

2011

Recommender Systems and their Effects on Consumers

Daniel Fleder

University of Pennsylvania, dfleder@wharton.upenn.edu

Kartik Hosanagar

University of Pennsylvania, kartikh@wharton.upenn.edu

Andreas Buja

University of Pennsylvania, buja@wharton.upenn.edu

Follow this and additional works at: <http://aisel.aisnet.org/mcis2011>

Recommended Citation

Fleder, Daniel; Hosanagar, Kartik; and Buja, Andreas, "Recommender Systems and their Effects on Consumers" (2011). *MCIS 2011 Proceedings*. 59.

<http://aisel.aisnet.org/mcis2011/59>

This material is brought to you by the Mediterranean Conference on Information Systems (MCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in MCIS 2011 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Recommender Systems and their Effects on Consumers

Daniel Fleder
OPIM Department
The Wharton School
University of Pennsylvania
dfleder@wharton.upenn.edu

Kartik Hosanagar
OPIM Department
The Wharton School
University of Pennsylvania
kartikh@wharton.upenn.edu

Andreas Buja
Statistics Department
The Wharton School
University of Pennsylvania
buja@wharton.upenn.edu

ABSTRACT

Recommender systems are becoming integral to how consumers discover media. The value that recommenders offer is personalization. In environments with many product choices, recommenders personalize the experience to each user's taste. Popular applications include product recommendations at online retail sites and online newspapers' automatically selecting articles to display based on the current reader's interests. This ability to focus more closely on one's taste and filter all else out has spawned criticism that recommenders will fragment users. Critics say recommenders cause consumers to have less in common with one another and that the media should do more to increase exposure to a variety of content. Others, however, contend that recommenders do the opposite: they may homogenize users because they share information among those who would otherwise not communicate. These are opposing views, discussed in the literature for over fourteen years, for which there is not yet empirical evidence. We present an empirical study of recommender systems in the music industry. In contrast to concerns that users are becoming more fragmented, we find that in our setting users become more similar to one another in their purchases. This increase in similarity occurs for two reasons, which we term volume and taste effects. The volume effect is that consumers simply purchase more after recommendations, increasing the chance of having more purchases in common. The taste effect is that, conditional on volume, consumers buy a more similar mix of products after recommendations. When we view consumers as a similarity network, the network becomes denser and smaller (characterized by shorter inter-user distances) after recommendations are introduced. These findings suggest that for this setting, recommender systems are associated with an increase in commonality among users and that concerns of fragmentation may be misplaced.

Keywords

Recommender systems, personalization, fragmentation, collaborative filtering, long tail

1. INTRODUCTION

Recommender systems are becoming integral to how consumers discover media. The value that recommenders offer is personalization. For example, a personalized radio station plays music not for the general public but for each particular user; a personalized newspaper does not show the same front page to everyone but customizes it for each reader. A retailer arranges its online shelves based on who is browsing at that moment. Such personalization is valuable in modern media markets, which can have millions of products to choose from, and the trend toward personalization is occurring for all major types of media, including books, music, movies, news, and television.

Along with the benefits of personalization, however, a debate has emerged as to whether it has drawbacks. Personalizing websites means that we may no longer see the same newspaper articles, television shows, or books as our peers. Critics thus argue that recommender systems will create fragmentation, causing users to have less and less in common with one another. An alternative view contends that recommenders may do the opposite: recommenders may have homogenizing effects because they share information among users who otherwise would not communicate. This paper presents empirical evidence for the debate on whether recommenders fragment versus homogenize users.

The motivation is two-fold. First, from a technology-policy perspective, the literature has expressed concern that fragmentation is a negative consequence. These critics suggest that the media and government should do more to increase exposure to a variety of content. For example, proposals include having highly partisan sites hyperlink to sites with differing points of view, to offset the effects of recommenders and ensure users are "alerted to materials other than those [recommended]" ([11] p. 187). In contrast, finding evidence of homogenization would suggest that such policies and regulation of the internet are not warranted. Second, the fragmentation question has marketing implications. If more fragmentation occurs, narrow, targeted marketing policies appear more justified. If not, consumers may prefer a range of experiences that narrow targeting does not deliver.

We find, in an empirical study of a major music recommendation service, that the service's recommendations are associated with an increase in commonality among consumers, as defined by similarity in their purchases. This increase in similarity occurs for two reasons, which we term volume and taste effects. The volume effect is that consumers simply purchase more after recommendations, increasing the chance of having purchases in common with one another. The taste effect is that consumers buy a more similar mix of products after recommendations, conditional on volume. When we view consumers as a similarity network before versus after recommendations, we find that the network becomes denser and smaller, or characterized by shorter inter-user distances.

