

LAPORAN TUGAS AKHIR
INFORMATION RETRIEVAL
“ Social Network Analysis ”



Nama Anggota Kelompok :

Ong Gabriel Riverine Susanto	205314111
Kurniawan Ronaldi Purnama	205314148
Paulus Caesario Dito P. Hartono	205314159

PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS SANATA DHARMA
YOGYAKARTA

2022

DAFTAR ISI

DAFTAR ISI	2
PEMIKIRAN DASAR	3
METODOLOGI	4
HASIL DAN ANALISIS	5
REFERENSI	6
LAMPIRAN	7

PEMIKIRAN DASAR

1. Social Network Analysis

Social Network Analysis (SNA) merupakan bidang studi yang memfokuskan pada struktur dan pola hubungan antara orang atau organisasi. Ini melibatkan penggunaan teori jaringan dan graf untuk menganalisis struktur sosial dan mengidentifikasi pola dan tren dalam cara orang atau organisasi berkomunikasi. Tujuan analisis jaringan sosial adalah untuk memahami bagaimana hubungan antara individu atau organisasi mempengaruhi perilakunya dan, pada akhirnya, untuk dapat memprediksi perilaku individu atau kelompok berdasarkan hubungan mereka dengan orang lain. Analisis jaringan sosial biasanya digunakan di bidang seperti sosiologi, antropologi, psikologi, dan bisnis.

a. Manfaat Social Network Analysis

Social Network Analysis memiliki banyak kegunaan dalam berbagai bidang. Di bidang sosiologi, misalnya, Social Network Analysis dapat digunakan untuk memahami struktur hubungan antarindividu dalam sebuah masyarakat atau kelompok dan bagaimana struktur tersebut mempengaruhi perilaku individu. Ini juga dapat digunakan untuk mengidentifikasi pola hubungan yang mempengaruhi kelompok sosial seperti kelompok kriminal atau kelompok teroris.

Di bidang bisnis, analisis jaringan sosial dapat digunakan untuk memahami hubungan antara perusahaan dan individu-individu yang mempengaruhi keputusan mereka, seperti investor, konsumen, atau mitra bisnis. Ini juga dapat digunakan untuk mengidentifikasi pola hubungan yang mempengaruhi keberhasilan sebuah bisnis, seperti jaringan distribusi atau jaringan komunikasi.

Secara umum, analisis jaringan sosial dapat digunakan untuk memahami struktur hubungan dalam sebuah sistem sosial dan bagaimana struktur tersebut mempengaruhi perilaku individu atau kelompok. Hal ini dapat membantu para peneliti dan pengambil keputusan untuk memprediksi perilaku dan mengembangkan strategi untuk mengatasi masalah sosial atau bisnis.

METODOLOGI

Data yang kami gunakan diambil dari Kaggle yaitu berupa data Organisasi Github. di dalam data tersebut berisi daftar kontributor dari repository organisasi teratas yang ada di Github. kemudian dari data - data tersebut kami melakukan beberapa algoritma untuk memproses data tersebut :

1. Page Rank

PageRank merupakan algoritma yang dikembangkan oleh Google untuk menentukan relevansi halaman web dalam mesin pencari. Algoritma ini mengukur pentingnya suatu halaman web dengan menghitung jumlah dan kekuatan tautan masuk (inbound links) yang terkait dengan halaman tersebut. Semakin banyak dan semakin kuat tautan masuk yang terkait dengan suatu halaman, maka semakin tinggi peringkat PageRank yang akan diperoleh halaman tersebut. Peringkat PageRank ini kemudian digunakan oleh mesin pencari untuk menentukan urutan hasil pencarian yang ditampilkan kepada pengguna.

2. Betweenness centrality

Betweenness centrality merupakan sebuah algoritma untuk mengukur seberapa penting sebuah node dalam jaringan tersebut. Node yang memiliki nilai betweenness centrality yang tinggi akan menunjukkan bahwa node tersebut sering menjadi jalur lalu lintas antar node lain dalam jaringan. Algoritma untuk menghitung betweenness centrality dapat diterapkan dengan beberapa cara, salah satunya adalah dengan menggunakan rumus berikut:

$$\text{betweenness centrality}(\text{node}) = \sum_{(s,t) \in V} (\delta_{st}(\text{node})) / \sum_{(s,t) \in V} (\delta_{st})$$

Keterangan :

- node adalah node yang ingin dihitung nilai betweenness centrality-nya
- V adalah kumpulan semua node dalam jaringan
- δ_{st} adalah jumlah jalur terpendek antara node s dan node t
- $\delta_{st}(\text{node})$ adalah jumlah jalur terpendek antara node s dan node t yang melewati node node

Pada rumus di atas, nilai betweenness centrality suatu node dihitung dengan menjumlahkan proporsi jalur terpendek antara semua pasangan node yang melewati node tersebut, kemudian dibagi dengan jumlah semua jalur terpendek antara semua pasangan node. Node dengan nilai betweenness centrality yang tinggi menunjukkan bahwa ia sering melewati jalur terpendek antara node-node lain dalam jaringan.

