Mining Knowledge Graphs of Social Network Interactions for Robust Content Classification

ABSTRACT

Social networks vastly changed the way we communicate and disseminate information. In this paper, we investigate the impact of rich social network interactions analysis at scale on message content classification. The large number of users, messages, and tags makes it difficult to separate interesting and meaningful conversation threads from the spread of fake news, malicious accounts, background noise, or irrelevant trolling. We construct several knowledge graphs from Twitter network interactions around COVID-19. We mine the graphs and analyze, and summarize the joint sentiments of Twitter communities by analyzing millions of Tweets. We provide a comprehensive answer to the following questions: 1. can the content of Tweets be classified based on interactions alone; and 2. how well can community classification without analyzing content predict the content class, and which bonds are the most indicative ones. We evaluate the proposed methods and discuss the proposed evaluation measures using ground truth from the MediaEval 2020 Fake News challenge data.

ACM Reference Format:
. 2021. Mining Knowledge Graphs of Social Network Interactions for Robust Content Classification. In *Proceedings of ACM Symposium of Applied Computing Track on Knowledge Graphs (SAC-KG).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Social media has shown increasing influence on shaping public opinions and policies in the past decade, notably playing a prominent role in forming public opinions on the COVID-19 pandemic. One positive outcome is that the development of public campaigns about masking, social distancing, and vaccinations was greatly supported by social media presence and topic spread; however, social media also gave unchecked rise to the circulation of conspiracy theories and misinformation. Many conspiracy theories about the COVID-19 pandemic have circulated on Twitter since the initial outbreak of the virus, seriously undermining compliance with public health measures. The Twitter platform provides a fast exchange of information, opinions, and data, and Twitter's discussions and topic trends have been responsible for shaping the spread and keeping fake news and misinformation alive.

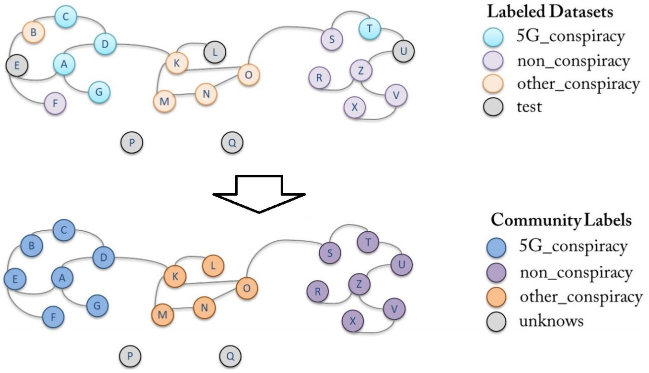


Figure 1: Community attribute enrichment: analyze labeled data set in a network graph and extract community labels from the graph analysis of the network

Can we classify the Tweet as conspiracy or non conspiracy without knowing the content tweet? In this paper we explore the social network context, a Twitter's rich network of interaction i.e., connections, tags, retweets, and mentions. 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. 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 [16–18]. 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 [19]. 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.

2 RELATED WORK

Twitter data has been used to understand the influence of fake news during the 2016 US presidential election [5]; it has also been used to analyze the COVID-19 and the 5G Conspiracy Theory [1]

SAC-KG, Apr 2022, Brno, Czech Republic

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

and the COVID-19 Twitter narrative among U.S. governors and cabinet executives [25]. Using logistic regression to classify Tweets based on topic [7] 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 [26]. Community-based modeling of social networks that leverages the spread of information in social media through retweets and comments has been shown to improve NLP-based modeling [26]. Two major directions of leveraging community information are adapting deep learning techniques to learn the underlying characteristics of the Tweets in communities (e.g., [9]) or exploring the structural and sharing patterns of the topic (e.g., [8]). Deep learning-related work has made significant progress in detecting the spread of news and topics on Twitter [15] but relies heavily on labeled training data. In this paper, we focus on the contributions of community network graph analysis to mine the intent and similarities of users.

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. Some research is focused on improving and extending the preprocessing of content and the network of Tweets to adapt the *Latent Dirichlet Allocation (LDA)* topic modeling tool from text analysis to Tweet format [12]. Twitter analysis studies usually focus on the specific project and fail to generalize the end-to-end process. Theses studies work for a small subspace and specific tasks of the Twitterverse, but they fail to generalize to scalable and subject agnostic systems. Researchers have proposed a data science pipeline for a limited data set for the specific application of integrating machine learning modules for fake news identification [24]. While they achieved scalability for the Twitter data set analysis by parallelizing the process, the data set and specific analysis of fake news limit the extension and generalization of the pipeline [24]. Graph-based approaches focus on biclique identification, graph-based feature vector learning, and label spreading on Twitter [11], but they do not scale well to the number and heterogeneity of the topics examined. In this paper, we use a scalable approach to gather, discover, analyze, and summarize joint sentiments of Twitter communities, extract community and network features, and improve the lexical-based baseline for Tweet classification using community information [18].

