Contextualizing Online Conversational Networks

Thomas Magelinski and Kathleen M. Carley

Center for Computational Analysis of Social and Organizational Systems Carnegie Mellon University School of Computer Science 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213 tmagelin@cs.cmu.edu

Abstract

Online social connections occur within a specific conversational context. Prior work in network analysis of social media data attempts to contextualize data through filtering. We propose a method of contextualizing online conversational connections automatically and illustrate this method with Twitter data. Specifically, we detail a graph neural network model capable of representing tweets in a vector space based on their text, hashtags, URLs, and neighboring tweets. Once tweets are represented, clusters of tweets uncover conversational contexts. We apply our method to a dataset with 4.5 million tweets discussing the 2020 US election. We find that even filtered data contains many different conversational contexts, with users engaging in multiple conversations. While users engage in multiple conversations, the overlap between any two pairs of conversations tends to be only 30-40%, giving very different networks for different conversations. Even accounting for this variation, we show that the relative social status of users varies considerably across contexts, with $\tau = 0.472$ on average. Our findings imply that standard network analysis on social media data can be unreliable in the face of multiple conversational contexts.

Introduction

Social network analysis relies on proper contextualization of social interactions. For offline social networks, contextualization has been traditionally done by scoping the measurement of social interactions within a physical space: a conference, an office, a school, etc. Situating a social network within a single context both provides a clean dataset and allows for interpretable analysis.

Central members in a network of interactions within an office place can be seen as information brokers within the office. Including information about how the workers interact outside of the office provides more information but can also muddy the analysis. Adding out-of-office connections to the initial network is likely to affect who the central actors are, what the community structure is, and the general topology of the network. While this denser network encodes more information, it also conflates two types of edges; workers will interact with each-other differently according to where they are. Thus, it is more appropriate to study the *contextualized* networks and the relationship between them. This may include studying changes in centrality and community structure from one context to another. Contextualization not only improves the specificity of the claims that can be made from the original network analysis, but also adds new information about the relationship between contexts.

This problem is illustrated in Figure 1. This simple scenario details 3 different social contexts with one having much different community structure than the others. A standard or decontextualized analysis mixes all of the edges together and hides *all* community structure. Networks analyses are known to be sensitive to data quality, making this an important problem (Borgatti, Carley, and Krackhardt 2006).

Methods for contextualizing social networks to date have leveraged simple property of offline networks: people can only be in one place at one time. The result of this obvious fact is that offline social contexts occur in sequences. For example, someone might go to work, then go to a restaurant to meet their friends, and finally return home to their family. This then creates 3 sequential social contexts: interactions among co-workers, friends, and family. Dynamic network analysis methods have leveraged this sequential structure can be leveraged to identify network "states," effectively contextualizing these interactions.

Online communication is different. Social media platforms such as Twitter are designed for users to engage in vastly different discussions simultaneously¹, ruining the sequential structure of contexts. Without this sequential structure, existing dynamic network approaches are inapplicable.

Researchers attempt to account for this by scoping the data collection to a specific topic or event. On Twitter, the available filters for data collection include keywords, specific users, and geographical bounding boxes. After applying these filters, researchers can obtain reasonably contextualized datasets about a certain event or topic. However, a related area of research, story-detection, has demonstrated that multiple events or "stories" have separate discussions

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹Technically, people can only send one tweet at a time, so they are actually oscillating between conversations rather than simultaneously being engaged in them. The distinction for online interactions is the time-scale of state changes. Online discussions play-out over hours, while users switch between conversations within minutes. This mismatch in timescales creates the ability for users to be in multiple discussions "simultaneously"



Figure 1: An illustration of different social contexts having different network structure. Before accounting for context, no community structure is apparent. After contextualization, all contexts appear to have community structure, with context 3's structure being very different from that of 1 and 2.

occurring even within filtered datasets.

Properly contextualizing online conversational networks is critical given their importance within the field of Social Cybersecurity (Carley et al. 2018; Carley 2020). Online conversational networks are deeply integrated into the methods for studying and understanding information operations (Lazer et al. 2018; Grinberg et al. 2019; Uyheng et al. 2019). Thus, accurate representations of the conversational networks are necessary to understand the information space.

This brings us to our first research question. **RQ1: Given a set of Tweets, can we cluster them into different conversations?** Methods that answer this question may be similar to topic or story detection. However, these approaches do not typically ensure that direct interactions like replies or quotes are given the same assignment as the content they interact with. At the same time, the work in story detection has shown the importance of URLs for capturing the context of a conversation. Thus, the method that we develop to answer RQ1 must account for the text, conversational structure, and the URLS involved. We develop such a method which works with multiple languages, is unsupervised, and also accounts for hashtags.

The contextualization model enables us to build and study the contextualized networks shown in Figure 1. Our primary objective is to test whether this separation of network data affects the outcome of a network analysis. To meet this objective, we ask two more research questions: **RQ2: Do different conversational contexts have the same active users?** and **RQ3: How does user status vary across contexts?** The answers to these questions will show whether contextualized analysis is actually needed to understand large social media data.

By applying our methodology on a real Twitter dataset consisting of 4.5 million tweets, we show that multiple conversational contexts are present in the data and that the combination of these contexts leads to misleading measurements of user centrality and network structure. Thus, standard network analyses may be unreliable in the face of the multitude of conversations present in large social media datasets.

Related Work

Social connections must occur in the same context for social network analysis to work effectively. What constitutes the "same context" depends on the study. For example, if a study seeks to understand the spread of information in the workplace, the inclusion of connections outside the workplace may be inappropriate. If the study instead was looking to measure epidemic spreading, all interactions are appropriate to include. In many settings, and particularly for offline networks, this is an extremely easy requirement to meet which is easily satisfied though data collection processes such as observing connections in a specific place.

For offline networks, dynamic analysis methods have been developed to detect sequences of network states, finding that datasets observed over longer time periods contain multiple contexts (Peixoto and Rosvall 2017; Masuda and Holme 2019; Magelinski and Carley 2019). In one example, changes can be observed from how students interact at lunch compared to in the classroom (Peixoto and Rosvall 2017; Masuda and Holme 2019). Students interact differently in the lunch context than they can in the classroom context. In another example, changes are observed in how Ukrainian legislatures cooperate before and after the Euromaidan revolution (Magelinski and Carley 2019). An upheaval in sociopolitical context disrupted friendships and rivalries between politicians. These studies find that the community structure and central actors can be very different from context to context, and that combining contexts leads to an inaccurate representation of the network. Accounting for contexts has also led to improvements in the modeling of processes occurring on the networks (Peixoto and Gauvin 2018).

