Understanding Political Polarization via Jointly Modeling Users, Connections and Multimodal Contents on Heterogeneous Graphs

Hanjia Lyu
University of Rochester
hlyu5@ur.rochester.edu

Jiebo Luo
University of Rochester & Meta AI
jlhuo@cs.rochester.edu

ABSTRACT
Understanding political polarization on social platforms is important as public opinions may become increasingly extreme when they are circulated in homogeneous communities, thus potentially causing damage in the real world. Automatically detecting the political ideology of social media users can help better understand political polarization. However, it is challenging due to the scarcity of ideology labels, complexity of multimodal contents, and cost of time-consuming data collection process. Most previous frameworks either focus on unimodal content or do not scale up well. In this study, we adopt a heterogeneous graph neural network to jointly model user characteristics, multimodal post contents as well as user-item relations in a bipartite graph to learn a comprehensive and effective user embedding without requiring ideology labels. We apply our framework to online discussions about economy and public health topics. The learned embeddings are then used to detect political ideology and understand political polarization. Our framework outperforms the unimodal, early/late fusion baselines, and homogeneous GNN frameworks by a margin of at least 9% absolute gain in the area under the receiver operating characteristic on two social media datasets. More importantly, our work does not require a time-consuming data collection process, which allows faster detection and in turn allows the policy makers to conduct analysis and design policies in time to respond to crises. We also show that our framework learns meaningful user embeddings and can help better understand political polarization. Notable differences in user descriptions, topics, images, and levels of retweet/quote activities are observed. Our framework for decoding user-content interaction shows wide applicability in understanding political polarization. Furthermore, it can be extended to user-item bipartite information networks for other applications such as content and product recommendation.

CCS CONCEPTS
• Information systems → Social networks; • Applied computing → Sociology.

KEYWORDS
political polarization, multimedia, heterogeneous graph, user-content interaction

ACM Reference Format:
Hanjia Lyu and Jiebo Luo. 2022. Understanding Political Polarization via Jointly Modeling Users, Connections and Multimodal Contents on Heterogeneous Graphs. In Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3503161.3547898

1 INTRODUCTION
A 2016 Pew Research study [14] found that of all U.S. adults, 67% use social media platforms with 44% using the platforms to discover news. Social media is found to shape political discourse in the U.S. and the whole world [8, 37, 45, 53]. The extent to which the opinions on an issue are opposed is termed “political polarization” [9]. The formation of such political polarization is not necessarily a serious problem, but the concern is that the opinions may become increasingly polarized when they are shared, viewed and discussed in a homogeneous community [8]. For example, on March 16, 2020, the former U.S. President Donald Trump posted a tweet calling COVID-19 “Chinese Virus”.1 After that, this term was frequently used among the right-leaning Twitter users, and in most cases, they were carried with negative sentiment against Chinese and Asian people [4, 32]. More importantly, the harm caused by biased and false news has a substantial negative impact across the entire society [53]. Fake or extremely biased news regarding COVID-19 vaccines is found to be negatively correlated with the state-level COVID-19 vaccination rate [35]. The growing political divide in the U.S. also extends to the #StopAsianHate movement with Trump’s followers being less supportive to the Asian-American victims of the racially motivated hate crimes [33]. Therefore, understanding political polarization on social platforms is important.

To better understand political polarization on social platforms, we also need to measure political ideology which refers to the political stance of people [55]. It is noteworthy that political ideology is a complicated concept and may often be used at multiple levels [10]. Similar to previous studies [6, 7, 55], we frame the stance detection problem as ideology detection. By identifying people’s ideology, we would be able to investigate questions such as what opinions they hold or what characteristics they share. There are two major challenges in estimating political ideology in real-world applications. First, unlike some politicians who explicitly show their ideology, the political ideology of most ordinary citizens remains unknown or hidden. Although social platforms such as Twitter
We apply our framework to user-tweet bipartite graphs on Twitter, which makes it possible to learn the ideology of the users who participate in political discussions and understand their opinions. Notable differences in user descriptions, topics, images, and levels of political polarization are investigated. Our framework leverages the multimodal post contents to tackle the ideology detection problem. Recent studies have shown promising performance in predicting people’s ideology by modeling the relations (e.g., follow) among users in a heterogeneous network. Our framework also builds upon a heterogeneous network design. However, our approach fundamentally differs from prior work [55] in that (1) our framework takes in both unimodal and multimodal information, (2) considers only one type of node (i.e., user node), (3) models user-item relations and user-user relations while collecting user-user relations requires an extra data collection process, and (4) is trained with political ideology labels, while their approach (1) takes in unimodal information, (2) considers only one type of node (i.e., user node), (3) models user-item relations and user-user relations while collecting user-user relations requires an extra data collection process which is time-consuming, and (4) is trained with political ideology labels.

