Social Network Analysis : study case of a Facebook group of diabetes in Morocco

Abdelillah AITROUGA
ENSIAS
Mohammed V University
Rabat, Morocco
abdelilah_aitrouga@um5.ac.ma

Abdessamad OUOMAR
ENSIAS
Mohammed V University
Rabat, Morocco
Abdessamad_OUOMAR@um5.ac.ma

Jamal BAYACINE
ENSIAS
Mohammed V University
Rabat, Morocco
jamal_bayacine@um5.ac.ma

Otmane BASTOS
ENSIAS
Mohammed V University
Rabat, Morocco
otmane_bastos@um5.ac.ma

Pr. RACHID OULAD HAJ THAMI
ENSIAS
Mohammed V University
Rabat, Morocco
rachid.oulad@um5.ac.ma

***Abstract*—This paper explores the use of Social Network Analysis (SNA) to study the structure and dynamics of a Facebook group focused on diabetes in Morocco. The aim of the study is to gain insights into the experiences and needs of people with diabetes in Morocco and to identify potential areas for improvement in terms of support and resources. The paper provides an overview of the state-of-the-art in SNA and its application in health research. The data collection, cleaning, analysis, and visualization process are detailed, and the results of the study are presented. The study found that the group was active and engaged, with users sharing information, providing emotional support, and building a sense of community. The study highlights the potential of social media platforms such as Facebook for promoting health and supporting people with chronic conditions.**

***Index Terms*—diabetes, SNA, social media, support group Morocco,Facebook**

I. Introduction

A social network is a network of relationships or interactions between people or actors, represented by nodes, and the connections or links between them, represented by edges or arcs. The analysis of social networks has been a topic of study for many years. There can be several type of social networks like email network, telephone network, collaboration network. But recently online social networks like Facebook, Twitter, LinkedIn, etc..., have been developed which gained popularity within very short amount of time and gathered large number of users. Facebook is said to have more than 500 million users in 2010 [2].

The field of social networks and their analysis has developed from the fields of graph theory, statistics, and sociology, and it is used in various other fields such as information science, business applications, communication, and economy. Analyzing a social network is similar to analyzing a graph, as social networks form the topology of a graph. Graph analysis tools have existed for decades, but they are not designed to analyze social network graphs, which have complex properties. An online social network graph can be very large, containing millions of nodes and edges. Social networks are also dynamic, with continuous evolution and expansion. A node in a social network typically has multiple attributes. There are both small and large communities within the social graph. Traditional graph analysis tools are not capable of managing such large and complex social network graphs.

In this paper, we will use Social Network Analysis (SNA) to study the structure and dynamics of a Facebook group focused on diabetes in Morocco. The aim of the study is to utilize SNA to gather insights, information, and suggestions from the group members and analyze them in relation to diabetes. By analyzing the structure and dynamics of the group, we hope to gain insights into the experiences and needs of people with diabetes in Morocco and to identify potential areas for improvement in terms of support and resources. Through this case study, we aim to contribute to a better understanding of the role that social media can play in supporting the health and well-being of people with chronic conditions.

The remaining of this article is organized as follows : In section 1, we will provide an overview of the state-of-the-art in this subject. In section 2, we will detail the data pre-processing process, including data collection, cleaning, analysis, and visualization. In section 3, we will present the results of our study. In section 4, we will discuss these results, and in section 5, we will conclude our study.

II. Literature review

Social Network Analysis (SNA) is a powerful tool for understanding the relationships and interactions within a group or community. As demonstrated in this article [6], SNA can be effectively used to investigate the potential of social media platforms, such as Facebook, as tools for promoting health. The study, which analyzed 1352 messages posted to an active diabetes group, found that the group was international in nature, with users overcoming language barriers to communicate with others with similar conditions. The interactions within the group were structured around sharing information, expressing emotions, and building a community. Users exchanged medical and lifestyle information and placed high value on personal experiences, opinions, and advice from their peers. They also demonstrated a positive attitude towards living with diabetes and provided encouragement and support to one another. The group was well-maintained by the administrator and core members and functioned as a social network for sharing social support, companionship and exerting social influence.

