ORIGINAL ARTICLE

Social network analysis of Twitter interactions: a directed multilayer network approach

Austin P. Logan[1] · Phillip M. LaCasse[2] · Brian J. Lunday[2]

Abstract

Effective employment of social media for any social influence outcome requires a detailed understanding of the target audience. Social media provides a rich repository of self-reported information that provides insight regarding the sentiments and implied priorities of an online population. Using Social Network Analysis, this research models user interactions on Twitter as a weighted, directed network. Topic modeling through Latent Dirichlet Allocation identifies the topics of discussion in Tweets, which this study uses to induce a directed multilayer network wherein users (in one layer) are connected to the conversations and topics (in a second layer) in which they have participated, with inter-layer connections representing user participation in conversations. Analysis of the resulting network identifies both influential users and highly connected groups of individuals, informing an understanding of group dynamics and individual connectivity. The results demonstrate that the generation of a topically-focused social network to represent conversations yields more robust findings regarding influential users, particularly when analysts collect Tweets from a variety of discussions through more general search queries. Within the analysis, PageRank performed best among four measures used to rank individual influence within this problem context. In contrast, the results of applying both the Greedy Modular Algorithm and the Leiden Algorithm to identify communities were mixed; each method yielded valuable insights, but neither technique was uniformly superior. The demonstrated four-step process is readily replicable, and an interested user can automate the process with relatively low effort or expense.

Austin P. Logan, Phillip M. LaCasse and Brian J. Lunday have contributed equally to this work.

✉ Phillip M. LaCasse
  phillip.lacasse@afit.edu

  Austin P. Logan
  austin.logan.2@us.af.mil

  Brian J. Lunday
  brian.lunday@afit.edu

[1] Directorate of Plans, Programs, and Requirements, Air Combat Command, 129 Andrews Street, Langley Air Force Base, VA 23665, USA

[2] Department of Operational Sciences, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson Air Force Base, OH 45433, USA

1 Introduction

Social media provides an open environment for individual users to send and receive information within the greater online community, simultaneously influencing and being influenced by it. This influence may take the form of mass content dissemination to a wide audience. It could also consist of targeted, tailored advertising to individuals based on their online behaviors. A less benign example is the use of propaganda or misinformation to shape public opinion or sow discord.

Twitter is a compelling source for social network analysis because its content is heavily text-based compared other common, image-based platforms, such as Instagram, Snapchat, or Tik Tok. Moreover, the unique features of Tweet

sharing and hashtag labeling distinguish Twitter from sites such as Facebook where post sharing is less emphasized.

For the most part, the capability for widespread social influence is restricted to entities with the resources and technical skills necessary to perform large-scale analysis of open source data. This research sets forth and demonstrates a set of analyses that are readily usable by smaller entities without a large capital investment.

This research makes four contributions to the larger discipline of social network analysis.

1. We propose and demonstrate a framework to analyze Twitter data consisting of the following four steps:

   - Topic discovery
   - Construction of the multilayer network
   - Identification of influential users and topics
   - Detection of communities within the network

2. We propose and demonstrate a multilayer network structure consisting of a user layer and a topics layer. This structure leverages relationships between and among the layers to provide meaningful insight into the network and its potential influence. As far as we can determine, the proposed structure is unique to this research.

3. We propose and demonstrate two new SNA arc weighting techniques: one between users based on the inverse of the long-term proportion of interactions and the other between topics based on cosine similarity.

4. We evaluate four alternative methods to identify influential users and two alternative algorithms to discover communities.

The remainder of the manuscript is organized as follows. Section 2 reviews the technical literature related to topic modeling and social network analysis (SNA), including methods previously used in these fields that can answer key research questions. Section 3 explains the methodology to conduct the social network analysis, to include creation of the multilayer network and the metrics to evaluate it. Section 4 presents the results of applying the proposed SNA process to a population of Tweets and discusses both the algorithmic results and instance-specific implications. Finally, Sect. 5 concludes the work and recommends how one may best apply the process and leverage the study's insights.

2 Literature review

This study leverages existing research related to social network analysis, influencer identification, sentiment analysis, community detection, and both directional relationship and multilayer modeling between entities in SNA.

On some level, social ties and connections have been worthy of study dating back to antiquity, as evidenced by the presence of genealogies in ancient texts such as the Bible or Greco-Roman poems and histories. Freeman (2004) traces the development of social network analysis from early sociometric studies at Sing Sing prison (Moreno 1932) and the Hudson School for Girls (Moreno 1933) through its formal establishment as a rigorous discipline in the 20th century. In particular, four features characterize modern social network analysis: structural intuition based on ties linking social actors, grounding in systemic empirical data, employment of graphic imagery, and quantification via rigorous mathematical modeling. (Freeman 2004)

Scott and Carrington (2011) defines social network analysis as the specific logic behind the relationships that people choose to form and maintain, resulting in a social configuration that can be represented graphically. This SNA approach leverages connections between entities to construct a graphical representation of a network comprised of nodes and arcs, wherein the arcs convey the relative strength of connections (Legradi 2009), as determined by mathematical analysis. Within this context, Allard (1990) outlines two main goals of SNA: understanding the factors that affect relationships and their correlations, and ascertaining the effects of these relationships, including the possible identification of an informal leader.

An important aspect of social media culture research related to this work is the phenomenon of influencers. Zhang and Vos (2015) examine social media culture in depth and conclude that the most effective way to spread a message on social media is through highly influential users known as influencers. Influencers have acquired the reputation of being compelling and reliable sources of information and are connected to large numbers of users who follow, comment on, and share their messages.

Given a social network representation of people and their interactions, it is therefore important to identify these disproportionately influential individuals. Several studies (e.g., Bakshy et al. (2011); Erlandsson et al. (2016); Dewi et al. (2017); Bhavnani et al. (2021)) research methods to characterize the influence of nodes within networks. Such methods range in approach from direct observations such as counting the average number of interactions by others for a user's Tweets (Erlandsson et al. 2016), to indirect inferences such as representing physical interactions among people as a network and modeling how quickly a virus would spread from an individual (Doerr et al. 2013). Of note, whereas many of the most influential users in a social network have a large number of followers, follower count alone is not a strong enough metric for quantifying influence (Erlandsson et al. 2016). Additionally, Pudjajana et al. (2018) identifies several centrality metrics useful to compare the influence of nodes, which Jin (2020) well demonstrated. Also related to

influential node identification, Sheth et al. (2022) and Venkatesan and Prabhavathy (2019) study methods to discover anomalous users within social networks.

However, it is not sufficient simply to identify influencers; one must also characterize the messages that they share with followers. Sentiment analysis is a technique to label a message as either positive or negative, and it can effectively monitor users' emotions towards a topic over time. Tsugawa and Ohsaki (2015) and Salehi et al. (2018) outline many of the methods to perform sentiment analysis. Featherstone and Barnett (2020) employ self-reported attitude scores to validate sentiment scores obtained from a comprehensive study on public opinion towards genome editing. Results are promising, although the strength of the relationship between attitude score and sentiment did vary between the subgroups sampled (Featherstone and Barnett 2020).

Moreover, sentiments can affect consumer behavior. As examples, both Gazdaggyori (2021) and Hamraoui and Boubaker (2022) apply sentiment analysis to Tweets related to financial stock performance. Although the growth of the studied stocks was inconsistent and relatively short-lived, both studies demonstrate that the sentiment of social media users reflects changes in consumer attitude that influence investor behaviors. The same phenomenon is observed in the pro-vaccine and anti-vaccine communities; each community contains its stable of influencers whose messages and sentiments produce the predictable effects on vaccination coverage in children (Featherstone et al. 2020).

Other aspects of SNA of interest to this research are topic modeling and community discovery. Topic modeling determines the most frequent topics of discussion in a collection of Tweets. The two primary topic modeling methods are Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA). Kalepalli et al. (2020) directly compare LDA and LSA, and both Rahmadan et al. (2020) and Yang et al. (2021) demonstrate the efficacy of LDA for topic modeling using Twitter Data. In a related work, Jiwanggi and Adriani (2016) detail methods for extracting a summary of topics from a collection of Tweets. Topic modeling is frequently useful to model the evolution of public discourse on topics of interest such as vaccination (Featherstone et al. 2020) or gene editing (Ji et al. 2022). Community discovery methods seek to identify the closely connected groups of users in a network, whether by individual communications or communications related to a common topic such as vaccine hesitancy (Ruiz et al. 2021). Various user features can help detect communities of connected users, as Pacheco et al. (2021) recently studies.

Of interest to this research is the modeling of social networks with directional representation of interactions. Aiello et al. (2010) characterizes representation decisions for constructing social networks and the links between users. Relationships and communication are often asymmetric, and

directed social networks better represent these interactions. However, directed social networks are not a modeling panacea, and both Malliaros and Vazirgiannis (2013) and Tsopze and Domgue (2021) describe the challenges associated with their modeling and analysis.

For SNA, a multilayer network can better represent the complexities of interactions better than a single-layer network model. Multilayer networks can include entity-specific layers and include edges or directed edges (i.e., arcs) to represent different interactions and the relative strength thereof. Figure 1 depicts an example of a multi-layer network with a user layer, and topic layer, and inter-nodal arcs both within and between layers.

Some SNA techniques are unique to multilayer networks, directed networks, or the combination of both. Kolda et al. (2005) define methods to quantify node influence in multilayer networks. Tang and Liu (2011) propose methods to detect communities in directed networks. This research leverages each of these contributions.

3 Methodology

Section 3.1 describes the analyzed datasets. Sections 3.2 and 3.3 contain the approaches for creating the user and topic layers in the multilayer network. Finally, Sects. 3.4 and 3.5 present the methods used to discover influential users and communities.