For online social networks, however, contextualization is not an easy task. Two related fields have shown that social media data often contains multiple entangled contexts: topic modeling and story detection. Topic modeling seeks to uncover a selection of different semantic contexts, or "topics" which occur within a collection of documents (Blei, Ng, and Jordan 2003). Traditional topic models such as LDA are poorly suited for the extremely short documents in Twitter data, leading to topic models specifically designed for short texts (Hong and Davison 2010; Zuo et al. 2016; Cheng et al. 2014). Alvarez-Melis and Saveski found tweets can aggregate information from their conversational context to improve topic representation (Alvarez-Melis and Saveski 2016). Other topic detection models have been developed which specifically leverage the hashtag feature of Twitter data to obtain topics (Wang et al. 2014; Magelinski, Bartulovic, and Carley 2020; Feng et al. 2015). Methods differ, but all of these works successfully demonstrate the presence of multiple semantic contexts in Twitter conversations.

Topic modeling demonstrates that entirely different things may be discussed in the same Twitter dataset, while storydetection shows that different contexts can occur even within very similar topics. Story detection seeks to uncover "stories" or discussions tied to specific events (Petrović, Osborne, and Lavrenko 2010; Srijith et al. 2017; Alshaabi et al. 2021a,b). First-story detection and event-detection are very related, as they seek to identify the first tweets breaking the news of a story developing, compared to more general storydetection, which detects all the tweets in the discussion of that story (Walther and Kaisser 2013; Petrović, Osborne, and Lavrenko 2010; Osborne et al. 2012). In any case, detected stories are separate contexts which could otherwise be considered the same topic. For example, story detection applied to Donald Trump's twitter timeline can distinguish withinparty arguments from between-party arguments, which both belong to the topic of federal US politics (Dodds et al. 2021). Another example applies story detection to the Twitter discussion following the police killing of Michael Brown (Srijith et al. 2017). Here, fine distinctions of context are made, such as the difference in discussion of the police-lead smear campaign against Michael Brown from the discussion of the robbery that Brown committed early in the day of the shooting. This is to say that both topic modeling and story detection develop methods of uncovering discrete conversational contexts on social media, and thereby demonstrate that these contexts exist. These works do not, however, investigate the implications of this finding for social network analytics.

While dynamic analysis can leverage the sequential structure of human movement in offline networks, this is not possible with online networks. The studies in topic modeling and story detection show that conversations within these conflicts can occur simultaneously, with users rapidly switching between contexts. And while methods from topic modeling and story detection can be used to uncover conversational contexts, existing methods don't typically leverage all of the available indicators of context simultaneously: tweet text, hashtags, URLs, and the conversational graph. Further, existing methods classify tweets with discrete labels, rather than represent them in a continuous space. Discrete labeling is useful for network analysis, but gives no way of measuring things like distance between contexts.

Advancements in graph neural networks enable us to develop a new architecture for unsupervised Tweet representation which leverages all of the available data and places tweets in a continuous space. Older methods of unsupervised node representation relied on random walks or "surfs" to obtain local information which can be encoded in node vectors (Grover and Leskovec 2016; Perozzi, Al-Rfou, and Skiena 2014; Cao, Lu, and Xu 2016). These methods do not rely on node features to obtain their representations, in contrast to the graph convolutional networks that are typically used in the semi-supervised or supervised setting (Kipf and Welling 2016; Hamilton, Ying, and Leskovec 2017). Node features are necessary for tweet representation because they are used to represent the actual contents of a tweet, the tweet's text.

Methods leveraging node features have been used applied to model social media users in a number of supervised settings, including the detection of hateful users, and the prediction of locations. (Pan and Ding 2019; Ribeiro et al. 2018; Do et al. 2018). Perhaps the closest related model to ours is that of Nguyen et al., who used unsupervised embedding methods such as BigGraph for users, hashtags, and URLs, before combining them in a supervised retweet prediction model (Nguyen et al. 2020; Lerer et al. 2019).

While models leveraging node features have been developed for social media, a mechanism for training them in an unsupervised manner was not available. Deep Graph Infomax (DGI) filled this gap by outlining an unsupervised training procedure for feature-leveraging approaches through the principle of mutual information (Velickovic et al. 2019). Similar to Stuctural Deep Network Embedding (SDNE), DGI derives an objective function in the unsupervised setting so that the architecture has something to optimize (Wang, Cui, and Zhu 2016).

Because DGI is a methodology for training, the specific architecture for node embedding is customizable, similar to the HARP procedure (Chen et al. 2017). In this work, we develop a custom GCN-based architecture for representing tweets, which uses the conversational network, hashtags, and URLs. The architecture is then trained with DGI on a real dataset. We use the obtained tweet representations to contextualize user-to-user interactions and demonstrate the importance of contextualized network analysis.

Data

Data Collection

The data collection strategy for this study is intended to match the typical procedures used in the field of Social-Cybersecurity, which relies heavily on network analysis of social media discussions (Uyheng et al. 2019). Thus, the data was captured using a keyword-based stream of Twitter's API from November 2 2020 to November 8 2020. This allowed for the capture of data one day before election night, which was November 3 2020, and one day after major news outlets declared Joe Biden the winner on November 7 2020. The keywords² were selected in order to maximize conversation around the election. This includes general hashtags, campaign hashtags, and mentions of prominent election figures such as Trump, Pence, Biden, and Harris. It also includes hashtags relating to anticipated electionrelated issues, such as the Black Lives Matter movement, US Sanctions on Iran, issues with voting-by-mail, and claims of voter fraud. The collection resulted in 4.5M tweets, 75k hashtags, and 47k URLs. The dataset approximately contains 2M retweets, 1.3M quotes, and 1.3M replies. Hashtags were used in tweets 886k times, whereas URLs were used in tweets 75k times.

Data Cleaning

Tweet text was cleaned by first removing all URLs, hashtags, and mentions. Next, punctuation was removed. Finally, text was tokenized in preparation for the text embedding discussed in the Methodology section.

²#election2020, #presidentialelection, #JoeBiden, #Biden, #BidenHarris2020, #MAGA, #KAG, #democrats, #republicans, #VoteByMai, #USPS, #SaveTheUSPS, #voterfraud, #reopen, #reopenamerica, #BLM, #BlackLivesMatter, #QAnon, #WWG1WGA, #IranSanctions, "natural born," and relevant politician's handles.

The procedure for URL normalization was as follows. First, text before the domain name was removed. Next, URL parameters were removed for links with domains other than "facebook", "google", and "youtube." These parameters commonly store information about the user who shared the link, among other things. The presence of these parameters prevents direct matching between URLs. For "facebook", "google", and "youtube," however, these parameters are used to point to the actual destination, so cannot be removed. "Amp" links were converted to non-amp links. Lastly, youtube.com, and yout.be links were all converted to the yout.be format.

