Graph Neural Networks Learn Twitter Bot Behaviour

Albert M. Orozco Camacho alorozco53@mila.quebec
Sacha Lévy sacha.levy@mail.mcgill.ca
Reihaneh Rabbany reihaneh.rabbany@mila.quebec

School of Computer Science, McGill University
Mila - Québec AI Institute

Abstract

- Social media trends are increasingly taking a significant role for the understanding of modern social dynamics. Platforms like *Twitter* enable democratic interaction within users, thus, providing an opportunity to inject influential ideas within formed communities.
- In this work, we take a look at how the Twitter landscape gets constantly shaped by automatically generated content.
- *Twitter bot* activity can be traced via network abstractions which, we hypothesize, can be learned through state-of-the-art graph neural network techniques.
- We employ a large *bot* database, continuously updated by *Twitter*, to learn how likely is that a user is mentioned by a bot, as well as, for a hashtag. Furthermore, we model this likelihood as a link prediction task between the set of users and hashtags.
- Moreover, we contrast our results by performing similar experiments on a crawled data set of real users.

Data Set Generalities

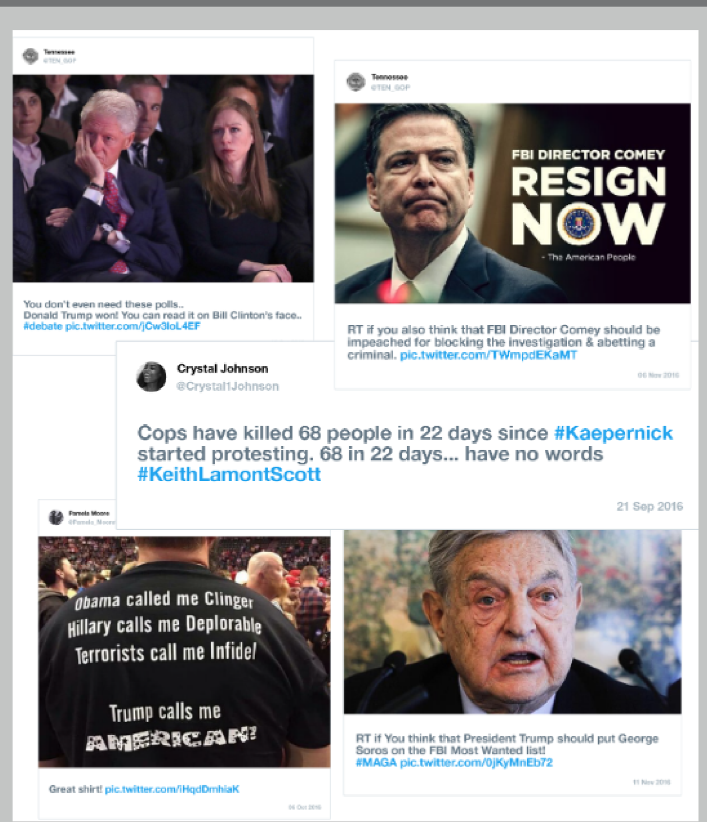

Figure 1: Sample tweets posted by accounts from the Twitter Election Integrity data set. Taken from https://about.twitter.com/en_us/values/elections-integrity.html#data

- Tracking malicious activity in Twitter is a major problem nowadays, in particular, to democracy [2].
- Recently, Twitter Operations has been releasing a huge sample of state-backed bots in order to foster research within the community (Figure 1). We refer to this set of accounts as the **Twitter Election Integrity** (TEI) dataset
- Table 1 summarizes the data set statistics. We also make use of a set of **REAL** users that we have crawled during the covid-19 pandemic, for comparison purposes.

Data Set Statistics

Table 1: General Data set Statistics

	Data set	
Statistic	TEI	REAL
Total number of hashtags	$804K$	$36K$
Total number of mentioned users	$2.56M$	$171K$
Total number of fake accounts	$22K$	–
Total number of real accounts	–	$67K$

Network Modeling

- To encode each data set's information, we extract directed graphs. We propose two ways of reconstructing a large-scale social network from the provided data.
- Our **Mention Graph** links every bot on TEI, or every existing user in REAL, to their mentioned users.
- On the other hand, we define a **Hashtag Graph** to include hashtags as nodes, rather than attributes. Figure 2 shows visually how both networks would look like.