3.1 Description of datasets

This research employs four traditional datasets studied within the SNA literature, listed in Table 1, along with customized datasets created by sampling Tweets from each of the four named sets. The Tweets are almost exclusively
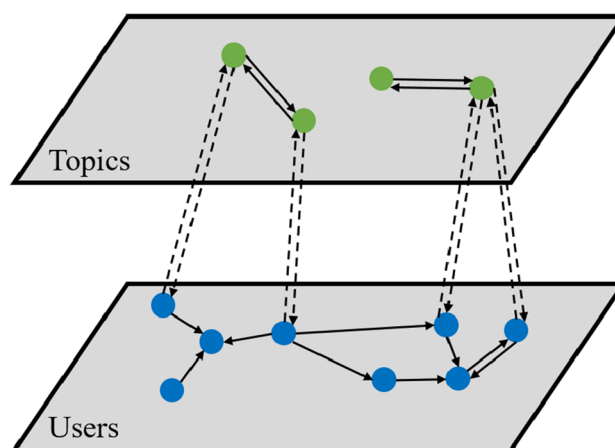

Fig. 1 Multilayer Social Network Representation of Entity-specific Layers and Directed-arc Representation of Interactions

Table 1 Summary of datasets and key features

Dataset name	# Tweets	Retweets	Replies
2018 World Cup	530,000	Yes	No
Game of Thrones	760,660	No	Yes
2016 US Election	42,013	Yes	Yes
COVID-19	179,108	No	Yes

written in the English language, and the information associated with each Tweet consists of username, number of followers, date of account creation, and verification status. The topics of conversation are sports, entertainment, politics, and health, respectively.

The first step in this analysis is to demonstrate the analytical techniques on the well understood datasets in Table 1. Such analysis can reveal differences in results pertaining both to the identification of influential users and to general network metrics. It can also provide insight into Tweet query practices and their effect on the resulting social network representation.

Next, we apply the techniques to customized datasets created by sampling a uniform number of Tweets from each of the named datasets. These customized datasets include more diverse topics and conversations, and they better represent the topical diversity of the Twittersphere. Moreover, such datasets help validate community identification and topic modeling methods because it is reasonable to expect topics and communities identified in the composite dataset to map to those in the four original datasets.

As mentioned in Sect. 1, the intended beneficiaries of this framework are entities without the budget required for more expensive APIs that collect large amounts of data about individual Tweets. As such, analytic techniques herein consider only the Tweet data available from a low-cost accessible API: Tweet text, username, and verification status. As is typical with most analyses, some minor data cleaning was necessary to ensure that subsequent analysis only considered Tweets with an associated username.

3.2 User network (layer) creation

Preliminary to the four-step analytic process, it is first necessary to generate a user network – or network layer, in the case of a multilayer network – for a dataset of Tweets. Within such a layer, nodes represent users and edges represent relationships between users. As Sect. 2 discusses, this research adopts a directed network with arcs to model Twitter relationships because it better represents the directional relationships between users. The directed network models actions taken by users to show their relationships with others as an outbound arc and Tweets about a user or in response to a user's Tweet using an inbound arc. In doing so, the directed

network can distinguish a celebrity who does not Tweet frequently but has many followers from a bot or spammer that Tweets frequently about other users.

Edges or arcs within a social network have associated weights to convey the strength of the connection implied by interactions between users. All of the datasets within Table 1 include three types of user interactions: a user mentioning another user, a user Retweeting another user's Tweet, and a user replying to another user's Tweet. Although no formal direct measure of user relationships exists, these interactions can inform a proxy metric that represents the relative, implied strength of relationships via arc-specific weights.

Some interactions imply a closer relationship between users. This is evident by observing the ratio of likes to Retweets for nearly every Tweet in the Twittersphere. Tweets consistently have far more likes than Retweets (e.g., Perdana and Pinandito (2018)), implying that a Retweet conveys a stronger engagement with a Tweet than a like. Moreover, Tweets typically have fewer mentions in new Tweets than Retweets, and fewer replies than mentions, indicating increasing degrees of engagement.

For this reason, the inverse of the frequency with which replies, mentions, and Retweets occur can provide a suitable proxy for the strength of connection implied by an interaction. For example, if the distribution of replies, mentions, and Retweets was uniform, then any action would contribute the same weight (i.e., $1/0.3\bar{3} = 3$) to an arc from one user to the author of the original Tweet. If the distribution were 14.3%, 28.6%, and 57.2%, a reply would contribute twice as much weight to the arc (i.e., 7) as a mention (i.e., 3.5) and four times as much as a Retweet (i.e., 1.75). If a user interacts with another user several times within a dataset, the net contributions of the interactions to the arc weight are additive. Among the responses to Tweets in Table 1 datasets, the distribution of interactions consisted of 6.97% replies, 39.49% mentions, and 53.54% Retweets. Thus, each reply, mention, and Retweet contributes 14.35, 2.53, and 1.87 to an arc weight, respectively, for the generation of the user networks or user network layers.

3.3 Topic modeling and integration as a network layer

As Sect. 1 outlines, the first two of the four steps in process are the discovery of topics and the construction of a topic-focused, directed multilayer network. The goal of implementing topic modeling alongside social network analysis is two-fold. First, topic modeling provides a general overview of the discussions contained within a dataset. Second, topic modeling can be used in conjunction with the existing network to connect users through the conversations which they are having, a feature that cannot be extracted directly from Twitter data. Thus, the inclusion of a topical layer with

the user network layer, combined with arcs indicating user participation in the topics, induces a multilayer network to more accurately represent both the direct and indirect connections among users participating in the discourse on the social media platform.

For the first step, this research conducts topic modeling using Latent Dirichlet Allocation (LDA). Pritchard et al. (2000) set forth LDA as an unsupervised clustering method to assign individual creatures (e.g., birds, people) to populations (e.g., species, tribes) based on genotype similarities. Blei et al. (2003) first applied LDA to topic modeling for text-based documents. The authors describe LDA as "a generative probabilistic model of a corpus"; it synthesizes a user-defined number of topics and populates them with the words that have the highest probabilities of belonging to them. The three key elements of an LDA model are the topics, documents, and corpus. Within this research, topics are the clusters to which statistical analysis will assign words from the Tweets. Documents are the individual Tweets appearing in the data, and the corpus is the complete collection of documents. Given k topics and V unique words in the corpus, LDA creates a $k \times V$ probability matrix β, where β_{ij} represents the probability that topic i includes word j.

Preprocessing of Tweets removes stop words, tokenizes the remaining text, and lemmatizes the individual words (i.e., tokens) to ensure only relevant text remains for topic discovery via LDA. Stop word removal deletes common words that provide no contextual meaning, such as articles, conjunctions, prepositions, and pronouns. Tokenization partitions the remaining text into words. Lemmatization replaces different forms of a word (e.g., runner, running, runs) with a common root word (e.g., run).

Only a user-defined number of topics k is necessary to apply LDA to preprocessed Twitter data. Although the optimal number of topics depends on the data, a coherence metric can assess the effectiveness of a k-topic LDA model by assigning a score to the set of highest probability words in each topic based on their similarity and interpretability by a human. Thus, a line search on k can identify the best number of topics to maximize coherence. Although several researchers (e.g., see Bouma (2009); Newman et al. (2010); Mimno et al. (2011)) have developed alternative coherence metrics, Röder et al. (2015) conducted an extensive, comparative study of such metrics and introduced two new metrics, identifying a superlative coherence metric the authors denoted as C_V. This research uses their recommended metric.

Once LDA is complete, the second step generates the directed multilayer network representation. An analyst names the topics after inspecting the highest probability words associated with each topic. Such a task is not arduous, given familiarity with the language-of-origin for the Tweet. Although this manual naming process is not strictly necessary because LDA identifies the topics, it does provide useful context for analysis. The new topic layer consists of a node for each topic, and the LDA results (i.e., β) inform arc creation between the user layer and the topic layer.

The two aspects of arc creation are which arcs to generate and how to weight those arcs. This research generates user-to-topic arcs for only the strongest Tweet-to-topic relationship for each of the user's Tweets, as measured by a similarity index of words within a Tweet to each of the k topics. For a given Tweet and a vector s of length V, wherein s_v is the number of times token v appears within the Tweet, the Tweet-to-topic similarity index for each topic i equals $\beta_i \cdot s$. As an aside, although it is possible for a single Tweet to relate to multiple topics rather than only its most relevant topic, such an alternative is perhaps a compelling sequel to this work, albeit a more computationally burdensome endeavor.

Additionally, arc generation only creates Tweet-to-topic arcs if the similarity index was in the top 25% of all such maximal indices for the corpus of Tweets. Doing so avoids establishing weak connections between users and topics. Although no formal research exists to determine such a threshold, future work could utilize labeled training data with a machine learning approach to explore better decisions in this space.

For this directed multilayer network, this research creates a pair of user-to-topic and topic-to-user arcs as determined by the similarity index. Inducing the opposite-direction topic-to-user arc represents scenarios wherein users scroll through a topic of conversation on Twitter and find another user via their topic-specific Tweets. The weight for each of the arcs in the generated pair is equal to the value of the similarity index.

After establishing connections between the user layer and the topic layer of the multilayer network, the final step to complete the multilayer network model is to represent the connections between the topics in the topic layer. For a given topic, the word probability vector is a vector of length V that indicates in each entry the likelihood that a word is in that topic. From this definition, the cosine similarity between two topics is the angle between their word probability vectors, as Equation (1) calculates for two vectors A and B.

$$\cos^{-1}\left(\frac{A^T B}{\|A\|\|B\|} \right) \tag{1}$$

Although the theoretical domain of Equation (1) is $[-1, 1]$, only a range of $[0, 1]$ is feasible for cosine similarities between topics; each vector is in the non-negative orthant because every element is a non-negative probability.

