



# Graph Attention Network-based User Intent Identification Method for Social Bots

**Yuxin Wu**

College of Engineering, Carnegie Mellon University, Moffett Field, CA 94035, USA.

**How to cite this paper:** Yuxin Wu. (2025) Graph Attention Network-based User Intent Identification Method for Social Bots. *Advances in Computer and Communication*, 6(4), 200-205.  
DOI: 10.26855/acc.2025.10.008

**Received:** August 28, 2025  
**Accepted:** September 25, 2025  
**Published:** October 24, 2025

\***Corresponding author:** Yuxin Wu, College of Engineering, Carnegie Mellon University, Moffett Field, CA 94035, USA.

---

## Abstract

A study was conducted on a user intent identification method for social bots based on graph attention networks, addressing the challenges of information manipulation and content abuse caused by automated accounts on social platforms. The method constructs a heterogeneous graph structure from users, content, and interaction relationships, integrating textual semantics, behavioral patterns, and network topology features. A multi-head attention mechanism is employed to achieve differentiated aggregation of neighborhood information. In the experimental design, publicly available international social media datasets were selected for comparative testing. The results demonstrate that the proposed method outperforms traditional text classification and graph convolutional models in multi-class intent recognition tasks, achieving higher accuracy and stability. It also exhibits stronger robustness in long-tail category detection and adversarial scenarios. The findings indicate that the graph attention network-based framework can effectively enhance the precision and interpretability of social bot intent identification, providing feasible technical support for platform governance and content regulation.

## Keywords

Graph Attention Network; Social Bots; User Intent Identification; Multimodal Features; Robustness Optimization

---

## Introduction

With the global proliferation of social media, automated accounts have increasingly become critical vehicles for disseminating misinformation and manipulating public discourse. Such accounts often influence collective opinion through high-frequency reposting, patterned commenting, and cross-platform content replication, thereby undermining public trust. Existing identification approaches mainly focus on account detection or content classification; however, these methods struggle to capture the genuine intent of users in the presence of complex relational networks and dynamic behavioral patterns. In recent years, graph neural networks have attracted significant attention due to their ability to model the structural relationships between users and content. Among them, graph attention networks demonstrate particular advantages in selective neighborhood information aggregation and relational differentiation. The identification of user intent thus emerges not only as a new academic challenge but also as a pressing technical requirement for ensuring governance and compliance in social platforms.

## 1. Design of the Graph Attention Network-based Identification Method

### 1.1 Intent Classification and Task Definition for Social Bots

The primary task in user intent identification is to establish a clear labeling system and scope of classification. In social media environments, the intents exhibited by social bots mainly include categories such as advertising

promotion, agenda manipulation, deceptive interaction, automated responses, and information relaying. These behaviors often overlap with or mimic those of genuine users, making them difficult to distinguish. To ensure the scientific validity of the classification system, behavioral features from publicly available datasets are incorporated to define intent as a multi-class classification problem, while also allowing for fine-grained label extensions. For example, advertising promotion may be further divided into commercial product marketing and malicious spamming, whereas agenda manipulation can be differentiated into political discourse infiltration and trending-event steering. The task is formalized as a semi-supervised multi-class learning problem, where limited labeled samples guide the inference of unlabeled nodes by leveraging both structured and unstructured data, thereby forming a transferable recognition framework [1].

## 1.2 Graph Structure Modeling and Multimodal Feature Construction

The identification of social bot intent relies on modeling complex interaction relationships. Users, content, hashtags, and hyperlinks are represented as nodes, while retweets, replies, follows, and mentions are abstracted as directed or undirected edges, forming a heterogeneous graph structure. Node attributes are described using multimodal features, including textual semantic features, behavioral rhythm features, and network topology indicators. Maria et al. (2020) proposed extracting features from user metadata and recent tweets and applying random forest classification, demonstrating the value of multidimensional features in bot detection. Wu et al. (2021) further extracted 30 features from four dimensions—metadata, interaction, content, and temporal behavior—to distinguish bots from normal users, yet their method showed limitations in capturing dynamic relationships and camouflaged behaviors. Abreu et al. (2021) classified Twitter accounts using basic platform functionalities and four machine learning algorithms, underscoring the importance of structural and behavioral information but lacking cross-modal integration. Hu et al. (2023) developed a multidimensional dynamic feature detection model optimized with a random forest algorithm based on the AUC metric, which improved detection accuracy to some extent but remained vulnerable to hidden or falsified features [2].

