

# LUD Project 3

2022/3/5

LUD lab

Herun Wan

# #1 Introductory Lectures

- #0 python basics
- #1 machine learning basics
- #2 deep learning and pytorch basics
- Project 1: pytorch practice
- #3 NLP basics
- #4 GNN basics
- Project 2: CORA node classification
- #5 NLP advanced
- #6 GNN advanced
- Project 3: BotRGCN Reproduction
- #7 Knowledge Graph
- #8 Computer Vision
- #9 Miscellaneous
- Project 4: Final Project



# So far ...

*Lec -2: Deep learning*

MLPs, model, loss, optimize

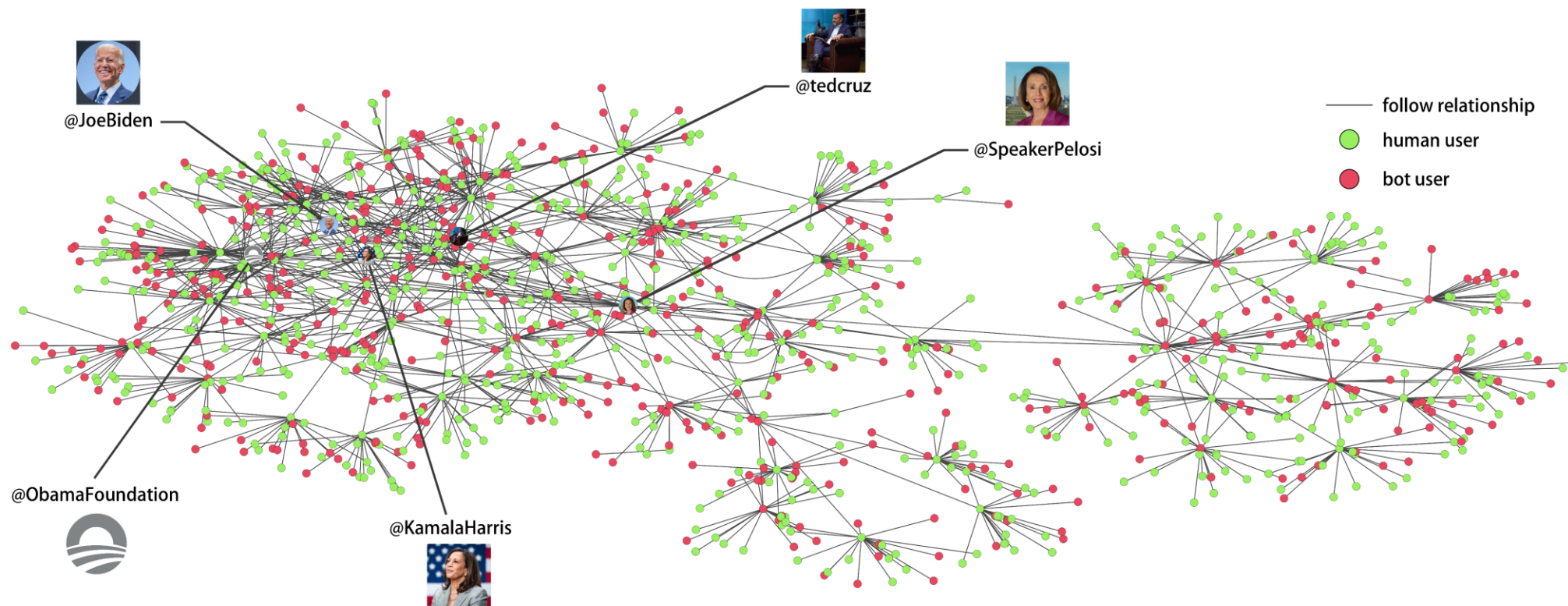
*Lec -3 & 5: NLP basic & advanced*

BERT, RoBERTa

*Lec -4 & 6: GNN basic & advanced*

GCN, R-GCN

# Project 3: BotRGCN Reproduction (on TwiBot-20)



heterogeneous information graph

229,573 users, 8,723,736 properties, 33,488,192 tweets

11,826 users with label

download: by fsb or tzx

# Project 3: BotRGCN Reproduction (on TwiBot-20)

TwiBot20:

