2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Applying Social Network Embedding and Word Embedding for Socialbots Detection

1st I-Hsien Ting
National University of Kaohsiung
Kaohsiung, Taiwan
iting@nuk.edu.tw

2nd Kazunori Minetaki
Kindai University
Osaka, Japan
kminetaki@bus.kindai.ac.jp

3rd Mei-Yun Hsu
National Museum of Taiwan History
Tainan, Taiwan
myhsu@nmth.gov.tw

4th Chia-Sung Yen
National Chinyi University of Technology
Taichung, Taiwan
csyen@ncut.edu.tw

***Abstract*—With the growth of social networking website, social media has become a major platform for marketing, such as social business, political manipulation, influence and brand management, etc. However, social media marketing is very different to traditional marketing. Social media marketing needs to face large number of users and need to repeat same process frequently. It is therefore a very human power consuming task. Under this situation, it is the reason why Robotic Process Automation and Social-bots is now very popular in many social networking websites. However, there are many negative effects when applying social-bots for social marketing. Therefore, more and more researchers are devoting on propose efficient ways to detect social-bots. In this paper, we proposed an approach to detect social-bots by considering the content that users posted as well as the behavior and features when using social networking website. In this approach, we adopt the concept of word embedding and social network embedding. Convolutional neural network is used as the main techniques to train the model for social-bots detection. The experimental results show that the proposed combination approach has better detection accuracy than only social network embedding or word embedding approach as well as it reaches 92% detection accuracy by using our dataset.**

***Index Terms*—Social-bot, Social Network Embedding, Social Networks Analysis, Neural Network**

I. Introduction

According to the Digital in 2022 Global Overview jointly released by We Are Social and Hootsuite (2022), the report indicates that the number of people using the internet has surpassed half of the global population, making the ”internet” a significant part of people’s lives. The usage of social networks also increased by more than 4% in 2021, with nearly 4.6 billion people worldwide using social networks at least once a month. Data from the Global Web Index highlights that as people spend more time on social networks, the daily average time spent on social platforms has reached 2 hours and 27 minutes [38].

With such shifts in habits, many social platforms have experienced remarkable user growth, underscoring the increasing reliance of people on these platforms as sources of important information. Statistics reveal that over half of active social platform users are on Facebook, and in 2022, Facebook’s monthly active users surpassed 2.9 billion, highlighting its prominent status and significance within the realm of social platforms. In Taiwan, Facebook also reigns as the primary social platform, with a monthly active user count reaching 18 million in 2022. Nearly 80% of the population utilizes Facebook, and according to a survey provided by Facebook, it stands as the social media platform with the longest daily usage time among Taiwanese individuals [38].

Due to the substantial user base on social platforms, the discourse and content within these platforms can simultaneously influence a large number of users, yielding highly effective outcomes in marketing. As a result, various fields have gradually emphasized advertising and marketing operations on social platforms, encompassing areas such as recruitment, product promotion, endorsements by celebrities and internet personalities, and even politics, where it has been demonstrated to impact election results [2]–[4], [8]. This has propelled social platforms into becoming a vital and mainstream marketing medium today.

Unlike traditional marketing, the distinctive feature of social marketing lies in its extensive audience reach and the need for 24/7 responsiveness and interaction. This demands a significant amount of manpower to carry out repetitive actions. Furthermore, for specific targeted social marketing efforts, it may even involve the creation of numerous virtual accounts to facilitate one-on-one marketing or to boost engagement metrics like likes, shares, and comments on social platform posts. These accounts often require separate device logins to avoid being flagged as fake accounts and subsequently banned by the platform. As depicted in Figure 1, the traditional physical device group control system follows a one-account-one-device

ASONAM ’23, November 6-9, 2023, Kusadasi, Turkey

ACM ISBN 979-8-4007-0409-3/23/11…$15.00
http://dx.doi.org/10.1145/3625007.3627306

approach. However, this method incurs high costs and requires manual intervention. The latest group control systems have evolved into software-based solutions that simulate mobile devices, allowing for account binding and manipulation, as shown in Figure 2 . [42], [43].

