ORIGINAL ARTICLE

Who is mentioning COVID-19 articles on twitter? Classifying twitter users in the context of scholarly communication

Yingxin Estella Ye[1] · Jin-Cheon Na[1]

Abstract
This study aims to examine the demographics of participants engaged in scholarly communication on Twitter, which has been rebranded as X. Firstly, based on a dataset of tweets citing COVID-19 publications, it proposed a more precise classification system consisting of eleven user categories for individuals who tweeted academic publication. Secondly, it explores the effectiveness of graph neural network models (GNNs) in combination with a transformer-based text classification model (specifically, BERT) to classify these newly defined user categories. The findings of this research highlight that GNNs can effectively interpret the social networks within scholarly communication, and complement text classification models in characterizing user types. The best-performing model achieved an accuracy rate of 84.05 percent in classifying user categories for a dataset of 10,048 labeled users. Subsequently, this model was employed to analyze 393,030 tweeters in our dataset. The analysis revealed that relevant scholarly discussion on Twitter was dominated by members from the general public (over 71 percent). Academic researchers and institutions constituted 12.48 percent, while health science professionals and institutions made up 7.35 percent of the contributors to relevant scholarly discussions on Twitter. Notably, academic publishers and research feed accounts exhibited aggressive tweeting behaviors and were responsible for the highest volume of tweets on average. This study also demonstrates the active involvement of various non-academic members, including commercial businesses, mass media outlets, public authorities, politicians, and civil society organizations, in Twitter scholarly communication.

Keywords Altmetrics · Twitter · Scholarly communication · User classification · Graph neural networks · Social networks · Machine learning

1 Introduction

Altmetrics are metrics that track and quantify the attention given to a scholarly work or researcher through online platforms (Priem and Hemminger 2010). Twitter (rebranded as X), a major source of altmetrics with the potential to trace up-to-date conversations about academic literature (Hassan et al. 2017; Holmberg and Vainio 2018; Özkent 2022), has received significant attention from researchers.

However, research communities have raised concerns about the validity of Twitter metrics as research impact indicators. Researchers have highlighted significant issues, such as obsessive duplicate tweeting from social media management tools and bots, unselective and mechanical tweeting that expresses no original thoughts and self-retweets (Cao et al. 2023; Hassan et al. 2017; Robinson-Garcia et al. 2017, 2018). Additionally, the lack of a robust understanding of Twitter metrics may have undermined research communities' confidence in utilizing them for research impact evaluations.

To advance our knowledge of Twitter altmetrics, it is critical to gain a comprehensive understanding of Twitter scholarly communication. Demographic analysis of participants discussing scientific publications on Twitter could be the first step.

A main objective of this study is to facilitate Twitter metrics research on a large scale by constructing classification models that can effectively identify user types of participants in the context of Twitter scholarly communication. Specifically, this article makes the following contributions:

✉ Yingxin Estella Ye
   yingxin001@e.ntu.edu.sg

   Jin-Cheon Na
   tjcna@ntu.edu.sg

1   Wee Kim Wee School of Communication and Information, Nanyang Technological University, 31 Nanyang Link, Singapore, Singapore 637718, Singapore

(1) Firstly, we propose a refined user classification scheme tailored to the context of Twitter altmetrics. This scheme covers a variety of participants engaged in scholarly communication on Twitter.

(2) User classification models are developed to facilitate user analysis in the context of Twitter metrics on a large scale. The best-performing model significantly achieved an accuracy of 84.05 percent and up to around 4 percent relative improvement compared to the fine-tuned BERT model, the state-of-art text classification approach.

(3) We demonstrate how graph neural network models (GNNs), in conjunction with a transformer-based text classification model (i.e., BERT), enhance user attributes of tweeters by supplementing social network information.

This article contributes to the current literature on demographics in Twitter scholarly communication by conducting a user analysis using a large dataset of tweeters discussing COVID-19 publications. The following section in this paper reviews related works and introduces the design of the proposed user classification models. The proposed models are evaluated in the next section. The final section presents a straightforward user analysis of a sample of tweeters mentioning COVID-19 publications, based on the predicted results generated by a selected model.

2 Related works

2.1 User analysis of twitter scholarly communication

The extant literature has illustrated the diversity of demographics in Twitter scholarly communication. Altmetric.com, a leading altmetrics service provider, categorizes Twitter users into four groups: members of the public, researchers, health science practitioners, and science communicators (e.g., journal publishers and editors) (Altmetric.com 2021). This classification scheme has been commonly adopted by researchers. For example, Yu (2017) discovered that over 85 percent of 3,903,054 academic publications had been tweeted by the general public. Researchers, clinical science practitioners, and science communicators accounted for 33.12 percent, 16.19 percent, and 16.43 percent of tweeted publications, respectively. Similarly, the study of Díaz-Faes et al. (2019) reported that approximately 86.3 percent of tweeters discussing academic articles between 2011 and 2017 were members of the general public. In a recent study, Abhari et al. (2022) developed a supervised learning model to classify users who tweeted about COVID-19 publications.

Their user classification results indicated that researchers dominated the discussion, contributing 23.29 percent of relevant tweets and likes (including both non-retracted and retracted articles), followed by the general public (18.63 percent), and science communicators (9.98 percent).

Some studies delve into detailed user categorization when examining tweeters who mention academic articles. For instance, Ferguson et al. (2014) identified a diversity of participants involved in scholarly discussion on Twitter, including faculty members, research fellows, communication officers, cardiologists, professional societies, and others, during a scientific conference held by the Cardiac Society. In another study, Didegah et al. (2018) examined 6388 tweets from multiple disciplines and classified Twitter users participating in scholarly communication into 12 categories: (1) individual researchers; (2) individual citizens; (3) individual journalists; (4) individual professionals; (5) research organizations; (6) funding organizations; (7) public authorities; (8) civil society organizations; (9) publishers/journals; (10) media; (11) businesses and (12) others. They found that individual citizens tweeted most frequently about social sciences & humanities and physics & engineering, while individual researchers dominated discussions in math & computers and biomedical and health sciences. Civil society organizations played a significant role in discussions related to life & earth studies.

Vainio and Holmberg (2017) analyzed 100 profiles of tweeters from four research domains and divided them into seventeen user types, including students, researchers, postdoc/Ph.D. researchers, professors, health care professionals, expertise, writers/editors/journalists, other professions, companies, entrepreneurs, publishers, publications, non-profit organizations/non-profit groups, government organizations/universities, opinions/propaganda, librarians/other academics, and others. As their study concluded, the discussions were primarily driven by researchers in fields such as agricultural, engineering, and technology sciences (19 percent) and natural sciences (16 percent). About 25 percent of tweeters discussing medical and health sciences publications were professionals, while company profiles accounted for the majority of tweeters mentioning publications in social sciences and humanities (14 percent).

Existing studies have provided substantial evidence of the diversity of Twitter users engaged in scholarly communication. Nevertheless, significant research gaps remain. Firstly, large-scale studies often employ broad user classification schemes, potentially leading to an overestimation of the general public's presence. Secondly, due to the cost of manually labeling users, studies with precise classification schemes tend to have limited sample sizes. To address these challenges, it is critical to refine user categories within the context of scholarly communication on Twitter and automate the user classification process.

Hence, this study proposes an adjusted user classification scheme that divides tweeters mentioning scientific publications into eleven categories (refer to Table 1). Research feeds are included to supplement the list of user categories, as they were observed to be prevalent in the context of Twitter altmetrics (Haustein et al. 2016). This classification scheme encompasses a wide range of participants in scholarly communication on Twitter. For example, it could be remapped to align with user categories adopted by Altmetric.com: (1) Researchers, including academic and non-academic researchers and institutions; (2) Health Science Practitioners, covering health science professionals and institutions; (3) Science Communicators, such as academic publishers and research feeds; (4) Members of the Public, which includes other user categories in the scheme. With more precise categories, it can offer a more comprehensive overview of tweeters involved in scholarly communication on Twitter. Though this classification scheme may not be fully compatible for Twitter altmetrics data outside of COVID-19 or public health domains, it can be easily adapted to different contexts.

2.2 Exploiting social network patterns in user classification task

Twitter users may provide limited information or leave their profiles empty, relying solely on textual information from user profiles may not be sufficient for user classification. In the study of Abhari et al. (2022), approximately 33.53 percent of relevant tweets and likes came from tweeters without profile descriptions. In another study, nearly 10 percent of human accounts (n = 11,241) were found to have empty description fields, while this percentage was 47.6 percent among 11,768 bot accounts (Hayawi et al. 2022).

The social web, exemplified by platforms such as Twitter, has significantly reshaped scholarly communication, promoting engagement and interactions within academia and with a broad range of audiences, including the general public. Considering that users' activities in scholarly communication on Twitter are influenced by their specific goals and fields of study (Holmberg and Thelwall 2014), they could exhibit distinct patterns of social networking. Hence, to provide a more holistic depiction of Twitter users one potential approach is to supplement their user profiles with social network information. The benefits of incorporating social network patterns as input features into user classification models have been widely established in the existing scholarship.

