Social Bot Detection through Structural Holes and Centrality Measures: Analyzing the Position of Humans and Bots in Networks

The presence of bot users on social media has become a significant issue over the past decade, posing threats to online discourse and public trust. As shown in problem section in *Fig1*, these automated accounts have been widely used to amplify disinformation, manipulate public opinion, interfere in elections, promote extremist ideologies, and spread conspiracy theories at an unprecedented scale (Ferrara et al. 2016; Shao et al. 2018; Cui et al. 2020; Wang et al. 2020). Their ability to coordinate in networks, evade detection, and mimic human behavior makes them a persistent challenge for both researchers and platform regulators (Cresci 2020). Studies have shown that bots can systematically distort engagement metrics, create artificial trends, and influence decision-making in political and financial contexts (Bessi and Ferrar 2016). Therefore, analyzing the positional differences between human and bot users is crucial for developing an effective detection model. While existing studies have examined various characteristics of bots, such as behavior (Feng et al. 2021; Lee, Eoff, and Caverlee 2011) and content(Wei and Nguyen 2019), few have explored whether there are significant differences in the structural positions of bots and humans within the network. This gap presents a significant challenge, as understanding these differences could provide valuable insights into how bots and humans function differently in online spaces, ultimately helping to make detection models more effective and accurate. Addressing this gap is essential for advancing the understanding of social bot dynamics and enhancing detection strategies.

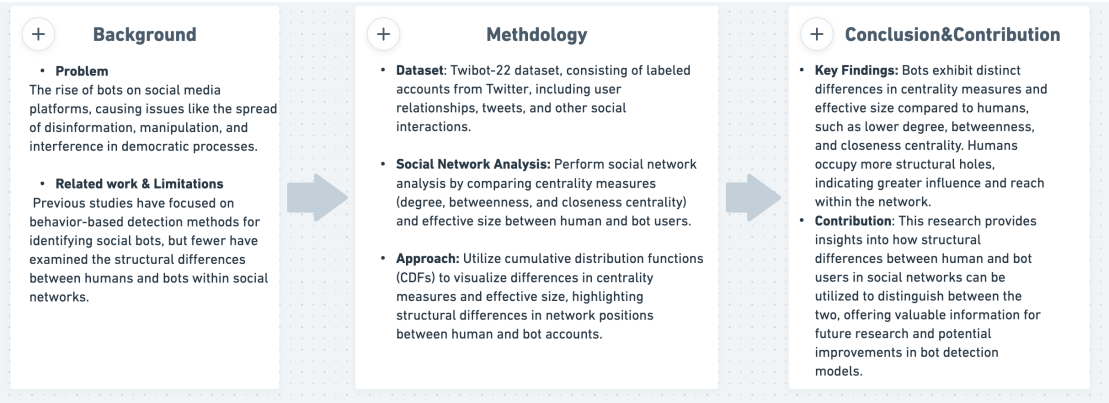

Figure1. **Flowchart of proposal.** The flowchart contains the Background, Methodology and Conlusion&Contribution sections.

This project seeks to answer the following research question: *How do the structural positions of human and bot users differ within social networks?* This question is important from both social

science and machine learning perspectives. From a social science viewpoint, understanding the positional differences between humans and bots can provide insights into their distinct roles and behaviors in online interactions. From a machine learning perspective, exploring these differences helps determine whether incorporating network-based features can improve a model's ability to distinguish between human and bot users, leading to more accurate detection models.

The dataset used in this project, Twibot-22, is derived from the Twitter (Feng et al. 2022). It is highly relevant to fields such as social media analysis and cybersecurity. Twibot-22 contains labeled data on both human and bot users, including user relationships and tweets. This makes the dataset well-suited to address the research question on the positional differences between human and bot users in social networks. By analyzing these differences, it provides valuable insights into their distinct roles and behaviors within online spaces.

