Evaluating Content Exposure Bias in Social Networks

1st Nathan Bartley
Information Sciences Institute
Marina Del Rey, United States
nbartley@isi.edu

2nd Keith Burghardt
Information Sciences Institute
Marina Del Rey, United States

3rd Kristina Lerman
Information Sciences Institute
Marina Del Rey, United States

***Abstract*—Online social platforms employ personalized feed algorithms to gather and prioritize messages from accounts followed by users, which distorts content's perceived popularity prior to personalization. We call this "exposure bias," and our research focuses on quantifying it using diverse exposure bias metrics, and we evaluate recommendation algorithms through various content ranking heuristics. Similarly we simulate activity in a network to assess the influence of such ranking heuristics on exposure bias. Furthermore, we are working on agent-based model simulations to comprehend the impact of ranking schemes, with the ultimate goal of exploring intervention effects over time. Our empirical findings reveal that users exposed to popularity-based feeds experience significantly lower exposure bias compared to chronologically-ordered feeds.**

***Index Terms*—Auditing, Exposure Bias, Recommendations**

I. Introduction

Online social networks (OSNs) play a pivotal role in information dissemination, encompassing a wide spectrum of knowledge and news, from innovations to job opportunities, and even during emergencies like natural disasters [1]–[3]. However, the burgeoning popularity of OSNs has led to an unprecedented surge in user-generated content, resulting in information overload. To manage this deluge, OSNs have built recommendation systems to curate personalized feeds for users, wherein content is reordered to optimize various metrics, e.g., engagement and user time spent on the platform. While personalized feeds offer significant advantages, recent evidence suggests that recommendation systems may inadvertently constrict the range of information users encounter, potentially fostering echo chambers [4] and intensifying partisanship [5], [6].

A frequently overlooked facet of personalized feeds is their potential to skew users' perception of content popularity. For instance, a user follows a set of accounts where some may be much more active than others, leading to a skewed inventory of possible tweets to serve a user. With a skewed inventory, there will likely be an over-representation of these highly-active users' opinions in a user's feed. Moreover, users themselves differ in when and how much time they allocate to social media, resulting in varying degrees of content consumption during their sessions. These disparities, coupled with the recommendation algorithm's content ordering within the feed, invariably impacts the information users encounter. We introduce the term "content exposure bias" or simply "exposure bias" to describe these potential feed distortions stemming from the content posted by followed accounts. In prior research, we delineated and quantified this bias within the context of an algorithmic sock-puppet audit [7].

This report describes a series of projects addressing the following research questions:

RQ1. Can we observe exposure bias in empirical online social networks like Twitter?
RQ2. Do different feed recommendation algorithms affect exposure bias?
RQ3. Do different session lengths affect the level exposure bias?
RQ4. Do interventions in recommender algorithms change the level of exposure bias over time in a simulated online social network?

II. Related Work

A. Cognition & Social Networks

In the realm of cognitive sciences, extensive research has illuminated the prevalence of cognitive biases in human decision-making. One particularly intriguing bias is the salience bias, which compels individuals to pay more attention to surprising or unexpected stimuli. In social psychology this salience bias often leads to the heightened perception of minority groups, resulting in an overestimation of their prominence. This phenomenon closely resembles a statistical bias known as the "majority illusion," which emerges within network structures, distorting the visibility of uncommon opinions [8]. This highlights a connection between perceptual biases and the dynamics of social networks, namely information diffusion. Past studies have unveiled the role of repeated stimulus exposure in fostering simple contagion, the significance of high-degree nodes in facilitating information spread, and the impact of cognitive overload on information transmission [9]–[11]. While these insights are valuable, the current research

ASONAM '23, November 6-9, 2023, Kusadasi, Turkey
© 2023 Association for Computing Machinery.
ACM ISBN 979-8-4007-0409-3/23/11…$15.00
http://dx.doi.org/10.1145/3625007.3627724

endeavors shift focus toward examining different feed ”queue” strategies and exploring the intricacies of information exposure in session-based social networks.

Building upon these insights, recent work in the statistical and theoretical domains has further expanded our understanding. Experiments were carried on with complex contagion model of information diffusion on Twitter data, demonstrating that repeated exposures from diverse sources offer a more comprehensive explanation of Twitter behavior than simplistic contagion models [12]. Meanwhile, other lines of theoretical work explored a graph theoretical perspective to address the majority illusion in social networks, including grappling with the NP-hardness of both its identification and elimination, and the classification of various types of networks that experience the effect [13], [14]. In the current set of studies, the focus shifts toward partial observations within session-based social networks and the upstream facets of information diffusion, particularly the dynamics of information exposure.

