An Approach to De-Homogenizing Recommender Systems

by

Christie Hong

S.B., Massachusetts Institute of Technology (2019)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Author . . . .
Department of Electrical Engineering and Computer Science
May 14, 2020

Certified by . . . .
Alexander 'Sandy' Pentland
Professor
Thesis Supervisor

Accepted by . . . .
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

An Approach to De-Homogenizing Recommender Systems

by

Christie Hong

Submitted to the Department of Electrical Engineering and Computer Science
on May 14, 2020, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Recommender systems are all-pervasive on every online medium we interact with today. Yet, they impose a substantial problem of homogenization of users over time, leading to lack of visibility of the discriminated. This problem stems from the fundamental literature and traditional approaches around recommender systems and the primary focus on user-centric accuracy. In practice, this results in popularity bias, filter bubbles and "down the rabbit holes," where popular content in "your circle" continues to be recommended, and less popular content becomes discriminated against. And how does one measure popularity? Typically by some data metric of ratings: likes on a post, views on a video, the number of 5 star ratings. This ultimately results in locally optimal recommendations, where the users are content and comfortable with taking these results, but globally sub-optimal diversity, as the ecology of users grows more and more homogeneous over time. In this paper, I propose an approach to de-homogenizing recommender systems via an ecological values-based recommender system, as well as different metrics to evaluate on. This system encompasses both personalization, as well as diversification at both a user, and ecosystem of users, level. With this research, I demonstrate that this approach can introduce diversity in applicable ways, and reveals the weaknesses in current traditional models in tackling the problem of homogeneity. These insights can be used to guide future recommender systems and continue the conversation of developing more diverse and impactful recommendations.

Thesis Supervisor: Alexander 'Sandy' Pentland
Title: Professor

Acknowledgments

Thank you MIT for being my home for the past five years, and giving me the opportunity of a life time to pursue my passions and meet the most spectacular people. Working on this thesis in the midst of COVID-19 has been quite the experience, and while this page is not nearly enough to thank all the friends, mentors, and professors I've met during this journey, I'd like to acknowledge a few people:

First and foremost, I'd like to thank Alex Pentland, for without his guidance and wisdom this thesis would not have been possible. To Dan Calacci, thank you for supporting my ideas since the beginning, engaging in conversations about discrimination, and brainstorming the possibilities to bring more diversity. To Aaron Wubshet and Benjamin Oberlton, thank you for sticking by my side through thick and thin since CPW - I could not be more proud to call you two my best friends. To Nilai Sarda, thank you for being my rock during the hardest of times; you always push me to be a better student, a better person.

Finally, to my parents and sister, thank you for your endless love and support. I would not be where I am without you guys, and I love you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Recommender systems are an integral part of how consumers discover a wide variety of artifacts. In today's day in age, recommender systems are ubiquitous, and its impact on users is high, long-lasting, and most notably - unnoticeable to its users day-to-day. From Facebook newsfeed, to Netflix's 'Top Picks for You,' Spotify's 'Discover Weekly,' to tailored front-page news articles on GoogleNews, all of these recommendations impact a user's data consumption, and thus indirectly, a user's thoughts and experiences.

The value of recommender systems is personalization. The customization of the consumer's experience from specially tailored music playlists, to social feeds, to online market displays is extremely valuable in modern markets. These recommender systems have high business value in: user retention, user usage/involvement, and profits. For example: GoogleNews stated that recommendations increased article views by 38% [8]; McKinsey Reported that 35% of Amazon sales stem from recommendations [11]; Spotify reported that over 30% of overall listening in 2017 was due to recommendations [27]. To translate that to a monetary value, GoogleNews made $4.7 billion in revenue from news content and advertisements in 2018 [24]; Amazon made $141.2 billion in revenue from Amazon Market Place in 2019 [15]; Spotify made $1.946 billion in revenue [32]. Recommender systems contribute significantly to these number by keeping user retention and interactions high.

While personalization seems like a major win-win with consumers and companies both benefiting from it, Fleder and Hosanger recently ran a comprehensive study on the 'fairness' of these recommenders. They found that recommenders are detrimental to diversity - that modern recommenders homogenize users over time [12]. They discovered that when viewing consumers as a "similarity network before versus after recommendations" they found that the network becomes "denser and smaller." The essence of this study outlines that current recommender systems are *locally optimal* (in that individual users are happy and benefiting off the recommendations), but *globally sub-optimal* (when users are viewed as a group/ecosystem, the effects are damaging). This paper attempts to design personalization algorithms for ecosystem-level goals that is locally optimal *and* globally optimal.

The motivation is two pronged. First, from a technology and policy lens, recent literature has proven that recommender systems are homogenizing users over time [12, 13]. Yet, very little work has been done to find a balance between optimizing recommender algorithms, as well as preserving and growing the diversity of its users. Second, the effects of recommender systems on homogeneity over time has serious implications to social diversification, particularly in physical spaces. Fleder and Hosanger found that networks of users became "denser and "smaller" over time, and Turner and Rawlings, in their extensive study in "Promoting Neighborhood Diversity", state with substantial evidence that "residential segregation undermines the well-being of individuals, communities, and the American society as a whole" [35]. Given the literature around recommender systems and urban studies, recommendations for physical spaces from apps such as GoogleMaps or Yelp have high potential in negatively affecting segregation in cities. Moreover, recommenders are fundamentally built off the idea that "likes like likes" and systemically calculate similarity distances between users and items to determine the best recommendations. By that logic, the best recommendations are gleaned from those similar to you. However, judging similarity to the behavior of people like you, today's recommender systems build echo chambers. By intentionally adding in variation to get people to visit other places that are semantically similar but not often visited by "your crowd" we can break up echo chambers and promote diversity. In other words, it's not necessary to recommend the same coffee shop that everyone else in "your crowd" goes to, instead we can try to pick a different nearby coffee shop that is patronized by different sorts of people.

Thus, this proposal aims to provide a recommender system that combats homogeneity via an ecological values-based approach, focused primarily on increasing user diversity.

Chapter 2

Motivation: A Critical Analysis of Recommender Systems

The severity of homoegeneity and the problems that incur because of it will be explored in the following section. This section will highlight the challenges of recommender systems, the effects of those challenges, and what homogenity looks like beyond just clustering of users over time.

"Youtube's Related Video Algorithm Helpful to Child Predators"

"Facebook's ad systems seems to discriminate by race and gender"

"How Twitter's algorithm is amplifying extreme political rhetoric"

These headliners all discuss a phenomenon called "down the rabbit hole effect" where the recommender algorithms from Youtube, Facebook, and Twitter keep recommending the same content and lead users down one path; or in other words: homogenization of content and ultimately users. I investigate the effects of recommender systems in social mediums to provide examples on how current recommender systems are locally optimal and globally extremely suboptimal, and why change is necessary.

The common problem amongst social media recommender systems is the "black box" problem. The primary concern with the "black box" problem is the lack of transparency of the algorithm, which results in a huge difficulty in filtering out the problems, and if unattended, the algorithm can become very biased or even "evil" in some cases (like the ones above).

2.1 Challenges of Recommender Systems

In this part of the analysis, I will explain the challenges around recommender systems: content diversity, data sparsity, result metrics, and creator influence. These challenges are derived from recent news and literature to prove that these issues exist and pose problems today. The next part will explain why these are critical problems, and the serious implications that can emerge from them.

2.1.1 Content Diversity

Fleder and Hosanger found that most recommender systems recommend items from the same categories, leading to a decrease in content diversity and ultimately, homogeneity in users over time [12]. This occurrence is traced to the phenomenon of concentration bias - a recommender system with concentration bias has an inclination for certain recommendation content over others and is thus unfair to other contents [2].

In later years this idea was refined to an idea of popularity bias by Abdollahpouri et. al [1]. This term further defines “concentration” more specifically as a high density of views, likes, clicks, ratings, etc. versus just general inclination as defined by Adampoulous et al. Thus a recommender system with population bias keeps recommending content that is already popular and thus causes prejudice towards less popular content. However, remember the value of recommender systems - personalization. From its inception, the goal of a recommender system was not to be fair and have a great balance in its recommendation contents; instead, it was to provide its user with the most personalized content which has led to Adampoulous et. al and Abdollahpouri et. al's work in the consequences of over-specialization, concentration bias, and popularity bias.

Fledar and Hosanger point out that several recommender systems have caused a decrease in user diversity because many of the result metrics of recommender systems focus primarily on accuracy of the recommendations. Since then, there has been much work done to combat this problem of content diversity and user diversity. Kunaver et. al provides a comprehensive overview on the evolution of these recommender system experiments - but this overview confirms that content diversity still remains a tough challenge in recommender systems today [16].

2.1.2 Data Sparsity

Recommender systems rely on users' historical data, whether it be ratings, site visits, successful purchases, or likes of social media posts. However, the heavy reliance on prior historical data results in a big problem when users are new to a platform. We refer to this problem as data sparseness. Lack of data is applicable to two categories: new users and new content [19]. Data sparseness spawns a bigger issue in recommender systems known as the "cold start problem" [19]. The "cold start problem" occurs for new users when the platform does not have any given ratings or explicit preferences. Hence, the first few recommendations are, most of the time, pulled from popularity bias of the masses, and/or not very accurate, which can decrease user satisfaction [2]. The "cold start problem" for content is similar in that, when new content does not yet possess any ratings, popularity bias prevents this content from being recommended easily and therefore becomes less noticed by many users.