In this study, we conduct a social network analysis on the Twibot-22 dataset, focusing on the differences in centrality measures (degree centrality, closeness centrality, betweenness centrality) and structural hole measures (effective size) between human and bot accounts within the following network. Social network analysis is well-suited for this research because it captures the relational structure between users, revealing how humans and bots occupy different positions in the network that cannot be captured by numerical features alone. We construct the network as an undirected graph, where users are represented as nodes, and the edges denote following relationships. Centrality measures such as degree centrality, closeness centrality, and betweenness centrality help characterize the relative importance and influence of a user within the network (Freeman 2002; Wasserman 1994). Additionally, structural hole measures like effective size capture how users control information flow by occupying unique positions in the network, with fewer redundant connections(Lin et al.2021). By comparing the distribution of these centrality and structural hole measures, we expect to observe distinct differences between human and bot users, which we will visualize using cumulative distribution functions (CDFs). This analysis will provide insight into the structural differences in how human and bot accounts interact within the network (see methodology section in *Fig1*).

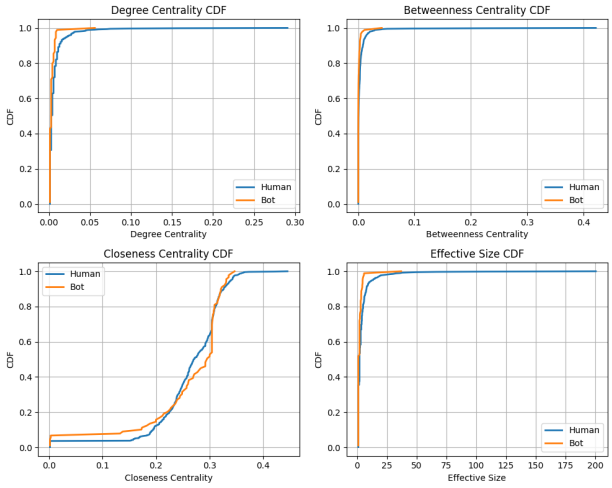

Figure2. **Cumulative Distribution Functions of Network Metrics: Human vs Bot.** The CDFs demonstrate the structural difference between human and bot measured by centrality measures and structure hole measure.

As shown in *Fig* 2, the CDF plots compare the distributions of degree centrality, betweenness centrality, closeness centrality, and effective size between human and bot accounts in the Twibot-22 dataset. In terms of degree centrality, both humans and bots show low values, with most users having connections near zero. However, bots (represented in orange) tend to accumulate connections faster, indicating that they have fewer links compared to humans. This suggests that bots are less integrated into the network and occupy less connected positions. For betweenness centrality, the distribution is highly skewed, with most users having low values. Bots generally have lower betweenness, meaning they are less likely to act as intermediaries or control the flow of information within the network. This indicates that bots are not central hubs of communication in the network. Regarding closeness centrality, bots accumulate values more rapidly at the lower end of the scale, meaning they are farther from other users and more isolated. In contrast, human accounts are generally more connected and have higher closeness centrality, indicating that they are better connected within the network. In terms of effective size, humans generally have larger values, indicating that they occupy more structural holes in the network. This means that humans tend to have more diverse, non-redundant connections. Bots, on the other hand, show smaller effective sizes, suggesting that their connections are more clustered and redundant, limiting their reach and influence. These findings highlight key differences between human and bot accounts. Bots tend to have lower degree, betweenness, and closeness centrality, making them less integrated and influential in the network. The effective size measure further reveals that humans bridge different parts of the network, while bots form more redundant connections. This suggests that network metrics, particularly effective

size, can provide valuable insights for bot detection.

This research extends existing literature on social bot detection by focusing on network-based features, an area that has been underexplored compared to behavioral features such as interaction patterns or content analysis (e.g., Ferrara et al., 2016; Shao et al., 2018; Mazza et al. 2019). While previous studies have identified bots based on user behavior, they often neglect the structural differences in how bots and humans occupy different positions within a network. By incorporating centrality measures like degree, betweenness, and closeness centrality, as well as effective size, this study provides a new perspective on bot detection through network structure. This work also highlights the potential of using structural insights to distinguish between bots and humans, offering a complementary approach to behavioral methods. However, limitations such as the use of a single dataset (Twibot-22) and a narrow set of network features suggest opportunities for future research. One promising direction is to explore multimodal data, examining how bots and humans differ in their engagement with images and videos. Combining network-based features with visual using ResNet(He et al. 2016) or multimedia analysis could provide a richer understanding of bot behavior. Expanding the dataset to include other platforms would also improve the generalizability of the findings.

