



Identification and role of opinion leaders in information diffusion for online discussion network

Ateeq Ur Rehman¹ · Aimin Jiang¹ · Abdul Rehman² · Anand Paul²  · Sadia din² · Muhammad Tariq Sadiq³

Received: 1 August 2019 / Accepted: 4 December 2019 / Published online: 24 January 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Social networks are playing a vibrant role in spreading information among its users. With the ongoing boom in technology, social networks are becoming very popular nowadays and are made up of a multitude of users. Some of these users may have a strong influence on the other network users depending on their uncommon elevated values of betweenness centrality (BC). In the online discussion network such as Twitter, the extremely important users are called Opinion Leaders, who play a vital role in the spread of information in an efficient and fast way and keep the isolates interested in the online discussion network. One of the most significant problems in the associated sector is the identification of opinion leaders. In this paper, opinion leaders are chosen based on various centrality measures. The central users are identified based on their in-degree and out-degree links and are ranked within the network by their BC values. Furthermore, we analyze community evolution by using the standard Louvain algorithm. The experiment is performed on publicly available Higgs Boson data from Twitter. The conversation starter and influencer have been observed as an opinion leader for each network. These users have been observed to play a crucial part in the dissemination of information in an online discussion network.

Keywords Social networks · Complex networks · Opinion leaders · Centrality · Twitter

1 Introduction

The use of the online social network (OSN) has grown exponentially in the recent decades. These OSNs including Twitter (<http://www.twitter.com>), LinkedIn (<http://www.linkedin.com>), Google+ (<http://plus.google.com>), Facebook (<http://www.facebook.com>), Sina Weibo (<http://www.weibo.com>), and Tumblr (<http://www.tumblr.com>) have hundreds of millions of active users. The use of OSNs is not only connecting users with their friends and family members, but it is also a fast way to keep users in touch with the latest news and events happening locally as well as globally. According to a recent study by Cherepnalkoski et al., Twitter has been nominated as the most widely used internet microblogging

platform with more than 0.3 billion active monthly users and 0.5 billion daily tweets (Cherepnalkoski et al. 2016). Politicians, scientists, and celebrities, who use Twitter as a medium to share their private profile and opinions, are the most prominent active users (De Domenico et al. 2013; Stever and Lawson 2013; Adalat et al. 2018). The interweaving of a variety of users involved in each kind of ongoing online discussions produces a very huge amount of unstructured data. The processing of these data may be helpful in identifying opinion leaders, their communication/connection patterns, and the flow of opinion of users participating in online communication.

OSNs can be clearly visualized as graphs. Many researchers have done work by using large-scale data in various domains such as SIoTs (Abdul et al. 2018; Rehman et al. 2019), small-world networks (Farhan et al. 2019), link prediction (Liben-Nowell and Kleinberg 2007), medical (Mitchell and Ross 2016; Sadiq et al. 2019) and opinion formation (O’Sullivan et al. 2017) to name a few. The primary purpose of this study is the identification of multiple kinds of significant users and visualization of their associated connection patterns. A directed Twitter dataset called Higgs-Twitter

✉ Anand Paul
paul.editor@gmail.com

¹ College of Internet of Things Engineering, Hohai University, Changzhou 213022, China

² Department of Computer Science and Engineering, Kyungpook National University, Daegu 41566, South Korea

³ School of Automation, Northwestern Polytechnical University, 127 West Youyi Road, Xi’an 710072, China

accessible at (Jure Leskovec) is used to validate the proposed work.

Many researchers worked for the identification of opinion leaders in social networks (Xiaoqing et al. 2013; Khan et al. 2015; Aleahmad et al. 2016). In addition, Feng contributed to this domain by finding the most central user from the network #RaceTogether Campaign (Feng 2016). There are many reasons to identify opinion leaders, which may include political intentions (Cherepnalkoski et al. 2016; Adalat et al. 2018), sports (Lamirán-Palomares et al. 2019), new technologies adaptation (Rogers and Everett 2003), updating brand knowledge (Keller and Berry 2003), and direct marketing and advertising (Trusov et al. 2010). The identification of opinion leaders is very important as these leaders play a significant role in the rapid spread of information within the network.

This paper presents a study based on various centrality measures and a community-detection-based approach for identifying opinion leaders for a specific scientific discovery trend on twitter. The twitter dataset is used for experimentation purposes as twitter has quick information flow with a huge impact on opinion formation of the public. The dataset used in this research contains tweets information about the discovery of a new particle with the features of elusive Higgs Boson. A novel community detection approach is proposed to identify the most influential users in the network. Furthermore, in-degree, out-degree, and betweenness centrality based measures are also used to find key users. The ultimate purpose of this study is two-fold. Firstly, we are interested to identify the most influential users which are termed as opinion leaders and explore the flow of information among these users in the online discussion network. Secondly, this study unveils the connection pattern of these identified most influential users of the network.

The organization of the paper is as follows. Section 2 reviews the literature. Different types of influential participants are included in Sect. 3. Section 4 presents the details about experimental analysis and results. Section 5 demonstrates the use of community detection for finding opinion leaders in the online discussion network. A brief discussion part is included in Sect. 6. Finally, Sect. 7 concludes the paper.

2 Related work

Identifying opinion leaders is of vital importance in many ways, such as marketing (Clow and Baack 2004), community health campaigns (Mitchell and Ross 2016) and, more interestingly, studying the flow of information in a distinctive event (De Domenico et al. 2013). This work is also inspired by such an application in several domains. Katz and Lazarsfeld found in (Katz and Lazarsfeld 2006) that the

information is not directly transferred from mass media to common people. Instead, in most cases it is first interpreted by opinion leaders. They called it a “two-step flow communication”. The final recipient can obtain the information directly from mass media without the intervention of opinion leaders or the information can be relayed via opinion leaders to the common people. Compared to mass media, the general impact of opinion leaders on an individual’s choice is enormous. In (Kotler 2007), Kotler characterized opinion leaders as being able to affect other community users on the basis of their unique characteristics, including knowledge, personality, and other unique features. Rogers highlighted in (2003) opinion leaders based upon three characteristics which include prominent social status, significant social responsibility, and major social involvement.

Burt in (1999) identified three major functionalities of opinion leaders, including information-seeker, information-provider, and strong social interaction. Feng proposed a methodology in Feng (2016) for identifying influential users in the online communication network. Huanhuan et al. suggested a new way for opinion leaders to identify the synthesized centrality (SC) (Xiaoqing et al. 2013). The authors calculated the synthesized centrality by multiplying the betweenness centrality (BC) with the normalized degree centrality and then dividing the resultant by the closeness centrality. Sina, one of China’s largest microblogging websites, was used as a network for experimental purposes. It consists of over four thousand users forming a local area network called SHU-LAN from Shanghai University. The authors over this network computed and compared the PageRank, HITS, and SC. The final findings showed that SC is better in comparison with PageRank and HITS to identify opinion leaders.

Twitter data was used by Eugene in (2017) to analyze the online communities while Baek and Kim in (2015) and Akar and Mardikyan in (2018), Akar et al. (2018) worked to identify the role of users in the online community. Cataldi et al. in (Cataldi et al. 2013) proposed an unsupervised scheme for predicting a user’s influence in a community. They classified all tweets and created a domain exchange graph for each class by deploying N-gram classifier; then they analyzed the diffusion of information in these graphs and assessed a user’s influence on each community. Lee et al. did work on how the Twitter network is being used by scholars for informal communication (Lee et al. 2017). Khan et al. used the Twitter network to demonstrate the differences between the government departments of USA and South Korea (Khan et al. 2014).

