

ANALISIS DAN IMPLEMENTASI ALGORITMA AGGLOMERATIVE HIERARCHICAL CLUSTERING UNTUK DETEKSI KOMUNITAS PADA MEDIA SOSIAL FACEBOOK

ANALYSIS AND IMPLEMENTATION OF AGGLOMERATIVE HIERARCHICAL CLUSTERING ALGORITHM FOR COMMUNITY DETECTION IN SOCIAL MEDIA FACEBOOK

Intan Widya Rahayu¹, Imelda Atastina, S.Si., M.T.², Anisa Herdiani, S.T., M.T.³

^{1,2,3}Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom

¹intaaaaan@student.telkomuniversity.ac.id, ²imelda@telkomunivesity.ac.id, ³anisaherdiani@gmail.com

Abstrak

Deteksi komunitas merupakan proses pencarian dan identifikasi komunitas dalam suatu jaringan, salah satunya pada jaringan sosial. Terdapat dua karakteristik data dalam mendeteksi komunitas yaitu hubungan antar individu dan kesamaan (*similarity*) antar individu. Algoritma deteksi komunitas biasanya hanya berfokus pada salah satu karakteristiknya saja. Pada penelitian ini penulis membahas pembentukan komunitas menggunakan algoritma *Agglomerative Hierarchical Clustering* (AHC) berbasis *enhanced similarity* pada data Facebook, yang mana merupakan algoritma yang secara berturut-turut memodelkan pengelompokan yang menggunakan nilai *threshold* sebagai alat pengelompokan, dua simpul yang memiliki nilai *similarity* lebih besar dibandingkan nilai *threshold* maka akan dijadikan satu komunitas dengan memperhitungkan hubungan yang dimiliki simpul. Dalam pengujiannya, penulis menggunakan *modularity* sebagai alat penentu kualitas *cluster*, yang mana kualitas *cluster* dikatakan baik jika memiliki hubungan yang pada di dalam satu komunitasnya dibandingkan hubungan di luar komunitasnya. Hasil dari Algoritma AHC mampu menunjukkan bahwa AHC dapat mendeteksi komunitas dengan kualitas *cluster* (*modularity*) mencapai 0.4453 pada jaringan nyata.

Kata kunci: *Agglomerative Hierarchical Clustering, Deteksi Komunitas, Jaringan Sosial, Similarity, Modularity*

Abstract

Community detection is the process of finding and identifying a community within a network, one of them on social networks. There are two characteristics of data in detecting the community that is the relationship between individuals and similarity between individuals. The community detection algorithm usually focuses on only one of its characteristics. In this research the author discusses the identify of the community using the algorithm Agglomerative Hierarchical Clustering (AHC) based on enhanced similarity on the Facebook data, which is an algorithm that sequentially modeling the grouping using the threshold value as a grouping tool, two nodes that have similarity value greater than threshold value then it will be made into a community by taking into account the relationship of the node. In the evaluation, the authors use modularity as a cluster quality measure, in which the quality of the cluster is said to be good if it has more solid relationship inside the community than the relationship outside the community. The results of Algoritma AHC are able to show that AHC can detect community with cluster quality (modularity) reach 0.4453 in real network.

Keywords: *Agglomertive Hierarchical Clustering, Community Detection, Social Networks, Similarity, Modularity*

1. Pendahuluan

Seiring perkembangan teknologi, media sosial telah berevolusi menjadi sumber informasi, media komunikasi, dan media promosi. Berdasarkan statistik tahun 2016, pengguna sosial media melampaui 2 miliar [1]. Hal ini dikarenakan manusia adalah makhluk sosial yang mana sekarang sulit untuk bertemu teman di dunia nyata tetapi lebih mudah mencari teman *online* dengan minat yang sama [2]. Media sosial terpopuler di seluruh dunia adalah Facebook dengan 1,37 miliar pengguna aktif harian dan 2,07 miliar pengguna aktif bulanan [3]. Profil pribadi pengguna Facebook dapat mencakup banyak informasi yang dapat dikelompokkan ke dalam jaringan sosial yang sama yang disebut sebagai komunitas [4] [5].

