Cross-Subreddit Behavior as Open-Source Indicators of Coordinated Influence: A Case Study of r/Sino & r/China

Manon Pilaud[1][0009-0002-7586-6944] and Ian McCulloh[2][0000-0003-2916-3914]

[1] Johns Hopkins University, Baltimore, MD 21218, USA
mpilaud1@jhu.edu
[2] Johns Hopkins University, Baltimore, MD 21218, USA
imccull4@jhu.edu

Abstract. This study explores potential indicators of coordinated influence among users active in both r/Sino and r/China—two ideologically opposed Reddit communities focused on Chinese political discourse. Using topic modeling and sentiment analysis, we construct a user–topic sentiment matrix to compare individual sentiment patterns against broader community baselines. We then apply behavioral profiling to full user histories, incorporating features such as account age, posting frequency, lexical diversity, and karma distribution. Users exhibiting multiple behavioral anomalies are further analyzed within a subreddit co-participation network to assess structural positioning and cross-community influence. Our integrated approach reveals patterns of sentiment deviation, rhetorical consistency, and anomalous engagement that suggest strategically structured or inauthentic behavior. These findings demonstrate the value of combining content and activity-based signals for detecting influence operations in open-source environments.

Keywords: Bot Detection, China, Coordinated Messaging, Online Political Discourse, Reddit, Sentiment Analysis, Topic Modeling, OSINT.

1 Introduction

Online platforms increasingly serve as arenas for geopolitical contestation, where narratives about states, societies, and conflicts are shaped, diffused, and strategically manipulated. Among these platforms, Reddit occupies a unique position due to its semi-anonymous structure and community-driven moderation model. Its decentralized architecture fosters ideological diversity, but also creates fertile ground for covert influence operations. This paper examines two prominent Reddit communities—r/Sino and r/China—which represent ideologically divergent spaces focused on Chinese political discourse. While r/China often features critical perspectives and Western-aligned discourse, r/Sino promotes narratives consistent with pro-government and nationalist positions, reinforced through selective moderation and content curation.

This ideological divergence provides a valuable setting to examine cross-community influence behavior. Users who actively participate in both communities may serve as narrative bridges—or, alternatively, as vectors for coordinated messaging efforts aimed at shaping or reframing contested discourse. Identifying the behaviors, sentiment patterns, and thematic alignments of these users offers insight into how influence operations may manifest in open-source environments. Such research aligns with emerging frameworks in social cybersecurity [1], which emphasize the need to detect and understand influence campaigns that exploit sociotechnical systems.

We build upon existing work in social network analysis and information operations detection [1]-[11], applying a combined methodology of topic modeling, sentiment analysis, and behavioral profiling to users active in both subreddits. By integrating content- and activity-based signals, our approach contributes to the growing literature on open-source intelligence (OSINT) methods for identifying coordinated online behavior in politically sensitive domains. By integrating linguistic and behavioral indicators, this research contributes to the identification of potential coordinated influence behavior and to methodological efforts for open-source monitoring of politically sensitive online spaces.

2 Background

Reddit has been the subject of increasing scholarly attention as a platform for ideological polarization, echo chambers, and strategic messaging [8]-[9],[12]. Much of the current literature has focused on U.S.-centric political topics or health misinformation [13]-[14], with relatively limited attention paid to Reddit's role in geopolitical influence operations. In contrast, platforms like Twitter and Facebook have been the primary focus of studies examining state-sponsored disinformation and bot behavior [7],[10],[15]-[16].

Yet Reddit presents unique methodological opportunities—and challenges—for influence detection. Its semi-anonymous architecture, decentralized moderation, and long-form discourse differentiate it from other social platforms and make traditional bot detection heuristics less reliable [8]-[9],[12]. Previous efforts to detect inauthentic activity on Reddit have employed indicators such as posting frequency, temporal regularity, lexical redundancy, and karma asymmetry [8]-[9],[12],[17]. However, most of these studies analyze users within isolated subreddits or narrowly defined thematic domains, without considering cross-community behavioral coherence.

