



# *Data Analytics* **PORTFOLIO**

SOCIAL MEDIA ANALYSIS



# *Table of* **OUTLINE**

---

**ABOUT ME**

**EDUCATIONAL BACKGROUND**

**TOOLS**

**DATA DESCRIPTION**

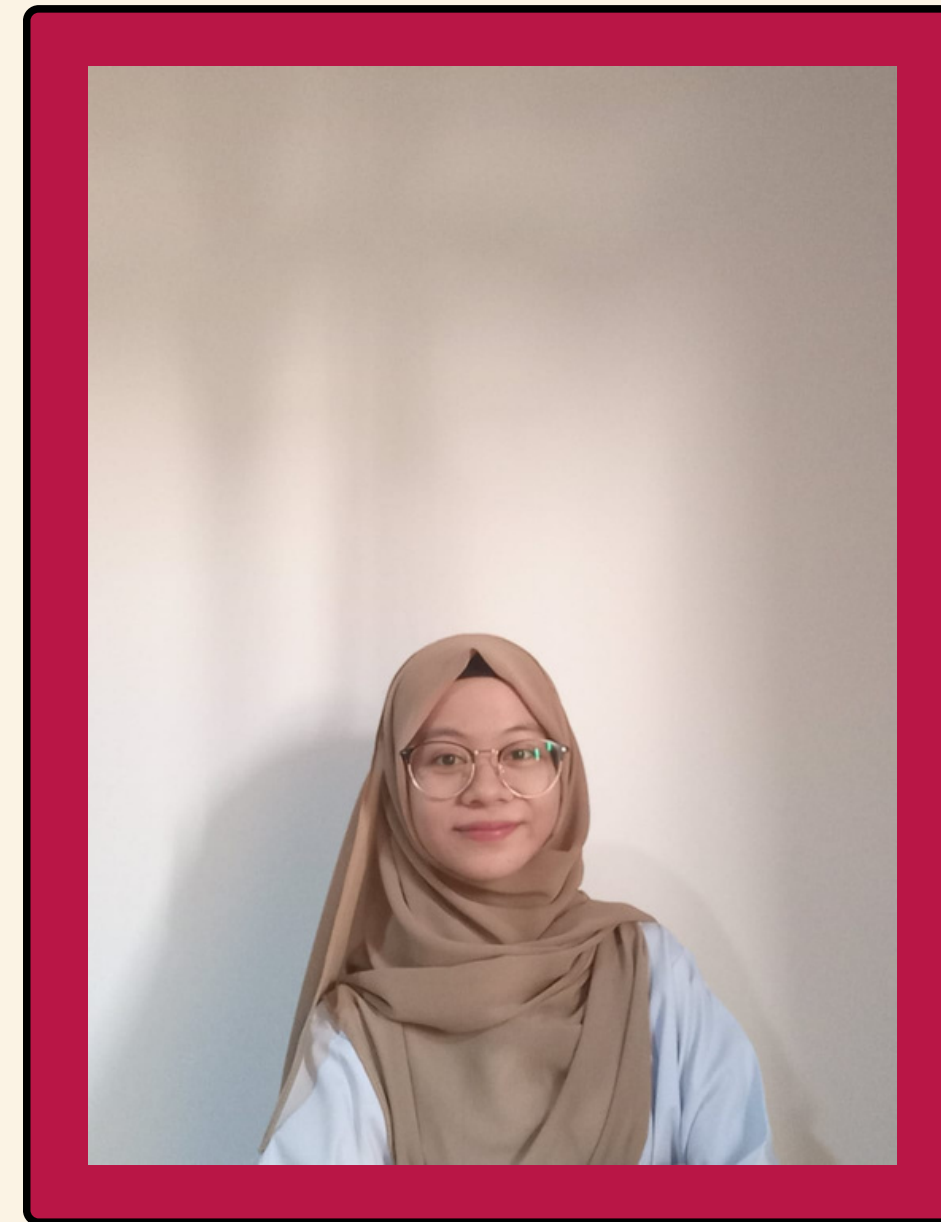
**SIMPLE TEXT VISUALIZATION**

**TEXT MINING & NLP**

**SOCIAL NETWORK ANALYSIS**

# *introduce* **ABOUT ME**

I am an undergraduate student in University of Indonesia majoring in Statistics who is steadfast with my commitment to doing the best in all circumstances. I am curious by nature and constantly researching ways to improve my skills and performance. I am looking forward to grow as Data Analyst as well with my enthusiastic to give my best effort for my works.



**WIDDYADARI  
DHIYA  
PARAHITA**

# EDUCATION BACKGROUND



# TOOLS



**PYTHON**



**VOYANT TOOLS**



**ORANGE 3**



# DATA DESCRIPTION

---

# DATASET

## **Dataset used:**

- Dataset with the keyword 'pildun' to perform Social Network Analysis. Data is taken from December 18–30, 2022.
- Dataset with the keyword 'pildun'. Data is taken from December 18–19, 2022. In the Text Mining and NLP section, this data is used to analyze sentiment before and after the 2022 Qatar World Cup Final.
- Dataset with keywords 'messi' and 'mbappe'. The data was taken from December 18, 2022 22:00 – December 19, 2022 1:30. This dataset was used to determine the sentiment of Twitter users towards the two players in the 2022 World Cup Final.