3. Community Detection

Community detection atau deteksi komunitas adalah salah satu metode yang digunakan dalam analisis jaringan untuk mengidentifikasi subjaringan atau komunitas dalam jaringan. Metode ini bertujuan untuk menemukan struktur dasar jaringan dan pola-pola khusus dalam jaringan, seperti kelompok-kelompok individu yang saling terkait atau topik-topik yang sering dibicarakan dalam jaringan.

Terdapat berbagai macam algoritma yang dapat digunakan dalam community detection, salah satunya adalah algoritma modularity maximization. Algoritma ini bekerja dengan cara mencari komunitas yang memiliki nilai modularity tertinggi, dimana nilai modularity mengukur seberapa baik suatu komunitas terpisah dari komunitas lain dalam jaringan. Algoritma ini menggunakan rumus berikut untuk menghitung nilai modularity:

$$Q = \sum_{(i,j) \in V} [A_{ij} - (k_i * k_j) / (2 * m)] * \delta(c_i, c_j)$$

Keterangan :

- Q adalah nilai modularity
- A_{ij} adalah nilai adjacency matrix untuk node i dan node j, dimana $A_{ij} = 1$ jika node i dan node j terhubung, dan $A_{ij} = 0$ jika tidak terhubung
- k_i adalah jumlah tetangga dari node i
- k_j adalah jumlah tetangga dari node j
- m adalah jumlah semua tautan dalam jaringan
- $\delta(c_i, c_j)$ adalah fungsi Kronecker delta, dimana $\delta(c_i, c_j) = 1$ jika node i dan node j tergabung dalam satu komunitas, dan $\delta(c_i, c_j) = 0$ jika tidak tergabung dalam satu komunitas

Dengan menggunakan rumus di atas, algoritma modularity maximization akan mencari kombinasi komunitas yang memberikan nilai modularity tertinggi. Setelah komunitas ditemukan, algoritma akan mengulang prosesnya dengan memisahkan komunitas yang telah ditemukan menjadi sub-komunitas yang lebih kecil, dan mencari kombinasi sub-komunitas yang memberikan nilai modularity tertinggi. Proses ini dilakukan secara iteratif hingga tidak ditemukan lagi sub-komunitas yang dapat dibentuk dari komunitas yang telah ditemukan sebelumnya.

HASIL DAN ANALISIS

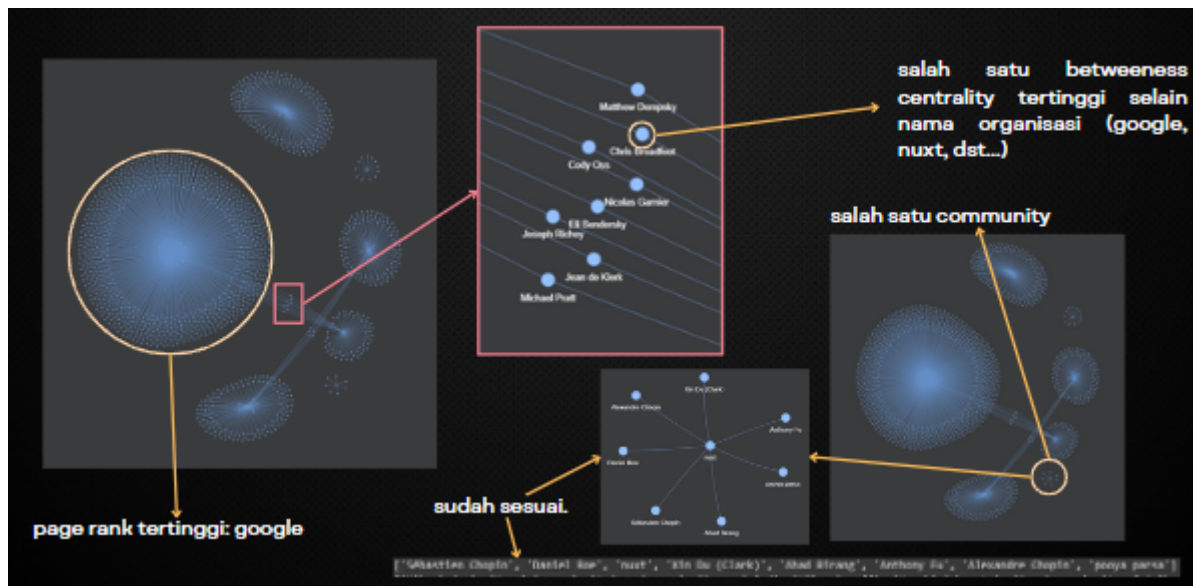
Pada dataset yang kami ambil (dari kaggle¹) terdapat 2 buah kolom yaitu organisasi dan member (di dalam laporan ini kami menyebutkan member juga sebagai kontributor), organisasi berisikan nama organisasi yang terdapat di github, untuk member merupakan kontributor yang berperan dalam organisasi tersebut. organisasi yang ada di dalam dataset kurang lebih sebagai berikut :

