

Deteksi Komunitas dalam Perbincangan "Techbros" di Twitter Menggunakan Algoritma Girvan-Newman

1st Hidayat Taufiqur Rahmah Achmad
S1 Informatika
Telkom University
Bandung, Indonesia
hidayattaufiqur@student.telkomuniversity.ac.id

Abstrak—Twitter merupakan salah satu media sosial paling populer sebagai sarana untuk berbagi informasi dan berinteraksi dengan orang lain. Sejak diluncurkan pada tahun 2006, Twitter telah menjadi platform microblogging online yang sangat diminati. Data interaksi pengguna di Twitter dapat dianalisis menggunakan metode closeness centrality analysis untuk mengidentifikasi aktor-aktor utama dalam suatu perbincangan. Penelitian ini dilakukan untuk menganalisis perbincangan seputar istilah "techbros" di Twitter. Istilah ini merujuk pada komunitas para pekerja di bidang teknologi, khususnya di kawasan Silicon Valley, yang sering dikritik atas gaya hidup dan budaya kerja mereka. Dataset yang digunakan dalam penelitian ini diperoleh dengan melakukan crawling terhadap tweet-tweet yang mengandung kata kunci "techbros" selama periode 01-01-2024 hingga 03-04-2024. Analisis ini menggunakan algoritma Girvan-Newman. Total dataset yang dianalisis sebanyak 328 tweet.

Keywords—Kata kunci: *techbros*, *Twitter*, *centrality analysis*, *closeness centrality*, *media sosial*, *analisis jaringan sosial*.

I. INTRODUCTION

Fenomena "techbros" adalah sebuah slang yang sedang marak digunakan di lingkungan pekerja teknologi, baik di Indonesia ataupun secara global. Belakangan ini, makin banyak pelaku di industri teknologi Indonesia, khususnya di bidang Software Engineering yang menggunakan slang techbro tersebut, baik dalam konteks positif, sarkas, ataupun negatif. Hal ini menarik untuk diteliti dikarenakan fenomena ini dapat menjadi representasi sebuah subkultur baru di dalam industri teknologi [1].

Media sosial, khususnya Twitter, telah menjadi platform utama bagi pengguna untuk mengekspresikan pandangan dan dukungan mereka, termasuk dalam diskusi tentang berbagai topik terkait teknologi dan budaya kerja dalam industri teknologi. Menurut Pew Research Center, 70% pengguna Twitter di AS menggunakan platform ini untuk mengikuti berita terbaru dan terlibat dalam diskusi sosial dan politik [2]. Penelitian lainnya menunjukkan bahwa Twitter dapat memperluas jangkauan komunikasi ilmiah dan mempengaruhi interaksi dengan konten akademis [3]. Selain itu, artikel di *Inquiries Journal* membahas bagaimana media sosial telah mengubah cara berita disebarkan dan dikonsumsi, yang relevan untuk memahami pembentukan komunitas di sekitar topik seperti "techbros" [4].

Penelitian ini menggunakan algoritma Girvan-Newman untuk mendeteksi komunitas dalam jaringan sosial tersebut. Algoritma Girvan-Newman bekerja dengan mengidentifikasi edge yang menjadi penghubung antara kelompok dalam jaringan, dan kemudian menghapus edge-edge tersebut secara iteratif [5]. Proses ini menghasilkan pemisahan jaringan menjadi kelompok-kelompok yang lebih

kecil atau komunitas, memungkinkan peneliti untuk mengidentifikasi dan menganalisis struktur komunitas dalam jaringan sosial yang membahas fenomena techbros.

KAJIAN TEORI

A. Algoritma Girvan-Newman

Algoritma Girvan-Newman adalah metode yang digunakan untuk mendeteksi struktur komunitas dalam jaringan. Algoritma ini bekerja dengan menghapus iteratif edge yang memiliki nilai betweenness centrality tertinggi, yang mengukur jumlah jalur terpendek yang melewati edge tersebut. Dengan menghapus edge dengan betweenness tertinggi, jaringan secara bertahap terbagi menjadi komunitas-komunitas yang lebih kecil [6].

