Heterogeneity-aware Twitter Bot Detection with Relational Graph Transformers

Shangbin Feng

LUD Lab, Xi'an Jiaotong University

wind_binteng@stu.xjtu.edu.cn

July 3, 2022

Table of Contents

Twitter Bot Detection

Definition

Twitter bots are Twitter users operated by automated programs.

- We detect them because they pursue **malicious goals**:

Election interfernece

- "Twitter Bots involve in the elections in the United States and Europe." WWW'19
- "Are 'bots' manipulating the 2020 conversation? Here's what's changed since 2016." The Washington Post.

Extreme ideology

- "Researchers: Nearly Half Of Accounts Tweeting About Coronavirus Are Likely Bots." NPR.
- Berger et al., "The Brookings project on US relations with the Islamic world."

Task Challenges

Generalization [CIKM 2021]

- Different kinds of Twitter Bots

Adaptation [CIKM 2021]

- The evolution of Twitter Bots

Community [ASONAM 2021]

- Bots attack in groups and seem genuine individually

Disguise [ASONAM 2021]

- Bots imitate the behavior and profile of genuine users

Related Work

Phase 1: feature engineering

User profile [ICWSM 2011], timeline [IEEE Intel. Systems 2016], URL redirection [TDSC 2013], mentioned websites [S&P 2011], efficient features [AAAI 2020], . . .

Phase 2: deep learning

RNN and word embeddings [TPS-ISA 2019], combine RNNs with features [Information Sciences 2018], GAN [IJCAI 2019], representation learning [CIKM 2021], . . .

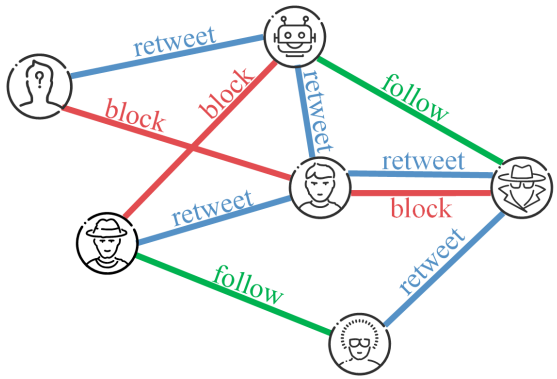
Phase 3: graph mining

GCN [WWW 2019], GraphHist [AAAI 2020], relational GNNs [ASONAM 2021], . . .

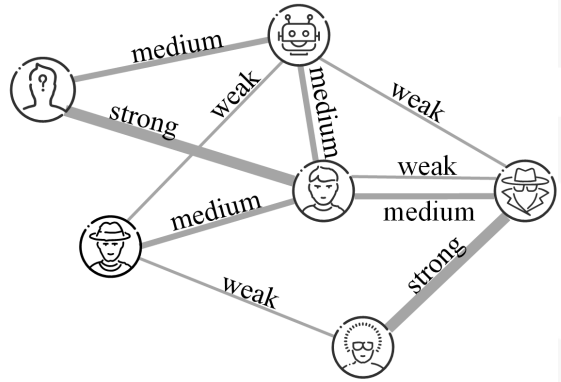
However

Previous graph-based approaches fail to leverage heterogeneities of online social networks!

