

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. Komunitas dapat dideteksi dari data interaksi pengguna di Twitter dengan menggunakan algoritma Girvan-Newman. Penelitian ini dilakukan untuk mendeteksi komunitas di Twitter berdasarkan perbincangan yang menggunakan istilah "techbros". 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 yang mengandung kata kunci "techbros" selama periode 01-01-2024 hingga 03-04-2024. Total dataset yang dianalisis sebanyak 774 tweet.

Keywords— *Kata kunci: techbros, twitter, deteksi komunitas, algoritma girvan-newman, 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 id tiap percakapan di Twitter, yang menjadi dasar untuk membentuk graf interaksi.

2. Pra-pemrosesan Data:

Data diproses untuk menghapus duplikasi dan nilai null. Hanya kolom relevan (username dan id_str) yang digunakan dalam proses deteksi komunitas ini.

3. Pembuatan Graf:

Graf dibuat menggunakan *library* 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.

kolom yang relevan, yaitu username dan id_str, yang dipilih untuk dianalisis lebih lanjut. Pemilihan kolom ini dilakukan karena mereka secara langsung mewakili interaksi antar pada tiap percakapan yang ada di dataset, 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* dilakukan, dataset kini 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"There are {num_tweets} number of tweets after preprocessing.")
```

There are 337 number of tweets after preprocessing.

Fig 2. Total data setelah pra-pemrosesan

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 atau tweet. 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.

tweet_id	user_id	text	retweet_count	reply_count	like_count
1734567890123456789	1234567890123456789	techbros are the future of the internet	15	5	20
1734567890123456790	1234567890123456790	the future is now with techbros	10	3	15
1734567890123456791	1234567890123456791	techbros are changing the world	20	8	25

Fig 1. Hasil crawling data

2. Pra-pemrosesan Data:

Dataset yang diperoleh telah menjalani proses pra-pemrosesan untuk memudahkan analisis. Hanya

3. Pembuatan Graf:

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

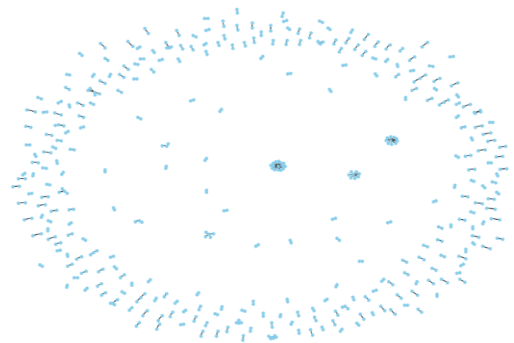


Fig 3. Graf NetworkX

Graf yang terbentuk memiliki total 337 node dan 631 edge.

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.

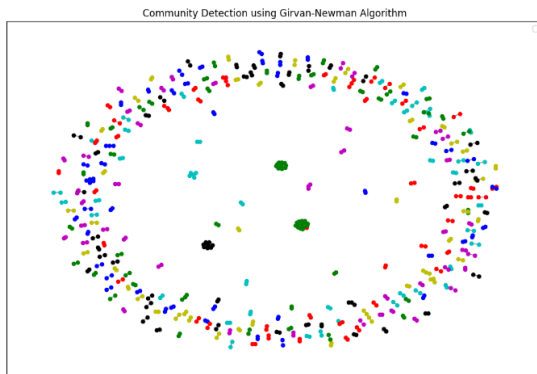