Data sparsity is a key component of content diversity and can and should be seen as a major challenge for recommender systems.

2.1.3 Result Metrics

Result metrics are usually defined by the developer of a recommender system. As the metrics define the optimization objective, creators have significant influence over how these systems fundamentally behave and ultimately affect their users. Historically, much of the literature in this field uses accuracy of recommendations to determine the impact of the system. Accuracy typically is a measurement of popularity: the number of views on a video or news article, the number of likes or retweets on a post, etc. Social media platforms, in particular, do not disclose the exact metrics of their recommender systems to avoid social manipulation attacks [17].

We must critically ask ourselves if this is really the correct way to determine accuracy of recommendations. With results metrics relying so heavily on popularity there are two implicit and problematic assumptions being made. Firstly, popular content is universally good/accepted content, and secondly, content of the other side (contrasting content) is not of interest and/or displeases the user. The first assumption leads to popularity bias and for systems to keep recommending content that is already popular and cause prejudice towards less popular content; the second assumption systematically deters diversity in recommendations.

Nevertheless, researchers have proven that metrics on several different platforms are indicative of content popularity. Zhou et al. did a study on Youtube and found that much of its recommender system is built on the idea of "strong correlation between the view count of a video and the average view count of its top referrer videos," meaning that a video has a higher chance to become popular when it is placed on the related video recommendation lists of popular videos [44]. Zhang et al. ran studies on Twitter and its recommender system for "trending" tweets [43]. They found that a topic starts "trending" and being recommended to its users due to four key factors: popularity, coverage, potential coverage, and reputation.

The challenge here lies in two places: 1) the default and dominant popularity metric, and 2) the lack of transparency in these metrics. The first problem results in potential exploitation and manipulation, as we will observe in a case study. The second problem results in paternalistic systems in which developers wield all the power to assume a user's value function.

2.1.4 Creator Influence

We briefly introduced the influence of the creator in the previous section with regards to metrics. In this section, we will examine the creator from two lenses: the operator, and the businessman.

Operator

The operator can be thought of as the company itself. The company influences the recommender system and the content on the platform, as discussed prior, and potentially introduces human bias and judgement on what content is provided. In particular, the operator has the power to block content, which is the ultimate exertion of bias and judgement in recommender systems.

For example, in recent years, Facebook has come under fire for preventing free speech [10] and removing more than 800 US political pages and accounts [36]. In 2019, the White House hosted a Social Media Summit where more than 200 conservatives and right-wing activists presented overwhelming evidence that Facebook and Twitter were censoring their messages on these platforms [20, 31]. These articles illustrate the influence that the company operator has on the content on its platforms.

Businessman

Recommender systems are a part of a business model for most social media platforms. Typically, profit is generated from advertisements as seen on Facebook, Twitter, and Amazon feeds, and sprinkled throughout Youtube videos. Due to their prioritization through financial contracts and incentives, advertisements find their way into recommendations and influence recommender systems.

With these recommender systems already being so black-boxed and paternalistic, it becomes difficult to discern the motive behind recommendations - is it for the benefit of the user, or the benefit of the businessman. In this thesis, I will attempt to answer the following question: can the objective be balanced to the benefit of both parties?

2.2 Case Study: 2016 U.S. Presidential Election

Clearly there is literature evidence of problems in today's recommender systems and their potential effects on users. But let us discuss real life effects with a particular case study of the 2016 U.S. Presidential election.

2.2.1 Description

The 58th United States presidential election was held on November 8, 2016, in which Republican Donald Trump won the election and took office as the 45th president of the United States over Democratic Hillary Clinton.

The aftermath of the election seemed to shock thousands of Americans, many asking themselves how they missed the popularity of Donald Trump. One answer is given by Eli Pariser, in which he coins the concept of a filter bubble: the idea that "personalization tools from Facebook, Google, etc. have isolated its users from opposing viewpoints" leading democrats, republicans, conservatives, and liberals to feel like they are in their own island of thought.

During the 2016 US presidential election, social media platforms were utilized to their full potential by both candidates. Spanning from Facebook to Twitter, Instagram, Snapchat, to Youtube, campaign teams on both sides heavily used these mediums to connect with American voters and win their votes. However, recently, there has been much scrutinization over how these social platforms have influenced the 2016 election, and the beginnings of filter bubble challenges.

Wired reporter El-Bermawy breaks down how a user of a social media platform can get stuck in their own filter bubble, namely himself. Identifying as a liberal, El-Bermawy explains how in the beginning of the 2016 presidential election his Facebook and Twitter newsfeed was filled with #ImWithHer or #FeelTheBern content. As the debates began his feeds moved to discussions of Trump scandals, and why Americans should support Clinton; moreover, he saw articles from liberal media such as the New York Times and the Washington Post. However, he grew skeptical of his recommended content, and ventured to other less liberal sites and found that the second most popular article shared on social media in the last six months leading to November 8, 2016 with words "Donald Trump" in the headline was "Why I'm Voting for Donald Trump" with over 1.5 million shares. Yet he had never seen that story on his Facebook newsfeed, and when asking many of his liberal friends, they had all said that they too had never seen it [9].

Another influence on the election has come from Youtube recommendation videos. According to Paul Lewis of The Guardian, he ran a study on Youtube's recommended videos with one user account starting with a search history of Trump, and another with Clinton. Despite one account searching for Trump and another for Clinton, he noticed that his Trump focused account had a habit of recommending extreme right-wing videos to him, and the Clinton account recommending left conspiracy theories, and pro-Trump videos. He found that Youtube was "six times more likely to recommend videos that aided Trump than his adversary" and that Youtube may be "leading people down hateful rabbit holes." Youtube responded that they "strongly disagree" with his statement and their "search and recommendation systems reflect what people search for, the number of videos available, and the videos people choose to watch on Youtube. That's not a bias towards any particular candidate; that is a reflection of viewer interest" [18]. Youtube seems to be saying that its recommender system is a neutral mirror of the interests of its users, however Lewis asks an important question: "how exactly does Youtube interpret "viewer interest" - and aren't "the videos people choose to watch" influenced by what the company shows them?" For example, when offered a choice in the moment, users may subconsciously click on a video of a dead man in a Japanese forest (Logan Paul), or a man faking his girlfriend's death (ImJayStation), or a fake news clip claiming Hillary Clinton paid Jay Z and Beyonce $62 million dollars to perform at a rally in Cleveland before the election. Zeynep Tukekci, a sociologist, believes that Youtube's recommender system has "probably figured out that edgy and hateful content is engaging" [22]. She draws comparisons and says that "it is a bit like an autopilot cafeteria in a school that has figured out children have sweet teeth, and also like fatty and salty foods. So you make a line offering such food, automatically loading the next plate as soon as the bag of chips or candy in front of the young person has been consumed" [34]. So the food gets higher and higher in sugar, fat and salt - similarly to how Youtube's recommended videos become more and more controversial and hateful.

In addition to filter bubbles and rabbit holes, the presidential election was hit with controversy as heavy inquiry into Russian interference of the 2016 US presidential election began. An investigation by the New York Times revealed mechanisms by which Russian operators used platforms like Twitter and Facebook to spread anti-Clinton messages and promote pro-Trump material [30]. On September 6, 2017, Facebook officials disclosed that they had shut down several hundred accounts that they believe were created by a Russian company linked to the Kremlin, and additionally bought $100,000 worth of ads supporting Trump and pushing divisive issues during the 2016 campaign [28]. The New Yorker reported "of the 470 Facebook accounts known to have been created by Russian hackers and trolls during the campaign, a mere six of them generated content that was shared at least 340 million times." For example, a fake Facebook page created by a Russian bot, Blacktivist, stoked racial tension amongst Americans by posting militant slogans and uncensored videos of police violence against African-Americans. This fake page gathered more hits than the Facebook page for Black Lives Matter [21]. It is now widely understood that Russia's influence was far larger than social media companies initially acknowledged. Facebook initially claimed that Russian disinformation had most likely reached only 10 million users, however that figure has been amended to over 126 million impacted users. Twitter recently acknowledged the influence too, revealing that it hosted more than 50,000 imposter Russian accounts [21]. The New York Times investigated found that of these accounts, many of them were bots that fired off identical messages seconds apart, and had immense power to get certain hashtags trending. For example, Russian bots managed to get the hashtag #HillaryDown to get on the top trending list and catch on amongst real Twitter users.

2.2.2 Consequences

Content Diversity

The problem of content diversity is clear in the articles about Facebook. It stems from a vicious cycle of popularity bias; popular content will get recommended, which in turn leads to the already popular content being recommended to even more users. We see this cycle occur on Facebook, in particular a fake page run by Russian bots, Blacktivist, that garnered extreme popularity and had real users interacting with the page, resulting in more hits than Black Lives Matter. However, this creates a second problem of the filter bubble, where users will mostly see content that agrees with one opinion and does not display content from the "other" opinion. Filter bubbles have occurred on newsfeeds, as we saw this El-Bermawy's experience during the presidential election, filtering only stories that the user identifies with and never showing the opposing view.