B. Algorithmic Audits of Online Social Networks

Since the pioneering work by Sandvig, Hamilton, Karahalios, & Langbort in 2014 there has been a growing interest in algorithmic audits, where researchers investigate and probe online algorithms for potential discriminatory behaviors, spanning various sectors, from e-commerce platforms [15] to search engines [16] and online social networks (OSNs) [7], [17]–[19]. On Youtube, Tomlein et al., 2021 evaluated a ”filter bubble popping” strategy to reconfigure recommended videos [20]. Although our work doesn’t pertain to YouTube, incorporating exposure biases alongside measures like Search Engine Results Pages (SERP) could complement their analyses. Similarly, Beattie, Taber & Cramer, 2022 examined social ”Who to Follow” recommendations on Twitter, proposing a method to break echo chambers through user embeddings, offering valuable intervention tools [21]. In contrast, Ramaciotti et al., 2021 employed simulations and empirical Twitter data to examine different recommender systems’ effects on the social graph. Through this they identified the co-evolution of the social graph and users’ ideological positions [22]. Our work is more focused on user activity and connecting it to their positions within the social graph.

In a study related to ours, Huszár et al., 2022 investigated the algorithmic amplification of different countries’ political parties on Twitter [23]. They considered different global sets of users and relevant political tweets that those users left lingering impressions on, whereas our concern primarily revolves around the ratio of exposure within an individual user’s feed.

C. Practical Concerns

Previous work has been carried out by Twitter itself into its own content recommendation system [24], detailing both the components of the system and how they measure the behavior of the system overall [25]. Lazovich et al., 2022 in particular outlines different criteria practitioners might use to decide how to measure bias within their own recommendation systems. Our work differs in that this work does not choose metrics that explicitly account for the social graph structure of the system: the metrics instead focus on tweets being categorized as ”in-network” or ”out-of-network”. However, the criteria of adjustability, scale invariance, and interpretability are relevant for our analyses [24].

III. Technical Challenges

There have been several technical challenges pertaining to this line of work. Originally we were focused on expanding the sock puppet audits, and while creating the sock puppets and running them was relatively scaleable, getting them to operate from within university-related computing infrastructure was terribly difficult, and justifying the costs for externally hosted VMs was difficult.

In addition to empirical data we are pursuing simulations in two ways, both of which have posed scaling problems: the first way is by statically generating activity for each user in a network and constructing their feeds from a global perspective. This has posed scaling problems primarily in the number of nodes in the network as the number of potential friends to sort through and construct feeds for increases. Similarly, inserting recommendations into simulations rather than just simple ranking has not been performant so far.

The second way is using agent-based simulations, where each node interfaces with a model to get served content from other nodes. The Repast library lends itself well to parallelizing this sort of simulation, however it requires high-performance computing resources.

Clean experimentation has also posed somewhat difficult: not only for the sock-puppet auditing and comparing black-box systems appropriately, but also for simulations as well. We try to control as any factors as possible comparing two different feed-sorting approaches like graph size, assortativity, allocation of binary traits across the population, and degree-attribute correlation (i.e., correlation between binary trait X and degree sequence).

Finally, the last of the larger technical challenges has been getting performant recommendations. If we are trying to simulate recommender systems in OSNs, it makes intuitive sense to emulate the production recommender systems as much as possible, however doing so while simulating user activity and the like is not straight-forward. Recommender system libraries used for benchmarking systems also do not tend to provide models appropriate for constructing news feeds (i.e., online model that can rank content unseen in the training set); the most relevant we could find involve suggesting new connections (e.g., Who to follow suggestions).

IV. Originality

We believe this is original work as it suggests the utility of incorporating the passive exposure to information as a part of the overall measure of bias in an OSN. Not only does the structure of a network manner in determining who you get exposed to, but the activity of your friends (and through that the work of the News Feed mediating your exposure to said activity). We do not believe that this gets adequate treatment in the study of OSNs especially in the context of echo chambers and (mis)information diffusion.

We believe that this work also suggests ample opportunity for intervention from a platform level for addressing concerns of misinformation. Rather than having to adjust the visibility of certain accounts across the entire platform, a dial could be set for each individual account (potentially controlled by the user as well) adjusting who they see post-recommendations (perhaps client-side).