2 RELATED WORK

2.1 Political ideology detection

Social platforms have rich multimodal information. Most studies rely on textual information to detect the political ideology of social media users [6, 7, 41]. However, the unimodal information is usually noisy on social media [25] and can hinder the detection performance. Inspired by the findings of previous work that multimodal information is helpful when the unimodal signal is noisy [18, 51, 57], our framework leverages the multimodal post contents to tackle the ideology detection problem. Our framework also builds upon a heterogeneous network design. However, ours differs from Xiao et al. [55] fundamentally as our framework (1) takes in both unimodal and multimodal information, (2) considers only one type of node (i.e., user node), (3) models user-item relations and user-user relations while collecting user-user relations requires an extra data collection process, and (4) is trained with political ideology labels, while their approach (1) takes in unimodal information, (2) considers only one type of node (i.e., user node), (3) models user-item relations and user-user relations while collecting user-user relations requires an extra data collection process which is time-consuming, and (4) is trained with political ideology labels.

2.2 Political polarization understanding

Most prior works on learning political polarization on social platforms focus on user behavior and unimodal contents [3, 8, 16, 17, 21, 46, 49]. Borge-Holthoefer et al. [3] analyze the Egyptian political polarization on Twitter by a set of manually labeled hashtags and the retweet network constructed by hand-labeled seed users. Hosseini et al. [21] identify several distinct communities of news consumers, including “far-right” and “anti-woke” by examining the consumption of radical content on YouTube. Conover et al. [8] investigate how social media shape the networked public sphere and facilitate communication between communities with different political orientations. In particular, they use the 2010 U.S. congressional midterm elections as a case study and focus on the retweet and mention networks. Waller and Anderson [49] quantify the online political polarization using community-level embeddings. In contrast, our individual-level embeddings learned by jointly modeling user characteristics, multimodal contents, and user-item relations allow us to conduct a more fine-grained analysis.

2.3 Graph neural networks

An example of different node types and relation types on Twitter is illustrated in Figure 1. The network is composed of two types of nodes - users and tweets. The relations between a user and a tweet include post, retweet, quote, etc. A user can follow another user and they can also follow each other. The tweet node has textual information and sometimes it has other media information such as explicit.
video, image, and link. The user node has user characteristics such as number of followers, verified status, and profile description. The heterogeneity of information and complex relations among the nodes make it challenging to uncover insight. Recently, representing graph nodes in a low dimensional vector space based on factorization methods, random walks, and deep learning has been proposed to address these issues [15]. Graph neural networks have shown promising results in multiple applications including medical image computing [29], rumor detection [23, 30], sentiment analysis [22, 50], and recommender systems [44]. More specifically, multiple heterogeneous GNN architectures have been proposed to address the challenge of heterogeneity [31, 58]. Motivated by the need to handle heterogeneity, we adopt a heterogeneous GNN architecture to effectively fuse the multimodal features of heterogeneous nodes and explicitly model the user-item relations in social networks. We then apply our framework to ideology detection and polarization understanding tasks on social platforms.

The objects collected are tweets. Each tweet is associated with multiple fields such as tweet ID, tweet content, and the information of the user who interacts with it (e.g., post, retweet, quote), including user ID, and user’s social capital (e.g., number of followers). Moreover, if a user A retweets/quotes the tweet of another user B, then the tweet information of the original tweet and the user characteristics of the original user (i.e., B) will also be associated with this retweet/quote. Hence, the retweet/quote relations are implied during the data collection process. In contrast, user-user following relations are not obtained during this phase. There are two challenges in collecting them. First, an extra data collection process is required, which is prohibitively time-consuming. On Twitter, the number of users a user is following is referred to as the number of friends. The average number of friends of the users in the inflation and vaccine datasets are 4,688 (SD = 12,411) and 7,822 (SD = 17,789). Using the Twitter API, it takes 9 days and 14 hours to collect all the user-user following relations for the inflation dataset, and 3 days and 21 hours for the vaccine dataset.

As of February 2022, there are millions of daily active users on Twitter. It is difficult to collect all the user-user following relations of the users-of-interest in a short period. Second, following relations are sometimes unavailable due to the privacy settings of Twitter users. Therefore, to increase the applicability of our framework and to conduct downstreaming tasks such as ideology detection and polarization analysis in a more timely manner, we do not crawl user-user following relations.

Figure 1: An example of different node types and relation types on Twitter. It is particularly noteworthy that our framework does not need user-user following relations. Collecting the user-user following relations requires an extra time-consuming process. Consequently, without the following relations, the general heterogeneous network is effectively reduced to a bipartite heterogeneous graph.

### 3 MATERIAL AND METHOD

#### 3.1 Datasets

Politically polarized opinions have been observed in online debates with respect to multiple areas such as economy and public health [34]. We employ two large-scale Twitter datasets that record such online discussions to demonstrate the effectiveness of our framework in learning user embeddings without any human annotated ideology labels. For simplicity, we refer to these two datasets as the inflation dataset and vaccine dataset in the remainder of this paper. Table 1 shows the descriptive statistics of the datasets. The total size (N = 11,038) of these two datasets is significantly larger than most public datasets for ideology detection (e.g., N = 213 for Chen et al. [6], N = 2,976 for Xiao et al. [55]). For more details on data collection, please refer to Appendix A where we also describe how we obtain the pseudo “ground truth” political ideology of these social media users. Note that these “ground truth” labels are only used for evaluation and not used during representation learning.

| Dataset     | # unique users | # unique tweets | # tweets |
|-------------|----------------|----------------|---------|
| inflation   | 8,824          | 22,661         | 42,297  |
| vaccine     | 2,214          | 10,998         | 20,331  |

#### 3.2 Graph Representation Learning

We model the user characteristics, multimodal contents, and the user-item relations in a bipartite heterogeneous graph. We discuss our framework in the following sections.