3 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 [14].

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, 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 [14]. In this paper, we use Logistic Regression as it has been shown to perform the best for the content-based classification in [14]. To account for the imbalance in data (see Table 2 for details), we experimented with data augmentation. Generating fake Tweets using the most predictive or most common terms for each class led to over-fitting of most classifiers. We took a different route and have adjusted class weights to account for imbalanced data when possible.

4 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 together in the same Tweet. We have developed an in-house scalable package *pytwanalysis* [16–18] to collect and save information-rich Twitter data, create networks, and discover communities in the data.

4.1 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 [2, 4] 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 1. 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 1. 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 3). 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 1.

#	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 1: Community attributes as explained in 4.1

4.2 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; 2. *user_friends_count* (Fig. 3; 3. *user_statuses_count* (Fig. 4); 4. *user_verified* (Fig. 6); 5. *tweet_age (days since creation)* (Fig. 5). 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 1 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, 3, 4, 5, and 6.

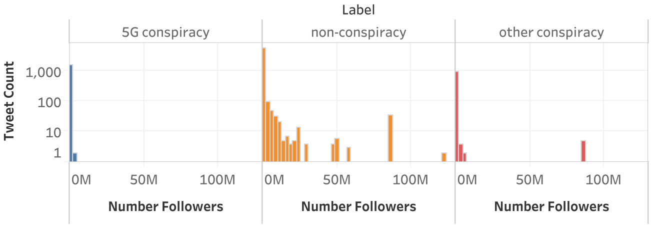


Figure 2: Distribution of the feature *user_followers_count* for the different class labels (5G, non, other)

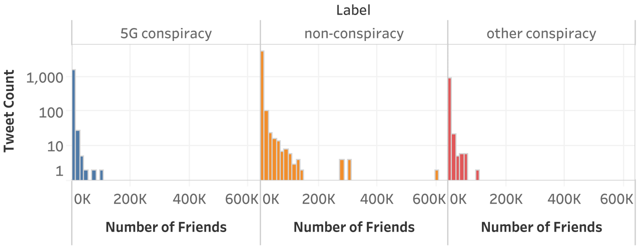


Figure 3: Distribution of the feature *user_friends_count* for the different class labels (5G, non, other)

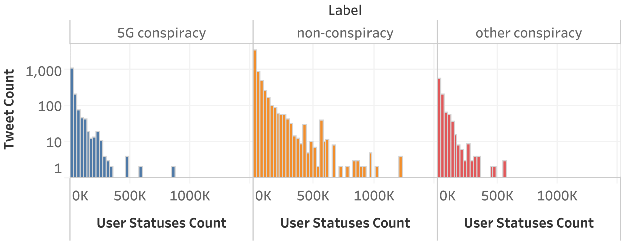


Figure 4: Distribution of the feature *user_statuses_count* for the different class labels (5G, non, other)

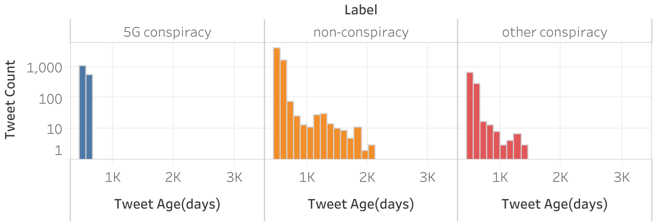


Figure 5: Distribution of the feature *tweet_age* for the different class labels (5G, non, other)

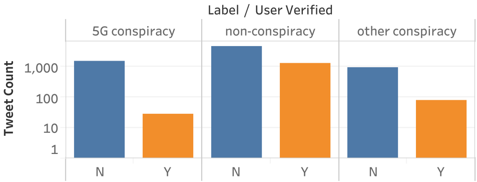


Figure 6: Distribution of the feature *user_verified* for the different class labels (5G, non, other)

5 EXPERIMENTAL SETUP

5.1 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 2). 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 2. 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 [17]. 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).

5.2 Measures

We measured the performance of the proposed methods on a very small labeled subset of test data in Table 2. 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) [3, 6, 19]. 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.

Dataset	Tweet Count	User Count
1. Fake News [19]	**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 [19]		**3,385,981**
3. COVID-19 (+) [17]	**771,203**	**657,785**

Table 2: 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.

