

Social Network Analysis

INFORMATION RETRIEVAL

The Team



Kurniawan

205314148



**Caesario
Dito**

205314159



Ong Gabriel

205314111

Overview

1

Pendahuluan

penjelasan, tujuan,
manfaat

2

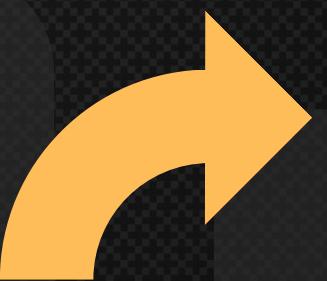
Metodologi

data, metode, tata cara

3

Hasil & Analisis

rekap hasil, keterbatasan,
kesimpulan





Pendahuluan

Pendahuluan

Social network analysis (SNA) adalah bidang studi yang berfokus pada hubungan antara perorangan atau organisasi

tujuannya : untuk memahami / memprediksi bagaimana hubungan antar individu dapat mempengaruhi perilaku individu



Manfaat SNA



Sosiologi

untuk mengidentifikasi pola hubungan yang mempengaruhi suatu kelompok



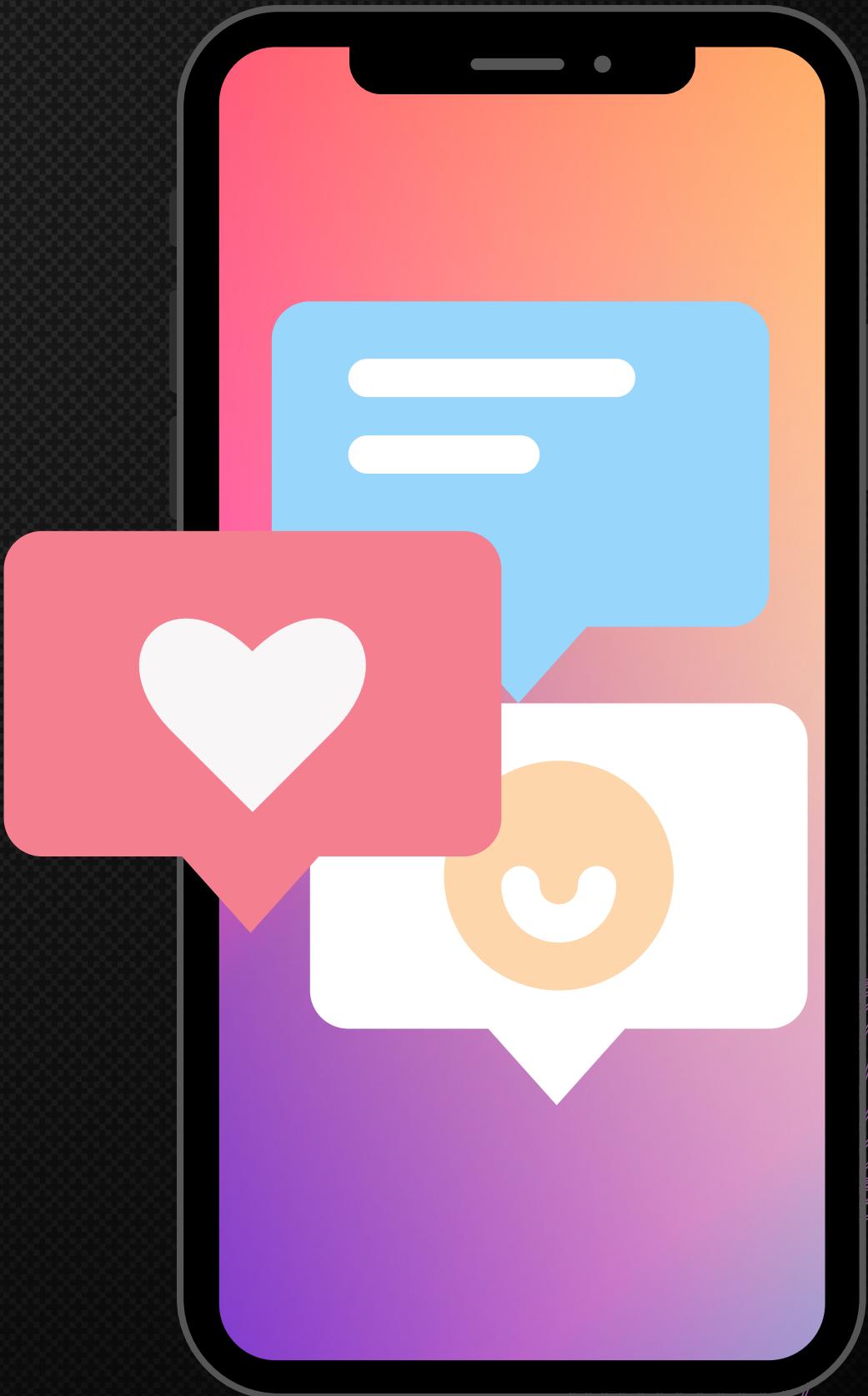
Bisnis

mengidentifikasi pola hubungan yang mempengaruhi keberhasilan bisnis



Umum

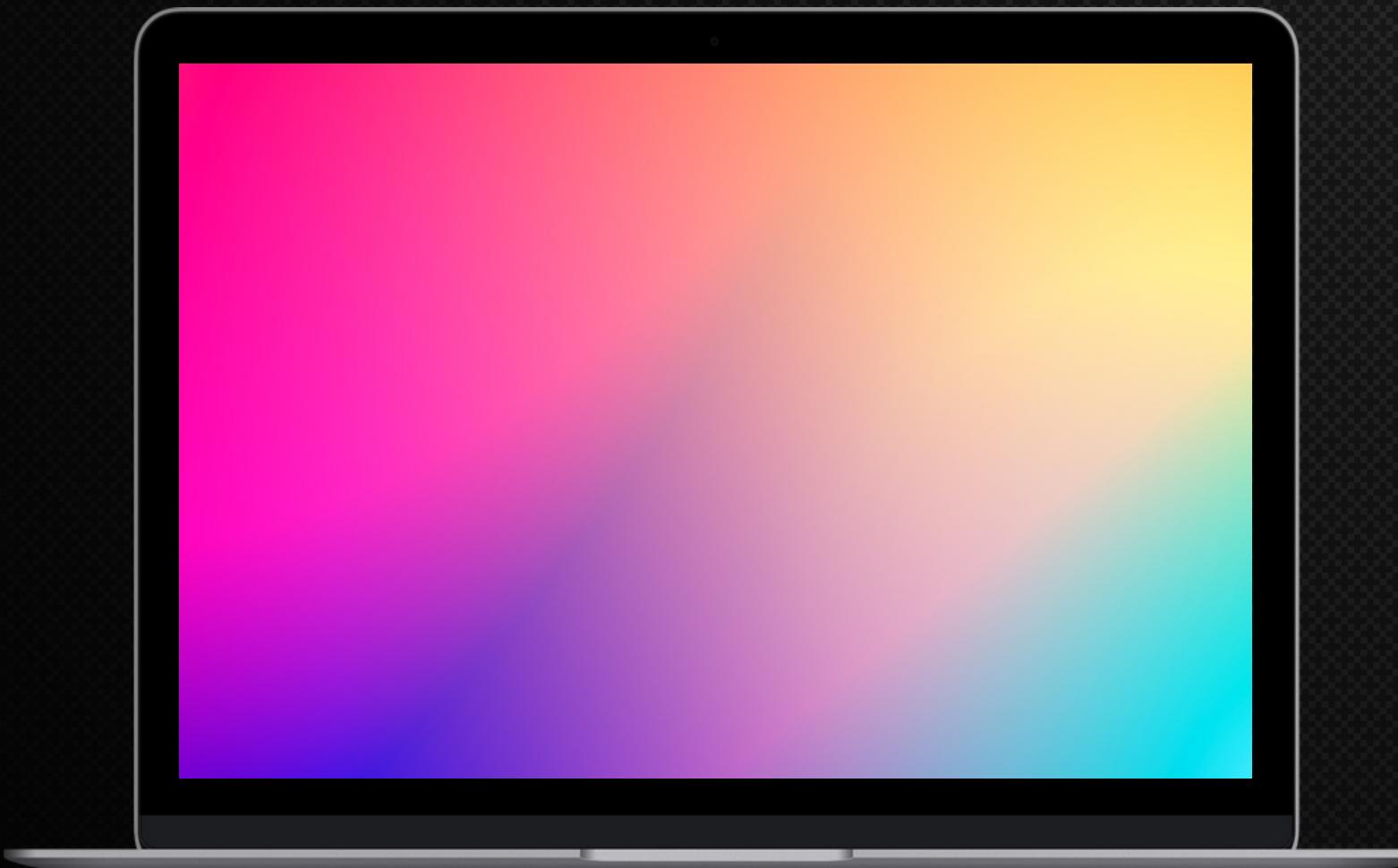
memahami struktur hubungan dalam suatu pola sistem sosial





Metodologi

Data



Data kami ambil dari kaggle berupa data organisasi github.

lebih spesifiknya data mengenai daftar kontributor dari repository organisasi teratas yang ada di Github.

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

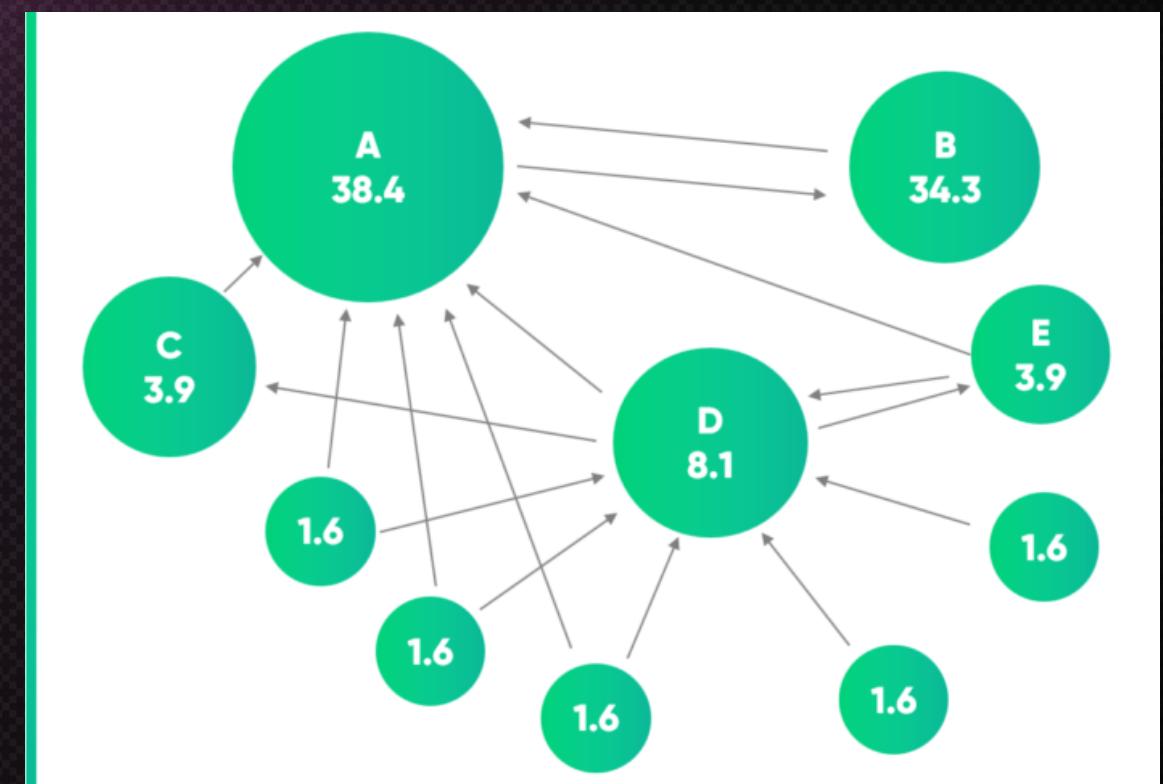
metode

Untuk menganalisis, metode yang digunakan adalah sebagai berikut:

