BotRGA: Neighborhood-Aware Twitter Bot Detection with Relational Graph Aggregation

Weiguang Wang [1,2], Qi Wang[3], Tianning Zang[1,2], Xiaoyu Zhang[1,2], Lu Liu[4], Taorui Yang[5], Yijing Wang[1]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3] National Computer Network Emergency Response Technical Team/Coordination Center of China
[4] China Assets Cybersecurity Technology CO., LTD
[5] Department of Accounting, The School of Business, Durham University, Durham, UK
{wangweiguang, zangtianning, zhangxiaoyu,wangyijing}@iie.ac.cn
{qwaterblue, liuluvaliant}@126.com
{cqlm57}@durham.ac.uk

Abstract. With the rapid development of AI-based technology, social bot detection is becoming an increasingly challenging task to combat the spread of misinformation and protect the authenticity of online resources. Existing graph-based social bot detection approaches primarily rely on the topological structure of the Twittersphere but often overlook the diverse influence dynamics across different relationships. Moreover, these methods typically aggregate only direct neighbors based on transitive learning, limiting their effectiveness in capturing the nuanced interactions within evolving social Twittersphere. In this paper, we propose BotRGA, a novel Twitter bot detection framework based on inductive representation learning. Our method begins with extracting the semantic features from Twitter user profiles, descriptions, tweets and constructing a heterogeneous graph, where nodes represent users and edges represent relationships. We then propose a relational graph aggregation method to learn node representations by sampling and aggregating the features from both direct and indirect neighbors. Additionally, we evaluate the importance of different relations and fuse the node's representations across diversified relations with semantic fusion networks. Finally, we classify Twitter users into bots or genuine users and learn model parameters. Extensive experiments conducted on two comprehensive Twitter bot detection benchmarks demonstrate that the superior performance of BotRGA compared to state-of-the-art methods. Additional studies also confirm that the effectiveness of our proposed relational graph aggregation, semantic fusion networks, and strong generalization ability to new and previously unseen user communities.

Keywords: Twitter Bot Detection, Relational Graph Aggregation, Semantic Representation Learning.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

1 Introduction

Along with the rapid development of artificial intelligence and Natural Language Processing (NLP) technology, social network bots have been widely used in various social network platforms, posing great challenges to the authenticity and information security of social networks. These social bots can realistically mimic human social behavior and language habits, and are used by malicious operators to spread disinformation, manipulate public sentiment and political interference. For example, in the past few years, Twitter bots have participated in US presidential election intervention [1], spread false information [2], and promote extremist ideologies [3]. Moreover, with the emergence of ChatGPT, the detection of social bots has become an urgent problem to be solved [4].

Earlier machine learning based Twitter bot detection methods generally utilize feature engineering to extract features from user profiles, and then use traditional machine learning algorithms to classify social robots [5-6]. However, the heavy reliance on analytical experience and subjective judgment in feature engineering leads to significant limitations for such methods in detecting the sophisticated and diversified social robots. Deep learning based methods use social network users profiles and posting contents as input to the neural network, and identify social bots by building a series of convolutional, recurrent neural network and other deep learning models [7-11]. These methods only consider the user profile and textual information, without utilizing the relations in social networks, making it difficult to achieve effective results in detecting the constantly evolving social bots. With the rapid development of graph neural networks, more and more methods have been proposed for detecting social bots by using graph neural networks for deep analysis of social network structures [12,13,14,15]. The above graph-based methods have achieved high recognition accuracy in social bot detection task, but they overlook the different influence weights of diversified relationship types in social networks and only consider the direct relationships between nodes in graph. Moreover, these methods generate node representations with transductive learning, fail to achieve strong generalization in evolving real-world social networks with dynamically added new nodes and relationships.

In light of these challenges, we propose a novel Twitter bot detection framework BotRGA (**Bot** Detection with **R**elational **G**raph **A**ggregation). Specifically, we construct a heterogeneous relational graph to present the Twitter social networks and adopt an inductive learning method to obtain user semantic representations by sampling and aggregating information from one's direct and indirect neighbors in the local neighborhood, then comprehensively integrate the user representations by semantic fusion networks across diversified relationships and conduct bot detection. Our main contributions are summarized as follows：

- We propose to comprehensively leverage the local neighborhood information and diversified relations in heterogeneous relational graphs constructed from real-word Twittersphere, and adopt an inductive method to learn user representation.
- On this basis, we propose BotRGA: a novel social bot detection framework. It is an end-to-end bot detector that uses relational graph aggregation to learn user

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

representation under different relations, then obtain the final node's representations by semantic fusion networks across diversified relations and conduct bot detection.
- We have conducted sufficient experiments to evaluate our proposed BotRGA and compared our method with state-of-the-art baseline methods. Experimental results demonstrate that our proposed method is more efficient and generalized than baseline methods.