Betweenness centrality adalah ukuran yang menilai pentingnya edge dalam jaringan berdasarkan seberapa sering edge tersebut muncul dalam jalur terpendek antara pasangan node. Freeman (1977) pertama kali memperkenalkan konsep betweenness untuk vertex, yang kemudian diperluas ke edge oleh Anthonisse (1971) dan dikembangkan lebih lanjut oleh Girvan dan Newman untuk digunakan dalam deteksi komunitas [7].

B. Crawling Data

Crawling adalah sebuah metode pengambilan data yang efektif, yang sering digunakan dengan memanfaatkan Search API pada platform Twitter untuk mengumpulkan informasi secara sistematis [8]. Proses ini melibatkan penggunaan bot atau script yang otomatis untuk mengakses dan mengunduh data dari Twitter, yang kemudian dapat digunakan untuk berbagai analisis data, seperti sentiment analysis, trend detection, atau community detection. Melalui Search API, pengguna bisa meminta kumpulan tweet yang memenuhi kriteria tertentu, seperti kata kunci, hashtag, atau berdasarkan lokasi geografis, yang memungkinkan peneliti untuk mengumpulkan dataset yang relevan dengan topik atau pertanyaan penelitian yang sedang dijajaki.[

DESAIN KERANGKA

Desain kerangka penelitian ini bertujuan untuk mengidentifikasi aktor kunci dan pola komunikasi dalam perbincangan tentang "techbros" di Twitter. Proses awal melibatkan pengumpulan data, di mana atribut dari akun Twitter dikumpulkan dan dimasukkan ke dalam dataset. Setelah itu, dilakukan preprocessing untuk mempersiapkan

data yang akan digunakan. Tahapan-tahapan selanjutnya melibatkan analisis sentralitas untuk mengidentifikasi aktor kunci dan deteksi komunitas menggunakan algoritma Girvan-Newman [5,6].

1. Pengumpulan Data:

Data dikumpulkan menggunakan teknik crawling pada tweet yang mengandung kata kunci "techbros" dengan bantuan *library* tweet-harvest [10]. Dataset yang dikumpulkan mencakup kolom username dan username yang di-reply, yang menjadi dasar untuk membentuk graf interaksi.

2. Pra-pemrosesan Data:

Data diproses untuk menghapus duplikasi dan nilai null. Hanya kolom relevan (username dan in_reply_to_screen_name) yang digunakan dalam analisis ini.

3. Pembuatan Graf:

Graf dibuat menggunakan pustaka NetworkX, di mana node mewakili pengguna dan edge mewakili interaksi antara mereka.

4. Algoritma Deteksi Komunitas:

Algoritma Girvan-Newman diterapkan pada graf untuk mendeteksi komunitas. Algoritma ini secara iteratif menghapus edge dengan betweenness centrality tertinggi, sehingga memisahkan jaringan menjadi komunitas yang lebih kecil.

IMPLEMENTASI DAN ANALISIS

A. Implementasi

Implementasi dari desain kerangka ini melibatkan beberapa tahapan berikut:

1. Pengumpulan Data:

Data dikumpulkan dari media sosial Twitter menggunakan teknik crawling, dengan bantuan *library* Tweet Harvest selama periode 01-01-2024 hingga 03-04-2024, mencakup tweet yang mengandung kata kunci "techbros". Dataset yang dikumpulkan terdiri dari 774 cuitan. Cuitan-cuitan ini kemudian akan digunakan sebagai *edges* dalam jaringan yang merepresentasikan interaksi antar pengguna Twitter. Analisis ini bertujuan untuk memahami pola komunikasi dan keterlibatan pengguna Twitter terkait dengan topik "techbros," serta bagaimana istilah tersebut digunakan dalam percakapan di platform media sosial.

