Multimodal Mining of Twitter Networks for Improved Label Propagation

Muhieddine Shebaro[1†], Lia Nogueira de Moura MSc[1†], Jelena Tešić PhD[1*]

[1*]Computer Science, Texas State University, San Marcos, 78666, TX, USA.

*Corresponding author(s). E-mail(s): jtesic@txstate.edu; Contributing authors: m.shebaro@txstate.edu; lia.lnm@gmail.com;
†These authors contributed equally to this work.

The rise of social networks has brought about a transformative impact on communication and the dissemination of information. However, this paradigm shift has also introduced many challenges in discerning valuable conversation threads amidst fake news, malicious accounts, background noise, and trolling. In this study, we address these challenges by focusing on propagating fake news labels. We evaluate the efficacy of community-based modeling in effectively addressing these challenges within the context of social network discussions using the state-of-art benchmark. Through a comprehensive analysis of millions of users engaged in discussions on a specific topic, we unveil compelling evidence demonstrating that community-based modeling techniques yield precision, recall, and accuracy levels that are comparable to those achieved by lexical classifiers. Remarkably, these promising results are achieved even without considering the textual content of tweets beyond the information conveyed by hashtags. Moreover, we explore the effectiveness of fusion techniques in tweet classification and underscore the superiority of a combined community and lexical approach, which consistently delivers the most robust outcomes and exhibits the highest performance measures. We illustrate this capability with specific network graphs constructed based on Twitter interactions related to the COVID-19 pandemic, showcasing the practicality and relevance of our proposed methodology.

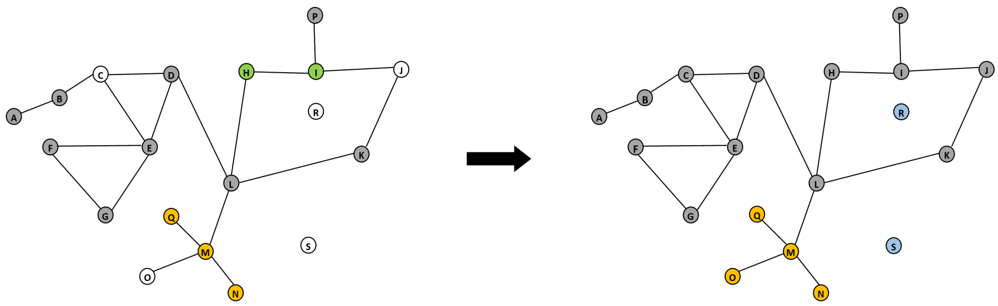

Fig. 1 Community attribute enrichment: analyze labeled data set in a network graph and extract community labels from the graph analysis of the network. Gray nodes are nodes with non-conspiracy content. Yellow nodes are promoting/discussing 5G conspiracy topics. White nodes are test nodes. Light blue are unknown nodes (indeterminant). Green nodes discuss other conspiracies.

1 Introduction

The emergence of social networks has bestowed considerable significance upon these platforms as primary sources of news consumption for a substantial portion of the population. The interconnected nature of online users within these networks facilitates the rapid dissemination of information, surpassing the conventional reach of traditional news media outlets such as newspapers and television. However, this inherent connectivity also amplifies the ease with which false and misleading information can propagate, particularly within the context of users' social network connections. This paper aims to delve into the investigation of whether the structural characteristics of social network user connections can serve as valuable assets in the detection and mitigation of fake news, specifically within the realm of Twitter.

Can we classify the Tweet without knowing the content tweet? In this paper, we explore the social network context, Twitter's rich network of interaction i.e., connections, tags, retweets, and mentions, and how they influence the labeling of the content. We test the observation that people in the same social network group or discussion thread tend to quote and discuss similar resources and have shared topic items, shed new light on the challenges posed by social network dynamics, and offer an effective means of tackling them through community-based modeling. By revealing the comparable performance of community-based approaches to traditional lexical classifiers, we contribute to advancing tweet classification methodologies. Our research opens up exciting avenues for further exploration and application, paving the way for more sophisticated network selection and fusion methods that leverage both community attributes and lexical modeling to enhance the accuracy and effectiveness of tweet classification in the ever-evolving landscape of social networks. Our findings carry substantial implications for understanding the dynamics of social networks and advancing methodologies for tweet classification. By harnessing the power of community attributes and models, our research uncovers the invaluable contextual information embedded within social network interactions involving tweet authors and objects. Furthermore, we present tangible evidence of our ability to capture comprehensive information by constructing network graphs that encapsulate key features such as retweets, mentions, replies, and quote networks.

To this end, we propose an enrichment of Tweet classification with a network-based analysis of the Twitter network, as illustrated in Figure 1. We relate the content of the Tweets using multimodal lexical analysis, employ community discovery by building a network of retweets, mentions, and hashtags, and employ network analysis on structural data mined from Twitter. To this end, we have developed a robust lexical-based analysis for Tweet content that takes into account colloquialisms, abbreviations, and OCR text in images if available. We have also developed a scalable data science package that downloads, saves, and analyzes Twitter data at scale, providing robust content analysis of noisy communities on Twitter [1–3]. We evaluate the approach in the MediaEval 2020 FakeNews task benchmark data set and COVID-19 (+) Twitter data set. We demonstrate the value of author's network in content classification on the MediaEval Fake News Detection Task 2020, which offers two Fake News Detection subtasks on COVID-19 and 5G conspiracy topics [4]. More specifically, they detect misinformation claims that the construction of the 5G network and the associated electromagnetic radiation trigger the SARS-CoV-2 virus. This benchmark challenge looked only at Tweet classification of COVID-19-related Tweets in two ways: (1) multi-class labeling: 5G-Corona_Conspiracy, Other_Conspiracy, and Non-Conspiracy, and (2) binary labeling: Unknown-or-Non-Conspiracy and Any-Conspiracy. In this paper, we show that tweet classification on author's network only (without analyzing tweet content) offers similar performance to tweet content classification.

Content: *Does #5G cause #COVID2019 #coronavirus? No, of course not! Does non-ionizing #wireless radiation accelerate viral replication and contribute to #AntibioticResistance? Yes.*
Ground Truth: **5g_corona_conspiracy**
Lexical model Prediction: **non_conspiracy**
Reply connection network majority prediction: **5g_corona_conspiracy**
of edges in labeled 5g_corona_conspiracy set: **11**
of edges in the other_conspiracy dataset: **0**
of edges in the non_conspiracy conspiracy dataset: **0**
% of tweets in the detected community that are from 5g_corona_conspiracy dataset: **100%**
% of tweets in the detected community that are from other_conspiracy dataset **0%**
% of tweets in the detected community that are from non_conspiracy dataset **0%**

Table 1 Tweet by a user with strong 5G Corona Conspiracy community ties. Community based detection identified the group and augmented the lexical classification.