In the literature, many researchers concentrated on a kind of follower–followee sort of relationship between users (Java et al. 2007; Kwak et al. 2010; Takhteyev et al. 2012). However, this sort of relationship does not provide a reasonable approximation for the real relationships among

users. On Twitter, a single user may follow dozens of users and their tweets appear in the user’s news feed that follows them, despite the fact that many users do not have the real interaction among them. In brief, there is a need to analyze the tweets that they produce to know this kind of emerging social relation, and interaction among Twitter users. Mainly, there exist three kinds of interactions among Twitter users shown in Fig. 1 and that we will focus on in our research.

Retweet (RT) A user endorses and broadcasts the information shared by the other users to his/her own followers. This implies that a user is retweeting the tweet of another user.

Reply (RP) This indicates a user-to-user interchange as a reaction to the information confined in a user’s tweet.

Mention (MT) This signifies that the user has shared some information by referencing another user in the tweet.

Cha et al. in (2010) proposed that a user can have three different types of influence (i.e. an in-degree influence, retweet influence, and mention influence) and concluded that users can be grouped based on their sort of influence. Also, after analyzing these influences, Cha et al. suggested that in-degree influence is related to popularity. Kwak et al. in (2010) compared the retweets, number of followers and PageRank results, and reassure that each criterion will lead to distinct user groups.

The work was done to measure the blogger’s influence in the social community and various sorts of influential bloggers were identified (Agarwal et al. 2008). The study was further extended to aggregate the individual blogs and bloggers were ranked from individual and community blogs (Agarwal and Nitin 2008). The authors of Moon and Han (2010) found influential users via the similarity between the bloggers and by the flow of information within the network.

Also, they categorized popular and influential bloggers and measured the influence of the blogger in social community. Furthermore, various kinds of influential bloggers were identified. The research was also conducted to classify the online users into followers and leaders by using the information that who they interact. The study was conducted by using the Facebook data (Shafiq et al. 2013).

Different approaches exist in the literature for the detection of opinion leaders. In Canali and Lancellotti (2012), principal component analysis (PCA) was deployed to select and combine user attributes for finding influential users. Bamakan et al. (2019) made a comparison of different approaches systematically including data mining and learning methods, hybrid content mining, descriptive, statistical and stochastic and topological measures, etc. Furthermore, the authors discussed the pros and cons of these methods to provide an understanding of the current research challenges (Bamakan et al. 2019). In this proposed research, we used a topological based approach to identify opinion leaders. A comparative study of centrality measures including degree centrality, closeness centrality, betweenness centrality, eigenvector centrality is discussed by Arrami et al. (2018). Our research work uses the degree centrality and betweenness centrality as a metric to identify the opinion leaders.

3 Key users in online discussion network and their interaction patterns

Five kinds of key users along with their interaction patterns are enlisted in this section.

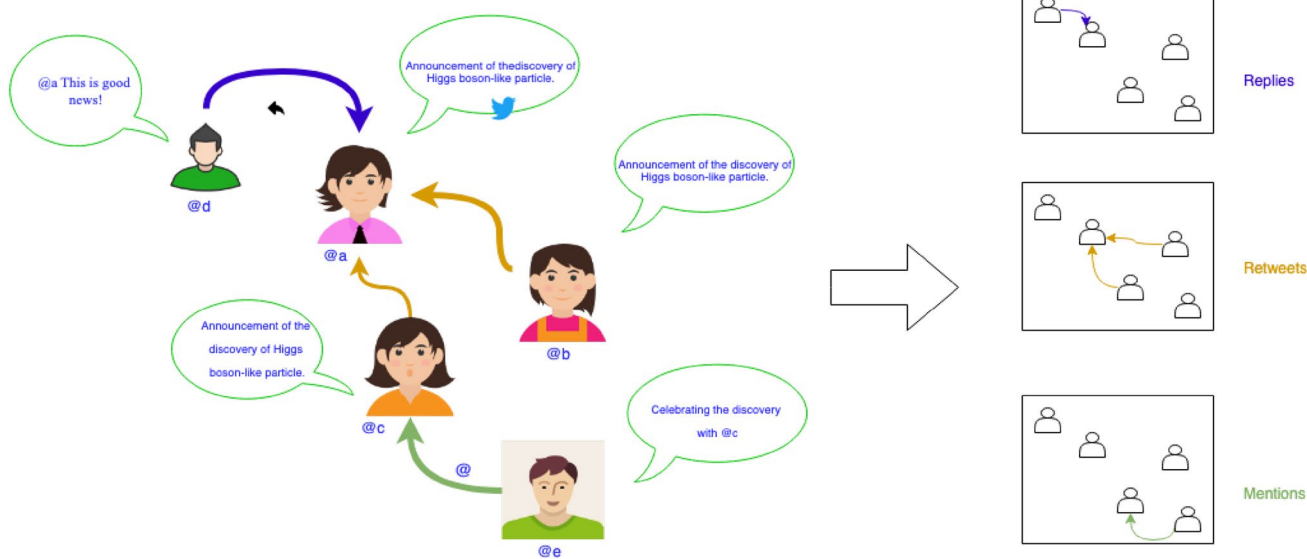


Fig. 1 Representation of different interactions among Twitter users. Different colors are used for different actions

3.1 Conversation starter

A user with a lot of ‘in-degree’ connections and a very small or no ‘out-degree’ link is appointed in online social networks as a conversation starter (see Fig. 2a). This user is the one who originally begins the discussion on that particular subject and is also accountable for the network flow information.

3.2 Influencer

As shown in Fig. 2b, in a network an influencer has many ‘in-degree’ connections and some ‘out-degree’ connections. By generating countless tweets that are retweeted by numerous other users, an influencer has a lot of impacts on the other network users. In addition, an influencer connects many isolates that otherwise have no network link. An influencer acts as an opinion leader in a network, since other users in the network retweet and mentions the influencer in their tweets.

3.3 Active engager

A user with an abundance of ‘out-degree’ and few or none ‘in-degree’ is an active engager in the network (see Fig. 2c). It disseminates the information in the network by connecting to other network users. However, as few users are retweet and mention active engager in their tweets, it is not the opinion leader.

3.4 Network builder

Despite getting some ‘out-degree’ connections and few or none ‘in-degree’ connections, network builder still has a significant role in the network. Network builder connects two or more network influencers. Figure 3a depicts the interaction pattern for network builder, which connects two influencers.

3.5 Information bridge

The information bridge contains some ‘in-degree’ and ‘out-degree’ connections in the network. The key role of an information bridge is to connect an active engager and influencer.