Komunitas adalah sekumpulan orang yang disatukan oleh hubungan dari dua orang atau lebih dengan kepentingan bersama seperti hobi, minat, tujuan, proyek, gaya hidup, lokasi maupun profesi di dalam suatu jaringan sosial [6]. Terdapat dua jenis komunitas yang terbentuk dalam media sosial yaitu *explicit* dan *implicit community* [2]. *Explicit community* adalah komunitas yang memang sudah disediakan oleh media sosial, sedangkan *implicit* adalah komunitas yang terbentuk berdasarkan interaksi sosial yang terjadi. Proses pembentukan dan pengidentifikasian komunitas dalam jaringan sosial disebut dengan *community detection* [7]. Penerapan deteksi komunitas dapat bermanfaat sebagai informasi dasar untuk strategi pemasaran, sistem rekomendasi, system pemodelan dunia nyata dan lain sebagainya [2] [8].

Pengelompokan simpul ke dalam komunitas dipandang sebagai masalah pada identifikasi komunitas. Simpul dapat memiliki beberapa komunitas karena simpul berbagi atribut yang sama dan memiliki hubungan diantara simpul lainnya. Sehingga ada dua karakteristik data yang bisa digunakan dalam deteksi komunitas, yaitu data tentang objek (simpul dan atributnya) dan hubungan antar objek (relasi) [9].

Beberapa algoritma pernah digunakan dalam mendeteksi komunitas sebelumnya, namun dalam pengelompokan biasanya hanya fokus pada salah satu karakteristiknya [10]. Dengan hanya mempertimbangkan salah satu karakteristiknya saja memungkinkan algoritma gagal dalam memperhitungkan struktur data dan pengelompokan komunitas. Oleh karena itu untuk mempertimbangkan kedua karakteristik data dalam pembentukan komunitas yang baik digunakan algoritma *Agglomerative Hierarchical Clustering* (AHC) berbasis *enhanced similarity*.

AHC merupakan algoritma pengelompokan yang menggunakan nilai *threshold* sebagai alat pengelompokan, yang mana dua individu yang memiliki nilai kemiripan lebih besar dibandingkan nilai *threshold* maka akan dijadikan satu komunitas dan sebaliknya. Hasil akhir dari penelitian ini adalah mengetahui performansi algoritma AHC dalam pembentukan komunitas dengan menggunakan perhitungan *modularity*.

2. Dasar Teori

2.1. Clustering

Clustering merupakan proses mengelompokkan objek data menjadi beberapa kelompok berdasarkan ukuran kemiripan informasi yang terdapat pada data yang menggambarkan objek dan hubungannya. Tujuannya adalah mengetahui bahwa objek ada di dalam suatu kelompok dengan objek lain yang serupa dan berbeda dari objek di dalam kelompok lain. Semakin besar kesamaan objek dalam satu kelompok, semakin besar perbedaan antar kelompok [11]. Kelompok yang dihasilkan disebut *cluster* [12]. Untuk lebih memahami, gambar 2-1 menampilkan data sebelum proses *clustering* dengan data yang sudah melalui proses *clustering*.

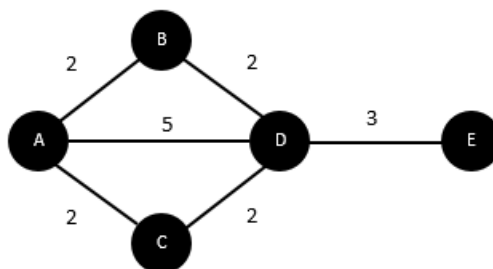
2.1.1. Graph Clustering

Graph adalah struktur yang dibentuk oleh suatu set simpul (*node*) dan satu set sisi (*edge*) yang merupakan hubungan antara pasangan simpul. *Graph clustering* adalah proses mengelompokkan graf ke dalam kelompok dengan mempertimbangkan struktur sisi sehingga harus ada banyak sisi dalam setiap *cluster* dan relatif sedikit diantara kelompok [12]. Gambar 2-2 merupakan contoh *graph clustering*.

2.2. Jaringan Sosial

Jaringan sosial adalah kumpulan objek (individu, organisasi atau entitas lain) yang dihubungkan oleh serangkaian hubungan [13]. Semakin berkembangnya internet dan tersedianya web 2.0 membuat jaringan sosial, situs media sosial menghubungkan beberapa orang di dunia maya. Pertukaran pesan elektronik (*e-mail*), *chat room*, berbagi momen merupakan contoh aktifitas jaringan sosial pada dunia maya.

Jaringan sosial biasanya diilustrasikan sebagai graf ($G=V,E,W$), yang mana *vertex* (V) menunjukkan individu, *edge* (E) menunjukkan hubungan antar individu, dan *weight* (W) menunjukkan bobot hubungan antar individu. Merujuk pada gambar 2-3. dapat dilihat bahwa *vertex* A,B,C,D,E merupakan individu yang dihubungkan oleh sisi dengan bobot masing-masing.