All links to twitter.com were not considered to be typical URLs, as they are either links to media or quotes of other tweets. Links to media were not included, while the metadata from quote-links was used to add the appropriate quoteedges in the tweet-tweet network discussed below. Hashtags were lower-cased, as case does not affect their functionality.

Heterogeneous Network Construction

We construct our heterogeneous Twitter network with three node types (tweet, hashtag, and URL) and three edge types (tweet-URL, tweet-hashtag, and tweet-tweet). When a URL or hashtag is used in a tweet, an edge is drawn between them.

The third relationship, tweet-tweet, occurs through replies or quotes. While these are slightly different operations, they both create the effect of continuing the conversation with a new tweet connected to the original. Future work may investigate leveraging slight differences between replies and quotes. Edges between tweets can be modeled as directed or undirected as a setting within the model. A directed edge allows the reply or quote's representation to be affected by the original tweet's representation while keeping the original tweet's representation isolated. This is an intuitive modeling approach; however, modeling this relation with an *undirected* edge allows for base tweets to obtain some context, which can push similar but disconnected conversations closer together. To find out which approach worked best, we quantitatively compare them in the Model Selection section.

Retweets are simply copies of tweets, so they will provide no additional information from a tweet-representation point of view. Worse, they are such a large fraction of the dataset that they could have adverse effects on the training process. Instead, we give retweets the same representation as their original tweet. Thus, retweets will always be considered in the same context as the original tweet, where the user-to-user implications of retweets are studied.

Methodology

Tweet Text Embedding

Graph convolutional networks require some form of nodefeatures. We derive features for tweets using the tweet text. To limit the scope of analysis to our proposed architecture and to enable the use of multi-language text embedding, we used the pre-trained³ and language-aligned vectors trained using fastText on the Wikipedia corpus (Joulin et al. 2018; Bojanowski et al. 2017). The use of language-aligned vectors allows us to place similar tweets in the same discussion, even if they are tweeting in different languages.

We rely on the Twitter language detection output for the classification of tweet language. Many tweets, however, do not have an available language label. This often occurs when tweets do not have text, but instead only have URLs, emojis, images, and sometimes hashtags. In our case, 15.6% of tweets in the dataset do not have an available label, and therefore cannot be yet embedded.

For each tweet with a label, we perform a normalized tfidf weighting of the fastText word vectors to obtain a 300dimensional tweet-text embedding. The tf-idf weights were calculated within-language to prevent language-based abnormalities. We use this procedure to embed tweets in Arabic, English, French, German, Hebrew, Italian, Portuguese, Romanian, Russian, Spanish, and Turkish, covering over 95% of the reachable tweets.

The initial embeddings for hashtags and URLs are obtained similarly. Aggregating the text from all the tweet's using a hashtag has been shown to be useful in topic modeling (Steinskog, Therkelsen, and Gambäck 2017). Following this result, we concatenate all same-language texts for each hashtag and URL. Next, for each hashtag or URL, a weighted average of the word vectors was applied to each language's document to obtain a language-specific vector. Finally, the language-specific vectors were averaged to give a single vector for each hashtag and URL. The same tf-idf weighting scheme applied to Tweets performed poorly in this context due to the length of the documents. In the case of #election2020, the top 25 terms had tf-idf weights one or more orders of magnitude higher than the following 15000. However, because the 25 terms were so outnumbered, their sum only accounted for 10% of the final representation. Raising the weights to a higher power, p = 3 lessened this effect such that the important 25 terms accounted for over 90% of the final representation. Because of the difficulty that graph convolution has representing nodes with very high degree, we keep these representations fixed in training.

Finally, we use feature propagation to obtain a feature vector for the remaining tweets (Rossi et al. 2022). Feature propagation holds known feature vectors fixed while iterative updating unknown feature vectors. In each iteration, each node with an unknown feature vector updates its vector by taking the average features of its neighbors. Nodes with unknown features which have not been reached by the propagation are not counted in the update step. After few iterations, all features converge. Rossi et al. demonstrate that this approach yields good results in downstream tasks such as node classification even in the face of extreme missing data, when 99% of nodes are featureless. The task of filling in features for 15.6% of nodes is much less formidable. Feature vectors converged within 40 iterations on our dataset.

Deep Tweet Infomax

We propose an unsupervised approach for tweet representation. The flow of information in 1 step of the graph neural network architecture can be seen in Figure 2. A tweet aggregates information from tweets that it is connected to (replies,

³https://fasttext.cc/docs/en/aligned-vectors.html



Figure 2: The flow of information in one layer of DTI.

or quotes), hashtags, and URLS, where the representations of Hashtags and URLs were obtained by aggregating from the tweets that they are used in. This model is trained using Deep Graph Infomax, leading to the informal name of our approach of Deep Tweet Infomax (DTI). The architecture will now be described in detail.

Let t, u, and h represent nodes of the type tweet, URL, and hashtag, respectively. They will be indexed using subscripts, e.g., t_i corresponds to the i^{th} tweet. Feature vectors are represented with the letter x, using subscripts to indicate the corresponding node and superscripts to indicate the layer. For example, $x_{t_i}^0$, represents the 0th-layer vector (otherwise known as the feature vector) for the i^{th} tweet. We will make use of a neighborhood function \mathcal{N} , which takes in a node and returns the set of its neighbors. Subscripts of the neighborhood function allow for the return of only a specific type of neighbor. For example, $\mathcal{N}_u(t_i)$ returns all of the URLs connected to the i^{th} tweet.

Tweets aggregate from their heterogeneous neighborhoods as shown in Equation 1, where AGG is a learnable aggregation function, and σ is an activation function. Separate aggregation functions are learned for the tweets, hashtags, and URLs that a tweet is connected to, which are then averaged, and an activation function is applied. Note that although a simple average is taken in Equation 1, it is possible that the learned aggregation schemes scale the components such that certain nodesets, e.g. URLs, play a larger role in the representation.

$$x_{t_i}^1 = \sigma(\frac{1}{3}(\operatorname{AGG}(\{x_{h_i}^0, \forall h_i \in \mathcal{N}_h(t_i)\}) + \operatorname{AGG}(\{x_{u_i}^0, \forall u_i \in \mathcal{N}_u(t_i)\}) + \operatorname{AGG}(\{x_{t_i}^0, \forall t_i \in \mathcal{N}_t(t_i) \cup \{t_i\}\})))$$
(1)