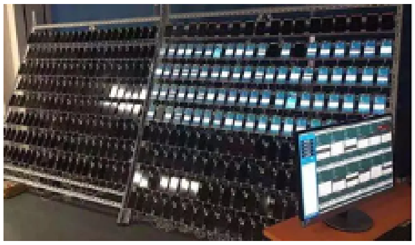

Fig. 1. Group control system by using physical mobile devices

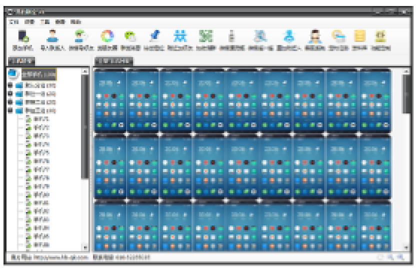

Fig. 2. Group control system by using software based virtual machines

The current development of group control systems has become increasingly diverse, with various functionalities incorporating the concept of Robotic Process Automation (RPA) [24]. This allows systems to automatically perform tasks that typically require substantial manual effort [30], such as automated liking, commenting, and browsing, among others. The concept of RPA is now widely applied across various business sectors, including banking, financial accounting, marketing, industrial processes, and more, primarily aiming to automate processes and reduce the need for extensive manpower or complex workflows [25]. When applied to social platforms, RPA is often referred to as Social Bots [6].

The technology behind Social Bots can provide significant benefits for social marketing. However, if used with improper motives, it can lead to substantial negative effects and issues. For instance, it can be involved in improper political operations and manipulation of public opinion, affecting government policies and even election outcomes [41]. There are also concerns about orchestrated attacks resembling ”web armies,” the spread of fake news and misinformation [7]. These problems erode trust in information and interactions on social platforms, and the widespread behavior and activities of Social Bots can lead to user attrition from these platforms [32]. As a response, social platforms are putting great effort into identifying and removing these Social Bots to enhance the credibility of their platforms.

In the context described above, many scholars have initiated research into the distinctions between behaviors of human users and social bots. Studies have revealed that differences can be discerned in various aspects, including social interaction behaviors (such as interaction frequency, speed, sequence, etc.), content of discourse, character profiles, and the behaviors of human users versus social bots [6]. However, the behaviors of social bots are not static. To evade detection by social platforms, they are constantly updated and maintained. If traditional detection methods are based solely on deducing behaviors, they may struggle to keep up with the evolution speed of social bots [25], [30].

Social Networks Embedding is often employed in social network analysis techniques to transform multi-dimensional features of social networks into vector spaces, thus reducing data dimensionality [23]. This approach is suitable for classifying nodes within networks and is thus well-suited for detecting social bots [39]. Through social networks embedding, the integration of social network analysis and artificial intelligence becomes possible. For example, neural networks and deep learning can be utilized for node classification [22].

The concept of Word Embedding is similar to social networks embedding. By creating word vectors that define the correspondence between words and dimensions, a word matrix is established. This method considers word order and is highly efficient for natural language processing tasks that require sensitivity to word sequence for meaning, especially in text analysis where word order significantly impacts the interpretation [29]. Word embedding is well-suited for text classification and is expected to effectively differentiate between bot and human speech after classification, if applied to detecting social bots.

Based on the aforementioned context and motivations, if relevant information technology can be applied to detect social bots, it would effectively restore the integrity of social platform information and prevent the negative consequences these bots may cause to communities and societies. Therefore, the title of this paper is defined as ”Research on Utilizing Social Embedding and Word Embedding for the Detection of Social Bots.”

II. Literature Review and Related Works

A. Social Networks Analysis

In the field of social network analysis, there is a long history of relevant research. As early as 1925, Lewin introduced the interpretation of interactions between individuals through simple graphs of nodes and edges. Wasserman and Faust provided an initial definition of social network analysis in 1994, in which they considered it a sociological methodology. They proposed that social network analysis involves analyzing patterns of relationships and interactions among social actors to uncover potential social network structures [46]. As a result, many terminologies and definitions related to social network analysis have become well-established and widely applied across various domains, including sociology, management and business, biology, and information science [19], [49].