These two problems together have a serious impact on the user as the content delivered plays a part in shaping the opinion of the users.

Data Sparsity

We see the problems of data sparsity on Youtube with Lewis' study with his Trump history account. After watching a few videos involving Trump, his recommended 'up next' videos after watching Trump videos are popular extreme right wing content, and more popular pro-Trump videos. It's clear that the lack of data on this particular account, in coalition with the problem of content diversity, leads to lack of other content, say positive videos on Clinton, to be shown to Lewis' account.

Youtube's heavy lean into the popularity bias to recommend up next videos because of the cold-start problem, only supports the vicious cycle of popularity bias, and the bubble that it forms around a user.

Result Metrics

The consequence of result metrics, in that it heavily optimizes for popularity, is evident in the Twitter articles with Russian bots. Twitter revealed that there were over 50,000 Russian accounts, and these bots had the power to dispatch the same identical tweets, just seconds apart. These bots managed to get #HillaryDown on the top trending list, catching the attention of real US users and consequentially, real accounts using this hashtag in their tweets as well.

The ability to exploit this metric, and get a negative troll hashtag to the top of the trending list, has serious implications. Not only does it send a negative message to Twitter's 126 million users [29], it spreads fake news and influences the opinion of the user.

Creator Influence

Lastly, we see creator influence in both the Facebook and Youtube articles. On Facebook, we see the businessman influence as Russian accounts linked to the Kremlin had spent more than $100,000 on Trump Facebook ads. Those ads then reached a projected 126 million users; it is not inadmissible that payment plays a particular role in Facebook's recommender system.

For Youtube, we see the operator influence with the bizarre recommendations from Lewis' Clinton focused account. This particular account had watched Clinton rallies, campaigns, and other related content with her name in the title. One would expect from then on, that Youtube would recommend videos with more democratic messages, positive leftist views, and more pro-Clinton content; however, this was not the case, and instead Lewis was recommended videos on leftist conspiracy theories, and pro-Trump videos. These recommendations are contrasting to all the other problem consequences from content diversity, data sparsity, or result metrics, and it displays a potential operator influence and bias from Youtube itself that inevitably influenced millions of Youtube users.

2.3 Key Takeaway

Current recommender systems have four key problems: content diversity, data sparsity, result metrics, and creator influence. These all contribute to the effects of "homogenization," "filter bubbles," and "down the rabbit holes" that illustrate how recommender systems are not impacting the larger group of users in a positive way due to the pinpoint focus on only bettering recommendation accuracy at a singular user level. With recommender systems functioning off the assumption that "likes like likes," people are being recommended the same content that their friends, and others like them are interacting with.

By simply judging similarity to the behavior of people like you, this results in the problems like filter bubbles. By intentionally adding in variation to get people to visit other places that are semantically similar to your interests, but not often visited by people similar to you, we can burst these bubbles and promote diversity. In other words, I aim to recommend a different nearby Thai restaurant that is visited by different types of people, rather than the same Thai restaurant that everyone in "your group" goes to.

Thus, the goal of this research is to provide a recommender system that not only targets an ecology of users, but also seeks to widen the scope of "accuracy" measures of these systems as well.

Chapter 3

Background

From a computer science lens, a recommender system is defined as a decision making strategy for users under complex information environments [23]. From a marketing lens, a recommender system is defined as a tool that helps users search through their records to extract the users' preferences [26]. Recommender systems have been defined from various lenses, but it is clear that these systems are designed to handle users' problems of information overload by providing them with personalized and exclusive content. Recently, several approaches have been developed to utilize collaborative filtering, content-based filtering, and hybrid filtering.

3.1 Collaborative Filtering

Traditional collaborative filters are solely dependent on historical user data. It does not use any information on items, and simply needs a user's historical preference on a set of items. Collaborative filtering recommenders thus recommend items by identifying other users with similar taste, and then use their opinions to recommend items to a selected user. Because collaborative filtering is based on historical data, the driving assumption here is that users who have agreed in the past, tend to also agree in the future.

When thinking about user preference, the literature has expressed it in two categories: explicit ratings and implicit ratings. Explicit ratings can be anything from a

1-5 star rating, commented feedback - anything that provides direct feedback on an item. Implicit rating, on the other hand, is a measure of user preference indirectly. This can be represented as page clicks, skips on a music track, or repeated purchase records. In our case of recommendations of physical spaces, we track implicit ratings, specifically location check-ins per user.

Using historical user data, the systems then finds similarity between all pairs of users, where the similarity between users is derived from the explicit and implicit ratings. The system then returns a list of recommendations from the set of items that are disjoint from the user in question's historical items.

Collaborative filtering techniques can be divided into two groups: memory-based and model-based.

3.1.1 Memory based techniques

Users historical data play a large role in identifying other users that share the same interests. Once another user is found, different algorithms can be used to combine the preferences of other users to generate recommendations. This can be done either through user-based or item-based techniques. In this research, I utilize user-based collaborative filters.

User-based collaborative filters compute similarity between users by comparing ratings on the same content. By observing a selected user, the system creates weights representing the similarity between selected user and other users. Then it computes a predicted rating of an item for the selected user by taking a weighted average of the ratings of the item by the users most similar to the selected user.

Several types of similarity measures are used to compute this similarity. The two most popular similarity measures are Pearson's correlation and cosine-similarity.

3.1.2 Model based techniques

Model based collaborative filters exploit the previous ratings of items to learn a model in order to improve the systems performance. Majority of these models involve machine learning that are similar to neighborhood-based recommender systems. Some examples of these techniques are: matrix completion, singular value decomposition, and regression and clustering.

3.2 Content-based Filtering

On the other hand, content-based filters, like it's name suggests, is mainly dependent on the content description of items. In this technique, recommendations are made based on features extracted from the content of items a user has interacted with or rated. Items that are most related to the positively rated items from the user are recommended. Similarity between items is typically derived from description of items using Term Frequency-Inverse Document Frequency (TF-IDF), or with probabilistic models such as Naive Bayes Classifier, decision trees, or even neural networks.

Then with a user's item history (ratings, purchases, visits, etc.), the system generates a list of recommended items most similar to those items of the user's item history. For this research, I will be focusing on the methodology of TF-IDF.

The following is a description of TF-IDF:

3.2.1 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF is a sub-field of Natural Language Processing (NLP), and is used for feature extraction purposes. In short, TF-IDF counts the occurrence of each word in a document, weights the importance of each of these words, and calculates a score for that document.

Term Frequency is the measure of the frequency of a word in a document to the total number of words in a document. The higher the frequency of the word, the higher the weight of the word, thus it is divided by document length to normalize.

$$TF(t) = \frac{\text{Frequency occurrence of term t in document}}{\text{Total number of terms in document}}$$

Inverse Document Frequency is the total number of documents to the number of document containing a word. The less frequent a word shows in the collection of documents, the higher the IDF score. It helps in giving a higher score to rare terms in the documents.

$$IDF(t) = log_{10}\Big(\frac{\text{Total number of documents}}{\text{Number of documents containing term t}}\Big)$$

Ultimately, TF-IDF is a measure of how important a word is to a document in a document corpus, where the importance of a word increase proportionally to its frequency in a document, but is offset by the frequency of the word in the corpus.

3.3 Known Limitations

While these two methods on their own have been incredibly successful, several limitations exists. With collaborative filters, there are issues with sparsity and scalability. The biggest disadvantage with this technique is known as the "cold-start" problem in which the system does not have enough historical data on a user or item to make adequate predictions. Similarly, some of the key issues with content-based filters are sparsity of data, overspecialization and limited content analysis. The major disadvantage of this technique is the necessity for in-depth and expansive knowledge on the descriptions of the features of items.

In order to assuage some of these issues, hybrid filtering has been proposed.

3.4 Hybrid Recommender Systems

Hybrid recommender systems combine two or more recommendation strategies in different ways to generate better recommendations. There are several different hybridization designs: parallel use of several systems, monolithic exploitation of different features, pipelined invocation of different systems. All of these can be classified based on their operations such as: mixed, hybrid, weighted hybrid, feature-augmented hybrid, feature-combination hybrid, and much else.

A comprehensive review is provided in Cano and Morisio [6].

Chapter 4

Related Works

4.1 Personalization of Recommender Systems

There have been several works done in the space of recommender systems to improve personalization. Most notably was the Netflix Prize, an open competition for the best recommendation algorithm to improve recommendation accuracy for users by 10%.

Much like the personalization of movies, music, and news, there have been several works in the personalization of recommended spaces. There have been several studies incorporating geographical user conditions [25, 33], sentiment enhancements [25, 37], and social networks [39, 40]. Takeuchi et. al [33] and Savage et. al [25] designed recommender systems that considered time geographies as well as transportation limitations to better hone physical space recommendations. Savage et. al [25] and Yang et. al [37] took into consideration the influence of emotions, and how user sentiment play into choosing locations. Furthermore, Ye et. al [39, 40] developed a social network approach to learn more from just a user's historical data, and leverage his/her social medias to gain more prior. Further explorations into data manipulation with semantics and hierarchies have also been explored [7, 3] to improve the accuracy of recommender systems.

