

Deteksi Komunitas pada Pengguna Aplikasi Kencan Berdasarkan Kebiasaan dan Perilaku Menggunakan Metode Louvain

Tarisma Dwi Putri Alifianti
Program Studi Sains Data
Telkom University Surabaya
Surabaya, Indonesia

tarismaalifianti@student.telkomuniversity.ac.id

Aisya Mufidah Najwa
Program Studi Sains Data
Telkom University Surabaya
Surabaya, Indonesia

aisyanaajwa@student.telkomuniversity.ac.id

Ikfan Putra Maesru Dwi Pradana
Program Studi Sains Data
Telkom University Surabaya
Surabaya, Indonesia

ikfanputramaesrudwip@student.telkomuniversity.ac.id

Abstrak— Pertumbuhan digitalisasi yang pesat telah mengubah berbagai aspek kehidupan, khususnya dalam interaksi sosial, dengan memfasilitasi komunikasi daring yang melampaui batas geografis. Penelitian ini menyoroti aplikasi kencan sebagai fenomena sosial yang signifikan, di mana individu dari berbagai latar belakang dapat terhubung berdasarkan minat dan preferensi mereka. Dengan menggunakan pendekatan analisis jaringan sosial, penelitian ini menerapkan algoritma deteksi komunitas Louvain untuk mengidentifikasi kelompok pengguna dalam platform OkCupid. Jaringan dibangun berdasarkan kemiripan antar pengguna menggunakan cosine similarity, yang mencerminkan preferensi dan pola interaksi mereka. Hasil analisis mengungkap sepuluh komunitas dengan karakteristik demografis, gaya hidup, dan preferensi yang beragam. Skor modularitas yang tinggi (0,686) menunjukkan keterhubungan internal yang kuat dalam komunitas. Temuan ini memberikan wawasan mendalam tentang struktur sosial dalam aplikasi kencan serta membuka peluang untuk pengembangan fitur personalisasi dan sistem rekomendasi yang lebih efektif. Penelitian ini berkontribusi pada pemahaman yang lebih luas mengenai dinamika jejaring sosial dalam konteks digital dan pentingnya analisis sosial berbasis data.

Kata Kunci—aplikasi kencan, Louvain, cosine similarity, deteksi komunitas, rekomendasi personalisasi, modularitas

I. PENDAHULUAN

Pertumbuhan digitalisasi semakin berkembang pesat dan mencakup berbagai bidang lembaga maupun kehidupan. Digitalisasi menjadi penghubung pada setiap aspek kehidupan terutama pada hubungan sosial [1]. Teknologi tersebut telah membawa perubahan besar secara global, menghapus batasan geografis dan menciptakan ruang baru [2]. Interaksi yang awalnya terbatas secara fisik kini bertransformasi menjadi komunikasi daring yang fleksibel dan berskala luas [3].

Digitalisasi yang berskala luas membentuk fenomena sosial yang beragam salah satunya seperti munculnya aplikasi untuk mencari pasangan atau sering disebut dating app yang memfasilitasi pertemuan dan interaksi antara individu dari berbagai latar belakang. Dating app memungkinkan pengguna untuk saling berkenalan, berkomunikasi, dan menjalin hubungan dengan orang baru berdasarkan preferensi dan minat masing-masing [4]. Dalam praktiknya, aplikasi ini menghasilkan jaringan sosial yang kompleks dan dinamis, menciptakan struktur hubungan sosial yang tidak selalu tampak secara eksplisit [5].

Pendekatan utama dalam analisis jejaring sosial memiliki berbagai metode salah satunya adalah deteksi komunitas (community detection), yaitu proses untuk menemukan kelompok-kelompok dalam jaringan yang memiliki keterkaitan erat antar anggotanya. Komunitas dalam konteks ini bisa merepresentasikan kelompok pengguna dengan preferensi, lokasi, atau pola interaksi yang serupa. Deteksi komunitas memberikan wawasan penting dalam memahami struktur sosial dan membantu dalam pengembangan fitur-fitur yang lebih personalisasi pada aplikasi.

Metode Louvain merupakan salah satu algoritma deteksi komunitas yang paling efisien dan banyak digunakan, terutama dalam jaringan berukuran besar. Metode ini bekerja dengan mengoptimalkan nilai modularitas, yaitu ukuran seberapa baik jaringan dibagi menjadi komunitas. Dengan menggunakan pendekatan hierarkis, Louvain dapat membagi jaringan sosial menjadi kluster-kluster dengan konektivitas internal tinggi dan koneksi eksternal yang rendah. Efisiensi komputasinya menjadikannya ideal untuk menganalisis data besar yang dihasilkan oleh aplikasi kencan [6].

Penelitian ini bertujuan untuk menerapkan metode Louvain dalam mendeteksi komunitas pengguna pada data interaksi aplikasi kencan. Penelitian ini menggabungkan cosine similarity untuk membentuk jaringan berdasarkan kemiripan antar pengguna dan algoritma Louvain untuk mendeteksi struktur komunitasnya. Dengan menganalisis struktur jaringan dan mengidentifikasi komunitas yang terbentuk, penelitian ini diharapkan dapat memberikan pemahaman yang lebih dalam mengenai pola sosial pengguna, serta membuka peluang pengembangan sistem rekomendasi dan fitur interaktif yang lebih adaptif terhadap perilaku dan preferensi pengguna.

II. KAJIAN PUSTAKA

A. Penelitian Terdahulu

Penelitian yang dilakukan oleh van der Heide et al. (2018) memberikan kontribusi dalam pemahaman struktur komunitas dalam platform kencan daring melalui pendekatan jejaring sosial. Dalam studi tersebut, penulis menyoroti bahwa interaksi pengguna dalam aplikasi kencan tidak hanya didasarkan pada preferensi yang sempit seperti usia atau

lokasi, tetapi juga dibentuk oleh dinamika jaringan yang lebih luas, termasuk pola keterhubungan dan eksklusivitas interaksi antar pengguna.

Penelitian ini memanfaatkan analisis jaringan sosial untuk mengidentifikasi batas-batas komunitas yang tidak selalu tampak secara kasat mata. Dengan menganalisis data dari sebuah platform kencan daring, peneliti menemukan bahwa komunitas terbentuk bukan hanya karena kesamaan demografis, tetapi juga karena adanya segmentasi sosial yang bersifat struktural dalam jaringan komunikasi. Temuan ini menunjukkan bahwa walaupun pengguna dapat berinteraksi secara bebas, sistem algoritma dan dinamika jaringan menciptakan pembentukan komunitas yang cenderung tertutup dan tersegmentasi. Studi ini juga menjadi landasan dalam menyoroti perlunya teknik deteksi komunitas yang efisien dan akurat guna mengungkap pola-pola komunikasi sosial yang tersembunyi dalam jaringan daring berskala besar [6].

