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

Echo Chambers within the Russo-Ukrainian War: The Role of Bipartisan Users

Peixian Zhang, Ehsan-Ul Haq, Yiming Zhu,* Pan Hui,* and Gareth Tyson
Hong Kong University of Science and Technology (GZ)
Email: pzhang041@connect.hkust-gz.edu.cn {euhaq, yzhucd}@connect.ust.hk {panhui, gtyson}@ust.hk

***Abstract*—The ongoing Russia-Ukraine war has been extensively discussed on social media. One commonly observed problem in such discourse is the emergence of echo chambers, where users are rarely exposed to opinions outside their own worldview. Prior literature on this topic has assumed that such users hold a single consistent view. However, recent work has revealed that complex topics often trigger bipartisanship among certain people. With this in mind, we study the presence of echo chambers on Twitter related to the Russo-Ukrainian war. We measure their presence and identify an important subset of bipartisan users who vary their opinion during the invasion. We then explore the role they play in the communications graph and their impact on echo chambers.**

***Index Terms*—Russo-Ukrainian war, Echo chambers, Twitter**

I. Introduction

With the increasing influence of social media on public discourse, individuals' opinions are gaining growing visibility [1, 2]. On the 24^{th} February 2022, Russia invaded Ukraine. The invasion has become the subject of significant online debate and has evoked diverse and polarized opinions [3]. Prior research demonstrates that opinions on the conflict vary across regions, with Western Europe and the United States holding different views from those of Eastern Europe and Asian countries [4]. One particular concern is that these trends can result in echo chambers [5, 6], segregating people with opposing stances. Echo chambers generally refer to individuals or groups predominantly interacting with those with similar viewpoints, reinforcing and amplifying their existing stances. This has been shown to create numerous societal issues [7], *e.g.* the attitude to refugees from Ukraine [8].

Prior literature has analyzed users' stances by exploring echo chambers on social media during specific events, *e.g.* the 2018 Brazilian Presidential election [9, 10], US Presidential election [11] and COVID-19 [2, 12]. However, closer inspection reveals that many users do not hold a consistent single stance [2, 12]. Building on this prior work, we aim to explore the presence of echo chambers in online discourse about the Russian invasion of Ukraine. Additionally, we investigate whether users display inconsistencies in their polarity as pro-Russia or pro-Ukraine. Specifically, we explore two research questions (RQs):

RQ1: Are there *bipartisan users* who exhibit both pro-Russia and pro-Ukraine polarity? Bipartisan users are found in various election discussions [5]. We explore how such users differ from those with a consistent partisan polarity.

RQ2: Do *bipartisan users* help bridge polarized communities and decrease the prominence of echo chambers? We study this because partisan users are demonstrated to increase their prominence in the masses by being more partisan [13].

II. Related Work

Polarization on Social Media: Prior work has focused on polarization, bipartisan users, and the properties associated with echo chambers [14, 15]. The appearance of echo chambers can lead to several issues that undermine users' communications. These include rumor cascades [16], fake/partisan news propagation [17, 18], and hate speech [19]. Moreover, several researchers have observed that political echo chambers emerge on mainstream social media [20]. This could result in a surge of propaganda and partisan content among online communities [21]. For this, Garrett and Kelly look at how users' news consumption relates to the appearance of echo chambers in online political news sharing [6]. The authors show that an awareness of the echo chamber's early-stage formation can help administrators reduce the spread of extreme ideologies.

Detection of Echo Chambers: There has also been work looking at the detection of echo chambers. Typically, a combination of features are used for classification, including textual features (*e.g.* tweet text and hashtags [22]), as well as social feedback like retweets and mentions [10]. Other properties can be used to detect echo chambers, *e.g.* homophily [23] and social influence [24]. Some researchers have treated the detection of echo chambers as a network influence estimation problem, leveraging the network structure to identify communities with echo-chamber characteristics [5, 25]. There has also been work exploring the role of online news within echo chambers [26].

Our Contribution: In contrast to prior work, we focus on *bipartisan users* in a specific domain: the Russo-Ukrainian

*Yiming Zhu is also with The Hong Kong University of Science and Technology, and Pan Hui is also with The Hong Kong University of Science and Technology and University of Helsinki.

