

Blockbuster Culture's Next Rise or Fall:
The Impact of Recommender Systems on Sales Diversity

Daniel Fleder and Kartik Hosanagar
The Wharton School, University of Pennsylvania
{dfleder,kartikh}@wharton.upenn.edu

First draft: 2007 (Net Institute Working Paper 07-10)

Final version: to appear in *Management Science* Vol. 55, No. 5, May 2009, pp. 697-712

Abstract:

This paper examines the effect of recommender systems on the diversity of sales. Two anecdotal views exist about such effects. Some believe recommenders help consumers discover new products and thus increase sales diversity. Others believe recommenders only reinforce the popularity of already popular products. This paper seeks to reconcile these seemingly incompatible views. We explore the question in two ways. First, modeling recommender systems analytically allows us to explore their path dependent effects. Second, turning to simulation, we increase the realism of our results by combining choice models with actual implementations of recommender systems. We arrive at three main results. First, some well known recommenders can lead to a reduction in sales diversity. Because common recommenders (e.g., collaborative filters) recommend products based on sales and ratings, they cannot recommend products with limited historical data, even if they would be rated favorably. In turn, these recommenders can create a rich-get-richer effect for popular products and vice-versa for unpopular ones. This bias toward popularity can prevent what may otherwise be better consumer-product matches. That diversity can decrease is surprising to consumers who express that recommendations have helped them discover new products. In line with this, result two shows that it is possible for individual-level diversity to increase but aggregate diversity to decrease. Recommenders can push each person to new products, but they often push users toward the same products.. Third, we show how basic design choices affect the outcome, and thus managers can choose recommender designs that are more consistent with their sales goals and consumers' preferences.

The authors thank Yannis Bakos, Eric Bradlow, Terry Elrod, Pete Fader, Greg Linden, Robin Pemantle, David Schweidel, Michael Steele, Christophe Van den Bulte, and seminar participants at Carnegie Mellon University, New York University, Stanford University, the University of Connecticut, the University of Pennsylvania, University of Washington, Seattle and WISE for their comments. This work is a much extended version of a conference paper by the same authors (2007) from the ACM Conference on Electronic Commerce, and we thank their three anonymous reviewers. The authors also thank Barrie Nault, an associate editor, and two anonymous reviewers for their valuable suggestions. The paper benefited thoroughly from their comments, and any remaining errors are the authors'. Finally, we thank the NET Institute (www.NETinst.org) and Ackoff Fund of the Wharton Risk Management and Decision Processes Center for financial support.

1. INTRODUCTION

Media has historically been a “blockbuster” industry (Anderson 2006). Of the many products available, sales have concentrated among a small number of hits. In recent years, such concentration has begun to decrease. The last ten years have seen an extraordinary increase in the number of products available (Brynjolfsson et al. 2006; Clemons et al. 2006), and consumers have taken to these expanded offerings. Many believe this increased variety allows consumers to obtain more ideal products, and if it continues it could amount to a cultural shift from hit to niche products. One difficulty that arises, however, is how consumers find such niche products among seemingly endless alternatives.

Recommender systems are considered one solution to this problem. These systems use data on purchases, product ratings, and user profiles to predict which products are best suited to a particular user. These systems are commonplace at major online firms such as Amazon, Netflix, and Apple’s iTunes Store. In author Chris Anderson’s view, “The main effect of filters, [which include online recommender systems], is to help people move from the world they know (‘hits’) to the world they don’t (‘niches’)” (2006, p. 109).

While recommenders have been assumed to push consumers toward the niches, we present an argument why some popular systems might do the opposite.¹ Anecdotes from users and researchers suggest recommenders help consumers discover new products and thus increase diversity (Anderson 2006). Others believe several recommender designs might reinforce the position of already popular products and thus reduce diversity (Mooney & Roy 2000; Fleder & Hosanagar 2007). This paper attempts to reconcile these seemingly incompatible views. Holding supply-side offerings fixed, we ask whether recommenders make media consumption more diverse or more concentrated.

We explore this question in two ways. First, modeling recommender systems analytically allows us to explore their path dependent effects. Second, using simulation, we increase the realism of our results by combining choice models with actual implementations of recommender systems. Our main result is

¹ With so many different recommenders employed by firms, one cannot state a universal result for all. Instead this paper picks several recommenders we believe are commonly used in industry and focuses on them.

that some popular recommenders can lead to a reduction in diversity. Because common recommenders (e.g., collaborative filters) recommend products based on sales or ratings, they cannot recommend products with limited historical data, even if they would be viewed favorably. These recommenders create a rich-get-richer effect for popular products and vice-versa for unpopular ones. Several popular recommenders explicitly discount popular items, in an effort to promote exploration. Even so, we show this step may not be enough to increase diversity.

That diversity can decrease is surprising to consumers who express that recommendations have helped them discover new products. The model provides two insights here. First, we find it is possible for individual-level diversity to increase but aggregate diversity to decrease. Recommenders can push each person to new products, but they often push similar users toward the same products. Second, if recommenders are simply replacing best-seller lists, diversity can increase by cutting out what is an even more popularity-biased tool.

The results have implications for firms and consumers. For retailers, we show how design choices affect sales and diversity. For consumers and niche content producers, we show how a recommender's bias toward popular items can prevent what would otherwise be better consumer-product matches. We find that recommender designs that explicitly promote diversity may be more desirable.

The rest of the paper is organized as follows. Section 2 reviews prior work. Section 3 gives a formal problem statement. Section 4 presents the analytic model, which is stylized but still able to show how sales information can bias recommenders. To increase the realism of our setting, and in particular incorporate actual recommender designs, a complementary simulation is developed in Sections 5-7. The simulation combines consumer choice models with actual recommender algorithms. Section 8 discusses the implications for producer and consumer welfare. Section 9 concludes, reviewing the findings and offering directions for future work.

2. PRIOR WORK

Recommender systems help consumers learn of new products and select desirable products among myriad choices (Resnick & Varian 1997). A simplified taxonomy divides recommenders into content-based versus collaborative filter-based systems. Content-based systems use product information (e.g., author, genre) to recommend items similar to those a user rated highly. Collaborative filters, in contrast, recommend what similar customers bought or liked. Perhaps the best-known collaborative filter is Amazon.com's, with its tagline, "Customers who bought this also bought..."

The design of these systems is an active research area. Reviews are provided in Breese et al. (1998) and Adomavicius and Tuzhilin (2005). For business contexts, Ansari et al. (2000) describes how firms can integrate other data sources (e.g., expert opinions) into recommendations. Work by Bodapati (2008) places recommender systems into a profit-maximizing framework. For industry applications, implementations at firms such as Amazon.com and CDNOW are described by Schafer et al. (1999) and Linden et al. (2003). Although there is a large body of work on building these systems, we know less about how they affect consumer choice and behavior.

Studies have recently begun to examine individual-level, behavioral effects. In marketing, Senecal and Nantel (2004) show experimentally that recommendations do influence choice. They find that online recommendations can be more influential than human ones. Cooke et al. (2002) examine how purchase decisions under recommendations depend on the information provided, context, and familiarity.

While the above studies ask how recommenders affect individuals, our interest is the aggregate effect they have on markets and society. In particular, we are interested in how recommenders affect sales diversity. In related work, Brynjolfsson et al. (2007) find that a firm's online sales channel has slightly higher diversity than its offline channel. They suggest demand-side causes, such as active tools (search engines) and passive tools (recommender systems), but do not isolate the specific effect of recommenders. In contrast, Mooney and Roy (2000) suggest collaborative filters may perpetuate homogeneity in choice but do not study it formally.

Given our focus on aggregate effects, the streams of work on information cascades and Internet balkanization are also related. The information cascades literature has looked at aggregate effects of observational learning and resulting convergence in behavior, or “herding” (Bikhchandani et al. 1998). The Internet balkanization literature asks whether the Internet creates a global community freed of geographic constraints. Van Alstyne and Brynjolfsson (2005) find that while increased integration can result, the Internet can also lead to greater balkanization wherein groups with similar interests find each other. Although our problem is different, we see these papers as complementary in highlighting the social implications of technologies that share information among users.

This prior work reveals four themes. One, recommender systems research in the data mining literature has focused more on system design than understanding behavioral effects. Two, the marketing literature is just beginning to examine such behavioral effects. Three, of the existing behavioral work, the focus has been more on individual outcomes than aggregate effects. Four, regarding aggregate effects, there are opposing conjectures as to the effect of recommenders on sales diversity.

3. PROBLEM DEFINITION

3.1 Focus on Collaborative Filters

The current work focuses on collaborative filtering recommender systems which appear to be more common than content based ones. The diversity debate focuses specifically on collaborative filters because these systems use historical sales data to generate recommendations. Content based systems do not use historical data and so do not naturally raise the question of whether positive feedback cycles could emerge and lower diversity. For ease of exposition, throughout the paper recommender system is synonymous with collaborative filter.

3.2 Measure of Sales Diversity

Our context is a market with a single firm selling one class of good (e.g., music versus movies).

Before examining recommender systems’ effects, it is necessary to distinguish between sales and product diversity. *Product diversity*, or product variety, typically measures how many different products a firm

offers. It is a supply-side measure of breadth. In contrast, we use *sales diversity* to describe the concentration of market shares conditional on firms' assortment decisions. To measure sales diversity, we adopt the Gini coefficient. The Gini is a common measure of distributional inequality. It has been applied to many problems, the most common being perhaps wealth inequality (e.g., Sen 1976).

Let $L(u)$ be the Lorenz curve denoting the percentage of the firm's sales generated by the lowest $100u\%$ of goods sold during a fixed time period. Further, let $A = \int_0^1 (u - L(u))du$ and $B = 0.5 - A$. The Gini coefficient is defined $G := A / (A + B)$. Figure 1 illustrates this. Thus $G \in [0,1]$, and it measures how much the $L(u)$ deviates from the 45° line. A value $G = 0$ reflects diversity (all products have equal sales), while values near 1 represent concentration (a small number of products account for most of the sales).

