

Analisis dan Implementasi Algoritma CNM-Centrality untuk Deteksi Komunitas pada Dataset Facebook

Farah Rahmadiani¹, Imelda Atastina, S.Si.,M.T.², Anisa Herdiani,S.T.,M.T.³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹farahrahmadiani@students.telkomuniversity.ac.id, ²ImeldaAtastina@telkomuniversity.ac.id,

³AnisaHerdiani@telkomuniversity.ac.id

Abstrak

Deteksi komunitas menjadi salah satu *task* dari *social network analysis* yang banyak dibahas. Hal ini disebabkan dengan pendeteksian komunitas tersebut dapat digunakan untuk berbagai hal seperti skema pemasaran produk, mendeteksi isu di masyarakat, dan sebagainya.

Masalah dalam mendeteksi komunitas yaitu bagaimana cara mengelompokkan simpul yang baik untuk mengidentifikasi suatu kelompok. Salah satu algoritma untuk mendeteksi komunitas adalah CNM-Centrality. CNM-Centrality merupakan algoritma Cluset-Newman-Moore yang dikombinasi dengan algoritma PAGERANK. Algoritma PAGERANK digunakan untuk menentukan simpul terpenting sedangkan algoritma CNM digunakan untuk mendeteksi komunitas. Algoritma ini dianggap algoritma dengan kecepatan *running time* yang baik dan nilai *modularity* yang baik pada pengujian sebelumnya. Namun, pada pengujian sebelumnya hanya menguji beberapa data saja seperti data jazz network dengan 198 simpul dengan relasi sebanyak 2742, belum diketahui bagaimana performansi algoritma CNM-Centrality jika menggunakan dalam jumlah yang besar.

Algoritma CNM-Centrality mempunyai parameter yaitu *damping factor*. *Damping factor* merupakan parameter yang digunakan untuk menghitung nilai *rank* dan bernilai antara 0 sampai 1. Pada pengujian ini akan menggunakan dataset Facebook dengan tiga jenis *damping factor* yang berbeda untuk mengetahui pengaruh *damping factor* terhadap nilai *modularity* dan juga keandalan performansi algoritma CNM-Centrality untuk dataset Facebook. Penelitian ini bermanfaat untuk menilai apakah algoritma CNM-Centrality mempunyai nilai *modularity* dan *running time* yang baik dengan dataset yang berbeda dan *damping factor* yang berbeda.

Hasil dari pengujian ini adalah nilai *damping factor* 0,15 memperoleh nilai *rank* tertinggi daripada dua *damping factor* yang diujikan. Namun, *damping factor* 0,85 memperoleh *running time* yang baik untuk menentukan simpul terpenting dan mendeteksi komunitas dibandingkan dua *damping factor* yang diuji. Selain itu, *modularity* terbaik yang dihasilkan pada pengujian ini adalah 0,8318 yang diperoleh oleh *damping factor* 0,85.

Kata kunci : Deteksi Komunitas, Algoritma CNM-Centrality, *Damping factor*, *Modularity*

Abstract

Community detection becomes one of the tasks of social network analysis that is widely discussed because with the detection of the community can be used for various things such as product marketing schemes, detect issues in the community, and so forth.

The problem in detecting communities is how to group good nodes to identify a group. One of the algorithms for detecting communities is CNM-Centrality. CNM-Centrality is a Cluset-Newman-Moore algorithm combined with PAGERANK algorithm. PAGERANK algorithm is used to determine the most important node while the CNM algorithm is used to detect the community. This algorithm is considered an algorithm with good running time speed and good modularity value in previous test. However, in the previous test only tested some data only, not yet known how CNM-Centrality algorithm performance if using in large amount.

The parameter used in CNM-Centrality Algorithm is damping factor. Damping factor is a parameter used to calculate rank values and values between 0 and 1. In this test will use the dataset up with three different damping factor to determine the effect of damping factor on the value of modularity and also the reliability of CNM-Centrality algorithm performance for Facebook dataset . This research is useful to assess whether the CNM-Centrality algorithm has good modularity and running time values with different datasets and different damping factors.

The result of this test is the damping value factor 0.15 obtains the highest rank value rather than the two damping factors tested. However, the damping factor of 0.85 obtains good running time to determine the most important node and detect the community over the two damping factors tested. In addition, the best modularity produced in this test is 0.8318 obtained by the damping factor of 0.85

Keywords: Community Detection, CNM-Centrality Algorithm, Damping factor, Modularity

1. Pendahuluan

Latar Belakang

Internet merupakan salah satu media yang banyak digunakan untuk berkomunikasi. Menurut data pada tahun 2017, sebanyak 3,578 miliar masyarakat dunia menggunakan internet [1]. Perkembangan ini menyebabkan banyaknya muncul *social network* dan situs media sosial yang mana ini juga akan mengakibatkan masyarakat menjadi lebih mudah terhubung satu dengan yang lain dalam *cyberspace* [2].