Two prominent books, ”Social Network Analysis: A Handbook” and ”Social Network Analysis: Methods and Applications" [37], [46], provide comprehensive introductions to the fundamental concepts and methods of social networks and social network analysis. These include definitions of various roles within social networks, definitions and calculations of relationships in social networks (such as centrality, closeness, betweenness, network clusters, network diameter, etc.), and the positioning of nodes and structural holes within social network structures [5], [48]. These well-defined terminologies, methods, and metrics in social network analysis establish a solid foundation for scholars venturing into this field.

In the realm of social network analysis within information science, the focus is on efficiently and rapidly processing vast amounts of data while building upon the foundation of traditional sociological social network analysis [47]. Two common approaches are the ego-centric and whole-network approaches to social network analysis. Presently, the research direction in information technology for social network analysis not only emphasizes the effective and comprehensive analysis of data-rich social networks [13], [14], [27] but also contemplates how existing information technology and data analysis techniques can be applied to social network analysis. Examples include HITS (Hypertext Induced Topics Selection), Semantic Web, PageRank, and more [10], [44]. Matsuo et al. utilized integrated techniques for analyzing web data and links, incorporating user interaction analysis, user description analysis, and web mining to construct social networks [26].

Among the visible techniques for social network analysis and construction, the utilization of data mining and web mining techniques is commonplace for data analysis and processing. As a result, data mining and social network analysis are closely intertwined [1]. Web mining is classified into three categories based on the processed data: web content mining, web usage mining, and web structure mining. In online social network analysis, message classification or clustering is a primary task, and web content mining can aid in determining users' preferences for messages. Usage patterns can be transformed into relational data for social network construction using web usage mining [21]. In terms of the overall network structure, web structure mining can analyze path lengths, reachability, and detect structural holes within the network.

B. Social Networks Embedding, Words Embedding and Social Bots Detection

1) Social Networks Embedding: Social Networks Embedding is a technique commonly used in social network analysis to transform multi-dimensional features of social networks into vector spaces, thus reducing data dimensionality [23].

User embedding based on social networks is one of the methods for user embedding and represents node features and social network structures as vectors in a vector space. The most common social network embedding methods include DeepWalk, Node2Vec, and Non-Negative Matrix Factorization [31].

The DeepWalk method primarily learns latent features of nodes in a network through random walks. This type of network embedding can be applied in various fields, such as user interest prediction and anonymous detection [35]. Node2Vec improves upon DeepWalk's shortcomings by using Biased Random Walk, which better preserves the so-called first-order and higher-order estimates in the social network [15]. Node2Vec has been applied to interest prediction and classification of Twitter's social network structure [11]. The last method, Non-Negative Matrix Factorization (NFM), utilizes matrix factorization to represent node connections in the network using an adjacency matrix, reducing the dimensionality of network features. NFM can preserve both first-order and second-order estimates in the network [45].

2) Words Embedding: The concept of Word Embedding is similar to Social Networks Embedding. It involves creating word vectors that define the correspondence between words and dimensions, forming a matrix of words. This approach considers the order of word occurrences and is particularly effective for natural language processing tasks where word sequence significantly impacts meaning [29].

Word Embedding has three main methods: Word2Vec, Doc2Vec, and GloVe. Word2Vec primarily uses neural networks to predict words corresponding to context inputs, such as the context above and below the target word [28]. Doc2Vec is an unsupervised algorithm that calculates distances to determine the similarity between sentence segments or documents, often employed in text classification tasks [36]. GloVe (Global Vectors for Word Representation) is considered one of the fastest training algorithms and is capable of processing large corpora efficiently. It performs well even with smaller corpora and vector sizes [34].

3) Social Bots Detection: In recent years, scholars have started using the techniques of Social Networks Embedding and Word Embedding for the detection of social bots. In related research, social network structure and interactions with friends are the most common methods for detection. Kolomeet et al. proposed an analysis method based on friend network relationships, achieving an accuracy rate of up to 91% [20]. Li et al. presented a more comprehensive framework that utilizes social network embedding and employs similarity measurement techniques for detection [22]. Dehghan et al. combined natural language processing and social network embedding using Twitter as their dataset [9]. Guo et al. introduced a method using BERT combined with neural networks for text-based bot detection, achieving over 90% accuracy [16].