2 Related Work

2.1 Graph Neural Network

Graph Neural Network is a deep learning based method for processing graph domain information. The notion of graph neural networks was initially outlined by Gori et al. [16] and further elaborated in Scarselli et al. [17]. These early studies fall into the category of recurrent graph neural networks (RecGNNs), and Li et al. [18] proposed an gated graph sequence neural networks to solve the challenges in previous research. In follow-up works, Kipf et al. [19] further simplify the graph convolutions through a localized first-order approximation and present graph convolutional networks (GCN). Hamilton et al. [20] propose the GraphSAGE framework based on node sampling and features aggregating. It can efficiently generate node embeddings by leveraging neighbor features. Inspired by the attention mechanisms, Velickovic et al. [21] present the graph attention networks (GAT) for node classification. Schlichtkrull et al. [22] proposed Relational Graph Convolutional Networks (RGCN) and apply GCN framework to modeling relational data. In recent years, due to the powerful expressive power of graph structures, GNN is widely used in social network analysis such as node classification, link prediction and graph community detection.

2.2 Twitter Bot Detection

Twitter bots are automated accounts run by software, pose a serious threat to the authenticity and integrity of online platforms. How to effectively detect social bots is a difficult challenge. Existing Twitter bot detection methods mainly fall into three categories: feature-based methods, text-based methods, and graph-based methods.

Feature-based methods generally focused on manually designed features and combined them with traditional machine learning classifiers. These methods conduct feature engineering based on handcrafted user features extracted from user profiles. Yang et al. [6] utilize minimal account metadata and labeled datasets to detect social bots. Davis et al. [8] leverage more than 1,000 user features to evaluate the extent to which a Twitter account exhibits similarity to the known characteristics of social bots. Wu et al. [9] adopt user behavioral sequences and characteristics as features for classifiers to detect bots. However, evolving bots can evade the detection of feature-based approaches by creating deceptive accounts with manipulated metadata and stolen tweets from genuine users [2].

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

Text-based methods adopt NLP techniques to detect Twitter bots with their historical tweets and user descriptions. Kudugunta et al. [7] proposed a bot detection framework based on contextual LSTM (Long Short-Term Memory) and exploits both user tweet content and account metadata. Wei et al. [10] use word embeddings to encode user historical tweets and adopt a Bi-directional LSTM to distinguish Twitter bots from human accounts. David et al. [29] present a BERT (Bidirectional Encoder Representation from Transformers) based bot detection model to analyze tweets written by bots and humans. However, text-based methods are easily deceived when advanced bots post stolen tweets and descriptions from genuine users [2].

Graph-based methods utilize the graph structure to represent the various users and diversified relationships in Twitter social networks, and attempt to separate bots and humans based on the graph structure. Ali Alhosseini et al. [14] adopt graph convolutional networks to aggregate the features of user node and conduct bot detection. Feng et al. [13] proposed BotRGCN framework and represent Twitter networks and uses R-GNNs for social bot detection. Lei et al. [15] propose a Twitter Bot detection framework BIC and employs a text-graph interaction module to enable information exchange across modalities in the learning process. The above works used graph convolutional networks to achieve higher detection accuracy than traditional classification methods in robot detection tasks, but both they ignored the different influence weights of diversified relationship types in social node classification. To address the aforementioned issues, Feng et al. [14] construct heterogeneous information networks and propose relational graph transformers to model influence intensity with the attention mechanism and learn node representations to detect social bot, and achieves state-of-the-art performance among graph-based methods. However, this method only considers node's one hop direct relationship in the relationship graph, and the processing of node features is relatively universal, without considering the theme and emotional features of tweet text. Recent years, new advanced AI-based social bots appeared on social platforms from time to time, with the ability to imitate humans and evade detection. The above approaches could not naturally generalize to those new unseen nodes of bots in Twittersphere, because they generate node embeddings by transductive learning and prediction on nodes in a single and fixed graph [20].

3 Methodology

3.1 Overview

Fig. 1 presents an overview of our proposed neighborhood-aware and relational graph aggregation Twitter bot detection framework BotRGA. Specifically, we first extract the semantic features encoding from Twitter user profiles, and construct a heterogeneous graph with users as nodes and relationships as edges. We learn node representations by sampling and aggregating its neighbor features under each relationship with our proposed relational graph aggregation. After that, we evaluate the importance of different relations and fuse the node's representations across diversified relations with semantic fusion networks. Finally, we classify Twitter users into bots or genuine users and learn model parameters.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