## 1.3 Structure and Implementation of the Graph Attention Network

In this task, the role of the graph attention network is to assign differentiated weights to heterogeneous neighborhoods in order to capture the relationships most indicative of user intent. In practice, distinct attention parameters are allocated to different types of relationships, and a multi-head attention mechanism is employed to enhance semantic representations during aggregation. For input features, the network applies a linear transformation followed by a nonlinear activation to initialize node embeddings, after which attention coefficients are calculated to perform weighted integration of neighbor information. During training, cross-entropy loss is adopted to handle class imbalance, combined with label smoothing to improve generalization. In the inference phase, neighborhood sampling is used to reduce computational complexity and prevent excessive overhead on large-scale graphs [3]. This mechanism ensures that the model maintains global consistency while effectively capturing local critical relationships, thereby enhancing both the accuracy and interpretability of user intent classification.

## 2. Key Technical Implementations and Optimization Strategies

### 2.1 Attention-based Aggregation of Heterogeneous Relations

In social bot intent recognition tasks, users, content, and interactions form diverse heterogeneous relations, represented as various edge types such as retweets, replies, follows, and mentions. These relations provide crucial cues for distinguishing intent but also introduce the risk of feature homogenization when aggregated through multiple layers of graph convolution. A well-known issue is over-smoothing, where node embeddings converge to similar values as network depth increases, thus weakening discriminative power [4]. As illustrated in Fig. 1, after one convolutional layer, node 1 and node 2 still retain distinct neighborhood information, but after two layers, their aggregated neighbors nearly overlap, resulting in highly similar embeddings. This effect is particularly pronounced in bot detection, as automated accounts often mimic surface-level user behavior, making neighborhoods indistinguishable.

To address this issue, an LSTM-based attention residual connection is introduced, mathematically defined as:

$$n' = \sum (\alpha^T \cdot [n_1, n_2, \dots, n_k]) \quad (1)$$

$$\alpha = \text{soft max}(W \cdot (\vec{h}, \hat{h}) + b) \tag{2}$$

Here,  $n'$  denotes the final node representation;  $\alpha$  is the attention weight vector, assigning different importance to neighbors;

$W$  and  $b$  are trainable parameters and  $\hat{h}$  represent forward and backward hidden states from the bidirectional LSTM, capturing sequential dependencies among neighbors [5]. This mechanism ensures that node embeddings retain heterogeneity by selectively aggregating neighborhood information, thereby mitigating over-smoothing and improving both accuracy and interpretability in intent recognition.

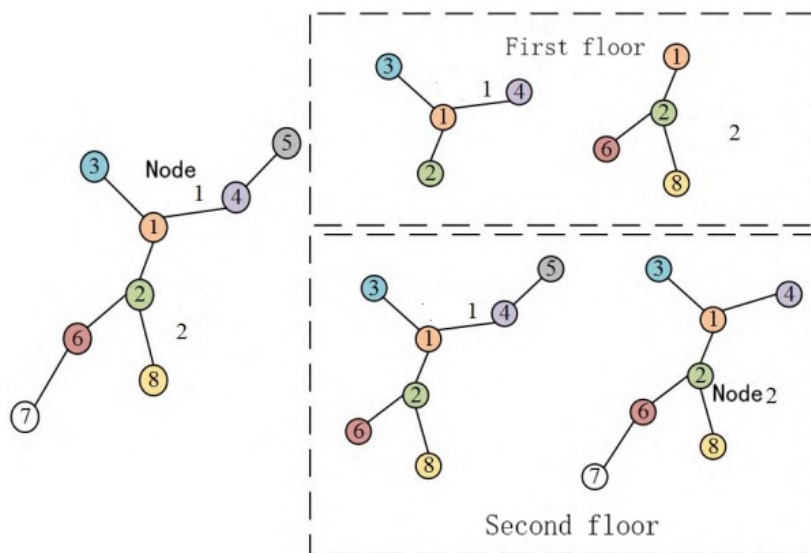


Figure 1. Diagram of over G smoothing.

### 2.2 Coordinated Alignment of Textual Semantics and Structural Features

Intent recognition for social bots cannot rely solely on graph topology; textual semantics also play a critical role. Semantic features reveal stylistic and emotional tendencies in user-generated content—advertising bots often use repetitive promotional language, while manipulative bots embed emotionally charged or polarizing terms within trending topics. Structural features, in contrast, capture relational patterns in the interaction graph, such as degree centrality, retweet depth, and community clustering. While both dimensions are valuable, naive fusion can dilute discriminative information, leading to what is termed “feature dilution.”