# PREPROCESSING

## **Import Module:**

- First, twint and tweepy modules were installed to retrieve data from Twitter. Then, import other modules for further data processing.
- The installation of the Sastrawi module to convert Indonesian words and also the unidecode module to process symbols and non-Latin characters.

## **Scrapping:**

- Scrapping data from twitter using twint module with keyword "pildun" and configuration limit of 20000.
- Obtained 20009 lines of twitter data containing the word
- Defining the cleantext function, the data is cleaned from affixes, symbols, and non-Latin characters.
- Then, the previously clean twitter data is saved into csv format to be processed at a later stage.



# DATA OBTAINED

	date object ▾	tweet object ▾	username object ▾	clean_tweet object ▾
0	2022-12-29 19:21:29	@MadridisWay @Tresnadika07...	NorthBoysGaruda	madrdisway tresnadika07...
1	2022-12-29 19:20:28	@ChovaHerman Event AFF aja...	akustikinesis	chovaherman event aff aja kaya...
2	2022-12-29 19:15:48	Di pildun brazil gatot &gt; yg juar...	bagustrchyo	pildun brazil gatot juara malah...
3	2022-12-29 19:02:25	Neymar otw pildun	almufid20	neymar otw pildun
4	2022-12-29 18:57:18	@benwaee_ @jebreetmedia...	Ramdhan__Wahyu	benwaee jebreetmedia...