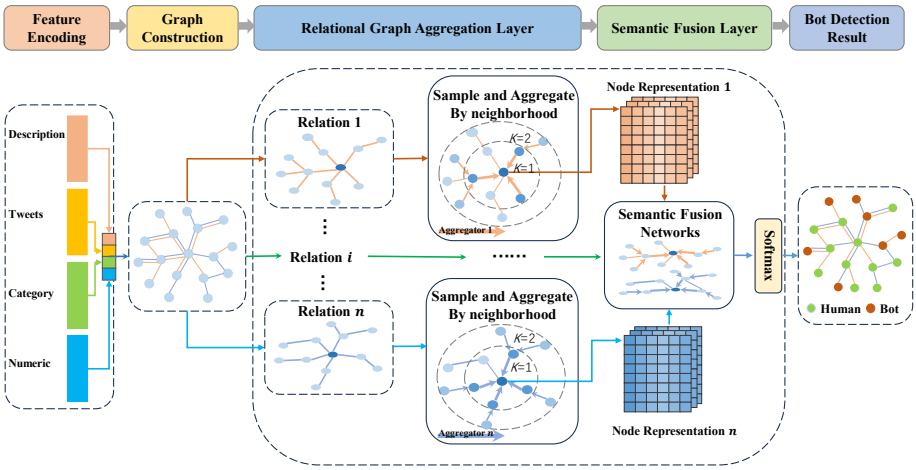

Fig. 1. Overview of our Twitter bot detection with relational graph aggregation framework.

3.2 Feature encoding

Similar to the Twitter feature extraction method in [13], we separately extracted Twitter user profile metadata and textual embedding, the metadata includes user categorical and numerical metadata and textual embeddings include the semantic representation of user description and historical tweets, and then we use a concatenate function to fuse the four types of features to form the initial encoding of user nodes, which can be defined as:

$$x_{init}^{i} = CONCAT(r_{des}; r_{tweet}; r_{cat}; r_{num}) \in \mathbb{R}^{D\times 1} \tag{1}$$

where x_{init}^{i} denotes the initialization encoding of the $i\text{-}th$ user node, $r_{des}, r_{tweet}, r_{cat}, r_{num}$ are respectively denotes the representation of description, tweet, category and numerical information extracted from a Tweet user, and D is the user embedding dimension, the detailed encoding strategy is as follows.

Since users of the Twittersphere come from different countries and regions, they usually have different language habits for personalized expressions to share their life experiences and observations on Twittersphere, so the description and tweets in user profiles often contain different languages, a quantity of emoji and named-entity, those textual content imply rich semantics topics and personal sentiments. In order to fully utilize the semantic information form user profiles, different from previous related works using a universal NLP model BERT to handle tweet text, we adopt TweetNLP [23] to learn the semantic embedding on user description and historical tweets, which is a Transformer-based language models and specialized on Twitter social media text. Furthermore, we use TweetNLP to extract topics and sentiments contained in user tweets to form numerical features, which will be introduced in later.

For numerical features, the method of directly using the current number of posts, likes, followers, and following in user profiles as features ignores the factor of user

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

survival time. Therefore, we combine the above values with user survival time as numerical features. In addition, with the support of AI based chatbot technologies such as ChatGPT, it is becoming increasingly difficult to distinguish a tweet is published by human or social bots in terms of syntax and semantics. However, there are still significant differences in posting emotions, topic scope, and other aspects. Therefore, we analyze and extract the topics and sentiments of user tweets, and abstract them into numerical features. The numerical features we extracted are shown in Table 1.

Table 1. Numerical user properties adopted in BotRGA.

Feature Name	Description
Active_days	number of Twitter account active days
Followers_growth_rate	followers_count/active_days
Followings_growth _rate	followings_count/active_days
Tweets_growth_rate	tweets_count/active_days
Status_growth_rate	status_count/active_days
Name_digit_length	number of digits in screen name
Name_upper_length	number of upper case in screen name
Screen name length	number of screen name character
Tweet_neutral_sentiments	number of user tweets with neutral sentiment
Tweet_topics	number of topics in user tweets

3.3 Graph Construction

Relation Graph can be expressed formally as $G(V, E, R)$, where V denotes the user nodes in Twittersphere social graph, E denotes the edges which connecting different user nodes, and R denotes the diversified relationships between user nodes.

We then construct a heterogeneous information network to represent the Twittersphere, which take Twitter users as nodes and take diversified relations types as different edges to connect different user in the social relational graph. We denote the set of relations in the network as R while our framework supports multiple type of relation settings.

In order to better fuse and utilize the four types of feature vectors, we transform the initialization value of the node encoding $x^{(0)}$ with a fully connected layer to serve as initial features in the GNNs, i.e.,

$$x^{(0)} = \sigma(W_0 \cdot x_{init}^{i} + b_0) \tag{2}$$

where W_0 and b_0 are learnable parameters, σ denotes nonlinearity and we use leaky-relu as σ.