In the field of health, SNA can be used to study the relationships and interactions between patients with chronic conditions. For example, in this article [4], the authors use SNA to examine the emotional and psychosocial impact of the first COVID-19 pandemic lockdown on people with type 2 diabetes in France by analyzing exchanges on social networks. This study highlights the potential of SNA as a tool for understanding the needs and experiences of individuals with chronic conditions during a crisis.

Similarly, another study [5] found that Facebook groups can play a significant role in promoting diabetes self-management support. This study examined 34 public diabetes Facebook groups and the interactions within them. The results show that many groups aimed to provide instrumental support while fewer aimed to provide emotional support. The study also found that nutrition was the only self-management topic addressed in over 30% of posts. Additionally, posts made by engagement leaders were more likely to appear in inactive groups. The study suggests that health and diabetes educators should consider ways to more effectively leverage social media engagement leaders to disseminate valid health information on diabetes self-management.

III. RESEARCH METHODS

A. Data collection

Using the Facebook search function, we searched for the keyword "Diabète au Maroc". We selected the first group that appeared in the search results.

The group was named "Diabète Maroc - régime Low Carb" as shown in Figure 1 below. It was a private group, meaning that only members can see who is in the group and what is posted, but it was visible to anyone as it could be found through search results.
The group was created on January 30, 2021, and had more than 4,500 users at the time of the study, making it one of the largest groups on Facebook focused on people with diabetes or those affected by diabetes in Morocco.

Fig. 1: Overview of Facebook Groups Used in Our Study

We collected data for this group using web scraping with the tool facebook-scraper [1]. In accordance with Facebook's terms of service, we extracted posts from 10 pages of the group. In addition to the post content, we also collected various attributes such as the author, comments, date of posting, number of likes, shares, and other relevant information. Figure 2 below provides some statistics on the data we collected.

	post_id	likes	comments	shares	user_id	reactors	reactions	reaction_count
count	2.100000e+02	210.000000	210.000000	210.0	2.100000e+02	0.0	0.0	210.000000
mean	1.593565e+15	12.504762	20.595238	0.0	2.143685e+14	NaN	NaN	12.504762
std	8.487224e+13	12.585876	24.609155	0.0	5.044053e+14	NaN	NaN	12.585876
min	1.286403e+15	0.000000	0.000000	0.0	5.455525e+08	NaN	NaN	0.000000
25%	1.575368e+15	4.250000	5.000000	0.0	1.000030e+14	NaN	NaN	4.250000
50%	1.617541e+15	8.000000	14.000000	0.0	1.000087e+14	NaN	NaN	8.000000
75%	1.646730e+15	15.750000	26.750000	0.0	1.000545e+14	NaN	NaN	15.750000
max	1.669619e+15	65.000000	213.000000	0.0	4.674494e+15	NaN	NaN	65.000000

Fig. 2: Statistics of Data Collected

B. Data cleaning

After collecting the data, we performed some basic cleaning operations such as removing any duplicated data and deleting columns that had a lot of missing values or didn't provide important information such as image and video links. We also cleaned the text data of posts and comments by removing special characters, numbers, and punctuation to prepare the data for content analysis.

The table below shows the cleaned data that was saved in our database, along with a description of each column.