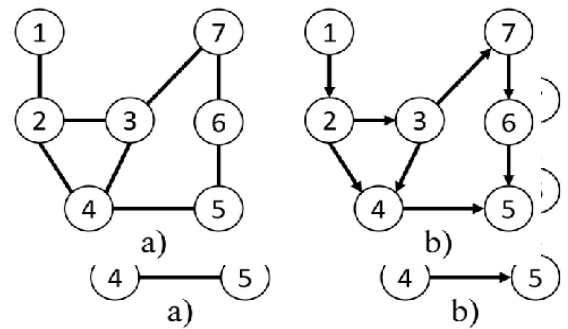
Penelitian selanjutnya oleh Akhmad Irsyad dan Nur Aini Rakhmawati (2020) menunjukkan keberhasilan deteksi komunitas pada media sosial Twitter dengan menggabungkan Cosine Similarity dan algoritma Louvain. Mereka membangun graf berdasarkan kemiripan isi tweet antar pengguna, lalu mengelompokkan akun-akun ke dalam komunitas yang memiliki pola komunikasi serupa. Hasilnya menunjukkan bahwa Louvain mampu mendeteksi komunitas dengan modularitas tinggi, meskipun pada jaringan besar dan kompleks [7]. Namun, studi tersebut berfokus pada media sosial umum seperti Twitter, tanpa mengeksplorasi konteks spesifik aplikasi kencan yang memiliki dinamika sosial lebih intim dan berbasis preferensi personal.

Selain itu, meskipun pendekatan kombinasi Louvain dan cosine similarity telah dibuktikan mampu menghasilkan partisi komunitas dengan modularitas tinggi, masih sedikit penelitian yang mengaitkan hasil deteksi komunitas ini secara langsung dengan sistem rekomendasi interaksi antar pengguna di aplikasi kencan.

Penelitian ini mengisi celah tersebut dengan menerapkan algoritma Louvain berbasis cosine similarity pada data pengguna aplikasi kencan OkCupid, yang memiliki karakteristik unik dalam interaksi sosial dan menghubungkan hasil deteksi komunitas dengan potensi pengembangan sistem rekomendasi yang lebih personal dan adaptif berdasarkan kesamaan sosial dan gaya hidup pengguna.

B. Social Network Analysis (SNA)

Social Network Analysis atau yang biasa disingkat SNA merupakan pendekatan analitis yang digunakan untuk memahami struktur dan pola hubungan antar entitas dalam sebuah jaringan sosial. Dalam SNA, individu direpresentasikan sebagai simpul (*nodes*) dan interaksi antar mereka sebagai sisi (*edges*) [8]. Metode ini dapat membantu mengidentifikasi struktur sosial, keterhubungan dan juga potensi pengaruh antar individu. Dua *node* dianggap terkoneksi jika ada sebuah garis yang menghubungkan keduanya. Representasi sebuah relasi dalam graf, yang disimbolkan oleh *edges* atau garis, dapat dilakukan dengan dua metode, yaitu [9]:



1. Undirected Graph (Graf Tak Terarah)

Graf tak terarah atau *undirected graph* merupakan graf sederhana yang menghubungkan setiap individu dengan menggunakan tepi atau garis (*edges*) atau garis tanpa diberi tanda panah.

2. Directed Graph (Graf Terarah)

Graf terarah atau *directed graph* merupakan graf yang digunakan untuk memperjelas hubungan karena setiap koneksi diwakili oleh *edges* atau garis yang ditampilkan dengan menggunakan panah.

C. Algoritma Louvain

Metode Louvain adalah algoritma deteksi komunitas yang mengoptimalkan modularitas jaringan secara efisien pada graf berukuran besar [10]. Modularitas Q diukur sebagai perbedaan antara bobot sisi aktual dan bobot sisi ekspektasi dalam model acak, yaitu

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

di mana A_{ij} adalah bobot sisi antara simpul i dan j , k_i dan k_j adalah derajat (bobot total sisi) simpul i dan j , m adalah setengah jumlah total bobot sisi, dan $\delta(c_i, c_j)$ sama dengan 1 jika i dan j berada dalam komunitas yang sama, serta 0 jika tidak [11].

Algoritma berjalan dalam dua tahap berulang:

1. Optimalisasi Modularitas

Setiap simpul dipertimbangkan satu per satu dan dipindahkan ke komunitas tetangganya yang menghasilkan peningkatan ΔQ terbesar, dihitung sebagai

$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right]$$

di mana \sum_{in} adalah jumlah bobot sisi dalam komunitas target, \sum_{tot} total bobot sisi yang terhubung ke komunitas itu, dan $k_{i,in}$ bobot sisi simpul i ke komunitas tersebut.

2. Agregasi Komunitas

Setelah tidak ada perpindahan yang meningkatkan Q , tiap komunitas dianggap sebagai satu simpul baru, dan bobot antar-komunitas dijumlahkan untuk membentuk graf tereduksi. Kemudian tahap pertama diulang pada graf yang telah disederhanakan.

Proses kedua tahap ini terus berputar hingga modularitas tidak dapat lagi ditingkatkan. Hasil akhir adalah partisi komunitas yang memaksimumkan modularitas, sehingga merepresentasikan struktur klaster dalam jaringan secara optimal.

D. NetworkX

NetworkX merupakan salah satu *library* Python *open-source* yang dirancang untuk membangun, memanipulasi, dan menganalisis struktur jaringan kompleks [12]. Library ini menyediakan struktur data yang fleksibel untuk merepresentasikan berbagai jenis graf, baik berarah maupun tak berarah, termasuk multigraf. Setiap simpul dalam jaringan dapat berupa berbagai tipe data, seperti teks atau gambar, sedangkan setiap sisi dapat menyimpan atribut tambahan seperti bobot atau data deret waktu. Fleksibilitas ini menjadikan NetworkX sebagai alat yang serbaguna dalam analisis jaringan.

Selain mendukung representasi graf, NetworkX juga dilengkapi dengan berbagai algoritma standar untuk analisis jaringan, seperti penghitungan sentralitas, pencarian jalur terpendek, dan deteksi komunitas. Kemampuan visualisasi jaringan yang dimilikinya turut membantu dalam menampilkan struktur jaringan secara informatif dan intuitif. Dengan dukungan bahasa Python, NetworkX memungkinkan prototipe cepat, mudah digunakan lintas platform, dan banyak dimanfaatkan di berbagai bidang seperti ilmu sosial, biologi, dan ilmu komputer [13].