Gambar 1. Contoh Jaringan Sosial

Untuk mempermudah pemrosesan dan pengolahan data, graf digambarkan dalam bentuk matriks berdimensi $n \times n$ [7], yang mana n menunjukkan jumlah individu dalam jaringan sosial. individu yang saling berhubungan akan diberi bobot, sedangkan yang tidak saling berhubungan akan diberi bobot 0.

2.3. Community Detection

Community detection atau deteksi komunitas merupakan salah satu *task* dari *social network analysis* (SNA). Kata *community detection* sendiri terdiri dari dua kata yaitu deteksi dan komunitas. Komunitas memiliki arti sekumpulan orang yang disatukan oleh kepentingan bersama seperti hobi, minat, tujuan, proyek, gaya hidup, lokasi maupun profesi yang sama [6] dan deteksi yang berarti usaha, proses menemukan dan menentukan keberadaan sesuatu. Dengan kata lain pengertian deteksi komunitas adalah usaha serta proses menemukan dan menemukan sekumpulan orang yang memiliki kepentingan bersama.

2.4. Facebook

Facebook merupakan salah satu situs jaringan sosial terbesar yang didirikan pada tahun 2004 oleh Mark Zuckerberg. Dalam misinya yaitu "Give people the power to build community and bring the world closer together." membuat Facebook memiliki layanan yang mana penggunaanya tetap terhubung dengan teman dan keluarga [3]. dengan dapat mengirim komentar, berbagi foto, dan berbagi aktifitas lainnya yang dapat diakses secara publik atau hanya dapat dibagikan diantara sekelompok orang atau hanya dengan satu orang saja [14]. Selain itu Facebook juga memiliki layanan yang mana penggunaanya dapat mendapatkan informasi maupun berita yang sedang terjadi di dunia. Meskipun mulai banyak bermunculan situs jaringan social lainnya, Facebook tetap memiliki 1,37 miliar pengguna aktif harian dan 2,07 miliar pengguna aktif bulanan (terhitung bulan September 2017) [3].

2.5. Algoritma Agglomerative Hierarchical Clustering (AHC)

Hierarchical clustering adalah metode pengelompokan yang mana hasil pengelompokan direpresentasikan dalam bentuk pohon atau hirarki yang biasa disebut dendrogram. algoritma *hierarchical clustering* dibagi menjadi dua kelas, yaitu *divisive (top-down)* dan *agglomerative (bottom-up)*.

Algoritma *agglomerative hierarchical clustering* (AHC) adalah algoritma *clustering* dengan pendekatan *bottom-up* yang mana membangun hirarki dimulai dari masing-masing simpul sebagai *singleton cluster* dan secara berturut-turut menggabungkan simpul menjadi *cluster* yang lebih besar sampai hanya terdapat satu *cluster* yang tersisa [15] [13]. Untuk menggabungkan *simpul* menjadi satu *cluster*, biasanya AHC menghitung *similarity* sebagai ukuran jarak antar *simpul*, dengan menggunakan tiga jenis teknik pemilihan jarak yang sering digunakan yaitu *Single-linkage* (jarak terdekat), *complete-linkage* (jarak terjauh), *average-linkage* (jarak rata-rata) [13].

2.5.1. Similarity Measure

Matriks *similarity* adalah pengukuran yang digunakan untuk mengukur kemiripan antara dua objek data yang memiliki satu atau beberapa atribut. Biasanya nilai *similarity* tidak negatif dan biasanya pula nilai *similarity* berada pada antara nilai 0 dan 1, yang mana 0 berarti tidak memiliki kemiripan sama sekali dan 1 berarti data dianggap sangat mirip.

Pengukuran cosine adalah pengukuran kemiripan antar dua vektor non-biner. Biasanya digunakan dalam data dokumen yang memiliki atribut pada masing-masing data. Kesamaan struktural antara dua individu pada kasus deteksi komunitas berdasarkan atribut juga dapat diukur menggunakan pengukuran *cosine*. Semakin besar nilai cosine maka semakin mirip dua data. Perhitungan dirumuskan sebagai berikut :

$$\cos \theta = \frac{x \cdot y}{\|x\| \|y\|} \quad (1)$$

atau

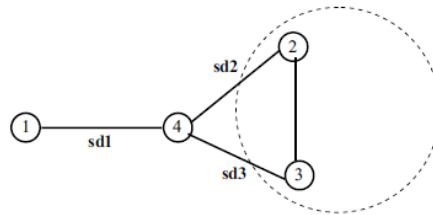
$$\cos \theta = \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}} \quad (2)$$

