Using Social Network Analysis to Understand Public Discussions: The Case Study of #SaudiWomenCanDrive on Twitter

Zubaida Jastania[1], Rabeeh Ayaz Abbasi[3*]
Kawther Saeedi[4]
Faculty of Computing and Information Technology
King Abdul-Aziz University
Jeddah, Kingdom of Saudi Arabia

Mohammad Ahtisham Aslam[2]
Faculty of Computing and Information Technology
King Abdul-Aziz University
Jeddah, Kingdom of Saudi Arabia

***Abstract*—Social media analytics has experienced significant growth over the past few years due to the crucial importance of analyzing and measuring public social behavior on different social networking sites. Twitter is one of the most popular social networks and means of online news that allows users to express their views and participate in a wide range of different issues in the world. Expressed opinions on Twitter are based on diverse experiences that represent a broad set of valuable data that can be analyzed and used for many purposes. This study aims to understand the public discussions that are conducted on Twitter about essential topics and developing an analytics framework to analyze these discussions. The focus of this research is the analytical framework of Arabic public discussions using the hashtag #SaudiWomenCanDrive, as one of the hot trends of Twitter discussions. The proposed framework analyzed more than two million tweets using methods from social network analysis. The framework uses the metrics of graph centrality to reveal essential people in the discussion and community detection methods to identify the communities and topics used in the discussion. Results show that @SaudiNews50, @Algassabinasser, and @Abdulrahman were top users in two networks, while @KingSalman and @LoujainHathloul were the top two users in another network. Consequently, "King Salman" and "Loujain Hathloul" Twitter accounts were identified as influencers, whereas "Saudi News" and "Algassabi Nasser" were the leading distributors of the news. Therefore, similar phenomena could be analyzed using the proposed framework to analyze similar behavior on other public discussions.**

Keywords*—Social network analysis; twitter; public discussion; network science

I. Introduction

Understanding human social behavior is important to understand the sophisticated processes that happen in real life. Social networks, as media of human communication, can be considered as a primary source of social behavior analysis [1]. Social network analysis (SNA) that was originated from graph theory has become one of the essential tools for studying social movements. Twitter is one of the most popular social networks and means of online news that allows users to express their views and participate in a wide range of different issues in the world. In the contemporary news flow, Twitter is among the channels that allow individual users to interact in a more engaged and self-driven way. Twitter has also become a platform by which different views are shared and spread through society, and it plays a vital role in current social events and processes.

The "Saudi Women Driving" was a debate for a long time [2][3], in which researchers study and analyze to see the trends and factors that affect people's decisions [4]. The purpose of this analysis is to find which of the users who shared their opinion were more critical in this discussion; whether they formed groups based on their opinion; and whether the more dominant influence is popularity—for example, in terms of the number of followers. The analysis also digs into further analysis of user behavior, including what is retweeted, or who is co-mentioned.

The rest of the paper is organized as follows: Section II discusses the background and related work, Section III presents the methodology used in this paper, Section IV presents the results, Section V discusses the results and Section VI concludes the paper.

II. Background and Related Work

Social networks form an environment where people express their opinions and collaborate to instigate discussions which affect governments to take requested decisions. Moreover, it is also an environment where people express their gratitude and thankfulness to others. In addition, they are also used for marketing. Therefore, studying and analyzing social networks could reveal information flow and patterns that could be used for decision making [5][6][7]. Often Social networks are studied based on graph theory [8].

In mathematics, graph theory is the study of graphs, used to model pairwise relations between objects. Therefore, a graph is made up of vertices (or nodes), which are connected by edges (or links). Based on the graph theory many concepts have emerged that were used to resolve issues in Social Network Analysis. The centralization of any network is a measure of how central its most central nodes are [9]? There are several metrics for centralization including degree, eigenvectors, closeness, and PageRank. The in-degree and out-degree give an indication to the centralization in a directed graph.

Twitter is a way to express patterns and significant topics discussed online. Twitter information can be classified based

*Corresponding Author.

on patterns of information flow [2][10]. Researchers used the metrics of density, modularity, centralization, and the fraction of isolated users [2]. Consequently, they identified categories which include Broadcast and Support; Polarized and In-Group, and Brand and Clustered Community networks. However, the output depends on the search string, and the model was tested on a limited set of topics including health, politics, leisure, academic, and commercial.

The Twitter network structure is affected by geographical and language diversity. Based on a network analysis of 2,500 users, the model in [3] shows the structure of the network's relationships and users' position in the discussion online. They found that people who discuss the same language tend to have similar clusters. They found that defining digital humanities help understand the behavior of people, where a small number of individuals and institutions are discussing. However, this model, using centrality metrics of in-degree, out- degree, betweenness, and eigenvector does not include other features of retweet networks which limits the gatekeeper and hubs of users' information.

AlFarasani et al. [11] proposed a road incident prediction model using an Arabic lexicon. The predicted incident classes from the city of Riyadh, Saudi Arabia over 10,000 tweets provide an accuracy of 82%. Kuroshima1 and Tian [12] mined 10,000 tweets to demonstrate a practical case of utilizing social media in identifying customer opinions to build an effective drug detection system. They applied a set of machine learning algorithms (Decision Trees, Random Forest, Support Vector Machines, K-Nearest Neighbors, and Naïve Bayes) to classify a set of selected drugs (Advil, Aleve, Motrin, Tylenol) based on a set of keywords features. The work in [13] used the concept of homophily to design an affiliation graph to identify the most popular community. They have used data mining techniques and embeddedness, betweenness, and graph occupancy to detect the evolutionary dynamics of social behavior.

The use of case studies in a research like the one proposed herein is strategic for multiple reasons. Reviewing different articles that are conducting similar research allow for the utilization of information that would normally not be easily attainable. Twitter, for example, is such a big data source, with a myriad of topics and issues being addressed, extracting specific data would require a specific framework in order to comprehend the analytical results. Therefore, the use of case studies and different literature materials submitted by different researchers who use different social media platforms as sources of their data provide consequential information that would have otherwise been difficult to obtain. This includes research conducted by Thom and Kruger [4], who were investigating the possibility whether Twitter can save lives. To achieve their objective, they considered the 2013 German Floods as their case study and collected all the relevant Twitter data on the floods. Even though the researchers would be able to acquire paramount data on the floods, the case study is still relatively very general, as it would yield an insurmountable amount of data that would still require filtering. A related research explored the use of Twitter for saving lives by analyzing blood donation requests on Twitter [14]. A recent study used machine learning to identify tweets related to the empowerment of women in Saudi Arabia by a startup and determining the sentiments of these tweets [15].