Social network adalah suatu interaksi sosial antar individu baik itu berupa interaksi dalam hal berbagi informasi, ide dan sebagainya dalam jaringan. *Social network* dapat direpresentasikan berupa graf, $G(V,E)$. G menyatakan graf, V menyatakan *vertex* yang merepresentasikan simpul atau aktor, dan E menyatakan *edge* yang merepresentasikan koneksi atau relasi. Facebook merupakan salah satu contoh *social network* yang diminati oleh masyarakat di dunia. Terbukti, pada tahun 2017 sebanyak 2,2 miliar masyarakat menggunakan Facebook [3]. Studi yang mempelajari tentang hubungan antar individu dalam suatu jaringan termasuk mempelajari tentang struktur sosial, menganalisis peranan suatu individu (aktor), dan posisi sosial yaitu *Social Network Analysis* [2].

Salah satu *task* dari *Social Network Analysis* adalah deteksi komunitas yang bertujuan untuk mengidentifikasi komunitas melalui struktur jaringan. Adanya pendeteksian suatu komunitas pada jaringan menggambarkan suatu komunitas. Pendeteksian komunitas tersebut dapat digunakan untuk berbagai hal seperti skema pemasaran produk, mendeteksi isu dimasyarakat, dan sebagainya [4].

Masalah dalam mendeteksi komunitas yaitu bagaimana cara mengelompokkan simpul yang baik untuk mengidentifikasi suatu kelompok. Salah satu algoritma untuk mendeteksi komunitas adalah Algoritma CNM-Centrality. Algoritma CNM-Centrality merupakan algoritma Clauset-Newman-Moore yang dikombinasi dengan algoritma Pagerank. Algoritma Pagerank digunakan untuk menentukan simpul terpenting sedangkan algoritma CNM digunakan untuk mendeteksi komunitas. Pada Algoritma Pagerank terdapat parameter yang mempengaruhi hasil dari nilai *rank*, yaitu *damping factor*. Nilai *damping factor* telah ditetapkan dari 0 hingga 1 [5]. Nilai *rank* yang tertinggi akan menjadi node terpenting yang akan digunakan algoritma CNM-Centrality untuk mendeteksi komunitas. Algoritma CNM-Centrality dianggap mempunyai kecepatan *running time* dan nilai *modularity* yang baik oleh FangHu,dkk [6]. Namun, pada pengujian sebelumnya hanya menguji beberapa data saja seperti data jazz network yang hanya memiliki 198 simpul dengan relasi sebanyak 2742. Pada penelitian ini akan menggunakan dataset Facebook yang digunakan sebagai data uji untuk mengetahui keandalan performansi algoritma CNM-Centrality pada dataset yang lebih besar.

Topik dan Batasannya

Dari latar belakang yang telah dipaparkan, didapatkan beberapa rumusan masalah pada penelitian ini, yaitu pengaruh simpul terpenting terhadap terbentuknya komunitas, pengaruh nilai *damping factor* terhadap *running time* Algoritma Pagerank, pengaruh nilai *damping factor* terhadap perhitungan nilai *modularity*, analisis performansi penggunaan algoritma CNM-Centrality dilihat dari hasil perhitungan kualitas komunitas dengan menggunakan *modularity*.

Adapun batasan masalah dalam pengerjaan Tugas Akhir ini, yaitu : Dataset yang digunakan merupakan dataset Facebook dari situs <http://snap.stanford.edu/data>, tidak mendefinisikan nama komunitas yang terbentuk, data *network* digunakan adalah *network* yang tidak berarah dan tidak berbobot, pengukuran kualitas komunitas hanya dengan menggunakan *modularity*.

Tujuan

Terkait permasalahan yang dipaparkan di atas, dapat ditarik tujuan yang akan dicapai dalam Tugas Akhir ini adalah menganalisis pengaruh *damping factor* terhadap *running time* Algoritma Pagerank, menganalisis pengaruh simpul terpenting terhadap terbentuknya komunitas, menganalisis pengaruh nilai *damping factor* terhadap nilai *modularity* suatu komunitas, dan menganalisis performansi Algoritma CNM-Centrality dalam mendeteksi komunitas dengan menghitung nilai *modularity*.

