Twitter Bot Detection

SWS3023 Project Showcase

Shangbin Feng, Herun Wan, Qingyue Zhang, Zhaoxuan Tan

School of Electronic and Information Engineering

Xi'an Jiaotong University

July 3, 2022

Table of Contents

What is social media?

Definition

Social media are interactive technologies that allow the creation or sharing/exchange of information, ideas, career interests, and other forms of expression via virtual communities and networks.

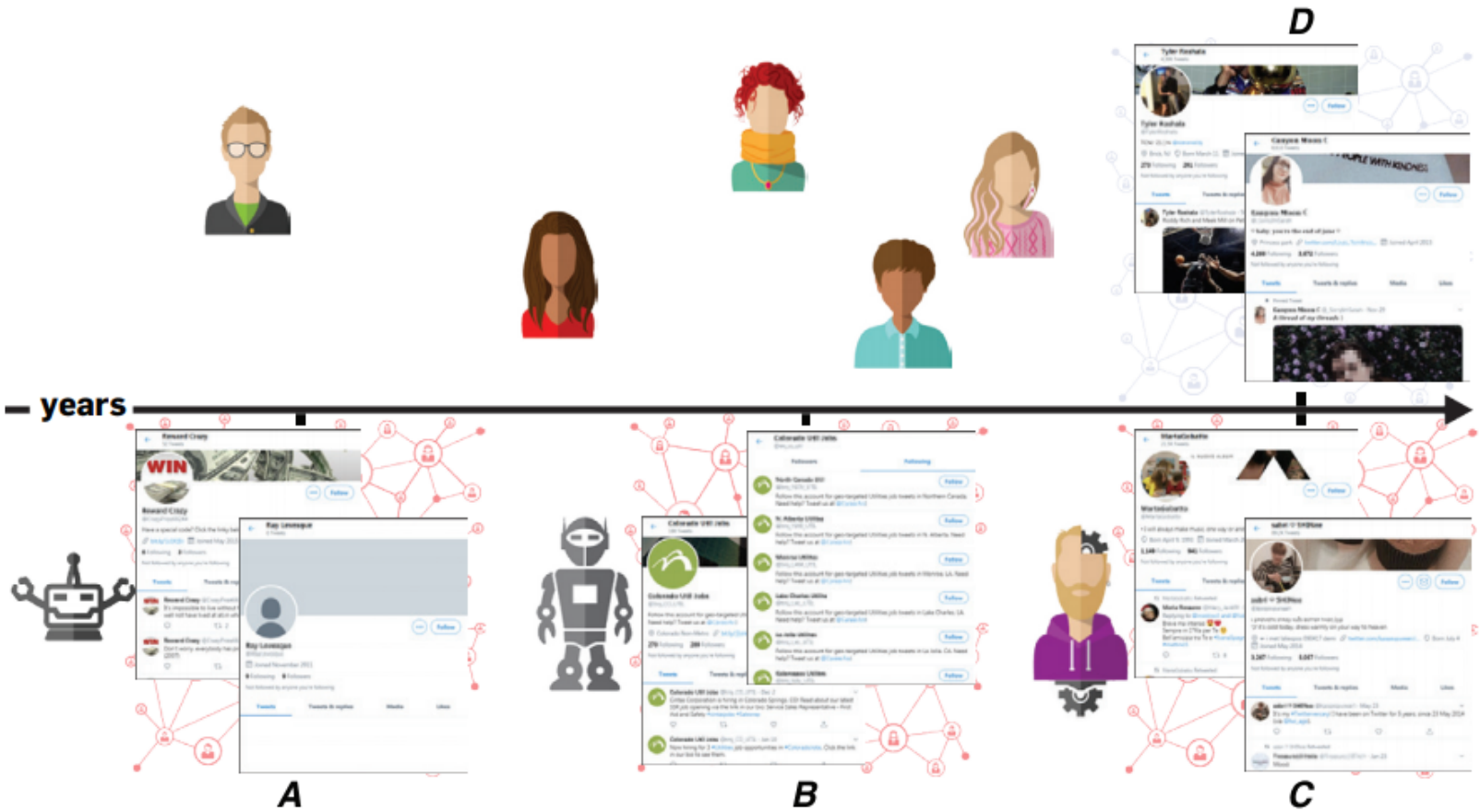
What is Twitter?

A social media platform with 330 million monthly active users.
"Twitter, its what's happenning."

What is a Twitter bot?

Definition

A **Twitter bot** is a type of bot software that controls a Twitter account via automated programs and the Twitter API.



Why Twitter bot detection?

benign vs. malicious, negative social impact

Are 'bots' manipulating the 2020 conversation? Here's what's changed since

Twitter Bots Are a Major Source of Climate Disinformation

Twitter Bots Are Spreading Massive Amounts of COVID-19 Misinformation

Task: Twitter bot detection

Definition

The task of **Twitter bot detection** aims to identify automated users with the help of their semantic, property and neighborhood information.

- semantic: natural language text in tweets
- property: numerical and categorical features such as follower count and whether the user is verified
- neighborhood: follow relations and the graph structure they form

How did previous models conduct Twitter bot detection?