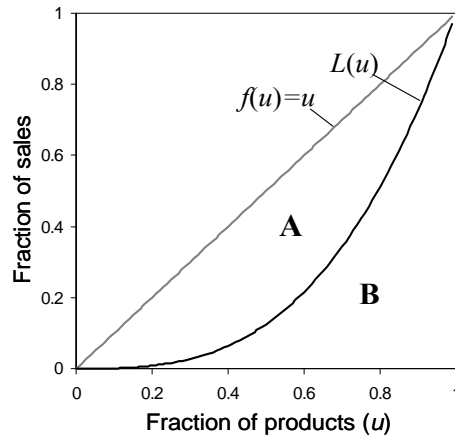


Figure 1. Lorenz curve

3.3 Problem Statement

Consider a firm with I customers c_1, \dots, c_I and J products p_1, \dots, p_J . Define a recommender system as a function r that maps a customer c_i and database onto a recommended product p_j . Typically the database records consumer purchases and/or ratings. Consider next a set of different recommenders r_1, \dots, r_k . Each r_i reflects certain design choices. For example, r_i might be the “standard” user-to-user collaborative filter, while r_j might be a variant that explicitly gives low weight to popular items. Denote by G_0 the Gini coefficient of the firm's sales during a fixed time period in which a recommender system was *not* used. In contrast, let G_i be the Gini coefficient of the firm's sales in which r_i was employed with all else equal.

Definition. *Recommender bias.* Recommender r_i is said to have a concentration bias, diversity bias, or no bias depending on the following conditions:

$$\begin{cases} \text{Concentration bias} & G_i > G_0 \\ \text{Diversity bias} & G_i < G_0 \\ \text{No bias} & G_i = G_0 \end{cases}$$

For various recommenders, we examine whether a bias exists and its direction.

4. ANALYTICAL MODEL

4.1 Assumptions and Model

Collaborative filters can operate on purchase or ratings data. To fix a context, our model considers purchases. We consider a set of customers making purchases sequentially.

Assumption 1. *Each consumer buys one product per time step.*

The customer's decision is which product to buy and not whether to buy. For example, at a subscription media service, this could reflect customers who decide to consume an item (e.g., a movie or song) but have not yet chosen which.

Assumption 2. *We assume there are only two products, w and b (white and black).*

This assumption is for tractability, but it still allows us to illustrate how the use of sales information affects diversity.

Assumption 3. *Consumers have purchase probabilities $(p, 1-p)$ for (w, b) in the absence of recommendations.*

We do not model the decision process that generates these purchase probabilities.

Assumption 4. *At each occasion, the firm recommends a product, which is accepted with probability r .*

Assumption 5. *The firm's recommendation is generated using a function $g(X_t) \in \{w, b\}$, where X_t is the segment share of w just before purchase t .*

The recommender's inputs are segment shares (market shares within a segment of similar users), and its output is a product. The system modeled recommends the product with higher segment share. This choice of g has a parallel with collaborative filters, which identify similar customer segments and

recommend the most popular item within them (e.g., “people who bought X also bought Y ”). This recommender can be represented by the step function

$$g(X_t) := P(w \text{ recommended} \mid X_t) = \begin{cases} 0 & , X_t < \frac{1}{2} \\ \frac{1}{2} & , X_t = \frac{1}{2} \\ 1 & , X_t > \frac{1}{2} \end{cases} \quad (1)$$

where $X_t \in [0,1]$. Figure 2 plots this. When $X_t = \frac{1}{2}$ and the products have equal shares, the recommendation is determined by a Bernoulli($\frac{1}{2}$) trial. To start, the recommender does not favor either product, and we assume $X_1 = \frac{1}{2}$.

Assumption 6. *The segment of consumers constituting X_t is pre-selected and does not change over time.*

This segment of similar consumers is identified based on past behavior, possibly from purchases of products in other categories. The assumption that the group does not evolve is for tractability, but it does have a parallel with business practice. In industry, real-time updating of segments is often computationally prohibitive, so many firms update segments periodically.²

The process defined by these assumptions can be illustrated by an urn model. Consider the two urn system of Figure 4. Urn 1 contains balls representing products w and b . A fraction p of the balls in urn 1 are white; this fraction is the consumer’s purchase probability for w in the absence of recommendations. Urn 2 is the recommender: its contents reflect the sales history within the segment, and it produces recommendations according to $g(X_t)$, where X_t is the fraction of w in urn 2 just before t . To start, urn 2 contains one w and one b . At time $t=1$, a ball is drawn *with* replacement from urn 1 representing the consumer’s choice before seeing the recommendation. Next, a ball is drawn *with* replacement from urn 2 according to $g(X_t)$, representing the recommended product. With probability r , the consumer accepts the recommendation, and with probability $(1-r)$ the consumer retains the original choice. The ball chosen represents the actual product purchased; a copy of it is added to urn 2, which is equivalent to updating the

² Section 5 presents an alternate approach where we relax these assumptions. Specifically, we model the consumer’s decision process, consider multiple products with a no-purchase option, and allow segments to evolve over time.

recommender's sales history (e.g., the firm's database). Consumer 2 then arrives, and the process repeats (p and r are the same, but X_2 is used instead of X_1), and so on for other customers.

From these assumptions, the probability that w is purchased at time t is

$$\begin{aligned}
 f(X_t) &:= P(w \text{ chosen on occasion } t \mid X_t) \\
 &= p(1-r) + g(X_t)r \\
 &= \begin{cases} p(1-r) & , X_t < \frac{1}{2} \\ \frac{[p(1-r)] + [p(1-r) + r]}{2} & , X_t = \frac{1}{2} \\ p(1-r) + r & , X_t > \frac{1}{2} \end{cases} \quad \begin{matrix} "l" \\ "m" \\ "h" \end{matrix}
 \end{aligned} \tag{2}$$

Figure 3 plots an example of f . The labels in (2) “ l ”, “ m ”, “ h ” are short-hand; they visually refer to the low (l), middle (m), and high (h) portion of f 's shape in Figure 3. The geometry of this figure helps illustrate the results derived next.

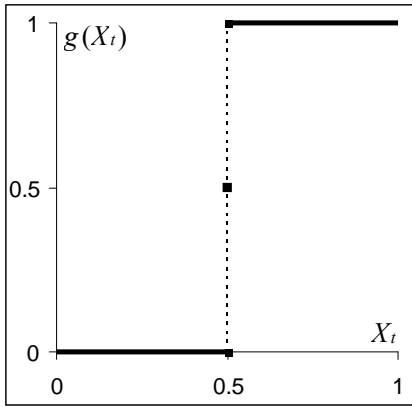


Figure 2. Recommender $g(X_t)$

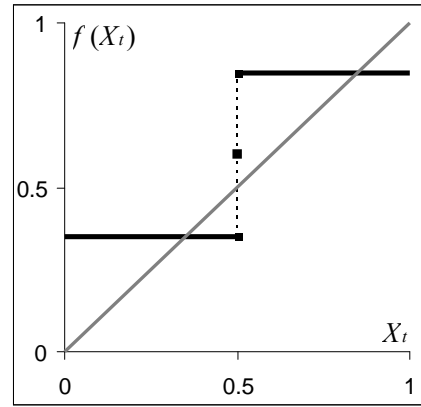


Figure 3. $f(X_t)$ and 45° line ($p=0.7, r=0.5$)

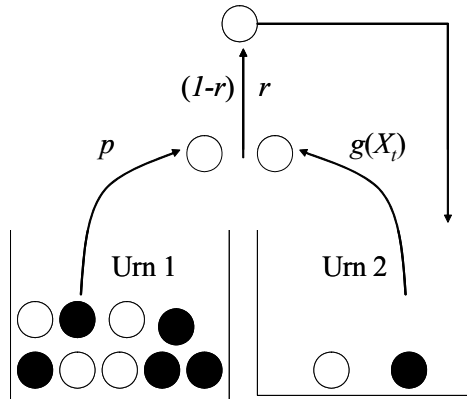


Figure 4. A two-urn model for recommender systems

4.2 Model Results

4.2.1 Theoretical results

The following results are derived in a random walks framework by examining the difference $w - b$ over time. All proofs are in the online appendix.

Without recommendations, shares converge to $(p, 1-p)$. The first proposition asks how this is affected by the presence of a recommender. As $t \rightarrow \infty$, $\{X_t\}$ converges to one of two values. These limiting values depend on the consumer's initial p and recommender's influence r and are given by

Proposition 1. *Support points. As $t \rightarrow \infty$, X_t converges to*

Case	Support point 1	Support point 2
1. $p \leq (\frac{1}{2} - r)/(1 - r)$	l	n/a
2. $(\frac{1}{2} - r)/(1 - r) < p < \frac{1}{2}/(1 - r)$	l	h
3. $p \geq \frac{1}{2}/(1 - r)$	h	n/a

where the shorthand l and h are from equation (4), $p \in [0, 1]$, and $r \in (0, 1)$ ($r = 0$ or 1 is trivial).

The cases in Proposition 1 have an attractive geometric interpretation: The support points are simply the intersections of $f(X_t)$ with the 45° line in Figure 3. That is, the support points are $\{x : f(x) = x\}$.³ Visually, p and r shift and stretch the step function; as a result, it has either one intersection occurring below $f(X_t) = 0.5$ (Case 1), one intersection occurring above $f(X_t) = 0.5$ (Case 3), or both (Case 2).

Corollary 1. *Chance and winning the market. In Case 2, $P(\lim_{t \rightarrow \infty} X_t < 1/2) > 0$ and $P(\lim_{t \rightarrow \infty} X_t > 1/2) > 0$.*

This is evident because $l < 0.5$ and $h > 0.5$ are both support points. This shows an interesting aspect of Case 2: regardless of the initial p , either product can obtain and maintain the majority share.

With the limiting value(s) of $\{X_t\}$ known, we ask whether they reflect higher or lower concentration. Let the term “increased concentration” define shares that are less equal than they would be without recommendations. Increased concentration means $\lim_{t \rightarrow \infty} X_t > p$ when $p > 1/2$ and $\lim_{t \rightarrow \infty} X_t < p$ when $p < 1/2$. The effect on concentration is given by the following proposition.