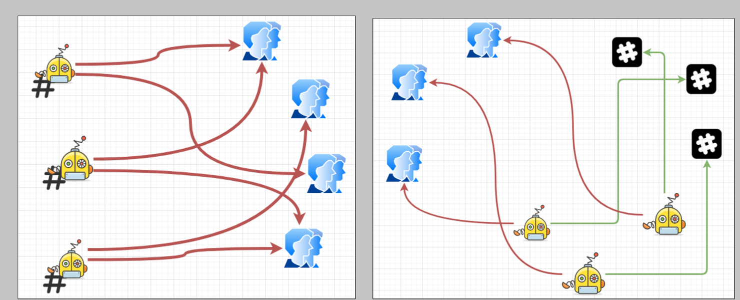

Figure 2: On the left, a visual representation for our **Mention Graph**: bots, along with hashtag node attributes, are matched with their mentioned users. On the right, a similar explanation of how we structure our **Mention-Hashtag Graph**: bots are connected to both other users and hashtags.

Methodology

- First, we sample a set of tweets from a predefined interval. Note that this process can be performed either on both TEI and REAL datasets.
- We experiment using node2vec embeddings [3] as a way to characterize each node with the help of its neighbours. We also approach this step by predefining labeling random walks, as introduced in [1] (metapath2vec).
- Finally, we use the SEAL framework [4] to learn a link prediction model. This methodology extracts a local subgraph per each target edge. Figure 3 depicts the overall neural processing pipeline.