2 Related Work

In this section, we review the related work on fake news detection on Twitter. The prevalence of "fake news" raises significant concerns. Recent research shows that fake news sharing is fueled by the same psychological motivations that drive other forms of partisan behavior, including sharing partisan news from traditional and credible news sources [5]. Given the widespread proliferation of misinformation online and the growing reliance on social media for news consumption, it is essential to comprehend how people evaluate and engage with posts of low credibility. This study examines users' responses to fake news posts on their Facebook or Twitter feeds, seemingly originating from accounts they follow. To explore this phenomenon, we conducted semi-structured interviews with 25 participants who regularly employ social media for news consumption. Employing a browser extension unbeknownst to the participants, we temporarily introduced fake news into their feeds and observed their subsequent interactions. Through this process, participants provided insights into their browsing experiences and decision-making processes. Our findings highlight various reasons individuals refrain from investigating posts of low credibility, including a tendency to accept content from trusted sources at face value and a reluctance to invest additional time in verification. Additionally, we document the investigative methods adopted by participants to ascertain the credibility of posts, incorporating both platform affordances and ad-hoc strategies. Drawing upon our empirical insights, we present design recommendations to support users in their efforts to evaluate the credibility of low-credibility posts [6]. Twitter data has been used to understand the influence of fake news during the 2016 US presidential election [7]; it has also been used to analyze the COVID-19 and the 5G Conspiracy Theory [8] and the COVID-19 Twitter narrative among U.S. governors and cabinet executives [9]. Using logistic regression to classify Tweets based on topic [10] shows that the content of the Tweet dominates in correct Tweet classification. Writing style and frequency of word usage emerged as relevant features in the lexical analysis [11]. Two major directions of leveraging community information are adapting deep learning techniques to learn the underlying characteristics of the Tweets in communities (e.g., [12]) or exploring the structural and sharing patterns of the topic (e.g., [13]).

2.1 Context Through Connections

Community-based modeling of social networks that leverages the spread of information in social media through retweets and comments has improved NLP-based modeling [11]. Structural and sharing patterns in the Twitterverse are rich, and the definition of communities on Twitter is multi-dimensional. Users in the community can share geographic proximity and interconnections with mutual friends, groups, and topics of interest. Careful mapping of psychological profiles of over 2,300 American Twitter users linked to behavioral sharing data and sentiment analyses of more than 500,000 news story headlines finds that the individuals who report hating their political opponents are the most likely to share political fake news and selectively share content that is useful for derogating these opponents [5]. Factual News Graph (FANG) was proposed as a graphical social context representation and learning framework for fake news detection that focuses on representation learning. FANG has captured. social context to a degree if the topic is well represented, and has generalized to related tasks, such as predicting the factuality of reporting of a news medium [14]. Similar unsupervised graph embedding methods on the graphs from the Twitter users' social network connections are used to find that the users engaged with fake news are more tightly clustered together than users only engaged in factual news [15]. Graph-based approaches focus on biclique identification, graph-based feature vector learning, and label spreading on Twitter [16], but they do not scale well to the number and heterogeneity of the topics examined. Schroeder and al. developed a framework for capturing and analysing huge amounts of Twitter data. It consists of the main data capturing component (Twitter API), the proxy, the storage, and experiment wrappers which are connected to storage and to the proxy. The proxy provides quota leasing, an external API to allow users to execute calls having the same syntax, and request caching. The storage supports diverse types of databases and file storages. And experiment wrappers constitute a setup for analytical tasks as well as collecting data. Example experiments include follower analysis of an account for fake followers detection and network analysis of a user for determining the position of an account with respect to its surrounding network [17].

2.2 Lingustics Aspects

Beyond the utilization of lexical and community features, other novel avenues of harnessing a tweet have been explored in various tasks. The prediction of linguistic aspects of tweets with the #MeToo hashtag, a movement that has recently emerged against sexual assault and advocating women's rights, has been performed by capitalizing on both textual and visual modalities but contextual embeddings and transformer language models weren't employed due to the sole reason of them being computationally expensive [18]. Many similar works dealing with these same type of modalities, however, has put the unadapted version of BERT and a generic Deep Neural Network (DNN) to use for feature extraction significantly boosting generalization performance such as [19] for ultimately developing a profiling system to identify anonymous and potentially nefarious users' genders and [20] for finding disaster tweets. The concept of multi-modal tweet fusion is even introduced in the context of geosciences [21] where the authors proposed to incorporate contextual hydrological information to effectively classify flood-related tweets. This proposal has yielded promising success along several metrics and sheds light on the importance of not restricting models, regardless of their type, in feeding on lexical data and not neglecting other discriminative information. Another pivotal and salient modality is observed in the form of location features of geo-tagged (longitude and latitude) tweets for sentiment analysis [22]. The tweets' word embeddings were obtained and merged with the vectorized location features to create a set of hybrid representations. These representations led to an enhanced accuracy in classifying sentiment compared to the baseline GloVe model using a CNN and a bi-directional LSTM.

Graph Neural networks perform well in multi-modal contexts. For instance, Gao et al. presented MM-GNN, a novel framework aimed at addressing inquiries by providing information from images. MM-GNN incorporates visual, semantic, and numeric modalities to represent an image as a graph. Next, the node features are refined by leveraging contextual information from these modalities (using message passing) which contributes to improved performance in question answering tasks [23].

There are also a multitude of state-of-the-art GNN variants that has then been developed to resolve current issues of vanilla baseline GNNs. For instance, SelfSAGCN was created to alleviate over-smoothing and when labeled data are severely scarce using "Identity Aggregation" and "Semantic Alignment" techniques [24]. Due to the limited memory resources when loading the entire attributed graph into network for processing in current GNNs, Bi-GCN was designed in which it binarizes both the network parameters and input node features and produces comparable results as baseline vanilla models such as GraphSage and GCN [25]. In addition, sparsely and noisily labeled graphs have been dealt with via the novel NRGNN variant [26]. Another notable GNN framework, Tail-GNN, is based on the concept of neighborhood translation in the structurally rich head nodes to be transferred to the structurally limited tail nodes to enhance their representations and uncover missing neighborhood nodes [27].

Unsupervised graph clustering approaches have considered merely structural information, and in recent years attributed graph clustering has gained strong attention: it integrates additional attribute data about vertices into the clustering task to enhance its result. State-of-the-art graph neural networks suffer from training data bias and vertex feature dependency [28].

Social media platforms have become a vital source of information during the outbreak of the pandemic (COVID-19). The phenomena of fake information or news spread through social media have become increasingly prevalent and a powerful tool for information proliferation. Detecting fake news is crucial for the betterment of society. Existing fake news detection models focus on increasing the performance which leads to overfitting and lag generalizability. Hence, these models require training for various datasets of the same domain with significant variations in the distribution. In our work, we have addressed this overfitting issue by designing a robust distribution generalization of transformers-based generative adversarial network (RDGT-GAN) architecture, which can generalize the model for COVID-19 fake news datasets with different distributions without retraining. Based on our experimental findings, it is evident that the proposed model outperforms the current state-of-the-art (SOTA) models in terms of performance

Social media provides a rapid, simple, and accessible platform for people to communicate and share news through the Internet. However, the information published on this platform is not always trustworthy. As a result, malicious actors often use social media to disseminate fake news or mislead news readers, such as with personal or political attacks that could spark protests or riots. In this paper, we propose a learning technique for detecting fake news sources (i.e., fake users) on the Twitter platform. Three main types of features—tweet content, published time, and social graph—have been defined and extracted from Twitter to create a deep neural network (DNN) as a predictive model. We conducted experiments on PolitiFact, a standard FakeNewsNet dataset. The results show that the proposed approach outperforms traditional baselines with 98.7% accuracy [29].

3 Methodology

In this paper, we use a scalable approach to gather, discover, analyze, and summarize joint sentiments of Twitter communities, extract community and network features,

Content: *Explaining why beneficial effects from cannabis on intestine inflammation conditions like ulcerative colitis and Crohn's disease have been reported often. If the endocannabinoid isn't present, inflammation isn't kept in balance; the body's immune cells attack the intestinal lining.*
Ground Truth: **non_conspiracy**
Lexical model Prediction: **5g_corona_conspiracy**
All connections network majority prediction: **non_conspiracy**
of connections in the 5g_corona_conspiracy dataset: **0**
of connections in the other_conspiracy dataset: **129**
of connections in the non_conspiracy conspiracy dataset: **185**
% of tweets in the community that are from 5g_corona_conspiracy dataset: **10%**
% of tweets in the community that are from other_conspiracy dataset: **25%**
% of tweets in the community that are from non_conspiracy dataset: **65%**

Table 2 Tweet content has all the words, and lexical approach misclassified it. Community approach provided enough attributes for the fusion run to identify it correctly.

and improve the lexical-based baseline for Tweet classification using community information [2].