The interpretation of these values is as follows. A similarity of 0 means the vectors are orthogonal, so no inter-topic relationship exists; a value close to 1 results from nearly parallel vectors, indicating similar relative distributions of

word-to-topic probabilities for the two topics. The threshold to generate a pair of directed arcs between two topics is if their cosine similarity exceeds 0.5. Differing from the user-to-topic connections, arcs may connect a single topic to multiple other topics.

The resulting directed multilayer network includes respective user and topic layers; weighted, directed arcs connecting users based on replies, mentions, and Retweets; pairs of weighted, directed arcs connecting users to topics based on the vocabulary of a user's Tweets, with at most one user-to-topic connection formed by a single Tweet; and pairs of weighted, directed arcs connecting topics based on the cosine similarities of their respective word probability distributions.

Complementing the first two topic-focused steps of the process is the summarization of Tweets linked to each topic. This research creates extractive summaries rather than abstractive summaries, favoring the former for its simplicity. Moreover, the ability of an abstractive summary to generate unique thoughts is mitigated by the methods by which Twitter creates trending topics; they often present the most relevant Tweets within a conversation, an outcome similar to an extractive summary (Rudrapal et al. 2018). This research uses the TextRank algorithm to create extractive summaries. Created by Mihalcea and Tarau (2004), it applies a graph-based ranking technique that induces a graph wherein nodes represent sentences (i.e., Tweets) and edges are weighted by a user-defined sentence similarity metric. This work utilizes the better performing metric (i.e., BM25) proposed by Barrios et al. (2015) in lieu of the alternatives originally set forth by Mihalcea and Tarau (2004). In comparison, BM25 considers the inverse frequency of words within a document to increase the relative similarity metric for documents containing words that are rare in the topic-specific corpus.

For the TextRank generated graph, PageRank subsequently identifies the most important sentences for inclusion in the extractive summary of each topic. Page et al. (1999) proposes the PageRank algorithm to calculate the most important sentences for extractive summaries. The algorithm repeatedly applies an extended random walk on a graph to determine the long-term probabilities of residing at each node. In a random walk, a simulated entity sequentially travels from one node to an adjacent node with a probability equal to the arc weight, relative to the total weights of arcs emanating from the current node. The authors modify the adjacent step probabilities of the random walk to create small, nonzero probabilities of traversing from a given to any other (i.e., non-adjacent) node in the network to mitigate the effect of disconnected network components on long-term probability calculations. The application of multiple random walks from initial entity locations determined via a uniform distribution over the nodes more notably mitigates that effect. Augmenting the list of the highest probability words

for each topic identifiable via β, this extractive summarization of the most relevant topic-specific Tweets provides additional context regarding topics and reduces the creative, cognitive labor required to analyze Twitter data scrapes.

3.4 Influential user identification

For the third step in the process, it is worth noting that there are myriad methods to identify influential nodes within a network. Although many such methods produce reasonable results for smaller, highly-connected networks, the relative performance of the methods depends notably on the network. Given that this research examines larger networks expected to be relatively disconnected, it is relevant to evaluate alternative methods to identify influential nodes. Testing within Sect. 4 compares rankings via PageRank algorithm, Hyperlink-Induced Topic Search (HITS) algorithm, betweenness centrality, and eigenvector centrality. For each of these techniques, a higher computed value indicates greater influence.

As described in Sect. 3.3, the PageRank algorithm uses long-term node visit probabilities for a random walk to rank order the users and infer a relative degree of influence.

Kleinberg (1999) modifies the PageRank algorithm to create the HITS algorithm to identify influential nodes. The author conjectures a conceptual shortcoming of the PageRank algorithm for directed network analysis; whereas PageRank readily identifies *authority* nodes having many inbound arcs, it can underestimate the influence of *hub* nodes having many outbound arcs. It is arguably influential to direct connections, not just to be the directed target of connections. Accounting for both authority and hub behaviors of nodes, the HITS algorithm identifies a root set of nodes via a targeted search query and augments it with all nodes adjacent via outgoing arcs from the root set. For this larger subgraph, the algorithm iteratively updates each node's authority and hub scores to be equal to the sum of the hub and authority scores of nodes respectively connected to or from the node, until convergence. The HITS algorithm yields two metrics, one each for authority and hub rankings. In practice, researchers often average these scores to enable a direct comparison with other influential node identification methods, and this research does likewise.

As a third method to identify influential nodes, betweenness centrality (Freeman 1977) computes for a given node v the frequency with which it is on one of the shortest paths between a pair of nodes (s, t), considered over all node pairs $s, t \in V$, as per Equation (2).

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s, t \mid v)}{\sigma(s, t)} \tag{2}$$

For social network analysis, betweenness centrality computations use the inverse of edge weights as edge distances

because larger weights indicate strong connections that would conceptually correspond to a shorter distance (i.e., an edge more likely to be traversed). A notable downside to this method is that it requires calculating the shortest paths between all pairs of nodes. Although either a repeated application of Dijkstra's Algorithm or the Floyd Warshall algorithm can run in $O(n^3)$ time (Ahuja et al. 1993), such effort remains computationally expensive for larger networks, and both algorithms require modification to identify alternative optima for shortest (s, t)-paths.

Finally, eigenvector centrality (Landau 1895) provides another alternative to identify influential nodes. This metric leverages the idea that nodes of high influence are adjacently connected to other nodes of high importance. Given an $N \times N$ node adjacency matrix $\mathbf{A}$, wherein $\mathbf{A}_{ij}$ is equal to the weight of the connection between nodes i and j, solve the eigenvector equation $\mathbf{A}x = \lambda x$. Designating λ as the largest eigenvalue, the corresponding vector x indicates the respective influence scores for each of the nodes. This metric is conceptually simple and easy to calculate.

3.5 Community detection

For the fourth step in the proposed SNA process, multiple methods for community detection exist in the literature. Among them, this research tests and compares the Greedy Modularity Algorithm (GMA) and the Leiden algorithm (Traag et al. 2019).

Before discussing these methods, it is important to note the characteristics of data that inform such choices. In this research, communities are identified using only the information contained within the directed multilayer network's nodes and arcs. Given the nature of social media data, especially data gathered via broad queries of Tweets among many unique users, the resulting social network structures tend to be fragmented. Even when querying data by common keywords, the likelihood of capturing a back-and-forth conversation via Tweets between two or more users is exceedingly small, considering the millions of Tweets posted daily. Accordingly, the idealized version of a social network community as clique subgraph having k nodes and $k(k-1)/2$ edges (or $k(k-1)$ directed arcs) is elusive. Rather, community detection methods must consider the implicit networks of users who are not in direct conversation with each other, but who share the same topics of conversation or common connections with other users. This low level of direct connectivity motivates the use of agglomerative community detection methods, wherein every node begins as a sole member of its own community, and an algorithm iteratively conjoins smaller communities to improve the collective strength of the respective communities, as measured via a customized metric.

The first agglomerative technique this research uses is the Greedy Modularity Algorithm (GMA). The GMA is a modification of the Clauset–Newman–Moore (CNM) algorithm set forth by Clauset et al. (2004). Like CNM, GMA is a heuristic approach to maximize a modularity metric that measures the strength of community classification. Whereas Clauset et al. (2004) designed the CNM algorithm for undirected networks, the GMA seeks to maximize the modularity metric in Equation (3), adapted for directed networks when implemented via the NetworkX library (Hagberg et al. 2008) for the Python programming language.

$$Q = \sum_{c=1}^{n} \left(\frac{L_c}{m} - \gamma \left(\frac{k_c^{\text{in}} k_c^{\text{out}}}{2m} \right)^2 \right) \tag{3}$$

Therein, L_c is the number of arcs within community c; m is the total number of edges in the graph; k_c^{in} and k_c^{out} are the sums of the respective in-degree and out-degree weights in community c; and γ is a positive, user-defined resolution parameter to balance the importance of edges within a community and edges connecting communities. Smaller γ-values yield fewer, larger communities, and larger γ-values yield more, smaller communities (Newman 2016). At initialization, there are $n = N$ communities, and $Q \leq 0$ because $L_c = 0$ for $c = 1, ..., n$. Within an iteration, the CNM algorithm calculates the net change to network modularity that would result from conjoining pairs of communities via an edge connecting them. If the maximal such change to modularity is positive, the edge is added and the algorithm proceeds to the next iteration; otherwise, GMA terminates with identified community structures.

The Leiden algorithm is an agglomerative community detection algorithm for undirected, fully connected networks. Fortunately, Malliaros and Vazirgiannis (2013) discussed transformations one can apply to a directed network to enable the application of the Leiden algorithm. First, edges replace pairs of equal-weight, opposite direction arcs between nodes. Second, edges replace singular arcs between nodes, with the same total edge weight. This transformation implies two-way connections that do not exist in the data, but it allows for the exploration of a larger number of community detection methods for which the results can be validated in comparison to the original social network structure. Third, a minimal number of low-weight edges augment the social network to ensure all nodes are fully connected. These artificial connections minimally modify the network representation in a manner that should be negligible but for which the results of any community discovery algorithm should be validated against the original network.