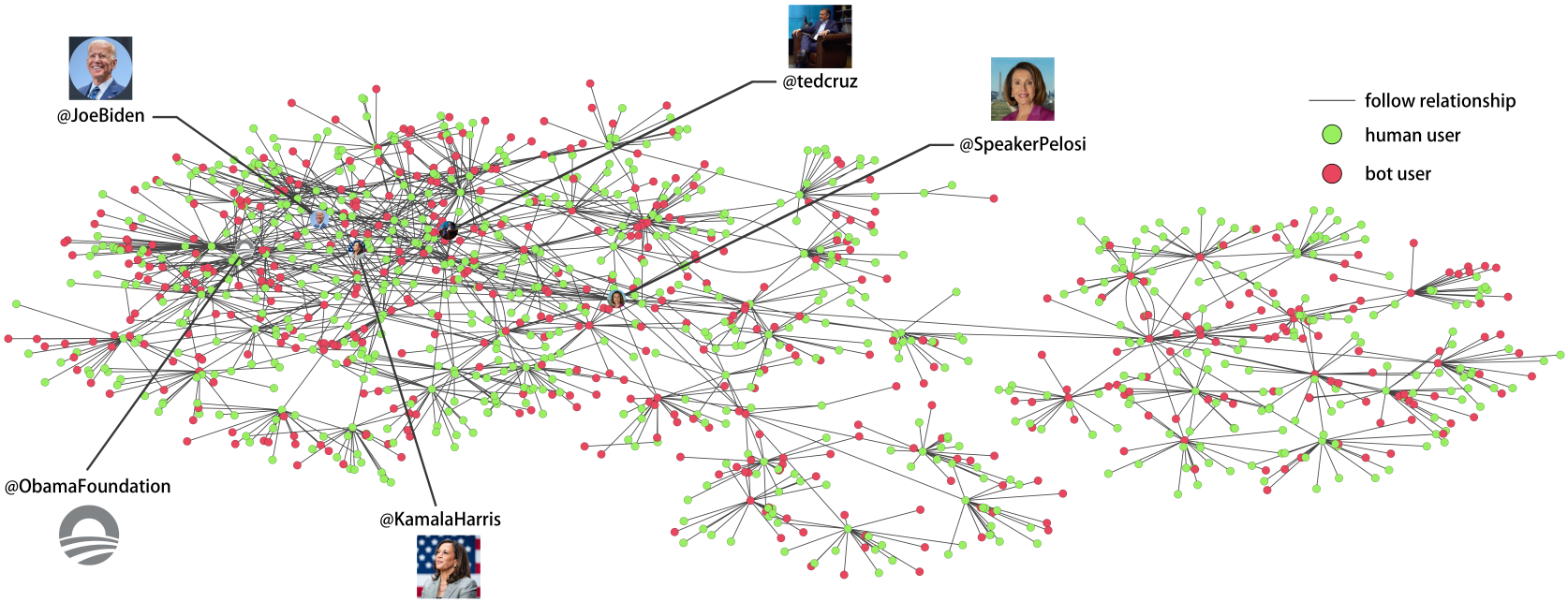
Related Work

Previous Twitter bot detection methods could be divided by two properties: user information and supervision.

- semantic + supervised
- semantic + unsupervised
- property + supervised
- property + unsupervised

Neighborhood?

such rich graph structure, but...



Motivation

traditional web-based classification: isolated data points

- email spam detection
- movie review sentiment analysis
- social media user profiling

social media such as Twitter: interconnected entities

- follow
- favourites
- retweet

Motivation: We should build Twitter bot detection systems that effectively leverage neighborhood information!

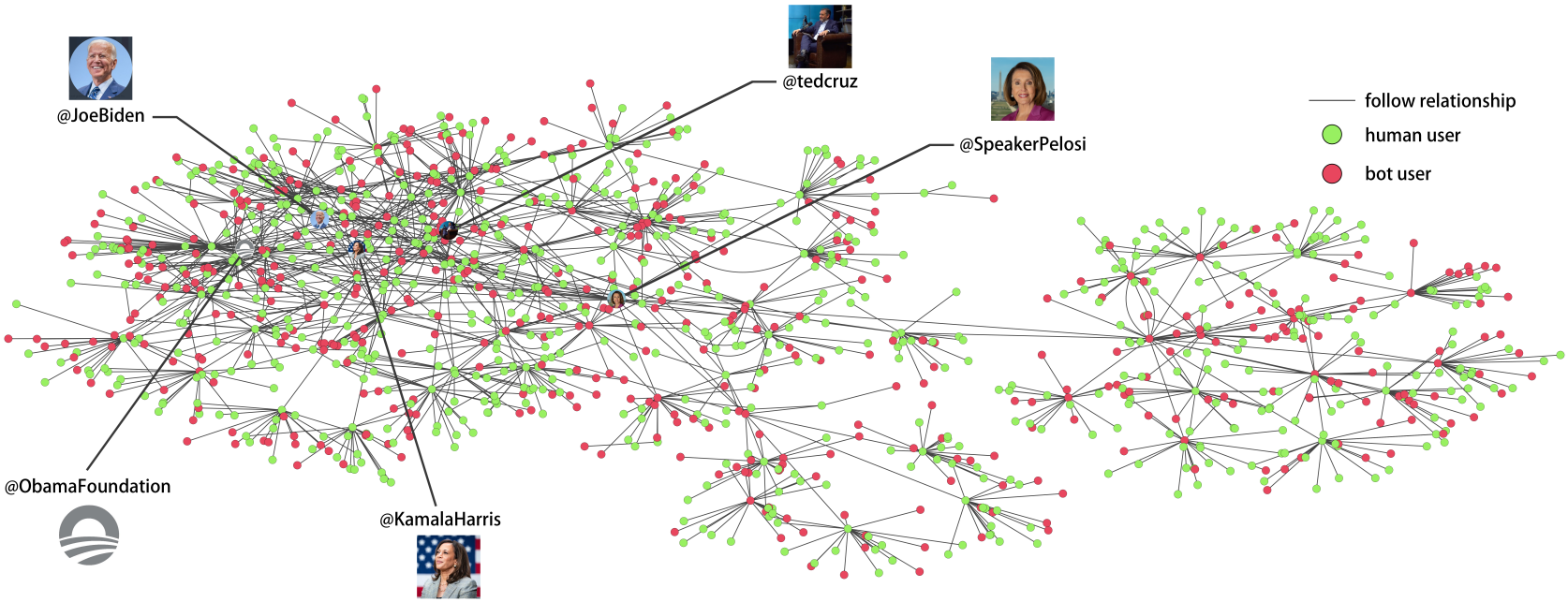
Preparation

To implement a graph-based Twitter bot detection system, we need...

- A Twitter bot detection dataset with the graph structure.
- Effective deep learning techniques of graph analysis.

Preparation #1: TwiBot-20 Dataset

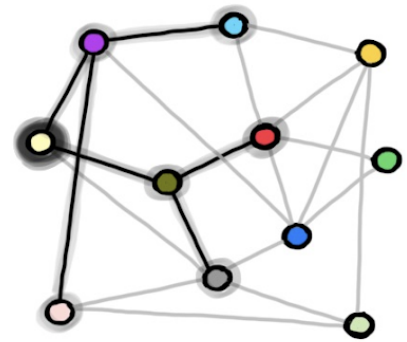
"To the best of our knowledge, TwiBot-20 is the first publicly available bot detection dataset that includes user follow relationships."

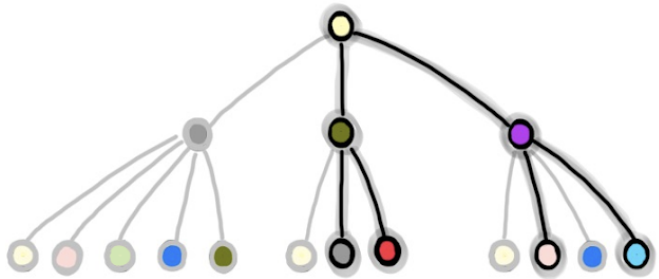


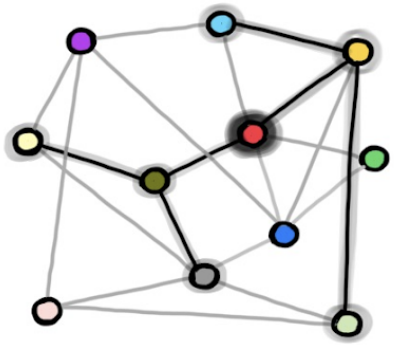
We supplement TwiBot-20 with semantic and property information to result in a dataset with 229,573 Twitter users, 8,723,736 user property items, 33,488,192 tweets and 455,958 follow links.