For instance, Marco and Popescu (2011) used gradient boosted decision trees (GBDTs) to predict tweeters' political leanings based on their social circles of followship and interaction, achieving an 80 percent accuracy. Li et al. (2019) inputted the "following" relationship among Weibo users in their user attribute classification tasks, and found that integrating text and social network features into neural networks improved the prediction of users' age, gender, and geographical location. Campbell et al. (2014) supplied various social graphs of Twitter user interactions, including mentions, retweets, and hashtag usage, into their account verification algorithm. By considering both social network information and Twitter content, their decision tree-based

Table 1 Category of twitter users

Academic researchers and institutions	Academic researchers and scientists (e.g., faculty members, lecturers, research fellows, and postgraduate students) and academic institutions, such as universities, research labs, and researcher networks
Academic publishers	Publishers that distribute academic research and scholarly works, as well as editors working in the academic publishing industry
Business	Commercial organizations in non-health science domain
Civil society organizations	Non-profit, voluntary groups, e.g., associations of patients, societies based on common interest, social media communities, churches, etc
Funding organizations	Public agencies that allocate funding and resources to researchers and research organizations
Health science professionals and institutions	Professionals that work in health science (physicians, pharmacists, and nurses) and related organizations such as hospitals and clinics
Mass media	Mass media organizations such as TV channels, magazines, and books, as well as personnel working in the industry, such as journalists, reporters, and talk show hosts
Non-academic researchers and institutions	Researchers and scientists who are affiliated with non-academic institutes (e.g., research departments in industries or governments) and non-academic research institutes
Public authorities and politicians	Public administration agencies (e.g., Ministry of Health, National Department of Statistics), international organizations (e.g., United Nation, World Health Organization), political parties, and affiliated individuals (e.g., ministers, advisors, and politicians)
Research feeds	Feeds or alerts for scientific publications (e.g., automated or human-curated updates from academic databases)
Others	Users from the general public who do not fall into any of the other categories

model achieved an average area under the ROC curve (AUC) of 0.76, surpassing the baseline model that used only Twitter content (AUC of 0.67). Jiang et al. (2022) developed network-based models by constructing retweet networks, combining Bidirectional Encoder Representations from Transformers (BERT) and GraphSAGE (a type of Graph Neural Networks or GNNs) to identify political leaning with over 90 percent accuracy.

It is also common among researchers to employ interaction network features to detect social bots on Twitter, including metrics like the number of nodes and edges, the size of the largest connected component, average degree, and the count of isolated nodes (Aljabri et al. 2023; Dehghan et al. 2023). In the realm of Twitter scholarly communication, Aljohani et al. (2020) utilized a graph convolutional neural network (GCN) model for their bot prediction task. Their model effectively captured social interaction patterns from the undirected user network of retweets and @mentions, achieving an impressive accuracy of 71 percent in distinguishing between humans and bots among a dataset of 16,264 users, including 64 labeled as bots.

Therefore, GNNs, a class of methods that represent social interactions and connections as graphs, could play a useful role in developing an effective approach for classifying diverse participant types within the context of Twitter metrics.

2.3 Related classification models

GNNs, which model social interactions and connections as graphs, can potentially capture tweeters' social network information and enhance the input features for user classification models. This study proposes to classify user types in Twitter scholarly communication by combining the transformer-based text classification model (BERT) with GNNs.

2.3.1 The BERT model

The Bidirectional Encoder Representations from Transformers (BERT) model stands as one of the most successful deep learning methods for text classification. BERT employs the Masked Language Model (MLM) to learn language representations bidirectionally, both from left to right and right to left. It consists of a multi-layer, multi-head self-attention mechanism that harnesses transformers, allowing it to capture word relationships effectively and generate attention maps (Devlin et al. 2018; Lu et al. 2020). Vaswani et al. (2017) illustrated that the transformer-based encoder-decoder structure comprises three key components: (1) a multi-head attention mechanism, (2) layer standardization and residual connections, and (3) position-wise feed-forward networks.

Researchers have applied BERT models to various text classification tasks on Twitter. These include classifying tweets into specific domains or topics, identifying tweets containing specific information, and categorizing user attributes based on their profiles (Basile et al. 2019; Dukic et al. 2020; Müller et al. 2020). Hence, our Twitter user classification task can also leverage BERT's capability to extract distinctive features from users' profiles.

2.3.2 Graph neural networks

Graph structures, which model data through nodes and edges, are widely used to capture relationships within various data types, such as genome sequences, social networks, and image representations. Graph Neural Networks (GNNs) are deep learning methods that extend traditional neural networks into the realm of graphs. Unlike conventional machine learning approaches, which simplify graph structures into numerical vectors, GNNs retain the topological relationships between nodes when encoding graph structures. GNNs can be tailored for node-level, edge-level, and graph-level prediction tasks (Scarselli et al. 2009; Wu et al. 2021; Zhou et al. 2020).

Zhou et al. (2020) identified three crucial components of GNN models. The first component is the propagation module, which allows GNNs to capture underlying features and topological information within a given graph. It comprises three primary functions: (1) Message functions that transform node features for transmission to adjacent nodes; (2) Aggregation functions, such as sum, mean, and max, which compile information received from neighboring nodes; and (3) Update functions that update a node's state by integrating the aggregated message with its previous state. GNNs define a target node based on its neighborhood. For example, as illustrated in Fig. 1, target node i can be represented by its neighbors j, k, l, as well as their respective neighbors. GNNs update the representation of node i through message passing, transmitting messages (i.e., embeddings) from i's neighbors and its neighbors to i along edges using specific aggregation functions.

The second component involves a sampling module, which can be helpful when working with large-scale graphs. These modules effectively mitigate the growth in neighbor nodes, which occurs as a result of multiple stacks in GNN layers. Various types of sampling, including node sampling, subgraph sampling, and layer sampling, can be applied at different levels.

The third component is the pooling module. A pooling layer is usually added after the convolutional layer in graph convolutional networks to summarize node representations and generate an abstract representation of high-level graphs or subgraphs.

Fig. 1 Neighborhood aggregation in GNNs
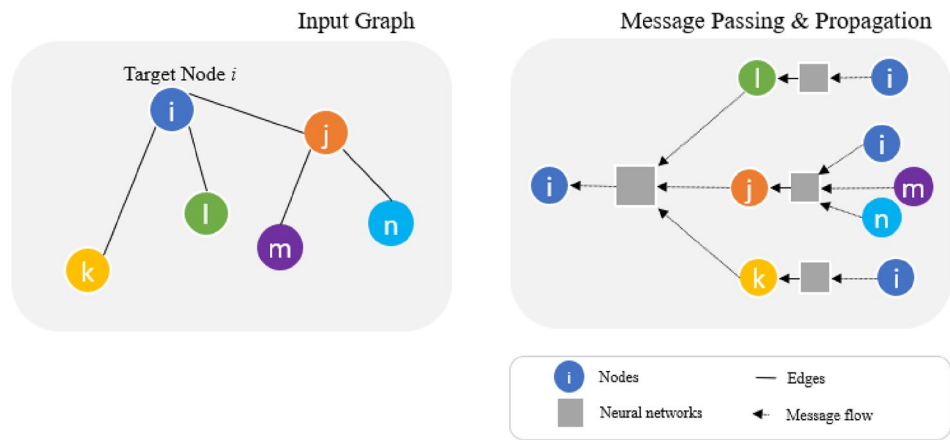
Therefore, GNNs could be a feasible method for improving user classification models in the context of Twitter scholarly communication, as they have the capability to represent not only the features of target tweeters but also their social networks by aggregating features of neighboring tweeters with whom they have interacted. This approach offers flexibility in representing target tweeters at different levels. For instance, a tweeter can be chracterized by updated features that combine their own profile with information from their neighbors. We could also employ pooling modules to profile target tweeters by summarized features of all neighbors within their Twitter interaction networks.