For an undirected multilayer representation of the directed multilayer network via the aforementioned steps, the Leiden algorithm (Traag et al. 2019) can detect

communities via an agglomerative, modularity-focused approach. It also begins by assigning each node to its own community. Each iteration consists of three steps: moving nodes locally, refining a partition of the network, and aggregating nodes within the network. The first step reassigns individual nodes to the community that yields the largest increase in network modularity, partitioning the network into larger, potential communities. The second step refines each of the partitions by re-agglomerating its nodes via stochastic, metric-improving assignments. The third step aggregates nodes within each component of the refined partition. The iteration terminates by assigning the aggregate nodes to their aligned component in the unrefined partition from the first step.

Modularity is not the only metric-of-interest to assess community detection algorithms. Other useful metrics include partition coverage and partition performance. Partition coverage is the ratio of the number of intra-community edges to the number of edges in the graph. The partition coverage metric favors community partitions with few edges connecting communities. The partition performance metric is the ratio of the combined number of intra-community edges and possible inter-community non-edges to the total possible edges in the graph. For the networks in this research, both partition coverage and partition performance scores should be high because social networks representing Twitter data tend to be highly disconnected; most graph partitions would detect communities which isolate many of the fragmented components. Moreover, such networks are not often dense, corresponding to a much larger number of potential edges than actual edges.

4 Testing, results, and analysis

In presenting and discussing the results of applying the SNA process Sect. 3.2 proposes, Sect. 4.1 initially presents visualizations for selected named datasets from Table 1 to derive high-level insights regarding the user layers. Section 4.2 details the results of applying LDA with respect to the number of topics and the corresponding LDA coherence, and it subsequently illustrates a topic layer for an aggregated dataset. Section 4.3 compares the four methods for identifying influential users (i.e., PageRank, HITS, betweenness centrality, eigenvector centrality). Section 4.4 compares GMA and Leiden for detecting communities, both for a single-layer user network and the multi-layer network this research proposes for SNA. Section 4.5 concludes with an examination of query size on the identification of influential users via this process, highlighting the practical implications thereof.

4.1 User network layer creation

A sampling of 15,000 Tweets from each of the named datasets in Table 1 yields the information necessary to create user network layers. Table 2 presents the summary statistics for the user network layers, wherein arcs correspond to replies, mentions, and Retweets.

As a first observation, it is possible to have more nodes and arcs than sampled Tweets if Tweets convey more than one relationship between users (e.g., if a reply to one user's Tweet mentions another user). Such is the case with the 2018 World Cup dataset and nearly the case with the 2016 US Election dataset. By contrast, both the Game of Thrones and COVID-19 datasets yield fewer users and arcs for the same number of Tweets, implying a difference in the nature of communications. Visual depictions of the user network layer can help garner insight in this regard.

While network visualizations can be misleading since node placement is stochastic and semi-arbitrary, they can help identify users with many strong relationships with others. Figure 2 depicts the user network layer for a 5,000 Tweet sample from the 2016 US Election dataset. The graphical depiction results from the Fruchterman-Reingold force-directed algorithm, which represents nodes connected by an arc closer together, reducing arc crossover. Red nodes depict the five most influential users as per the PageRank metric, yellow nodes represent the five users with the highest number of incoming arcs (i.e., authority nodes), green nodes are the users with the highest number of outgoing arcs (i.e., hub nodes), and blue nodes represent all other users.

Although Sects. 4.3 and 4.4 will formally identify influential users and detect user communities, the user network layer depiction does provide preliminary insights. Large clusters of users in these graphs reliably imply influence and community membership. Visible within Fig. 2, clusters of users surround the influential nodes, indicating the strength of their connections and implying a community their communications can affect rapidly. Many nodes also surround the authority nodes, although the implied community is less dense. In contrast, the hub nodes are all in the relative center of the graphical depiction; this graphical depiction understates the potential influence of hub nodes that the HITS algorithm seeks to characterize.

Table 2 User network layer characteristics for 15,000 sampled tweets

Sampled dataset	# Nodes	# Arcs
2018 World Cup	15,073	16,522
Game of Thrones	6,416	4,432
2016 US Election	12,565	13,603
COVID-19	8,457	6,565

Fig. 2 User Network Layer for 5,000 Tweets Sampled from the 2016 US Election Dataset, Depicted via the Force-Directed Algorithm
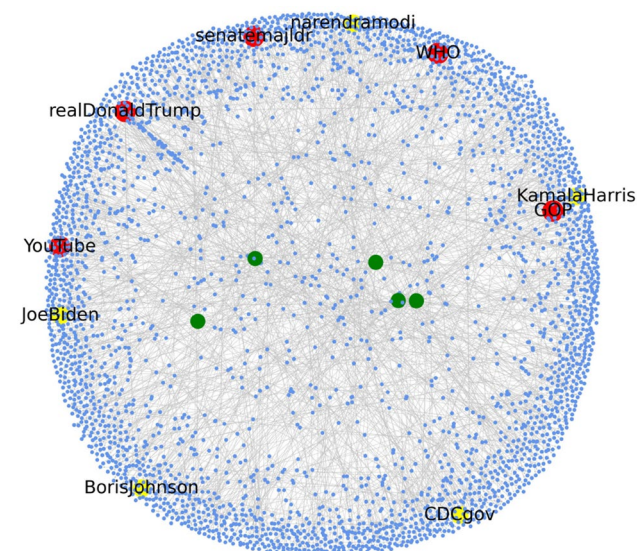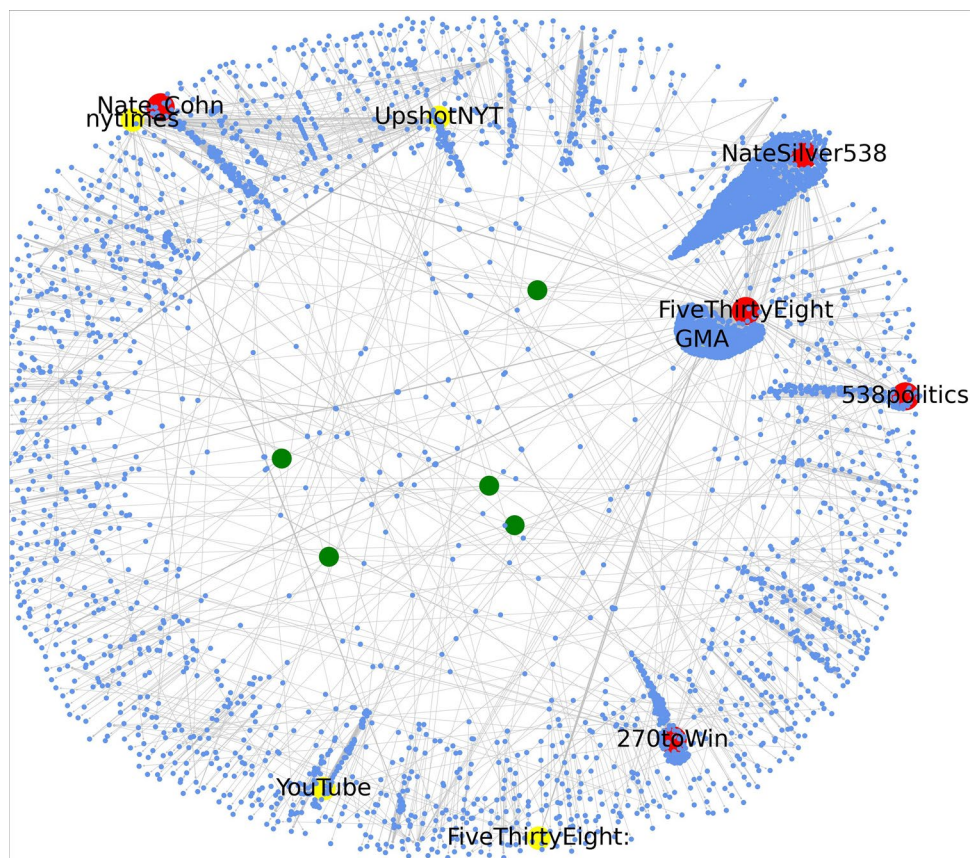

Fig. 3 User Network Layer for 5,000 Tweets Sampled from the COVID-19 Network, Depicted via the Force-Directed Algorithm

For comparison, Fig. 3 depicts the user network layer for the 15,000 Tweet sample from the COVID-19 Network dataset. Similar to Fig. 2, influential users and authority nodes are on the outside of the graph, with many nodes

Table 3 User network layer statistics

Sampled dataset	Network density	Average degree	Average weighted degree
2018 World Cup	7.36×10^{-5}	2.190	0.295
Game of Thrones	1.08×10^{-4}	1.363	0.628
2016 US Election	8.63×10^{-5}	2.169	0.375
COVID-19	8.97×10^{-5}	1.533	0.338

surrounding them. In contrast to Fig. 2, Fig. 3 contains a densely packed outer shell of nodes, which suggests a far more fractured network of individual users connecting with a small number of other users. Figure 2, however, reflects a network by which a relatively small number of influential nodes reaches a large number of users. This is visually evident by the node clusters surrounding the influential users and authority nodes. Of note, the top five hub nodes in Figs. 2 and 3 are near the center of the graph with fewer adjacent nodes, seemingly undervaluing their potential influence.

To illustrate the need for both visualizations and quantitative analysis using established metrics, consider the summary of network statistics in Table 3. Despite having similar

network densities, the user network layers for the 2016 US Election and COVID-19 datasets exhibit very different user behaviors in Figs. 2 and 3. The average degree metric better conveys what the user network layer visualizations depict. The higher degree in the 2016 US Election network shows that, on average, each user interacts with many other users, relative to the COVID-19 network. Comparing the average weighted degrees for those two networks, the relatively close values indicate that, although COVID-19 network users interact with fewer other users, their connections with them are stronger, on average, than the 2016 US Election users. Even absent a visualization of the Game of Thrones user network layer, the statistics in Table 3 convey that users have very strong connections with very few other users, relative to the other datasets.

