Zhou and Elejalde *EPJ Data Science* (2024) 13:28
https://doi.org/10.1140/epjds/s13688-024-00469-y

EPJ.org

EPJ Data Science
a SpringerOpen Journal

RESEARCH Open Access

Unveiling the silent majority: stance detection and characterization of passive users on social media using collaborative filtering and graph convolutional networks

Zhiwei Zhou[1*] and Erick Elejalde[1*]

[*] Correspondence:
zzhou@l3s.uni-hannover.de;
elejalde@l3s.uni-hannover.de
[1] L3S Research Center, Leibniz Universität Hannover, Appelstraße 9A, Hannover, Germany

Abstract

Social Media (SM) has become a popular medium for individuals to share their opinions on various topics, including politics, social issues, and daily affairs. During controversial events such as political elections, active users often proclaim their stance and try to persuade others to support them. However, disparities in participation levels can lead to misperceptions and cause analysts to misjudge the support for each side. For example, current models usually rely on content production and overlook a vast majority of civically engaged users who passively consume information. These "silent users" can significantly impact the democratic process despite being less vocal. Accounting for the stances of this silent majority is critical to improving our reliance on SM to understand and measure social phenomena. Thus, this study proposes and evaluates a new approach for silent users' stance prediction based on collaborative filtering and Graph Convolutional Networks, which exploits multiple relationships between users and topics. Furthermore, our method allows us to describe users with different stances and online behaviors. We demonstrate its validity using real-world datasets from two related political events. Specifically, we examine user attitudes leading to the Chilean constitutional referendums in 2020 and 2022 through extensive Twitter datasets. In both datasets, our model outperforms the baselines by over 9% at the edge- and the user level. Thus, our method offers an improvement in effectively quantifying the support and creating a multidimensional understanding of social discussions on SM platforms, especially during polarizing events.

Keywords: Collaborative filtering; Recommendation system; Graph convolutional networks; Stance prediction

1 Introduction

In recent years, digital social media networks have become increasingly important for social studies as they provide researchers with new ways of understanding users' attitudes and social interactions. However, such social analyses have relied almost exclusively on active participation and user-generated content. Still, it is known that a large percentage of social media users rarely engage [1]. According to a Pew Research report [2], in the U.S., 97% of all tweets come from only 25% of the users. This illustrates a broader, globally observable trend on Twitter, where the majority of the users restrict themselves to passively consuming the information produced by others [3, 4]. Therefore, content-based inferences may fail to generalize to real-world populations or even online communities of less active users [5]. Studies ranging from mass media attention, through stock market movements, to political election predictions are affected by these generalization issues as they increasingly rely on social media as a proxy for audience attention and opinion [6–8]. Social media's power to shape the informational landscape and its ever-more important role as a tool for policymakers render this problem very relevant in our society.

To tackle this issue, we must design models that consider broader inclusivity and reduce social biases in public discussion analyses. Multiple factors can root the user's online behavior or, in this case, self-censorship [9]. Among others, a user may remain silent online to avoid arguing with others, is insecure about their opinion, or fears negative evaluation [9–11]. This can further widen the gap in participation and inclusion between the dominant sociopolitical establishment view and vulnerable or minority communities [12]. For example, misogyny and harassment on social media platforms can have a silencing effect on women [13], including stopping expressing their opinions on specific issues [14]. Nevertheless, despite silent users (also called lurkers) rarely posting on social media, they still have valid views about socially discussed topics. By overlooking their latent views, our analyses' predictions could amplify social biases.

Most previous approaches to understanding users' viewpoints depend on identifying filters to retrieve relevant content for the topic(s) of interest and running a sentiment polarity analysis for the resulting documents [15, 16], or a keyword-based or other rule-based stance inference [17]. However, filtering users based on keywords or engagement in a particular topic may introduce selection bias. Moreover, it disregards information from discussions on other subjects that can also shed light on the opinions for the matter in question, e.g., through collaborative filtering [16, 18]. Finally, when decontextualized for document-wise analysis, the sentiment of short social media messages may not reflect a particular stance [19]. In our analysis, we look beyond the users' active engagement with the target topic and leverage other human digital traces that reflect their preferences.

Previous studies have shown that by looking into social interactions such as retweets, following, and other network associations, it is also possible to capture the lean of users [20]. Therefore, we can examine silent users' other activities to capture some signals of their preferences [21, 22]. Our analysis of silent users' stance prediction uses graph convolutional networks (GCN) to combine several types of social interactions and participation across several topics and estimate users' affinities to hashtags [23]. This method allows us to assign users to a stance represented by a (set of) hashtag(s), even when users are "silent" for the topic of interest τ (also referred to as τ*-silent* [5]). Since the model is topic agnostic, it requires limited human annotation and is only for the final analysis stage. Moreover, the users' opinions can go in many directions and represent multiple perspectives [24]. The proposed model also contributes to the state of the art by representing users' positions in a dense multidimensional space that allows us to gain subtle insights concerning public discussions rather than just a bipolar in favor/against position on a topic.

Using two recent Chilean national referendums as case studies for our target topic, we are interested in characterizing *referendum*-silent users' stances and how representative the active discussions are of the entire opinion space. Correspondingly, we use tweet datasets representing the public discussion in Santiago in the months leading to the referendums.

Our main contributions can be summarized as follows:

- We proposed a new τ-silent users' stance analysis method based on a collaborative filtering hashtag affinity prediction. This method also requires minimal supervision, making it easy to adopt in practical applications.
- We propose an encoding method to represent users' stance distributions on particular topics within an embedded space. This continuous representation enables the exploration of opinions at different granularity levels (e.g., stance, user, community).
- Our research delves into the contrasts between active users and their τ-silent counterparts, examining these differences through multiple lenses, such as tones, associated interests, and representativeness in the space of discussion. This analysis enriches our understanding of user engagement and behavior in the context of particular topics.

Collectively, these contributions advance the field of stance analysis by providing new, more inclusive tools and insights for understanding user opinions and interactions on online platforms.

2 Background and related works

As mentioned, we center our analysis on those accounts that *never* expressed their opinions explicitly for a particular topic τ of interest [5]. Following the notation in the literature, we refer to them as τ*-silent*. For example, in our case study (see Sect. 3.1), if a user has not voiced a stance on the referendum-related discussions, we call them a *referendum-silent* user.[1] Note that they may still follow, retweet, or mention (@) more active users on this or other subjects (see Fig. 1b and 1c). However, these user interactions may have different motivations (e.g., irony, criticism, or sharing information/bringing it to attention). Given this ambiguity, they could be understood as an implicit expression of interest rather than direct evidence of a user's stance on τ. In contrast, an *active user* is an account that has directly engaged with the topic τ by posting tweets that make an explicit reference to the subject. In our dataset, this is identified by the inclusion of one of the referendum-related hashtags (see Sect. 3.3). We leverage the content produced by active users and their multiple connections with τ-silent users (i.e., using collaborative filtering) to identify both active and silent users' stances on topic τ.

To approach this problem, we leverage previous works on opinion mining, recommendation systems, and collaborative filtering.

2.1 Stance detection

With most social discussions taking place online, e.g., on social media platforms, the problem of stance detection has attracted much interest over the past years. This problem is commonly defined as automatically extracting a user's position (in favor, neutral, or against) towards a particular target entity. Gauging users' opinions has multiple applications, from political analysis [25, 26], to content summarization [27], to market-trends

[1] Following, we use the term 'τ*-silent*' when referring to the abstract problem of silent users' stance detection and the more specific instance '*referendum-silent*' to refer to our case study.

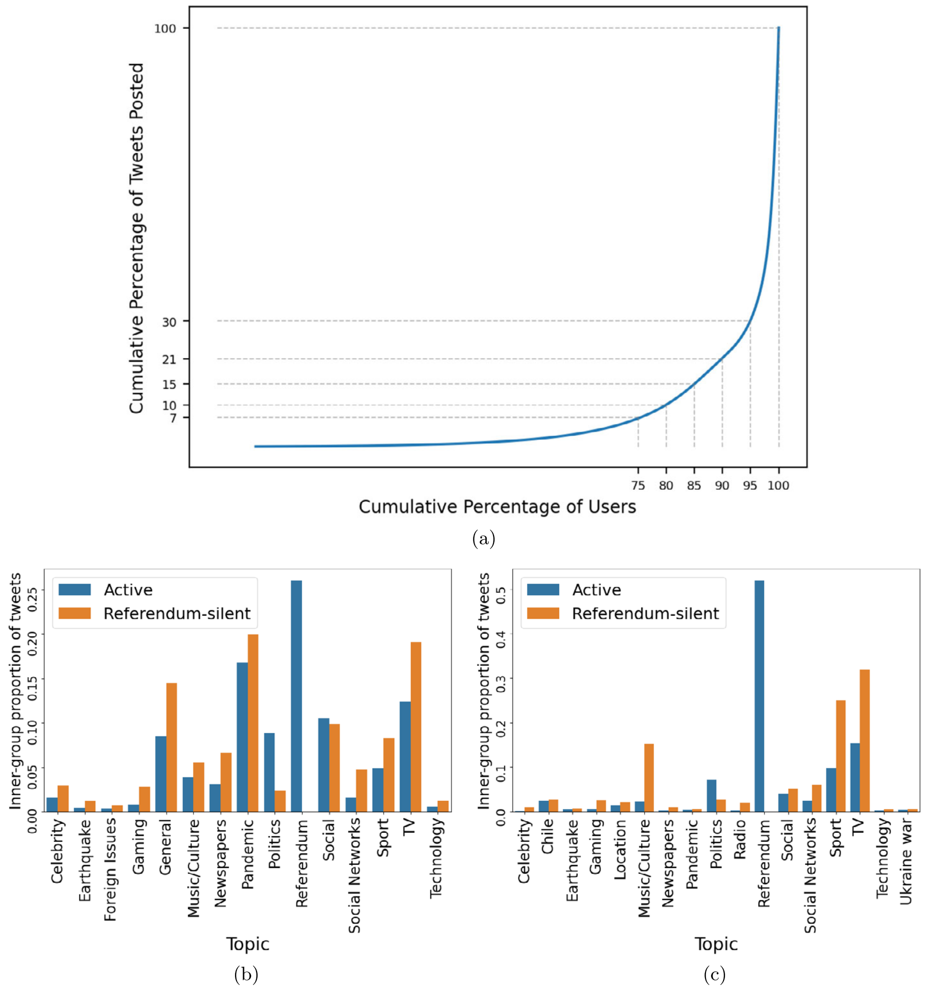


Figure 1 User Engagement and Topic Interest Distribution. 1a shows the accumulative percentage of tweets contributed by users in different activity levels. 1b and 1c shows topics relative interest of active and referendum-silent users. The proportion of attention to each topic is normalized for each group (i.e., active/silent). The topics are annotated based on the top 500 most popular hashtags. Figure 1a: Cumulative Tweet contributions by User Activity Level. Figure 1b: Topic Engagement Distribution Among Active and Silent Users in Entry_DS. Figure 1c: Topic Engagement Distribution Among Active and Silent Users in Exit_DS

prediction [28]. Depending on the application, the stance's target can be, for example, a public figure, a new government policy, or vaccination during the COVID-19 pandemic [29].

An automatic and reliable approach to this problem offers a solution to analyzing large volumes of unstructured data. Twitter, in particular, is (or used to be) an appealing data source among researchers and practitioners due to its popularity (353M active users in 2023[2]). However, short documents, informal language, and slang commonly used in social media channels pose new challenges for traditional models of opinion mining [30]. Therefore, recent studies have focused on other characteristic elements of social media. For example, methods like concatenating comments to their associated tweet to gain extra context [31] or recognizing emojis [17] have proved to help in this task.

[2] https://www.statista.com/statistics/303681/twitter-users-worldwide/.

Similarly, hashtags have been used to identify opposite stances [8, 32–34]. Communities defending a viewpoint will usually adopt a (set of) hashtag(s) that represents their position (e.g., #TrudeauMustGo or #Trudeau4MoreYears, #ISISisNotIslam or #DeportAllMuslims) [35, 36]. However, they typically start by identifying the relevant hashtags and filtering the documents based on those. Our analysis also exploits these distinguishing hashtags from different camps to profile Twitter users. Yet, we use the information from all the hashtags (related to the topic or not) and let the model simultaneously learn embeddings for users and hashtags unsupervised.

These approaches discussed above are all content-based and, thus, rely on users' active participation in the topic of interest. This makes them unsuitable to address our problem of predicting the stance for τ-silent users.

2.2 User homophily

Given the scarcity of content for τ-silent users, we need to appeal to other features that will allow us to infer their opinions in a given matter. For example, the social network's topology has been shown to provide additional information to create classifiers concerning a user's preference, even when the choices are very similar (e.g., Pepsi vs. Coca-Cola, Hertz vs. Avis or McDonald's vs. BurgerKing) [37]. These features work under the principle of homophily [22], where we assume that social entities will associate with similar others. For example, in Twitter, researchers have experimented with the 'following' relationship (both unidirectional and bidirectional) [21]. Moreover, these relations can be extended to indirect or second-order co-following [37], e.g., two accounts that do not share a single follower can still be considered similar if their followers are highly connected.