3.4 Relational Graph Aggregation

In order to comprehensively utilize the user node's neighbor features information and its own features to reveal the deep semantic representations under relation r ($r \epsilon R$), at the same time, to ensure has high generalization and performance in social graph analysis scenarios with large amounts of data and dynamic updates, we propose relational

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

graph aggregation mechanism, a GNN architecture that separately learns embeddings by sampling and aggregating features from a node's local neighborhood on different relationships, formulated as:

$$h_{SN_i^r}^{r(k)} = AGGREGATE_k^r(h_u^{r(k-1)}, \forall u \in SN_i^r) \tag{3}$$

$$h_i^{r(k)} = sigmoid(W_k^r \cdot CONCAT(h_i^{r(k-1)}, h_{SN_i^r}^{r(k)}) + b_k^r) \tag{4}$$

where SN_i^r denotes the sampled i-th node's neighbors set under relation i, $h_u^{r(k-1)}$ denotes the representation of $k-1$ depth neighbor u in sampling set SN_i^r under relation i, $h_{SN_i^r}^{r(k)}$ denotes aggregated representation of the i-th node's neighbors, $h_i^{r(k)}$ denotes the learned representation of i-th node, $AGGREGATE_k^r$ is the aggregation function of k depth under relation i, W_k^r and b_k^r denote learnable parameters.

The size of SN_i^r is set to 25, the depth k is set to 2 by default following GraphSAGE [20]. We use max-pooling aggregator as the aggregation function due to empirical performance. After aggregating $depth\text{-}K$ neighborhood node information, we obtain a new node embedding $h_i^{r(k)}$。

In order to obtained smooth representation learning results, we adopt the gate mechanism to obtain the representation of node i by:

$$h_i^{r(K)} = tanh(h_{SN_i^r}^{r(k)}) \odot h_i^{r(k)} + x_i^{(0)} \odot (1 - h_i^{r(k)}) \tag{5}$$

where $\odot$ denotes the Hadamard product operation, $x_i^{(0)}$ denotes the initial features of node i in the GNNs, and $h_i^{r(K)}$ is the learned representation of node i for relation r in $depth\text{-}K$.

3.5 Semantic Fusion Networks

To aggregate more comprehensive semantic information, the multiple features needed to be revealed by different relation-paths. Moreover, the weights of relationships are different, treating each relationship equally weakens the semantic features which are aggregated by some more important relationships.

In order to address these issues, we propose a novel relation-based attention mechanism to obtain the importance of different relation-paths then utilized to aggregated various semantic information across different relationship to learn the node's semantic representation, defined as:

$$\alpha_d^{r(l)} = \sigma(\frac{1}{|V^r|}\sum_{i \epsilon V^r}(q_d^{(l)^T} \cdot \tanh(W_d^{r(l)} \cdot h_i^{r(l)} + b_d^{r(l)}))) \tag{6}$$

where $\alpha_d^{r(l)}$ denotes the learned importance of relation r at the d-th attention head, $|V^r|$ denotes the number of nodes under relation r, σ is sigmoid function, $q_d^{(l)^T}$ is semantic attention vector of relation r at the d-th attention head in layer l, $q_d^{(l)^T}$, $W_d^{r(l)}$, $b_d^{r(l)}$ are learned parameters.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

We then aggregate node information based on edge relationships with different weights, the formula is as follows:

$$x_i^{(l)} = \frac{1}{D}\sum_{d=1}^{D}[\sum_{r\in R}\alpha_d^{r(l)} \cdot h_i^{r(l)}] \tag{7}$$

where $x_i^{(l)}$ denotes the learned representation of node i aggregated from different relations in layer l, $h_i^{r(l)}$ denotes the results of relational graph transformers and D is the number of attention heads.

3.6 Learning and Optimization

After L layers of GNNs messages passing, we obtain the final node representations $x^{(L)}$, and transform them with an output layer and a softmax layer to get Twitter bot detection result, *i.e.*,

$$\hat{y}_i = softmax(W_O \cdot \sigma(W_L \cdot x_i^{(L)} + b_L) + b_O) \tag{8}$$

where $\hat{y}_i$ is out model's prediction of user i, all W and b are learnable parameters. We then adopt supervised annotations and a regularization term to train out model, formulated as:

$$Loss = -\sum_{i\epsilon Y}[y^i log(\hat{y}_i) + (1-y^i)log(1-\hat{y}_i)] + \lambda\sum_{\omega\epsilon\theta}\omega^2 \tag{9}$$

where Y is the annotated user set, y^i is the ground-truth labels, θ denotes all trainable parameters in the model and λ is a hyperparameter.

4 Experiments

4.1 Dataset

In order to verify the effectiveness of our proposed model, the data set needs to have a certain graph structure type. We conducted our experiments on two public data sets with topological relationships, TwiBot-20 [24] and TwiBot-22 [25], which are more representative of the current social network environment.