The process thus far defines the network over which features are passed and the order in which to pass them. The selection of the aggregation function, AGG, is the main topic of debate within graph neural network research. In future work, AGG, can be expected to be substituted for the new state-of-the-art aggregation schemes. We first employ the GraphSAGE aggregation, which is the initial aggregation scheme applied in the Deep Graph Infomax work (Hamilton, Ying, and Leskovec 2017). This aggregation scheme is detailed for the tweet-to-tweet relationship in Equation 2, where **W** are trainable weight matrices, and **b** is a trainable bias vector.

$$x_{t_i}^1 = \mathbf{W}_1 x_{t_i}^0 + \frac{1}{|\mathcal{N}(t_i)|} \sum_{t_j \in \mathcal{N}(t_i)} \mathbf{W}_2 x_{t_j}^0 + \mathbf{b} \qquad (2)$$

GraphSAGE is a relatively simplistic aggregation scheme, where all neighbors are treated equally. More recent aggregation schemes add attention, which allows for a weighted average of neighbors. Thus, we compare results with those from the attention function detailed by Brody, Alon, and Yahav, which improved on the original graph-attention network from Veličković et al (Brody, Alon, and Yahav 2021; Veličković et al. 2017). This aggregation function is shown in equations 3 and 4, where Θ is a matrix of learnable weights, **a** is a learnable vector, and α gives the attention score between two tweets. The comparison between aggregation functions is given in the Model Selection section.

$$\begin{aligned} x_{t_i}^1 &= \sum_{t_j \in N_t(i) \cup \{t_i\}} \alpha_{t_i^0, t_j^0} \Theta x_{t_j}^0 \tag{3} \\ \alpha_{t_i, t_j} &= \frac{\exp(\mathbf{a}^T \text{LeakyReLU}(\Theta[x_{t_i}||x_{t_j}]))}{\sum_{t_k \in N(i) \cup \{t_i\}} \exp(\mathbf{a}^T \text{LeakyReLU}(\Theta[x_{t_i}||x_{t_k}]))} \end{aligned}$$

Finally, we must select a nonlinear activation function. Again following the original Deep Graph Infomax work, we use the PReLU, activation function (He et al. 2015). Afterwards, the vectors are L2 normalized, to enable easy comparison with cosine similarity.

The process up to here details a single-layer of the architecture, where Tweets will only obtain information from 1hop away. Stacking these layers enables further information spread. Due to experiments not shown for space considerations, we chose a depth of 3. On Twitter, the vast majority of replies are replies to a base-tweet, rather than replies to replies. So, 3 layers capture nearly all the context of a Tweet.

As a final step, the hashtags and URLs were projected into the same space as the Tweets by applying their respective final-layer aggregation functions in Equation 1.

To train this architecture, we use Deep Graph Infomax (DGI), an approach for learning unsupervised node representations by maximizing mutual information between patch representations and corresponding high-level summaries of graphs (Velickovic et al. 2019). We note that a version of DGI has been developed specifically for heterogeneous networks (Ren et al. 2019). Because of our focus on tweet representations and the lack of features available for URLs and hashtags, we proceed with the original formulation of DGI.

The DGI training process involves four steps. First, a normal forward pass on the data is performed, giving tweet representations, \mathbf{x}_t . Next, a readout function is applied to give a graph-level summary vector, s. Velickovic et al. applied a sigmoid function to a simple averaging of the node vectors but suggest that more sophisticated methods such as the Set2Vec method could perform better on larger graphs (Vinyals, Bengio, and Kudlur 2015). So for the summarization step, we compare the mean function with Set2Vec using 5 processing steps: $\mathbf{s} = \sigma(\text{Set2Vec}(\{x_{t_i} \forall t_i\}))$, where σ is the sigmoid function. For the third step, a forward pass is performed on corrupted data, giving corrupted tweet representations, $\tilde{\mathbf{x}}_t$. We use the same corruption function as the original work, a shuffling of the tweet features while keeping edges intact. Finally, to classify tweets as corrupted or not a scoring function is given as $d_{t_i} = \sigma(x_{t_i}^T \mathbf{W}s)$, where \mathbf{W} is a trainable scoring matrix and σ is the sigmoid function, providing a score between 0 and 1. Binary cross entropy loss was used on the score, d, and the label (corrupted or not) to train the model.

The model was implemented using the PyG library (Fey and Lenssen 2019). All hidden and output layer dimensions were set to 300 to match the initial features. The model was trained using mini-batches of 2500 due to limited GPU memory. PyG's "NeighborLoader" was used to handle the neighborhood sampling within minibatches, where 20 neighbors of each edge-type were sampled for 3 iterations. The ADAM optimizer was used during gradient descent with an initial learning rate of $\alpha = 0.001$ for 50 epochs with early stopping⁴ (Kingma and Ba 2014).

Clustering

Once tweets are represented in a continuous space through DTI, they can be clustered with a variety of clustering algorithms. Tweet clusters, then, are the discrete contexts that conversational networks can be studied within. Given the size of Twitter datasets k-Means clustering is one of the only available choices though it requires a set number of clusters (Lloyd 1982). The "elbow method" heuristic is a useful way of selecting this number, wherein the mean cluster distance is plotted against the number of selected clusters, and the "elbow" or point of diminishing returns is selected (Thorndike 1953). The number of clusters can also be determined externally. Here, we will use the number of clusters obtained through a hand-annotation of the data explained in the following section. For smaller datasets, methods like HDBSCAN are a better choice since they can find the optimal number of clusters (McInnes, Healy, and Astels 2017; McInnes and Healy 2017).

Results

Model Selection

We have detailed 4 levels of design choices: directed vs. undirected edges, GraphSage vs. GAT aggregation, and mean vs. set2vec summarization. This leads to 8 possible configurations for our model. We evaluate these configurations based on their ability to capture the relationships in the heterogeneous conversation network. This ability can be quantitatively evaluated by first calculating the cosine similarity of neighbors in the network. Then, non-edge pairs are sampled and the cosine similarity of these pairs is calculated. Finally, the average difference is taken from edge pair similarity and non-edge pair similarity; the higher the difference the better the model. This task quantifies how well the firstorder neighbors of the heterogeneous network are captured by the embeddings, which is necessary to construct the conversation networks detailed in Figure 1.

Results are shown in Table 1. The best performing model is the directed network aggregated with Graph-Attention and summarized with Set2Vec. We select this as the model going forward. This model is clearly the top performer for the most important relationship, tweet-tweet. At the same time, it performs reasonably well in capturing the relationship between Tweets and the hashtags and URLs they use.

Validity of Embeddings

To answer RQ1 with a "yes," the embeddings from Deep Tweet Infomax must be validated. The model selection results of the previous section provide an initial test of validity. For all edge-types, the embeddings clearly differentiate edges from non-edges, as evidenced by their positive scores. Thus, the structure of the conversational network is wellcaptured in the embeddings.

