Social Network Timeline Bias and Amplification

NATHAN T. BARTLEY, Information Sciences Institute, USA

Algorithmic systems mediate many of our interactions in online social networks (OSNs) making them necessary objects of study for their role in the spread of information. Recommender systems in online social networks in particular are suspected sources of potential radicalization and amplification of mis- and dis-information.

A key feature of recommender systems is their ability to sort and select for content that users will find useful. In the context of a online social network these filtering systems would naturally interact with individual users' cognitive biases, especially those around perception of their peers. Under the social cognitive theory of human functioning, these social perception biases can then inform how users might behave; in an online context any behavior would then turn around and inform the OSN's timeline system.

To be able to contest these complex ecosystems of models, external researchers and users ought to be able to observe and measure model outputs in (near) real-time. Here we describe a generalized man-in-the-middle proxy server that can be constrained and focused on specific OSNs for recommender system crowdsourced data audits.

Additional Key Words and Phrases: algorithmic audits, recommender systems, blackbox, proxy

ACM Reference Format:

Nathan T. Bartley. 2018. Social Network Timeline Bias and Amplification. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Algorithmic audits are an increasingly fruitful area of research as individuals and organizations are becoming more invested in understanding how their data-powered systems behave, especially in an tightened regulatory environment. While companies strive to build internal infrastructure it is increasingly important for external researchers and organizations to find ways to assess how these tools work from the other side of the public-private interface. This is apparent as online social network (OSN) platforms change and/or remove access to their APIs for researchers, as seen with FaceBook/Instagram, Twitter, and TikTok.

Sandvig et al. (2014), details different forms of algorithmic audits, including the crowdsourced audit in which confederates are either hired or recruited to donate data on how the algorithm in question behaves [4].

Here we describe how a man-in-the-middle proxy server can be generalized across different OSNs for crowdsourced data audits of production AI systems (namely recommender systems).

2 MATERIALS AND METHODS

As most online social networks utilize HTTPS, data donation of real-user OSN sessions is somewhat constrained to three options:

(1) Install a plugin to read user data once it has been rendered
(2) Nest a user's connection in another website and acquire data once it has rendered

Manuscript submitted to ACM

(3) Break the HTTPS encryption and intercept OSN specific packets to acquire data before it has been rendered

Installing a plugin can be costly to build and train users to use [3], and psychologically for uninformed users can seem like a burden on their privacy. As such, the barrier to entry for plugins in data donations is not insignificant. Nesting a user's connection in another site through the use of iframes or analogous frameworks is a potential security risk for the websites being targeted, meaning it is often blocked [2].

Breaking HTTPS encryption with full user consent is an ethically dubious approach to data donation. However, in the context of a fair-use research project with appropriate treatment of data it is legally more sound, especially if run through an IRB where possible. In order to intercept traffic bound towards and from any specific OSN, (e.g. Twitter) we suggest making use of a man-in-the-middle proxy like mitm-proxy [1].

Man-in-the-middle proxies can operate in different modes that are relevant for this work:

- *Transparent mode* where relevant traffic is routed through the proxy server. Encryption is established between user and proxy, as well as, proxy and target OSN, allowing for analyzing target network data.
- *Regular mode* where a client actively configures their connection to the proxy server, and the server will only intercept traffic destined for specific OSNs. This requires a higher level of trust as it needs a system/browser level configuration.
- *Reverse proxy* mode where a client connects to a proxy server that in turn forwards the connection to a pre-determined OSN.

Transparent and reverse proxy modes are the most relevant modes of operation as they can theoretically require no client configuration of proxies, and should only require logging into a researcher's website that can forward the connection to the researcher's target OSN of choice. These proxied connections would behave mostly as normal (at least when it comes to staying on the OSN being studied), and allow for a landing page for users to see before consenting to donate their session data. We believe user cookies could be configured and sent through the proxy server.

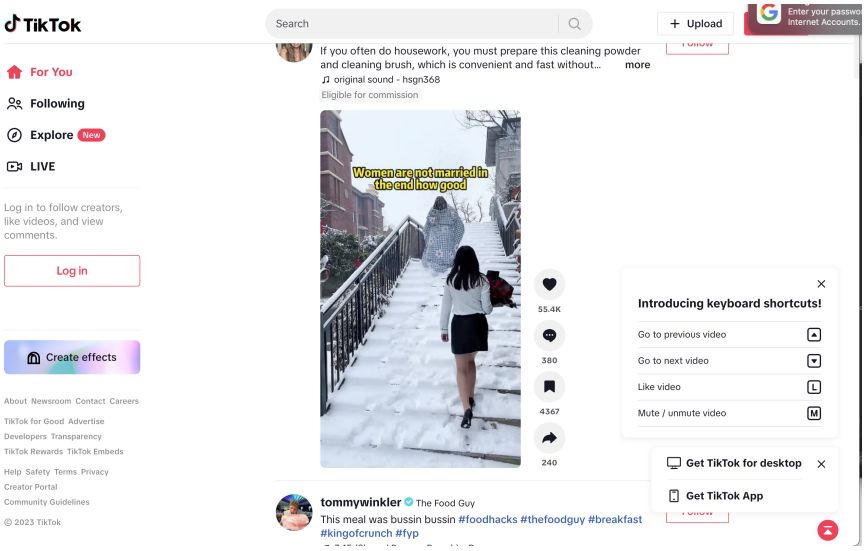
An additional benefit to the man-in-the-middle proxy approach is that it is platform agnostic and readily generalizable to different OSNs. Similarly, it could feasibly be used to analyze apps on smart phones, which can be problematic for data donation purposes (especially on video-focused apps like TikTok and YouTube). Frida has been demonstrated to be helpful at analyzing API calls for apps on Android. An example of how such a server might work is described in Fig. 1a.

A motivating idea for this work is that it can and should be easier for users to donate their data. If we can collect and manage this data in an ethical manner than it is possible to collect blackbox AI system output without API access. With an appropriate data collection methodology it should be possible to construct studies not only of the information but also people users get exposed to. It would be very interesting to be able to interrogate the differences in communities and their experiences online.

REFERENCES

[1] Aldo Cortesi, Maximilian Hils, and Thomas Kriechbaumer. [n. d.]. contributors. 2010–. mitmproxy: A free and open source interactive HTTPS proxy.
[2] Niels Provos, Panayiotis Mavrommatis, Moheeb Rajab, and Fabian Monrose. 2008. All your iframes point to us. (2008).
[3] Ronald E Robertson, David Lazer, and Christo Wilson. 2018. Auditing the personalization and composition of politically-related search engine results pages. In *Proceedings of the 2018 World Wide Web Conference*. 955–965.
[4] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry* 22, 2014 (2014), 4349–4357.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009



(a) Example screenshot of TikTok web browser without logging-in.

(b) Packet captured from TikTok with first video information highlighted.

(c) Packet captured from TikTok with second video information highlighted.

Fig. 1. An example of how a proxy server can intercept data from a local machine.