IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 2, 2021, accepted June 18, 2021, date of publication June 22, 2021, date of current version July 1, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3091537

Social Network Analysis and Visualization of Arabic Tweets During the COVID-19 Pandemic

AMAL A. AL-SHARGABI AND AFEF SELMI
Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia

Corresponding author: Amal A. Al-Shargabi (a.alshargabi@qu.edu.sa)

This work was supported by Qassim University through the Deanship of Scientific Research, during the academic year 1441 AH/2020 AD, under Grant coc–2020-1-1-L-10058.

ABSTRACT Twitter data have been used widely in times of pandemics and crises. Currently, the world is suffering from the outbreak of new coronavirus disease 2019 (COVID-19), in which the COVID-19 virus has infected more than one hundred and twenty million people worldwide with more than two million deaths. Consequently, people tend to share COVID-19-related content on social media extensively. As a new pandemic, only a small number of studies have been conducted to analyze COVID-19-related tweets, and even fewer were meant for Arabic tweets. This research explores the influence of the COVID-19 pandemic on Saudi users' tweeting behavior. In particular, the research adopts a social network analysis (SNA) for COVID-19 Arabic tweets. This approach is interesting, as it is based on analysis of social structures, such as Twitter users and the relationships among them, through the use of networks and graph theory without the contents of the tweets themselves. Based on 8905 collected Arabic tweets, this research resulted in three main contributions: 1) a visualization of the social network for COVID-19 tweets of Saudi users, 2) an identification of information sources that Twitter users employ during the COVID-19 pandemic, and 3) an identification of the most popular influencers among users of COVID-19 tweets. The results of this study may help identify the most popular Twitter influencers and those who deliver the information to Twitter users, utilizing them to increase awareness and deliver information and instructions to overcome the COVID-19 pandemic.

INDEX TERMS COVID-19, social network analysis, twitter analysis.

I. INTRODUCTION

New coronavirus disease 2019 (COVID-19) is an infectious disease that recently appeared in December 2019 as lung syndrome. The symptoms mostly include fever, cough, dyspnea and viral pneumonia [1]. The first case was reported in Wuhan, China and soon after, it spread rapidly all around the world. At the time of this writing, 219 countries and territories have been affected, with approximately 130 million reported cases (https://www.worldometers.info/coronavirus/). On March 11, 2020, the World Health Organization (WHO) declared the outbreak of the new coronavirus a global pandemic. As of May 21, 2021, there have been 165,069,258 confirmed cases of COVID-19 worldwide, including 3,422,907 deaths. (https://covid19.who.int).

Consequently, governments all over the world were and are still looking for ways to contain the disease and alleviate its disastrous consequences to their people's health and economy.

Several preventive measures have been implemented by governments to stop the spread of the virus. These include social distancing, school and business closures, travel bans and others. As a result, the daily practices of people have been affected, and most of the social interactions and conversations have moved online on social media platforms.

Social media outlets, such as Twitter, are real-time communication platforms that allow individual users to create and share content with anyone on Earth or with many people simultaneously [2]. This powerful tool has become central during the COVID-19 pandemic so that people can find the latest information and follow what is happening in real-time [3].

With the enormous number of tweets associated with COVID-19 shared and retweeted daily, different research directions have been proposed to make use of these loads of relevant tweets. One line of research was to investigate

The associate editor coordinating the review of this manuscript and approving it for publication was Camelia Delcea.

the behavior of people and how they access information of COVID-19 via Twitter [4]–[7]. For example, Yum [4] conducted a study in the United States to explore the key players of COVID-19 based on a collected Twitter dataset. Additionally, Jain and Sinha [5] have identified the influential users on Twitter using COVID-19-related tweets. Moreover, Schild *et al.* [6] studied the emergence of Sinophobic behavior on Twitter during the COVID-19 pandemic. The second line of research was to apply text mining methods to social network data. These include sentiment analysis, e.g., [8]–[14], and misinformation modeling, forecasting and detection, e.g., [15]–[19]. The third line of research was to diagnose COVID-19 patients using machine and deep learning methods, e.g., [20]–[24].

This research falls in the first research direction in which the study explores how people access information on COVID-19 via Twitter social networks using Arabic tweets. In other words, the study focused on revealing social activity patterns of the Arabic world. According to [25], there were 4,514,136 Arabic tweets related to COVID- 19 that were available for public research and analysis by the beginning of April 2020 [25]. Nevertheless, very few studies have been published and used these enormous loads of relevant Arabic tweets.

To this end, the study collected Twitter data streams in Saudi Arabia since it was ranked globally in 2021 as the eighth in terms of the number of Twitter users, with more than 12 million active users. Similar to other countries, COVID-19 has infected many people in all Saudi districts. By the end of March 2021, Saudi Arabia had confirmed cases of more than 390,000 and more than 6500 deaths. This study is in line with other COVID-19 studies that have been conducted based on region-specific data, such as [8] in Brazil, [26] in Romania, and [27] in Egypt, Kuwait, and Saudi Arabia.

However, this study followed a social networking analysis (SNA) approach to investigate how COVID-19 influences Saudi Twitter users' behavior and how public key players such as news channels play an essential role in communications.

SNA is the process of examining social structures through network usage and graph theory [28]. SNA describes network structures in terms of nodes and edges. The nodes represent the people within the network, and the edges represent the relationships or interactions that connect them. Such networks are mostly visualized graphically via sociograms in which nodes are represented as points and edges are represented as lines. These visualizations provide a means of qualitative assessment of the network using attributes of interest [28]. This approach is special as it gives an overview of the current situation using simple data, i.e., users and interactions among them.

In this study, the tweet owners are analyzed and visualized to form a directed social network graph using different centrality measures, including degree, closeness, betweenness, eigenvector, and page rank.

Directed graphs do not base on symmetric edges between vertices. In a directed graph, if a and b are two vertices connected by an edge (a, b), this does not necessarily mean that an edge connecting (b, a) also exists. Directed graphs are most common as they do not impose the restrictive assumption of symmetry in the connection demonstrated by the edges [28]. This kind of graphs is suitable to model twitter data as connections among different users may occur from one side only.

Based on 8905 collected tweets, this study identifies information sources and influencers in social networks through centrality measures. Moreover, the study provides visualization for the entirety of social networks and communities within the network and uses word cloud analysis of the interests and topics of individuals for COVID-19. To the best of our knowledge, this is the first work that explores Arabic COVID-19 tweets using a social network analysis approach.

The main contributions of this study are as follows:

1) A visualization of a social network for COVID-19 tweets of Saudi users.
2) An identification of information sources that Twitter users are using during the COVID-19 pandemic.
3) An identification of the most popular influencers among COVID-19 tweet users.

By identifying the key information sources and influencers on Twitter, the Saudi Ministry of Health and other authorities can understand the awareness of tweeter behaviors. Additionally, the identified information sources and influencers can be used by the government to deliver instructions and safety protocols to the public.

This paper is structured as follows: Section II presents the related work. Section III describes the proposed method. Section IV provides the experimental results. Section V discusses the main findings. Section VI concludes the paper.

II. RELATED WORK

Numerous COVID-19 twitter studies focused on the sentiment analysis of the tweets' text, e.g., [12]. In the present study, our focus is on the social network analysis, and thus, we summarize the studies that follow this approach. Social network analysis applies graph theory to explore the social structure within a network using nodes and edges [28]. People in the network are represented using nodes and connections among them are represented using edges.

Table 1 shows a summary of the related studies that applied social network analysis in the context of COVID-19.

Jain and Sinha [5] have focused on merging the profile activity and underlying network topology to designate online users with combined effects. That study aims to build a weighted correlated influence (WCI) approach to compute each user's influence scores in the Twitter network. The study results have shown the performance of the proposed WCI compared to existing methods.

Tahmasbi *et al.* [6] collected two datasets to analyze how online Sinophobia has evolved as a COVID-19 pandemic.

TABLE 1. A summary of twitter and COVID-19 studies.

Study	Location	Dataset			Language	Type of analysis		Measures and algorithms of SNA
		Timeframe	Network type	Size		SNA used?	Other analysis	
[5]	_	Real-time Twitter dataset	Users, Interaction, Hashtag	Two datasets (15018 nodes, 21509 edges and 18473 nodes, 22833 edges)	English	Yes	Profile activity	- Degree - Betweenness - PageRank - Eigenvector - Weighted Correlated Influence
[6]	China	Nov. 01, 2019 to Mar. 22, 2020	Terms	222,212,841 tweets	Chinese	Yes	Content analysis	
[7]	China	Mar. 17 to Apr. 28, 2020	Target words (antisocial keywords)	40,385,257 tweets	English	Yes	Lexicon-based Perspective API	
[9]	Korea	Feb. 29, 2020	Text, Hashtags	43,832 users and 78,233 relationships	Korean	Yes	Content analysis	- Diameter (geodesic path) - Modularity
[29]	United Kingdom	7 days	Hashtag	10,140 tweets	English	Yes	Content analysis	- Betweenness - Clauset-Newman-Moore
[30]	_	Jan. 21 to May 03, 2020	Users	2,558,474 Tweets	English	Yes	_	- Degree - Betweenness - Density
[31]	South Korea	Between Feb.10 and 14, 2020	Users	30,168 tweets	Different languages	Yes	Content analysis	_
[32]	_	Mar. 24 to Apr. 09, 2020	Tweets	23,830,322 tweets	English	Yes	Keyword trend Topic modeling	- Coherence Scores - Density - Degree
[33]	United States	Between Apr. 16 and Apr. 22, 2020	Key players	2,864 users, 2,775 relationships	English	Yes	Content and sentiment analysis	- Degree - Word Cloud - Clauset-Newman-Moore
[34]	Spain	From March 16, 2020–April 27, 2020	Organizations, types	The list of 123 active media firms	Spanish	Yes	Social services approach	- Density - Diameter - Average path - Length, Transitivity - Centralization - Edges, Dyad - Degree, Centrality - Ego-density
[35]	China	From December 31, 2019, to February 20, 2020	public opinion	4056 topics	Chinese	Yes	Content Analysis	- Trend analysis - Content analysis - Social Network Analysis and Visualization (High-frequency keywords)
[36]	Different countries	From 12/31/19 to 05/15/20	Countries	195 countries	English	Yes	_	- Closeness Centrality - Degree centrality - Betweenness Centrality - Diameter of the network - Clustering coefficient
[37]	**Saudi Arabia**	**Feb. 2020**	**Tweets with relevant hashtags**	**53,127 tweets**	**Arabic**	**No**	**Sentiment analysis Naïve Bayes**	_
[38]	**Saudi Arabia**	**Dec. 2019 to Apr. 2020**	**Tweets**	**1,000,000 tweets**	**Arabic**	**No**	**K-means, Logistic regression, Support vector Naïve Bayes**	_
[39]	**Saudi Arabia**	**Jan. 01 to Apr. 30, 2020**	**Tweets**	**4,514,136 tweets**	**Arabic**	**No**	**HITS and PageRank**	_