Apart from the foundational approaches to social bot detection mentioned above, numerous studies have combined social network embedding with neural network techniques for detection. Many of these studies have emerged within the past couple of years, highlighting the significance of this issue. Fazil et al. proposed a framework that employs CNN to process user behaviors after social network embedding [12]. Yang et al. introduced a similar framework, utilizing Twitter bot data and Graph Neural Network (GNN) techniques for social network embedding [50]. Peng et al. also focused on GNN, targeting music-based social network data [33]. Hayawi et al. developed a comprehensive framework for Twitter bot detection, incorporating textual factors and numerical user behavior data, although interactions and network structure were not considered. Their detection was accomplished using neural network techniques [18]. Skorniakov et al. tested various embedding methods and classification approaches using data from social platforms VK and Twitter, achieving detection accuracy ranging from 60% to 80% [39].

Upon reviewing the literature, it becomes evident that CNN is often the most effective technique tested. Therefore, in this research project, CNN will also serve as the core technology for classification training and judgment. However, among these studies, Twitter is the most commonly used data source due to its accessible datasets of known bots. There is limited availability of Facebook datasets. Additionally, most of the mentioned studies tend to focus on a single factor as the measuring criterion, such as user data, social network structure, or text. Comprehensive consideration of all potential factors is currently less common in the literature.

III. Social Bots Detection Approach

Based on the aforementioned background and literature review, this paper proposes a framework for social bot detection and classification, utilizing the concepts of social network embedding and word embedding, with convolutional neural networks as the core technology. The schematic diagram of the research framework is illustrated in Figure 3. The upcoming sections of this paper will provide detailed explanations for each step.

1) Data Sources: The data source for this research project is Facebook, covering the period from December 1, 2021, to November 25, 2022 (during the 2022 Taiwan local elections). A total of 4,235 known social network bot accounts and 4,000 genuine user accounts were collected during this period, including their personal information and all activity data. The data was obtained using the publicly available API provided by Facebook (https://developers.facebook.com/docs/graph-api/). As the research team currently possesses account access for these profiles, all data collected falls within the authorization scope granted by Facebook.

2) Training and Testing Data: All collected data in this study is divided into training data and testing data in a 2:1 ratio. The training dataset consists of 2,823 social network bot accounts and 2,667 genuine user accounts. The testing dataset includes 1,333 social network bot accounts and 1,412 genuine user accounts.

3) Features Extraction: In this step, social data features are extracted. This project will extract social network features as shown in Table 1. The feature categories include interaction, friendship, posting, personal, network size, and centrality. The attributes for each category are presented in the table and are derived from literature analysis and this study's synthesis. These attributes are potentially capable of distinguishing between social network bots and genuine users [7], [18], [50].

The social network hierarchy in this study is divided into nodes and sub-networks. Sub-network segmentation will be performed using the Random Walk approach. This information will be provided to the social network embedding step for processing. The data in Table I can be directly obtained, while some data will be acquired after undergoing preliminary calculations. All data formats are of numeric type.

TABLE I
Social Network Features that Used in This Research

Category	Features	Level
Interaction	Frequency, Ration, Maximum Time Interval , Minimum Time Interval, Total Time, Medium, Time Point	Node
Friendship	Friend List, Number of Friends, Sequence, Interval, Time Point, Speed	Node
Post	Frequency, Ratio, Maximum Time Interval , Minimum Time Interval, Total Time, Medium, Time Point	Node
Personal	Age, Gender, Location, Education, Information Completeness	Node
Network Size	Size, In-degree, Out-degree, Density, Diameter	Sub-network
Centrality	Degree-centrality, Closeness-centrality, Betweenness-centrality	Sub-network

4) Natural Language Processing: Textual data types collected in this project, such as personal descriptions, posts, and comments, will undergo standard natural language processing procedures, including sentence segmentation, word tokenization, part-of-speech tagging, syntactic analysis, and sentiment analysis. The processed data will be used for word embedding purposes.