Yet, across the board, these papers echo the need to "improve the accuracy" of recommendations, and more specifically at the individual user level. Once again, the question of social harm that recommenders bring have not been answered in lieu of personalization. This paper seeks to find that balance between personalization and diverse social impact the user level (*locally*)and the larger user group level (*globally*).

4.2 Fairness in Recommender Systems

Recently, there has been a growing awareness about the impact of social harm that machine learning algorithms have on human decisions. In response, there have been several attempts at tackling the problem of fairness in machine learning.

Fairness, in a general sense, means to not discriminate against individuals or groups. There have been several proposals and measures in fairness for machine learning methods such as classification [14, 41], and ranking [4, 42]. Both classification and ranking examine fairness from an individual user level, and a group user level as well.

In comparison to classification and ranking, not much work has explored the fairness question in recommender systems. Recently, 2017 FATML Multisided Fairness for Recommendations [5], found that recommendation systems predicting user preferences over items would have to consider fairness from two sides: *subjects* and *objects*. Broadly speaking, subjects are the perspective of the *user* receiving the recommendation, and objects are from the perspective of the *item* being recommended. Recent work from Yao et. al [38] has investigated group-level fairness in recommender systems based on prediction accuracy across different groupings of users or items.

Our contribution in this research is to create and deploy a framework to understand and increase individual *and* group fairness in diversity for all users in physical space recommender systems, which is, until now, an understudied problem.

Chapter 5

Methodology

In this section, I propose a new algorithm to traditional recommender systems. The key changes made are to how "similarity" is defined and calculated for both collaborative and content-based filters. This recommender system focuses on values, particularly social segregation of users, and proves to decrease segregation scores by 40% more than traditional recommender systems do. We utilize Cuebiq data for the analysis. The dataset consists of anonymized records of GPS locations from users that opted-in to share the data anonymously in the Boston metropolitan area over a period of 6 months, from October 2016 to March 2017. Data was shared in 2017 under a strict contract with Cuebiq through their Data for Good program where they provide access to de-identified and privacy-enhanced mobility data for academic research and humanitarian initiatives only. All researchers were contractually obligated to not share data further or to attempt to de-identify data. Mobility data is derived from users who opted in to share their data anonymously through a General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) compliant framework.

5.1 Semantic Location Hierarchy

Before any filtering is done, we first and foremost begin with creating a semantic location hierarchy. This semantic location hierarchy is necessary because the data collected from users will be inherently sparse, as chances of users visiting the same locations in metropolitan areas is quite low. Hence, we build a semantic location hierarchy to generalize locations, eliminate the problem of sparsity, and ultimately offer better recommendations to users.

5.1.1 Categorizing Locations

There are two potential ways of categorizing locations:

1. by generalization of location descriptions

2. by how people use space

Most hierarchies categorize by generalization of location descriptions; we will be doing the same. However for future research, it is important to note the value to knowing how our users utilize spaces to give them the best recommendations. For example, coffee shop descriptions primarily are related to food, yet many users utilize coffee shop spaces as a location for studying, or holding meetings.

5.1.2 Hierarchy Layers

Regardless of how we choose to categorize our locations, our hierarchy will consist of three layers:

1. Layer 1: All individual locations visited by all users

2. Layer 2: Location sub-category (bottom layer)

3. Layer 3: Location category (upper layer)

To break down this hierarchy, in Figure 1, the 'Locations' layer consists of all the individual locations visited by all users. The 'Bottom Layer' consists of sub-categories of the groupings of various locations, while the 'Upper Layer' has more general categories encompassing many sub-categories.

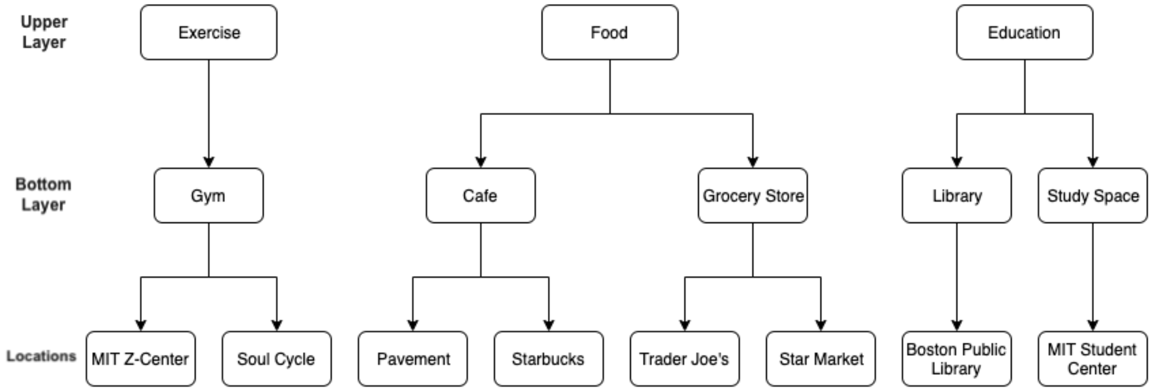

Figure 5-1: Semantic hierarchy example categorized by generalization of location descriptions

5.2 Hybrid Filtering

With the location data tiered out into more general categories, we can begin the recommendation algorithm. I construct a hybrid recommender system, using both collaborative filtering and content-based filtering. In both algorithms I investigate finding "trends" and "tiers" of users and locations. I will develop these terms as we discuss each respective algorithm, but identifying the "trends" and "tiers" of the users and locations is critical to the recommender system as these measures allow me to find equilibrium in recommending locations that a user is not only genuinely interested in, but also diversifies their experiences, thus lowering his/her social segregation.

5.3 Collaborative Filtering

Collaborative filtering models work on the assumption that similar users like similar items. With this assumption I can identify similarity in user trends, where "trends" is simply defined as a user's location habits - i.e. where he/she visits and spends time the most. In addition to identifying similarity of trends, we must also identify similarity "tiers" of users.

I define "tiers" of users as groups of users with similar segregation scores. These groups of users, or tiers, will be important when filtering as this is the first point of difference in this algorithm versus traditional recommender systems. In traditional recommender systems, users are grouped with users that are similar to them, and from those similar users, traditional recommender systems will build recommendations. Because this is a values-based recommender system focusing on diversity and social segregation, it is crucial that users are *not* grouped with similar users. Hence, when choosing what users to glean recommendations from, the "tiers" of users will prove to be critical in delivering recommendations that lower segregation scores.

The key takeaway here is that it is not enough to merely identify users with similar trends; we must additionally identify users with similar trends *and* desirable segregation scores.

5.3.1 Similarity Trends of Users

Similarity Weights at the Bottom Layer

Similarity calculation is a necessary step in predicting a set of locations that a user would find interesting and be inclined to visit. We calculate the similarity weight between users in each layer of our semantic location hierarchy.

Let C be the set of all sub-categories visited by all users. Now assume that User A and user B visited locations in sub-category C_j, m_j and m'_j times, respectively. Let w_j and w'_j be a weight value to user A and user b to C_j, respectively. We then calculate the weights:

$$\begin{aligned} w_j &= IDF_j \times m_j \\ w'_j &= IDF'_j \times m'_j \end{aligned} \tag{5.1}$$

where IDF is inverse document frequency to normalize the popularity of different sub-categories. In this case, our sub-category is synonymous to a document, and the users who check-in to a sub-category are considered terms.

This will create the following vectors for user A and user B:

$$u_{subcat_A} =< w_1, w_2, ..., w_j >$$

$$u_{subcat_B} =< w'_1, w'_2, ..., w'_j >$$

With these vectors we can use Pearson Similarity to calculate the similarity weight between subcategories of users A and B:

$$SimilarityWeightSubCategory_{A,B} = \frac{\sum_i^j (w_i - \overline{u_{subcat_A}})(w'_i - \overline{u_{subcat_B}})}{\sqrt{\sum_i^j (w_i - \overline{u_{subcat_A}})^2 (w'_i - \overline{u_{subcat_B}})^2}} \quad (5.2)$$

Category Propagation Significance Score

Typically users frequent locations under a particular category. Consider user A who frequents locations such as coffee shops and sit down restaurants, but rarely visit locations such as night clubs and bars. We can infer that this user is more interested in the categories of food than in nightlife. With this in mind, I calculate significance scores, denoted as $SigScore$, of each location category. Let $SigScore_A(X)$ be the significance score of Category X of user A.

$$SigScore_A(X) = \frac{Category_A(X)}{\sum_i^K Category_A(i)} \quad (5.3)$$

where $Category_A(X)$ is the total number of visits user A visited locations under Category X, and $\sum_i^k Category_A(i)$ is the total number of visits user A has across all k categories.