4.2 Topic modeling and integration as a network layer

Although Latent Dirichlet allocation (LDA) discovers k topics of discussions among a corpus, an analyst must identify the optimal number of topics to extract. This number depends on the data; a collection of Tweets gathered via a very focused, topic-related query manifests fewer topics. As Sect. 3.3 discussed, topic coherence quantifies the performance of LDA. A model exhibiting larger coherence should yield the higher interpretability upon inspection. To illustrate this effect, Figs. 4 and 5 present the LDA coherence for $k = 2, ..., 25$ for the respective samples from the Game of Thrones and 2018 World Cup datasets.

Within Fig. 4, the coherence scores for the Game of Thrones data range from 0.14 to 0.30, and they generally increase over the full range of k explored. In contrast, the coherence scores for the 2018 World Cup exhibit higher average values within range of 0.25 to 0.33, but the effect of the number of topics is more nuanced. There is not a readily
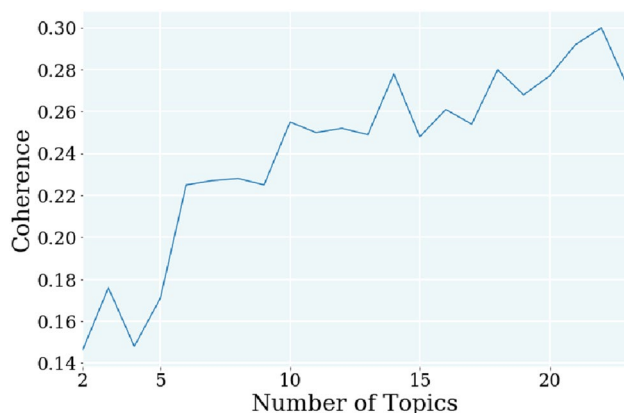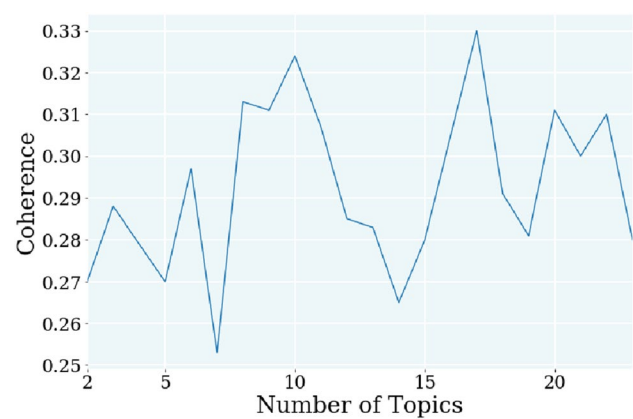
Fig. 5 LDA Coherence versus number of topics for 15,000 tweets sampled from the 2018 World Cup Dataset

discernible trend, indicating a more exhaustive line search on k such as simultaneous search is appropriate when tuning LDA performance. Parsimony suggests $k = 10$ topics as reasonable for the 2018 World Cup dataset.

With LDA as an unsupervised machine learning technique, one can at best conjecture about the reason for the difference in its performance in Figs. 4 and 5. For example, the Game of Thrones dataset tends to manifest opinionated reactions to specific episodes of the show, resulting in relatively differentiated, topic-focused language across the Tweets. In contrast, the 2018 World Cup dataset contains Tweets about a sequence of games, but user descriptions of the actions and players from game-to-game will be less variable. That is, the plot varies among television show episodes more than football matches. To visually depict this difference, Fig. 6 shows the intertopic distances of LDA models for both datasets when $k = 10$ (Sievert and Shirley 2014). The lack of clear topic separation present in the 2018 World Cup Dataset can be seen through the cluster of overlapping topics, especially when contrasted with greater relative distance between topics present in the Game of Thrones Dataset.

To further convey this effect, Table 4 presents the top 10 highest topic-specific probability words for five of the topics modeled by LDA when $k = 10$. Whereas words within each topic exhibit some intuitive relationship, the reuse of some words in several of the topics suggests that many of the Tweets, regardless of their underlying message, use the same verbiage. As a result, the LDA model struggles to clearly discern distinct topics of discussion. Since the best coherence scores for both datasets is approximately 0.3, a low performance for LDA models, topics will be more cryptic and difficult for an analyst to manually label.

Additionally, if coherence is low for the LDA model, Tweets are less likely to exhibit a strong connection with a topic because the words with the highest probability of topic membership will have less semantic connection with each
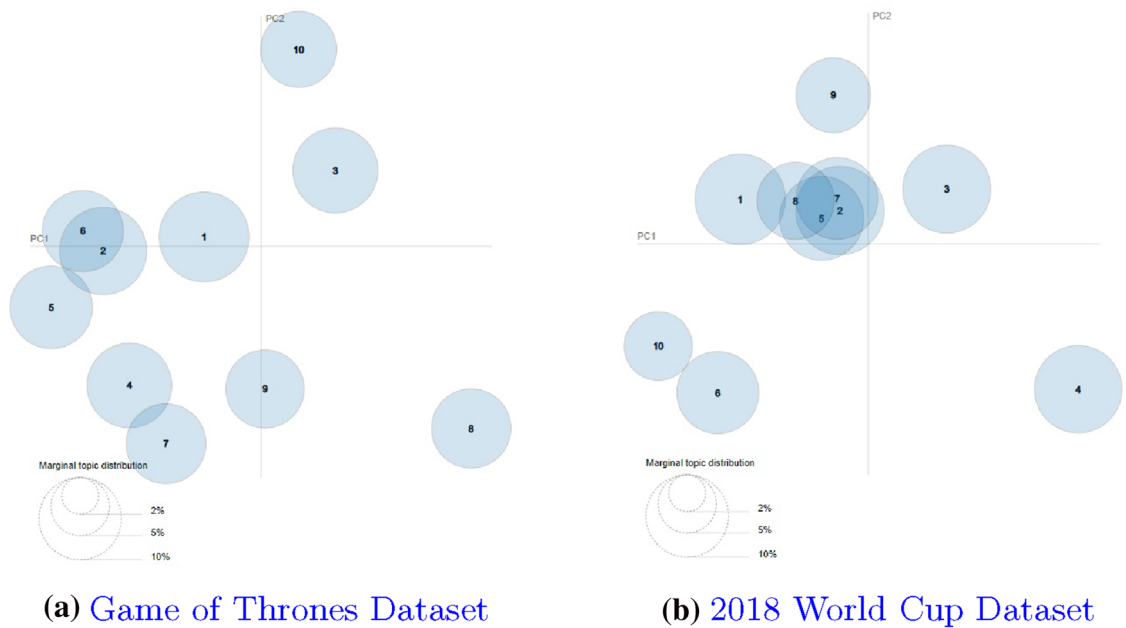
Fig. 4 LDA Coherence versus number of topics for 15,000 tweets sampled from the Game of Thrones Dataset

(a) Game of Thrones Dataset   **(b)** 2018 World Cup Dataset

Fig. 6 Intertopic distance maps of LDA models

Table 4 Selection of topics from LDA model of 2018 World Cup Data

Topic no	Top 10 highest probability words
Topic 0	World, cup, good, sorry, Russia, champion, fifa, congratulation, fra, win
Topic 1	Eng, ronaldo, messi, world, final, England, complete, premierleague, cup, paulpogba
Topic 3	Penalty, frabel, win, team, save, proud, France, eng, time, threelion
Topic 4	Mbappe, fra, player, Kylian, young, award, golden, fifa, score, ball
Topic 9	Team, France, congratulation, win, African, dear, Khaledbeydoun, cut, racism, xenophobi

other. As a direct result, Tweets may be categorized into topics for which the fit is not ideal. Revisiting Table 4, the Tweet "Kylian Mbappé will donate everything he earns playing for France at the World Cup to charity" was connected with Topic 0. Intuitively, this Tweet seems better suited for membership in Topic 4 because it refers to a specific player and his country. However, the tokens 'world' and 'cup' exhibited stronger connection to Topic 0. Such counter-intuitive topic modeling results can affect both influential user identification and community detection, and it motivates the use of broad queries to facilitate higher topic coherence scores for LDA, i.e., more discernible topic modeling.

To determine the effectiveness of LDA on a dataset more representative of a generic query of Tweets, samples were taken from each dataset and conjoined into an aggregate collection of Tweets, hereafter denoted as the Joint dataset. The higher diversity of word usage and topic discussion in the Joint dataset enabled LDA models to attain higher coherence values, as exhibited in Fig. 7.

The range of coherence scores in Fig. 7 generally increases with k, manifesting higher coherence scores just
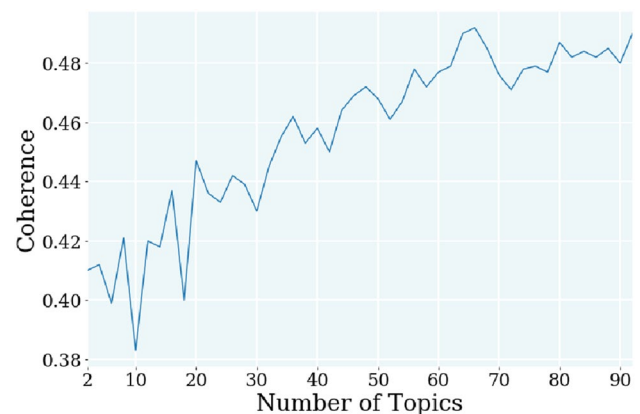

Fig. 7 LDA coherence versus number of topics for the joint dataset

below 0.50. The increase in coherence past 50 topics indicates that, as conjectured, a larger amount of topic separation is possible with a more diverse dataset of Tweets. This finding is important when creating a directed multilayer network that includes a topic layer to help identify influencers and

communities; larger coherence scores better justify connections from users to topics via their Tweets.