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

ACM ISBN 979-8-4007-0409-3/23/11…$15.00
http://dx.doi.org/10.1145/3625007.3627475

war. Although users who share inconsistent information are explored in prior work [2, 5, 12], this has not focused on echo chambers. Further, prior work often focuses on a single country. This means the topic usually pertains to politics and elections. In contrast, the global discourse surrounding the Russo-Ukrainian War has not yet been studied in depth.

III. Dataset and Annotation

A. Tweet Dataset

We utilize the Twitter dataset on the Russo-Ukrainian war from [27], which is collected with war-related keywords using the Twitter Streaming API. We utilize a subset of features provided by this dataset – tweet id, username, tweet text, number of retweets, number of likes, and the ID of referenced tweet (the tweet that is retweeted by current tweet).[1]

We extract the 16,889,957 English language tweets from the dataset (from 337,302 unique users). 1.22% of these are from verified accounts. The tweets cover a period of the first ten days of the war starting from 24^{th} February to 5^{th}, March 2022. We observe a long-tail distribution of the number of tweets per user ($min = 1, max = 1858, \mu = 4.38, mid = 1$). Out of the total tweets, 7.42% are original tweets, and 78.94% are retweets. The remaining tweets consist of quotes and replies.

B. Manual Annotation of Tweets

We next label a subset of tweets with their polarity. We use the term *polarity* to refer to the stance of the tweet's author. We categorize tweets into one of three polarity labels – pro-Russia, pro-Ukraine, or not-sure. To select tweets that mention Russia and Ukraine, we first search for tweets in our dataset that contain any relevant keywords or hashtags: "Russia", "Ukraine", #IStandwithRussia, #StopRussia, #IstandwithPutin and #RussiaUkraineWar. We then randomly sample 2,205 tweets to manually annotate their polarity in text. However, we observe that not all tweets directly state their polarity towards Russia or Ukraine. Instead, they might indicate their leaning based on other entities, such as politicians. For instance, here is a pro-Russia tweet indicating: *"Dear my president, president Vladimir Putin, keep strong protect your nation against the evil NATO and America! #IstandwithPutin"*. To capture this, we also consider users' stance (pro or anti) towards relevant entities mentioned in the text. For example, "pro Putin", "anti NATO", and "anti US" are assigned to the above example tweet, and our final decision for its stance is pro-Russia. We use the following entity-related attitudes to annotate tweets' polarity:

- ***Pro-Russia*:** pro Putin, anti Biden, anti US, anti Trump, anti Lukashenko, anti Carlson, pro Russia(n), anti Kamala, anti Ukraine, anti Ukrainian, anti Zelensky, anti NATO, anti GOP (Grand Old Party), anti Democrats, anti West
- ***Pro-Ukraine***: pro US, pro Zelensky, pro Ukraine, pro Ukrainian, pro Biden, anti Putin, anti Russia(n), anti Oligarch, anti Belarus, pro Trump.

Overall, the manually annotated data includes 539 tweets labeled as *pro-Russia*, 938 tweets labeled as *pro-Ukraine*, and 728 tweets labeled as *not-sure*.

IV. Quantifying Polarity

We next use our annotated dataset to train a classifier that predicts the polarity of the rest of the tweets in the dataset.

A. Predicting Tweet Polarity

We experiment with five commonly used text embedding methods – BERT [28], Sentence Transformer [29], Universal Sentence Encoder [30], Word2Vec [31], and TF-IDF [32] to train several classifiers to select the best performing one. We preprocess the tweet text before applying the embedding methods. Our process involves: (*i*) Removing any mentions for users (`@{username}`) or retweeting (`RT@{username}`). (*ii*) Removing all hyperlinks and emojis. (*iii*) Removing all non-alphanumeric characters, including punctuation and special symbols (*e.g.* #, @, $).