V. Outline of Proposed Solutions

We propose the following outline for a research agenda pertaining to exposure bias in OSNs:

1) For RQ1 we measure the existence of exposure bias in a production OSN ecosystem with sock-puppet audits. Additional work would consist of more sock-puppet accounts and the use of data donations from users.
2) For RQ2 and 3 we use simulations and Twitter data from 2014 before any production recommender system was implemented, gathered by Alipourfard et al., 2020 [26]. Among several metrics we extend and utilize a measure of local perception bias described in their work [26]: $B_{\text{local}} = \mathbb{E}\{q_f(\mathrm{X})\} - \mathbb{E}\{f(\mathrm{X})\}$ where $\mathbb{E}\{q_f(\mathrm{X})\} = \bar{d} * \mathbb{E}\{f(U)A(V)|(U,V)\ \text{Uniform}(E)\}$, $f(U)$ is the value of the binary trait of friend U, and $A(V)$ is the attention node V pays to friend U.
3) For RQ4 we use agent-based simulations and control recommendations to see dynamics of exposure bias in artificial scenarios. We can use real data to generate the graphs and activity patterns and see if interventions can significantly change different measures of exposure bias.

VI. Preliminary Experimental Evaluation

A. RQ1

For RQ1 we refer to a previous report [7]. To summarize, we can observe a skew in popularity of what is observed from the inventory of tweets between the personalized and reverse chronological feeds, but a study with eight sockpuppets do not identify a skew in exposure bias between the activity network and the networks observed in the feeds (measured by Gini coefficient). However, we do observe a measurable effect in Gini coefficient in results not yet published from a study with more sockpuppets following more users.

B. RQ2

For RQ2 we utilize both empirical Twitter data and simulations of user activity and feed-construction schemes to evaluate the effect of different algorithms on exposure bias. We observe in the empirical data described in Figure 1 and 2 significant difference in the bias especially as we vary the degree-attribute correlation ρ_{kx} (t-statistic: -4.96, p-value: $< 10^{-10}$).

As we could not reasonably model reverse chronological timelines with the static numerical simulations, we simulate other timelines and assess their difference under a Mann-Whitney U test. We find the correlated and popularity timelines presented are both significantly different than a random baseline (not reported here). We also examine other metrics in Figure 3 including the fraction of positive friends seen each day (i.e., $X_i = 1$ for friend i) and Gini coefficient as a measure of visibility inequality.

C. RQ3

For RQ3 we utilize similar empirical and simulated data to evaluate different lengths. Variance in the results seem to suggest that only in a few conditions do finite length sessions demonstrate different effects of exposure bias. Intuitively, in the empirical data, longer sessions seem to present less bias according to the local bias measure in both Figure 1 and 2, which can be explained by the saturation of observed edges leading to an less biased estimator.

D. RQ4

We are currently evaluating agent-based models to assess the impact of interventions in the level of exposure bias over time. We are utilizing similar ranking heuristics to construct feeds and plan to incorporate a distilled version of the candidate-generation and heavy ranker models described in the Twitter recommendation system GitHub release.

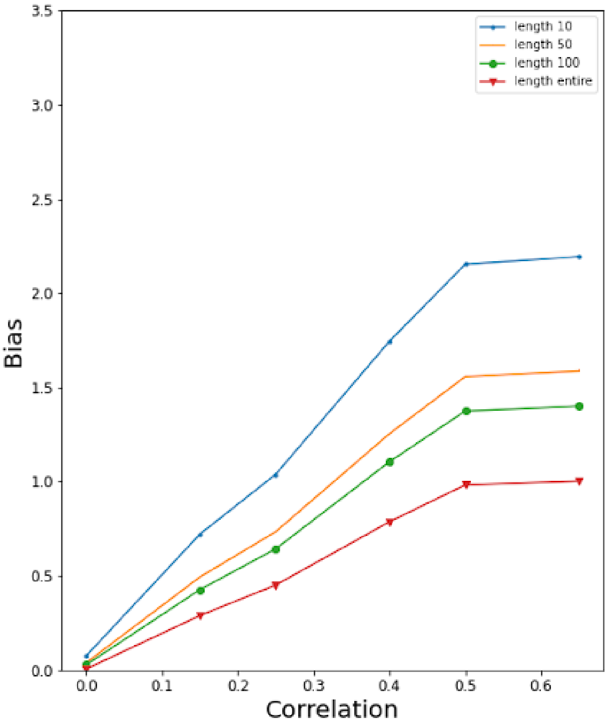

Fig. 1. Empirical Popularity Local Bias versus degree-node correlation. Bias refers to the measurement of Local Bias described in [26].