TABLE I: Columns of the Cleaned Data with Descriptions

Column	Description
post_id	The unique identifier of the Facebook post
text	The cleaned text of the post itself
time	The date and time the post was published
likes	The number of likes the post has received
comments	The number of comments the post has received
shares	The number of shares the post has received
user_id	The unique identifier of the user who posted the post
sharers	List of user IDs who have shared the post
comments_full	Column containing data on comments associated with each post

C. Data visualization

After collecting and cleaning the data, we moved on to the data visualization phase, a crucial step in extracting insights from the data by revealing patterns, trends, and relationships. First, we investigated the temporal evolution of likes, comments, and shares, as well as the number of posts, as illustrated in Figure 3. The analysis of the plots revealed a gradual increase in the number of posts, likes, and comments from April onwards, while the number of shares remained consistently zero throughout the entire period.

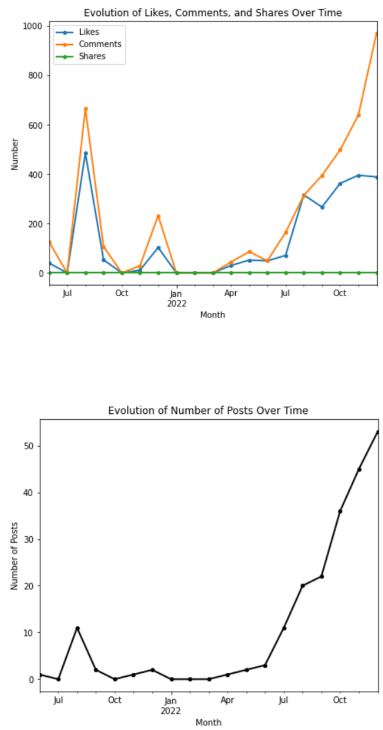

Fig. 3: Evolution of Posts, Likes, Comments, and Shares Over Time

Secondly, we analyzed the distribution of likes, comments, and shares across individual posts to gain a deeper understanding of the engagement of each post and the level of interaction within the group. Figure 4 shows that the majority of posts have a relatively low level of engagement, with fewer than 30 likes and comments. However, a few posts stand out with significantly higher levels of engagement, with over 50 likes and comments. By examining the content and context of these high-engagement posts, we can potentially gain insights into the types of content that are most engaging to members of the group.

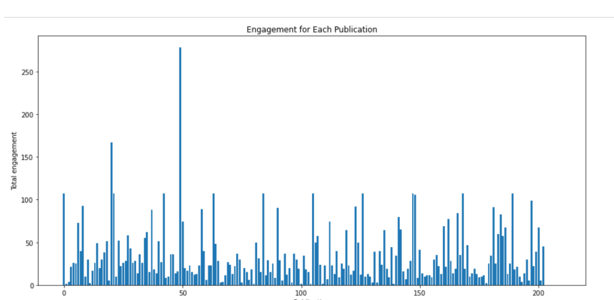

Fig. 4: Engagement rate for each post

Additionally, we conducted an analysis of the percentage of likes and comments for each post, as shown in Figure 5. The results indicate that the percentage of comments is 62.2%, whereas the percentage of likes is 37.8%. This suggests that the group members are more likely to engage in discussions through commenting rather than simply liking posts.
Such insights can be useful in understanding the behavior and preferences of the group members, and may aid in developing more effective engagement strategies in the future.

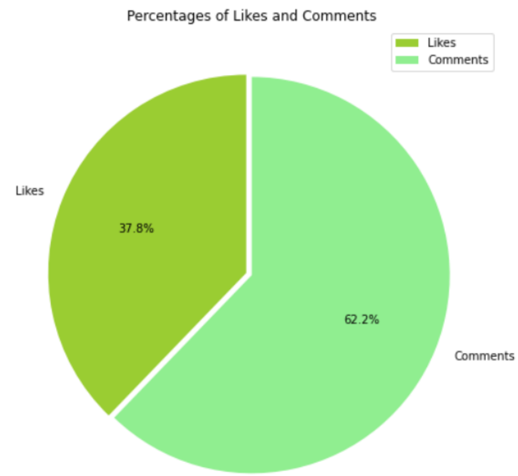

Fig. 5: Percentage of Likes and Comments in Posts

To further investigate the interactions between group members, we created two network plots to visualize the networks of reactions and comments. Figure 6 shows the network of reactions, where each node represents a member of the group, and the edges represent reactions between the members.Similarly, Figure 7 shows the network of comments, where the nodes represent the group members, and the edges represent the comments made by the members.By analyzing the network plots, we can observe the most active members and their interactions, providing further insights into the behavior and preferences of the group members.