- > Page Rank
- > Betweenness Centrality
- > community detection

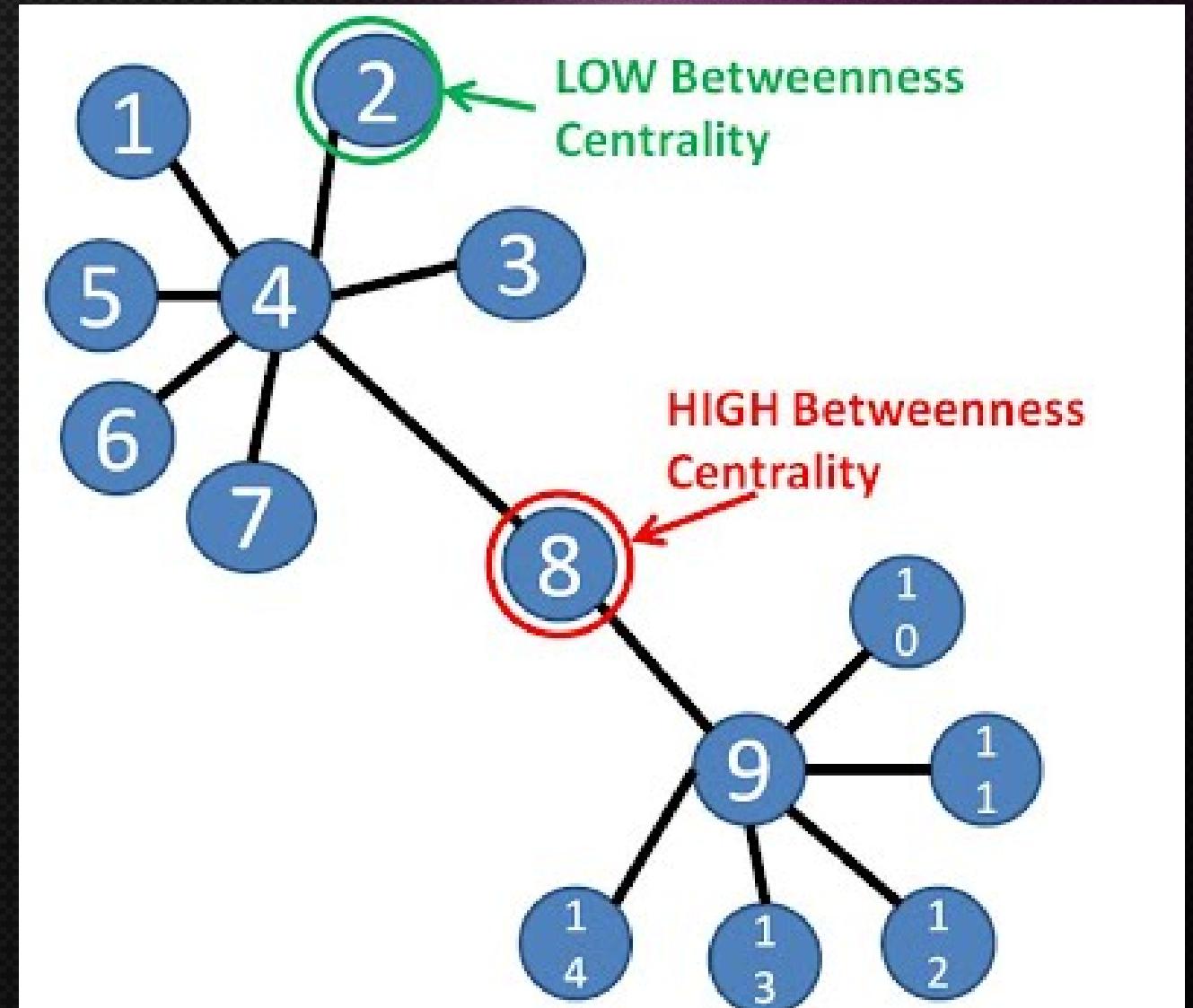
Page Rank

- > menentukan relevansi halaman web dalam mesin pencarian
- > mengukur pentingnya suatu halaman web dengan menghitung jumlah tautan masuk
- > semakin tinggi page rank maka urutan pada mesin pencarian juga semakin tinggi



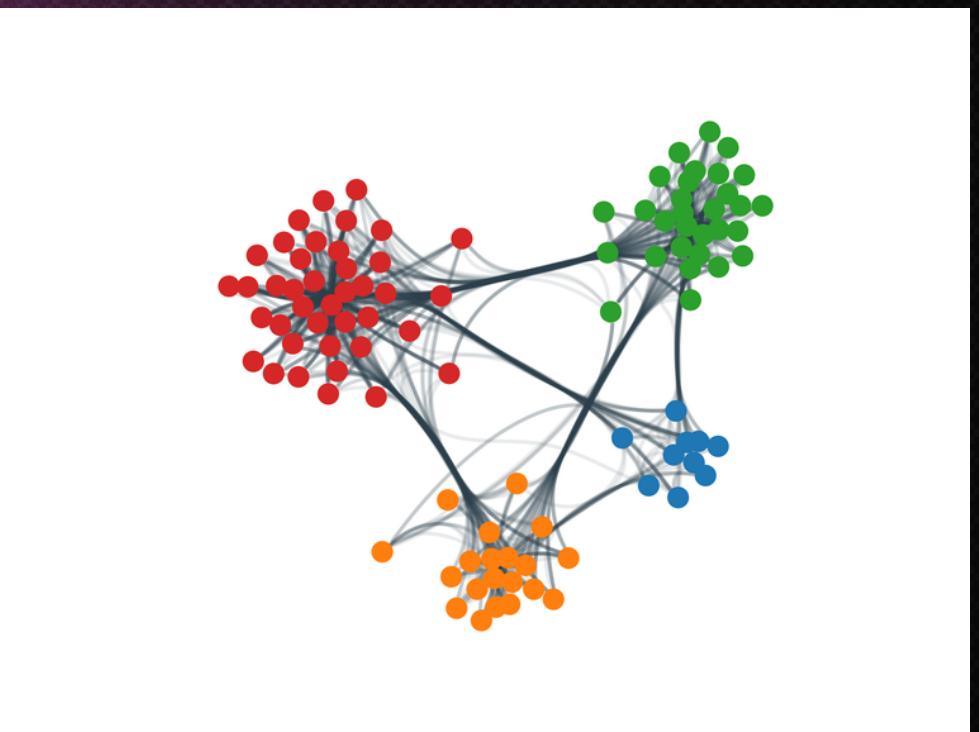
Betweenness Centrality

- > untuk mengidentifikasi individu yang memiliki peran penting dalam jaringan
- > mengukur banyak suatu node, berperan sebagai jalur transisi antara node-node lain dalam jaringan
- > Node yang memiliki betweeness centrality yang tinggi maka node tersebut memiliki akses yang luas ke informasi jaringan dan merupakan penghubung node lain