Proposition 2. *Relation of limit points to concentration. For any (p,r) , the effect on concentration is*

Case	Support points	Effect on concentration relative to p
1	L	Increased concentration
2	2	Case 2A: $p \in (\frac{1-r}{2-r}, \frac{1}{2-r})$. Increased concentration for both support points Case 2B: $p \notin (\frac{1-r}{2-r}, \frac{1}{2-r})$. Increased concentration for one support point; decreased for the other
3	1	Increased concentration

These cases are shown in Figure 5. For Cases 1 and 3, there is a single outcome and that outcome always has increased concentration. These are areas of the $p \times r$ space where consumers have strong initial probability (p) relative to the recommender's strength (r); as a result, the recommender's effect is to reinforce this tendency even more. For example, if consumers have a fairly strong tendency to buy w with $p = .90$ and the recommender is fairly influential with $r = .25$, the recommender creates a positive feedback loop, reinforcing the popularity of w and giving it a limit share of $0.93 > 0.90$. Product w was initially bought more, which made it recommended more, which made it bought more, and so on.

Case 2A occurs where the recommender's influence (r) is high relative to the initial probability (p). This has two implications, one at the sample-path and one at the aggregate level. At the sample path level, *either* product can win the market, regardless of p . For example, $p = 0.55$ and $r = 0.75$ imply limiting market shares of $(w,b) \in \{(0.89,0.11), (0.14,0.86)\}$. In the first outcome, w wins the market. In the second, b wins, even though $p = 0.55$ initially favored w (c.f. Corollary 1). This occurs because r is large relative to p , and the recommender reinforces whichever product does well early on without too much resistance from p . This leads to the finding that recommenders can create hits. Some product will become a winner with a permanent, majority share, but we cannot say which beforehand. At the aggregate level, concentration always increases. We do not know which of w or b will win, but we know that one will and whichever does will be an outcome with greater concentration. Although they start with different models, a similar phenomenon occurs in other contexts (e.g., studies of firm location). Arthur (1994) provides an overview of applications, while earlier mathematical results are in Hill (1980).

³ The visual interpretation applies only to where f 's line segments intersect the 45° line (not the single point at $X_i = 0.5$).

Last, in Case 2B, neither the initial probability (p) nor the recommender's influence (r) is strong relative to one another. As a result, two outcomes are possible. The tendency p can be reinforced by the recommender. This increases concentration. Or, the recommender can give whichever product was not favored a small majority. This decreases concentration. For example, if $p = .60$, which is mild, and $r = .25$, the limit points are $.70$ and $.45$. Often w has more early successes and the recommender reinforces this, leading to less diverse $.70$ outcome. In some cases, if b is chosen enough early on, the recommender reinforces b leading to the $.45$ outcome, which entails less concentration than the initial share of $.40$.

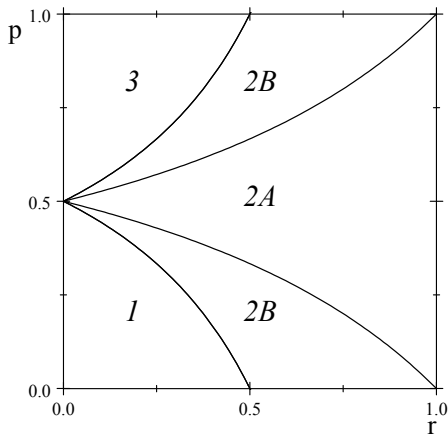


Figure 5. Relating the $p \times r$ space to concentration effects (numbers refer to cases in Proposition 2).

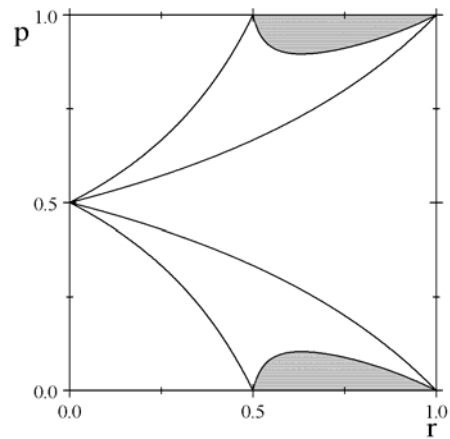


Figure 6. Concentration increases in white areas and decreases in shaded ones.

While both outcomes are possible in 2B, they are not equally likely. Next we determine the probability of arriving at each. This, in turn, allows us to calculate the expected effect on concentration.

Proposition 3. *The distribution of $\lim_{t \rightarrow \infty} X_t$ is*

Case	Support point 1	Support point 2	$P(\lim_{t \rightarrow \infty} X_t = \text{support point 1})$	$P(\lim_{t \rightarrow \infty} X_t = \text{support point 2})$
1. $p \leq (\frac{1}{2} - r)/(1 - r)$	l		1	0
2. $(\frac{1}{2} - r)/(1 - r) < p < \frac{1}{2}/(1 - r)$	l	h	γ	$1 - \gamma$
3. $p \geq \frac{1}{2}/(1 - r)$	h		1	0

where $\gamma = \frac{(1-m) \cdot (1 - \frac{l}{1-l})}{m \cdot (1 - \frac{1-h}{h}) + (1-m) \cdot (1 - \frac{l}{1-l})} \in (0,1)$. This proposition will be applied subsequently.

4.2.2 Graphical example

A graphical example helps illustrate the results. For sake of illustration, take $p = .70$ and $r = .50$. Figure 7 plots 10 realizations of this process over time. The left part of the figure shows these paths converging to two outcomes. One sees that the limits are in accord with Proposition 1, which says the process converges to a random variable whose support is $\{0.35, 0.85\}$. At right, the figure shows the frequencies of arriving at the lower versus upper outcome approach .27 and .73, in accord with Proposition 3.

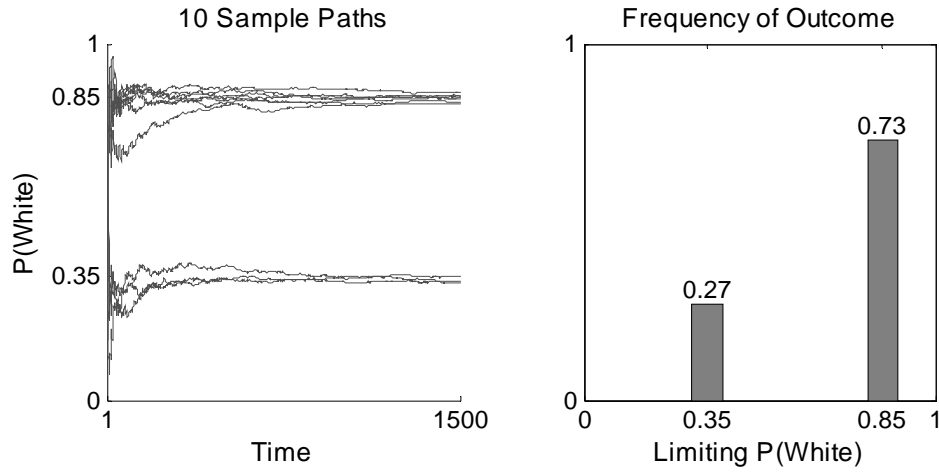


Figure 7. The two limiting outcomes for our example $f(x)$

4.2.3 Net effect on sales concentration

With the limiting distribution of $\{X_t\}$ known, we complete the connection to sales concentration. For two products with shares p and $1-p$, the Gini coefficient is proportional to (Sen 1976):

$$G(p) = |p - 1/2|. \quad (3)$$

With recommendations, we define

$$\begin{aligned} G_{p,r} &= E[G(\lim_{t \rightarrow \infty} X_t) \mid p, r] \\ &= G(l)P(\lim_{t \rightarrow \infty} X_t = l) + G(h)P(\lim_{t \rightarrow \infty} X_t = h). \end{aligned} \quad (4)$$

The net effect on concentration is given by $G_{p,r} - G_{p,0}$, which is >0 (<0) when concentration increases (decreases). Substituting into (3) and (4) terms from the previous propositions gives

Case	$G_{p,r}$	$G_{p,0}$
1 $p \leq (\frac{1}{2} - r)/(1 - r)$	$ l - 1/2 $	$ p - 1/2 $
2 $(\frac{1}{2} - r)/(1 - r) < p < \frac{1}{2}/(1 - r)$	$ l - 1/2 \gamma + h - 1/2 (1 - \gamma)$	$ p - 1/2 $
3 $p \geq \frac{1}{2}/(1 - r)$	$ h - 1/2 $	$ p - 1/2 $

The above gives a closed-form expression for the change in Gini coefficient. Figure 6 shows this graphically. For most of the $p \times r$ square, concentration increases. This is true, of course, for areas under Case 1, 2A, and 3, where the only possibility was increased concentration. It is also true for most areas where both outcomes were possible (Case 2B). In extreme cases, it is possible for a net decrease to occur, as shown by the shading. These areas are largely an artifact of the initial conditions assumed for urn 2, which place one w and one b in a high r recommender even when $p \approx 0$ or ≈ 1 .⁴

Summarizing, under recommendations the shares converge to either one or two limiting outcomes depending on (p, r) . When there is one outcome, it always reflects increased concentration: the recommender reinforces the popularity of the initially preferred product. In the two outcome cases, either both outcomes have greater concentration or one has greater concentration and the other has less. For the latter, a net effect must be calculated. This typically has greater concentration, although for extreme (p, r) , as discussed, increased diversity may occur. Thus the recommender seems to increase concentration among a set of similar users.

5. SIMULATIONS

5.1 Rationale for Simulation

Simulation offers three benefits for this problem. First, while actual recommender algorithms are difficult to represent analytically, they can be implemented in simulation. Second, heterogeneity in user preferences is easily accommodated. Third, more complex choice processes can be represented.

5.2 Choice Model and Simulation Design

We now turn to a simulation that combines a choice model with actual recommender systems. Repeat purchases are permitted in the simulation. Examples of contexts with repeat purchases could include music and video streaming from a subscription service (e.g., Rhapsody).

⁴ An example illustrates how this is related to initial conditions. Suppose $p = 0.99$, and $r = 0.99$, which is in the shaded region. Since $X_1 = .50$, $P(b \text{ on first purchase}) \approx 0.50$. If b is chosen, the recommender next suggests b ; since $r = 0.99$ the next consumer is almost certain to pick b too, and so on for the remaining consumers even though $p = 0.99$ favors w . If the initial conditions are determined by k Bernoulli(p) trials, diversity decreases even more: the shaded areas of Figure 6 begin to turn white even for small k . (These additional experiments are available from the authors on request.)