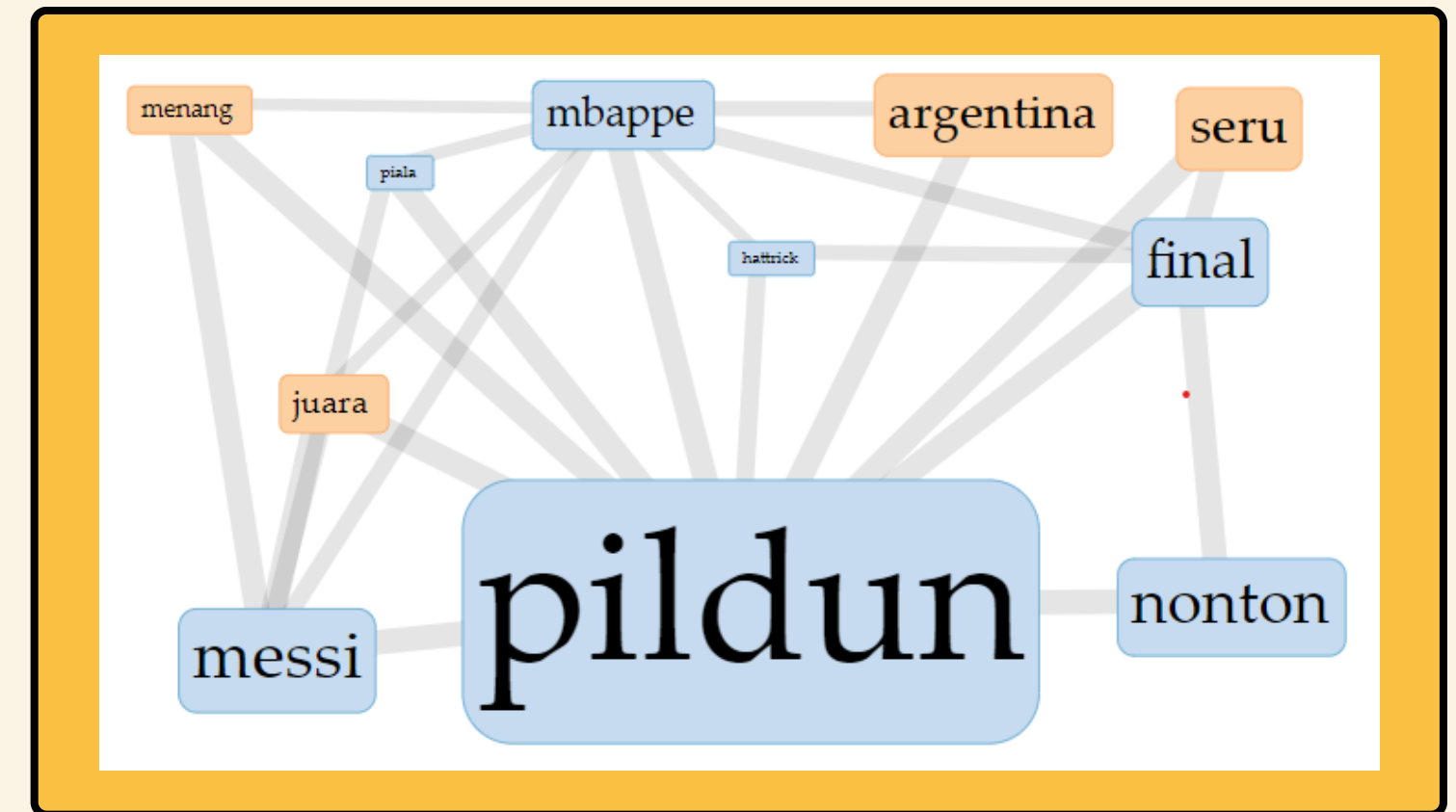
# **SIMPLE TEXT VISUALIZATION**

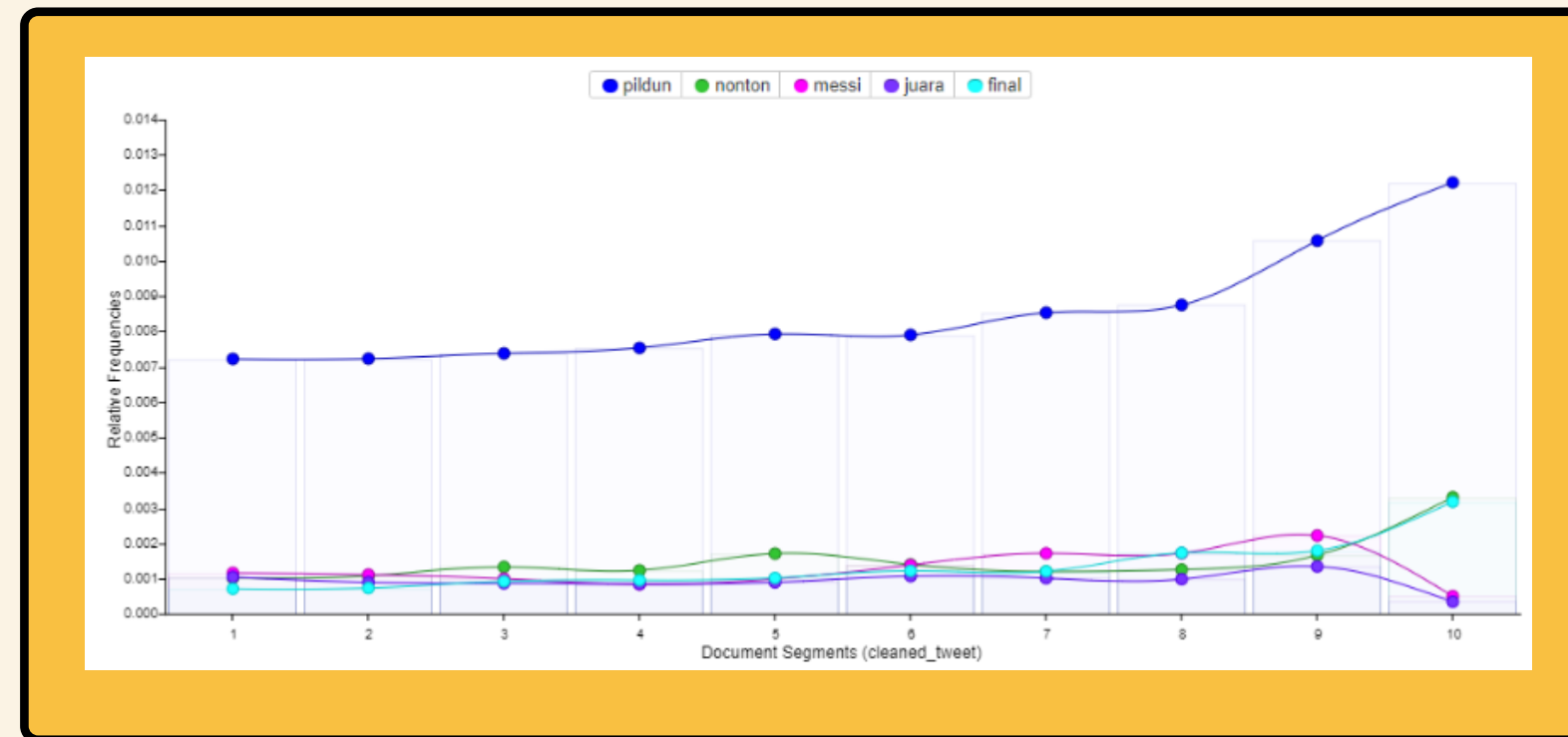
---

- Using Word Cloud, we found out the words that are often discussed by netizens on twitter about the world cup.
- The words most often expressed to describe the world cup are watch, final, messi, win, Argentina, and champion.
- Netizens often associate pildun with another famous French soccer player, Mbappe.
- This means that when netizens discuss the topic of the world cup, they often also discuss the countries that are included in the final qualifications to important figures who play a role, such as Messi and Mbappe.



- Using link visualization, it is known that several words are often associated with "Pildun", such as "nonton", "messi", "juara", "mbappe", "argentina", and "menang".
- If we investigate further about the word "messi", most people associate it with the pildun champion.