Our work addresses this gap by examining discourse and behavior across ideologically opposed communities. We draw on social cybersecurity theory [1] to frame the potential for coordinated influence campaigns that operate beneath the surface of overt

bot activity, using more nuanced strategies like tone modulation, narrative blending, and identity-driven engagement. This approach resonates with concepts from engagement manipulation and cognitive security articulated in military SNA contexts [9]-[11], where influence is operationalized not solely through automation, but through strategic manipulation of discourse networks and ecosystems.

We also position this work within the broader trajectory of computational propaganda research [18]-[19], which has highlighted the evolving nature of influence operations from crude spam campaigns to more sophisticated hybrid efforts involving both automated and human actors. Our focus on dual-subreddit users enables the detection of subtle alignment behaviors—users who may amplify specific narratives, maintain consistent sentiment profiles across topics, or exhibit stylistic irregularities that suggest non-organic participation.

By integrating topic and sentiment analysis with behavioral heuristics, we contribute a modular framework for identifying indicators of coordinated influence within Reddit. This study thus extends prior work in OSINT and influence detection by highlighting the importance of cross-community dynamics, affective profiling, and narrative evolution in the detection of strategic online messaging.

3 Methodology

We employed a multi-method approach to analyze discourse, sentiment, and behavioral patterns among users active in both r/Sino and r/China. Posts and comments were collected from r/Sino and r/China using the Reddit API, yielding 999 posts and 6,036 comments from r/Sino and 930 posts and 7,641 comments from r/China. Users with at least three contributions in each subreddit were retained, resulting in a sample of 63 dual-subreddit participants. To analyze discourse patterns, topic modeling was performed using Latent Dirichlet Allocation (LDA) on the dual-user corpus to identify dominant themes. A second LDA model trained on the full r/Sino and r/China dataset provided a comparative baseline for thematic alignment. Sentiment analysis was conducted using TextBlob, with post-level polarity scores aggregated by user and dominant topic. These scores were compared against both dual-user and global topic-level baselines to identify affective deviations. Behavioral profiling incorporated features such as account age, posting frequency, karma distribution, lexical diversity, and language detection. Users exhibiting two or more heuristic-based anomalies were flagged for further analysis. Finally, a subreddit co-participation network was constructed to assess the structural positioning of flagged users and evaluate potential cross-community influence patterns.

4 Findings

Out of 5,054 users active in either r/China or r/Sino, 63 met the criteria for dual participation. While most engaged sporadically, a small subset exhibited disproportionately high activity, particularly in r/Sino. To assess discourse patterns, we applied Latent Dirichlet Allocation (LDA) to two corpora: one from dual-subreddit users and one from the full population. The dual-user model revealed six thematically diffuse topics, blending informal rhetoric with ideological and geopolitical commentary. The all-user model displayed cohesive, topically segmented discussions, with themes like trade, propaganda, and diplomacy clearly separated. Tables 1 and 2 summarize the top keywords and thematic interpretations for each model.

Sentiment analysis revealed three notable trends among dual-subreddit users: 1) Positive outliers—users who expressed unusually favorable sentiment on topics typically discussed neutrally (e.g., trade); 2) Negative outliers—a larger group that consistently conveyed more negative sentiment across multiple topics; 3) Low-variance users—individuals who maintained a flat, neutral tone across all themes, potentially signaling rhetorical control or strategic moderation. These patterns suggest non-random

