Detecting Fake Users on Social Media with a Graph Database

Yichun Zhao[1] and Jens Weber

yichunzhao@uvic.ca

Abstract

Social media has become a major part of people's daily lives as it provides users with the convenience to connect with people, interact with friends, share personal content with others, and gather information. However, it also creates opportunities for fake users. Fake users on social media may be perceived as popular and influential if not detected. They might spread false information or fake news by making it look real, manipulating real users into making certain decisions. In computer science, a social network can be treated as a graph, which is a data structure consisting of nodes being the social media users, and edges being the connections between users. Graph data can be stored in a graph database for efficient data analysis. In this paper, we propose using a graph database to achieve an increased scalability to accommodate larger graphs. Centrality measures as features were extracted for the random forest classifier to successfully detect fake users with high precision, recall, and accuracy. We have achieved promising results especially when compared with previous studies.

Keywords: fake users; graph database; machine learning; random forest; fake news

[1] I thank Dr. Jens Weber for his expertise and guidance, the peer reviewers and editors of *The Arbutus Review* for their suggestions and revisions, and the Jamie Cassels Undergraduate Research Awards for their support.

Detecting Fake Users on Social Media with a Graph Database

Social media has become a major part of people's daily lives as it provides users with the convenience to connect with people, interact with friends, share personal content with others, and gather information. However, it also creates opportunities for fake users. They might present unreliable information or fake news that is not fact-checked, or they might spam others with automated account interactions. Fake users can manipulate people into false beliefs and cause negative societal impacts by posting harmful links repeatedly and abusing the reply and sharing functions to make topics appear trendier than they actually are (Ersahin et al., 2017).

The objective of this research project is to improve the accuracy of detecting fake users reported by a previous study (Mehrotra et al., 2016). By using a graph database, the scalability of the approach is increased to billions of nodes compared to an in-memory approach (Pokorný, 2015). Different centrality measures supported by the Neo4j graph database are used as features for the random forest classifier to train and test. This detection method should also be tested with more datasets. The rest of this paper is structured as follows: The next section introduces related work regarding the detection of fake users on social media. Section 3 introduces the methods used in our study. Section 4 summarizes our results and discusses them. Section 5 concludes our study.

Related Work

Mehrotra et al. (2016) have proposed a method that uses graph based centrality measures including betweenness centrality, eigenvector centrality, in-degree centrality, out-degree centrality, katz centrality, and load centrality to detect fake followers in a social graph network. Three classifiers including artificial neural network, decision tree classifier, and random forest classifier have been compared. The random forest classifier yields the highest accuracy on fake follower detection. This method is generic and can be applied regardless of which social network is used.

Cresci et al. (2015) have used the random forest classifier and features including number of friends, number of tweets, content of tweets, and relation between number of friends and followers to detect fake users. High accuracy was obtained.

Narayanan et al. (2018) have devised a browser plugin to recognize fake accounts on Twitter by also using the random forest classifier and features such as number of friends, followers, and statuses with a precision of 95%, recall of 94%, and accuracy of 94.1%.

Gupta et al. (2017) have applied 12 supervised machine learning classification algorithms on a Facebook dataset and evaluated them to find the best performing classifiers which are mostly decision tree and decision rules classifiers. An accuracy around 79% is achieved by the random forest and JRip classifier. Features are also evaluated, and comments-related and likes-related attributes show effectiveness in detecting fake accounts.

Zenonos et al. (2018) have compared the effectiveness of using centrality measures versus characterization measures (centralization, density, and reciprocity) to detect fake influencers on Twitter. The results show that both types of measures are almost equivalently effective. Kagan et al. (2018) have created a generic unsupervised learning algorithm to detect anomalous vertices. The algorithm creates a link predication classifier based on the graph's topology in its first iteration to predict the possibility of existence of an edge with high accuracy. In the second iteration, metafeatures based on features created by the link prediction classifier are generated and used to construct an anomaly detection classifier. The algorithm has been evaluated using 10 networks that are fully simulated, semi-simulated, or real world, showing good performance with high Area Under the Curve (AUC) measure.

Investigating the related area of rumour detection, Zubiaga et al. (2018) have provided an overview of research into rumours on social media. They define rumours as circulating misinformation or disinformation whose accuracy is yet to be verified. Rumour analysis can be categorized into detection, tracking, stance classification, and veracity classification. Although recent research has focused more on the latter two categories and research in tracking has been limited, there are studies conducted in all four categories.