We then use six commonly used machine learning algorithms (SVM, KNN, Decision Tree, Random Forest, Naive Bayes, and Logistic Regression) [33] and train multiple classifiers using the above-mentioned text features. We use grid search to identify the best hyper-parameters [33]. We evaluate each combination using 5-fold cross-validation. The SVM classifier with Sentence Transformer achieves the highest performance (F1-Score = 0.70). Finally, we select this classifier to predict the rest of the tweets' polarity in our dataset. The classifier outputs probabilities for the three labels (*pro-Russia*, *not-sure*, *pro-Ukraine*), where each probability represents the likelihood of a tweet leaning towards the corresponding polarity.

B. Quantifying Users' Polarity

We then quantify users' polarity based on their tweets' polarity. We first encode tweets' polarity into numerical notations: *pro-Ukraine* $= 1$, *not-sure* $= 0$, and *pro-Russia* $= -1$. Given a tweet t, we quantify its polarity s_t as: $s_t = \sum_l p_l * l$, where l denotes the polarity label ($l \in [-1, 0, 1]$) and p_l denotes the probability output by the classifier for polarity l. Following this, for a user u, we quantify the user's polarity g_u as the average of the tweets' polarity posted by this user:

$$g_u = \frac{\sum_{i=1}^{n} s_{t_i}}{n}$$

where $t_i (i \in [1, n])$ denotes the tweet posted by u.

When g_u is close to 1, it indicates that the user is pro-Ukraine, and a score close to -1 indicates that the user is more pro-Russia. The polarity distribution of all users ranges from -0.99 to 0.97, with a mean of 0.27 and a median of 0.288.

V. Results & Analysis

The above dataset and classifier allow us to estimate the polarity of each user. We next exploit our data to answer our research questions.

[1]https://developer.twitter.com/en/docs/twitter-api/tweets/lookup/api-reference/get-tweets

A. RQ1: Presence of bipartisan users

We first investigate if users share tweets that include both pro-Ukraine and pro-Russia material.

Identifying Bipartisan Users: We utilize the labels obtained from the classifier to categorize the users as pro-Russia or pro-Ukraine. We focus on users who have submitted multiple tweets (> 1), as users with only one tweet inherently hold a consistent attitude. For analysis, we classify the users into two groups below:

- **Bipartisan Users:** A user is classified as *bipartisan user* if user has at least 20% pro-Russia and at least 20% pro-Ukraine tweets. 20% is the minimum thresholds to distinguish *bipartisan users* and partisan users in several topics [5]. We choose these conservative thresholds to minimize the impact of mis-classification errors.
- **Partisan Users:** A user is classified as *partisan* if only pro-Russia or pro-Ukraine tweets cover more than 20% of a user's total tweets individually. A partisan user is further categorized into *pro-Russia group* or *pro-Ukraine group*.
- **Not Sure:** If a user is neither a *bipartisan* nor a partisan, this user is classified as *not sure*.

Results: We observe users exhibiting bipartisan polarity, including both pro-Russia and pro-Ukraine tweets. Our dataset contains 130,170 bipartisan users, 65,380 pro-Russia users and 1,276,671 pro-Ukraine users. We note that bipartisan users are the second largest in the dataset, covering 8.84% users and 4.41% tweets. In addition, we also observe a strong bias towards pro-Ukraine among Twitter users, where the pro-Ukraine group consists of 86.72% users and 94.23% tweets. Pro-Russia group is a minor class, constituting only 4.44% users and 1.37% of tweets. Moreover, we find that the pro-Ukraine group (1.64%) contains more verified users than pro-Russia group (0.61%) and bipartisan users (0.66%).