References

[1] M. S. Granovetter, "The strength of weak ties," *American journal of sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[2] E. M. Rogers, *Diffusion of innovations*. Simon and Schuster, 2010.

[3] P. Panagiotopoulos, J. Barnett, A. Z. Bigdeli, and S. Sams, "Social media in emergency management: Twitter as a tool for communicating risks to the public," *Technological Forecasting and Social Change*, vol. 111, pp. 86–96, 2016.

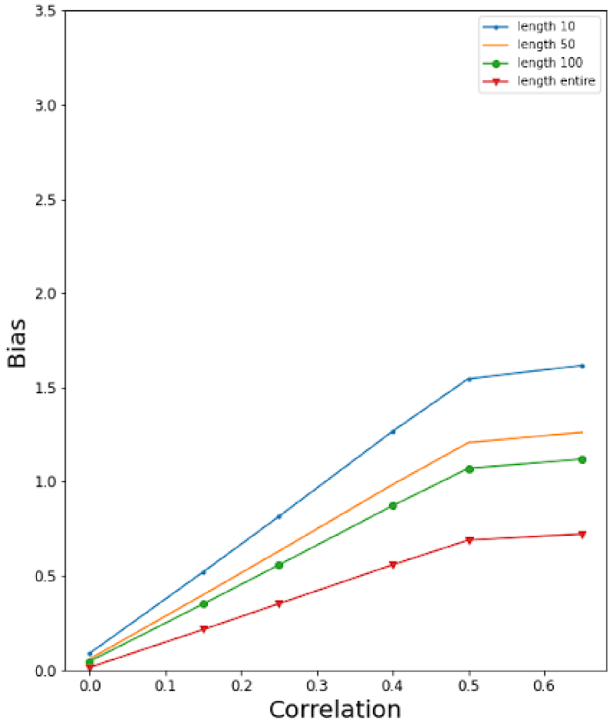

Fig. 2. Empirical Reverse Chronological Local Bias versus degree-node correlation. Bias refers to the measurement of Local Bias described in [26].

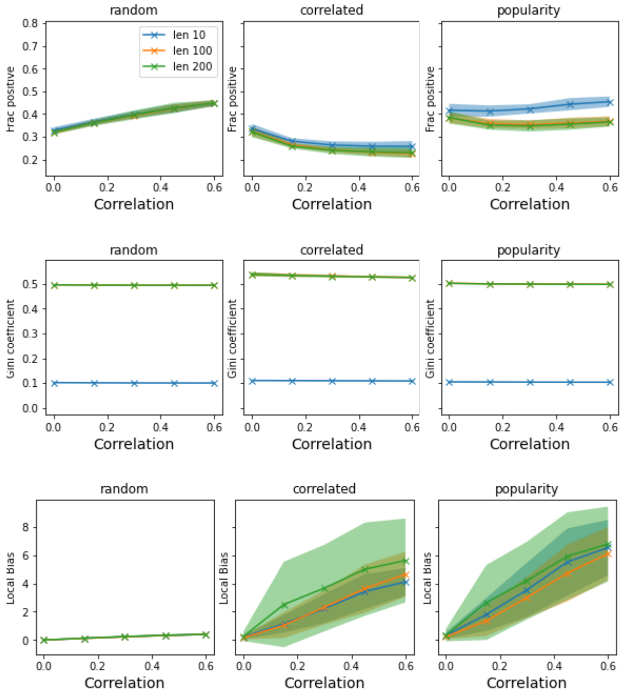

Fig. 3. **Power-law simulations**. Plotted are the mean values of each metric, with bars being one standard deviation. We plot F_{ti} and not F_{ui} as they reported nearly identical data.

[4] D. Nikolov, A. Flammini, and F. Menczer, "Right and left, partisanship predicts (asymmetric) vulnerability to misinformation," *Harvard Kennedy School (HKS) Misinformation Review*, 2021.

[5] M. H. Ribeiro, R. Ottoni, R. West, V. A. Almeida, and W. Meira Jr, "Auditing radicalization pathways on youtube," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 2020, pp. 131–141.

[6] W. Chen, D. Pacheco, K.-C. Yang, and F. Menczer, "Neutral bots reveal political bias on social media," *arXiv preprint arXiv:2005.08141*, 2020.

[7] N. Bartley, A. Abeliuk, E. Ferrara, and K. Lerman, "Auditing algorithmic bias on twitter," in *13th ACM Web Science Conference 2021*, 2021, pp. 65–73.