Keterangan :

$\cos \theta$ = jarak antara data x dan data y
 x = jumlah dari nilai bobot *sisi* yang bertetangga dengan simpul x
 y = jumlah dari nilai bobot *sisi* yang bertetangga dengan simpul y

2.4.2. Enhanced Similarity

Menggabungkan simpul pada Algoritma *Agglomerative Hierarchical Clustering* sebenarnya hanya memilih salah satu penentu *similarity* (kesamaan) antar simpul [16]. Namun dalam beberapa kasus, hanya menggunakan penentu jarak *similarity* antar simpul mungkin tidak masuk akal



Gambar 2. Contoh kasus *enhanced similarity*

Sebagai contoh, pada gambar 2-3, terdapat 4 simpul, yang mana simpul 1 terhubung dengan simpul 4, dan 4 terhubung dengan simpul 2 dan 3. Simpul 2 dan 3 menunjukkan satu komunitas, *Similarity* antar simpul menunjukkan bahwa $sd1 > sd2$ dan $sd1 > sd3$. Jika dilihat dari gambar 2-3, memungkinkan simpul 1 dan 4 akan menjadi satu *cluster* dilangkah selanjutnya karena simpul 4 dianggap terhubung komunitas 2 dan 3. Padahal secara nyata, dua individu yang memiliki kemiripan lebih besar akan memiliki interaksi lebih besar pula. *Enhanced similarity* yang dirumuskan sebagai berikut :

$$I_{AB} = (1 - \prod_{i=1}^k (1 - I_i)) \sqrt{\frac{k}{m \times n}} \quad (3)$$

Keterangan :

- A, B = komunitas dengan anggota $\{a_1, a_2, \dots, a_m\}$ dan $\{b_1, b_2, \dots, b_n\}$
- m, n = jumlah anggota komunitas A dan B.
- k = sisi antara dua simpul
- I_i, I_k = *similarity* antar dua simpul

2.6. Modularity

Modularity merupakan perhitungan untuk mengukur kualitas pengelompokan jaringan. Pengelompokan akan dikatakan optimal jika nilai modularitasnya mencapai maksimal [17]. Nilai modularitas berada pada kisaran -1 sampai 1, yang mana nilai modularitas yang mendekati 1 menunjukkan kualitas yang baik. *Modularity* dilakukan dengan membandingkan jumlah sisi yang berada didalam *cluster* dan antar *cluster* [18]. Perhitungan modularitas dirumuskan dalam:

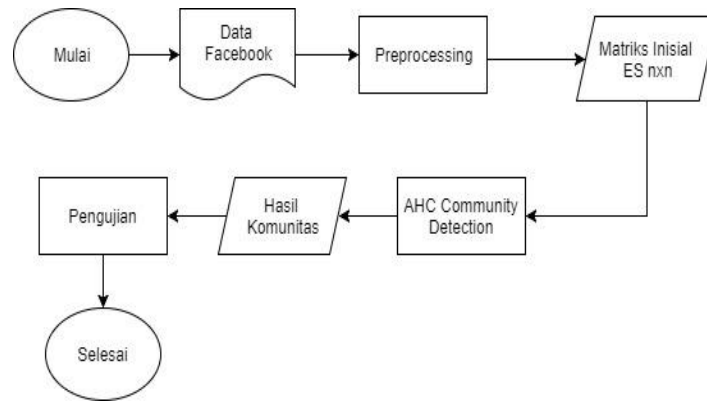
$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \quad (4)$$

Keterangan:

- m = total *sisi* pada *cluster*
- A_{ij} = nilai relasi antar dua simpul i dan j dengan nilai 1 jika memiliki hubungan dan 0 jika tidak memiliki hubungan
- k_i, k_j = jumlah derajat yang dimiliki oleh simpul ke- i atau ke- j
- c_i, c_j = komunitas yang di dalamnya terdapat simpul i dan simpul j
- $\delta(c_i, c_j)$ = fungsi delta yang mana akan bernilai 1 jika c_i, c_j dalam satu komunitas dan akan bernilai 0 jika sebaliknya