TwiBot-20 and TwiBot-22 are comprehensive Twitter bot detection benchmarks and provide user follow relationships to support graph-based methods. TwiBot-20 is proposed in 2020 and includes 229,573 Twitter users, 33,488,192 tweets and 455,958 follow relationships. TwiBot-22 is the largest public dataset to date for Twitter bot detection and includes 1,000,000 Twitter users, 88,217,457 tweets and 170,185,937 follow relationships. An overview of the datasets is provided in Table 2.

Table 2. Database Overview.

Dataset	Account	Bot	Human	Tweets	Edges
TwiBot-20	229,573	5,273	6,589	33,488,192	33,716,171
TwiBot-22	1,000,000	139,943	860,057	88,217,457	170,185,937

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

4.2 Baselines and experiment setting

In order to validate the effectiveness of our proposed Twitter bot detection model, we compare our graph-based approach with the following methods:

Yang et al. (2020) [6] use random forest classifier with minimal user metadata and derived features.

Kudugunta et al. (2018) [7] propose to jointly leverage user tweet semantics and user metadata.

Botometer (2016) [8] is a bot detection service that leverages more than 1,000 user features.

Wei et al. (2019) [10] use recurrent neural networks to encode tweets and classify users based on their tweets.

Alhosseini et al. (2019) [14] use graph convolutional networks to learn user representations and conduct bot detection.

BotRGCN (2021d) [13] constructs a heterogeneous graph to represent the Twittersphere and adopts relational graph convolutional networks for representation learning and bot detection.

Feng et al. (2022) [14] propose relational graph transformers to model heterogeneous influence between users and use semantic attention networks to aggregate messages across users and relations and conduct heterogeneity-aware Twitter bot detection.

BotBuster (2022) [13] is a social bot detection system that processes user metadata and textual information to enhance cross-platform bot detection.

We use pytorch [26], pytorch lightning [27], torch geometric [28] for an efficient implementation of our proposed Twitter bot detection framework. We conduct all experiments on a server with 2 Tesla V100 GPUs with 32 GB memory, 32 CPU cores, and 300GB CPU memory. To directly and fairly comparing with previous works, we follow the same train, valid and test splits provided in the benchmark.

4.3 Main Results

We then benchmark these bot detection models on TwiBot-20 [24] and TwiBot-22 [25] and present results in Table 3, which demonstrates that:

- BotRGA consistently and significantly outperforms all baseline methods across the two datasets. Specifically, compared with the previously state-of-the-art method proposed by Feng et al [14], BotRGA achieves 1.2 % higher accuracy and 1.1% higher F1-score on TwiBot-20, and also provides a gain of 1.7% F1-score compared with the second best method on TwiBot-22.
- Graph-based methods for Twitter bot detection, such as BotRGA (Ours), BotRGCN [13], and Feng et al [14], demonstrate higher classification effectiveness compared to traditional non-graph methods like Yang et al [6] and Kudugunta et al [7]. This underscores the critical importance of leveraging the topological structure for node classification tasks in social networks.
- We propose the first relation-based and neighborhood-aware bot detection frameworks, which achieves the best performance on a comprehensive benchmark. Our results highlight the necessity of aggregating semantic information from diverse user

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

relationships, validating the effectiveness of our approach in addressing this challenge. Additionally, our method outperforms existing approaches, emphasizing its potential for advancing Twitter bot detection capabilities.

Table 3. Accuracy and binary F1-score of Twitter bot detection systems on two datasets. Bold indicates the best performance, underline the second best. This table indicates that the results of our method BotRGA is significantly better than the second best baseline.

Model	TwiBot-20		TwiBot-22	
	Accuracy	F1-score	Accuracy	F1-score
Yang et al.	0.8191	0.8546	0.7508	0.3659
Kudugunta et al.	0.8174	0.7515	0.6578	0.5167
Botometer	0.4801	0.6266	0.4990	0.4275
Wei et al.	0.7126	0.7533	0.7020	0.5360
Alhosseini et al.	0.6813	0.7318	0.4772	0.3810
BotRGCN	0.8462	0.8707	0.7887	0.5499
Feng et al.	0.8664	0.8821	0.7650	0.4294
BotBuster	0.7724	0.8118	0.7406	0.5418
BotRGA(Ours)	**0.8783**	**0.9035**	**0.7947**	**0.5671**

4.4 Ablation Study

In order to effectively identify social bot on the Twittersphere, we propose a novel graph-based social bot detection method BotRGA, which comprehensively utilizes and integrates user information and topology in social networks. Especially, we adopt the follower and following relationships between different users as edges to construct a heterogeneous relational graph.