- Relation heterogeneity
- Influence heterogeneity



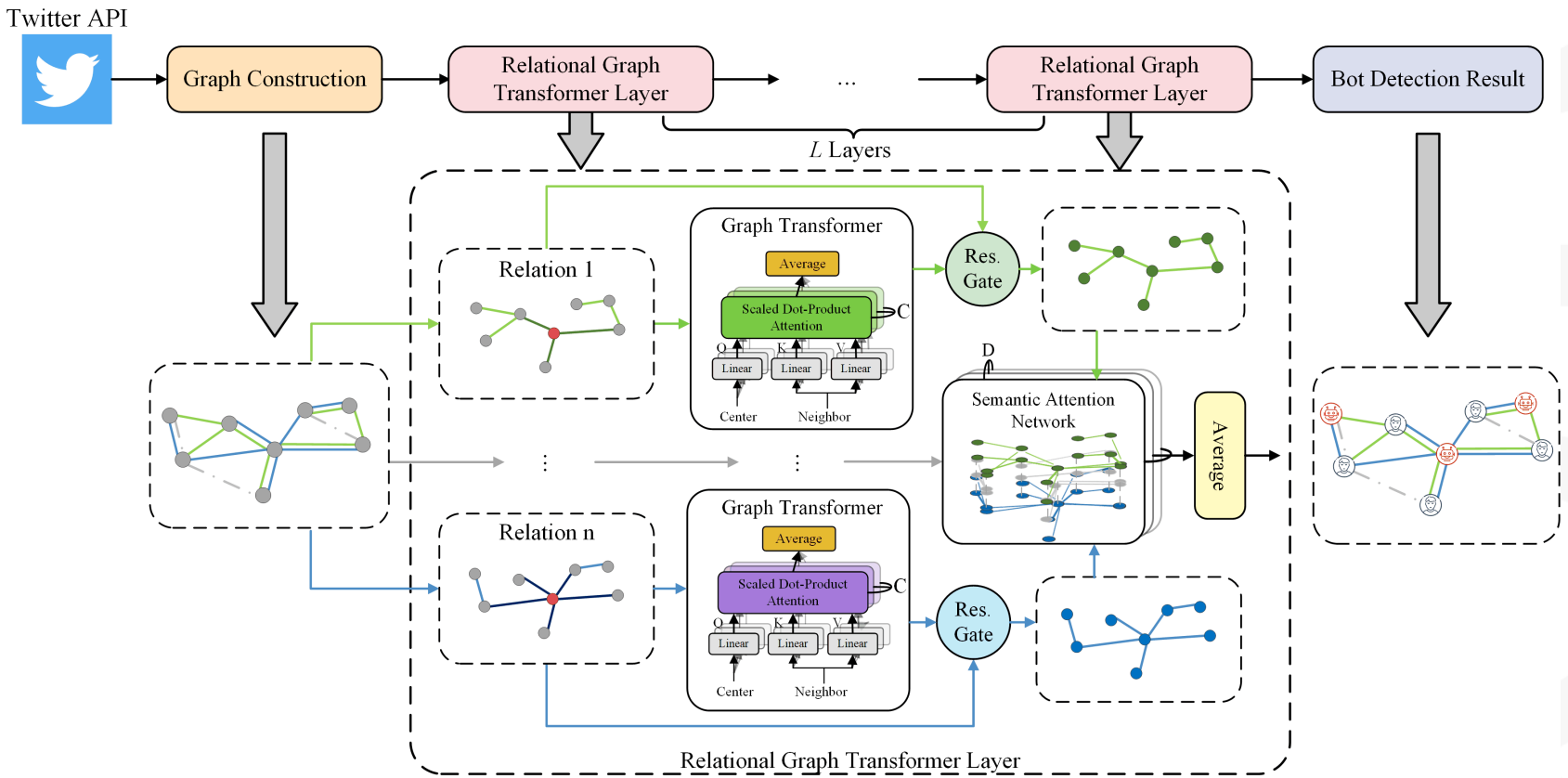
Relation Heterogeneity



Influence Heterogeneity

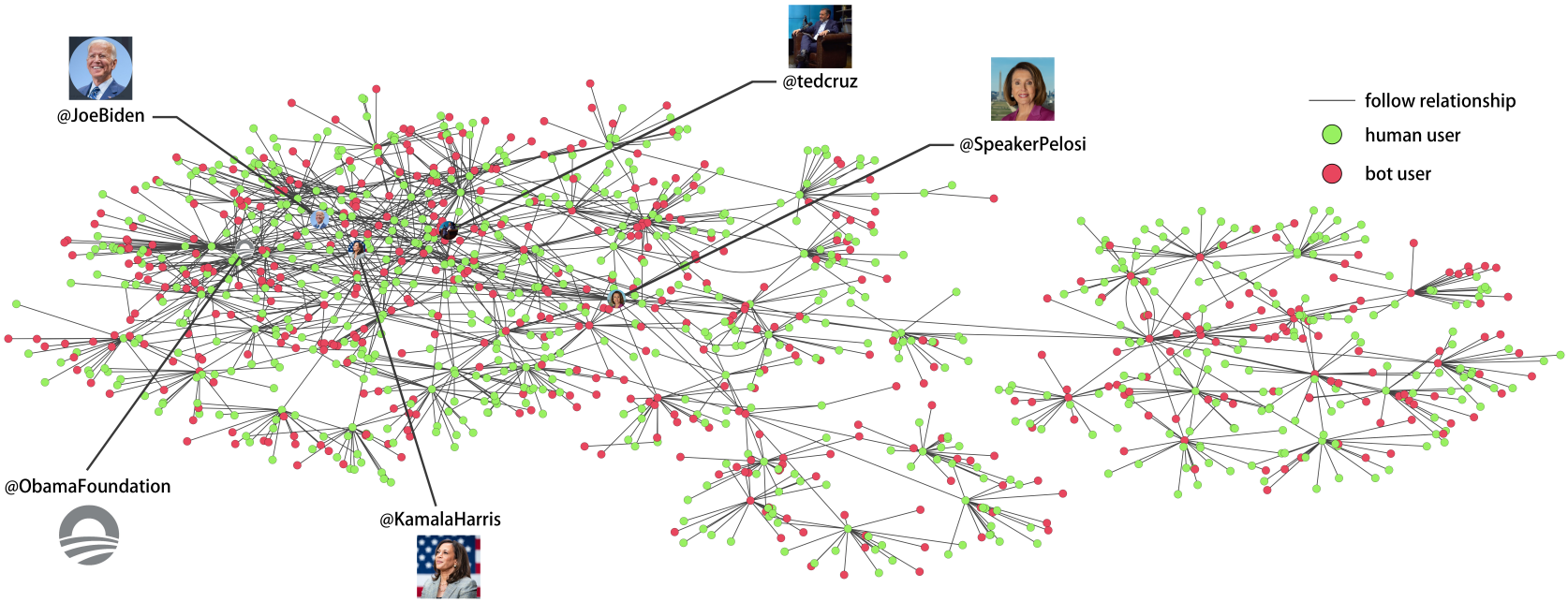
- Construction a **HIN** to represent Twitter
- Propose **relation graph transformers**

Overview



- Relational Graph Transformers
- Semantic Attention Networks

Graph Construction



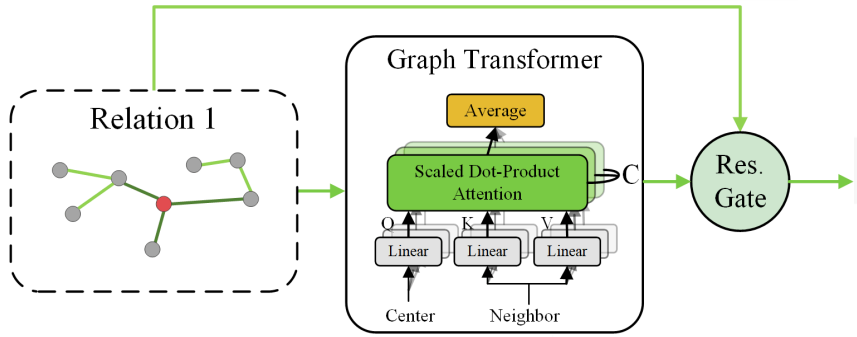
Relational Graph Transformers

- Relational Graph Transformers

$$q_{c,i}^{r}{}^{(l)} = W_{c,q}^{r}{}^{(l)} \cdot x_i^{(l-1)} + b_{c,q}^{r}{}^{(l)}$$
$$k_{c,j}^{r}{}^{(l)} = W_{c,k}^{r}{}^{(l)} \cdot x_j^{(l-1)} + b_{c,k}^{r}{}^{(l)}$$
$$v_{c,j}^{r}{}^{(l)} = W_{c,v}^{r}{}^{(l)} \cdot x_j^{(l-1)} + b_{c,v}^{r}{}^{(l)}$$

$$\alpha_{c,ij}^{r}{}^{(l)} = \frac{\langle q_{c,i}^{r}{}^{(l)} k_{c,j}^{r}{}^{(l)} \rangle}{\sum_{u \in N^r(i)} \langle q_{c,i}^{r}{}^{(l)}, k_{c,u}^{r}{}^{(l)} \rangle}$$

$$u_i^{r(l)} = \frac{1}{C} \sum_{c=1}^{C} \left[\sum_{j \in N^r(i)} \alpha_{c,ij}^{r}{}^{(l)} \cdot v_{c,j}^{r}{}^{(l)} \right]$$