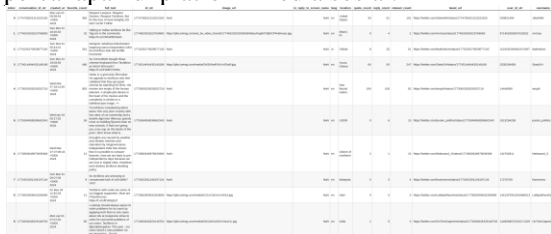


Fig 1. Hasil crawling data

2. Pra-pemrosesan Data:

Dataset yang diperoleh telah menjalani proses pra-pemrosesan untuk memudahkan analisis. Hanya kolom yang relevan, yaitu username dan in_reply_to_screen_name, yang dipilih untuk analisis lebih lanjut. Pemilihan kolom ini dilakukan karena mereka secara langsung mewakili interaksi antar pengguna, yang merupakan fokus utama dari analisis ini. Setelah itu, operasi *drop_duplicates* dilakukan untuk menghapus data duplikat, memastikan dataset bersih sebelum melanjutkan ke tahap analisis lebih lanjut. Selanjutnya, operasi *dropna* diterapkan untuk menghapus baris yang mengandung nilai kosong, memastikan integritas dataset. Setelah penerapan *drop_duplicates* dan *dropna*, dataset terdiri dari 328 tweet. Data ini akan diproses lebih lanjut dengan algoritma yang telah ditentukan untuk mencapai hasil yang optimal. Tujuan dari pra-pemrosesan ini adalah memastikan dataset bebas dari duplikasi dan nilai kosong, sehingga siap untuk analisis lanjutan yang dapat memberikan wawasan yang lebih akurat dan bermakna. Langkah-langkah ini penting untuk memastikan bahwa data yang digunakan dalam analisis valid dan representatif, memungkinkan pengambilan keputusan yang lebih baik berdasarkan hasil analisis yang dilakukan.

```
# Show how many rows of data are there after preprocessing
num_tweets = len(df)
print(f"Jumlah tweet dalam dataframe adalah {num_tweets}.")
```

Jumlah tweet dalam dataframe adalah 328.

Fig 2. Total data setelah pra-pemrosesan

3. Pembuatan Graf:

Graf dibuat menggunakan *library* NetworkX, di mana node mewakili pengguna dan edge mewakili interaksi antara mereka.

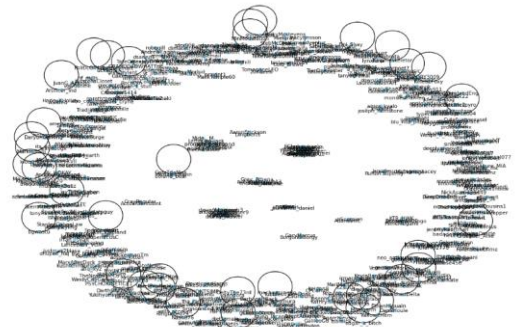


Fig 3. Graf NetworkX

4. Algoritma Deteksi Komunitas:

Algoritma Girvan-Newman diterapkan untuk mendeteksi komunitas dalam graf. Proses ini melibatkan penghitungan nilai betweenness untuk setiap edge dan penghapusan edge dengan nilai betweenness tertinggi secara iteratif hingga jaringan terbagi menjadi komunitas yang lebih kecil. Proses ini juga memungkinkan analisis terperinci terhadap komunitas-komunitas dalam graf, mengungkapkan informasi penting tentang bagaimana node-node dalam komunitas saling terhubung dan berinteraksi. Visualisasi yang dihasilkan tidak hanya memudahkan pemahaman tentang struktur komunitas, tetapi juga membantu dalam

mengidentifikasi pola-pola penting yang mungkin tidak terlihat dalam analisis data mentah. Dengan demikian, Community Layout menjadi alat yang sangat berguna untuk mengeksplorasi dan memahami dinamika serta hubungan antar node dalam sebuah jaringan.