6 RESULTS AND ANALYSIS

6.1 Lexical Analysis Pipeline

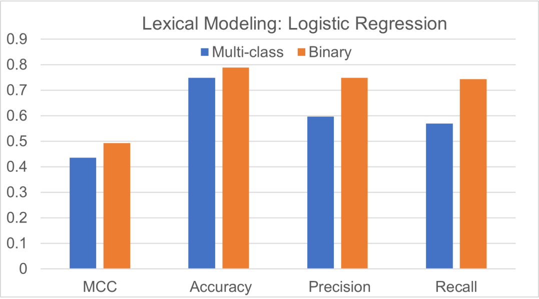


Figure 7: Multi-class and binary labeling scores (MCC, Accuracy, Precision, Recall) for 2020MediaEval Test Set. Model abbreviations: LR for logistic regression; LR-OCR for logistic regression with OCR

Figure 7 shows the metrics for the multi-class and binary predictions using the Logistic Regression classifier [14]. 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 [14] 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.

6.2 Community Discovery Methodologies Comparison

The Louvain method was the main algorithm used for community discovery because of its scalability, low execution time, and the ability to find communities in disconnected networks. In order to compare our chosen method with other state-of-the-art community discovery methods when applied to our task, we ran an experiment using the Karate Club[23] library implementation of the following non-overlapping community detection algorithms: GEMSEC[22], EdMot[13], SCD[20], and LabelPropagation[21]. For the comparison, we used the small network created from user mentions without the additional COVID-19 (+) data, containing 6,193 vertices and 6,654 edges.

Figure 8 shows the metrics for the multi-class predictions using the community majority assignment for the mention network in the five different models. The results are for the predictions excluding the unknowns. The Louvain method shows comparable performance with the other models, but with the lowest execution time. The SCD model performed the best of the five models. For small networks, the SCD method seems better suited for this task. Since we are using a large network in the analysis of over 2 million vertices and 3 million edges, the execution time is an important factor in the decision of what method to use.

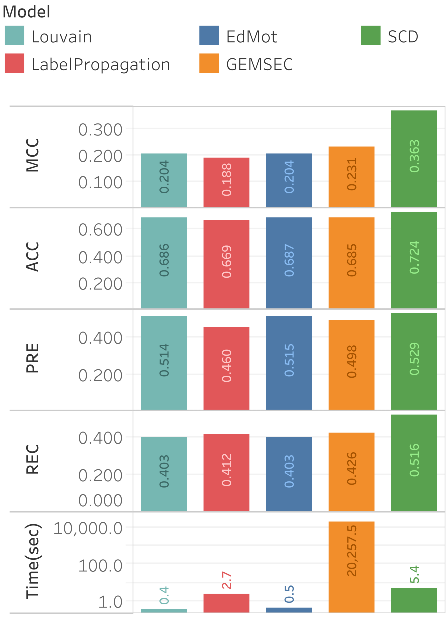


Figure 8: Comparison of different community discovery methods for the mention network. Performance measures (execution time in seconds, MCC, Precision, Recall, and Accuracy) were computed for every model. Metrics were evaluated excluding the unknown predictions.

6.3 Community Analysis Pipeline

Table 3 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 3. We separate the 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 3 and in Figure 9.

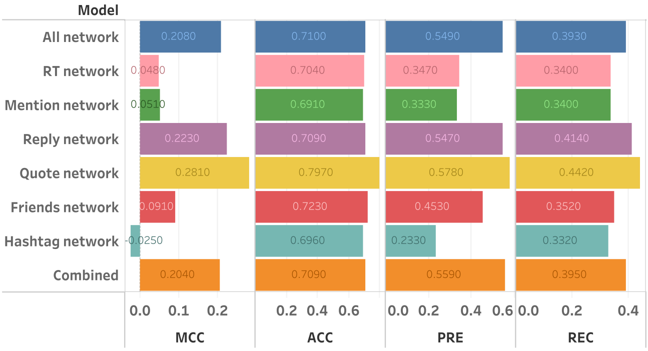


Figure 9: Comparison of the multi-class community majority assignment excluding the unknowns for the different types of networks, as detailed in section *Multi-class without Unknowns* in Table 3

The results in Table 3 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 5.1. Community based analysis with the auxiliary data brought the value of community connections to this analysis; compare model and model+ in Table 3. 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 3. The classifier can assign values

			Multi-class With Unknowns				**Multi-class Without 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 3: 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 Tweets that could not be classified with the community majority assignments, since it uses additional features apart from the community features; see Section 4.2.