We further demonstrate the validity of embeddings in two ways. First, we use a simplistic data annotation scheme and see how 5 categories of tweets fall within the embedded space, where we find that the clusters in the tweetembedding space are well-correlated with the annotated groups. Then, we perform a nearest-neighbors analysis and confirm that that the nodes closest in the embeddings space are similar, even when comparing Tweets across languages.

Validation With Hand-Labeled Data The 100 URLs receiving the most cumulative retweets in our dataset were hand-labeled with the story that they pertain to, as were the 100 Tweets with the most likes were hand annotated according to their conversational context. This procedure resulted in 38 conversational contexts, the top 5 of which were Claims of Fraud, Spam, Election Updates Biden Campaign and Trump Campaign. Spam is seen as a popular category because spammers often tweet the same URL many times. In Claims of Fraud, users discussed false accusations about election Fraud carried out by Democrats. Election Updates included live dashboards and other news breaking about the election process. The respective campaign discussions included advertisements, endorsements, and information about the candidate's platforms. The 6th context, which was used to replace Spam, was Vote Info, where information about how to vote was discussed.

The labeled tweets and URLs were then used to annotate tweets through a 2-step label propagation. This process assumes that tweets using a URL are part of the discussion that URL refers to, and that tweets directly connected to a tweet with URL are also part of that discussion. The assumption is similar for replies. Further propagation is possible, but the assumption that tweets further and further from the initial URL should still have the same label becomes harder to justify. Tweets that have conflicting labels were not included, though these made up less than 1% of the tweets. Each labeled tweet was projected into a 2-dimensional space using t-SNE on their initial text embeddings and their final embeddings in Figures 3a and 3b, respectively, where tweets are colored by their label (Van der Maaten and Hinton 2008).

The text embedding in Figure 3a, is similar to Sia et al.'s

 $^{^{4}}$ For reference, the model could be trained in about 12 hours on an Intel E5-2687W v3 CPU.

	S-D-M	S-D-S	S-U-M	S-U-S	A-D-M	A-D-S	A-U-M	A-U-S
Tweet	0.013	-0.002	0.278	0.006	0.599	0.273	0.293	0.192
Hashtag	0.095	0.065	0.110	0.067	0.084	0.086	0.081	0.094
URL	0.248	0.152	0.249	0.158	0.216	0.215	0.185	0.223

Table 1: Model selection results. The configuration keys are as follows: S is GraphSage, A is GraphAttention, D is Directed, U is Undirected, M is Mean, and S is Set2Vec. The best results are emboldened.



Figure 3: 2-dimensional t-SNE representation of tweets before and after DTI training. These are the tweets within 2-hops of a hand-annotated URL or Tweet. The tweets are colored with their associated hand-annotation.

approach to topic modeling and is used as a baseline (Sia, Dalmia, and Mielke 2020). We observe that the text embedding is mostly unable to recover the conversational contexts we set out to find. This is likely due to the facts that these contexts have similar word distribution and that a text-only approach cannot leverage replies, hashtags, or URLs. Next, we observe in Figure 3b that there are a number of wellformed tweet clusters which correspond to different conversational contexts. We observe that some of these clusters form tight balls, almost perfect circles in the embedded space. Investigation into these regions finds that this occurs when many tweets reply or quote a popular tweet. The original tweet anchors the conversation, while the additional information in replies or quotes move these secondary quotes in different directions within the embedded space, but not far from their neighboring tweet. We also see that not all of the clusters are so simple, pointing to more interesting contextual structure.

Importantly, the majority of tweet clusters have homogeneous labels. We see many clusters with 100% label agreement labeled with *Claims of Fraud* and *Trump Campaign*. Many of the instances of *Vote Info* or *Biden Campaign* tweets are also surrounded by *Claims of Fraud*. This follows from the observation that many tweets about the election or statements in support of Biden were replied or quoted with lies about voter fraud and the Democrat's efforts to steal the election.

Hashtag 1	Hashtag 2
#bidencrimesyndicate	#laptopfromhell
#jewsfortrump	#womenfortrump
#wethepeople	#wwg1wgaworldwide
#japanisready*	#returnoftheusa
#presidenttrumpwins	#returnoftheusa

Table 2: Pairs of hashtags that are closest in the embedded space. Only the top 500 hashtags are considered, and perfect matches are excluded. #japanisready was translated from Japanese.

Nearest-Neighbor Analysis We list the nearest neighbors in the embedding space for each nodeset in Tables 2, 3, and 4. In the case of Tweets, we show the neighbors for offlanguage pairs to highlight the method's ability to work in the multi-lingual setting

We observe that the hashtag pairs have similar meaning or usage. Both #bidencrimesyndicate and #laptopfromhell are anti-Biden hashtags. Similarly, #jewsfortrump and #womenfortrump express support form Trump from groups outside his base. The hashtags #japanisready, #returnoftheusa, and #presidenttrumpwins, stem from a bot-driven conversation mostly in Japanese falsely stating that Trump has won the election and that Japan is welcoming this result.

The nearest URLs are also extremely similar. The closest URLs are the NBC News election dashboards for Ari-

URL 1	URL 2
https://www.nbcnews.com/politics/2020-elections/	https://www.nbcnews.com/politics/2020-elections/
arizona-results	arkansas-results
https://www.nbcnews.com/politics/2020-elections/	https://www.nbcnews.com/politics/2020-elections/
california-results	arkansas-results
https://www.nbcnews.com/politics/2020-elections/	https://www.nbcnews.com/politics/2020-elections/
california-results	arizona-results
https://www.vote.org/polling-place-locator/	https://vote.gop/
https://www.newsweek.com/why-i-will-vote-trump-	https://vote.gop/
opinion-1543803	

Table 3: Pairs of URLs that are closest in the embedded space. Only the top 500 URLs are considered, and perfect matches are excluded.

Tweet
QT: BREAKING: Michigan USPS Whistleblower Details Directive From Superiors: Back-Date Late Mail-
In-Ballots As Received November 3rd, 2020 So They Are Accepted "Separate them from standard letter
mail so they can hand stamp them with YESTERDAY'S DATE & put them through" #MailFraud
this is outright voter fraud. twitter will no doubt block direct video evidence.
where is the doj???
this is outright voter fraud. twitter will no doubt block direct video evidence.
President Trump needs to talk about this. Game changer.
election fraud alleged by whistleblower in michigan. btw, do we still have a justice department?