3 Methods

3.1 Data collection

First, we obtained a set of academic publications related to COVID-19 using the Scopus Search API, using the query string constructed by Kousha and Thelwall (2020). We refined the search results to include only English-written journal articles published in Q4 2021 (n = 39,487). We retrieved articles mentioned on Twitter through Altmetric. com by querying journal identifiers (including ISSNs and e-ISSNs) of publications collected from Scopus as of September 17, 2022. This resulted in two data files: (1) a list of 14,845 publications (37.58%) cited on Twitter and (2) a collection of 1,172,349 tweets citing these publications. Subsequently, using the IDs of tweets that mentioned selected publications captured by Altmetric.com, we retrieved 871,611 tweets from the Twitter API as of September 19, 2022. We further collected profiles of 393,030 tweeters, the creators of these tweets, from the Twitter API using their Twitter user IDs. It's worth noting that tweets may become inaccessible in various situations, such as when a tweet is removed by its creator or when the creator's account is suspended or set to private. We expanded the tweet dataset by tracing tweets that directly reacted to selected tweets and their sourced tweets (n = 968,820). Additionally, we included users who interacted with the selected tweets, including through retweets, replies, or mentions, in our dataset (Fig. 2).
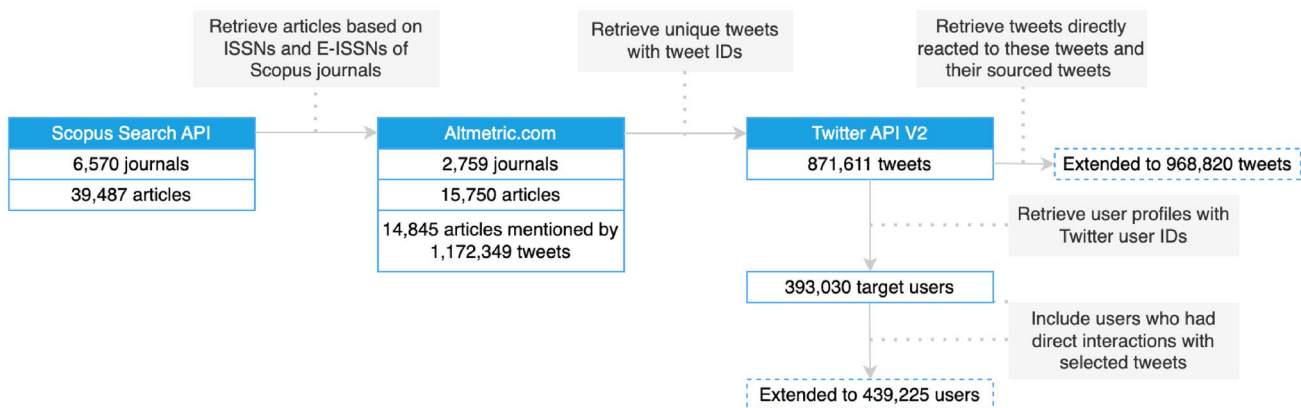

Fig. 2 Data collection process

3.2 Data sampling

In this study, we adopted a random sampling approach, aiming for a sample size of 10,000. As our primary focus was on users with profiles written in English, we initially selected 10,200 tweeters to ensure that we would have a sufficient number of samples even after excluding non-English profiles. To detect the language of each user profile, we employed the Python package Google Translate API. Subsequently, we manually removed non-English profiles from the pool of 10,200 selected tweeters, resulting in a final dataset comprising 10,048 sampled tweeters.

Model building and evaluation were carried out by splitting the labeled users into a training set (70 percent), a validation set (15 percent), and a test set (15 percent). Stratified random sampling was applied to ensure that the three sets of data had a consistent distribution of user types.

3.3 Data labeling

Three coders, including one of the authors, were recruited for the data labeling exercises. The other two coders consisted of an engineering graduate with prior experience using Twitter and a postgraduate student from the Information Studies program. They labeled 4894, 3814, and 3746 tweeters, respectively. Table 2 presents Cohen's Kappa coefficients, which were calculated based on labels for overlapping tweeter profiles assessed by our coders. The Cohen's Kappa statistics (> 0.8) indicate acceptable intercoder reliability.

To tag a tweeter, our coders examined the Twitter user page to identify clues about their occupations and affiliations, as exemplified in Table 1. They observed the tweeter's Twitter activities, including tweets, replies, likes, and posted media. When applicable, external profiles, such as scholarly or faculty profiles or LinkedIn profiles, were also accessed. The coders considered the following information for judgment: (1) Twitter user name and screen name, (2) description in the user profile and the URL of the personal page, (3) age of the account (as of 20 September, 2022); (4) numbers of followers and friends, (5) statuses count, (6) whether the Twitter account is verified, (7) location of the user, (8) the number of tweets that the user had created in our dataset, (9) the number of unique articles that the user had tweeted, and (10) the number of tweets per article.

A tweeter's profile can only be classified into a category that best describes its user type. Users with both academic and non-academic status will be labeled as academic users. Consistent with the study by Hayawi et al. (2022), approximately 10.39 percent of sampled users have empty descriptions in their profiles. Figure 3 presents the demographics of labeled users, with roughly half of them categorized as others, likely representing members of the general public. Academic researchers and institutions account for 20.55 percent, and health science practitioners make up 10.12 percent. Academic publishers constitute 2.7 percent, while a notable presence of civil society organizations (4.44 percent) and non-academic researchers & institutions (3.79 percent) was also observed.

4 Preliminaries

This study characterizes Twitter users through two layers: (1) Textual Content: This primarily includes the name, description, and URLs found in their Twitter profiles. (2) Ego Graph of Twitter Interactions: We created social network graphs based on Twitter interactions, encompassing retweets, replies, and @mentions, derived from the dataset of tweets (n = 968,820). An ego graph was utlized to illustrate a tweeter's social network patterns. It is a subgraph that centers on a node of the target tweeter, and the connections that it has with other neighbors. This section describes the process of transforming textual content into text embeddings and how ego graphs of tweeters were constructed (Fig. 4).

Table 2 Inter-coder reliability

	Coder 1	Coder 2	Coder 3
Coder 1		0.84 (n = 718)	0.81 (n = 713)
Coder 2	0.84 (n = 718)		0.83(n = 718)
Coder 3	0.81(n = 713)	0.83 (n = 718)	

Fig. 3 Demographics breakdown of labelled users

Category	Value
Academic publishers	271 (2.70%)
Academic researchers & institutions	2,065 (20.55%)
Business	132 (1.31%)
Civil society organizations	446 (4.44%)
Funding organizations	52 (0.52%)
Health science professionals & institutions	1,017 (10.12%)
Mass media	181 (1.80%)
Non-academic researchers & institutions	381 (3.79%)
Public authorities & politicians	100 (1.00%)
Research feeds	92 (0.92%)
Others	5,311 (52.86%)

A Tweeter

Content of User Profile

Name: Anna H.
Description: Associate Professor studying global health
URLs: https://www.xxxx.edu/faculty/anna_h

Ego Graph of Social Interactions
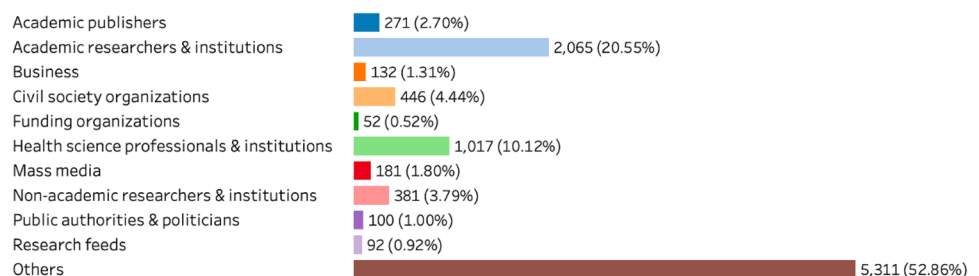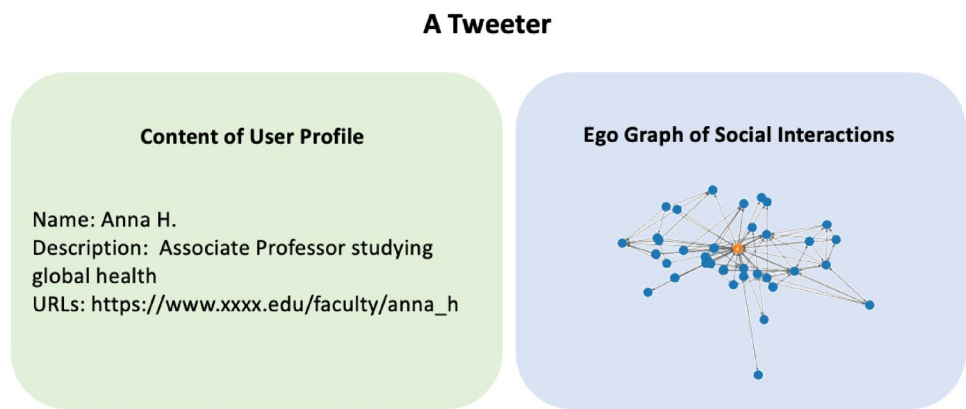
4.1 Text embeddings

This study adopts both bag-of-words (BoW) and BERT embeddings to represent the textual content in tweeters' profiles. For non-English profiles, we employed the Python package google-cloud-translate 3.7.0 to translate them into English. Emojis within the text were decoded using the emoji 1.2.0 package.

4.1.1 BoW embeddings of nodes in GNNs

Text was preprocessed for the generation of embeddings through the following steps: (1) removal of English stop words using NLTK 3.6.2 packages, (2) removal of 'http://' and 'https://' from URLs, 3) tokenization using non-alphanumeric character separators. Based on the training data, a corpus dictionary (vocabulary size: 3884) was constructed, including word tokens that occurred three times or more.

Two approaches were used to represent the occurance of word: (1) Binary: The user profile of a user i was represented as a document vector $\boldsymbol{a}_i \in \{0 or 1\}^{3884}$, where a value of 1 was assigned if the word occurred in the profile content, while a value of 0 indicated its absence. (2) TF-IDF: The Term Frequency-Inverse Document Frequency (TF-IDF) metric was calculated for each word appearing in the user's profile content. Accordingly, the user profile of a user i was represented as a document vector $\boldsymbol{a}_i \in \{t\}^{3884}$, t refers to the normalized TF-IDF score of the word, ranging between 0 and 1. The Inverse Document Frequency (IDF) was calculated based on the training data set.