An overview of the process is as follows. There are I consumers and J products positioned in an attribute space. Consumers are not aware of all products. Each consumer knows most of the center products and a small number of products in his own neighborhood. Every period, a consumer either purchases one of the products or makes no purchase at all. To model this choice, a multinomial logit is used for J products plus an outside good. Just before choosing a product, a recommendation is generated. The recommender has two effects. First, the consumer becomes aware of the recommended product if he was not already. This increase in awareness is permanent. Second, the salience of the recommended product is increased temporarily, raising the chance that the recommended product is purchased in that purchase instance. The next consumer makes a purchase in a similar manner, and the process repeats after all consumers have purchased. After a predetermined number of iterations, the Gini is computed. The Gini is then compared to a benchmark G_0 , the Gini from an equivalent period in which recommendations were not offered.

We now discuss each of the simulation components: (i) the map of products and consumers, (ii) the recommender r , (iii) the awareness distribution, (iv) the choice model, and (v) the salience factor δ .

(i) *Map of product and consumer points.* The map of products and ideal points is the input for the choice model. Plotting consumer points and product locations goes back at least to Hotelling (1929) and is commonly used in marketing (e.g., Elrod & Keane 1995). Our consumers and products are points in a two-dimensional space. The use of two dimensions is for simplicity and visualization; for contexts with more than two attributes, the maps can be considered dimensionality-reduced versions, as is common in marketing research. We take both ideal points and products to be standard bivariate normal. The normality assumption for consumers is common in factor-analytic market maps (e.g., Elrod & Keane 1995). Our base case uses 50 consumers and 50 products, an example of which is in Figure 8.⁵

⁵ We have tested sensitivity to different numbers of consumers and products, higher dimensions, and other distributions (e.g., uniform, normal, and Pareto for each combination of consumers and products). The specific Gini values vary, but the conclusions are qualitatively similar. The main sensitivity results are in the appendix.

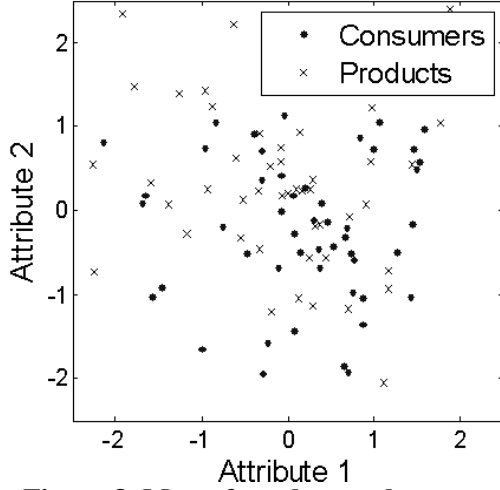


Figure 8. Map of product and consumer points

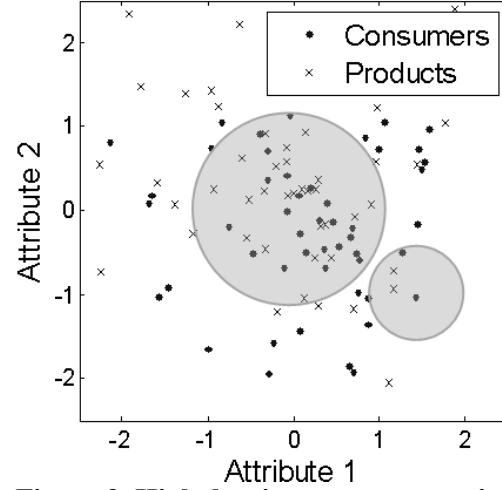


Figure 9. High density awareness regions shaded for one customer

(ii) *The recommender system.* An advantage of simulation is the ability to test real recommender systems. Our base case examines sales diversity under two systems, termed here r_1 and r_2 . In the taxonomy of Adomavicius and Tuzhilin (2005), both are memory-based, collaborative filters. Recommender r_1 is the most basic collaborative filter: for a given user, it first finds the set N^* of the n most similar customers by using cosine similarity to compare vectors of purchase counts. It then recommends the most popular item among this group.⁶ Formally, let $sales$ be an I users \times J items matrix of purchase counts, with $sales_{ij}$ the (i, j) element and $sales_i$ the row vector of c_i 's purchase counts. For a given user c_i , let

$$N^* := \operatorname{argmax}_N \sum_{c_j \in N} \cos(sales_i, sales_j) \text{ s.t. } |N| = n, i \neq j. \quad (5)$$

The system then recommends product

$$r_1: j^* = \operatorname{argmax}_j \sum_{c_i \in N^*} sales_{ij}. \quad (6)$$

Recommender r_2 has one difference. When selecting the most popular product among similar users, candidate items are first discounted by their overall popularity in the entire population:

$$r_2: j^* = \operatorname{argmax}_j \left[\sum_{i=1}^I sales_{ij} \right]^{-1} \sum_{c_i \in N^*} sales_{ij}. \quad (7)$$

The motivation for r_2 's popularity discounting is a belief that popular items are so obvious they should not be suggested. This was described to us in industry interviews as common practice. For example, if a

consumer is expected to buy or be aware of a product with high probability, the firm should recommend something else. Note, r_2 is not the same as applying “term-frequency inverse-document frequency” weights (*tf-idf*) to algorithm r_1 . *tf-idf* would insert discounting in the user similarity calculation (Breese et al. 1998), whereas r_2 inserts it in the final argmax of (7). In Section 7, we test other recommenders, including one with *tf-idf* weights, and show the results are directionally the same.

(iii) *Awareness*. Recommenders are valuable to consumers because they help overcome information asymmetry: the seller and other users may know of a product, but the given consumer may not. Recommenders share this information across the population. We assume each consumer is aware of a subset of the J products, and only items in this awareness set can be purchased. Once an item is recommended to a consumer, he is always aware of it in future periods. At the start, consumers are aware of many of the central products on the map plus a few items in their own neighborhood. These initial awareness states for each consumer-product pair are sampled according to

$$P(c_i \text{ aware of } p_j) = \lambda e^{-\text{distance}_{0j}^2 / \theta} + (1 - \lambda) e^{-\text{distance}_{ij}^2 / \kappa \theta}, \quad (8)$$

where distance_{0j} and distance_{ij} are respectively the Euclidean distances from the origin to product p_j and from consumer c_i to product p_j . The higher is λ , the more users are aware of central, mainstream products (left term), and the higher is $1 - \lambda$ the more users are aware of products in their neighborhood. θ and $\kappa\theta$ determine how fast awareness decays with distance. Note that users are not aware of the same products: they are likely to overlap in their awareness of the central products but less so in the local products.

The awareness model for one consumer is shown in Figure 9 for $\lambda=.75$, $\theta=.35$, and $\kappa=1/3$. We use these values for our base case. Setting $\lambda=.75$ creates a market with consumers more aware of mainstream goods than niche ones. This assumption is consistent with a market in which mass advertising makes consumers aware of the center, mainstream products. Under the opposite ($\lambda < .5$), the base-case is already a market of niches and it only strengthens later results that diversity can decrease. θ determines how many central products users know. Setting $\theta=.35$ creates an easy to understand “radius 1” rule: $e^{-1/.35} = .057 \approx 0$.

⁶ An alternative is to use correlation (i.e. cosine on mean-centered data). This does not qualitatively affect the results.

In other words, outside a radius of 1, the consumer is unlikely to be aware of the product. In our maps, about 40% of the products are within 1 unit from the origin; it is on this 40% of products that consumers are likely to overlap most in their awareness. The value κ determines awareness in the consumer's own neighborhood. The value $\kappa=1/3$ creates roughly a 0.5 radius rule. Outside the 0.5 radius, the consumer is unlikely to know about products, unless they are the central ones. The approach in selecting these parameters was to create an interpretable base case. In sensitivity analysis, we find the Gini can change for other parameter values but the results are directionally the same.⁷

(iv) *Choice model.* At each step of the simulation, a consumer either purchases an item in his awareness set or makes no purchase at all. We model this using the multinomial logit. The logit is well established in economics and marketing and has an axiomatic origin in random utility theory (for a Marketing application, see Guadagni & Little 1983). Consumer c_i 's utility for product p_j at time t is defined as $u_{ijt} := v_{ijt} + \varepsilon_{ijt}$, where v_{ijt} is a deterministic component and ε_{ijt} is an i.i.d. random variable with extreme value distribution. Under these assumptions

$$P(c_i \text{ buys } p_j \text{ at } t \mid c_i \text{ aware of } p_j \text{ at } t) = \frac{e^{v_{ijt}}}{\sum_{k=1}^J e^{v_{ikt}}} . \quad (9)$$

The unconditional probability is defined $P(c_i \text{ buys } p_j \text{ at } t) = P(c_i \text{ buys } p_j \text{ at } t \mid c_i \text{ aware of } p_j \text{ at } t) P(c_i \text{ aware of } p_j \text{ at } t)$. If a consumer is unaware of a product, the rightmost term is zero, and he cannot buy it.

The deterministic component v_{ijt} is often modeled as a linear combination of a brand intercept, product attributes, and covariates (e.g., price, promotion). In our context, since all relevant variables up to white noise are encompassed in the map, we define the logit's deterministic portion as

$$v_{ijt} := \text{similarity}_{ij} = -k \log \text{distance}_{ij} , \quad (10)$$

⁷ If consumers know only the central products ($\lambda=1$) the results are directionally the same. If consumers are aware of all products ($\theta \rightarrow \infty$), the results are the same direction as well. The same holds if awareness is Pareto distributed instead of normal.

where $distance_{ij}$ is the Euclidean distance between consumer c_i and product p_j . Our choice of a log transformation from distance to similarity is consistent with prior research (e.g., Schweidel et al. 2007).⁸

The parameter k determines the consumer's sensitivity to distance on the map. The higher is k , the more the consumer prefers the closest products. For our base case, as k ranges from 1 to 40, the Gini increases from .68 to .75. This range is consistent with several prior estimates of market concentration in media and e-commerce settings. An estimate for a major online clothing retailer is 0.70 (Brynjolfsson et al. 2007), an estimate for the music sales of debut albums is 0.724 (Hendricks & Sorensen 2007)⁹, and an estimate for the online book market is also near 0.75 (Chevalier & Goolsbee 2003)¹⁰. To fix a base case, we use $k = 10$ because the 0.72 Gini it produces matches the average of the estimates above. This k forms our base case. For other values, the results change in magnitude but not direction.