Our approach also takes advantage of other topological features that can help expand similarity networks [38]. Specifically, we rely on (i) *Social Graphs* – including different social circles such as friend or mention; (ii) *Entity-Centric Graphs*, based on co-following relations between the users around a particular type of entity such as political candidates or news outlets; and (iii) Geo-Centric Graphs, grouping users with a given geopolitical profile, e.g., as self-reported in their biographies [38]. Our analysis explores the influence of multiple relations between users on predicting the users' stances for different levels of activities. Furthermore, by considering users' activities and interactions beyond the specified topic τ, we can exploit user-user relationships and similarities [21, 38, 39] that might shed some light on the characteristics of less active member accounts.

However, in our case, observable connections might still be sparse as τ-silent users tend to have low participation and a limited number of other interactions. For example, the number of followers/followees is usually positively correlated with the influential role of the user [40]. To address this problem, we include the analysis of another relevant set of latent relations between users referred to as meta-paths [41]. A meta-path represents a sequence of relations between different object types where the first and last objects in the sequence are similar. The authors note an improved performance of meta-paths compared to random-walk-based methods [41].

2.3 Collaborative filtering

Most of the previous work on opinion mining focuses on training a classifier. However, this task can also be framed from the recommendation system (RS) perspective. In this case, we are interested in predicting user-topic affinity, or more precisely, user-[opinion on a topic] affinity. There are two popular approaches in the area of RS, namely Matrix Factorization methods and user-item graph structures analysis. Matrix factorization (MF) projects the ID of a user u and an item f into a lower-dimensional embedding vector H_u and W_f, respectively [42]. During the training process, W_f and H_u are iteratively updated to minimize the discrepancy between observed user-item interactions and those predicted by the model. The ultimate goal is for the inner product of H_u and W_f to accurately estimate the missing or unobserved interactions, thereby revealing underlying patterns in the data that can inform recommendations or insights. Some frameworks have tried to extend MF, e.g., by combining it with a multilayer perceptron (named neural collaborative filtering – NCF) [18]. However, with the proper setting, the original MF method outperformed the NCF framework and other methods in most cases [43, 44]. Also, Wang et al. proposed a coupled sparse matrix factorization (CSMF) approach to collaborative filtering in the prediction of sentiments towards topics [16]. The authors relied on manually selected and annotated topics and used the accuracy of the sentiment polarity predictions to evaluate the model.

Alternatively, RS can be approached by exploiting the user-item bipartite graph structure. This creates a mapping from the RS to the link prediction problem. Motivated by the strength of graph convolution, Wang et al. proposed a Neural Graph Collaborative Filtering (NGCF) framework that captured collaborative signals in high-hop neighbors and integrates them into the embedding learning process [45]. However, further studies showed that NGCF demonstrates higher training loss and worse generalization performance with nonlinear activation and feature transformation [46]. As a result, the authors proposed a simplified model named Light Graph Convolution Network (LightGCN). Other works have leveraged LightGCN by aggregating information from different aspect-level graphs [47] (e.g., adding a user-director graph on a user-movie recommendation to guide the embedding learning process). These RS models typically aggregate information by averaging data from neighbors. Alternatively, attention mechanisms have also been proposed to capture the importance of different relationships between users and items [48].

Finally, in earlier work, we tested the effectiveness of adding weights to a LightGCN-based model [23]. Here, we expand on this work by comparing different weighting strategies. We compare the performance of the previous user-level normalized weights (local) against a simple count of interactions (without normalization) and a TF-IDF normalized weight (global). Furthermore, we evaluate another method of aggregating information from different input graphs. In contrast to the previous averaging approach, we implemented a "projection & fusion" layer that significantly improves the model performance across multiple scenarios.

2.4 Silent users

Some previous studies have directly addressed the stance prediction of silent users [5, 16, 32]. A combination of network and content features is a common strategy in these cases. Collaborative filtering has been proposed to transfer information from active to silent users [16]. However, they depend on a manually annotated set of topics or rely on supervised learning, making it difficult/costly to scale and deploy in practical scenarios. Our model achieves state-of-the-art performance with relatively few annotated hashtags for each stance. Moreover, most approaches in the literature only provide a stance polarity classification (i.e., in favor – against), and it is usually inferred from sentiment analysis (which does not necessarily equal stance [19, 31]). Our methodology contributes to this line of research by predicting the τ-silent users' stance in a continuous higher-dimensional space, thus allowing a finer-grain stance analysis (i.e., potentially applicable to discussions with more than two opinions). In this study, we use our case study to show how this learned continuous space of discussion and the users' positioning can help to further characterize users and communities. These aspects of the study include several new valuable dimensions to our research.

Finally, we contribute to understanding the dynamics and representativeness of opinions expressed by users with different levels of engagement [34]. Expanding on the previously introduced WLGCN [23], the present work includes an in-depth analysis of the proposed model's performance across groups of users with various activity levels. This should help understand how overall users' behavior can affect the task of predicting their stances.

3 Datasets

This section describes the datasets used to train and validate our models. We start by presenting the case study and contextualizing the collected data. Then, we define the collection process and the filters applied to the data, resulting in our final corpus.

Following the FAIR data principles, we make our datasets available on GitHub.[3] However, to comply with Twitter's terms and conditions, we only share tweet IDs that can be rehydrated.

3.1 Case-study: Chilean constitutional referendum

In 2019, Chile saw one of its biggest popular uprisings following a perceived increase in economic hardship and social inequalities. After weeks of protest, lawmakers agreed to hold a referendum on the nation's dictatorship-era constitution. The constitutional referendum was demarked by two plebiscites: the first plebiscite (25 October 2020[4]) asked whether a new constitution should be drafted; the second plebiscite (4 September 2022) was to vote on whether the people agreed with the text of the new constitution drawn up by the Constitutional Convention. These are popularly known in Chile as "entry plebiscite" (plebiscito de entrada) and "exit plebiscite" (plebiscito de salida).

In the entry plebiscite, the "Approve" side won by a large margin, with over 78% agreeing to draft a new constitution. However, after two years of intense political campaigns from both sides, including heated social media discussions, the new text was rejected in the exit plebiscite with almost 62% of the votes for "Reject". These campaigns were especially active on social media and, as expected, made extensive use of hashtags denoting the corresponding position of each camp (e.g., #Apruebo (I approve) or #Rechazo (I reject)).

3.2 Data collection

Twitter (now X) is one of the most popular social media platforms in Chile[5] for news consumption and where millions of Chileans discuss trending topics. Thus, we used Twitter to collect topics and users' information through the official API.

[3] https://github.com/imzzhou/StanceInferenceInTwitter.git.

[4] The plebiscite was initially set for 26 April 2020. However, due to the COVID-19 pandemic, it was rescheduled for October of that year.

[5] https://www.statista.com/topics/6985/social-media-usage-in-chile/#dossierKeyfigures.

We start from a database of 384 news outlets with an active social media presence and targeting a Chilean audience [49]. Then, we collect tweets and profiles from these news outlets' followers. By focusing on people who consume their news from this media system, we target informed users who probably have a formed opinion on the discussed topics. As with many other studies on social media platforms like Twitter, we are limited to the positive actions of the users in the system. This means we cannot distinguish in our analysis between users who only read and those who have not yet seen the content. However, we can confirm that the topic of the referendum received ample coverage by the mainstream media in Chile. By focusing on newspaper followers who were also active on Twitter during the observed period (i.e., they interacted with other topics), we assume that, with a high probability, they have been exposed to the referendum discussion (our target topic). We expect these users to leave online traces of their stand, even if not explicitly shared. For example, users who do not participate in the referendum topic may participate in other discussions and align with other users in different stances.

We start from a collection of 9.2 million unique followers. We need to apply several filters to make our social graphs manageable and remove potential noise. We try to restrict our analysis to accounts that represent regular users and most likely reflect the popular range of stances. We empirically found that accounts following more than ten news outlets have characteristics often associated with automated accounts or representing businesses (e.g., high follower-to-following ratio)(see Appendix A). Such accounts tend to exhibit markedly different patterns from those of regular users. Thus, their engagement with one or another hashtag is probably not driven by a genuine interest in the topic but rather by a commercial interest or a preplanned political agenda. Since our model learns the τ-silent users' stances from other more active users, including these accounts in our analysis could lead to results not representative of the behavior of the average human user. So, we filter users who simultaneously follow at most ten news outlets, excluding potential automatic accounts.

To further eliminate potential noise in the opinions, we remove hyperactive accounts that, e.g., might be managed by automatic processes (i.e., bots) or work as part of an information campaign. As mentioned before, these accounts usually do not represent real individuals and will not convey a genuine personal instance within a controversial discussion. So, we introduce an additional filter based on the average daily number of tweets an account posts. We empirically chose at most three tweets per day on average as a reasonable activity for a regular personal account.

More advanced methods of bot detection (e.g., [50]) may help in the collection of a larger sample or in identifying more sophisticated bots that mimic the behavior of human users (e.g., social bots [51]). However, with our filters, we still retain a sizeable amount of relevant users, and while our current dataset may contain some advanced bots, we considered their presence tolerable for our analysis. We assume that they will not significantly affect topic interaction patterns as they usually engage with fewer topics, which limits potential interactions in our graph (see Sect. 4). We also assume they do not represent a significant percentage of the community, such as to bias affinity patterns.

We also want to reduce possible bias introduced by geographic and social factors. For this, we use the *location* field in the users' profiles to restrict the network to followers self-geolocated in one city, i.e., the capital of Chile, Santiago. Outlets like, e.g., *El Mercurio* have a national scope and, thus, will have many followers that do not match this filter. Previous studies have shown that many user profiles contained a genuine location, mainly at the city level [49, 52, 53]. We specifically targeted profiles that included "Chile" and "Santiago" in their *location* field.

After applying the filters above, we considerably reduced the number of users to under 40K accounts. In comparison, previous studies addressing a similar task worked with 6K users [16]. Our first dataset (*Entry_DS*) comprises 34,412 users with 915,672 associated tweets (between Jan 1st and October 24th, 2020) containing 189,115 hashtags. This dataset tries to capture the popular discussions during the political campaigns for the "entry plebiscite." For our second dataset, our final *Exit_DS* comprises 39,239 users. For each account, we collected all tweets between Jan 1st and September 3rd, 2022. This resulted in 2,161,806 associated tweets containing 69,892 hashtags. Equivalent to the first dataset, *Exit_DS* tries to capture the popular discussions during the political campaigns for the "exit plebiscite".

In our datasets, we also observe an engagement pattern similar to the one described in the literature [2], where the vast majority of the content comes from a small group of very active users. In our case, the 25% most active users authored 93% of the tweets (see Fig. 1a). This further strengthens the relevance of τ-silent users' stance detection in Chilean online discussion.

3.3 Hashtag annotation

We identified the 500 most used hashtags for each dataset and annotated them into several general topics. This gives us an idea of the relative interest of users in other issues during these periods (see Sect. 7.4).

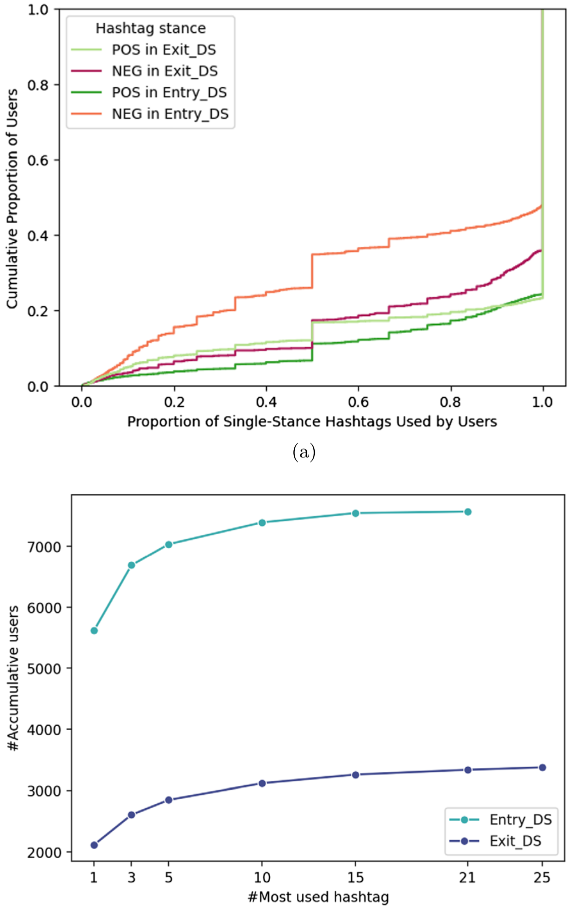
Concerning our main topic, we split the hashtags related to the Chilean referendum into three groups: "*POS*" indicating a favorable stance, "*NEG*" indicating a rejecting stance, and "*NEUTRAL*" indicating interest or engagement but with a neutral stance. See Appendix B for the list of referendum-related hashtags. For the annotation of referendum-related hashtags, two independent coders (one of the authors and another external researcher, both native speakers) annotated each hashtag as belonging exclusively to one of the three stances (*POS*,*NEG*, *NEUTRAL*) or others (i.e., not related to the referendum). In a second stage, coding disagreements were solved through a negotiation among coders in order to improve inter-rater reliability. The final classification was established with high agreement, Cohen's Kappa $\kappa = 0.962$ and $\kappa = 0.993$ for the *Entry_DS* and *Exit_DS*, respectively.