- | | |
|------------------|----------------|
| 1. Ethereum | 21. reduxjs |
| 2. Pytorch | 22. NVIDIA |
| 3. udacity | 23. redis |
| 4. firebase | 24. rails |
| 5. flutter | 25. grafana |
| 6. deepmind | 26. jquery |
| 7. reactjs | 27. ansible |
| 8. huggingface | 28. graphql |
| 9. google | 29. babel |
| 10. docker | 30. atom |
| 11. apache | 31. prometheus |
| 12. vuejs | 32. django |
| 13. freeCodeCamp | 33. opencv |
| 14. airbnb | 34. mongodb |
| 15. golang | 35. python |
| 16. nodejs | 36. bitcoin |
| 17. elastic | 37. plotly |
| 18. laravel | 38. serverless |
| 19. aws | 39. nuxt |
| 20. Azure | 40. facebook |
| | 41. microsoft |

dari daftar organisasi tersebut kami hanya mengambil beberapa data organisasi yang menarik perhatian kami seperti facebook, reactjs, NVIDIA, nuxt, aws, google, dan golang.

¹ <https://www.kaggle.com/datasets/anshulmehtakaggl/github-organizations-social-network-analysis>

kemudian dari beberapa data organisasi yang menarik perhatian kami, kami lakukan proses PageRank, Betweenness Centrality, Community Detection. berikut hasil visualisasi data dari proses tersebut :



dari gambar tersebut kita dapat menyimpulkan bahwa hasil dari pagerank, google menjadi organisasi yang memiliki kontributor terbanyak. kemudian, dengan menggunakan metode betweenness centrality kita melihat terdapat node tertinggi setelah organisasi, yaitu terdapat kontributor yang menjadi penghubung di antara 2 organisasi. kita juga bisa mengasumsikan bahwa kontributor tersebut hadir dalam 2 buah organisasi atau lebih dalam proses pengembangan di dalam organisasi tersebut.

dengan community detection kita bisa mengelompokkan antara node - node dalam jaringan untuk mencari kemiripan dalam topik atau aktivitas, dan mengelompokkannya ke dalam komunitas yang berbeda. sehingga kita bisa melihat data visual tersebut sudah dikelompokkan sesuai dengan organisasi masing - masing.

KESIMPULAN DAN SARAN

Dari hasil analisa kami dengan batasan data yang digunakan, kami menyimpulkan bahwa dengan menggunakan metode pagerank, google menjadi organisasi tertinggi. kita juga dapat melihat bahwa kebanyakan “member” atau kontributor hanya berfokus pada satu organisasi saja. kemudian dengan metode betweenness centrality ditemukan beberapa "member" yang berkontribusi terhadap 2 atau lebih organisasi. hasil community detection berpatok pada organisasi yang ada. Dengan penerapan fungsi, data yang diambil dapat dipilah sebelum melakukan visualisasi untuk menghasilkan graf yang lebih menarik.

Saran untuk kedepannya, data yang diproses dapat digunakan secara menyeluruh, sehingga hasil visualisasi lebih menarik. Analisa lebih lanjut dapat dilakukan menggunakan berbagai fungsi graf lain.

REFERENSI

- [1] Rogers, Ian. "The Google Pagerank algorithm and how it works." (2002).
- [2] Freeman, Linton. "The development of social network analysis." *A Study in the Sociology of Science* 1.687 (2004): 159-167.
- [3] Li, Yeqing, et al. "Large-scale multi-view spectral clustering via bipartite graph." *Twenty-ninth AAAI conference on artificial intelligence*. 2015.
- [4] Brandes, Ulrik. "A faster algorithm for betweenness centrality." *Journal of mathematical sociology* 25.2 (2001): 163-177.
- [5] Lancichinetti, Andrea, and Santo Fortunato. "Community detection algorithms: a comparative analysis." *Physical review E* 80.5 (2009): 056117.

LAMPIRAN

program :

<https://colab.research.google.com/drive/1C-ZqWyE2Mn9LcHng6X9VVbw145MMghqV?usp=sharing>

dataset -

<https://www.kaggle.com/datasets/anshulmehtakaggl/github-organizations-social-network-analysis>