Last, as noted, consumers may choose not to purchase. This is modeled by an outside good with equal distance to all users. This approach is one common specification for modeling a no-purchase option (e.g., Chintagunta 2002). Our base cases uses a distance of .75 for this option, which implies the outside good's proximity is about 90th percentile (.87) for each consumer. That is, for each person, the outside good is closer than roughly 90% of the other goods. This means consumers have a fairly good outside option. If the outside good is farther, consumers substitute farther products for the outside good and diversity increases. The change in Gini under recommendations, however, is in the same direction.

(v) *Salience* δ . The term δ is the amount by which a recommended product's salience is temporarily increased in the consumer's choice set. The impact of the salience boost is that the purchase probability for the recommended item j is the same as that for an item j' with $v_{ij'} = v_{ij} + \delta$. The functional form is analogous to the modeling of store displays in marketing (e.g., Guadagni & Little 1983), which might be

⁸ Other transformations have been used, and the literature does not have a single standard: for example, $-k \cdot distance_{ij}$ in Elrod (1988); $(distance_{ij})^k$ in DeSarbo and Wu (2001); and $-k \cdot \log(distance_{ij})$ in Schweidel et al. (2007) with k a scaling parameter. While our base case uses the log transformation (e.g., Schweidel et al. (2007) and other references contained therein), we have tested sensitivity to the other specifications, and the results are not substantively different.

⁹ The .724 could underestimate concentration because the authors' data excludes less successful artists. This may not affect their objective, which differs from that in this paper.

considered an offline example of recommendations. The resulting choice probability is $P(c_i \text{ buys } p_j \text{ at } t \mid c_i \text{ aware of } p_j \text{ at } t) = e^\delta e^{v_{ijt}} \left(\sum_{k \neq j} e^{v_{ikt}} + e^\delta e^{v_{ijt}} \right)^{-1}$.

When $\delta = 0$, the recommender has only an awareness effect. Recommended items enter the awareness set if not there already. When $\delta > 0$, the recommender also has a salience effect, which increases the probability of buying the item (conditional on awareness). The salience effect exists for several reasons. First, consumers aware of many goods may have difficulty comparing all of them; recommended items become more salient in this comparison. Second, the salience boost may reflect the ease of clicking a recommended item versus continuing to search through a firm's website. Last, salience may capture persuasive effects. Recommendations often show an item's packaging and artwork, akin to a persuasive advertisement. We assume the combined effect is to increase the salience by δ . Experiments have begun to demonstrate that recommendations can have influential effects beyond awareness (Senecal & Nantel 2004). This simultaneity of both effects, awareness and salience, has parallels with advertising's informative and persuasive effects (e.g., Narayanan et al. 2005).

The salience term δ is a key parameter because it controls the strength of the recommender. For this reason, the paper's main results are shown for a range of δ and not a single point. To give some intuition for δ , consider the purchase probability of the 75th percentile closest item on the map (with 50 products, this is the 13th closest item). In our normal maps, if $\delta = 0$ the user chooses item 13 with $<10^{-4}$ probability. Item 1 is purchased with probability 0.85. If the 75th percentile item is recommended, for $\delta = (1, 5, 10, 15)$ the item takes on purchase probability ($<10^{-3}$, $<.01$, $.15$, and $.48$) respectively. Thus $\delta = 0-1$ is low, for it has little effect on purchase probability. A value $\delta=15$ is high, for it makes a close item (100th percentile) and far item (75th percentile) equal in probability.

¹⁰ The Zipf formulation can be equated to a power law, and from the power law a closed form expression for the Gini can be derived. A rank-on-sales coefficient of 1.17 in a power law implies a Gini of $(2 \times 1.17 - 1)^{-1} = 0.75$.

6. RESULTS

We now present simulation results for the two real-world recommenders.¹¹ We use 50 consumer points and 50 products sampled from a bivariate normal distribution $N_2(\mathbf{0}, \mathbf{I})$ with $k = 10$.

6.1 Example of a Single Sample Path

Before presenting overall results, we illustrate the process with one sample run. At first, recommendations are disabled and customers make purchases for 200 periods. Then r_1 is enabled and customers make purchases for an additional 200 periods. For sake of illustration, $\delta = 5$, but more general results follow.

The Lorenz curves and Ginis from both periods are shown in Figure 10. The example shows $G_1 - G_0 = 0.82 - 0.72 = 0.10 > 0$, and hence r_1 increases concentration here. This is for one sample path, and a more systematic comparison is given below.

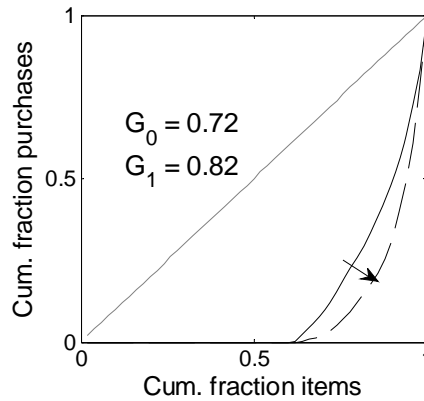


Figure 10. One sample path before and after recommendations (r_1 , $\delta=5$)

6.2 Simulation Results

With the same parameters as above, we average results across 1000 experiments/maps each for r_1 and r_2 . After, we generalize the findings beyond the base case of $\delta = 5$.

As Table 1 shows, both recommenders have a concentration bias on average, as reflected by $\bar{G}_1 > \bar{G}_2 > \bar{G}_0$ ($0.81 > 0.74 > 0.72$). The “standard” collaborative filter r_1 has the larger bias. It is not surprising that $\bar{G}_1 > \bar{G}_2$ because r_2 explicitly discounts popularity. However, we do find it surprising that $\bar{G}_2 > \bar{G}_0$: beforehand, we could not rule out the possibility of r_2 ’s discounting leading to lower concentration. In

fact, in a small number of runs (17%), r_2 increases diversity, but in the majority of runs (83%) and on average it reduces diversity. A t-test of paired differences for unequal means (pre versus post recommendations) shows the differences are significant.

For r_1 , this is partly explained by (6), in which popularity determines what product is recommended. This creates a self-reinforcing cycle: popular items are recommended more, items recommended more are purchased more, purchased items are recommended more, and so on. Despite this, the increased concentration was not readily obvious: recommendations are generated in many local user groups, making a priori conclusions difficult. Although r_2 dampens the popularity bias, the result also originates from using only sales data to make recommendations. Products with limited historical sales have little or no chance of being recommended even if they would be favorably received by the consumer.¹²

Figure 11 shows the change in Gini for a range of δ . When the recommender has both awareness and salience effects, concentration increases in δ . The effect is most pronounced at high δ , where by construction the recommender has a bigger effect. In the special case $\delta = 0$, the recommender has only awareness effects. System r_1 continues to increase concentration, although by much less (+1.4%), as seen in Figure 11. The feedback loop is weaker: even if popular items are recommended more, recommended items are not necessarily purchased more because $\delta = 0$. As a result, the Gini's increase is attenuated. With r_2 , diversity increases under $\delta = 0$, although the magnitude is small (-1.4%) as shown in Figure 11. The deliberate exploration of r_2 coupled with low salience of recommendations increases diversity.

To summarize, when recommenders have both effects, diversity generally decreases. When recommenders affect only awareness, diversity decreases slightly for r_1 and increases slightly for r_2 . The $\delta = 0$ case is of conceptual interest, although it may not be commonplace. It is difficult to show consumers information without influencing them. As an example, the experiments of Senecal and Nantel (2004) show recommendations are influential even when consumers are aware of all products.

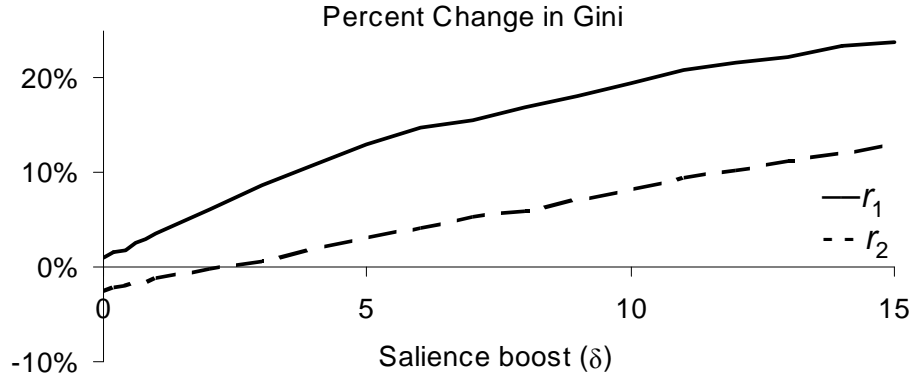
¹¹ The simulation code is available from the authors upon request

¹² With content based recommenders, we would not expect the same dynamics because past sales no longer affect which product is recommended. Studying the diversity question for content based systems is an interesting question but beyond the current scope.

Table 1. Comparison of r_1 and r_2 for $\delta = 5$ (1000 experiments; parentheses give standard errors)

	Average Gini		Average unique items aware of per person (AUIAP)		Average unique items bought per person (AUIBP)	
	r_1	r_2	r_1	r_2	r_1	r_2
Before	0.72 (.05)	0.72 (.05)	5.98 (1.33)	5.98 (1.33)	1.67 (0.32)	1.67 (0.32)
After	0.81 (.03)	0.74 (.04)	6.33 (1.34)	6.79 (1.41)	1.47 (0.17)	1.83 (0.26)
Change*	+0.09 (.03)	+0.02 (.02)	+0.37 (0.14)	+0.79 (0.24)	-0.20 (0.23)	+0.15 (0.25)

* Reports the average paired difference, and so this row may differ from the average after minus the average before. T-test of paired differences is significant ($<.05$) for the bottom row.

