

Laporan Tugas Akhir

Program KMMI Social Network in R (Batch 3)



Kelompok 11 :

1. Ikhwan Arif Fauzan (A11.2020.12867)
2. Irvan Cahaya Putra (A11.2019.12148)
3. Yuliette Rindani (A12.2018.06075)

Fakultas Ilmu Komputer

Universitas Dian Nuswantoro

Tahun 2021

Pendahuluan

1. Latar belakang

Analisis data bertujuan mendapatkan informasi yang terkandung di dalam data dan menggunakan hasilnya untuk memecahkan suatu permasalahan. Berbagai metode dapat diimplementasikan guna pelaksanaan analisis data, baik data dengan skala kecil maupun skala besar. Salah satu metode yang sering digunakan dalam analisis data yakni *Betweenness Centrality*.

Dalam teori graf, *betweenness centrality* adalah ukuran sentralitas (pemusatan) dalam graf berdasarkan jalur terpendek. Untuk setiap pasangan simpul dalam graf terhubung, terdapat setidaknya satu jalur terpendek antar simpul sedemikian rupa sehingga jumlah sisi yang dilalui jalur (untuk graf tak berbobot) atau jumlah bobot sisi (untuk graf berbobot) diminimalkan. *betweenness centrality* dari setiap simpul adalah jumlah jalur terpendek yang melewati simpul tersebut.

Dalam teori graf dan analisis jaringan, indikator *Centrality* menetapkan angka atau peringkat ke *node* dalam grafik yang sesuai dengan posisi jaringannya. pengaplikasiannya termasuk untuk mengidentifikasi orang yang paling berpengaruh dalam jaringan sosial, simpul infrastruktur utama di Internet atau jaringan perkotaan, penyebarpenyakit, dan jaringan otak. Konsep *Centrality* pertama kali dikembangkan dalam analisis jaringan sosial, dan banyak istilah yang digunakan untuk mengukur *Centrality*.

Seperti contoh di atas, Analisis Data dapat dilakukan dalam berbagai kelompok jaringan. Dalam program KMMI SNR ini, Analisis Data dilakukan guna mengetahui berbagai informasi dari suatu data, seperti tokoh, networking dan lain sebagainya yang dapat digunakan sebagai alat tracking untuk memperoleh informasi yang akurat.

Metode analisis data yang digunakan ialah melalui visualisasi data dengan berbagai metode seperti contohnya *betweenness centrality*, *Modularity*, *Eigen vector* dan lain sebagainya untuk melihat kedetailan dari data tersebut dari berbagai sudut pandang, terutama dengan sudut pandang analisis sosial seperti program yang berlangsung ini.

2. Rumusan Masalah

Dalam penyusunan tugas akhir ini, permasalahan yang dibahas ialah bagaimana visualisasi data yang dilakukan melalui Tools Gephi dengan berbagai metode visualisasi berdasarkan data yang telah dianalisis sebelumnya dengan topik warga / masyarakat. Serta menampilkan hasil analisis data berupa tokoh yang sangat berpengaruh (menginfluence) dengan topik yang sedang dibahas.