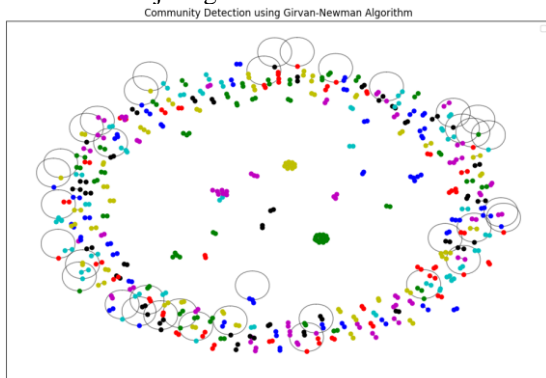


Fig 4. Graf NetworkX Hasil Algoritma Girvan-Newman

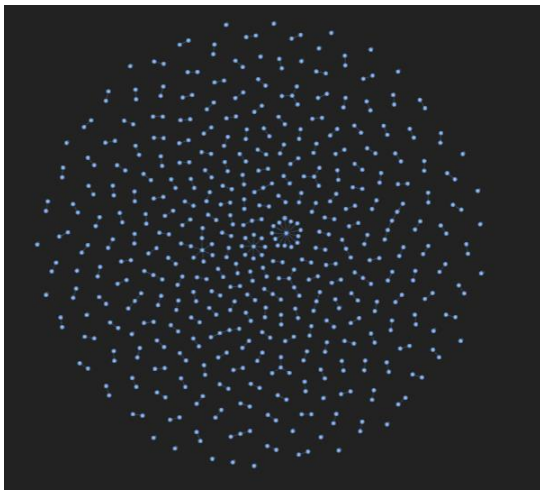


Fig 5. Graf Pyvis Hasil Algoritma Girvan-Newman

B. Analisis

Setelah menerapkan metode Girvan-Newman, data menghasilkan 270 komunitas yang memerlukan analisis lebih lanjut. Graf pada *fig. 3* menunjukkan struktur jaringan dari interaksi pengguna Twitter yang membahas topik "techbros". Setiap node mewakili pengguna, dan setiap edge mewakili interaksi (reply) antara dua pengguna. Graf pada *fig. 4* dan *fig. 5* menunjukkan deteksi komunitas menggunakan algoritma Girvan-Newman. Setiap warna atau node yang berkumpul mewakili komunitas yang berbeda. Komunitas ini terbentuk karena pengguna dalam satu komunitas lebih sering berinteraksi satu sama lain daripada dengan pengguna di luar komunitas. Kemudian dilakukan juga analisis sentralitas menggunakan beberapa algoritma.

Rank	Username	Degree Centrality
1	Jaytech	0.0234
2	kvngNath	0.0141
3	TechBros_7	0.0108
4	MichaelEMann	0.0072

5	JonLamArt	0.0072
---	-----------	--------

Tabel 1. Hasil Analisis Sentralitas Degree

Pengguna Jaytech memiliki degree centrality tertinggi, menunjukkan bahwa ia memiliki koneksi langsung paling banyak dalam jaringan. Ini berarti Jaytech adalah pengguna yang paling aktif berinteraksi dengan pengguna lain dalam diskusi tentang "techbros".

Rank	Username	Betweenness Centrality
1	Jaytech	0.0005
2	kvngNath	0.0002
3	TechBros_7	0.0002
4	nnadibrendan4	4.55e-05
5	Letter_to_Jack	4.55e-05

Tabel 2. Hasil Analisis Sentralitas Betweenness

Pengguna Jaytech memiliki betweenness centrality tertinggi, menunjukkan bahwa ia sering berada di jalur terpendek yang menghubungkan pengguna lain dalam jaringan. Ini berarti Jaytech adalah penghubung penting yang memungkinkan informasi mengalir antara komunitas yang berbeda.

Rank	Username	Closeness Centrality
1	Jaytech	0.0234
2	kvngNath	0.0141
3	SChukwuakalo	0.0121
4	Ekitipikin	0.0121
5	clinton_pase	0.0121

Tabel 3. Hasil Analisis Sentralitas Closeness

Pengguna Jaytech juga memiliki closeness centrality tertinggi, menunjukkan bahwa ia berada di posisi strategis yang memungkinkan dia untuk mencapai pengguna lain dengan lebih cepat. Ini mengindikasikan bahwa Jaytech dapat menyebarkan informasi dengan efisien dalam jaringan.

Rank	Username	Eigenvector Centrality
1	Jaytech	0.7071
2	SChukwuakalo	0.1961
3	Ekitipikin	0.1961
4	clinton_pase	0.1961
5	darkskinned_kei	0.1961

Tabel 4. Hasil Analisis Sentralitas Eigenvector

Pada betweenness centrality, pengguna Jaytech menjadi yang tertinggi, menunjukkan bahwa ia sering berada di jalur terpendek yang menghubungkan pengguna lain dalam jaringan. Ini berarti Jaytech adalah penghubung penting yang memungkinkan informasi mengalir antara komunitas yang berbeda.