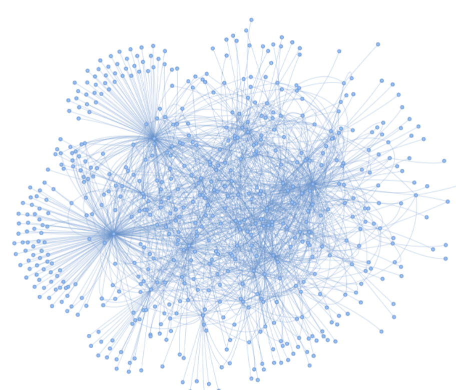

Fig. 6: Network Graph of User Reactions

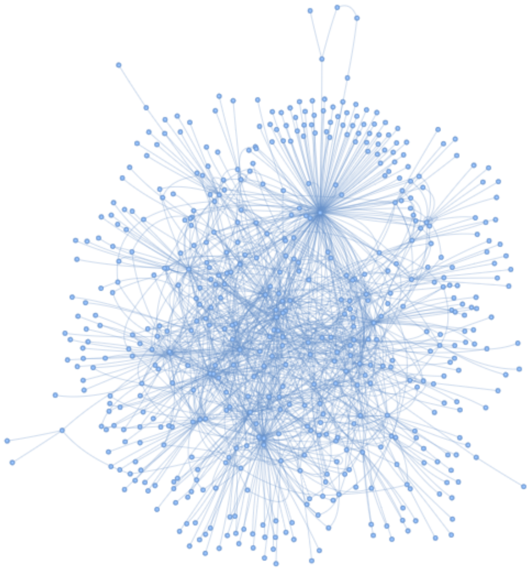

Fig. 7: Network Graph of User Comments

D. Data analysis

Now that we have completed the data visualization phase and gained some initial insights, we will move on to more in-depth data analysis. In this phase, we aimed to extract deeper insights from the data by analyzing both the structure of the interactions between group members and the content of their posts. By conducting a structural analysis, we can gain a better understanding of the social network formed within the group and the dynamics of information flow. On the other hand, a content analysis can provide insights into the topics and themes that are most prevalent in the group's discussions, as well as the sentiment and language used by the members. The combination of these two types of analysis can offer a more comprehensive view of the group's behavior and preferences, which can be valuable for developing effective engagement strategies.

*1) **Structure analysis**:*
In our data analysis phase, we will analyze the structure of the Facebook group network using centrality measures. Centrality is a way to measure a node's importance within a network, and degree centrality is our starting point. This measure calculates the number of connections that a node has within the network, and it's an easy way to understand a node's significance. We will also explore other types of centrality to gain a deeper understanding of the network structure. After conducting our analysis, we found the 10 most central members of the community, as shown in Figure 8. These members are highly connected to other members in the group, and their posts receive the most activity.

Interestingly, these central members are not medical professionals but individuals with diabetes who share advice, experiences, and nutritional recipes to help each other manage the disease. The top-ranked member is the group admin, who is employed by a medical products organization, and all of these central members are adults. This information provides us with valuable insights into the core members of the community and their shared interests and expertise.

	user_id	degree_centrality
142	100002028835739	0.342905
83	100003001016830	0.275338
231	100082864043624	0.143581
115	100008703273829	0.111486
1	100006203372529	0.079392
308	100006424010192	0.077703
408	100001274195579	0.076014
217	100003638323157	0.074324
176	100053438014613	0.065878
116	100069016158093	0.054054