Organisasi Tulisan

Pada penulisan Tugas Akhir ini akan dibagi menjadi beberapa bagian yaitu :

1. Pendahuluan : Pada bagian ini menjelaskan latar belakang, Topik dan Batasan, Tujuan dan Organisasi Tulisan untuk Tugas Akhir ini.
2. Studi Tekait : Pada bagian ini berisi teori/studi literatur yang mendukung Tugas Akhir.
3. Sistem yang Dibangun : Pada bagian ini menjelaskan rancangan dan sistem yang akan dibangun.
4. Evaluasi : Bagian ini berisi dua sub bagian yaitu : Hasil Pengujian dan Analisis Hasil Pengujian.
5. Kesimpulan : Bagian kesimpulan memuat kesimpulan dan saran yang diambil dari hasil dan analisis hasil pengujian

2. Studi Terkait

Algoritma CNM-Centrality

Algoritma CNM-Centrality didefinisikan sebagai berikut : Simpul penting (*central node*) dalam sebuah jaringan diidentifikasi oleh algoritma *Pagerank* dengan menggunakan persamaan dibawah ini

$$x = \text{div} \left((I - P * G * D) \right)^* a \quad (1)$$

Dimana , I merupakan matriks identitas, p merupakan parameter damping factor, G merupakan matriks ketetanggaan, D merupakan *sparse matrix* dan a merupakan matriks berdimensi satu. Setelah melakukan identifikasi simpul penting, hasilnya akan diurutkan berdasarkan dari nilai tertinggi ke nilai terendah. Flowchart dari langkah-langkah untuk menentukan simpul terpenting dapat dilihat pada lampiran a. Setiap simpul akan diinisialisasi sebagai komunitas, dan *modularity* akan dihitung sebagai nilai awal. Setelah menghitung nilai awal, nilai maksimum dari *modularity* akan didapatkan dengan membangun *max-heap*. *Max-heap* merupakan urutan nilai *modularity* yang telah diurutkan dari nilai *modularity* tinggi ke nilai *modularity*. Nilai maksimal dari *modularity* setiap relasi antar simpul dapat dihasilkan pada *max-heap*. Setelah itu, jika saat penggabungan komunitas terdapat simpul terpenting pada salah satu komunitas atau bahkan tidak memiliki simpul terpenting, maka dua komunitas tersebut dapat digabungkan. Saat dua komunitas ini merupakan simpul terpenting, algoritma CNM-Centrality akan menghitung nilai *SNN Similarity* untuk menentukan apakah kedua komunitas tersebut dapat digabungkan atau tidak. Jika nilai *Similarity* antara dua komunitas kurang dari 0,5, kedua komunitas tersebut tidak dapat digabungkan dan memilih nilai *modularity* yang terbesar lainnya yang berada pada *max-heap*. Setelah penggabungan, nilai *modularity* akan diperbarui. Algoritma ini akan berhenti saat tidak ada komunitas yang dapat digabungkan, dan kemudian menghasilkan *output*. Adapun langkah-langkah perhitungan untuk Algoritma CNM-Centrality yang dapat dilihat pada paper acuan [6]. Flowchart dari langkah-langkah dalam algoritma ini dapat dilihat pada lampiran a dan b.

Modularity

Ada beberapa kriteria penilaian yang digunakan untuk mengetahui nilai hasil deteksi komunitas. Salah satunya yaitu *modularity*. *Modularity* merupakan ukuran untuk mengetahui seberapa baik komunitas yang terbentuk [7].

Persamaannya, sebagai berikut :

$$Q = \sum_{s=1}^K \left[\frac{m_s}{m} - \left(\frac{d_s}{2m} \right)^2 \right] \quad (2)$$

Dengan Keterangan :

K = Jumlah komunitas pada *network*.

m = Jumlah dari relasi pada *network*.

ms = Jumlah dari relasi pada komunitas s.

ds = Jumlah *degree* pada semua simpul di komunitas s.

Algoritma Shared Nearest Neighbor (SNN)

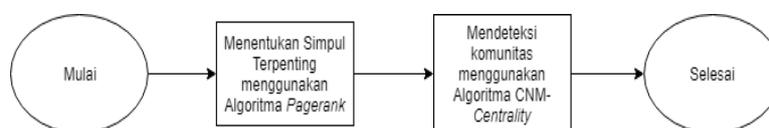
Pendekatan SNN merupakan pendekatan yang memiliki prinsip sebagai berikut : Jika dua titik memiliki kesamaan terhadap titik yang sama banyak, maka kedua titik tersebut memiliki kesamaan satu sama lain, bahkan jika pengukuran kesamaan tidak menunjukkan kesamaan tersebut. Ide kunci dari algoritma ini adalah mengambil jumlah dari titik-titik data untuk menentukan pengukuran kesamaan. Kesamaan dalam algoritma SNN didasarkan jumlah tetangga yang dimiliki secara bersama-sama selama kedua obyek terdapat dalam daftar tetangga terdekat masing-masing [8].

$$\sigma(v, w) = \frac{|\tau(v) \cap \tau(w)|}{\sqrt{|\tau(v)| \times |\tau(w)|}} \times \log(x(w)) \quad (3)$$

Dengan keterangan, $\tau(v)$ merupakan set dari relasi simpul v dan x(p) merupakan nilai *rank* untuk simpul w [6].