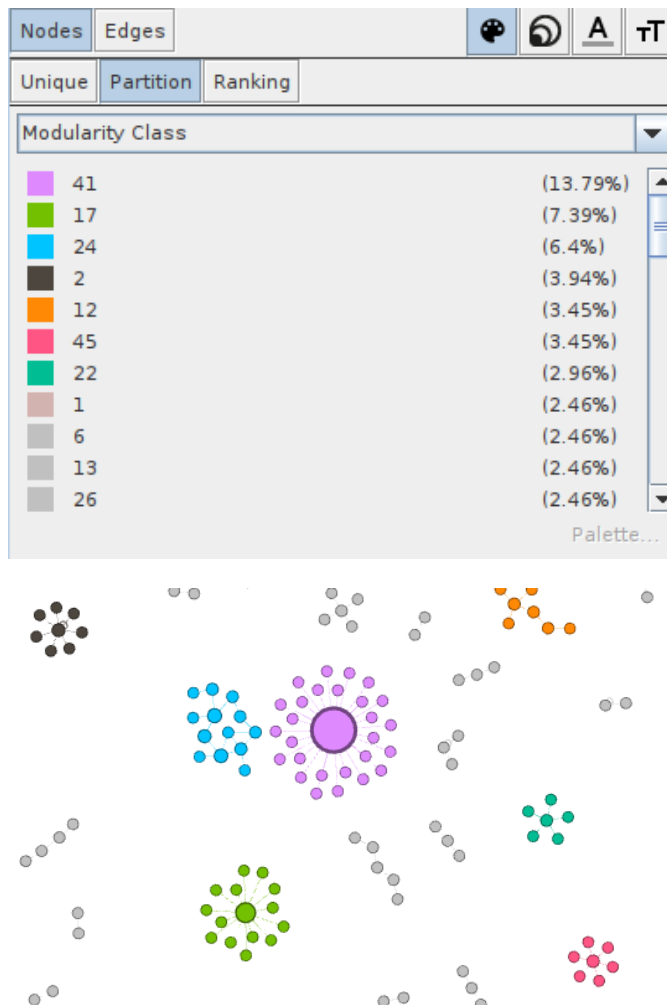
3. Tujuan

- a. Mempelajari proses analisis data yang dilakukan dengan menggunakan Tools Rstudio
- b. Mempelajari Visualisasi data yang dilakukan dengan menggunakan Tools Gephi.
- c. Menganalisis data yang telah divisualisasikan dengan menggunakan Tools Gephi sehingga memperoleh berbagai informasi

Metode dan Teknik





Pada analisis 25000 data dari covid19_tweet.csv dengan kata kunci masyarakat dan warga, digunakan metode dan teknik sebagai berikut.

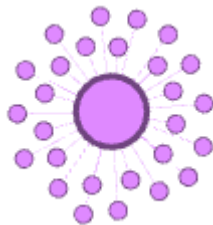
1. Modularity Class



Metode ini dilakukan dengan mengelompokkan warna node menggunakan nilai modularity sebagai parameternya. Hal ini bertujuan untuk memperlihatkan kelompok atau komunitas yang terlibat dalam topik terkait kata kunci.





2. Betweenness Centrality

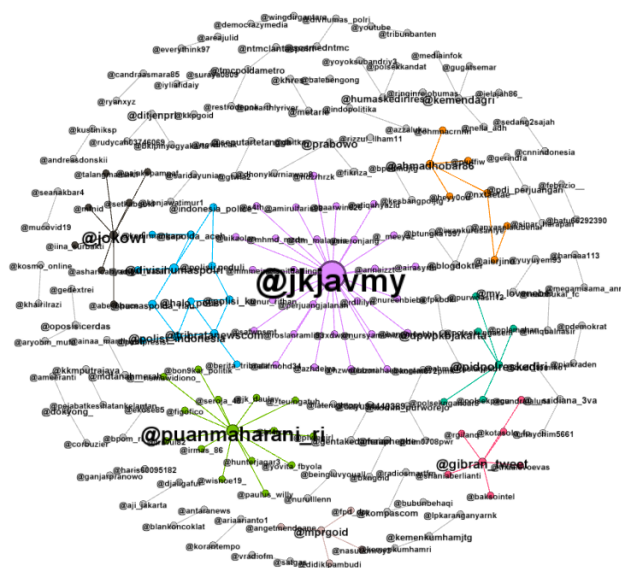
Nodes	Edges				
Unique	Ranking				
Betweenness Centrality					
Min size:		<input type="text" value="15"/>	Max size:		<input type="text" value="60"/>



Metode ini dilakukan dengan memperlihatkan node yang menjadi penghubung informasi. Hal ini digunakan untuk mengidentifikasi orang yang berperan sebagai jembatan antara dua komunitas.

3. Degree Centrality

Nodes	Edges				
Unique	Ranking				
Degree					
Min size:		<input type="text" value="1"/>	Max size:		<input type="text" value="4"/>



Metode Degree Centrality ini dilakukan dengan menunjukkan jumlah koneksi yang dimiliki sebuah node. Jumlah koneksi digambarkan oleh label, yaitu semakin besar ukuran label maka semakin banyak jumlah koneksi yang dimiliki node tersebut.

4. Filtering

```
df <- filter (df, str_detect(text, "masyarakat|warga"))
```

Teknik Filtering digunakan untuk menyaring data yang sesuai dengan kata kunci atau ketentuan yang diperlukan.