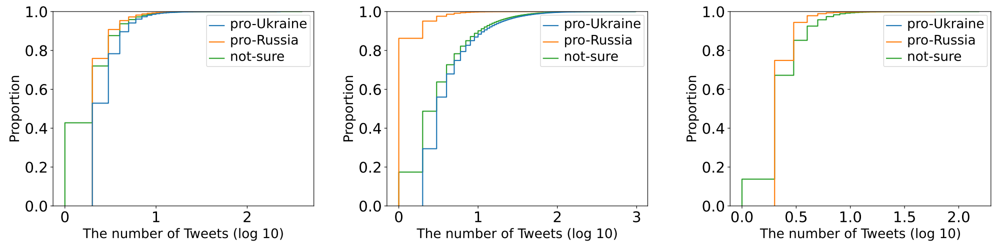
After detecting these bipartisan users, we inspect the difference in posting behavior among different user groups. Thus, we examine the distribution of tweets' polarity within each group. Figure 1 displays the cumulative distribution function (CDF) for the number of tweets with polarity in the three groups. The x-axis shows the logarithm base 10 of the number of tweets$+1$. Figure 1a illustrates that for bipartisan users, pro-Ukraine tweets have the highest frequency, followed by not-sure tweets, while pro-Russia tweets have the lowest frequency. Moreover, we find that the distribution of tweets' polarity among bipartisan users appears to be distinct from those in other groups by posting relatively more not-sure tweets (25.71%). Interestingly, we also notice in Figure 1b that 13.76% of pro-Ukraine users have pro-Russia tweets. However, none of the pro-Russia users have ever posted pro-Ukraine tweets in Figure 1c. Furthermore, based on our dataset, while some pro-Ukraine users share tweets with opposing polarity, none of the pro-Russia users exhibit similar activities.

B. RQ2: Impact of bipartisan users on echo chambers

We next investigate if bipartisan users mitigate the prominence of echo chambers. To study this, we remove such users from the interaction graph and study the impact on echo chambers.

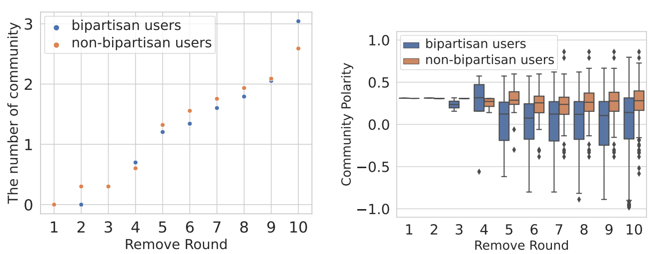
Impact of Removing Bipartisan Users: We first take the largest community in the retweet graph detected by the Louvain algorithm with a resolution of 0.1. This contains 68.30% of the nodes in the network of the retweet network. The largest community consists of 79.15% pro-Ukraine users, 2.68% pro-Ukraine users, 6.96% bipartisan users, and the remaining portion comprising not sure users. For a baseline comparison, we select a control group consisting of non-bipartisan users. For a given user in bipartisan users, we take the degree of the user (node) and find another user in the non-bipartisan users who has the closest matching degree. Out of all bipartisan users, only 11 (0.01%) users cannot find non-bipartisan users who have the same degree. The largest difference in degree between a bipartisan user and a non-bipartisan user is 25, where the bipartisan node has a degree of 4504. We use these non-bipartisan users for comparison against bipartisan users to characterize the role of the latter in polarized communities. We compute the average polarity of each community determined as the average individual polarity of the users in the community [34] and take communities with absolute polarity larger than 0.5 as polarized communities. In order to understand whether bipartisan users help bridge polarized communities, we next take the network and systematically remove a set of users from the largest community. We do so separately for the non-bipartisan nodes and bipartisan nodes. We start by removing nodes in ascending order of degree, in deciles upon each iteration. This process is repeated for ten iterations for each group until all the nodes are removed. Upon each iteration, we check whether the retweet network contains more echo chambers based on the homophily of communities. Specifically, we assess whether a new community emerges as separate from the existing community and whether this new community exhibits polarity close to 1 and -1. We use the same community detection approach (Louvain method with the same resolution) and echo chamber analysis as before.

Results: Figure 2 presents the progressive outcomes following each round of node removal. Figure 2a displays the number of communities in the network after each removal step. The y-axis, on a logarithmic scale base 10, quantifies the number of communities at each removal stage. Figure 2b also shows the distribution of community polarity resulting from the removal of nodes. As nodes are removed from the graph, the communities become more fragmented. Following the removal of bipartisan nodes, the largest community contains 92.93% of the total nodes, and the remaining nodes are divided into 1,229 communities, with a proportion of 98.86% consisting of single-node communities. After the removal of non-bipartisan nodes, the largest community contains 93.01% of nodes and the remaining nodes are partitioned into 404 communities, with a proportion of 99.75% communities only having one



(a) *bipartisan Users* (b) *pro-Ukraine group* (c) *pro-Russia group*

Fig. 1: Display the cumulative distribution function for the number of polarity of tweets in the three groups, respectively. The x-axis shows the logarithm base 10 of the number of tweets, where each value is incremented by 1 to mitigate the impact of zero.