To overcome this, a coordinated alignment strategy is employed. Text embeddings are derived from pretrained language models, while structural metrics are extracted from the graph. Both are projected into a unified latent space, ensuring comparability across modalities in terms of scale and dimensionality. An attention module is then applied to assign adaptive weights to features, enabling the model to emphasize the most relevant cues [6]. For instance, in advertising bots, semantic signals carry higher weight, whereas in agenda-manipulation bots, structural anomalies dominate. This alignment not only enhances cross-modal coupling but also reduces redundancy, enabling the model to isolate subtle yet critical signals of intent. Experiments on public social media datasets demonstrate that coordinated alignment significantly improves classification accuracy, particularly in multi-class tasks and in detecting long-tail categories where either semantics or structure alone is insufficient.

### 2.3 Time-sensitive Dynamic Intent Modeling

Behavioral patterns of social bots often display strong temporal characteristics, such as bursts of high-frequency posting or long-term maintenance of rigid intervals. Ignoring these signals risks misclassifying automated accounts as legitimate users. To integrate temporal dynamics, time features are treated as critical attributes alongside textual and structural representations. Position encodings and temporal windowing are introduced to differentiate short-term bursts from sustained behaviors, ensuring that the model captures both event-driven spikes and long-term regularities.

The time-gated update process is formulated as:

$$\gamma_t = \sigma(W_h \cdot h_t + W_\tau \cdot \tau_t + b) \quad (3)$$

$$h_t' = \gamma_t \Theta h_t + (1 - \gamma_t) \Theta h_{t-1} \quad (4)$$

Here,  $h_t$  represents the hidden state of a node at time  $t$ ,  $\tau_t$  denotes the time feature vector,  $W_h$  and  $W_\tau$  are trainable matrices,  $b$  is a bias term, and  $\gamma_t$  is the time gate that controls the balance between current and historical states. The updated representation  $h_t'$  emphasizes recent activity while gradually down-weighting outdated behaviors. In practice, this mechanism enables the model to capture manipulative bots that launch coordinated posting campaigns during the onset of major events. By selectively emphasizing temporally salient interactions, the framework strengthens its capacity to detect dynamic manipulation intents that static models often miss.

### 3. Experimental Design and Evaluation

#### 3.1 Dataset Construction and Annotation Protocol

To evaluate the graph attention network-based method for social bot intent identification, experiments were conducted on publicly available international social media datasets, including the Twitter Bot Repository, Cresci-2017, and the more recent TwiBot-20 dataset. These datasets contain both genuine users and automated accounts, accompanied by multimodal information such as textual content, user metadata, and interaction networks. To ensure scientific rigor, a unified data cleaning process was applied: accounts with severe missing values or redundancies were removed, while samples retaining text, relational, and temporal features were preserved [7].

For labeling, a combination of existing annotations and manual verification was employed. Multiple annotators independently reviewed samples, and inter-rater reliability was validated using Cohen's Kappa, with values greater than 0.8 considered acceptable. To align with the research objectives, intent categories were reorganized to encompass four major types: advertising promotion, agenda manipulation, automated responses, and information relaying.

The datasets were divided in a 7:2:1 ratio into training, validation, and test sets, ensuring fairness and reproducibility of evaluation. This partitioning strategy enabled consistent benchmarking across different models and facilitated the assessment of generalization capability (see Table 1).

Table 1. Basic Statistics of Datasets

Dataset	Number of Users	Number of Tweets	Bot Ratio
Cresci-2017	8,000	3.2M	47%
Twitter Bot Repo	12,500	4.8M	42%
TwiBot-20	11,830	5.3M	41%

#### 3.2 Evaluation Metrics and Baseline Models

The evaluation framework was designed with multiple dimensions to ensure comprehensiveness and reproducibility. Classification performance was measured using Accuracy, Macro-Precision, Macro-Recall, and Macro-F1, which mitigate the effects of class imbalance. In addition, the Area Under the ROC Curve (AUC) was introduced to assess performance on long-tail categories.

In terms of experimental procedure, the LSTM-based attention residual connection defined in equations (1) and (2) was specifically tested for its ability to mitigate over-smoothing and preserve node distinctiveness, while the time-gated mechanism in equations (3) and (4) was examined for its contribution to dynamic behavior modeling. To this end, ablation studies were conducted by selectively removing the residual connection module and the time-gated module, and comparing performance differences to determine their impact within the overall framework.

The baseline models covered diverse approaches: Support Vector Machines (SVM) and BiLSTM for text-driven classification, Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT) for structural modeling, and Heterogeneous Attention Networks (HAN) for multimodal integration. All models were trained and tested under identical data splits and feature preprocessing conditions, ensuring fair comparison. This experimental design

effectively validated the advantages of the proposed approach in multi-class intent recognition and highlighted the central role of attention aggregation and temporal modeling in improving performance.