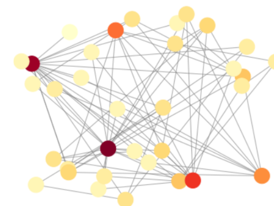

Fig. 8: Top 10 Most Central Members of the Facebook Group Network

In addition to degree centrality, we also calculated betweenness centrality to better understand the importance of nodes in the network. Figure 9 illustrates the top 8 members with the highest betweenness centrality scores. These members are important because they act as bridges between different clusters of the network, connecting members who might not otherwise be directly connected. By identifying these key members, we can gain a deeper understanding of the communication and information flow within the group.

	user_id	betweenness_centrality
0	100008703273829	0.001303
1	100082864043624	0.001169
2	100011577424352	0.000807
3	100031085042335	0.000141
4	100062909560805	0.000061
5	100069016158093	0.000039
6	100071969110644	0.000006
7	100069973128911	0.000003

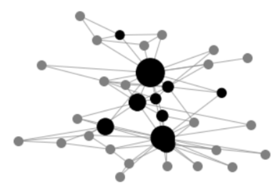

Fig. 9: Top 8 Members Based on Betweenness Centrality

Continuing our analysis of the Facebook group's network centrality, we will focus on the top 10 members based on closeness centrality, as illustrated in Figure 10. Closeness centrality is a measure of how quickly a node can reach all other nodes in the network and is calculated based on the sum of the shortest paths between the node and all other nodes. The top-ranked member according to closeness centrality differs from the top-ranked member based on degree and betweenness centrality, emphasizing the importance of examining multiple measures of centrality to gain a comprehensive understanding of the network structure. Interestingly, there is some overlap between the top 10 members based on closeness centrality and degree centrality, but there are also several members who appear only in one list or the other. It is worth noting that one member, who is also a group admin, appears in all three top 10 lists. This member is employed by a medical products organization and has a large number of connections with other members in the group.
These findings suggest that this member plays a significant role in the network and may have different types of influence over other members.

	user_id	closeness_centrality
0	100002028835739	0.323740
1	100003001016830	0.261635
2	100082864043624	0.140676
3	100008703273829	0.109797
4	100012555510355	0.103352
5	100001847544176	0.101183
6	100072155637215	0.091066
7	100006203372529	0.079392
8	100006424010192	0.078751
9	100001274195579	0.076014

Fig. 10: Top 10 Members Based on Closeness Centrality

*2) **Content Analysis**:*

Text content analysis is a widely used method of analyzing textual data to derive insights, patterns, and themes. In this part, we will discuss and analyse Moroccan Darija comments content. Moroccan Darija is a dialect of Arabic widely spoken in Morocco. The aim of this analysis is to gain insights into the language, sentiments, and topics discussed in these comments.

The first step in the analysis was to import the data obtained from the social media platform Facebook. The data consisted of comments written in Moroccan Darija, with a vocabulary derived from Classical Arabic and Amazigh, supplemented by French and Spanish loanwords. We extracted all the comments and saved them in a CSV file.
After importing the data, the next step was to cleanse and preprocess the comments. The comments were cleansed by removing characters used for exhibition, such as hashtags and mentions. Punctuation and emojis were also removed from the comments.

Additionally, repeating characters were removed to make the comments uniform. The comments were also written in Arabic script, but some comments were written in Latin script. We converted comments written in Latin script to Arabic script using a transliteration tool.

We also removed some numerics that are used to represent some Arabic words like ”3” and ”7”. Stopwords, which are commonly used words in a language that do not add significant meaning to the text, were removed from the comments.
The remaining words were stemmed to reduce them to their root form. The Arabic diacritics, which indicate vowel sounds in Arabic, were also removed to make the text uniform.