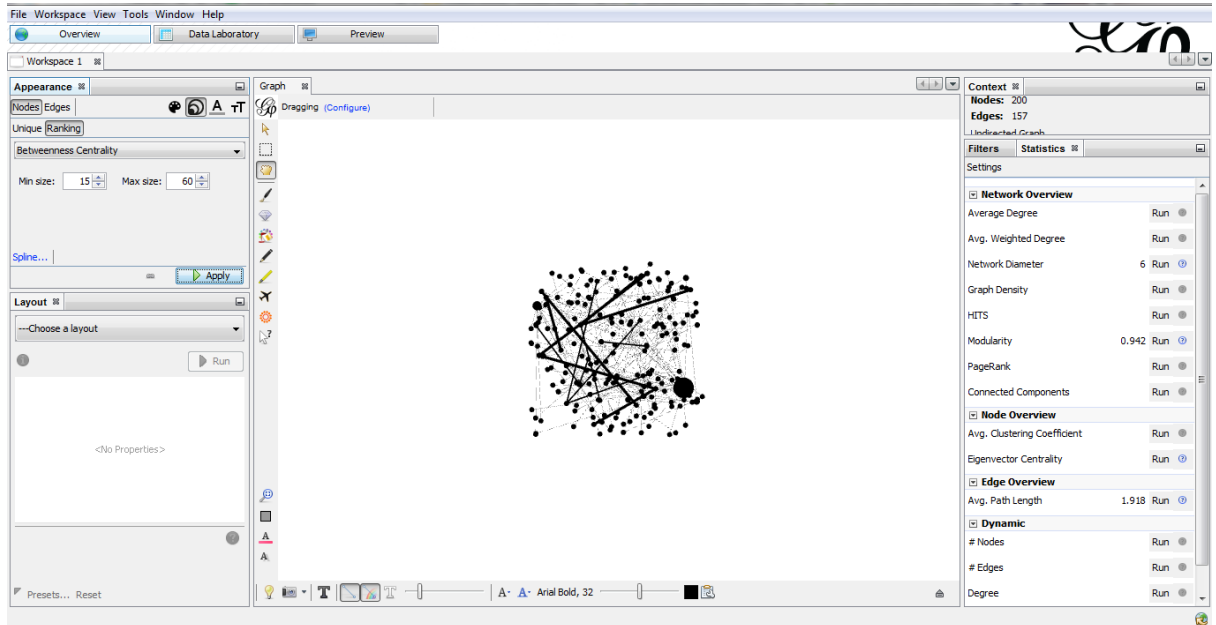
Langkah-langkah filtering

1. Pertama memanggil Library yang dibutuhkan, lalu mengimport dataset yang telah disediakan, yaitu covid19_tweet.csv
2. Dataset covid19_tweet memiliki 73 variabel atau kolom dengan 125577 entri.
3. Melakukan filter jumlah data, data yang kami gunakan 25000
4. Melakukan filtering pada kolom text dengan str_detect untuk mendeteksi word yaitu word "masyarakat".dan "warga"
5. Melakukan select untuk memilih 2 kolom pada dataset covid19_tweet, yaitu kolom user_screenshot dan kolom text.
6. Menambahkan @ ke username
7. Menggabungkan 2 kolom
8. Mengambil hanya user (@...)
9. Menghitung jumlah user dalam 1 baris
10. Menghapus baris yang berisi hanya 1 user
11. Membuat pasangan source-target
12. Memisahkan ke dalam 2 kolom
13. Menambah @ di depan username

14. Export file

Langkah-langkah visualisasi data

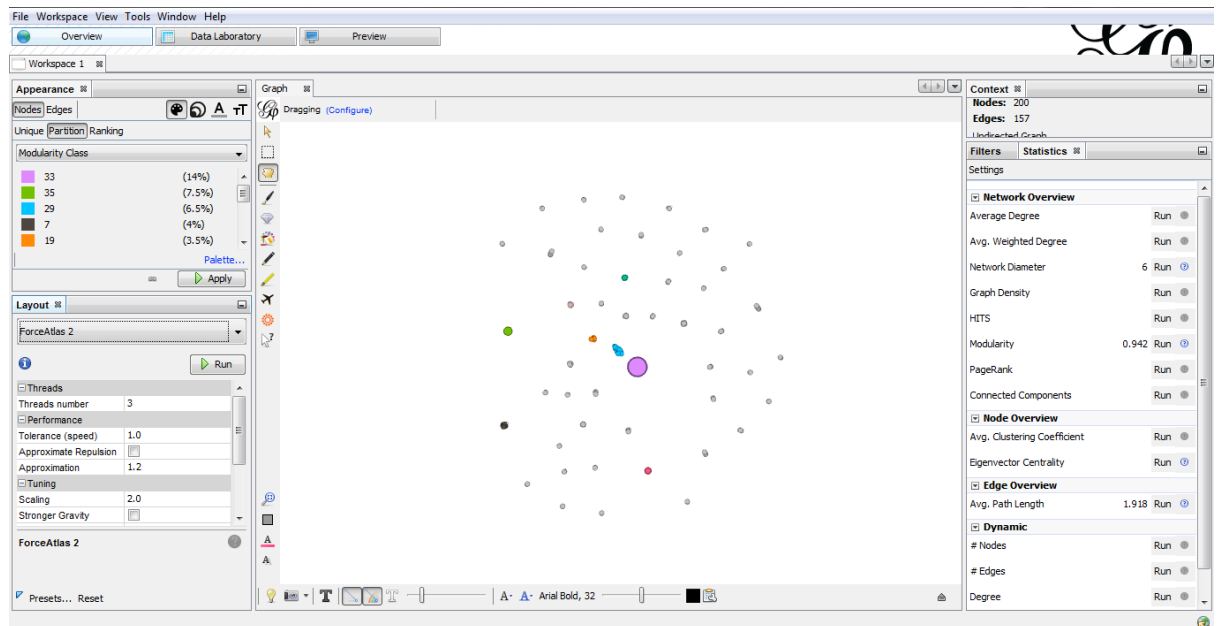
1. Membuka file gexf melalui aplikasi Gephi. Buka Gephi > Open File
2. Kemudian, pilih file gexf yang telah disimpan tadi dan ubah graph type menjadi undirected dan klik OK.