3. Sistem yang dibangun

Pada bab ini akan membahas tentang perancangan sistem yang akan dibangun. Berikut adalah tahapan-tahapan dari perancangan sistem ini:



Gambar 1. Gambaran Umum Sistem

Menentukan Simpul Terpenting menggunakan Algoritma Pagerank

. Pada Algoritma *Pagerank* membutuhkan matriks identitas, sparse matriks, matriks ketetanggaan (dan juga matriks a. Selain itu, terdapat nilai p yang merupakan nilai *damping factor*. Nilai parameter tersebut dapat dipilih antara 0 sampai 1.

Matriks jarang atau *sparse matrix* merupakan matrik n x n. Nilai pada setiap elemen diagonal pada matriks ini adalah $\frac{1}{\text{jumlah nilai 1 pada matriks relasi (G)}}$ dan elemen yang lain bernilai 0. Berikut contoh matriks jarang.

Tabel 1. *Sparse Matrix*

x/y	1	2	3
1	0,5	0	0
2	0	0,5	0
3	0	0	0,333

Matriks a merupakan matriks satu dimensi dan bernilai 1 pada setiap elemennya. Sehingga, jika memiliki tiga buah simpul, maka matriks a akan berukuran 3 x 1 dengan nilai tiap elemennya adalah 1. Setelah membangun matriks identitas, matriks ketetanggaan, *sparse matrix* dan matriks a, maka akan dihitung nilai *rank* sesuai persamaan 1.

Setelah mendapatkan nilai *rank*, nilai *rank* tersebut akan diurutkan berdasarkan nilai terbesar ke nilai terkecil. Simpul yang merupakan simpul terpenting akan didapat berdasarkan nilai *rank* terbesar dan banyaknya simpul terpenting dapat dihitung seperti contoh dibawah ini :

$$\text{Node Terpenting} = \sqrt{10/2} = 2$$

Dari 10 simpul hanya terdapat 2 simpul yang dianggap sebagai simpul terpenting. Maka, simpul terpenting adalah simpul yang nilai *rank* terbesar pada peringkat ke satu dan peringkat dari hasil nilai *rank* yang telah diurutkan. Kemudian, simpul tersebut akan menjadi simpul pusat saat proses penggabungan.

4. Evaluasi

Dalam pengujian ini, dataset yang digunakan merupakan dataset facebook yang didapat dari SNAP Stanford University berupa simpul yang saling berelasi. Data tersebut merepresentasikan pengguna facebook (simpul) yang saling terhubung dan merupakan dataset graf yang tidak berarah yang berarti jika simpul satu berelasi dengan simpul dua maka simpul dua juga berelasi dengan simpul satu [9]. Pada pengujian ini akan digunakan 10 dataset dengan jumlah simpul dan relasi yang berbeda untuk mengetahui pengaruh jumlah data terhadap pembentukan komunitas. Berikut jumlah simpul dan relasi yang digunakan pada pengujian ini.

Tabel 2. Dataset Pengujian

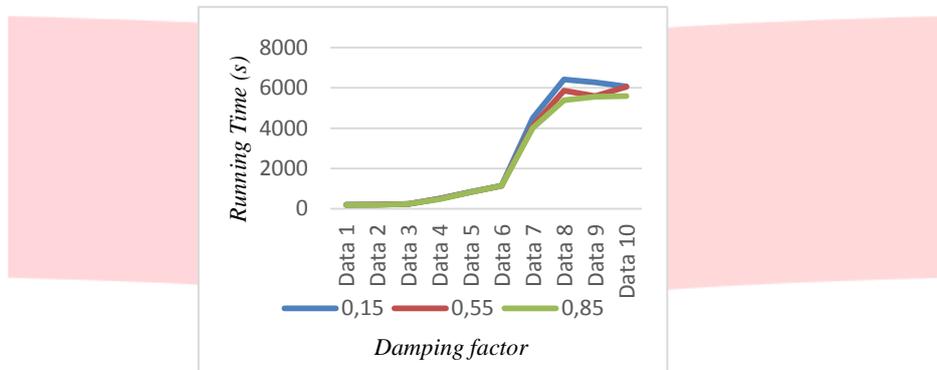
Data	Jumlah Simpul	Jumlah Relasi
1	729	7676
2	730	6232
3	733	7044
4	942	11286
5	1123	15914
6	1588	29426
7	1967	45506
8	2165	51188
9	2198	53520
10	2265	57308