Although it is possible that other recommender designs will have different effects, our results based on a real-world implementation at a major online firm clearly show that recommendations can create homogeneity and the argument that they creating fragmentation is not universally true.

2. PRIOR WORK

Although a large body of work exists on designing recommender systems, we know much less about how they affect the market and society. This is despite the hundreds of papers presenting new recommender algorithms and millions of transactions occurring through them. This paper continues a stream of work in that direction. Recent work ask how recommenders affect *products*: which products gain versus lose sales due to recommenders and whether recommenders increase the market for niche goods (or "long tail") [6][7][9][10]. This paper asks the complementary question of how recommenders affect *consumers*: whether they cause consumers to have more or less in common with one another.

A range of views exist as to whether recommenders will fragment versus homogenize users. The strongest proponent of the fragmentation view is perhaps Cass Sunstein from the law community. Sunstein argues that recommenders create fragmentation by limiting users' media exposures to their predefined, narrow interests. These fragmentation effects, he argues, are undesirable. "In a democracy people do not live in echo chambers or information cocoons. They see and hear a wide range of topics and ideas, ... even if they did not ... choose to ... in advance" [11]. While Sunstein is clear in explaining why fragmentation may be undesirable, the antecedent, that recommenders create fragmentation to begin with is ultimately an assumption.

A second view suggesting fragmentation is that of Pattie Maes, creator of one of the first recommender systems. Maes says that recommenders can have a "narrow-minded" and "hyperpersonalized" aspect. "You don't want to see a movie just because you think it's going to be good. It's also because everyone [else is] ... talking about it, and you want to be able to talk about it too" [12]. Consuming the same media and products "is a way of participating in society," and this could be lost on account of recommender systems (paraphrased in [12]).

Sunstein and Maes both appear to view recommenders as causing fragmentation, but they differ in their views as to why this is undesirable. Sunstein argues that a democracy requires citizens to have a range of experiences and viewpoints. For example, in news programming, users should be exposed to multiple views on a topic, not just the one that reinforces their existing beliefs – which he believes will be the case as recommenders become more prevalent. Maes' has a different criticism: a product's popularity has a positive externality, and recommenders may cause us to forfeit this. If there is a benefit to reading the same books as others or seeing popular movies (e.g., by being able to discuss the experience with others) we should be wary of recommenders because these benefits could disappear.

Van Alstyne and Brynjolfsson [13] formalize a mixed view in an economic model. They ask whether internet technologies like recommender systems will lead to fragmentation versus homogenization – in their terms, a cyber-Balkans versus a global village. Fragmentation is measured both by physical interaction and consumers' knowledge overlap. They show that as technology lowers search costs and communication costs, either outcome can occur. Which outcome occurs in their model depends on a parameter representing consumers' taste for specialization. This parameter is difficult to specify, and so complementary empirical work is needed.

Similar mixed views were shared by the creators of early collaborative filters. At the Berkeley Collaborative Filtering Workshop in 1996, a time at which research on recommender systems was just beginning, Paul Resnick asked if the "global village [would] fracture into tribes." John Riedl, co-inventor of one of the first recommenders, asked if collaborative filtering would "democratize ... information ... or result in social fragmentation" [2].

3. PROBLEM FORMULATION

This section defines the problem formally. While many authors have discussed the fragmentation question qualitatively, the empirical question has not been posed in concrete terms.

3.1 Research Questions

Our goal is to study whether recommenders make users more or less similar to one another. We divide the question in two components:

1. Aggregate level: overall, are consumers farther or closer to one another?
2. Individual level: are there differential effects at the individual level, by which some users become closer and others farther?

The meaning of far and close will soon be made precise.

3.2 Two Group Design

The analysis design throughout is analogous to a two-group experiment. One group is “treated” with recommendations and their behavior compared before versus after. A control group is not treated with recommendations, and their behavior is compared over the same period. The data are in fact observational, as we will discuss, but the terminology of experiments simplifies the writing.

Let O_{it} denote an observation on group i during time period t . O_{it} is a list of tuples (user, artist, # songs purchased) for all users in group i during period t . Group $i = 1$ is the treated group, which is unexposed to the recommender during $t = 1$ but exposed to the recommender during $t = 2$. Group $i = 2$ is the control, which is unexposed to the recommender during both time periods. The time periods are the same for both groups. The schematic below shows this setup, where X denotes exposure to recommendations.

Schematic of the Two Group Design

Treated:	O_{11}	X	O_{12}
Control:	O_{21}		O_{22}

Using this design, we can compare the treated group before and after recommendations. We can also compare the treated group to the control over the same period. The control accounts for factors such as time trends and maturation that might be confounded with recommender usage in a one group pre-post design [4].