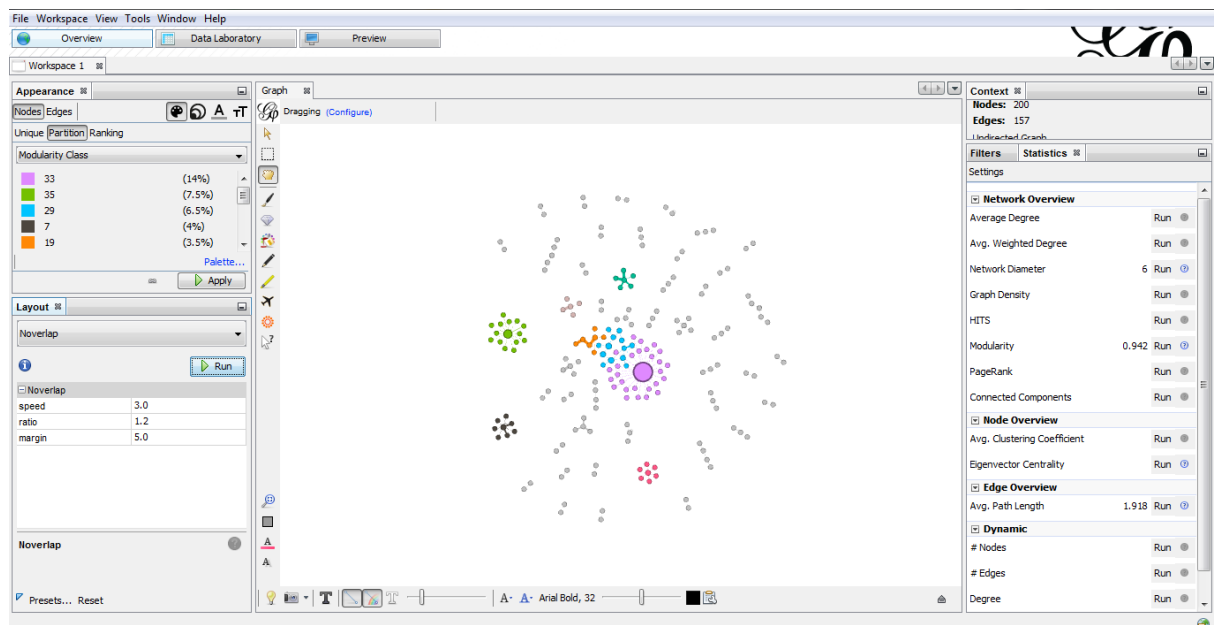
Untuk visualisasi yang lebih jelas, akan dibuat visualisasi dengan modularity yang diwakili oleh perbedaan warna dan penggunaan skor betweenness centrality yang diwakili oleh ukuran font. Semakin besar ukuran font, maka semakin tinggi skor centralitynya. Pertama - tama, hitung skor centrality. Run beberapa opsi pada menu statistic di sebelah kanan seperti Modularity, Network Diameter dan Eigenvector Centrality. Modularity untuk mengukur dan mengelompokkan node berdasarkan komunitas, Network Diameter untuk mendapatkan skor degree, dan Eigenvector Centrality untuk menunjukkan bobot yang lebih tinggi pada node yang terhubung dengan node yang juga memiliki keterhubungan tinggi.

3. Kemudian, mengatur warna berdasarkan modularity. Pada bagian Appearance, klik Nodes > Ikon palette > Partition > Betweenness Centrality >