To compare the topic preferences between *referendum*-silent and active users, we calculate the inner-group proportion of tweets associated with each topic (see Fig. 1b and 1c). For example, out of all the tweets from *referendum*-silent users in Entry_DS, 0.20 is related to the pandemic. On the other hand, out of all the tweets from active users in the same dataset, approximately 0.17 reference the pandemic. Note this only shows the relative proportional interest within each group (e.g., even when *referendum*-silent users have a higher relative interest in sports than active users, they may have fewer tweets on the topic). As expected, *referendum*-silent users are involved in other topics.

Notably, other topic annotations and insights are only intended to contextualize the case study and provide a more detailed characterization of the collected data. Only the referendum hashtags annotation is necessary for the users' classification.

Figure 1b and 1c already shows interesting characteristics of *referendum*-silent users. For example, they have relatively low participation in other political discussions. However, they share common interests with active users in topics related to the pandemic, sports, and TV. Within these shared interests, we can also find polarized discussions that help the model find patterns and categorize similar users. This illustrates the advantage of extending our observation to other concurrent topics and the collaborative filtering approach for analyzing τ-silent users.

To determine users' consistent affinity for specific hashtags, we examined user-generated tweets, focusing on the annotated hashtags related to the referendums. Figure 2a effectively illustrates a clear pattern: users predominantly align with hashtags that reflect one or the other stance, with positive and negative curves in each dataset showing a steep rise as the proportion increases. Notably, the POS curves for both the Exit_DS and Entry_DS datasets remain under 20% until the end, suggesting a more concentrated user base using exclusively positive stance hashtags. This analysis indicates that, regardless of the referendum outcomes, a substantial cohort of users predominantly employ hashtags that resonate with a single stance, underscoring the importance of hashtag stance in user



(a)

(b)

Figure 2 Insights into Hashtag Usage Patterns Across Datasets. 2a represents the proportion of users who predominantly use hashtags of a single stance (either POS or NEG) in both *Entry_DS* and *Exit_DS* datasets. A sharp rise towards the right indicates many users favoring hashtags exclusively from one stance. 2b plots the cumulative number of users who interacted with the most used hashtags: Fig. 2a: Cumulative Distribution of User Preferences by Hashtag Stance. Figure 2b: Cumulative User Engagement with Top Referendum-Related Hashtags

engagement patterns. These results also align with previous findings on "echo chambers" in social media [54].

The pattern seen in Fig. 2a is also significant because it suggests that instances of hashtag hijacking – where users co-opt hashtags for purposes contrary to their original intent – would be relatively rare in our dataset. If present, the majority of the community did not seem to have adopted the hijacked hashtags. The alignment of user behavior with specific hashtag classes, as shown in the cumulative distribution plot, suggests that most users interact with hashtags in a manner consistent with their original context and meaning.

Finally, regarding the representativeness of the most popular hashtags for the users' set, Fig. 1a represents the cumulative number of users who interacted with the most used hashtags. The graph shows that after 5–10 hashtags, most of the active users are represented in the discussion. This pattern is consistent for both datasets. In practice, an analyst could identify and annotate these highly occurring hashtags, e.g., based on trending topics reported by the platform, without having to go through all the documents related to the topic.

4 Methodology

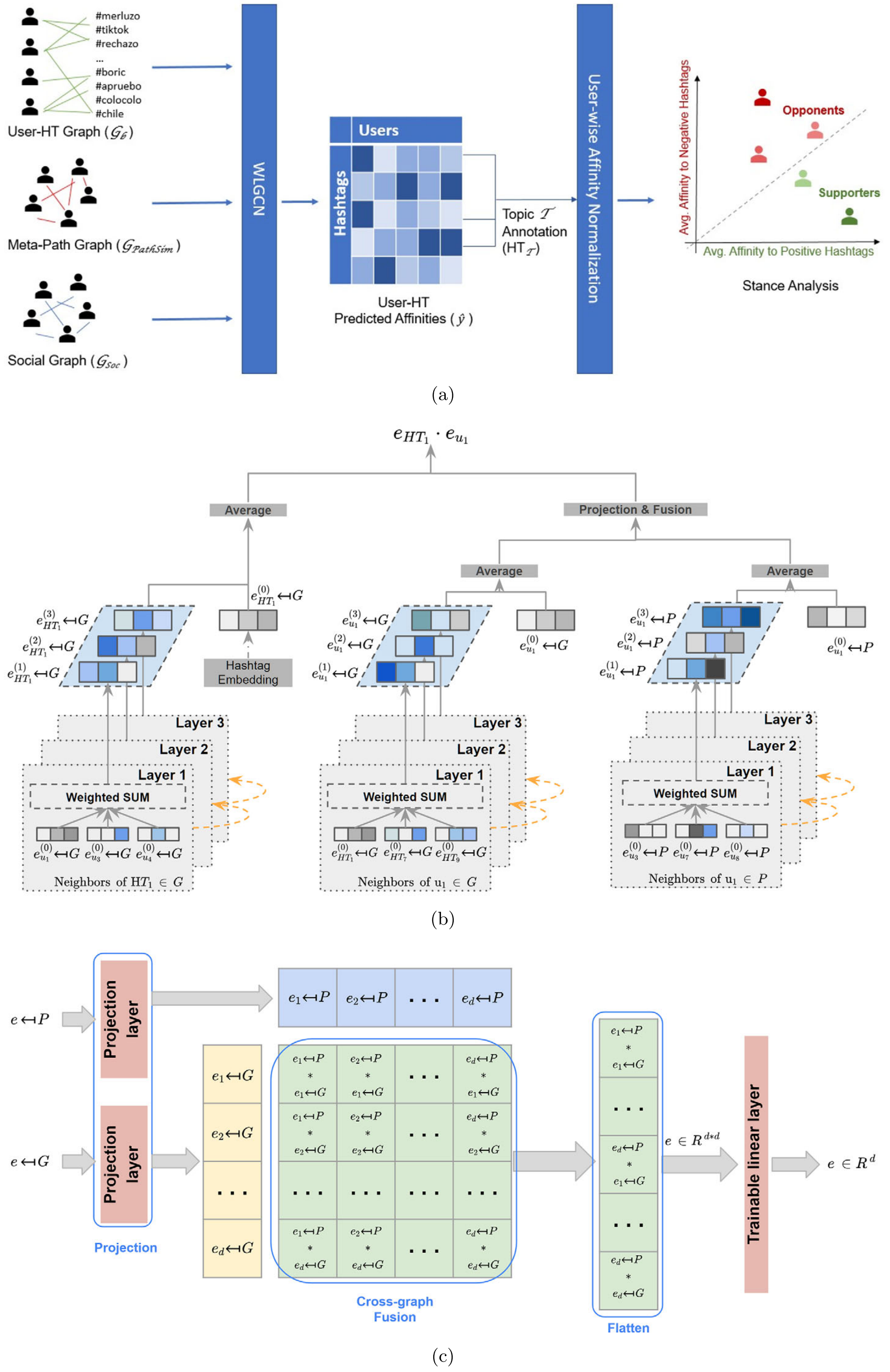
Our approach to τ-silent users' stance analysis relies on identifying their engagement patterns with different topics and similarities to other, more active users. Using collaborative filtering, we aim to predict users' association with a topic and the various perspectives within each discussion. However, instead of depending on the content sentiment, we leverage the semantic information provided by hashtags and users' affinity to these hashtags to differentiate between stances [32]. Hashtags play a crucial role in categorizing, discovering, and contextualizing content on social media platforms facilitating user engagement. By learning the hashtags' embeddings we hope to get representations that capture their semantic similarities.

We represent the User-Hashtag relationship as a bipartite graph $\mathcal{G}_b$. The graph consists of two classes of nodes $\mathcal{V}_U$ and $\mathcal{V}_{\text{HT}}$, which represent the users and hashtags, respectively. A set of weighted edges $\mathcal{E}$ is defined to represent the interactions between users and hashtags. Then, each edge only connects nodes from different classes. We define the weight of an edge as $e_{i,j} = T_{i,j}$, where $T_{i,j}$ is the number of times user i used the hashtag j.

We test alternative weighting approaches based on various normalizations. In previous work, we experimented with edge weights normalized by users [23]. Specifically, the edge weight was $e_{i,j} = \frac{T_{i,j}}{\sum_j T_{i,j}}$. This normalization aims to identify hashtags that are more important to individual users. Here we also experiment with a global strategy. Recognizing that some hashtags may be used by many users and thus reduce their discriminative value, we introduce a strategy based on the term frequency-inverse document frequency (TF-IDF) principle. Specifically, TF-IDF normalized weights are defined as $e_{i,j} = \frac{T_{i,j}}{\sum_j T_{i,j}} \times \log \frac{N}{\sum_{u=1}^{N} I_u(j)}$ where N is the number of users and $I_u(j)$ is an indicator function that is 1 if an edge exists between user i and hashtag j, and 0 otherwise. For comparison, we try the models with each weighted strategy. The experimental setting details can be found in Sect. 5.1.

In the following sections, we first describe the pre-processing steps and hashtag classification. Then, we present the model's general architecture and discuss the integration of user-hashtag interactions. We also take on other types of information from social media interactions, including hashtag embeddings and inferred user-user relationships. Finally, we introduce optimizing the objective function of our model.

Based on the predicted affinities toward a relatively small set of annotated topic-specific hashtags, we propose a data-driven user stance classification that can be applied to active and τ-silent users. Figure 3a gives an overall view of our approach.


#merluzo
#tiktok
#rechazo
...
#boric
#apruebo
#colocolo
#chile
User-HT Graph ($\mathcal{G}_{\mathcal{E}}$)
Meta-Path Graph ($\mathcal{G}_{PathSim}$)
Social Graph ($\mathcal{G}_{Soc}$)
WLGCN
Users
Hashtags
Topic $\mathcal{T}$ Annotation (HT$_{\mathcal{T}}$)
User-HT Predicted Affinities ($\hat{y}$)
User-wise Affinity Normalization
Avg. Affinity to Negative Hashtags
Avg. Affinity to Positive Hashtags
Opponents
Supporters
Stance Analysis
$e_{HT_1} \cdot e_{u_1}$
Average
Projection & Fusion
Average
Average
$e^{(0)}_{HT_1} \leftarrow G$
$e^{(3)}_{HT_1} \leftarrow G$
$e^{(2)}_{HT_1} \leftarrow G$
$e^{(1)}_{HT_1} \leftarrow G$
Hashtag Embedding
$e^{(3)}_{u_1} \leftarrow G$
$e^{(2)}_{u_1} \leftarrow G$
$e^{(1)}_{u_1} \leftarrow G$
$e^{(0)}_{u_1} \leftarrow G$
$e^{(3)}_{u_1} \leftarrow P$
$e^{(2)}_{u_1} \leftarrow P$
$e^{(1)}_{u_1} \leftarrow P$
$e^{(0)}_{u_1} \leftarrow P$
Layer 3
Layer 2
Layer 1
Weighted SUM
$e^{(0)}_{u_1} \leftarrow G$ $e^{(0)}_{u_3} \leftarrow G$ $e^{(0)}_{u_4} \leftarrow G$
Neighbors of HT$_1$ $\in G$
$e^{(0)}_{HT_1} \leftarrow G$ $e^{(0)}_{HT_7} \leftarrow G$ $e^{(0)}_{HT_9} \leftarrow G$
Neighbors of u$_1$ $\in G$
$e^{(0)}_{u_3} \leftarrow P$ $e^{(0)}_{u_7} \leftarrow P$ $e^{(0)}_{u_8} \leftarrow P$
Neighbors of u$_1$ $\in P$
$e \leftarrow P$
$e \leftarrow G$
Projection layer
Projection layer
Projection
$e_1 \leftarrow P$
$e_2 \leftarrow P$
. . .
$e_d \leftarrow P$
$e_1 \leftarrow G$
$e_2 \leftarrow G$
$e_d \leftarrow G$
$e_1 \leftarrow P * e_1 \leftarrow G$
$e_2 \leftarrow P * e_1 \leftarrow G$
$e_d \leftarrow P * e_1 \leftarrow G$
$e_1 \leftarrow P * e_2 \leftarrow G$
$e_2 \leftarrow P * e_2 \leftarrow G$
$e_d \leftarrow P * e_2 \leftarrow G$
$e_1 \leftarrow P * e_d \leftarrow G$
$e_2 \leftarrow P * e_d \leftarrow G$
$e_d \leftarrow P * e_d \leftarrow G$
Cross-graph Fusion
$e_d \leftarrow P * e_1 \leftarrow G$
$e_d \leftarrow P * e_d \leftarrow G$
Flatten
$e \in R^{d*d}$
Trainable linear layer
$e \in R^{d}$

(a)

(b)

(c)

Figure 3 Comprehensive Overview of the Methodology and WLGCN Model Components. Figure 3a: Schematic Overview of Stance Classification Framework. Figure 3b: Architecture of the WLGCN Model. Figure 3c: Projection and Fusion Mechanism in WLGCN

4.1 Data preprocessing

From the collected datasets, we normalize and standardize the hashtags encoding into UTF-8 and get 185,965 and 68,331 unique hashtags for *Entry_DS* and *Exit_DS*, respectively. In addition to their selection by the users, hashtags' semantic information plays an important role. Therefore, we apply the following steps to process the content of the tweets:

- `standardization` of texts into UTF-8, replace the accented characters with regular ones (i.e., á → a), and lowercase the texts;
- `removal` of URLs, emojis, punctuation, stopwords, and personal information;
- `lemmatization and stemming` of tokens into declined forms;
- `word-embedding`: we use the cleaned content to train a word embedding model.[6] From this embedding, we keep only the hashtags' representation.