3.3 Hypotheses to Test

We wish to compare how the treated and control groups change over time. Let $T(O_{it})$ be some statistic of interest on O_{it} measuring fragmentation. As shorthand, we will write T_{it} . We define the following quantities of interest:

$$\text{Difference in treated: } D_1 \equiv T_{12} - T_{11}$$

$$\text{Difference in control: } D_2 \equiv T_{22} - T_{21}$$

$$\text{Difference-in-differences: } D \equiv D_1 - D_2$$

D_1 describes changes in the treated group. D_2 describes changes in the control. D describes how much changes in the treated group exceed those in the control. For this reason, D is termed the difference-in-differences estimator. For example, suppose that independent of recommendations, a time trend is occurring in the music industry that affects both groups. Thus observing $D_1 \neq 0$ does not mean recommendations have an effect on consumers because the same trend will affect D_2 . However, the difference-in-differences estimator D can identify changes in the treated group beyond the time trend by subtracting the change in the control. Let $\mu \equiv E[D]$, where D 's distribution is not known to us. The central questions of this paper take the form:

$$H_0: \mu \equiv E[D] = 0$$

$$H_a: \mu \equiv E[D] \neq 0$$

4. FORMULATION SPECIFICS

This section defines the quantities of interest $T(O_{it})$. To facilitate this, we take the intermediate step of defining a network $G(O_{it})$ among the firm's consumers and making $T(G(O_{it}))$ a function of that network. At first glance, introducing networks appears to complicate the analysis by adding an extra step. In contrast, we will see this provides a great service for interpreting the data.

Mathematically, our network is a graph made of nodes and edges. Users are the nodes, and the edge weights describe the similarity between user pairs, as defined by commonality in purchases. For notation, we can interpret O_{it} as a users \times artists matrix of purchase counts. An element $(O_{it})_{xy}$ is the number of songs user x purchased of artist y . We define an unweighted network in which users x and y have an edge between them if they purchase at least one artist in common.

$$\begin{aligned} & \text{Unweighted} \\ (G_{it})_{xy} & \equiv \begin{cases} 1, & \text{if users } x \text{ and } y \text{ have } \geq 1 \text{ artist in common } ((O_{it})_x \cdot (O_{it})_y \geq 1) \\ \text{Unconnected,} & \text{otherwise} \end{cases} \end{aligned}$$

The \cdot symbol indicates the vector dot product. There are many other ways to construct the network. In the online appendix, we present results for other network definitions, both unweighted and weighted.¹

With the network $G(O_{it})$ defined, we next define summary statistics of the network's properties, $T(G(O_{it}))$. T summarizes a particular network property and thus facilitates comparisons of the network over time. We define three such measures below.

Density. Here $T(G(O_{it})) \equiv$ the fraction of edges that exist out of the total number of edges possible. Higher density means users have more connections among them.

Median Degree. Here $T(G(O_{it})) \equiv$ the number of connections to other users that the typical (median) user has. The higher the median degree, the more similar users are to one another. This represents the median of the degree distribution, whereas the density is the degree distribution's average.

Path Length. Here $T(G(O_{it})) \equiv$ the shortest distance between any two users, averaged over all users in the network. If users x and y are connected, the shortest distance is 1, the edge between them. Otherwise, the path is through other users. The shorter this distance, the "smaller" the network is said to be, in the terminology of Watts & Strogatz [14], who studied "small world" networks.

5. DATA

5.1 Data Source

We study the fragmentation question using data from a major online music recommendation service, referred to here as Service. Service is a free software add-on to Apple's iTunes. iTunes, in turn, is the music player that allows users to buy music from Apple's iTunes store, the largest music retailer in the U.S. [1]. When users listen to music in iTunes, Service recommends other songs that the user may like. The recommendations appear in a window appended to iTunes, where the user can sample the recommended songs and opt to purchase them. If a purchase results, Service earns a commission. Service also provides a website where users can view the play histories of other Service users with similar taste. These play histories are uploaded automatically by the plugin to Service's website on a continual basis. The recommendations are based on the artist currently playing (i.e., the query to obtain recommendations is the current artist). Based on the current artist, Service's content-based recommender identifies the 6 most similar artists and populates the window with this list.

5.2 Novelty of the Data

To study the effects of recommenders, a contrast is needed between users exposed and unexposed to recommendations. The data collected by most retailers (e.g., Amazon, Netflix) is inadequate because retailers only observe consumers after they arrive at their website and hence after exposure to recommendations. This may be the reason, we speculate, that others have not been able to study the fragmentation question. Our data are novel in this regard. When a user registers for Service, a history file is extracted from the user's iTunes player. This history file contains the names and timestamps of all songs ever added to that user's music library, and thus it provides a record of the user's behavior prior to joining Service. The user's post-registration purchases are also observed by Service because the plugin notifies Service via the internet of all songs

¹ We simplify the exposition in this way, since all of the networks tested yield the same qualitative conclusions. This appendix of results appears at <http://assets.wharton.upenn.edu/~dfleder/ec-appendix.pdf>

added to the user's iTunes library, whether bought at the iTunes store or not. This combination of the history file and continued communication via the plugin thus gives us a before and after view of the user's behavior.