Pengujian ini juga akan diuji 3 *damping factor* yang berbeda yaitu, mendekati 0, antara 0,5-0,6 dengan 0,85-0,9. Menurut hasil 3 pengujian sebelumnya graf dengan CitaSeer *citation* memiliki nilai yang baik menggunakan *damping factor* antara 0,5-0,6 [10], graf dengan Power network memiliki nilai yang baik menggunakan *damping factor* mendekati 0 [11] dan graf dengan Karate network memiliki nilai yang baik jika menggunakan *damping factor* 0,85-0,9 [6]. Kemudian, pada pengujian ini akan diujikan setiap *damping factor* dengan 3 nilai batas SNN *Similarity* untuk proses penggabungan yaitu : 0,1, 0,5, dan 0,9. Ini disebabkan menurut hasil 3 pengujian sebelumnya, 0,1 mendapatkan hasil 85,05% pada dataset Pendigits, 0,5 mendapatkan hasil 96,61% pada dataset Wine dan 0,9 mendapatkan hasil 100% pada dataset Iris. [8] Dengan tiga *damping factor* tersebut akan digunakan untuk menganalisa pengaruh *damping factor* terhadap *running time* pada algoritma *Pagerank*, dan

juga pengaruh *damping factor* dan nilai terhadap *modularity* dan *running time* pada algoritma CNM-Centrality. Setelah mendapatkan simpul terpenting yang dihasilkan pada algoritma *Pagerank*, akan dianalisis pada hasil CNM-Centrality untuk melihat pengaruh simpul terpenting terhadap *modularity per cluster*.

4.1 Hasil Pengujian

1. Hasil pengujian pengaruh *damping factor* terhadap *running time* dalam menentukan simpul terpenting. Pada gambar 1, dapat dilihat kecepatan *running time* untuk *damping factor* 0,15 lebih lama dibandingkan dua *damping factor* yang telah diujikan. Untuk hasil tabel pengujian pada gambar 1 dapat dilihat pada tabel yang



telah dilampirkan pada lampiran c.

Gambar 2. Pengaruh *damping factor* terhadap *running time* dalam menentukan simpul terpenting

2. Hasil pengujian pengaruh simpul terpenting terhadap *modularity per cluster*. *Modularity per cluster* merupakan *modularity* yang dilihat dari tiap cluster yang terbentuk dalam satu data. Setiap data memiliki simpul terpenting yang didapatkan dari proses menentukan sebelumnya sehingga pada tabel ini akan menampilkan maksimal 5 simpul terpenting.

Berikut merupakan tabel hasil pengujian simpul terpenting yang telah dilakukan terhadap dataset ke sepuluh dan batasan *Similarity* 0,9 .

Tabel 3. Hasil pengujian pengaruh simpul terpenting terhadap *modularity per komunitas* pada dataset kesepuluh.

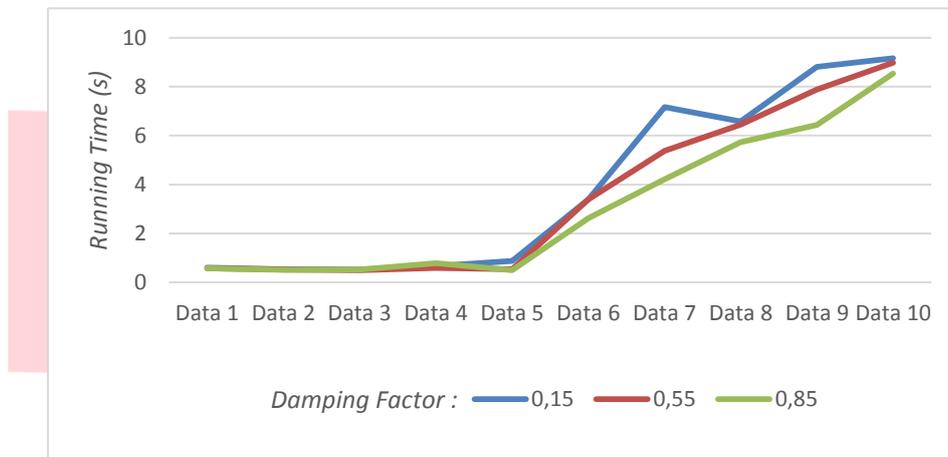
Data	<i>Damping factor</i>								
	0,15			0,55			0,85		
	Simpul penting	Jumlah Anggota	<i>Modularity per cluster</i>	Simpul Penting	Jumlah Anggota	<i>Modularity per cluster</i>	Simpul Penting	Jumlah Anggota	<i>Modularity per cluster</i>
10	1983	140	0,137881	1471	673	0,213498	3836	11	0,0015
				1912	380	0,154607	3325	10	0,00118
				1107	960	0,234005	3747	225	0,04998
				1238	1	3,02E-05	1440	655	0,28166
	2543	1934	0,588191	1835	1	3,48E-05	3075	1	3,49E-05
				1238	1	3,02E-05	1137	19	0,00413
				1791	1	3,12E-05	1602	838	0,253306
						3967	1	3,49E-05	