node.json

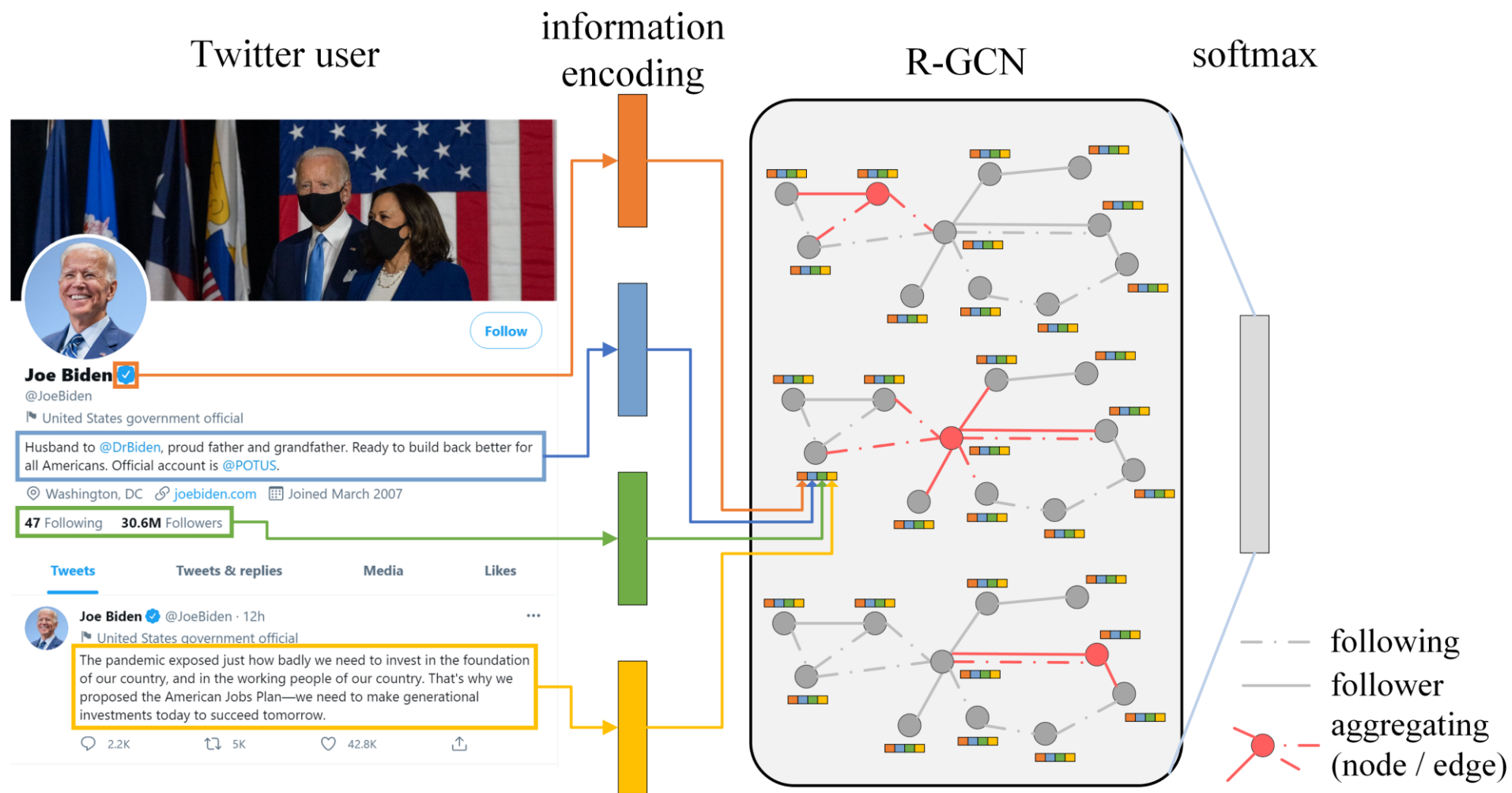
edge.csv

label.csv

split.csv

# Project 3: BotRGCN Reproduction (on TwiBot-20)

BotRGCN(acm doi/10.1145/3487351.3488336)



# Project 3: BotRGCN Reproduction (on TwiBot-20)

Feature set:

- description
  - RoBERTa encoding
- tweets
  - RoBERTa encoding
- numerical property
  - z-score
- categorical property
  - one-hot

TABLE I  
NUMERICAL USER PROPERTIES ADOPTED IN BOTRGCN.

Feature Name	Description
#followers	number of followers
#followings	number of followings
#favorites	number of likes
#statuses	number of statuses
active_days	number of active days
screen_name_length	screen name character count

TABLE II  
CATEGORICAL USER PROPERTIES ADOPTED IN BOTRGCN.

Feature Name	Description
protected	protected or not
geo_enabled	enable geo-location or not
verified	verified or not
contributors_enabled	enable contributors or not
is_translator	translator or not
is_translation_enabled	translation or not
profile_background_tile	the background tile
profile_user_background_image	have background image or not
has_extended_profile	have extended profile or not
default_profile	the default profile
default_profile_image	the default profile image

## Project 3: BotRGCN Reproduction (on TwiBot-20)

MLPs, R-GCN, Softmax and optimizer

- $d_f \rightarrow D / 4 \rightarrow D \rightarrow \text{R-GCN} \rightarrow \text{Softmax}$
- CE loss



# Tasks

1. Reproduce BotRGCN by yourself(include the dataset, model, and metrics)
2. Calculate the accuracy, f1-score and MCC(acc over 70)
3. Ablation study(feature set, r-gcn layers, gnn)

# Report

1. Your results
2. Your discovery
3. ...

# THX

2022/3/5

LUD lab

Herun Wan