Table 4: Pairs of different-language tweets that are closest in the embedded space. All shown tweets are closest to the following text translated from Japanese using Google Translate: "Breaking: At a post office in Michigan, a bureau clerk said, Whistleblowing. My boss instructed me to postmark the ballot that arrived at the post office today with yesterday's date. In this regard, it looks like an investigation will begin. As soon as I called my boss, I was cut off."

zona, Arkansas, and California. Clearly all of these URLs are very similar and fall under the *Election Updates* conversation. The next pair are two URLs giving information about how to find your polling place, one partisan and one not. Lastly, we see the pro-GOP voting information link is neighbors with a Newsweek editorial on why someone is voting for Trump. These links were shared in close proximity as Trump supporters attempted to convince others to vote for him and provided information on how to do so.

Lastly, the closest pairs of different language tweets in the embedding space are given in Table 4. Off-language pairs were chosen to highlight the method's ability to work in the multi-lingual setting, while the overall neighbors are omitted given space restrictions. All of the pairs included one Japanese tweet claiming there was fraud. The closest neighbor repeats the claim in English. The following neighbors are direct replies to the claim. These examples show that similar tweets across language are given similar representations and that conversational structure is captured.

User Overlap Across Conversational Contexts

We now turn to RQ2. Users are active within a conversational context if any of their Tweets or Retweets are clustered in that context. We calculate the pairwise percentage of overlap in membership and plot the results in Figure 4. Overall, there is low (25-35%) overlap of active users between clusters. There are exceptional pairs of contexts with 40-55% overlap, which is still low.



Figure 4: Pairwise overlap of active users in the conversational contexts. The diagonal is set to 0% for readability.

This finding has important ramifications for conversational network analysis. The presence of non-overlapping contexts highlight that *global* properties of conversational networks are being affected by context. Placing users from the first cluster of contexts in the same network as those in the second is a misleading representation of the data. It is possible that users from these different contexts may even be placed in the same component of a decontextualized network. As the number of active users increases, it becomes more likely that the two contexts will be merged into a single component under decontextualized analysis.



Figure 5: Pairwise Kendall-Tau correlation of user PageRank in the conversational contexts, accounting for differences in nodesets.

More importantly, there are contexts with low but not negligible overlap, around 15%. This means that the *local* properties of the de-contextualized network are affected. With 15% overlap, we can expect that about 15% of users will have connections from users in both contexts with no way of distinguishing them. This has negative effects on every aspect of network analysis. Path-based centralities, for example, will be calculated on paths that could not occur in the data because they span contexts. The impact of this is further studied in the following section.

Status Correlation Conversational Contexts

Finally, we address RQ3. From the previous analysis, we have seen that contextualized networks are very different than uncontextualized nodesets. The most basic property, the nodeset, has extreme variation from one context to the next. However, for many social media network studies, the users with high status are of primary concern. For example, two conversations may have many different users in them, but if they both have the same 5 influencers dominating the discussion the other differences may not matter. We now ask how users' social status varies between conversations.

To answer our third research question, we measure the correlation of the social status ranking of users across conversations. If correlation is low, then we can again say that contextualized analysis provides different results than of non-contextualized analysis.

To measure status correlation, we first control for differences in nodesets. For a pair of conversations, we construct the conversational sub-graphs only considering users that are active in both conversations. From there, PageRank centrality is used to quantify user status in both contexts (Page et al. 1999). Lastly, the correlation of status rankings is quantified using Kendall-Tau correlation. Non-significant results (those with p < 0.05) were set to 0.

The pairwise correlations are shown in Figure 5. The mean correlation between contexts is 0.472. As we would expect, users who are important in one conversation tend to be important in the next. However, with a correlation below 0.5 and with many pairwise instances below 0.5, this is far from a one-to-one mapping. This difference can be es-

pecially important for those performing qualitative analysis on the top n users according to their status. With correlations below 0.5, the top n users in the uncontextualized setting may vary significantly from the important users in the conversations of interest to the researcher. We controlled for the vastly different nodesets between contexts but comparison of the full networks would likely lead to much lower correlations, due to the lack of robustness of centrality measures (Borgatti, Carley, and Krackhardt 2006). This is to say that status varies from one conversation to the next and performing a contextualized analysis is the only way of fully accounting for this variation.

Discussion

Tweet representation through Deep Tweet Infomax appears to be a well-validated approach to clustering Tweets into their respective conversations. This validation stems from a 3-part analysis, where tweets, URLs, and hashtags were all tested. The embeddings were shown to capture all three edge relationships in the heterogeneous social graph. From there, the embedding clusters were demonstrated to agree with hand-labeled data of conversational contexts. Lastly, examples of multi-lingual nearest neighbors showed that Tweets with similar content across languages are given similar representations. With this validation in mind, we move to discussion of the network results.

When considering both the overlap in active users and the correlation of status between conversational contexts, we note that there appears to be two higher-level clusters in the dataset. The conversations within these clusters have slightly higher status correlation and user overlap than those between clusters. The presence of these clusters implies the existence of hierarchical conversational contexts. Prior work in story detection, which the present work builds on, have also acknowledged the hierarchical nature of stories (Srijith et al. 2017). Here, we identify 2-layers of this hierarchy, which is of use when qualitatively analyzing the data within these contexts. Future work that explicitly models the hierarchical nature of conversational contexts is of interest.

Next, we find that combined network analysis of multiple contexts severely corrupts measurements of important actors. The average correlation of status between different conversations is below 0.5. So, while important users in one conversation tend to be important in others, there are many exceptions to this trend. The conditions under which users carry status across contexts is of interest in future work. The development of centrality measures which account for node position within and between contexts could quantitatively address this question. These may be developed in a similar manner to community-aware centrality measures, which account for node position within and between communities (Magelinski, Bartulovic, and Carley 2021; Rajeh et al. 2021).

We have demonstrated that each conversation in a social media dataset may have unique and important conversational structure. To understand the entire dataset, it is clear that the information given by these networks is complimentary. For example, understanding the active and important users in the *Claims of Fraud* and *Trump Campaign* discussions help us understand the overall dataset. Methods that explicitly leverage this complimentary information to answer questions about the overall conversational network structure are of interest for future work.

Limitations and Future Work

There are a number of limitations to this work. First, the data was collected using a keyword-based collection, resulting in only a sample of the true conversational network. Twitter's new API provides functionality to build out the full conversational graph after the initial sample is collected⁵. This is a powerful data collection tool that will be leveraged in future work; obtaining the full conversational network will provide more edges for information to flow between.

Additionally, the initial feature representations of tweets are derived from a relatively simple language embedding scheme which does not include attached images or video. The scheme was selected due to its scalability and its ability to embed tweets written in different languages within the same vector space. This approach embedded tweets from 11 languages, but 4% of the reachable tweets were still not reached. The lack of image or video representation is the more important limitation, particularly because many tweets with images or video do not have text. While this is largely the case for replies and not original tweets, the full space is affected due to the transfer of information from reply to base tweet and vice-versa. Even though a pre-trained model can be used to obtain image or video representations, the process of including this information in initial tweet feature representation is unclear. Most tweets lack images, so feature concatenation will not be effective. Combining the feature vectors is also not straightforward because the vector spaces are not aligned. A process which gives a feature representation of both text and images is left for future work.