[8] R. Kardosh, A. Y. Sklar, A. Goldstein, Y. Pertzov, and R. R. Hassin, "Minority salience and the overestimation of individuals from minority groups in perception and memory," *Proceedings of the National Academy of Sciences*, vol. 119, no. 12, p. e2116884119, 2022.

[9] N. O. Hodas and K. Lerman, "The simple rules of social contagion," *Scientific reports*, vol. 4, no. 1, p. 4343, 2014.

[10] V. Kumar, D. Krackhardt, and S. Feld, "Network interventions based on inversity: Leveraging the friendship paradox in unknown network structures," *Yale University, Tech*, 2018.

[11] M. G. Rodriguez, K. Gummadi, and B. Schoelkopf, "Quantifying information overload in social media and its impact on social contagions," in *Proceedings of the international AAAI conference on web and social media*, vol. 8, no. 1, 2014, pp. 170–179.

[12] B. Mønsted, P. Sapieżyński, E. Ferrara, and S. Lehmann, "Evidence of complex contagion of information in social media: An experiment using twitter bots," *PloS one*, vol. 12, no. 9, p. e0184148, 2017.

[13] U. Grandi, L. Kanesh, G. Lisowski, M. Ramanujan, and P. Turrini, "Identifying and eliminating majority illusion in social networks," 2023.

[14] M. Los, Z. Christoff, and D. Grossi, "On the graph theory of majority illusions," *arXiv preprint arXiv:2304.02258*, 2023.

[15] P. Juneja and T. Mitra, "Auditing e-commerce platforms for algorithmically curated vaccine misinformation," in *Proceedings of the 2021 chi conference on human factors in computing systems*, 2021, pp. 1–27.

[16] P. Sapiezynski, W. Zeng, R. E Robertson, A. Mislove, and C. Wilson, "Quantifying the impact of user attentionon fair group representation in ranked lists," in *Companion proceedings of the 2019 world wide web conference*, 2019, pp. 553–562.

[17] C. Sandvig, K. Hamilton, K. Karahalios, and C. Langbort, "Auditing algorithms: Research methods for detecting discrimination on internet platforms," *Data and discrimination: converting critical concerns into productive inquiry*, vol. 22, no. 2014, pp. 4349–4357, 2014.

[18] J. Bandy and N. Diakopoulos, "More accounts, fewer links: How algorithmic curation impacts media exposure in twitter timelines," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW1, pp. 1–28, 2021.

[19] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine bias," in *Ethics of data and analytics*. Auerbach Publications, 2016, pp. 254–264.

[20] M. Tomlein, B. Pecher, J. Simko, I. Srba, R. Moro, E. Stefancova, M. Kompan, A. Hrckova, J. Podrouzek, and M. Bielikova, "An audit of misinformation filter bubbles on youtube: Bubble bursting and recent behavior changes," in *Proceedings of the 15th ACM Conference on Recommender Systems*, 2021, pp. 1–11.

[21] L. Beattie, D. Taber, and H. Cramer, "Challenges in translating research to practice for evaluating fairness and bias in recommendation systems," in *Proceedings of the 16th ACM Conference on Recommender Systems*, 2022, pp. 528–530.

[22] P. Ramaciotti Morales and J.-P. Cointet, "Auditing the effect of social network recommendations on polarization in geometrical ideological spaces," in *Proceedings of the 15th ACM Conference on Recommender Systems*, 2021, pp. 627–632.

[23] F. Huszár, S. I. Ktena, C. O'Brien, L. Belli, A. Schlaikjer, and M. Hardt, "Algorithmic amplification of politics on twitter," *Proceedings of the National Academy of Sciences*, vol. 119, no. 1, p. e2025334119, 2022.

[24] T. Lazovich, L. Belli, A. Gonzales, A. Bower, U. Tantipongpipat, K. Lum, F. Huszar, and R. Chowdhury, "Measuring disparate outcomes of content recommendation algorithms with distributional inequality metrics," *Patterns*, vol. 3, no. 8, p. 100568, 2022.

[25] A. El-Kishky, T. Markovich, S. Park, C. Verma, B. Kim, R. Eskander, Y. Malkov, F. Portman, S. Samaniego, Y. Xiao *et al.*, "Twhin: Embedding the twitter heterogeneous information network for personalized recom-

mendation," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 2842–2850.

[26] N. Alipourfard, B. Nettasinghe, A. Abeliuk, V. Krishnamurthy, and K. Lerman, "Friendship paradox biases perceptions in directed networks," *Nature communications*, vol. 11, no. 1, p. 707, 2020.