

Klasterisasi Pahlawan Nasional Indonesia Berdasarkan Daerah Asal Menggunakan Algoritma *Community Detection*

Luh Putu Gayatri Widiastuti*, Nida Inayah Maghfirani, Nur Aini Rakhmawati

Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas,
Institut Teknologi Sepuluh Nopember,

Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota SBY, Jawa Timur 60111

*e-mail: gayatri.putri@gmail.com

(received: 15 September 2020, revised: 4 November 2020, accepted: 3 Januari 2021)

Abstrak

Kisah sejarah sampai kemerdekaan Indonesia tak pernah lepas dari kurikulum pendidikan wajib belajar 12 tahun. Namun seiring dengan perkembangan zaman, hipotesis-hipotesis sejarah Indonesia juga ikut berkembang. hipotesis-hipotesis juga membantu sejarawan untuk melihat sejarah Indonesia menjadi lebih utuh. Untuk mendukung perkembangan tersebut, penulis mencoba mengimplementasikan algoritma *Community Detection* untuk mencari keterkaitan antar pahlawan nasional. Sumber data pahlawan nasional yang digunakan pada penelitian ini adalah *id.dbpedia*. Namun akibat inkonsistensi dan sedikitnya data yang disimpan di *id.dbpedia*, maka pengumpulan data tidak berhenti di *id.dbpedia* saja. Salah satu sumber data yang jumlah datanya lebih banyak dari *id.dbpedia* adalah Wikipedia. Data yang diperoleh kemudian dilakukan standarisasi secara manual, dan dibuatkan graf di Jupyter. Algoritma *Community Detection* digunakan untuk menemukan kluster-kluster komunitas pahlawan nasional Indonesia berdasarkan data daerah dan tahun. *Community Detection* menunjukkan bahwa data yang diolah dengan menggunakan Algoritma Greedy Modularity memiliki 16 komunitas atau kelompok. Selain itu, sebuah pola ditemukan pada 6 komunitas yang ada berdasarkan Algoritma Girvan Newman Sehingga dapat disimpulkan bahwa Algoritma Greedy Modularity menghasilkan lebih banyak klasifikasi komunitas.

Kata Kunci: algoritma, klasterisasi, deteksi komunitas, pahlawan nasional indonesia, daerah, tahun

Abstract

The story of history Indonesia until Indonesia's independence has never been separated from the 12-year compulsory education curriculum. However, along with the times, historical hypotheses in Indonesia also developed. Hypotheses also help historians to see Indonesian history more fully. To support this development, the author tries to implement the Community Detection algorithm to find linkages between national heroes. The data source for the national heroes used in this study is id.dbpedia. However, due to inconsistencies and the lack of data stored on id.dbpedia, data collection does not stop at id.dbpedia only. One of the data sources with more data than id.dbpedia is Wikipedia. After the data was obtained, then it standardized manually, and graphs were made in Jupyter. The Community Detection algorithm is used to find clusters of Indonesian national hero communities based on regional and year data Community Detection shows that the data processed using the Greedy Modularity Algorithm generates 16 communities or groups. In addition, a pattern was found in 6 existing communities based on the Newman Girvan Algorithm. So it can be concluded that the Greedy Modularity Algorithm produces more community classifications.

Keywords: algorithm, clustering, community detection, indonesian national heroes, region, years

1 Pendahuluan

Indonesia telah menyimpan banyak sejarah bahkan sebelum dapat memerdekakan negaranya. Dalam memperjuangkan kemerdekaan Indonesia, ada banyak pihak yang ikut turun tangan membantu perlawanan penjajah sampai ketika teks proklamasi dikumandangkan. Pihak-pihak itu kini disebut dengan Pahlawan Nasional Indonesia, yang namanya tercatat dan diceritakan dari satu generasi ke generasi sebagai apresiasi dari perjuangan tumpah darah mereka. Terlepas dari sejarah dan jumlah para pahlawan yang diakui, Indonesia merupakan negara yang luas berbentuk kepulauan. Pahlawan

<http://sistemasi.ftik.unisi.ac.id>

nasional juga seorang manusia yang pasti memiliki daerah dan tahun kelahiran, juga daerah dan tahun wafatnya. Selain itu, ada atribut lain unik yang hanya dimiliki oleh seseorang yang telah dinobatkan menjadi pahlawan nasional di Indonesia, yaitu tahun ditetapkannya beliau sebagai Pahlawan Nasional Indonesia oleh Pemerintah Republik Indonesia. Untuk mendukung perkembangan hipotesis sejarah pahlawan Indonesia, implementasi algoritma Community Detection dapat dilakukan untuk mencari keterkaitan antar pahlawan, relasi seperti apa yang ada di antara populasi ini dan daerah manakah yang cukup populer sebagai “daerah para pahlawan” dan mana yang tidak.

Algoritma Community Detection adalah sebuah bidang penelitian yang menarik dalam Complex Networks. Algoritma ini bekerja dengan cara mengelompokkan *nodes* atau *vectors* menjadi kelompok-kelompok komunitas [1]. Algoritma inilah yang dipilih untuk mengklasterisasi data pahlawan nasional Indonesia. Hal ini dikarenakan kemampuan algoritma Community Detection dikenal mampu mengelompokkan data. Sehingga data pahlawan nasional yang telah diperoleh dapat dikelompokkan menjadi komunitas-komunitas berdasarkan daerah dan tahun.