(a) *The number of community* (b) *The distribution of community polarity*

Fig. 2: The result of the nodes removing with 10% of the nodes being removed each time in ascending order of degree are considered from the number of communities and the distribution of the community polarity. The x-axis of 2 and 2a represent the round of removal. The y-axis of 2a represents, on a logarithmic scale base 10 quantifies the number of communities at each removal stage.

node. After removing the bipartisan nodes, the resulting communities show a clearer trend, including more polarized communities. This results in those smaller communities being identified. Following the removal of all bipartisan nodes, the resulting smaller communities consist of 30.68% pro-Russia users, 57.13% pro-Ukraine users, and the remaining portion comprising not sure users. However, the result of the non-bipartisan nodes, the resulting smaller communities consist of 5.40% pro-Russia users, 66.32% pro-Ukraine users, 6.94% bipartisan users, and the remaining portion comprising not sure users. The removal of bipartisan nodes leads to increased fragmentation of pro-Russia communities, as more pro-Russia nodes form new communities.

The polarity of communities indicates echo chambers, characterized by frequent interactions within the community and the sharing of similar polarity. The above results suggest that the presence of bipartisan nodes do play a role in preventing the emergence of polarized communities. Removing bipartisan nodes leads to an increase in the number of communities exhibiting polarity close to 1 and -1. We argue that the bipartisan nodes, therefore, induce the number of polarized communities, indicating that the existence of *bipartisan users* decreases the prominence of echo chambers.

VI. Conclusion

We have explored the presence of bipartisan users and echo chambers in online discussions related to the Russo-Ukrainian war. We detected a group of users who are bipartisan, who share information showing both pro-Russia and pro-Ukraine polarity. We investigated whether these users might mitigate the echo chamber effect. The result shows that, compared with the control group, removing bipartisan nodes from the retweet network leads to more communities with extreme polarity, suggesting that the bipartisan nodes connect the clusters of consistent users together to decrease the echo chamber.

A key limitation of our analysis is the inherent inaccuracies of the classifier used to categorize tweets. Naturally, inaccuracies in the classifier may mis-label a user as bipartisan. We tuned the thresholds to mitigate this but, nevertheless, this is a problem that must be revisited in future work. Our dataset is also limited to two weeks. We conjecture that a longer timeframe may expose greater bipartisanship, as users may vary their opinions with time.

References

[1] J. P. Chang, J. Cheng, and C. Danescu-Niculescu-Mizil, “Don’t let me be misunderstood: Comparing intentions and perceptions in online discussions,” in *WWW*, 2020.

[2] D. Naskar, S. R. Singh, D. Kumar, S. Nandi, and E. O. d. l. Rivaherrera, “Emotion dynamics of public opinions on twitter,” *ACM TOIS*, vol. 38, no. 2, pp. 1–24, 2020.

[3] S. Sazzed, “The dynamics of ukraine-russian conflict through the lens of demographically diverse twitter data,” in *IEEE Big Data 2022*, 2022, pp. 6018–6024.

[4] H. Vahdat-Nejad, M. G. Akbari, F. Salmani, F. Azizi, and H.-R. Nili-Sani, “Russia-ukraine war: Modeling and clustering the sentiments trends of various countries,” *arXiv preprint arXiv:2301.00604*, 2023.

[5] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis, “Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship,” in *Proceedings of WWW 2018*, 2018, pp. 913–922.