Preparation #2: Graph Neural Networks

Graph neural networks expand node neighborhood to form computation graphs, aggregates information from it to learn node representation.

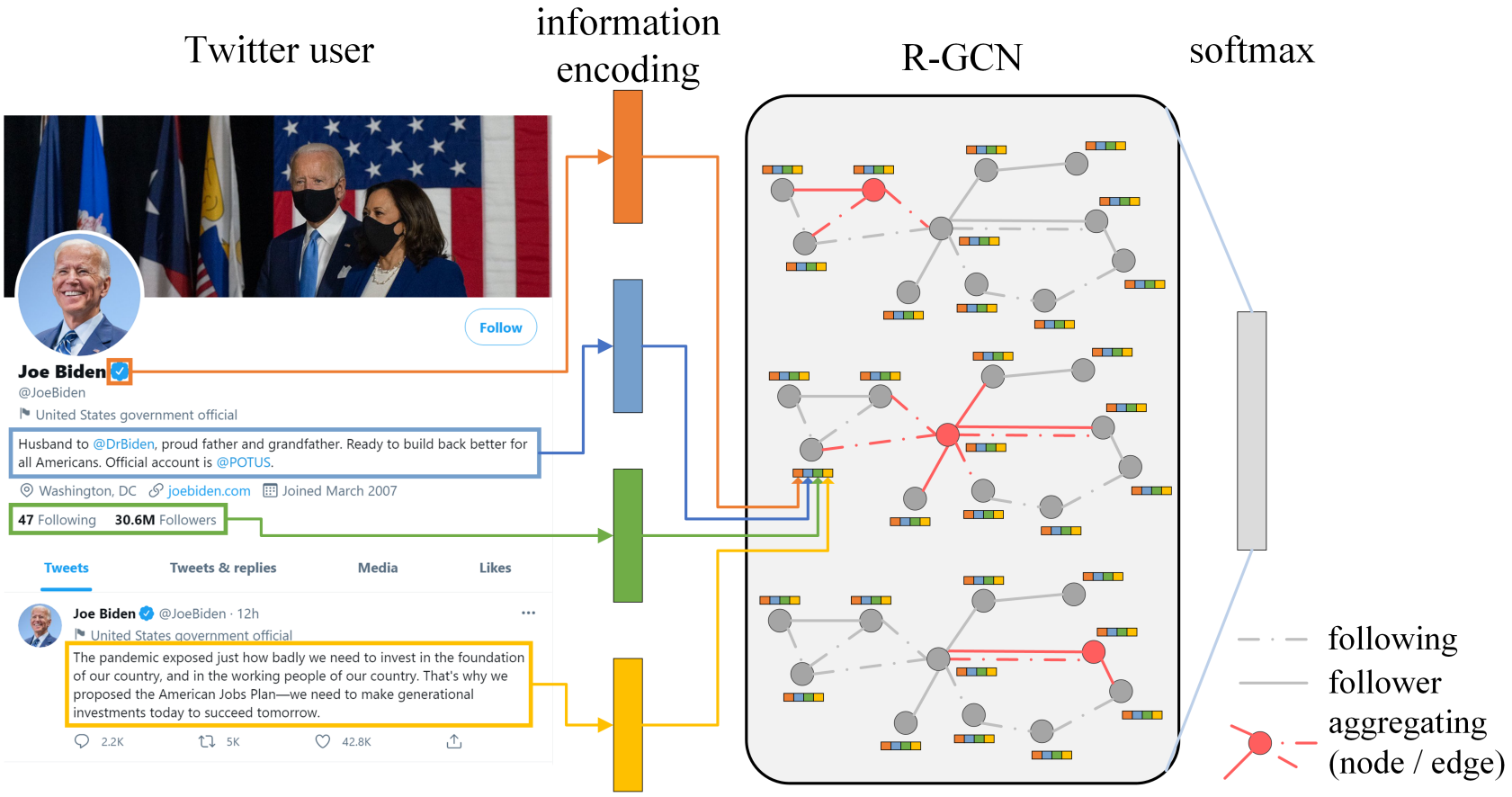






Our Proposal

We propose the pipeline of identifying Twitter bots with graph neural networks and multi-modal user information.



Our Proposal

We study the problem of Twitter bot detection from four different perspectives, building upon the TwiBot-20 dataset and graph neural networks (GNNs).

- **Feature Engineering**: Which features are important?
- **Graph Operator**: Which GNN architectures are effective?
- **Graph Heterogeneity**: Which types of relations exist on Twitter?
- **Contrastive Learning**: How to detect bots with few labels?

Perspective 1: Feature Engineering

Q: Which features are important for Twitter bot detection?
We propose four user feature sets,

	feature set	examples or description
1	numerical features	user follower count
2	categorical features	whether the user is verified
3	user description	RoBERTa encoding of user description
4	user tweet	RoBERTa encoding of user tweets

As a result, we represent a Twitter user with 5, 11, 768 and 768 features in four different feature engineering settings and conduct Twitter bot detection with graph convolutional networks.

Perspective 2: Graph Operator

Q: Which graph neural network architectures are effective in the task of Twitter bot detection?

We adopt these following GNNs and test out their performance on Twitter bot detection:

1. GCN
2. GAT
3. GraphSAGE
4. GraphConv
5. Gated GraphConv
6. ChebConv
7. ResGatedGraphConv
8. TransformerConv
9. ClusterGCNConv
10. ARMAConv

Perspective 3: Graph Heterogeneity

Q: Which types of inter-user relations exist on Twitter?
We adopt three different criteria to divide user into two types and divide edges between users into three types.
As a result, we obtain a heterogeneous graph of the orginal Twitter-sphere.

User Types	Description
verified	whether users are verified
follower	whether users have more than 7,000 followers
status	whether users have more than 10,000 tweets

We conduct experiments on homogeneous graphs with GAT, heterogeneous graphs with R-GCN and compare their performance.

Perspective 4: Contrastive Learning

Q: How to detect Twitter bots with few annotated users?
Data annotation in bot detection is time-consuming and subject to bias.

Definition

Self-supervised learning is a representation learning method where a supervised task is created out of the unlabelled data.

Definition

Contrastive learning is a machine learning technique used to learn the general features of a dataset without labels by teaching the model which data points are similar or different.

We conduct contrastive learning on tweets with SimCSE and on the graph with GRACE. We test out whether introducing the self-supervised contrastive learning will lead to better task performance.

Evaluation Metrics

We train our models with the TwiBot-20 dataset. We evaluate them on the test set and report accuracy and F1-score results.

Definition

Let TP, TN, FP, FN be the true positives, true negatives, false positives and false negatives.