- Residual Gate

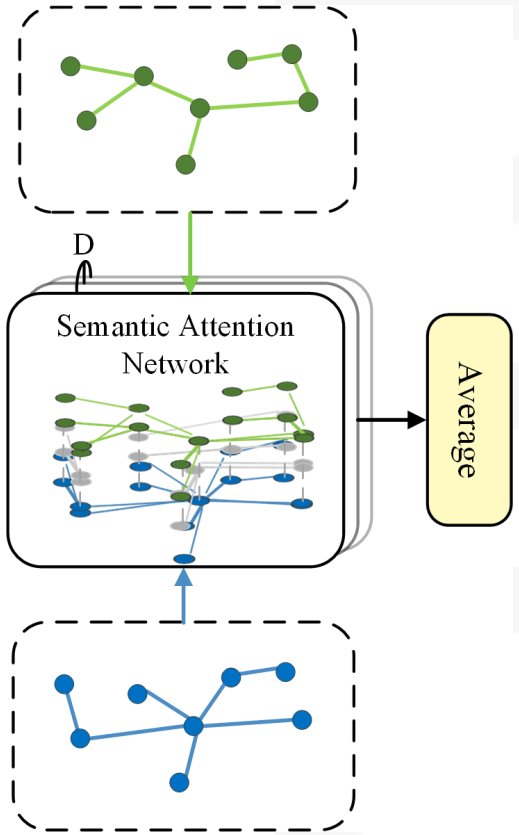
$$z_i^{r(l)} = sigmoid(W_A^r \cdot [u_i^{r(l)}, x_i^{(l)}] + b_A^r),$$
$$h_i^{r(l)} = tanh(u_i^{r(l)}) \odot z_i^{r(l)} + x_i^{r(l)} \odot (1 - z_i^{r(l)})$$

- Semantic Attention Networks

$$w_d^{r(l)} = \frac{1}{|V|} \sum_{i \in V} q_d^{(l)^T} \cdot tanh(W_{d,s}^{(l)} \cdot h_i^{r(l)} + b_{d,s}^{(l)})$$

$$\beta_d^{r(l)} = \frac{\exp(w_d^{r(l)})}{\sum_{k \in R} \exp(w_d^{k(l)})}$$

$$x_i^{(l)} = \frac{1}{D} \sum_{d=1}^{D} \left[\sum_{r \in R} \beta_d^{r(l)} \cdot h_i^{r(l)} \right]$$



- Learning and Optimization

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

$$Loss = -\sum_{i \in Y} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] + \lambda \sum_{w \in \theta} w^2$$

Experiment Settings

Dataset

- TwiBot-20 is the comprehensive benchmark. Feng *et al.*, CIKM 2021.

Baselines

- **Feature-based**
  Lee et al., Yang et al., Cresci et al., Miller et al., Botometer
- **Deep learning-based**
  Kudugunta et al., Wei et al., SATAR
- **Graph-based**
  Alhosseini et al., BotRGCN, Ours

SOTA Performance

Table 1: Characteristic and performance of different Twitter bot detection methods. Deep, interactive, representative, graph-based and heterogeneity-aware denotes whether the method involves deep learning, leverages user interactions, learns user representation, involves graph neural networks or leverages Twitter heterogeneity.

Method	Deep	Interactive	Representative	Graph-based	Heterogeneity-aware	Accuracy	F1-score
Lee *et al.*						0.7456	0.7823
Yang *et al.*						0.8191	0.8546
Cresci *et al.*						0.4793	0.1072
Kudugunta *et al.*	✓					0.8174	0.7515
Wei *et al.*	✓					0.7126	0.7533
Miller *et al.*		✓				0.4801	0.6266
Botometer		✓				0.5584	0.4892
SATAR	✓	✓	✓			0.8412	0.8642
Alhosseini *et al.*	✓	✓	✓	✓		0.6813	0.7318
BotRGCN	✓	✓	✓	✓		0.8462	0.8707
Ours	✓	✓	✓	✓	✓	**0.8664**	**0.8821**