Fig 4. Graf NetworkX Hasil Algoritma Girvan-Newman

Terdapat total 295 komunitas yang terdeteksi setelah menerapkan algoritma Girvan-Newman pada dataset. Jumlah tersebut terlalu besar dan sebagian besar dari komunitas-komunitas tersebut tidak efektif untuk diteliti lebih lanjut dikarenakan isinya yang terlalu kecil. Sehingga, proses selanjutnya adalah mengurutkan komunitas-komunitas tersebut berdasarkan banyaknya member komunitasnya, lalu mengambil 10 komunitas teratas. Tiap warna merepresentasikan komunitas yang memiliki label 0 hingga 9.

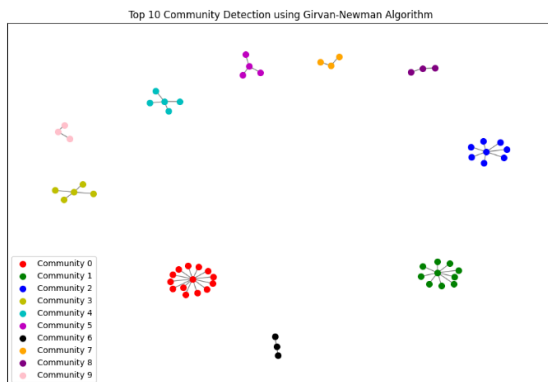


Fig 5. Graf NetworkX Dari Top 10 Komunitas Teratas

Dapat kita lihat bahwa 10 komunitas tersebut dikategorikan berdasarkan warnanya masing-masing, sehingga kita dapat melihat komunitas yang dideteksi dengan lebih jelas. Kita juga dapat melihat graf dari *library* Pyvis di bawah ini untuk melihat lebih jelas node yang ada pada masing-masing komunitas, dimana node yang berada di tengah merepresentasikan orang yang melakukan percakapan, sedangkan node-node yang terhubung ke orang tersebut merepresentasikan cuitan di Twitter.

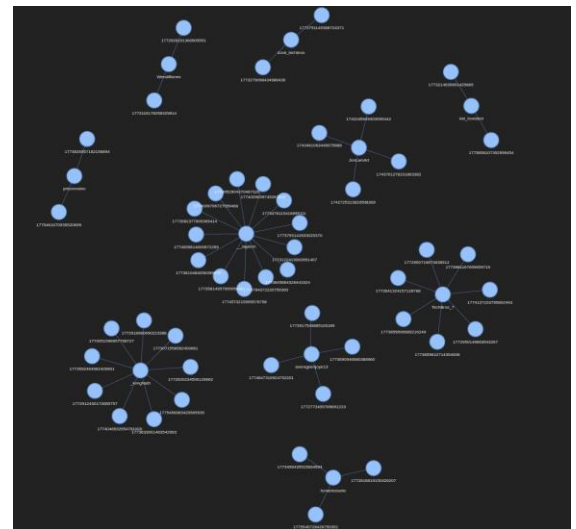


Fig 6. Graf Pyvis Hasil Algoritma Girvan-Newman

Selanjutnya, dari data 10 komunitas teratas tersebut, kita dapat membuat bar chart untuk melihat komunitas mana yang memiliki member terbesar hingga member terkecil berdasarkan label yang telah diberikan sebelumnya.

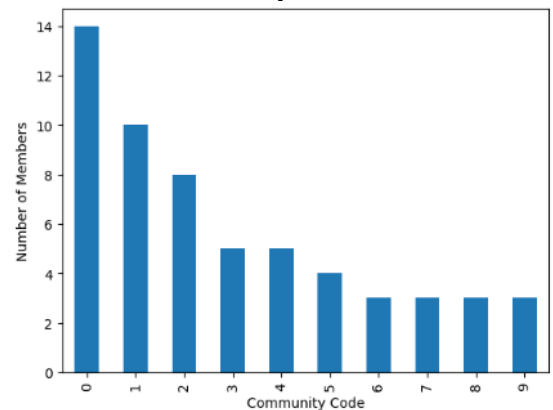


Fig 7. Barchart 10 komunitas teratas

B. Analisis

Setelah menerapkan algoritma Girvan-Newman ke dataset, terdapat 295 komunitas yang berhasil dideteksi. Graf pada *fig. 3* dan *fig. 4* menunjukkan struktur jaringan dari interaksi pengguna Twitter yang membahas topik "techbros". Setiap node mewakili pengguna, dan setiap edge mewakili interaksi pengguna pada tiap cuitan. Untuk melihat lebih jelas komunitas yang terbentuk, kita dapat melihat graf pada *fig. 5* dan *fig. 6* yang menunjukkan 10 komunitas teratas yang dikategorikan berdasarkan label serta setiap warna atau node yang berkumpul mewakili komunitas yang berbeda. Dapat kita lihat juga pada barchart bahwa komunitas dengan label 0 merupakan komunitas terbesar dengan total 14 member, sedangkan komunitas dengan label 6 hingga 9 merupakan komunitas terkecil dengan total 3 member.

Selanjutnya, kita dapat melihat *wordcloud* dari masing-masing komunitas untuk mendapatkan wawasan terkait topik yang dibahas oleh masing-masing komunitas.