Pada tabel 3, dapat dilihat bahwa beberapa simpul terpenting memiliki nilai yang rendah dan juga memiliki jumlah relasi yang sedikit contohnya saja simpul terpenting 1835 yang hanya memiliki satu buah anggota komunitas dan memiliki nilai *modularity* sebesar 3,48E-05. Namun, ada beberapa simpul terpenting yang memiliki nilai tertinggi dengan jumlah anggota komunitas yang banyak contohnya simpul 2543 yang memiliki 1290 anggota dengan nilai *modularity* 0,435849. Pada dataset ke satu sampai ke sembilan pun memiliki hasil yang sama bahwa nilai *modularity* akan besar jika cluster tersebut merupakan simpul terpenting dan memiliki hasil yang relasi yang banyak dibandingkan *modularity cluster* yang lain sehingga tabel 3 merupakan contoh pengaruh simpul terpenting terhadap *modularity per cluster*.

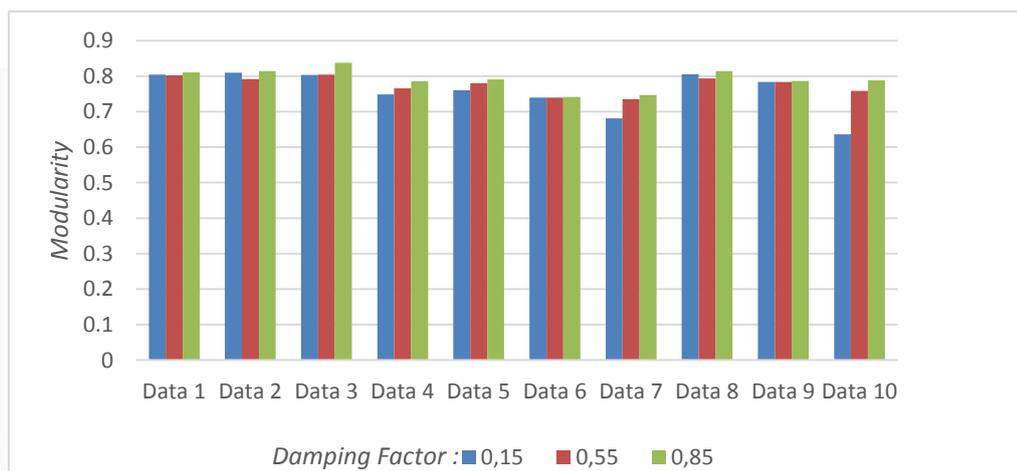
Pada tabel 3, dapat dilihat dari jumlah simpul terpenting hanya ada beberapa simpul terpenting yang menjadi pusat. Ini disebabkan adanya simpul terpenting yang dapat digabungkan dengan simpul yang lainnya

3. Hasil pengujian pengaruh *damping factor* terhadap nilai *modularity* dan *running time* dalam mendeteksi komunitas.

Hasil pengujian pengaruh *damping factor* dan nilai batasan SNN Similarity : 0,9 terhadap *modularity* pada setiap data yang diujikan.



Gambar 4. Hasil pengujian pengaruh *damping factor* terhadap *modularity* dengan nilai batasan SNN-Similarity =0,9



Gambar 3. Hasil pengujian pengaruh *damping factor* terhadap *modularity* dengan nilai batasan SNN Similarity = 0,9

Pada Gambar 3, dapat dilihat bahwa *modularity* tertinggi diperoleh oleh data ketiga dengan *damping factor* 0,85 dengan nilai *modularity* sebesar 0,838059 dengan batasan SNN Similarity sebesar 0,9. Hasil untuk SNN Similarity 0,1 dan 0,5 dapat dilihat pada lampiran d. Pada Gambar 4, dapat dilihat bahwa *running time* terbaik diperoleh dengan *damping factor* 0,85 dengan nilai batasan SNN-Similarity sebesar 0,9 dibandingkan dua *damping factor* dan dua nilai batasan SNN-Similarity yang diujikan.