To prove the effectiveness of our graph construction method, we remove different types of edges and obtain results under the ablation setting in Table 4. It is illustrated that the graph with both follower and following edges outperforms any reduced settings. The results also prove the effectiveness of our proposed method to construct heterogeneous relational graph of Twittersphere.

Table 4. Ablation studying removing different relationship of our constructed Relational Graph.

Ablation Settings	Accuracy	F1-score
only follower relationship	0.8531	0.8623
only following relationship	0.8587	0.8649
follower + following relationship(homogeneous)	0.8652	0.8726
follower + following relationship(heterogeneous)	**0.8783**	**0.9035**

Upon obtaining a Heterogeneity graph, we propose relational graph aggregation to learn node representation by fusion relation-based local neighborhood property information. To prove the effectiveness of our proposed GNN architecture, we conduct ablation study on relational graph aggregation and semantic fusion networks separately,

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

and report the results under different settings in Table 5. It is illustrated that the values of accuracy and F1-score drop significantly when removing our proposed Relational Graph Aggregation and Semantic Fusion Networks independently. Furthermore, when adopting 3 popular GNN models and 4 commonly used fusion algorithms to replace our proposed GNN architecture respectively, the values of accuracy and F1-score have increased but are still lower than our model. The experiment results proved that the relational graph aggregation and semantic fusion networks are all essential parts of our proposed GNN architecture.

To sum up, both our constructed relational graph and our proposed GNN architecture are effectiveness and contribute to our model's outstanding performance.

Table 5. Ablation study of our proposed GNN architecture.

Ablation Settings	Accuracy	F1-score
remove Relational Graph Aggregation	0.8571	0.8691
remove Semantic Fusion Networks	0.8605	0.8758
replace Relational Graph Aggregation with GAT	0.8646	0.8853
replace Relational Graph Aggregation with GCN	0.8625	0.8812
replace Relational Graph Aggregation with RGCN	0.8655	0.8854
Summation as Semantic Fusion Networks	0.8663	0.8862
Mean pooling as Semantic Fusion Networks	0.8651	0.8825
Max pooling as Semantic Fusion Networks	0.8676	0.8863
Min pooling as Semantic Fusion Networks	0.8545	0.8801
full model	**0.8783**	**0.9035**

4.5 Generalization Study

As Twitter bots are constantly evolving in real world [2], the research on effective detection of bots calls for models to better generalize to unseen user accounts. To this end, we evaluate BotRGA's generalization ability to detect bots in unseen user nodes in TwiBot-20 [24] benchmark. Specifically, TwiBot-20 [24] has collected Twitter user accounts which created between 2008 and 2020, so we divide the dataset by year and take the accounts created in 2020 as testing set, the accounts before 2020 as training set and validation set. This division ensures that the user accounts in the testing set are unseen in training. After that, we use BotRGA and baseline models to conduct generalization studies in the above dataset. The result is presented in Fig. 2, the graph-based models BotRGA, Feng et al. [14] and BotRGCN [13] generally outperform the feature-base models. From the result we can see, our proposed BotRGA achieves the highest accuracy of 0.869 and F1-score of 0.8907 among these models, and its accuracy outperforms the second-best model of Feng et al. [14] by 2.7% on unseen accounts. This indicates that incorporated with relational graph aggregation and semantic fusion networks, BotRGA could better generalize to previously unseen user accounts.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

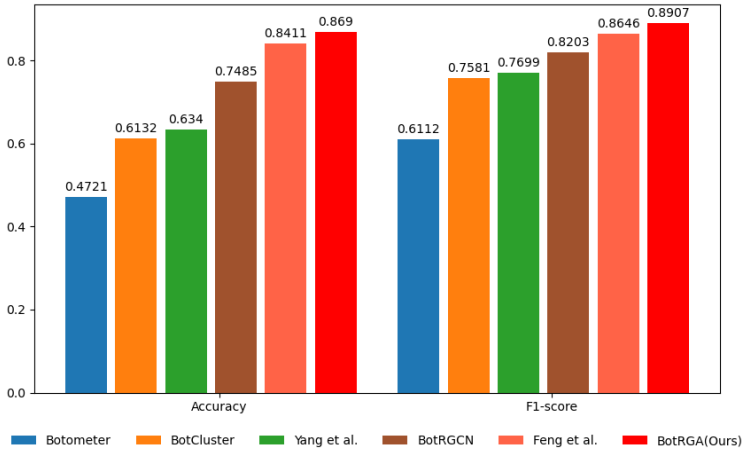

Fig. 2. Generalization study on TwiBot-20[24] indicates that we proposed BotRGA is better at generalizing to unseen user nodes than baselines approaches.

