-masing sebesar 0.930233 dan 0.816327. Tingginya nilai kedekatan ini menunjukkan bahwa simpul penyakit tersebut berada di posisi makro paling tengah (pusat) dari topologi graf. Secara struktural, jarak lintasan terpendek dari *Dengue* untuk menjangkau seluruh simpul penyakit lain di dalam jaringan adalah yang paling minimal, sehingga menjadikannya sebagai simpul yang paling mudah diakses atau dipengaruhi oleh penyebaran karakteristik gejala dari kluster mana pun.

C. Analisis Komunitas Hasil Algoritma Louvain

Setelah metrik sentralitas dihitung, Algoritma Louvain berhasil mengekstraksi struktur pengelompokan penyakit secara otomatis. Algoritma ini membagi 41 penyakit ke dalam lima kelompok komunitas yang berbeda, dengan total nilai modularitas global sebesar 0.1743. Nilai modularitas yang positif dan relatif tinggi ini membuktikan bahwa pembagian kluster oleh algoritma memiliki struktur pengelompokan yang kuat (padat di dalam, renggang di luar).

Menariknya, meskipun Algoritma Louvain murni bekerja menggunakan perhitungan matematika topologi graf tanpa dibekali pengetahuan medis sama sekali, kelompok-kelompok yang terbentuk secara otomatis ini ternyata sangat sejalan dengan klasifikasi spesialisasi medis di dunia nyata. Berikut adalah rincian beberapa komunitas yang terbentuk:

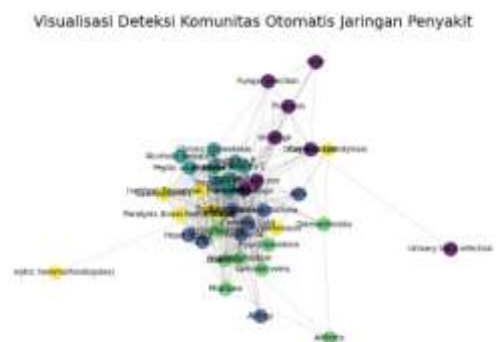
- *Komunitas 1 (Rumpun Infeksi Sistemik dan Dermatologi)*: Komunitas ini mencakup total 8 penyakit, di antaranya adalah *Fungal infection*, *Drug Reaction*,

Chicken pox, *Dengue*, dan *Acne*. Secara klinis, penyakit-penyakit ini berkumpul dalam satu kluster karena berbagi irisan gejala permukaan kulit (seperti ruam atau gatal) serta manifestasi sistemik akut seperti demam tinggi dan kelelahan (*fatigue*).

- *Komunitas 2 (Rumpun Penyakit Kronis, Imun, dan Respirasi)*: Terdiri dari 8 penyakit, termasuk *Allergy*, *GERD*, *AIDS*, *Bronchial Asthma*, dan *Tuberculosis*. Pengelompokan otomatis ini terjadi karena adanya tumpang tindih pada gejala-gejala gangguan pernapasan (seperti batuk dan sesak napas) serta penurunan kondisi fisik secara kronis yang menjadi indikator utama dari penyakit-penyakit tersebut.
- *Komunitas 3 (Rumpun Hepato-Gastroenterologi)*: Kluster ini diisi oleh 9 penyakit, seperti *Chronic cholestasis*, *Peptic ulcer disease*, *Jaundice*, *Hepatitis A*, dan *Hepatitis B*. Komunitas ini merupakan salah satu hasil partisi yang paling akurat, karena secara otomatis menyatukan penyakit-penyakit gangguan hati dan saluran cerna yang terikat kuat oleh gejala spesifik yang sama, yaitu menguningnya kulit/mata (*jaundice*), mual, muntah, dan nyeri abdomen.
- *Komunitas 4 (Rumpun Endokrin dan Gangguan Metabolik)*: Meliputi 8 penyakit, di antaranya *Diabetes*, *Migraine*, *Varicose veins*, *Hypothyroidism*, dan *Hyperthyroidism*. Penyakit-penyakit ini dikelompokkan bersama karena memiliki kesamaan karakteristik gejala pada gangguan fungsi metabolisme tubuh, seperti fluktuasi berat badan yang ekstrem, letargi, dan kelelahan konstan.
- *Komunitas 5 (Rumpun Neurologi dan Kondisi Akut)*: Terdiri dari 8 penyakit, seperti *Gastroenteritis*, *Hypertension*, *Cervical spondylosis*, *Paralysis (brain hemorrhage)*, dan *Malaria*. Kluster ini mengagregasikan gangguan neuro-vaskular berat dan infeksi akut yang saling bertumpuk pada indikator gejala seperti sakit kepala hebat, gangguan tekanan darah, hingga gangguan motorik.