Thus, if user A frequents many locations under Category 'Food,' and few in category 'Nightlife,' then this user would have a higher significance score in the former category than the latter. When observing more users, the following vectors are created:

$$\begin{aligned} u_{cat_A} &=< SigScore_A(1), SigScore_A(2), ..., SigScore_A(k) > \\ u_{cat_B} &=< SigScore_B(1), SigScore_B(2), ..., SigScore_B(k) > \end{aligned} \quad (5.4)$$

Similarity Weights at the Upper Layers

As mentioned before, simply using $SimilarityWeightSubCategory_{A,B}$ will not be sufficient because the likelihood that majority of users visit the same subcategories in a metropolitan area is quite low. Hence, I use the derived significance scores to calculate the similarity weights of the categories in the upper layers (i.e. more general categories) of my semantic location hierarchy.

$$SimilarityWeightCategory_{A,B} = \frac{\sum_i^k (SigScore_A(i) - \overline{u_{cat_A}})(SigScore_B(i) - \overline{u_{cat_B}})}{\sqrt{\sum_i^k (SigScore_A(i) - \overline{u_{cat_A}})^2 (SigScore_B(i) - \overline{u_{cat_B}})^2}} \tag{5.5}$$

Computing Similarity Matrix between Users

With our two matrices, I apply a min-max normalization over both of them. We now can take into consideration both similarity weights found at the sub-category and category levels to determine the similarity of trends between users using α. α can be thought of as a weight parameter that scales how much to take into consideration the two similarity weights.

I shorthand $SimilarityWeightSubCategory_{A,B}$ to $SWSubCategory_{A,B}$ and $SimilarityWeightCategory_{A,B}$ to $SWCategory_{A.B}$:

$$\begin{aligned} TrendSimilarityScore_{A,B} = \alpha \times SWSubCategory_{A,B} + \\ (1 - \alpha) \times SWCategory_{A,B} \end{aligned} \tag{5.6}$$

5.3.2 Similarity Tiers of Users

The underlying motivation of our algorithm is to increase one's social diversity, synonymous to decreasing his/her's segregation score. Therefore it is important to not only identify users with similar visiting habits, but also identify those users who also have *lower* social segregation scores to derive recommendations from. I group users into tiers to introduce diversity in a more steady manner. I chose to do this because introducing a user who is highly socially segregated to a very diverse location may not be a comfortable experience for that user, and could potentially deter them from taking further recommendations. By grouping users by tiers, I can slowly introduce moer diverse locations to users in high segregation tiers by selecting recommendations from users who are slightly less segregated, but have similar location trends.

This concept is analogous to that of gradient descent, and grouping of tiers can be thought of as increasing/decreasing our step function.

A user's segregation score can be decreased over time as a step-wise function. For example, say user A's segregation score was calculated to be 85% after some period of time; we might say in the next month or so, we generate recommendations to get the user to the 75% tier. Over time a user's "goal tier" will be adjusted and decreased - let us call that t'. I find the distance between a user's goal tier, t' and the tier of all other users current tiers, t to generate tier similarity scores:

$$TierSimilarityScore_{A,B} = |t'_A - t_B| \tag{5.7}$$

Users with small distances away from our user A's desired tier, t', would be considered our top candidates.

5.3.3 Total Collaborative Filtering Similarity Matrix

Now that we have calculated the similarity score across users for both trends and tiers, I apply min-max normalization. Then we can combine the two to calculate our total similarity score, using a weight parameter β, where β scales how much to weight the consideration of similarity scores of trends amongst users, versus the consideration of similarity scores of tiers amongst users.

$$TotalSimilarityScore^{CF}_{A,B} = \beta \times TrendSS_{A,B} + (1 - \beta) \times TierSS_{A,B} \tag{5.8}$$

Where SS simply stands for "similarity score." Thus creating the total similarity score for collaborative filtering (TSS^{CF} for short) matrix across users:

$$\begin{bmatrix} 0 & TSS^{CF}_{A,B} & ... & TSS^{CF}_{A,J} & ... & TSS^{CF}_{A,N} \\ TSS^{CF}_{B,A} & 0 & ... & TSS^{CF}_{B,J} & ... & TSS^{CF}_{B,N} \\ ... & ... & ... & ... & ... & ... \\ TSS^{CF}_{J,A} & TSS^{CF}_{J,B} & ... & 0 & ... & TSS^{CF}_{J,N} \\ TSS^{CF}_{N,A} & TSS^{CF}_{N,B} & ... & TSS^{CF}_{N,J} & ... & 0 \end{bmatrix} \tag{5.9}$$

5.4 Content-Based Filtering

Content-based filtering models work on the assumption that users enjoy similar items to the items that they have previously interacted with. In this particular scenario, 'items' are our locations. I identify similarity of location trends as the following: "trends" are indicative of a locations qualitative factors, expensiveness, and much else. Additionally I identify similarity of "tiers" of location, where "tiers" is defined as how socially segregated a particular location is based off of socio-economic distributions of its visitors. Once again, the value to examining both trends and tiers of our collection of locations allows the recommender system to not only offer locations that a user would genuinely be interested in, but offer recommendations that would increase one's diversity experience and inherently decrease his/her segregation score.

5.4.1 Similarity Trends of Locations

Each location will have a vector of similarity using word embeddings. In this case, the FourSquare dataset is the most dense in the description of subcategories of locations. Using word embedding vectors from NLP library SpaCy, I then measure the similarity between locations P and Q's subcategory word embedding using cosine similarity.

As seen in Figure 5-2, the word embddings have successfully clustered these subcategories. For example, there is clustering of food related categories such as bars, restaurants, and categories of foods like sandwiches, seafood, and candy; buildings related to academia and community services; transportation; and recreation areas.

5.4.2 Growth of Tiers of Locations

Similarly to the collaborative filtering portion of this system, it is necessary to consider both the trends and tiers of the locations. I define "tier" of location as a classification of externality; the externality in this case being the distribution of socio-economics per location. However, instead of looking for locations with similar externality measures, I instead look for a "growth of tiers", meaning I am interested in locations that have incremental growth in diversity to user A.

With each location having it's own segregation score derived from the distribution of visitors socio-economics, I calculate the growth of tiers between locations p and q for user A using the following distance function:

$$\begin{split} GrowthOfExternalities(A)_{p,q} = SegregationScoreImpact(q, A) - \\ SegregationScoreImpact(p, A) \end{split} \tag{5.10}$$

$SegregationScoreImpact$ is a function that measure how much a location impacts a user's segregation score. $GrowthOfExternalities$ then takes the difference between those impact scores to identify how much positive or negative impact location p has over location q for user A's diversity.

5.4.3 Total Content-Based Filtering Similarity Matrix

We conduct a very similar calculation to that of finding collaborative filtering similar users matrix:

$$TotalSimilarityScore^{CBF}_{p,q} = \gamma \times SimOfLocs_{p,q} + (1-\gamma) \times GrowthOfExt_{p,q} \tag{5.11}$$

Where γ is a weight parameter that scales the consideration of $SimOfLocs$ against $GrowthOfExt$ in calculating the total similarity score of all locations.

Thus creating the total similarity score for our content-based filtering algorithm

(TSS^{CBF} for short) matrix across all locations:

$$\begin{bmatrix} 0 & TSS_{a,b}^{CBF} & ... & TSS_{a,j}^{CBF} & ... & TSS_{a,m}^{CBF} \\ TSS_{b,a}^{CBF} & 0 & ... & TSS_{b,j}^{CBF} & ... & TSS_{b,m}^{CBF} \\ ... & ... & ... & ... & ... & ... \\ TSS_{j,a}^{CBF} & TSS_{j,b}^{CBF} & ... & 0 & ... & TSS_{j,m}^{CBF} \\ TSS_{m,a}^{CBF} & TSS_{m,b}^{CBF} & ... & TSS_{m,j}^{CBF} & ... & 0 \end{bmatrix} \tag{5.12}$$

5.5 Giving the recommendation

Using the two total similarity matrices, one from collaborative filtering, and the other from content-based filtering, we can gather the set of potential locations to recommend.

5.5.1 Top Locations from TSS^{CF} (Collaborative Filtering)

The TSS^{CF} is a n-users by n-users matrix, with similarity calculations between each pair of users. From this matrix I pull user A's top 10 most "similar" users, and take the total set of each similar user's top 10 locations, $TopLocations_{CF}$. From $TopLocations_{CF}$, I calculate the respective top similar user's rating of that location based off of how frequently they visited that location. Finally, I sort these locations based off their ratings, and return the top 20 locations from our TSS^{CF}, $Top20^{CF}$.

5.5.2 Top Locations from TSS^{CBF} (Content-Based Filtering)

The TSS^{CBF} is a m-locations by m-locations matrix, with similarity calculations between each pair of locations. From this particular matrix, I pull the most impactful and similar locations to each location, creating a total set of top locations, $TopLocations_{CBF}$. I sort these locations based off their scores, and return the top 20 locations from our TSS^{CBF}, $Top20^{CBF}$

5.5.3 Combining to the Final Recommendation

With two recommended lists of locations, $Top20^{CF}$ and $Top20^{CBF}$, I then use min-max scaling to normalize the scores of both lists. After normalizing, I use the following to combine the scores of the lists of locations.

$$FinalRecommendations = \delta \times Top20^{CF} + (1 - \delta) \times Top20^{CBF} \tag{5.13}$$

where δ is once again a weight parameter that can be thought of as a weighting the scale of collaborative filter and content-based filter recommendations.

With the scores normalized and weighted once more, I sort the scores one last time, and return the top ten locations as the final recommendation of locations to the user.