5) Social Network Embedding: This project will utilize the Node2Vec algorithm for conducting social network embedding. Through social network embedding, the social network features from Table I can be effectively transformed into a vector space, thereby reducing dimensionality. As shown in Figure 4, the vector space of node features after social network embedding is presented. Nodes of the same color in the table represent being part of the same sub-network. Fi represents the node's social network features.

As shown in Figure 5, the vector space of sub-network features after social network embedding is presented. The colors in the table represent different sub-networks, and SFi represents the social network features of the sub-network. The features of the sub-network are represented using the representative feature values of nodes within the same sub-network. This approach helps in reducing the dimensionality of the entire network.

6) Word Embedding: In the literature review, the GloVe method for word embedding has been verified to have better

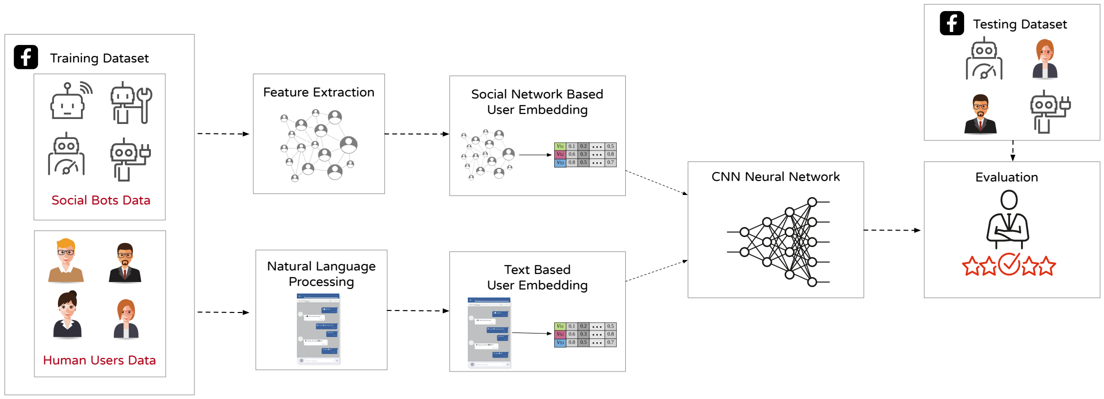

Fig. 3. The Research Architecture of The Paper

Node	F_1	F_2	F_3	⋯	F_i
1	0.2	0.6	0.6	⋯	0.3
2	0.4	0.9	0.5	⋯	0.2
3	0.3	0.8	0.9	⋯	0.1
4	0.5	0.7	0.9	⋯	0.1
5	0.7	0.3	0.4	⋯	0.1
.	.	.	.	⋯	.
N-1	0.4	0.7	0.5	⋯	0.5
N	0.8	0.5	0.4	⋯	0.8

Fig. 4. The Vector Space of Nodes After the Process of Social Network Embedding

Sub-network	SF_1	SF_2	SF_3	⋯	SF_i
SN1	0.2	0.9	0.5	⋯	0.2
SN2	0.5	0.7	0.9	⋯	0.1
SN3	0.4	0.7	0.5	⋯	0.5

Fig. 5. The Vector Space of Nodes After the Process of Social Network Embedding

performance. Therefore, this project also plans to adopt this word embedding approach. This approach represents a word as a vector composed of numerical values, enabling the depiction of similarities and analogies. Through distance calculations, semantic similarity between words can be computed.

7) Convolutional Neural Network; CNN: After the collected data preprocessing and word embedding as well as social network embedding, it will enter the model training phase. In this stage, the project will import all the data into a convolutional neural network (CNN) for training. Through a multi-layer CNN deep learning algorithm, the word vector features and social network vector features distinguishing real users from social bots will be successfully classified. The trained model will then be used for subsequent discrimination and result validation. Continuous adjustments of relevant parameters and data distributions will be made to achieve optimal accuracy.

8) Evaluation: In the final phase of this project, the testing dataset will be used to validate whether the model trained through the convolutional neural network can accurately determine whether a user is a real user or a social bot. As shown in Figure 6, it illustrates the architecture for detecting social bots using convolutional neural network techniques.