Definition

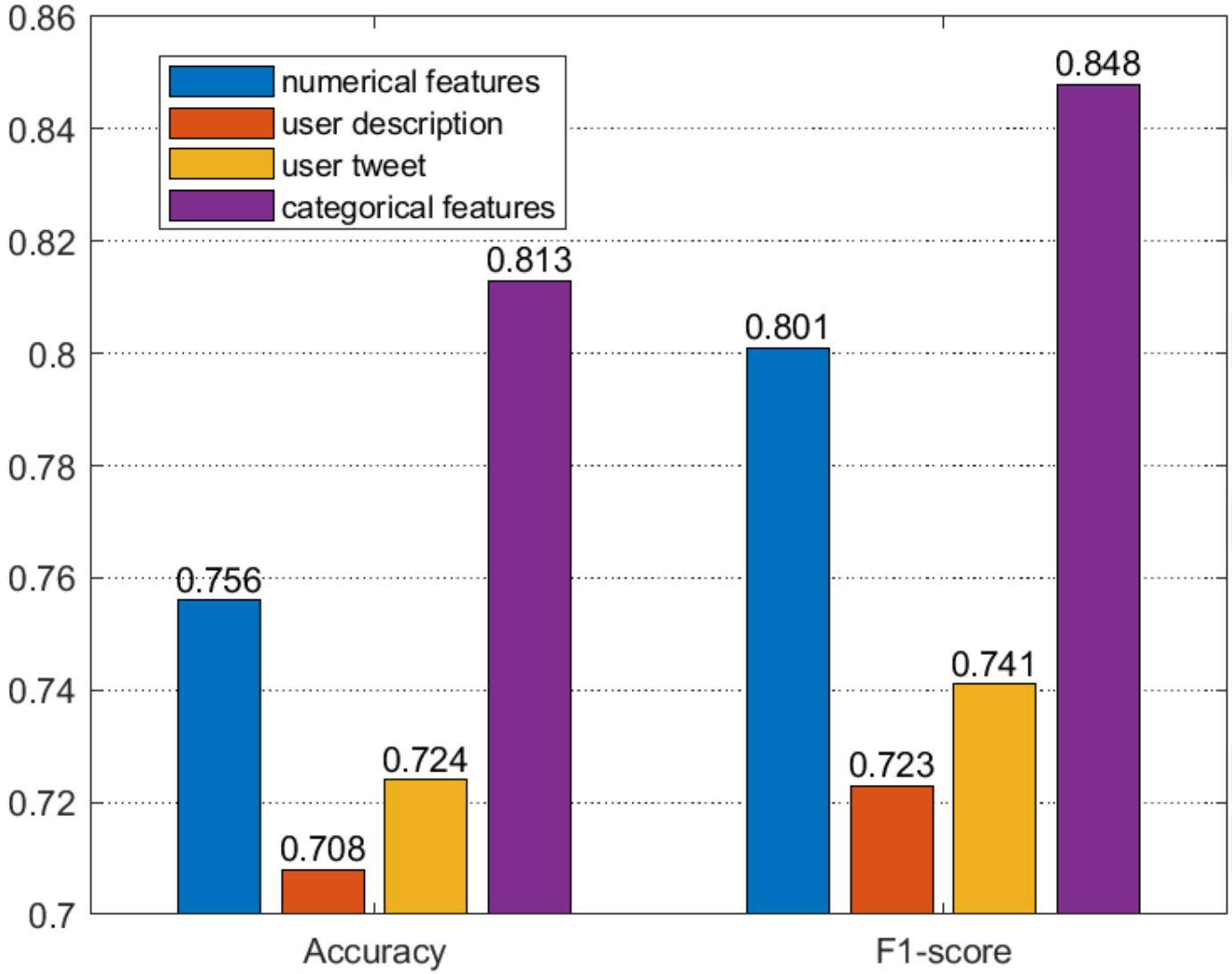
$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Definition

$$\text{F1-score} = \frac{2\times precision\times recall}{precision+recall} = \frac{TP}{TP+\frac{1}{2}(FP+FN)}$$