All tweets are treated equally in current methods, however, social signals such as the number of retweets or favorites a tweet gets could inform more sophisticated aggregation schemes for GNNs. This is left for future work. Further, methods which incorporate URLs domain name could improve embeddings but are also left for future work.

Another limitation of the current analysis is the lack of mention representation. Mentions are a core feature of Twitter, allowing for users to directly tag other users in their Tweets. Incorporating mention nodes into Deep Tweet Infomax should improve tweet representation, since mentions are so commonly used to tag major players in a discussion. This was not done in the present work because of the quality of the data available. In the first version of Twitter's API, replying to a tweet adds a "mention" of the user that is being replied to, and often adds "mentions" to several other user higher in the conversation tree. These are not actual "mentions," just artifacts of the already modeled tweet-tweet graph, so their inclusion could harm our results. This problem is resolved in the new API. Future work applying DTI to datasets collected with the new API will model mentions in a similar way to hashtags and URLs.

The last major limitations of the approach are that the method still derives *discrete* conversational contexts and that these are compared with noisy human annotations. Specifically, the fact that our annotations were given by a single annotator poses a limitation. Next, we have seen that interactions can be represented as occurring in a *continuous* context space. However, all existing network approaches assume discrete context spaces. Given the appropriate methods, the continuous context space could be used to measure things such as conversational drift and contextual persistence of links. Thus, methods directly operating in continuous space are of interest in future work.

Broader Perspectives and Ethics

Contextualizing data allows for more accurate representation of user's importance within a discussion. Social media analysis can have high stakes when it is used to determine the importance, or presence, of users within information operations. While this work moves closer to properly attributing users to the conversations that they are actually active in, there is a question of interpretability. The move towards deep graph neural networks makes interpretability challenging. While the initial layer of our model is easy to interpret, this becomes more challenging as layers are stacked. We have tried to validate that our model is appropriately representing the data by checking even the intermediate node representations of hashtags and URLs, but if this work is to be applied to qualitative work looking to attribute accounts a high-stakes setting, much more in-depth checks about how specific users fit into a conversation must be taken.

Conclusion

We provide a method of contextualizing conversational networks on Twitter. This method represents tweets in a vector space using their text, hashtags, URLs, and the conversational network. Vectorized tweets are then clustered into conversational contexts. We apply our approach to a dataset of 4.5 million tweets and validate the results through inspection of nearest-neighbors in the embedded space, and by comparison with a label propagation procedure.

Conversational contexts have been shown to have low overlap in user participation. Thus, distinguishing between contexts allows for more accurate analysis about which users are participating in the same discussions. Further, we see that it allows for more accurate analysis of which users are important within conversations, as contextualized networks are then demonstrated to have different central actors. This points to an area of future work: quantification of user importance within and between conversational contexts.

While we analyzed Twitter data, the approach is readily extensible to other mediums. We hope that this work has demonstrated the importance of uncovering the context in which social connections are made online and will spark future work both in detecting and understanding the implications of such contexts.

⁵https://blog.twitter.com/developer/en_us/topics/tools/2020/ introducing_new_twitter_api

References

Alshaabi, T.; Adams, J. L.; Arnold, M. V.; Minot, J. R.; Dewhurst, D. R.; Reagan, A. J.; Danforth, C. M.; and Dodds, P. S. 2021a. Storywrangler: A massive exploratorium for sociolinguistic, cultural, socioeconomic, and political timelines using Twitter. *Science advances*, 7(29): eabe6534.

Alshaabi, T.; Arnold, M. V.; Minot, J. R.; Adams, J. L.; Dewhurst, D. R.; Reagan, A. J.; Muhamad, R.; Danforth, C. M.; and Dodds, P. S. 2021b. How the world's collective attention is being paid to a pandemic: COVID-19 related n-gram time series for 24 languages on Twitter. *Plos one*, 16(1): e0244476.

Alvarez-Melis, D.; and Saveski, M. 2016. Topic modeling in twitter: Aggregating tweets by conversations. In *Tenth international AAAI conference on web and social media*.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3: 993–1022.

Bojanowski, P.; Grave, E.; Joulin, A.; and Mikolov, T. 2017. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5: 135–146.

Borgatti, S. P.; Carley, K. M.; and Krackhardt, D. 2006. On the robustness of centrality measures under conditions of imperfect data. *Social networks*, 28(2): 124–136.

Brody, S.; Alon, U.; and Yahav, E. 2021. How Attentive are Graph Attention Networks? *arXiv preprint arXiv:2105.14491*.

Cao, S.; Lu, W.; and Xu, Q. 2016. Deep Neural Networks for Learning Graph Representations. In *AAAI*.

Carley, K. M. 2020. Social cybersecurity: an emerging science. *Computational and mathematical organization theory*, 26(4): 365–381.

Carley, K. M.; Cervone, G.; Agarwal, N.; and Liu, H. 2018. Social cyber-security. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 389–394. Springer.

Chen, H.; Perozzi, B.; Hu, Y.; and Skiena, S. 2017. HARP: Hierarchical Representation Learning for Networks. *arXiv:1706.07845 [cs]*. ArXiv: 1706.07845.

Cheng, X.; Yan, X.; Lan, Y.; and Guo, J. 2014. Btm: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering*, 26(12): 2928–2941.

Do, T. H.; Nguyen, D. M.; Tsiligianni, E.; Cornelis, B.; and Deligiannis, N. 2018. Twitter user geolocation using deep multiview learning. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 6304–6308. IEEE.

Dodds, P. S.; Minot, J. R.; Arnold, M. V.; Alshaabi, T.; Adams, J. L.; Reagan, A. J.; and Danforth, C. M. 2021. Computational timeline reconstruction of the stories surrounding Trump: Story turbulence, narrative control, and collective chronopathy. *PloS one*, 16(12): e0260592. Feng, W.; Zhang, C.; Zhang, W.; Han, J.; Wang, J.; Aggarwal, C.; and Huang, J. 2015. STREAMCUBE: Hierarchical spatio-temporal hashtag clustering for event exploration over the Twitter stream. In 2015 IEEE 31st international conference on data engineering, 1561–1572. IEEE.

Fey, M.; and Lenssen, J. E. 2019. Fast graph representation learning with PyTorch Geometric. *arXiv preprint arXiv:1903.02428*.