This paper builds upon the study by Mehrotra et al. (2016) by using a graph database for analysis, including more training features in the classifier, and including more datasets for external validation.

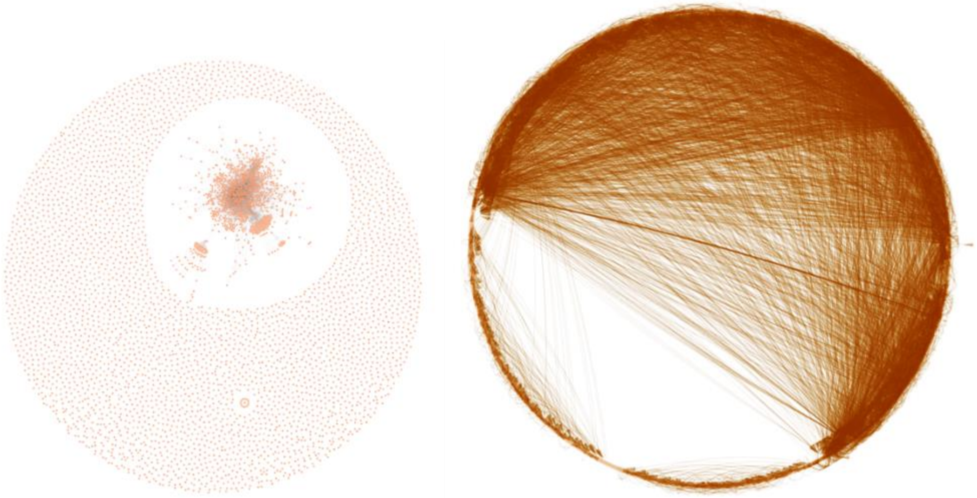
Methods and Procedures

Data Preparation

Three different datasets are used in this research project. All of them are labelled and found from previous studies on similar subject matter. The first dataset was collected by a study on detecting fake Twitter users (Cresci et al., 2015) and used in Mehrotra et al.'s (2016) study. It consists of five different split datasets of Twitter users: data from the fake project, which is all humans; data from elections 2013, which is all humans; data from intertwitter, which is all fake users; data from fastfollowerz, which is all fake users; and data from twittertechnology, which is also all fake users. The fake users were purchased from sellers on the Internet. Mehrotra et al. (2016) combined these five datasets and merged the followers and friends in their study. For the sake of simplicity, we call the combined dataset the "fake project."

Figure 1

Graphs of the Fake Project Dataset



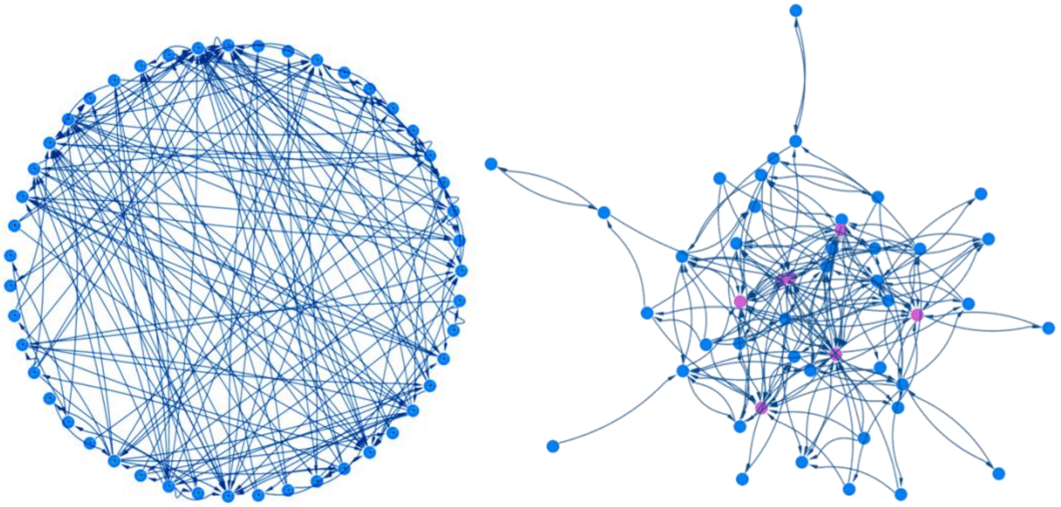
(a) Original (b) Filtered circular arrangement