D. Visualisasi Spasial Jaringan Penyakit

Hasil akhir dari pembagian komunitas otomatis ini direpresentasikan secara visual melalui gambar graf berdimensi dua menggunakan metode *force-directed layout (spring layout)*, seperti yang ditunjukkan pada Gambar 3. berikut.



Gambar 3. Visualisasi jaringan proyeksi penyakit berbobot. Simpul-simpul diwarnai berdasarkan nomor kelompok komunitas yang dideteksi oleh Algoritma Louvain.

Pada Gambar 3. dapat dilihat dengan jelas bahwa simpul-simpul penyakit yang memiliki warna yang sama akan otomatis berkumpul dan mengelompok pada area spasial yang berdekatan. Jarak yang dekat antar-simpul di dalam satu warna ini disebabkan oleh tingginya nilai bobot garis (banyaknya irisan gejala) yang menarik simpul-simpul tersebut agar saling mendekat secara visual. Sebaliknya, simpul penyakit yang tidak memiliki kesamaan gejala akan saling tolak-menolak dan berada di posisi yang berjauhan. Visualisasi ini membuktikan bahwa pendekatan teori graf mampu memetakan kedekatan karakteristik penyakit secara intuitif dan mudah dipahami.

V. PENUTUP

A. Kesimpulan

Berdasarkan hasil eksperimen dan analisis jaringan yang telah dilakukan, dapat disimpulkan bahwa pendekatan teori graf melalui metode proyeksi satu mode (*one-mode projection*) dan Algoritma Louvain terbukti sangat efektif dalam memetakan serta mengelompokkan karakteristik penyakit berdasarkan tumpang tindih (*overlap*) gejala klinisnya. Transformasi dari graf bipartit awal menghasilkan Jaringan Penyakit berbobot yang terdiri dari 41 simpul penyakit, di mana nilai bobot pada setiap sisi berhasil merepresentasikan kuantitas eksak dari gejala-gejala yang saling beririsan antar-penyakit.

Melalui analisis metrik sentralitas kompleks, penelitian ini berhasil mengidentifikasi peran strategis dari setiap simpul penyakit di dalam jaringan topologi. Penyakit Dengue dan Typhoid ditemukan sebagai simpul paling dominan dengan nilai *Degree Centrality* dan *Closeness Centrality* tertinggi, yang menandakan bahwa kedua penyakit infeksi sistemik ini didominasi oleh gejala-gejala yang sangat umum sehingga berada tepat di pusat makro dari struktur graf. Di sisi lain, penyakit seperti GERD, Drug Reaction, dan Impetigo berhasil diidentifikasi sebagai simpul jembatan (*clinical bridge*) karena memiliki nilai *Betweenness Centrality* yang tinggi meskipun nilai derajatnya rendah. Hal ini membuktikan bahwa penyakit-penyakit tersebut memiliki kombinasi gejala spesifik yang bertindak sebagai penghubung kritis antar-rumpun spesialisasi medis yang berbeda, sehingga menjadi titik yang paling rawan memicu kesalahan diagnosis jika tidak diidentifikasi dengan hati-hati.

Selain itu, penerapan Algoritma Louvain berhasil melakukan deteksi komunitas secara otomatis dan membagi 41 penyakit ke dalam 5 kelompok komunitas yang terpisah berdasarkan optimasi nilai modularitas global. Meskipun algoritma ini bekerja murni menggunakan perhitungan matematis struktur graf tanpa dibekali basis pengetahuan medis, partisi yang dihasilkan terbukti memiliki akurasi dan keselarasan yang sangat tinggi dengan rumpun spesialisasi klinis di dunia nyata, seperti pemisahan yang rapi pada kluster Hepato-Gastroenterologi dan Infeksi Sistemik. Secara keseluruhan, integrasi pemodelan graf dan deteksi komunitas ini berhasil memberikan sudut pandang baru yang terukur dalam analisis data medis terstruktur, serta memiliki potensi besar untuk

dikembangkan lebih lanjut menjadi sistem pendukung keputusan klinis (*clinical decision support system*) guna membantu meminimalkan risiko ambiguitas dan galat diagnosis di dunia medis.