A more specific case study yielded valuable big data for analysis in the case of the Brexit referendum. Grčar et al. [16] investigated the influence of Twitter users regarding the Brexit referendum, which showed a possibility to establish the relationship between the Twitter mood and the outcome of the referendum. Despite following a single topic, these researchers faced specific challenges concerning the demographics of Twitter users concerning that of the real voters. These concerns are genuine and warrant the need for a case study for a practical framework to analyze social media data to be developed. Similar challenges are distinguished by Wan and Gao [17], who were analyzing Twitter data for airline services analysis. They concluded that for an overall accuracy while analyzing Twitter data, a classification system adoption would be necessary. Such a classification system for the research at hand would analyze a hashtag as the case study.

A number of Twitter analysis models could be used to analyze the impact on a patient's life. A study of patient-driven innovation in diabetes management resulted in building and sharing knowledge around a do-it-yourself (DIY) open-source artificial pancreas systems (OpenAPS). The developers OpenAPS [18] examined Twitter data to understand how patients, caregivers, and care partners perceive OpenAPS. They found a different cluster of people has several forms that provide a widely accepted way to manage day-to-day glucose levels and quality of life. However, the work did not provide details of social network metrics calculations. Also, in the medical domain, the study of people with dementia was analyzed as some people with dementia publicly share their experiences on Twitter [19]. Talbot et al. [19] have manually identified six themes relating to collaboration, experience, community, and stories of dementia. While the study raises awareness, challenge stigma, it was limited to 3 countries with only 2,774 tweets.

As part of the role of education network reform in the US, Rosenberg [20] studied the debate changes in teaching and learning by analyzing tweets of Next Generation Science Standards chat hashtag (#NGSSchat). They examined users' profile data, locations, and tweets with regression model regression analysis for the exploration of influence to study the diverse participants on the topic in terms of occupation, location, and activity. Therefore, they identified what explains the significant interaction between conversing and endorsing participants. As a result, they found that administrators and teachers appear to be central, though there does not appear to be robust clustering between groups.

Based on the graph theory, many concepts have emerged that were used to resolve issues in Social Network Analysis (SNA). Social network analysis has been used in a variety of applications including but not limited to archeology [21], education [22][23], influence propagation [24], [25], event detection [26]–[29], altmetrics [30], disaster management [31], community detection [32], linguistics [33], analyzing money laundering [34], tourism [35], and software engineering [36]. The betweenness-centrality is a measure of centrality based on shortest paths [9], where the shortest path is the minimum

number of edges between two vertices. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex. In twitter, betweenness centrality finds those who are on the most paths between others in the network [37]. Therefore, people with high betweenness tend to be the innovators and brokers in any network [38].

III. Materials and Methods

A. Dataset

The data was crawled using the search API for the Arabic keywords related to the topic "Saudi women can drive" (Arabic hashtags were used related to the discussions). As a result, the dataset contained two million tweets, with more than 680,000 users. The statistics of the dataset are shown in Table I. Sample tweets from the dataset are given in Table II. After getting the tweets, networks were extracted. To illustrate the networks extracted from the data, Fig. 1 to 3 show various networks. Fig. 1 shows the retweet (RT) network. According to the figure, each user who has retweeted someone else appears as the initiator of a link – between him/her and the retweeted one (with the end pointing towards the retweeted one).

Fig. 2 shows the mention network extracted from the sample tweets of Table I. The figure shows which users are mentioned by the initial user in his/her tweet, thus creating a link between both (again, the point is directed towards the one mentioned). Similarly, the hashtag network is constructed based on hashtag co-occurrence, i.e., when two hashtags appear in a tweet, thus generating a link between them. An example of hashtag network is shown in Fig. 3. It is an undirected network. Similar to the hashtag network, the co-mention network is based on the co-occurrences of the people mentioned together in a tweet.

Therefore, with many types of networks, the analysis of Twitter data is comprehensive. For example, the Retweet network can be a measure of how much attention the retweeted user attracted through his post. To what extent he/she provoked a reaction in the discussion in general. The Mention network is pointing to some participants that might be interested in or used as an example in proving a point. The hashtags network, for example, can be an estimate of how many themes were intertwined and how this discussion was formed by the common phrases of the Saudi women driving.

B. Networks

In the course of the analysis, the original dataset was transformed into several types of networks:

- Retweet: a link is formed when a user retweets another user (directed network);
- Mention-Network: a link is formed when a user mentions another user (directed network);
- Co-Mention-Network: when two users are mentioned in the same tweet (undirected network); and
- Hashtag-Network: when two hashtags appear in the same tweet, and there will be an edge between them (undirected network).

Table III shows the total number of nodes and edges in each network.

C. Methods

Analyzing large volume of information can be a considerable effort, so, whenever possible, scientists look for ways that can reveal clues of the bigger picture. Essentially, this is the case with the current study, which tries to get at the essence of more than two million tweets to highlight patterns, draw out expected behaviors, and reveal indispensable persons and areas for further analysis. The centralization of any network is a measure of how central are its nodes [9]. There are several metrics for centralization, including the degree, eigenvector, closeness, and PageRank. The in-degree and out-degree indicate the centralization of a Twitter user, while the eigenvector is a measure of node importance in a network based on a node's connections. When information flows in a hierarchal way, it provides information about hierarchical or egalitarian, which indicates the level to which people are hubs to sharing information or gatekeepers of information. Table IV summarizes commonly used SNA metrics.