The same approach is carried out in our research study. First, a shell script was executed to combine the five datasets. Then, the Neo4j graph database was used to import the whole fake project dataset to create the Twitter user nodes (5,301 nodes). Relationships between the nodes were created from the merged data of followers and friends. However, as seen in Figure 1a, only the center of the graph is clustered with nodes that have relationships between them. Surrounding the cluster are the isolated nodes that do not have relationships. The graph was then filtered by removing these isolated nodes in the Neo4j graph database to have a smaller dataset. Figure 1b shows us the filtered graph with a different arrangement of the user nodes. There are 1,222 real users and 698 fake users in the final dataset.

The second dataset is a friendship network of a school class from 1880/81 (Kagan, 2017). It is likely the earliest known social media dataset and it is used in the study by Kagan et al. (2018). We call it the "class" dataset. It was then imported into Neo4j which constructed the graph seen in Figure 2a. There are no isolated nodes. There are 40 real users and 7 fake users. Figure 2b is another arrangement of the graph, which shows us the fake nodes coloured purple.

Figure 2

Graphs of the Class Dataset



(a) Random arrangement (b) Circular arrangement of nodes with purple fake nodes

The third dataset contains data about Twitter users obtained using an application programming interface (API) crawler (Kagan, 2017), and it is also used in the study by Kagan et al. (2018). The crawler capped the maximum number of friends and followers to 1,000 for each user because the number of friends and followers is unlimited. We call it the "twitter" dataset. However, the dataset was too large (5,384,162 nodes) for the Neo4j database to calculate the centrality measures using its graph algorithms. Random sampling of the dataset was performed using Python to downsize it to 119,661 nodes. Because there are isolated nodes without any relationships, the graph was filtered to remove these nodes. There are 22,912 real users and 9,694 fake users in the final sampled and filtered dataset.

Feature Extraction

These centrality measures were calculated and extracted for the three datasets using the graph algorithms library from Neo4j version 3.5.6:

- Betweenness centrality is the number of shortest paths passing through a node, showing its influence over the flow of information in the network (Needham & Hodler, 2019).
- Closeness centrality is the number of shortest paths from a node to all other nodes, showing how efficient it is to spread information or how central the node is in the network (Needham & Hodler, 2019).
- In-degree is the number of incoming relationships of a node.
- Out-degree is the number of outgoing relationships of a node.
- Page-rank calculates the number and quality of links to a node, showing its transitive influence (Needham & Hodler, 2019).
- Eigenvector also shows a node's transitive influence (Neo4j, 2019).

Classification

Since the random forest classifier was proven to yield the highest accuracy compared to other classifiers (Mehrotra et al., 2016), the random forest classifier was performed with 10-fold cross-validation for training and testing and with centrality measures as its features. The idea of 10-fold cross-validation is to divide the whole dataset into 10 pieces, taking 1 piece as the testing dataset and the other 9 as the training dataset. The classifier is trained and tested 10 times using a different piece as the testing set each time. After the learning with 10-fold cross-validation, the final classifier was then run on the whole dataset to achieve the final results. Weka was used for this classification.

Feature Evaluation

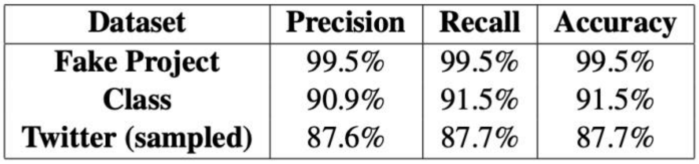
For the fake project dataset, the features were evaluated using correlation-based feature selection, information-gain-based feature selection, and learner-based feature selection in Weka to identify the most influential feature.

Results and Discussion

The most influential feature for the fake project dataset is the closeness centrality measure. Compared to the results from the previous study by Mehrotra et al. (2016) using the fake project dataset (precision = 89.0%; recall = 100%; accuracy = 95%), a significant increase in precision and accuracy is observed in Table 1.