**Figure 11. Change in Gini by recommender and salience level (δ)**

6.3 Further Discussion of the Results

This section examines three ideas beyond the main results: aggregate versus individual-level diversity, product-level effects, and consumer-level effects.

(i) *Aggregate versus individual effects.* Table 1's middle section shows the average unique items aware of per person (AUIAP). This quantity increases under recommendations. Systems r_2 , as expected, creates a bigger increase, but in general both inform consumers of new products. Combining this observation with the change in Gini is revealing. Individually, consumers learn of more products (higher AUIAP), yet in aggregate diversity can decrease (higher Gini). This may explain user perceptions that recommenders create diversity even when an aggregate statistic, the Gini, reports less. A similar effect is seen in Table 1's right panel. This panel reports the average unique items bought per person (AUIBP). Under r_1 , consumers buy a narrower range of items, as seen by the lower AUIBP. Under r_2 , the outcome is different: AUIBP increases. Consumers are pushed toward products that are not necessarily popular,

which means they are less likely to have bought them previously. The Gini, however, still increases. This again leads to the finding that individual diversity can increase while aggregate diversity decreases. Consumers are discovering new products, but they are discovering the same products others have bought.

(ii) *Product-level view*. Figure 12 shows how the market share of particular products is affected by the recommender. Each point is a product, with the x coordinate giving the product's market share before recommendations and the y -coordinate its share after. The concentration bias is especially clear with r_1 . There is a systematic dispersion off the 45-degree line: low share products become even lower, and high share products become even higher. This reflects a 'poor get poorer' and 'rich get richer' phenomenon, both of which contribute to the increased Gini. The lower portion is related to the "cold-start" problem of collaborative filters, in which unpurchased/unrated items cannot be recommended (Schein et al. 2002). While the bias is not as acute with r_2 , the high share products are likely to gain more share in this case as well. It is interesting to note that the feedback effect of recommendations can turn some medium selling products into high selling ones, which is consistent with the findings from Section 4. As long as a product has modest sales, recommendations have the potential to make it more successful.

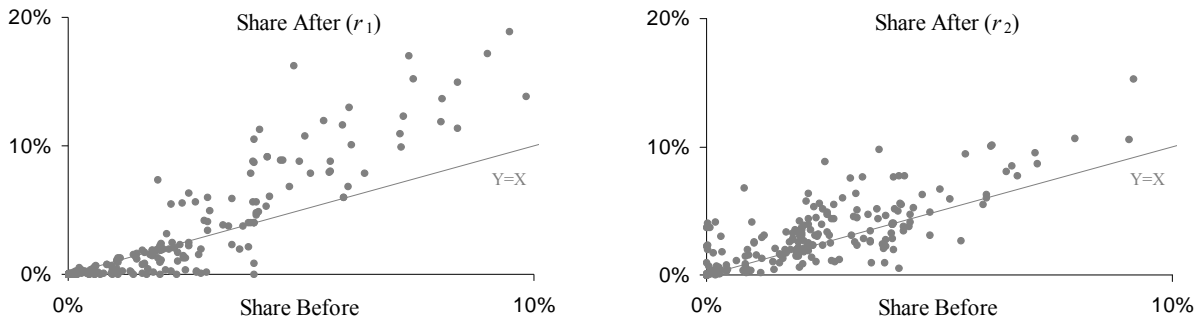


Figure 12. Market shares by product ($\delta=5$). Each point is a product whose coordinates give its market share before versus after recommendations. (Data pooled across 10 experiments)

(iii) *Consumer-level view*. The recommender systems push consumers toward the same products, and thus make consumers more similar in their purchases. This is illustrated in Figure 13. In the graph, consumers are nodes equally spaced on the perimeter of a circle. An edge joins consumers (c_i, c_j) if $\text{correlation}(\text{sales}_i, \text{sales}_j) > 0$. The edge's thickness is proportional to the correlation. Comparing the

graphs, the increased density at right shows that consumers have become more similar in their purchases. The figure alone does not imply the Gini has increased. For example, a correlation of 1 among all users could occur if everyone bought a single product (Gini=1), but it could also occur if all users bought all items equally (Gini=0). On its own, the figure shows consumers have become more alike. Combining this with the increased Gini, we see the complete picture: users are more similar (from Figure 13), and the items they purchase come from a smaller, more popular set (Ginis in Table 1).

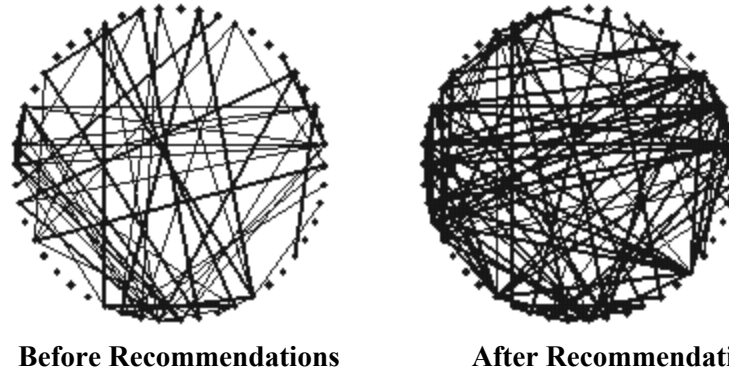


Figure 13. Each point is a user, and edge thickness is proportional to the pair's similarity

7. SENSITIVITY ANALYSIS

We approach sensitivity analysis in four parts: additional recommenders; best-seller lists in the base case; variety seeking in the utility specification, and alternate parameter values.

7.1 Alternate Recommender Systems

The base case examined two recommenders that were considered representative of industry practice. This section tests additional systems. A comparison of eight recommenders r_i ($i=1,\dots,8$) is given in Table 2.

The recommenders tested are as follows. r_1 and r_2 are as before. r_3 is another popularity-discounting variation on r_1 (Breese et al. 1998). It places discounting in the user similarity calculation but not the product selection calculation. (i.e. r_2 and r_3 add discounting in opposite places). Specifically, in (5) the user-item frequencies are multiplied by the inverse of each item's total sales (known as the "inverse document frequency" (*idf*) in the field of information retrieval); once the similar user group is determined, the undiscounted argmax of (6) is used. This still leads to an increase in the Gini. The magnitude is similar to r_1 's increase for the following reason. The intention of r_3 is to prevent latently different users

with little purchase history from being grouped together (e.g., two users who each bought Harry Potter and one very different item). Because of the initialization period, our users have several purchases, and so the similar user-groups under r_1 and r_3 are often similar (and hence $\overline{G}_1 \approx \overline{G}_3$). System r_4 is a combination of r_2 and r_3 : discounting is performed in both the user similarity calculation and argmax. As with its parents, r_4 also lowers diversity.

To build context for these comparisons, we tested four other designs ($r_5 - r_8$). System r_5 recommends the lowest sales product. As expected, it decreases the Gini. System r_6 recommends the median selling product. It also reduces the Gini because it diverts attention from otherwise higher selling products. System r_7 recommends the best-selling product and as expected increases the Gini. We highlight that the Gini under r_7 is not higher than under r_1 . A single product, the best seller, cannot be close to everyone. As a result, fewer users accept r_7 's recommendations, limiting its influence. In contrast, r_1 recommends local best-sellers, which are closer to each user and thus accepted more. r_8 is a best-seller list, which recommends the top 5 selling items. This system has the highest concentration: it shows the most popular items, and by showing multiple items increases the chance that at least one is close to the user. Similar results were confirmed experimentally by Salganik et al. (2006).

Table 2. Comparison of additional recommenders (1000 experiments).

	r_1	r_2	r_3 ($r_1 + idf$ weights)	r_4 ($r_2 + r_3$ combined)
\overline{G}_i	0.81 (0.03)	0.74 (0.05)	0.81 (0.03)	0.74 (0.05)
$\overline{G}_i - \overline{G}_0$ *	+0.09 (0.03)	+0.02 (0.02)	+0.09 (0.03)	+0.02 (0.02)
	r_5 (lowest)	r_6 (median)	r_7 (highest)	r_8 (top-five sellers)
\overline{G}_i	0.45 (0.10)	0.61 (0.03)	0.81 (0.04)	0.85 (0.03)
$\overline{G}_i - \overline{G}_0$ *	-0.27 (0.10)	-0.11 (0.04)	+0.09 (0.02)	+0.14 (0.04)

* Significant at the 0.05 level (2-sided, paired differences t-test for unequal means). For all cases, $\overline{G}_0 = 0.72$ (0.05).

7.2 Best-seller Lists in the Base Case

Without recommenders, consumers might obtain product suggestions from best-seller lists. We model this by introducing a best-seller list in the base case. This is equivalent to r_8 from the previous subsection – but now r_8 is the base case and r_1 or r_2 the treatment. Viewed this way (Table 2), the Gini decreases:

$\bar{G}_1 < \bar{G}_8$ ($0.82 < 0.85$) and $\bar{G}_2 < \bar{G}_8$ ($0.75 < 0.85$). If recommenders are simply replacements for best-seller lists, diversity can increase by cutting out what is an even more popularity-biased tool. Although it is unlikely that best-seller lists drive purchase decisions in all product categories, it seems feasible that best-seller lists affect purchase decisions in some categories. If so, this implies the role of recommenders is misunderstood. Relative to an ‘older’ world of best-seller lists, recommenders may reduce concentration, by virtue of cutting out the even more popularity-biased tool ($\bar{G}_1, \bar{G}_2 < \bar{G}_8$). But relative to a world without such lists, recommenders may increase concentration ($\bar{G}_1, \bar{G}_2 > \bar{G}_0$).

7.3 Modifying the Utility Specification: Variety Seeking

Since the choice model allowed for repeat purchases, we ask whether the concentration results are affected if consumers seek variety across purchase occasions. The concept of state dependence has a long history in choice models (e.g., McAlister 1982). “Structural state dependence” (Seetharaman 2004) is the extent to which prior purchases of a product affect its future purchase propensity; positive dependence is termed inertia, while negative dependence is termed variety seeking.