3.2.1 Graph definition. In the graph $G = (V, E, X)$ of our study, $G$ denotes the graph, $V$ denotes the set of nodes, $E$ denotes the set of edges, and $X$ represents the attributes of a node. There are two types of nodes including users $V_{user} \subset V$ and tweets $V_{tweet} \subset V$ ($V_{user} \cup V_{tweet} = V$, $V_{user} \cap V_{tweet} = \emptyset$) as shown in Figure 1. An edge $e \in E$ between a user node $v_{user} \in V_{user}$ and a tweet node $v_{tweet} \in V_{tweet}$ is created if this user interacts with this tweet. In our study, the interactions between a user and a tweet include posting, retweeting and quoting. Since we do not consider the user-user following relations, there is no edge between two user nodes. The retweet and quote relations are represented as a star topology where the tweet that is retweeted/quoted by multiple users serves as the hub of the topology and is connected to these users. No edge is drawn between tweet nodes as well. As a result, the general heterogeneous network is effectively reduced to a bipartite heterogeneous graph. The attributes $X_{user} \subset X$ of the user

---

https://www.omnicoreagency.com/twitter-statistics/ [Accessed 2022-04-05]
nodes are user characteristics including the number of followers, friends, listed memberships, statuses, favorites, verification status, as well as the profile description. These publicly available user characteristics are crawled during the data collection. The attributes $X_{tweet} \subset X$ of the tweet nodes have two components: (1) multimodal contents $M_{tweet}$ and (2) the user characteristics of the original author $X_{author}$ ($X_{tweet} = (M_{tweet}, X_{user,author})$). Multimodal contents such as text including hashtags, and image (if any) are used. The user characteristics of the original author of each tweet node are also used to embed the tweet node. In this way, the user-tweet relations are modeled explicitly as each tweet is associated with a vector indicating the information of the original author.

3.2.2 Feature embeddings. Textual and visual feature embedding methods are employed to capture the semantic meanings of the user and tweet nodes. We use the state-of-the-art framework - SentenceBERT [42], to represent the user profile description and tweet text content. Similar to word embeddings [36], it converts sentences into a vector space where the vectors of two semantically similar sentences will be closer. The output vectors have 384 dimensions. Inspired by Zhang et al. [59], an image is encoded into a feature vector space using ResNet-50 [19], which is pre-trained on the ImageNet dataset [28]. The 2048-dimensional image features from the last pooling layer are extracted. In addition, if a tweet is associated with videos, they can be mapped into the feature space via different feature embedding methods. The framework is the same. In our study, we focus on text and image.

3.2.3 Representation learning. To capture the heterogeneous attributes or contents associated with each node, we adopt the Heterogeneous Graph Neural Network (HetGNN) [58] which first uses random walks with restart to generate neighbors for nodes and capture the structural information, and then leverages bi-directional LSTM (Bi-LSTM) [20] and the attention mechanism [48] to aggregate the structural information, and then leverages bi-directional heterogeneous Graph Neural Network (HetGNN) [58] which first uses random walks with restart to generate neighbors for nodes and cap-

$$f_1(v) = \frac{\sum_{i \in X_v} [LSTM(h(x_i)) \oplus LSTM(h(x_i))]}{|X_v|}$$

It is noteworthy that $h$ denotes a feature transformer which can be identity or fully connected neural network, etc.

Next, we aggregate the neighbor information of node $v$ via two steps: (1) using Bi-LSTM to aggregate embeddings of the neighbor nodes of the same type, and (2) aggregating the embeddings of the neighbor nodes of different types through the attention mechanism. By employing the random walk with restart strategy, for each node $v \in V$, we generate the user-type sampled neighbor set $N_{user}(v)$ and the tweet-type sampled neighbor set $N_{tweet}(v)$. The aggregated t-type neighbor embeddings $f_2^t(v) \in \mathbb{R}^{d \times t}$, $t \in \{user, tweet\}$ ($d$: aggregated attribute embedding dimension) are then calculated as follows:

$$f_2^t(v) = \frac{\sum_{i \in N_t(v)} [LSTM(f_i(v')) \oplus LSTM(f_i(v'))]}{|N_t(v)|}$$

where $v'$ denotes the neighbor nodes in the t-type sampled neighbor set $N_t(v)$. The output embeddings $E_v \in \mathbb{R}^{d \times 1}$ of node $v$ are computed as follows:

$$E_v = a_{0,0}f_1(v) + a_{0,User}f_{user}^v + a_{0,Tweet}f_{tweet}^v$$

where $a_{0,\cdot}$ represents the attention coefficients of (1) content embeddings of $v$ or (2) the aggregated neighbor embeddings to node $v$. These two sets of embeddings are grouped and denoted as $F(v, i) \in \mathbb{R}^{d \times 1}$ which is defined as follows:

$$F(v, i) = \begin{cases} f_1(v) & i = v \\ f_2^t(v) & i = t, \text{where } t \in \{user, tweet\} \end{cases}$$