Besides comparing users' purchase histories before and after registering, we can also compare these users with a control group. The control data are obtained by again exploiting the history files of Service users. For users who register after our study, their history files allow us to look backward at their Service-uninfluenced behavior during the same time period. This use of eventual Service users for the control affords a measure of similarity between the groups. Thus the new data source enables a before-after recommendations contrast as well as data on a control group for the same period.

5.3 Data Inclusion Criteria

This section describes the process for setting up the data in the two-group design introduced earlier. Figure 1 summarizes the details of this process. (*For reference, all figures and tables appear at the end of this paper.*)

The data are collected via Service's plugin that is installed on each user's machine. The plugin relays to Service in near real-time the timestamp and product information of any song added to that user's iTunes library. For ease of writing, we refer to songs as purchases, but our data in fact capture all songs added to a user's library, whether purchased from Apple's iTunes store, purchased from another firm, or downloaded elsewhere online.

The data comprise users who registered for Service between January – July 2007. We define the treated group as those users who registered sometime during March 2007. March is chosen because it was Service's month with the highest number of registrations during this period and thus affords the largest sample size. The time periods for the before-after comparison are the two month windows January-February and March-April.² The control group is defined by users who registered for Service sometime from May on. We observe the control's Service-unaffected behavior over January-April because upon their eventual registration, sometime from May on, we extract their iTunes history files and look backward at the January-April period.

A criterion for inclusion in the study is that each user began using iTunes in August 2006 or earlier. Upon installing iTunes or buying an iPod, users often load their CD collections onto their computers. We do not want to treat loading of old music as new purchases. Thus the criterion of installing iTunes in August 2006 or earlier creates a buffer of at least four months (September-December 2006) between installing iTunes and our analysis. This is conservative because the loading of old CDs typically occurs within the first month of iTunes/iPod use.

A limitation of the data collection is that assignment to the treated versus control group is not randomized (i.e. the data are observational and not a controlled experiment). Since registration is the user's choice, the analysis cannot account for selection on unobservables. For example, it is possible that registration is a response to increased demand for music rather than a cause of it. However, detailed sensitivity analysis in the online appendix shows this is unlikely. (The appendix's URL was given in footnote 1.)

The resulting data set is summarized in Table 1. In terms of number of users, the treated group is larger than the control. This occurs because we chose March as the time of registration, which determined the number of treated users. The number of control users is determined by how many people register from May–July on (when data collection ends) such that we can look backward at their history files. The table next shows that purchases increase after recommendations. This increase was anticipated, although the size of roughly 50 percent is larger than expected. This increase in songs added is not seen in the control, where in fact the number of songs added decreases. Last, the table shows the number of artists for whom at least one song was purchased. This figure also increases for the treated group, indicating that users explore a wider range of artists under recommendations.

6. RESULTS ON THE OBSERVED DATA

This section shows how the consumer network changes when recommendations are introduced.

² That some users registered in late March could dampen the results' magnitude because it allows some Service-unaffected data to enter the post-recommendations period. One cannot circumvent this by centering each user's before-after data exactly on his registration date; since each user differs in this date, there would be no well-defined period for constructing the control. We are conservative in accepting this tradeoff: possibly dampening the results in order to have a well defined control.

6.1 Aggregate Analysis on the Observed Data

Using the two group design, we construct the four networks – before and after recommendations for the treated and control – and calculate the summary measures $T(\cdot)$ on each. Then, for each summary statistic T , we calculate the changes over time $D_1 = T_{12} - T_{11}$, $D_2 = T_{22} - T_{21}$, and the difference-in-differences estimator $D = D_2 - D_1$.

Table 2 shows the results. Across the columns are the three T metrics: density, median degree, and path length. Across the vertical dimension is the treated group (row “T”) and control group (row “C”). The table’s elements show the values of T before and after recommendations. The column D_i lists the difference for each group. Last, the column D/p lists the difference-in-differences estimate D with the p -value below it from a test that $D = 0$. (The “/” symbol indicates separate rows.) To test the hypothesis that $D = 0$, we use the non-parametric method of permutation tests [8].