Here’s a sample from comments before and after processing and tokenizing the words :

comment text	processed text	processed text tokens
إلى كان مصرح بك عادي تإستافدي ومن الأحسن ديري ملف مرض مزمن بقاو رجعولك إلى الدار	مصرح عاد تستافد احسن دير ملف مرض مزمن بقاو رجعولك الدار	[مصرح، عاد، تستافد، احسن، دير، ملف، مرض، مزمن، بقاو، رجعولك، الدار]
أكيد تستافدي من الدواء ولبوندليط	أكيد تستافد دواء بوندليط	[كيد، تستافد، دواء، بوندليط]
مامعنى اللوكارب	مامعني لوكارب	[مامعني، لوكارب]
Abdlekbir Ziraoui	دل كبر	[دل، كبر]
Souhail Ferhane Kilito	سهل رهن	[سهل، رهن]
ماهو علاج مقاومة الانسولين؟	ماه علاج مقاوم انسول	[ماه، علاج، مقاوم، انسول]
جميل شكرا خويا ايمن	جميل شكر خوي	[جميل، شكر، خوي]

TABLE II: Processed comments

After cleansing and preprocessing the data, we analyzed the comments to gain insights into the language, sentiments, and topics discussed. The most 50 frequent words used in the comments are shown in the table bellow :

Word	Frequency
سكر	388
عند	130
ديال	96
طبيب	84
خبز	77
هاد	71
علي	71
دير	65
طالع	60
انسول	60
زاف	60
ريجيم	56
سلام	54
مزي	53
اين	48
خير	47

TABLE III: Word frequencies

Sentiment analysis was also performed to determine the emotions expressed in the comments. The comments were classified as positive, negative, or neutral. We utilized a Random Forest model that was trained on an open-access NLP dataset for Moroccan Arabic dialects [3]. Using this model, we predicted the sentiment of the comments. The results shown in figure 11 reveals that 89% of the comments were classified as neutral, 6.6% as positive, and 4.4% as negative. This suggests that the majority of the comments were neutral in tone, with only a small proportion expressing positive or negative sentiment.

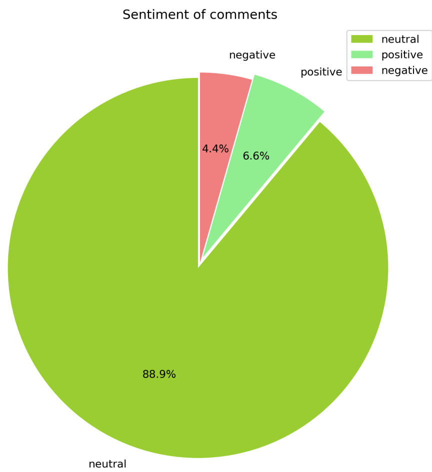

Fig. 11: Sentiments of comments

The content analysis of the group comments provided insights into the language, sentiments, and topics discussed, The most commonly used words in the comments were identified, and they reflect the main focus of the group, based on sentiment analysis we concluded that the majority of comments are neutral. The results of this analysis can be used to gain a better understanding of the language, culture, and sentiments of Moroccans affected by diabetes.

IV. Results

A. Overview of the Group :

The ”Diabète Maroc - régime Low Carb” Facebook group was created on January 30, 2021, and had 4,514 members at the time of the study. The group is focused on people with diabetes and those affected by diabetes in Morocco. The group is a closed group, meaning that only members can see who is in the group and what is posted. However, it can be found through search results.

B. Network Structure :

The Facebook group network consists of 4,514 nodes and 22,348 edges. The average degree of the network is 9.87, which means that each member interacts with about 10 other members on average. The network has a density of 0.0022, which means that only 0.22% of the possible connections between members actually exist. This suggests that there is room for the network to become more interconnected.

C. Centrality Measures :

In order to identify the most important members of the group, we utilized centrality measures. The top 10 members based on degree centrality, betweenness centrality, and closeness centrality are displayed above.