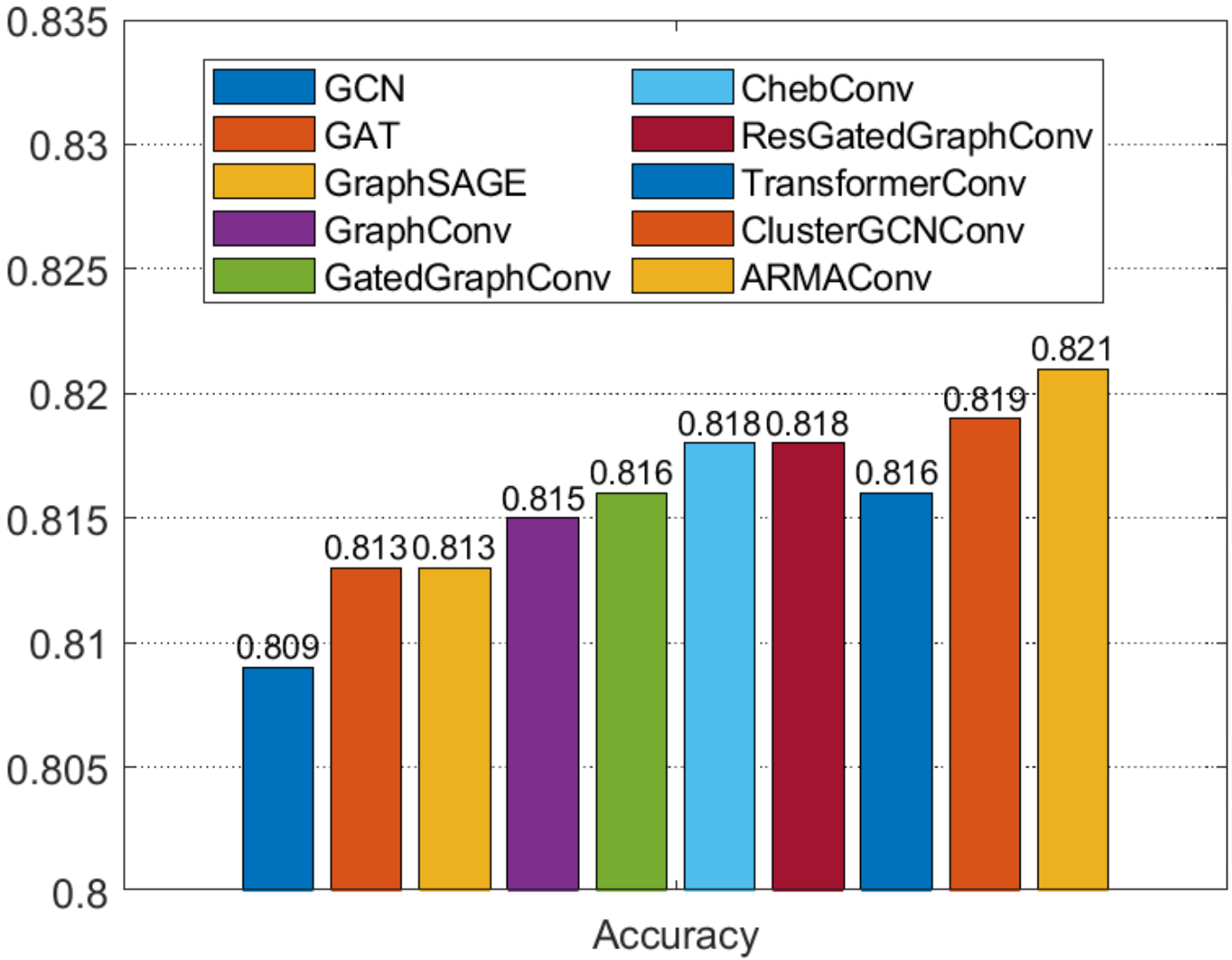
Results: Feature Engineering

Model performance with different set of user features.



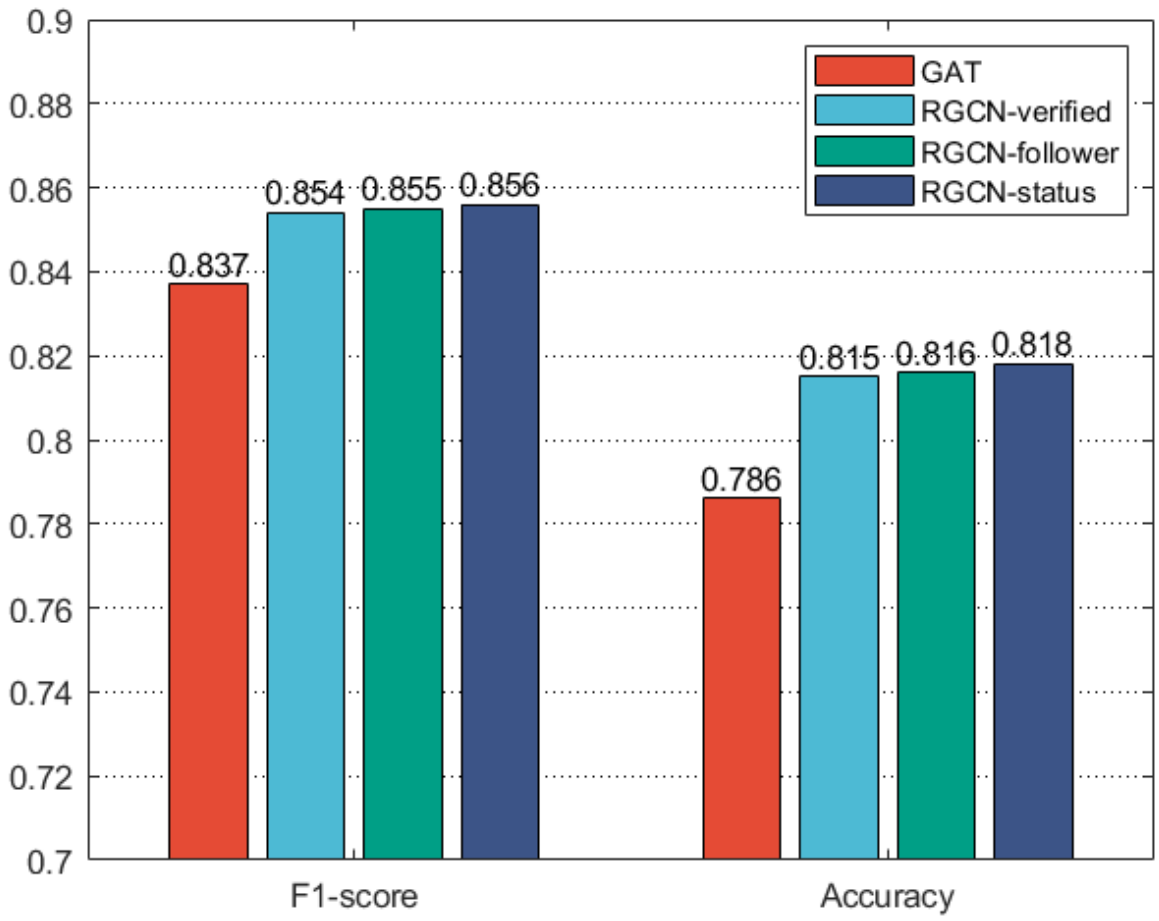
Results: Graph Operator

Model performance with different graph neural network architectures.



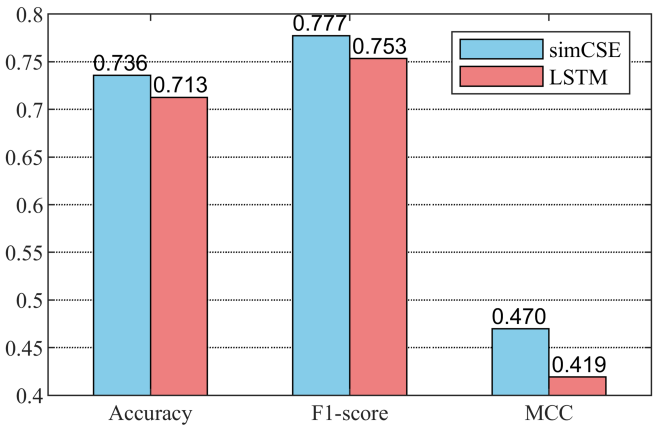
Results: Graph Heterogeneity

Model performance with different types of user relations.

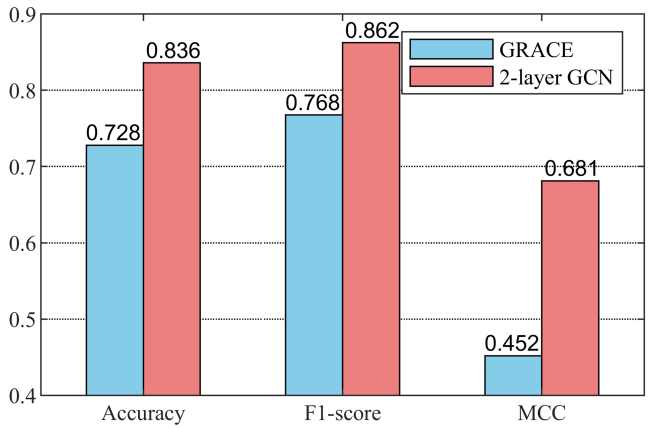


Results: Contrastive Learning

Model performance with contrastive learning on tweets and graphs.