Awal *et al.* [7] have proposed an automatic online antisocial behavior annotation framework to annotate COVID-19 tweets. This study conducts an empirical analysis to identify the targets of antisocial behaviors and factors that influence the dissemination of online antisocial content during the COVID-19 pandemic.

The study of Park *et al.* [9] has applied social network analysis in conjunction with content analysis to analyze how COVID-19-related issues circulate on Twitter in Korea. The study was based on three approaches: social network analysis, news channel classification, and content analysis. The study results have shown that the spread of information transmission was faster on the COVID-19 network than the other networks. A study conducted in the United Kingdom [29] has used a COVID-19 dataset to produce sentiment analysis for COVID-19 Tweets. The study results have shown that most of the studied tweets have positive content, while only 15% percent have negative content. The study results have shown that Twitter users have grouped in social network graph clusters, and influential users are ranked using the betweenness centrality metric.

Sunmoo *et al.* [30] have applied social network analysis techniques on COVID-19 tweets mentioning cannabis and opioids to detect the similarity between groups and identify the stakeholders on the topics within Twitter. The study results have shown that the betweenness centrality algorithm could better reveal the spread of information and disinformation on illicit drug use during COVID-19 than the degree centrality algorithm.

Kim [31] conducted semantic network analysis and computer-assisted content analysis to evaluate the South Korean public's attention toward the COVID-19 epidemic. The content analysis identifies the social factors that reduce user's incivility comments about the COVID-19 situation. Semantic network analysis is used to analyze how people express their thoughts.

Ordun *et al.* [32] used uniform manifold approximation and projection (UMAP) and DiGraphs to assess the speed of information dissemination and network behaviors for COVID-19 tweets. The dimensionality reduction algorithm, UMAP, was applied to visualize latent Dirichlet allocation (LDA) to generate the topics that discuss case spread and healthcare workers. The study results have shown that UMAP has allowed a good clustering of different topics generated by LDA.

Another study was conducted on a Twitter dataset to investigate social networks among public key players for COVID-19 in the United States [33]. This study explored the in-degree centrality to examine the relationship between vital public players and social networks. It uses word cloud analysis to visualize the individual interest for COVID-19. The study results have shown that topic-based networks and person-based networks play different roles in the social network.

Belso-Martínez *et al.* [34] have focused on social network analysis techniques and the social services approach to study the COVID-19 pandemic in Spain (Valencian region). The study aims to analyze the roles and positions of the participating organizations in the ecosystem. The quadratic assignment procedure (QAP) regression method is used to find the critical factors that influence the relationship between the different organizations. The results have shown that SNA is a suitable tool for identifying and analyzing the relationship between organizations.

Zhao *et al.* [35] have used the Sina microblog hot search list to evaluate the attention of the Chinese public toward the COVID-19 epidemic. They used the ROST Content Mining System to analyze the collected data for segmentation and sentiment analysis and build a social network of public opinion content. The obtained results can help public health professionals provide better communication to the public in terms of identifying the public needs, making prevention measures to control the spread of COVID-19.

Montes-Orozco *et al.* [36] have proposed a methodology to identify COVID-19 spreaders in different countries. This study aims to use the vertex separator problem in multiplex complex networks to analyze the effect of socio-cultural and economic factors on the spreading of COVID-19 in each country. The results have shown that the methodology can help classify the spreader countries based on their numerical values in socioeconomics, population, health, etc.

In Saudi Arabia, only a few studies have been found. Alhajji *et al.* [37] have conducted a sentiment analysis for COVID-19 tweets in Saudi Arabia. The study used the naïve Bayes model to predict a Twitter user's sentiment based on their tweets. Most of the studies showed positive tweets and attitudes for the infection control measures to combat COVID-19.

Additionally, Alsuias *et al.* [38] have combined qualitative and quantitative studies to identify and analyze one million Arabic tweets related to the COVID-19 pandemic. Three types of analysis are performed to investigate the collected data: cluster, rumor detection, and source type prediction. The study results have shown that health professionals and academics write 60% of the rumor-related tweets. One more study was focused on large Arabic tweets collected during the COVID-19 pandemic to identify information super-spreaders of COVID-19 [39].

Alqurashi *et al.* have applied two ranking algorithms, hyperlink-induced topic search (HITS) and PageRank, to determine the influencers in the Arabic content of Twitter. The study results have shown that HITS and PageRank discovered a similar subset of spreaders in which 40% were verified as Twitter accounts, and 50% of the influential accounts were located in Saudi Arabia.

Based on our review, we can derive the following observations:

1) During the COVID-19 epidemic, social media, especially Twitter, has played an increasingly important role in public health emergencies. In particular, it can help study the spread of the disease and find the latest

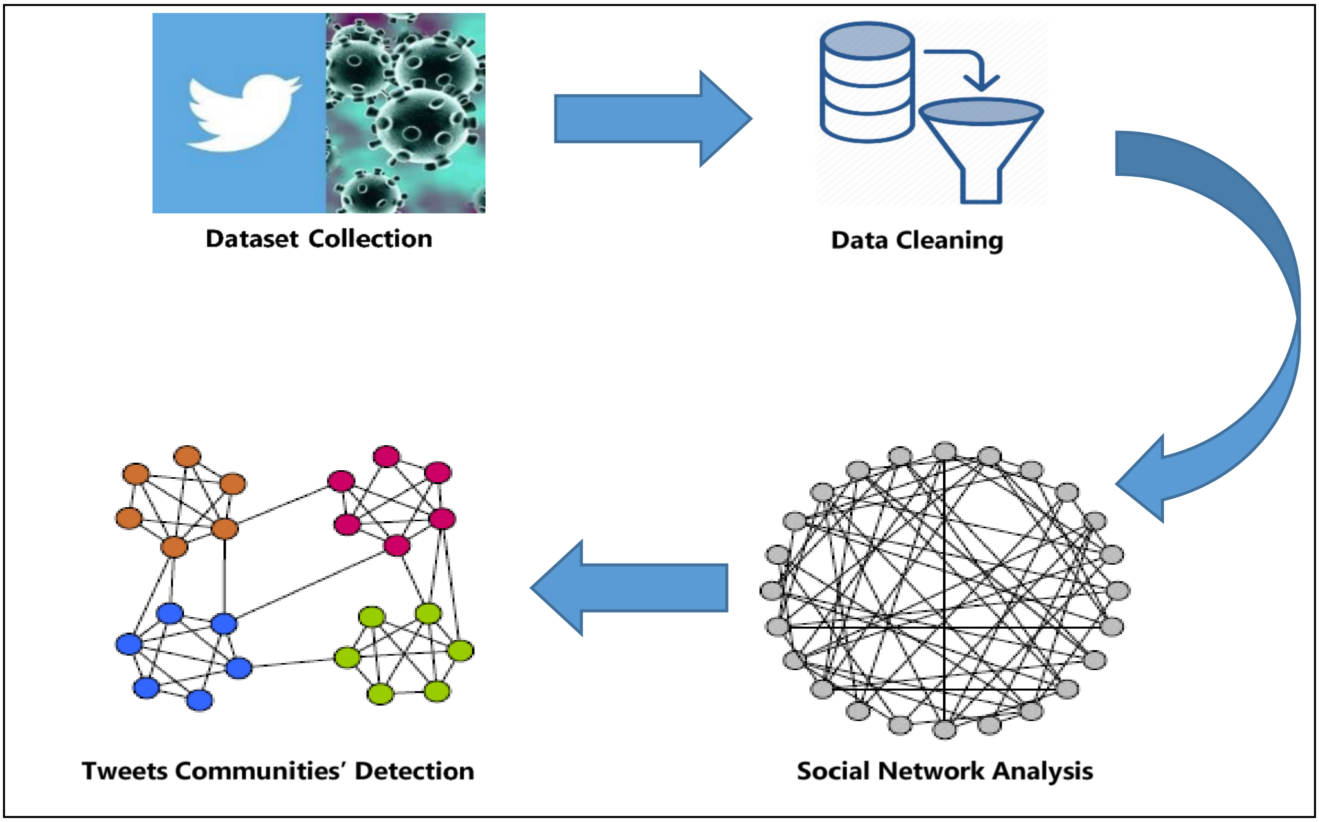

FIGURE 1. The methodology followed in the study. First, Arabic twitter data related to COVID-19 are stream-based data. Second, the collected data are filtered. Third, twitter data are analyzed and visualized to form a directed social network graph. Last, twitter data are clustered into several communities with similar characteristics.

information on COVID 19 and tweets from public health experts [3].