Since we are using a Bag-of-Word-based model in our experiments for learning the embedding of our hashtags, we decided to remove non-essential content such as punctuation marks, URLs, and stopwords. Together with our other preprocessing steps, this helps the model to focus on the more relevant information, reduce sparsity, and better differentiate the topics associated with the hashtags [56].

However, we use, among other features, sentiment analysis to characterize the communities (Sect. 7.4). In this case, contextual elements in the content, like emoticons, may play an important role, especially in social media communications. Thus, for the sentiment analysis, we start from the raw content and translate emojis into their textual form.

4.2 Weighted light graph convolutional network for hashtag-affinity prediction

Figure 3b presents the architecture of our proposed model. This represents an extension of the LightGCN that introduces weights to the relation graphs and various additional characteristic features of our social network. We will refer to our model as WLGCN (Weighted-LightGCN). The model's inputs include a user-hashtag interaction graph, hashtag embeddings, and the inferred relationship between users. The output represents the users' predicted affinity to the hashtags in the dataset. Through a series of graph convolutional layers, the model jointly updates the representations of users and hashtags by aggregating the neighbors' features. After K layers, the affinity score is calculated as the inner product of the users' and hashtags' embedded representation.

4.3 Graph convolutional network

The basic idea of Graph Convolutional Networks (GCN) is to learn representations of nodes by aggregating the neighbors' embeddings as the new presentation of the target node. The layer-k embeddings of the target node can be represented as:

$$\mathbf{h}_n^k = \mathrm{AGG}\big(\mathbf{h}_n^{(k-1)}, \{\mathbf{h}_i^{(k-1)} : i \in \mathcal{N}_n\}\big), \quad \mathbf{h}_n^0 = \mathbf{e}_n \tag{1}$$

where $\mathbf{e}_n$ represents the initial embeddings of a node n, $\mathcal{N}_n$ represents neighbors of this node, and AGG is a function used to aggregate the features of the neighbors. The other standard operations in a GCN layer (i.e., non-linear activation and feature transformation) have been shown to contribute little to the recommendation performance [46]. Therefore,

[6] We use FastText with CBOW [55].

we also skip these two operations and use the simple average aggregator instead. To illustrate, consider our interaction graph $\mathcal{G}_b$ with N users and M hashtags (HT),[7] the propagation rule in layer k can be defined as:

$$H^k = \left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)H^{k-1}, \qquad H^0 = \mathbf{E}^0 \tag{2}$$

where $H^k \in \mathbb{R}^{(N+M)\times d}$ is the User-HT graph embedding matrix after the kth propagation step; E^0 is the initial d dimensional embedding of users and HTs; D is a diagonal matrix, where $D_{i,i}$ equals to $\sum_j A_{i,j}$, A stands for the User-HT graph adjacency matrix and is defined as:

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{R} \\ \mathbf{R}^T & \mathbf{0} \end{pmatrix} \tag{3}$$

being $R \in \mathbb{R}^{N\times M}$ the User-HT interaction matrix, where $R_{i,j} = e_{i,j}$ (i.e., the weight of the edge connecting user i and hashtag j). After propagation, for the node n, which represents the user or the hashtag, we employ the weighted average to combine the embeddings learned through layers 1 to K, and the combination can be formulated as:

$$e_n = \frac{1}{K+1}\sum_{k=1}^{K} H_n^k \tag{4}$$

Finally, we calculate the affinity by applying the inner product operations to the user and hashtag embeddings:

$$\hat{y}(u, ht) = e_u e_{ht}^{\top} \tag{5}$$

Depending on the use case, normalized vectors that prevent an effect on the final result from their magnitude could be desirable. However, for our study, we believe occurrence counts are a feature since more popular/trending hashtags are typically adopted as banners by the different camps participating in the discussion and more confidently represent the stance of a community.

4.4 Inferred information

We complement the user-hashtag interaction graph with three additional types of data characteristic of our social network to help in the above learning process.

First, we add hashtag embeddings to capture their semantics. The aim is to complement the hashtags usage patterns at the user level, represented by the vanilla WLGCN, with the contextual information provided by the tweets' content. For our experiments, we trained a FastText model [55] with the pre-processed corpus introduced in Sect. 3. Then, we use the representation of the hashtags (E_{HT}) to initialize the hashtag embedding layers of our model.

The second type of information is user-user network interactions. The user-user graph is an instance of a *Social Graph* with heterogeneous connections (henceforth $\mathcal{G}_{\mathrm{Soc}}$). In $\mathcal{G}_{\mathrm{Soc}}$,

[7] We apply the same strategy to the other inferred graphs.

we include as links the mutual friend/follow relationship as well as mentions of other users in our network.

The last type of information is the user-user simulated path (PathSim [41]). The graph $\mathcal{G}_{\text{Soc}}$ mentioned above represents direct user-based connections observable from our Twitter dataset. However, in practice, these interactions are sparse in a network like ours. So, they would offer a limited contribution to the embedding learning process. To address this issue, we extend these observed relations with inferred pseudo-relations based on *meta-paths*. A *meta-paths* captures a sequence of relations connecting two users that may contain multiple steps. For example, users u_i and u_j are connected through a path "user-retweet-hashtag-tweet-user" (U-RT-HT-T-U) if u_i retweeted/quoted a tweet containing a hashtag that also appeared in a tweet of u_j. Given the *meta-paths* ($\mathcal{P} = U$-RT-HT-T-U), the similarity between u_i and u_j is defined as:

$$s(i,j) = \frac{2 \times |\{p_{i \rightsquigarrow j} : p_{i \rightsquigarrow j} \in \mathcal{P}\}|}{|\{p_{i \rightsquigarrow i} : p_{i \rightsquigarrow i} \in \mathcal{P}\}| + |\{p_{j \rightsquigarrow j} : p_{j \rightsquigarrow j} \in \mathcal{P}\}|} \tag{6}$$

where $p_{i \rightsquigarrow j}$ represent the path instance between u_i and u_j that follows the *meta-paths* $\mathcal{P}$. In our experiments we use the RT (retweet/quote) relation. However, the RT relation could be replaced with other content-based relations such as replies. These path instances define an additional, denser user-user graph ($\mathcal{G}_{\text{PathSim}}$). Since both graphs, $\mathcal{G}_{\text{Soc}}$ and $\mathcal{G}_{\text{PathSim}}$ contained additional user information, We assume these graphs could help updating the embeddings of users. Inside each graph, we also applied Equation (4) with K layers to extract the potentially useful information.

4.5 Graph projection and fusion

To effectively aggregate and align these vectors from different contexts, we adopt the projection layers and the cross-fusion technique [57] (see Fig. 3c). Traditional methods, such as simple averaging or concatenating embeddings from multiple graphs, often fail to capture the contextual relationships, especially when various user-user interactions are included in addition to the primary user-hashtag graph. Given a user u_i, for each user-user graph, we compute the outer product of the user's embedding from that graph with the user's embedding from the main user-hashtag graph. This results in an interaction representation that encapsulates the correlations between the contexts of these graphs and the primary user-hashtag graph. Specifically, given two embeddings $r_{u_i}^{\mathcal{G}_b}$ (from the user-hashtag graph $\mathcal{G}_b$) and $r_{u_i}^{\mathcal{G}_x}$ (from each user-user graph $\mathcal{G}_x$ such as $\mathcal{G}_{\text{PathSim}}$ or $\mathcal{G}_{\text{Soc}}$) of size d, the representation R is computed as $R = r_{u_i}^{\mathcal{G}_b} \otimes r_{u_i}^{\mathcal{G}_x}$. Flattening R results in a vector of length d^2. This is then reduced to a dimensionally consistent vector r of size d using a sequence of linear layers. Once we have obtained the user embeddings from all graphs, they are aggregated using a MEAN operation. This ensures a balanced inclusion of context information from different graphs and alignment with the hashtag embeddings.

Our strategy captures the complex relationships between embeddings from various interaction graphs. Thus, we leverage the rich contextual information across graphs, ensuring our aggregated representations are context-aware and carefully aligned.

4.5.1 Optimization

With Equation (4), our idea is to keep nodes connected with an edge close to each other in the latent space while pushing nodes without a shared edge farther apart. So, we adopt the Bayesian Personalized Ranking (BPR) loss [58] as objectives for training our model:

$$\text{Loss} = -\sum_{u=1}^{N}\sum_{i\in\mathcal{N}_u}\sum_{j\notin\mathcal{N}_u} \ln\sigma\left(\hat{y}(u,i) - \hat{y}(u,j)\right) + \lambda\left\|\mathbf{E}^0\right\|^2 \tag{7}$$

where σ is the sigmoid function, λ is the regularization parameter to avoid overfitting, and i, j represent the hashtags used or not used by the user u. We adopt the Adam algorithm [59] for model optimization. We sample a tuple of (u, i, j) for each mini-batch and update the embeddings.

4.6 Affinity-based stance classification

Manually inspecting and annotating all hashtags covering multiple topics can still be expensive and time-consuming. The proposed WLGCN model can leverage hashtag semantics to predict the user's viewpoint without any annotation. Thus, identifying a relatively small subset of hashtags related to the topic of interest τ should suffice to characterize the users' stance. Even if other τ-related hashtags exist in the dataset, their influence on the users will be captured through the hashtags embedding (E_{HT}) and convolutions on $\mathcal{G}_b$.

We use previously identified referendum-related hashtags (HT_τ) to characterize the topic discussion (40 and 56 hashtags for Entry_DS and Exit_DS, respectively). These are further assigned into three groups: $\mathrm{HT}_\tau^{\mathrm{POS}} \subset \mathrm{HT}_\tau$ expressing approval of a new constitution, $\mathrm{HT}_\tau^{\mathrm{NEG}} \subset \mathrm{HT}_\tau$ indicating a rejection of a new constitution, and $\mathrm{HT}_\tau^{\mathrm{NEUTRAL}} \subset \mathrm{HT}_\tau$ indicating interest or engagement but with a neutral stance (usually found in news media tweets). We use these annotations as ground truth in the validation step to measure the performance of the proposed approach.

Based on its affinities to hashtags in each class, we assign each user to one of the defined stances on a topic (i.e., *POS*, *NEG*, *NEUTRAL*). To decide the stance of user u_i on the topic τ, we normalize her affinities and then select the class with the highest average affinity (see Equation (8)).

$$\begin{aligned}&\text{stance}_\tau(u_i) = \underset{c\in C}{\arg\max}\left(\frac{1}{|\mathrm{HT}_\tau^c|}\sum_{j\in\mathrm{HT}_\tau^c}\frac{\hat{y}(u_i,j) - \hat{y}_{\min}(u_i,\mathrm{HT}_\tau)}{\hat{y}_{\max}(u_i,\mathrm{HT}_\tau) - \hat{y}_{\min}(u_i,\mathrm{HT}_\tau)}\right)\\ &C = \{\mathrm{POS}, \mathrm{NEG}, \mathrm{NEUTRAL}\}\end{aligned} \tag{8}$$

where $\hat{y}_{\min}$ ($\hat{y}_{\max}$) represents the minimum (maximum) predicted affinity for user u_i among the τ-related hashtags.

For evaluation purposes, we follow the convention for stance classification (i.e., POS, NEG, NEUTRAL). However, our method could be used for any set of user-defined stances determined by the selection and annotation of the related hashtags.

5 Experiments

In this section, we first describe our experimental setup and evaluation approach. We describe the selected baselines that represent state-of-the-art approaches to collaborative-filtering-based stance detection and recommendation systems.

5.1 Experimental setup

Before training, at the first embedding layer, we use the Xavier uniform [60] to initialize the embeddings of users. As for hashtags, the previously trained word embeddings are used for initialization. For comparison, we also tried hashtag representations with the Xavier uniform initializer in our experiments. For the number of convolutional layers K in our GCN, similar to previous works ([46, 47]), we use three layers to extract and aggregate information from neighbor nodes. Early stopping is performed to prevent overfitting, i.e., the training will stop if nDCG@20 on the validation data does not increase for 50 consecutive epochs.

We conducted several experiments to evaluate the contribution of different components to the τ-silent users' stance detection.

5.2 Baselines

We use two state-of-the-art methods as baselines to evaluate the performance of our proposal. In addition, we also use a Null-model to test whether the observed User-HT relations contain non-trivial information that helps in the identification of users' stances. Below, we summarize the included baselines:

- `Null-Model` [61]: We create a randomized User-HT interaction matrix. For each user, we get the number (N) of interactions with hashtags and randomly sample N interactions with replacement from a uniform distribution.
- `NGCF` [45]: Neural Graph Collaborative Filtering (NGCF) is an approach that integrates the collaborative filtering paradigm into the Graph Neural Network (GNN) framework. It builds an interaction graph based on the user-item interaction data and improves the collaborative signal by exploiting higher-order connections in the interaction graph. By propagating embeddings on this interaction graph, NGCF captures the collaborative filtering patterns and thus learns better user and item embeddings for recommendation.
- `LightGCN` [46]: This GCN-based method simplifies the standard design of GCN to make it more concise and appropriate for collaborative filtering and recommendation tasks. It jointly learns user and item embeddings through a user-item interaction graph. Unlike our proposed WLGCM, LightGCN uses binary user-item interactions, while ours uses weights.
- `CSMF` [16]: The proposed Coupled Sparse Matrix Factorization (CSMF) model is designed to infer the opinions of τ-silent users in online social networks. The model uses three matrices: one for the users' average sentiments on specific topics, another one for the collective average stance of communities on those topics, and a third one to describe individual user attributes. In our implementation of CSMF, community structures and user attributes are extracted using $\mathcal{G}_{\text{PathSim}}$ and $\mathcal{G}_{\text{Soc}}$. To optimize the loss function, we adopted the Adam algorithm [59].