This research has real-world implications for distinguishing between human and bot accounts on social media, helping to reduce disinformation and improve online trust. By enhancing the ability to distinguish bots from humans, the project promotes a healthier digital ecosystem. Potential applications include improving content moderation systems, enhancing user experiences, and informing public policy around digital governance. The machine learning approach in this project aligns with ethical AI principles by promoting responsible innovation, ensuring fairness, and mitigating bias. One potential risk is the mislabeling of data, which could lead to incorrect conclusions. To address this, in future work, multiple datasets can be used for validation to ensure accuracy and fairness. This project fosters inclusivity in AI development by focusing on distinguishing bots from humans equitably across different user groups. By helping to distinguish between bots and humans, the project contributes to SDG 16 (peace, justice, and strong institutions) and long-term societal well-being. It promotes responsible AI use in both public and private sectors, safeguarding democratic processes and supporting a more trustworthy online environment.

Supplementary Materials

Data and code is available at https://github.com/Rising-Stars-by-Sunshine/Boen-Final-Project/tree/main.

Bibliography

Bessi, Alessandro, and Emilio Ferrara. "Social bots distort the 2016 US Presidential election online discussion." First monday 21, no. 11-7 (2016).

Cresci, Stefano. "A decade of social bot detection." *Communications of the ACM* 63, no. 10 (2020): 72-83.

Cui, Limeng, Haeseung Seo, Maryam Tabar, Fenglong Ma, Suhang Wang, and Dongwon Lee. "Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation." In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 492-502. 2020.

Feng, Shangbin, Herun Wan, Ningnan Wang, Jundong Li, and Minnan Luo. "Satar: A self-supervised approach to twitter account representation learning and its application in bot detection." In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 3808-3817. 2021.

Feng, Shangbin, Zhaoxuan Tan, Herun Wan, Ningnan Wang, Zilong Chen, Binchi Zhang, Qinghua Zheng et al. "Twibot-22: Towards graph-based twitter bot detection." *Advances in Neural Information Processing Systems* 35 (2022): 35254-35269.

Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. "The rise of social bots." *Communications of the ACM* 59, no. 7 (2016): 96-104.

Freeman, Linton C. "Centrality in social networks: Conceptual clarification." *Social network: critical concepts in sociology. Londres: Routledge* 1 (2002): 238-263.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*,

pp. 770-778. 2016.

Lee, Kyumin, Brian Eoff, and James Caverlee. "Seven months with the devils: A long-term study of content polluters on twitter." In *Proceedings of the international AAAI conference on web and social media*, vol. 5, no. 1, pp. 185-192. 2011.

Lin, Zihang, Yuwei Zhang, Qingyuan Gong, Yang Chen, Atte Oksanen, and Aaron Yi Ding. "Structural hole theory in social network analysis: A review." *IEEE Transactions on Computational Social Systems* 9, no. 3 (2021): 724-739.

Mazza, Michele, Stefano Cresci, Marco Avvenuti, Walter Quattrociocchi, and Maurizio Tesconi. "Rtbust: Exploiting temporal patterns for botnet detection on twitter." In *Proceedings of the 10th ACM conference on web science*, pp. 183-192. 2019.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. "The spread of low-credibility content by social bots." *Nature communications* 9, no. 1 (2018): 1-9.

Wang, Youze, Shengsheng Qian, Jun Hu, Quan Fang, and Changsheng Xu. "Fake news detection via knowledge-driven multimodal graph convolutional networks." In *Proceedings of the 2020 international conference on multimedia retrieval*, pp. 540-547. 2020.

Wasserman, Stanley. "Social network analysis: Methods and applications." *The Press Syndicate of the University of Cambridge* (1994).

Wei, Feng, and Uyen Trang Nguyen. "Twitter bot detection using bidirectional long short-term memory neural networks and word embeddings." In *2019 First IEEE International conference on trust, privacy and security in intelligent systems and applications (TPS-ISA)*, pp. 101-109. IEEE, 2019.

Appendix

Machine Learning for Prediction:

We apply a Hypergraph Neural Network (HGNN) for binary node classification to distinguish between human and bot accounts, overcoming limitations of traditional graph-based methods which only capture the pairwise relationships but overlook group interaction. In our approach, we construct a hypergraph by connecting each user's followers (excluding themselves) via a single hyperedge, effectively capturing group behaviors within the social media network. The model uses user metrics (follower count, tweet count) and centrality measures (degree, betweenness, closeness) as features. By leveraging both the hypergraph structure and centrality features, this method enables the network to capture collective behavior and higher-order interactions that are often missed by traditional pairwise graph models.