E. Cosine Similarity

Metode *cosine similarity* merupakan suatu teknik untuk menilai derajat keserupaan antara dua entitas. Perhitungan metode ini pada umumnya berlandaskan pada ukuran kesamaan dalam ruang vektor. Dengan menggunakan kumpulan kata kunci dari suatu dokumen sebagai komponen vektor, metode kesamaan kosinus menghitung sejauh mana dua objek misalnya $D1$ dan $D2$ yang telah direpresentasikan dalam bentuk vektor, memiliki tingkat kemiripan [14].

Consine Similarity digunakan untuk melakukan perhitungan kesamaan dari dokumen. Rumus yang digunakan oleh consine similarity adalah [15]

$$\cos \alpha = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Keterangan :

A = Vektor A, yang akan dibandingkan kemiripannya

B = Vektor B, yang akan dibandingkan kemiripannya

$A \cdot B$ = *dot product* antara vektor A dan vektor B

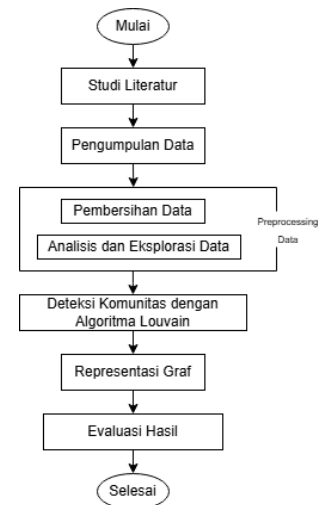
$|A|$ = Panjang vektor A

$|B|$ = Panjang vektor B

$|A||B|$ = *cross product* antara $|A|$ dan $|B|$

III. METODOLOGI PENELITIAN

Pada bab ini, akan dijelaskan secara rinci mengenai metodologi yang diterapkan dalam penelitian. Diagram alur yang menggambarkan tahapan pengerjaan penelitian dapat dilihat pada Gambar 2.



Gambar 1. Diagram alur penelitian

A. Studi Literatur

Pada tahapan ini akan dilakukan pengumpulan informasi dari berbagai sumber yang dapat menunjang pengerjaan penelitian ini yang mencakup penelitian-penelitian sebelumnya mengenai deteksi komunitas, analisis jaringan sosial, sistem rekomendasi, perilaku pengguna aplikasi kencan, serta penerapan Metode Louvain. Studi literatur bertujuan untuk membangun landasan teoretis yang kuat, mengidentifikasi celah penelitian (*research gap*), dan memperkaya pemahaman terkait metodologi yang akan digunakan.

B. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini bersumber dari profil publik 59.946 pengguna aktif OkCupid yang memenuhi kriteria berikut: (1) berdomisili dalam radius 25 mil dari San Francisco, (2) memiliki profil aktif selama periode tertentu pada tahun 2010-an, (3) melakukan aktivitas dalam platform dalam satu tahun terakhir, dan (4) memiliki minimal satu foto profil (Kim & Escobedo-Land, 2015). Data dikumpulkan melalui *web scraping* menggunakan skrip Python yang mengekstrak informasi dari profil publik pengguna, dengan mengabaikan data privat seperti riwayat pesan (*messaging*) yang tidak dapat diakses secara etis.

Dataset ini terdiri dari dua kategori variabel utama:

1. Data demografis terstruktur

- Informasi dasar pengguna seperti sex (jenis kelamin), sexual orientation (orientasi seksual), age (usia dengan penambahan random noise untuk anonimisasi), dan ethnicity (etnisitas).
- Karakteristik gaya hidup antara lain diet (pola diet), drinking habits (kebiasaan konsumsi alkohol), dan smoking habits (kebiasaan merokok).

2. Data tekstual semi-terstruktur

Deskripsi profil (essays) yang mencakup preferensi pribadi, minat, dan nilai hidup pengguna [16].

C. Preprocessing Data

Pada tahapan ini, *preprocessing data* digunakan untuk membersihkan dan juga untuk mengidentifikasi data sebelum data tersebut akan diolah, ada beberapa tahapan yang digunakan yakni

1. Pembersihan data

Pembersihan data merupakan tahapan awal yang bertujuan untuk meningkatkan kualitas dataset agar siap untuk analisis lebih lanjut. Pada tahap ini, dilakukan identifikasi dan penanganan nilai hilang (missing values) melalui teknik penghapusan data. Selain itu, duplikasi catatan dihilangkan untuk mencegah bias pada hasil analisis, sementara anomali ekstrem diperiksa secara mendetail untuk memastikan bahwa tidak terdapat data yang keliru akibat kesalahan input atau gangguan teknis. Transformasi format tanggal dan waktu juga dilakukan agar konsistensi struktur data terjaga. Semua langkah tersebut, termasuk konversi tipe data dan penataan kembali atribut-atribut yang relevan, dirancang agar dataset akhir merepresentasikan kondisi aktual sistem aplikasi kencana secara akurat dan minim gangguan.

2. Analisis dan eksplorasi data

Setelah dataset melewati proses pembersihan, tahap analisis dan eksplorasi data berfokus pada pemahaman karakteristik dan pola dasar dalam data pengguna aplikasi kencana. Peneliti melakukan eksplorasi deskriptif dengan menghitung statistik ringkasan seperti mean, median, modus, dan sebaran varians untuk setiap variabel penting. Visualisasi grafis, seperti histogram, diagram sebar (scatter plot), dan heatmap korelasi, digunakan untuk mengungkap distribusi dan hubungan antar-variabel misalnya hubungan antara frekuensi interaksi harian dan usia pengguna.

D. Deteksi Komunitas dengan Algoritma Louvain

Dengan fitur diekstraksi, algoritma Louvain diterapkan untuk mendeteksi komunitas pada data yang belum dalam bentuk graf penuh. Proses Louvain dimulai dengan mengasumsikan setiap entitas pengguna sebagai komunitas terpisah, lalu secara iteratif menggabungkan simpul-simpul yang meningkatkan nilai modularitas secara signifikan. Meskipun data sejatinya masih berupa representasi non-graf, tahapan ini menitikberatkan pada proses perhitungan aturan pemisahan berdasarkan matriks kemiripan yang dihasilkan dari feature engineering sebelumnya. Keluaran utama adalah pembagian awal simpul ke dalam komunitas dan nilai modularitas sementara yang menunjukkan kekuatan pembentukan klaster.