Neural Pipeline

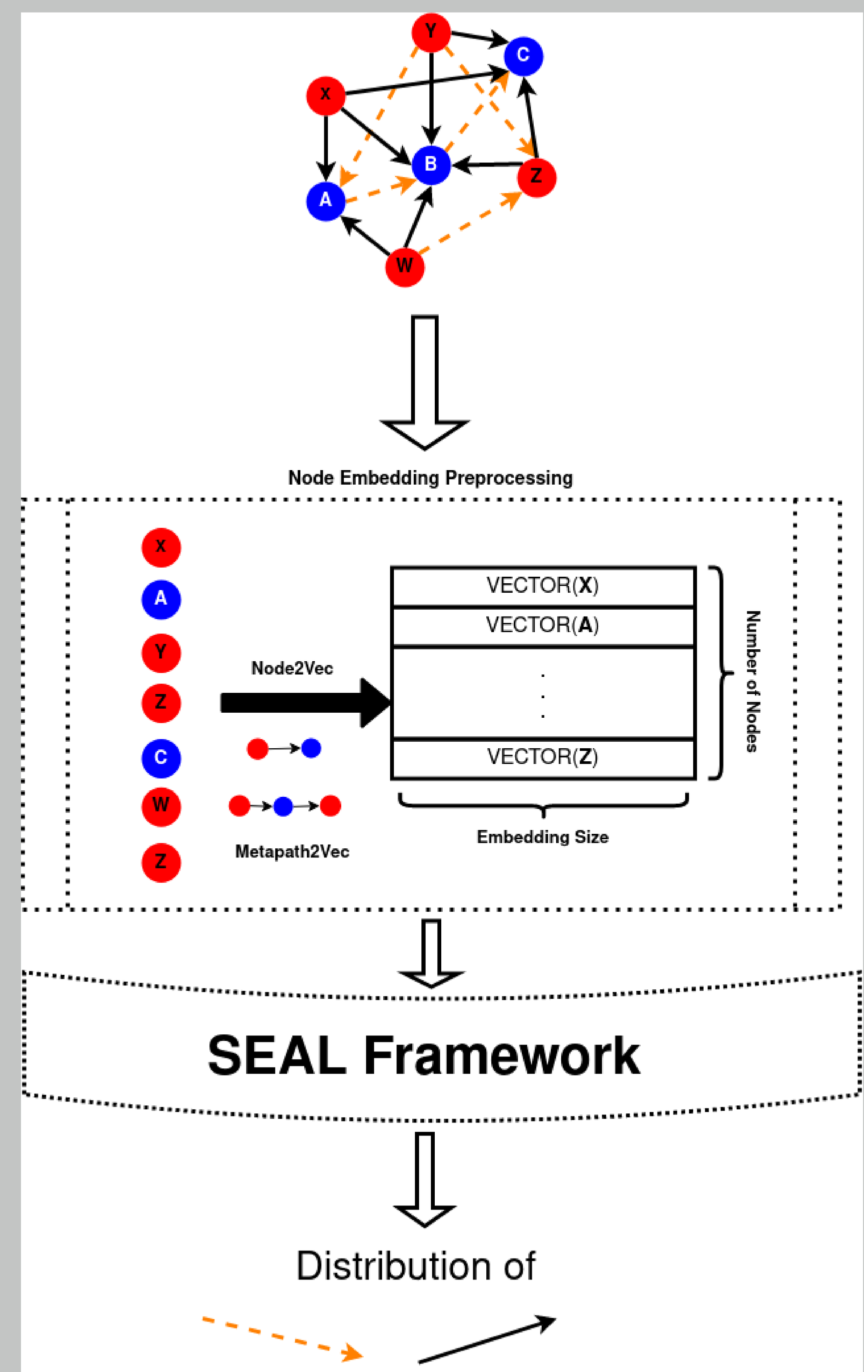

Figure 3: Our neural processing pipeline. Colour in nodes represents different types of nodes (e.g. users, hashtags, mentioned users); meanwhile, orange dashed links represent negative sampled links. For further information on how the SEAL framework looks like, refer to Figure 1 of [4].

Ongoing Experiments

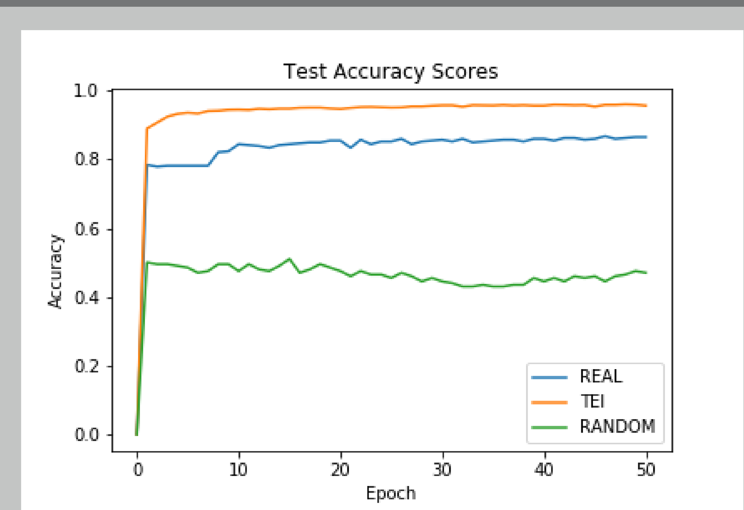

Figure 4: Test accuracy scores after each epoch run. We depict, for the sake of giving a comparison, three runs with a similar portion of training and testing data. We construct a random graph, whose set of nodes and links are uniformly sampled, to show our accuracy scores exceed the minimal baseline.

Discussion and Conclusions

- The current preliminary results (Figure 4) follow a trend in which any TEI learning accuracies exceed those from any REAL sample.
- Thus, we are capable of learning, in this case, which accounts will an *Internet bot* tend to follow, as well as the topics it is aiming, via a list of hashtags.
- These metrics show great potential on downstream applications; in particular, distinguishing coordinated activity in social media.
- Our results depict a constant pattern notwithstanding different samples, we argue for the use of continuous representations of data to capture time invariant features.

Future Work

- We conjecture that the proposed neural model's penultimate layer's link embeddings could provide a rich understanding on both language and behaviour of Internet actors.
- Despite our results, further research is still necessary to be performed for gaining a better insights on temporal changes of these data. For instance, all our training procedures have been assumed to be fixed over time.

References

[1] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami.
metapath2vec: Scalable representation learning for heterogeneous networks.
In *KDD '17*, pages 135–144. ACM, 2017.

[2] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini.
The rise of social bots.
Commun. ACM, 59(7):96–104, June 2016.

[3] Aditya Grover and Jure Leskovec.
node2vec: Scalable feature learning for networks.
CoRR, abs/1607.00653, 2016.

[4] Muhan Zhang and Yixin Chen.
Link prediction based on graph neural networks.
CoRR, abs/1802.09691, 2018.