We evaluate the performance of different variations of WLGCN against these baselines on the task of τ-silent users' stance detection for the target τ. We aim to analyze the impact of various methodological choices on the model's performance.

We propose two edge weight normalizations (see Sect. 4): one is the user-specific normalization as introduced in [23], highlighting hashtags that are particularly important to individual users. The second method, inspired by the TF-IDF principle, addresses the potential for some hashtags to be used ubiquitously, thus reducing their global value.

Also, we test model extensions by complementing the main user-hashtag interaction graph with various types of data, such as pre-trained hashtag embeddings, as well as observed and inferred user-user relations.

5.3 Evaluation protocol

Above, we introduced the WLGCN model and the datasets used in this study. Here, we evaluate the model's effectiveness in extracting useful information from the data. To this end, we test two key aspects: (1) the model's performance in predicting users' affinities toward each hashtag, thus reflecting their preferences within a topic, and (2) the model's accuracy in predicting each user's overall stance on a topic in the absence of explicit knowledge about users' opinions for this particular topic.

These two aspects above translate into an investigation of the prediction performance of our model at two levels: edge and user level. To test the second aspect (user-level prediction), we consider a specific topic: the Chilean constitutional referendum processes (2020 and 2022). For this, we rely on tweets that include referendum-related hashtags.

Given the nature of online social media platforms, we do not have a ground truth for actual τ-silent users' stances. For the user-level analysis, we first identify users who have engaged with referendum-related hashtags and remove their interactions with the referendum topic. In doing so, these users become *referendum*-silent for the model. However, we hypothesize that real silent users might behave differently than active users (e.g., links to other users or participation in other topics). We explore this potential effect by evaluating the model's performance with users at different activity levels. First, from the users that had interactions (i.e., edges in $\mathcal{G}_b$) with referendum-related hashtags (HT_τ), we divide them into three classes based on how many times they used these hashtags. That is, we group them into most active, least active, and middle (average activity). For testing, we select 10% of the users with a random sampling stratified over these three classes. All referendum-related interactions for these users are removed to artificially make them τ-silent. The removed edges are kept as ground truth. For the evaluation, we predict affinities for the removed interactions and use Equation (8) to compute the stances of users from both the ground truth and the predicted affinities. For the ground truth, the affinity $y(i,j) = T_{i,j}$ where $T_{i,j}$ is the number of times u_i used the hashtag j. We then report the accuracy of predicted users' stances.

Since we are representing the stances in a continuous space but evaluating the accuracy with discrete values/classes (i.e., negative, neutral, positive), we also measure the root mean square error (RMSE) [16]. RMSE helps us assess how close our predictions are to the ground truth classes before the transformation in Equation (8). Smaller RMSE indicates a better inference performance in the experiments.

For the edge level, we report the average performance of a 5-fold cross-validation analysis. We use the remaining edges in $\mathcal{G}_b$ (after removing the user-level-test set) and randomly select 80% for training and 20% for validation.

To assess the edge-level performance, we use normalized discounted cumulative gain (nDCG) and mean average precision (MAP) based on the top K recommendations with the highest affinities (K = 20 in our experiments).

6 Experiment results

We compared WLGCN's performance against the baselines and investigated the impact of integrating other features, such as pre-trained hashtags embedding and user-user relation graphs. The models are evaluated regarding nDCG@20 and MAP@20, as well as accuracy and RMSE for the three users' activity levels. Tables 1 and 2 summarize the results for the Entry_DS and Exit_DS datasets. *WLGCN* and *WLGCN (norm by user)*/*WLGCN (norm by*

Table 1 Edge- and User-level Performances in Entry_DS

Models	Edge		User				
	nDCG	MAP	Acc.	RMSE	Least	Middle	Most
Null Model	0.0176	0.0066	0.4628	0.7858	0.3672	0.4559	0.5474
CSMF	0.0457	0.0180	0.3929	0.7792	0.4061	0.4015	0.3717
NGCF	0.0539	0.0192	0.7192	0.5300	0.5862	0.7530	0.8184
LightGCN	0.0879	0.0377	0.7387	0.5079	0.5706	0.7389	0.9066
WLGCN (norm by user)	0.0728	0.0307	0.7445	0.4946	0.5955	0.7373	0.9006
WLGCN (norm by TF-IDF)	0.0668	0.0271	0.7399	0.4951	0.6004	0.7329	0.8864
WLGCN	0.0889*	0.0378*	0.7589	0.4891	0.6073	0.7593	0.9101*
WLGCN + (E_{HT})	0.0896*	0.0381*	0.7678*	0.4803*	0.6367*	0.7730*	0.8938
WLGCN + ($\mathcal{G}_{\mathrm{PathSim}}$)	0.0860*	0.0364*	0.7673*	0.4809*	0.6391*	0.7590*	0.9038
WLGCN + (E_{HT}, $\mathcal{G}_{\mathrm{PathSim}}$)	**0.0921***	**0.0393***	0.7915*	0.4540*	0.6772*	**0.8010***	0.8963*
WLGCN + ($\mathcal{G}_{\mathrm{Soc}}$)	0.0857$^+$	0.0357$^+$	0.7704	0.4790	0.6479$^+$	0.7569	0.9064
WLGCN + (E_{HT}, $\mathcal{G}_{\mathrm{Soc}}$)	0.0865*	0.0356*	0.7741*	0.4738$^+$	0.6314*	0.7761*	**0.9147***
WLGCN + ($\mathcal{G}_{\mathrm{PathSim}}$, $\mathcal{G}_{\mathrm{Soc}}$)	0.0838*	0.0357*	0.7757*	0.4720*	0.6539*	0.7710*	0.9022
WLGCN + (E_{HT}, $\mathcal{G}_{\mathrm{PathSim}}$, $\mathcal{G}_{\mathrm{Soc}}$)	0.0917*	0.0392$^+$	**0.7954***	**0.4507***	**0.6895***	0.7988*	0.8977*

We marked WLGCN variants' results that are statistically significant (t-test) compared to the baseline LightGCN: $*(p < 0.05)$, $+(p < 0.1)$.

Table 2 Edge- and User-level Performances in Exit_DS

Models	Edge		User				
	nDCG	MAP	Acc.	RMSE	Least	Middle	Most
Null Model	0.0358	0.0094	0.5596	0.6655	0.4583	0.5692	0.6514
CSMF	0.0557	0.0404	0.3955	0.7775	0.4211	0.3857	0.3806
NGCF	0.0728	0.0335	0.6312	0.6132	0.4833	0.5846	0.8257
LightGCN	0.0773	0.0388	0.6599	0.5766	0.5462	0.5956	0.8380
WLGCN (norm by user)	0.0754	0.0385	0.6543	0.5948	0.5083	0.5923	0.8624
WLGCN (norm by TF-IDF)	0.0776	0.0397	0.6494	0.5994	0.5333	0.5615	0.8532
WLGCN	0.0783	0.0395$^+$	0.6642$^+$	0.5860	0.5646$^+$	0.6015*	0.8267
WLGCN + (E_{HT})	0.0885$^+$	0.0449*	0.6702$^+$	0.5743$^+$	0.5622$^+$	0.6150*	0.8332
WLGCN + ($\mathcal{G}_{\mathrm{PathSim}}$)	0.0769$^+$	0.0392*	0.6795*	0.5658*	0.5757$^+$	0.6276*	0.8352
WLGCN + (E_{HT}, $\mathcal{G}_{\mathrm{PathSim}}$)	0.0902*	0.0458*	0.7065*	0.5422*	0.5733$^+$	0.6654*	0.8807*
WLGCN + ($\mathcal{G}_{\mathrm{Soc}}$)	0.0770$^+$	0.0390*	0.6896*	0.5578*	**0.5900**$^+$	0.6346*	0.8440
WLGCN + (E_{HT}, $\mathcal{G}_{\mathrm{Soc}}$)	**0.0908***	**0.0459***	**0.7118***	**0.5367***	0.5817*	0.6731*	0.8807*
WLGCN + ($\mathcal{G}_{\mathrm{PathSim}}$, $\mathcal{G}_{\mathrm{Soc}}$)	0.0749$^+$	0.0384*	0.6922*	0.5496*	0.5567*	0.6577*	0.8624*
WLGCN + (E_{HT}, $\mathcal{G}_{\mathrm{PathSim}}$, $\mathcal{G}_{\mathrm{Soc}}$)	0.0881*	0.0444*	0.7091*	0.5402$^+$	0.5567	**0.6808***	**0.8899***

We marked WLGCN variants' results that are statistically significant (t-test) compared to the baseline LightGCN: $*(p < 0.05)$, $+(p < 0.1)$.

TF-IDF) represent the vanilla version of our model using only the user-hashtag interaction graph.

First, we look at the effect of our weight normalization approaches on the performance of WLGCN. Intuitively, TF-IDF (i.e., the global strategy) seems better for the least active users, while user normalization (i.e., the local strategy) gives better results for users with middle and high activity. However, the integer weights strategy (i.e., count of interactions) without normalization (*WLGCN* in Tables 1 and 2) shows a more consistent performance across activity levels and datasets. Moreover, an improvement of WLGCN over LightGCN and the other baselines shows that adding weight to the user-hashtag graph contributes to predicting *referendum*-silent users' stances. Particularly for more challenging scenarios like in *Exit_DS*, where information for less active users is sparser, and the stances are closer to the center of the discussion, WLGCN offers a significant gain. Since the integer weights strategy shows the best performance, we proceeded to evaluate the contributions of the other features using this approach as our base.

Regarding users' activity level, we saw a steep increase in performance as we moved to higher levels. This marked difference in accuracy suggests that not only content production but also user behavior differs significantly between various activity levels, making more active users easier to classify. Simultaneously, these results highlight the ability of the graph-based models to extract relevant information from social affinities, other user-user interactions, and activity on other topics, if present. Interestingly, the matrix factorization-based model CSMF does not follow the same trend. Since CSMF puts more focus on the communities than on the individual users, our results suggest this strategy is unable to distinguish behaviors between users with different activity levels. Moreover, CSMF training requires considerably more resources (i.e., time and computational power) than the graph-based alternatives. Overall, the results show that WLGCN variants significantly outperform the baselines in terms of accuracy and RMSE for both Entry_DS and Exit_DS datasets.

Specifically, results show that incorporating pre-trained hashtag embeddings and user-user relationship graphs into the WLGCN model can improve its performance. For the Entry_DS dataset, combining WLGCN with E_{HT} and $\mathcal{G}_{\text{PathSim}}$ achieved the highest performance at the edge level. This reflects its superiority in ranking quality and accuracy of recommended hashtags. At the user level, WLGCN alone provides very accurate stance predictions for the most active accounts. Since these users are probably also relatively active in other topics, there is enough information in the user-hashtag interactions to estimate users' stances. Interestingly, E_{HT} improves performance for almost every metric. Moreover, we confirmed that integrating social interactions ($\mathcal{G}_{\text{PathSim}}$ or $\mathcal{G}_{\text{Soc}}$) boosts performance of the model for less active user groups. In particular, when we combine all features, we see the most accurate predictions for the overall and least active user groups. Overall, the tested additional features are particularly helpful in improving the predictions for the less active users compared to the vanilla version, adding over an 8% increase in performance.

As mentioned, the Exit_DS dataset is a more challenging case for most models. Although we collected more content (over twice as many tweets) for this period, we have significantly fewer hashtags. Figure 2b shows that in Exit_DS, fewer users posted tweets containing the annotated referendum hashtags. This lack of user interaction could subsequently decrease the model's predictive accuracy. In addition, discussions on topics parallel to the referendum appear to have been more diffuse (see Fig. 1b and 1c, which makes our interaction graphs sparser. Finally, the outcome of the second referendum showed that people's stances regarding the new constitution were more equally distributed and closer to the center compared to the first one. This may indicate more undecided users in the months before the vote.