4.2 Analisis Hasil Pengujian

1. Pada gambar 2 menjelaskan bahwa *damping factor* 0,85 memperoleh *running time* yang baik. Ini disebabkan nilai *rank* yang dihasilkan *damping factor* 0,85 kecil sehingga proses perhitungan pada sistem menjadi mudah dibandingkan dua *damping factor* lainnya. Selain itu, banyaknya jumlah simpul dan relasi mempengaruhi *running time* dalam menentukan simpul terpenting karena, dalam algoritma terdapat perhitungan matriks nxn. Sehingga, jika simpulnya banyak akan memakan waktu dalam proses perhitungan. Dengan dataset yang menggunakan *graph* yang tidak berarah dan jumlah relasi lebih banyak dari pada jumlah simpul, Nilai *rank* dari *damping factor* 0,15 menghasilkan nilai *rank* yang tinggi

- dibandingkan dua *damping factor* yang lain. Hal ini disebabkan adanya perhitungan invers matriks dalam algoritma *pagerank*. Nilai *rank* dapat dilihat pada lampiran c.
2. Pada tabel 3 menjelaskan bahwa dari beberapa simpul terpenting memiliki nilai *modularity* yang baik. Ini disebabkan jika proses penggabungan yang berdasarkan simpul terpenting. Simpul terpenting akan dijadikan simpul pusat dalam proses penggabungan. Jika simpul terpenting tidak memiliki relasi yang banyak, maka nilai *modularity* akan kecil karena proses penggabungan tidak hanya berdasarkan simpul terpenting tetapi jumlah relasi juga mempengaruhi nilai *modularity* per komunitas. Sehingga, jika simpul pusat merupakan simpul terpenting dan memiliki banyak relasi maka nilai *modularity* yang akan dihasilkan baik.
 3. Pada gambar 3 menjelaskan bahwa nilai *modularity* terbaik diperoleh oleh *damping factor* 0,85. Ini disebabkan jumlah komunitas yang dihasilkan oleh 0,85 lebih banyak dibandingkan dua *damping factor* lainnya. Pada proses penggabungan, jika dua komunitas yang ingin digabungkan merupakan simpul terpenting maka akan dihitung nilai *Similarity* menggunakan *SNN Similarity*. Dalam perhitungan *Similarity* dibutuhkan nilai *rank*. Jika hasil nilai *rank* kecil maka nilai *Similarity* mempunyai potensi untuk kurang dari 0,5 sehingga komunitas tersebut tidak dapat digabung. Hal ini menyebabkan jumlah komunitas yang terbentuk menjadi banyak. Karena nilai *rank* yang diperoleh *damping factor* 0,85 memiliki nilai terendah dibandingkan dua *damping factor* yang lain, *modularity* terbaik diperoleh oleh *damping factor* 0,85. Sedangkan gambar 4 menjelaskan bahwa nilai *running time* 0,85 juga memperoleh hasil yang baik dengan nilai 0,522s dibandingkan dua *damping factor* yang lain. Ini juga dipengaruhi dari perhitungan *Similarity* yang memperoleh nilai *rank* karena nilai *rank* 0,15 memiliki *rank* terbesar maka perhitungan akan menjadi kompleks sehingga mempengaruhi *running time* pada sistem ini. Selain itu, dari tiga nilai batasan *SNN-Similarity* yang diujikan, nilai 0,9 memiliki nilai *modularity* yang tinggi dibandingkan dua nilai batasan *SNN-Similarity* yang diujikan karena nilai tersebut akan membentuk komunitas lebih banyak dari pada dua nilai batasan *SNN-Similarity* yang diujikan. Ini disebabkan jika simpul dapat digabungkan nilai relasi pada setiap simpul tidak akan diperbaharui dan menyebabkan hasil *running time* yang baik dari pada dua nilai batasan *SNN-Similarity* yang diujikan. Banyaknya data yang digunakan untuk mengetahui performansi algoritma ini berdasarkan *modularity* tidak mempengaruhi *modularity* yang dihasilkan karena ada parameter lain yang mempengaruhi algoritma ini yaitu *damping factor* dan juga nilai batasan *SNN-Similarity*.
 4. Pengaruh *running time* untuk mendeteksi komunitas juga dipengaruhi juga oleh banyaknya jumlah relasi. Semakin banyaknya jumlah relasi akan membutuhkan waktu yang banyak untuk mendeteksi suatu komunitas. Namun, *running time* untuk mendeteksi komunitas, beberapa hasil *running time* yang memiliki banyak data tidak memperoleh *running time* yang lama ini disebabkan perhitungan yang kompleks yang dipengaruhi oleh nilai *damping factor*. Dari pengujian yang telah ditemukan bahwa adanya batasan jumlah relasi pada sistem yang telah dibangun sebanyak 57500. Hal ini disebabkan sistem yang dibangun memiliki operasi perhitungan yang banyak sehingga menyebabkan *memory* pada sistem yang dibangun *full* dan terjadi *error*.