4.6 Representation Learning Study

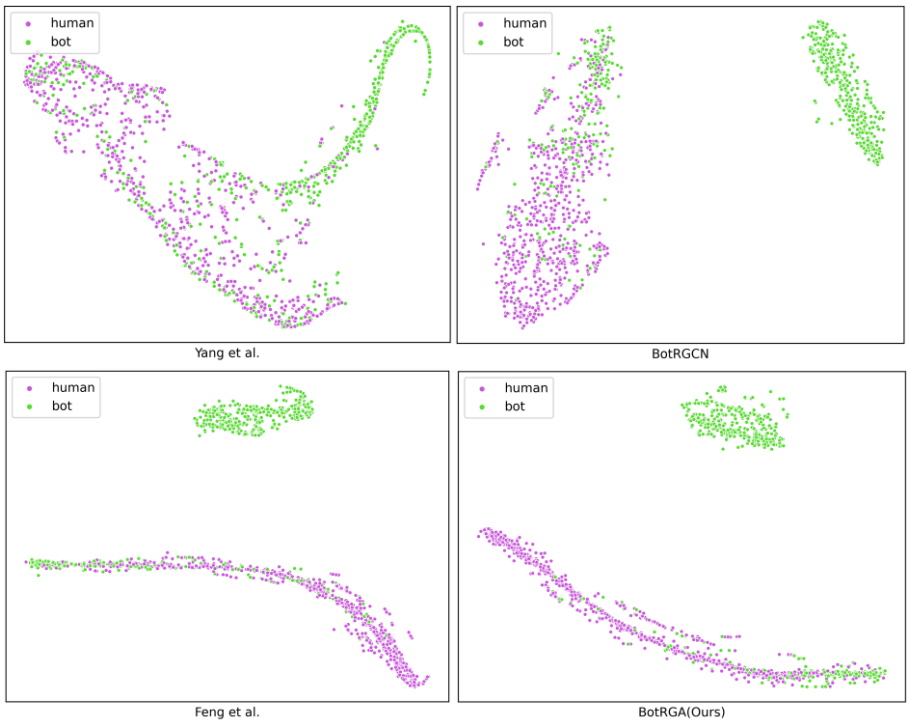

Fig. 3. Plots of Twitter user representations learned with our model and different baselines on TwiBot-20, the result indicates that the learned representations of BotRGA(Ours) have higher discrimination between the group of human and bots than baseline models.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

Using GNNs for social bot detection is essentially to learn the user's representation to distinguish humans and bots in social networks. In order to verify the effectiveness of our proposed social bot detection model for user representation learning, we present the T-SNE [30] plot of user representation of our method BotRGA and baselines using Twibot-20 [24] dataset in Fig. 3. The result illustrates that the learned representation of our proposed BotRGA and Feng et al [14] have clearer discrimination between the group of human and bots than the other two baselines on Twittersphere. Compared with Feng's model, the representations learned by our model have higher purity in the clusters of human and bots, it indicates that our proposed method could learn higher quality Twitter user representation to distinguish the group of social bots and human in Twittersphere.

5 Conclusion and Future Work

With the rapid development of AI-based technology, social bot detection has gradually to be an important and challenging task. We propose BotRGA, a novel Twitter bot detection framework with neighborhood-aware and relational graph aggregation to inductively learn the user deep semantic representations in heterogeneous social network. Extensive experiments demonstrate that BotRGA significantly advances the state-of-the-art on two Twitter bot detection benchmarks. Further studies demonstrate the effectiveness of our relation-based features aggregation strategy, and prove the superior generalization ability and representation learning efficiency of BotRGA.

Social network is a dynamically updated temporal network, and the user's survival status and mutual relationships continue to change over time. Especially with the support of large language models, such as ChatGPT, social bots have become more humanoid, how to detect social bots accurately and effectively becomes very difficult. In future work, we plan to experiment with more diversified ways to face the dynamic update scenario of social networks and extend our graph-based detection approach.

Acknowledgments. This work was supported by the Defense Industrial Technology Development Program (Grant JCKY2021906A001), and the National Natural Science Foundation of China (NSFC) (Grant 62376265).

References

1. Deb, A., Luceri, L., Badaway, A. and Ferrara, E.: Perils and challenges of social media and election manipulation analysis: The 2018 us midterms. In: Companion proceedings of the 2019 world wide web conference. pp. 237-247 (2019)
2. Cresci, S.: A decade of social bot detection. In: Communications of the ACM, 63(10), pp.72-83. (2020)
3. Berger, Jonathon M., and Jonathon Morgan.: The ISIS Twitter Census: Defining and describing the population of ISIS supporters on Twitter. (2015)
4. Ferrara, Emilio.: Social bot detection in the age of ChatGPT: Challenges and opportunities. In: First Monday (2023).