Community Detection

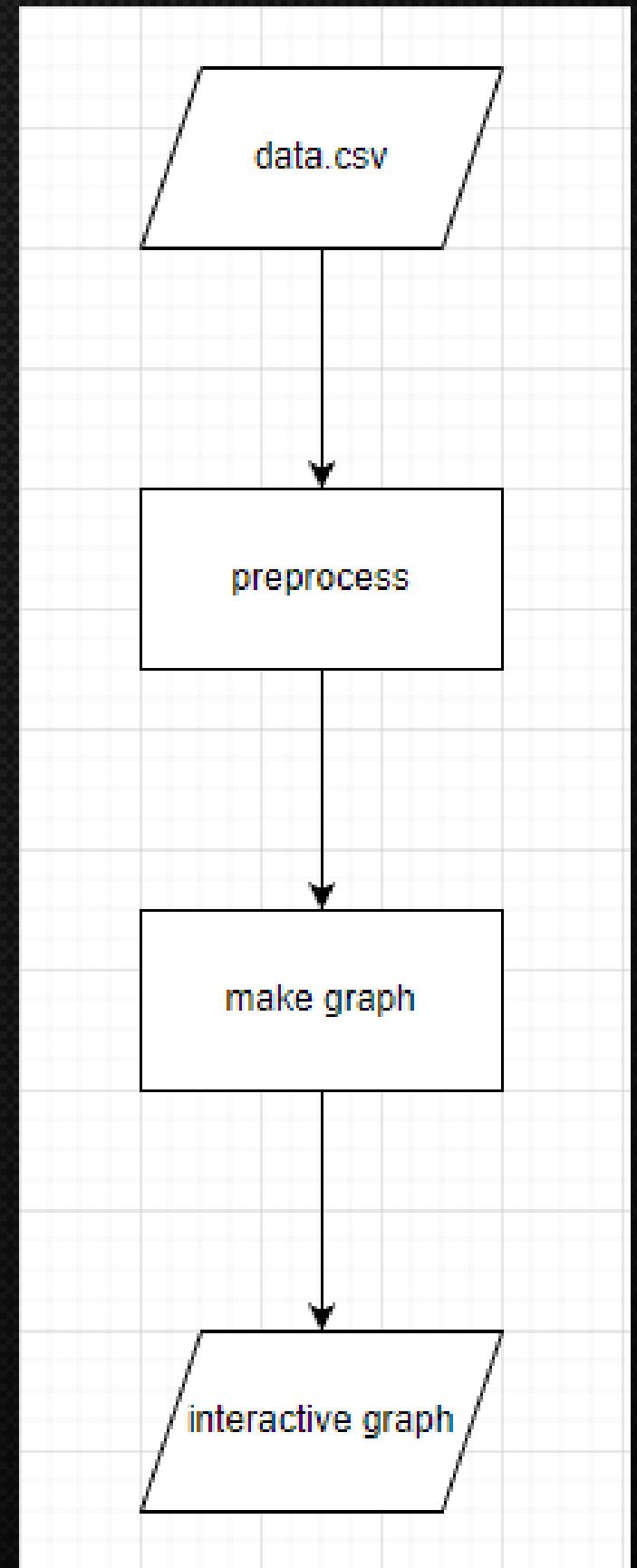
- > mengidentifikasi subjaringan / komunitas dalam jaringan
- > untuk menemukan struktur dasar jaringan dan pola - pola khusus seperti kelompok individu yang saling terkait dan topik topik yang sering dibicarakan
- > algoritma modularity maximization adalah salah satu contoh algoritma community detection



Tata cara

Adapun prosesnya dapat tergambar di diagram flowchart disamping.

Proses preprocess merupakan proses untuk membersihkan data. Pada kasus ini, terdapat sedikit data yang tidak lengkap, sehingga data tersebut perlu di-drop.



Tata cara - data awal

	organisation	member
0	ethereum	Oleh Aldekein
2	ethereum	Frank Szendzielarz
3	ethereum	Kolby Moroz Liebl
4	ethereum	Marius van der Wijden
5	ethereum	Mihai Alisie
...
10592	microsoft	Chris Colborne
10593	microsoft	Zachary Danz
10594	microsoft	zsy
10595	microsoft	Zuhair Parvez
10597	microsoft	Yiwen Zhu

9652 rows x 2 columns

Data di samping ini mengandung 2 kolom yaitu "organisation" dan "member".

Data ini menceritakan tentang suatu organisasi dan developer yang berkontribusi di organisasi tersebut.

Tata cara - preprocess data

	Organisation	member
1	ethereum	None
6	ethereum	None
17	ethereum	None
21	ethereum	None
27	ethereum	None
...
10540	microsoft	None
10551	microsoft	None
10557	microsoft	None
10568	microsoft	None
10596	microsoft	None

946 rows x 2 columns

Di dalam data tersebut terdapat "member" dengan nilai "None". Hal ini dapat menghasilkan graf yang tidak sesuai pada output.

Data ini akan di-drop.