Nevertheless, for the Exit_DS dataset, WLGCN with different input combinations also outperformed the baselines. The consistent performance of WLGCN underscores its relevance and effectiveness in hashtag affinity recommendation. The inclusion of E_{HT} consistently improves the edge-level performance. In particular, the combination with E_{HT} and $\mathcal{G}_{\text{Soc}}$ offers the best results at this level with the highest nDCG and MAP scores. Again, social characteristics play a relevant role. Each of the two social graphs alone helps to make better predictions for the least active groups, but $\mathcal{G}_{\text{Soc}}$ offers the best accuracy for this

Table 3 Performance including/excluding the Projection & Fusion (P&F) layers

	nDCG	MAP	Acc.	RMSE	Least	Middle	Most
Dataset	Entry_DS						
WLGCN + ($\mathcal{G}_{\text{PathSim}}$) w/o P&F	0.0854	0.0355	0.7513	0.4984	0.6180	0.7333	0.9026
WLGCN + ($\mathcal{G}_{\text{PathSim}}$)	0.0860*	0.0364*	0.7673*	0.4809	0.6391*	0.7590*	0.9038
WLGCN + ($\mathcal{G}_{\text{Soc}}$) w/o P&F	0.0850	0.0353	0.7488	0.5009	0.6217	0.7294	0.8951
WLGCN + ($\mathcal{G}_{\text{Soc}}$)	0.0857*	0.0357*	0.7704*	0.4790*	0.6479*	0.7569	0.9064*
WLGCN + ($\mathcal{G}_{\text{PathSim}}$, $\mathcal{G}_{\text{Soc}}$) w/o P&F	0.0832	0.0343	0.7476	0.5022	0.6180	0.7333	0.8914
WLGCN + ($\mathcal{G}_{\text{PathSim}}$, $\mathcal{G}_{\text{Soc}}$)	0.0838	0.0357	0.7757*	0.4720	0.6539*	0.7710	0.9022
Dataset	Exit_DS						
WLGCN + ($\mathcal{G}_{\text{PathSim}}$) w/o P&F	0.0741	0.0378	0.6569	0.5924	0.5083	0.6000	0.8624
WLGCN + ($\mathcal{G}_{\text{PathSim}}$)	0.0769^{+}	0.0392*	0.6795^{+}	0.5658	0.5757*	0.6276^{+}	0.8352
WLGCN + ($\mathcal{G}_{\text{Soc}}$) w/o P&F	0.0772	0.0392	0.6605	0.5901	0.5083	0.5923	0.8807
WLGCN + ($\mathcal{G}_{\text{Soc}}$)	0.0770^{+}	0.0390	0.6896^{+}	0.5578*	0.5900^{+}	0.6346^{+}	0.8440
WLGCN + ($\mathcal{G}_{\text{PathSim}}$, $\mathcal{G}_{\text{Soc}}$) w/o P&F	0.0742	0.0380	0.6552	0.5860	0.5420	0.5914	0.8321
WLGCN + ($\mathcal{G}_{\text{PathSim}}$, $\mathcal{G}_{\text{Soc}}$)	0.0749*	0.0384*	0.6922^{+}	0.5496^{+}	0.5567^{+}	0.6577	0.8624

Note: "w/o P&F" means the model doesn't include the P&F layers. We marked WLGCN variants' results that are statistically significant (t-test) compared to the same variant but without the P&F layers: $*(p < 0.05)$, $+(p < 0.1)$.

group. The robustness of the ($\mathcal{G}_{\text{Soc}}$) input was evident, emphasizing its relevance in user stance prediction.

The analytical results from both datasets illustrate the strengths of the WLGCN model, especially when augmented with diverse inputs. At the edge level, the inclusion of E_{HT} in combination with $\mathcal{G}_{\text{PathSim}}$ or $\mathcal{G}_{\text{Soc}}$ offers significant advantages in hashtag recommendation quality. When shifting to a user-level stance prediction, especially for users with lower activity levels, the inclusion of relational data (e.g., $\mathcal{G}_{\text{Soc}}$) appears to be a prevalent influencing factor.

6.1 Projection & fusion of user embeddigs

Table 3 presents an analysis of the performance of the WLGCN model, comparing variants that include and exclude the Projection & Fusion (P&F) layers across our two datasets. On average, including the P&F layers in the WLGCN model leads to consistent improvements across the board. Note that this layer aims to align the user embeddings from the various graphs. As discussed before, the groups with lower activity levels benefit the most from including these complementary social graphs. Correspondingly, the P&F mechanism tends to show a stronger effect for lower activity levels. Thus, this layer helps us to improve the performance of the WLGCN model further, underscoring its importance in the overall architecture.

7 Analysis and discussion

So far, we evaluated our models' performance on various metrics to establish its competitiveness. In this section, we use the model's predictions alongside the learned embeddings of users and hashtags to confirm some of our key assumptions. First, we verify that learned hashtag embeddings can accurately capture the polarization inherent in topics such as referendums. We also investigate the impact of hashtag manual annotation efforts on the accuracy of model predictions. Furthermore, we verify that the model's predicted user-hashtag affinities are reliable for differentiating between referendum opponents and supporters, providing insights from different perspectives. Our analysis aims to reveal the

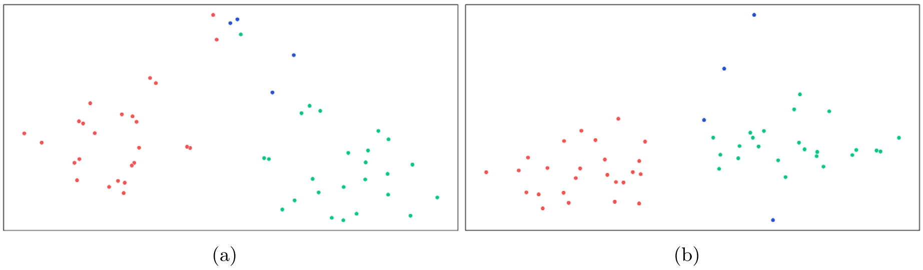


Figure 4 Latent Space Distribution of Referendum-Related Hashtags in Exit_DS. Hashtag stances are indicated by color: Green for Positive, Red for Negative, and Blue for Neutral. Figure 4a: Initial Pre-trained Hashtag Embeddings in Exit_DS. Figure 4b: Hashtag Embeddings Post-WLGCN Training in Exit_DS

model's potential to segment users' stances based on their interactions with referendum-related content. Finally, we explore the differences between *referendum-silent* users and their active counterparts engaged in the discourse. This exploration aims to identify patterns in tone, related interests, and stance distribution that delineate these user groups. For each dataset, we use the WLGCN variation with the best overall performance at the user level (which also coincides with the best performance for the least active users).

7.1 Hashtag-based stance representation

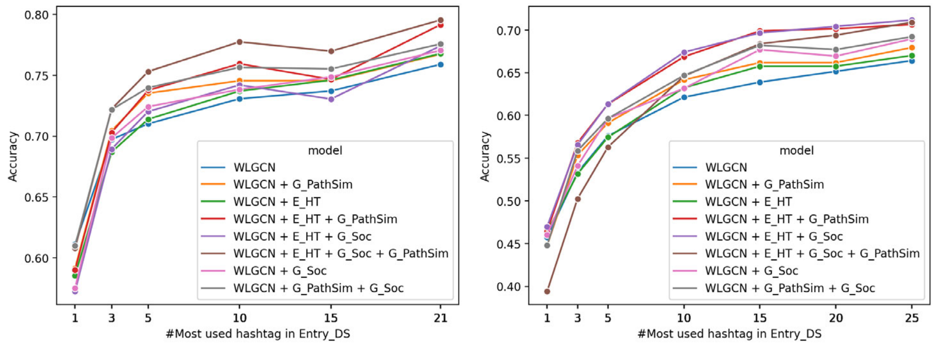
One key assumption in our study is that the hashtags used within a polarized topic like a referendum can be characteristic of different popular viewpoints. In our case, this premise is supported by a t-SNE projection [62] of the, otherwise 100-dimensional, learned representations of the referendum-related hashtags (see Fig. 4). For comparison, we analyzed the hashtags' pre-trained representations (Fig. 4a) and the fine-tuned embeddings learned by the WLGCN (Fig. 4b). We observed that the hashtag groups (POS, NEG, and NEUTRAL) were distinctly separated. Notably, the fine-tuned representation (and, thus, the one used for stance prediction) shows better grouping with clearer boundaries between classes. To further support the improvement after WLGCN training, we measured inter- and intra-group similarities for the hashtag classes regarding cosine similarities and distances before and after WLGCN training. Increased cosine similarities within each group (between 8% and 18%) indicate tighter clusters. Similarly, increased cosine distances between groups (between 36% and 40%) suggest a better separation among stances.

These observations substantiate our initial assumption that hashtags can serve as reliable indicators of users' stances since our unsupervised approach can correctly capture and differentiate their semantic meaning.

7.2 Annotation effort analysis

Another advantage of our approach compared to previous works (e.g., [16]) is the minimal annotation required. As presented before, only a set of hashtags related to the topic of interest must be identified. Still, the model can profit from other discussions and interactions potentially outside this topic. We experiment with user-level stance prediction and a growing number of annotated hashtags to further investigate the impact of an increased expert effort. The results are shown in Fig. 5.

For the results presented above, we used all annotated referendum-related hashtags to evaluate the model's performance (i.e., estimated affinity from each user to all hashtags).



(a) (b)

Figure 5 Impact of hashtag annotations on the user-level prediction accuracy. Figure 5a: Prediction Accuracy Using Entry_DS. Figure 5b: Prediction Accuracy Using Exit_DS

However, we also examined the accuracy variability in the proposed model at the user level when different numbers of hashtags are annotated. In Fig. 5, the x-axes represent a prediction of the users' stances when including only x annotated referendum-related hashtags for each stance class ($1 \leq x \leq \min(|\text{POS}|, |\text{NEG}|)$). Note that we have ($|\text{POS}| = 14$, $|\text{NEG}| = 21$, $|\text{NEU}| = 5$) in *Entry_DS* and ($|\text{POS}| = 25$, $|\text{NEG}| = 25$, $|\text{NEU}| = 6$) in *Exit_DS*. So, we don't consider the neutral ones because only a few were found.

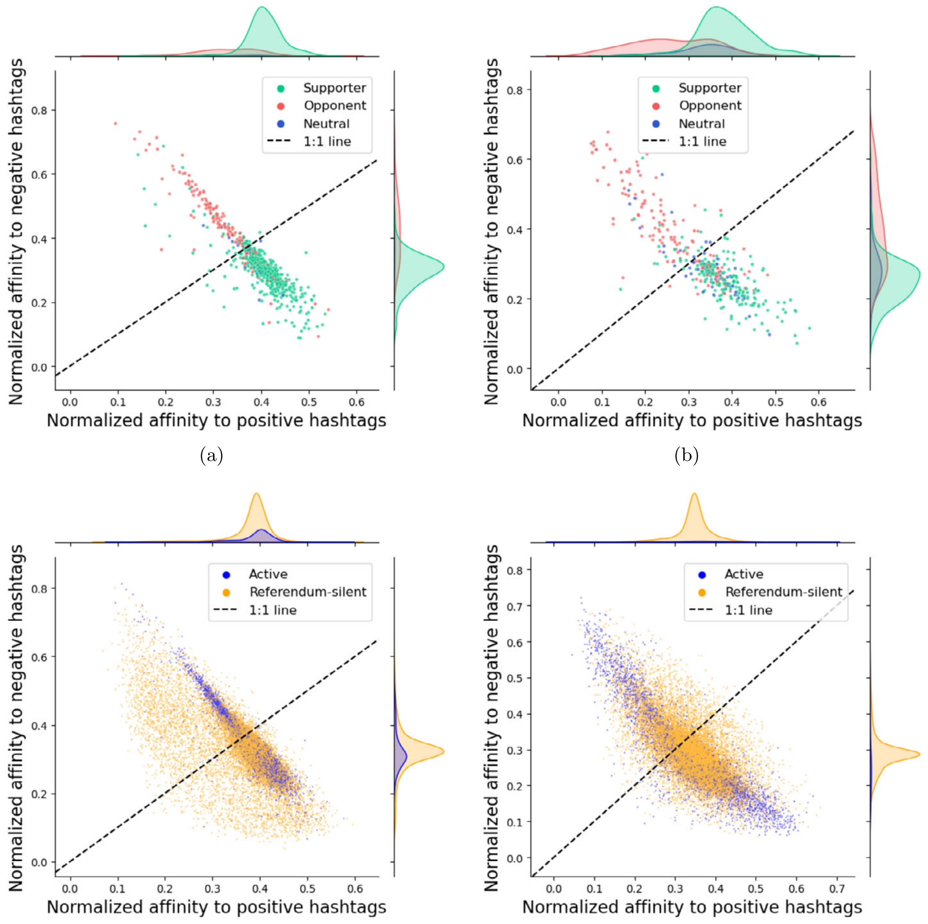
In each case, we select the top-x most used hashtags in each class. These should represent the easiest ones to identify by the experts and thus require the least effort.

Both figures show similar behavior. As expected, a higher number of annotations leads to higher accuracy. However, the increase in accuracy slows down after five hashtags and tends to become asymptotic as the number of annotations increases, especially in *Exit_DS*. This tendency strengthens the practical implications in the applicability of our model as it could further simplify our approach. For example, we might only know some related hashtags for a new topic, or they could evolve in time. An expert could only need to annotate a small sample of the most used hashtags related to that topic. As a result, the performance should remain stable without heavy annotation work.

7.3 Stance distribution

The main goal of our method is the prediction of τ-silent users' stances. However, since our model is able to represent users' stances in a continuous space of discussion, we can also evaluate how these τ-silent accounts' distribution compares to that of the actives users. This should shed some light into how representative a purely content-based analysis would be of the extended online community.

For the stance classification, we first applied user-wise min-max normalization to affinities of the *referendum*-related hashtags, as shown in Equation (8). We also calculate the average normalized affinity of each user to every stance class (i.e., POS, NEG, NEUTRAL). This allows us to represent users in a continuous space representing the *referendum* discussion on Twitter. In Fig. 6, we show the plane defined by the POS and NEG stances, where each dot represents a user. In Figs. 6a and 6b, we present the learned stances for the 10% of the users kept for testing from the Entry_DS and Exit_DS, respectively. The colors of the nodes in these figures represent the ground truth annotations. The proposed approach offers an effective classifier for users' stances. Furthermore, users on (or close



(a)

(b)

(c)

(d)

Figure 6 Users' referendum stance distributions. The affinities are normalized using Equation (8) for POS and NEG hashtags. Figure 6a: Entry_DS test users' referendum stance distribution. Figure 6b: Exit_DS test users' referendum stance distribution. Figure 6c: Entry_DS users' referendum stance distribution. Figure 6d: Exit_DS users' referendum stance distribution

to) the 1:1 diagonal may be considered undecided about the target, while those further from the diagonal line represent more extreme stances. Moreover, accounts closer to the origin may represent users with lower political engagement, while users in the diagonal but moving away from the origin may be undecided voters but more politically active.