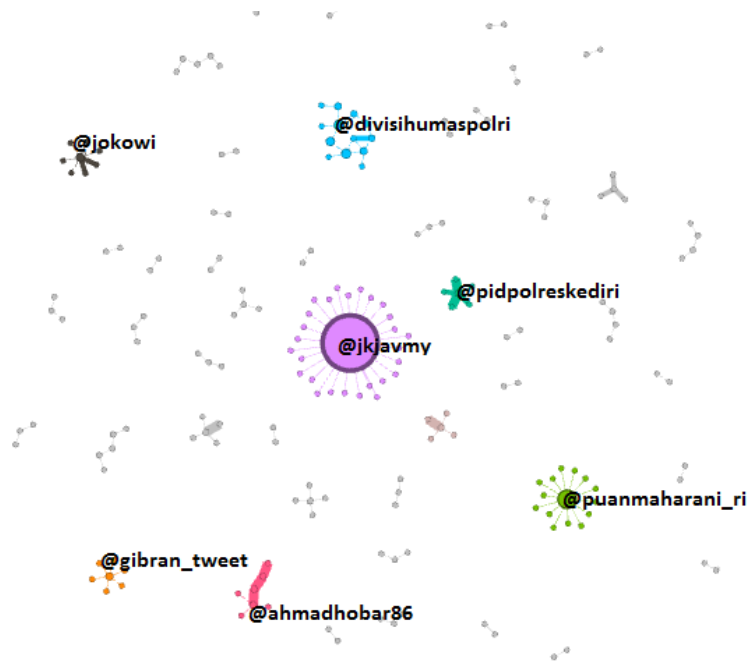
Apply. Lalu, Klik Nodes > Ikon Size > Ranking > Dropdown Modularity Class > Apply untuk menyesuaikan ukuran nodes sesuai dengan modularity class.



4. Selanjutnya, mengatur bentuk graf. Pada bagian Layout, pilih Noverlap, lalu Run. Apply. Klik ikon di bawah pada tab label centang pada bagian node dan hide non selected untuk memunculkan nama node.