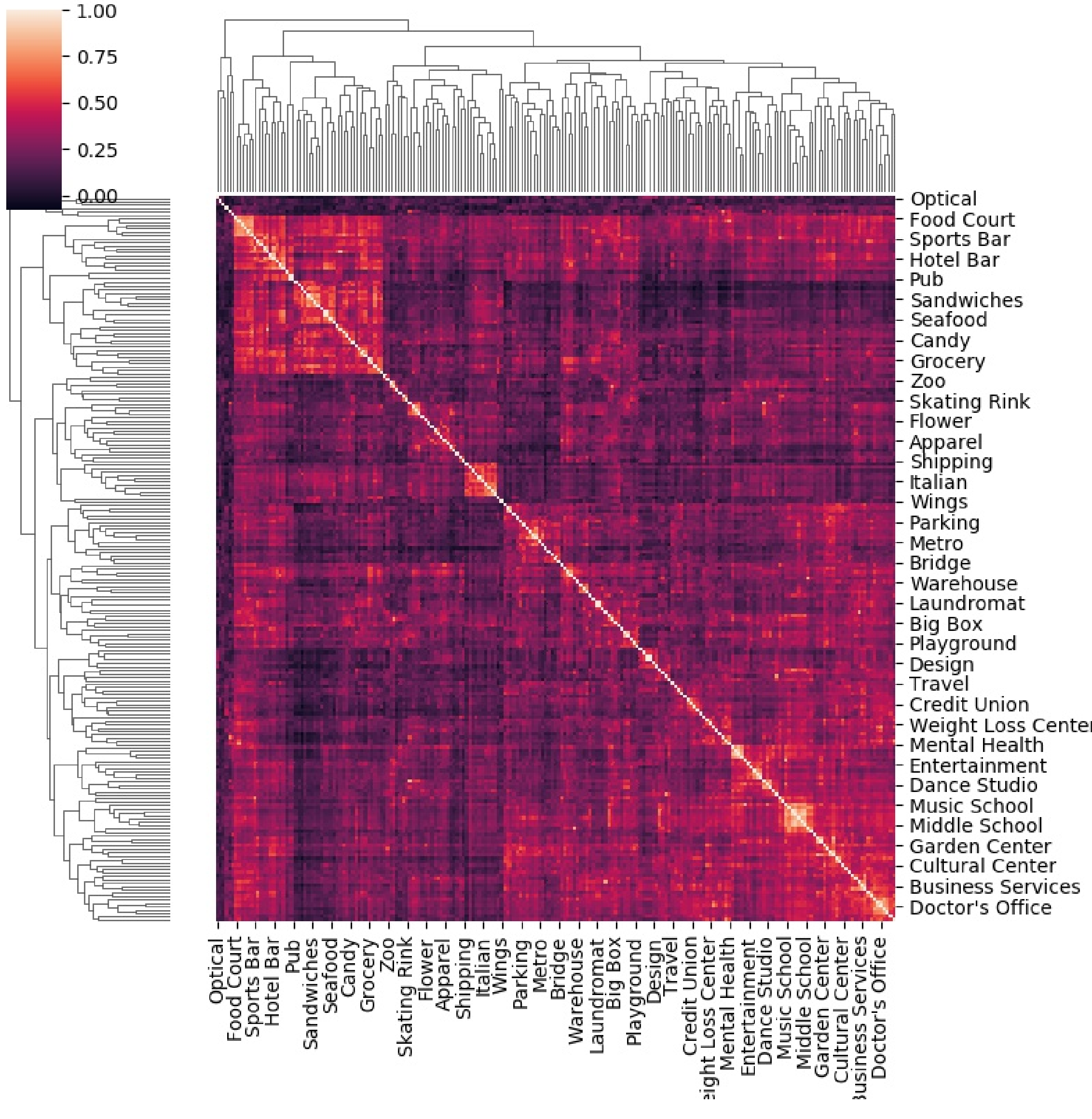

Figure 5-2: Clustering of subcategory description of locations

Chapter 6

Results

In this section, I will outline the results. The key takeaways are that my algorithm improves performance on decreasing social segregation drastically. Because the algorithm considers a set of users in a "goal tier" segregation score, manipulating what that goal tier is improves diversity scores against the baseline by as much as three-fold. While this is the case, it's important to keep in mind that while decreasing segregation scores is the motivation, we must maintain the balance of giving recommendations users would actually take. Hence, setting a user's goal tier to 40% less than their current score for recommendations, is not entirely reasonable. For this reason, I maintain a decrement of 10% which continuously introduces diversity without introducing too much discomfort, and decreases segregation scores by 20% more than the baseline algorithm. When taking a closer look at the parameters in the algorithm, I find that the consideration of externalities of locations are the most important, and even a 10% consideration of externalities can decrease segregation scores by 40% more than traditional recommender systems. This highlights that traditional content-based filters, which rely solely on similarity semantics of items, are clearly not enough to promote larger values on its users.

6.1 Methods

The analysis of my results were conducted on a dataset of 1,000 users' location histories. For privacy reasons, user locations are masked to the level of place categories. Segregation scores metrics were obtained from the Atlas of Inequality public data. We tag locations with categorical semantics, check-in history, and segregation score metrics based off of all visitors' socio-economics.

I chose a traditional memory-based collaborative filtering recommender to use as a baseline and compare the impact that the baseline and my algorithm have on decreasing user segregation scores. I sample 50 random users, receive the top 10 recommendations from both systems, and re-calculate the respective segregation score, assuming that each user takes the 10 recommendations.

6.2 Traditional Memory-Based CF (Baseline) vs. Ecological Values-Based System

In Section 5.3.2, I detail the calculation of 'Similarity Tiers of Users' in the collaborative filtering portion of my algorithm. The goal of this part was to add an additional feature when selecting "similar" users by putting more weight on users that were in a goal tier segregation score, i.e. a lower segregation score group. This metric is denoted as t', or how much lower than a user's current score do we want our goal tier to be.

Table 6.1: Percentage decrease of test user's segregation score

Recommender System	**Average**	**Std. dev.**
Baseline	11.8%	2.3%
Ecological Values-Based (t'=10%)	13.2%	2.4%
Ecological Values-Based (t'=20%)	21.4%	3.1%
Ecological Values-Based (t'=30%)	29.6%	2.9%
Ecological Values-Based (t'=40%)	35.9%	3.7%

The results above are expected when t' is set to such a high number as this effectively weights users in a far lower segregation tier with much more significance. While the goal is to decrease social segregation of users, there are caveats with giving recommendations that are too different, too diverse, too soon. The biggest being that users will not take the recommendation because it is out of their comfort zone, and/or don't align with their current interests. Hence, we want to introduce diversity incrementally over time such that users can ease into this change naturally. I choose to stick with a t' of 10% as it on average, decreases test users' segregation scores by 20% more than traditional memory-based collaborative filters, and doesn't risk introducing recommendations that a user would not be inclined to take.

Now let us take a closer look at the recommendation themselves to gain a better understanding of what types of recommendations are made. After sampling 50 users, I took a look at the recommendations generated from the baseline and our ecological values-based system. In the dataset, each location has data on it's own socio-economic distribution broken into four categories, ranging from those in the lowest-income bracket (p1a), to those in the highest (p4a). Using this information I calculated the average distribution of socio-economics of the recommended locations of each recommender system.

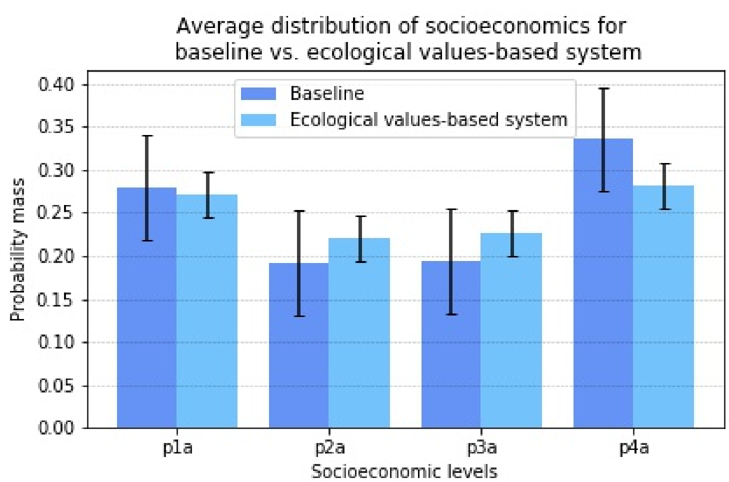

Figure 6-1: Average distribution of socioeconomics for top 10 baseline vs. ecological values-based system recommended locations for 50 users

It is clear from the figure above, that the ecological values-based recommender system does a much better job of recommending locations that have equal distribution and representation in its visitors than the traditional collaborative filtering recommender system. Figure 6-1 supports the notion that traditional ways of approaching recommender systems are not focused on supporting user diversity, and ultimately detract from promoting a more diverse group of users, or ecology of users.

