Biases in Facebook Friend Recommendation Services

Mahi Elango, Joshua Noel, Kristen Surrao, Francisca Vasconcelos

Originating in the early 2000s as a social media platform to connect Harvard University and other elite college students, Facebook has now become ubiquitous throughout communities across the world. Users can connect with friends, follow pages/people, join interest groups, discover nearby events, use the same login for other websites, make marketplace purchases, and find romantic partners. Facebook collects a significant amount of user data to drive its advertisement service and monetize the platform. Furthermore, in order to keep millions of users engaged, Facebook engineers its general recommendation services, such as friend, feed, and event recommendations, to optimize user satisfaction by personalizing each user's friend network. While this may be superficially beneficial for Facebook and the individual user, we believe that such recommendation systems detrimentally reinforce users' existing biases and ideologies. Furthermore, the reach of these recommendations extend far beyond the Facebook interface itself, skewing public perception and shaping real-world human networks.

Overview

In this paper, we aim to assess biases pervasive in Facebook recommendation systems, specifically as they pertain to the Facebook friend recommendation algorithm. We apply the notion of reinforcement described by Safiya Noble in "Missed Connections: What Search Engines Say About Women," as well as analysis of different sources of bias, such as feature selection and proxies, described by Solon Barocas and Andrew Selbst in "Big Data's Disparate Impact." However, neither paper addresses the full extent or implications of bias presented by the Facebook friend recommendation system. As a vehicle to understand the complexities of this algorithm, we will investigate a case study of friend recommendations for students at elite academic institutions. We conclude with an analysis of how the limited scope of a friend network creates intrinsic challenges in addressing bias.

Data Dump

Although Facebook's recommendation algorithms are not publicly available, we make use of the Facebook data dump tool to gain insight into data-types that may be relevant. Facebook lists explicitly provided personal information (i.e. gender, age, university, and employment), a Facebook activity history, and a list of, possibly inaccurate, labelled interests derived from the user's activity history. Interestingly we also see data that users may unknowingly provide Facebook, such as location history (gathered through devices used to access Facebook) and information gathered from advertisers/other businesses. This "Advertisers and Businesses" section presents a browsing history gathered from websites that integrate Facebook plugins and a system where companies can sell user lists to Facebook. This asymmetry in information sourcing begs the question: What biases arise in Facebook's recommendation systems? If so, will these biases always exist?

"Big Data's Disparate Impact"

In their paper, "Big Data's Disparate Impact", Barocas and Selbst analyze a number of sources of bias to which algorithmic inference systems are susceptible [1]. In some respects, Facebook's recommendation systems often protect themselves from bias by defining a clear *target variable* and *class labels*. This alleviates the need for one to engineer quantifiable, though possibly biased, proxy target variables and labels. For example, when deciding what new friends to recommend, Facebook aims to maximize one's

likelihood of knowing or sharing interests with the recommended individual. Although a difficult question, it does not require the designer to rephrase it to be quantifiable. This same lack of bias does not necessarily follow for bias that the authors state can be introduced through the collection and labelling of training data.

There exists a nested interaction between Facebook's recommendation systems: training data for one recommendation system may be based on how users have interacted with other recommendation systems. A successful recommendation is defined as one with positive engagement. Engagement can be characterized by how long a user looks at a recommendation or by the occurrence of an interaction. Such objective measurements of engagement minimize bias introduced through *labelling of training data*. Although the explicit purpose of a given recommendation system is not to influence the training data for another recommendation system, such interaction is likely, given the complex network of connected systems that Facebook employs. The *collection of training data* from nested recommendation systems could over-represent frequent users, since more interactions exist. The authors caution against overrepresentation in training data, as the model may overfit characteristics of the overrepresented population, making the trained decisions biased. The discussion of bias through data extends beyond labelled training data to data used as inputs to the model, which the previously discussed Facebook data dump grants insight into.