TABLE. I. Dataset Statistics

Entity	Frequency
Tweets	2,132,379
Hashtags	2,121,087
Retweets	1,616,543
Users	630,824

TABLE. II. Sample Tweets of the Dataset of Two Million Tweets

User	Tweet
user-1	RT @user-10: The only consistent in Saudi history is change. - Saudi FM @AdelAljubeir□□#SaudiWomenCanDrive #MyArabia https://t.co/07kNKkNW9Y
user-2	RT @FordMiddleEast: Hi @user-11, we'd like to give you your dream car. #MustangSahar #SaudiWomenCanDrive https://t.co/Tln8aiWUNU
user-3	RT @FordMiddleEast: Hi @user-11, we'd like to give you your dream car. #MustangSahar #SaudiWomenCanDrive https://t.co/Tln8aiWUNU
user-4	RT @user-12: Let's not forget who fought to make this possible @manal_alsharif #المرأة_لـ قـ يادة_يـ نـ تـ صر__الـ مـ لك #SaudiWomenDriving https:…
user-5	RT @user-13: Now we all know what it meant. @user-14 #المراه_ق ياده_يـ رفـ ض_الـ شعب #SaudiWomenDriving https://t.co/Tan1xH9GHp
user-6	RT @user-13: Now we all know what it meant. @user-14 #المراه_ق ياده_يـ رفـ ض_الـ شعب #SaudiWomenDriving https://t.co/Tan1xH9GHp
user-7	RT @user-15: Buckle up, ladies! @user-16 writing on the big news from #SaudiArabia this week https://t.co/PpoXWPnogC #SaudiWomenC…
user-8	RT @user-17: The day after #KSA announced that #SaudiWomenCanDrive, @SAP attracts young female professionals, to fuel next wave of growth.…
user-9	RT @user-13: Now we all know what it meant. @user-14 #المراه_ق ياده_يـ رفـ ض_الـ شعب #SaudiWomenDriving https://t.co/Tan1xH9GHp

Fig. 1. Retweet Network Extracted from the Sample Tweets Shown in Table I.

Fig. 2. Mention Network Extracted from the Sample Tweets Shown in Table I.

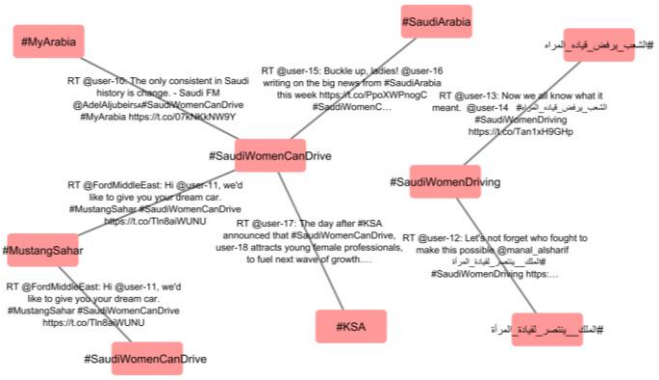

Fig. 3. Hashtag Network Extracted from the Sample Tweets Shown in Table I.

TABLE. III. NETWORKS STATISTICS

	Number of Nodes	Number of Edges
Retweet network	516,248	1,356,538
Mention network	520,670	1,384,388
Co-Mention network	5,875	7,536
Hashtag network	14,305	59,638

There are several ways to analyze Twitter content: social network metrics based on graph theory [18] [20] and machine learning models [39] [40]. Literature reported several types of community detection algorithms: discordant algorithms that detect inter-community links and remove them from the network [16], recursive algorithms that agglomerate similar communities recursively [17][41] and optimization-based algorithms that maximize of an objective function [42].

TABLE. IV. MOST COMMONLY USED SNA METRICS

Metric	Description	Objective
Degree Power Law Diameter	The degree of a node is the number of edges that are adjacent to the node. The Degree Power Law measures how closely the degree distribution of a network follows a power-law scale. The diameter metric measures the maximal distance between all pairs of nodes.	Node to node distance functions
Average Clustering Coefficient	The clustering coefficient when applied to a single node, is a measure of how complete the neighborhood of a node is. When applied to an entire network, it is the average clustering coefficient over all of the nodes in the network.	The clustering coefficient, along with the mean shortest path, can indicate a "small-world" effect.
Average Path Length	The average of shortest path lengths between all pairs of nodes.	How dense and lengthy is the graph. How long are chains of communications?
Eigenvector Centrality	A measure of node importance in a network based on a node's inwards connections.	Node centrality measure

IV. RESULTS

A. Influential People in the Discussion

This section answers the following research question.

RQ: How to identify influential people in the discussion?

It is crucial to consider the right measures to find out which are the most critical nodes taking into consideration the multitude of tweets and users in this conversation. Social network analysis suggests multiple approaches regarding this matter. To answer the research question, we used degree centrality in the retweet, mention, and the co-mention networks.

1) Retweet network: The retweet network is a directed, weighted graph where nodes represent Twitter users, while edges are formed whenever a retweet occurred. The direction of the link reflects the retweet mechanism, with the direction pointing to the user that was retweeted. Table V and Fig. 4 depict this network. Before constructing the retweet network, the number of followers was used to determine the importance of the node. A plausible hypothesis might be that a Twitter account with 100,000 followers is highly influential. However, if only 1,000 users retweet their tweets, then this user might be not so influential as a user with 20,000 followers but 7,000 retweets his/her message. The higher rate of retweeting suggests that the second user is better nested and more active in terms of Twitter communication.

The in-degree interval of this network falls between 0 and 4,430, and the out-degree is between 0 and 1,066. As the degree centrality measure suggests, we can first focus on the nodes with the highest in-degree (the user who has been retweeted the most) and highest out-degree (the user who retweets the most). Table VI shows the top ten in-degree users.

The data in Table VI is summarized with the highest in-degree users: @SaudiNews50 and @AjelNews24: the certified

accounts of well-known news sources in Saudi Arabia. @Algassabinasser: the official account of Nasser Al Gassabi, a Saudi actor. @abdulrahman: Abdulrahman bin Musa'id bin Abdul Aziz is a Saudi Arabian-French businessman. @FordMiddleEast: official account for Ford in the Middle East. @AzzamAlDakhil: the official account of Azzam Al Dakhil, Minister of Education of Saudi Arabia, since 2015. @AmeerahAltaweel: The official account of Ameera Al-Taweel Al-Otaibi, a Saudi princess and philanthropist. Results showed that the highest in-degree user is SaudiNews50 (44,330), which is followed by more than 10 million followers. Other prominent nodes with high in-degree are algassabinasser, with an in-degree of 22,827 and more than 1 million followers, and abdulrahman, with an in-degree of 15,891 and more than 7 million followers.

