

BAB I PENDAHULUAN

1.1 Latar belakang masalah

Twitter telah berkembang dengan pesat menjadi situs *social network* populer dan telah memiliki jutaan pengguna yang tersebar di berbagai belahan dunia. Sekarang ini pengguna *twitter* telah mencapai 271 juta pengguna aktif perbulan dan perharinya ada 500 juta *tweets* yang dikirim. *Twitter* saat ini tidak hanya sekedar digunakan untuk media pertemanan, namun seringkali juga digunakan sebagai media bertukar informasi, media iklan sebuah produk, media kampanye partai-partai politik, dan media propaganda bagi kelompok-kelompok yang memiliki tujuan tertentu.

Perkembangan *social network* yang pesat menjadi bahan penelitian yang menarik di kalangan penggiat ilmu informatika, salah satunya adalah *community detection* pada jaringan kompleks seperti *social network* [12]. Dalam *community detection*, tujuannya adalah untuk membagi jaringan ke dalam daerah-daerah pada graph. Daerah tersebut biasanya sesuai dengan entitas-entitas yang memiliki hubungan erat, dan dapat dikatakan sebagai satu komunitas [10]. Pembentukan komunitas ini kemudian dapat digunakan untuk berbagai keperluan seperti menargetkan skema pemasaran produk, menghitung kepopuleran suatu barang, mendeteksi isu di masyarakat, mendeteksi jaringan sel teroris, dan masih banyak lagi [10]. Interaksi antar user pada *twitter* terbagi menjadi 4 macam yaitu interaksi *mention*, *reply*, *follower* dan *followed*. Interaksi inilah yang akan dijadikan nilai dasar untuk pembentukan komunitas pada *community detection*.

Beberapa algoritma yang sering digunakan untuk penelitian tentang *community detection* diantaranya adalah algoritma *Girvan-Newman*, *K-L algorithm* dan *hierarchical clustering method based on similarity measurement*. Algoritma tersebut telah menghasilkan komunitas yang baik, namun algoritma tersebut belum mampu menghasilkan komunitas dengan penentuan jumlah komunitas yang objektif. Maka dari itu dibutuhkan algoritma yang dapat membentuk komunitas dengan penentuan jumlah komunitas yang objektif dalam jaringan. Salah satu algoritma yang dapat digunakan untuk *community detection* adalah Algoritma yang di dasarkan pada *Genetic Algorithm*. Suatu algoritma mengharuskan pengelompokan komunitas menetapkan batasan sejauh mana suatu kelompok dapat dianggap sebagai komunitas [12]. *Genetic Algorithm* dipilih dalam penelitian ini karena dapat menemukan komunitas terbaik dengan memaksimalkan modularitas jaringan. Algoritma ini memiliki $O(e)$ kompleksitas waktu dan tidak perlu mengetahui terlebih dahulu tentang jumlah komunitas atau nilai ambang batas, yang membuat algoritma berguna dalam jaringan di kehidupan nyata sehingga algoritma ini lebih objektif dalam menentukan banyaknya komunitas dalam suatu jaringan [12].

1.2 Perumusan masalah

Perumusan masalah dari penelitian ini adalah :

- a. Bagaimanakah pengaruh interaksi *mention*, *reply* dan *follow* antar user di *social network twitter* dalam perhitungan *similarity*?
- b. Bagaimanakah pembentukan komunitas menggunakan *Genetic Algorithm*?
- c. Bagaimanakah nilai *modularity* dari pembentukan komunitas menggunakan algoritma yang didasarkan pada *Genetic Algorithm*?

1.3 Tujuan

Tujuan dari penelitian ini adalah :

- a. Menganalisis pengaruh interaksi *mention*, *reply* dan *follow* antar user di *social network twitter* dalam perhitungan *similarity*
- b. Menganalisis proses pembentukan komunitas yang dibentuk menggunakan *Genetic Algorithm*
- c. Menganalisis nilai *modularity* dari pembentukan komunitas menggunakan algoritma yang didasarkan pada *Genetic Algorithm*

1.4 Batasan masalah

Batasan masalah pada penelitian ini adalah :

- a. Perhitungan kedekatan antar user didasarkan pada interaksi *mention*, *reply*, *follower* dan *followed*
- b. Isi *tweet*, hash tag, dan lokasi tidak diproses dalam penelitian ini
- c. Pembentukan komunitas tidak mendefinisikan nama komunitas yang terbentuk
- d. Tidak menjelaskan proses penentuan topik pada komunitas yang terbentuk

1.5 Metodologi penelitian

Metodologi penelitian pada penelitian ini menggunakan tahapan pengerjaan sebagai berikut :

1. Studi pustaka
Pada tahapan ini penulis mempelajari sumber tertulis seperti buku, literature, artikel, jurnal, dan internet yang berhubungan dengan community detection menggunakan *Genetic Algorithm*
2. Pengumpulan Data
Pengambilan data di social network *twitter* menggunakan *software NodeXL*.
3. Analisis kebutuhan perancangan sistem
Melakukan analisis terhadap kebutuhan sistem dan melakukan perancangan proses implementasi sistem.

4. Implementasi Metode
 - a. *Preprocessing* data
 - Parsing
 - Pembuatan matriks $n \times n$
 - Pembuatan matriks *similarity*
 - b. Pembentukan dan perhitungan komunitas menggunakan *Genetic Algorithm*.
5. Analisis hasil dan kesimpulan
 - a. Analisis hasil perhitungan nilai *similarity*
 - b. Analisis hasil komunitas yang dibentuk menggunakan *Genetic Algorithm*
 - c. Penarikan kesimpulan
6. Pembuatan Laporan

Melakukan dokumentasi penelitian berupa tahap-tahap dari setiap kegiatan yang dilakukan.