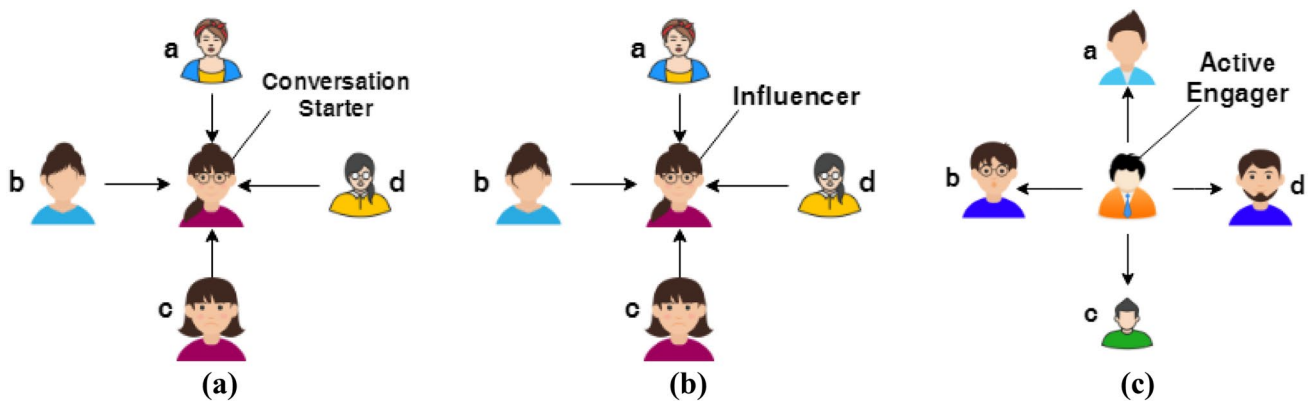


Fig. 2 Links association of a conversation starter, b influencer, c active engager, in the network

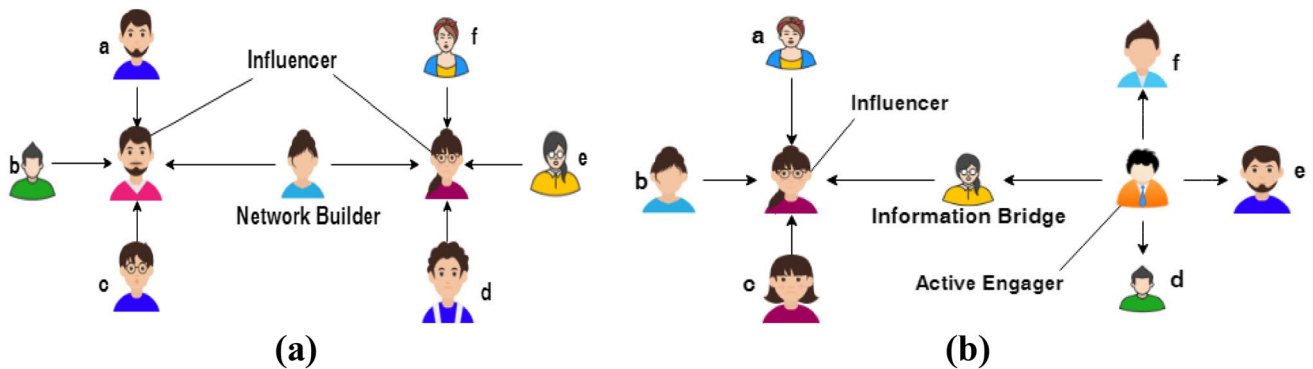


Fig. 3 Links associations of a network builder, b Information Bridge, in the network

Figure 3b shows the Information Bridge interaction pattern linking an influencer and an active engager.

4 Experimental analysis and results

4.1 Dataset description

A directed Twitter network named as Higgs-Twitter available publicly at (Jure Leskovec Stanford Large Network Dataset Collection 2019) is used in our experimentation. Previously, the dataset was also used in De Domenico et al. (2013), Omodei et al. (2015), Al-garadi et al. (2017) for the experimentation. The dataset was collected on July 4, 2012. It contains the dataset from 1 to 7 July 2012. The available dataset covers the time span before, during and after the discovery of the particle having the resemblance with the Higgs Boson. The identity of the users is not revealed to keep them nameless and ID is allocated to each user. In addition, users found in each dataset are allocated the same IDs. Mainly, the data are the Retweet network consisting of 256,491 nodes and 328,132 edges, mention network of 116,408 nodes and 150,818 edges and reply network consisting of 38,918 nodes and 32,523 edges.

4.2 Conversation starter user

According to Borgatti et al. centrality is the main idea in the theory of social network (Borgatti et al. 2018). It is linked with the participant’s ability to influence the inner dynamics of the network owing to the place of that participant. The degree centrality and betweenness centrality are among the different measures of centrality. The conversation starter is ranked first or second among all the users for all three types of data sets which are RP, MT, and RT, based on the betweenness centrality as shown in Table 1. A user is characterized in the network by the degree of centrality which evaluates the user’s individual standing in the network by looking at the edges connecting to that specific node. This can be further distinguished by in-degree (when the node starts the interaction) and out-degree (when the interaction is addressed to the node). The in-degree of conversation starter is 1206, 11,953 and 14,060 for the reply, mention and retweet network respectively. Whereas the out-degree is 7, 7 and 3 for these networks. The degree centrality of the user i is given below

$$D_i = \sum_{j=1}^n a_{ij}, \quad i \neq j \tag{1}$$

where a_{ij} is the amount of contacts between i and j . The connection pattern for a conversation starter for all networks reveals a similar sort of behavior where all users along with isolates are attracted toward the conversation starter.

A conversation starter is ranked at position 1 or 2 in all the three networks including RP, MT, and RT based on the high value of BC i.e. 1,216,103.466, 51,202,891.041, and 44,671,510.507 respectively. Whereas, the BC of the user i is given by Eq. (2)

$$B_k = \sum_i^n \sum_j^n \frac{g_i k_j}{g_{ij}}, \quad i \neq j \neq k \tag{2}$$

where n is the number of points in the graph, g_{ij} is the number of routes from i to j and $g_i k_j$ represents the amount of routes through a single k (Freeman 1977, 1978). The connection pattern of the conversation starter for reply, mention and retweet network respectively are shown in Table 1. In summary, the conversation starter user acts as a hub in this network. The simulations were performed in Gephi (Bastian and Heymann 2009; Cherven 2015) and MultiGravity ForceAtlas2 Layout (Jacomy et al. 2014) was deployed. It has the capability to visualize the complex graphs resulting from the large-scale dataset.

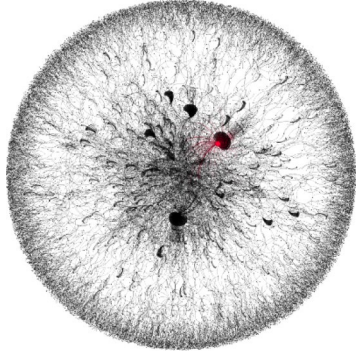
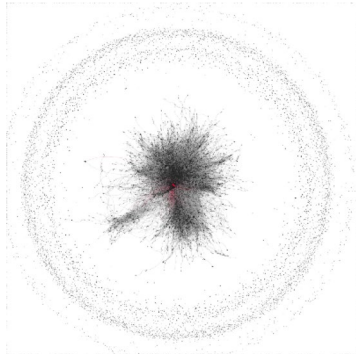
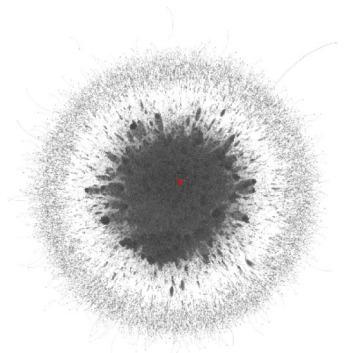
4.3 Influencer user

For the reply network, the influencer user is the most central user with the top position among all the nodes in the network because it has the highest betweenness centrality value of 1,427,809.475. In addition, the online discussion network can have more than one influencer. Four influencer users with the in-degree of 436, 773, 1587, 3906 and out-degree of 23, 21, 5, 14 are identified for the mention network. These users are ranked at 2, 3, 10, 11 based on their BC values as the results are shown in Table 2. Similarly, two major influencer users with the in-degree values of 5613, 4335 and the out-degree links 8 and 2 were identified for the retweet network. These users rank at positions 8 and 10 in the network based on their BC values as shown in Table 2. From the subgraphs of influencers for all three kinds of networks, a similar sort of pattern is depicted.