4. Kesimpulan

Setelah melakukan pengujian dan menganalisis hasil yang didapat maka dapat disimpulkan bahwa :

1. *Damping factor* mempengaruhi *running time* untuk menentukan simpul terpenting. *Damping factor* 0,85 memperoleh *running time* tercepat dengan hasil 187,001 detik. Ini disebabkan pada perhitungan nilai *rank*, 0,85 mempunyai perhitungan yang tidak terlalu kompleks dan menghasilkan nilai *rank* yang kecil dibandingkan dua *damping factor* yang lain. Jumlah simpul dan relasi juga mempengaruhi *running time* untuk menentukan simpul terpenting. Semakin banyaknya simpul dan juga relasi akan semakin lama proses dari algoritma *Pagerank*. Hasil dapat dilihat pada gambar 1 atau tabel yang telah dilampirkan pada lampiran b.
2. Nilai *modularity* per komunitas tidak hanya dipengaruhi oleh simpul terpenting saja. Namun, jumlah relasi juga mempengaruhi nilai *modularity* per komunitas. Sehingga, jika komunitas tersebut merupakan simpul terpenting dan mempunyai relasi yang banyak, maka akan menghasilkan nilai *modularity* per komunitas yang baik dibandingkan komunitas yang lain.
3. *Damping factor* dan nilai batasan *SNN-Similarity* mempengaruhi nilai *modularity* untuk mendeteksi komunitas dan juga mempengaruhi jumlah anggota yang terbentuk. Namun banyaknya data tidak mempengaruhi nilai *modularity* yang dihasilkan.
4. Pada hasil percobaan 3, Performansi algoritma *CNM-Centrality* dilihat dari nilai *modularity* dan kecepatan *running time* berdasarkan pengujian yang telah dilakukan mendapatkan hasil yang baik dengan nilai 0.838059 dengan batasan *SNN-Similarity* 0,9. Namun, sistem ini hanya mampu menangani 57500 simpul saja. Nilai *modularity* yang dihasilkan tidak dipengaruhi oleh jumlah data

yang diuji melainkan parameter yang terdapat pada algoritma ini yaitu nilai batasan SNN-*Similarity*, *Damping factor* dan juga jenis dari *graph* yang digunakan.

Adapun saran untuk mengembangkan penelitian selanjutnya mengenai community detection adalah :

1. Untuk menentukan simpul terpenting juga dapat menggunakan metode lain seperti *Degree Centrality*, *Betweenness Centrality* dsb.
2. Penggunaan graf dengan jenis yang berbeda seperti *directed-graph*.
3. Untuk menentukan kualitas dari komunitas yang terbentuk dapat juga menggunakan selain modularity seperti NMI (*Normalized mutual information*), *clustering coefficient*, *graph density* ataupun perhitungan lainnya

Daftar Pustaka

- [1] International Telecommunication Union, "ICT Facts Figures 2017," 06 Januari 2018. [Online]. Available: <https://www.key4biz.it/wp-content/uploads/2017/05/ICTFactsFigures2017.pdf>. [Accessed 05 Maret 2018].
- [2] L. Huan and T. Lie, "Graph Mining Application to Social Network Analysis," in *Managing and Mining Graph Data, Advances in Database Systems 40*, New York, Springer Science+Business Media, 2010.
- [3] Statista, "Number of monthly active Facebook users worldwide as of 4th quarter 2017," Statista, Januari 2018. [Online]. Available: <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>. [Accessed 06 Maret 2018].
- [4] I. A. Nur, *Community Detection menggunakan Genetic Algorithm dalam Social Network Twitter*, Bandung: Universitas Telkom, 2014.
- [5] P. Boldi, M. Santini and S. Vigna, *PageRank as a Function of the Damping Factor*, New York: Proceedings of the 14th international conference on World Wide Web, 2005.
- [6] H. Fang and L. Yuhua, *A new algorithm CNM-Centrality of detecting communities based on node centrality*, China, 2015.
- [7] M. Newman, "Modularity and community structure in networks," *The National Academy of Sciences of the USA*, New York, 2006.
- [8] R. F. Zinal and A. Djunaidy, "Algoritma Shared Nearest Neighbor berbasis Data Shrinking," Fakultas Teknologi Informasi, Institut Teknologi Sepuluh November, Surabaya, 2016.
- [9] J. McAuley and J. Leskovec, "Learning to Discover Social Circles in Ego Networks," *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems*, vol. 1, 2012.
- [10] M. Bressan and E. Peserico, "Choose the damping, choose the ranking?," vol. 8, no. 2, 2010.
- [11] K. Avrachenkov, N. Litvak and K. S. Pham, "A singular perturbation approach for choosing PageRank damping factor," *Cornel University Library*, California, 2006.
- [12] Statista, "Most famous social network sites worldwide as of September 2016, ranked by number of active users (in millions)," Statista, 02 September 2016. [Online]. Available: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>. [Accessed 11 November 2016].
- [13] A. Clauset, M. Newman and C. Moore, "Finding Community Structure in Very Large Network," *The American Physical Society*, Michigan, 2004.