E. Representasi Graf

Setelah komunitas awal diidentifikasi secara konseptual melalui Louvain, hasil tersebut dimodelkan pada setiap pengguna sebagai simpul (node) dan relasi kemiripan sebagai sisi (edge) dalam sebuah graf terstruktur. Bobot sisi diukur berdasarkan skor kesamaan yang dihitung dari vektor-vektor fitur perilaku. Hasilnya adalah *adjacency list* atau matriks bobot lengkap yang menggambarkan keterkaitan antarpengguna secara kuantitatif. Representasi ini memungkinkan visualisasi jaringan sosial secara eksplisit dan menjadi fondasi untuk evaluasi hasil deteksi komunitas serta tahap rekomendasi pertemanan berikutnya.

F. Evaluasi Hasil

Tahap evaluasi memeriksa sejauh mana komunitas yang terbentuk pada fase sebelumnya memenuhi kriteria kualitas partisi. Peneliti menghitung skor modularitas akhir dari graf untuk menilai koherensi internal komunitas dan pemisahan eksternal antarkelompok. Jika tersedia data ground truth, metrik seperti *Normalized Mutual Information* (NMI) atau *Adjusted Rand Index* (ARI) dapat digunakan untuk membandingkan hasil Louvain dengan pembagian sebenarnya. Visualisasi graf yang diberi pewarnaan berdasarkan label komunitas juga diterapkan untuk meninjau konsistensi klaster secara kualitatif. Hasil evaluasi ini menjadi acuan validitas sebelum membangun rekomendasi pertemanan.

IV. HASIL DAN PEMBAHASAN

A. Pre-Processing Data

Dataset yang digunakan dalam penelitian ini berasal dari platform kencan daring dan terdiri atas 59.946 entri dengan 31 atribut. Setiap baris merepresentasikan satu pengguna, dengan informasi demografis, kebiasaan, serta opini yang diekspresikan melalui esai singkat. Namun, tidak semua kolom digunakan dalam penelitian ini. Pemilihan kolom dilakukan berdasarkan relevansi terhadap karakteristik sosial, tingkat kelengkapan data, serta hasil korelasi awal terhadap pembentukan komunitas. Setelah melalui proses seleksi, dipilih 12 kolom yang dianggap paling informatif, yaitu: education, job, religion, orientation, drinks, location, status, diet, offspring, smokes, drugs, dan age.

Data kemudian dibersihkan dari nilai kosong (missing values) pada kolom-kolom terpilih. Seluruh baris yang mengandung nilai kosong pada salah satu kolom tersebut dihapus, sehingga jumlah data menyusut dari 59.946 menjadi 9.922 baris. Pembersihan ini bertujuan untuk menjaga kualitas hasil analisis dan menghindari bias akibat imputasi yang tidak akurat. Selain itu, dilakukan penyaringan pada kolom usia (age) karena terdapat outlier dengan nilai hingga 110 tahun. Berdasarkan analisis interkuartil (IQR), batas atas ditentukan pada usia 53 tahun. Oleh karena itu, hanya pengguna berusia antara 18 hingga 53 tahun yang disertakan dalam analisis.

Untuk keperluan kategorisasi, usia kemudian dikelompokkan ke dalam enam rentang interval 6 tahun, yaitu 18–23, 24–29, 30–35, 36–41, 42–47, dan 48–53. Setiap pengguna diberikan label kelompok usia (age_group) yang sesuai. Transformasi ini memudahkan dalam melakukan segmentasi serta analisis komunitas berbasis demografi. Selain itu, informasi location pengguna yang semula dalam format "city, state" dipisahkan, dan hanya bagian negara bagian (state) yang diambil untuk penyederhanaan. Hasil

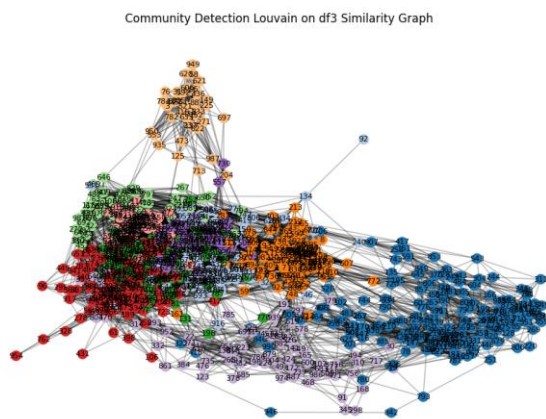
akhir dari pra-pemrosesan ini adalah data bersih sebanyak 9.128 baris dengan 12 kolom terpilih, siap digunakan untuk tahap analisis lebih lanjut.

Untuk mengurangi beban komputasi dan memudahkan visualisasi data dalam tahap eksplorasi, dilakukan pengambilan sampel sebanyak 1.000 data secara acak dari total 9.128 baris hasil pembersihan. Pengambilan sampel ini menggunakan metode random sampling tanpa pengelompokan khusus, dengan asumsi bahwa sampel tetap merepresentasikan sebaran populasi secara umum.

B. Hasil

Fokus analisis ditujukan pada individu-individu yang berasal dari California karena setelah ditelusuri, terdapat 995 individu yang berasal dari negara bagian tersebut. Sementara itu, sisanya, hanya lima individu, berasal dari lokasi yang beragam seperti Switzerland, New York, Arizona, Netherlands, dan Ohio. Keberadaan lima individu ini justru membentuk dua komunitas outlier yang muncul karena lokasi mereka sangat berbeda dari mayoritas data. Komunitas outlier ini tidak merepresentasikan struktur sosial yang signifikan karena anggotanya sangat sedikit dan tersebar. Oleh karena itu, analisis dilanjutkan dengan memfokuskan data hanya pada individu yang berasal dari California. Setelah dilakukan penyaringan data, jumlah individu yang dianalisis menjadi 995, dan jumlah komunitas yang awalnya 12 menyusut menjadi 10 komunitas utama yang lebih relevan untuk dianalisis lebih lanjut.

Penelitian ini berhasil mengidentifikasi 10 komunitas dari jaringan sosial yang terbentuk berdasarkan 12 atribut (termasuk state) dan hubungan kemiripan (cosine similarity) antarindividu dalam dataset. Untuk membangun jaringan tersebut, digunakan pendekatan cosine similarity antar pengguna, dengan hanya menghubungkan pasangan individu yang memiliki nilai kemiripan di atas 0,7. Proses deteksi komunitas dilakukan menggunakan algoritma Louvain, yang secara efektif memaksimalkan modularitas untuk menemukan kelompok dengan keterkaitan internal yang tinggi.