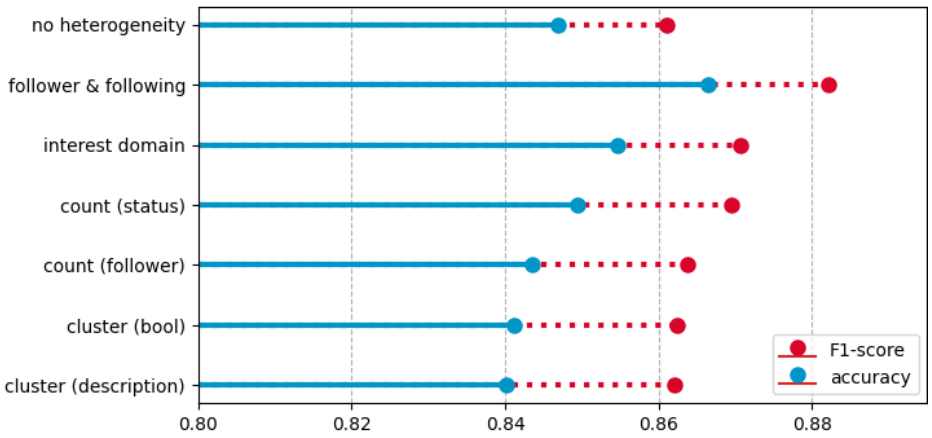
Relational Graph Transformers

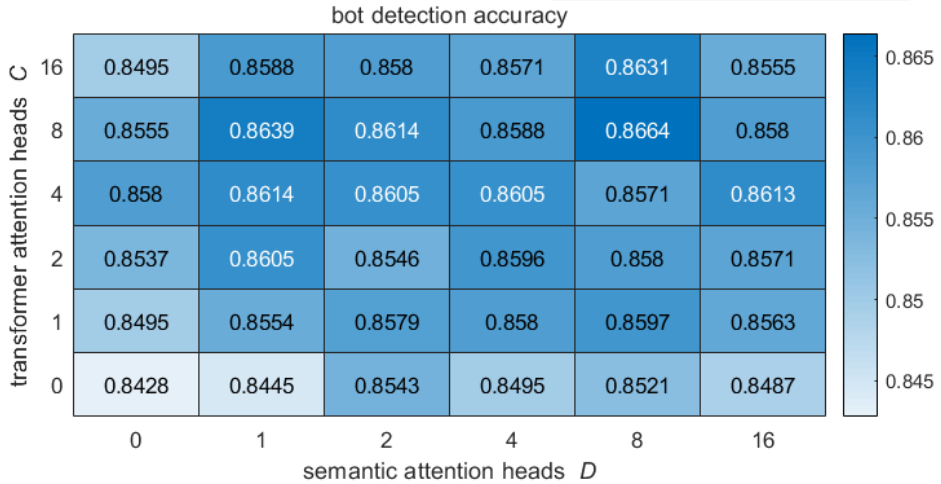
- We proposed them, are they good?

Table 3: Ablation study of our proposed GNN architecture. RT and SA denote relational transformers and semantic attention networks respectively.

Ablation Settings	Accuracy	F1-score
full model	**0.8664**	**0.8821**
remove transformer in RT	0.8521	0.8679
remove gated residual in RT	0.8478	0.8646
replace RT with GAT	0.8571	0.8726
replace RT with GCN	0.8444	0.8619
replace RT with SAGE	0.8546	0.8687
summation as SA	0.8512	0.8654
mean pooling as SA	0.8512	0.8663
max pooling as SA	0.8495	0.8629
min pooling as SA	0.8555	0.8704