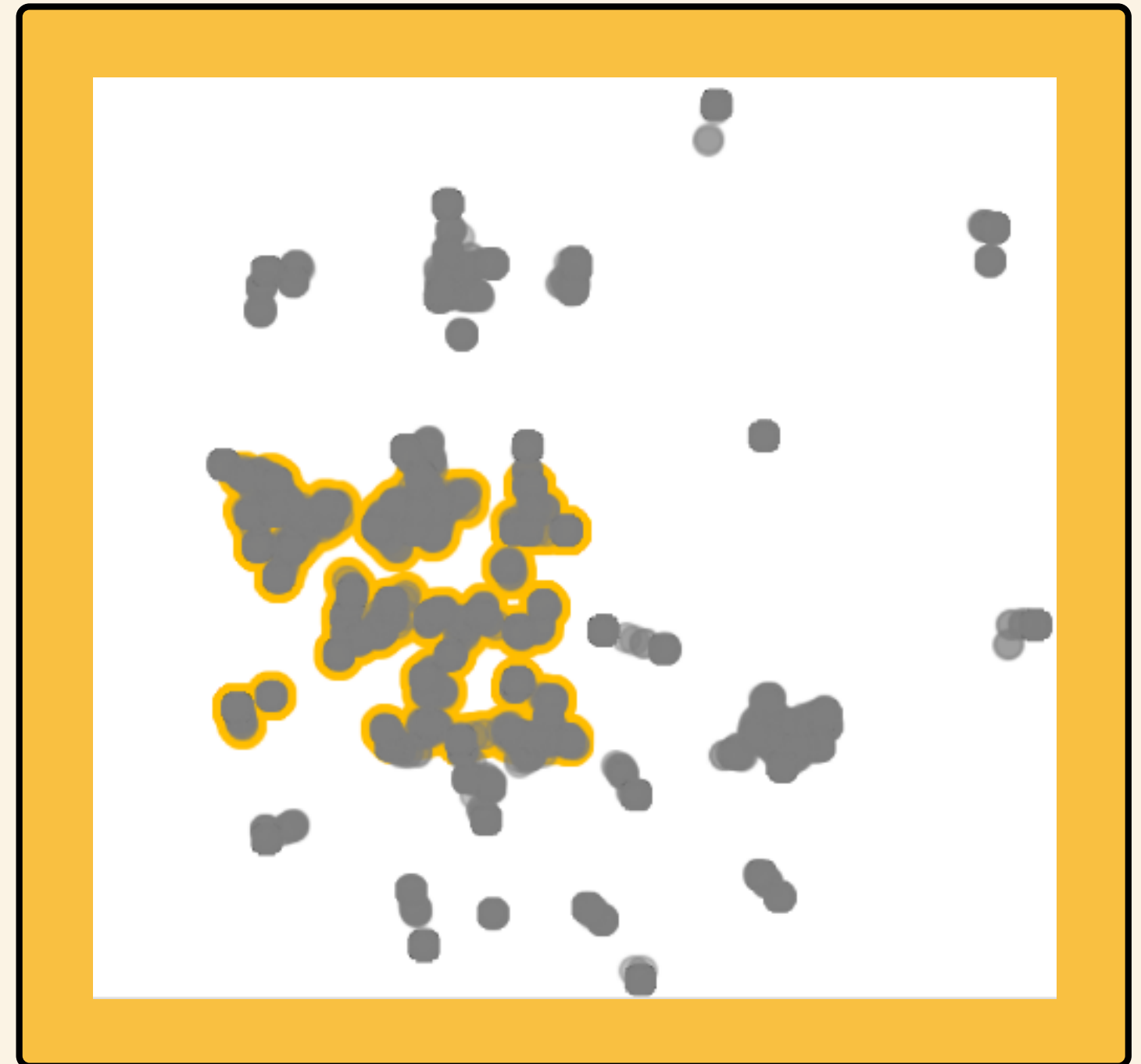
- The visualization above shows the trend of the 5 words with the highest frequency of occurrence.
- The word "pildun" has the highest frequency and the trend is increasing.
- In the words "watch" and "final", both are seen to have an increasing trend at the end in line with many netizens discussing the final at the end of the world cup period.
- For the words "messi" and "champion", both appear to have the same trend pattern, indicating a strong correlation between them. This is likely because many people associate Messi with winning.

# TEXT MINING AND NLP

---

# CLUSTERING T-SNE

- Only 1 cluster stands out
- What are these tweets talking about?
- To answer the above question in finding the most significant keywords, we use the Extract Keywords widget.