### 3.3 Performance and Ablation Analysis

The proposed framework was evaluated across three publicly available social media datasets, focusing on multi-class intent recognition. Comparative results revealed that the model consistently outperformed baseline methods in terms of Macro-F1 and AUC, with particular improvements in detecting long-tail categories such as advertising promotion and agenda manipulation. This demonstrated that coordinated modeling of graph structure and multimodal information substantially enhanced discriminative capacity.

Ablation experiments further verified the contribution of each component. When the LSTM-based residual connection defined in equations (1) and (2) was removed, performance in deeper layers degraded, with node embeddings converging and Macro-F1 dropping by approximately five percentage points. This confirmed the module's role in alleviating over-smoothing and maintaining representational diversity. Similarly, when the time-gated mechanism defined in equations (3) and (4) was excluded, recall on event-driven datasets declined significantly, underscoring the importance of temporal sensitivity for detecting manipulative bots [8].

Moreover, the experiments highlighted the complementary effects of both mechanisms. With the full model, attention-based aggregation ensured selective structural information capture, while time-gating strengthened sequential pattern recognition. Their integration enabled stable performance in complex social media environments, confirming the practical effectiveness and extensibility of the proposed method.

### 3.4 Interpretability Cases and Error Analysis

Interpretability was examined through attention weight distributions. For an advertising bot on Twitter, the model—via the residual attention mechanism in equations (1) and (2)—assigned higher weights to frequently reposted commercial link nodes, while down-weighting ordinary interactions, accurately exposing its promotional intent. In the case of an agenda-manipulation bot, the time-gated mechanism in equations (3) and (4) captured its concentrated reposting behavior during the onset of a major event, revealing hidden manipulation strategies.

Despite stable overall performance, several error types were observed. Some automated response bots generated text that closely resembled human language, making semantic differentiation difficult. Conversely, certain legitimate users engaged in high-frequency activity during short time windows, which led to misclassification as manipulative bots. These findings indicate that the model still faces challenges with borderline cases. Case reviews suggest that stronger cross-modal coordination and richer contextual modeling could reduce error rates and further improve the interpretability of complex behavioral patterns.

## 4. Conclusion

Addressing the complexity and dynamism of social bot intent recognition, this study proposed a framework based on graph attention networks. By employing differentiated aggregation of heterogeneous relations, coordinated alignment of multimodal features, and time-sensitive modeling, the approach effectively mitigated over-smoothing while enhancing performance on long-tail categories and dynamic scenarios. Experimental results on international benchmark datasets confirmed high accuracy, stability, and strong interpretability. Case analyses further demonstrated how attention weights and time-gating mechanisms revealed critical behavioral patterns, offering reliable support for platform governance and risk control. Future work should explore cross-platform adaptation and adversarial robustness to improve the generalizability and practical utility of this approach.

## References

- [1] Pires RP, Almeida AT. Interact2Vec — An efficient neural network-based model for simultaneously learning users and items embeddings in recommender systems. *Appl Soft Comput.* 2025;181:113408.
- [2] Akhtar MM, Bhuiyan SN, Masood R, et al. BotSSCL: Social Bot Detection with Self-Supervised Contrastive Learning. *Online Soc Netw Media.* 2025;48:100318.
- [3] Li L, Liu G, Liu Y, et al. A novel rumor detection method focusing on social psychology with graph attention network. *Neuro-computing.* 2025;626:129609.

- 
- [4] Chythanya KN, Bh. P, D. K, et al. A proposal of bidirectional grid long short term memory based model for user intention identification in on-line search query using text. *Mater Today Proc.* 2020;[prepublish].
  - [5] Ko HJ, Bae HJ, Hong D. Variable Impedance Control and Fuzzy Inference Based Identification of User Intension for Direct Teaching of a Mobile Robot. *J Korean Soc Precis Eng.* 2016;33(8):647-54.
  - [6] Wang S. A Cross-Distribution User Intention Identification Based on Topic Transfer. *IOP Conf Ser: Mater Sci Eng.* 2019;490(6).
  - [7] Brahma B, Kumari ALG, Panigrahi SB, et al. Interpretable Temporal-Spatial Graph Attention Network with Hyperbolic Sine Optimizer Algorithm for Alzheimer's Disease Diagnosis Through Multiscale Feature Modeling. *Biomed Mater Devices.* 2025;[prepublish]:1-24.
  - [8] Luo X, Shu P, Liu N, et al. DG-MSGAT: A Biologically-informed Differential Gene Multi-Scale Graph Attention Network for predicting neoadjuvant therapy response in rectal cancer. *Comput Methods Programs Biomed.* 2025;271:108974.