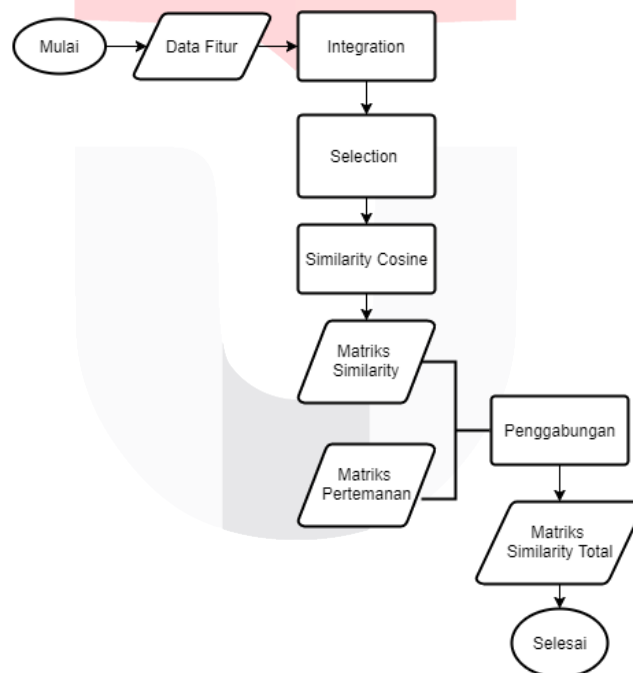
3. Perancangan Sistem

Sistem yang dibangun pada penelitian tugas akhir ini adalah sistem dengan hasil keluaran berupa pengelompokan komunitas (kumpulan individu) pada dataset social media *Facebook*. Perancangan sistem digambarkan pada gambar 3.



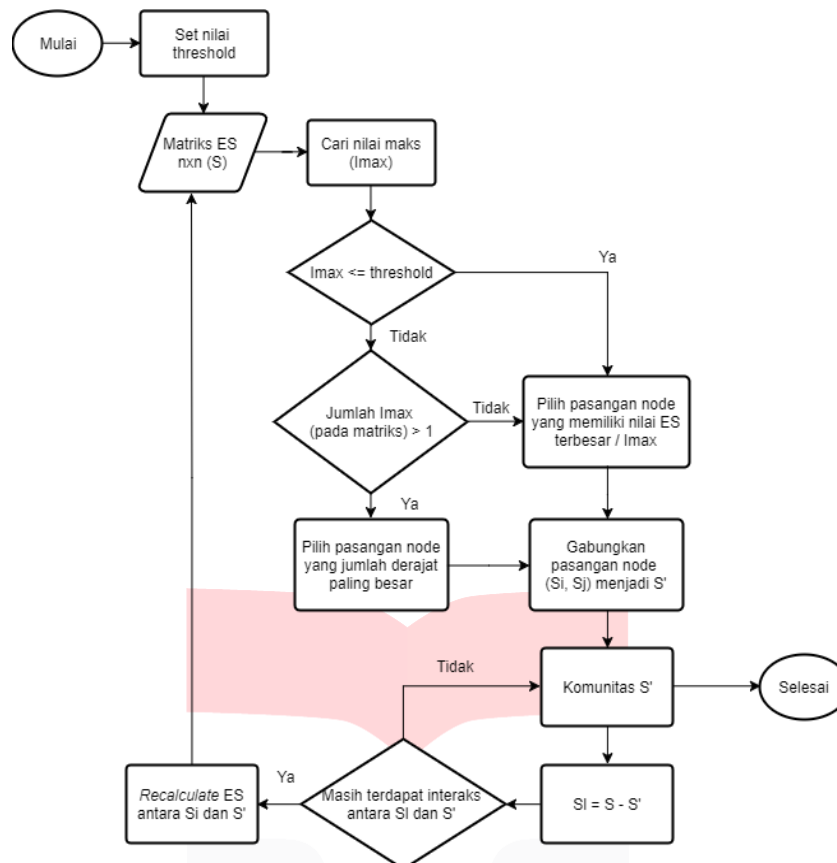
Gambar 3 : Gambaran umum sistem

Merujuk gambar 3. Dataset Facebook akan menjadi masukan dalam program, yang mana data di-*preprocessing* terlebih dahulu. *Preprocessing* ini bertujuan mendapatkan matriks inisial ES nxn . Setelah data sudah menjadi matriks inisial ES nxn , data siap untuk di kelompokkan menggunakan algoritma AHC. Setelah didapatkan pengelompokan komunitas, dihitung pengujian dari hasil sistem dan dianalisis. Tahap awal yang dilakukan yaitu pengambilan dataset Data yang digunakan dalam penelitian diambil dari situs <https://snap.stanford.edu/data/egonets-Facebook.html> dengan data yang digunakan adalah relasi pertemanan dan atribut profil pengguna. Dengan dua jenis data yang masih sulit untuk dipahami sistem, kedua data diatas diolah terlebih dahulu. Gambar 4 menunjukkan tahapan-tahapan dalam *preprocessing* :



Gambar 4 : Flowchart Preprocessing

Setelah melewati *preprocessing* data, selanjutnya hasil *preprocessing* yang berupa matriks nxn akan diproses dalam *Agglomerative Hierarchical Clustering* untuk dilakukan pengelompokan data. Gambar 5 menunjukkan cara kerja algoritma AHC.