3.1 Content Analysis, Transformation, and Feature Selection

Twitter restricts Tweet content to 280 characters, a limit that tends to produce a writing style that differs from most corpora. To achieve brevity, users employ a lexicon that includes abbreviations, colloquialisms, *hashtags*, and *emoticons*, and Tweets may contain frequent misspellings. The context of a Tweet is also richer, as it resides in a rich network of retweets and replies. To this end, we employ lexical-based analysis and community analysis for Tweet content and context. The **Lexical Analysis Pipeline** implements the transformation of Twitter content, feature extraction, and modeling to make predictions for the NLP-based task [30].

In the *transformation* step, we tested several pre-processing, tokenization, and normalization techniques. We measured the influence of each transformation approach to predict performance on the part of the development set, turning off the feature and comparing the performance using 5-fold measures. Removing punctuation, preserving URLs, and normalizing several specific terms (e.g., 'U.K.' to 'UK') in the Tweet contributed to better content classification, as expected for the short tweet content. Stemming did not influence the classification recall on this small development set, and neither did lemmatization. We speculate that the Tweet content was too short and the data was too small to derive any meaningful conclusion, and therefore we did not apply either.

Feature extraction from Tweet content was implemented two ways: encoding terms as vectors representing either the occurrence of terms in text (*Bag-Of-Words*) or the impact of terms on a document in a corpus (TF-IDF). We attempted to use the TF-IDF vectorizer in order to capture the importance of terms, but we found better results using a *Bag-Of-Words* model, perhaps due to the high occurrence and variety of colloquialisms and abbreviations. We extended the feature set in the Tweets using *Optical Character Recognition* (OCR) of embedded images.

We tested the lexical *classification* pipeline incorporating a variety of classifiers: Naive Bayes, Support Vector Machine, Random Forest, Multilayer Perceptron,

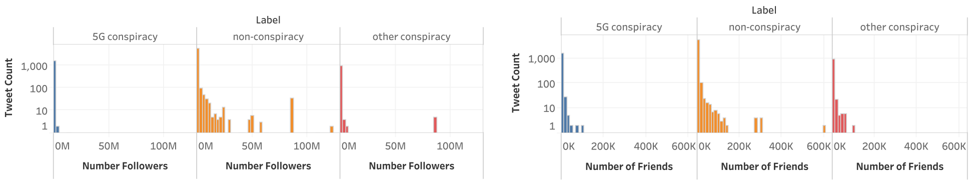

Fig. 2 Distribution of the feature *user_followers_count* (left) and *user_friends_count* (right) for the different class labels (5G, non, other).

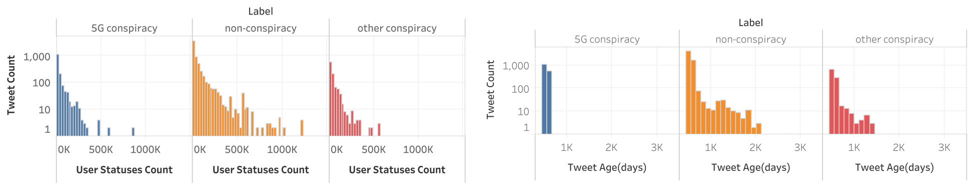

Fig. 3 Distribution of the feature *user_statuses_count* (left) and *tweet_age* (right) for the different class labels (5G, non, other).

Stochastic Gradient Descent, and a Logistic Regression classifier. We compared the performance of the classifiers on validation sets, both for the multi-class and binary classification subtasks. The Logistic Regression classifier showed the best results in [30]. In this paper, we use Logistic Regression as it has been shown to perform the best for the content-based classification in [30]. To account for the imbalance in data (see Table 4 for details), we experimented with data augmentation. Generating fake Tweets using the most predictive or most common terms for each class led to overfitting of most classifiers. We took a different route and have adjusted class weights to account for imbalanced data when possible.

3.2 Rich Graph Network Analysis

We apply the **Community Analysis Pipeline** for community discovery in networks created from user and hashtag connections to construct seven different networks from the raw Twitter data: *All Users Connections*, a network created from the labeled data set, with each vertex in the network being a user, and each edge of the network being the connection between two users by either a retweet, quote, reply, mention, or friendship; *Retweet Connections*, which is similar to *All Users Connections*, but with each edge being the connection between two users by retweets only; *Mention Connections* which is similar to *All Users Connections*, but with each edge being the connection between two users by mentions only; *Reply Connections*, which is similar to *All Users Connections*, but with each edge being the connection between two users by replies only; *Quote Connections*, which is similar to *All Users Connections*, but with each edge being the connection between two users by quotes only; *Friends Connections*, which is similar to *All Users Connections*, but with each edge being the connection between two users by friendship only; and *Hashtag Connections*, a network created from the labeled data set with each vertex in the network being a hashtag and each edge of the network being the connection between two hashtags when they were used

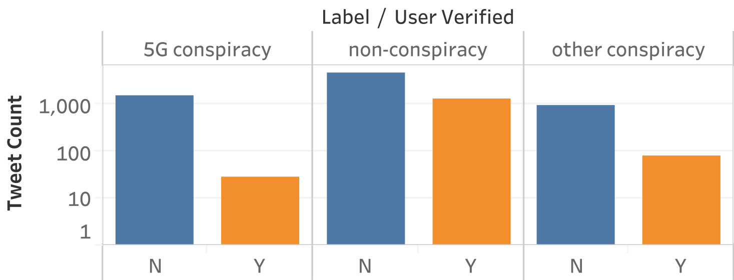

Fig. 4 Distribution of the feature *user_verified* for the different class labels (5G, non, other)

together in the same Tweet. We have developed an in-house scalable package *pytwanalysis* [1–3] to collect and save information-rich Twitter data, create networks, and discover communities in the data.

3.2.1 Community Labeling

We utilized all networks to learn the attributes of the user and Tweet that were relevant to the community and topic. Using an adapted Louvain [31, 32] method, we found communities and labeled each community with one of the three conspiracy categories (5G, non, other) based on the majority of the Tweets for that community. If we found a community with more Tweets with the *5G* label as opposed to *non* or *other*, we assigned the *5G* label to unlabeled Tweets in that community. Figure 1 demonstrates a simplification of this method. We applied the method to all seven networks for community discovery and assigned seven community labels (from seven networks) to each Tweet, listed as features 1 through 7 on Table 3. For the *Hashtag Connections* network, because one Tweet can have multiple hashtags, then one Tweet could belong to multiple hashtag communities. In that case, the majority logic selects the most common community found for that Tweet. The remaining Tweets that did not belong to any community or that belonged to a community with Tweets strictly originating from the test data set were assigned as *Unknown*. Many *Unknowns* were found because a large number of Tweets did not have any connections with other users in the labeled data sets (i.e., no retweets, replies, quotes, mentions, friends, or hashtags). An additional combined label was created with a combination of the other seven labels, listed as feature 8 on Table 3. The combined label first uses the label from the quote network; if the quote network has an unknown value, it uses the value from the reply network, followed by the mention, then all user connections, then retweets, then friends, and then hashtag networks. The order of use for each network in the combined label was decided based on the evaluation metrics for the predictions coming from each network (Table 6). The community discovery approach can be useful for data sets in which users are well-connected to each other.

User connectivity was also extracted from the graphs created from the development data sets. *User connectivity* is a feature that shows the degree of connectivity between each user in the *All Users Connections* network for each of the provided classification labels, driven by the observation that if vertices are well-connected, their content is similar. See features 9 through 12 on Table 3.