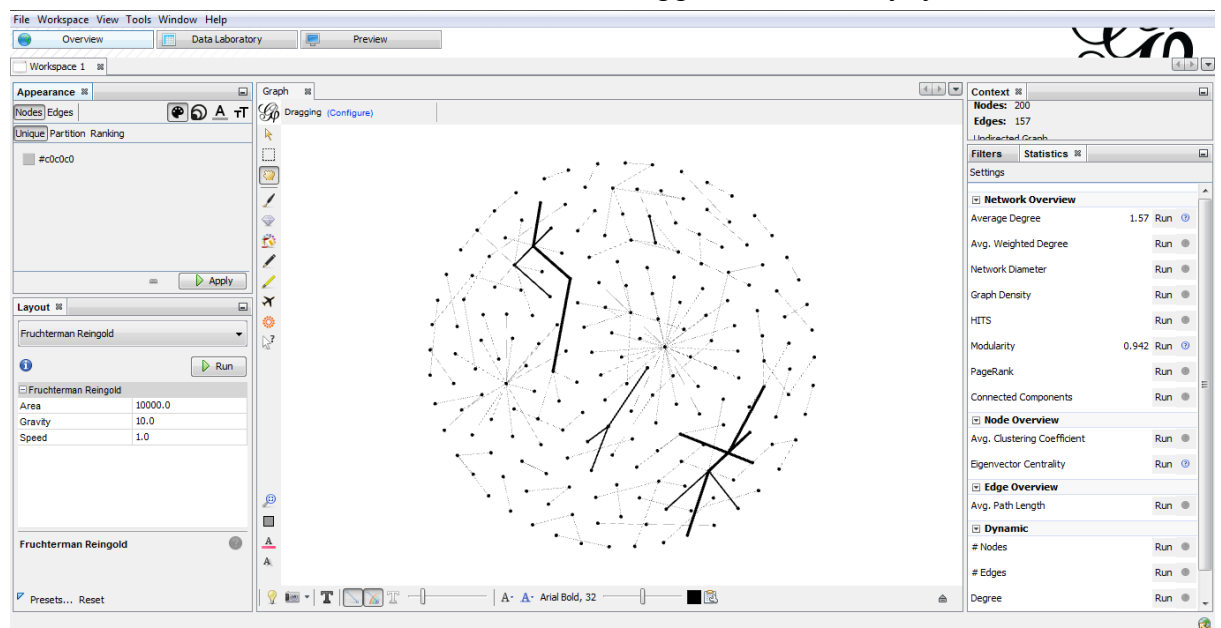


5. Hasil akhir visualisasi



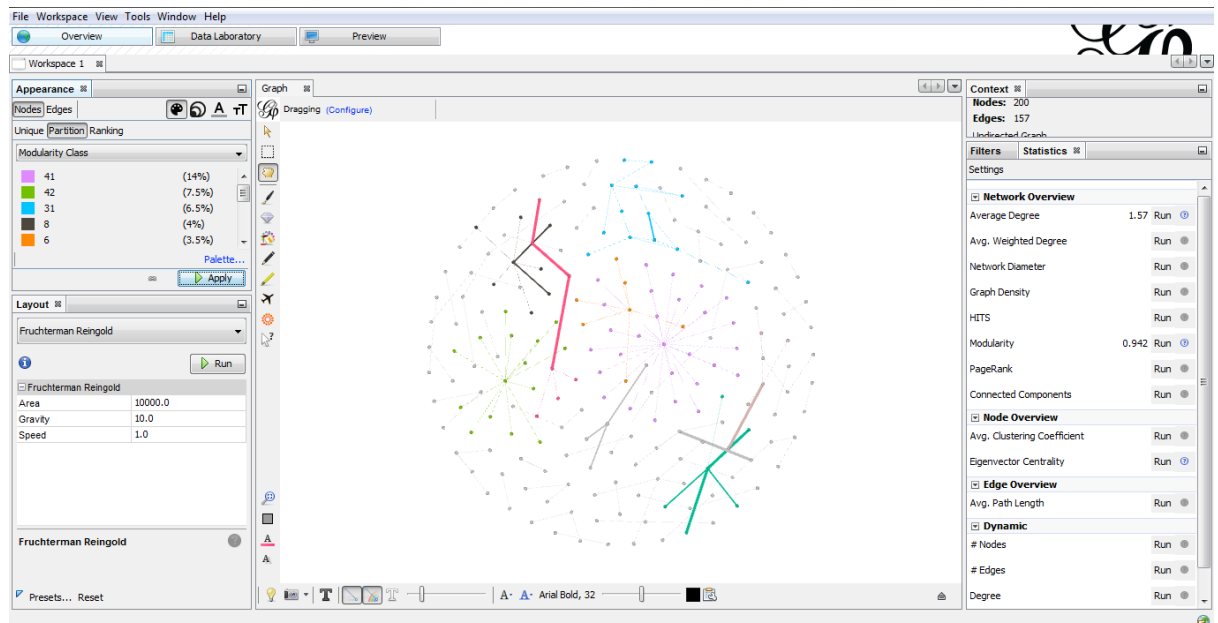
6. Visualisasi Degree Centrality

akan dibuat visualisasi dengan modularity yang diwakili oleh perbedaan warna dan penggunaan skor degree centrality yang diwakili oleh ukuran font. Semakin besar ukuran font, maka semakin tinggi skor centralitinya.

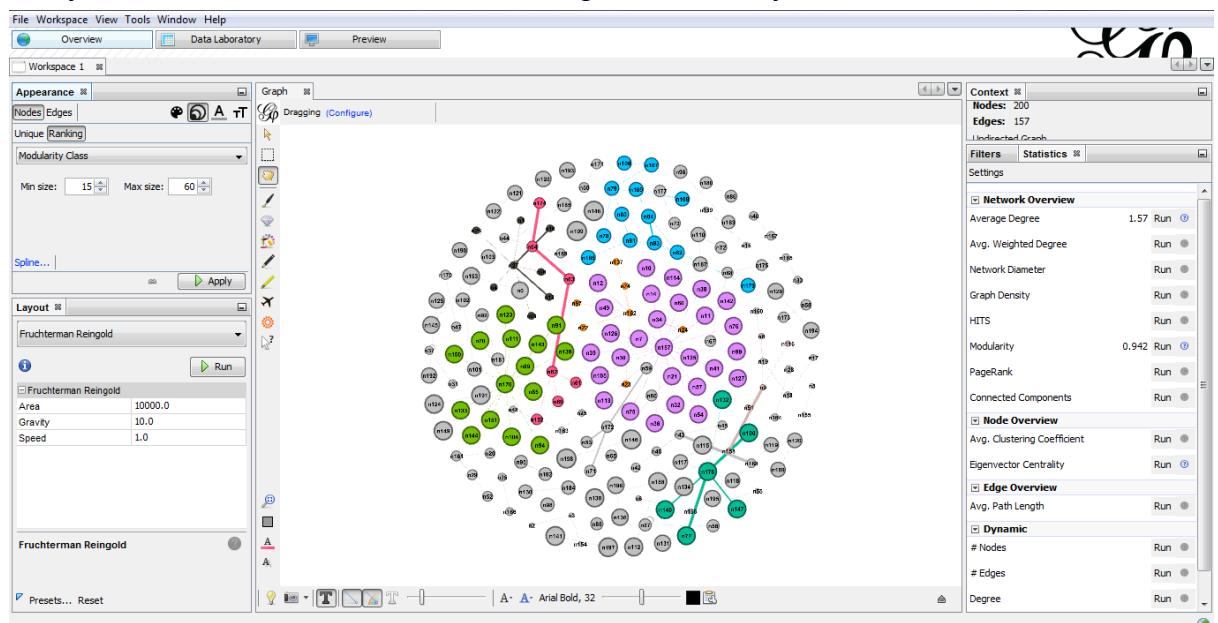


Run beberapa opsi pada menu statistic di sebelah kanan seperti Modularity, Average Degree dan Eigenvector Centrality. Modularity untuk mengukur dan mengelompokkan node berdasarkan komunitas, Average Degree untuk mendapatkan skor degree, dan Eigenvector Centrality untuk menunjukkan bobot yang lebih tinggi pada node yang terhubung dengan node yang juga memiliki keterhubungan tinggi.