Gambar 5. Flowchart Algoritma AHC

Setelah mendapatkan bobot *similarity* (nilai inisial *enhanced similarity*), selanjutnya ditetapkan nilai *threshold* untuk pembentukan komunitas. Di setiap iterasi dilakukan pemilihan nilai *similarity* maksimal dan pengecekan jumlah nilai *similarity* maksimal. Setelah didapatkan nilai *similarity* maksimal, dua simpul atau lebih dengan nilai *similarity* terbesar akan digabungkan dalam satu komunitas dan dilakukan perhitungan ulang nilai *enhanced similarity* (ES) antara simpul yang sudah dikelompokkan dengan simpul yang tersisa dengan menggunakan rumus pada persamaan (3) hingga nilai *similarity* maksimal kurang dari nilai *threshold*.

4. Pengujian

Pengujian yang dilakukan pada penelitian ini adalah untuk mengetahui performansi algoritma AHC dalam pembentukan komunitas berdasarkan beberapa parameter yaitu ukuran data dan perubahan nilai *threshold* menggunakan perhitungan *modularity*. Adapun skenario pengujian dalam penelitian ini:

1. Pengujian dilakukan dengan 5 kali percobaan pada nilai *threshold* yang sama untuk 3 ukuran data simpul dengan masing-masing data memiliki jumlah relasi yang berbeda.
2. Pengujian dilakukan secara 3 kali dengan nilai *threshold* yang berbeda yang mana diambil dari rentang perubahan nilai *threshold* pada skenario pengujian 1 yang sekiranya dapat menghasilkan komunitas dengan nilai *modularity* terbaik.

4.1. Analisis Hasil Pengujian Skenario 1

Pengujian skenario 1 merupakan skenario pengujian yang dilakukan dengan 5 kali percobaan pada nilai *threshold* yang sama untuk 3 ukuran data simpul yang berbeda dengan masing-masing data memiliki 2 jumlah relasi yang berbeda. Pengujian dilakukan untuk mengetahui pengaruh jumlah *node*, jumlah sisi (kepadatan hubungan) dan nilai *threshold* terhadap komunitas yang terbentuk dari data pengguna facebook. Hasil pengujian dijelaskan pada tabel 1.

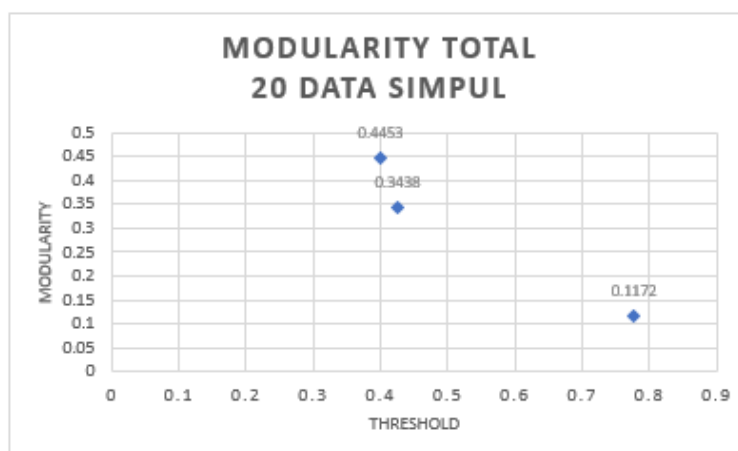
Tabel 1. Tabel Hasil Pengujian Skenario 1.

Threshold	Hasil Implementasi Algoritma AHC					
	50 node		100 node		1000 node	
	Jumlah Sisi	Jumlah Komunitas	Jumlah Sisi	Jumlah Komunitas	Jumlah Sisi	Jumlah Komunitas
Percobaan 1						
0.1	8	16	354	37	18960	151
	99	1	2755	1	998039	10
Percobaan 2						
0.3	8	16	354	47	18960	242
	99	2	2755	1	998039	10
Percobaan 3						
0.5	8	18	354	57	18960	321
	99	3	2755	3	998039	10
Percobaan 4						
0.7	8	18	354	68	18960	436
	99	7	2755	15	998039	11
Percobaan 5						
0.9	8	20	354	95	18960	631
	99	12	2755	50	998039	14