#	Community Feature
1	lv_comty_usr_all(majory_label)
2	lv_comty_usr_rt(majory_label)
3	lv_comty_usr_mention(majory_label)
4	lv_comty_usr_reply(majory_label)
5	lv_comty_usr_quote(majory_label)
6	lv_comty_usr_friend(majory_label)
7	lv_comty_usr_ht(majory_label)
8	lv_comty(majory_label)_combined
9	usr_degree_in_5g_corona_conspiracy
10	usr_degree_in_non_conspiracy
11	usr_degree_in_other_conspiracy
12	usr_degree_combined

Table 3 Community attributes as explained in 3.2.1

3.2.2 Attribute Labeling

User Attributes in the Tweets are also extracted from the Twitter data. The produced networks can contain a number of disconnected Tweets, so we expand the suite of network features and extract four additional user attributes and one Tweet attribute as follows: 1. *user_followers_count* (Fig. 2 (left)); 2. *user_friends_count* (Fig. 2 (right)); 3. *user_statuses_count* (Fig. 3 (left)); 4. *user_verified* (Fig. 4); 5. *tweet_age (days since creation)* (Fig. 3 (right)). Since the community majority selection predictions generated a large number of unknown assignments, we used an additional classifier to help in predicting labels for Tweets that were disconnected from the network. Since we have different types of features, we used the versatile Random Forest classifier that can work well with a mixture of categorical and numerical features. Community features 1 through 12 from Table 3 and user features 1. to 5. listed above are used as input to the Random Forest classifier. The distribution of data for the features in the labeled data is shown in Figures 2 (left), 2 (right), 3 (left), 3 (right), and 4.

Community features 8 through 20 from Table 3 and user features aforementioned from 1 through 5 are used as input to the multilabel (5G, non, other) Random Forest classifier. Because of the amount of unknown predictions from the community assignments, this additional classifier helps in predicting labels for Tweets that were disconnected from the network. Since we have different types of features, we used the versatile Random Forest classifier that can work well with a mixture of categorical and numerical features.

First, we create three different networks from the raw data: *User Connections* from provided data: vertex is a user, and each edge is the connection between two users by either a retweet, quote, reply, or mention; *Hashtag Connections* from provided data: vertex in the network is a hashtag, and edge exists between two hashtags if they were used together in the same tweet; and *User Connections 8M*: a network created from provided data and the auxiliary dataset of over 8M tweets, where vertices and edges of the network created the same way as the *User Connections* network. Next, we extract the degree of connectivity for each of the provided conspiracy labels (5G, non, and other) driven by observation that if vertices are well connected their content is similar. We employ the *Louvain Community* discovery method to discover communities in all three networks, and apply to specific tweets information from each network analyzed [3]. We labeled each community with one of the three conspiracy categories (5G, non, other), based on majority of the labels for that community associated with the tweet label. If we found a community where 5G labels are larger than non or other, we will use 5G label to assign the label to unlabeled tweets in that community. These assignments were done based on the combination of communities found in all three networks. Tweets that did not belong to any community, or belonged to a community with tweets strictly originating from the test dataset, were assigned based on their degree of connectivity, and the remaining were assigned as *Unknown*. Many unknowns were found because a large number of tweets did not have any connections with other users in the given datasets (no retweets, replies, quotes, mentions, or hashtags).

3.3 Multi-Modal Tweet Overlap Analysis

In this section, we aim to explore and determine whether the communities derived from different modalities exhibit low overlap, signifying complementary information, or if there is a significant overlap, suggesting redundancy or similar underlying structures. By quantifying this measure, one may identify the modalities that contribute the most unique information and design fusion methods accordingly. For example, it can allow researchers to determine which modalities should be given more weight to achieve the best performance in classification tasks.

3.3.1 Network Construction

A network has been constructed from the COVID-19 (+) data set, which consists of 8 million tweets, after undergoing multiple preprocessing steps. First, replies, quotes, and retweets are the selected connection modes of the network. Unlike in the case of quotes and retweets, we have found that there are no elaborate information present (full_text, media_url...etc.) replied by tweets in COVID-19 (+). Hence, we removed any edges, constructed in the replies connection mode, where the target node is not found within the 8 million tweets due to the inability to extract textual and visual features from it. To reduce sparsity in the network, every target node should be connected to at least 10 nodes. Otherwise, the isolated nodes or the nodes' connections falling under this threshold are pruned. Moreover, isolated nodes and duplicate edges were eliminated and the first occurrence of any duplicate was kept. As a result, the total number of nodes and edges dropped to 3,407,903 and 3,316,523 respectively. For simplicity, every node id, designated by the its tweet ID, was mapped to values ranging from 0 to 3,407,902. Concerning the labeled MuMiN tweets, the connection mode of the network is Replies and Quotes.

3.3.2 Visual and Textual Feature Extraction

We find that 154,923 tweets had images in COVID-19 (+). Some of the tweets were suspended impeding some of the retrieval of the images. We also assigned the name of each image to its corresponding tweet ID preserving the link between the tweet and the image. *VGG16* model pretrained on *ImageNet* was employed as a feature extractor for all the images. On the other hand, textual embeddings were produced by a pretrained adapted version of BERT for COVID Tweets called *BERTweet* by VinAIResearch [33]. We utilized the baseline normalizations as elaborated below in subsection 3.1 but with a few alterations that includes removing usernames, all special characters, hashtags, contractions, non-English Tweets if present, links (which not only incorporates "https://t.co/", but also "http" and "www"), and emojis. These additional textual normalizations were applied and BERTweet features were subsequently extracted. On the other hand, the existence of built-in embeddings for both modalities in MuMiN allows us to skip this feature extraction phase.

3.3.3 Augmented Network Construction

We seek to obtain an infused network that is comprised of the aforementioned network as well as a visual similarity graph. The latter is built by computing the cosine similarity between each node's image DNN features in the preprocessed network. The edges are hence, formed between each node and its 5 most visually similar nodes. The number of edges bumped up to 4,091,138 in our processed COVID (+) network. The motive behind this is that the GNN is going to aggregate features from the neighboring nodes of not only those from replies, quotes, retweets (replies and quotes in MuMiN), but also from the nodes that has an image that's visually like it.

3.3.4 Graph Neural Network Training

In an attempt to leverage all modalities and aggregate features from neighborhood nodes, both the adjacency and the feature matrices are fed to an unsupervised GNN framework. The selected model for training the GNN is GraphSage [34] which produces an embedding output of size 50 dimensions. The hyperparameters are epoch = 1 , the batch size = 50, layer size = 50, and the learning rate = 0.001 with Adam as an optimizer. The choice of this variant of GNN is ascribed to the fact that GraphSage utilizes the neighborhood sampling concept which is renders it to be scalable. Moreover, GraphSage has been trained separately on both the constructed network as well as the visually infused network with the same textual feature matrix representing the nodes' features.

3.3.5 Clustering

Both networks have been clustered using the Louvain Algorithm [35]. However, the rest has been clustered using HDBSCAN (Hierarchical DBSCAN) [36]. It is faster than regular DBSCAN. And the hyperparameters used are default except that the minimum cluster size has been set to 10. Due to the memory constraints associated with clustering high dimensional textual embeddings and large data, the number of dimensions of the text has been reduced to 10 using the PCA method. However, the dimensions are intact when generating GNN embeddings.

