
Graph Neural Networks Learn Twitter Bot Behaviour

Albert M. Orozco Camacho

School of Computer Science, McGill University
Mila - Québec AI Institute
alorozco53@mila.quebec

Sacha Lévy

School of Computer Science, McGill University
Mila - Québec AI Institute
sacha.levy@mail.mcgill.ca

Reihaneh Rabbany

School of Computer Science, McGill University
Mila - Québec AI Institute
reihaneh.rabbany@mila.quebec

Abstract

Social media trends are increasingly taking a significant role for the understanding of modern social dynamics. 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. Thus, 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.

1 Introduction

The use of automatic tweet spreaders (**social bots**¹) remains a very common practice to rapidly alter a group of user’s opinion [2]. We approach this issue by modelling this behaviour in *user-follower-hashtag* networks, which we hypothesize contain encoded information capable to be learned by state-of-the-art *graph neural network techniques*. Therefore, we learn to *predict the links* within such networks as a means to understand bot activity (Section 2).

Recently, *Twitter* has decided to release a large database of accounts, whose origin is presumed to be backed-up by state operations. These data, that has been collected since 2016, include account information, as well as, tweets starting from 2011 all the way to 2020. Under the company’s *Transparency* efforts², this collection, referred to as the **Twitter Election Integrity (TEI)** data set, includes over 800K hashtags, 2M users; all of which comes from around 22K suspended fake accounts.

To contrast with TEI’s automatic user’s activity, we crawl 67K currently existing accounts that have made use over 36K hashtags and 171K mentions, mainly from the last months of 2019 and the

¹We call *social bots* to automatized accounts, usually by a computer program behind.

²<https://transparency.twitter.com/en/reports/information-operations.html>

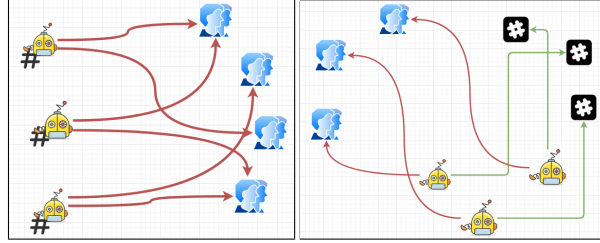


Figure 1: On the left, a visual representation for our **Mention Graph**: bots, along with hashtag node attributes, are matched with their mentioned users. On the right, a similar explanation of how we structure our **Mention-Hashtag Graph**: bots are connected to both other users and hashtags.

beginning of 2020. Our purposes with this **REAL** data set is to give a comparison between the probability of who a *real account* would most likely follow and which hashtags would be used. To ensure this data contains real accounts, we shortlist those that are verified by Twitter; yet also include and ensure all accounts are running during experiment time.

To encode each data set’s information, we extract directed graphs. Firstly, given a *user* u (which can either be a bot or a real one) we link it to every other *mentioned user* v . Note that v need not to be a bot or a real user (in fact, the TEI data set’s set of mention users include real accounts, as well). Furthermore, we compute node attributes per each v , based on the utilized hashtags. We refer to such network type as our **Mention Graph**. On the other hand, we define a **Hashtag Graph** to include hashtags as nodes, rather than attributes. Figure 1 shows visually how both networks would look like.

2 Methodology

Our proposed approach requires, as inputs, embedding both, the node and link sets of any given network, into vector representations. For starters, 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*), in order to preserve node context in heterogeneous graphs.

Next, we use the *SEAL* framework [5] to learn a link prediction model. This methodology extracts a local subgraph per each target edge. A graph neural network (GNN) is used to learn the structure (and possibly, a set of heuristics) that best characterize each edge.

Because the large size of hashtags and users makes it not tractable to build a single graph out of both data sets, we perform a series of experiments during various time frames. Thus, we firstly extract all the data for a defined time frame, to then build a desired graph, and finally process it as mentioned previously. Moreover, the train and test split is defined over the list of all possible links for the constructed network.

3 Discussion and Conclusions

The current preliminary results follow a trend in which any TEI learning accuracies exceed those from any REAL sample. Although our research experiments are still ongoing, these metrics show great potential on downstream applications; in particular, distinguishing coordinated activity in social media. Finally, since our results depict a constant pattern notwithstanding different samples, we argue for the use of continuous representations of data to capture time invariant features.

For future work, we are verifying empirically that our numerical results reflect on the neural pipeline’s learned representations. We conjecture that this line of research promises a rich understanding on both language and behaviour of Internet actors. Such level of comprehension would be determined by how an input network is defined. Furthermore, all our training procedures have been assumed to be fixed over time. On the other hand, there are open research lines in favour of training GNNs over temporal networks [4]. Besides, our proposed link feature extraction techniques show a rather promising approach to learn explainable signals that justify each user interaction.

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] Mahmudur Rahman, Tanay Kumar Saha, Mohammad Al Hasan, Kevin S. Xu, and Chandan K. Reddy. *Dylink2vec*: Effective feature representation for link prediction in dynamic networks. *CoRR*, abs/1804.05755, 2018.
- [5] Muhan Zhang and Yixin Chen. Link prediction based on graph neural networks. *CoRR*, abs/1802.09691, 2018.