2) The majority of studies have focused on analyzing the English content of Twitter. Only limited studies [37]–[39] have been conducted in analyzing COVID-19-related Arabic tweets. Most of the studies examining Arabic tweets have focused on sentiment analysis and topic modeling of Twitter data. None of them used social network analysis for COVID-19-related tweets to explore the pandemic's influence on Saudi Twitter users' tweeting behavior.

III. THE PROPOSED METHOD

Analyzing Twitter data allows for understanding public behavioral responses, and longitudinal tracking allows for identifying changes in responses. By applying social networking analysis for tweets about COVID-19, various informative resources of tweets are identified. Moreover, people's interest and awareness about COVID-19 can be detected. Twitter is particularly suitable for network analysis. The brief tweets that users openly share with followers contain a wealth of data as they share their thoughts and knowledge and provide links to other sources of information.

The followed method to solve this problem uses a similar effort to the research work of Bermudez and colleagues on social network analysis (SNA) [40]. In particular, the methodology of the current study consists of four main phases (Fig. 1): dataset collection, data cleaning, social network analysis, and tweet community detection.

A. DATASET COLLECTION AND ANALYSIS

The dataset collection phase aims to obtain the raw data from Twitter to be used in the social network analysis process. Existing datasets for COVID-19 Arabic Twitter data, such as the dataset available in [41], cannot be used to visualize the social network. It contains tweet texts only but not users and the relationships among them. We attempted to obtain the network inputs from the dataset mentioned above with the hydrator tool's help. However, the retrieved Twitter objects did not contain source, target, or strength attributes essential for SNA.

In this study, data was collected and analyzed using Gephi software (https://gephi.org). Gephi is an open-source and free visualization and exploration software for graphs and networks. It was used mainly in our study to collect the data, explore it, and apply the social network analysis.

To collect our data, an application was submitted to the Twitter Developers Website. After the application was approved, suitable COVID-19 hashtags were entered with a chosen geographical coordinate using Gephi Software and the Twitter Streaming Importer plugin.

The tweets were collected from Riyadh city, the capital of Saudi Arabia. The reason behind choosing Riyadh was and still is that the city has the highest infection rate of COVID-19.

To obtain an idea of the number of cases in Riyadh and other cities, Table 2 shows the number of cases recorded on 07/18/2020 for the high infected Saudi cities according to the Saudi Ministry of Health (https://www.moh.gov.sa).

TABLE 2. The number of confirmed cases in the different Saudi cities as of 07/18/2020.

City	Number of confirmed cases
Riyadh	51331
Jeddah	28618
Mecca	27016
Dammam	15701
Medina	15676
Alhufuf	14414
Altaief	8344
Alqatif	7091
Alkubar	6324

On this day, Riyadh city was the highest infected city with 51,331 cases, followed by Jeddah and Makkah with a number of 28618 and 27016 confirmed cases, respectively. The hashtags used were the most common ones provided by the Saudi Ministry of Health and the hashtags used in related Saudi studies.

The settings used in data collection are presented in Table 3.

TABLE 3. Parameter settings used to collect tweets.

Parameter	Value	
Network type	User Network	
Location	Riyadh Southwest point (24.281790, 46.349903) Northeast point (25.169916, 47.388111)	
Timeframe	Between 16- 30 July 2020 (Average of new cases was 2000++ per day)	
Hashtags	#الصين (#China)	#كوفيد_19 (# COVID19)
	#الصحة (#the_health)	#فيروس_كورونا_المستجد
	#كورونا (#Corona)	(#novel_coronavirus)
	#فيروس_كورونا	#كلنا_مسؤول
	(#Corona_virus)	(#We_are_all_responsible)
	#الكورونا	#نعود_ بحذر #
	(#The_Coronavirus)	Go_back_with_caution)
	#كورونا_الجديد	#عش_بصحة (#Live_healthy)
	(#New_Corona)	#الوقاية_من_كورونا
	#فايروس_كورونا	(#Coronavirus_prevention)
	(#Corona_virus)	#أسئلة_كورونا
		(#Corona_questions)
Language	Arabic	

1) DATASET FEATURES DESCRIPTION

The collected dataset contains Arabic tweets of X features as follows:

- ID: Twitter user ID
- Label: Twitter user label
- Timeset: Tweet date and time
- Created_at: The place where the Twitter user was created
- Friends_count: Number of friends
- Followers_count: Number of followers
- Location: Place where the Twitter user was when writing the tweet

B. DATASET PREPROCESSING

To be able to obtain accurate social network visualization, data preprocessing shall take place. Thus, after data were loaded into the Gephi environment, tweets were filtered, and unrelated tweets were excluded from the analysis. Ajao *et al.* [42] indicated that locations that can infer Twitter are of three types: user home residence, tweeting location, and message context.

In this work, the first type was considered. As shown in Fig. 2, only accounts of users who reside in Saudi Arabia were included (colored in yellow), while others who interact and respond with Saudi residences from outside Saudi Arabia were excluded (colored in blue).

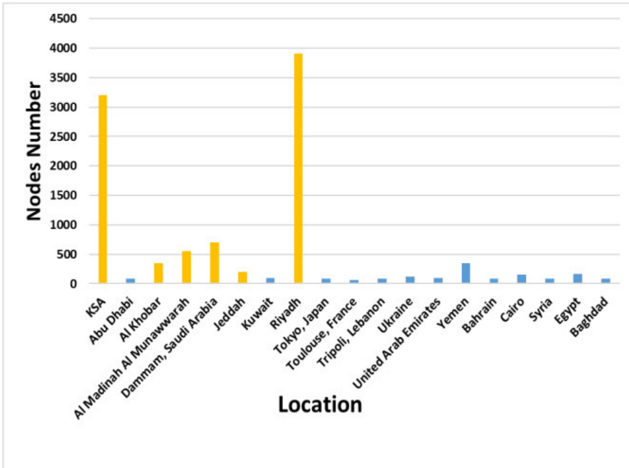

FIGURE 2. Number nodes in the different locations. Only users who reside in Saudi Arabia were included (colored in yellow), while others were excluded from the analysis.

The initial number of nodes was 10391 and after excluding the unrelated tweets, the network's resulting dataset became 8905.

1) AN OVERVIEW OF THE STREAMED DATASET

The collected tweets were of three types: mentions, quotes, and retweets. Some of the tweets may have more than one type. To obtain an idea about the collected tweets, Fig. 3 shows the number of tweets in each type. As the figure shows, most of the tweets were of the mention type followed by retweets and quotes.

Fig. 4 depicts the users' followers number range. The figure shows that the most common followers' range was 0-500 followers with a total number of 2844 users. This is followed by the 1001-4000 range with 1399 users.

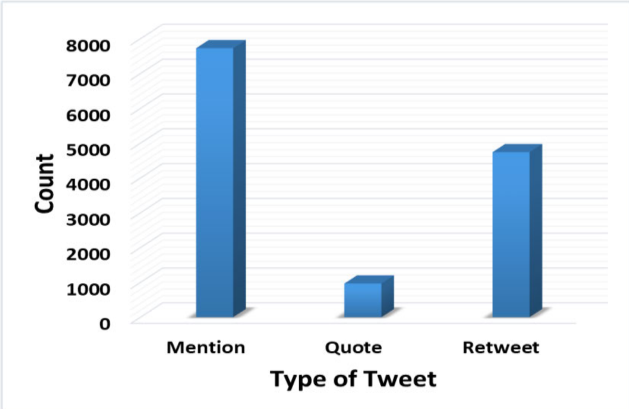

FIGURE 3. Counts of tweet types. Mention tweets are the most common tweets followed by retweets.

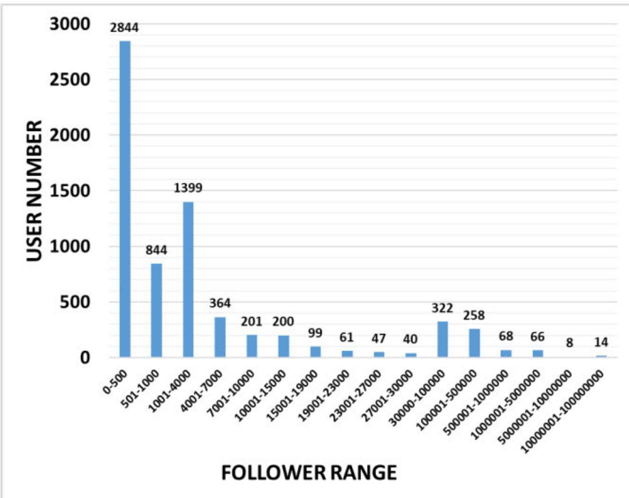

FIGURE 4. Number range of user followers. Most of the users (approximately 2800 users) have less than 500 followers. Many of them (approximately 1400 users) have between 1001 and 4000 followers.

C. SOCIAL NETWORK ANALYSIS

In the social network analysis phase, tweet owners were analyzed and visualized to form a directed social network graph. A few algorithms were used to obtain a better view of the network used to gain insights from the visualization of the tweet owners and their connections.

One of the key characteristics of social network analysis is finding prominent and influential nodes in the social network. Centrality measures are among the most commonly used indices based on network data. They usually represent the importance of a node; this can be its status, visibility, structural power, or prestige. Among the popular centrality measures are degree, closeness, betweenness, eigenvector, and page rank. In the following, these measures are explained briefly.