Tata cara - make graph

seluruh data

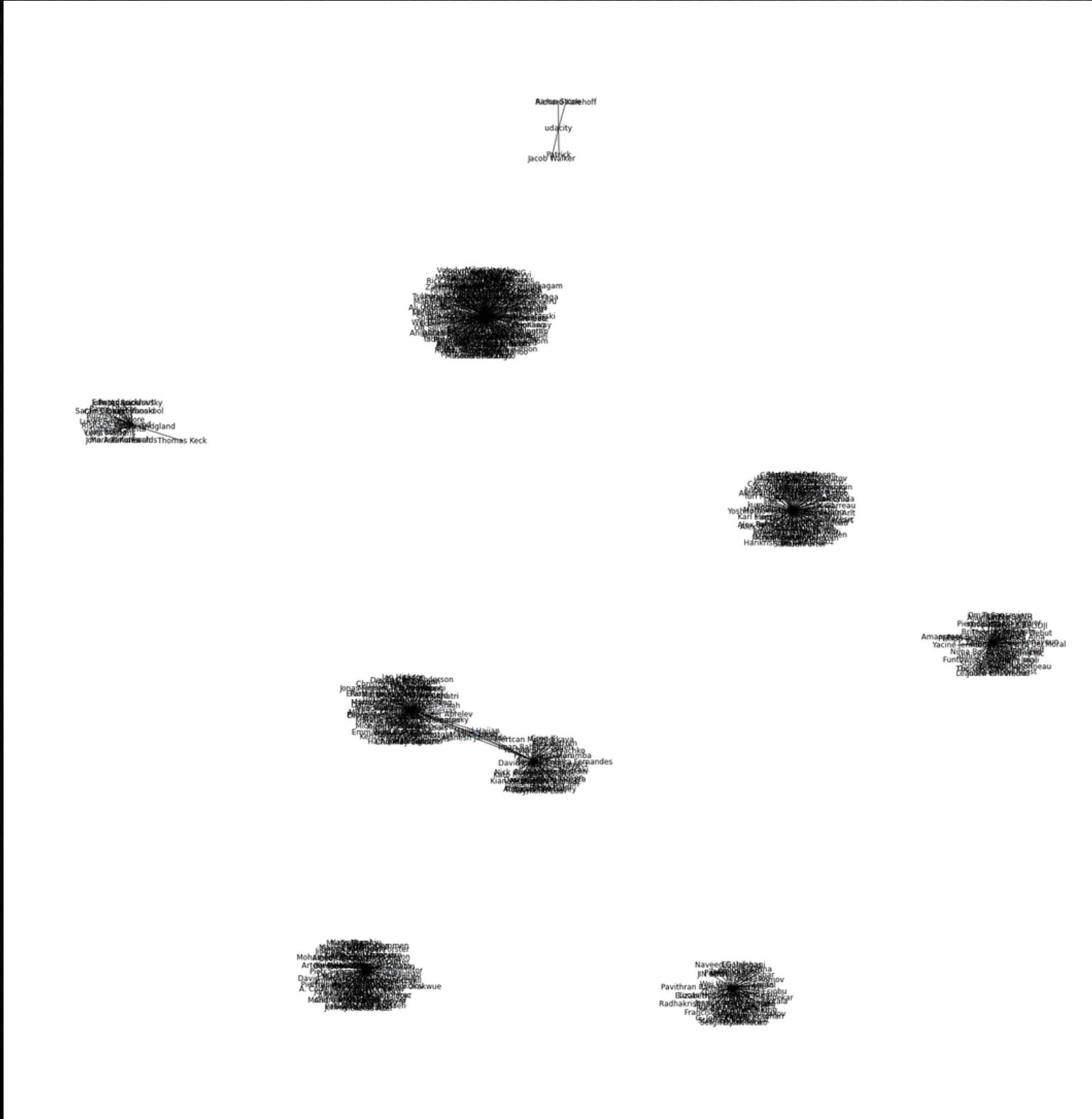


Dari data yang sudah di drop/preprocess, akan dihasilkan graf seperti di samping ini.

Hal ini dihasilkan dengan library python yang bernama networkx.