The analysis of Table VI shows that the number of followers is not the only factor contributing to a user being retweeted because there are nodes with significantly fewer followers that received many retweets. It is also notable that the top in-degree nodes retweeted others rarely, somewhat marginally. Therefore, users suggest that they act more as authority and information generators. These accounts represent the most influential nodes in the retweet network, as they have a huge follower base, and their posts are the most retweeted in the network. Another layer of the data is revealed through the top out-degree nodes. By definition, these would be the participants that retweeted most, which means they acted as transmitters of messages produced by others. Fig. 5 and 6 shows the in-and out-degree distribution of this network, which follows the power-law distribution.

2) Mention network: The mention network is constructed by observing when a user mentions another user. The direction of the link is towards the mentioned user. Being a directed network, which reveals the in-degree versus out-degree statistics of the network. The mention network gives a clue of who might have more significant influence. Table VII and Fig. 7 represent this network.

In and out-degree of this network reflect important people in the discussion. Given that the mention network also consists of a large number of nodes, hence it requires to filter the network. If we focus on the meaning of this representation of discourse, we can conclude that the higher the edge weight, the more semantic value it has, considering that the most mentions/retweets occur between the two accounts involved. Similarly, nodes with lower in-degree value are contributing less to our analysis, since a small number of accounts mentions them. Consequently, we decided to exclude the nodes and edges of lower semantic or analytical significance from our graph model. Using subjective judgment, we set the threshold to keep edge weights and in-degrees higher than 2. Table VIII shows the statistics of the network before and after filtering, while Fig. 8 shows the filtered mention network. Fig. 9 and 10 show the in-degree and out-degree distribution of this network, which follow a power law distribution, similar to the retweet network. Fig. 11 shows the prominent users in the mention network.

As shown in Table IX, the list of discussion participants with the high in-degree measure is substantially different from that of the ones with high out-degree (as it was with the retweet network). This difference is because, in the context of Twitter communication, high in-degree means that more observers are following and mentioning the specific member's tweets, hence this member has more influence on others. On the other hand, high out-degree measure means that a user is mentioning a large number of accounts, thus only consuming information and not creating it or influencing others in a significant way, which is similar to the finding in [13]. The most distinctive feature of the two groups is that the high in-degree group is formed by news channels and public figures, while individual users form the high out-degree. Here is a summary of them: @SaudiNews50, @Sabqorg, @AjelNews24, and @Alekhbariyatv: these are all news accounts. @Sabqorg: the official account of the electronic magazine Sabq. @Alekhbariyatv: the official account of the Saudi channel Al Ekhbariya. @KingSalman: the official account of King Salman. @SaudiNews50, @Sabqorg, @AjelNews24, and @Alekhbariyatv: these are all news accounts. @Sabqorg: the official account of the electronic magazine Sabq. @Alekhbariyatv: the official account of the Saudi channel Al Ekhbariya. @KingSalman: the official account of King Salman. @Algassabinasser: the official account of Nasser Al Gassabi, a Saudi actor.

In order to reveal the opinion-makers and to filter users, we filtered the graph by in-degree and number of followers. Fig. 11 shows the users with high in-degree and the highest number of followers. The dark red nodes have high number of followers, and the size refers to in-degree. There are many well-known accounts for tweeting and spreading the latest and important news, such as: @SaudiNews50, one of the major players; @AjelNews24, with fewer followers than @SaudiNews50; and @SabqOrg, with the highest number of followers but less in-degree than the others.

TABLE. V. RETWEET-NETWORK STATISTICS

Retweet-Network	
Number of nodes	516,248
Number of edges	1,356,538
Average degree	2.628
Avg. weighted network	3.131
Network diameter	25
Modularity	0.71
Number of communities	1,033

Fig. 4. Retweet Network

TABLE. VI. DATA LABORATORY OF THE HIGHEST IN-DEGREE USERS IN THE RETWEET-NETWORK

Label	User_name	In-degree	Out-degree	Followers count
SaudiNews50	أخبار السعودية	44,330	3	44,333
algassabinasser	ناصر القصبي	22,827	1	22,828
abdulrahman	عبدالرحمن بن مساعد	15,891	5	15,896
FordMiddleEast	Ford Middle East	13,214	-	13,214
AjelNews24	خبر عاجل	12,283	3	12,286
AzzamAlDakhil	عزام الدخيّل	10,033	2	10,035
Al_khalden8	الخالدي	9,568	-	9,568
AmeerahAltaweeL	أميرة الطويل	8,796	3	8,799
badughaish	Faisal BaDughaish	7,838	1	7,839
faare8	فارس الهلال	7,498	1	7,499

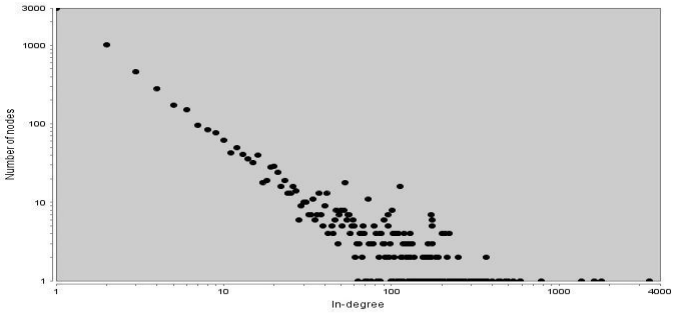

Fig. 5. In-Degree Plot of the Retweet Network.