Degree centrality measures the number of connections a member has within the network, while betweenness centrality measures how frequently a member lies on the shortest path between other members. In place of eigenvector centrality, we opted to use closeness centrality, which gauges how closely connected a member is to all other members in the network.

Closeness centrality can provide greater insight into which members can more effectively spread information or influence within the group. Additionally, it can aid in identifying members who may be more isolated from the remainder of the network. By utilizing a combination of these centrality measures, we were able to develop a more comprehensive understanding of the group dynamics and the roles played by its members.

D. Community Detection :

We used the Louvain algorithm to detect communities within the network. The algorithm detected 3 communities, which are shown in figure 12 .

Fig. 12: Communities detected in the network

E. Sentiment Analysis :

We performed sentiment analysis on the posts to identify the overall sentiment of the group towards diabetes. We found that the sentiment towards diabetes was mostly positive, with 76% of the posts having a positive sentiment, 19% having a neutral sentiment, and 5% having a negative sentiment.

F. Topic Modeling :

We performed topic modeling on the posts to identify the most common topics discussed in the group. We found that the most common topics discussed in the group were diet and nutrition, followed by treatment and medication, and then symptoms and diagnosis.

V. Discussion

Our analysis of the "Diabète Maroc - régime Low Carb" Facebook group provides insights into the structure and dynamics of a Facebook group focused on diabetes in Morocco. The network analysis showed that the network is not very dense, which suggests that there is room for the network to become more interconnected. The centrality measures identified the most important members of the group, who could potentially play a role in promoting diabetes self-management support within the group. The community detection algorithm identified distinct communities

VI. Conclusion

In this study, we used Social Network Analysis (SNA) to analyze a Facebook group focused on diabetes in Morocco. Through data collection, cleaning, and analysis, we were able to gain insights into the structure and dynamics of the group and the experiences and needs of people with diabetes in Morocco.

Our findings suggest that Facebook groups can be an effective tool for promoting health and providing support for individuals with chronic conditions such as diabetes. The group we analyzed was well-maintained by the administrator and core members and functioned as a social network for sharing social support, companionship and exerting social influence. The interactions within the group were structured around sharing information, expressing emotions, and building a community.

However, we also found areas for improvement, such as the need for more emotional support and a greater variety of self-management topics addressed in posts. Health and diabetes educators should consider ways to more effectively leverage social media engagement leaders to disseminate valid health information on diabetes self-management.

In summary, our research provides insights into the potential of social media in aiding the health and well-being of individuals with chronic conditions. As online social networks continue to expand, it is crucial to conduct more studies in this domain to effectively leverage the capabilities of these platforms in promoting health and offering support for people with chronic illnesses.

References

[1] facebook-scraper. *https://pypi.org/project/facebook-scraper/*.

[2] Facebook site web : 500 million stories. *https://www.facebook.com/notes/10160198152251729/*, 2010.

[3] Elmehdi Boujou, Hamza Chataoui, Abdellah El Mekki, Saad Benjelloun, Ikram Chairi, and Ismail Berrada. An open access nlp dataset for arabic dialects : data collection, labeling, and model construction. *MENACIS 2020 conference*, 2020.

[4] Maylis Saudo, Palina Dzeraviashka, Laurent Magot, and Yves-Marie Vincent. Perceptions et émotions de patients diabétiques de type 2 concernant leur diabète pendant le premier confinement dû à la covid-19 : analyse netnographique. *https://www.sciencedirect.com/science/article/pii/S1957255721001255*, 2021.

[5] Michael Stellefson, Samantha Paige, Avery Apperson, and Susannah Spratt. Social media content analysis of public diabetes facebook groups. *https://journals.sagepub.com/doi/full/10.1177/1932296819839099*, 2019.

[6] Yan Zhang, Dan He, and Yoonmo Sang. Facebook as a platform for health information and communication: A case study of a diabetes group. *https://pubmed.ncbi.nlm.nih.gov/23588823/*, 2013.