4.1.2 Bert embeddings in BERT

This study adopts the tokenizer from the "BERT-Base, Uncased: 12-layer, 768-hidden, 12-heads, 110 M parameters" and "BERT-Large, Uncased, 24-layer, 1024-hidden, 16-heads, 340 M parameters" pre-trained models. A user profile is represented by a list of token IDs and associated attention masks. For a user i, the user profile $\boldsymbol{tok}_i$ can be denoted as $\left(tok_{i_{[CLS]}}, \ldots, tok_{i_{28}}, tok_{i_{[SEP]}}\right)$. $tok_{[CLS]}$ is a token positioned at the start of the text input and is meant for sentence-level classification, while $tok_{[SEP]}$ is a separator token that marks the end of the sentence. Given that 99 percent of labeled users contained fewer than 31 tokens, the maximum sequence length was set to 30 tokens.

4.2 Graph construction

An ego graph is constructed for each tweeter with interactions extracted from tweets they contributed to and related tweets, including tweets they reacted to and tweets that replied to them. We denote the ego graph G_i of the tweeter i, the ego node, as $G_i = (V_i, E_i)$, where V_i represents the tweeter and their neighbors, including direct neighbors. A directed link $(V_i, V_j) \in E_i$, is drawn if user i has ever retweeted, replied to, or @mentioned user j (see Fig. 5). The radius of the ego graph is set to 2. Social networks of isolated users, accounted for around 8 percent among sampled users, were represented by ego graphs containing only self-loops.

G_i can be represented by the adjacency matrix $A_i \in \mathbb{R}^{|V_i| \times |V_i|}$, where $|V_i|$ is the number of nodes in the ego graph. Self-connections were added to the graph, and $A_{i_{ii}}$ was always set to 1. We let X_i be the nodes' features of the ego graph. In the feature matrix $X_i = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_{|V_i|})$, each user ($e.g., \boldsymbol{x}_1$) is represented as a one-hot vector, the BoW representation of the user profile.
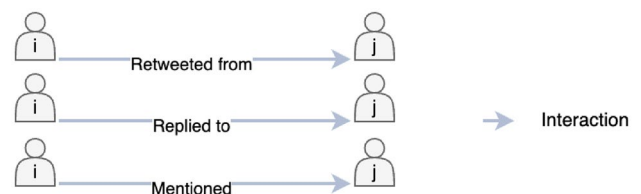

Fig. 5 Directions of user edges

5 Methods

First, we examined models that use only textual information from tweeters' profiles as baseline models. Next, we tested GNNs that utilize social network information captured from user interactions. In the last section, we explored combined models that classify user types based on both text and social network information.

The model was developed using the framework of PyTorch 1.13.1 and PyTorch Geometric 2.0.4 on GPU. The BERT model was imported from Transformers 4.28.0. For BERT models, the pretrained model was adopted in accordance with the tokenizer respectively. Table 3 describes the notations used in this article.

5.1 Text-based models

We employed a linear model, referred to hereafter as the BoW model, and a fine-tuned BERT model to assess the capability of text content in predicting user backgrounds in the context of Twitter scholarly communication.

5.1.1 BoW

The BoW model is a two-layer neural network with one hidden layer containing 32 hidden units. This model captures correlations between input and output features through linear transformation functions (see Eq. 1). The BoW representation of a user i is represented by the input $\boldsymbol{x}_i \in \mathbb{R}^{3884}$. Meanwhile, in the hidden layer, $\boldsymbol{W}^1 \in \mathbb{R}^{32 \times 3884}$ is the weight matrix and $\boldsymbol{b}^1 \in \mathbb{R}^{32}$ is the bias vector. The output $\boldsymbol{h}_i^1 \in \mathbb{R}^{32}$ from the *ReLU* activation and *dropout* function moves to the output layer. $\boldsymbol{W}^2 \in \mathbb{R}^{11 \times 32}$ is the weight matrix, and $\boldsymbol{b}^2 \in \mathbb{R}^{11}$ refers to

the bias vector. The dropout probability p is set to 0.5. In later sections, we use *Linear* as the linear transformation function.

$$\boldsymbol{h}_i^1 = dropout\big(ReLU\big(\boldsymbol{W}^1 \boldsymbol{x}_i + \boldsymbol{b}^1\big)\big) \tag{1a}$$

$$\boldsymbol{o}_i = W^2 h_i^1 + \boldsymbol{b}^2 \tag{1b}$$

5.1.2 Fine-tuned BERT

The model first converts the BERT representation of a user profile $\big(tok_{i_{[CLS]}}, \ldots, tok_{i_{28}}, tok_{i_{[SEP]}}\big)$ to a matrix of input embeddings $\boldsymbol{EM}_i = (\boldsymbol{em}_{i_{[CLS]}}, \ldots, \boldsymbol{em}_{i_{28}}, \boldsymbol{em}_{i_{[SEP]}})$, $\boldsymbol{EM}_i \in \mathbb{R}^{30 \times 768}$. Each input embedding $\boldsymbol{em}_{i_n} \in \mathbb{R}^{768}$ combines position embeddings that interpret the position of the words within the sentence and the token embeddings that explain the token vocabulary. Next, the model learns the contextual word representations of input words and predicts user types based on the pooled output of the learned classifier token $(\boldsymbol{t}_{i_{[CLS]}} \in \mathbb{R}^{768})$. This process is followed by a dropout function with a p of 0.5. A linear transformation layer is then stacked on top of the BERT model to generate an output with a size of eleven.

$$\boldsymbol{o}_{pooled_i} = dropout\big(\boldsymbol{t}_{i[CLS]}\big) \tag{2a}$$

$$\boldsymbol{o}_i = Linear\big(\boldsymbol{o}_{pooled_i}\big) \tag{2b}$$

Table 3 Notations used in this article

Notation	Description
$\boldsymbol{x}_i$	Input vector of node i
$\boldsymbol{o}_i$	Output vector of node i
$\hat{\boldsymbol{y}}_i$	Predicted output vector of node i
$\boldsymbol{h}_i^l$	Hidden state of node i in the lth hidden layer
$\mathcal{N}_i$	Neighborhood set of node i, inclusive of the self-loop
G_i	Ego graph of a node i, $G_i = \big(V_i, E_i\big)$, where V_i is the set of nodes and E_i is the set of edges among nodes within the ego graph
$|V_i|$	Number of nodes in node i's ego graph G_i
$|E_i|$	Number of edges in node i's ego graph G_i
dropout	A dropout function with a default dropout probability of 0.5
GAP	A global average pooling layer
ReLU	Rectangular linear unit (ReLU) activation function
BatchNorm	Layer-wise batch normalization function
OR	OR operator
$||$	Vector concatenation
flow	Flow direction of the message passing (ST: source to target, TS: target to source)

5.2 GNNs based on ego graphs of social interactions

We examined three popular convolutional GNNs to our user classification task: (1) GAT (Brody et al. 2021); (2) GraphSAGE (Hamilton et al. 2017), and (3) GIN models (Xu et al. 2018).

5.2.1 Propagation modules of selected models

5.2.1.1 GAT

GATv2Conv Is the implementation of the GAT approach of Brody et al. (2021) in the PyTorch Geometric package. GATs attend each node to all other neighbor nodes with attention coefficients. A vector of attention coefficients $\alpha_{i,j}$ indicates the importance of the neighbor node j's features to node i (Veličković et al. 2017). W is the weight matrix associated with the linear transformation. GATs adopt a multi-layer attention strategy to update the representation of node i by applying k number of independent attention head matrices. Through this process, GATs generate multiple hidden states, followed by a concatenation of the resulted features (see Eq. 3b). *LeakyRELU* was applied as the nonlinearity function σ.

$$\alpha_{i,j} = \frac{\exp\left(a^\top \sigma\left(W\left[x_i \| x_j\right]\right)\right)}{\sum_{u \in \mathcal{N}_i} exp\left(a^\top \sigma\left(W\left[x_i \| x_u\right]\right)\right)} \tag{3a}$$

$$h_i^{l+1} = \|_{k=1}^K \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W^k h_j^l\right) \tag{3b}$$

5.2.1.2 GraphSAGE

SAGEConv In PyTorch Geometric packages was utilized to implement the GraphSAGE model. As the ego graphs in our study are generally small, we decided to skip the node sampling process described by Hamilton et al. (2017) when carrying out user classification tasks. Equation 4 shows the propagation step used in GraphSAGE. W denotes a learnable weight matrix, and the features of i's neighbors are aggregated using the aggregation function, *AGG*, with a mean aggregator.

$$h_i^{l+1} = W_i^{l+1} h_i^l + W_{\mathcal{N}_i}^{l+1} \cdot AGG\left(\left\{h_j^l, \forall j \in \mathcal{N}_i\right\}\right) \tag{4}$$

5.2.1.3 GIN

This study adopts *GINConv* to implement GIN models. The propagation process is described in Eq. 5, with ϵ serving as a learnable floating point value. Multi-layer perceptron (MLP) is used in the subsequent layer of the first iteration to aggregate features of i's neighbors $j \in \mathcal{N}_i$.

$$h_i^{l+1} = MLP^{l+1}\left(\left(1 + \in^{l+1}\right)h_i^l + \sum_{j \in \mathcal{N}_i} h_j^l\right) \tag{5}$$

5.2.2 Inputs of models

Our GNN models take ego graphs of target tweeters as inputs. The input of an ego graph with a target node i has two parts: (1) a feature matrix X_i containing BoW representations extracted from the user profiles of i and its neighbors within the ego graph G_i, $X_i \in \mathbb{R}^{|V_i| X 3884}$; (2) an edge index E_i which reflects the directed links between nodes within the ego graph G_i, $E_i \in \mathbb{R}^{2 X |E_i|}$.