The results show that on all three measures, users become more similar to one another after recommendations. First, the treated network becomes denser, showing that users have more connections among themselves. Before recommendations, 23% of the edges are filled in, and after 46% are present, yielding $D_1 = 23\%$. This is a large increase in density. Over the same period, the control has no noticeable change and $D_2 \approx 0$. The difference-in-differences estimate is $D = 23\% > 0$, indicating that the treated network does become more similar relative to the control. This difference is significant, as the hypothesis $D = 0$ is rejected ($p < 0.01$).³ On the other two metrics, we also observe greater similarity after recommendations. The median degree increases, $D > 0$, indicating that the typical user has more connections to others. Similarly, the path length decreases, $D < 0$, indicating that on average users are fewer hops away from one another. All of the results are significant ($p < 0.01$).

6.2 Individual-Level Analysis on the Observed Data

The above analysis showed that in aggregate users are more similar after recommendations. This section asks if there are differential effects at the individual level. For example, could close users become closer but far ones become farther – in such a way that the aggregate result masks this? If true, even though the network is more similar in aggregate, far users becoming farther would be evidence of fragmentation.

This question is related to asking whether users form tighter clusters after recommendations. Clustering is a useful way to phrase this question but not a profitable way to answer it. For every type of clustering method, there is no “true” number of clusters. Thus one cannot say whether users are more clustered after recommendations – one would have to analyze this for all possible numbers of clusters and it is difficult to draw firm conclusions.

Thus to assess such individual-level effects, we compare the distance of all user pairs before versus after recommendations and examine whether there are sub-populations that become farther. Table 3 presents these results. The table plots the path length between all nC_2 user pairs. The horizontal axis is the number of hops before recommendations, and the vertical axis is the number of hops after recommendations. The values in the table are the percent of user-pairs falling in each cell. User-pairs becoming farther lie above the diagonal. A distance of infinity means there is no path between the two users.⁴

The control group appears stable (right side), as it has roughly equal weight above and below the diagonal. In contrast, the treated group (left side) shows a different pattern. First, the aggregate effect toward similarity is evident: there are more user-pairs becoming closer (36.9% weight below the diagonal) than there are becoming farther (9.2% weight above the diagonal). This is consistent with the aggregate findings above. Second, the increase in similarity appears uniform: all types of users become closer to one another. Users who were close became closer, *and* users who were initially far became closer too. There does not appear to be evidence of a differential effect.

Note that some users do grow farther, but this is not a differential effect. We expect some chance fluctuation: users who were by chance closer revert to being farther, and users who were by chance farther revert to being closer. This is seen in the control group, where 11.3% went from 1 to 2 hops while 11.1% went from 2 to 1 hops. This level of mixing is roughly equal. In the treated group, some pairs do become farther (7.3% go from 1 to 2 hops) but many more become closer (27.7% went from 2 to 1 hops). This difference of 20.4% is large as well, since it is a fraction of $nC_2 \approx 300,000$ user pairs. To summarize, the trend toward greater similarity exists at all initial path lengths, and so we do not see evidence of a differential effect.

³ All the networks have one large, connected component containing nearly all users with few unconnected users outside it. Thus the density, degree, and path lengths are not biased due to changes in the size of the main component.

⁴ In Table 3, a very small number of pairs are four or five hops away. This number is so small ($\approx 0.04\%$) that for clarity we omit them from the presentation (but not the analysis) to avoid rows and columns of nearly all zeros..

7. VOLUME EQUALIZATION

The results thus far show that similarity increases after recommendations. We also know that purchase volume increased after recommendations (from Table 1). This fact raises the question of whether the volume alone is responsible for creating more network edges and hence more similarity (i.e. simply buying more products increases the chance of having something in common with another user). We thus want to decompose the recommender’s effects into *taste* and *volume* components. The taste component is the portion of D due to changes in the assortment of artists users buy, with volume held equal. Both are valid ways for recommenders to affect similarity, but we wish to distinguish them.

We decompose these effects by carefully equalizing the volume before versus after recommendations. To equalize the volume before versus after, we use the bootstrap [5]. Instead of comparing O_{11} and O_{12} , we compare O_{11} and O_{12}^* , where O_{12}^* is sampled randomly with replacement from O_{12} and has sample size $|O_{11}|$. In other words, we are sampling for the empirical distribution of O_{12} and limiting the sample size. This procedure assumes the observations are i.i.d. over time, which is a common assumption in many statistical models of purchase data (e.g., latent-class multinomial models). For consistency, we also equalize the volume in the control group before versus after. (This is for consistency but likely unnecessary because in the control $|O_{21}| \approx |O_{22}|$ anyway.) Last, for consistency, we equalize the volume across O_{11} and O_{21} : before recommendations, we will see, the control has slightly more purchases per user than the treated group. To prevent this difference from affecting the results, we reduce $|O_{21}|$ to $|O_{11}|$ in the same manner. Thus in the volume-equalized case, we have four data sets O_{11} , O_{12}^* , O_{21}^* , and O_{22}^* , all with the number of purchases equal to $|O_{11}|$. This sampling introduces a source of variation in the results, and thus all results are averaged over repeated trials (1000 simulations).