Algoritma Community Detection telah banyak digunakan untuk mendeteksi jaringan sosial [2]. Biasanya, untuk mendeteksi jaringan sosial, peneliti menggunakan data dari Twitter [3] [4]. Dalam kasus demikian, umumnya yang dicari adalah kedekatan antar akun pengguna. Tolak ukurnya, bisa jadi dari relasi *follow*, *reply*, *retweet*, *direct message*, atau dari kesamaan akun yang diikuti. Kegunaan Community Detection tidak hanya diperuntukkan jaringan sosial saja, terbukti pada penelitian sebelumnya [5] yang memanfaatkan algoritma ini untuk merancang DMA (*District Metered Area*).

Tujuan dari penelitian ini adalah untuk melihat data persebaran pahlawan di setiap daerah untuk melihat pembentukan grup berdasarkan daerah masing-masing pahlawan. Dengan metode *clustering*, data yang diperoleh akan dikelompokkan berdasarkan interaksi/relasi yang sama yaitu kota asal dan/atau kota wafat yang sama [6].

2 Tinjauan Literatur

Pada bab ini akan dijelaskan terkait teknologi dan algoritma yang digunakan dalam penelitian ini yaitu SPARQL, Networkx, algoritma *community detection*, dan *link prediction*.

2.1 SPARQL

SPARQL yang merupakan akronim dari Simple Protocol and RDF Query Language, adalah rangkaian spesifikasi yang menyediakan bahasa dan protokol untuk menanyakan dan memanipulasi konten grafik RDF di *web* atau di RDF Store [7]. Melalui SPARQL, nama-nama pahlawan itu dikumpulkan menjadi sebuah dataset yang dapat terhubung antar satu sama lain di *web*. Melalui ini, dataset yang berisi nama pahlawan beserta tempat lahir yang menjadi asal daerah pahlawan didapatkan dengan mudah.

2.2 NetworkX

Penelitian ini akan berfokus ke analisis jaringan sosial. Maka dari itu, dibutuhkan *tools* untuk membantu menganalisis. NetworkX adalah sebuah *tools* analisis jaringan sosial yang tersedia gratis dan dapat mengatasi data berukuran besar. NetworkX adalah sebuah *library* yang bekerja dengan bahasa Python. NetworkX mampu menampung hingga 1 juta *nodes* dengan kecepatan komputasi yang cepat [8]

Untuk mulai menggunakan NetworkX, pertama-tama kita perlu memiliki Python Language System dan *package* NetworkX dari Python. Untuk melakukan impor adalah dengan baris kode perintah

```
>>> import networkx as nx[9]
```

Dalam penelitian ini, NetworkX digunakan untuk membuat *graph* dan memvisualisasikan *nodes* yang populer dengan warna oranye. Ini dilakukan dengan menggunakan banyak baris kode perintah, termasuk memanfaatkan *library* lain yaitu Pandas, Matplotlib, Seaborn.

Penggunaan *library* NetworkX juga untuk menghitung *network density*, yang mana sangat penting dalam penelitian ini untuk mengklasterisasi data menjadi komunitas-komunitas. *Network density* atau yang dapat diartikan sebagai kepadatan jaringan, adalah sebuah ukuran yang menyatakan seberapa padatnya area tertentu oleh *nodes/vertex* pada suatu jaringan.

Perhitungan *network density* dianggap sebagai pengukuran agregat, yaitu mengukur suatu data melalui *mean* atau rata-rata dari atribut yang dimiliki oleh data [10]. Berikut adalah baris kode perintah yang digunakan untuk menghitung *density*, dengan *g* adalah *graph*

```
>>> density = nx.density(g)
```

2.3 Algoritma Community

Algoritma yang digunakan untuk mendeteksi komunitas berdasarkan interaksi yang sama. Dalam penelitian ini, klusterisasi data pahlawan nasional dengan dua algoritma Community Detection, yaitu dengan menggunakan algoritma Girvan Newman dan algoritma Greedy Modularity. Dua algoritma tersebut adalah algoritma yang digunakan untuk mendeteksi komunitas berdasarkan interaksi yang sama [11].

2.3.1 Algoritma Girvan Newman

Algoritma Girvan Newman adalah sebuah algoritma klusterisasi *divisive hierarchical* yang sudah dikenal dan digunakan luas untuk Community Detection. Algoritma ini bekerja dengan konsep *edge betweenness* [12]. Algoritma Girvan Newman memiliki keterbatasan dalam menangani jumlah data yang besar. Proses uji algoritma ini dimulai dengan langkah berikut [13]:

1. Hitung *edge betweenness* untuk setiap *edge* dalam grafik.
2. Hilangkan *edge* yang merupakan *edge betweenness* tertinggi
3. Hitung selisih *edge* yang tersisa
4. Ulangi langkah 2-4 sampai semua *edge* dihilangkan

Setelah itu, hasil dari uji algoritma ini nantinya akan divisualisasikan dengan *library NetworkX*.

2.3.2 Algoritma Greedy Modularity

Secara definisi, modularitas adalah karakteristik numerik dari sebuah jaringan, tingkat akurasi dari dibuatnya suatu partisi [14]. Algoritma ini diajukan oleh Newman juga, sebagai metode lain klusterisasi hierarki aglomeratif [15].