5.2.3 Model building

Models were constructed with two GNN layers, which facilitate the transmission of messages among nodes located two steps away. For each model, we considered the effects of flow direction of message passing, *flow* $\in$ {εsource-to-target", εtarget-to-source"} on the performance of models. In addition to source-to-target (hereafter referred to as ST) and target-to-source (hereafter referred to as TS) approaches, we tested the bidirectional approach (hereafter referred to as BI), a combination of both ST and TS propagations.

With the GAT layer function $GAT(\cdot)$, we denoted GAT models in this study in Eq. 6. For any given target node i, feeding the input graph into the first GAT layer with eight attention heads (see Eq. 4) provides a hidden state $H_i^1 \in \mathbb{R}^{|V_i| \times 512}$ as an output. The second GAT layer, which has a single attention head, outputs updated user features, $X_i' \in \mathbb{R}^{|V_i| \times 11}$.

$$H_i^1 = dropout\left(ReLU\left(GAT\left(X_i, E_i, flow, K_1\right)\right)\right), \\ where \quad K_1 = 8 \tag{6a}$$

$$X_i' = GAT\left(H_i^1, E_i, flow, K_2\right), where \; K_2 = 1 \tag{6b}$$

With the GraphSAGE layer function $GS(\cdot)$, Eq. 7 depicts how node features in the input graph of a node i are updated. The first GraphSAGE layer leads to a hidden state $H_i^1 \in \mathbb{R}^{|V_i| \times 64}$. After the hidden state is revealed, another GraphSAGE layer updates the features of nodes to $X_i' \in \mathbb{R}^{|V_i| \times 11}$.

$$H_i^1 = dropout\left(ReLU\left(GS\left(X_i, E_i, flow\right)\right)\right) \tag{7a}$$

$$X_i' = GS\left(H_i^1, E_i, flow\right) \tag{7b}$$

With the GIN layer function $GIN(\cdot)$, we computed the outputs of the GIN layers as expressed in Eq. 8. In our GIN models, the MLP is a block of sequential layers comprising two linear layers. A *ReLU* function is applied to the hidden state. With the hidden state $H_i^1 \in \mathbb{R}^{|V_i| \times 64}$ generated from the 1st GIN layer, the features of nodes are updated as $X_i' \in \mathbb{R}^{|V_i| \times 11}$.

$$H_i^1 = dropout\left(ReLU\left(GIN\left(X_i, E_i, flow\right)\right)\right) \quad (8a)$$

$$X_i' = GIN\left(H_i^1, E_i, flow\right) \quad (8b)$$

Models adopting a bidirectional message passing approach fed the input graph GNN layers as described in Eqs. 6–8. Each layer was expected to generate a feature matrix $X_{STi}' \in \mathbb{R}^{|V_i| \times 11}$ (from the ST approach) or $X_{TSi}' \in \mathbb{R}^{|V_i| \times 11}$ (from the TS approach). The concatenation of X_{STi}' and X_{TSi}' will be inputted into a linear layer to generate the updated features of nodes $X_i' \in \mathbb{R}^{|V_i| \times 11}$.

$$X_i' = Linear\left(\left[X_{STi}' || X_{TSi}'\right]\right) \quad (9)$$

Two approaches were considered for the target node representation in the output layer.: 1) Target node approach (TN): The target node i, namely the ego node of the extracted graph, is represented by its updated node features x_i' withdrawn from X_i' (*i.e.*, $o_i = x_i'$), 2) Ego graph approach (EG): The updated features of nodes X_i' are connected to a global average pooling (GAP) layer to generate a representation of the ego graph G_i (see Eq. 10).

$$o_i = GAP\left(X_i'\right) \quad (10)$$

5.3 Combined models: BERT + GNN

The proposed combined models consisted of two major modules. Firstly, a fine-tuned BERT model processed the text information extracted from the target user's Twitter profile. Simultaneously, we applied a GNN model, which was the best-performing GNN model from the above-mentioned experiments, to capture the characteristics of the target user using social network information.

In the fine-tuned BERT model, we computed the output vector o_{iBERT} of a node i by applying Eqs. 2a and 2b. We set the output size of the linear transformation layer (Eq. 2b) to 32 to obtain $o_{iBERT} \in \mathbb{R}^{32}$. For the GNN model, the output vector o_{iGNN} of a node i was generated using Eqs. 6–10, depending on the GNN algorithm selected.

We set the output of the second GNN layer to 32. This allowed us to obtain an output vector $o_{iGNN} \in \mathbb{R}^{32}$ for GNN models with a non-bidirectional approach, or $o_{iGNN} \in \mathbb{R}^{64}$ for models that followed a bidirectional approach.

Next, we concatenated both the outputs and a *BatchNorm* function was applied to them (Eq. 11a). We then inputted o_{cat} to a linear transformation function to generate the output vector o_i. Finally softmax is applied to o_i to get final prediction outputs (Fig. 6).

$$o_{cat} = BatchNorm\left(\left[o_{iBERT} || o_{iGNN}\right]\right), where\, o_{cat} \in \mathbb{R}^{64}\, OR\, o_{cat} \in \mathbb{R}^{96} \quad (11a)$$

$$o_i = Linear\left(o_{cat}\right), where\, o_i \in \mathbb{R}^{11} \quad (11b)$$

5.4 Loss function

For all models, *softmax* outputs a vector of values $\hat{y}_i \in \mathbb{R}^{11}$ (which have a sum of 1) that can be interpreted as the probability of membership for each user category.

$$\hat{y}_i = softmax(o_i) \quad (12)$$

The focal loss function was adopted to address the class imbalance in our dataset. Lin et al. (2017) denoted a focal loss function as shown in Eq. 13. This function focuses on hard misclassified targets by multiplying a modulating factor $\left(1 - p_t\right)^\gamma$ by a focusing parameter ($\gamma \geq 0$) to the standard cross-entropy criterion $log(p_t)$ where $p_t \in [0, 1]$ is the model's estimated probability for the class membership. α_t is a learnable weight vector. We set γ to 2 in this study because this was the optimal value in the researchers' previous experiments. The loss is computed based on the output vectors of target nodes.

$$FL(p_t) = -\alpha_t\left(1 - p_t\right)^\gamma log\left(p_t\right) \quad (13)$$

5.5 Model training

To initiate comparisons at an early stage, we selected the same learning rate for all models, with the exception of the fine-tuned BERT. In our experiments, we tested learning rates of 1e−3, 5e−3, and 5e-4. The learning rate was defaulted to 5e−3, as it allowed selected models to achieve relatively good performance. For the BERT models, we set the learning rate to 2e−5 because it yielded better performance compared to the rates of 5e−5 and 1e−5. The weight decay was set to 5e−4. The dropout probability defaulted to 0.5. The Adam optimizer was adopted in all experiments.