The aggregate analysis is repeated on the volume equalized data, and Table 4 shows the results. The same conclusion of greater similarity after recommendations emerges. However, the magnitudes are smaller, as expected, because of volume equalization. For example, the treated network’s density increases from 23% to 27%. This magnitude is smaller than on the observed (unequalized) data, where it increased from 23% to 46%. Though the magnitude is smaller, it is still a significant increase compared with the control group ($p = 0.03$). The other measures show the same conclusions: the median degree increases, showing users have more connections to one another, and the average path length decreases, showing that users are closer to one another and the network is “smaller.” In every case we reject $D = 0$ ($p \leq .05$), providing evidence of a standalone taste effect. We also repeated the individual-level analysis and find the results hold with volume-equalized data. (These latter results appear in the online appendix.)

8. SIMULTANEOUS TWO-GROUP ANALYSIS

The previous analyses mirrored an experimental design: recommendations were introduced, and changes were examined in the treated versus control groups. We found that treated users became more similar to one another, whereas the control showed almost no change.

We expand the analysis now in two ways, examining changes between the groups and changes in the population as a whole. Figure 2 shows this graphically. The previous analysis considered the *Treated* and *Control* regions of Figure 2. We enlarge the analysis to include the between group similarity (*Between*), which describes how close the entire treated group is to the entire control, and the overall similarity (*Overall*), which treats all users as a single population and describes the change in similarity within it.

Mathematically, consider a network built on the combined data $\{O_{1t}, O_{2t}\}$, which combines the treated and control groups. The set of all edges in the network, termed *Overall*, can be partitioned into three groups: *Treated*, *Control*, and *Between*. *Treated* is the set for which both nodes are in the treated group. *Control* is the set for which both nodes are in the control group. *Between* is the set for which one node is in the treated group and the other is in the control. This section extends the analysis to *Overall* and *Between*.

The motivation is two-fold. First, although recommender systems are becoming increasingly common, it is possible that not everyone will be exposed to them at all times. In this case, the state of the world under recommendations may reflect *Overall* more than *Treated*. This situation is unlike many experiments in statistics (e.g., clinical drug trials), in which if a new method is effective we envision treating everyone. Second, the *Between* analysis tests for another type of fragmentation in which treated users become self-similar but distant from control users. For example, suppose half the population uses recommenders; if *Treated* users became more similar to each other but *Treated* and *Control* moved apart, this would be another form of fragmentation. The *Between* analysis tests for this.

Using the same network definition and same statistics $T(\cdot)$, we repeat the analysis on *Overall* and *Between*. Table 5 presents the results. For ease of comparison, the *Treated* and *Control* results are reproduced from earlier. Examining *Between*, one

sees that the treated and control groups do become closer to each other. There are more edges between the groups after recommendations than before, and the path length between users in different groups decreases. Thus the treated group has not moved away from the control; rather, they are becoming closer. Examining *Overall* shows a similar result: the population as a whole is becoming more similar after recommendations. The density increases after recommendations and the path length decreases. This could be expected for *Overall*, since if *Treated and Between* exhibit more similarity and *Control* shows little change, then *Overall* will show a weighted average of this trend.

As before, we examine whether this result is due solely to volume or has a standalone taste component. Table 6 presents the results after equalizing the volume post-recommendations. As before, the magnitudes are dampened, but the results are the same: the treated and control groups move closer to one another and the overall population of users becomes more similar, as seen by the higher density, higher degree, and lower path length.

To summarize, in previous sections we found that treated users became more similar to one another, whereas the control showed almost no change. This section showed that the treated and control groups as a whole become closer too. This additional finding rules out another form of fragmentation in which the treated group, despite its becoming more self-similar, could have moved in entirety away from the control. Thus at several levels we observe a trend toward more similarity: within the treated group, between treated and control groups, and in the population as a whole.

9. CONCLUSIONS

This paper asked whether recommender systems fragment versus homogenize users. Using data from the music industry, we found that a network of users becomes more similar to one another after recommendations.

At the aggregate level, we found that users exposed to recommendations appear more similar afterward in their purchases. This finding occurred for two reasons. Users shifted their purchases toward more similar items, the taste effect; and, users simply bought more under recommendations, the volume effect, which increased the likelihood of co-purchases with others.

At the individual level, we looked for fragmentation in terms of a differential effect – namely, close users becoming closer and far ones becoming farther – but did not find evidence of this. This helped rule out the possibility that users were fragmenting into groups. If users were splitting into groups, a form of fragmentation, we should have seen far users become farther, which we did not.