Therefore, Eq. (3) is re-formulated as:

$$E_v = \sum_{i \in P} a_{0,i}F(v, i)$$

The attention coefficients $a_{0,i}$ are built by a single-layer feedforward neural network parametrized by a weight vector $u \in \mathbb{R}^{2d \times 1}$, with the LeakyReLU nonlinearity. To make the coefficients comparable, we normalize them across all choices of $i$ using the softmax function. Thus, the coefficients are expressed as:

$$a_{0,i} = \frac{\exp(\text{LeakyReLU}(u^T[F(v, i) \oplus F(v, j)]))}{\sum_{j \in P} \exp(\text{LeakyReLU}(u^T[F(v, i) \oplus F(v, j)]))}$$

where $\text{T}$ represents transposition and $\oplus$ is concatenation.

We do not require any ideology-related “supervision” (i.e., ideology labels) during the training process. The goal of the representation learning of our framework is to maximize the similarity between the embeddings of two nodes if they have similar attribute features and/or are geometrically closer to each other in the graph.
otherwise, minimize the similarity. Therefore, we apply the negative sampling technique [36]. The goal can be interpreted as, for each \( v \in V \), to minimize the cross entropy loss as follows:

\[
\log(p(E_{pos} \cdot E_v)) - \log(p(E_{neg} \cdot E_v))
\]

where \( \sigma \) represents the Sigmoid function, \( p_{pos} \) and \( p_{neg} \) are the positive and negative sample nodes. Zhang et al. [58] defines the positive sample nodes of the node \( v \) as the nodes that can be reached by node \( v \) via a random walk, and the negative sample nodes of the node \( v \) as any nodes in the graph. Inspired by the findings of Ying et al. [56] that importance-based neighborhoods can help improve the representation learning of GNNs, we propose a social-platform based negative sampling strategy. We call it social-platform based because it is based on our observation of the online social platforms, in this case, Twitter. We attempt to measure the likelihood of a user interacting with a tweet. If the likelihood is high, but the user never interacts with the tweet, then this is considered as an important negative pair. The social-platform based negative sampling strategy is designed to capture this kind of negative pairs.

More specifically, we sample the nodes based on the social capital features. This component is only considered when the node type of the negative sample is user. If the node type of the negative sample is tweet, we apply random sampling. The higher the normalized number of statuses a user has posted since the creation of the Twitter account, the more chance the user will be sampled. The motivation is that the number of statuses indicates the probability of a user’s posting/retweeting/quotting behavior. In other words, it indicates the probability of a user interacting with a tweet.

We acknowledge that this strategy has limitations. For example, we aggregate the number of statuses of each user since the creation of the Twitter account. This implicitly assumes that the user behavior does not change significantly over time. Although it would be better to measure the user behavior in a more fine-grained and more dynamic fashion, it is beyond the scope of this study and left to future work.

Similarly to Zhang et al. [58], the dimension of the final feature space is set to be 128. For random walk, the walk length is 30 and window size is 5.

While there are other feature embedding methods and heterogeneous graph representation learning methods that could potentially improve the expressive power of the overall representation [31], our ultimate goal is not to perfectly characterize the users, tweet contents and relations. Instead, we aim to show that jointly modeling the information in a bipartite heterogeneous graph can benefit the ideology detection task and help understand online political polarization.

4 RESULTS

4.1 Political Ideology Detection

We refer to our framework as MBPHGNN (Multimodal Bipartite Heterogeneous Graph Neural Network). To evaluate the quality of the embeddings learned via MBPHGNN, we compare the performance on a user ideology detection task on two aforementioned Twitter datasets using the learned embeddings and multiple other embeddings that are associated with the users. The “ground truth” ideology label is obtained as discussed in Appendix A. If the political score is greater than or equal to zero, the label is right-leaning (accounts for 34%/27% in inflation/vaccine), otherwise it is left-leaning (accounts for 66%/73% in inflation/vaccine). Ideology detection is a binary classification task. We expect our models to predict whether a user is left- or right-leaning. The embeddings are fed into a logistic regression classifier and a random forest classifier. For the logistic regression model, the elastic net regularization is used \((l_1\text{-}ratio = 1, C = 0.5)\). For the random forest model, the number of trees is set to 100. The performance is measured using 10-fold cross-validation.

We compare MBPHGNN with the following state-of-the-art embedding methods to demonstrate the effectiveness of our framework over unimodal frameworks, early/late fusion strategies, and homogeneous GNN frameworks. The GCN [26] and GAT [48] baselines consider both user and tweet nodes the same type. The user and multimodal information is preprocessed using the well-studied principal component analysis [54]. The InfoNCE loss [38] is used to train the GCN and GAT baselines. The embeddings we compare include:

- **User info**: The concatenation of the number of followers, friends, listed memberships, statuses, favorites, verification status and the sentence embeddings [42] of the profile description is used.
- **Textual**: Only the sentence embeddings [42] of the tweet textual content are used.
- **Visual**: Only the image embeddings [19] are used. If a tweet is not associated with an image, the visual embeddings will be filled with zeros.
- **Textual + Visual**: The concatenation of text and image embeddings is used.
- **User info + Textual + Visual**: The concatenation of user characteristics, text and image embeddings is used.
- **Homogeneous GNN**: The embeddings learned by GCN [26] and GAT [48] are used.
- **TIMME** [55]: The embeddings learned via the link prediction task of the TIMME framework are used.