Gambar 1. Graf Komunitas Pengguna OkCupid

Distribusi jumlah anggota per komunitas dapat dilihat pada Tabel 1 berikut:

Komunitas	Jumlah Anggota
0	195
1	90
2	117
3	40

4	96
5	59
6	145
7	81
8	85
9	87

Struktur komunitas yang terbentuk menunjukkan adanya pola keterhubungan yang cukup kuat di dalam tiap kelompok, dengan nilai modularity sebesar 0,686. Nilai ini menunjukkan bahwa struktur jaringan memiliki modularitas yang tinggi, mengindikasikan bahwa node-node dalam jaringan lebih banyak terhubung di dalam komunitasnya masing-masing dibandingkan dengan komunitas lain.

C. Pembahasan

Berdasarkan hasil analisis karakteristik tiap komunitas, terlihat bahwa data yang muncul pada setiap kolom sangat bervariasi dan tidak selalu terdistribusi secara dominan pada satu nilai saja. Banyak komunitas menunjukkan keberagaman preferensi dan gaya hidup yang cukup tinggi, mencerminkan kompleksitas individu dalam komunitas tersebut. Oleh karena itu, untuk menggambarkan karakteristik komunitas secara lebih representatif dan mudah dipahami, interpretasi yang disajikan sebelumnya didasarkan pada tiga nilai terbanyak (top 3) dari masing-masing kolom kategori yang paling menonjol di komunitas tersebut.

Pendekatan ini dipilih karena hasil distribusi nilai di tiap kolom umumnya tidak terpusat hanya pada satu kategori, melainkan menyebar ke beberapa pilihan. Dengan demikian, penggunaan top 3 memungkinkan kita tetap menangkap ciri khas komunitas tanpa mengabaikan keragaman yang ada.

Berikut ini adalah pembahasan karakteristik dari masing-masing komunitas berdasarkan hasil tersebut.

1. Komunitas 0 (195 Anggota)

- education: graduated from college/university (75), working on college/university (32), graduated from masters program (22)
- job: other (35), student (23), sales / marketing / biz dev (22), artistic / musical / writer (22)
- religion: atheism (53), agnosticism (43), other (39)
- orientation: straight (191), gay (4)
- drinks: socially (126), often (36), rarely (11), very often (11)
- state: california (195)
- status: single (192), seeing someone (2), married (1)
- diet: mostly anything (97), strictly anything (33), anything (25)
- offspring: doesn't have kids (69), doesn't have kids, but might want them (44), doesn't want kids (29)
- smokes: no (72), when drinking (47), yes (37)
- drugs: sometimes (138), never (47), often (10)
- age_group: 24-29 (90), 18-23 (42), 30-35 (37)

Komunitas ini adalah yang paling besar dengan 195 anggota. Mayoritas berusia 18–29 tahun, dengan latar belakang pendidikan yang sudah lulus kuliah atau sedang kuliah. Pekerjaannya beragam, mayoritas mahasiswa, seni, atau marketing. Kebanyakan ateis atau agnostik. Hampir semua berorientasi heteroseksual dan mayoritas berstatus single. Kebiasaan minum tergolong tinggi, sementara penggunaan obat-obatan cukup tinggi. Kebanyakan mengikuti pola makan "anything" dalam berbagai variasi. Kebanyakan tidak merokok, namun cukup banyak juga yang merokok saat minum atau merokok aktif. Komunitas ini menggambarkan kelompok muda yang sangat terbuka, sekuler, dan berpola hidup sosial aktif. Mereka cenderung memiliki gaya hidup bebas, baik dalam pergaulan maupun kebiasaan rekreasi (alkohol dan narkoba). Ini menunjukkan komunitas urban yang liberal dan berjiwa muda, dengan banyak di antaranya dalam masa transisi pendidikan atau awal karier.

2. Komunitas 1 (90 Anggota)

- education: graduated from college/university (45), graduated from masters program (22), graduated from space camp (4)
- job: sales / marketing / biz dev (11), other (10), computer / hardware / software (10)
- religion: atheism (25), agnosticism (24), christianity (22)
- orientation: straight (89), gay (1)
- drinks: socially (53), often (14), rarely (12)
- state: california (90)
- status: single (87), seeing someone (3)
- diet: strictly anything (48), mostly vegetarian (18), strictly vegetarian (11)
- offspring: doesn't have kids (35), doesn't have kids, but might want them (23), doesn't have kids, but wants them (18)
- smokes: no (84), sometimes (5), trying to quit (1)
- drugs: never (87), often (3)
- age_group: 24-29 (34), 30-35 (26), 36-41 (18)

Didominasi oleh kelompok usia 24–35 tahun. Latar pendidikan cukup tinggi, dengan banyak yang sudah lulus kuliah dan S2. Kebanyakan bekerja di bidang pemasaran, komputer, atau bidang lainnya. Mayoritas ateis, agnostik, atau Kristen, dengan orientasi heteroseksual dan status single. Sebagian besar tidak merokok atau menggunakan narkoba, serta pola minum mereka cenderung sosial atau jarang. Pola makan didominasi oleh diet "strictly anything" dan vegetarian. Komunitas ini tampaknya terdiri dari profesional muda dengan gaya hidup sehat dan konservatif dalam konsumsi zat. Mereka cenderung rasional, terdidik, dan memiliki kecenderungan untuk menghindari risiko. Komunitas ini juga cukup seragam dari sisi status dan preferensi hidup, mencerminkan kelompok stabil di awal fase karier.

3. Komunitas 2 (117 Anggota)

- education: working on college/university (52), working on masters program (17), working on ph.d program (12), working on two-year college (12)

- job: student (65), science / tech / engineering (12), hospitality / travel (7)
- religion: atheism (27), other (23), christianity (22)
- orientation: straight (112), gay (5)
- drinks: socially (88), rarely (23), not at all (4)
- state: california (117)
- status: single (116), seeing someone (1)
- diet: mostly anything (67), anything (13), strictly anything (13)
- offspring: doesn't have kids (42), doesn't have kids, but might want them (25), doesn't have kids, but wants them (24)
- smokes: no (109), sometimes (8)
- drugs: never (115), often (2)
- age_group: 18-23 (50), 24-29 (43), 30-35 (20)

Mayoritas berusia 18–29 tahun yang sedang menempuh pendidikan, terutama di tingkat S1 dan S2. Sebagian besar adalah mahasiswa, dan secara agama mereka tersebar antara ateis, Kristen, dan lainnya. Status dominan adalah single dan orientasi heteroseksual. Hampir semuanya tidak merokok dan tidak menggunakan narkoba. Gaya hidup mereka cenderung sehat dengan pola minum sosial yang tidak berlebihan dan pola makan "anything". Komunitas ini adalah tipikal mahasiswa yang fokus pada studi dan memiliki gaya hidup terkontrol. Minimnya penggunaan zat dan tingginya status lajang menunjukkan kelompok yang masih dalam masa pengembangan diri. Mereka cukup netral secara religius dan cenderung hidup di lingkungan yang terbuka tapi terstruktur.