1) DEGREE

Degree centrality is the most straightforward centrality measure to compute. A node's degree is simply a count of how many social connections it has, i.e., edges. The degree centrality for a node is its degree. Separate in- and out-degree measures are suitable for nonsymmetric data. A node's degree centrality $C_D(i)$ reflects its number of relationships [43] as shown in (1).

$$C_{\mathrm{D}}(i) = \sum\nolimits_{\substack{j=1 \\ (i \neq j)}}^{N} x_{ij} \tag{1}$$

2) CLOSENESS

A closeness measure conceives of a node as central to the extent that it is related to other nodes via short geodesics. Closeness centrality $C_C(i)$ is defined only for sets of nodes that are mutually related via finite geodesic distances. Nodes linked to others via short geodesics have comparatively little need for intermediary nodes and thus possess relative independence in managing their relationships [43]. Refer to (2).

$$C_{\mathrm{C}}(i) = \frac{N-1}{\sum_{j=1}^{N} d_{ij}} \tag{2}$$

3) BETWEENNESS

Betweenness centrality is a widely used measure that captures a person's role in allowing information to pass from one part of the network to another. A widely used measure $C_B(i)$ reflects node i's "betweenness" [43]. Refer to (3).

$$C_{\mathrm{B}}(i) = \sum\nolimits_{\substack{j=1 \\ (j \neq i)}}^{N} \sum\nolimits_{\substack{k=1 \\ (k \neq i)}}^{j-1} \frac{g_{jk}(i)}{g_{jk}} \tag{3}$$

where g_{jk} is the number of geodesic paths linking nodes j and k, and $g_{jk}(i)$ is the number of those geodesics on which node i occupies an intermediary location.

4) EIGENVECTOR AND PAGE RANK

Eigenvector centrality measures a node's importance while considering the importance of its neighbors. It is determined by performing a matrix calculation to determine what is called the principal eigenvector using the adjacency matrix. A variant of eigenvector centrality is the PageRank algorithm, which is used to rank web pages.

D. TWEETS' COMMUNITY DETECTION

Once the social network was visualized, tweet owners were clustered into several communities with similar characteristics and a stronger relationship. Clauset, Newman, and Moore's (CNM) method [44] is one of the most important methods for community detection in networks, and, currently, it is one of the most studied methods for this purpose (Fig. 5). The method is a heuristic method aiming at the fast identification of communities suited for large-scale networks. As CNM is a greedy heuristic method, and its application may lead to partitions that differ from the optimal solution. In many cases, the modularity obtained is much lower than the values found by other methods.

As such, tweet owners are divided into distinct communities. Each community performs according to a related and

CNM algorithm
Input: A network $G = (V, E)$
Output: A community structure $C = \{C_a, a = 1, \ldots, nc\}$
(1) Calculate the initial values for $\boldsymbol{M}(10)$;
(2) Calculate the initial modularity value Q;
(3) $nc \leftarrow n$;
(4) **repeat**
(5) Join the pair of communities C_a and C_b corresponding to the highest value of $\boldsymbol{M}(\max(\boldsymbol{M}) : \boldsymbol{M}_{ab}$;
(6) Update matrix $M(12)$;
(7) $nc \leftarrow n - 1$;
(8) $Q = Q + M_{ab}$;
(9) **until** $\boldsymbol{max}(\Delta \boldsymbol{Q}) < \boldsymbol{0}$

FIGURE 5. Steps of CNM algorithm. CNM greedily optimizes the modularity to detects community structure in a network of V vertices and E edges using a hierarchical agglomeration approach.

connected tweet owner. Hence, insights related to the most trusted resources in the network were perceived.

IV. EXPERIMENTAL RESULTS

In the following, the results are presented in terms of the description of the extracted social network, the centrality measures results, the detected communities, and the trend of analysis.

A. OVERVIEW OF THE NETWORK

The diameter is representative of the linear size of a network. If nodes A-B-C-D are connected, going from A->D, this would be the diameter of 3 (3-hops, 3-links). For our studied network, the diameter was calculated as 4 hops. For the preliminary overall graph visualization, Fig. 6 demonstrates the social network during the day of streaming.

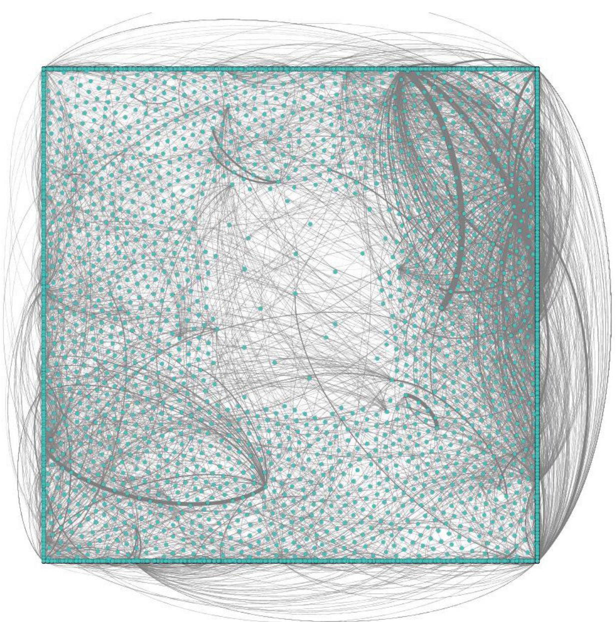

FIGURE 6. Preliminary overall social network structure of twitter users. The formed network shows the users and the interactions among them.

The network layout contains the nodes (users) and the interactions among them (edges). The layout was created based on the ForceAtlas2 algorithm.

According to Fig. 6, we see that the interactions among some users are high, such as the interactions shown in the right side of the figure, while they are low for other users, such as the interactions in the top left of the figure. To show the general overall graph metrics, Table 4 summarizes the overall graph metrics of the Twitter social network.

TABLE 4. Overall graph metrics of twitter users.

Metrics	Values
Graph total edges	13435
Connected components (Strongly connected)	8865
Weekly connected vertices	1546
Graph total vertices	8905
Average geodesic distance	1.429
Maximum geodesic distance (diameter)	4
Modularity	0.946
Average Clustering Coefficient	0.265
Total triangles	685

B. NETWORK MEASURES

This section focuses mainly on the centrality measure, which is used to identify the prominent nodes in the network. Specifically, the section reports the characteristics of every vertex based upon in-degree and out-degree, closeness, betweenness, eigenvector, and PageRank measures.

1) IN-DEGREE AND OUT-DEGREE CENTRALITY RESULTS

Table 5 represents the in-degree and out-degree centrality of the social network. As shown in Table 5, the highest in-degree in the Riyadh network is 283, which indicates that the maximum number of interactions with a node is 283. Moreover, the highest out-degree is 29, and this indicates the number of interactions a node has. Moreover, there are some nodes that interacted with other nodes while their owners have no tweets, whereas some nodes have some created tweets with no interactions with other nodes. Furthermore, the average degree for the network is 1.5, which indicates that normal Twitter users who have a moderate active tweeting and reacting behavior own most of the nodes.

TABLE 5. Degree overview of twitter user network.

Degree Centrality Measures	Values
Maximum In-degree	283
Minimum In-degree	0
Maximum Out-degree	29
Minimum Out-degree	0
Average Degree	1.509

To obtain a better idea of the in-degree and out-degree measures of the created network, Fig. 7 and Fig. 8 show the counts of each in-degree/out-degree value in the analyzed dataset.

As shown in Fig. 7, for in-degree values, most of the nodes (approximately 8400) have a small number of in-degree values (less than 5). Additionally, there is a small number of nodes (9 nodes) with a high number of in-degree count (more than 100).

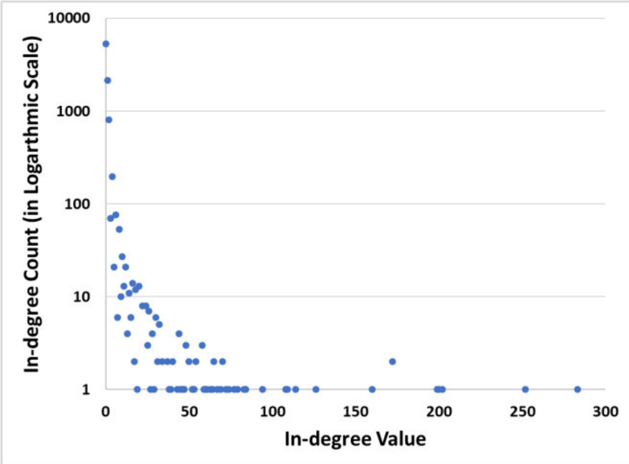

FIGURE 7. In-degree values and counts. Most of the nodes have a small number of in-degree values.

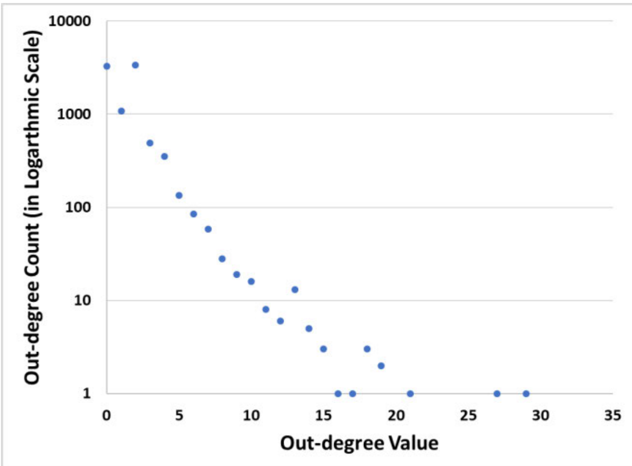

FIGURE 8. Out-degree values and counts. A noticeable number of nodes have small number of out-degree values.

For the out-degree values, as shown in Fig. 8, there is also a noticeable number of nodes (more than 7500) with a small number of out-degree values (0, 1, or 2). Moreover, there are approximately 1000 nodes with an out-degree count of 3,4, or 5, and approximately 200 nodes with an out-degree count range of 6-10. The remaining nodes (45 nodes) have an out-degree count of more than 10.

