

1. Pendahuluan

1.1 Latar belakang

Social network Analysis (SNA) adalah sebuah ilmu yang memfokuskan pada analisis di sebuah *network* untuk melihat relasi yang ada antara 1 *node* dengan *node* lainnya. Fungsi dari SNA sendiri sekarang banyak diterapkan di kehidupan sehari-hari seperti mencari aktor yang paling penting dalam sebuah komunitas, mencari grup atau komunitas yang ada pada sebuah web, dll. Pengimplementasian itu dapat diterapkan dalam perusahaan seperti untuk mencari karyawan yang akan dinaikan jabatannya, pencarian target pasar dengan sasaran komunitas tertentu, dll. Dalam Media sosial sendiri, secara tidak langsung dapat dikelompokkan menjadi komunitas-komunitas yang memiliki kesamaan atau similarity. Contoh dari kesamaan yaitu misal kesamaan dalam hobi, kesamaan keluarga, teman sekolah, dll. Tidak menutup kemungkinan seseorang menjadi anggota dari 2 atau lebih komunitas, hal tersebut biasa disebut dengan *Overlapping*. Telah muncul algoritma untuk mencari dan mendeteksi *Overlapping* yang ada pada sebuah *network*.

Beberapa algoritma yang sudah ada antara lain CPM (Clique Percolation Method), FCM (Fuzzy C-Means), NMF (Non Negative Matrix Factorization). Clique Percolation Method atau yang lebih dikenal dengan CPM adalah sebuah algoritma yang diciptakan oleh Gergely Palla, Imre Derényi, Illés Farkas, dan Tamás Vicsek [18] untuk menentukan *Overlapping* yang ada didalam komunitas yang random. Algoritma ini memfokuskan pada penganalisan pada *k-cliques* yaitu subgraph yang ada pada sebuah komunitas yang terdiri dari *vertice*. Setelah diketahui *k-cliques* kemudian akan dilakukan proses pengclusteran. Sebuah aplikasi bernama Cfinder adalah sebuah aplikasi yang dilandasi oleh algoritma CPM [1]. Algoritma selanjutnya adalah FCM atau Fuzzy C-Means Clustering pertama kali diperkenalkan oleh Dunn di tahun 1973 dan dikembangkan lagi oleh Jim Bezdek pada tahun 1981 yang bertujuan untuk mengklusterkan tiap data dimana 1 *node* bisa menjadi anggota 2 atau lebih kelompok (cluster) [27].

Salah satu algoritma yang ada adalah CONGA atau Cluster-Overlap Girvan Newman Algorithm adalah sebuah algoritma pendeteksi overlap dengan metode memisahkan atau menghapus relasi antar *vertices* dengan mempertimbangkan perhitungan *edge betweenness* dan *betweenness centrality*. CONGA sendiri adalah sebuah algoritma pengembangan dari algoritma yang sebelumnya sudah sangat terkenal yaitu GN (*Girvan Newman*) Algoritma. Perbedaan yang ada adalah pada proses pemisahan *vertices* karena pada metode ini lebih memspesifikasi kapan dan bagaimana memisahkan *vertices* yang ada.

Dalam *paper* ini akan dijelaskan dan dipaparkan bagaimana pengimplementasian serta penganalisan dari sebuah algoritma yang bernama CONGA untuk mencari *Overlapping* pada jejaring sosial Twitter. Data yang diambil akan berasal dari twitter karena saat ini twitter adalah sebuah *microblogging* yang sedang *trend* di kalangan masyarakat. Penulis ingin membuktikan performansi dari sebuah algoritma CONGA dengan twitter.

1.2 Perumusan masalah

Melihat dari latar belakang yang dijelaskan pada halaman sebelumnya, maka dapat dijelaskan masalah yang akan dibahas dalam Tugas Akhir ini adalah:

1. Bagaimana cara mendeteksi overlap yang ada menggunakan algoritma CONGA?
2. Apa saja faktor yang menentukan overlap dengan menggunakan algoritma CONGA?
3. Bagaimana performansi algoritma CONGA dalam pengujian di twitter

Adapun yang menjadi batasan masalah dalam Tugas Akhir ini adalah :