Proses uji algoritma ini dimulai dengan langkah berikut [16]:

1. Hitung *betweenness* untuk semua *edge* dalam jaringan
2. Hilangkan ujungnya dengan jarak tertinggi
3. Hitung ulang antara semua tepi yang terkena dampak dan hilangkan
4. Ulangi langkah-langkah 2 dan 3 sampai tidak ada tepi yang tersisa dalam jaringan

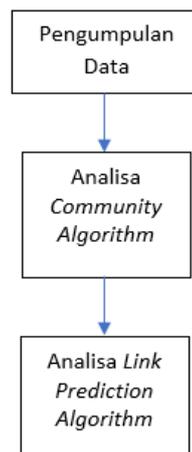
2.4 Link Prediction

Kerangka kerja paling mudah dari algoritma Link Prediction didasarkan pada kesamaan algoritma. Setiap pasangan simpul *x* dan simpul *y*, yang telah ditetapkan untuk simpul ini adalah fungsi *similarity* (*x*, *y*), fungsi ini didefinisikan sebagai fungsi *similarity* antara simpul *x* dan *y*. Selanjutnya, mengurutkan pasangan *node* sesuai dengan nilai-nilai fungsi dari yang terbesar ke yang terkecil, semakin besar nilai dari fungsi *similarity*, semakin besar probabilitas *link* di dalam *node* [17].

Link Prediction berdasarkan *similarity index* salah satunya adalah *jaccard coefficient*, dimana [18] *node-node* yang memiliki kesamaan akan dihitung 1. Kesamaan ini dapat diatur yakni sebesar >90%.

3 Metode Penelitian

Metode penelitian yang dilakukan untuk menyelesaikan permasalahan di atas adalah menggunakan metode pada Gambar 1.



Gambar 1. Diagram Penelitian

3.1 Metode Pengumpulan Data

Sumber data yang digunakan dalam penelitian ini adalah data hasil kueri SPARQL id-dbpedia serta Wikipedia. Teknik pengumpulan data adalah dengan mengumpulkan data profil para pahlawan ke dalam satu tabel berformat CSV. Proses akuisisi data dimulai dengan menggunakan kueri SPARQL. Data diambil bernama “Pahlawan nasional Indonesia” dari dbpedia [19] dengan kueri yang ditunjukkan di bawah ini. Setelah kueri berhasil, data disimpan di dalam bentuk CSV. Beberapa potong data hasil dari kueri SPARQL dapat dilihat di Tabel 1.

```

SELECT distinct*
WHERE {
  ?pahlawan <http://purl.org/dc/terms/subject>
  <http://id.dbpedia.org/resource/Kategori:Pahlawan_nasional_Indon
  esia> .
  ?pahlawan <http://xmlns.com/foaf/0.1/name> ?nama_pahlawan .
  ?pahlawan <http://id.dbpedia.org/propertybirthPlace> ?kota_lahir
  .
  ?pahlawan <http://id.dbpedia.org/propertydeathPlace> ?kota_wafat
  .
}
    
```

Pada proses pengumpulan data dengan SPARQL, data yang berhasil diperoleh hanya sebanyak 58 data. Akan tetapi, agar dapat menjadi pembelajaran untuk penelitian ke depannya, penulis akan menunjukkan langkah-langkah yang telah dilakukan.

Tabel 1. Sampel Data Pahlawan Dari id.dbpedia.org

pahlawan	nama	kota_lahir	Kota_wafat
http://id.dbpedia.org/resource/Ahmad_Dahlan	"Kyai Haji Ahmad Dahlan"@id	http://id.dbpedia.org/resource/Kota_Yogyakarta	"Yogyakarta"@id
http://id.dbpedia.org/resource/Ahmad_Yani	"Ahmad Yani"@id	"border link=Hindia-Belanda 22pxPurworejo, Hindia Belanda"@id	"Jakarta, Indonesia"@id
http://id.dbpedia.org/resource/Amir_Hamzah	"Amir Hamzah"@id	http://id.dbpedia.org/resource/Kabupaten_Langkat	http://id.dbpedia.org/resource/Kota_Binjai
http://id.dbpedia.org/resource/Amir_Hamzah	"Amir Hamzah"@id	http://id.dbpedia.org/resource/Hindia_Belanda	http://id.dbpedia.org/resource/Kota_Binjai
http://id.dbpedia.org/resource/Amir_Hamzah	"Amir Hamzah"@id	http://id.dbpedia.org/resource/Sumatera_Timur	http://id.dbpedia.org/resource/Kota_Binjai

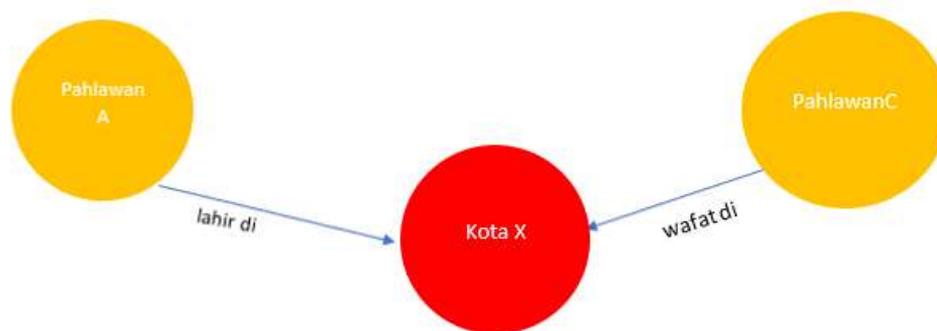
Dalam penelitian ini, data diperoleh dengan pendekatan manual. *Scraping* dari Wikipedia sendiri berhasil memperoleh 186 data pahlawan yang nantinya akan dicocokkan dengan data yang diperoleh melalui SPARQL untuk dilakukan pengubahan manual. Hasil pencocokan dapat dilihat pada Zenodo [20]. Data yang dikumpulkan adalah nama pahlawan, daerah (provinsi) kelahiran, tahun kelahiran, daerah (provinsi) wafat, tahun wafat, dan tahun ditetapkannya sebagai pahlawan nasional. *Scraping* dilakukan menggunakan bantuan *scraping tools* seperti ParseHub, namun pengkoreksian dan standarisasi data tetap dilakukan secara manual. Berikut adalah hasil yang diperoleh dalam bentuk *sample* atau *head data* yang dapat dilihat pada Tabel 2.