Figure 6 shows that users are distributed on a range of stances rather than in a bipolar grouping. Generally, most users appear to be situated along a diagonal axis that seems to characterize polarized discussions.[8] In contrast to the more linear active discussion from the Exit_DS (see Fig. 6d), the Entry_DS shows an elongated disc shape (see Fig. 6c) that may represent some multifaceted debates in public forums (e.g., for the first plebiscite the people should vote not only for a new constitution but also on what kind of legislative body would write it).

[8] Note that, in our case, this effect is emphasized by our enforcement of a three-class stance analysis during the affinity normalization.

Interestingly, these more complex opinion patterns are more prominent for *referendum*-silent users. As mentioned, although we work with the three standard stance classes, our approach should generalize to hashtags annotated with more complex topics/stances, enabling a deeper understanding of users' preferences. For example, a third pole of stances can be represented in the same way by adding another dimension (i.e., 3-D plot). We leave a more in-depth analysis of multipolar cases to future work.

Finally, from Figs. 6c and 6d, we observe the distribution differences between active and *referendum*-silent users. Although active users cover most of the discussion range, the central tendencies and variability can differ from lurkers, reinforcing our initial motivation of potential biases introduced by relying solely on active users.

7.4 Characteristic tone and other interests

We are also interested in how our WLGCN can help further characterize online communities and users aligned with a particular stance. For this, we analyze the sentiment associated with various groups of users to illustrate other differences in how users engage online. To perform sentiment analysis on our Spanish tweet collection, we used an off-the-shelf Spanish model from Hugging Face,[9] which is explicitly trained on Spanish tweets [63]. To emphasize the contrasts in the tone from various groups (i.e., opponents and supporters in various activity levels), we calculate the relative sentiment for each referendum-related tweet by subtracting the average sentiment of all referendum tweets [64]. Then, we associate an average relative tone to each hashtag based on the corresponding tweets' relative sentiment. Finally, to account for *referendum*-silent users' positions, we find the characteristic tone as the average relative sentiment of the hashtags supporting each stance weighted by their average affinities among users in a group. Figure 7 shows a general trend for both supporters and opponents that associates a relatively more positive (or less negative) tone with positive hashtags. Note that we keep more contextual information (e.g., emojis and stopwords) for the affective analysis, and the language proper to each stance is already emotionally-charged (i.e., rejection vs. acceptance). This emphasizes the unreliability of sentiment analysis for determining stances. However, we see that Opponents tend to have less positive tone for the positive hashtags and less negative for the negative hashtags. As expected, the sentiment polarity aligns with the users' stance for neutral hashtags representing the more general discussion.

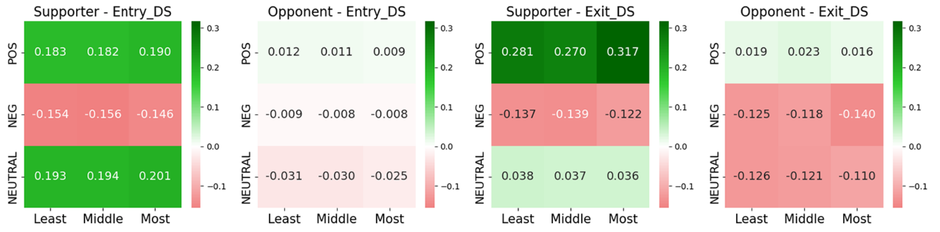


Figure 7 Sentiment Analysis of Referendum Supporters and Opponents Across Activity Levels. This set of heatmaps presents a sentiment analysis of supporters and opponents towards referendum-related hashtags categorized by POS (Positive), NEG (Negative), and NEUTRAL stances. Each heatmap is divided into three activity levels—Least, Middle, and Most Active—illustrated across three rows for hashtag stances and three columns for user activity levels, providing a comprehensive view of sentiment distribution

[9] https://huggingface.co/pysentimiento/robertuito-sentiment-analysis.

Also, we see a relatively stable tone when comparing users with different activity levels. An interesting observation is the shift in the opponents' tone between plebiscites when using NEG hashtags. This suggests a change in the discourse from a more engaging discussion in the first stage to a more hostile narrative for the second plebiscite.

Another way to characterize different groups is by analyzing what other hashtags (outside the referendum) have the highest average affinity. We computed the top 50 hashtags with the highest average predicted affinity for four groups: Silent-Opponents, Active-Opponents, Silent-Supporters, and Active-Supporters. For example, for the Entry_DS, supporters of the referendum are also in favor of the social movement that preceded it (#chiledesperto), in favor of other social reforms (e.g., to the pension system (#nomasafp) and gender equality (#8m2020)), and against the conservative government of president Piñera (#renunciapiñera). In particular, Active-Supporters are more interested in other political and social issues, while Silent-Supporters tend to have higher affinities to everyday concerns such as the pandemic, salaries, and sports. On the other hand, opponents are against the social movement and in favor of the police, which usually clashed with the protesters (#yoapoyocarabineros), and seem more conservative (#fueracomunistasdechile). Also, Silent-Opponents are less focused on politics and more interested in TV, news, and sports than Active-Opponents. For the Exit_DS, referendum's opponents become part of the opposition to the new liberal government (#renunciaboric). Notably, the highest affinity hashtags reflect the shift in the discourse for the opponents compared to 2020. We see an increase in their affinity to derogatory hashtags that is especially prominent for Active-Opponents (e.g., #merluzoinepto).

These elements give insights into the composition of the groups. They help us identify differences but also their commonalities. For example, supporters and opponents of the referendum favored lockdown measures during the COVID-19 pandemic (#quedateencasa). This information can be beneficial for reaching all sides.

8 Conclusions

Our analysis reveals that the WLGCN model, in combination with our normalized affinity-driven classification, improved performance over baseline models in predicting τ-silent user stances. This shows the potential of using social media activity beyond the topic of interest to address this task. Furthermore, this approach allows us to make a multidimensional depiction of different user groups and understand their other preferences, characteristic tones, and coverage of the discussion space.

An interesting observation is the majority stance difference between our two datasets. While for *Entry_DS* the majority of the users are predicted as in favor of the referendum, *Exit_DS* shows a distribution centered much closer to the center of the discussion space. These results closely reflect the actual output of the referendum process. Given that both datasets are focused on the same topic and cover periods only two years apart, they highlight the dynamic nature of public opinion on social and political issues and, thus, the importance of a readily applicable, large-scale, and accurate representation of stances in public discussions online.

Since hashtags are a standard feature across multiple social media platforms, our approach should also be extensible to other sources. Multi-source and longitudinal dynamics of public opinions are interesting lines of research for future work.

Note, however, that hashtag hijacking poses a challenging problem for a hashtag-based approach to stance detection. Although it does not seem to have affected our case study (see Fig. 2a), this limitation has to be considered case-by-case, as our model may only capture some of its nuances.

Another imperative aspect to consider while interpreting social media-based results is the coverage error associated with online platforms [65], as users of various platforms may have distinct preferences and behaviors that could bias our findings [66]. For example, some demographic groups may be underrepresented, including age, socioeconomic status, or geographic location [67]. By extending the analysis to τ-silent users through collaborative filtering, the model tries to reach a larger user base. However, this might not be enough if the entire platform's demographics are biased. For example, in the U.S., Twitter tends to have a larger ratio of younger adults and a more liberal leaning compared to the general public [68]. Considering these demographic dimensions could add valuable insights into the characterization of the communities. Nevertheless, privacy and ethical considerations should not be overlooked. Demographic information is usually considered among the most sensitive personal data. Following the principle of data minimization, we limited ourselves to the collection of personal information that was directly relevant and necessary to accomplish the specified purpose of this study. Since our model does not require features such as gender or age for the stance inference, we prefer not to include them in our analysis. In specific cases, analysts and practitioners should consider using multiple sources and longitudinal analyses to mitigate the impact of coverage error and account for potential stance shifts, thereby gaining a more representative and comprehensive understanding of the case study.

Similarly, we recognize that user behavior on social media can vary based on regional and cultural contexts. Therefore, although previous works have proven that lurkers constitute a significant share of online communities [4, 69], general statistics of user behavior on Twitter may not directly correspond to the specific dynamics of Twitter use in a particular region. In our analysis, we also had to make practical choices when filtering and processing the data (e.g., the number of news outlet friends or the average number of daily tweets). These thresholds aim to strike a balance between the inclusivity of average users and computational efficiency by reducing the spatial complexity of our methods. Further research with region-specific data would help draw more accurate conclusions about the Chilean (or any targeted region's) Twitter landscape.

In general, our findings underscore the potential of the proposed methods in various scenarios related to stance detection and social media analysis while also highlighting the importance of addressing the social media coverage error and considering dynamic changes in public opinion to ensure more generalizable conclusions.

Appendix A: Distribution of the number of Chilean news outlets followed by Twitter users

Figure 8 shows the distribution of the number of Chilean news outlets followed by users in our dataset. This distribution is heavily skewed, with almost 60% of users following ten or fewer newspapers (mean: 11.05, std: 7.99). The median in our dataset is nine news outlets.

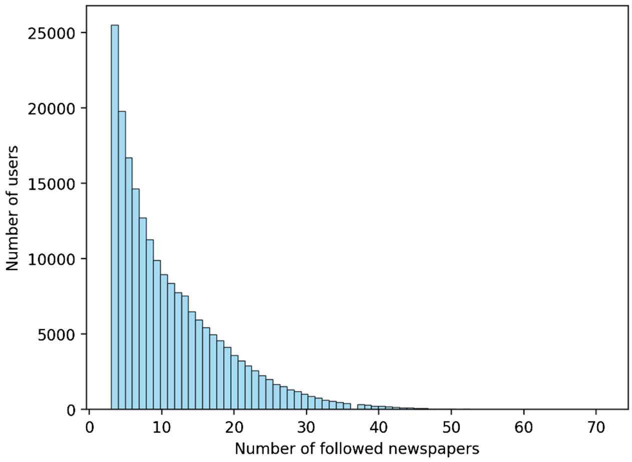


Figure 8 Distribution of the number of Chilean news outlets followed simultaneously among media followers on Twitter

Appendix B: Annotated referendum-related hashtags

Table 4 Annotated referendum-related hashtags

Stance	*Entry_DS*
POS	apruebo, apruebo26abril, apruebocc, apruebochiledigno, aprueboconvencionconstitucional, aprueboganaenoctubre, apruebonuevaconstitucion, apruebosinmiedo, nuevaconstitucionparachile, yoapruebo, yoapruebocc, yoapruebolanuevaconstitucion, yoapruebonuevaconstitucion, yovotoapruebo
NEG	lacallerechaza, noalanuevaconstitucion, porchileyorechazo, rechazo, rechazocrece, rechazoganaenoctubre, rechazoganasivotamos, rechazoganasivotamostodos, rechazonuevaconstitucion, rechazoporchile, rechazosalvaachile, rechazosalvachile, rechazosinmiedo, rechazotutongo, rechazoynulo, retrazo, votarechazo, votorechazo, yorechazo, yorechazonuevaconstitucion, yovotorechazo
NEUTRAL	convencionconstitucional, convencionconstituyente, nuevaconstitucion, plebiscito2020, plebiscitochile
Stance	*Exit_DS*
POS	aprobamosfelices, aprobareshumano, apruebaserahermoso, apruebaxchile, apruebazo, apruebo, apruebo4deseptiembre, aprueboconesperanza, apruebocrece, apruebodesalida, aprueboel4deseptiembre, apruebofeliz, apruebonuevaconstitucion, aprueboparaquenuncamasenchile, aprueboplebicitodesalida, apruebosincondiciones, apruebosinmentiras, apruebosinmiedo, apruebounchilemejor, aprueboxamor, chilevotaapruebo, laconvencionsedefiende, mivotonocambia, yoapruebo, yoapruebofeliz
NEG	circoconstituyente, convencionculia, rechazo, rechazoconesperanza, rechazoconfuerza, rechazocontodos, rechazocrece, rechazodesalida, rechazodesalida2022, rechazoel4deseptiembre, rechazoelmamarracho, rechazoelmamarrachocomunista, rechazoelplurimamarracho, rechazoganael4deseptiembre, rechazoladestrucciondechile, rechazopopular, rechazoporamorachile, rechazoporchile, rechazosalvaachile, rechazosalvachile, rechazotransversal, rechazoxamorachile, rechazoxchile, rechazoypunto, yorechazo
NEUTRAL	100indecisos, convencionconstitucional, convencionconstituyente, nuevaconstitucion, plebiscitodesalida, radiografiaconstitucional

Abbreviations

SM, Social Media; RS, recommendation system; MF, matrix factorization; GCN, graph convolutional networks; CSMF, coupled sparse matrix factorization; NGCF, Neural Graph Collaborative Filtering; LightGCN, Light Graph Convolution Network; WLGCN, Weighted Light Graph Convolution Network; TF-IDF, term frequency-inverse document frequency; nDCG, normalized discounted cumulative gain; MAP, mean of the average precision.