Fig 8. Wordcloud komunitas 0



Fig 9. Wordcloud komunitas 1



Fig 10. Wordcloud komunitas 2



Fig 11. Wordcloud komunitas 3



Fig 12. Wordcloud komunitas 4

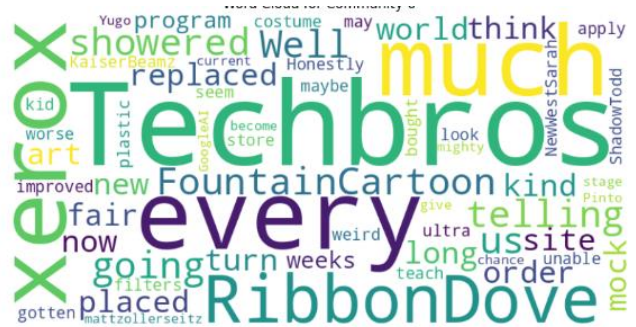


Fig 13. Wordcloud komunitas 5

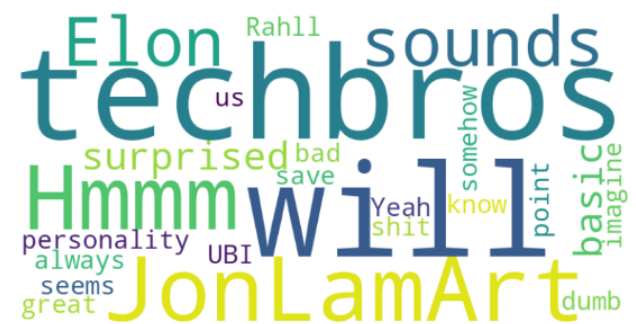


Fig 14. Wordcloud komunitas 6

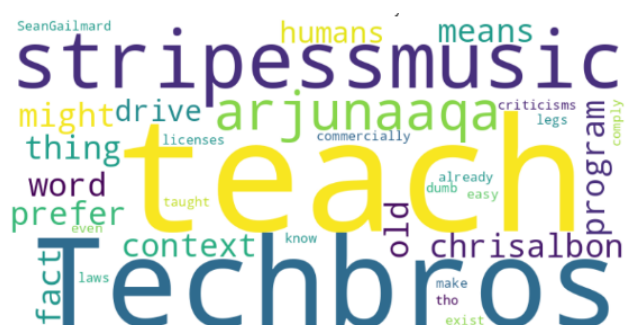


Fig 15. Wordcloud komunitas 7



Fig 16. Wordcloud komunitas 8



Fig 17. Wordcloud komunitas 9

KESIMPULAN

Deteksi komunitas menggunakan algoritma Girvan-Newman berhasil mengidentifikasi cukup banyak komunitas dalam dataset yang ada, dengan setiap komunitas menunjukkan interaksi yang kuat di antara anggotanya. Komunitas ini terbentuk karena tiap pengguna dalam satu komunitas lebih sering berinteraksi satu sama lain daripada pengguna di luar komunitas, hal ini memberikan kita wawasan tentang struktur dan dinamika percakapan "techbros" di Twitter serta mengidentifikasi aktor-aktor kunci yang berkontribusi signifikan dalam diskusi tersebut.

Dari hasil *wordcloud* tiap komunitas, kita dapat mendapatkan gambaran terkait topik yang sering di bahas pada masing-masing komunitas. Contohnya di komunitas 0, tidak banyak kata-kata yang mengandung arti khusus di bidang teknologi, selain Figma. Dari sini, kita dapat melihat bahwa komunitas 0 lebih sering berinteraksi hal-hal umum atau mungkin proyek-proyek antar satu sama lain, dimana kata-kata seperti "Ekitipikin" dan "Clinton" bisa jadi merupakan username atau nama dari member komunitas tersebut. Lalu terdapat banyak kata-kata suportif seperti "Thanks" dan "Awesome", sehingga komunitas 0 kemungkinan adalah komunitas UI/UX desainer yang sering berbagi ke komunitas menggunakan Figma/Ghost. Komunitas 1 ini lebih berfokus pada interaksi sosial dan apresiasi satu sama lain. Untuk komunitas lain yang terdeteksi menunjukkan interaksi yang positif dan suportif dengan banyak nama member yang muncul sebagai pusat percakapan. Tidak banyak fokus khusus pada teknologi tertentu selain beberapa referensi umum. Ini menunjukkan bahwa komunitas "techbros" di Twitter cenderung memiliki suasana yang ramah dan kolaboratif.

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%20automatis%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 Classifier." [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>