Table 5 presents the top 10 highest topic-specific probability words for five of the topics modeled by LDA when $k = 10$. The improved topic separation is apparent when inspecting word membership in topics. In this LDA model, words strongly associated with the different topics appear to come from each of the different datasets, and the greater topic separation is apparent.

High LDA coherence scores and the well separated nature of the identified topics allows for meaningful extractive topic summarization. This activity reduces the work required of an analyst to infer meaning for a topic. For example, within Table 5, both Topics 3 and 8 appear to discuss the forecast of the 2016 US Election, but differentiation of the topics is elusive using only the highest probability words. Extractive topic summaries characterized Topic 3 as "Now, 95% for Trump: Live Presidential Forecast – Election Results 2016 – The New York Times.", whereas it characterized Topic 8 as "RT @DrewLinzer: My final 2016 presidential election forecast: Clinton 323 - Trump 215." These summaries provide added context to convey that Topic 3 is mainly concerned with the conversation surrounding a forecast projecting Trump to win, whereas Topic 8 is discussing a different poll projecting Clinton as the winner. That insight resulting from summaries obviates the need for an analyst to conduct a manual inspection of Tweets. In addition, the intertopic distance map in Fig. 8 reveals that these two topics do occupy distinct spaces despite initially appearing similar. Summaries also provide context when the word membership in a topic makes the topic difficult to identify, in general. For example, the summary of Topic 5 is, "80% of your team is African, cut out the racism and xenophobia. Africa did not win the #Worldcup France did. Africa did not even win it for France", revealing both the controversy aligned with Topic 5 and the countering stances of the users engaged in the discourse.

By augmenting the user layer with the topic layer in a multilayer network, analysis may discover connections between users through the topic layer in the absence of direct connection between them. This modeling characteristic more accurately depicts the dynamics of a social network because users engaged in similar conversations have a higher likelihood of seeing each other's Tweets; such connections are not
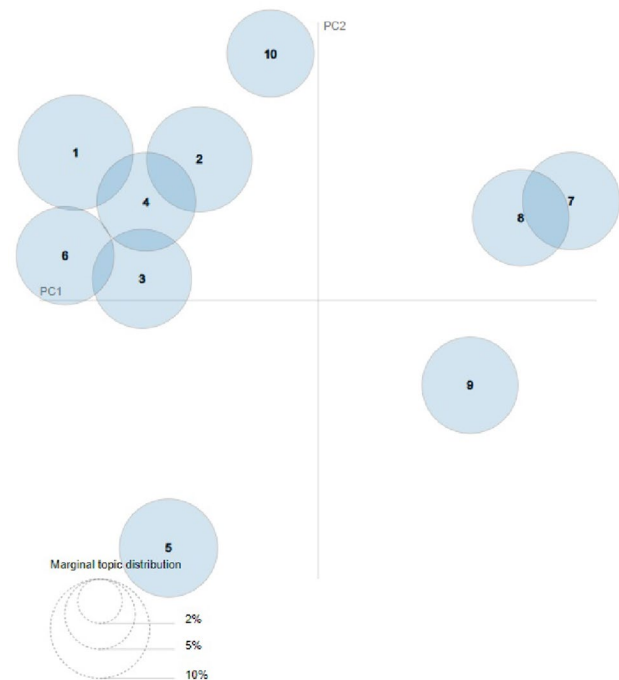

Fig. 8 Intertopic distance map for the joint dataset

direct, but are justified in the model to represent the weaker, more distant relationships.

A visualization of the topic layer for the LDA model referenced in Table 5 can be seen in Fig. 9. Of note, Topics 2 and 4 are a part of the topic layer, but they are not depicted in Fig. 5 because they are not connected to any other topics, as per the methodology set forth in Sect. 3.3. The weights on the edges 9 represent the cosine similarities between topics and are color mapped to show the relative strength of topical similarities. Interestingly, Topics 6 and 8 are not directly connected, but they are connected through other topics with which they are similar. This illustration demonstrates the ability of the topical layer to model relational intricacies in the conversations.

4.3 Influential user identification results

Preliminary analysis applied PageRank and eigenvector centrality to identify influential users, both with and without

Table 5 Selection of topics from LDA model of joint data

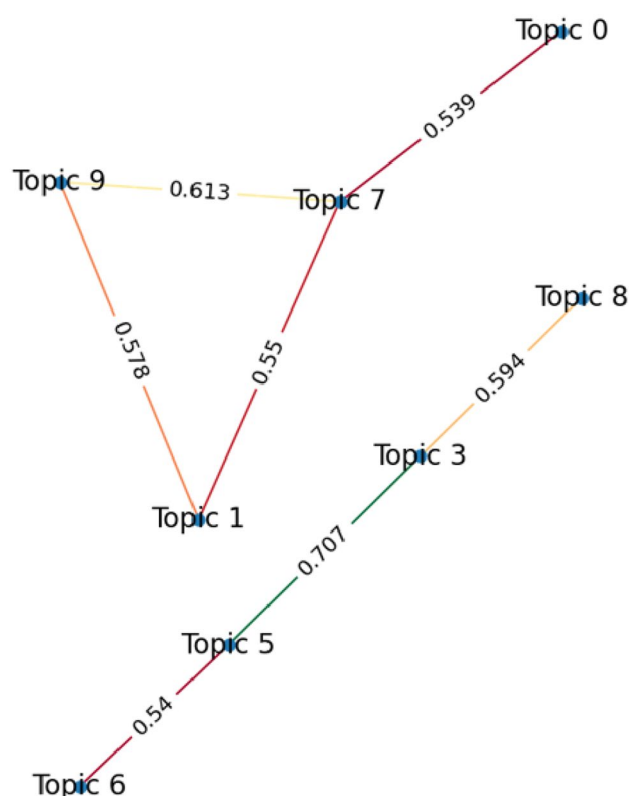| Topic No	Top 10 Highest Probability Words
Topic 1	Like, watch, season, come, episode, time, atch, fifaworldcup, people, go
Topic 3	New, election, forecast, york, presidential, result, times, final, live, fra
Topic 5	France, win, team, election, fra, fifaworldcup, congratulation, forecast, African, trump
Topic 8	clinton, forecast, chance, poll, election, gt, good, trump, fivethirtyeight, vote
Topic 10	Forecast, case, update, election, late, poll, death, new, today, coronavirus

Fig. 9 Topic layer connections from Joint data show affiliations of topics and the strength of their connections

influential user characteristics can assist in determining their relative performance. For example, both PageRank and eigenvector centrality identify several highly influential bots in the multilayer network. This result is counter-intuitive because these bots either scrape data or share news articles but do not actively engage in discussion or offer views to stimulate conversation from other users. Moreover, many of these bots reply with information (e.g., a requested statistic) to users who mentioned them. This dynamic induces an artificially high number of connections with other users. While this information can be useful to an analyst, bots such as these often cannot hold opinions and are therefore of less interest to this research.

Applying PageRank or eigenvector centrality to only the user network layer identified a number of high-profile politicians and government organizations as being influential. This outcome is logical, given the nature of data concerning COVID-19. Although these results are similar, eigenvector centrality identifies as its most influential user an unverified user with fewer than 1000 followers. Such a conclusion seems conceptually unlikely, and the PageRank outcomes do not comport with it.

Whereas topic inclusion exhibited a negative impact on influential user identification when the LDA model was poor, results are more promising for the Joint dataset, which has better coherence and topic separation. For the user layer only and the multilayer network, respectively, Tables 7 and 8 present the top ten identified influential users for the Joint dataset, as determined by PageRank, HITS, betweenness centrality, and eigenvector centrality.

The effect of both the topic layer and the method of influence ranking is evident. Within Table 7, identifying influential users via only the user network layer yields no users common to every ranking, two users (i.e., "Five Thirty Eight" and "GMA") common to three rankings, and four users (i.e., "538Politics", "Ginger_Zee", and "Author") common to two rankings. Different node properties influence the various ranking methods, and all but the HITS algorithm

the topic layer, to assess its impact. Noting the quality of topic identification via LDA can affect the identification of influential users, Table 6 presents the top ten identified influential users for the COVID-19 dataset, for which the topic coherence scores were low and the topic separation was relatively weak.