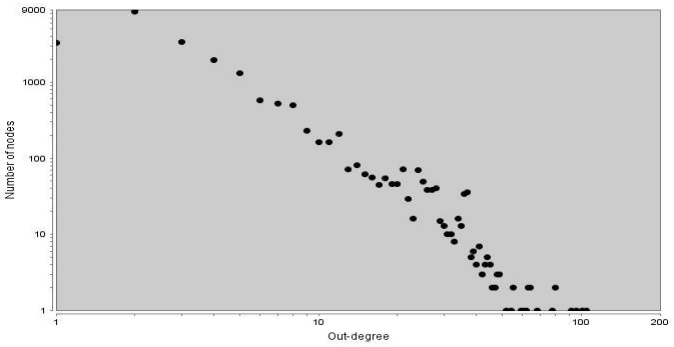

Fig. 6. Out-Degree Plot of the Retweet Network.

TABLE. VII. MENTION NETWORK STATISTICS

Mention-Network	
Number of nodes	520,670
Number of edges	7,536
Average degree	2.565
Avg. weighted network	28.823
Network diameter	17
Modularity	0.808
Number of communities	1,268

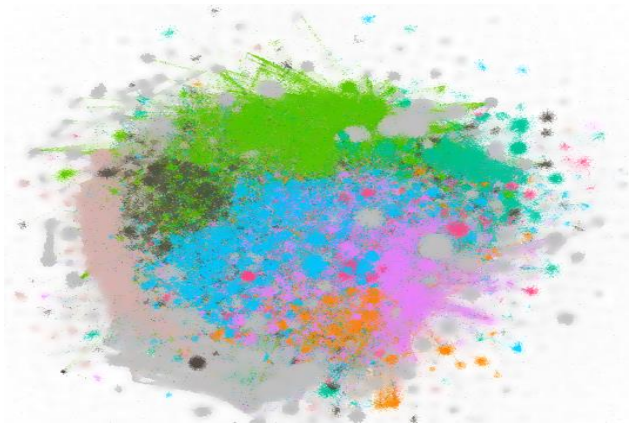

Fig. 7. Mention Network.

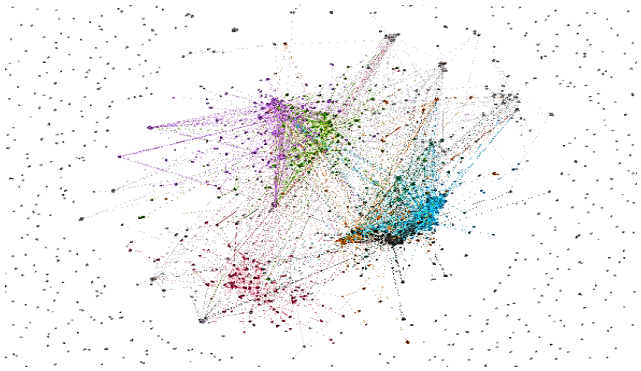

Fig. 8. Mention-Network Graph after Filtering.

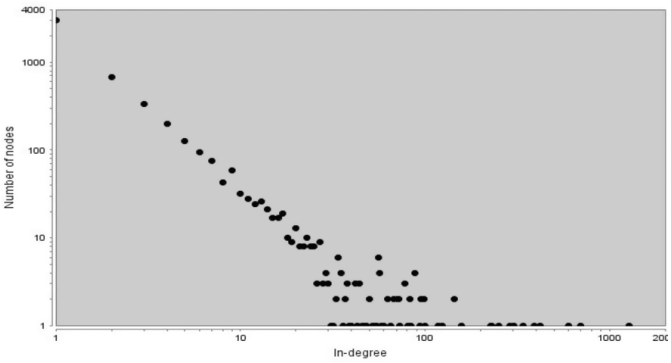

Fig. 9. In-Degree Plot of the Mention Network.

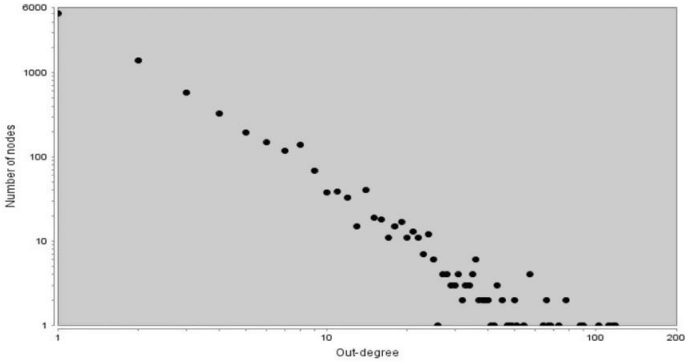

Fig. 10. Out-Degree Plot of the Mention Network.

TABLE. VIII. FILTERED MENTION NETWORK

Mention-Network – Original		Mention-Network – Filtered	
Average degree	2.565	Average degree	3.666
Avg. weighted network	2.596	Avg. weighted network	1.971
Network diameter	28	Network diameter	22
Graph Density	0	Graph Density	0.0
Modularity	0.712	Modularity	0.731
PageRank	0.85	PageRank	0.85

TABLE. IX. IMPORTANT NODES IN THE MENTION NETWORK

Top in-degree nodes			Top out-degree nodes		
#	*Label*	*In-degree*	*#*	*Label*	*Out-degree*
1	SaudiNews50	1274	1	alkinha505	119
2	Algassabinasser	697	2	Asss2019	116
3	Abdulrahman	600	3	mode_smart2030	112
4	Sabqorg	419	4	Waeel05550Waeel	103
5	AjelNews24	393	5	sltan_656	90
6	KingSalman	342	6	asd_sunah	88
7	KSA_620	303	7	Aaaa06259528	78
8	hosbah_tweet	289	8	asd025056	78
9	Alekhbariyatv	251	9	TqU6lSOGEFjRJcu	73
10	faare8	229	10	l680352	68

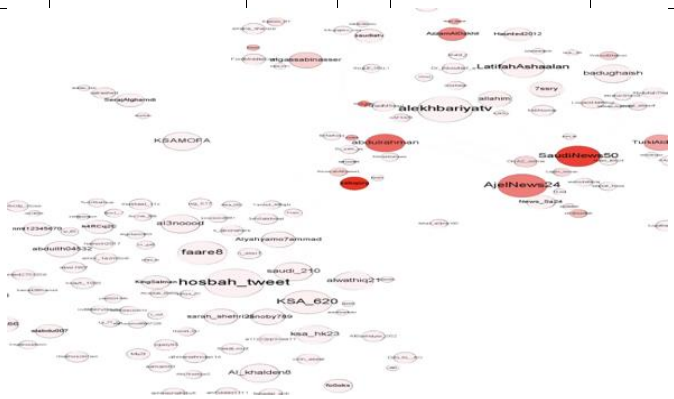

Fig. 11. Network of the users with a High in-Degree and Highest Number of Followers in Mention Network.