4.4 Active engager user

An active engager has very less in-degree links and plenty of in-degree. For mention network, a user with an in-degree value of 24 and out-degree value of 169 is an active engager. It is ranked at the fourth position in the online discussion network based on the value of its betweenness centrality which is 26,829,538.854 as in shown Table 3. In a retweet network, the user is active engager with the betweenness value of 8,065,069.849. It has the in-degree value of 11 and out-degree links are 48 as demonstrated in Table 3. Furthermore, the user is ranked at the 23rd position based on the value of its BC.

Table 1 Presentation of ‘tweet type’, ‘node id’, ‘subgraph’, ‘in-degree’, ‘out-degree’ and ‘betweenness centrality’ for nodes that play a part as ‘Conversation Starter’ within the Higgs network

Tweet type	ID	In-degree	Out-degree	Betweenness	Rank	Subgraph
Reply network	677	1206	7	1216103.466	2	
Mention network	88	11953	7	51202891.041	1	
Retweet network	88	14060	3	44671510.507	2	

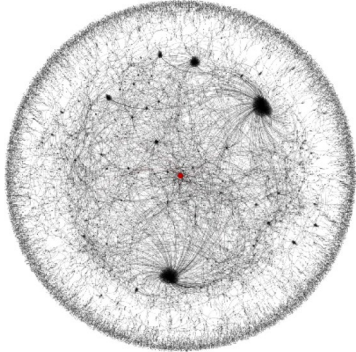
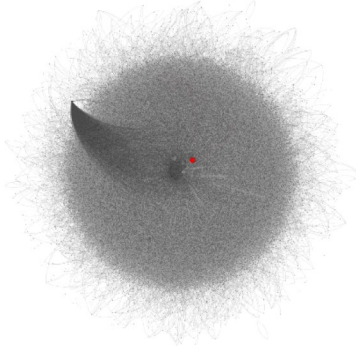
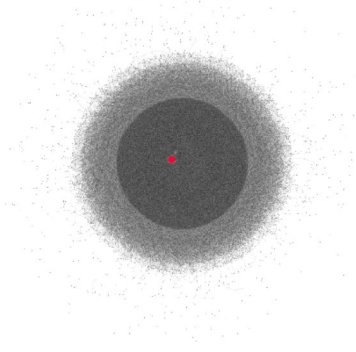
4.5 Network builder and information bridge user

For Retweet Network, a center node is a network builder as shown in Fig. 4a. The size of the node in Fig. 4a indicates its degree centrality, a node with a high value of degree centrality is bigger in size and vice versa. The arrow going out from a node indicates its out-degree link, while the arrow coming towards the node indicates its in-degree link. The central node has only 2 in-degree and 19 out-degree connections. Although, the network builder has only 2 in-degree connections and 19 out-degree connections, but is ranked 23 among 256,491 nodes in the network. The reason that it has got a major position in the network is that it is connecting two influencers with degree 2803 (in-degree = 2802,

out-degree = 1) and 1677 (in-degree = 1668, out-degree = 9), respectively.

For the mention network, a node with ID 110,903 acts in the capacity of both network builder and as well as an information bridge. In spite of having very less value of degree centrality which is 70 (in-degree = 39, out-degree = 31). This node is ranked 7th in the network with the value of betweenness centrality 22,887,429.871. While it acts as a network builder, it connects two influencers which are ranked at the 3rd and 10th positions in the network. The node acts as an information bridge as it connects an active engager ranked at the 4th position and an influencer ranked at the 3rd position in the network. The connection pattern for the node with ID 110,903 is shown in Fig. 4(b). Furthermore, nodes having

Table 2 Presentation of ‘tweet type’, ‘node id’, ‘subgraph’, ‘in-degree’, ‘out-degree’ and ‘betweenness centrality’ for nodes that play a part as ‘Influencer’ within the Higgs network

Tweet type	ID	In-degree	Out-degree	Betweenness	Rank	Subgraph
Reply network	13808	46	9	1,427,809.475	1	
Mention network	64911	436	23	5027,9287.657	2	
Retweet network	677	5613	8	18,914,597.087	8	

a lot of in-degree and out-degree links with one another are shown with a thick line among them.

5 Community detection for analysis of overall networks

Figure 5a demonstrates the pattern in which the users interact with each other in the reply network. There are certain prominent communities (Akar and Mardikyan 2018) surrounding conversation starter and influencers. These detected communities help in the fast dissemination of information in the network (Lambiotte and Panzarasa 2009). Furthermore, many users having lower degree value

are connected with conversation starter and influencer, we may say that these opinion leaders (conversation starter and influencers) have the tendency to keep connected many low degree users. The low degree users connected with opinion leaders do not share the information with each other. However, they are in connection with opinion leaders because they want to get the updated information from the highly influential users. In Fig. 5a, the graph visualization for the complete reply network is demonstrated by using a standard community detection algorithm called Louvain (Blondel et al. 2008). The deployment of the Louvain algorithm results in 10,696 communities. Many isolates can be seen at the edge of the graph as they are not linked with any key node in the network. Communities around conversation

Table 3 Presentation of ‘tweet type’, ‘node id’, ‘subgraph’, ‘in-degree’, ‘out-degree’ and ‘betweenness centrality’ for nodes that play a part as ‘Active Engager’ within the Higgs network

Tweet type	ID	In-degree	Out-degree	Betweenness	Rank	Subgraph
Mention network	89805	24	169	26,829,538.854	4	

Retweet network	27705	11	48	80,650,69.849	21	
-----------------	-------	----	----	---------------	----	--