B. Saran

Meskipun sistem yang dibangun telah berhasil memetakan jaringan penyakit dan mendeteksi komunitas klinis secara otomatis dengan akurasi topologi yang tinggi, masih terdapat beberapa pengembangan yang dapat dilakukan pada penelitian selanjutnya. Berikut adalah beberapa saran yang dapat diterapkan untuk meningkatkan kualitas dan fungsionalitas penelitian ini:

- *Pengembangan Bobot Sisi Menggunakan Pendekatan Probabilistik*: Pada penelitian ini, bobot sisi graf proyeksi hanya dihitung berdasarkan kuantitas murni (*absolute count*) dari jumlah gejala yang beririsan. Untuk penelitian selanjutnya, disarankan menggunakan metode pembobotan yang lebih kompleks seperti *Jaccard Similarity Index* atau pendekatan probabilistik. Hal ini penting agar gejala yang sangat umum (seperti demam) tidak mendistorsi nilai kedekatan antar-penyakit secara berlebihan dibandingkan dengan gejala yang bersifat langka dan spesifik.
- *Integrasi Atribut Demografis dan Riwayat Medis Pasien*: Dataset biner yang digunakan saat ini hanya berfokus pada relasi gejala-penyakit secara statis. Penelitian ke depan dapat menyematkan atribut tambahan pada simpul atau sisi graf, seperti faktor usia, jenis kelamin, faktor risiko genetik, atau riwayat perjalanan pasien, guna menghasilkan visualisasi kluster yang jauh lebih personal dan akurat secara klinis.
- *Implementasi ke Dalam Antarmuka Pengguna (User Interface)*: Hasil pemetaan dan visualisasi graf dari penelitian ini disarankan untuk diintegrasikan ke dalam sebuah aplikasi web atau *dashboard* interaktif (misalnya menggunakan pustaka *Streamlit* atau *Dash*). Antarmuka tersebut akan sangat membantu tenaga medis dalam menginput gejala pasien secara dinamis dan langsung melihat visualisasi penyakit terdekat yang rawan terjadi salah diagnosis secara *real-time*.

LAMPIRAN

Implementasi program dan dataset kesehatan dapat diakses melalui repositori GitHub berikut:

<https://github.com/zidaneuland/Relasi-Gejala-dan-Penyakit>

UCAPAN TERIMA KASIH

Puji Syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas Rahmat dan karunia-Nya makalah ini dapat diselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Prof. Dr. Ir. Rinaldi Munir, M.T selaku dosen mata kuliah Matematika Diskrit yang telah memberikan ilmu, arahan, dan bimbingan selama proses penyusunan makalah ini. Penulis juga mengucapkan terima kasih kepada keluarga dan teman-teman

yang telah memberikan dukungan, bantuan, serta motivasi selama proses pengerjaan penelitian ini.

Penulis menyadari bahwa makalah ini masih memiliki berbagai kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun agar penelitian selanjutnya dapat dilakukan dengan lebih baik.

Semoga makalah ini dapat memberikan manfaat dan menambah wawasan bagi para pembaca.

DAFTAR PUSTAKA

- [1] P. Chawla, "An Interdisciplinary Review on Application of Graph Theory," *Journal of Computational Analysis and Applications*, vol. 25, no. 1, pp. 1–13, 2018.
- [2] B. Buhaerah, Z. Busrah, dan H. Sanjaya, *Teori Graf dan Aplikasinya*. 2022.
- [3] N. Nisrina, *Analisis Centrality dan Modularity pada Graf untuk Identifikasi Relasi Penelitian Indeks Topologi*. 2023.
- [4] Mutiara, S. Wirawan, T. Yusnitasari, "Expanding Louvain Algorithm for Clustering Relationship Formation," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, 2023.
- [5] Kaggle, "Disease Prediction Using Machine Learning," Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/kaushil268/disease-prediction-using-machine-learning>. [Accessed: 13-May-2026].

- [6] R. Munir, "Graf (Bagian 1)," Sekolah Teknik Elektro dan Informatika ITB, 2026. [Online]. Available: <https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2025-2026/20-Graf-Bagian1-2026.pdf>. [Accessed: 18-May-2026].

PERNYATAAN

Dengan ini, saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 18 Mei 2026



Zidane Uland Fakhry

13525132