Table 4 summarizes the correct classification results that the network modeling produces that lexical one does not. Tweets that have ambiguous text but strong and well-separated connections in the network can potentially perform better than the lexical model, as shown in Table 6. The community approach can be useful when the users in the available data are well-connected, as shown in the Table 5 example. The community predictions perform comparably for cases in which the Tweet was not isolated from the network. Figure 10 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.

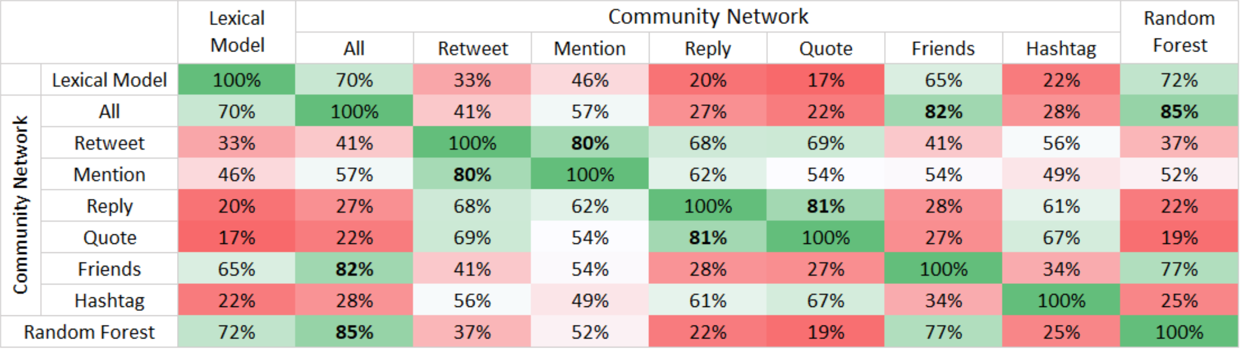
6.4 Combining Community and Lexical Attributes

In this experiment, we combine the logic of the lexical pipeline, as described in Section 3, and the community pipeline, as described in Section 4. We use the prediction of the lexical pipeline as a new input feature for the community pipeline that uses the RandomForest 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 11 and Table 7. 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 [10], and our community and lexical binary captures this observation well, as it outperforms across 4 different measures of classification efficiency; see Figure 11 for details.

7 DISCUSSION AND OUTLOOK

In this paper, we have shown that community behavior speaks *almost* as loud as the tweet content for tweet classification. We have proposed a community based approach to tweet classification and mined six different community network knowledge graphs to correctly classify the tweet content. We have demonstrated how the lexical baseline for Tweet classification can benefit from community attributes and community models. Community networks provide context for tweet content classification, and we demonstrate that community only modeling is comparable to lexical modeling, as it contains a lot of useful information on the social network interactions with the author and the tweet object. Community modeling in a large real network achieved comparable precision recall and accuracy to lexical classifier *without* considering tweet content beyond hashtags. We have also demonstrated that basic fusion improves



		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%

Figure 10: 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.

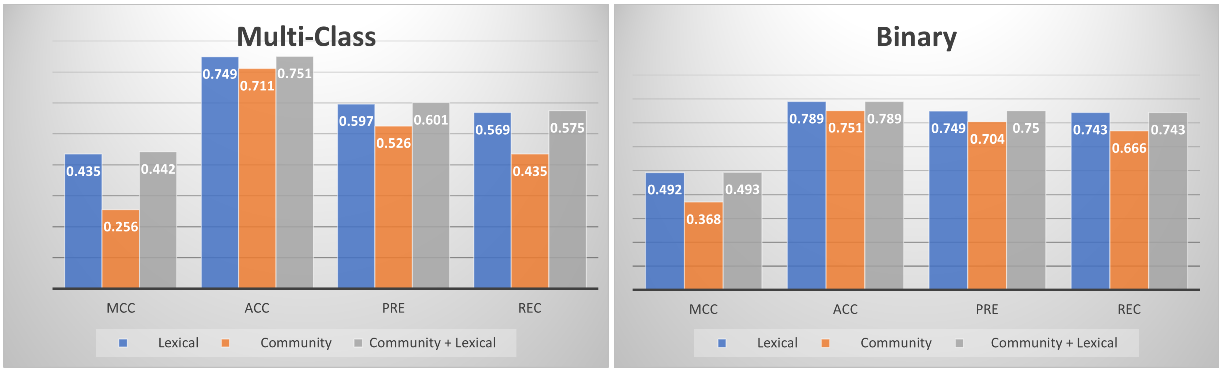


Figure 11: Modeling comparisons on multi-class and binary results for the test set for Multi-Class (left) and Binary (right) classification. Note that community-only classification offers comparable precision and accuracy without even considering tweet text. Fusion of the lexical and community method offers the best performance across the board.

over both lexical, and network baselines, and community+lexical approach to tweet classification provides the most robust outcome and best measures, as reported on MediaEval 2020 FakeNews task. Figure 10 indicates we can create complex knowledge graph from retweet, mentions, reply, and quote networks that captures all network information. We will explore better network selection and fusion methods with Lexical Modeling and Friends Network.