Tabel 2. Head Data Pahlawan Dari id.wikipedia.org

Nama	Daerah Lahir	Tahun Lahir	Daerah Wafat	Tahun Wafat	Tahun Penetapan
Abdul Halim	Sumatera Barat	1911	Jakarta	1988	2008
Abdul Halim Majalengka	Jawa Barat	1887	Jawa Barat	1962	2008
Abdul Haris Nasution	Sumatera Utara	1918	Jakarta	2000	2002
Abdul Kadir	Kalimantan Barat	1771	Kalimantan Barat	1875	1999
Abdul Kahar Mudzakkir	Yogyakarta	1907	Yogyakarta	1973	2019

Data nama pahlawan yang diperoleh tidak diubah, dengan maksud menjaga orisinalitas dari sumber yaitu Wikipedia. Sama halnya dengan data tahun penetapan status pahlawan nasional dan tidak ditemukan kejanggalan yang dirasa perlu untuk diubah. Sedangkan untuk data daerah kelahiran dan daerah wafat banyak dilakukan perubahan untuk menyesuaikan nama daerah yang dimaksud dengan nama daerah saat penelitian ini dilakukan. Terdapat berbagai penyebutan nama kota dari jaman penjajahan, nama kerajaan yang saat ini sudah tidak ada, dan semacamnya. Untuk tahun kelahiran dan tahun wafat, kurang dari sepuluh data dilakukan perubahan untuk data yang tidak seragam. Standarisasi data daerah juga dilakukan dengan cara mencatat provinsinya saja. Namun ada pengecualian bagi daerah di luar wilayah Indonesia, hanya nama negaranya saja dicatat. karena melakukan *pre-processing* secara manual, inkonsistensi tahun kelahiran dan tahun wafat banyak ditemukan sehingga tidak bisa dipastikan 100% benar. Beberapa di antaranya masih menjadi perdebatan para ahli, dan beberapa lainnya tidak memiliki sumber yang bisa menjelaskan dengan rinci tentang kelahiran dan kematian tokoh pahlawan yang dimaksud.

Berdasarkan proses yang dilakukan, tiga kolom utama yang didapatkan dibuat menjadi dua buah *node* yaitu "Nama Pahlawan" dan "Asal Kota". dari tahap ini, ditemukan popularitas suatu kota dilihat dari jumlah atau keberadaan relasi dengan tokoh pahlawan. *Node* pahlawan dengan *node* kota dihubungkan melalui sebuah relasi "lahir di" yang mana akan menunjukkan kota asal pahlawan tersebut. Ilustrasi *graph* dari proses data di tahap ini dapat dilihat pada Gambar 2.