We could say that these drives and, to a certain extent, shape the whole conversation by being the top Arabic news sources on Twitter. The distribution of the users varies due to location and language on Twitter. Most of the users do not mention their locations. Also, the language used can be either Arabic or English, which adds to the disparity in the distribution of the users. From Fig. 12, we can see that most of the users do not state a location (indicated by the color purple). Others are from Saudi Arabia and England.

In summary, there are many similarities regarding the leading players in the degree centrality of the retweet and Mention networks. These results are because the dataset is the same. So, even though the links are formed differently, the leading players re-appear.

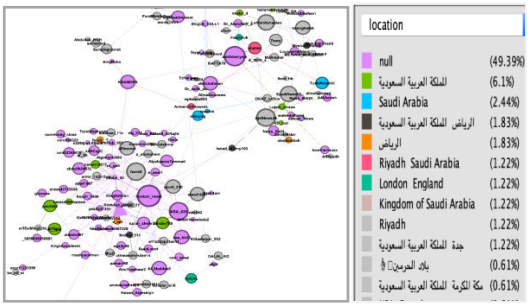

Fig. 12. Location of the users in the Mention Network.

3) Co-mention network: The co-mention network is created based on the fact that users are mentioned together in tweets. For example: if @user1 and @user2 are both mentioned in a tweet, they will be connected through an edge in the network. The co-users appearing together does not represent information flow, and this determines this network as undirected. Table X and Fig. 13 represent the co-mention network.

Table XI shows the top ten users as follows:@KingSalman, the official account of the King of Saudi Arabia. @LoujainHathloul is a Saudi women's rights activist, a social media figure. @protectmax is a corporate account that announced a prize for users retweeting tweets with specific hashtags related to Women Driving, so, logically, the company received many mentions. @algassabinasser, or Nasser Algassabi is a famous Saudi actor. @SaudiNews50 and @KSA24 are personal accounts for spreading news, but they are highly prevalent in Saudi Arabia. @MOISaudiArabia is the official account of the Ministry of the Interior in Saudi Arabia. @ssa_at is the official account of the Council of Senior Scholars. @abdulrahman: Abdulrahman bin Musa'id bin Abdul Aziz is a Saudi Arabian-French businessman. @manal_alsharif: Manal al-Sharif is a Saudi Arabian women's rights activist who helped start a women's right to drive campaign in 2011.

It is clear from Fig. 14 that @KingSalman has edges with many other top accounts. Two of them are the official accounts @MOISaudiArabia and @ssa_at. These form a triangle. We can also observe that @manal_alsharif and @LoujainHathloul have a thick edge between them because they are considered endorsers of the decision, and it is highly probable that users mention both of them in their tweets. @protectmax does not have an edge with any of these most prominent accounts, but it is still one of the top users because of the interest the company triggered by encouraging users to cite particular hashtags about women driving.

TABLE. X. CO-MENTION NETWORK STATISTICS

Co-Mention network	
Number of nodes	5,875
Number of edges	1,384,388
Average degree	2.659
Avg. weighted network	3.131
Network diameter	28

TABLE. XI. THE TOP TEN DEGREE USERS IN THE CO-MENTION-NETWORK

Label	User_name	Degree	Count
KingSalman	ع بدالـ عزيز ز ب ن سلمان	286	5,155
LoujainHathloul	الهذلـ ول هذلـ ول لـ ج ين	111	687
protectmax	ماكس بـ روتـ كت شركة	99	361
algassabinasser	الـ قصـ بي نـ اصر	87	1,031
SaudiNews50	الـ سعوديـ ة أخـ بار	70	718
manal_alsharif	الـ شريـ ف مـ سعودـ منال	70	658
MOISaudiArabia	الـ داخـ لـ يةـ وزارة	68	449
KSA24	الأخـ بار موجز SA	66	122
ssa_at	الـ عـ لماءـ كـ بار هيـ ئة	66	519
abdulrahman	مـ ساعدـ بـ ن عـ بدالـ رحمن	57	521

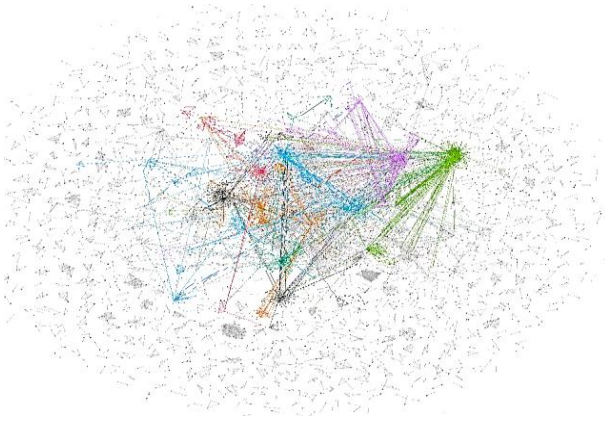

Fig. 13. Co-Mention Network.

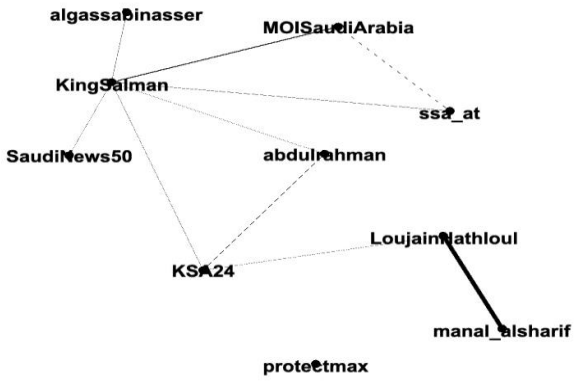

Fig. 14. Top Ten users in the Co-Mention Network.