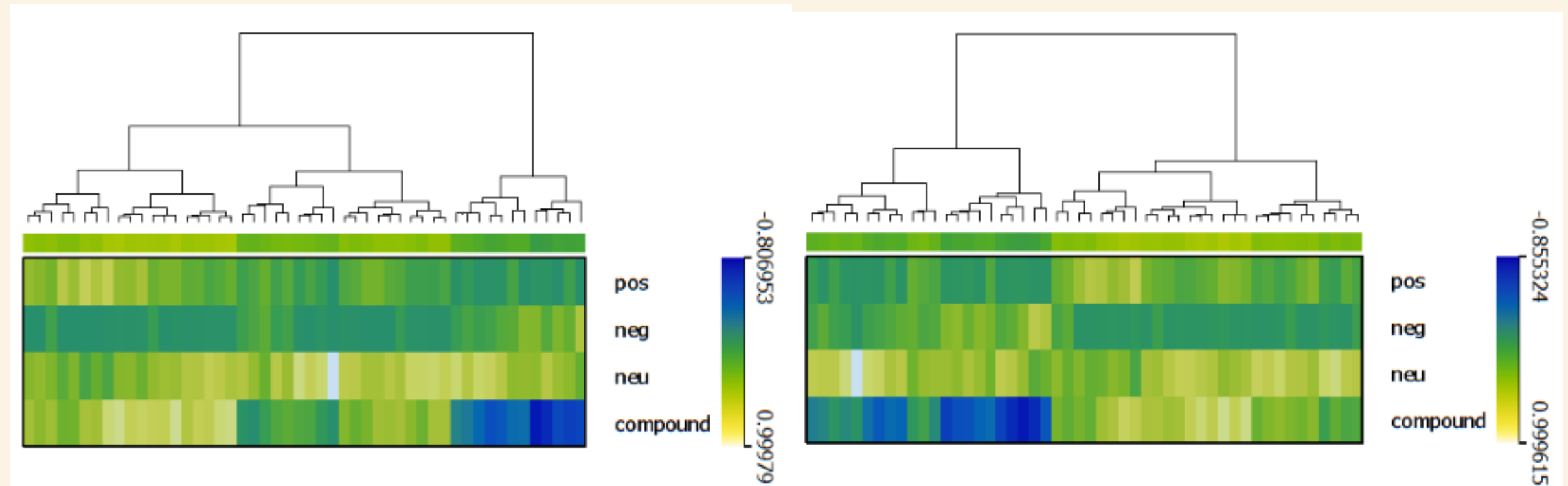
# TF-IDF

- The top words were 'messi' (Captain of Team Argentina) and 'champion', followed by 'mbappe' (Striker of Team France).
- Of the many tweets collected in the 2022 world cup time period, the most significant keywords are expected to be words that talk about the World Cup Final (Argentina vs France).
- 'idextratime' juga merupakan kata yang kerap di-tweet oleh user twitter yang membicarakan tentang Piala Dunia.
- idextratime (@idextratime) merupakan akun sepak bola non-formal yang aktif dalam memposting tweets tentang sepak bola secara keseluruhan.

Word	TF-IDF
messi	0.188
juara	0.185
mbappe	0.134
idextratime	0.134
argentina	0.129
main	0.127
udah	0.124
menang	0.124



# SENTIMENT ANALYSIS



- The number of tweets talking about Messi and Mbappe can be due to both being the Top Player of their respective National Teams, as well as the fierce battle with a score of 3-3 where Messi contributed 2 goals for Argentina and Mbappe 3 goals for France.
- What were the emotions/sentiments of twitter netizens towards Messi or Mbappe during the fierce battle? Twitter conversations about Mbappe tend to be neutral. This is different from Messi who tends to be negative based on the data of each cluster
- Crawling data with the keywords 'messi' and 'mbappe' during the 2022 World Cup Final (18/12/2022 22.00 – 19/12/2022 01.30)

# SENTIMENT ANALYSIS: COLLOCATION

- Twitter netizens talked more about the 2022 World Cup Final with top 3 scores about the victory of the Argentine national team and Messi.
- The 4th and 5th scores with the words "seru bgt" and "final seru" are basically positive sentiments.
- It is likely that the sentiment towards the 2022 World Cup is mostly positive.

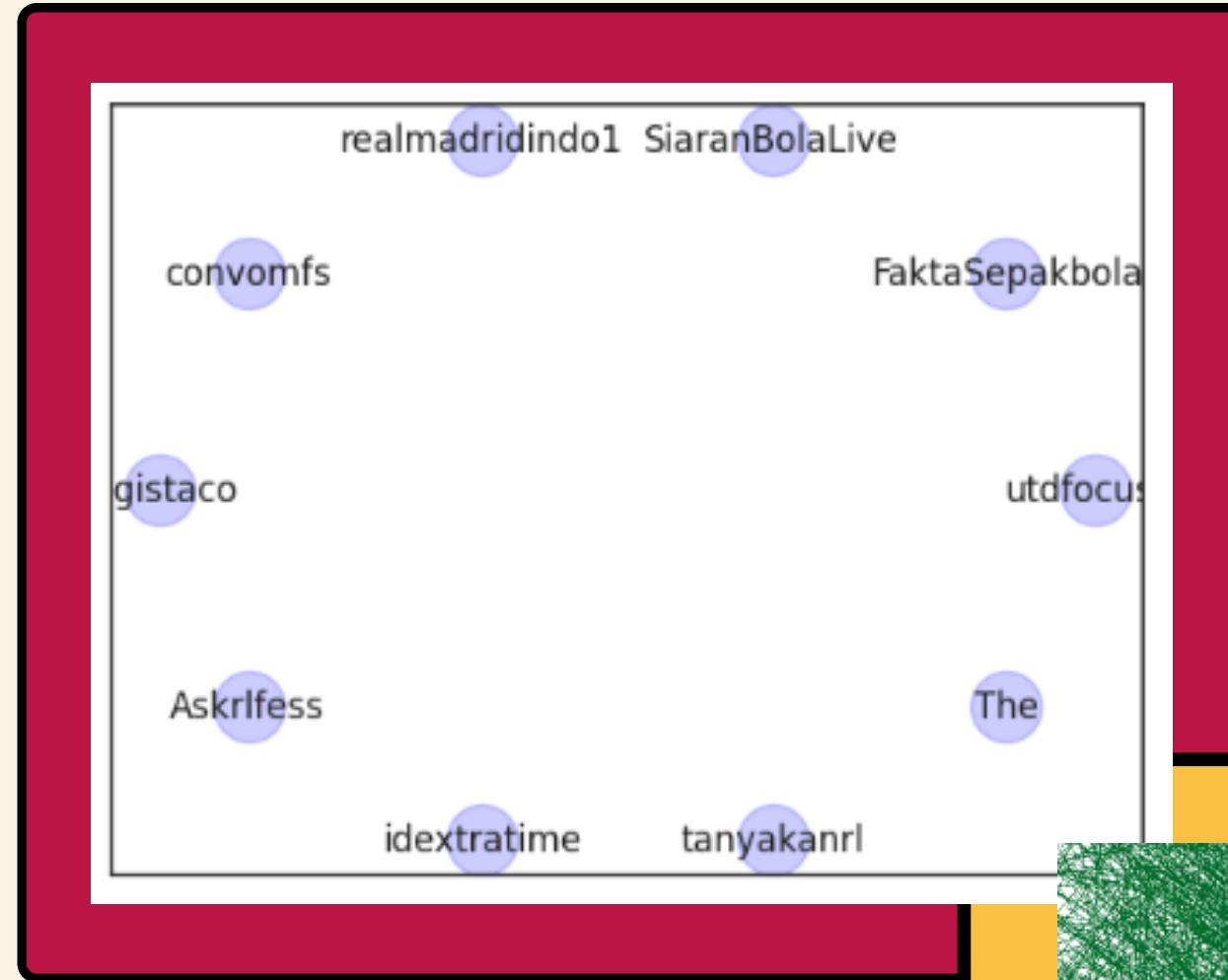
	Scores	Collocations
1	argentina menang	0.19212
2	messi juara	0.178108
3	argentina juara	0.176796
4	seru bgt	0.160864
5	final seru	0.158809