In addition, we construct a Late fusion model. The predicted ideology label of each user is calculated from the weighted outcomes of User Info, Textual and Visual models. Weights are assigned based on the F1 score of each single model.

Each user may post multiple tweets. The embeddings of these tweets are fed to the classifier, respectively. The predicted ideology label of this user is voted by the predicted outcome of each tweet.

Since during representation learning, no ideology labels are provided for the frameworks, the embeddings learned by TIMME [55] are obtained via the link prediction task. In addition, TIMME does not compute the embeddings of isolated nodes. Therefore, the comparison between MBPHGNN and TIMME is conducted for the set of nodes with at least one neighbour. The experiment results are reported in separate tables.

Tables 2 and 3 show the ideology detection performance of the logistic regression and the random forest classifiers. We can observe that, our framework performs well, which achieves an overall accuracy of 84% and 93%, and outperforms other baselines by a margin of at least 9% and 11% absolute gain in AUROC on the inflation and vaccine datasets, respectively. Textual + Visual achieves a higher recall in the inflation dataset. After further investigation, we find
that it tends to assign the default class label to all users, resulting in a higher recall but a lower precision. It does not perform well in distinguishing users of different groups. F1 and AUC still suggest that the embeddings learned by MBPHGNN are overall better in ideology detection. Further, the aid of visual information by simple concatenation seems inconsistent across two datasets, which might be because concatenation cannot capture the high-level correlations among different modalities.

Although Twitter allows users to describe themselves in the profile description section, simply reading the user profile only is not reliable enough to detect political ideology as shown in the rows of User Info in Tables 2 and 3. The overall accuracy is around 65% and 84% in the inflation and vaccine datasets. The low performance of the GCN and GAT frameworks, which consider the user and tweet nodes as the same type, suggests the importance of addressing heterogeneity. Interestingly, the choice of classifier does not make much difference on the inflation dataset. On the other hand, except for MBPHGNN, the performance of random forest is worse than logistic regression on the vaccine dataset, which might be related to the number of noise variables [27], again suggesting that the embeddings learned via MBPHGNN are more robust.

Tables 4 and 5 show that MBPHGNN achieves a better performance than TIMME [55]. More importantly, we find that the prediction performance of MBPHGNN on connected nodes is better than the prediction performance on isolated nodes (e.g., Table 2 versus Table 4), indicating the effect of modeling connections.
Table 4: Results using the logistic regression model between TIMME and MBPHGNN (The best results are highlighted in bold).

| Dataset    | Representation | Accuracy ±/− 0.02 | Precision ±/− 0.02 | Recall ±/− 0.03 | F1 ±/− 0.02 | AUROC ±/− 0.02 |
|------------|----------------|------------------|-------------------|----------------|-------------|----------------|
| Inflation  | TIMME [55]     | 0.90 ±/− 0.02    | 0.86 ±/− 0.02     | 0.85 ±/− 0.03  | 0.86 ±/− 0.02 | 0.89 ±/− 0.02  |
|            | MBPHGNN        | 0.90 ±/− 0.02    | 0.87 ±/− 0.02     | 0.87 ±/− 0.04  | 0.87 ±/− 0.02 | 0.90 ±/− 0.02  |
| Vaccine    | TIMME [55]     | 0.96 ±/− 0.01    | 0.93 ±/− 0.04     | 0.92 ±/− 0.04  | 0.92 ±/− 0.03 | 0.95 ±/− 0.02  |
|            | MBPHGNN        | 0.96 ±/− 0.01    | 0.93 ±/− 0.04     | 0.93 ±/− 0.02  | 0.93 ±/− 0.02 | 0.95 ±/− 0.01  |

Table 5: Results using the random forest model between TIMME and MBPHGNN (The best results are highlighted in bold).

| Dataset    | Representation | Accuracy ±/− 0.01 | Precision ±/− 0.02 | Recall ±/− 0.03 | F1 ±/− 0.02 | AUROC ±/− 0.02 |
|------------|----------------|------------------|-------------------|----------------|-------------|----------------|
| Inflation  | TIMME [55]     | 0.90 ±/− 0.01    | 0.87 ±/− 0.02     | 0.84 ±/− 0.03  | 0.85 ±/− 0.02 | 0.88 ±/− 0.02  |
|            | MBPHGNN        | 0.91 ±/− 0.01    | 0.88 ±/− 0.02     | 0.87 ±/− 0.04  | 0.87 ±/− 0.02 | 0.90 ±/− 0.02  |
| Vaccine    | TIMME [55]     | 0.96 ±/− 0.01    | 0.94 ±/− 0.04     | 0.91 ±/− 0.05  | 0.92 ±/− 0.04 | 0.94 ±/− 0.03  |
|            | MBPHGNN        | 0.96 ±/− 0.01    | 0.95 ±/− 0.04     | 0.91 ±/− 0.03  | 0.93 ±/− 0.03 | 0.95 ±/− 0.02  |

To have a more intuitive understanding of the relations between the embeddings and political ideology, we apply t-distributed stochastic neighbor embedding method (t-SNE) which can visualize high-dimensional data by giving each data point a location in a two or three-dimensional map. t-SNE is capable of capturing the local as well as the global structures to reveal information of the presence of clusters [47]. Figure 3 shows the t-SNE outputs of the embeddings of MBPHGNN.

