4 Assignment 4: Graph Analytics On Social Network Data

The advent of social media platforms has fundamentally transformed the landscape of political communication, enabling direct and immediate interaction between politicians and the public. This assignment leverages data extracted from X (formerly Twitter) to analyze the social media activity of the 150 members of the Belgian federal parliament. The dataset, collected from January 2024 to early May 2024, includes tweets and replies involving these politicians and many other attributes.

By employing Cypher queries in Memgraph and visualization tools like Gephi, we uncovered some interesting insights into the behavior and influence of Belgian politicians and their corresponding political parties on X. Our methodology and their final results will be discussed in what follows.

4.1 Cypher Queries and Memgraph

After extensive research into the field of Cypher queries and becoming well-versed in their application, we managed to uncover some intriguing relationships within the dataset stored in Memgraph. The process involved delving deep into the powerful graph query language used by Memgraph, which allowed us to efficiently query and manipulate the graph data. The analysis revealed several key findings, consisting of the following elements:

4.1.1 Politicians with the Most Replies

Using Cypher queries, we pinpointed which politicians received the highest number of replies to their tweets. This metric is crucial in understanding the engagement and interaction levels certain politicians elicited from the public, indicating their prominence or the contentious nature of their posts. Using the following query:

```
MATCH (p:Party)<-[e1]-(u:User)-[e2]->(t:Tweet)
RETURN u.screen_name AS Politician, SUM(t.reply_count) AS TotalReplies
ORDER BY TotalReplies DESC
```

Figure 42: Most Replies

We found the following politicians with their corresponding amount of replies:

#	Politician	TotalReplies
1	FranckenTheo	25,531
2	SafaiDarya	16,350
3	egbertlachasert	7,499
4	emir_kir	7,245
5	JorisVDBroucke	6,783
6	MichaelFreilich	6,142
7	RaoulHedebouw	5,442
8	MelissaDepr	5,213
9	peter_mertens	4,872
10	SanderLoones	3,772
11	WouterDeVriendt	3,316

(a) Politicians Replies 1

126	KurtRavyts	3
127	SophieThemont	2
128	VicaireAlbert	2
129	DewulfNathalie	1
130	LeslieLeoni_PS	1
131	danielsenesael	0
132	ErikGilissen	0
133	scoumeau	0
134	jmdelizee	0
135	olivier_vajda	0
136	PhTison	0
137	StefaanVanhecke	0

(b) Politicians Replies 2

Figure 43: Politicians Replies

We see that the amount of replies is a lot higher for controversial public figures like Theo Franken and the leaders of the political parties, which is expected as they receive more media attention.

4.1.2 Politicians Receiving the Most Community Notes

By querying the dataset, we identified which politicians' tweets were most frequently flagged with community notes. These notes are indicative of tweets that the community has deemed needing additional context or correction, highlighting which politicians' posts were most scrutinized or controversial. Using the following query:

```
MATCH (p:Party)<-[e1]-(u:User)-[e2]->(t:Tweet)
WHERE t.has_birdwatch_notes = true
RETURN u.screen_name AS Politician, count(t) AS CommunityNotesCount
ORDER BY CommunityNotesCount DESC
```

Figure 44: Most Receiving

We found the following politicians with their corresponding amount of community notes:

#	Politician	CommunityNotesCount
1	FranckenTheo	8
2	ducarmedenis	4
3	MichaelFreilich	3
4	peter_mertens	3
5	SimonMoutquin	3
6	SafaiDarya	3
7	McMarghem	1
8	JorisVDBroucke	1
9	RaoulHedebouw	1
10	BuystKim	1
11	Sofie_Merckx	1

Figure 45: Politician Receiving

It is no surprise that the person with the most replies also has the highest amount of community notes. This shows us that tweets that gain a lot of attention and replies are often the ones where additional context is needed, due to their controversial nature.

4.1.3 Parties Receiving the Most Community Notes

Extending the analysis to political parties, we were able to determine which parties had the highest number of community notes across their members' tweets. This provided insights into which parties were most often involved in contentious or heavily debated discussions on the platform. Using the following query:

```
MATCH (p:Party)<-[e1]-(u:User)-[e2]->(t:Tweet)
WHERE t.has_birdwatch_notes = true
RETURN p.ident AS Party, count(t) AS CommunityNotesCount
ORDER BY CommunityNotesCount DESC
```

Figure 46: Parties Receiving

We found the following political parties and their corresponding amount of community notes:

#	Party	CommunityNotesCount
1	N-VA	15
2	MR	5
3	PVDA-PTB	5
4	Ecolo-Groen	4
5	ONAFH	1
6	Vooruit	1

Figure 47: Parties Most Receiving

4.1.4 Engagement metrics for tweets with and without community notes

To understand the impact of community notes on tweet engagement, we analyzed tweets with and without these notes. We calculated average favorites, retweets, replies, and views for both categories. This analysis helps determine whether community notes influence user interactions with tweets.Using the following query:

```
MATCH (t:Tweet)
WHERE t.has_birdwatch_notes = true
RETURN AVG(t.favorite_count) AS avgFavoritesWithNotes,
       AVG(t.retweet_count) AS avgRetweetsWithNotes,
       AVG(t.reply_count) AS avgRepliesWithNotes,
       AVG(t.views_count) AS avgViewsWithNotes;

MATCH (t:Tweet)
WHERE t.has_birdwatch_notes = false
RETURN AVG(t.favorite_count) AS avgFavoritesWithoutNotes,
       AVG(t.retweet_count) AS avgRetweetsWithoutNotes,
       AVG(t.reply_count) AS avgRepliesWithoutNotes,
       AVG(t.views_count) AS avgViewsWithoutNotes;
```

Figure 48: Metrics Query

We were able to create the following bar chart in R.