Gambar 2. Ilustrasi Node Dan Relasi Pada Graf

3.2 Metode Analisis Data

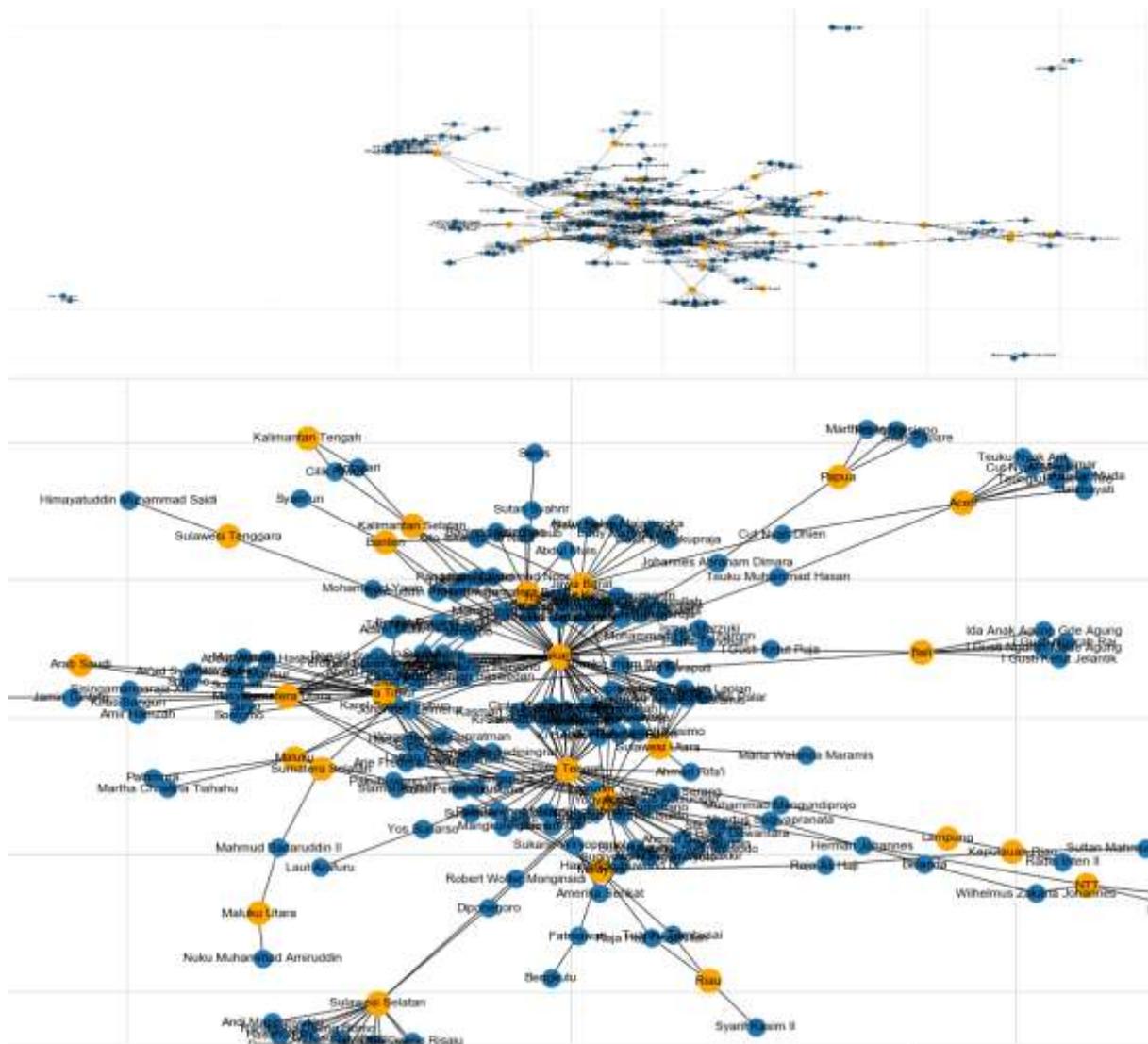
Teknik analisa data menggunakan analisa statistik deskriptif berupa graf, yaitu melakukan uji melalui perbandingan hasil klasifikasi Girvan Newman dengan Greedy Modularity yang keduanya merupakan Algoritma Community. Dengan adanya library NetworkX untuk mendeteksi dan mengelompokkan data ke dalam komunitas-komunitas agar visualisasi graf dapat terlihat. Dikarenakan pentingnya jumlah data pada tahap ini, maka tahap ini menggunakan data yang diperoleh dari Wikipedia dan kemudian disimpan dalam bentuk CSV.

Perhitungan selanjutnya adalah menentukan komunitas. Komunitas pada sekelompok simpul diukur dengan density setiap edge di antara mereka. Density inilah yang menentukan seberapa besar kepadatan membuat suatu komunitas. Bahkan jika menghasilkan grafik secara acak, beberapa kelompok simpul akan memiliki density yang lebih tinggi daripada yang lain, bisa jadi mereka adalah komunitas [11]. Node-node dalam graf dapat diprediksi dengan menggunakan Jaccard Coefficient untuk menghitung kesamaan node yang ada dengan node lain [17].

4 Hasil dan Pembahasan

Berikut adalah hasil plot untuk Daerah Populer dengan ditandai warna jingga seperti pada Gambar 3.

Dari Gambar 3, didapatkan daerah populer yaitu Jakarta, Sumatera Barat, Jawa Barat, Jawa Timur, Sulawesi Utara, Jawa tengah, Yogyakarta, Sumatera Utara, Malaysia (luar Indonesia), Arab Saudi (luar Indonesia), Maluku, Sumatera Selatan, Maluku Utara, Sulawesi Selatan, Riau, Lampung, Kepulauan Riau, NTT, Bali, Aceh, Papua, Kalimantan Tengah, Kalimantan Selatan, Banten, dan Sulawesi Tenggara. Hasil perhitungan *density* berdasarkan nama pahlawan, tempat lahir, dan tempat wafat adalah sebesar 0.011865079365079365. Hal ini berarti tidak semua node terhubung karena nilai density yang sangat rendah.