We conducted mini-batch training in all experiments, using a batch size of 32. For each training instance, the maximum number of epochs was set to 200. However, training was set to terminate if the validation loss increases
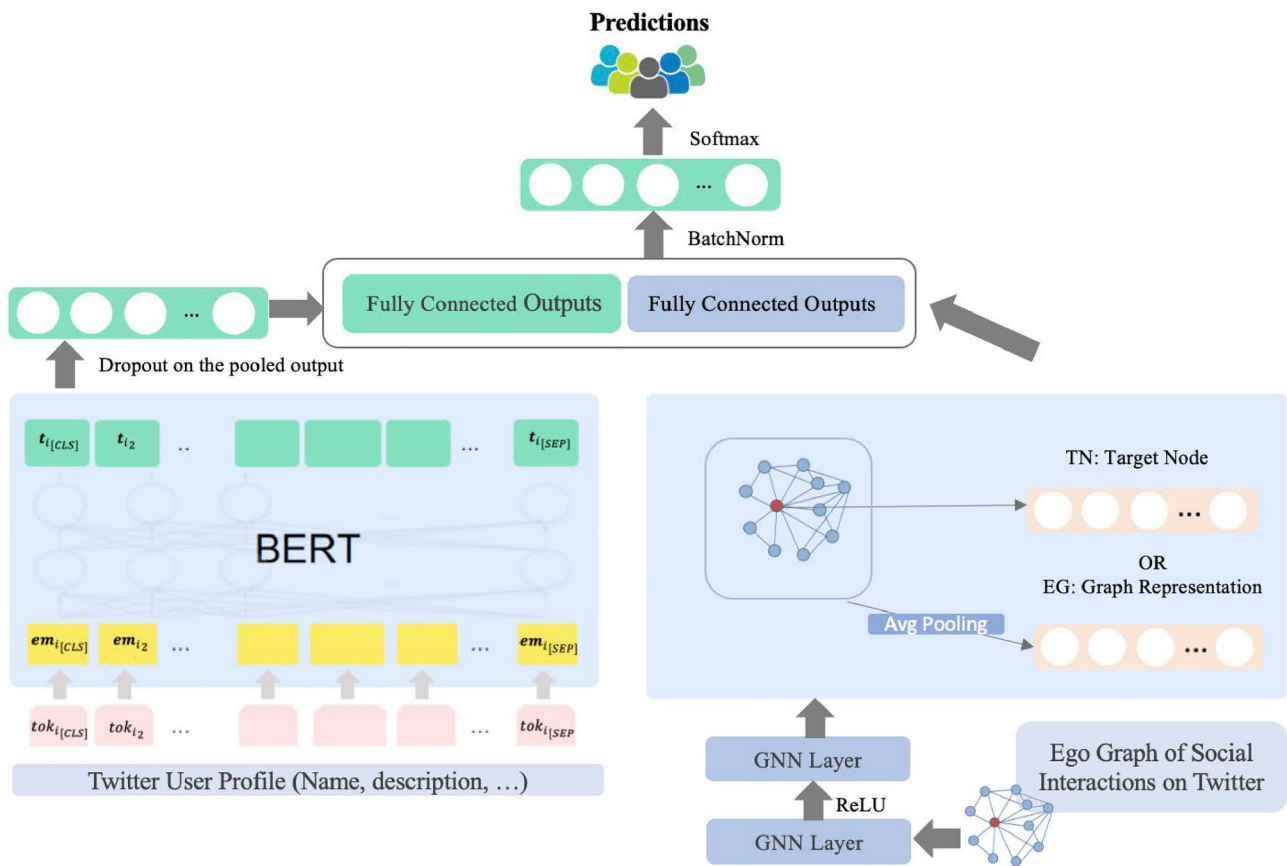
Fig. 6 The combined model

five consecutive times. We iterated training and testing processes 10 times to increase the reliability of the performance evaluation.

5.6 Model evaluation

We used accuracy and F1-score (hereafter referred to as F1) as performance indicators. Among the samples, a size of n_{total}, $n_{correct}$ refers to the number of samples in the test data that are correctly classified, whereas accuracy reflects the ratio between $n_{correct}$ and n_{total}. We provided a more comprehensive evaluation by computing F1 based on recall and precision. The number of classes or categories n was 11 in this study.

For any given category i, TP_i referred to the number of correctly identified samples, and FP_i referred to the number of incorrectly classified samples. Meanwhile, FN_i denoted the number of samples that should have belonged to one category but were wrongly tagged as belonging to a different category. The average recall and precision scores were used to calculate F1 as depicted in Eq. 14.

$$Accuracy = \frac{n_{correct}}{n_{total}} \tag{14a}$$

$$Recall = \frac{\sum_{i=1}^{n} \left(\frac{TP_i}{TP_i+FN_i} \right)}{n} \tag{14b}$$

$$precision = \frac{\sum_{i=1}^{n} \left(\frac{TP_i}{TP_i+FP_i} \right)}{n} \tag{14c}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{14d}$$

6 Results

Table 4 presents the performance of the text-based models. The fine-tuned BERT (BERT-Base-Uncased) model exhibited a significant advantage over the BoW-Binary model, with an average accuracy of 80.49 percent versus 79.61 percent, $t = 2.58$, $p < 0.01$. We decided not to continue examining the BoW (TF-IDF) model due to its comparatively lower performance. While BERT (BERT-Large-Uncased) showed

Table 4 Results of experiments—text-based models

Model	Accuracy (%)	F_1
BoW (binary)	79.61	0.77
BoW (TF-IDF)	69.56	0.63
BERT (BERT-base-uncased)	80.49	0.81
BERT (BERT-large-uncased)	80.33	0.81

similar performance to BERT (BERT-Base-Uncased), we opted for the latter as it is a more lightweight model.

As shown in Table 5, GNN models based on social network information did not perform as well as text-based models solely. In other words, stand-alone GNN models, or social network information, may not be sufficient enough for user classification tasks in the context of Twitter metrics. It is worth noting that models adopting the TN (Target Node) approach achieved higher accuracies ($M = 0.75$, $SD = 0.03$) when compared to those with the EG (Ego Graph) approach ($M = 0.73$, $SD = 0.03$) on average, $t = 4.28$, $p < 0.01$. A possible reason is that graph representations of target users, involving features of various users, might generate noises for the classification tasks.

Additionally, the direction of message passing had a significant impact on predicting user types of tweeters. In general, models implemented with message passing from the target to the source ($M = 0.76$, $SD = 0.02$) outperformed those using the source-to-target approach ($M = 0.71$, $SD = -0.04$), $t = 9.64$, $p < 0.01$. They also tended to achieve better accuracies than models employing a bi-directional approach ($M = 0.74$, $SD = 0.03$), $t = 4.74$, $p < 0.01$. Therefore, it can be assumed that tweeters of sourced tweets with whom a target user attempted to interact could better define their identities.

GAT-TS-TN is the best-performing model, with an accuracy of 78.77 percent. Among the GraphSAGE and GIN models, GraphSAGE-TS-TN and GIN-TS-TN achieved the highest accuracies of 77.74 percent and 74.63 percent, respectively. To finalize the integrated model of BERT and

GNN, we combined the fine-tuned BERT model with GAT-TS-TN, GraphSAGE-TS-TN, and GIN-TS-TN to evaluate their respective performances.

The final model combines GAT-TS-TN and the fine-tuned BERT model. It achieved an accuracy of 82.70 percent with an F1 score of 0.82 on average. This has 2.21 percent relative improvement when compared to the fine-tuned BERT model. The BERT-GIN-TS-TN model reached an accuracy of 80.96 percent with an average F1 score of 0.80, while BERT-GraphSAGE-TS-TN achieved an accuracy of 80.73 percent and an average F1 score of 0.80.

BERT-GAT-TS-TN was further optimized by testing the hyperparameters of the sub-models, as listed in Table 6. The selected hyperparameters are highlighted in bold. The best-performing model achieved an accuracy of 84.05 percent, with an F1 score of 0.83.

Table 7 presents the performance of the best models of BERT, GAT-TS-TN, and BERT-GAT-TS-TN approaches in accurately predicting each user type. Overall, the combined model, BERT-GAT-TS-TN, achived higher accuracies in predicting the majority of user types (as highlighted in bold in Table 7).

This model is good at identifying academic researchers and institutions, academic publishers, research feeds, and health science professionals and institutions. However, due to the limited sample size of user types such as public authorities & politicians and funding organizations, the

Table 6 Hyperparameters tested

Hyperparameters	Sub models	
	Fine-tuned BERT	GAT-TS-TN
Learning rate	**2e−5**	0.01, 1e−3, **5e−3**, 3e−3, 5e−4
Dropout probability	0.3, 0.4, **0.5**	0.3, 0.4, **0.5**
The size of the last hidden layer	**16**, 32, 64	16, **32**, 64
Weight decay	**5e−4**	**5e−4**

Table 5 Results of experiments—social network-based models

	GAT		GraphSAGE		GIN	
	Accuracy (%)	F_1	Accuracy (%)	F_1	Accuracy (%)	F_1
ST-EG	70.99	0.67	71.65	0.67	67.03	0.69
ST-TN	73.41	0.70	77.11	0.74	67.05	0.68
TS-EG	77.52	0.70	76.95	0.74	72.91	0.71
TS-TN	**78.77**	0.76	**77.74**	0.75	**74.63**	0.73
BI-EG	76.74	0.74	70.38	0.68	71.45	0.69
BI-TN	78.17	0.75	75.12	0.73	73.51	0.71

ST source-to-target message passing approach, *TS* target-to-source message passing approach, *BI* bidirectional message passing approach, *EG* ego-graph node representation approach, *TN* Target node node representation approach. For each model, the best-performing approach is highlighted in bold

Table 7 Accuracy (%) in predicting each category—BERT versus GAT-TS-TN versus BERT-GAT-TS-TN

User type	BERT	GAT-TS-TN	BERT-GAT-TS-TN
Academic researchers and institutions	82.93	80.48	**90.24**
Academic publishers	80.00	81.29	**84.84**
Business	20.00	10.00	**40.00**
Civil society organizations	65.67	40.30	**70.14**
Funding organizations	00.00	**37.50**	22.50
Health science professionals and institutions	72.54	48.37	**73.85**
Mass media	35.71	28.57	**42.86**
Non-academic researchers and institutions	46.56	17.24	**56.90**
Public authorities and politicians	13.33	6.67	**20.00**
Research feeds	78.57	78.57	**85.71**
Others	94.23	**97.99**	92.47

accuracies of some categories' classification do not appear to be ideal.