Figure 3: t-SNE visualization of the output of MBPHGNN (a: inflation, b: vaccine). The color indicates the political scores. The bluer the color is, the lower the political score is (more left-leaning). The redder the color is, the higher the political score is (more right-leaning). We highlight certain areas for further analysis on political polarization.

4.2 Political Polarization Understanding

In this section, we conduct a characterization study to show that MBPHGNN learns meaningful embeddings and can help better understand online political polarization. In Figures 3a and 3b, we observe that except for the clusters that are predominately blue (Area 1) or red (Area 3), there is a third cluster (Area 2) which is a mixture of the red and blue nodes. For better presentation, we highlight them in Figures 3a and 3b. For simplicity, we refer to the users of Areas 1, 2, 3 as Left, Middle, and Right, respectively.

4.2.1 Users with similar politics-related descriptions tend to be in the same cluster. To understand the differences among the users of Left, Middle, and Right, we visualize the user descriptions and the multimodal post contents that are related to the users of Left, Middle, and Right clusters of the inflation dataset. Panel (a) shows the word cloud of the user descriptions of the users in each group. The size of the word is proportional to the frequency of the appearance. Panel (b) shows the word cloud of the most popular topic of each group. The size of the word is proportional to the weight assigned by the LDA model. Panel (c) lists the representative images that are shared by the users of each group. See Appendix B for details on the word weights.
Almost all groups post images of political figures (e.g., Joe Biden, Donald Trump). In Left (lower left) and Middle (upper right), we observe more portraits, while in Right (upper left) there are more sarcasm images of Joe Biden. Interestingly, users of Right share more eye-catching images. We find users in Left share screenshots of long paragraphs and charts that explain the reasons for inflation (upper left and right in Left). Users in Middle use images of items that can reflect inflation (lower right in Middle) more often. However, users in Right use images that intend to depict inflation more vividly (lower left in Right).

4.2.4 The levels of retweet/quote activity are different. We calculate the number of users per unique tweet in each group. This value indicates the number of retweet/quote activities. The higher this value is, the more users retweet/quote the same tweet, suggesting more retweet/quote activities. There are more retweet/quote activities in the Left cluster (4.58/1.73 users per unique tweet in inflation/vaccine) and Right cluster (2.90/1.90 users per unique tweet in inflation/vaccine) than in the Middle cluster (1.03/1.02 users per unique tweet in inflation/vaccine). Fewer retweet/quote activities indicate a sparser network on which GNN normally performs poorly due to the limited knowledge gain from less representative neighbors [24].

For the inflation dataset, we also observe two sub-clusters composed of left-leaning users (highlighted in Areas 4 and 5 in Figure 3a). By analyzing the user characteristics and the multimodal contents, we find that it is mainly because of the difference in user descriptions. Unlike previous studies [8, 49] that focus on the political polarization in terms of two groups (left and right), by combining information of users, contents, and relations, we are able to conduct a more fine-grained analysis and provide insights into political polarization guided by the clusters found by our ideology-agnostic representation learning framework (i.e., no political labels during training). In our case, the three clusters are different regarding user descriptions, multimodal contents and retweet/quote activities.

5 CONCLUSION AND FUTURE WORK

We introduce an effective and efficient framework based on bipartite heterogeneous graph neural networks to decode user-content interactions by jointly modeling user characteristics, multimodal contents and user-item relations. The user embeddings learned via our framework help achieve improved performance in an ideology detection task on two social media datasets. In addition, our framework helps obtain a more fine-grained understanding of political polarization on social platforms.

There are a few limitations in our studies. Our study only focuses on the political polarization in the United States. Although the topics of the two datasets we use are diverse, the periods of the data collections are close. We can extend our study in a few directions. First, motivated by the findings of Waller and Anderson [49] that social media users became more polarized after 2016, we plan to apply our framework to the online discussions before and after 2016 of more countries (e.g., U.K., Canada, and France) to further investigate political polarization. Second, our framework is designed to decode user-content interactions on user-item bipartite information networks. We plan to apply our framework to other applications, such as content and product recommendation to users.

ACKNOWLEDGMENTS

We are grateful to Wei Xiong, Weijian Li, Wei Zhu, Dr. Yinglong Xia, Dr. Yan Zhu and Prof. Chen Ding for their constructive suggestions. This research was supported in part by a University of Rochester Research Award.
A SUPPLEMENTAL MATERIAL AND METHOD

A.1 Datasets

In our study, we employ two large-scale Twitter datasets that record online discussions on economy and public health to demonstrate the effectiveness of our framework in learning user embeddings without any human annotated ideology labels. For the economy topic, we choose to investigate the discussions about inflation during the COVID-19 pandemic as a case study. According to a Yahoo News/YouGov poll, about 90 percent of Republicans assign President Biden at least some blame for inflation while only 28 percent of Democrats do. In terms of public health topic, we focus on the discussions regarding COVID-19 vaccines, since evidence shows that there is an opinion divide in the presence of political affiliations.

