

1. Pendahuluan

Latar Belakang

Kemunculan platform jejaring sosial telah secara signifikan mengubah pola interaksi sosial. Selain memfasilitasi komunikasi, media sosial juga telah menjadi sarana pertukaran informasi yang efektif dan sumber ekonomi. Menurut DataReportal 2023, basis pengguna global situs jejaring sosial telah mencapai 4,88 miliar, setara dengan 60,6% dari populasi dunia[1]. Dampak besar penggunaan media sosial ini menghasilkan data yang besar dan beragam. Oleh karena itu, pemrosesan data jejaring sosial menjadi sangat penting untuk memperoleh pengetahuan berharga dari informasi yang terkandung dalam data tersebut.

Salah satu contoh masalah jejaring sosial adalah deteksi komunitas. Deteksi komunitas adalah proses mengidentifikasi kelompok atau komunitas dalam sebuah jaringan yang terdiri dari node-node yang saling terhubung[2]. Tujuan deteksi komunitas adalah untuk memahami struktur jaringan dan mengidentifikasi kelompok-kelompok yang memiliki pola koneksi yang serupa[3].

Untuk melakukan deteksi komunitas, diperlukan sumber daya dan algoritma yang efisien[4]. Saat ini, terdapat banyak *library* yang tersedia untuk memproses data jejaring sosial, seperti *NetworkX*, *iGraph*, *Scikit-Network*, *CDlib*, dan banyak lagi. Pemilihan *library* yang tepat merupakan hal yang sangat penting karena dapat mempengaruhi kinerja algoritma dalam deteksi komunitas.

Penelitian sebelumnya yang dilakukan oleh Thomas Bonald[5] telah menghasilkan *library* baru bernama *Scikit-Network*, yang dirancang untuk analisis graf. *Library* ini dibandingkan dengan *library* lainnya seperti *NetworkX*, *iGraph*, dan *Graph-tool* menggunakan algoritma seperti *Louvain*, *Pagerank*, *HITS*, dan *Spectral*, dengan fokus pada efisiensi *runtime* dan *memori*. Terdapat keterbatasan dalam penelitian ini, seperti *runtime* yang tidak tersedia di *NetworkX* ketika diimplementasikan dengan menggunakan algoritma *Louvain*.

Penelitian lain yang dilakukan oleh Giulio Rossetti telah mengembangkan *library* bernama *CDlib*[6], sebuah *library* deteksi komunitas untuk jejaring sosial berbasis *Python*. *CDlib* menyediakan akses ke 39 algoritma deteksi komunitas, seperti *Walktrap*, *Label Propagation*, *Girvan-Newman*, *Louvain*, dan *Infomap*. *CDlib* juga memiliki kemampuan evaluasi dan perbandingan, serta beragam kemampuan visualisasi. Hal ini menjadikan *CDlib* sebagai *library* yang komprehensif.

Penelitian ini mengevaluasi kinerja algoritma deteksi komunitas pada jejaring sosial menggunakan beberapa *library* berbasis *Python*, dengan mempertimbangkan *runtime*, *memory usage*, *modularity*, dan kemudahan penggunaan. Algoritma deteksi komunitas yang akan diimplementasikan adalah *Louvain*[7] dan *Label Propagation*[8]. *Library* yang akan digunakan adalah *NetworkX*, *iGraph*, *Scikit-Network*, dan *CDlib*.

Topik dan Batasannya

Penelitian ini membahas perbandingan performa algoritma deteksi komunitas *Louvain* dan *Label Propagation* dengan menggunakan berbagai *library*, yaitu *NetworkX*, *iGraph*, *Scikit-Network*, dan *CDlib*. Selain itu, penelitian ini juga mengevaluasi performa dari keempat *library* tersebut dalam memproses graf jejaring sosial berukuran besar. Batasan masalah dalam penelitian ini adalah sebagai berikut:

- a. Menggunakan fungsi yang tersedia pada *library* *NetworkX*, *iGraph*, *Scikit-Network*, dan *CDlib* untuk mengimplementasikan algoritma *Louvain* dan *Label Propagation*.
- b. Menggunakan tiga dataset bertipe *undirected* dengan ukuran yang berbeda, yang bersumber dari *public dataset*.

Tujuan

Tujuan dari penelitian ini adalah untuk mengetahui perbandingan performa algoritma deteksi komunitas dengan menggunakan *library* *NetworkX*, *iGraph*, *Scikit-Network*, dan *CDlib*. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi performa dari keempat *library* tersebut dalam memproses graf jejaring sosial berukuran besar berdasarkan kriteria *runtime*, *memory usage*, *modularity*, dan kemudahan penggunaan.

Organisasi Tulisan

Penulisan Tugas Akhir ini disusun dalam beberapa bagian yang dari setiap bagian, berisi data-data berikut ini: Bagian 1 – Pendahuluan, Bagian 2 – Studi Literatur, Bagian 3 – Uraian sistem yang dibangun, Bagian 4 – Evaluasi hasil pengujian, Bagian 5 – Kesimpulan & Saran.