Table 1. Dual-User LDA Topic Summary

Topic	Top Keywords	Interpretation
Topic 1	"china","trump","country","government","american","chinese","tariff", "state", "united", "deal"	US-China relations during the Trump era with economic emphasis.
Topic 2	"china", "chinese", "world", "people", "would", "trade", "market", "money", "tariff", "american"	Focuses on China's role in the global economy, highlighting trade, markets, tariffs, and financial influence. Emphasizes economic interactions and consumer dynamics over political discourse.
Topic 3	"china","war","chinese","american", "people","like","propaganda", "usa", "life","support","taiwan", "america", "russia", "russian", "current"	Focuses on geopolitical conflict and propaganda, especially relating to China, the U.S., and Taiwan. Highlights themes of war, media narratives, and international tensions.
Topic 4	"china", "chinese", "european", "economic", "system", "democracy", "collapse", "europe", "state", "used"	Centers on ideological and political comparisons, particularly between China and Western democracies, including perceived economic, government, and societal strengths or weaknesses.
Topic 5	"chinese", "china", "american", "people", "trump", "taiwan", "company", "buy", "work", "believe"	Captures informal, opinion-driven discussions involving China, the U.S., and related political economic themes. Conversational language suggests social media or everyday discourse tone.
Topic 6	"china", "western", "chinese", "american", "taiwanese", "social", "government", "hypocrisy", "power", "standard"	Explores cultural, ideological, and social contrasts between China, the West, and Taiwan. Includes themes of governance, societal values, and critique of power structures and perceived hypocrisy.

Table 2. All-User LDA Topic Summary

Topic	Top Keywords	Interpretation
Topic 1	"china", "people", "dont", "like", "chinese", "know", "even", "really", "say", "care"	Captures broad, conversational discourse about China and its people, with informal tone and general sentiment.
Topic 2	"china", "trump", "war", "taiwan", "trade", "tariff", "military", "russia", "ukraine", "japan"	Focuses on U.S.–China geopolitical tensions, highlighting Trump, Taiwan, trade, and global security.
Topic 3	"china", "government", "western", "state", "democracy", "propaganda", "party", "communist", "freedom", "political"	Explores political ideologies and systems, contrasting China's government with Western democratic values.
Topic 4	"china", "market", "company", "product", "tariff", "manufacturing", "buy", "price", "industry", "export"	Focuses on manufacturing, trade, and economic products related to China and the global market.
Topic 5	"china", "city", "university", "student", "visit", "school", "visa", "foreigner", "shanghai", "beijing"	Captures study abroad, travel, and visa-related experiences, particularly in Chinese cities and schools.
Topic 6	"chinese", "video", "post", "news", "article", "propaganda", "reddit", "asian", "deleted", "watch"	Focuses on social media discourse, online content, and digital propaganda related to Chinese identity.

affective behavior, potentially aligned with efforts to shape narrative tone or introduce dissonance in ideologically charged threads.

Behavioral profiling identified 5 users with two or more anomalies across features such as account age, karma asymmetry, language irregularities, and lexical redundancy. The most common flag was low lexical diversity, found in over 80% of users. Additionally, two accounts had been suspended by Reddit, reinforcing concerns about possible inauthentic behavior. The distribution of these sentiment patterns is summarized in Table 3, which presents the number of users in each group, their average sentiment scores by topic, and the degree of deviation from the global topic averages. These values highlight how certain users consistently diverge—either positively, negatively, or through affective flattening—from the broader Reddit discourse, reinforcing their potential role in modulating tone within contested narratives.

The subreddit co-participation network (Figure 3) shows that flagged dual users are embedded across a diverse range of politically and economically salient communities. They do not cluster within niche forums but occupy bridging positions—connecting Chinese discourse to broader ecosystems like r/worldnews, r/technology, r/Economics, and r/Futurology. This structural footprint supports the hypothesis that some users serve as narrative brokers, amplifying or reframing content across domains of strategic interest.