Findings within Table 6 vary by method and network type. Although there is not a single 'correct' answer to assess the quality of methods, some observations regarding

Table 6 Top Influential Users for COVID-19 Dataset via the User Network Layer and the Multilayer Network, using Selected Techniques

Rank	User network layer		Multilayer network	
	PageRank	Eigenvector	PageRank	Eigenvector
1	Donald Trump	Unverified user	Donald Trump	News Blog
2	YouTube	Donald Trump	ANI	ANI
3	Boris Johnson	Kamala Harris	Bot	Bot
4	WHO	Joe Biden	News Blog	Bot
5	Change	Tamara McCleary	Global Pandemic.NET	Journalist
6	thehill	Business Writer	Journalist	Data Bot
7	CDCgov	Journalist	Medical Journal	Data Bot
8	GOP	CPHO Canada	Data Bot	Medical Journal
9	Narendra Modi	Boris Johnson	Journalist	Global Pandemic.NET
10	Joe Biden	DrRP Nishank	Data Bot	Journalist

Table 7 Top Influential Users for the Joint Dataset via the User Network Layer

Rank	PageRank	HITS	Betweenness	Eigenvector
1	Nate Silver 538	Donald Trump	GMA	538politics
2	538Politics	Unverified user	Five Thirty Eight	matthewjdowd
3	Five Thirty Eight	GOP	Ginger_Zee	RobMarciano
4	FIFA World Cup	Unverified user	ringer	rickklein
5	Nate_Cohn	Unverified user	imarleneking	Five Thirty Eight
6	270toWin	Unverified user	SkyNews	Ginger_Zee
7	Khaled Beydoun	Unverified user	Professor	Peggynoonannyc
8	BBC MOTD	Unverified user	Lawrence	GMA
9	GMA	Unverified user	Author	Author
10	YouTube	Unverified user	ABC San Diego	rachel_handler

Table 8 Top Influential Users for the Joint Dataset via the Multilayer Network

Rank	PageRank	HITS	Betweenness	Eigenvector
1	FIFA World Cup	Unverified user	Five Thirty Eight	Unverified user
2	Nate Silver 538	Unverified user	Nate Silver538	Unverified user
3	Five Thirty Eight	Unverified user	Unverified user	Five Thirty Eight
4	Khaled Beydoun	Unverified user	Unverified user	Unverified user
5	BBC MOTD	Five Thirty Eight	Unverified user	Unverified user
6	ManUtd	Unverified user	katz	Unverified user
7	Donald Trump	Unverified user	Unverified user	Unverified user
8	YouTube	Unverified user	Unverified user	Unverified user
9	brfootball	Unverified user	Unverified user	Unverified user
10	HNS_CFF	Unverified user	Unverified user	Unverified user

identify top influential users who have verified Twitter accounts. Section further examines this phenomenon.

Within Table 8, the rankings determined via the multilayer network are notably different. Pagerank identifies six of the same top ten influential users that it found with only the user network layer. However, the remaining three methods generally identify low profile, unverified users as being highly influential.

When identifying influential users via either the user network layer only or the multilayer network, PageRank outperforms the other methods based on three factors. First, it exhibits relative consistency in identifying some influencers. Second, many of the users PageRank identifies have verified Twitter accounts. Third, many of the same users have hundreds of thousands if not millions of followers. Thus, these influential users have a high in-degree because other users frequently mention them or Retweet their Tweets.

Another characteristic difference between the two sets of rankings is that rankings leveraging only the user network layer tend to identify influential users related to politics, news, or entertainment, whereas the rankings from the multilayer network identify influential users related to politics and sports. This outcome implies that users Tweeting about sports are more likely to be connected via their topics of conversation than via direct conversations, and it reveals opportunities for marketing sports brands and merchandise that a company might otherwise overlook.

4.4 Community detection results

Whereas topic modeling can help find users having specific, topical interests, community detection finds the groups of users having more generally related interests. In doing so, one may design branding or product marketing material for a broader community rather than a topical interest group, thereby engaging with a larger set of potential customers. Of interest is the merit of the directed multilayer network model for detecting communities of users.

As discussed in Sect. 3.5, this research applies both the Greedy Modularity Algorithm (GMA) and the Leiden algorithm to detect communities of users, both for the user network layer only and for the multilayer network. That discussion noted the potential disadvantages of applying the Leiden algorithm to the directed multilayer network: the algorithm applies to undirected networks, so selected transformations are necessary that may reduce model efficacy.

For the 2016 US Election dataset, Table 9 reports the number of communities, modularity, partition coverage, and partition performance for the aforementioned combinations of network models and community detection methods.

Table 9 Community Detection Results for the 2016 US Election Dataset and Alternative Network Models & Detection Algorithms

Measure	User network layer		Multilayer network	
	GMA	Leiden	GMA	Leiden
Communities	1125	822	1059	172
Modularity	0.881	0.880	0.589	0.619
Coverage	0.969	0.921	0.861	0.853
Performance	0.901	0.901	0.922	0.924

Table 10 Community Detection Results for the Joint Dataset and Alternative Network Models & Detection Algorithms

Measure	User network layer		Multilayer network	
	GMA	Leiden	GMA	Leiden
Communities	3192	1990	2608	585
Modularity	0.969	0.973	0.799	0.824
Coverage	0.983	0.870	0.945	0.920
Performance	0.978	0.983	0.959	0.961

Recall that this dataset has a relatively low coherence for topic identification.

As reported in Table 9, the Leiden algorithm identified fewer communities than GMA for each type of network, and notably less for the directed multilayer network; the undirected network representation to enable the Leiden algorithm artificially connected more components. Otherwise, the GMA and Leiden results for other metrics were comparable.

Both algorithms identified fewer communities when applied to the multilayer network. The topic layer helped identify connections between nodes that would otherwise not be detected. In the user network layer alone, there are 1045 (disconnected) components, whereas the multilayer network has only 991. Thus, the connectivity between users modeled via the topic layer helps identify larger communities. The modularity and partition coverage metrics are worse for both GMA and Leiden when applied to the 2016 US Election dataset, a result consistent with degraded influential user identification via the multilayer network. Only the partition performance is elevated for the multilayer network, by about 2.5%.

For the Joint dataset, Table 10 reports the number of communities, modularity, partition coverage, and partition performance for both the user network layer and the multilayer network, when applying the GMA and Leiden algorithms. Relative to the 2016 US Election dataset, the Joint dataset has a higher coherence for topic identification.

Within Table 10, the Leiden algorithm again identified fewer communities than GMA for both types of network models. Despite the addition of low weight edges to connect the components of the network, the Leiden algorithm

yielded higher modularity scores than GMA. The significance of this improvement as it relates to the required graph transformations would require further research to ascertain, and we propose that exploration as a sequel to this research. The Leiden algorithm also yielded slightly lower partition coverage and marginally higher partition performance for both network models. Compared to their performance on the 2016 US Election dataset, both GMA and Leiden performed better on most metrics, with notably higher modularity for this dataset having high topic coherence. This result reinforces the merit of the multilayer network for modeling and analyzing user interactions attained via broad search queries.

4.5 Impact of dataset size on multilayer network approach by query type

Common to results in Sects. 4.2 and 4.4, the efficacy of methods vary by the type of query. LDA topic separation was better for the Joint dataset, yielding a coherence of 0.5. In turn, these results allowed the directed multilayer network approach to identify influential users via PageRank and detect communities using either GMA or the Leiden algorithm. By comparison, the directed multilayer network approach was not well suited to analyze datasets attained via topic-specific queries. LDA encountered challenges attempting to differentiate topics within the 2016 US Election dataset because, e.g., Tweets from different political parties will use much of the same language. PageRank and other methods can identify influential users for datasets culled using topic-specific queries, but the performance is better when applied to a single, user-layer network.

Redundancy is an aspect of Twitter data that compels an examination of the appropriate query size for data queries. For example, despite the 2016 US Election dataset containing over 42,000 Tweets, it has only 15,000 unique Tweets. The majority of its communications are Retweets. Four of its Tweets and the ensuing Retweets and replies account for over 1,000 of the dataset's instances. Although one would expect some data redundancy in Twitter, its existence is potentially beneficial. Smaller sized datasets may suffice for SNA.

To examine the potential reduction in dataset size, testing examines the process through the third step: the identification of influential users. As a benchmark for expectations, analysis identified the top ten influential users for the entire 2016 US Election dataset and for 50,000 observations sampled from the Joint dataset using only the user network layer. For various sample sizes, 50 trials of bootstrap sampling (with replacement) and analysis of data from each dataset identified the top ten influential users. Table 11 reports for each of the sample sizes the average percentage of top influential users from the 50,000 observation samples found by the smaller samples.

Table 11 Average (%) of Top Influential Users from a 50,000 Tweet dataset found by 50 Samples Each of Smaller Datasets

Sample Size	2016 US election data (%)	Joint data (%)
250	61.6	38.2
500	65.0	40.6
1000	65.6	44.2
2500	67.6	46.0
5000	71.0	46.8

Observable in Table 11, smaller datasets produce similar results for specific queries, but more general queries that collect data from different conversations require more data to accurately identify influential users.

5 Conclusions and recommendations

This research proposed a four-step process for analyzing social networks to identify and target individuals and communities with brand and product-specific marketing. For such marketing, it is intuitively helpful to understand a target audience's interests, i.e., their topics of discussion. Within this context, this study set forth and tested a four-step process that leveraged a directed multilayer network approach for analysis. Augmenting traditional user network (layer) construction, the proposed process leverages Latent Dirichlet analysis (LDA) with extractive summarization to identify topics; constructs a directed multilayer network with a user layer, topic layer, and appropriate arcs to represent connections; identifies influential users (e.g., via PageRank); and detects the related communities of interest (e.g., via a Greedy Modularity Algorithm).

Testing these techniques for named datasets attained via specific queries and a more generally focused dataset sampled from the named datasets revealed several important findings. First, LDA better identified topics via the directed multilayer network approach when analyzing datasets attained via a broad query, enabling higher coherence scores and better topic separation. The proposed directed multilayer network approach was effective for identifying influential users and communities for such datasets. In contrast, the proposed, four-step process was not effective for datasets attained via specific queries. LDA had difficulty identifying distinct topics of conversation, so the inclusion of a topic layer in the network degraded the processes of influential user identification and community detection.

Testing also revealed several procedural insights. The proposed weighting schemes to quantify directed user-to-user, directed user-to-topic, and undirected inter-topic relationships in the multilayer network are conceptually sound and easy to implement. PageRank is the superlative technique among those tested to identify influential users, regardless of the dataset or modeling approach. Twitter verification status strongly relates to influential user identification. For broad-query datasets analyzed via the directed multilayer network approach, larger samples than would be required by a topic-specific dataset are necessary for procedural accuracy. Finally, both GMA and the Leiden algorithm are useful for community detection, regardless of dataset query type or network modeling approach.