- Selanjutnya, mengatur bentuk graf. Pada bagian Layout, dropdown dan pilih FruchtermanReingold, lalu Run. Apply. Klik ikon T di bawah untuk memunculkan nama node. Apply. Klik ikon T di bawah untuk memunculkan nama node.

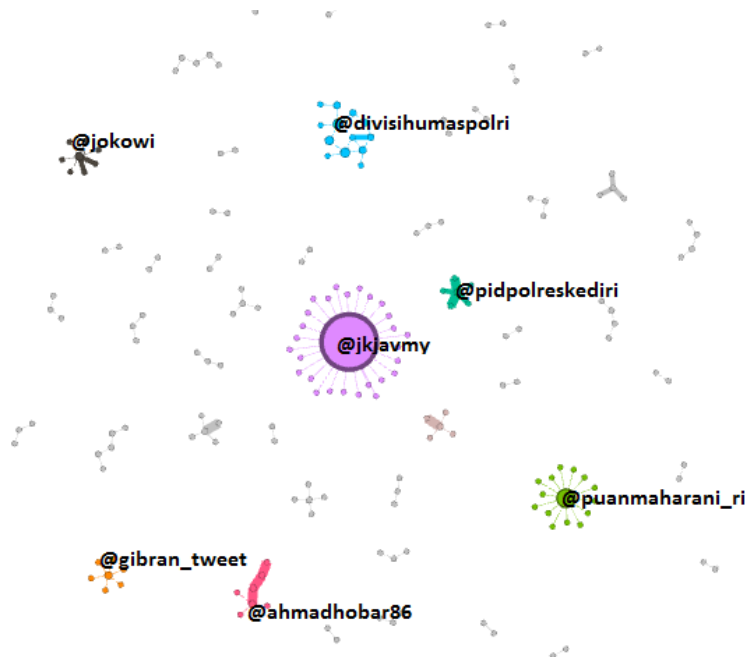


- Mengatur warna berdasarkan modularity. Pada bagian Appearance, klik Nodes > Ikon palette > Partition > Dropdown Modularity Class > Apply. Lalu, Klik Nodes > Ikon Size > Ranking > Dropdown Modularity Class > Apply untuk menyesuaikan ukuran nodes sesuai dengan modularity class.

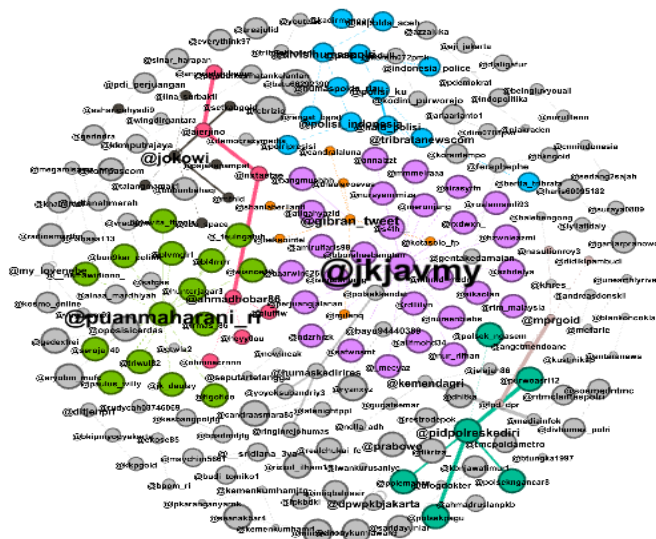


Analisa dan Penutup

Pemrosesan data yang dilakukan menggunakan data covid19_tweet.csv dengan jumlah data sebanyak 25000. Teknik filtering pada kata masyarakat dan warga dengan menggunakan dua metode yaitu Betweenness centrality dan Degree Centrality didapatkan hasil sebagai berikut:



Betweenness Centrality



Degree Centrality

Dengan hasil result modularity

Modularity : 0,942

Modularity with resolution: 0.942

Number of Communities: 49

Most Influential: @jkjavmy, @puanmaharani_ri, @divisihumaspolri, @jokowi, @gibran_tweet, @pidpolreskediri, @ahmadhobar86

Keterangan Akun

1. @jkjavmy



2. @puanmaharani_ri



3. @divisihumaspolri

twitter.com/DivisiHumasPolri

DivisiHumasPolri_RBX
0 Tweet

DivisiHumasPolri_RBX
@DivisiHumasPolri

Welcome to Divisi Humas Polri ROBLOX official Twitter account.
We are not affiliated with real @HumasPolri
[Terjemahkan bio](#)