V. DISCUSSION

The most influential people in the discussion are listed in Table XII and Table XIII shows the top 5 users in the Mention, Retweet, and Co-mention networks:

The most critical users in mention and Retweet network are SaudiNews50, Sabqorg, AjelNews24, and Alekhbariyatv, who are news accounts. While KingSalman's account (the King of Saudi Arabia) has the highest degree in the co-mention network since his account was mentioned along with other users to discuss the topic of women driving decision, namely Loujain Hathloul. Loujain Hathloul is one of the biggest supporters of Women Driving in Saudi Arabia, which makes her one of the top co-mentioned accounts. Similarity, algassabinasser, presumably got many mentions due to his popularity and to his stance in favor of allowing Saudi women to drive.

TABLE. XII. MOST SIGNIFICANT PEOPLE IN THE STUDIED DATASET

Account owner	Accounts
The official account of King Salman.	@KingSalman
The official account of the news TV channel	@AlArabiya_Brk , @Alekhbariyatv
certified accounts of well-known news sources in Saudi Arabia	@SaudiNews50 , @AjelNews24, and @KSA24
Islamic scholar	@Dr_alqarnee, @MohamadAlarefe and @Dr_almosleh
Among the first supporters of the "Women Driving" decision, even before the decision's official declaration.	@LoujainHathloul, @manal_alsharif

TABLE. XIII. TOP FIVE USERS IN ALL NETWORKS

Top users in Mention-Network	Top users in Retweet-Network	Top users in Co-Mention-Network
SaudiNews50	SaudiNews50	KingSalman
Algassabinasser	Algassabinasser	LoujainHathloul
Abdulrahman	Abdulrahman	protectmax
Sabqorg	AjelNews24	algassabinasser
AjelNews24	Alekhbariyatv	manal_alsharif

However, there are some differences between all the three networks. For example – SaudiNews50 is among the top users in the mention and retweet networks, which means that many users referred to it as a news source by either for sharing the decision or proving a point. In this sense – both retweet and mention networks reveal the most important news outlets or information disseminators. However, the co-mention network is more targeted towards connected persons, as in its top users, there are no news outlets. Therefore, we can say that when it comes to sharing of information, retweet and mention networks tend to provide the most influential persons, including news sources. Whereas influential people closely associated with the discussions can be identified through the co-mention network, it may exclude popular users like news dissemination accounts. Results showed that @SaudiNews50, @Algassabinasser, and @Abdulrahman were top users in both retweet and co-mention networks, while @KingSalman and @LoujainHathloul were the top two users in the co-mention network. Consequently, KingSalman and LoujainHathloul were identified as influencers, whereas SaudiNews and Algassabinasser were the leading distributors of the news.

VI. CONCLUSIONS AND RECOMMENDATIONS

This research analyzed a dataset of two million tweets discussing the topic of women driving cars in Saudi Arabia. Using a large dataset of #SaudiWomenCanDrive, this research has proposed a framework that consists of four networks: hashtag, retweet, mention, co-mention networks. The framework helps in identifying influential people. For example, the framework showed that the account of KingSalman admired people by his decisive decision; therefore, the account appears on top in co-mention network. Moreover, the research identified the collective behavior of people while expressing

their opinions on different topics. This research found that people's influence was related to their effect on social media in terms of Twitter network structure as distributors or hubs of news. The proposed framework could provide to analyze similar behavior on other discussions.

REFERENCES

[1] W. Tan, M. B. Blake, I. Saleh, and S. Dustdar, "Social-network-sourced big data analytics," IEEE Internet Comput., vol. 17, no. 5, pp. 62–69, 2013.

[2] A. Alotaibi, "Why the panic? Gendered moral panics and the Saudi ban on women driving," The George Washington University, 2017.

[3] E. Alhussein, "Triangle of change: the situation of women in Saudi Arabia," Exec. Summ., 2014.

[4] I. Chaudhry, "Arab Revolutions: Breaking fear|# hashtags for change: Can Twitter generate social progress in Saudi Arabia," Int. J. Commun., vol. 8, p. 19, 2014.

[5] M. Boukes, "Social network sites and acquiring current affairs knowledge: The impact of Twitter and Facebook usage on learning about the news," J. Inf. Technol. Polit., vol. 16, no. 1, pp. 36–51, 2019.

[6] D. Knoke and S. Yang, Social network analysis, vol. 154. SAGE Publications, Incorporated, 2019.

[7] L. Vega and A. Mendez-Vazquez, "Detecting of topic-specific leaders in social networks," Procedia Comput. Sci., vol. 151, pp. 1188–1193, 2019.

[8] J. A. Barnes and F. Harary, "Graph theory in network analysis," 1983.

[9] L. C. Freeman, "A set of measures of centrality based on betweenness," Sociometry, pp. 35–41, 1977.

[10] J. A. Morente-Molinera, G. Kou, K. Samuylov, R. Ureña, and E. Herrera-Viedma, "Carrying out consensual Group Decision Making processes under social networks using sentiment analysis over comparative expressions," Knowledge-Based Syst., vol. 165, pp. 335–345, 2019.

[11] A. LFarasani, T. AlHarthi, and S. AlHumoud, "ATAM: Arabic Traffic Analysis Model for Twitter," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 3, 2019.

[12] D. Kuroshima and T. Tian, "Detecting Public Sentiment of Medicine by Mining Twitter Data," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 10, 2019.

[13] H. Al-Qaheri and S. Banerjee, "Measuring Homophily in Social Network: Identification of Flow of Inspiring Influence under New Vistas of Evolutionary Dynamics," in International Journal of Advanced Computer Science and Applications (IJACSA), Special Issue on Extended Papers from Science and Information Conference, 2013.

[14] R. A. Abbasi et al., "Saving lives using social media: Analysis of the role of twitter for personal blood donation requests and dissemination," Telemat. Informatics, vol. 35, no. 4, 2018.