In Figure 6, all testing dataset will undergo social network feature extraction and natural language processing, followed by word embedding and social network embedding. Finally, the model obtained from Step 7 of this project will be used for determination. The project will conclude by employing a common technique in the field of machine learning, the Confusion Matrix, for model validation and evaluation [40].

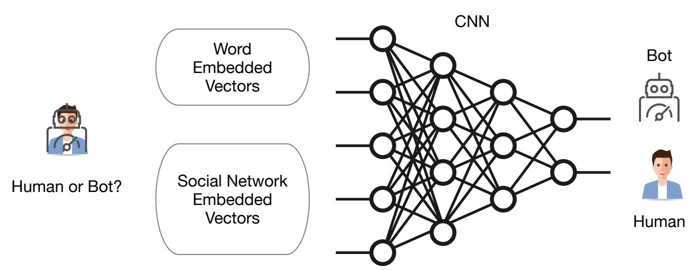

Fig. 6. The Architectural Diagrams of Social Bots Detection Using Convolutional Neural Network Techniques

IV. Evaluation

In our dataset, there are 2745 testing data. Among the testing data, there are 1333 data made by genuine users and 1412 by social bots. After performing the proposed social

TABLE II
THE CONFUSION MATRIX FOR PERFORMANCE EVALUATION

	Positive	Negative
Positive	1198	250
Negative	145	1162

bots detection approach. The confusion matrix of performance evaluation is shown in Table II.

From the confusion matrix, the Presicion=0.83, Recall=0.89 and F1 Score=0.86. The performance evaluation therefore shows a very ideal results that more than 80% of the data made by social bots can be detected.

V. CONCLUSION

In this paper, we present a novel method for detecting social bots. Our approach integrates social network embedding and word embedding techniques with a CNN neural network, enabling us to train a robust model using a Facebook dataset. Our performance evaluation demonstrates the efficacy of our approach, achieving a detection rate of over 80% for content generated by social bots. This holds significant promise for its application in the realm of social marketing.

Looking ahead, our research will encompass an expansion of the dataset, encompassing data from diverse social networking websites. Furthermore, we intend to refine our approach for real-world application, transcending reliance solely on the dataset.

REFERENCES

[1] Agrawal, R., Rajagopalan, S., Srikant, R., & Xu, Y. (2003). Mining Newsgroups Using Networks Arising From Social Behavior. World Wide Web 2003 Conference, 529–535.

[2] Arora, A., Bansal, S., Kandpal, C., Aswani, R., & Dwivedi, Y. (2019). Measuring social media influencer index- insights from facebook, Twitter and Instagram. Journal of Retailing and Consumer Services, 49, 86–101. https://doi.org/10.1016/j.jretconser.2019.03.012

[3] Bhattacharyya, S., & Bose, I. (2020). S-commerce: Influence of Facebook likes on purchases and recommendations on a linked e-commerce site. Decision Support Systems, 138, 113383. https://doi.org/10.1016/j.dss.2020.113383

[4] Bossetta, M. (2018). The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. Journalism & Mass Communication Quarterly, 95(2), 471–496. https://doi.org/10.1177/1077699018763307

[5] Burt, S. P. (1992). Structural Holes. Harvard University Press.

[6] Cai, M., Luo, H., Meng, X., & Cui, Y. (2022a). Differences in Behavioral Characteristics and Diffusion Mechanisms: A Comparative Analysis Based on Social Bots and Human Users. Frontiers in Physics, 10. https://doi.org/10.3389/fphy.2022.875574

[7] Cai, M., Luo, H., Meng, X., & Cui, Y. (2022b). Differences in Behavioral Characteristics and Diffusion Mechanisms: A Comparative Analysis Based on Social Bots and Human Users. Frontiers in Physics, 10. https://doi.org/10.3389/fphy.2022.875574

[8] Chugh, R., & Ruhi, U. (2018). Social media in higher education: A literature review of Facebook. Education and Information Technologies, 23(2), 605–616. https://doi.org/10.1007/s10639-017-9621-2

[9] Dehghan, A., Siuta, K., Skorupka, A., Dubey, A., Betlen, A., Miller, D., Xu, W., Kamiński, B., & Prałat, P. (2022). Detecting Bots in Social-networks using Node and Structural Embeddings. Proceedings of the 11th International Conference on Data Science, Technology and Applications, 50–61. https://doi.org/10.5220/0011147300003269

[10] Dingt, C. H., Zha, H., Husbands, P., & Simont, H. D. (2004). Link Analysis: Hubs and Authorities on the World Wide Web. SIAM Review, 46(2), 256–268.