Tata cara - make graph - 2

500 data pertama

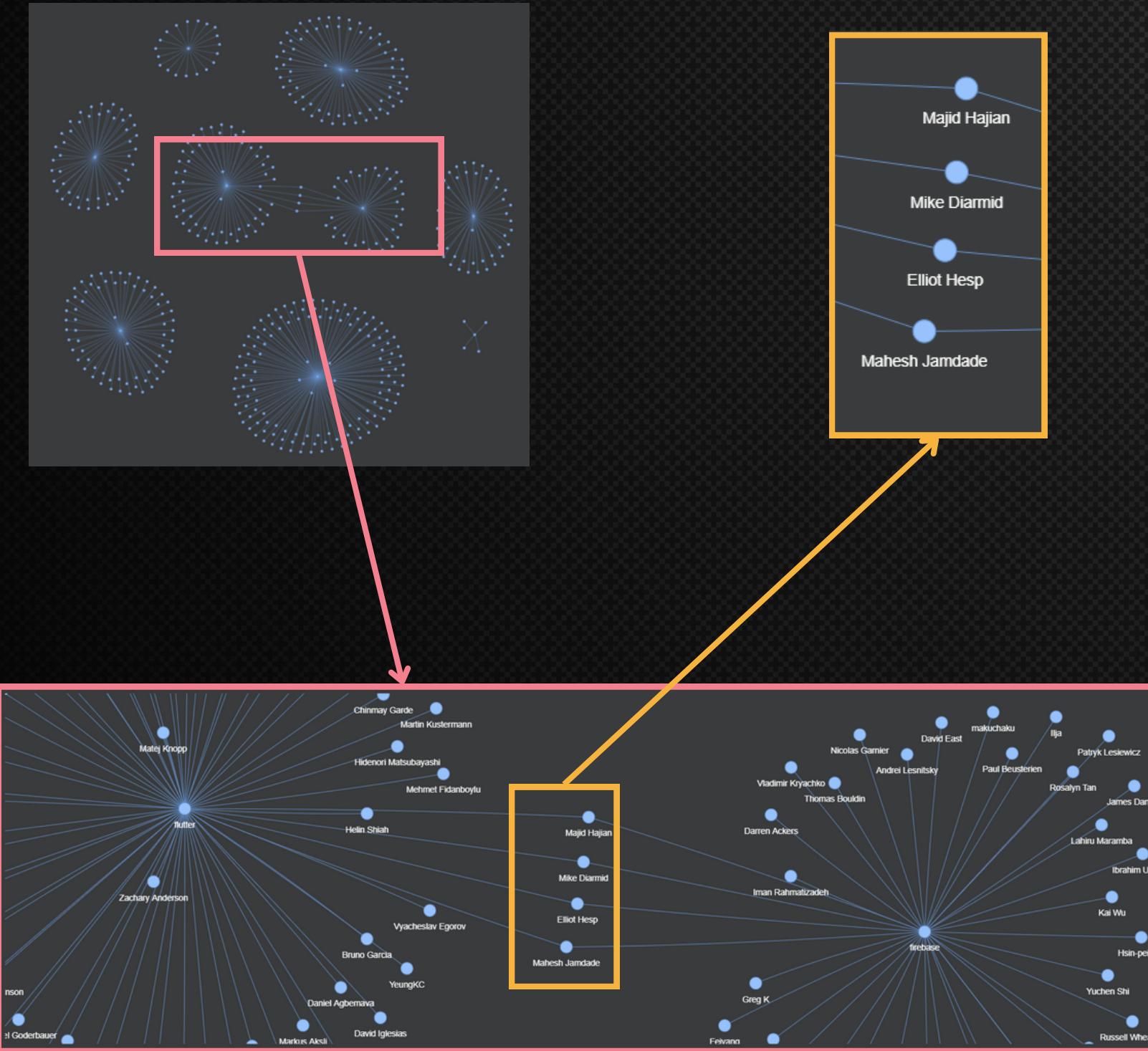


Untuk visualisasi yang lebih jelas,
diputuskan untuk memperkecil
data yang digunakan.

Data yang diambil adalah 500 data
pertama pada tabel.

Dari 500 data tersebut, dihasilkan
graf sebagai berikut.

Tata cara - interactive graph



Dari gambar grafik sebelumnya, belum terlihat dengan jelas nama dari masing - masing node yang menunjukkan keterkaitan satu dengan yang lainnya.

Oleh karena itu, dibuat grafik interaktif menggunakan library python pyvis.



Hasil & Analisis

Pemilihan Data

list organisasi

```
array(['ethereum', 'pytorch', 'udacity', 'firebase', 'flutter',
       'deepmind', 'reactjs', 'huggingface', 'google', 'docker', 'apache',
       'vuejs', 'freeCodeCamp', 'airbnb', 'golang', 'nodejs', 'elastic',
       'laravel', 'aws', 'Azure', 'reduxjs', 'NVIDIA', 'redis', 'rails',
       'grafana', 'jquery', 'ansible', 'graphql', 'babel', 'atom',
       'prometheus', 'django', 'opencv', 'mongodb', 'python', 'bitcoin',
       'plotly', 'serverless', 'nuxt', 'facebook', 'microsoft'],
      dtype=object)
```

```
interested = ['facebook', 'reactjs', 'NVIDIA', 'nuxt', 'aws', 'google', 'golang']
```

	index	Organisation	member
0	273	reactjs	Imed Jaber
1	274	reactjs	Aissaoui Ahmed
2	275	reactjs	Greg Myers
3	276	reactjs	Bunhouth
4	277	reactjs	Bunlong VAN
...
1289	6231	facebook	Andres Suarez
1290	6232	facebook	Jetsada Machom
1291	6233	facebook	Ziyad Bazel
1292	6234	facebook	Paul O'Shannessy
1293	6235	facebook	Gerard Rovira

1294 rows × 3 columns

Untuk melakukan analisis, diputuskan untuk mengambil organisasi yang menurut kami menarik.

Gambar disamping menunjukkan organisasi - organisasi yang ada dan organisasi yang dipilih untuk dianalisis.

Fungsi

Page Rank - Top 5

```
[('google', 0.29349016656727717),  
 ('aws', 0.06215504462820502),  
 ('reactjs', 0.050200466921470045),  
 ('facebook', 0.0388260886653768),  
 ('golang', 0.021203989781147892)]
```