# SOCIAL NETWORK ANALYSIS

---

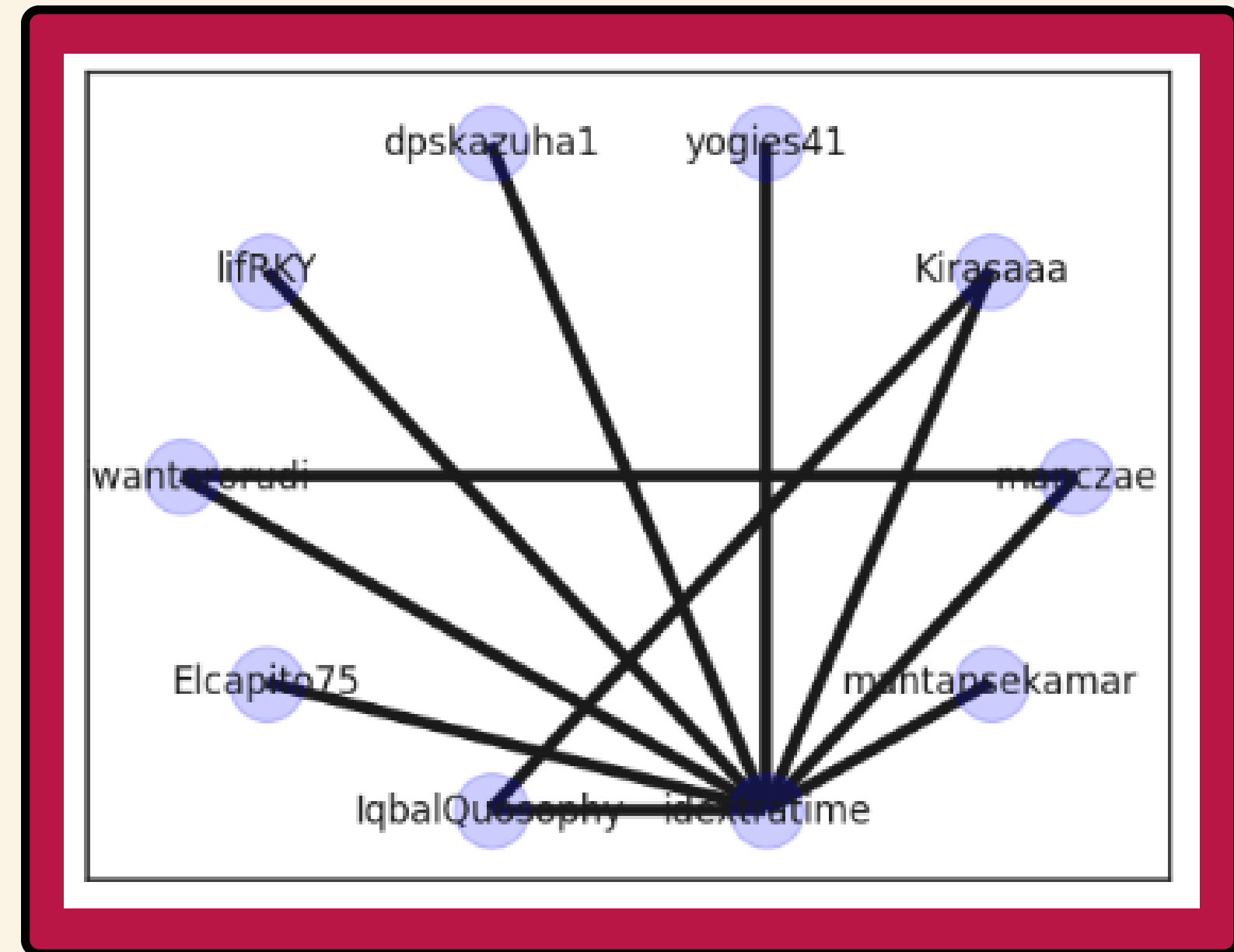
# DEGREE CENTRALITY

There are 10 nodes that have the most number of connections or interactions among the other nodes, where idextratime is the node with the highest number of connections followed by Broadcast Bola Live in second place.