[11] Do, T. H., Nguyen, D. M., Tsiligianni, E., Cornelis, B., & Deligiannis, N. (2018). Twitter User Geolocation Using Deep Multiview Learning. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 6304–6308. https://doi.org/10.1109/ICASSP.2018.8462191

[12] Fazil, M., Sah, A. K., & Abulaish, M. (2021). DeepSBD: A Deep Neural Network Model with Attention Mechanism for SocialBot Detection. IEEE Transactions on Information Forensics and Security, 16, 4211–4223. https://www.statista.com/statistics/278414/number-of-

[13] Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. Proceedings of ICWSM 2007.

[14] Goodreau, S. M. (2007). Advances in exponential random graph (p *) models applied to a large social network. Sociological Methodology, 29(39), 231–248. https://doi.org/10.1016/j.socnet.2006.08.001

[15] Grover, A., & Leskovec, J. (2016). node2vec. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 855–864. https://doi.org/10.1145/2939672.2939754

[16] Guo, Q., Xie, H., Li, Y., Ma, W., & Zhang, C. (2022). Social bots detection via fusing bert and graph convolutional networks. Symmetry, 14(1). https://doi.org/10.3390/sym14010030

[17] Hajli, N., Saeed, U., Tajvidi, M., & Shirazi, F. (2022). Social Bots and the Spread of Disinformation in Social Media: The Challenges of Artificial Intelligence. British Journal of Management, 33(3), 1238–1253. https://doi.org/10.1111/1467-8551.12554

[18] Hayawi, K., Mathew, S., Venugopal, N., Masud, M. M., & Ho, P. H. (2022). DeeProBot: a hybrid deep neural network model for social bot detection based on user profile data. Social Network Analysis and Mining, 12(1). https://doi.org/10.1007/s13278-022-00869-w

[19] June, T., Kim, J., Kim, B. J., & Choi, M. Y. (2006). Consumer Referral in A Small World Network. Social Networks, 28(3), 232–246.

[20] Kolomeets, M., Chechulin, A., & Kotenko, I. (2021). Bot detection by friends graph in social networks. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 12(2), 141–159. https://doi.org/10.22667/JOWUA.2021.06.30.141

[21] Lento, T., Welser, H. T., Gu, L., & Marc, S. (2006). The Ties that Blog: Examining the Relationship Between Social Ties and Continued Participation in the Wallop Weblogging System. Proceedings of the 15th International World Wide Web Conference, 23.

[22] Li, S., Zhao, C., Li, Q., Huang, J., Zhao, D., & Zhu, P. (2022). BotFinder: a novel framework for social bots detection in online social networks based on graph embedding and community detection. World Wide Web. https://doi.org/10.1007/s11280-022-01114-2

[23] Lizi Liao, Xiangnan He, Hanwang Zhang, & Tat-Seng Chua. (2018). Attributed Social Network Embedding. Social and Information Networks.

[24] M. H. Jansen-Vullers, M. Netjes, & H. A. Reijers. (2004). Business process redesign for effective e-commerce. The 6th International Conference OnElectronic Commerce, 382–391.

[25] Martínez-Rojas A., Sánchez-Oliva J., López-Carnicer J.M., & Jiménez-Ramírez A. (2021). AIRPA: An Architecture to Support the Execution and Maintenance ofAI-Powered RPA Robots. In González Enríquez J., Debois S., Fettke P., PlebaniP., van de Weerd I., & Weber I. (Eds.), Business Process Management: Blockchainand Robotic Process Automation Forum. BPM 2021. Lecture Notes in BusinessInformation Processing (Vol. 428).

[26] Matsuo, Y., Mori, J., & Hamasaki, M. (2006). POLYPHONET: An Advanced Social Network Extraction System from the Web. Proceedings of 2006 International World Wide Web Conference.