The model shows a moderate performance, with a test accuracy of 0.5 and a ROC AUC score of 0.61. Although this suggests that the model is somewhat effective, there is considerable room for improvement. The limited features may not fully capture the complex interactions in the network, and incorporating additional relationships could enhance the model’s performance. Future work could include expanding the types of interactions considered, such as co-retweeting, co-hashtag usage, and co-liking, to capture a broader range of group behaviors and potentially improve bot detection accuracy.

Machine Learning for Causal Inference:

We apply a Regression Discontinuity (RD) design to investigate the impact of Twitter's subscription program on user engagement, using the 2000 follower threshold as the cutoff for eligibility. The outcome variable is tweet count, which serves as a measure of user engagement. Our analysis finds that eligibility for the subscription program significantly reduces tweet count by approximately 4935 tweets. Additionally, follower count has a positive effect on tweet count, but this effect weakens after eligibility for the subscription program. The interaction between subscription status and follower count is negative, indicating that the positive influence of followers diminishes once users qualify for the program.

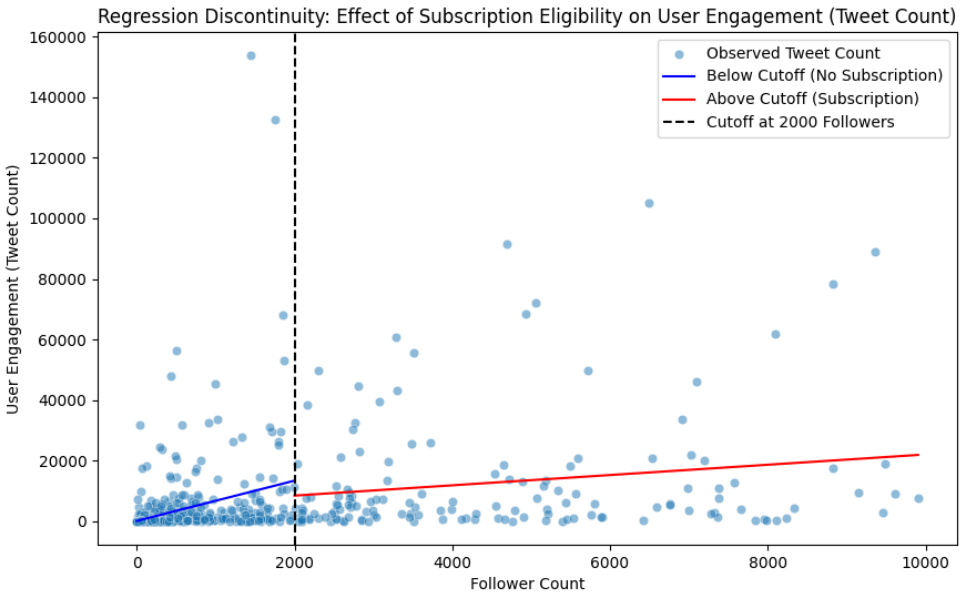

Figure A1. Plot of Regression Discontinuity Design

The application of the Regression Discontinuity (RD) design reveals several key findings regarding the impact of Twitter's subscription program on user engagement. Specifically, being eligible for the subscription program significantly reduces tweet count by approximately 4935 tweets, indicating that the program may discourage frequent tweeting. On the other hand, follower count has a positive effect on tweet count, with an increase of about 6.64 tweets per additional follower. However, the interaction term between subscription eligibility and follower count is negative, suggesting that the positive influence of follower count on tweet count weakens once users become eligible for the program. This suggests that the program may diminish the incentive for users to tweet frequently despite having a higher number of followers. The statistical model demonstrates significance, with an R-squared value of 0.156, indicating that the model explains a reasonable portion of the variation in tweet count. We hypothesize that eligibility for the program may reduce the incentive to tweet frequently, leading to a shift toward different forms of engagement, such as long threads or private messages. To mitigate the potential decrease in tweet count, we recommend adjusting the cutoff threshold, introducing incentives for continued engagement, and further investigating long-term effects on user behavior.