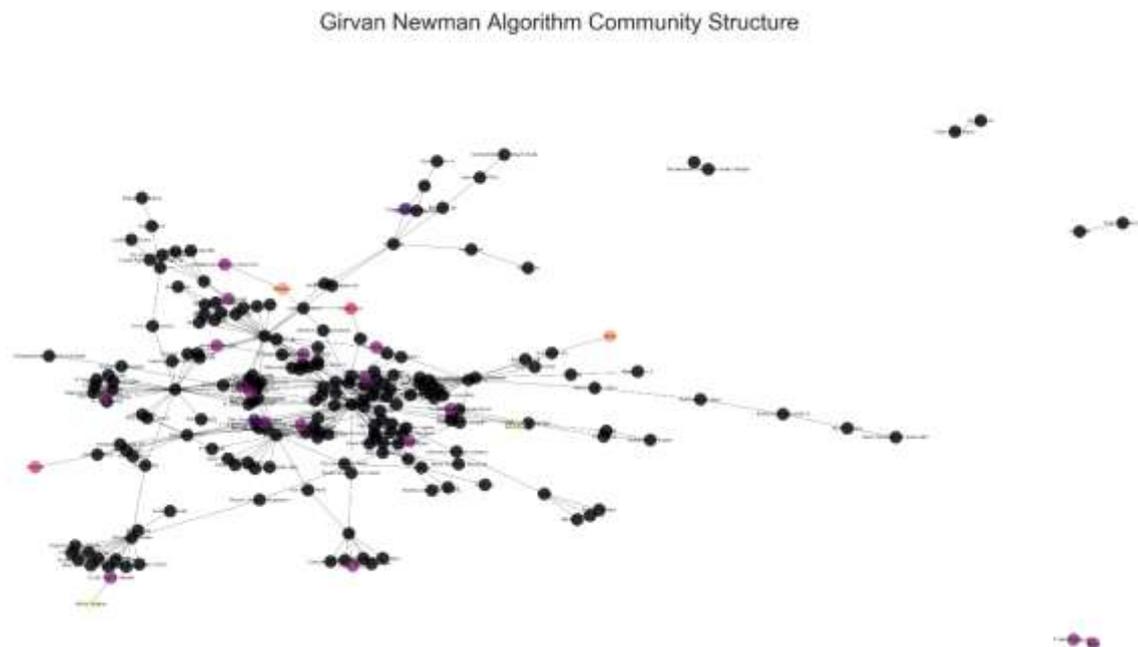
Gambar 3. Plot Daerah Populer

Tabel 2. Klasifikasi Komunitas Berdasarkan Algoritma Girvan Newman

		Komunitas
[0]	Wage Rudolf Supratman	0
[1]	Mohammad Hatta	0
[2]	Saharjo	0
[3]	Hazairin	0
[4]	Karel Satsuit Tubun	0
...
[220]	NTB	3
[221]	Nani Wartabone	4
[222]	Gorontalo	4
[223]	Jambi	5
[224]	Thaha Syaifuddin	5

Pada Tabel 2, algoritma Girvan Newman menghasilkan sejumlah 6 komunitas. “Wage Rudolf Supratman” berada di komunitas 0, “NTB” berada di komunitas 3, “Nani Wartabone” berada di komunitas 4, dst. Dari data ini kemudian dibuatkan struktur komunitas yang memperlihatkan

node-node yang saling berhubungan dalam graf dengan ditujukan melalui 6 warna *node* utama yang memperlihatkan komunitasnya yang dapat dilihat pada Gambar 4.



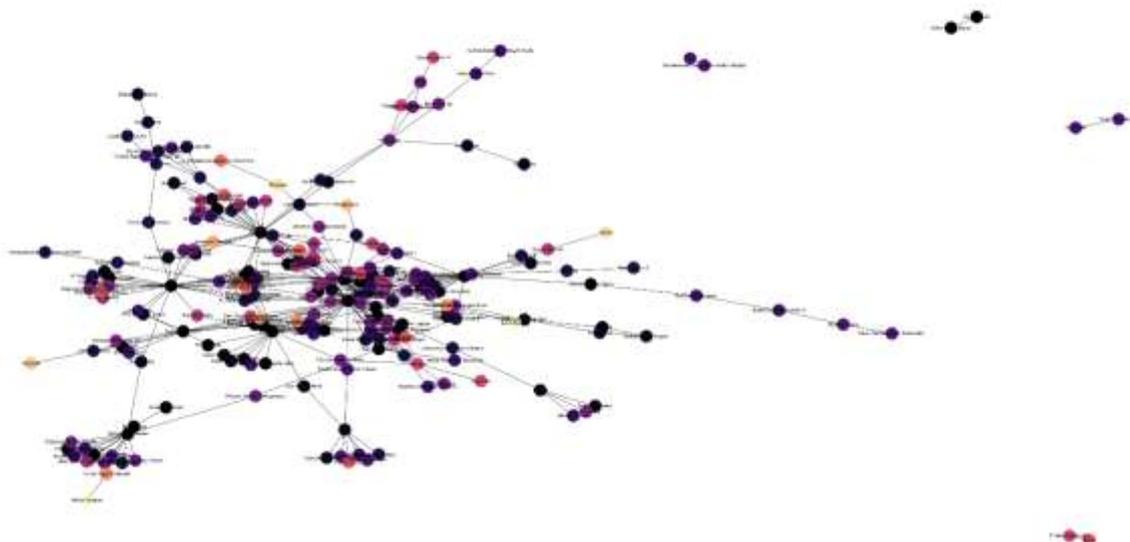
Gambar 4. Struktur Komunitas Berdasarkan Algoritma Girvan Newman

Tabel 3. Klasifikasi komunitas berdasarkan algoritma Greedy Modularity

		Komunitas
[0]	Mas Mansur	0
[1]	Wage Rudolf Supratman	0
[2]	Arab Saudi	0
[3]	Harun Bin Said	0
[4]	Fatmawati	0
...
[220]	NTB	14
[221]	Nani Wartabone	15
[222]	Gorontalo	15
[223]	Jambi	16
[224]	Thaha Syaifuddin	16

Algoritma Greedy Modularity mendapatkan 17 komunitas. “Wage Rudolf Supratman” berada di komunitas 0, “NTB” berada di komunitas 14, “Nani Wartabone” berada di komunitas 15, dst. Dari data ini kemudian dibuatkan struktur komunitas yang memperlihatkan *node-node* yang saling berhubungan dalam graf dengan ditujukan melalui 17 warna *node* utama yang memperlihatkan komunitasnya yang dapat dilihat pada Gambar 5.