Grinberg, N.; Joseph, K.; Friedland, L.; Swire-Thompson, B.; and Lazer, D. 2019. Fake news on Twitter during the 2016 US presidential election. *Science*, 363(6425): 374–378.

Grover, A.; and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 855–864.

Hamilton, W. L.; Ying, R.; and Leskovec, J. 2017. Inductive representation learning on large graphs. In *Proceedings* of the 31st International Conference on Neural Information Processing Systems, 1025–1035.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, 1026–1034.

Hong, L.; and Davison, B. D. 2010. Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics*, 80–88.

Joulin, A.; Bojanowski, P.; Mikolov, T.; Jégou, H.; and Grave, E. 2018. Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion. In *Proceedings* of the 2018 Conference on Empirical Methods in Natural Language Processing.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Lazer, D. M.; Baum, M. A.; Benkler, Y.; Berinsky, A. J.; Greenhill, K. M.; Menczer, F.; Metzger, M. J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. 2018. The science of fake news. *Science*, 359(6380): 1094–1096.

Lerer, A.; Wu, L.; Shen, J.; Lacroix, T.; Wehrstedt, L.; Bose, A.; and Peysakhovich, A. 2019. Pytorch-biggraph: A large scale graph embedding system. *Proceedings of Machine Learning and Systems*, 1: 120–131.

Lloyd, S. 1982. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2): 129–137.

Magelinski, T.; Bartulovic, M.; and Carley, K. M. 2020. Canadian Federal Election and Hashtags that Do Not Belong. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 161–170. Springer.

Magelinski, T.; Bartulovic, M.; and Carley, K. M. 2021. Measuring Node Contribution to Community Structure with Modularity Vitality. *IEEE Transactions on Network Science and Engineering*, 8(1): 707–723. Magelinski, T.; and Carley, K. M. 2019. Community-based time segmentation from network snapshots. *Applied Network Science*, 4(1): 25.

Masuda, N.; and Holme, P. 2019. Detecting sequences of system states in temporal networks. *Scientific reports*, 9(1): 1–11.

McInnes, L.; and Healy, J. 2017. Accelerated Hierarchical Density Based Clustering. In *Data Mining Workshops* (*ICDMW*), 2017 *IEEE International Conference on*, 33–42. IEEE.

McInnes, L.; Healy, J.; and Astels, S. 2017. hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11): 205.

Nguyen, T. T.; Singh, K.; Verma, S.; Wadhwa, H.; Vimal, S.; Dheekollu, L.; Lui, S. J.; Gupta, D.; Jin, D. Y.; and Wei, Z. 2020. Word and Graph Embeddings for COVID-19 Retweet Prediction. In *CIKM AnalytiCup 2020*, 5–8.

Osborne, M.; Petrovic, S.; McCreadie, R.; Macdonald, C.; and Ounis, I. 2012. Bieber no more: First story detection using Twitter and Wikipedia. In *Sigir 2012 workshop on time-aware information access*, 16–76. Citeseer.

Page, L.; Brin, S.; Motwani, R.; and Winograd, T. 1999. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Pan, S.; and Ding, T. 2019. Social Media-based User Embedding: A Literature Review. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 6318–6324.

Peixoto, T. P.; and Gauvin, L. 2018. Change points, memory and epidemic spreading in temporal networks. *Scientific reports*, 8(1): 1–10.

Peixoto, T. P.; and Rosvall, M. 2017. Modelling sequences and temporal networks with dynamic community structures. *Nature communications*, 8(1): 1–12.

Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 701–710.

Petrović, S.; Osborne, M.; and Lavrenko, V. 2010. Streaming first story detection with application to twitter. In *Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics*, 181–189.

Rajeh, S.; Savonnet, M.; Leclercq, E.; and Cherifi, H. 2021. Characterizing the interactions between classical and community-aware centrality measures in complex networks. *Scientific Reports*, 11(1): 1–15.

Ren, Y.; Liu, B.; Huang, C.; Dai, P.; Bo, L.; and Zhang, J. 2019. Heterogeneous deep graph infomax. *arXiv preprint arXiv:1911.08538*.

Ribeiro, M. H.; Calais, P. H.; Santos, Y. A.; Almeida, V. A.; and Meira Jr, W. 2018. Characterizing and detecting hateful users on twitter. In *Twelfth international AAAI conference on web and social media*.

Rossi, E.; Kenlay, H.; Gorinova, M. I.; Chamberlain, B. P.; Dong, X.; and Bronstein, M. M. 2022. On the unreasonable effectiveness of feature propagation in learning on graphs with missing node features. In *Learning on Graphs Conference*, 11–1. PMLR.

Sia, S.; Dalmia, A.; and Mielke, S. J. 2020. Tired of Topic Models? Clusters of Pretrained Word Embeddings Make for Fast and Good Topics too! In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1728–1736. Online: Association for Computational Linguistics.

Srijith, P.; Hepple, M.; Bontcheva, K.; and Preotiuc-Pietro, D. 2017. Sub-story detection in Twitter with hierarchical Dirichlet processes. *Information Processing & Management*, 53(4): 989–1003.

Steinskog, A.; Therkelsen, J.; and Gambäck, B. 2017. Twitter topic modeling by tweet aggregation. In *Proceedings of the 21st nordic conference on computational linguistics*, 77– 86.

Thorndike, R. 1953. Who belongs in the family? *Psychometrika*, 18(4): 267–276.

Uyheng, J.; Magelinski, T.; Villa-Cox, R.; Sowa, C.; and Carley, K. M. 2019. Interoperable pipelines for social cybersecurity: assessing Twitter information operations during NATO Trident Juncture 2018. *Computational and Mathematical Organization Theory*, 1–19.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *arXiv* preprint arXiv:1710.10903.

Velickovic, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep Graph Infomax. *ICLR* (*Poster*), 2(3): 4.

Vinyals, O.; Bengio, S.; and Kudlur, M. 2015. Order matters: Sequence to sequence for sets. *arXiv preprint arXiv:1511.06391*.

Walther, M.; and Kaisser, M. 2013. Geo-spatial event detection in the twitter stream. In *European conference on information retrieval*, 356–367. Springer.

Wang, D.; Cui, P.; and Zhu, W. 2016. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 1225–1234.

Wang, Y.; Liu, J.; Qu, J.; Huang, Y.; Chen, J.; and Feng, X. 2014. Hashtag graph based topic model for tweet mining. In *2014 IEEE International Conference on Data Mining*, 1025–1030. IEEE.

Zuo, Y.; Wu, J.; Zhang, H.; Lin, H.; Wang, F.; Xu, K.; and Xiong, H. 2016. Topic modeling of short texts: A pseudo-document view. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2105–2114.