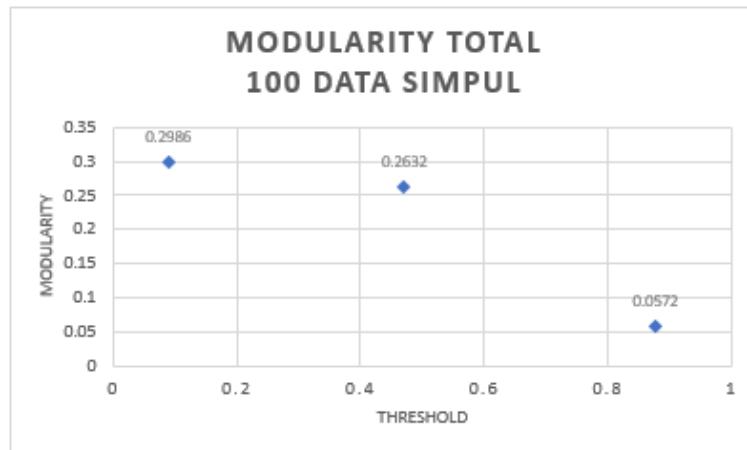
Hasil tersebut menunjukkan bahwa semua simpul yang nilai *similarity*-nya lebih dari nilai *threshold* akan dikelompokkan menjadi satu komunitas dan kepadatan sisi dalam pembentukan komunitas akan menghasilkan komunitas yang lebih padat keanggotaannya sehingga menghasilkan jumlah komunitas yang lebih sedikit dibandingkan data dengan relasi yang renggang.

4.2 Analisis Hasil Pengujian Skenario 2

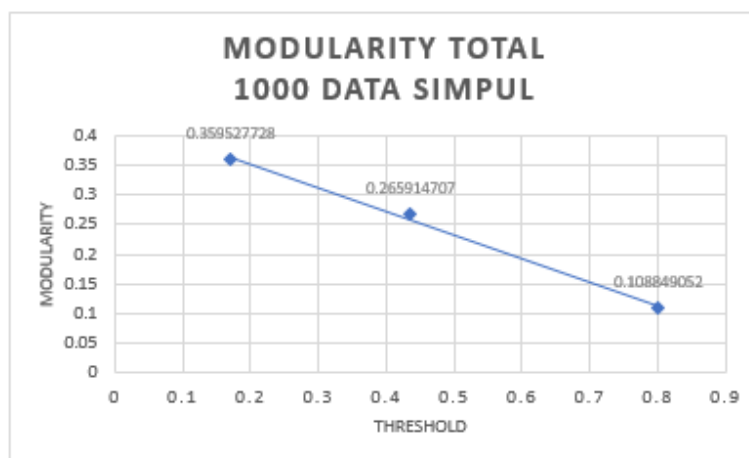
Pengujian skenario 2 merupakan skenario pengujian yang dilakukan pada 3 ukuran data simpul yaitu 20, 100 dan 1000 dengan 3 ukuran data relasi yaitu 8, 354 dan 2775 yang mana menunjukkan kepadatan hubungan yang relatif renggang yang diuji menggunakan 3 nilai *threshold* yang berbeda. Pengujian dilakukan untuk mengetahui pengaruh perubahan nilai *threshold* terhadap waktu proses berjalannya algoritma AHC dan kualitas *cluster*. Hasil Pengujian digambarkan pada gambar 6, gambar 7 dan gambar 8.



Gambar 6. Grafik modularity total dengan 20 data simpul



Gambar 7. Grafik modularity total dengan 100 data simpul



Gambar 8. Grafik modularity total dengan 100 data simpul

Hasil di atas ini menunjukkan bahwa kualitas *cluster* tidak hanya dipengaruhi oleh nilai *threshold* namun juga relasi yang terhubung antar anggota di luar dan di dalam komunitas. Hal ini terbukti jika terdapat komunitas yang memiliki satu anggota, maka kualitas *cluster* sama dengan 0, dengan kata lain tidak adanya relasi antar simpul lain didalam komunitasnya maupun diluar komunitasnya.