Betweenness Centrality

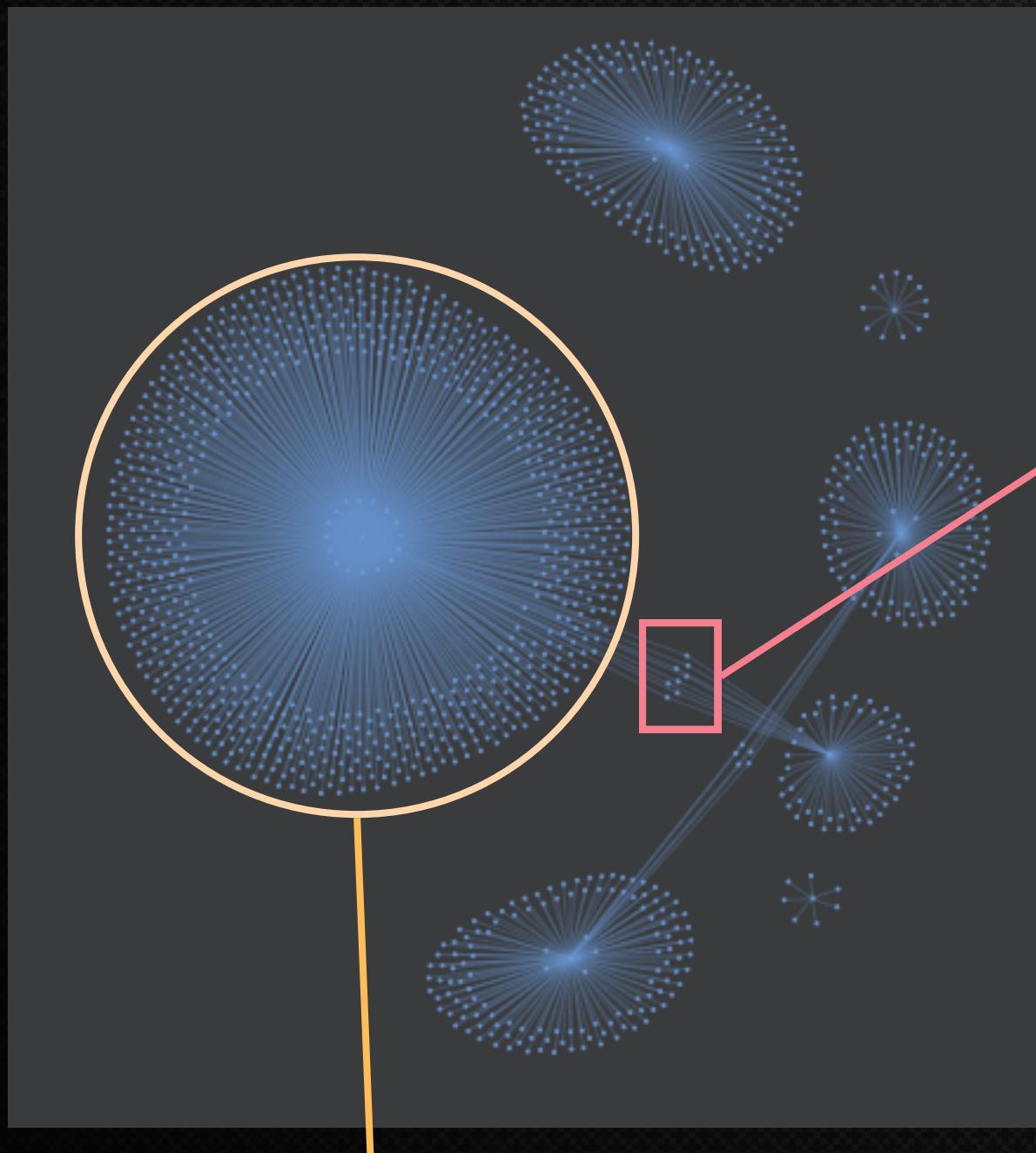
```
[('google', 0.4389740206225432),  
 ('golang', 0.052077178049184276),  
 ('reactjs', 0.027970819572685827),  
 ('facebook', 0.0233631928810778),  
 ('aws', 0.016951425971643702),  
 ('Chris Broadfoot', 0.006364488285172578),  
 ('Cody Oss', 0.006364488285172578),  
 ('Eli Bendersky', 0.006364488285172578),  
 ('Jean de Klerk', 0.006364488285172578),  
 ('Joseph Richey', 0.006364488285172578)]
```

Community -> 7

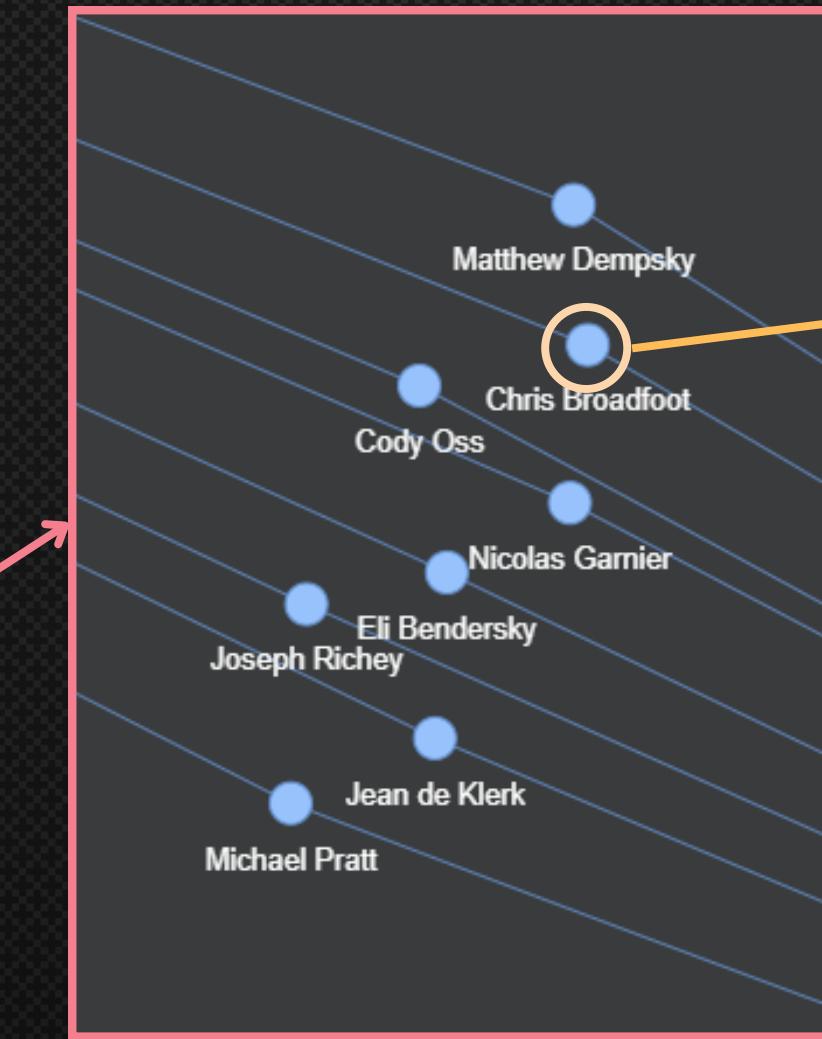
```
['Anirudh Bangalore Srinivas', 'Sureshkumar Selvanayagam', 'Keyan Zhang',  
 ['Luiz Antonio', 'Joey Parrish', 'Adam Rice', 'Brandon Weeks', 'Tom Hennig',  
 ['Jean de Klerk', 'Joseph Poirier', 'Paul Jolly', 'mattn', 'Todd Neal', 'sa',  
 ['Kyle Knapp', 'Daniel Budris', 'Jackson West', 'David LaBissoniere', 'Brad',  
 ['Oleksii Kuchaiiev', 'Felix Abecassis', 'Vinay Deshpande', 'Alex Aizman', 'A',  
 ['Sébastien Chopin', 'Daniel Roe', 'nuxt', 'Xin Du (Clark)', 'Ahad Birang',  
 ['Mike Dodge', 'Naomi Reeves', 'Andres Suarez', 'Jason Fried', 'Niles Rogof
```