4 Experimental Setup

4.1 Data Sets

Dataset	Tweet Count	User Count
1. Fake News [4]	**8,854**	**7,475**
Development Labels	**Tweet Count**	**User Count**
5g_corona_conspiracy	1,120	1,053
other_conspiracy	688	638
non_conspiracy	4,138	3,643
Total	**5,946**	**5,197**
Test Labels	**Tweet Count**	**User Count**
5g_corona_conspiracy	532	512
other_conspiracy	346	334
non_conspiracy	2,030	1,832
Total	**2,908**	**2,639**
2. Friends of Fake News [4]		**3,385,981**
3. COVID-19 (+) [3]	**771,203**	**657,785**

Table 4 MediaEval 2020, COVID-19 (+), and friendship data sets. For MediaEval 2020, note that the number of users in each set does not add up to the total number of users, as the same user can have tweets in different data sets.

The MediaEval Fake News Detection Task 2020 looks into Tweets for misinformation claims that the construction of the 5G network and the associated electromagnetic radiation trigger the SARS-CoV-2 virus. We have received a labeled data set of approximately 6,000 Tweets related to COVID-19, 5G, and their corresponding metadata; see details in Table 4). Note that all of our training was done using the development set which contains 1,120 Tweets labeled for 5G-COVID conspiracy, 688 Tweets for other conspiracy, and 4,138 for non-conspiracy Tweets, as shown in Table 4. This data set is small in nature and very imbalanced. Thus, we extended the labeled data set with a new COVID-19 (+) data set that contains Tweets related to #Coronavirus, #Covid19, and #Covid-19, collected from March through September 2020, with over 3.2 million users and 8 million Tweets [3]. From the 8 million Tweets, we filtered only the Tweets that can make a connection in the existing networks created from the labeled data. After applying the filter, we ended with a total of 771,203 COVID-19 Tweets. The COVID-19 (+) data set was used to augment the feature space for classification. We also extended knowledge about the relationships between users by using the Twitter API to retrieve a list of friends for each user in the labeled data set. A total of 3,385,981 users were retrieved, but that number does not include 100% of the users in the friendship list, as some of the previously existing users are not accessible anymore (e.g., account is suspended).

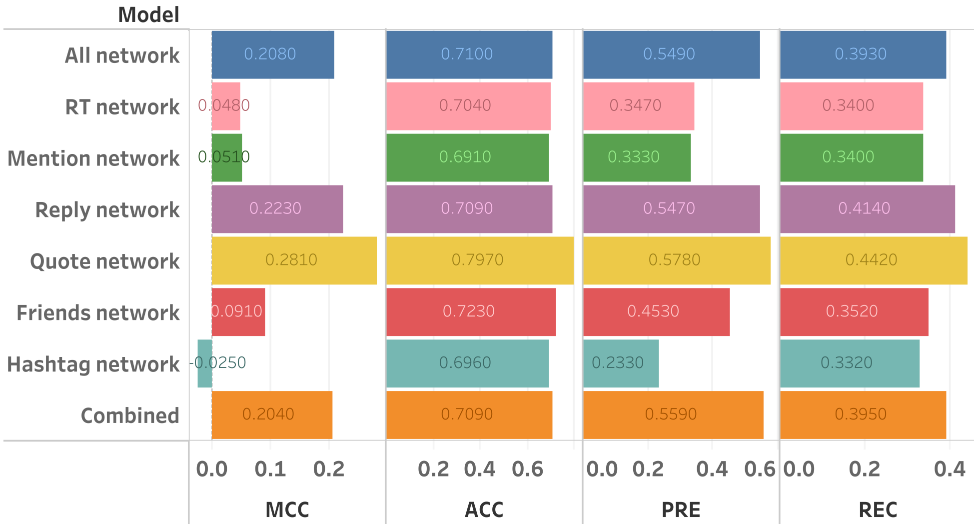

Fig. 5 Comparison of the multi-class community majority assignment excluding the unknowns for the different types of networks, as detailed in section *Multi-class without Unknowns* in Table 6

4.2 Measures

We measured the performance of the proposed methods on a very small labeled subset of test data in Table 4. MediaEval officially reported that the metric used for evaluating the multi-class classification performance was the multi-class generalization of the Matthews correlation coefficient (MCC) [4, 37, 38]. MCC has shown to have advantages in bioinformatics over F1 and accuracy, as it takes into account the balance ratios of the four confusion matrix categories (true positives, true negatives, false positives, and false negatives). In a social network analysis, we are more interested in missed Tweets (false negatives) and true positives. For this reason, we discuss our results from the perspective of precision, recall, and accuracy.

5 Results and Analysis

5.1 Lexical Analysis Pipeline

Figure 7 shows the metrics for the multi-class and binary predictions using the Logistic Regression classifier [30]. The baseline results for the lexical analysis pipeline used in this paper improves upon Data Lab's best multi-class logistical regression (LR) model MediaEval 2020 submission [30] using cross-validation and regularization. The new best MCC result for the LR used in this paper is **0.435** for multi-class and **0.492** for binary classification.

Multi-class				
Model	MCC	ACC	PRE	REC
LR	0.435	0.749	0.597	0.569
LR-OCR	0.379	0.706	0.459	0.384
Binary				
Model	MCC	ACC	PRE	REC
LR	0.492	0.789	0.749	0.743
LR-OCR	0.492	0.789	0.749	0.742

Table 6 Multi-class and binary labeling scores (MCC, Accuracy, Precision, Recall) for MediaEval 2020 Test Set. Model abbreviations: LR for logistic regression; LR-OCR for logistic regression with OCR.

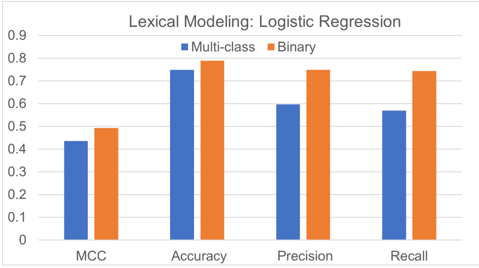

Fig. 7 A histogram visualizing the evaluation performance of the LR model in Table 5.1.

Evaluation Set		Test	Development			
Ternary	Model	MCC	MCC	Prec	Recall	Acc
001	LR	**0.431**	0.431	0.624	0.510	0.766
002	LR-OCR	0.363	0.465	0.599	0.565	0.767
003	CL	0.081	0.170	0.388	0.229	0.281
004	LR-CL	*0.363*	0.442	0.462	0.430	0.725
Binary	Model	MCC	MCC	Prec	Recall	Acc
011	LR	**0.437**	0.487	0.770	0.720	0.856
012	LR-OCR	0.428	0.516	0.780	0.737	0.862
013	CL	0.091	0.219	0.604	0.615	0.748
014	LR-CL	0.091	0.244	0.613	0.631	0.743

Table 5 Ternary (runs 001 - 004) and binary (runs 011 - 014) labeling scores returned by benchmark engine (MCC), and our analysis on development set released ground-truth (MCC, Precision, Recall, Acc). Model abbreviations: LR for logistic regression; LR-OCR for logistic regression w OCR; CL for community labeling; LR-CL for fusion run. Team has places second in the competition.

5.2 Community Analysis Pipeline

Table 6 shows the metrics for the multi-class and binary predictions using the Louvain community majority assignment for each type of network with and without the COVID-19 (+) data set. Results are intuitive, as community majority assignments using the combined connections network with the COVID-19 (+) data set perform the best over the range of measures. The table also shows the number of Tweets that were classified as unknown when they did not belong to any community. The additional results for the Random Forest classifier are included in the table for comparison. Note that the total for each model is always 2,908, which is the number of labeled Tweets in the test set.