4. Komunitas 3 (40 Anggota)

- education: graduated from college/university (22), graduated from masters program (4), graduated from high school (3)
- job: artistic / musical / writer (5), education / academia (5), other (5), medicine / health (5)
- religion: atheism (15), agnosticism (14), other (5)
- orientation: straight (27), bisexual (7), gay (6)
- drinks: socially (21), rarely (8), often (7)
- state: california (40)
- status: available (34), married (3), seeing someone (3)
- diet: mostly anything (20), anything (9), mostly vegetarian (6)
- offspring: doesn't want kids (10), doesn't have kids, and doesn't want any (8), doesn't have kids (5)
- smokes: no (34), yes (3), sometimes (3)
- drugs: never (22), sometimes (16), often (2)
- age_group: 36-41 (13), 24-29 (11), 30-35 (5), 42-47 (5), 18-23 (5)

Mayoritas 36–41 tahun. Kebanyakan lulusan S1 dan S2, dengan pekerjaan yang tersebar di seni, akademik, dan kesehatan. Banyak yang ateis atau agnostik, dan orientasi

seksualnya beragam. Pola minum tergolong moderat, merokok dan penggunaan narkoba cukup bervariasi. Pola makan kebanyakan "mostly anything" dan vegetarian. Status dominan adalah "available" dan menikah. Komunitas ini mencerminkan kelompok usia dewasa dengan latar belakang kreatif dan intelektual, yang lebih terbuka terhadap keragaman orientasi dan gaya hidup. Mereka kemungkinan punya preferensi personal yang kuat dan lebih eksploratif.

5. Komunitas 4 (96 Anggota)

- education: graduated from college/university (58), graduated from masters program (21), graduated from law school (4)
- job: other (17), computer / hardware / software (11), artistic / musical / writer (10)
- religion: other (66), judaism (29), hinduism (1)
- orientation: straight (93), gay (3)
- drinks: socially (76), rarely (17), very often (2)
- state: california (96)
- status: single (95), seeing someone (1)
- diet: mostly anything (46), mostly vegetarian (14), anything (12)
- offspring: doesn't have kids (30), doesn't have kids, but might want them (28), doesn't have kids, but wants them (20)
- smokes: no (92), sometimes (3), trying to quit (1)
- drugs: never (96)
- age_group: 30-35 (31), 24-29 (31), 42-47 (14)

Mayoritas orang dewasa usia 24–35 tahun. Mereka sangat terdidik (banyak lulusan S2 dan hukum) dan kebanyakan bekerja di bidang teknologi atau seni. Mayoritas beragama non-arus utama seperti "other" dan Yahudi. Mereka cenderung heteroseksual, lajang, dan tidak merokok atau menggunakan narkoba. Pola makan mereka cenderung fleksibel dengan variasi sehat. Komunitas ini menunjukkan karakter profesional urban dengan nilai-nilai inklusif dan latar belakang pendidikan tinggi. Mereka rasional, fokus pada karier, dan menjaga gaya hidup sehat. Pola hidup mereka konservatif terhadap penggunaan zat, mencerminkan kelompok yang stabil dan berorientasi masa depan.

6. Komunitas 5 (59 Anggota)

- education: graduated from college/university (31), graduated from masters program (9), working on college/university (3), graduated from high school (3)
- job: other (11), computer / hardware / software (7), artistic / musical / writer (7)
- religion: christianity (16), agnosticism (12), other (10)
- orientation: straight (59)
- drinks: not at all (35), often (22), rarely (1), desperately (1)

- state: california (59)
- status: single (57), seeing someone (2)
- diet: mostly anything (45), anything (8), mostly vegetarian (3), mostly other (3)
- offspring: doesn't have kids (18), doesn't have kids, but might want them (14), doesn't have kids, but wants them (14)
- smokes: no (57), trying to quit (1), sometimes (1)
- drugs: never (59)
- age_group: 24-29 (17), 30-35 (15), 42-47 (11)

Mayoritas berusia 24–35 tahun. Mereka sebagian besar sudah lulus S1 atau S2 dan bekerja di berbagai bidang, terutama teknologi, seni, dan lainnya. Mereka mayoritas Kristen atau agnostik, tidak merokok, dan tidak menggunakan narkoba. Pola minum relatif konservatif, dengan banyak yang tidak minum sama sekali. Status dominan adalah lajang. Komunitas ini mencerminkan kelompok konservatif, berpendidikan, dan teratur. Mereka kemungkinan memiliki latar belakang budaya atau keluarga yang menekankan gaya hidup sehat dan disiplin.

7. Komunitas 6 (145 Anggota)

- education: graduated from college/university (56), graduated from masters program (40), graduated from two-year college (14)
- job: other (29), executive / management (20), medicine / health (19)
- religion: agnosticism (34), catholicism (33), christianity (29)
- orientation: straight (142), gay (3)
- drinks: socially (113), rarely (21), often (5)
- state: california (145)
- status: single (143), seeing someone (2)
- diet: mostly anything (88), anything (23), strictly anything (13)
- offspring: has a kid (30), has kids (27), doesn't have kids (16)
- smokes: no (135), sometimes (6), when drinking (3)
- drugs: never (142), sometimes (3)
- age_group: 36-41 (46), 48-53 (44), 42-47 (37)

Mayoritas individu dewasa berusia 36–53 tahun. Mereka sangat terdidik dan bekerja di manajemen, kesehatan, atau bidang umum lainnya. Agama mereka bervariasi antara agnostik, Katolik, dan Kristen. Mereka umumnya heteroseksual, banyak yang memiliki anak, dan tetap lajang. Pola hidup mereka relatif sehat; sebagian besar tidak merokok dan tidak menggunakan narkoba. Pola makan mereka fleksibel. Komunitas ini merupakan kelompok profesional mapan, dengan pengalaman hidup yang lebih banyak dan tanggung jawab keluarga. Mereka stabil secara emosional dan finansial, dengan gaya hidup konservatif dan fokus pada kehidupan sehari-hari yang tertib.