Acknowledgements

We want to thank our colleagues from the multidisciplinary CRiTERIA team for the valuable discussions and feedback ranging from ethics to practical applications of silence users' stance detection. We also thank the anonymous reviewers for their constructive feedback and insightful comments.

Author contributions

ZZ and EE conceived and designed the analysis. ZZ collected data and conducted the experiments. ZZ and EE performed the analysis and wrote the paper. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This paper is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101021866 (CRiTERIA).

Data availability

The datasets supporting the conclusions of this article are available in a Github repository at https://github.com/imzzhou/StanceInferenceInTwitter.git.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 13 November 2023 Accepted: 22 March 2024 Published online: 04 April 2024

References

1. Benevenuto F, Rodrigues T, Cha M, Almeida V (2009) Characterizing user behavior in online social networks. In: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement. IMC '09. Assoc. Comput. Mach., New York, pp 49–62. https://doi.org/10.1145/1644893.1644900
2. McClain C, Widjaya R, Rivero G, Smith A (2021) The behaviors and attitudes of u.s. adults on twitter. Internet & Tech. Pew Research Center, Available from https://www.pewresearch.org/internet/2021/11/15/the-behaviors-and-attitudes-of-u-s-adults-on-twitter/ (Accessed 04-Apr-2023)
3. Antelmi A, Malandrino D, Scarano V (2019) Characterizing the behavioral evolution of Twitter users and the truth behind the 90-9-1 rule. In: Companion proceedings of the 2019 World Wide Web Conference. WWW '19. Assoc. Comput. Mach., New York, pp 1035–1038. https://doi.org/10.1145/3308560.3316705
4. Gong W, Lim E-P, Zhu F (2015) Characterizing silent users in social media communities. In: Ninth international AAAI conference on web and social, Media
5. Gong W, Lim E-P, Zhu F, Cher PH (2016) On unravelling opinions of issue specific-silent users in social media. In: Proceedings of the international AAAI conference on web and social media, vol 10, pp 141–150
6. Elejalde E, Ferres L, Herder E (2018) On the nature of real and perceived bias in the mainstream media. PLoS ONE 13(3):1–28
7. Paul D, Li F, Teja MK, Yu X, Frost R (2017) Compass: spatio temporal sentiment analysis of us election what Twitter says! In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1585–1594
8. Di Giovanni M, Brambilla M (2021) Content-based stance classification of tweets about the 2020 Italian constitutional referendum. In: SocialNLP@ NAACL 2021, pp 14–23
9. Hampton KN, Rainie H, Lu W, Dwyer M, Shin I, Purcell K (2014) Social media and the 'spiral of silence'. PewResearchCenter, Washington
10. Sleeper M, Balebako R, Das S, McConahy AL, Wiese J, Cranor LF (2013) The post that wasn't: exploring self-censorship on Facebook. In: 2013 conference on Computer Supported Cooperative Work. CSCW '13. Assoc. Comput. Mach., New York, pp 793–802. https://doi.org/10.1145/2441776.2441865
11. Shin D-I, Lim Y-W, Kwahk K-Y (2022) Sns users' opinion expression: focusing on suppression effect in spiral of silence. Telemat Inform 72:101859
12. Mizan AS, Ishtiaque Ahmed S (2019) Silencing the minority through domination in social media platform: Impact on the pluralistic bangladeshi society. ELCOP Yearbook of Human Rights (2018)
13. International A (2018) Toxic Twitter: the silencing effect. https://www.amnesty.org/en/latest/news/2018/03/online-violence-against-women-chapter-5-5/
14. Dhrodia A (2018) Unsocial media: a toxic place for women. IPPR Progress Rev 24(4):380–387
15. Hoang T-A, Cohen WW, Lim E-P, Pierce D, Redlawsk DP (2013) Politics, sharing and emotion in microblogs. In: 2013 IEEE/ACM international conference on Advances in Social Networks Analysis and Mining (ASONAM 2013). IEEE, Los Alamitos, pp 282–289
16. Wang L, Niu J, Liu X, Mao K (2019) The silent majority speaks: inferring silent users' opinions in online social networks. In: The World Wide Web Conference. WWW '19. Assoc. Comput. Mach., New York, pp 3321–3327. https://doi.org/10.1145/3308558.3313423
17. Graells-Garrido E, Baeza-Yates R, Lalmas M (2020) Every colour you are: stance prediction and turnaround in controversial issues. In: 12th ACM conference on web science, pp 174–183
18. He X, Liao L, Zhang H, Nie L, Hu X, Chua T-S (2017) Neural collaborative filtering. In: Proceedings of the 26th international conference on world wide web, pp 173–182
19. Bestvater SE, Monroe BL (2022) Sentiment is not stance: target-aware opinion classification for political text analysis. Polit Anal, 1–22

20. Xiao Z, Song W, Xu H, Ren Z, Sun Y (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, pp 2258–2268
21. Tan C, Lee L, Tang J, Jiang L, Zhou M, Li P (2011) User-level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1397–1405
22. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27(1):415–444
23. Zhou Z, Elejalde E (2023) Stance inference in Twitter through graph convolutional collaborative filtering networks with minimal supervision. In: Companion proceedings of the ACM web conference 2023. WWW '23 companion. Assoc. Comput. Mach., New York, pp 1030–1038. https://doi.org/10.1145/3543873.3587640
24. Quraishi M, Fafalios P, Herder E (2018) Viewpoint discovery and understanding in social networks. In: Proceedings of the 10th ACM conference on Web Science. WebSci '18. Assoc. Comput. Mach., New York, pp 47–56. https://doi.org/10.1145/3201064.3201076
25. Burfoot C, Bird S, Baldwin T (2011) Collective classification of congressional floor-debate transcripts. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, pp 1506–1515
26. Reyero TM, Beiró MG, Alvarez-Hamelin JI, Hernández L, Kotzinos D (2021) Evolution of the political opinion landscape during electoral periods. EPJ Data Sci 10(1):31
27. Sridhar D, Getoor L, Walker M (2014) Collective stance classification of posts in online debate forums. In: Joint workshop on social dynamics and personal attributes in social media, pp 109–117
28. Conforti C, Berndt J, Pilehvar MT, Giannitsarou C, Toxvaerd F, Collier N (2022) Incorporating stock market signals for Twitter stance detection. In: Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: long papers), pp 4074–4091
29. Kalimeri K, Beiró MG, Urbinati A, Bonanomi A, Rosina A, Cattuto C (2019) Human values and attitudes towards vaccination in social media. In: Companion proceedings of the 2019 world wide web conference, pp 248–254
30. Baldwin T, Cook P, Lui M, MacKinlay A, Wang L (2013) How noisy social media text, how diffrnt social media sources? In: Proceedings of the sixth international joint conference on natural language processing. Asian Federation of Natural Language Processing, Nagoya, pp 356–364
31. Wildemann S, Niederée C, Elejalde E (2023) Migration reframed? A multilingual analysis on the stance shift in Europe during the Ukrainian crisis. In: Proceedings of the ACM web conference 2023. WWW '23. ACM, New York. https://doi.org/10.1145/3543507.3583442
32. Magdy W, Darwish K, Abokhodair N, Rahimi A, Baldwin T (2016) #isisisnotislam or #deportallmuslims? Predicting unspoken views. In: Proceedings of the 8th ACM conference on Web Science. WebSci '16. Assoc. Comput. Mach., New York, pp 95–106. https://doi.org/10.1145/2908131.2908150
33. Kobellarz JK, Broćić M, Graeml AR, Silver D, Silva TH (2022) Reaching the bubble may not be enough: news media role in online political polarization. EPJ Data Sci 11(1):47
34. Vilella S, Lai M, Paolotti D, Ruffo G (2020) Immigration as a divisive topic: clusters and content diffusion in the Italian Twitter debate. Future Internet 12(10):173
35. Jackson SJ, Foucault Welles B (2015) Hijacking# mynypd: social media dissent and networked counterpublics. J Commun 65(6):932–952
36. Xu S, Zhou A (2020) Hashtag homophily in Twitter network: examining a controversial cause-related marketing campaign. Comput Hum Behav 102:87–96
37. Garimella VRK, Weber I (2014) Co-following on Twitter. In: Proceedings of the 25th ACM conference on hypertext and social media, pp 249–254
38. Volkova S, Coppersmith G, Van Durme B (2014) Inferring user political preferences from streaming communications. In: Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: long papers), pp 186–196
39. Yang J, McAuley J, Leskovec J (2013) Community detection in networks with node attributes. In: 2013 IEEE 13th international conference on data mining. IEEE, Los Alamitos, pp 1151–1156
40. Riquelme F, González-Cantergiani P (2016) Measuring user influence on Twitter: a survey. Inf Process Manag 52(5):949–975
41. Sun Y, Han J, Yan X, Yu PS, Wu T (2011) Pathsim: meta path-based top-k similarity search in heterogeneous information networks. Proc VLDB Endow 4(11):992–1003
42. Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. Computer 42(8):30–37
43. Rendle S, Krichene W, Zhang L, Anderson J (2020) Neural collaborative filtering vs. matrix factorization revisited. In: Fourteenth ACM conference on recommender systems, pp 240–248
44. Anelli VW, Bellogín A, Di Noia T, Pomo C (2021) Reenvisioning the comparison between neural collaborative filtering and matrix factorization. In: 15th ACM conference on recommender systems, pp 521–529
45. Wang X, He X, Wang M, Feng F, Chua T-S (2019) Neural graph collaborative filtering. In: 42nd international ACM SIGIR conference on research and development in information retrieval, pp 165–174
46. He X, Deng K, Wang X, Li Y, Zhang Y, Wang M (2020) Lightgcn: simplifying and powering graph convolution network for recommendation. In: Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval, pp 639–648
47. Mei D, Huang N, Li X (2021) Light graph convolutional collaborative filtering with multi-aspect information. IEEE Access 9:34433–34441
48. Fan W, Ma Y, Li Q, He Y, Zhao E, Tang J, Yin D (2019) Graph neural networks for social recommendation. In: The world wide web conference, pp 417–426
49. Elejalde E, Ferres L, Schifanella R (2019) Understanding news outlets' audience-targeting patterns. EPJ Data Sci 8(1):16
50. Yang K-C, Ferrara E, Menczer F (2022) Botometer 101: social bot practicum for computational social scientists. J Comput Soc Sci 5(2):1511–1528
51. Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. Commun ACM 59(7):96–104

52. Hecht B, Hong L, Suh B, Chi EH (2011) Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In: Proceedings of the SIGCHI conference on human factors in computing systems. CHI '11. Assoc. Comput. Mach., New York, pp 237–246. https://doi.org/10.1145/1978942.1978976
53. Field A, Park CY, Theophilo A, Watson-Daniels J, Tsvetkov Y (2022) An analysis of emotions and the prominence of positivity in# blacklivesmatter tweets. Proc Natl Acad Sci 119(35):2205767119
54. Mejova Y, Crupi G, Lenti J, Tizzani M, Kalimeri K, Paolotti D, Panisson A (2023) Echo chambers of vaccination hesitancy discussion on social media during covid-19 pandemic XX ISA World Congress of Sociology (June 25-July 1, 2023). ISA
55. Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. Trans Assoc Comput Linguist 5:135–146
56. HaCohen-Kerner Y, Miller D, Yigal Y (2020) The influence of preprocessing on text classification using a bag-of-words representation. PLoS ONE 15(5):0232525
57. Kumar GK, Nandakumar K (2022) Hate-clipper: multimodal hateful meme classification based on cross-modal interaction of clip features. arXiv preprint. arXiv:2210.05916
58. Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L (2012) Bpr: bayesian personalized ranking from implicit feedback. arXiv preprint. arXiv:1205.2618
59. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint. arXiv:1412.6980
60. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: Thirteenth international conference on artificial intelligence and statistics, pp 249–256
61. Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69(2):026113
62. Van der Maaten L, Hinton G (2008) Visualizing data using t-sne. J Mach Learn Res 9(11)
63. Pérez JM, Furman DA, Alemany LA, Luque F (2021) Robertuito: a pre-trained language model for social media text in spanish. arXiv preprint. arXiv:2111.09453
64. Rudra K, Backfried G, Shaltev M, Niederée C, Elejalde E (2021) My eu = your eu? Differences in the perception of European issues across geographic regions. IEEE Trans Comput Soc Syst 8(6):1475–1488
65. Blank G (2017) The digital divide among Twitter users and its implications for social research. Soc Sci Comput Rev 35(6):679–697
66. Tufekci Z (2014) Big questions for social media big data: representativeness, validity and other methodological pitfalls. In: International AAAI conference on web and social media, vol 8, pp 505–514
67. Hargittai E (2015) Is bigger always better? Potential biases of big data derived from social network sites. Ann Am Acad Polit Soc Sci 659(1):63–76
68. Wojcik S, Hughes A (2019) Sizing up twitter users. Internet & Tech. Pew Research Center. Available from https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ (Accessed 20-Apr-2023)
69. Soroka V, Rafaeli S (2006) Invisible participants: how cultural capital relates to lurking behavior. In: Proceedings of the 15th international conference on World Wide Web. WWW '06. Assoc. Comput. Mach., New York, pp 163–172. https://doi.org/10.1145/1135777.1135806

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.