These findings were observed for a variety of similarity measures and network definitions. For the setting of the music industry and our firm, it thus appears that recommender systems are associated with an increase in commonality rather than fragmentation.

The findings have policy and business implications. Each, in turn, introduces directions for future work. Regarding policy, we began with the question of whether recommender systems create fragmentation. The fragmentation outcome, should it exist, would be undesirable in our view and the view of many others'. Sunstein argued that the effects of recommender systems have connections to democracy itself: "it is highly desirable for a democracy to contain a kind of 'social architecture' that offers both shared experiences and unanticipated exposures," (p. 206) and there is concern that recommenders could weaken this if they show people only what they already like and know. On balance, the internet allows people to access more sources of information than ever. However, to the extent technology design choices undermine the above goal, we might ask how we can build better recommender systems that offset such fragmentation effects. We did not, however, find evidence of fragmentation, despite multiple ways of trying to identify it. In the absence of such effects, there is not cause, based on this study, to modify the architecture of e-commerce or the web. That said, we believe research on this question should continue in order to cover more product categories and recommendation technologies.

Regarding business, the study provides a window onto the ongoing trend of targeted marketing. Recommender systems lower search costs, so one interpretation of the post-recommendations data is that it better reveals preferences. Why then did we not observe consumers clustering into the hyper-specialized groups that some targeted marketers and advertisers might expect? It is possible consumers are not yearning for more of the same from their same segment but looking for variety and commonality. If so, consumers may prefer a range of experiences that narrow targeting does not deliver. A difficulty in making this inference is separating the effects of reduced search costs from the influence of the recommender itself. The recommender lowers search costs, but it also influences users in choosing what items to show them. This influence is unavoidable and will exist for any recommender system. An interesting empirical question is thus to separate these effects: when we observe greater commonality post recommendations, how much is due to consumers' preference for variety and commonality (revealed to us by reduced search costs) versus the bias of what the recommender selects for the user.

In his book *The Big Sort*, Bill Bishop [3] documents how over the last thirty years Americans sorted themselves into politically like-minded neighborhoods. This paper asks a similar question about the web. While many predict the web will further this trend of fragmentation, the evidence for the industry and firm studied here is to the contrary. As this is the first empirical study on the topic, we look forward to additional studies on the matter.

10. REFERENCES

- [1] Apple. 2008. iTunes Store Top Music Retailer in the US. Company website. <http://www.apple.com/pr/library/2008/04/03itunes.html>
- [2] Arnheim, A. 1996. Summary of the proceedings of the U.C. Berkeley Collaborative Filtering Workshop. March 16. www2.sims.berkeley.edu/resources/collab/collab-report.html.
- [3] Bishop, B. 2008. *The Big Sort*. New York: Houghton Mifflin.
- [4] Campbell, D. T. and J. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin Company.
- [5] Efron, B. and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1(1):54-75.
- [6] Fleder, D. and K. Hosanagar. 2007. Recommender systems and their impact on sales diversity. *Proc. of the 8th ACM conference on Electronic Commerce*: 192-199.
- [7] Fleder, D. and K. Hosanagar. 2009. Blockbuster culture's next rise or fall: the impact of recommender systems on sales diversity. *Management Science* 55(5):697-712.
- [8] Good, P. 1994. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. New York: Springer-Verlag.
- [9] Hervas-Drane, A. 2007. Word of mouth and recommender systems: a theory of the long tail. NET Institute Working Paper, No. 07-41. <http://ssrn.com/abstract=1025123>
- [10] Oestreicher-Singer, G. and A. Sundararajan. 2009. Recommendation networks and the long tail of electronic commerce. <http://ssrn.com/abstract=1324064>.
- [11] Sunstein, C. R. 2001. *Republic.com*. Princeton: Princeton University Press.
- [12] Thompson, C. 2008. If you liked this, you're sure to love that. *The New York Times Magazine*. November 23.
- [13] Van Alstyne, M. and E. Brynjolfsson. 2005. Global village or cyber-Balkans? Modeling and measuring the integration of electronic communities. *Management Science* 51(6):851-868.
- [14] Watts, D. J. and S. H. Strogatz. 1998. Collective dynamics of small-world networks. *Nature* 393:440-44.