This study leverages the publicly available tweets collected using the Twitter API. We use lists of keywords and hashtags to acquire tweets. To construct the inflation dataset, unlike Angelico et al. [1] who use “inflation”, “inflationary”, “price”, “expensive”, “cheap”, and “expensive bill” as the search keywords, we only use one keyword - “inflation” and one hashtag - “#inflation” to collect the tweets regarding inflation because using their list of keywords may collect many false positive tweets. For example, “Pakistanis will surely pay the price of sending terrorists to Afghanistan.” would have been collected because of the word “price”. However, it is apparently not related to the inflation issue. The dataset of Lyu et al. [34] is used as the vaccine dataset in our study. A detailed description of the data collection process can be found in Lyu et al. [34]. Additionally, our study focuses on the political polarization in the United States. Therefore, we only include the Twitter users that are located in the United States by extracting the location information from public user profiles.

A.2 Political Ideology

One of the applications of our framework is to detect the political ideology of social media users. To evaluate the performance, we need the “ground truth” labels. However, most ordinary citizens’ political ideology is unknown or hidden while human annotated labels would be costly to obtain as aforementioned. Therefore, motivated by the findings of Golbeck and Hansen [13] that Twitter users mostly follow accounts that share the same political ideology, we infer each user’s political ideology by examining the number of Twitter accounts with known political leaning that a given user follows. Several previous studies on selective exposure to political information suggest that people look for information from people with similar political views [11, 12]. The experimental results of both quantitative and qualitative validations also confirm that calculating the political scores based on the following behavior yields an accurate estimate of people’s true political leaning.

The accounts with known political leaning include media accounts, journalists and political figures. The political leanings of the media accounts and the journalists are judged and assigned by allsides.com and politico.com respectively. This set of Twitter accounts is roughly balanced with around 400 right-leaning and 400 left-leaning. Prominent politicians are excluded since Twitter users may just follow them because of their popularity and importance. Using the Twitter API, we collect the Twitter accounts that the Twitter users follow and count the number of right- and left-leaning accounts, denoted by $N_R$ and $N_L$, respectively. The political score $P_S$ is calculated as follows:

$$P_S = \frac{N_R - N_L}{N_R + N_L}$$

The political score has a range of $[-1, 1]$ with 1 meaning the most extreme right and $-1$ meaning the most extreme left. To increase the robustness of inference, only the political scores of the Twitter users who follow at least five accounts from either side (i.e., right- and left-leaning) are calculated. In this way, we calculate the political scores of 8,824 Twitter users in the inflation dataset and 2,214 Twitter users in the vaccine dataset. Figure 6 shows the distribution of political scores. More Twitter users are left-leaning, which is consistent with the findings of the Pew Research Center, i.e., Twitter users are more likely to identify as Democrats than Republicans.

A.3 Topic Inference

We apply a Latent Dirichlet Allocation (LDA) model [2] to the tweets of each group to extract the topics. The hyper-parameters are selected based on a grid search and the coherence score. By assigning the dominant topic label to each tweet, we obtain the topic distributions. We refer to the topic with the highest proportion as the most popular topic among the tweets of each group.

References

[32] Yu Wang, Jiebo Luo, and Xiyang Zhang. 2017. When follow is just one click away: Understanding twitter follow behavior in the 2016 us presidential election. In International Conference on Social Informatics. Springer, 409–425.

[33] Galen Weld, Maria G lenski, and Tim Althoff. 2021. Political bias and factuality in news sharing across more than 100,000 Online Communities. arXiv preprint arXiv:2102.08537 (2021).

[34] Svante Wold, Kim Ebensen, and Paul Geladi. 1987. Principal component analysis. Chemometrics and intelligent laboratory systems 2, 1-3 (1987), 37–52.

[35] Zhuping Xiao, Weiping Song, Haoyan Xu, Zhiceng Ren, and Yizhou Sun. 2020. Timme: Twitter ideology-detection via multi-task multi-relational embedding. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2258–2268.

[36] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William I. Hamilton, and Jure Leskovec. 2018. Graph convolutional neural networks for web-scale recommender systems. In Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 974–983.

[37] Ben P Yuhas, Moise H Goldstein, and Terrence J Sejnowski. 1989. Integration of acoustic and visual speech signals using neural networks. IEEE Communications Magazine 27, 11 (1989), 65–71.

[38] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. 2019. Heterogeneous graph neural network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 793–803.

[39] Zhongqiang Zhang, Tianlang Chen, Zheng Zhou, Jiaxin Li, and Jiebo Luo. 2018. How to become Instagram famous? Post popularity prediction with dual-attention. In 2018 IEEE international conference on big data (big data). IEEE, 2383–2392.