Melihat hasil fungsi menggunakan graf

graf organisasi yang dipilih



page rank tertinggi: google

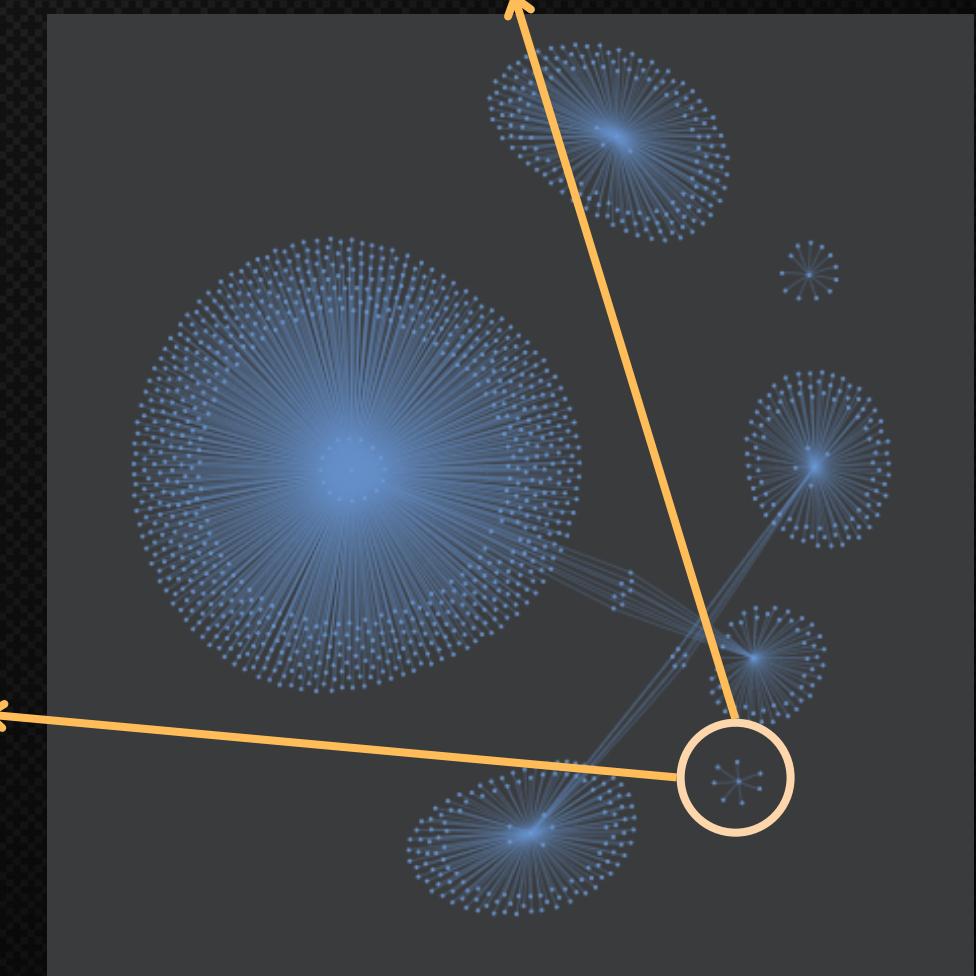


sudah sesuai.

```
[ 'Sébastien Chopin', 'Daniel Roe', 'nuxt', 'Xin Du (Clark)', 'Ahad Birang', 'Anthony Fu', 'Alexandre Chopin', 'pooya parsa' ]
```

salah satu betweenness centrality tertinggi selain nama organisasi (google, nuxt, dst...)

salah satu community



Analisis Fungsi Betweeness Centrality dengan seluruh data

Hasil perhitungan - Top 8

```
('Daniel', 0.03972454290640497),  
('Ryan Wilson', 0.02921219607488707),  
('Glaucia Lemos', 0.023724403050407284),  
('Bing Hu', 0.02354607111301818),  
('Seth', 0.02354607111301818),  
("Yegor", 0.02354607111301818),  
('Dr Christian Geuer-Pollmann', 0.0227355193),  
('Sergey Grebnov', 0.022735519330738585),
```

Pencarian member dengan nama "Ryan Wilson"

Organisation	member
173	firebase Ryan Wilson
1171	google Ryan Wilson
6547	microsoft Ryan Wilson

Dilakukan perhitungan betweeness centrality dengan seluruh data yang ada.

Dari hasil di samping ini, dapat dilihat bahwa Ryan Wilson berkontribusi terhadap 3 organisasi yang berbeda.

Hasil Diskusi

- Organisasi tertinggi terletak pada google yang memiliki page rank tertinggi (29%).
- Terdapat beberapa "member" yang berkontribusi terhadap 2 organisasi.
- Hasil community detection berpatok pada organisasi yang ada.
- Dengan penerapan fungsi, data yang diambil dapat di-filter/pilah sebelum melakukan visualiasi untuk menghasilkan graf yang lebih menarik (contoh: mempunyai banyak jaringan).

Keterbatasan

- Dataset yang digunakan terbatas dan hanya berfokus pada "organisasi" saja. Sehingga tidak dapat melakukan analisis terhadap "member" dari organisasi tersebut. Hal ini mengakibatkan cabang yang dihasilkan di graf menjadi sedikit.
- Node dan Edges sebanding banyaknya, sehingga membuat proses komputasi memakan waktu yang lama dan tidak memungkinkan untuk diproses dengan device yang tersedia.
- Device yang digunakan terbatas, maka proses visualisasi mengalami gangguan (frame drop, low fps).

Kesimpulan

- Data yang diambil hanya sebagian.
- Dari data yang diambil, banyak pekerja yang fokus pada satu organisasi saja. Ditunjukkan dari banyak node tunggal/daun/leaf di dalam graf.
- Dari fungsi yang digunakan, hasilnya valid (dilihat dari graf).

Saran

- Data yang digunakan tidak menyeluruh, hal ini dapat dikembangkan dengan melakukan filtering/pemilihan data dengan fungsi (betweeness centrality, page rank, dll) untuk memilih data yang dapat memvisualisasikan social network yang lebih kompleks (banyak jaringannya).
- Pengambilan data untuk visualisasi dapat dilakukan secara menyeluruh.
- 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.



Terima kasih