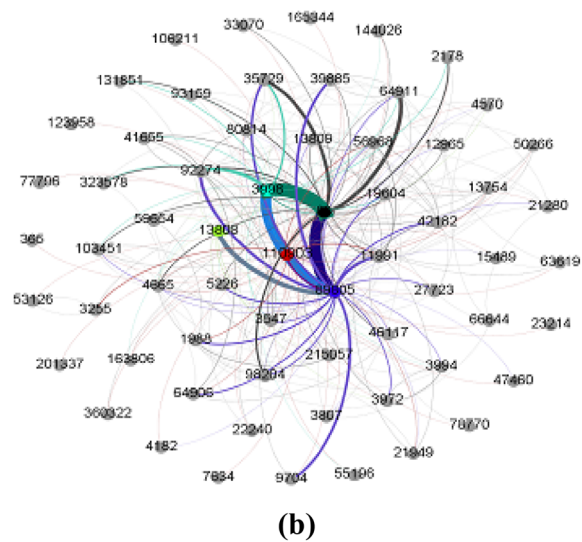
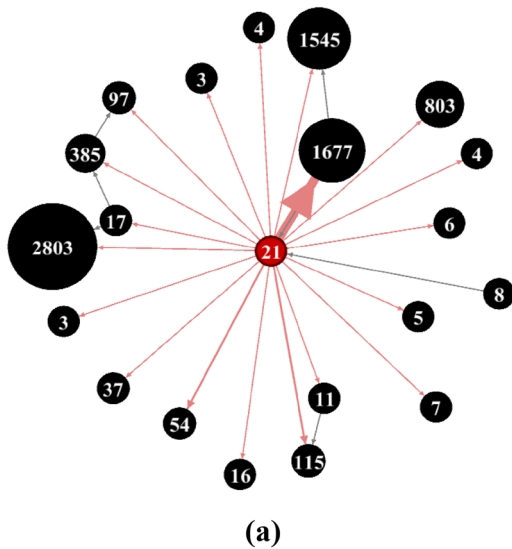
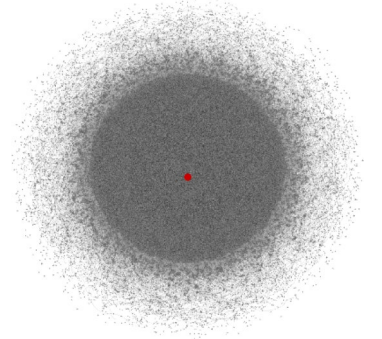
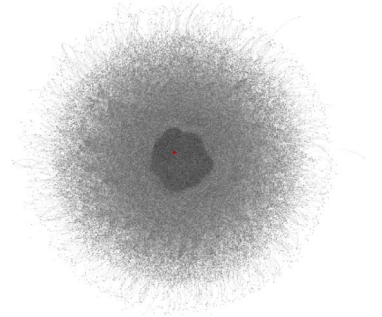


Fig. 4 a Network builder’s connection pattern for the retweet network. The middle node is the Network Builder. **b** Connection pattern for the Network Builder & Information Bridge for the Mention Network. The middle node is acting as both the Network Builder and Information Bridge

starter and influencer are identified and due to the presence of conversation starter and influencer, the diffusion of information within the community is very fast. In Fig. 5b, the network is filtered for better visualization and only eight communities are remaining. After filtering, the isolates and

weak communities are removed from the network. Although there are less than 1% of communities left, it still consists of more than 15% of the nodes and 21% of the edges. In Fig. 5c, the left column shows the number assigned by the Gephi (Bastian and Heymann 2009; Cherven 2015) to identify that

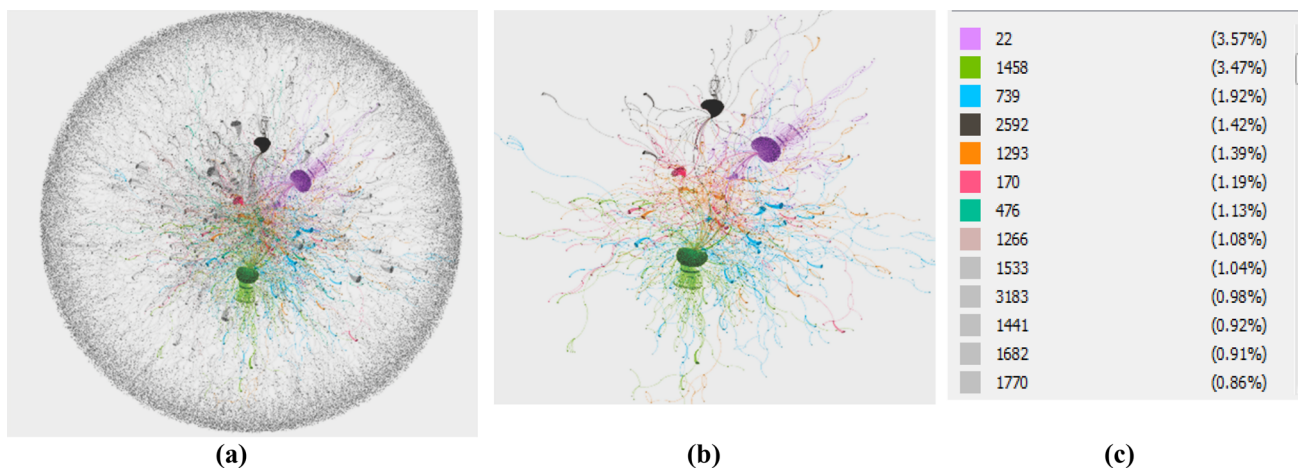


Fig. 5 Graph visualization of Reply Network by deploying Yifan Hu Algorithm in Gephi. **a** Whole network, **b** filtered network with only 8 Communities. **c** Constitution of different Modularity classes

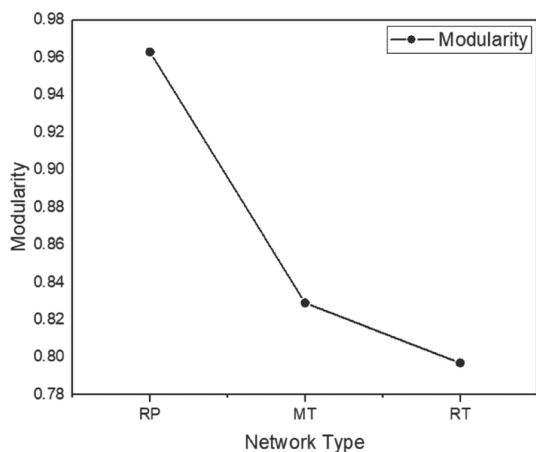


Fig. 6 Modularity values of the networks (i.e. Reply, Mention and Retweet)

specific community and in the right column, it shows the percentage covered by that specific community.

The Louvain method is used to calculate the modularity. The modularity value of all three network types is shown in Fig. 6. The modularity values for the reply, mention and retweet network are 0.963, 0.829 and 0.797, respectively. The modularity algorithm searches for the nodes that are attached to each other more densely than to the remainder of the network. The strength of information dissemination is high when the modularity value is high. Moreover, when all communities have the same attractiveness, there is no diffusion of information within the network (Cui and Zhao 2017). A similar sort of pattern is identified for the mention and retweet network as shown in Figs. 7 and 8, respectively. All three kind of networks shows a similar sort of behavior.