KESIMPULAN

Berdasarkan hasil analisis sentralitas, pengguna Jaytech menonjol sebagai aktor kunci dalam jaringan percakapan "techbros" di Twitter, ditandai dengan nilai tertinggi dalam semua ukuran sentralitas: degree centrality (0.0234), closeness centrality (0.0234), betweenness centrality (0.0005), dan eigenvector centrality (0.7071). Hal ini menunjukkan bahwa Jaytech sangat aktif, memiliki posisi strategis untuk menyebarkan informasi dengan cepat, sering berperan sebagai penghubung penting antara komunitas, dan memiliki pengaruh besar dalam jaringan. Pengguna lain seperti kvngNath, TechBros_7, SChukwuakalo, dan Ekitipikin juga memainkan peran penting dalam percakapan ini dengan nilai sentralitas yang tinggi di berbagai ukuran. Deteksi komunitas menggunakan algoritma Girvan-Newman berhasil mengidentifikasi beberapa komunitas dalam jaringan ini, dengan setiap komunitas menunjukkan interaksi yang kuat di antara anggotanya. Analisis ini memberikan wawasan mendalam tentang struktur dan dinamika percakapan "techbros" di Twitter serta mengidentifikasi aktor-aktor kunci yang berkontribusi signifikan dalam diskusi tersebut.

REFERENCES

- [1] E. K. Crandall, R. H. Brown, and J. McMahon, "Magicians of the Twenty-first Century: Enchantment, Domination, and the Politics of Work in Silicon Valley," *Theory & Event*, vol. 24, no. 3, pp. 841-873, Jul. 2021. doi: 10.1353/tae.2021.0045.
- [2] E. Shearer and A. Mitchell, "News on Twitter: Consumed by Most Users and Trusted by Many," Pew Research Center, 2021. [Online]. Available: <https://www.pewresearch.org/fact-tank/2021/11/15/news-on-twitter-consumed-by-most-users-and-trusted-by-many/>.
- [3] S. Samoilenko, P. Kim, and J. L. Bosley, "Whose research benefits more from Twitter? On Twitter-worthiness of communication research and its role in reinforcing disparities of the field," *PLOS ONE*, vol. 16, no. 4, pp. e0250041, Apr. 2021. [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0250041>.
- [4] S. H. Anwar, "The Social Media Revolution: Exploring the Impact on Journalism and News Media Organizations," *Inquiries Journal*, vol. 2, no. 3, pp. 1-6, 2010. [Online]. Available: <http://www.inquiriesjournal.com/articles/202/the-social-media-revolution-exploring-the-impact-on-journalism-and-news-media-organizations>.
- [5] S. Devi and M. Rajalakshmi, "Community Detection by Node Betweenness Using Optimized Girvan-Newman Cuckoo Search Algorithm," *Information Technology and Control*, vol. 52, no. 1, pp. 53-67, Mar. 2023, doi: 10.5755/j01.itc.52.1.31535.
- [6] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, pp. 026113, 2004. doi: 10.1103/PhysRevE.69.026113.
- [7] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, pp. 7821-7826, 2002. doi: 10.1073/pnas.122653799.
- [8] PT. Ivonesia Solusi Data Website Team (2022) Data crawling: Pengertian, Tujuan, Dan Cara Kerjanya, PT. Ivonesia Solusi Data Official Website. Available at: <https://ivosights.com/read/artikel/data-crawling-pengertian-tujuan-dan-cara-kerjanya#:~:text=itu%20Data%20Crawling%3F-,Data%20crawling%20adalah%20proses%20otomatis%20untuk%20mengumpulkan%20dan%20mengindeks%20data,dan%20mengambil%20informasi%20yang%20dibutuhkan>. (Accessed: April 1, 2024).
- [9] N. T. Luchia et al., "SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Analysis of Twitter User Sentiments for the Aplikasi TikTok Application Using Naïve Bayes Classifier Algorithm Analisis Sentimen Pengguna Twitter Terhadap Aplikasi TikTok Menggunakan Algoritma Naïve Bayes Clasifier." [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [10] H. Satria, "helimisatria/tweet-harvest." Apr. 02, 2024. Accessed: Apr. 03, 2024. [Online]. Available: <https://github.com/helimisatria/tweet-harvest>