1. Penelitian Tugas Akhir ini menangani anggota komunitas yang ikut dalam 2 komunitas atau lebih, yang biasa disebut dengan tumpang tindih (*overlap*)
2. *Microblogging* yang digunakan adalah twitter
3. Penelitian nilai keterhubungan 2 *vertices* menggunakan *Edge Betweenness*
4. Perhitungan nilai *vertices* overlap menggunakan *Split Betweneess*
5. Pengukuran optimalisasi menggunakan nilai *modularity*
6. Pembobotan hanya untuk menggambarkan kedekatan antar *node*
7. Parameter pembobotan ditentukan dari beberapa faktor, antara lain *followed*, *reply*, dan *mention* dan tidak menggunakan aktifitas didalam tweet tersebut seperti *hash tag*, lokasi, dan isi tweet
8. Graf yang digunakan adalah graf yang tidak berbobot dan tidak berarah (*unpartite and undirect graph*)
9. Komunitas yang terbentuk berdasarkan kedekatan pada waktu tertentu
10. Tidak diberikan pendefinisian setiap komunitas yang terbentuk

1.3 Tujuan

Tujuan dari penelitian Tugas Akhir ini adalah sebagai berikut :

1. Mendeteksi *Overlapping* yang ada dengan menggunakan algoritma CONGA
2. Menentukan faktor-faktor yang mempengaruhi hasil nilai overlap dalam algoritma CONGA
3. Menganalisis performansi algoritma CONGA dalam pengujian di Twitter

1.4 Metodologi penyelesaian masalah

Metode penelitian yang saya gunakan adalah :

1. Studi Pustaka
Dengan membaca, mempelajari dan memahami literatur-literatur yang ada dan berkaitan dengan *community detection*, *CONGA Algorithm*, dan *modularity*. Literatur yang ada berupa antara lain artikel, jurnal-jurnal gratis yang ada internet, video pengajaran, dll.

2. Pengambilan dan Pengumpulan Data
Pengambilan data diambil dari jejaring sosial Twitter, lalu datanya akan diambil menggunakan aplikasi *nodeXL* yang dapat diunduh secara gratis
3. Analisis dan Perancangan Sistem
Membuat analisis dan rencana perancangan sistem yang akan dibuat. Serta membuat analisis proses implementasi yang akan digunakan seperti bahasa pemrograman, algoritma yang akan digunakan, dan fungsionalitas yang dibutuhkan.
4. Implementasi metode *preprocessing* data
 - a. Pengambilan data tidak berbobot (*unweighted graph*) dalam bentuk matriks adjecancy berdasarkan faktor *followed/following*, *mention*, dan *reply*
 - b. Pembobotan matriks *followed/following*, *reply*, *mention* dengan menggunakan aturan yang telah ditentukan
 - c. Penggabungan matriks *followed/following*, *mention*, *reply* menjadi matriks hasil pembobotan
5. Tahapan Implementasi metode
 - a. Pembentukan komunitas dan *Overlapping* dengan menggunakan *CONGA Algorithm*
 - b. Perhitungan nilai *modularity*
6. Pengambilan Kesimpulan dan Penyusunan Laporan Akhir
Pengambilan kesimpulan berdasarkan hasil dari tiap tahap yang dilakukan kemudian disusun menjadi sebuah laporan Tugas Akhir yang berisi dokumentasi dari semua hasil tersebut.

1.5 Sistematika Penulisan

Adapun sistematika penulisan Tugas Akhir ini adalah sebagai berikut:

1. BAB 1 Pendahuluan
Pada bab 1 dijelaskan isi dan rencana pengerjaan penelitian Tugas Akhir secara keseluruhan yang meliputi latar belakang, rumusan masalah, tujuan, batasan masalah dan metode penyelesaian yang diterapkan.
2. BAB 2 Dasar Teori
Bab 2 memaparkan dasar-dasar teori yang berkaitan dengan *Social Network Analysis*, *Social Media*, *Algoritma CONGA*
3. BAB 3 Perancangan dan Implementasi
Perancangan sistem, analisis kebutuhan sistem, dan contoh implementasi dari sistem yang akan dibangun akan dipaparkan pada bab ini. Selanjutnya akan dilakukan proses implementasi.
4. BAB 4 Pengujian dan Analisis
Pada bab ini akan dibahas scenario dan hasil pengujian yang dilakukan pada hasil implementasi sistem.
5. BAB 5 Kesimpulan dan Saran
Bab ini berisi kesimpulan dan saran yang didapatkan dari hasil implementasi sistem secara keseluruhan.