The *Community Contribution Analysis* MediaEval 2020 development set is small, and it only captures fragments of the community. The number of unknown community assignments is large and skews the use of community attributes, as shown by the low performance in section *Multi-class with Unknowns* in Table 6. We separate the

		Lexical Model	Community Network							Random Forest
			All	Retweet	Mention	Reply	Quote	Friends	Hashtag	
	Lexical Model	100%	70%	33%	46%	20%	17%	65%	22%	72%
Community Network	All	70%	100%	41%	57%	27%	22%	**82%**	28%	**85%**
	Retweet	33%	41%	100%	**80%**	68%	69%	41%	56%	37%
	Mention	46%	57%	**80%**	100%	62%	54%	54%	49%	52%
	Reply	20%	27%	68%	62%	100%	**81%**	28%	61%	22%
	Quote	17%	22%	69%	54%	**81%**	100%	27%	67%	19%
	Friends	65%	**82%**	41%	54%	28%	27%	100%	34%	77%
	Hashtag	22%	28%	56%	49%	61%	67%	34%	100%	25%
Random Forest		72%	**85%**	37%	52%	22%	19%	77%	25%	100%

Fig. 8 Overlap in the community multi-class predictions by method: the percentage shows the overlap between the predictions of two methods out of the 2908 test records.

evaluation in the multi-class community majority assignment into evaluation including the unknowns and evaluation excluding the unknowns. The metrics without the unknowns were calculated separately so that we could evaluate how well we can classify the Tweets that did belong to a community, as shown in section *Multi-class without Unknowns* in Table 6 and in Figure 5. Results calculated without the unknowns show comparative performance with lexical pipeline.

The results in Table 6 show that the performance of community modeling is **comparable** to the lexical model if unknown assignments are excluded and the quality of the predictions in different types of networks are broken down. Networks created from *quotes* and *replies* seem to yield the best results. Our initial premise is that similar topics and news are shared with the people that quote each other or participate in the same thread of a discussion, so this finding confirms the value of that correlation. The predictions from the hashtag network, on the other hand, do not provide great results, as many of the same hashtags are used in both conspiracy and non-conspiracy labeled data.

Labeling Considerations: The main challenge of the community approach is scale; the annotations and the topic should be prevalent in the data set to truly benefit from the community based analysis. The COVID-19 (+) data set was obtained by finding an **intersection** of our originally mined data set of 8 million Tweets; see Section 4.1. Community based analysis with the auxiliary data brought the value of community connections to this analysis; compare model and model+ in Table 6. The COVID-19 (+) data set improved the connectivity in the network, which consequently improved the number of Tweets that were able to be classified. The number of unknowns from the all connection network (All) decreased from 198 (All) to 108 (All+) when an analysis of the same labeled data was done within the larger network, and the MCC score jumped from 0.089 to 0.180. The use of the Random Forest classifier over community and attribute labels improves the overall performance of the classification; see Table 6. The classifier can assign values for Tweets that could not be classified with the community majority assignments, since it uses additional features apart from the community features; see Section 3.2.2.

Table 7 summarizes the correct classification results that the network modeling produces that lexical one does not. The community predictions perform comparably

			Multi (Unknowns)				Multi (No Unknowns)				Binary predictions			
Community Predictions - Majority Selection														
Description	**Total**	**Unknowns**	**MCC**	**ACC**	**PRE**	**REC**	**MCC**	**ACC**	**PRE**	**REC**	**MCC**	**ACC**	**PRE**	**REC**
All network	2908	198	0.089	0.664	0.425	0.249	0.101	0.713	0.566	0.352	**0.276**	**0.733**	**0.694**	**0.598**
RT network	2908	2908									0.000	0.698	0.349	0.500
Mention network	2908	2095	0.027	0.192	0.386	0.084	0.204	0.686	0.514	0.403	0.123	0.703	0.632	0.529
Reply network	2908	2474	0.036	0.098	0.361	0.051	0.234	0.654	0.481	0.448	0.137	0.706	0.644	0.533
Quotes network	2908	2659	0.064	0.067	**0.457**	0.035	**0.461**	**0.783**	**0.609**	**0.597**	0.110	0.704	0.663	0.518
Friends network	2908	390	0.091	0.627	0.405	0.232	0.074	0.724	0.540	0.346	0.231	0.722	0.680	0.574
Hashtag network	2908	2158	-0.002	0.174	0.326	0.065	0.070	0.675	0.434	0.345	0.058	0.699	0.636	0.506
Combined	2908	154	**0.142**	**0.675**	0.391	**0.270**	0.161	0.713	0.522	0.377				
Community Predictions - Majority Selection - COVID-19 (+) Dataset														
Description	**Total**	**Unknowns**	**MCC**	**ACC**	**PRE**	**REC**	**MCC**	**ACC**	**PRE**	**REC**	**MCC**	**ACC**	**PRE**	**REC**
All network +	2908	108	0.180	0.683	0.412	0.283	0.208	0.710	0.549	0.393	**0.345**	**0.743**	0.692	**0.655**
RT network +	2908	1636	0.012	0.308	0.261	0.112	0.048	0.704	0.347	0.340	0.231	0.724	**0.700**	0.567
Mention network +	2908	1107	0.006	0.428	0.250	0.157	0.051	0.691	0.333	0.340	0.209	0.716	0.661	0.568
Reply network+	2908	2107	0.040	0.195	0.410	0.085	0.223	0.709	0.547	0.414	0.134	0.704	0.632	0.534
Quote network +	2908	2296	0.075	0.168	0.433	0.070	**0.281**	**0.797**	**0.578**	**0.442**	0.138	0.707	0.668	0.528
Friends network +	2908	392	0.101	0.625	0.340	0.235	0.091	0.723	0.453	0.352	0.243	0.725	0.682	0.581
Hashtag network +	2908	2076	-0.001	0.199	0.174	0.071	-0.025	0.696	0.233	0.332	-0.017	0.697	0.349	0.500
Combined +	2908	80	**0.180**	**0.689**	**0.419**	**0.288**	0.204	0.709	0.559	0.395				
ML Classifier														
Description	**Total**	**Unknowns**	**MCC**	**ACC**	**PRE**	**REC**	**MCC**	**ACC**	**PRE**	**REC**	**MCC**	**ACC**	**PRE**	**REC**
Random Forest	2908	0	0.256	0.711	0.526	0.435					0.368	0.751	0.704	0.666

Table 6 Predictions for the community labeling using MediaEval development data and Auxiliary COVID-19 (+) data set. Performance measures (MCC, Precision, Recall,Accuracy) were computed for every type of network for multi-class classification including the unknown predictions, for multi-class classification excluding the unknown predictions, and for binary classification.

for cases in which the Tweet was not isolated from the network. Figure 8 illustrates the overall multi-class detection overlap by method. The highest overlap occurs between the *all connections* network predictions and the Random Forest model, which is expected since the network predictions were used as features for the Random Forest model. The lexical model has the highest overlap with the *all connections* network predictions and Random Forest. Other methods that have high overlap in their predictions are the *all connections* network with the *friends* network, the *retweet* network with the *mention* network, and the *quote* network with the *reply* network.

5.3 Combining Community and Lexical Attributes

Multi-class				
Model	MCC	ACC	PRE	REC
Lexical-(LogisticRegression)	0.435	0.749	0.597	0.569
Community-(RandomForest)	0.256	0.711	0.526	0.435
Community + Lexical	**0.442**	**0.751**	**0.601**	**0.575**
Binary				
Model	MCC	ACC	PRE	REC
Lexical-(LogisticRegression)	0.492	0.789	0.749	0.743
Community-(RandomForest)	0.368	0.751	0.704	0.666
Community + Lexical	**0.493**	**0.789**	**0.750**	**0.743**

Table 9 Modeling comparisons on multi-class and binary results for the test set, illustrated in Figure 10 (Yellow color is community-(random forest). Red color is lexical-(logistic regression). Blue color is both lexical and community.