[15] B. Alotaibi, R. A. Abbasi, M. A. Aslam, K. Saeedi, and D. Alahmadi, "Startup Initiative Response Analysis (SIRA) Framework for Analyzing Startup Initiatives on Twitter," IEEE Access, vol. 8, pp. 10718–10730, 2020.

[16] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, "Defining and identifying communities in networks," Proc. Natl. Acad. Sci., vol. 101, no. 9, pp. 2658–2663, 2004.

[17] P. Pons and M. Latapy, "Computing communities in large networks using random walks.," J. Graph Algorithms Appl., vol. 10, no. 2, pp. 191–218, 2006.

[18] M. L. Litchman, D. Lewis, L. A. Kelly, and P. M. Gee, "Twitter analysis of# OpenAPS DIY artificial pancreas technology use suggests improved A1C and quality of life," J. Diabetes Sci. Technol., vol. 13, no. 2, pp. 164–170, 2019.

[19] C. V Talbot, S. T. O'Dwyer, L. Clare, J. Heaton, and J. Anderson, "How people with dementia use twitter: A qualitative analysis," Comput. Human Behav., vol. 102, pp. 112–119, 2020.

[20] J. Rosenberg, "Exploring the Next Generation Science Standards Chat (# NGSSchat) Professional Network on Twitter Through Social Network Analysis," 2019.

[21] B. J. Mills, "Social Network Analysis in Archaeology," Annu. Rev. Anthropol., vol. 46, no. 1, pp. 379–397, Oct. 2017.

[22] F. Haneef et al., "Using network science to understand the link between subjects and professions," Comput. Human Behav., vol. 106, 2020.

[23] R. Isba, K. Woolf, and R. Hanneman, "Social network analysis in medical education," Med. Educ., vol. 51, no. 1, pp. 81–88, Jan. 2017.

[24] N. E. Friedkin, A. V. Proskurnikov, R. Tempo, and S. E. Parsegov, "Network science on belief system dynamics under logic constraints," Science (80-.)., vol. 354, no. 6310, pp. 321–326, Oct. 2016.

[25] N. Alduaiji, A. Datta, and J. Li, "Influence Propagation Model for Clique-Based Community Detection in Social Networks," IEEE Trans. Comput. Soc. Syst., vol. 5, no. 2, pp. 563–575, Jun. 2018.

[26] Z. Saeed, R. A. Abbasi, I. Razzak, O. Maqbool, A. Sadaf, and G. Xu, "Enhanced Heartbeat Graph for emerging event detection on Twitter using time series networks," Expert Syst. Appl., vol. 136, pp. 115–132, 2019.

[27] Z. Saeed et al., "What's Happening Around the World? A Survey and Framework on Event Detection Techniques on Twitter," J. Grid Comput., vol. 17, no. 2, pp. 279–312, Jun. 2019.

[28] Z. Saeed, R. A. Abbasi, A. Sadaf, M. I. Razzak, and G. Xu, Text stream to temporal network - A dynamic heartbeat graph to detect emerging events on twitter, vol. 10938 LNAI. Springer Verlag, 2018, pp. 534–545.

[29] Z. Saeed et al., "Event Detection in Twitter Stream Using Weighted Dynamic Heartbeat Graph Approach," IEEE Comput. Intell. Mag., vol. 14, no. 3, pp. 29–38, Aug. 2019.

[30] A. Said, T. D. Bowman, R. A. Abbasi, N. R. Aljohani, S.-U. Hassan, and R. Nawaz, "Mining network-level properties of Twitter altmetrics data," Scientometrics, Apr. 2019.

[31] J. Kim and M. Hastak, "Social network analysis: Characteristics of online social networks after a disaster," Int. J. Inf. Manage., vol. 38, no. 1, pp. 86–96, Feb. 2018.

[32] A. Said, R. A. Abbasi, O. Maqbool, A. Daud, and N. R. Aljohani, "CC-GA: A clustering coefficient based genetic algorithm for detecting communities in social networks," Appl. Soft Comput. J., vol. 63, 2018.

[33] C. S. Q. Siew and M. S. Vitevitch, "The phonographic language network: Using network science to investigate the phonological and orthographic similarity structure of language," J. Exp. Psychol., vol. 148, no. 3, pp. 475–500, 2019.

[34] A. Fronzetti Colladon and E. Remondi, "Using social network analysis to prevent money laundering," Expert Syst. Appl., vol. 67, pp. 49–58, Jan. 2017.

[35] C. Casanueva, Á. Gallego, and M.-R. García-Sánchez, "Social network analysis in tourism," Curr. Issues Tour., vol. 19, no. 12, pp. 1190–1209, Oct. 2016.

[36] J. Kanwal, O. Maqbool, R. Abbasi, and A. Q. A. Q. Abbasi, "Network analysis of software change history for understanding software evolution," in 17th IEEE International Multi Topic Conference 2014, 2014, pp. 229–234.

[37] U. Brandes, "A faster algorithm for betweenness centrality," J. Math. Sociol., vol. 25, no. 2, pp. 163–177, 2001.

[38] F. Imran, R. A. Abbasi, M. A. Sindhu, A. S. Khattak, A. Daud, and T. Amjad, "Finding research areas of academicians using clique percolation," in 2018 14th International Conference on Emerging Technologies, ICET 2018, 2019.

[39] M. Congosto, P. Basanta-Val, and L. Sanchez-Fernandez, "T-Hoarder: A framework to process Twitter data streams," J. Netw. Comput. Appl., vol. 83, pp. 28–39, 2017.

[40] D. M. Best, J. Bruce, S. Dowson, O. Love, and L. McGrath, "Web-based visual analytics for social media," in Sixth International AAAI Conference on Weblogs and Social Media, 2012.

[41] Z. Liu and Y. Ma, "A divide and agglomerate algorithm for community detection in social networks," Inf. Sci. (Ny)., vol. 482, pp. 321–333, 2019.

[42] S. Rahimi, A. Abdollahpouri, and P. Moradi, "A multi-objective particle swarm optimization algorithm for community detection in complex networks," Swarm Evol. Comput., vol. 39, pp. 297–309, 2018.