To incorporate variety and inertia in the specification, we use a common approach and define

$$\begin{aligned} v_{ijt} &:= -k \log \text{distance}_{ij} + \beta X_{ijt} \\ X_{ijt} &:= \alpha X_{ijt-1} + (1 - \alpha) I(c_i \text{ bought } p_j \text{ on } t - 1). \end{aligned}$$

X_{ijt} is an exponential smooth of purchase indicators $I(\cdot)$, and thus it summarizes how often and recently c_i has bought p_j . The parameter $\alpha \in (0,1)$ determines how much weight is placed on recent versus distant purchase occasions. β determines the effect strength, with $\beta < 0$ for variety seeking, and $\beta > 0$ for inertia. This approach has been used frequently in the literature (e.g., see Guadagni & Little 1983; Seetharaman 2004). Past empirical studies have found consistent values of α in the range 0.70-0.80, and thus we set $\alpha = 0.75$ (Guadagni & Little 1983; Lattin 1987; Seetharaman 2004). For β , we consider a range of values to explore both variety-seeking and inertia. The β term is not applied to the outside good, which by definition has the same distance to all consumers at all times.

Table 3 shows the Gini under state dependence. Under inertia ($\beta > 0$), the findings are directionally the same as before: concentration increases. Under high inertia, consumers do not want to deviate from their choices in the pre-recommendation period, and so the recommender's influence becomes limited. Under variety seeking ($\beta < 0$), concentration still increases for r_1 but by less. r_1 suggests heavily purchased items, which are less likely to provide variety. As a result, users ignore recommendations that are too similar, and the change in Gini is lessened. For r_2 , at moderate levels of variety seeking (e.g., $\beta = -5$) concentration still increases. At strong levels of variety seeking, the diversity can increase. For example, at $\beta = -20$, the Gini drops .03 points. We note that this level of variety seeking is high. Suppose c_i buys p_j semi-frequently so that $X_{ijt} = 0.5$ at some time. $\beta = -20$ implies $\beta X_{ijt} = (-20)(0.5) = -10$, which is twice as strong as the $\delta = 5$ salience effect of recommendations. Under such high variety seeking, the Gini decreases because users ignore recommendations of popular items and selectively accept recommendations of less popular ones. Whereas r_1 cannot supply these (r_1 focuses on past hits), r_2 makes this possible. Users want items not purchased recently, and r_2 's discounting meets this goal.¹³

The variety seeking results have an interesting interpretation. If consumers turn to recommendations only in their most variety-seeking moments, diversity increases under r_2 . However, as recommenders become ubiquitous, consumers are affected by them all the time – e.g., as with sites users visit regularly, such as personalized news, personalized radio, and personalized retail. In these instances, diversity decreases even under r_2 .

Table 3. Gini values under state dependence at $\delta=5$. For variety seeking $\beta < 0$, and for inertia $\beta > 0$

β	-30	-20	-10	-5	0	5	10	20	30
$\overline{G_1 - G_0}$	+.04	+.05	+.07	+.08	+.09	+.07	+.04	+.02	+.02
$\overline{G_2 - G_0}$	-.04	-.03	-.01	+.01	+.02	+.03	+.02	+.02	+.02

¹³ Letting variety seeking go to $-\infty$, we test the case where repeat purchases are not allowed. To implement this, we ensure the number of products is more than the number of simulation iterations. The results are similar to those in Table 3. Concentration still increases under r_1 . Consumers always buy new products, but the recommender still pushes them toward products that others purchased previously. With r_2 , concentration can decrease at such a high level of variety seeking as discussed above.

7.4 Altering Other Simulation Parameters

We also examine sensitivity to other simulation parameters (e.g., number of consumers and products, map distributions). The main sensitivity results appear in the online appendix, and others are available on request. In general, we find that varying these parameters affects the degree of the results (e.g., the Gini may increase more versus less), but the substantive findings remain the same.

8. WELFARE IMPLICATIONS

Thus far we have examined how recommenders affect concentration. We next ask whether these changes leave firms and consumers better off.

For firms, we examine the change in sales. For consumers, we examine the change in product fit. Consumers' product fit is defined as the average of $-k \log \text{distance}_{ij} + \varepsilon_{ijt}$ over all purchases, including those of the outside good. This measure reflects the map distance between consumers and purchased products. Figure 14 shows these quantities, plotting the numbers in percent change so that both firm and consumer effects can be plotted together. When $\delta = 0$, the recommender has a pure awareness effect. The firm's sales are higher, and consumers find products closer to them. The gains for both parties are larger under r_2 . The deliberate exploration of r_2 helps consumers find better products, which translates into higher sales (fewer no-purchases) for the firm. When $\delta > 0$, recommenders have both awareness and salience effects. At low δ , the results are the same as $\delta = 0$. At high δ , firms always sell more: the greater the salience δ , the more likely the consumer is to buy the recommended product than the outside good. For consumers, high δ increases the average map distance of purchases: consumers may forgo a slightly closer product if the recommended product has increased salience. A slightly better song or news article may be available deeper in the website, but the recommendation's salience makes it easier to click.

Does this mean consumers are worse off if $\delta \gg 0$? If the salience effect is simply a momentary increase in purchase probability but does not contribute to post-purchase satisfaction, then consumers are worse off because their purchases are farther away. However, a more complete answer considers additional factors. First, it is possible that δ , or part of it, should be included in the consumer's utility.

This is the case if recommendations add value to the choice occasion. In this case, the consumer effect in Figure 14 becomes positive and increasing (not shown for clarity). This view is consistent with several logit applications in marketing in which a store display adds utility to the choice occasion (e.g., Guadagni & Little 1983). For example, a display means the user does not have to walk down the aisle to get the product or price information. Similarly, choosing the recommended item may

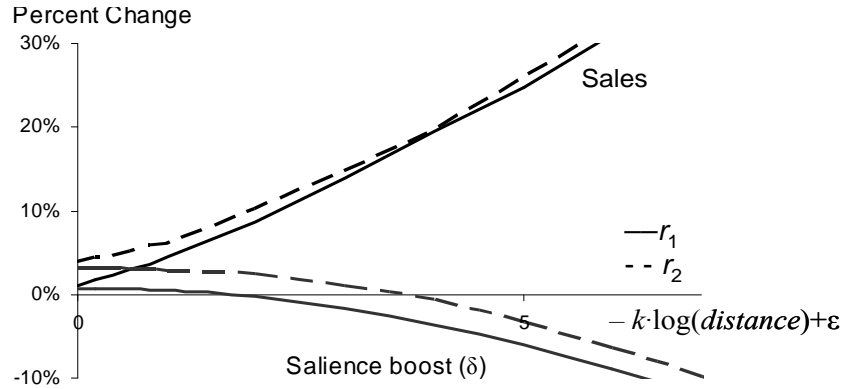


Figure 14. Percent changes in consumer and producer surplus for varying levels of salience (δ)

save time browsing the site or effort in making product comparisons. Second, to the extent media products have positive externalities, these may offset the increased distance. For example, watching the same movies as others is valuable if it permits discussion. In this case, the recommender serves a coordinating role whose value is not fully accounted for by measuring map distances. For firms, we measured the change in sales. To the extent changes in concentration simplify inventory management, these factors are also unaccounted for. Last, recommenders may have welfare implications at the societal level. Sunstein (2001) discusses the risk of “filters” creating a fragmented society. He asks whether en masse filtering of all but one's exact interests will reduce people's ability to understand one another. Such considerations are beyond the current scope, but we raise them to show that an exhaustive analysis of welfare implications would involve more than changes in sales and map distances.

9. CONCLUSIONS AND FUTURE WORK

This paper examined the effect of popular recommender designs on sales concentration and offered evidence that recommenders do influence sales diversity. Several common recommenders were found to

exert a concentration bias. Thus the traditional view that recommenders increase diversity may not always hold. The work also demonstrated that some designs may be associated with greater bias than others. The results have important managerial and consumer implications. We find that recommenders can increase sales, and recommenders that discount popularity appropriately may increase sales more. For consumers, we showed that the awareness effects of recommenders can inform consumers of better (closer) products. However, if recommendations are highly salient, popularity-influenced recommendations may displace what would otherwise be better product matches. Future, empirical work would be valuable for determining the relative strength of the awareness versus salience effects.

Given these findings, why do consumers feel that recommendations have increased the range of media they consume? We offered several explanations. The first is that diversity can increase at the individual level but still decrease in aggregate. This was borne out under r_2 , in which each user became aware of more items and purchased more unique items, but the Gini still increased. Individuals may be exploring more choices, but they are being pushed toward the same choices. Second, if recommenders are simply replacing best-seller lists, diversity can increase by virtue of cutting out an even more popularity-biased tool. A final possibility is that the effect of increased product offerings outweighs the effect of recommenders. Increased offerings may lower concentration (Anderson 2006; Brynjolfsson et al. 2007), while recommenders could temper but not reverse the effect. Examining the simultaneous effects of recommenders and increased offerings is an interesting question for future work.

A final interesting aspect arose to the extent that externalities exist for media goods. If, for example, there is a benefit to reading popular books or seeing popular movies (e.g., by increasing the likelihood of being able to discuss the experience with others), then consumer utility involves a tradeoff between a Hotelling-like similarity and the externality from a popular product. To the extent such externalities are strong, it would be interesting to see if they pose a limit, or upper bound, on the degree of diversity consumers would ever prefer. If this were the case, a concentration bias may be more desirable than previously considered. We hope to explore these questions in future work as well.