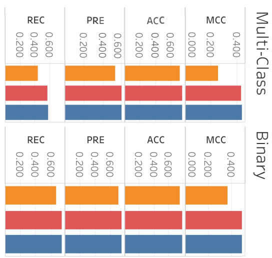

Fig. 10 Modeling comparisons on multi-class and binary results for the test set: (i) all subfigures need to have the same y-axis span; (ii) see Figure 7 for legend placement.

In this experiment, we combine the logic of the lexical pipeline, as described in Section 3.1, and the community pipeline, as described in Section 3.2. We use the prediction of the lexical pipeline as a new input feature for the community pipeline that uses the Random Forest classifier. The combination of features that provided the best results were the following: lexical_prediction, user_followers_count, user_friends_count, user_statuses_count, user_verified, tweet_age, lv_comty_usr_all(majory_dataset), and lv_comty(majory_dataset)-combined.

Community modeling does not consider the content of the tweet beyond hashtags: it models the interactions with the tweet (mentions, quotes, retweets, reply), and with the author (friends). The model trained on community-based and lexical-based features achieved the highest MCC score on the test set, as shown in Figure 10 and Table 9. Binary lexical and community classifications (non-conspiracy vs. conspiracy) have superior performance over the lexical multi-class baseline. Recent work has shown different dispersion patterns regardless of the conspiracy topic [39], and our community and lexical binary captures this observation well, as it outperforms across 4 different measures of classification efficiency; see Figure 10 for details.

Lexical Model vs Community Predictions				
Lexical Model **Multi-class**: correct 2,177; incorrect 731				
	Equal to Lexical		Unique	
Model	Correct	Incorrect	Correct	Incorrect
All network	1726	470	261	451
RT network	799	635	96	1378
Mention network	1106	592	139	1071
Reply network	499	662	69	1678
Quote network	443	686	45	1734
Friends network	1604	517	214	573
Hashtag network	523	671	60	1654
Random Forest	1772	434	297	405
Lexical Model **Binary**: correct 2,293; incorrect 615				
	Equal to Lexical		Unique	
Model	Correct	Incorrect	Correct	Incorrect
All network	1810	265	350	483
RT network	1783	292	323	510
Mention network	1767	299	316	526
Reply network	1737	305	310	556
Quote network	1746	304	311	547
Friends network	1788	295	320	505
Hashtag network	1705	319	296	588
RandomForest	1855	286	329	438

Table 7 Comparison of the predictions between the community models and lexical model. Test data set has 2,908 labeled Tweets. ***Equal to lexical*** is the number of predictions for that model that were classified the same as the lexical model. ***Unique*** is the number of predictions that the model predicted differently than the lexical model.

5.4 Communities and Degree of Overlap

Tables 11, 12, 10,12 show that multiple modalities seem to capture specific information and it is not relevant for community discovery at a global scale. This is due to the negligible overlap between the modalities. However, communities produced by each modality might have value for specific discovery and mining tasks. Moreover, ground truth labeling is missing in COVID(+) to conclude.

ARI	Network	BERTweet	GNN	Network-V	GNN-V
Network	1.0	0.084	0.0002	0.124	0.001
BERTweet	0.084	1.0	0.0004	0.053	0.0266
GNN	0.0002	0.00036	1.0	0.0001	-0.001
Network-V	0.124	0.0533	0.0001	1.0	0.0138
GNN-V	0.001	0.0265	-0.00091	0.01376	1.0

Fig. 11 ARI between various multi-modal modes in processed COVID (+)

ARI	Network	Text-Emb	GNN	Network-V
Network	1.0	0.0046	0.18	0.065
Text-Emb	0.0046	1.0	0.0013	0.0041
GNN	0.18	0.0013	1.0	0.015
Network-V	0.065	0.0041	0.015	1.0
GNN-V	-0.0083	-0.033	-0.0066	0.0033

Fig. 12 ARI between various multi-modal modes in MuMiN - Small

Mode	# of Communities
Network	91,380
BERTweet	81,252
GNN	30,995
Network-V	67,146
GNN-V	87,505

Table 8 # of communities in various multi-modal modes in processed COVID (+)

Mode	# of Communities
Network	108
Text-Emb	3
GNN	3
Network-V	23
GNN-V	161

Table 9 # of communities in various multi-modal modes in MuMiN - Small

ARI	Network	Text-Emb	GNN	Network-V	GNN-V
Network	1.0	0.014	0.032	0.036	-0.0014
Text-Emb	0.014	1.0	0.0021	0.078	-0.020
GNN	0.032	0.021	1.0	0.0044	0.00096
Network-V	0.036	0.0078	0.0044	1.0	0.0014
GNN-V	-0.014	-0.020	0.00096	0.0014	1.0

Table 10 ARI between various multi-modal modes in MuMiN - Medium

Mode	# of Communities
Network	238
Text-Emb	3
GNN	3
Network-V	20
GNN-V	376

Table 11 # of communities in various multi-modal modes in MuMiN - Medium

ARI	Network	Text-Emb	GNN	Network-V	GNN-V
Network	1.0	0.00028	0.000052	0.016	0.000052
Text-Emb	0.00028	1.0	0.00066	0.0044	0.00018
GNN	0.000052	0.00066	1.0	0.000052	0.000052
Network-V	0.016	0.0044	0.000052	1.0	0.99
GNN-V	0.000052	0.00066	0.00012	0.99	1.0

Table 12 ARI between various multi-modal modes in MuMiN - Large

Mode	# of Communities
Network	655
Text-Emb	10
GNN	3
Network-V	21
GNN-V	2

Table 13 # of communities in various multi-modal modes in MuMiN - Large

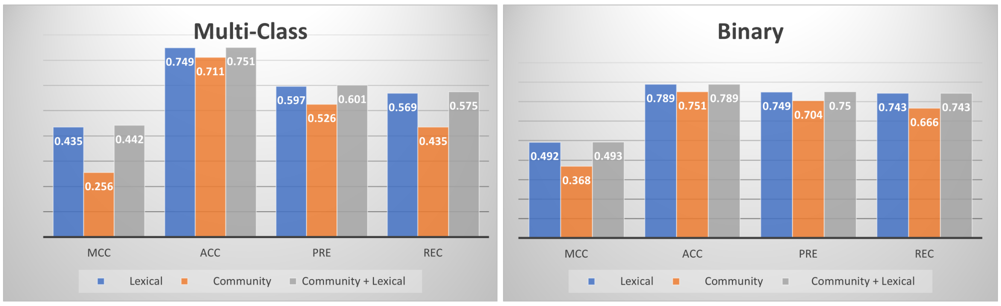

Fig. 13 Modeling comparisons on multi-class and binary results for the test set for Multi-Class (left) and Binary (right) classification. Note that community-only classification offers comparable precision and accuracy without even considering tweet text. Fusion of the lexical and community method offers the best performance across the board.

6 Discussion and Outlook

In conclusion, this research highlights the significant influence of community behavior in tweet classification, suggesting that it carries a comparable weight to tweet content. By introducing a community-based approach to tweet classification, we successfully utilized six distinct community network knowledge graphs to accurately classify tweet content. Our findings demonstrate the advantages of incorporating community attributes and models into the lexical baseline for tweet classification. Notably, community networks offer valuable contextual information for understanding tweet communication, and our study reveals that community-only modeling is as informative as content modeling, as it encompasses crucial details regarding social network interactions with the tweet object. Remarkably, our community modeling techniques, implemented on a large-scale real network, achieved comparable precision, recall, and accuracy to a lexical classifier, even without considering tweet content beyond hashtags. Furthermore, we have shown that basic fusion techniques outperform lexical and network baselines. In contrast, the combination of community and lexical approaches produces the most robust outcomes and superior performance measures, as evidenced by the results of the MediaEval FakeNews task. The complex knowledge graph depicted in Figure 8, which encompasses retweet, mentions, reply, and quote networks, illustrates our ability to capture and incorporate comprehensive network information. Moving forward, we plan to explore enhanced network selection and fusion methods in conjunction with Lexical Modeling and Friends Network, aiming further to improve the effectiveness and accuracy of tweet classification.