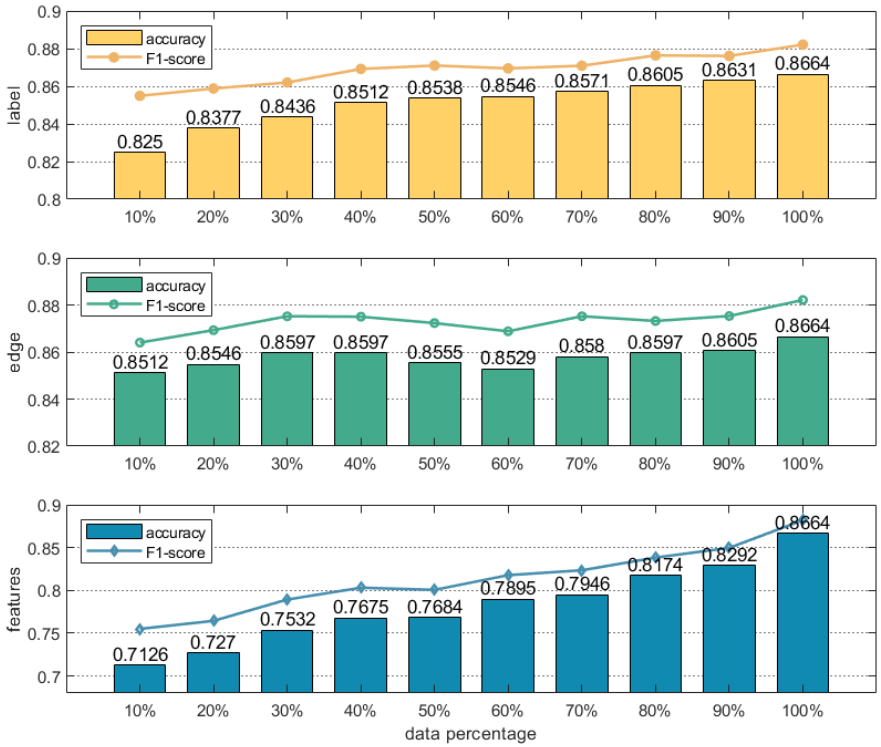
Heterogeneity

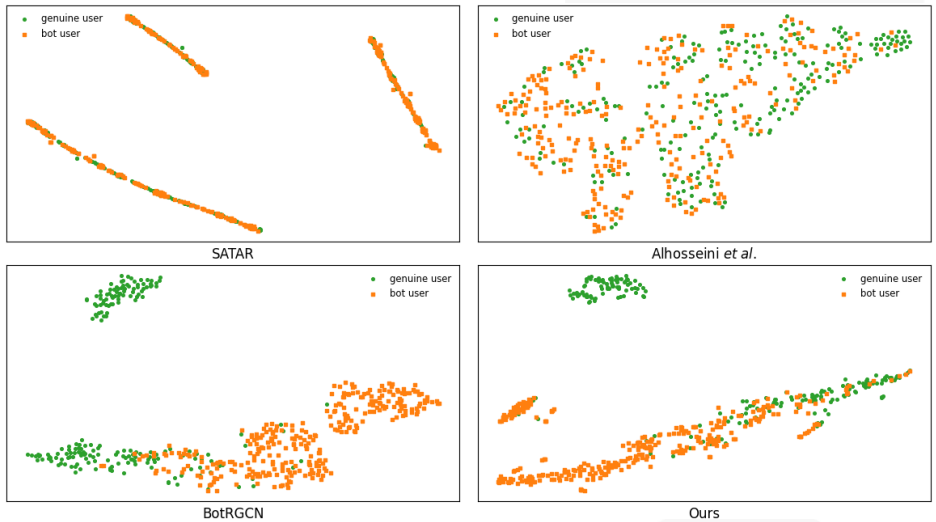




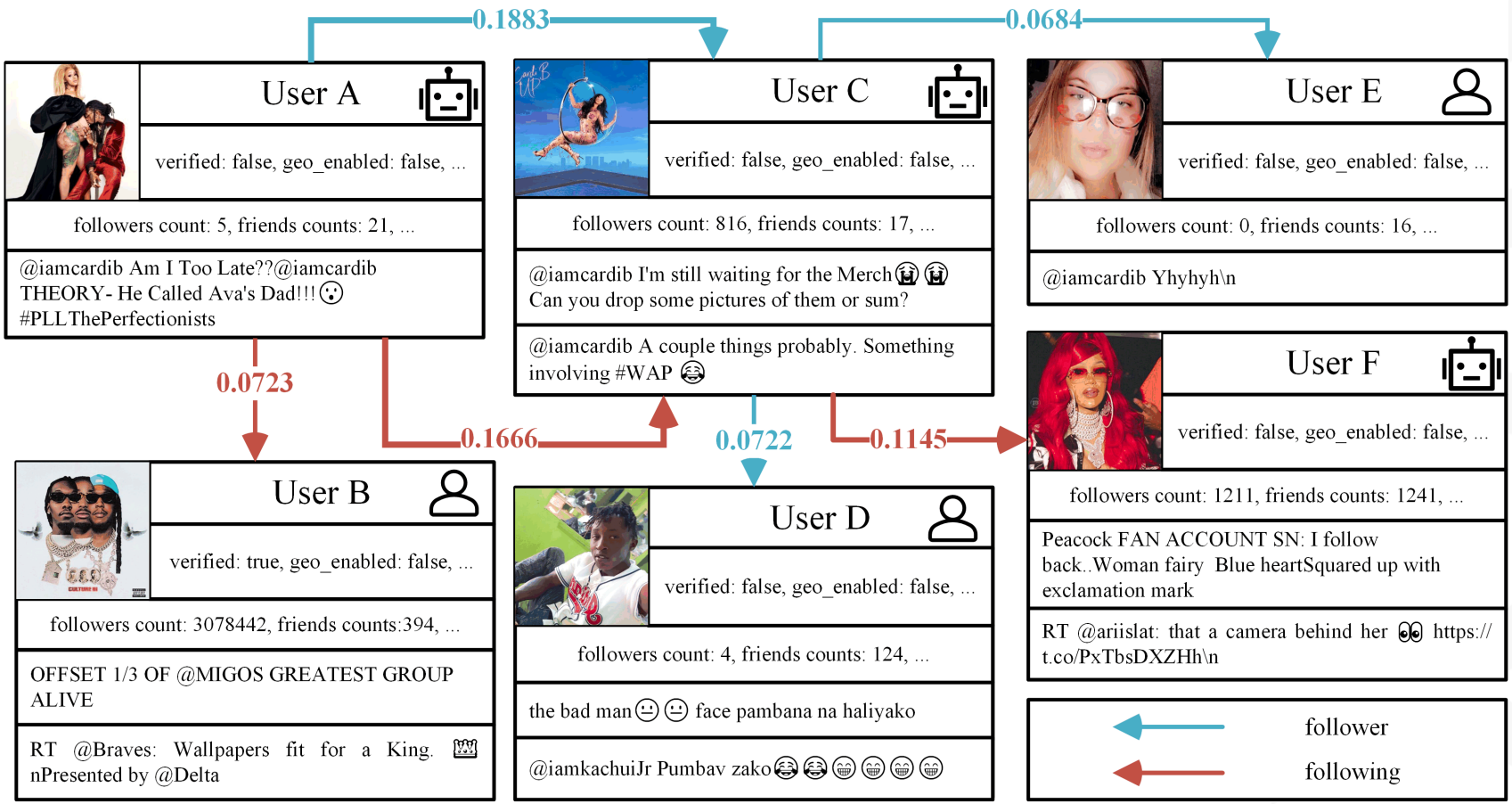
Besides

- Our approach is **data-efficient** and **learns good representations**!





Case Study


0.1883
0.0684
User A
verified: false, geo_enabled: false, ...
followers count: 5, friends counts: 21, ...
@iamcardib Am I Too Late??@iamcardib THEORY- He Called Ava's Dad!!!😮 #PLLThePerfectionists
User C
verified: false, geo_enabled: false, ...
followers count: 816, friends counts: 17, ...
@iamcardib I'm still waiting for the Merch😭😭 Can you drop some pictures of them or sum?
@iamcardib A couple things probably. Something involving #WAP 😂
User E
verified: false, geo_enabled: false, ...
followers count: 0, friends counts: 16, ...
@iamcardib Yhyhyh\n
0.0723
0.1666
0.0722
0.1145
User B
verified: true, geo_enabled: false, ...
followers count: 3078442, friends counts:394, ...
OFFSET 1/3 OF @MIGOS GREATEST GROUP ALIVE
RT @Braves: Wallpapers fit for a King. nPresented by @Delta
User D
verified: false, geo_enabled: false, ...
followers count: 4, friends counts: 124, ...
the bad man😐😐 face pambana na haliyako
@iamkachuiJr Pumbav zako😂😂😁😁😁😁
User F
verified: false, geo_enabled: false, ...
followers count: 1211, friends counts: 1241, ...
Peacock FAN ACCOUNT SN: I follow back..Woman fairy Blue heartSquared up with exclamation mark
RT @ariislat: that a camera behind her 👀 https://t.co/PxTbsDXZHh\n
follower
following

Conclusion

In this paper, we

- leverage relation and influence heterogeneities for HIN-based Twitter bot detection
- propose relational graph transformers to model these two heterogeneities
- achieve state-of-the-art performance on TwiBot-20 and bear out the necessity of analyzing Twitter heterogeneity

Resources

We make the code and model of Relational Graph Transformers available at

- `https://github.com/BunsenFeng/BotHeterogeneity`

For the datasets we used to train and test Relational Graph Transformers:

- TwiBot-20: `https://github.com/BunsenFeng/TwiBot-20`

Thank You !

Shangbin Feng
LUD Lab, Xi'an Jiaotong University
wind_binteng@stu.xjtu.edu.cn
July 3, 2022