Table 3. Summary of Sentiment Outlier Patterns by Topic Among Dual-Subreddit Users

Group	Topic	Num Users	Avg Sentiment	Global Avg Sentiment	Sentiment Deviation
Low Variance Users	0	3	-0.007	0.066	-0.073
Low Variance Users	1	1	-0.045	0.119	-0.164
Low Variance Users	2	2	0.001	0.060	-0.059
Low Variance Users	3	1	0.000	0.094	-0.094
Low Variance Users	4	1	-0.050	0.054	-0.104
Low Variance Users	5	4	0.011	0.081	-0.071
Neg. Sent. Outliers	0	7	-0.236	0.066	-0.302
Neg. Sent. Outliers	1	12	-0.016	0.119	-0.135
Neg. Sent. Outliers	2	5	-0.190	0.060	-0.250
Neg. Sent. Outliers	4	12	-0.213	0.054	-0.266
Neg. Sent. Outliers	5	6	-0.167	0.081	-0.249
Pos. Sent. Outliers	3	7	0.427	0.094	0.333

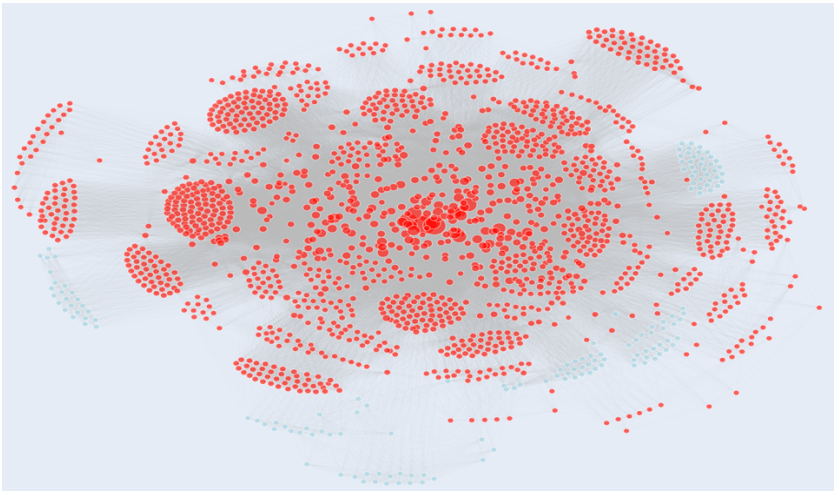

Figure 3. Network visualization of full sample of 1,819 subreddits. Subreddits with flagged user activity indicated in red.

5 Conclusion and Future Work

This study investigated user behavior across r/Sino and r/China—two ideologically opposed Reddit communities engaged in discourse about China—to surface indicators of coordinated inauthentic behavior. By focusing on users who participated in both communities, we identified a subset exhibiting linguistic, affective, and behavioral anomalies that deviate from the general Reddit population. These anomalies included topic-level sentiment shifts, unusually consistent or divergent affective profiles, and metadata patterns suggestive of non-organic engagement.

Our results indicate that dual-subreddit users may not simply act as ideological bridges, but could function as strategic amplifiers—modulating tone, reinforcing narratives, and diffusing discourse across communities. When considered in combination with subreddit co-participation patterns and engagement distribution, the findings suggest the possibility of engagement manipulation, where user behaviors (e.g., post timing, tone modulation, karma distribution) are leveraged to game platform dynamics and influence the visibility or perceived legitimacy of content. This aligns with broader concerns in social cybersecurity and computational propaganda research, which emphasize the evolving sophistication of influence campaigns that blend automated tools with coordinated human behavior.

The modular framework presented here—combining topic modeling, sentiment profiling, and heuristic-based behavioral analysis—offers a scalable approach for identifying suspicious patterns in open-source environments. It contributes both empirically and methodologically to the detection of subtle, persona-driven influence activity that may escape traditional bot-detection systems.

Future work should incorporate temporal dynamics, enabling the tracking of narrative evolution and user behavior over time to distinguish organic shifts from coordinated strategy. In addition, the flagged users identified in this study could serve as weak labels for developing supervised classifiers to detect coordinated inauthentic behavior more systematically. Integrating structural features (e.g., karma ratios), affective signatures (e.g., sentiment deviation from topic norms), and co-engagement networks could support the development of robust, generalizable detection pipelines.

As Reddit and similar platforms continue to shape geopolitical discourse, the ability to detect cross-community influence operations remains critical for platform integrity, public trust, and the broader aims of open-source intelligence.