6.3 Impact of parameters

In order to tune the weight parameters (α, β, γ and δ) and see the impact each parameter has independently on the test user's segregation score, I select 25 users at random to run the baseline and thesis algorithm on. The weight parameters are for the following equations:

1. $TrendSimilarityScore_{A,B} = \alpha \times SWSubCategory_{A,B} + (1-\alpha) \times SWCategory_{A,B}$
2. $TotalSimilarityScore^{CF}_{A,B} = \beta \times TrendSS_{A,B} + (1-\beta) \times TierSS_{A,B}$
3. $TotalSimilarityScore^{CBF}_{p,q} = \gamma \times SimOfLocs_{p,q} + (1-\gamma) \times GrowthOfExt_{p,q}$
4. $FinalRecommendations = \delta \times Top20^{CF} + (1-\delta) \times Top20^{CBF}$

where α is a weight parameter that that scales the category used to define semantic similarity; β scales the consideration of similarity of trends amongst users versus similarity of tiers amongst users; γ weighs the consideration of similarity of locations between the growth of externality of locations; and lastly, δ can be thought of as weighting the scale between our two sides of the recommender system.

While testing each of the parameters, I want to set the other parameters such that they do not contribute any more than the traditional recommender system implementation would. So, I hold the other parameters at the following: $\alpha = 1.0, \beta = 1.0, \gamma = 1.0$, and $\delta = 0.5$. Setting $\alpha, \beta, \gamma = 1$ ensures that the semantic location hierarchy tree, categorization of tiers of users, and the externality of locations, are all not accounted for. I set $\delta = 0.5$ to pull equally from both sides of the recommender system to not interfere with the measures of other parameter contributions.

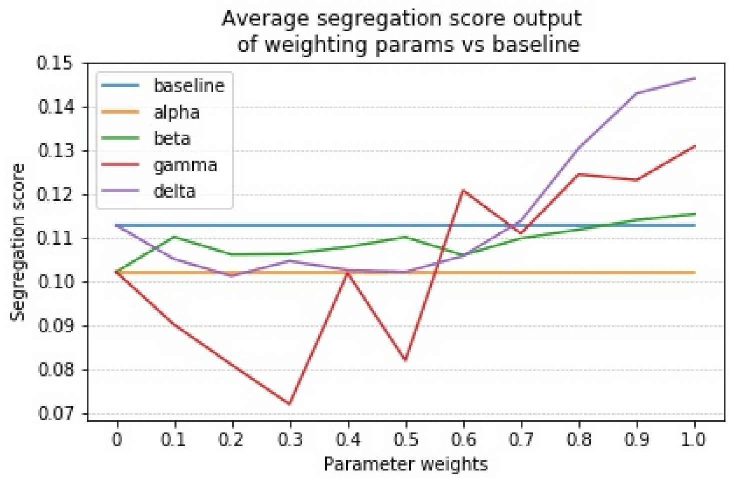

Figure 6-2: Average segregation score output when weighting params vs baseline

Figure 6-2 is a snapshot of the average segregation score output across 25 users when weighting the parameters independently. The key takeaways here are the impact of the γ and δ parameters on segregation scores. γ scales the weight that growth of externality of locations is taken into consideration, and the more it weighs the similarity of locations (higher γ) the less the algorithm is able to decrease segregation score. This plays into the performance of δ since α, β, γ are held constant, the lack of impact that the collaborative filter has independently, leads δ to perform worse as it approaches a weight of 1.

6.3.1 α, Semantic Hierarchy Location

α is a weight parameter that that scales the category used to define semantic similarity, As seen in Figure 6-3, there is no impact of the α parameter on the segregation score, as our ecological values-based system returns the same segregation score for the test users across all α's. Moreover, while our algorithm performs better than the baseline, it is only by the slightest of margins. This shows a weakness in existing memory-based collaborative filtering algorithms, with regards to preserving values, or in this case social segregation because it is unable to perform better than our system. Our system performs slightly better than baseline because while we are weighting α, we are keeping β, γ and δ constant. β, γ and δ in their constant state means that δ is

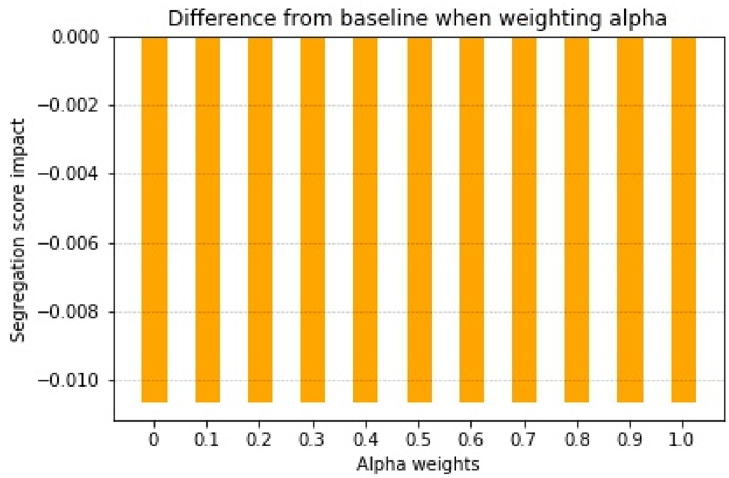

Figure 6-3: α impact with β, γ, δ held constant

held at 50%, and is taking consideration of outputs from both the collaborative filtering side and the content-based filtering side. The content-based filter, despite being just a normal content-based filter, has a slight edge in introducing more diversity, hence resulting in the ecological values-based system to outperform the baseline.

But this lack of impact suggests that semantic hierarchy of locations is not an advancement to improve diversity of results of recommender systems, but only an advancement in combating data sparsity. Therefore, The lack of impact that semantic hierarchies have only go to show the disregard that these "improvements to recommender systems" have on the diversity of its users.

6.3.2 β, Categorization of Tiers of Users

β scales the consideration of similarity of trends amongst users versus similarity of tiers amongst users. In the above figure I vary β, and as β approaches 0, the ecological values-based recommender performs better at lowering segregation score. It is clear in Figure 6-4 that β weight of 0 performs best against the traditional CF recommender in lowering segregation score. This means that the selection of users from a different segregation tier to draw recommendations from has a bigger impact on lowering segregation scores, than merely selecting users with the same trends of habit. This is a clear indication that without selecting users from a different segrega-

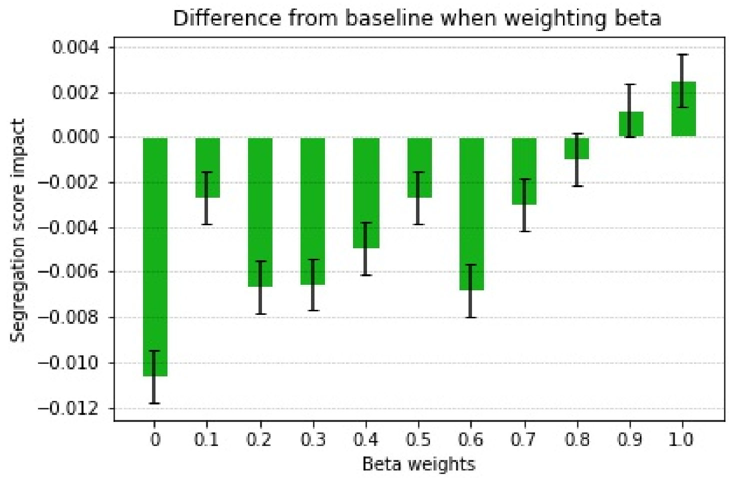

Figure 6-4: β impact with α, γ, δ held constant

tion tier and relying solely on similarity, existing memory-based collaborative filters are not particularly supporting the diversification of its users.

6.3.3 γ, Externality of Locations

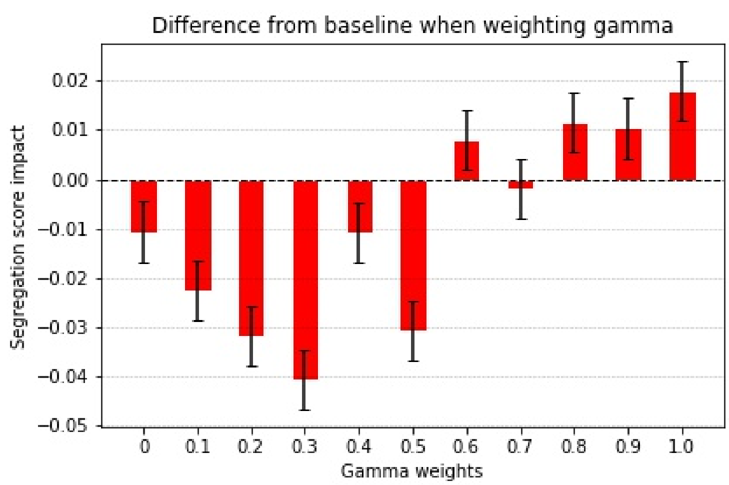

Figure 6-5: γ impact with α, β, δ held constant

γ weighs the consideration of similarity of locations between the growth of externality of locations. In Figure 6-5, it is clear that the similarity of locations generated by the word embeddings has very little impact in decreasing the test user's segregation score (i.e. when $\gamma = 1$). In fact, the more the recommender systems weights the externality of locations, the better it performs at decreasing social segregation, and effectively introducing the test user to more diverse locations. This result illustrates that traditional content-based filters, which rely solely on similarity of semantics of items, is clearly not enough to promote larger values on users. Furthermore these results speak to the necessity to re-examine how recommender systems are evaluated as "accurate." The drastic impact by examining externality data is show here, where taking even a 10% consideration of the $GrowthOfExternalities$ ($\beta = 0.9$) can decrease segregation scores by 40%.