Bergabung Juni 2019
1 Mengikuti 96 Pengikut
Tidak diikuti oleh siapa pun yang Anda ikuti

Tweet Tweet & balasan Media Suka

Anda mungkin suka

- Adam Sandi Toshiaki @Adamsan16 Ikuti
- Derajat @Derajat21654275 Ikuti
- HumasPolsek Liukung Kal @KalmasHumas Ikuti

[Tampilkan lebih banyak](#)

Tren untuk Anda

Populer di Indonesia
Pacarku
9.971 Tweet

#LoversOfTheRedSky
Nonton Lovers Of The Red Sky Sekarang
Dipromosikan oleh Vlu Indonesia

4. @gibran_tweet

twitter.com/gibran_tweet

Gibran Rakabuming
131 Tweet

Gibran Rakabuming
@gibran_tweet

Kebut vaksinasi
Kebut pemulihan ekonomi
Whatsapp Lapor Mas Wali 081225067171 (text only)
Surakarta, Jawa Tengah Bergabung Februari 2021

19 Mengikuti 7.091 Pengikut
Tidak diikuti oleh siapa pun yang Anda ikuti

Tweet Tweet & balasan Media Suka

Anda mungkin suka

- ORICON NEWS (オリコン) @oricon Ikuti
- IG chchotimah @ChusnulCh_ Ikuti
- Nita I @Leonita_Lestari Ikuti

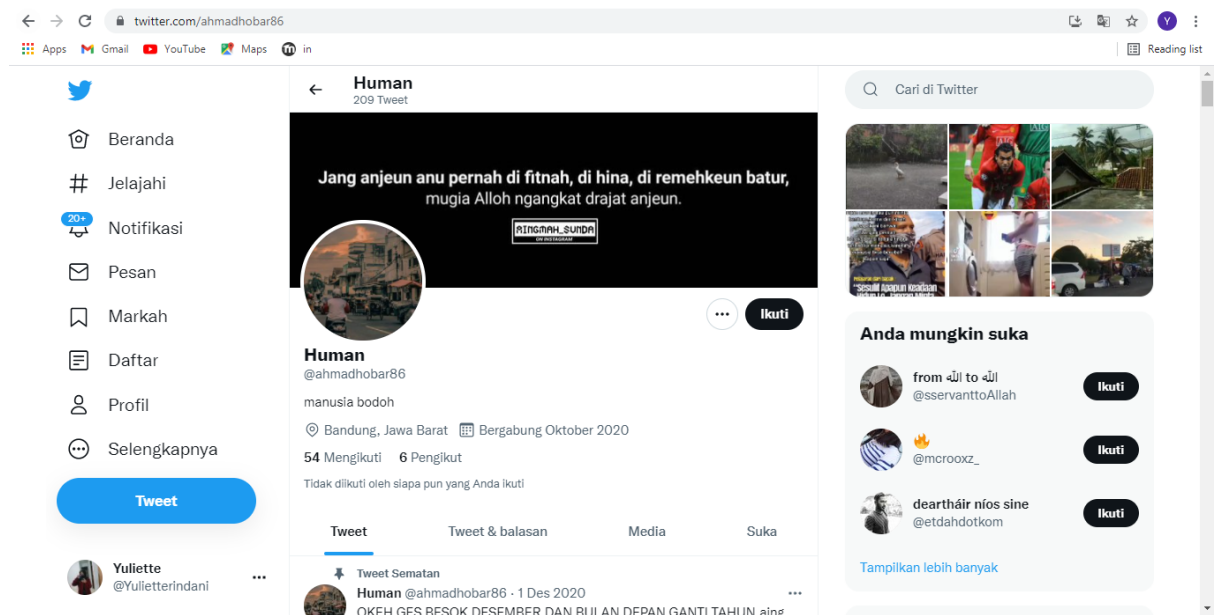
[Tampilkan lebih banyak](#)

5. @jokowi



6. @pidpolreskediri Akun dibekukan.

7. @ahmadhobar86



Kesimpulan

Pada data yang diambil dari dataset covid19_tweet.csv dengan memfilter kata “masyarakat” dan “warga” dihasilkan visualisasi seperti gambar di atas. Pada visualisasi tersebut, terdapat 49 komunitas - komunitas kecil. Node “jkjavmy” memiliki ukuran paling besar yang artinya memiliki skor betweenness dan degree centrality yang paling tinggi pada komunitas yang ditandai dengan warna ungu. Artinya, jkjavmy adalah node yang memiliki koneksi paling banyak diantara nodes lainnya.