5. Kesimpulan

Berdasarkan pengujian dan analisis pada bab 4, maka dapat diambil kesimpulan sebagai berikut:

- Nilai *threshold* terbukti berpengaruh dalam pembentukan komunitas dengan memperhitungkan hubungan yang dimiliki oleh simpul. Semakin besar nilai *threshold* akan semakin sedikit simpul yang dapat dikelompokkan, sedangkan semakin kecil nilai *threshold* maka hampir semua simpul akan dikelompokkan dalam satu komunitas.
- Ukuran data tidak mempengaruhi proses pembentukan komunitas. Sedangkan jumlah sisi mempengaruhi pembentukan komunitas, hal ini dibuktikan dengan semakin sedikit jumlah sisi, maka semakin banyak jumlah komunitas. Dengan kata lain, simpul yang tidak memiliki hubungan maka akan dianggap sebagai komunitas berjumlah 1 (satu) anggota.
- Kualitas *cluster* dipengaruhi oleh :
 - Threshold*, semakin padat hasil komunitas yang dibentuk oleh *threshold* pada algoritma AHC maka semakin baik kualitas *cluster*.
 - Jumlah sisi, jumlah sisi antar anggota di dalam satu komunitas harus lebih banyak dibanding dengan relasi antar anggota di luar komunitasnya agar menghasilkan kualitas *cluster* yang baik.
- Running time* tidak hanya dipengaruhi oleh jumlah simpul, namun juga proses pengelompokan berdasarkan nilai *threshold*. Semakin banyak simpul yang mirip, maka pengelompokan akan semakin lama *running time* dan sebaliknya.
- Kualitas *cluster* dikatakan baik jika nilai *modularity* mendekati 1, yang artinya jumlah sisi dalam komunitas lebih padat. Nilai *modularity* sama dengan 0 berarti simpul tidak memiliki hubungan dengan simpul lainnya. Sedangkan kualitas *cluster* dikatakan tidak baik jika nilai *modularity* mendekati -1.

Daftar Pustaka

- [1] Statista, "Social Media Statistics & Facts," Statista, [Online]. Available: <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/>. [Diakses 20 12 2017].
- [2] H. L. Lei Tang, "Community Detection and Evaluation," dalam *Community Detection and Mining in Social Media*, Morgan & Claypool, 2010.
- [3] Facebook, "Newsroom," Facebook, September 2017. [Online]. Available: <https://newsroom.fb.com/company-info/>. [Diakses 2017].
- [4] M. H. Andreas M. Kaplan, "The Challenges and Opportunities of Social Media," *Elsevier*, 2010.
- [5] J. L. Julian McAuley, "Social circles: Facebook," 2012. [Online]. Available: <http://snap.stanford.edu/data/>. [Diakses 2017].
- [6] M. Wu, "Community vs Social Network," Lithium Community, 06 June 2010. [Online]. Available: <https://community.lithium.com/t5/Science-of-Social-Blog/Community-vs-Social-Network/ba-p/5283>. [Diakses September 2017].
- [7] H. L. Lei Tang, "Graph Mining Applications to Social Network Analysis," *Spring Science+Business Media*, 2010.
- [8] J. S. Mini Singh Ahuja, "Future Prospect in Community Detection," *International Journal of Computer Science Engineering and Information Technology Research*, no. vol 4, Issues 5, 2014.
- [9] J. M. J. L. Jaewin Yang, "Community Detection in Networks with Node Attributes," *arXiv*, 2014.
- [10] L. Z. D. W. Zhang Shoaqian, "An Enhanced Community Detection Method Based on Neighborhood Similarity," *IEEE*, 2012.
- [11] M. S. V. K. Pang-Ning Tan, "Cluster Analysis: Basic Concept and Algorithm," dalam *Introduction to Data Mining*, 2015.
- [12] S. E. Schaeffer, "Graph Clustering," *Elsevier*, 2007.
- [13] L. Ma, "Social Network Analysis Using a Multi-agent System: A School System Case," Trinity University, 2013.
- [14] D. Nations, "What is Facebook," 6 Juni 2017. [Online]. Available: <https://www.lifewire.com/what-is-facebook-3486391>. [Diakses 7 Agustus 2017].
- [15] R. C. Donna Beers, Community Detection with Hierarchical, Command, Control, and Interoperability Center for Advanced Data Analysis.
- [16] L. L. L. Z. C. B. Bing Kong, "Hierarchical Agglomerative Algorithm of Community Detecting in Social Network Based on Enhanced Similarity," *IEEE*, 2014.
- [17] M. M. X. X. T. A. S. Nurcan Yuruk, "AHSCAN : Agglomerative Hierarchical Structural Clustering Algorithm for Networks," *IEEE*, 2009.
- [18] R. G. W. N. A. M. B. R. A. K. M. Yousra Asim, "Community Detection in Networks using Node Attributes and Modularity," *International Jurnal of Advanced Computer Science and Applications*, 2017.