contrastive learning on tweets



contrastive learning on graphs

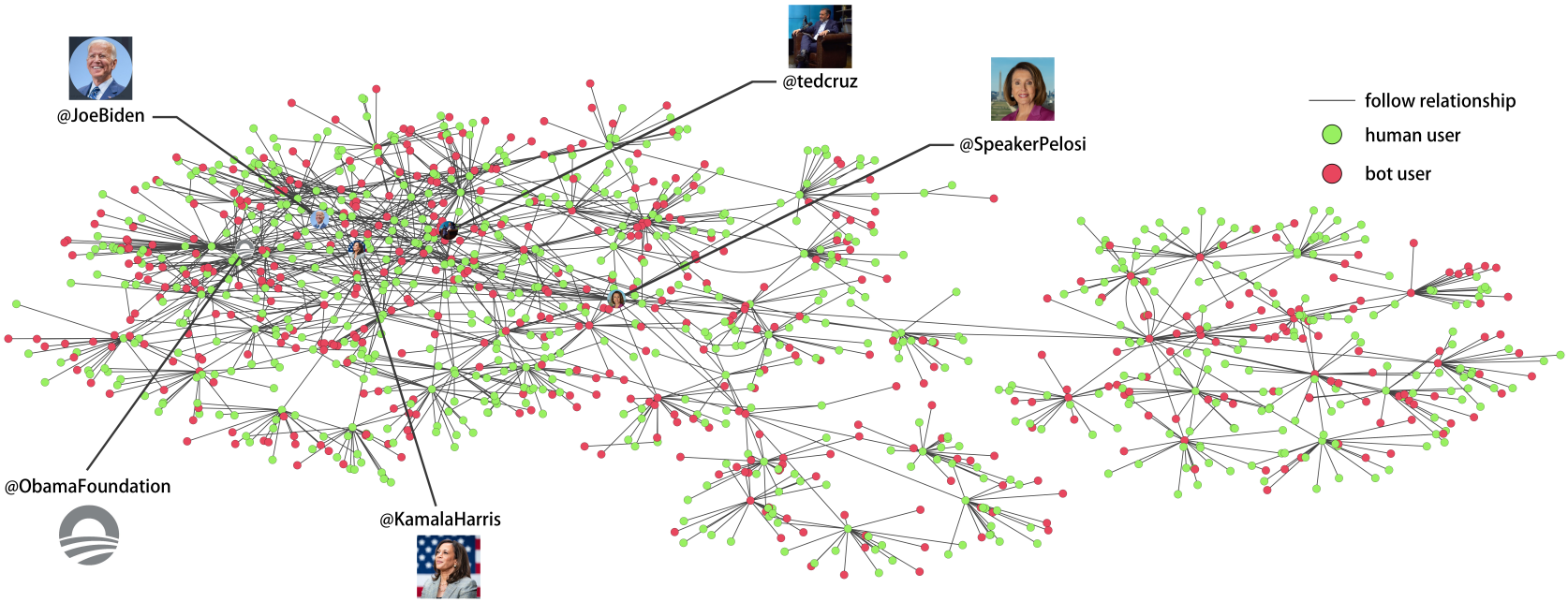
Findings

We conduct data visualization to examine experiment results and identify typical Twitter bot characteristics. Specifically, we present:

- **Graph structure** of the TwiBot-20 dataset.
- **Word cloud** of Twitter bots and genuine users.
- **User representation** vectors and their visualization.
- **User feature distribution** on the TwiBot-20 dataset.
- Typical Twitter **bot characteristics**.
- **Case study** of specific Twitter users and bots.

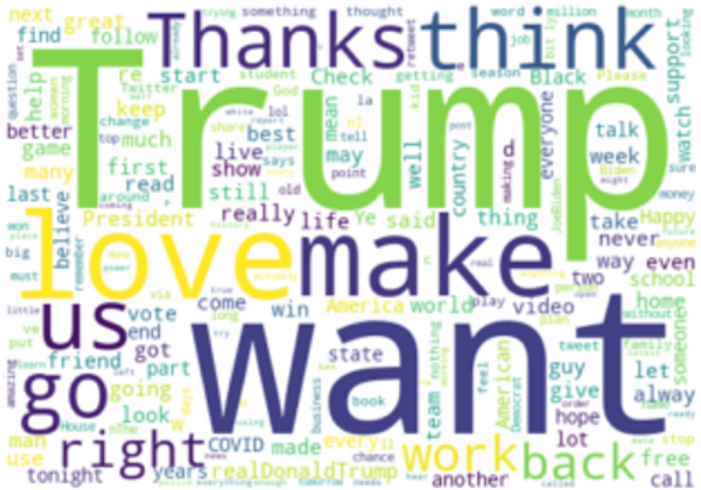
Graph Structure

We propose to leverage the Twittersphere structure with graph neural networks. We visualize the graph of dataset TwiBot-20:



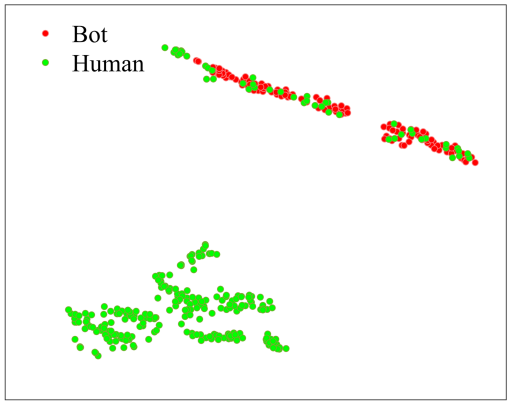
Word Cloud

We collect Twitter users' tweets and visualize how often certain words appear in tweets of Twitter bots and genuine users.

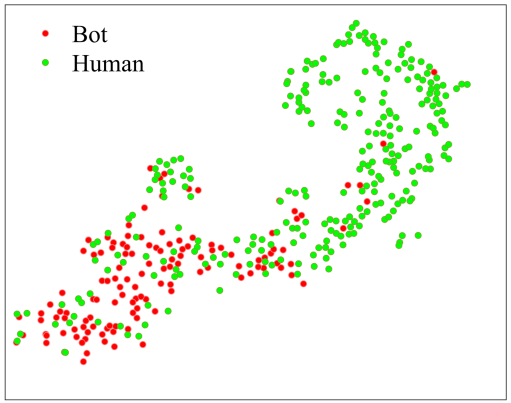


User Representation

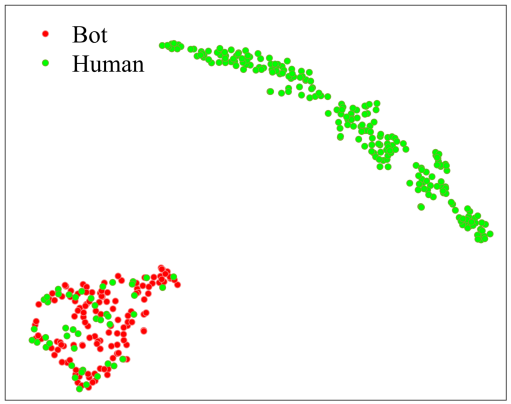
Graph neural networks turn nodes into representations. We use t-sne to visualize user representations under different settings.