7 Use cases

7.1 Demographic breakdown of sample users in our dataset

We used the best-performing model to predict unlabeled users in our dataset. Figure 7 displays the demographic breakdown of the 393,030 sample users who engaged in Twitter discussions about COVID-19 articles. According to the predicted results, scholarly communication on Twitter is primarily led by academic researchers and institutions (12.48 percent) and health science practitioners (7.36 percent). Interestingly, researchers in the industry or other non-academic research institutions (3.07 percent) also participated in conversations about academic literature. Serving the role of science communicators, academic publishers (0.40 percent) and research feeds (0.29 percent) also contributed to the relevant discussions. Given that COVID-19 is a public health issue, mass media (1.55 percent) was relatively active in disseminating relevant information. More than 71 percent of users fell under the category of others, likely representing members of the general public. Additionally, civil society

organizations (2.04 percent) were observed to participate in discussions on relevant topics.

Regarding tweeting frequency, research feeds and academic publishers exhibited highly active tweeting behavior related to COVID-19 articles (Table 8). This highlights their aggressive tweeting behavior, with an average of 6.91 and 7.12 tweets, respectively. Moreover, they cited a greater number of articles on Twitter, with averages of 5.22 and 5.92 articles, respectively, compared to others. Active participation of commercial businesses was observed, with each of them averaging 2.43 tweeted articles. Both academia and non-academic researchers, as well as institutions, were enthusiastically engaged in discussions about COVID-19 publications on Twitter. Funding organizations may have also promoted relevant articles on Twitter. Health science professionals and institutions displayed enthusiasm for tweeting relevant articles, contributing an average of 2.34 tweets per user. Additionally, COVID-19 publications garnered attention from public authorities & politicians, civil society organizations, and other users, likely representing the general public.

7.2 Profiling audience of publications

Predicted user labels can facilitate audience profiling at various levels. For instance, Fig. 8 illustrates the demographic
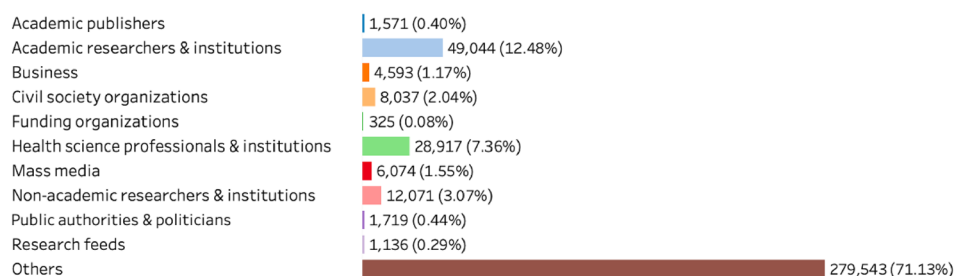
Fig. 7 Breakdown of demographics in sample dataset

Category	Value
Academic publishers	1,571 (0.40%)
Academic researchers & institutions	49,044 (12.48%)
Business	4,593 (1.17%)
Civil society organizations	8,037 (2.04%)
Funding organizations	325 (0.08%)
Health science professionals & institutions	28,917 (7.36%)
Mass media	6,074 (1.55%)
Non-academic researchers & institutions	12,071 (3.07%)
Public authorities & politicians	1,719 (0.44%)
Research feeds	1,136 (0.29%)
Others	279,543 (71.13%)

Table 8 Frequency of tweeting

User type	Avg. no. of tweets	Avg. no. of articles	Avg. no. of tweets per article
Academic publishers	**6.91**	**5.22**	**1.34**
Academic researchers and institutions	2.34	1.93	1.17
Business	**2.85**	**2.43**	1.18
Civil society organizations	2.26	1.86	1.14
Funding organizations	2.83	2.40	**1.25**
Health science professionals and institutions	2.75	2.33	1.11
Mass media	2.18	1.90	1.09
Non-academic researchers and institutions	2.44	2.04	1.14
Public authorities and politicians	2.01	1.77	1.09
Research feeds	**7.12**	**5.92**	**1.21**
Others	2.07	1.74	1.15

Fig. 8 Audience profiles of two research articles

segmentation of Twitter users discussing three articles. Despite a similar volume of Twitter mentions, these articles attract different audiences. Article A, centered on proteogenomics, is predominantly popular among researchers from both academic and non-academic backgrounds, as well as health science professionals. In contrast, Article B, addressing COVID-19 vaccination, has garnered more interest from the general public and sparked discussions among civil society organizations. Article C, which examines the impacts of COVID-19 lockdowns on cancer surgery operations, has

primarily engaged health science professionals and related institutions.

Predicted labels of tweeters can also help investigate the preferences of various types of users. Figure 9 presents an example of demographic breakdowns of tweeters discussing publications in different subject areas. For instance, except in the domain of social sciences and humanities, over half of tweeters citing COVID-19 publications appear to be members of the general public. As indicated by the percentages of tweeters, relevant Twitter scholarly communication was dominated by academic researchers & institutions and health science professionals & institutions in all five subject areas. Non-academic users have shown interest in publications from various areas. For example, over 5 percent of tweeters mentioning physical sciences and social sciences & humanities publications were non-academic researchers, whereas 4.1 percent of tweeters mentioning social sciences & humanities were civil society organizations.

8 Conclusion

Our study confirms that social network information can complement the text content from user profiles in identifying user types in the context of Twitter scholarly communication. While GNN models with social interaction graphs alone may not be sufficient to identify user categories, our results highlight the effectiveness of models that combine BERT and GNN mechanisms. Our optimized models had better performance than a stand-alone fine-tuned BERT model, which represents the state-of-the-art mechanism in text classification tasks. Evaluating GNN models with various node representation methods and message passing flows, our study confirms the value of user interactions in identifying user categories within the context of Twitter metrics.

Utilizing the proposed model, we identified academic researchers & institutions and health science professionals & institutions as the most significant contributors to Twitter scholarly communication in terms of number of participants, excluding the general public. Additionally, we observed the active engagement of non-academic entities such as mass media, industry researchers, and civil society organizations in discussions about scientific publications. This suggests that Twitter metrics may have the potential to indicate the translational impact of research. However, it is worth noting the aggressive tweeting behavior exhibited by academic publishers and research feeds. Their excessive tweets could compromise the reliability of Twitter metrics. Considering these findings, we recommend enhancing Twitter metrics by providing more granular demographic breakdowns to cater to the interests of diverse stakeholders. This approach would enable a comprehensive analysis of the impact of scholarly work from various perspectives.
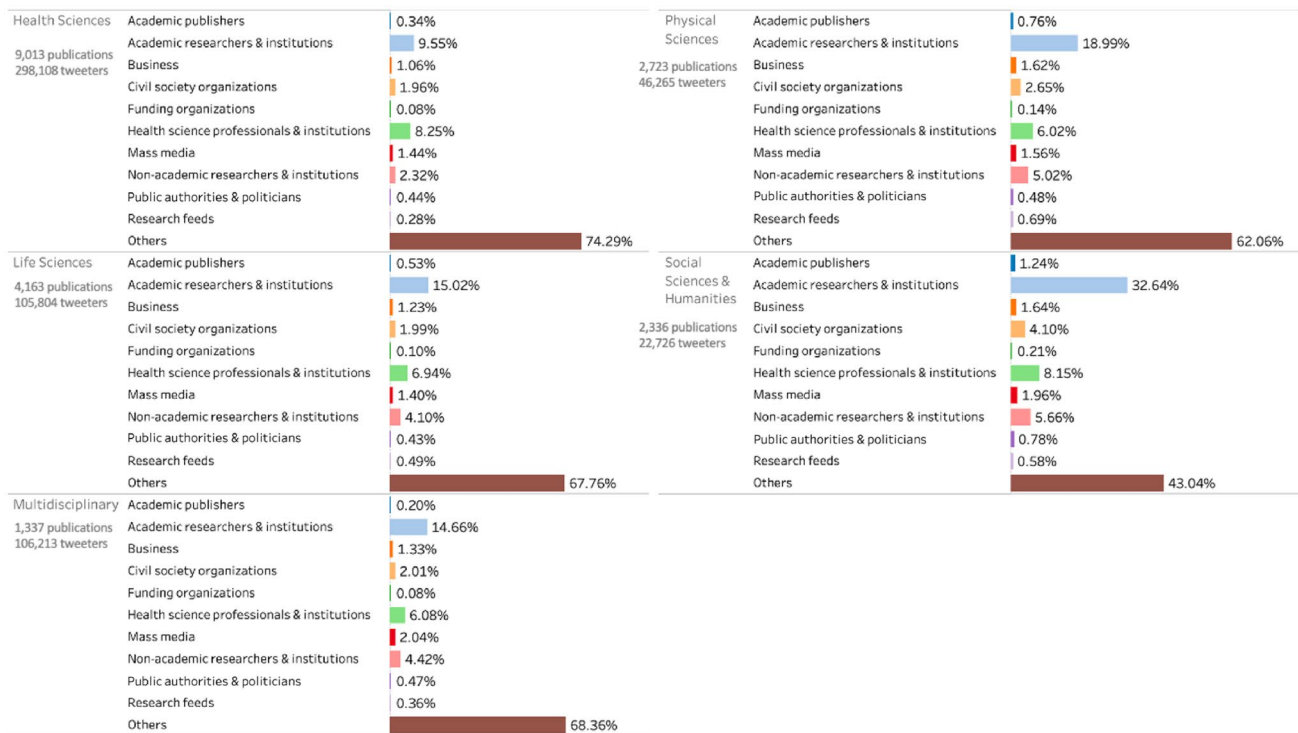

Fig. 9 User profiles by ASJC subject areas

One limitation of this study is that the classification model is trained solely on tweets mentioning COVID-19 articles, making its applicability primarily relevant to publications in similar domains, such as public health or medicine. To enhance the classification model's performance, researchers may consider incorporating additional user statistics related to tweeting behaviors such as the number of publications tweeted and the number of tweets, and metadata from users' profiles (i.e., the number of followers and friends, the number of statuses, the number of favorites, whether the account is verified) as input features. For studies focusing on online scholarly communication within academia, a more detailed classification scheme could be essential. It is beneficial to distinguish among various academic entities and individuals, such as higher education institutions, research institutes, academic associations, faculty members, research fellows, postgraduate students, and others. Future studies could delve into topics such as understanding the motivations driving various user categories to engage in tweeting about academic publications and investigating the dynamics of user interactions (i.e., communities involved in scholarly discussions on Twitter, information flow through relevant user interactions) within the context of Twitter scholarly communication. Relevant explorations would significantly contribute to advancing our comprehension of Twitter metrics.