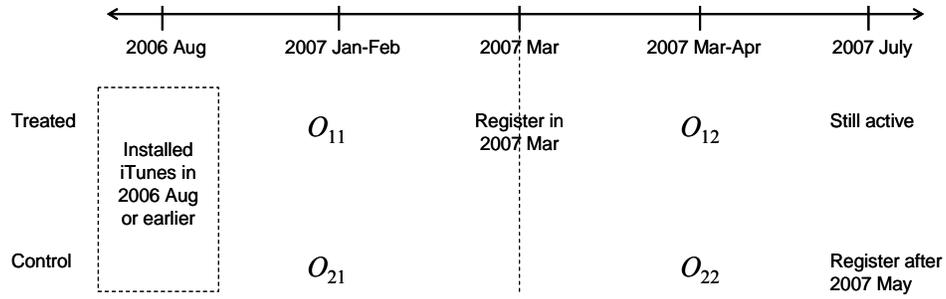


Figure 1. Data Setup and Analysis Design

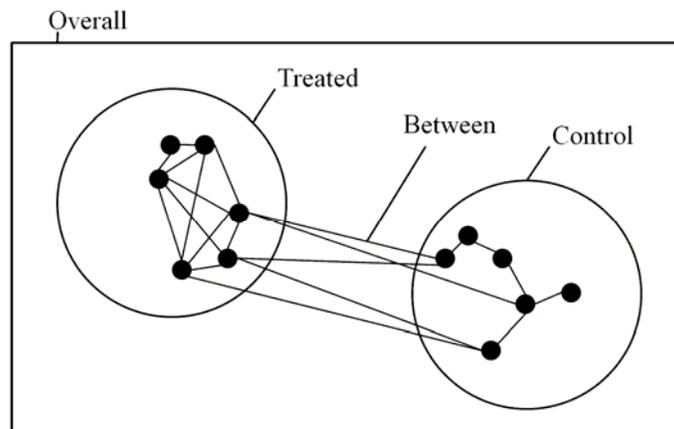


Figure 2. Edges in *Overall* can be partitioned into *Treated*, *Control*, and *Between*

Table 1. Summary statistics

	Treated		Control	
	Before	After	Before	After
Users	1,794	1,794	858	858
Songs purchased	215,749	326,640	106,431	97,553
Artists with at least one purchase	24,368	34,411	14,785	13,768

Table 2. Main results for the Unweighted Network – *Observed Data*

	Density				Median Degree				Path Length			
	Before	After	D_i	D/p	Before	After	D_i	D/p	Before	After	D_i	D/p
Treated	23%	46%	23%	22%	167	402	235	234	1.80	1.54	-0.26	-0.26
Control	19%	19%	0%	<0.01	134	135	1	<0.01	1.86	1.86	0.00	<0.01

Table 3. Path Lengths between all user pairs – *Observed data*.
Entries represent the percentage of all nC_2 user-pairs.

		Treated				Control			
# Hops After	∞	0.2	1.2	0.2	0.3	0.3	4.0	0.6	1.9
	3	0.0	0.3	0.1	0.1	0.2	3.8	1.2	0.8
	2	7.3	37.8	3.0	3.8	11.3	48.9	4.0	3.6
	1	15.5	27.7	1.0	1.3	7.4	11.0	0.4	0.3
			1	2	3	∞	1	2	3
		# Hops Before				# Hops Before			
Becoming closer (below diagonal)		36.9				20.1			
Becoming farther (above diagonal)		9.2				20.2			
No change (on diagonal)		53.7				59.4			

Table 4. Main results for the unweighted network – *Volume-equalized data*

	Density		Median Degree				Path Length					
	Before	After	D_i	D/p	Before	After	D_i	D/p	Before	After	D_i	D/p
Treated	23%	27%	0.04	4%	167	213	46	43.35	1.80	1.74	-0.07	-0.06
Control	12%	13%	0.00	0.03	79	82	3	<0.01	1.98	1.97	-0.01	0.05

Table 5. Simultaneous analysis for the unweighted network – *Observed data*

	Density		Median Degree				Path Length		
	Before	After	D_i	Before	After	D_i	Before	After	D_i
Treated	0.23	0.46	0.23	167	401	235	1.80	1.54	-0.26
Control	0.19	0.19	0.00	134	135	1	1.85	1.83	-0.02
Between	0.21	0.30	0.09	147	235	88	1.83	1.71	-0.12
Overall	0.21	0.31	0.1	295	503	207	1.83	1.69	-0.13

All values of D_1 and D_2 are significantly different from zero ($p < 0.05$)

Table 6. Simultaneous analysis for the unweighted network – *Volume-equalized data*

	Density		Median Degree				Path Length		
	Before	After	D_i	Before	After	D_i	Before	After	D_i
Treated	0.23	0.27	0.04	167	214	47	1.80	1.73	-0.07
Control	0.13	0.13	0.00	80	82	3	1.94	1.93	-0.01
Between	0.17	0.19	0.02	112	131	19	1.88	1.84	-0.04
Overall	0.17	0.19	0.02	230	275	45	1.87	1.83	-0.04

All values of D_1 and D_2 are significantly different from zero ($p < 0.05$)