References

1. Carley, K.M.: Social cybersecurity: an emerging science. Computational and Mathematical Organization Theory 26(4), 365-381 (2020)
2. Ng, L.H.X., Carley, K.M.: The dual personas of social media bots. arXiv2504.12498. (2025)
3. Ng, L.H.X., Carley, K.M.: A global comparison of social media bot and human characteristics. Scientific Reports, 15(1) (2025)

4. Marigliano, R. Ng, L.H.X., Carley, K.M.: Analyzing digital propaganda and conflict rhetoric: a study on Russia's bot-driven campaigns and counter-narratives during the Ukraine crisis. J. of Social Network Analysis and Mining 14(1) (2024).
5. Carragher, P., Williams, E.M., Carley, K.M.: Detection and discovery of misinformation sources using attributed webgraphs. In: Proc. Of Int'l AAAI Conference on Web and Social Media 18 (2024).
6. Kenny, R., Fischhoff, B., Davis, A., Carley, K.M., Canfield, C.: Duped by bots: why some are better than others at detecting fake social media personas. Human Factors 66(1) 88-102 (2024).
7. Ng, L.H.X., Carley, K.M.: Deflating the Chinese balloon: types of Twitter bots in US-China balloon incident. EPJ Data Science 12(1) (2023)
8. Lowetz, C., McCulloh, I.: Russian Invaders on the Internet's Front Page – A Survey of Behaviors in Ukraine-Related Subreddits. In: Proc. 2024 IEEE/ACM Foundations of Open Source Intelligence and Security Informatics. (2024)
9. McCulloh, I., Bergamini, R., Mackey, C.: Echoes of War: How Reddit Narratives Shape Sectarian Views of the Israel-Hamas War. In Proc. 2024 IEEE/ACM Foundations of Open Source Intelligence and Security Informatics. (2024)
10. Takacs, R., McCulloh, I.: Dormant bots in social media: Twitter and the 2018 US Senate election. In: Proc 2019 IEEE/ACM ASONAM (2019)
11. McCulloh, I., Armstrong, H., Johnson, A.N.: Social Network Analysis with Applications. Wiley. (2013)
12. Albanese, J., Morgan, D., Goldberg, D.: Detecting bots on Reddit using posting behavior and content features. In: Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), pp. 1–8. AAAI Press (2020).
13. Cinelli, M., Morales, G.D.F., Galeazzi, A., Quattrociocchi, W., Starnini, M.: The echo chamber effect on social media. Proc. Natl. Acad. Sci. 118(9), e2023301118 (2021). https://doi.org/10.1073/pnas.2023301118
14. Starbird, K., Arif, A., Wilson, T.: Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. Proc. ACM Hum.-Comput. Interact. 3(CSCW), 127 (2019). https://doi.org/10.1145/3359229
15. Varol, O., Ferrara, E., Davis, C.A., Menczer, F., Flammini, A.: Online human-bot interactions: detection, estimation, and characterization. ICWSM (2017)
16. Zannettou, S., Caulfield, T., Setzer, W., Blackburn, J.: Who let the trolls out? Towards understanding state-sponsored trolls. In: Proceedings of the 10th ACM Conference on Web Science (WebSci '19), pp. 353–362. ACM, Boston (2019). https://doi.org/10.1145/3292522.3326042
17. Kumar, S., Hamilton, W.L., Leskovec, J., Jurafsky, D.: Community interaction and conflict on the web. In: The Web Conference 2018 (WWW '18), pp. 933–943. ACM, Lyon (2018). https://doi.org/10.1145/3178876.3186141.
18. Keller, T.R., Klinger, U.: Social bots in election campaigns: Theoretical, empirical, and methodological implications. J. Political Communication 36(1) 171-189 (2019)
19. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. J. Communications of the ACM 59(7) 96-104 (2016)
20. Cresci, S.: A decade of social bot detection. Commun. ACM 63(10), 72–83 (2020). https://doi.org/10.1145/3363184