The in-degree value is the number of Twitter users retweeted, mentioned, or quoted from a user tweet. Based on the in-degree values generated by Gephi, the top users had over 100 arrows pointing toward them, which represent the most popular accounts in this context, as that indicates that there are many other nodes that have interactions with their tweets, as shown in Fig. 9.

By drilling down to the top 3 users with the highest in-degree, the following users have been recognized from highest to lowest as: 1) @cgcsaudi with an in-degree of 283 (highlighted in red), 2) @banderalrezohan with an in-degree of 252, and 3) @okaz_online with an in-degree of 202.

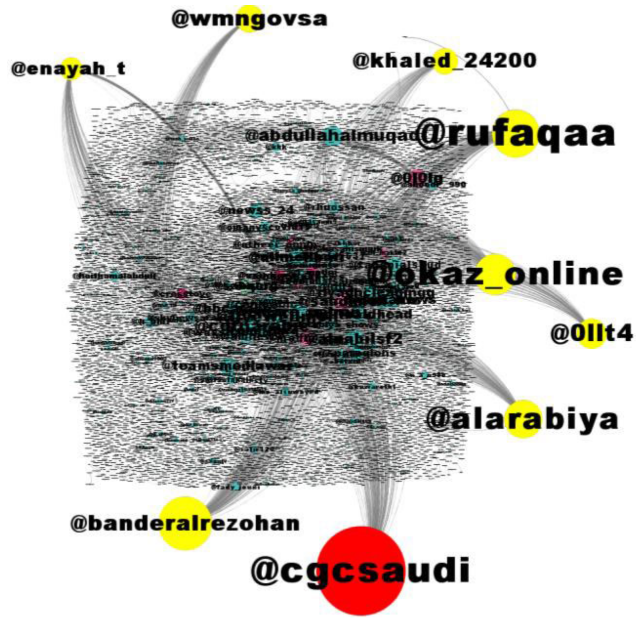

FIGURE 9. Over 100 in-degree users. The top three users are @cgcsaudi, @banderalrezohan, and @okaz_online.

Subsequently, @cgcsaudi, which is a government communication center for joint media, appears to be the most popular account in our case in terms of the in-degree count. Next in popularity are @banderalrezohan, a citizen who is a former player in the Al-Hilal Club, an exclusive critic of "24 media channel", and a member of the Saudi Journalists Authority and, finally, @okaz_online, which is a daily newspaper issued by the OKAZ Organization for Press and Publication.

Being influential is not only determined by in-degree centrality. Hence, out-degree centrality was considered, as well. The users have been filtered to show nodes with more than 15 out-degree links. Fig. 10 shows the generated graph for the top out-degree users. By narrowing down the list of retrieved users with the highest out-degree rank, the top three users found from highest to lowest are: 1) @sska1v with an out-degree of 29, 2) @turkifa91328107 with an out-degree of 27, and 3) @m055446_m with an out-degree of 21. Looking at the 3 top out-degree nodes, the three accounts belong to individuals who are very active in tweeting. They will be excluded from being influencers or information sources as they might have been tweeting much on the day of data streaming. Moreover, the out-degree does not reflect how people are reacting to one's content. Noticeably, the @enayah_t account has been found to be one of the top accounts in both in-degree and out-degree, even though it is not from the top 3 users. Surprisingly, the @enayah_t account is a Saudi shop specializing in personal care kinds of items, and being one of the top users during COVID-19 pandemic might be due to selling goods online that are intensively used to avoid the infection of the virus. Moreover, it might be because of lockdown caution rules and constraints that lead consumers to buy online.

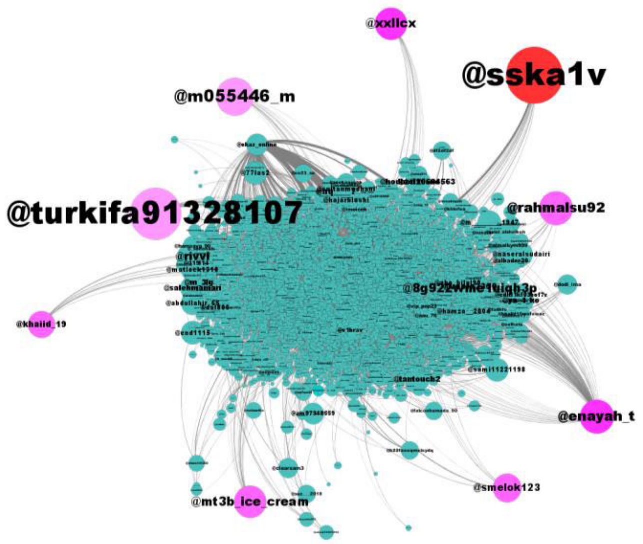

FIGURE 10. Over 15 out-degree user nodes. The top three users are @turkifa91328107, @sska1v, and @m055446_m.

2) CLOSENESS CENTRALITY RESULTS

Closeness centrality measures calculate the shortest paths between all nodes and then assign each node a score based on its sum of shortest paths. It can be viewed as the efficiency of each vertex (individual) in spreading information to all other vertices. The larger the closeness centrality of a vertex is, the shorter the average distance from the vertex to any other vertex, and thus the better positioned the vertex is in spreading information to other vertices. This type of centrality is used for finding the individuals who are best placed to influence the entire network most quickly. Therefore, closeness centrality can help find good 'broadcasters' in a social network. In the case of closeness centrality, or average shortest path length, lower values indicate more central nodes.

For the studied Twitter users, only 38.8% of the users have a closeness centrality value between 0 and 0.5. At the same time, 12% of the nodes have a closeness score between 0.51 and 1. Based on the results, and as a lower closeness centrality score designates a more central node, it can be assumed that the network's connectedness is complex but not significantly connected. Hence, another filter, friends count, was added to narrow down the most influential three nodes (see Fig. 11).

The top three users with the lowest closeness rank (equal to 0) and highest friends count are: 1) @alarabiya_brk with 19646094 friends count (highlighted in red), 2) @ajarabic with 15232214 friends count, and 3) @alarabiya with 14107107 friends count.

The nodes with the lowest closeness centrality nodes and highest friends count were all found to be accounts belonging to media channels. Therefore, they are considered as information sources in Riyadh during the Covid-19 pandemic.

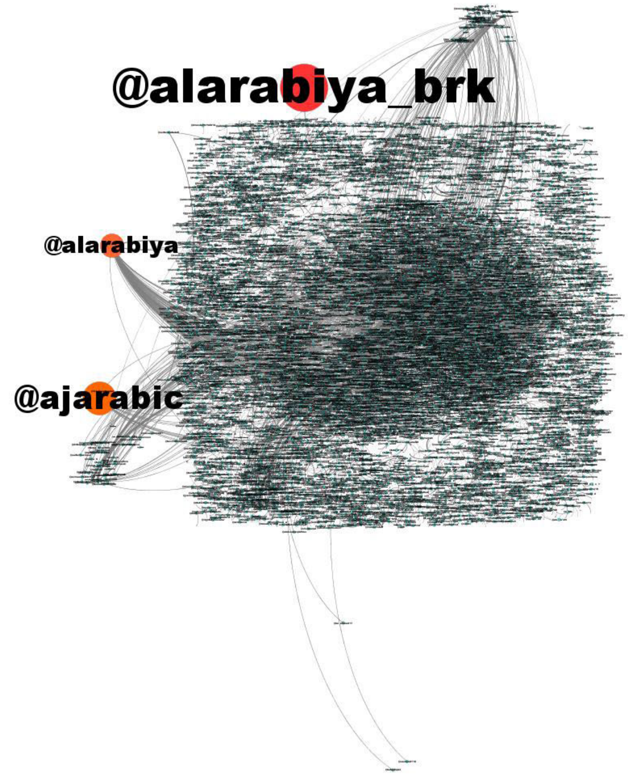

FIGURE 11. op nodes with the lowest closeness rank (equal to 0) and highest friend count. The top three users are @alarabiya_brk, @ajarabic, and @alarabiya.

3) BETWEENNESS CENTRALITY RESULTS

The betweenness centrality measure shows Twitter users that act as 'bridges' between vertices in the social media network by identifying all the shortest paths and then counting how many times each vertex falls on one. In Twitter, information spreads through relatively short paths. Consequently, these Twitter accounts on short paths control the information broadcasting through the social media network. Thus, Twitter accounts with many short paths have high betweenness centrality and are considered influential information gatekeepers.

Fig. 12 represents the betweenness centrality results of the studied Twitter user network. In our case, Twitter accounts with the highest betweenness centrality from highest to lowest were as follows: 1) @enayah_t with 1892 betweenness rank (highlighted in red), 2) @okaz_online with 1156.5 betweenness rank, and @danaati with 391.6 betweenness rank.

Notably, the @enayah_t account, which is a Saudi Shop specializing in personal care items, was found to score the highest betweenness value, followed by @okaz_online, which a press agency, and finally, the account @danaati, which belongs to one of the influencers that is a social activist, writer, author, and office supervisor.

4) EIGENVECTOR CENTRALITY RESULTS

Eigenvector centrality is regarded as a "higher-level" type of centrality. With eigenvector centrality, a Twitter user with fewer connections could hold a very high eigenvector

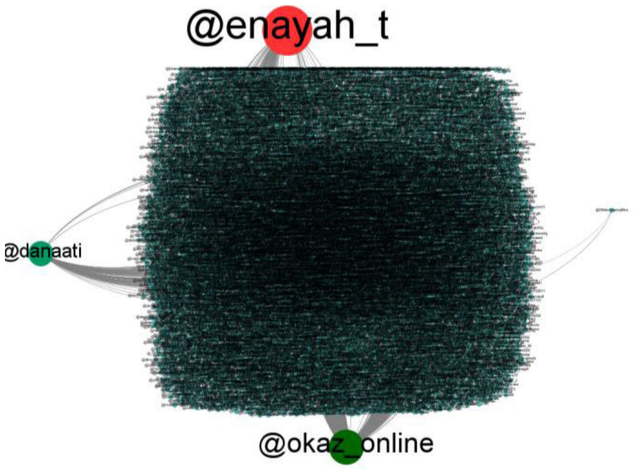

FIGURE 12. Top nodes with the highest betweenness values. The top three users are @enayah_t, @okaz, and @danaati.