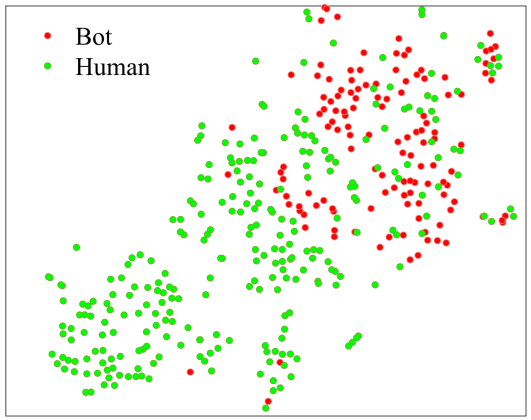
GAT (Homogeneity Score: 5.873×10^{-1})



simCSE (Homogeneity Score: 2.071×10^{-1})



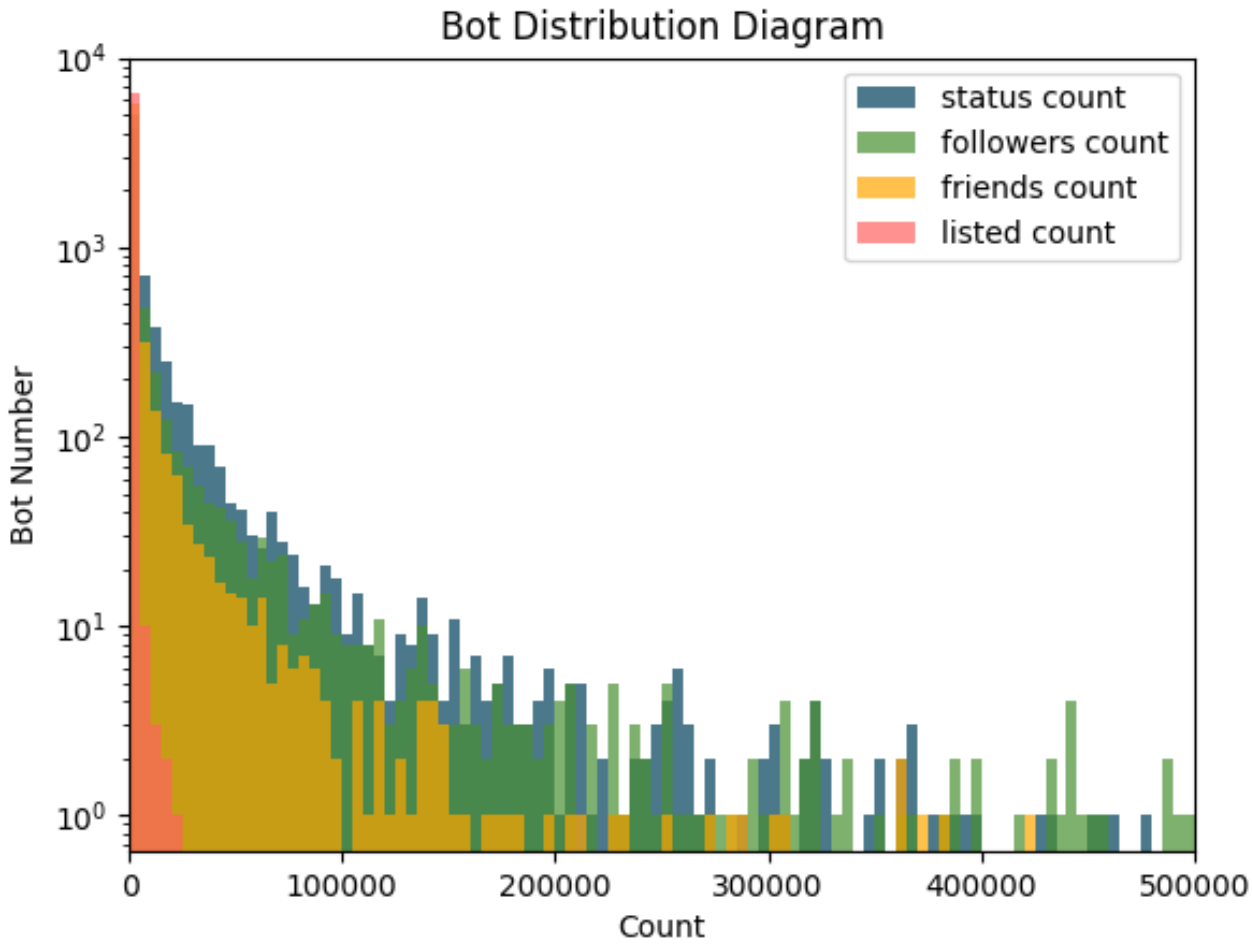
ClusterGCN (Homogeneity Score: 5.773×10^{-1})



Yang et al. (Homogeneity Score: 5.570×10^{-2})