Greedy Modularity Algorithm Community Structure



Gambar 5. Struktur Komunitas Berdasarkan Algoritma Greedy Modularity

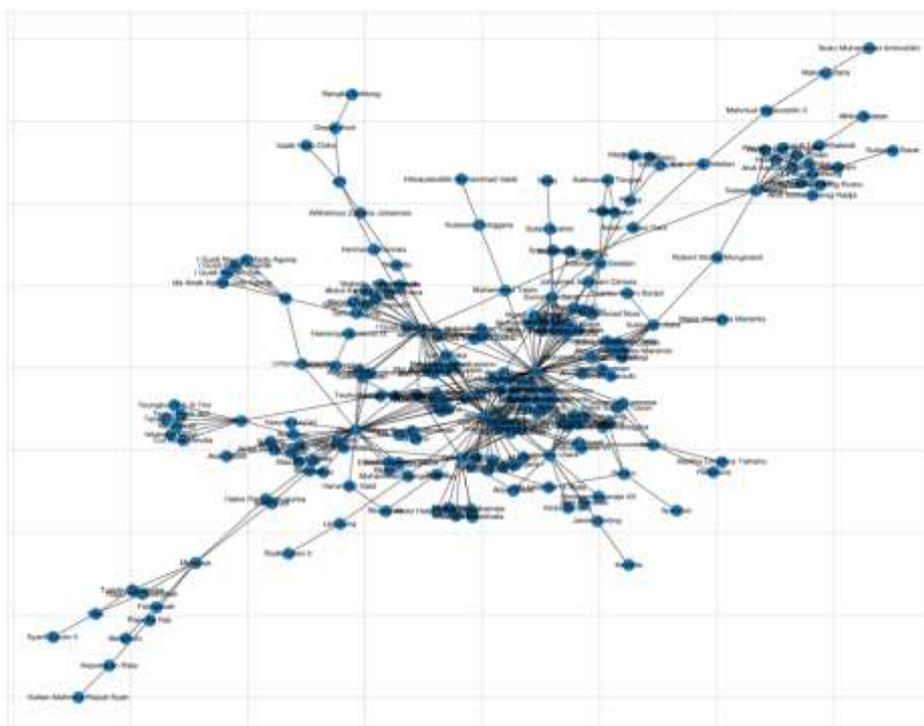
Prediksi kesamaan *node* menggunakan *similarity index* bernama Jaccard Coefficient, dimana jika ada dua *node* yang memiliki kesamaan lebih dari 90% akan bernilai 1. Hasil perhitungan Jaccard dapat dilihat pada Tabel 4.

Tabel 4 Hasil Jaccard Coefficient (sebagian)

Sutoyo Siswomiharjo Ahmad Yani 1.0
Sutoyo Siswomiharjo Alimin 1.0
Sutoyo Siswomiharjo Cipto Mangunkusumo 1.0
Sutoyo Siswomiharjo Gatot Subroto 1.0
Sutoyo Siswomiharjo Jatikusumo 1.0
Sutoyo Siswomiharjo Kasman Singodimedjo 1.0
Sutoyo Siswomiharjo Ki Sarmidi Mangunsarkoro 1.0
Sutoyo Siswomiharjo Saharjo 1.0
Sutoyo Siswomiharjo Siswondo Parman 1.0
Sutoyo Siswomiharjo Siti Hartinah 1.0
Sutoyo Siswomiharjo Supomo 1.0
Sutoyo Siswomiharjo Suprpto 1.0
Sutoyo Siswomiharjo Tirto Adhi Suryo 1.0
Tirto Adhi Suryo Ahmad Yani 1.0
Tirto Adhi Suryo Alimin 1.0
Tirto Adhi Suryo Cipto Mangunkusumo 1.0
Tirto Adhi Suryo Gatot Subroto 1.0
Tirto Adhi Suryo Jatikusumo 1.0
Tirto Adhi Suryo Kasman Singodimedjo 1.0
Tirto Adhi Suryo Ki Sarmidi Mangunsarkoro 1.0
Tirto Adhi Suryo Saharjo 1.0
Tirto Adhi Suryo Siti Hartinah 1.0
Tirto Adhi Suryo Supomo 1.0

Tirto Adhi Suryo Suprpto 1.0
...
Siswondo Parman Ahmad Yani 1.0
Siswondo Parman Alimin 1.0
Siswondo Parman Cipto Mangunkusumo 1.0
Siswondo Parman Gatot Subroto 1.0
Siswondo Parman Jatikusumo 1.0
Siswondo Parman Kasman Singodimedjo 1.0
Siswondo Parman Ki Sarmidi Mangunsarkoro 1.0
Siswondo Parman Saharjo 1.0
Siswondo Parman Siti Hartinah 1.0
Siswondo Parman Supomo 1.0
Siswondo Parman Suprpto 1.0

Algoritma Jaccard menghasilkan 372 pasang *node* yang memiliki kesamaan, seperti misalnya Siswondo Parman dengan Ahmad Yani yang akhirnya membentuk sebuah relasi baru.



Gambar 6. Hasil Jaccard Coefficient

5 Kesimpulan

Kesimpulan dari penelitian ini adalah klasifikasi menggunakan Algoritma Community dapat memperlihatkan *node* populer yaitu dengan nama “daerah populer” yang ditunjukkan dengan warna berbeda, yaitu jingga. Hasil klasifikasi menggunakan Algoritma Greedy Modularity dinilai lebih banyak menghasilkan komunitas yakni sebanyak 16 komunitas yang dapat terbentuk, sedangkan algoritma Girvan Newman hanya menghasilkan 6 komunitas. Untuk memprediksi berapa banyak *node* yang memiliki kesamaan, dengan Jaccard Coefficient ditemukan hingga 372 *node*.