[27] Mika, P. (2005). Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks. Web Semantics, 3(2–3), 211–223.

[28] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013, January 16). Efficient Estimation of Word Representations in Vector Space. ICLR 2013.

[29] Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2022). Deep Learning–based Text Classification. ACM Computing Surveys, 54(3), 1–40. https://doi.org/10.1145/3439726

[30] Noppen, P., Beerepoot, I., van de Weerd, I., Jonker, M., & Reijers, H. A. (2020). How to Keep RPA Maintainable? (pp. 453–470). https://doi.org/10.1007/978-3-030-58666-9-26

[31] Pan, S., & Ding, T. (2019). Social Media-based User Embedding: A Literature Review. Social and Information Networks. http://arxiv.org/abs/1907.00725

[32] Paquet-Clouston, M., Bilodeau, O., & Décary-Hétu, D. (2017). Can We Trust Social Media Data? Proceedings of the 8th International Conference on Social Media & Society - SMSociety17, 1–9. https://doi.org/10.1145/3097286.3097301

[33] Peng, H., Zhang, Y., Sun, H., Bai, X., Li, Y., & Wang, S. (2022). Domain-Aware Federated Social Bot Detection with Multi-Relational Graph Neural Networks. Proceedings of the International Joint Conference on Neural Networks, 2022-July. https://doi.org/10.1109/IJCNN55064.2022.9892366

[34] Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532–1543. https://doi.org/10.3115/v1/D14-1162

[35] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 701–710. https://doi.org/10.1145/2623330.2623732

[36] Quoc Le, & Tomas Mikolov. (2014). Distributed representations of sentences and documents. Proceedings of the 31st International Conference on International Conference on Machine Learning.

[37] Scott, J. (2000). Social Network Analysis: A Handbook. SAGE Publication.

[38] Simon Kemp. (2022, January 26). DIGITAL 2022: GLOBAL OVERVIEW REPORT. DataReportal. https://datareportal.com/reports/digital-2022-global-overview-report

[39] Skorniakov, K., Turdakov, D., & Zhabotinsky, A. (2018). Make Social Networks Clean Again: Graph Embedding and Stacking Classifiers for Bot Detection.

[40] Stehman, S. v. (1997). Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment, 62(1), 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7

[41] Ting, I. (2016). Detection of the Spiral of Silence Effect in Social Media. Journal of Universal Computer Science, 22(3), 438–452.

[42] Ting, I.-H., Yen, C. S., Kang, C.-C., & Yang, S.-C. (2022, November). An Empirical Study of Automatic Social Media Content Labeling and Classification based on BERT Neural Network. The 2022 The International Conference Series on Advances in Social Network Analysis and Mining. https://github.com/fxsjy/jieba

[43] Ting, I.-H., Yen, C.-S., & Yang, S.-C. (2021). RPA as A Tool for Social Media Marketing: The good, The Bad and The Ugly. 49–52. https://doi.org/10.1145/3504006.3504015

[44] Vise, D., & Malseed, M. (2005). The Google Story. Delacorte Press.

[45] Wang, X., Cui, P., Wang, J., Pei, J., Zhu, W., & Yang, S. (2017). Community Preserving Network Embedding. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1). https://doi.org/10.1609/aaai.v31i1.10488

[46] Wasserman, B., & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.

[47] Wellman, B. (1982). Studying Personal Communities. In M. P. & L. N. (Eds.), Social Structure and Network Analysis (pp. 61–80). SAGE Publication.

[48] Wellman, B., & Berkowitz, S. D. (1988). Social Structures: A Network Approach. Cambridge University Press.

[49] Wilson, D. S. (1989). Levels of Selection: An Alternative to Individualism in Biology and the Human Science. Social Networks, 11(3), 257–272.

[50] Yang, Y., Yang, R., Li, Y., Cui, K., Yang, Z., Wang, Y., Xu, J., & Xie, H. (2022). RoSGAS: Adaptive Social Bot Detection with Reinforced Self-Supervised GNN Architecture Search. ACM Transactions on the Web. https://doi.org/10.1145/3572403