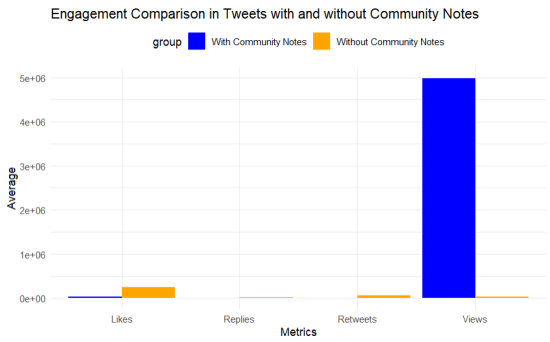

Figure 49: Engagement Comparison in Tweets with and without Community Notes

Almost no tweet with a community note receives a like, reply or retweet. However, we see that there is a very significant difference in the amount of views between the two types of tweets. This shouldn't be too much of a surprise, because it's a fact that controversial statements tend to get more attention in general.

4.1.5 Combining the elements

As we now have found the politicians and their parties with the most community notes, it is time to combine them to gain a better overview. We linked the politicians who received community notes to their corresponding party.

In the following code, we calculate and set community notes count for each politician. Afterwards we calculate and set community notes count for each political party and we finish by returning the nodes and relationships for visualization.

```
MATCH (p:Party)<-[e1]-(u:User)-[e2]->(t:Tweet)
WHERE t.has_birdwatch_notes = true
WITH p, u, count(t) AS CommunityNotesCount
SET u.community_notes_count = CommunityNotesCount
WITH p, u, CommunityNotesCount

MATCH (p:Party)<-[e1]-(u:User)
WITH p, u, CommunityNotesCount, count(u.community_notes_count) AS
PartyNotesCount
SET p.community_notes_count = PartyNotesCount
WITH p, u, CommunityNotesCount, PartyNotesCount

MATCH (p:Party)<-[e1]-(u:User)-[e2]->(t:Tweet)
WHERE t.has_birdwatch_notes = true
RETURN p, u, e1, e2, t
```

Figure 50: Combining Element Query

This gave us the following graph results in Memgraph.

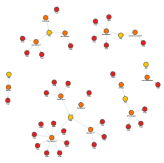

Figure 51: Overview of Combining Elements

As the graph is not very structured and easy to read, we used Gephi to upgrade our visualization to gain better insights.

4.2 Gephi

We exported the data from Memgraph in a JSON format. Using the provided python script, we transformed it to a graphML file, which Gephi is able to read.

'Force Atlas' was used to layout the graph, which pushed apart the non-linked nodes and attracted linked nodes to each other. Controlling the layout properties and many more settings made sure the representation was aesthetically pleasing.

The ranking module helped to configure the nodes colors and sizes.

- Orange for politicians with community notes.
- Purple for tweets with community notes

- Green for parties with community notes

The size of the orange nodes corresponds to the amount of community notes linked to that politician, which helps us gain a clear and quick understanding of which politicians are more 'controversial' than others (the bigger nodes are more controversial).

The labels of the political parties and politicians are shown so that they are able to be identified quickly. The id's of the tweets are hidden as they wouldn't provide extra useful information for our research.

The final results can be seen below and provided us with an interesting insight.

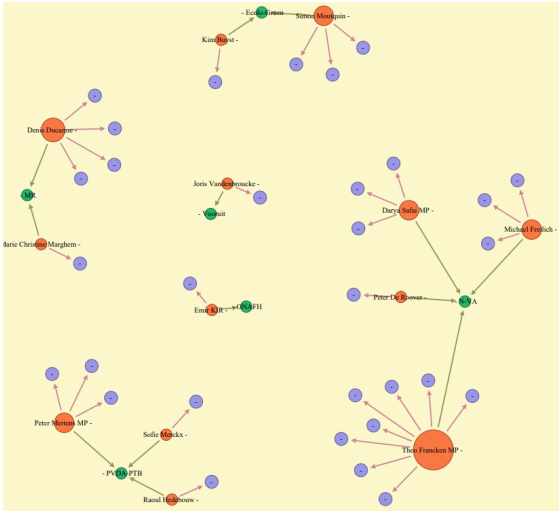

Figure 52: Politician Party Network

We clearly see that the political party N-VA is not only connected to the most amount of politicians with community notes (4), but it also contains some of the more controversial politicians (these politicians have many outgoing arcs). The political parties PVDA en MR show the same pattern but a bit less extreme.

This shows us that the political parties (that are currently present in the parlement) with a large amount of community notes are more on the extreme sides of the political spectrum and especially on the right side. In times of false information, deep fakes and an election, it is important to stay aware of these problems and to do your own research.

4.3 Conclusion

By using Cypher queries in Memgraph and visualizing the data with Gephi, we identified which politicians and political parties received the most replies and community notes, highlighting

their prominence and the contentious nature of their posts. The engagement metrics demonstrated that tweets with community notes, typically linked to controversial content, received significantly fewer likes, replies, and retweets but more views. Our analysis underscored that political parties with more extreme positions, particularly on the right, had higher incidences of community notes, emphasizing the importance of vigilance and independent research in an era of misinformation.