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

5. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A. and Tesconi, M.: DNA-inspired online behavioral modeling and its application to spambot detection. In: IEEE Intelligent Systems, 31(5), pp.58-64 (2016)
6. Yang, K.C., Varol, O., Hui, P.M. and Menczer, F.: Scalable and generalizable social bot detection through data selection. In: Proceedings of the AAAI conference on artificial intelligence, Vol. 34, No. 01, pp. 1096-1103 (2020)
7. Kudugunta, S. and Ferrara, E.: Deep neural networks for bot detection. In: Information Sciences, 467, pp.312-322 (2018)
8. Davis, C.A., Varol, O., Ferrara, E., Flammini, A. and Menczer, F.: Botornot: A system to evaluate social bots. In: Proceedings of the 25th international conference companion on world wide web, pp. 273-274 (2016)
9. Wu, Jun, Xuesong Ye, and Chengjie Mou.: Botshape: A novel social bots detection approach via behavioral patterns. arXiv preprint arXiv:2303.10214 (2023).
10. Wei, F. and Nguyen, U.T.: Twitter bot detection using bidirectional long short-term memory neural networks and word embeddings. In: 2019 First IEEE International conference on trust, privacy and security in intelligent systems and applications, pp. 101-109 (2019)
11. Ng, Lynnette Hui Xian, and Kathleen M. Carley.: Botbuster: Multi-platform bot detection using a mixture of experts. In: Proceedings of the International AAAI Conference on Web and Social Media. Vol. 17. (2023)
12. Ali Alhosseini, S., Bin Tareaf, R., Najafi, P. and Meinel, C.: Detect me if you can: Spam bot detection using inductive representation learning. In: Companion proceedings of the 2019 world wide web conference, pp. 148-153 (2019)
13. Feng, S., Wan, H., Wang, N. and Luo, M.: BotRGCN: Twitter bot detection with relational graph convolutional networks. In: Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 236-239 (2021)
14. Feng, S., Tan, Z., Li, R. and Luo, M.: Heterogeneity-aware twitter bot detection with relational graph transformers. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, No. 4, pp. 3977-3985 (2022)
15. Lei, Zhenyu, Herun Wan, Wenqian Zhang, Shangbin Feng, Zilong Chen, Jundong Li, Qinghua Zheng, and Minnan Luo.: Bic: Twitter bot detection with text-graph interaction and semantic consistency. arXiv preprint arXiv:2208.08320 (2022)
16. Gori, Marco, Gabriele Monfardini, and Franco Scarselli.: A new model for learning in graph domains. In: 2005 IEEE International Joint Conference on Neural Networks, Vol. 2 (2005)
17. Scarselli, Franco.: The graph neural network model. In: IEEE transactions on neural networks 20.1, pp. 61-80 (2008)
18. Li, Yujia, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel.: Gated graph sequence neural networks. In: arXiv preprint arXiv:1511.05493 (2015)
19. Kipf, Thomas N., and Max Welling.: Semi-supervised classification with graph convolutional networks. In: arXiv preprint arXiv:1609.02907 (2016)
20. Hamilton, Will, Zhitao Ying, and Jure Leskovec.: Inductive representation learning on large graphs. In: Advances in neural information processing systems 30 (2017)
21. Veličković, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio.: Graph attention networks. In: arXiv preprint arXiv:1710.10903 (2017)
22. Schlichtkrull, Michael, Thomas N. Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling.: Modeling relational data with graph convolutional networks. In: The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15, pp. 593-607. Springer International Publishing (2018)

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13

23. Camacho-Collados, J., Rezaee, K., Riahi, T., Ushio, A., Loureiro, D., Antypas, D., Boisson, J., Espinosa-Anke, L., Liu, F., Martínez-Cámara, E. and Medina, G.: Tweetnlp: Cutting-edge natural language processing for social media. arXiv preprint arXiv:2206.14774 (2022)
24. Feng, Shangbin, et al.: Twibot-20: A comprehensive twitter bot detection benchmark. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management (2021)
25. Feng, Shangbin, et al.: TwiBot-22: Towards graph-based Twitter bot detection. In: Advances in Neural Information Processing Systems 35 (2022)
26. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. and Desmaison, A.: Pytorch: An imperative style, high-performance deep learning library. In: Advances in neural information processing systems, 32 (2019)
27. Falcon, W.A.: Pytorch lightning. Homepage: https://lightning.ai
28. Fey, M. and Lenssen, J.E.: Fast graph representation learning with PyTorch Geometric. In: arXiv preprint arXiv:1903.02428 (2019)
29. Dukić, David, Dominik Keča, Dominik Stipić.: Are you human? Detecting bots on Twitter Using BERT. In: International Conference on Data Science and Advanced Ana-lytics (DSAA). (2020)
30. Van der Maaten, Laurens, and Geoffrey Hinton.: Visualizing data using t-SNE. In: Journal of machine learning research 9.11 (2008)

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_13