[40] Yu Wang, Jiebo Luo, and Xiyang Zhang. 2017. When follow is just one click away: Understanding twitter follow behavior in the 2016 us presidential election. In International Conference on Social Informatics. Springer, 409–425.
Table 7: Word frequency of the user descriptions of the users in the Left, Middle, and Right clusters in the vaccine dataset (top 10).

| Left word | count | Middle word | count | Right word | count |
|-----------|-------|-------------|-------|------------|-------|
| resist    | 192   | people      | 14    | maga       | 179   |
| mom       | 92    | blm         | 14    | trump      | 116   |
| blm       | 87    | conservative| 14    | conservative| 84    |
| love      | 71    | author      | 13    | love       | 84    |
| lover     | 68    | us          | 12    | kag        | 83    |
| trump     | 67    | maga        | 12    | god        | 68    |
| proud     | 66    | trump       | 11    | patriot    | 54    |
| theresistance | 64      | mom         | 11    | 2a         | 50    |
| bidenharris | 63     | love        | 11    | trump2020  | 46    |
| democrat  | 61    | writer      | 11    | family     | 44    |

Table 8: Top 15 keywords of the most popular topics of the Left, Middle, and Right clusters in the inflation dataset.

| Left word | weight | Middle word | weight | Right word | weight |
|-----------|--------|-------------|--------|------------|--------|
| high      | 0.011  | high        | 0.023  | people     | 0.020  |
| economic  | 0.011  | get         | 0.019  | high       | 0.016  |
| bill      | 0.010  | gas         | 0.017  | say        | 0.015  |
| well      | 0.009  | year        | 0.011  | mean       | 0.014  |
| economy   | 0.008  | rise        | 0.011  | gas        | 0.013  |
| inflationary | 0.008    | border   | 0.011  | pay        | 0.012  |
| get       | 0.007  | well        | 0.010  | go         | 0.011  |
| year      | 0.007  | say         | 0.009  | economic   | 0.010  |
| say       | 0.007  | covid       | 0.008  | work       | 0.010  |
| president | 0.007  | trump       | 0.008  | rise       | 0.010  |
| people    | 0.007  | crime       | 0.008  | economy    | 0.009  |
| much      | 0.007  | see         | 0.007  | bad        | 0.009  |
| ease      | 0.006  | come        | 0.007  | increase   | 0.009  |
| good      | 0.006  | go          | 0.007  | crisis     | 0.008  |
| plan      | 0.005  | big         | 0.006  | problem    | 0.008  |

Table 9: Top 15 keywords of the most popular topics of the Left, Middle, and Right clusters in the vaccine dataset.

| Left word | weight | Middle word | weight | Right word | weight |
|-----------|--------|-------------|--------|------------|--------|
| trump     | 0.057  | take        | 0.027  | take       | 0.029  |
| say       | 0.015  | get         | 0.026  | stop       | 0.022  |
| get       | 0.013  | safe        | 0.019  | trump      | 0.021  |
| debate    | 0.013  | people      | 0.016  | get        | 0.016  |
| people    | 0.011  | time        | 0.014  | say        | 0.016  |
| know      | 0.009  | work        | 0.014  | mask       | 0.014  |
| need      | 0.009  | say         | 0.014  | kamala     | 0.013  |
| take      | 0.009  | trump       | 0.012  | president  | 0.012  |
| trial     | 0.008  | think       | 0.012  | work       | 0.010  |
| safe      | 0.008  | believe     | 0.012  | testing    | 0.008  |
| china     | 0.008  | effective   | 0.012  | go         | 0.008  |
| make      | 0.008  | virus       | 0.009  | public     | 0.008  |
| see       | 0.008  | available   | 0.009  | administration | 0.008 |
| trust     | 0.007  | debate      | 0.009  | see        | 0.008  |
| administration | 0.007    | trust   | 0.009  | new        | 0.007  |

Figure 6: Political scores of (a) the users in the inflation dataset and (b) the users in the vaccine dataset, inferred by follower relationship on Twitter. Note that the left side of the plot corresponds to the politically left and the right side to the politically right.

A.4 Representative Image Selection

Similar to Chen et al. [5], we extract a 2048-dimensional feature vector for each image from the last "pool5" layer of ResNet-50. Next, we apply K-means clustering. The optimal number of clusters is chosen based on the Silhouette Coefficient [43]. Within each cluster, we plot the images that are closest to their corresponding cluster centers in Figure 4c.

B SUPPLEMENTAL RESULTS

Tables 6 and 7 show the word frequency of the user descriptions of the users in the Left, Middle, and Right clusters. Tables 8 and 9 show the top 15 keywords of the most popular topics of the Left, Middle, and Right clusters.

Table 6: Word frequency of the user descriptions of the users in the Left, Middle, and Right clusters in the inflation dataset (top 10).

| Left word | count | Middle word | count | Right word | count |
|-----------|-------|-------------|-------|------------|-------|
| resist    | 228   | fan         | 90    | maga       | 380   |
| blm       | 193   | love        | 80    | conservative| 261   |
| love      | 181   | former      | 76    | love       | 217   |
| mom       | 150   | writer      | 73    | trump      | 216   |
| lover     | 143   | proud       | 72    | god        | 161   |
| democrat  | 141   | maga        | 68    | patriot    | 146   |
| proud     | 137   | author      | 67    | proud      | 128   |
| writer    | 128   | conservative| 67    | life       | 127   |
| fan       | 122   | us          | 65    | christian  | 119   |
| retired   | 117   | opinions    | 63    | kag        | 109   |