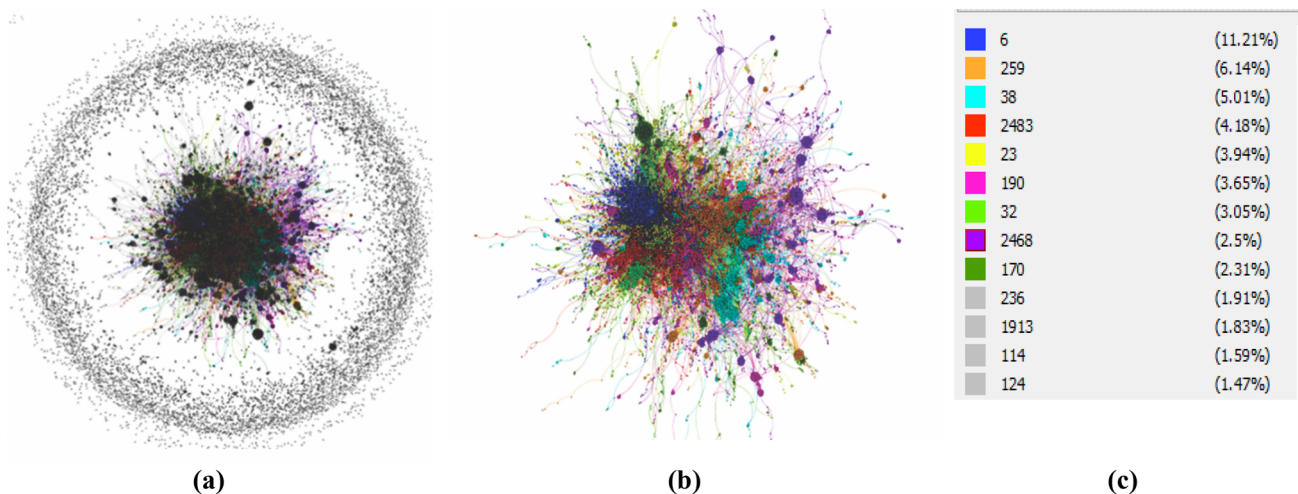


Fig. 7 Graph Visualization of Mention Network by deploying MultiGravity ForceAtlas 2 Algorithm in Gephi. **a** Whole network, **b** filtered Network with only 9 Communities. **c** Constitution of different Modularity classes

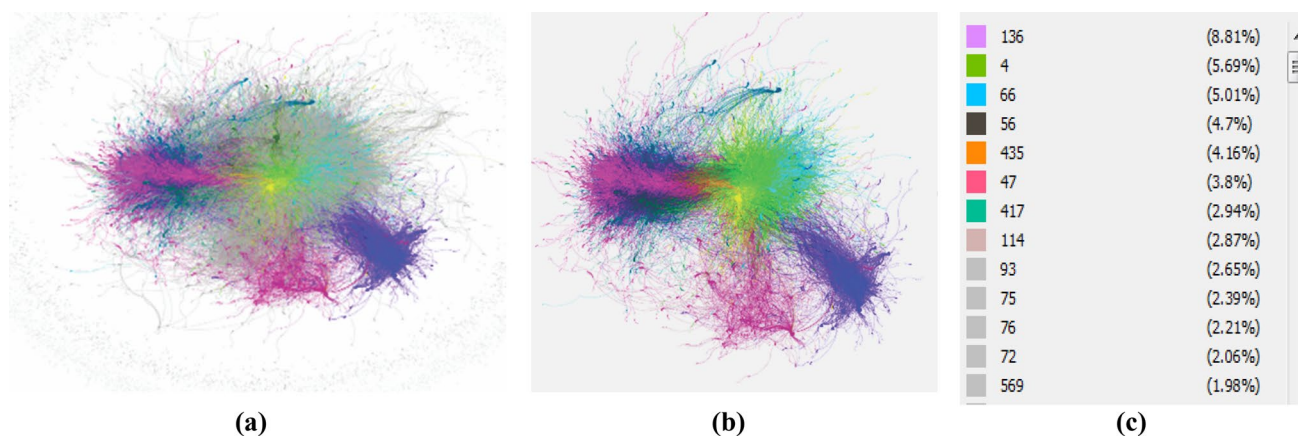


Fig. 8 Graph Visualization of Retweet Network by deploying MultiGravity ForceAtlas 2 Algorithm in Gephi. **a** Whole network, **b** filtered network with only 8 Communities. **c** Constitution of different Modularity classes

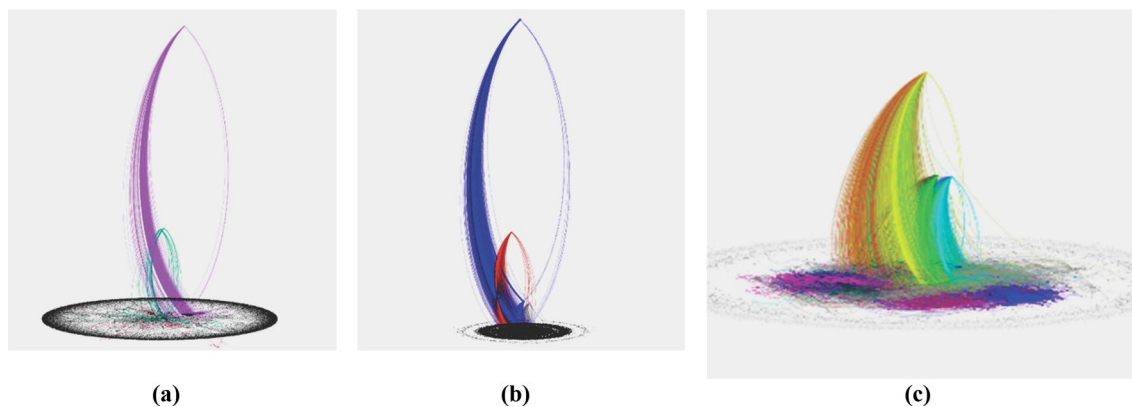


Fig. 9 Graph visualization by using network 3D splitter layout for **a** reply network, **b** mention network and **c** retweet network

The Network Splitter 3D (Barão 2014) applied after the Yifan Hu algorithm results in Fig. 9a and Network Splitter 3D applied after MultiGravity ForceAtlas2 Layout results in Fig. 9b, c. In Fig. 9a, the purple node having the highest peak is the conversation starter for the reply network. It has the ID 677 and degree centrality of 1213 (In-degree = 1206 and out-degree = 7) and lies at rank 2 in the network based on the high value of BC which is 1,216,103.466. The second peak is of the influencer. In Fig. 9b, the blue node mounting at the top is the conversation starter for the mention network and is ranked at 1 in the mention network based on the highest value of BC. Two influencers ranked at positions 2 and 10 also have minor blue peaks. In addition, one influencer with rank 3 occupies the grey peak. Finally, the red node with ID 677 mounting up is the influencer. As stated in Table 2, this node has an in-degree of 3906 and out-degree of 14. In addition, this node is at rank 11 based on the BC value of 16,531,206.031. In Fig. 9c, the node at the top of the highest peak is the conversation starter node for the retweet network. The node ID is

88 and stands at the second position in the network on the basis of the high value of BC which is 44,671,510.507. Two influencers already identified and shown in Table 2 occupy the other two peaks as shown in Fig. 9(c). These influencers with the ID 677 and 1988 have value of degree centrality of 5621 (In-degree = 5613, out-degree = 8) and value of degree centrality of 4337 (In-degree = 4335, out-degree = 2) respectively. Furthermore, these influencers rank at position 8 at 10 out of 256,491 nodes in the network.

6 Discussion

Our findings show that all kinds of users found by the Feng's may not exist in every network (Feng 2016). The detailed discussion about these identified users from the used dataset is demonstrated in this section. In addition, we have assessed the roles of these central users in the Twitter network.

Conversation starter user has the highest rank (first or second) in all kind of networks (RP, MT, and RT) and is the most central participant in all considered networks for this study as shown in Fig. 10. An influencer has many ‘in-degree’ connections and some ‘out-degree’ connections. Conversation starter and influencer are opinion leaders and many isolates mention them or retweet their tweets. Opinion leaders generate extensive content in order to influence other users’ opinions. However, the impact of opinion leaders is inconsistent. Making the comparison of Information diffusion among all three networks, information within the communities is diffused fastly for reply network due to the high value of modularity as shown in Fig. 6. The links among the communities are very less which shows weak ties among the communities. Users tend to cluster around conversation starter and influencers and soon the users become less independent for the reply, mention and retweet network as modularity value decreases (Xu et al. 2017).