Referensi

- [1] M. Tasgin, A. Herdagdelen, and H. Bingol, “Community Detection in Complex Networks Using Genetic Algorithms,” Nov. 2007, Accessed: May 22, 2020. [Online]. Available:

<http://sistemasi.ftik.unisi.ac.id>

- <http://arxiv.org/abs/0711.0491>.
- [2] Nurmasiyah, M. T. Imelda Atastina, S. Si, and M. T. Anisa Herdiani, S.T., “Analisis Dan Implementasi Community Detection Menggunakan Spectral Clustering Method Dalam Social Network | Nurmasiyah | eProceedings of Engineering,” *e-Proceeding of Engineering*, 2018. <https://libraryproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/6738/6638> (accessed May 22, 2020).
- [3] Itsna Alfin Nur, Moch. Arif Bijaksana, and E. D. S. T. . M. T. Ir. M.Tech., “Community Detection Menggunakan Genetic Algorithm Dalam Social Network Twitter | Nur | eProceedings of Engineering,” 2015, Accessed: May 22, 2020. [Online]. Available: <https://libraryproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/1287/1229>.
- [4] Lulu Alfi’a Rahma Ningsih, Imelda Atastina, and Anisa Herdiani, “Analisis Dan Implementasi Community Detection Menggunakan Algoritma Dbscan Pada Twitter | Ningsih | eProceedings of Engineering,” 2018. <https://libraryproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/6154/6133> (accessed May 22, 2020).
- [5] A. N. Annisa, “Studi Literatur Perencanaan dan Algoritma Pembentukan DMA (District Metered Area),” 2016.
- [6] M. A. Javed, M. S. Younis, S. Latif, J. Qadir, and A. Baig, “Community detection in networks: A multidisciplinary review,” *J. Netw. Comput. Appl.*, 2018.
- [7] “SPARQL 1.1 overview.,” *World Wide Web Consortium*, 2013. <https://www.w3.org/TR/sparql11-query/> (accessed May 06, 2020).
- [8] N. Akhtar, “Social network analysis tools,” in *Proceedings - 2014 4th International Conference on Communication Systems and Network Technologies, CSNT 2014*, 2014, pp. 388–392, doi: 10.1109/CSNT.2014.83.
- [9] A. Hagberg, P. Swart, and D. S. Chult, “Exploring network structure, dynamics, and function using networkx (Conference) | OSTI.GOV,” 2008, Accessed: May 22, 2020. [Online]. Available: <https://www.osti.gov/biblio/960616>.
- [10] P. V. Marsden, “The reliability of network density and composition measures,” *Soc. Networks*, vol. 15, no. 4, pp. 399–421, Dec. 1993, doi: 10.1016/0378-8733(93)90014-C.
- [11] Vinh-Loc Dao, C. Bothorel, and P. Lenca, “Community structure: A comparative evaluation of community detection methods,” vol. hal-019765, 2019.
- [12] S. Moon, J. G. Lee, and M. Kang, “Scalable community detection from networks by computing edge betweenness on MapReduce,” in *2014 International Conference on Big Data and Smart Computing, BIGCOMP 2014*, 2014, pp. 145–148, doi: 10.1109/BIGCOMP.2014.6741425.
- [13] L. Despalatovic, T. Vojkovic, and D. Vukicevic, “Community structure in networks: improving the Girvan-Newman algorithm,” in *Conference: 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2014, pp. 997–1002, doi: 10.1109/MIPRO.2014.6859714.
- [14] S. Lobanova and A. Chepovskiy, “Combined method to detect communities in graphs of interacting objects,” *Bus. Informatics*, vol. 2017, no. 4, pp. 64–73, Dec. 2017, doi: 10.17323/1998-0663.2017.4.64.73.
- [15] M. Chen, K. Kuzmin, and B. K. Szymanski, “Community detection via maximization of modularity and its variants,” *IEEE Trans. Comput. Soc. Syst.*, vol. 1, no. 1, pp. 46–65, Mar. 2014, doi: 10.1109/TCSS.2014.2307458.
- [16] A. F. Almukhtar and E. S. Al-Shamery, “Greedy Modularity Graph Clustering for Community Detection of Large Co-Authorship Network,” *Int. J. Eng. Technol.*, 2018.
- [17] Liyan Dong, Yongli Li, Han Yin, Huang Le, and Mao Rui, “The Algorithm of Link Prediction on Social Network,” *Math. Probl. Eng.*, 2013, Accessed: May 22, 2020. [Online]. Available: <https://www.hindawi.com/journals/mpe/2013/125123/>.
- [18] A. F. Al-Mukhtar and E. S. Al-Shamery, “Greedy Modularity Graph Clustering for Community Detection of Large Co-Authorship Network,” *Int. J. Eng. Technol.*, vol. 7, no. 4.19, p. 857, Nov. 2018, doi: 10.14419/ijet.v7i4.19.28058.
- [19] Dbpedia, “http://id.dbpedia.org/page/Kategori:Pahlawan_nasional_Indonesia.”

- [20] N. I. Maghfirani and L. P. G. Widiastuti, "Klasterisasi Pahlawan Nasional Indonesia Berdasarkan Daerah Asal menggunakan Algoritma Louvain," May 2020, doi: 10.5281/ZENODO.3808974.