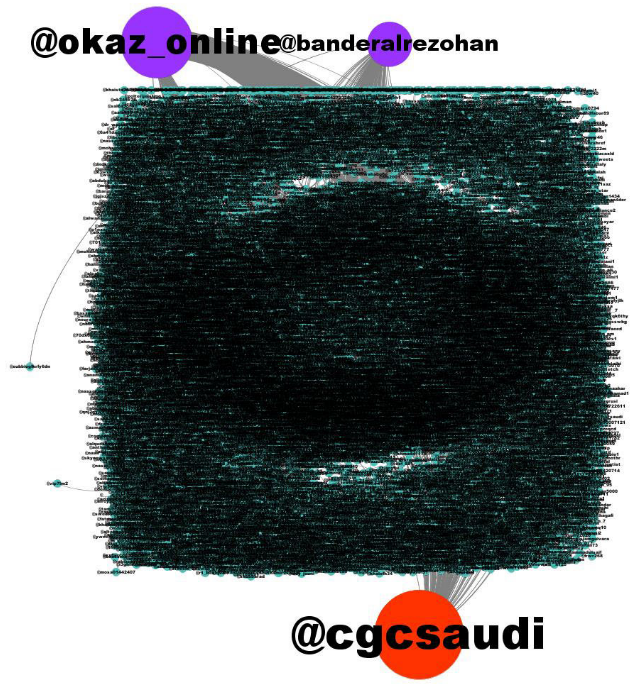

FIGURE 13. Three top nodes with highest eigenvector centrality. The top three users are @cgcsaudi, @okaz, and @banderalrezohan.

centrality. However, those few connections need to be very well linked to permit connections with a high variable value. This implies that connecting to certain vertices is more beneficial than a connection to others.

In our network, as shown in Fig. 13, the top nodes with the highest eigenvector centrality scores were: 1) @cgcsaudi with 1.0 eigenvector centrality (highlighted in red), 2) @okaz_online with 0.8 eigenvector centrality and @banderalrezohan with 0.72 eigenvector centrality.

Correspondingly, we noticed that the top nodes with the highest Eigenvector Centrality values are just the same nodes with the highest in-degree rank. Hence, these nodes could have a high possibility to be considered as the most popular and influential in our case.

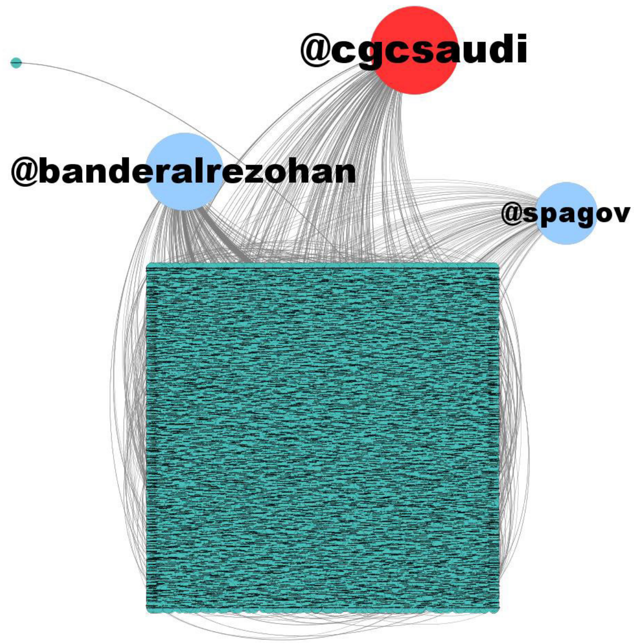

FIGURE 14. Top nodes with the highest pageRank values. The top three users are @cgcsaudi, @banderalrezohan, and @spagov.

5) PAGERANK RESULTS

PageRank is a variant of the eigenvector centrality measure, which assigns nodes a score based on their connections and their connections' connections. The difference is that PageRank also takes link direction and weight into account, so links can only pass influence in one direction and pass different amounts of influence. This measure uncovers nodes whose influence extends beyond their direct connections into the wider network. Because it takes into account direction and connection weight, PageRank can be helpful for understanding citations and authority.

As shown in Fig. 14, we found that the nodes with the highest PageRank value from highest to lowest are: 1) @cgcsaudi (highlighted in red), 2) @banderalrezohan, and @spagov.

Consistently, the nodes @cgcsaudi and @banderalrezohan have scored high centrality values for other centrality measures as well. Moreover, @spagov, which is the official Twitter account for the Saudi Royal Family, was uncovered to show its influence and centrality during the COVID-19 pandemic.

C. COMMUNITY DETECTION VISUALIZATION

For the purpose of detecting and visualizing communities, the Clauset-Newman-Moore algorithm was applied to display the network vertices and their connections. Modularity as network property was used in this algorithm to form a network distributed into communities.

Fig. 15 shows the visualization of the communities detected in the studied network.

As shown in Fig. 15, nodes were arranged into separate groups to present the nodes with the same modularity value in a separate group. The clusters are based on the parameters used in choosing the groups, which has generated 70 groups

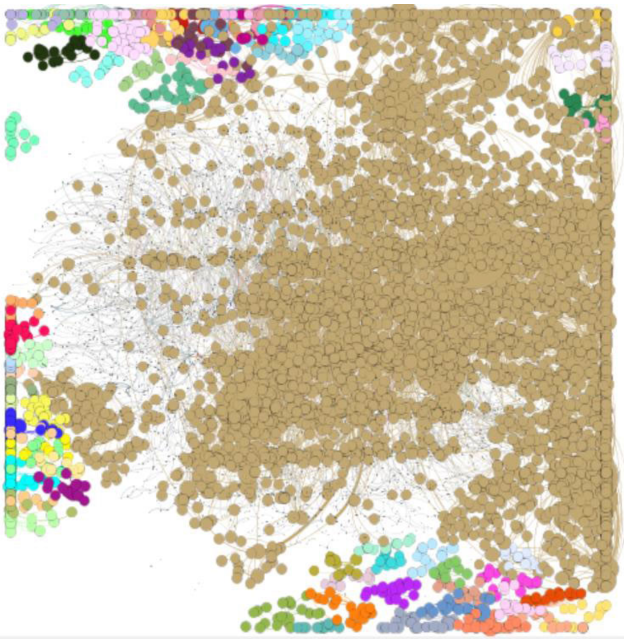

FIGURE 15. Network general communities. There is only one dominant cluster, highlighted in beige.

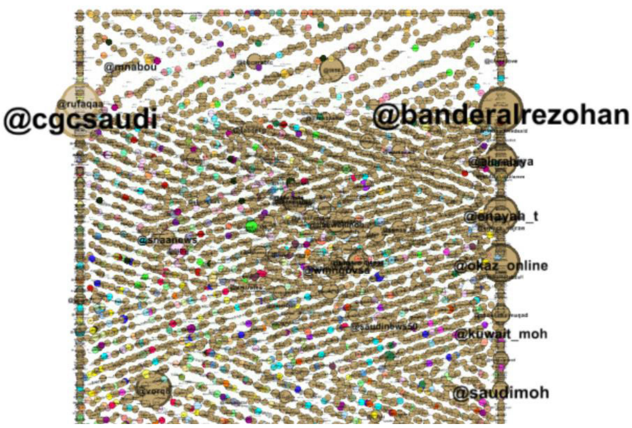

FIGURE 16. Network communities with pageRank. The dominant cluster contains all top influencers and information sources in the studied network.

displayed through various colors. Fig. 15 shows that the largest cluster is focused in the middle with references to many other nodes in the social network. In contrast, other clusters contain limited numbers of nodes with few connections in between.

For a further and more in-depth view, the nodes' labels have been shown and resized according to the PageRank as presented in Fig. 16. The central cluster in the studied network is the beige color cluster. Moreover, the more influencing nodes belong to that cluster too. That would conclude the nodes' strong connection in the network compared to the rest of the clusters.

D. TREND ANALYSIS

Word clouds are a text-based visual illustration of words, classically used to show the relative word frequency or importance by font size.

Fig. 17 and Fig. 18 show the overall word cloud based on all collected hashtags and the word cloud of hashtags with a count of more than 1, respectively.

FIGURE 17. Overall Word cloud of the hashtag network. The most often mentioned hashtags are related to COVID-19 term.

FIGURE 18. Word cloud of hashtags that have occurred multiple times. Beside COVID-19 hashtag, s ome words such as Riyadh "الرياض" and Newspapers "صحيفة" were in the trend too.

As shown in the figures, the most common keywords were Coronavirus "كورونا" and synonyms of this terms such as Novel Coronavirus "فيروس_كورونا_المستجد", Coronavirus Crisis "أزمة_كورونا". In addition, some other words not related to COVID-19, were in the trend as well. These included Riyadh "الرياض" and Newspapers "صحيفة". Twitter users share information about the highest infected city among others. They are also interested in the role of media and newspapers in understanding more about the pandemic.