An interested analyst or company can readily replicate and automate the proposed four-step process to gather information for marketing via social media. In doing so, it is important to use broad search queries and gather large datasets of Tweets. As a check on expected outcomes, analysis should proceed with the proposed directed multilayer network approach if the LDA topic coherence is approaching 0.5, at least.

The impact of this research would benefit from the following extensions. First, additional study should examine the thresholds for including user-to-topic and inter-topic relationships as arcs in the directed multilayer network. Second, it is relevant to examine more broad-query datasets to verify or refine the proposed threshold for LDA coherence. Third, the effect of required network transformations on the efficacy of the Leiden algorithm merits study, arguably using datasets with known community membership. Finally, the impact of Twitter's recent changes in user verification should be studied to determine if status has an effect on social influence.

As a caveat to the recommendations, it is important to note that relationships are not static, nor is user discourse. Although testing demonstrated the potential benefit of the proposed, four-step process for analyzing large, broad-query datasets, analysis supporting marketing must be an iterative process. Only by analyzing a market repeatedly, over time may one be aware not only of user interests but evolving user interests that allow a company to exercise marketing initiatives.

6 Disclaimer

The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, United States Army, United States Department of Defense, or United States Government.

Declarations

Conflict of interest The authors received no funds, grants, or other support for this research.

References

Ahuja RK, Magnanti TL, Orlin JB (1993) Network flows. Prentice-Hall Inc., Upper Saddle River, NJ

Aiello LM, Barrat A, Cattuto C, et al (2010) Link creation and profile alignment in the aNobii social network. In: 2010 IEEE Second International Conference on Social Computing, IEEE, pp 249–256

Allard K (1990) Command, control, and the common defense. Yale University Press, New Haven, CT

Bakshy E, Hofman JM, Mason WA, et al (2011) Everyone's an influencer: quantifying influence on Twitter. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. ACM, pp 65–74

Barrios F, López F, Argerich L, et al (2015) Variations of the similarity function of TextRank for automated summarization. In: 2015 Argentine Symposium on Artificial Intelligence, Sociedad Argentina de Informática e Investigación Operativa (SADIO), pp 65–72

Bhavnani V, Galphat Y, Bhawsinghka G, et al (2021) A survey on detecting influential user in social networking. In: 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), IEEE, pp 1–7

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3(Jan):993–1022

Bouma G (2009) Normalized (pointwise) mutual information in collocation extraction. In: 2009 Proceedings of the Biennial German Society for Computational Linguistics & Language Technology, vol 30. GSCL, pp 31–40

Clauset A, Newman ME, Moore C (2004) Finding community structure in very large networks. Phys Rev E 70(6):066–111

Dewi FK, Yudhoatmojo SB, Budi I (2017) Identification of opinion leader on rumor spreading in online social network twitter using edge weighting and centrality measure weighting. In: 2017 Twelfth International Conference on Digital Information Management (ICDIM), IEEE, pp 313–318

Doerr C, Blenn N, Van Mieghem P (2013) Lognormal infection times of online information spread. PloS ONE 8(5):e64-349

Erlandsson F, Bródka P, Borg A et al (2016) Finding influential users in social media using association rule learning. Entropy 18(5):164

Featherstone JD, Barnett GA (2020) Validating sentiment analysis on opinion mining using self-reported attitude scores. In: 2020 Seventh International Conference on Social Networks Analysis. Management and Security (SNAMS), IEEE, pp 1–4

Featherstone JD, Barnett GA, Ruiz JB et al (2020) Exploring childhood anti-vaccine and pro-vaccine communities on twitter-a perspective from influential users. Online Soc Netw Media 20(100):105

Featherstone JD, Ruiz JB, Barnett GA et al (2020) Exploring childhood vaccination themes and public opinions on twitter: A semantic network analysis. Telemat Inf 54(101):474

Freeman L (2004) The development of social network analysis. Stud Soc Sci 1(687):159–167

Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40:35–41

Gazdaggyori Z (2021) A case study of Gamestop. Bachelor's thesis, Aarhus University

Hagberg A, Swart P, S Chult D (2008) Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the 7th Python in Science Conference (SciPy2008). SciPy, Pasadena, CA, pp 11–15

Hamraoui I, Boubaker A (2022) Impact of Twitter sentiment on stock price returns. Soc Netw Anal Min 12(1):1–15

Ji J, Robbins M, Featherstone JD et al (2022) Comparison of public discussions of gene editing on social media between the united states and china. Plos one 17(5):e0267406

Jin X (2020) Exploring crisis communication and information dissemination on social media: social network analysis of Hurricane Irma tweets. J Int Crisis Risk Commun Res 3(2):179–210

Jiwanggi MA, Adriani M (2016) Topic summarization of microblog document in Bahasa Indonesia using the phrase reinforcement algorithm. Proc Comput Sci 81:229–236

Kalepalli Y, Tasneem S, Teja PDP, et al (2020) Effective comparison of LDA with LSA for topic modelling. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, pp 1245–1250

Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. J ACM 46(5):604–632

Kolda TG, Bader BW, Kenny JP (2005) Higher-order web link analysis using multilinear algebra. In: Fifth IEEE International Conference on Data Mining (ICDM'05), IEEE, pp 8–pp

Landau E (1895) Zur relativen wertbemessung der turnierresultate. Deutsches Wochenschach 11:366–369

Legradi J (2009) An exploratory social network analysis of military and civilian emergency operation centers focusing on organization structure. Master's thesis, Air Force Institute of Technology, Wright Patterson AFB, OH

Malliaros FD, Vazirgiannis M (2013) Clustering and community detection in directed networks: A survey. Phys Rep 533(4):95–142

Mihalcea R, Tarau P (2004) Textrank: Bringing order into text. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. ACL, pp 404–411

Mimno D, Wallach H, Talley E, et al (2011) Optimizing semantic coherence in topic models. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. ACL, pp 262–272

Moreno J (1933) Psychological and social organization of groups in the community. In: Proceedings & Addresses. American Association on Mental Deficiency

Moreno JL (1932) Application of the group method to classification. National committee on prisons and prison labor

Newman D, Lau JH, Grieser K, et al (2010) Automatic evaluation of topic coherence. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. ACL, pp 100–108

Newman ME (2016) Equivalence between modularity optimization and maximum likelihood methods for community detection. Phys Rev E 94(5):052–315

Pacheco D, Hui PM, Torres-Lugo C, et al (2021) Uncovering coordinated networks on social media: Methods and case studies. In: 2021 Proceedings of the AAAI International Conference on Web and Social Media (ICWSM). AAAI, pp 455–466

Page L, Brin S, Motwani R, et al (1999) The PageRank citation ranking: Bringing order to the web. Tech. Rep. SIDL-WP-1999-0120, Stanford University InfoLab, Stanford, CA

Perdana RS, Pinandito A (2018) Combining likes-retweet analysis and naive Bayes classifier within Twitter for sentiment analysis. Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 10(1-8):41–46

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959

Pudjajana AM, Manongga D, Iriani A, et al (2018) Identification of influencers in social media using social network analysis (SNA). In: 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), IEEE, pp 400–404

Rahmadan MC, Hidayanto AN, Ekasari DS et al (2020) Sentiment analysis and topic modelling using the LDA method related to the flood disaster in Jakarta on Twitter. In :2020 International Conference on Informatics. Multimedia, Cyber and Information System (ICIMCIS), IEEE, pp 126–130

Röder M, Both A, Hinneburg A (2015) Exploring the space of topic coherence measures. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. ACM, pp 399–408

Rudrapal D, Das A, Bhattacharya B (2018) A survey on automatic twitter event summarization. J Inf Process Syst 14(1):79–100

Ruiz J, Featherstone JD, Barnett GA (2021) Identifying vaccine hesitant communities on twitter and their geolocations: a network approach

Salehi A, Ozer M, Davulcu H (2018) Sentiment-driven community profiling and detection on social media. In: Proceedings of the 29th ACM Conference on Hypertext and Social Media. ACM, pp 229–237

Scott J, Carrington PJ (2011) The SAGE Handbook of Social Network Analysis. SAGE Publications, London, UK

Sheth A, Shalin VL, Kursuncu U (2022) Defining and detecting toxicity on social media: context and knowledge are key. Neurocomputing 490:312–318

Sievert C, Shirley K (2014) Ldavis: A method for visualizing and interpreting topics. In: Proceedings of Workshop on Interactive Language Learning, Visualization, and Interfaces, Association for Computational Linguistics, pp 63–70

Tang L, Liu H (2011) Leveraging social media networks for classification. Data Min Knowl Discov 23(3):447–478

Traag VA, Waltman L, Van Eck NJ (2019) From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 9(1):1–12

Tsopze N, Domgue FG (2021) Boolean factor based community extraction from directed networks with the non reciprocal link relationship. Inf Sci 569:544–556

Tsugawa S, Ohsaki H (2015) Negative messages spread rapidly and widely on social media. In: Proceedings of the 2015 ACM on Conference on Online Social Networks. ACM, pp 151–160

Venkatesan M, Prabhavathy P (2019) Graph based unsupervised learning methods for edge and node anomaly detection in social network. In: 2019 IEEE 1st International Conference on Energy. Systems and Information Processing (ICESIP), IEEE, pp 1–5

Yang Y, Hsu JH, Löfgren K et al (2021) Cross-platform comparison of framed topics in Twitter and Weibo: machine learning approaches to social media text mining. Soc Netw Anal Min 11(1):1–18

Zhang B, Vos M (2015) How and why some issues spread fast in social media. Online J Commun Media Technol 5(1):90–113