[6] R. K. Garrett, “Echo chambers online?: Politically motivated selective exposure among internet news users,” *Journal of computer-mediated communication*, 2009.
[7] E.-U. Haq, G. Tyson, T. Braud, and P. Hui, “Weaponising social media for information divide and warfare,” in *ACM Hypertext (HT)*, 2022.
[8] S. Wildemann, C. Niederée, and E. Elejalde, “Migration reframed? a multilingual analysis on the stance shift in europe during the ukrainian crisis,” in *Proceedings of WWW 2023*, 2023, p. 2754–2764.
[9] R. Recuero, F. B. Soares, and A. Gruzd, “Hyperpartisanship, disinformation and political conversations on twitter: The brazilian presidential election of 2018,” in *AAAI ICWSM*, 2020.
[10] F. B. Soares, R. Recuero, and G. Zago, “Asymmetric polarization on twitter and the 2018 brazilian presidential elections,” in *SM&S*, 2019.
[11] V. R. K. Garimella and I. Weber, “A long-term analysis of polarization on twitter,” in *AAAI ICWSM*, 2017.
[12] S. Dash, D. Mishra, G. Shekhawat, and J. Pal, “Divided we rule: Influencer polarization on twitter during political crises in india,” in *AAAI ICWSM*, 2022.
[13] M. Grömping, “‘echo chambers’ partisan facebook groups during the 2014 thai election,” *Asia Pacific Media Educator*, 2014.
[14] E. U. Haq, T. Braud, Y. D. Kwon, and P. Hui, “A survey on computational politics,” *IEEE Access*, 2020.
[15] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, “Political polarization on twitter,” in *AAAI ICWSM*, 2011.
[16] X. Wang and Y. Song, “Viral misinformation and echo chambers: The diffusion of rumors about genetically modified organisms on social media,” *IR*, 2020.
[17] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, and Y. Liu, “Combating fake news: A survey on identification and mitigation techniques,” *ACM TIST*, 2019.
[18] E.-U. Haq, Y. K. Lu, and P. Hui, “It’s all relative! a method to counter human bias in crowdsourced stance detection of news articles,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2, 2022.
[19] A. Bagavathi, P. Bashiri, S. Reid, M. Phillips, and S. Krishnan, “Examining untempered social media: analyzing cascades of polarized conversations,” in *ASONAM*, 2019.
[20] N. Gillani, A. Yuan, M. Saveski, S. Vosoughi, and D. Roy, “Me, my echo chamber, and i: Introspection on social media polarization,” in *ACM WWW*, 2018.
[21] S. Phadke, M. Samory, and T. Mitra, “What makes people join conspiracy communities? role of social factors in conspiracy engagement,” *Proceedings of the ACM CSCW*, 2021.
[22] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, “Modeling echo chambers and polarization dynamics in social networks,” *Physical Review Letters*, vol. 124, no. 4, p. 048301, 2020.
[23] F. Cinus, M. Minici, C. Monti, and F. Bonchi, “The effect of people recommenders on echo chambers and polarization,” in *AAAI ICWSM*, 2022.
[24] K. Sasahara, W. Chen, H. Peng, G. L. Ciampaglia, A. Flammini, and F. Menczer, “Social influence and unfollowing accelerate the emergence of echo chambers,” *Journal of Computational Social Science*, 2021.
[25] S. Kumar, W. L. Hamilton, J. Leskovec, and D. Jurafsky, “Community interaction and conflict on the web,” in *ACM WWW*, 2018.
[26] Y. Jeon, B. Kim, A. Xiong, D. Lee, and K. Han, “Chamberbreaker: mitigating the echo chamber effect and supporting information hygiene through a gamified inoculation system,” *ACM CSCW*, 2021.
[27] E.-U. Haq, G. Tyson, L.-H. Lee, T. Braud, and P. Hui, “Twitter dataset for 2022 russo-ukrainian crisis,” *arXiv preprint arXiv:2203.02955*, 2022.
[28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.
[29] N. Reimers and I. Gurevych, “Sentence-bert: Sentence embeddings using siamese bert-networks,” *arXiv preprint arXiv:1908.10084*, 2019.
[30] C. et al., “Universal sentence encoder for english,” in *EMNLP 2018 System Demonstrations*, 2018.
[31] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.
[32] C. D. Manning, *An introduction to information retrieval*. Cambridge university press, 2009.
[33] P. et al., “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, 2011.
[34] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, “The echo chamber effect on social media,” *PNAS*, 2021.