V. FINDINGS AND DISCUSSION

Based on the results presented earlier in the previous section, the following findings can be derived:

1) The information sources that Twitter users mostly referred to during the COVID-19 pandemic were: @cgcsaudi, @okaz_online, @alarabiya_brk, @ajarabic, @alarabiya, and @spagov.
   The @cgcsaudi account represents the official Saudi Government Communication Center for media on Twitter. Hence, its popularity among Riyadh Twitter users as a reference during the time of pandemic is justified. People might follow official announcements and

COVID-19 pandemic-related news from their country's official account. The @okaz_online account belongs to a newspaper issued by OKAZ Organization for Press and Publication. The Okaz Newspaper is considered the most popular paper in Hijaz and the third most popular in Riyadh. Riyadh Twitter users mostly referred to it during the COVID-19 pandemic. The @alarabiya-brk and @alarabiya accounts both belong to Al Arabiya Channel, which is a Saudi-owned free-to-air television news channel located in Dubai. It is broadcast in Modern Standard Arabic to a pan-Arab audience. Consequently, these accounts were among the most referred to accounts for information by Riyadh Twitter users during the COVID-19 pandemic to obtain the most trusted news and updates about the pandemic and other news. The @ajarabic account belongs to Al Jazeera Channel, an international Arabic news channel founded in Doha, Qatar. As it one of the most popular channels in the Arab World, it was also referred to by Riyadh Twitter users during the COVID-19 pandemic. The @spagov account represents the official account for the news of the Saudi Royal Family. It broadcasts their orders and official statements. As the COVID-19 pandemic caused countries to issue frequent new official statements regarding the operations of all government and private sectors, it is expected that people refer to the authorized account that provides them the latest updates regarding official statements during the COVID_19 pandemic. @spagov is considered to be trusted as an information source, whereas many other untrusted accounts exist to spread rumors and fake news.

2) The most popular influencers among COVID-19 Twitter users are @cgcsaudi, @banderalrezohan, and @okaz_online. As shown in Table 6, these were the common accounts among multiple network measures. The @cgcsaudi and @okaz_online accounts have been found to be among the top nodes based on different measures. As mentioned earlier, these two accounts represent media centers and channels for information in Saudi Arabia in general and specifically in Riyadh. As a result, they are dominant in Riyadh's Twitter Social Network and mostly referred to by Twitter users. The @banderalrezoha account belongs to a former football player in the Hilal Club, currently an exclusive critic of the # 24 channel. Moreover, he is a member of the Saudi Journalists Authority. Thus, it was found that his account has scored the highest in importance among other personal Twitter accounts in the Riyadh Twitter Network. Being a famous journalist made him an influencer during the COVID-19 pandemic. Moreover, as a former football player, he sometimes tweets about football-related analysis and opinions, and this is a favorite of youth Twitter users while staying home during the lockdown period.

TABLE 6. Top influencers and information sources based on the different measures.

In-degree	Out-degree	Closeness	Betweenness	Eigenvector	PageRank
@cgcsaudi	@turkifa91328107	@alarabiya_brk	@enayah_t	**@cgcsaudi**	**@cgcsaudi**
@banderalrezohan	@sska1v	@ajarabic	**@okaz_online**	**@banderalrezohan**	**@banderalrezohan**
@okaz_online	@m055446_m	@alarabiya	@danaati	**@okaz_online**	@spagov

3) Based on community detection results, there are numerous miniclusters scattered around the social network. However, there is only one dominant cluster, highlighted in beige color (Fig. 16), and this cluster contains all top influencers and information sources in the studied network.

Based on the findings above, two accounts were found to be the main key players in Twitter during COVID-19 pandemic: @okaz_online and @cgcsaudi. Both accounts were identified to be the most referred information sources and also the most influential users.

@okaz_online belongs to a Saudi daily newspaper, Okaz (www.okaz.com.sa). Based on the results of a media survey conducted by Ipsos Stat research company (https://www.ipsos.com/en), Okaz is the first in readership ratings, beating many other newspapers published in Saudi Arabia. According to Dubai Press Club (https://dpc.org.ae), Okaz newspaper, is mostly preferred by Saudi nationals and younger people. Okaz Twitter account is very popular among Saudi tweeters as it has around 1M followers.

@cgcsaudi, which belongs to the Government Communication Center (cgc.gov.sa), is one main platform of Saudi Media Ministry that connects the ministry with people other organizations and authorities. The Twitter account of this center is also popular as it has around 950k followers.

The popularity of the two accounts mentioned above explains at least partially why they were the main key players in Saudi Arabia during COVID-19 pandemic.

The Saudi Ministry of Health and other authorities can exploit the identified key information sources and influencers to disseminate instructions and safety protocols to the public, such as motivating them to obtain COVID-19 vaccination.

VI. CONCLUDING REMARKS

Twitter has been a highly useful tool in healthcare for rapid information dissemination and acquisition during the COVID-19 pandemic. It helped researchers study the spread of the disease, find the latest information on COVID 19 and tweets from public health experts, and follow what is happening in real time during the pandemic. This study studied the influence of the COVID-19 pandemic on Saudi Twitter users' tweeting behavior using social networking analysis (SNA).

As Twitter influencers have a worldwide scope and access to different age sectors, the Saudi government may invest them in reducing the severity of COVID-19 and increasing the awareness of the disease consequences. As indicated by [45], countries with lower COVID-19 cases have generated a higher load of tweets to contribute to their public awareness.

The analysis of the Twitter social network performed in this study can be used to anticipate the spread of the virus in Saudi Arabia. Such analysis gives a detailed behavioral overview of Twitter users towards COVID-19 Pandemic in Saudi Arabia. This can help the Saudi government to fight COVID-19 spread, as they gain a hint of problems that require governments' consideration. In addition, the government may make use of influential users in fighting outbreaks as their influence in tremendous domains and communities can help decision-making bodies.

In the following, we propose few practical implications for the government and health authorities in Saudi Arabia, in how the study findings can be adopted to reduce the pandemic consequences.

1. To cooperate with the identified key Twitter influencers and interactive information sources in Saudi Arabia, to spread awareness that helps in stopping COVID-19 at borders.
2. To enable partnering these Twitter influencers and information sources with governmental entities in fighting COVID-19 by spreading well-designed and reliable content regarding the current situation.
3. To encourage social media influencers to be among the first batch to be vaccinated in order to tackle the new vaccination skepticism.

As a part of future work, related data of other social media websites can be studied and analyzed. Machine-learning techniques can be applied to create a prediction model to predict the trend for tweets of COVID-19. Furthermore, text analysis can be used in addition to SNA. In addition, more locations and timeframes can be added for future result comparisons.

ACKNOWLEDGMENT

The author(s) gratefully acknowledge Qassim University, represented by the Deanship of Scientific "Research, on the financial support for this research under the number (coc- 2020- 1-1-L- 10058) during the academic year 1441 AH / 2020 AD".

REFERENCES

[1] S. A. Sarkodie and P. A. Owusu, "Investigating the cases of novel coronavirus disease (COVID-19) in China using dynamic statistical techniques," *Heliyon*, vol. 6, no. 4, Apr. 2020, Art. no. e03747, doi: 10.1016/j.heliyon.2020.e03747.

[2] T. M. Chan, K. Dzara, S. P. Dimeo, A. Bhalerao, and L. A. Maggio, "Social media in knowledge translation and education for physicians and trainees: A scoping review," *Perspect. Med. Educ.*, vol. 9, no. 1, pp. 20–30, Feb. 2020, doi: 10.1007/s40037-019-00542-7.

[3] H. Rosenberg, S. Syed, and S. Rezaie, "The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic," *Can. J. Emerg. Med.*, vol. 22, no. 4, pp. 418–421, Jul. 2020, doi: 10.1017/cem.2020.361.

[4] S. Yum, "Social network analysis for coronavirus (COVID-19) in the united states," *Social Sci. Quart.*, vol. 101, no. 4, pp. 1642–1647, Jul. 2020, doi: 10.1111/ssqu.12808.

[5] S. Jain and A. Sinha, "Identification of influential users on Twitter: A novel weighted correlated influence measure for covid-19," *Chaos, Solitons Fractals*, vol. 139, Oct. 2020, Art. no. 110037, doi: 10.1016/j.chaos.2020.110037.

[6] F. Tahmasbi, L. Schild, C. Ling, J. Blackburn, G. Stringhini, Y. Zhang, and S. Zannettou, "'Go eat a bat, chang!': On the emergence of sinophobic behavior on Web communities in the face of COVID-19," Apr. 2020, *arXiv:2004.04046*. [Online]. Available: http://arxiv.org/abs/2004.04046

[7] M. R. Awal, R. Cao, S. Mitrovic, and R. K.-W. Lee, "On analyzing antisocial behaviors amid COVID-19 pandemic," Jul. 2020, *arXiv:2007.10712*. [Online]. Available: http://arxiv.org/abs/2007.10712

[8] H. V. Ribeiro, A. S. Sunahara, J. Sutton, M. Perc, and Q. S. Hanley, "City size and the spreading of COVID-19 in Brazil," *PLoS ONE*, vol. 15, no. 9, Sep. 2020, Art. no. e0239699, doi: 10.1371/journal.pone.0239699.

[9] H. W. Park, S. Park, and M. Chong, "Conversations and medical news frames on Twitter: Infodemiological study on COVID-19 in South Korea," *J. Med. Internet Res.*, vol. 22, no. 5, May 2020, Art. no. e18897, doi: 10.2196/18897.

[10] N. K. Rajput, B. A. Grover, and V. K. Rathi, "Word frequency and sentiment analysis of Twitter messages during coronavirus pandemic," Apr. 2020, *arXiv:2004.03925*. [Online]. Available: http://arxiv.org/abs/2004.03925

[11] H. Dai and R. Charnigo, "A sentiment analysis OF MERS-CoV outbreak through Twitter social media monitoring," *JP J. Biostatistics*, vol. 15, no. 2, pp. 107–125, Oct. 2018, doi: 10.17654/bs015020107.

[12] T. Wang, K. Lu, K. P. Chow, and Q. Zhu, "COVID-19 sensing: Negative sentiment analysis on social media in China via BERT model," *IEEE Access*, vol. 8, pp. 138162–138169, 2020, doi: 10.1109/ACCESS.2020.3012595.