A user with an abundance of ‘out-degree’ and few or none ‘in-degree’ is an active engager in the network (Feng 2016). Active engager for the reply network is not found, whereas these are identified for the mention and retweet network. Active engager does not have the tendency to become an opinion leader as none of the users from the network mentioned active engager in his/her tweet.

A network builder has some ‘out-degree’ connections and few or none ‘in-degree’ connections (Feng 2016). The key role of the network builder is to connect two or more influential users. Although the in-degree of network builder is two, still its rank is 23 among the network of 256,491 nodes for the retweet network as shown in Fig. 10. For mention network, the rank of network builder is 7. Influencers users do not necessarily mention or retweet the network builder,

therefore, the contribution of network builder in opinion formation is petite.

The information bridge contains some ‘in-degree’ and ‘out-degree’ connections in the network. Information bridge is only identified for the mention network and connects an active engager and an influencer. In summary, the conversation starter and influencers are the sources of information within the network while the active engager, network builder and information bridge act as a bridge connecting communities with one another and diffuse the information among them.

The findings from our work are significant, but still some limitations of this study are that the used dataset consists of the tweets from 1 to 7 July 2012. The dataset covers the time span before, during and after the discovery of the particle having the resemblance with the Higgs Boson. The analysis of only 7 days dataset would not be enough to make the exact connection pattern analysis of the identified users. So, a dataset consisting of a longer duration is recommended for future studies. Furthermore, this study does not use out-degree centrality as a measure to rank the user. In future studies, more variables may be included to make a model that is associated with the out-degree links of the users.

In spite of these limitations, this work contributes and advances the knowledge to identify the central users within the online discussion network. Also, the study unveils the connection patterns of the identified users and provides the motivations for future research directions.

7 Conclusion

In this study, we identified the different kinds of prominent users in the online discussion network and their associated connection patterns are visualized. Betweenness centrality measure was used to rank the users within the network. The in-degree and out-degree connections of a user help to identify the position of a user in the network, the central users are identified based on these centrality measures. More importantly, this sort of study helps in spreading the information in an efficient way in the marketing campaigns by identifying opinion leaders like conversation starter and influencers. Although the opinion leader (conversation starter) cannot control the opinion of masses participating in the discussion but the discussion started by the conversation starter was rapidly distributed within the network by the other four central users (influencer, active engager, network builder and information bridge).

Influencer usually has a high level of credibility among the masses. Network builder mention or retweet other influencers and opinion leaders in their tweets. While the information bridge seeks information in a regular manner and acts as a source of information provider for other users.

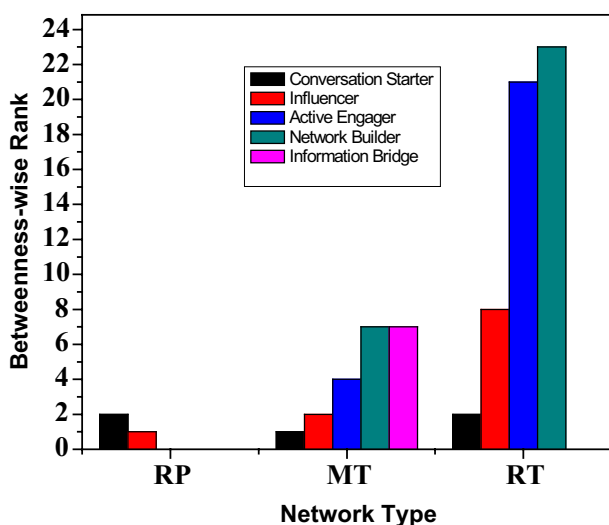


Fig. 10 Betweenness-wise rank of central users found in each network

In future studies, the contents of the tweets may be included to get a better idea of the relationship among the participants of the online discussion network. Furthermore, positive and negative opinions from the tweets can be segregated to construct the classified communities to show their trend. Also, we have identified and ranked the opinion users based on their centrality value, but in the future, some more variables can be added to this model.

Acknowledgements This work was supported in part by the National Nature Science Foundation of China under grants 61471157, 61772090, and 61801055, in part by the Fundamental Research Funds for the Central Universities under Grant 2018B23014, and in part by a National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2017R1C1B5017464).

References

- Abdul R, Paul A, Gul MJ et al (2018) Exploiting small world problems in a SIoT environment. *Energies*. <https://doi.org/10.3390/en11082089>
- Adalat M, Niazi MA, Vasilakos AV (2018) Variations in power of opinion leaders in online communication networks. *R Soc Open Sci* 5:180642. <https://doi.org/10.1098/rsos.180642>
- Agarwal N, Nitin (2008) A study of communities and influence in blogosphere. In: Proceedings of the 2nd SIGMOD PhD workshop on Innovative database research—IDAR'08. ACM Press, New York, New York, USA, pp 19–24
- Agarwal N, Liu H, Tang L, Yu PS (2008) Identifying the influential bloggers in a community. In: Proceedings of the international conference on Web search and web data mining—WSDM'08. ACM Press, New York, New York, USA, pp 207–218
- Akar E, Mardikyan S (2018) User roles and contribution patterns in online communities: a managerial perspective. *SAGE Open* 8:1–19. <https://doi.org/10.1177/2158244018794773>
- Akar E, Mardikyan S, Dalgic T (2018) User roles in online communities and their moderating effect on online community usage intention: an integrated approach. *Int J Hum Comput Interact* 35:495–509. <https://doi.org/10.1080/10447318.2018.1465325>
- Aleahmad A, Karisani P, Rahgozar M, Oroumchian F (2016) OLFinder: finding opinion leaders in online social networks. *J Inf Sci* 42:659–674. <https://doi.org/10.1177/0165551515605217>
- Al-garadi MA, Varathan KD, Ravana SD (2017) Identification of influential spreaders in online social networks using interaction weighted K-core decomposition method. *Phys A Stat Mech its Appl* 468:278–288. <https://doi.org/10.1016/j.physa.2016.11.002>
- Arrami S, Oueslati W, Akaichi J (2018) Detection of opinion leaders in social networks: a survey. Springer, Cham, pp 362–370
- Baek SI, Kim YM (2015) Longitudinal analysis of online community dynamics. *Ind Manag Data Syst* 115:661–677. <https://doi.org/10.1108/IMDS-09-2014-0266>
- Bamakan SMH, Nurgaliev I, Qu Q (2019) Opinion leader detection: a methodological review. *Expert Syst Appl* 115:200–222. <https://doi.org/10.1016/J.ESWA.2018.07.069>
- Barão A (2014) The Gephi Network Splitter 3D Layout. <http://www.relationalcapitalvalue.com/gephiplugins.html>. Accessed 10 Jun 2019
- Bastian M, Heymann S (2009) Gephi : an open source software for exploring and manipulating networks. In: Third International Conference on weblogs and social media. AAAI Press, Menlo Park, CA, pp 361–362
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 10:P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Borgatti SP, Everett MG, Johnson JC (2018) Analyzing social networks, 2nd edn. SAGE PublicationsSage CA, Los Angeles
- Burt RS (1999) The social capital of opinion leaders. *Ann Am Acad Pol Soc Sci* 566:37–54. <https://doi.org/10.1177/000271629956600104>
- Canali C, Lancellotti R (2012) A quantitative methodology based on component analysis to identify key users in social networks. *Int J Soc Netw Min* 1:27. <https://doi.org/10.1504/IJSNM.2012.045104>
- Cataldi M, Mittal N, Aufaure M-A (2013) Estimating domain-based user influence in social networks. In: 28th Symposium on applied computing (SAC). ACM, Coimbra, Portugal, pp 1957–1962
- Cha M, Haddadi H, Benevenuto F, Gummadi KP (2010) Measuring user influence in Twitter: the million follower fallacy. In: International conference on Weblogs and Social Media (ICWSM). AAAI Press, pp 10–17
- Cherepnalkoski D, Karpf A, Mozetič I, Grčar M (2016) Cohesion and coalition formation in the european parliament: roll-call votes and twitter activities. *PLoS One*. <https://doi.org/10.1371/journal.pone.0166586>
- Cherven K (2015) Mastering gephi network visualization. Packt Publishing Ltd., Birmingham
- Clow KE, Baack D (2004) Integrated advertising, promotion, and marketing communications, 6th edn. Pearson Education Inc, Upper Saddle River
- Cui X, Zhao N (2017) Modeling information diffusion in time-varying community networks. *Chaos*. <https://doi.org/10.1063/1.5002577>
- De Domenico M, Lima A, Mougél P, Musolesi M (2013) The anatomy of a scientific rumor. *Sci Rep* 3:1–9. <https://doi.org/10.1038/srep02980>
- Eugene C (2017) The bottom-up formation and maintenance of a Twitter community. *Ind Manag Data Syst* 115:612–624. <https://doi.org/10.1108/eb057530>
- Farhan A, Abbasi R, Rehman A, Gyu SC (2019) An advanced algorithm for higher network navigation in social internet of things using small-world networks. *Sensors* 19:1–20. <https://doi.org/10.3390/s19092007>
- Feng Y (2016) Are you connected? Evaluating information cascades in online discussion about the #RaceTogether campaign. *Comput Human Behav* 54:43–53. <https://doi.org/10.1016/j.chb.2015.07.052>
- Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40:35–41. <https://doi.org/10.2307/3033543>
- Freeman Linton C (1978) Centrality in social networks conceptual clarification. *Soc Netw* 179:215–239
- Jacomy M, Venturini T, Heymann S, Bastian M (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* 9:1–12. <https://doi.org/10.1371/journal.pone.0098679>
- Java A, Song X, Finin T, Tseng B (2007) Why we twitter: understanding microblogging usage and communities. In: 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis—WebKDD/SNA-KDD'07. ACM Press, New York, USA, pp 56–65
- Jure Leskovec Stanford Large Network Dataset Collection (2019) <http://snap.stanford.edu/data/>. Accessed 12 Apr 2019
- Katz E, Lazarsfeld PF (2006) Personal influence : the part played by people in the flow of mass communications. Transaction Publishers, Piscataway
- Keller E, Berry J (2003) The influentials: one American in ten tells the other nine how to vote, where to eat, and what to buy. Simon and Schuster, New York