REFERENCES

[1] Wasim Ahmed, Josep Vidal-Alaball, Joseph Downing, and Francesc López Seguí. Covid-19 and the 5g conspiracy theory: social network analysis of twitter data. *Journal of Medical Internet Research*, 22(5):e19458, 2020.

[2] Thomas Aynaud. python-louvain 0.14: Louvain algorithm for community detection. `https://github.com/taynaud/python-louvain`, 2020.

[3] Pierre Baldi, Søren Brunak, Yves Chauvin, Claus AF Andersen, and Henrik Nielsen. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, 16(5):412–424, 2000.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

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

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

[7] Indra et al. Using logistic regression method to classify tweets into the selected topics. In *Intl. Conf. on Advanced Computer Science and Information Systems (ICACSIS)*, page 385–390, NY, Oct 2016. IEEE.

[8] Kumar et al. An anatomical comparison of fake-news and trusted-news sharing pattern on twitter. *Computational and Mathematical Organization Theory*, 2020.

[9] Monti et al. Fake news detection on social media using geometric deep learning, 2019.

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

[11] Siva Charan Reddy Gangireddy, Deepak P, Cheng Long, and Tanmoy Chakraborty. Unsupervised fake news detection: A graph-based approach. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, HT '20, page 75–83, New York, NY, USA, 2020. Association for Computing Machinery.

[12] Elıas Jónsson and Jake Stolee. An evaluation of topic modelling techniques for twitter, 2015.

[13] Pei-Zhen Li, Ling Huang, Chang-Dong Wang, and Jian-Huang Lai. Edmot. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Jul 2019.

[14] Andrew Magill and Maria Tomasso. Fake news twitter data analysis. https://github.com/DataLab12/fakenews, 2020.

[15] Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, and Ryota Kobayashi. Modeling the spread of fake news on twitter. *PLOS ONE*,

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 4: 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.

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

16:1–16, 04 2021.

[16] Lia Nogueira. pytwanalysis package. https://pypi.org/project/pytwanalysis/.

[17] Lia Nogueira. Social network analysis at scale: Graph-based analysis of twitter trends and communities. Master's thesis, Texas State University, Dec 2020.

[18] Lia Nogueira and Jelena Tešić. pytwanalysis: Twitter data management and analysis at scale. In *International conference on Social Network Analysis Management and Security (SNAMS2021)*, Dec 2021.

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 6: Example 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.

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 7: Modeling comparisons on multi-class and binary results for the test set, illustrated in Figure 11.

[19] Konstantin Pogorelov, Daniel Thilo Schroeder, Luk Burchard, Johannes Moe, Stefan Brenner, Petra Filkukova, and Johannes Langguth. Fakenews: Corona virus and 5g conspiracy task at mediaeval 2020. In *Working Notes Proceedings of the MediaEval 2020 Workshop*. MediaEval, 2020.

[20] Arnau Prat-Pérez, David Dominguez-Sal, and Josep-Lluis Larriba-Pey. High quality, scalable and parallel community detection for large real graphs. In *Proceedings of the 23rd international conference on World wide web*, pages 225–236, 2014.

[21] Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 76(3):036106, 2007.

[22] Benedek Rozemberczki, Ryan Davies, Rik Sarkar, and Charles Sutton. Gemsec: Graph embedding with self clustering, 2019.

[23] Benedek Rozemberczki, Oliver Kiss, and Rik Sarkar. Karate Club: An API Oriented Open-source Python Framework for Unsupervised Learning on Graphs. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*, page 3125–3132. ACM, 2020.

[24] D. T. Schroeder, K. Pogorelov, and J. Langguth. Fact: a framework for analysis and capture of twitter graphs. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 134–141, 2019.

[25] Hao Sha, Mohammad Al Hasan, George Mohler, and P Jeffrey Brantingham. Dynamic topic modeling of the covid-19 twitter narrative among us governors and cabinet executives. *arXiv preprint arXiv:2004.11692*, 2020.

[26] Zhou and Zafarani. Fake news detection: An interdisciplinary research. In *WWW Proceedings*, page 1292, NY, 2019. ACM.