[13] C. K. Pastor, "Sentiment analysis of filipinos and effects of extreme community quarantine due to coronavirus (COVID-19) pandemic," *J. Crit. Rev.*, vol. 7, no. 7, pp. 91–95, Apr. 2020.

[14] A. D. Dubey, "Twitter sentiment analysis during COVID-19 outbreak," SSRN, Rochester, NY, USA, Tech. Rep. 3572023, Apr. 2020, doi: 10.2139/ssrn.3572023.

[15] M. S. Al-Rakhami and A. M. Al-Amri, "Lies kill, facts save: Detecting COVID-19 misinformation in Twitter," *IEEE Access*, vol. 8, pp. 155961–155970, 2020, doi: 10.1109/ACCESS.2020.3019600.

[16] R. Kouzy, J. A. Jaoude, A. Kraitem, M. B. El Alam, B. Karam, E. Adib, J. Zarka, C. Traboulsi, E. Akl, and K. Baddour, "Coronavirus goes viral: Quantifying the COVID-19 misinformation epidemic on Twitter," *Cureus*, vol. 12, no. 3, p. e7255, Mar. 2020, doi: 10.7759/cureus.7255.

[17] K. Sharma, S. Seo, C. Meng, S. Rambhatla, and Y. Liu, "COVID-19 on social media: Analyzing misinformation in Twitter conversations," 2020, *arXiv:2003.12309*. [Online]. Available: https://arxiv.org/abs/2003.12309 and https://covid-19.conacyt.mx/jspui/handle/1000/4350

[18] P. Pourghomi, M. Dordevic, and F. Safieddine, "The spreading of misinformation online: 3D simulation," in *Proc. 5th Int. Conf. Inf. Technol., Comput., Electr. Eng. (ICITACEE)*, Sep. 2018, pp. 299–304, doi: 10.1109/ICITACEE.2018.8576937.

[19] F. Safieddine, M. Dordevic, and P. Pourghomi, "Spread of misinformation online: Simulation impact of social media newsgroups," in *Proc. Comput. Conf.*, Jul. 2017, pp. 899–906, doi: 10.1109/SAI.2017.8252201.

[20] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103792, doi: 10.1016/j.compbiomed.2020.103792.

[21] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020, doi: 10.1109/TMI.2020.2993291.

[22] M. Nour, Z. Cömert, and K. Polat, "A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106580, doi: 10.1016/j.asoc.2020.106580.

[23] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101794, doi: 10.1016/j.media.2020.101794.

[24] P. K. Sethy, S. K. Behera, P. K. Ratha, and P. Biswas, "Detection of coronavirus disease (COVID-19) based on deep features and support vector machine," *Int. J. Math., Eng. Manage. Sci.*, vol. 5, no. 4, pp. 643–651, Aug. 2020, doi: 10.33889/IJMEMS.2020.5.4.052.

[25] (2020). *COVID-19-Arabic-Tweets-Dataset, GitHub*. [Online]. Available: https://github.com/SarahAlqurashi/COVID-19-Arabic-Tweets-Dataset

[26] M.-G. Hâncean, M. Perc, and J. Lerner, ''Early spread of COVID-19 in romania: Imported cases from italy and human-to-human transmission networks,'' *Roy. Soc. Open Sci.*, vol. 7, no. 7, Jul. 2020, Art. no. 200780, doi: 10.1098/rsos.200780.

[27] N. F. Omran, S. F. Abd-el Ghany, H. Saleh, A. A. Ali, A. Gumaei, and M. Al-Rakhami, ''Applying deep learning methods on time-series data for forecasting COVID-19 in Egypt, Kuwait, and Saudi Arabia,'' *Complexity*, vol. 2021, pp. 1–13, Mar. 2021, doi: 10.1155/2021/6686745.

[28] E. Otte and R. Rousseau, ''Social network analysis: A powerful strategy, also for the information sciences,'' *J. Inf. Sci.*, vol. 28, no. 6, pp. 441–453, Dec. 2002, doi: 10.1177/016555150202800601.

[29] W. Ahmed, J. Vidal-Alaball, J. Downing, and F. L. Seguí, ''COVID-19 and the 5G conspiracy theory: Social network analysis of Twitter data,'' *J. Med. Internet Res.*, vol. 22, no. 5, May 2020, Art. no. e19458, doi: 10.2196/19458.

[30] Y. Sunmoo, M. Odlum, P. Broadwell, N. Davis, C. H. O. Hwayoug, D. E. N. G. Nanyi, M. Patrao, D. Schauer, M. E. Bales, and C. Alcantara, ''Application of social network analysis of COVID-19 related tweets mentioning cannabis and opioids to gain insights for drug abuse research,'' *Stud. Health Technol. Inform.*, vol. 272, pp. 5–8, Jun. 2020, doi: 10.3233/SHTI200479.

[31] B. Kim, ''Effects of social grooming on incivility in COVID-19,'' *Cyberpsychol., Behav., Social Netw.*, vol. 23, no. 8, pp. 519–525, Aug. 2020, doi: 10.1089/cyber.2020.0201.

[32] C. Ordun, S. Purushotham, and E. Raff, ''Exploratory analysis of covid-19 tweets using topic modeling, UMAP, and DiGraphs,'' May 2020, *arXiv:2005.03082*. [Online]. Available: http://arxiv.org/abs/2005.03082

[33] S. Yum, ''Social network analysis for coronavirus (COVID-19) in the United States,'' *Social Sci. Quart.*, vol. 101, no. 4, pp. 1642–1647, Jul. 2020, doi: 10.1111/ssqu.12808.

[34] J. A. Belso-Martínez, A. Mas-Tur, M. Sánchez, and M. J. López-Sánchez, ''The COVID-19 response system and collective social service provision. Strategic network dimensions and proximity considerations,'' *Service Bus.*, vol. 14, no. 3, pp. 387–411, Jul. 2020, doi: 10.1007/s11628-020-00421-w.

[35] Y. Zhao, S. Cheng, X. Yu, and H. Xu, ''Chinese Public's attention to the COVID-19 epidemic on social media: Observational descriptive study,'' *J. Med. Internet Res.*, vol. 22, no. 5, May 2020, Art. no. e18825, doi: 10.2196/18825.

[36] E. Montes-Orozco, R.-A. Mora-Gutierrez, S.-G. De-Los-Cobos-Silva, E.-A. Rincon-Garcia, G.-S. Torres-Cockrell, J. Juarez-Gomez, B. Obregon-Quintana, P. Lara-Velazquez, and M.-A. Gutierrez-Andrade, ''Identification of COVID-19 spreaders using multiplex networks approach,'' *IEEE Access*, vol. 8, pp. 122874–122883, 2020, doi: 10.1109/ACCESS.2020.3007726.

[37] M. Alhajji, A. Al Khalifah, M. Aljubran, and M. Alkhalifah, ''Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19,'' Apr. 2020, doi: 10.20944/PREPRINTS202004.0031.V1.

[38] L. Alsudias and P. Rayson, ''COVID-19 and Arabic Twitter: How can arab world governments and public health organizations learn from social media?'' in *Proc. 1st Workshop NLP COVID-19 ACL*, 2020, pp. 1–9.

[39] S. Alqurashi, A. Alashaikh, and E. Alanazi, ''Identifying information superspreaders of Covid-19 from arabic tweets,'' Preprints, Basel, Switzerland, Tech. Rep. 202007.0228.v1, 2020, doi: 10.20944/preprints202007.0228.v1.

[40] I. Bermudez, D. Cleven, R. Gera, E. T. Kiser, T. Newlin, and A. Saxena, ''Twitter response to Munich July 2016 attack: Network analysis of influence,'' *Frontiers Big Data*, vol. 2, p. 17, Jun. 2019, doi: 10.3389/fdata.2019.00017.

[41] S. Alqurashi, A. Alhindi, and E. Alanazi, ''Large arabic Twitter dataset on COVID-19,'' Apr. 2020, *arXiv:2004.04315*. [Online]. Available: http://arxiv.org/abs/2004.04315

[42] O. Ajao, J. Hong, and W. Liu, ''A survey of location inference techniques on Twitter,'' *J. Inf. Sci.*, vol. 41, no. 6, pp. 855–864, Dec. 2015, doi: 10.1177/0165551515602847.

[43] J. Golbeck and J. L. Klavans, *Introduction to Social Media Investigation: A Hands-on Approach*, 1st ed. Boston, MA, USA: Elsevier, 2015.

[44] A. Clauset, M. E. J. Newman, and C. Moore, ''Finding community structure in very large networks,'' *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 70, no. 6, p. 66111, Dec. 2004.

[45] M. Saad, M. Hassan, and F. Zaffar, ''Towards characterizing COVID-19 awareness on Twitter,'' May 2020, *arXiv:2005.08379*. [Online]. Available: http://arxiv.org/abs/2005.08379

AMAL A. AL-SHARGABI received the master's and Ph.D. degrees from Universiti Teknologi Mara (UiTM), Malaysia. She is currently an Assistant Professor with the College of Computer, Qassim University. Her research interests include program comprehension, empirical software engineering, and machine learning. She was a recipient of a number of Qassim University's research grants, since 2018.

AFEF SELMI received the master's and Ph.D. degrees in computer science from the Faculty of Sciences of Tunis, University of Tunis El Manar, Tunisia, in 2011 and 2017, respectively. From 2011 to 2017, she worked as a Research Assistant with the University of Sousse, Tunisia. Since 2018, she has been an Assistant Professor with the College of Computer, Qassim University, Saudi Arabia. She has attended and participated in various international conferences. Her research interests include recommendation systems, data mining, and machine learning. She received the Publication Award from Qassim University, in 2020.