8. Komunitas 7 (81 Anggota)

- education: graduated from college/university (44), graduated from masters program (18), college/university (5)
- job: computer / hardware / software (25), banking / financial / real estate (14), artistic / musical / writer (12)
- religion: agnosticism (22), atheism (19), catholicism (18)
- orientation: straight (81)
- drinks: socially (71), rarely (10)
- state: california (81)
- status: single (81)
- diet: mostly anything (61), anything (19), mostly other (1)
- offspring: doesn't have kids (37), doesn't have kids, but wants them (19), doesn't have kids, but might want them (17)
- smokes: no (78), sometimes (3)
- drugs: never (81)
- age_group: 30-35 (37), 24-29 (30), 36-41 (10)

Mayoritas anggota muda dewasa (24–35 tahun), komunitas ini banyak bekerja di bidang teknologi dan finansial. Mereka mayoritas agnostik atau ateis, dan semuanya heteroseksual serta lajang. Gaya hidup sehat sangat dominan, dengan tidak adanya penggunaan narkoba dan kebiasaan merokok yang sangat rendah. Pola makan mereka didominasi oleh fleksibilitas. Komunitas ini adalah gambaran dari profesional muda perkotaan di bidang teknologi. Mereka sangat disiplin, fokus pada karier, dan menjaga gaya hidup sehat.

9. Komunitas 8 (85 Anggota)

- education: graduated from college/university (53), graduated from masters program (14), graduated from high school (4)
- job: science / tech / engineering (24), sales / marketing / biz dev (18), other (16)
- religion: catholicism (25), christianity (21), agnosticism (21)
- orientation: straight (85)
- drinks: socially (84), rarely (1)
- state: california (85)
- status: single (83), seeing someone (2)
- diet: mostly anything (59), anything (21), mostly other (5)
- offspring: doesn't have kids, but wants them (31), doesn't have kids (24), doesn't have kids, but might want them (19)
- smokes: no (83), sometimes (2)
- drugs: never (85)

- age_group: 24-29 (40), 30-35 (28), 36-41 (11)

Mayoritas berusia 24–35 tahun, banyak bekerja di sains dan pemasaran. Mereka mayoritas Katolik atau Kristen dan sangat sehat secara kebiasaan hidup—tidak merokok dan tidak menggunakan narkoba. Semua heteroseksual dan hampir semuanya lajang. Pola makan mereka fleksibel. Komunitas ini menunjukkan kelompok profesional sains yang stabil, religius, dan konservatif dalam gaya hidup. Mereka mengutamakan kesehatan dan memiliki visi hidup yang mungkin lebih terstruktur.

10. Komunitas 9 (87 Anggota)

- education: graduated from college/university (25), working on college/university (20), graduated from masters program (12)
- job: student (21), other (15), artistic / musical / writer (14)
- religion: atheism (22), agnosticism (19), other (17)
- orientation: bisexual (54), gay (33)
- drinks: socially (43), rarely (17), often (13)
- state: california (87)
- status: single (83), seeing someone (4)
- diet: mostly anything (41), anything (15), mostly vegetarian (7)
- offspring: doesn't have kids, but might want them (24), doesn't have kids, and doesn't want any (17), doesn't want kids (17)
- smokes: no (59), sometimes (12), yes (7)
- drugs: never (54), sometimes (32), often (1)
- age_group: 24-29 (25), 18-23 (24), 30-35 (18)

Komunitas ini sangat unik karena semua memiliki orientasi non-heteroseksual (biseksual dan gay). Mereka berusia 18–35 tahun, berpendidikan tinggi, dan cenderung mahasiswa atau bekerja di bidang seni. Penggunaan narkoba “kadang-kadang” cukup tinggi, dan pola minum mereka cukup bervariasi. Status mayoritas adalah lajang. Komunitas ini merupakan representasi dari kelompok LGBTQ+ muda yang ekspresif dan artistik. Mereka lebih terbuka terhadap eksplorasi identitas dan kebiasaan sosial, termasuk konsumsi zat. Komunitas ini mencerminkan keragaman dalam ekspresi diri.

D. Perbedaan Antar Komunitas

Analisis terhadap sepuluh komunitas yang teridentifikasi menunjukkan adanya variasi signifikan dalam karakteristik demografis, gaya hidup, dan preferensi personal. Perbedaan-perbedaan ini mencerminkan segmentasi natural dalam populasi berdasarkan tahap kehidupan, nilai-nilai, dan orientasi hidup. Setiap komunitas yang terbentuk merepresentasikan kelompok pengguna dengan kesamaan tertentu, sehingga dapat digunakan sebagai dasar dalam sistem rekomendasi, baik untuk pertemanan maupun pencarian pasangan. Misalnya, pengguna dalam komunitas yang sama cenderung memiliki kecocokan dalam minat dan gaya hidup, sehingga rekomendasi interaksi atau pertemanan dapat difokuskan pada individu dari komunitas tersebut.

Pendekatan ini memungkinkan aplikasi kencan untuk menyarankan koneksi yang lebih relevan dan potensial, serta meningkatkan pengalaman pengguna melalui interaksi yang lebih bermakna dan personal.

a. Perbedaan Berdasarkan Demografi Usia

Komunitas-komunitas yang teridentifikasi menunjukkan pola segmentasi yang jelas berdasarkan kelompok usia. Komunitas 0, 2, dan 9 didominasi oleh generasi muda berusia 18-29 tahun, yang umumnya merupakan mahasiswa atau profesional pemula. Kelompok ini menunjukkan karakteristik eksplorasi identitas dan gaya hidup yang lebih eksperimental. Di sisi lain, Komunitas 6 merepresentasikan kelompok dewasa mapan berusia 36-53 tahun yang telah memiliki tanggung jawab keluarga dan stabilitas karier. Komunitas 1, 4, 7, dan 8 berada di kategori menengah (24-35 tahun), mencerminkan fase transisi dari eksplorasi menuju stabilitas.

b. Variasi Orientasi Seksual dan Identitas

Perbedaan paling mencolok terlihat pada Komunitas 9 yang secara eksklusif terdiri dari individu dengan orientasi non-heteroseksual (biseksual dan gay). Komunitas ini menunjukkan karakteristik unik dalam hal ekspresivitas, keterbukaan terhadap eksperimen, dan keterlibatan dalam bidang artistik. Sebaliknya, sembilan komunitas lainnya didominasi oleh individu heteroseksual dengan variasi yang minimal dalam orientasi seksual.