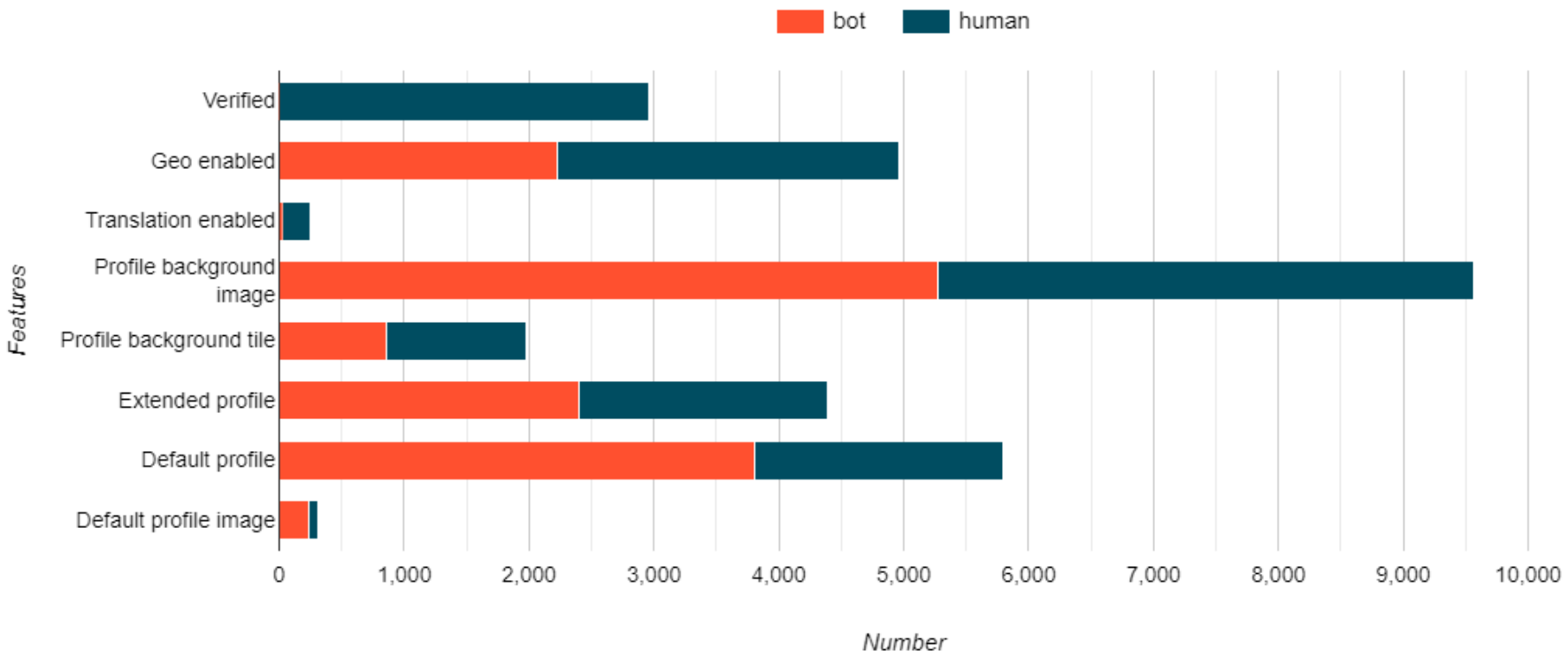
User Feature Distribution

We visualize the distribution of Twitter user numerical features.



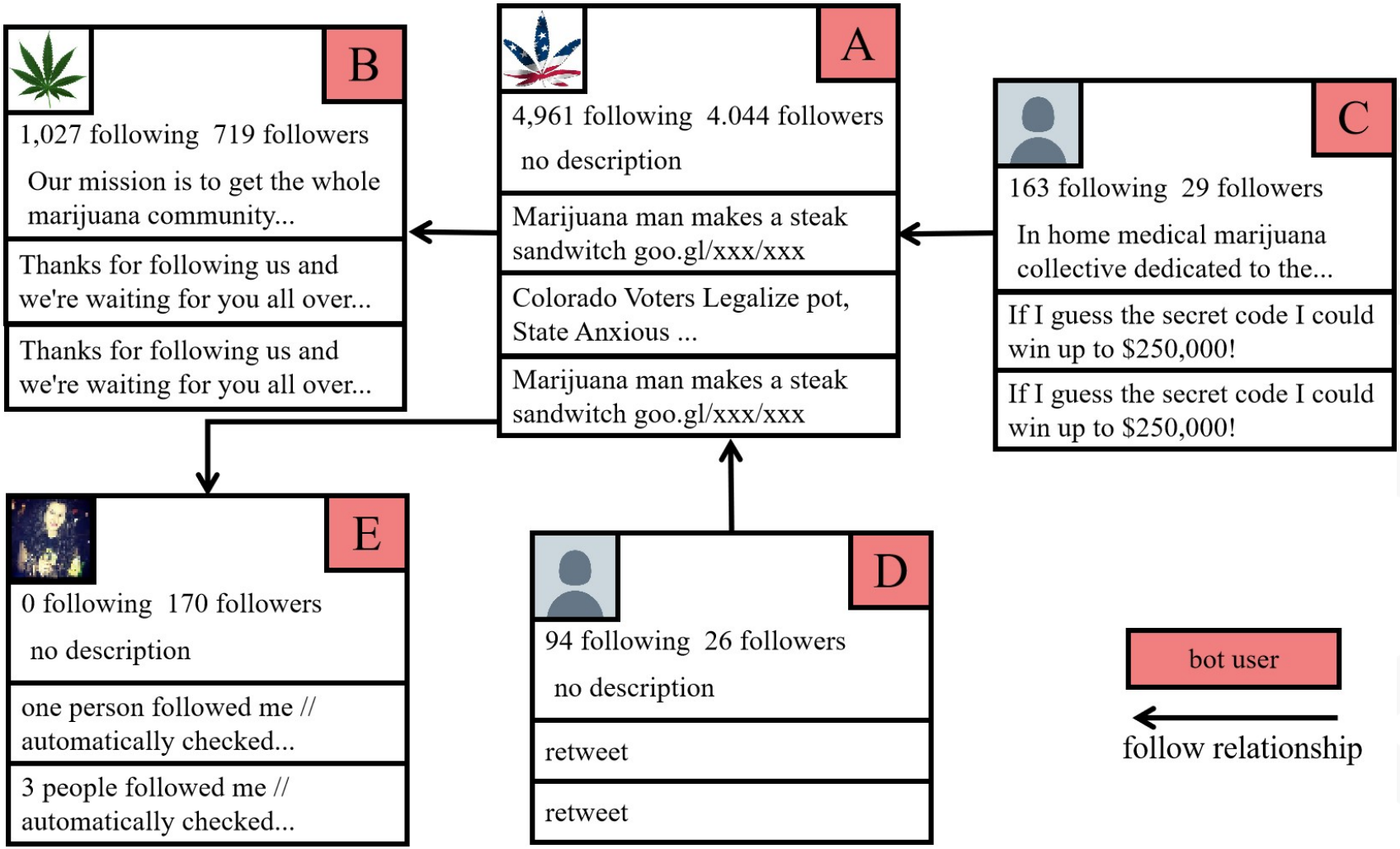
Bot Characteristics

We visualize the differences in user features between genuine Twitter users and Twitter bots.



Case Study

We select a bot subnetwork of five Twitter bots and visualize them.



Summary

- We **study Twitter bot detection** and previous works.
- We **learn about edge-cutting deep learning techniques** such as graph neural networks and contrastive learning.
- We propose to **leverage the graph structure** of Twitter for bot detection. Specifically, we study the problem from four perspectives: feature engineering, graph operator, graph heterogeneity and contrastive learning.
- We **supplement the TwiBot-20 dataset** with user tweets and properties and conduct extensive **experiments**.
- We conduct **data visualization** on the TwiBot-20 dataset and present our **findings** about Twitter bot characteristics.

Future Work

- Finalize our Twitter bot detection models and propose a comprehensive and effective solution.
- Build a real-world Twitter bot detection demo to help preserve the integrity of the online discourse.
- Extend our experiments, results and Twitter bot detection demo to a deep learning research paper.

Twitter Bot Detection

SWS3023 Project Showcase

Shangbin Feng, Herun Wan, Qingyue Zhang, Zhaoxuan Tan

School of Electronic and Information Engineering

Xi'an Jiaotong University

July 3, 2022