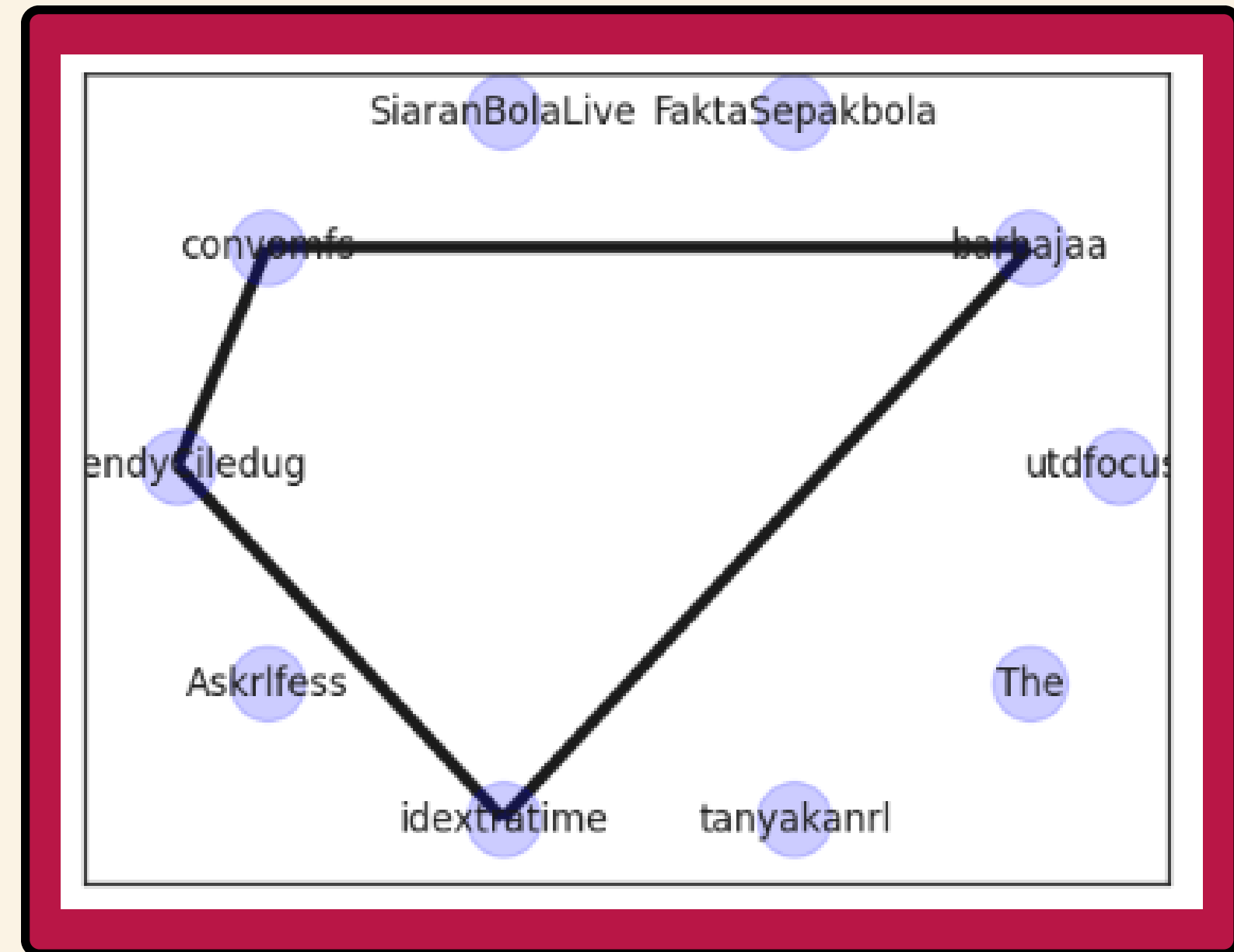
# CLOSENESS CENTRALITY

It can be seen in the side image that idextratime is a node that has interactions with all accounts. However, there are also many accounts that only interact with idextratime.



# BETWEENNESS CENTRALITY

idextratime, barbajaa, convomfs, and rendyCiledug serve as a bridge between the two communities.