c. Spektrum Gaya Hidup

Analisis pola konsumsi alkohol, narkoba, dan kebiasaan merokok menunjukkan spektrum gaya hidup yang luas. Komunitas 0 menempati posisi paling liberal dengan tingkat konsumsi alkohol dan narkoba yang tinggi, mencerminkan gaya hidup urban yang terbuka dan eksperimental. Komunitas 5, 7, dan 8 berada di ujung konservatif dengan tingkat abstinasi yang tinggi terhadap narkoba dan kebiasaan merokok. Komunitas 5 bahkan menunjukkan proporsi signifikan individu yang sama sekali tidak mengonsumsi alkohol, mengindikasikan pengaruh nilai-nilai religius atau budaya yang lebih ketat.

d. Pola Keagamaan dan Spiritualitas

Distribusi afiliasi keagamaan menunjukkan polarisasi yang menarik. Komunitas 0 dan beberapa komunitas muda lainnya didominasi oleh ateis dan agnostik, mencerminkan tren sekularisasi di kalangan generasi muda urban. Komunitas 4 menunjukkan keunikan dengan dominasi agama Yahudi dan kategori "lainnya", mengindikasikan diversitas religius yang lebih tinggi. Sebaliknya, Komunitas 6, 8, dan 9 menunjukkan prevalensi yang lebih tinggi untuk agama-agama tradisional seperti Katolik dan Kristen, meskipun dengan interpretasi dan praktik yang mungkin beragam.

V. KESIMPULAN

Penelitian ini berhasil menerapkan metode Louvain untuk mendeteksi komunitas pengguna pada aplikasi kencan OkCupid, dengan memanfaatkan cosine similarity untuk mengukur kemiripan antar pengguna berdasarkan karakteristik demografis dan perilaku. Melalui analisis terhadap 995 data sampel dari pengguna, teridentifikasi sepuluh komunitas yang menunjukkan pola keterhubungan

yang signifikan dan beragam dalam karakteristik demografis, gaya hidup, dan preferensi individu. Hasil analisis menunjukkan bahwa komunitas yang terbentuk tidak hanya dipengaruhi oleh faktor demografis seperti usia dan orientasi seksual, tetapi juga oleh dinamika sosial yang lebih kompleks, mencerminkan segmentasi alami dalam populasi pengguna.

Modularitas yang tinggi (0,686) menunjukkan bahwa pengguna dalam komunitas memiliki keterhubungan internal yang kuat, sementara perbedaan yang signifikan antar komunitas menggambarkan variasi dalam nilai, gaya hidup, dan pola interaksi. Karena anggota komunitas menunjukkan kemiripan dalam preferensi dan gaya hidup, pemetaan ini dapat digunakan sebagai dasar dalam sistem rekomendasi yang memprioritaskan kesamaan karakteristik sosial, sehingga potensi keberhasilan koneksi interpersonal meningkat. Temuan ini dapat memberikan wawasan berharga bagi pengembang aplikasi kencan dalam merancang fitur yang lebih personal dan adaptif, serta sistem rekomendasi yang lebih efektif berdasarkan pola perilaku pengguna. Dengan demikian, penelitian ini tidak hanya memperkaya pemahaman tentang struktur sosial dalam konteks digital, tetapi juga membuka peluang untuk pengembangan lebih lanjut dalam analisis jejaring sosial dan aplikasi berbasis data.

REFERENCES

- [1] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. 1892.
- [2] M. Castells, "The Rise of the Network Society Second edition With a new preface".
- [3] B. WELLMAN, A. Q. HAASE, J. WITTE, and K. HAMPTON, "Does the Internet Increase, Decrease, or Supplement Social Capital?," *American Behavioral Scientist*, vol. 45, no. 3, pp. 436–455, Nov. 2001, doi: 10.1177/00027640121957286.
- [4] G. Ranzini and C. Lutz, "Love at first swipe? Explaining Tinder self-presentation and motives," *Mob Media Commun*, vol. 5, no. 1, pp. 80–101, Jan. 2017, doi: 10.1177/2050157916664559.
- [5] G. David and C. Cambre, "Screened Intimacies: Tinder and the Swipe Logic," *Soc Media Soc*, vol. 2, no. 2, Apr. 2016, doi: 10.1177/2056305116641976.
- [6] D. H. Felmlee and D. A. Kreager, "The Invisible Contours of Online Dating Communities: A Social Network Perspective," *Journal of Social Structure*, vol. 18, no. 1, pp. 1–28, Jan. 2017, doi: 10.21307/joss-2018-004.
- [7] A. Irsyad and N. A. Rakhmawati, "Community detection in twitter based on tweets similarities in indonesian using cosine similarity and louvain algorithms," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 6, no. 1, pp. 22–31, 2020, doi: 10.26594/REGISTER.V6I1.1595.
- [8] M. Newman, "Networks: An Introduction," *Networks: An Introduction*, pp. 1–784, Sep. 2010, doi: 10.1093/ACPROF:OSO/9780199206650.001.0001.
- [9] N. A. W. Yoga, "Kajian Graph Social Network Analysis Untuk Identifikasi Relasi Mahasiswa Its Pada Komunitas Keilmiah Berdasarkan Data Usulan Pkm," 2017.

- [10] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, Oct. 2008, doi: 10.1088/1742-5468/2008/10/P10008.
- [11] F. Sains and D. Teknologi, "Analisis dinamika komunitas wargnet twitter terkait diskusi agama islam tahun 2009-2019," Jul. 2024, Accessed: Jun. 02, 2025. [Online]. Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/80161>
- [12] "NetworkX — NetworkX documentation." Accessed: Jun. 02, 2025. [Online]. Available: <https://networkx.org/>
- [13] A. Mubarak, "Analisis Jaringan Sosial dengan Menggunakan Graf dan NetworkX," 2023.
- [14] O. Nurdiana, J. Jumadi, and D. Nursantika, "Perbandingan Metode Cosine Similarity Dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Qur'an dalam Bahasa Indonesia," *Jurnal Online Informatika*, vol. 1, no. 1, pp. 59–63, Jun. 2016, doi: 10.15575/JOIN.V1I1.12.
- [15] "(PDF) Vector Similarity Measures of Simplified Neutrosophic Sets and Their Application in Multicriteria Decision Making." Accessed: Jun. 03, 2025. [Online]. Available: https://www.researchgate.net/publication/281546974_Vector_Similarity_Measures_of_Simplified_Neutrosophic_Sets_and_Their_Application_in_Multicriteria_Decision_Making
- [16] A. Y. Kim and A. Escobedo-Land, "OkCupid data for introductory statistics and data science courses," *Journal of Statistics Education*, vol. 23, no. 2, 2015, doi: 10.1080/10691898.2015.11889737.