Declarations

References

Abhari R, Vincent N, Dambanemuya HK, Bodon H, Horvát EÁ (2022) Twitter engagement with retracted articles: who, when, and how?. Retrieved from http://arxiv.org/abs/2203.04228.

Aljabri M, Zagrouba R, Shaahid A, Alnasser F, Saleh A, Alomari DM (2023) Machine learning-based social media bot detection: a comprehensive literature review. Soc Netw Anal Min 13(1):20. https://doi.org/10.1007/s13278-022-01020-5

Aljohani NR, Fayoumi A, Hassan S-U (2020) Bot prediction on social networks of Twitter in altmetrics using deep graph convolutional networks. Soft Comput 24(15):11109–11120. https://doi.org/10.1007/s00500-020-04689-y

Altmetric.com (2021) Source of Altmetrics-Twitter. Retrieved from https://help.altmetric.com/support/solutions/articles/6000235926-twitter

Basile V, Bosco C, Fersini E, Nozza D, Patti V, Pardo FM, Rosso P, Sanguinetti M (2019) SemEval-2019 task 5: multilingual detection of hate speech against immigrants and women in twitter. In: Proceedings of the 13th international workshop on semantic evaluation. p 54–63. https://doi.org/10.18653/v1/S19-2007

Brody S, Alon U, and Yahav, E (2021) How attentive are graph attention networks? ArXiv Preprint. Retrieved from https://arxiv.org/abs/2105.14491

Campbell WM, Baseman E, and Greenfield K (2014). Content + context networks for user classification in Twitter. In: Neural Information Processing Systems (NIPS) 2014 Workshop. Retrieved from https://aclanthology.org/W14-5909.pdf

Cao R, Liu XF, Fang Z, Xu X-K, Wang X (2023) How do scientific papers from different journal tiers gain attention on social media? Inf Process Manag 60(1):103152. https://doi.org/10.1016/j.ipm.2022.103152

Dehghan A, Siuta K, Skorupka A, Dubey A, Betlen A, Miller D, Wei X, Bogumil K, Prałat P (2023) Detecting bots in social-networks using node and structural embeddings. Journal of Big Data 10(1):119. https://doi.org/10.1186/s40537-023-00796-3

Devlin J, Chang M-W, Lee K and Toutanova K (2018) BERT: Pre-training of deep bidirectional transformers for language understanding. ArXiv Preprint

Díaz-Faes AA, Robison-García N, Bowman TD, and Costas R (2019) Towards a multidimensional classification of social media users around science on Twitter. In: 17th International Conference on Scientometrics and Informetrics (ISSI 2019). Retrieved from http://digital.csic.es/handle/10261/206040

Didegah F, Mejlgaard N, Sørensen MP (2018) Investigating the quality of interactions and public engagement around scientific papers on twitter. J Informet 12(3):960–971. https://doi.org/10.1016/j.joi.2018.08.002

Dukic D, Keca D, and Stipic D (2020) Are you human? Detecting bots on twitter using BERT. In: 2020 IEEE 7th international conference on data science and advanced analytics (DSAA). p 631–636. https://doi.org/10.1109/DSAA49011.2020.00089

Ferguson C, Inglis SC, Newton PJ, Cripps PJS, Macdonald PS, Davidson PM (2014) Social media: a tool to spread information: a case study analysis of Twitter conversation at the cardiac society of australia; New Zealand 61st annual scientific meeting 2013. Collegian 21(2):89–93. https://doi.org/10.1016/j.colegn.2014.03.002

Hamilton WL, Ying R, and Leskovec J (2017) Inductive Representation Learning on Large Graphs. Retrieved from http://arxiv.org/abs/1706.02216

Hassan S-UU, Imran M, Gillani U, Aljohani NR, Bowman TD, Didegah F (2017) Measuring social media activity of scientific literature: an exhaustive comparison of scopus and novel altmetrics big data. Scientometrics 113(2):1037–1057. https://doi.org/10.1007/s11192-017-2512-x

Haustein S, Bowman TD, Holmberg K, Tsou A, Sugimoto CR, Larivière V (2016) Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. J Am Soc Inf Sci 67(1):232–238. https://doi.org/10.1002/asi.23456

Hayawi K, Mathew S, Venugopal N, Masud MM, Ho P-H (2022) DeeProBot: a hybrid deep neural network model for social bot detection based on user profile data. Soc Netw Anal Min 12(1):43. https://doi.org/10.1007/s13278-022-00869-w

Holmberg K, Thelwall M (2014) Disciplinary differences in Twitter scholarly communication. Scientometrics 101(2):1027–1042. https://doi.org/10.1007/s11192-014-1229-3

Holmberg K, Vainio J (2018) Why do some research articles receive more online attention and higher altmetrics? Reasons for online success according to the authors. Scientometrics 116(1):435–447. https://doi.org/10.1007/s11192-018-2710-1

Jiang J, Ren X, and Ferrara E (2022) Retweet-BERT: political leaning detection using language features and information diffusion on social networks. Retrieved from http://arxiv.org/abs/2207.08349

Kousha K, Thelwall M (2020) COVID-19 publications: database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts. Quant Sci Stud. https://doi.org/10.1162/qss_a_00066

Li Y, Yang L, Xu B, Wang J, Lin H (2019) Improving user attribute classification with text and social network attention. Cogn Comput 11(4):459–468. https://doi.org/10.1007/s12559-019-9624-y

Lin T-Y, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2018.2858826

Lu Z, Du P, and Nie J-Y (2020) VGCN-BERT: augmenting BERT with graph embedding for text classification. In: lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics) pp 369–382). https://doi.org/10.1007/978-3-030-45439-5_25

Marco P, and Popescu A-M (2011) A Machine learning approach to twitter user classification. In: Proceedings of the fifth international AAAI conference on weblogs and social media

Müller M, Salathé M, and Kummervold PE (2020) COVID-Twitter-BERT: a natural language processing model to analyse COVID-19 content on twitter. Retrieved from http://arxiv.org/abs/2005.07503

Özkent Y (2022) Social media usage to share information in communication journals: an analysis of social media activity and article citations. PLoS ONE 17(2):e0263725. https://doi.org/10.1371/journal.pone.0263725

Priem J, Hemminger BH (2010) Scientometrics 2.0: new metrics of scholarly impact on the social Web. First Monday. https://doi.org/10.5210/fm.v15i7.2874

Robinson-Garcia N, Costas R, Isett K, Melkers J, Hicks D (2017) The unbearable emptiness of tweeting—about journal articles. PLoS ONE 12(8):e0183551. https://doi.org/10.1371/journal.pone.0183551

Robinson-Garcia N, van Leeuwen TN, Ràfols I (2018) Using altmetrics for contextualised mapping of societal impact: From hits to networks. Sci Public Policy 45(6):815–826. https://doi.org/10.1093/scipol/scy024

Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G (2009) The graph neural network model. IEEE Trans Neural Netw 20(1):61–80. https://doi.org/10.1109/TNN.2008.2005605

Vainio J, Holmberg K (2017) Highly tweeted science articles: who tweets them? An analysis of Twitter user profile descriptions. Scientometrics 112(1):345–366. https://doi.org/10.1007/s11192-017-2368-0

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need.

Veličković P, Cucurull G, Casanova A, Romero A, Liò P, and Bengio Y (2017) Graph attention networks. Retrieved from http://arxiv.org/abs/1710.10903

Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS (2021) A Comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst 32(1):4–24. https://doi.org/10.1109/TNNLS.2020.2978386

Xu K, Hu W, Leskovec J, and Jegelka S (2018) How powerful are graph neural networks? Retrieved from http://arxiv.org/abs/1810.00826

Yu H (2017) Context of altmetrics data matters: an investigation of count type and user category. Scientometrics 111(1):267–283. https://doi.org/10.1007/s11192-017-2251-z

Zhou J, Cui G, Shengding H, Zhang Z, Yang C, Liu Zhiyuan, Wang L, Li C, Sun M (2020) Graph neural networks: A review of methods and applications. AI Open 1:57–81. https://doi.org/10.1016/j.aiopen.2021.01.001