- Khan GF, Yoon HY, Park HW (2014) Social media communication strategies of government agencies: twitter use in Korea and the USA. *Asian J Commun* 24:60–78. <https://doi.org/10.1080/01292986.2013.851723>
- Khan NS, Ata M, Rajput Q (2015) Identification of opinion leaders in social network. In: 2015 International Conference on information and communication technologies (ICICT). IEEE, Karachi, pp 1–6
- Kotler P (2007) *Marketing management*, 12th edn. Pearson Education Inc, Upper Saddle River
- Kwak H, Lee C, Park H, Moon S (2010) What is Twitter, a social network or a news media? In: 19th International conference on World wide web. ACM, Raleigh, NC, pp 591–600
- Lambiotte R, Panzarasa P (2009) Communities, knowledge creation, and information diffusion. *J Informetr* 3:180–190. <https://doi.org/10.1016/j.joi.2009.03.007>
- Lamirán-Palomares JM, Baviera T, Baviera-Puig A (2019) Identifying Opinion Leaders on Twitter during sporting events: lessons from a case study. *Soc Sci* 8:141. <https://doi.org/10.3390/socsci8050141>
- Lee MK, Yoon HY, Smith M et al (2017) Mapping a Twitter scholarly communication network: a case of the association of internet researchers' conference. *Scientometrics* 112:767–797. <https://doi.org/10.1007/s11192-017-2413-z>
- Liben-Nowell D, Kleinberg J (2007) The link prediction problem for social networks. *J Am Soc Inf Sci Technol* 58:1019–1031. <https://doi.org/10.1002/asi.20591>
- Mitchell L, Ross JV (2016) A data-driven model for influenza transmission incorporating media effects. *R Soc open Sci* 3:160481. <https://doi.org/10.1098/rsos.160481>
- Moon E, Han S (2010) A qualitative method to find influencers using similarity-based approach in the blogosphere. In: 2010 IEEE Second International Conference on social computing. IEEE, pp 225–232
- O'Sullivan DJP, Garduño-Hernández G, Gleeson JP, Beguerisse-Díaz M (2017) Integrating sentiment and social structure to determine preference alignments: the Irish Marriage Referendum. *R Soc Open sci* 4:170154. <https://doi.org/10.1098/rsos.170154>
- Omodei E, De Domenico M, Arenas A (2015) Characterizing interactions in online social networks during exceptional events. *Front Phys* 3:1–8. <https://doi.org/10.3389/fphy.2015.00059>
- Rehman AU, Aimin J, Rehman A, Paul A (2019) Weighted based trustworthiness ranking in social internet of things by using soft set theory. In: 2019 IEEE 5th International Conference on computer and communications (ICCC). IEEE, Chengdu, China
- Rogers EM, Everett M (2003) *Diffusion of innovations*, 5th edn. Free Press, New York
- Sadiq MT, Yu X, Yuan Z et al (2019) Motor imagery EEG signals classification based on mode amplitude and frequency components using empirical wavelet transform. *IEEE Access* 7:127678–127692. <https://doi.org/10.1109/ACCESS.2019.2939623>
- Shafiq MZ, Ilyas MU, Liu AX, Radha H (2013) Identifying leaders and followers in online social networks. *IEEE J Sel Areas Commun* 31:618–628. <https://doi.org/10.1109/JSAC.2013.SUP.0513054>
- Stever GS, Lawson K (2013) Twitter as a way for celebrities to communicate with fans: implications for the study of parasocial interaction. *N Am J Psychol* 15:339–354
- Takhteyev Y, Gruzd A, Wellman B (2012) Geography of Twitter networks. *Soc Netw* 34:73–81. <https://doi.org/10.1016/j.socnet.2011.05.006>
- Trusov M, Bodapati AV, Bucklin RE (2010) Determining influential users in internet social networks. *J Mark Res* 47:643–658. <https://doi.org/10.1509/jmkr.47.4.643>
- Xiaoqing Y, Jing L, Huanhuan L (2013) Identifying top-n opinion leaders on local social network. In: IET International conference on smart and sustainable city 2013 (ICSSC 2013). IET, Shanghai, pp 268–271
- Xu WW, Park J, Park HW (2017) Longitudinal dynamics of the cultural diffusion of Kpop on YouTube. *Qual Quant* 51:1859–1875. <https://doi.org/10.1007/s11135-016-0371-9>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.