Table 1

Results Obtained Using the Random Forest Classifier on Each Dataset



Dataset	Precision	Recall	Accuracy
Fake Project	99.5%	99.5%	99.5%
Class	90.9%	91.5%	91.5%
Twitter (sampled)	87.6%	87.7%	87.7%

This might be due to the inclusion of the new feature, closeness centrality, which has the highest correlation with the output. The slight decrease in recall might be due to the exclusion of katz centrality, which was used in the previous study. Katz centrality measures the relative influence of a node by calculating the number of immediate neighbours and other nodes that connect to the node through these immediate neighbours. Since betweenness and load centrality measures have a high degree of correlation (Zenonos et al., 2018), the exclusion of load centrality should not have an impact on the results because betweenness centrality is still included.

The lower results achieved from using the class dataset might be due to its small size (47 nodes and 177 relationships). The twitter dataset originally has a maximum number of friends and followers capped for each user. Hence, it might not accurately represent the actual Twitter network. The random sample also might not represent the same state of the original graph and might not preserve certain graph properties. In addition, both the class and twitter datasets refer to their "fake" nodes as "anomalous," meaning these users might not be fake. This fact might also explain the lower results.

When we performed the graph algorithms, the need for sampling the twitter dataset even after proper indexing was surprising due to the expectation that Neo4j should be able to scale to billions of nodes (Pokorný, 2015). Further investigation is needed.

Conclusion

Neo4j was successfully used to prepare the data and extract the centrality measures as features and to increase scalability due to its easy and quick calculations. The random forest classifier with features including betweenness centrality, in-degree, out-degree, closeness centrality, page-rank, and eigenvector detected fake users with reasonable results. Compared to the study by Mehrotra et al. (2016), the use of different centrality measures increased the precision and accuracy significantly and achieved 99.5% in precision, recall, and accuracy.

Centrality measures are node properties. In the future, we can include edge properties like link prediction measures supported by Neo4j as new features to possibly achieve even better results.

References

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2015). Fame for sale: Efficient detection of fake twitter followers. *Decision Support Systems*, *80*, 56 –71. http://doi.org//10.1016/j.dss.2015.09.003

Erşahin, B., Aktaş, Kılınç, D., & Akyol, C. (2017). Twitter fake account detection. *2017 International Conference on Computer Science and Engineering (UBMK)*, 388–392.

Gupta, A., & Kaushal, R. (2017). Towards detecting fake user accounts in Facebook. *2017 ISEA Asia Security and Privacy (ISEASP)*, 1–6.

Kagan, D. (2017). anomalous-vertices-detection. *GitHub*. (https://github.com/Kagandi/anomalous-vertices-detection/tree/master/data)

Kagan, D., Elovichi, Y., & Fire, M. (2018, Apr 05). Generic anomalous vertices detection utilizing a link prediction algorithm. *Social Network Analysis and Mining*, *8*(1), 27. http://doi.org//10.1007/s13278-018-0503-4

Mehrotra, A., Sarreddy, M., & Singh, S. (2016). Detection of fake twitter followers using graph centrality measures. *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, 499–504. http://doi.org//10.1109/IC3I.2016.7918016

Narayanan, A., Garg, A., Arora, I., Sureka, T., Sridhar, M., & Prasad, H. B. (2018). Ironsense: Towards the identification of fake user-profiles on Twitter using machine learning. *2018 Fourteenth International Conference on Information Processing (ICINPRO)*, 1–7.

Needham, M., & Hodler, A. (2019). *Graph algorithms: Practical examples in apache spark and neo4j*. O'Reilly Media, Incorporated.

Neo4j. (2019). *The eigenvector centrality algorithm*. (https://neo4j.com/docs/graph-algorithms/current/labs-algorithms/eigenvector-centrality/)

Pokorný, J. (2015, September). Graph databases: Their power and limitations. *14th Computer Information Systems and Industrial Management (CISIM)*, *LNCS-9339*, 58–69. Retrieved from https://hal.inria.fr/hal-01444505 (Part 1: Full Keynote and Invited Papers). https://doi.org/10.1007/978-3-319-24369-6_5

Zenonos, S., Tsirtsis, A., & Tsapatsoulis, N. (2018). Twitter influencers or cheated buyers? *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and CyberScience and Technology Congress (DASC/ PiCom/ DataCom/ CyberSciTech)*, 236–242.

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018, February). Detection and resolution of rumours in social media: A survey. *ACM Comput. Surv.*, *51*(2). https://doi.org/10.1145/3161603