References

[1] Nogueira, L.: pytwanalysis Package. https://pypi.org/project/pytwanalysis/

[2] Nogueira, L., Tešić, J.: pytwanalysis: Twitter data management and analysis at scale. In: International Conference on Social Network Analysis Management and Security (SNAMS2021) (2021). https://emergingtechnet.org/SNAMS2021/

[3] Nogueira, L.: Social network analysis at scale: Graph-based analysis of twitter trends and communities. Master's thesis, Texas State University (Dec 2020). https://digital.library.txstate.edu/handle/10877/12933

[4] Pogorelov, K., Schroeder, D.T., Burchard, L., Moe, J., Brenner, S., Filkukova, P., Langguth, J.: Fakenews: Corona virus and 5g conspiracy task at mediaeval 2020. In: Working Notes Proceedings of the MediaEval 2020 Workshop. MediaEval, ??? (2020). http://ceur-ws.org/Vol-2882/

[5] Osmundsen, M., Bor, A., Vahlstrup, P.B., Benchmann, A., Petersen, M.B.: Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. American Political Science Review **115**(3), 999–1015 (2021) https://doi.org/10.1017/S0003055421000290

[6] Geeng, C., Yee, S., Roesner, F.: Fake news on facebook and twitter: Investigating how people (don't) investigate. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, pp. 1–14. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3313831.3376784 . https://doi.org/10.1145/3313831.3376784

[7] Bovet, A., Makse, H.A.: Influence of fake news in twitter during the 2016 us presidential election. Nature communications **10**(1), 1–14 (2019)

[8] Ahmed, W., Vidal-Alaball, J., Downing, J., Seguí, F.L.: Covid-19 and the 5g conspiracy theory: social network analysis of twitter data. Journal of Medical Internet Research **22**(5), 19458 (2020)

[9] Sha, H., Hasan, M.A., Mohler, G., Brantingham, P.J.: Dynamic topic modeling of the covid-19 twitter narrative among us governors and cabinet executives. arXiv preprint arXiv:2004.11692 (2020)

[10] al., I.: Using logistic regression method to classify tweets into the selected topics. In: Intl. Conf. on Advanced Computer Science and Information Systems (ICACSIS), pp. 385–390. IEEE, NY (2016)

[11] Zhou, Zafarani: Fake news detection: An interdisciplinary research. In: WWW Proceedings, p. 1292. ACM, NY (2019)

[12] al., M.: Fake News Detection on Social Media using Geometric Deep Learning (2019)

[13] al., K.: An anatomical comparison of fake-news and trusted-news sharing pattern on twitter. Computational and Mathematical Organization Theory (2020)

[14] Nguyen, V.-H., Sugiyama, K., Nakov, P., Kan, M.-Y.: Fang: Leveraging social context for fake news detection using graph representation. In: Proceedings of the 29th ACM International Conference on Information and Knowledge Management. CIKM '20, pp. 1165–1174. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3340531.3412046 . https://doi.org/10.1145/3340531.3412046

[15] Su, T.: Automatic fake news detection on twitter. PhD thesis, University of Glasgow (2022)

[16] Gangireddy, S.C.R., P, D., Long, C., Chakraborty, T.: Unsupervised fake news detection: A graph-based approach. In: Proceedings of the 31st ACM Conference on Hypertext and Social Media. HT '20, pp. 75–83. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3372923.3404783 . https://doi.org/10.1145/3372923.3404783

[17] Schroeder, D.T., Pogorelov, K., Langguth, J.: Fact: a framework for analysis and capture of twitter graphs. In: 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 134–141 (2019). https://doi.org/10.1109/SNAMS.2019.8931870

[18] Bansal, S.: A mutli-task mutlimodal framework for tweet classification based on cnn (grand challenge). 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM), Multimedia Big Data (BigMM), 2020 IEEE Sixth International Conference on, 456–460 (2020)

[19] Suman, C., Naman, A., Saha, S., Bhattacharyya, P.: A multimodal author profiling system for tweets. IEEE Transactions on Computational Social Systems, Computational Social Systems, IEEE Transactions on, IEEE Trans. Comput. Soc. Syst **8**(6), 1407–1416 (2021)

[20] Gao, W., Li, L., Zhu, X., Wang, Y.: Detecting disaster-related tweets via multimodal adversarial neural network. IEEE MultiMedia, MultiMedia, IEEE **27**(4), 28–37 (2020)

[21] Bruijn, J.A., Moel, H., Weerts, A.H., Ruiter, M.C., Basar, E., Eilander, D., Aerts, J.C.J.H.: Improving the classification of flood tweets with contextual hydrological information in a multimodal neural network. Computers and Geosciences **140** (2020)

[22] Lim, W.L., Ho, C.C., Ting, C.: Sentiment analysis by fusing text and location features of geo-tagged tweets. IEEE Access, Access, IEEE **8**, 181014–181027 (2020)

[23] Gao, D., Li, K., Wang, R., Shan, S., Chen, X.: Multi-modal graph neural network for joint reasoning on vision and scene text. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Computer Vision and Pattern Recognition (CVPR), 2020 IEEE/CVF Conference on, CVPR, 12743–12753 (2020)

[24] Yang, X., Deng, C., Dang, Z., Wei, K., Yan, J.: Selfsagcn: Self-supervised semantic alignment for graph convolution network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16775–16784 (2021)

[25] Wang, J., Wang, Y., Yang, Z., Yang, L., Guo, Y.: Bi-gcn: Binary graph convolutional network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1561–1570 (2021)

[26] Dai, E., Aggarwal, C., Wang, S.: Nrgnn : Learning a label noise resistant graph neural network on sparsely and noisily labeled graphs. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 227–236 (2021)

[27] Liu, Z., Nguyen, T.-K., Fang, Y.: Tail-gnn : Tail-node graph neural networks. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 1109–1119 (2021)

[28] Dai, E., Wang, S.: Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining, pp. 680–688. Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3437963.3441752

[29] Bhatia, T., Manaskasemsak, B., Rungsawang, A.: Detecting fake news sources on twitter using deep neural network. In: 2023 11th International Conference on Information and Education Technology (ICIET), pp. 508–512 (2023). https://doi.org/10.1109/ICIET56899.2023.10111446

[30] Magill, A., Tomasso, M.: Fake News Twitter Data Analysis. https://github.com/DataLab12/fakenews (2020)

[31] Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment **2008**(10), 10008 (2008)

[32] Aynaud, T.: python-louvain 0.14: Louvain algorithm for community detection. `https://github.com/taynaud/python-louvain` (2020)

[33] Nguyen, D.Q., Vu, T., Nguyen, A.T.: BERTweet: A pre-trained language model for English Tweets. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 9–14 (2020)

[34] Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 30. Curran Associates, Inc., ??? (2017)

[35] Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment **2008**(10), 10008 (2008) https://doi.org/10.1088/1742-5468/2008/10/p10008

[36] Campello, R.J.G.B., Moulavi, D., Sander, J.: Density-based clustering based on hierarchical density estimates. In: Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G. (eds.) Advances in Knowledge Discovery and Data Mining, pp. 160–172. Springer, Berlin, Heidelberg (2013)

[37] Chicco, D., Jurman, G.: The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. BMC genomics **21**(1), 1–13 (2020)

[38] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A., Nielsen, H.: Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics **16**(5), 412–424 (2000)

[39] al., V.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)