We can see bias in Facebook's recommendation systems when viewed through Barocas and Selbst's concept of bias through *feature selection*. For example, frequency of Facebook interactions can act as a proxy for features such as free-time or ease of Internet access. In another example, in the list of companies that have sold user lists to Facebook we see financial companies such as Experian. Facebook may match a user with specific financial institutions that have high correlation with economic status. These examples show how one's activity both inside and outside of Facebook can act as a *proxy feature*, as defined by the authors, for recommendation systems that aim to avoid discriminatory inputs, such as job advertisements. Further, as the friend recommendation system aims to connect people that know each other or share interests, selected features are limited to the available user-generated features, such as current friends and interests derived from likes, or third party tracking data. Let's consider the example of friend recommendations for students at elite colleges. Here, a prominent feature may be the college a user attends, which Facebook may inappropriately infer as representative of the user's academic interests. In recommending friends of similar academic interests, friend networks maybe have disproportionately high concentrations of friends from similarly elite schools, rather than a more general friend network truly representative of one's social network. If Facebook then utilizes the user's Facebook friend network as an input feature to other recommendation systems, the input will be inherently biased. Barocas and Selbst do not fully consider how to deal with sources of bias that are inherently unremovable, as will be further discussed after considering the multiplicative effect of this bias.

Noble and "Echo Chambers"

We now consider how Facebook's friend recommendation system propagates its biases as inputs into other recommendation systems. Extending the example of students at elite colleges, suppose Facebook's algorithm recommends two students from elite colleges to become friends. With two quick button clicks they become friends, interacting with each others' posts, shares, event activity, and more. Facebook uses these interactions to build up profiles of both students. Based on interaction with those of similar backgrounds, the students are recommended specific pages or ads reflecting that shared background.

Since there is a finite amount of recommended content, this disproportionately boosts the ratio of content reflective of their shared ideology and environment to otherwise unfamiliar content.

This ideological phenomenon, in which a person encounters only beliefs or opinions that coincide with their own, is known as that of "echo chambers." The result is the reinforcement of existing views and disregard of alternative ideas, which is reminiscent of arguments made by Noble [2]. Noble notes that "search engines don't only mask unequal access...as broken down by race, gender, and sexuality" but "they also maintain it" (Noble 39). To further extend the example of elite college students, if these students are only interacting with each other, then, for instance, their employment opportunity recommendations would be of a certain type, likely of higher wage and education, whereas others may not necessarily be presented with the same recommendations.

This idea of maintaining unequal access goes beyond our example of elite college students, infiltrating all sorts of demographic contexts. In fact, one study showed that recommendations for employment opportunities involving janitorial duties and taxi driving were shown to a higher fraction of racial minorities. Moreover, jobs for nurses and secretaries were shown to a higher fraction of women than men [3]. These occurrences highlight Noble's claim that initial biases propagate in a positive feedback loop, further stratifying groups according to particular characteristics.

Beyond Barocas, Selbst, and Noble

Both the Barocas & Selbst and Noble frameworks imply that addressing these various types of bias would guarantee a significantly less biased algorithm. However, both papers fail to address situations in which intrinsic properties of the algorithm limit the extent of bias removal. Specifically, in the context of Facebook's friend recommendation algorithm, users are unable to befriend everyone in the world and users desire personalization, explaining many biases previously outlined. It is impossible to recommend a set of friends, for a single user, that represents the diversity of the entire population. In fact, we argue that the only way to do so is through recommendation of all Facebook users as friends, thus sacrificing personalization. However, recognition that a perfect algorithm, which satisfies all intrinsic limitations *and* resolves biases, does not exist would empower Facebook engineers to design an algorithm that optimizes for these constraints.

Contributions

The work for this paper was equally divided among all members of the group. Several group discussions were held, but the paper was divided into three primary sections. Joshua wrote the analysis pertaining to the Barocas and Selbst paper. Kristen wrote the Data Dump section and analysis pertaining to the Noble paper. Francisca and Mahi focused on the introduction paragraphs, editing, cohesiveness, and concluding analysis of the paper.

Bibliography

1. Solon Barocas and Andrew D. Selbst, Big Data's Disparate Impact, 104 Calif. L. Rev. 671 (2016).
2. Noble, Safiya. "Missed Connections: What Search Engines Say About Women."
3. Hao, Karen. "Facebook's Ad-Serving Algorithm Discriminates by Gender and Race." *MIT Technology Review*, MIT Technology Review, 8 Apr. 2019, www.technologyreview.com/s/613274/facebook-algorithm-discriminates-ai-bias/.