6.3.4 δ, Final Recommendation Weighting

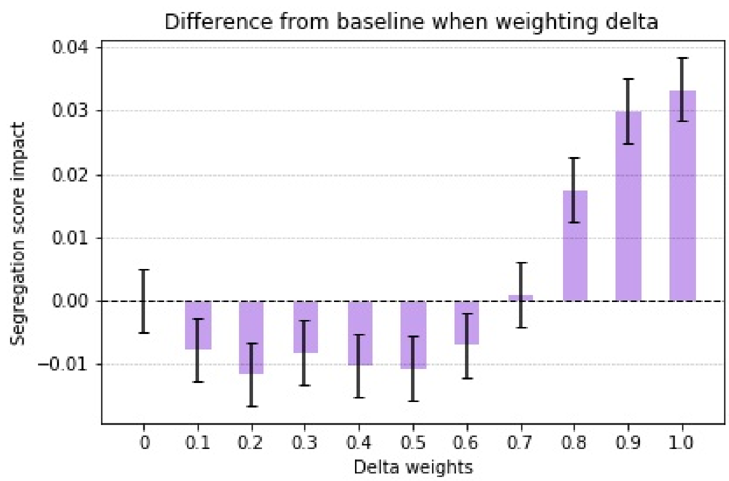

Figure 6-6: δ impact with α, β, γ held constant

δ is a scaling parameter that weighs the two sides of the recommender system. In Figure 6-6, a weighting of $\delta = 0.2$ proved to perform best against the traditional collaborative recommender. This means that in order to decrease segregation score, the system should weight more heavily the content-based filter than the collaborative filter. This follows in line with our results of the impact of α, or lack there of, and why our system would weight the results from the content-based filter higher.

While both collaborative and content-based filters are the two most common techniques in recommender systems, the results depict that without content-based filtering, existing collaborative systems cannot decrease segregation scores as effectively.

Chapter 7

Discussion

This ecological values-based recommender system has proven that traditional approaches to recommender systems are not equipped to tackle recommendations beyond the realm of popularity. The linear way of evaluating accuracy as simply how well matched recommendations are to a user's prior interactions, is extremely limiting.

Thus it is of utmost importance to start a discussion on redefining accuracy in recommender systems. Accuracy at the user level, while important, should not be the singular goal. Unfortunately, this rhetoric has rendered traditional recommender systems to be ill-conducive to diversifying an ecology of users, and homogenization poses a real threat with recommender systems affecting the way we consume news, information, and products constantly. Figure 6-3 is a prime example of an improvement that has been lauded in prior works for improving accuracy, but this feature does nothing to improve diversity.

Furthermore, because of the literature's heavy focus on measuring recommendation accuracy, rather than accuracy *and* values, there is very little research that deviates from this metric. Lack of deviation from the metric brings a lack of change in the ways we go about implementing recommender systems. All descriptions of traditional recommender systems are founded on the fundamental belief that "likes like likes." While true, it has been proven that this leads to homogenization of users, and a lack of overall diversity. The results from Figure 6-4 and Figure 6-5 are key, as they reveal the impact that changing the notion of similarity has on diversity for collaborative and content-based filters. These figures should encourage future research to rethink and redefine similarity, but additionally collect data beyond ratings and critically analyze different ways to generate recommendations beyond popularity of the masses.

With adjustments to different ways of rating content, redefining similarity of users and items, and focusing on bigger group impact, recommender systems can begin to shift towards really diversifying user experiences in a positive way.

Bibliography

[1] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. Managing popularity bias in recommender systems with personalized re-ranking. In *The Thirty-Second International Flairs Conference*, 2019.

[2] Panagiotis Adamopoulos and Alexander Tuzhilin. On over-specialization and concentration bias of recommendations: Probabilistic neighborhood selection in collaborative filtering systems. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 153–160, 2014.

[3] LSI aDepartament de Llenguatges i Sistemes Informàtics. Taking advantage of semantics in recommendation systems. In *Artificial Intelligence Research and Development: Proceedings of the 13th International Conference of the Catalan Association for Artificial Intelligence*, volume 220, page 163. IOS Press, 2010.

[4] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. Equity of attention: Amortizing individual fairness in rankings. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 405–414. ACM, 2018.

[5] Robin Burke. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093*, 2017.

[6] Erion Çano and Maurizio Morisio. Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*, 21(6):1487–1524, 2017.

[7] Siva Kumar Cheekula, Pavan Kapanipathi, Derek Doran, Prateek Jain, and Amit P Sheth. Entity recommendations using hierarchical knowledge bases. 2015.

[8] Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*, pages 271–280. ACM, 2007.

[9] Mostafa M. El-Bermawy. Your filter bubble is destroying democracy.

[10] Jon Evans. Facebook isn't free speech, it's algorithmic amplification optimized for outage.

[11] Daniel Faggella. The roi of recommendation engines for marketing.

[12] Daniel M Fleder, Kartik Hosanagar, Andreas Buja, et al. Recommender systems and their effects on consumers: the fragmentation debate. *EC*, 229:230, 2010.

[13] Pratik Gajane and Mykola Pechenizkiy. On formalizing fairness in prediction with machine learning. *arXiv preprint arXiv:1710.03184*, 2017.

[14] Moritz Hardt, Eric Price, Nati Srebro, et al. Equality of opportunity in supervised learning. In *Advances in neural information processing systems*, pages 3315–3323, 2016.

[15] Juzozas Kaziukenas. Amazon's retail business is now just 50% of the company's revenue.

[16] Matevž Kunaver and Tomaž Požrl. Diversity in recommender systems–a survey. *Knowledge-Based Systems*, 123:154–162, 2017.

[17] Juan Lang, Matt Spear, and S Felix Wu. Social manipulation of online recommender systems. In *International Conference on Social Informatics*, pages 125–139. Springer, 2010.

[18] Paul Lewis. How youtube's algorithm distorts truth.

[19] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4):2065–2073, 2014.

[20] Michael Martin. White house social media summit recap.

[21] Jane Mayer. How russia helped swing the election for trump.

[22] Colin Moynihan. The anarchists combatting the far right on youtube.

[23] Al Mamunur Rashid, Istvan Albert, Dan Cosley, Shyong K Lam, Sean M McNee, Joseph A Konstan, and John Riedl. Getting to know you: learning new user preferences in recommender systems. In *Proceedings of the 7th international conference on Intelligent user interfaces*, pages 127–134, 2002.

[24] Mariel Soto Reyes. Google made nearly as much ad revenue from news as us publishers made overall.

[25] Norma Saiph Savage, Maciej Baranski, Norma Elva Chavez, and Tobias Höllerer. I'm feeling loco: A location based context aware recommendation system. In *Advances in Location-Based Services*, pages 37–54. Springer, 2012.

[26] J Ben Schafer, Joseph Konstan, and John Riedl. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 158–166, 1999.

[27] Harvard Business School. How spotify is changing the way we consume music.

[28] Vindu Goel Scott Shane. Fake russian facebook accounts bought $100,000 in political ads.

[29] Hamza Shaban. Twitter reveals its daily active user numbers for the first time.

[30] Scott Shane. The fake americans russia created to influence the election.

[31] Daniel Kreiss Shannon C. McGregor. Conservatives say google and facebook are censoring them.

[32] Spotify. Spotify technology s.a. announces financial results for fourth quarter 2018.

[33] Yuichiro Takeuchi and Masanori Sugimoto. Cityvoyager: an outdoor recommendation system based on user location history. In *International Conference on Ubiquitous Intelligence and Computing*, pages 625–636. Springer, 2006.

[34] Zeynep Tufekci. Youtube, the great radicalizer.

[35] Margery Austin Turner and Lynette Rawlings. Promoting neighborhood diversity: Benefits, barriers, and strategies. *The Urban Institute*.

[36] Dan Tynan. Facebook accused of censorship after hundreds of us political pages purged.

[37] Dingqi Yang, Daqing Zhang, Zhiyong Yu, and Zhu Wang. A sentiment-enhanced personalized location recommendation system. In *Proceedings of the 24th ACM conference on hypertext and social media*, pages 119–128. ACM, 2013.

[38] Sirui Yao and Bert Huang. Beyond parity: Fairness objectives for collaborative filtering. In *Advances in Neural Information Processing Systems*, pages 2921–2930, 2017.

[39] Mao Ye, Peifeng Yin, and Wang-Chien Lee. Location recommendation for location-based social networks. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pages 458–461. ACM, 2010.

[40] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 325–334. ACM, 2011.

[41] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. Fairness constraints: Mechanisms for fair classification. *arXiv preprint arXiv:1507.05259*, 2015.

[42] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. Fa* ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1569–1578. ACM, 2017.

[43] Yubao Zhang, Xin Ruan, Haining Wang, Hui Wang, and Su He. Twitter trends manipulation: a first look inside the security of twitter trending. *IEEE Transactions on Information Forensics and Security*, 12(1):144–156, 2016.

[44] Renjie Zhou, Samamon Khemmarat, and Lixin Gao. The impact of youtube recommendation system on video views. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC '10, page 404–410, New York, NY, USA, 2010. Association for Computing Machinery.