# COMMUNITY DETECTION

From G\_karate, 3 communities were obtained with a total of 34 nodes and 74 edges. The community members include:

- Community 1 (Blue):

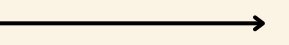
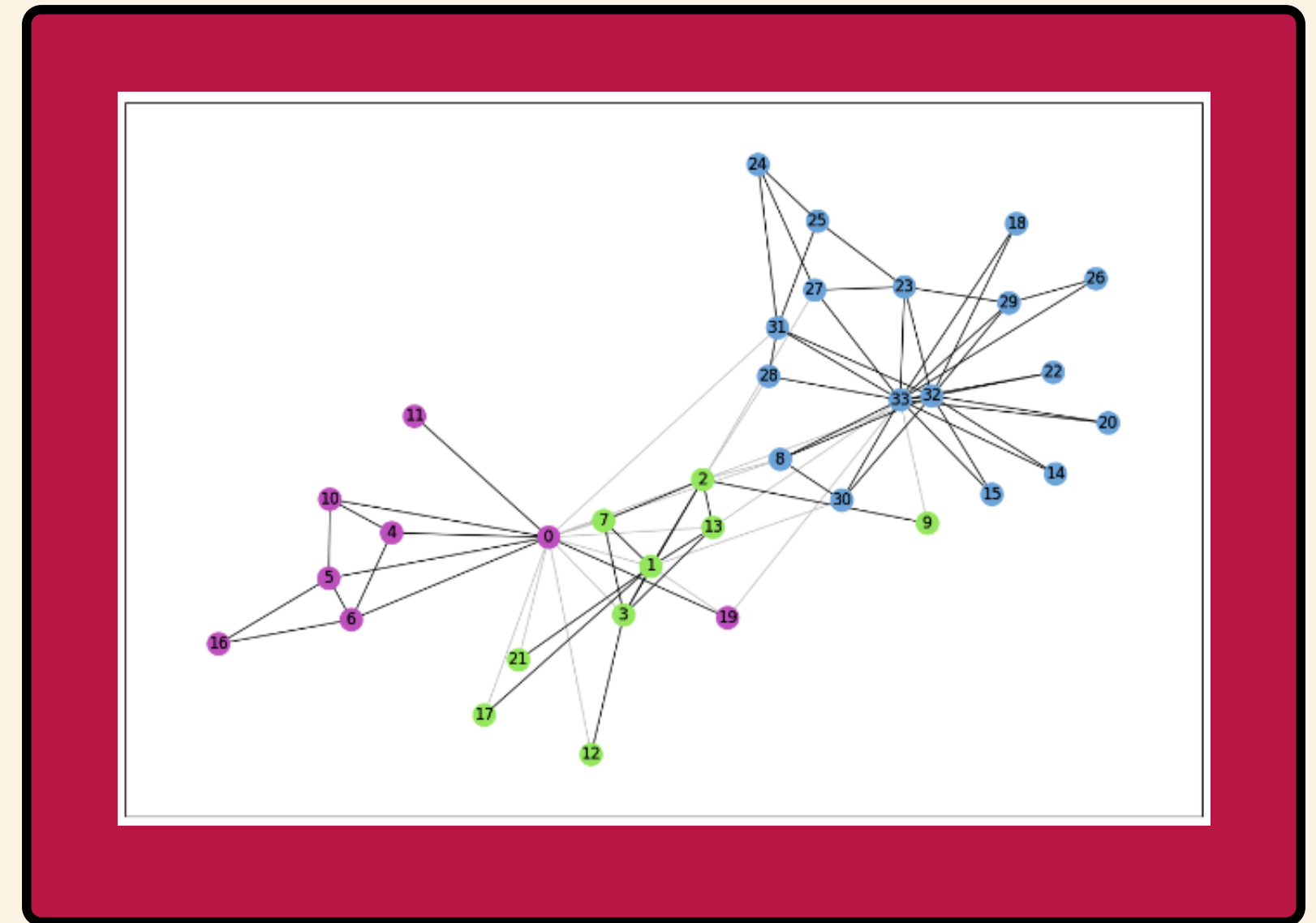
{8, 14, 15, 18, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33}

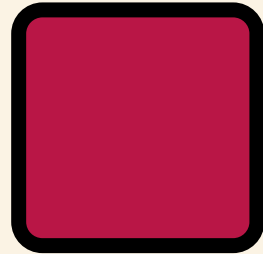
- Community 2 (Green):

{1, 2, 3, 7, 9, 12, 13, 17, 21}

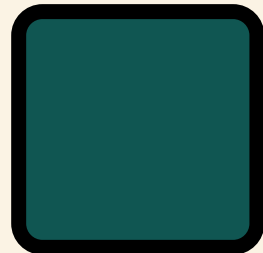
- Community 3 (Purple):

{0, 16, 19, 4, 5, 6, 10, 11}

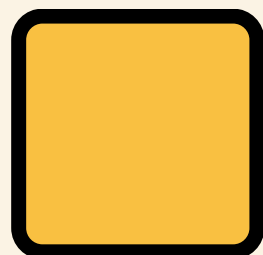




Twitter users, especially in Indonesia, often use informal words to interact on social media.



In copywriting, copywriters can use words found in social media analytics results (collocations, word clouds, etc.) to create advertising content to increase engagement and personal branding, especially on Twitter.



In marketing, companies can collaborate with people or social media accounts that have many connections to influence the company's products.

---

# CONCLUSION