10. REFERENCES

- Adomavicius, G. and A. Tuzhilin. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. Know. Data. Eng.* 17(6):734-749.
- Anderson, C. 2006. *The Long Tail*. New York: Hyperion.
- Ansari, A., S. Essegai, and R. Kohli. 2000. Internet recommendation systems. *J. Mktg. Res.* 37(3):363-375.
- Arthur, W. B. 1994. *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: University of Michigan Press.
- Bikhchandani, S., D. Hirshleifer, and I. Welch. 1998. Learning from the behavior of others: conformity, fads, and informational cascades. *J. of Economic Perspectives* 12:3 151-170.
- Bodapati, A. 2008. Recommendation systems with purchase data. *J. Mktg. Res.* 45(1):77-93.
- Breese, J., D. Heckerman, and C. Kadie. 1998. Empirical analysis of predictive algorithms for collaborative filtering. In *Proc. of the 14th Conference of Uncertainty in AI*.
- Brynjolfsson, E., Y. Hu, and D. Simester. 2007. Goodbye Pareto principle, hello long tail: the effect of search costs on the concentration of product sales. *SSRN eLibrary* 953587.
- Brynjolfsson, E., Y. Hu, and M. Smith. 2006. From niches to riches: the anatomy of the long tail. *Sloan Management Review* 47(4) 67-71.
- Chevalier, J. and A. Goolsbee. 2003. Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics* 1(2):203-222.
- Chintagunta, P. 2002. Investigating category pricing behavior at a retail chain. *J. Mktg. Res.* 39(2):141-54
- Clemons, E.K., G. G. Gao, and L.M. Hitt. 2006. When online reviews meet hyperdifferentiation. *J. of Management Information Systems* 23(2):149-171.
- Cooke, A., H. Sujan, M. Sujan, and B. Weitz. 2002. Marketing the unfamiliar: the role of context and item-specific information in electronic agent recommendations. *J. Mktg. Res.* 39(4):488-497.
- DeSarbo, W. S. and J. Wu. 2001. The joint spatial representation of multiple variable batteries collected in marketing research. *Journal of Marketing Research* 38(2):244-253.
- Elrod, T. and M. P. Keane. 1995. A factor-analytic probit model for representing the market structure in panel data. *J. of Marketing Research* 32(1):1-16.
- Fleder, D. and K. Hosanagar. 2007. Recommender systems and their impact on sales diversity. In *Proc. of the 8th ACM Conference on Electronic Commerce*, p. 192-199.
- Guadagni, P.M. and J.D.C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Science* 2(3):203-238.

- Hendricks, K. and A. T. Sorensen. 2007. Information spillovers in the market for recorded music. *NBER working paper W12263*. Last accessed 02/18/08 at <www.stanford.edu/~asorensen>.
- Hill, B. M., D. Lane, and W. Sudderth. 1980. A strong law for some generalized urn processes. *Annals of Probability* 8(2):214-226.
- Hotelling, H. 1929. Stability in competition. *The Economic Journal* 39:41-57.
- Lattin, J. M. 1987. A model of balanced choice behavior. *Marketing Science* 6(1):48-65.
- Linden, G., B. Smith, and J. York. 2003. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing* 7(1):76-80.
- McAlister, L. 1982. A dynamic attribute satiation model for choices made across time. *J. Cons. Res.* 9(3):141-150.
- Mooney, R. J. and L. Roy. 2000. Content-based book recommending using learning for text categorization. *5th ACM Conf. on Digital Libraries*, p.195-204.
- Narayanan, S., P. Manchanda, and P. K. Chintagunta. 2005. Temporal differences in the role of marketing communication in new product categories. *J Mktg. Res.* 42(3):278-290.
- Resnick, P. and H. Varian. 1997. Recommender systems. *Comm. of the ACM* 40(3):56-58.
- Salganik, M. J., P. S. Dodds, and D. J. Watts. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854-856.
- Schafer, J., J. Konstan, and J. Riedl. 1999. Recommender systems in e-commerce. *ACM EC*, p.158-166.
- Schein A. I., A. Popescul, L. H. Ungar, and D. M. Pennock. 2002. Methods and metrics for cold-start recommendations. *SIGIR*, p. 253-260.
- Schweidel, D. A., E. Bradlow, and P. Fader. 2007. Modeling the evolution of customers' service portfolios. *SSRN eLibrary* 985639
- Seetharaman, P. B. 2004. Modeling multiple sources of state dependence in random utility models: a distributed lag approach. *Marketing Science* 23(2):263-271.
- Sen, A. 1976. Poverty: an ordinal approach to measurement. *Econometrica* 44(2):219-231.
- Senecal, S. and J. Nantel. 2004. The influence of online product recommendations on consumers' online choices. *Journal of Retailing* 80:159-169.
- Sunstein, C. R. 2001. *Republic.com*. Princeton: Princeton University Press.
- 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.

Electronic Companion to *Blockbuster Culture's Next Rise or Fall:
The Impact of Recommender Systems on Sales Diversity*

Daniel Fleder and Kartik Hosanagar
The Wharton School, OPIM Dept.

Part I: Proofs for the Analytical Model

This section contains the proofs of all results from the main paper. The results are derived in a random walks framework. The process can also be described through the use of an urn function. Hill et al. (1980) derived a strong law for continuous urn functions and related the limiting distribution to the function's stationary points. This appendix derives results for a discontinuous case not covered by their results.

For a simple random walk on \mathbb{Z}^1 , let

$$\begin{aligned} S &: = \text{Event } \{\text{Particle at } i \text{ moves to } i + 1 \text{ on next move}\} \\ \theta &: = P(S) \\ i \rightarrow j &: = \text{Event } \{\text{particle at } i \text{ ever reaches } j\} \end{aligned}$$

Lemma 1 *One-away return probabilities*

$$\begin{aligned} P(i \rightarrow i + 1) &= \begin{cases} \theta(1 - \theta)^{-1} & , \theta < \frac{1}{2} \\ 1 & , \theta \geq \frac{1}{2} \end{cases} \\ P(i \rightarrow i - 1) &= \begin{cases} \theta^{-1}(1 - \theta) & , \theta > \frac{1}{2} \\ 1 & , \theta \leq \frac{1}{2} \end{cases} \end{aligned}$$

Proof. This is a basic result in stochastic processes (Durrett 2005, pg. 294) ■

Based on the main paper, we define the following:

$$\begin{aligned} p &:= P(\text{consumer picks } w \text{ on own}) \\ r &:= P(\text{consumer follows recommendation}) \\ W_t, B_t &:= \text{Total } w, b \text{ in Urn 2 prior to purchase } t \\ Z_t &:= W_t - B_t \\ X_t &:= W_t / (W_t + B_t), \text{ which is } w\text{'s share before } t \\ g(X_t) &:= P(w \text{ recommended at } t \mid X_t) \end{aligned}$$

The chance a consumer selects w at t is

$$\begin{aligned} f(X_t) &:= P(\text{consumer buys } w \text{ at } t | X_t) \\ &= p(1 - r) + g(X_t)r \end{aligned}$$

The function f is known as an urn function (Hill et al. 1980). It maps the unit interval into itself and defines a process that is Markov but can have nonstationary transition probabilities.

As defined in the paper, g is the step function

$$g(X_t) := \begin{cases} 0 & , X_t < \frac{1}{2} \\ \frac{1}{2} & , X_t = \frac{1}{2} \\ 1 & , X_t > \frac{1}{2} \end{cases}$$

This choice of g recommends the product with majority share.

Substituting $g(X_t)$ into $f(X_t)$ gives

$$f(X_t) = \begin{cases} p(1 - r), & X_t < \frac{1}{2} & \text{"}l\text{"} \\ p(1 - r) + r/2, & X_t = \frac{1}{2} & \text{"}m\text{"} \\ p(1 - r) + r, & X_t > \frac{1}{2} & \text{"}h\text{"} \end{cases}$$

The letters l , m , h are shorthand for the expressions at their left. They also have a geometric interpretation: f is a modified step function (shifted and stretched), and l , m , and h correspond to the height of f 's lower segment, middle point, and upper segment respectively. This interpretation was shown graphically in the main paper.

While the main paper states results about $\{X_t\}$, we can equivalently study $\{Z_t\}$: studying sales instead of shares carries the same information because X_t is a statistic of sales. This switch, however, is beneficial because Z_t changes by one unit each period and so is amenable to a random walks framework.

For any time τ at which $Z_\tau = 0$ (i.e. $W_\tau = B_\tau$), three events are possible

$$\begin{aligned} WB &:= \text{Event}\{Z_t > 0 \text{ for all } t > \tau | Z_t = 0\} \\ BW &:= \text{Event}\{Z_t < 0 \text{ for all } t > \tau | Z_t = 0\} \\ RTZ &:= \text{Event}\{Z_t = 0 \text{ for some } t > \tau | Z_t = 0\} \end{aligned}$$

In words, WB is the event that w leads b forever after the next time step; BW is the event that b leads w forever after the next time step; and RTZ is the event that Z_t returns to zero at some future time point.

We now have a random walk on \mathbb{Z}^1 beginning at the origin for which the transition probabilities of moving left versus right are $(l, 1 - l)$, $(m, 1 - m)$, and $(h, 1 - h)$ depending on whether the particle is left of zero, at zero, or right of zero.

Lemma 2 *Never Return Probabilities are Always Non-Zero*

For $p \in [0, 1]$ and $r \in (0, 1)$, either $P(WB) > 0$, $P(BW) > 0$, or both are > 0 .

Proof. By conditioning on the first event

$$\begin{aligned} P(WB) &= P(Z_{\tau+1} = 1)P(Z_t > 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = 1) \\ P(BW) &= P(Z_{\tau+1} = -1)P(Z_t < 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = -1) \\ P(RTZ) &= 1 - P(WB) - P(BW) \end{aligned}$$

Because the walk begins at the origin, the terms $P(Z_{\tau+1} = 1)$ and $P(Z_{\tau+1} = -1)$ follow immediately as m and $1 - m$. For the rightmost terms, Lemma 1 is needed.

To apply the lemma, three cases will need to be distinguished. The interpretation of these cases was given in the main paper. Here, we re-parameterize the cases from (p, r) notation to (l, h) notation to clarify how the lemma is applied.

The change of parameters assumes $r \in (0, 1)$, which is to say the recommender has some influence. The boundary case $r = 0$ or 1 is not of interest, for it does not concern recommender systems, but for completeness will be discussed afterward.

Case 1. $l < \frac{1}{2}, h \leq \frac{1}{2} \Leftrightarrow p \leq (\frac{1}{2} - r)(1 - r)^{-1}$

$$\begin{aligned} P(WB) &= P(Z_{\tau+1} = 1)P(Z_t > 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = 1) \\ &= m[1 - P(1 \rightarrow 0)] \\ &= m(1 - 1) \\ &= 0 \end{aligned}$$

$$\begin{aligned} P(BW) &= P(Z_{\tau+1} = -1)P(Z_t < 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = -1) \\ &= (1 - m)[1 - P(-1 \rightarrow 0)] \\ &= (1 - m) \left(1 - \frac{l}{1 - l} \right) \end{aligned}$$

$$\begin{aligned} P(RTZ) &= 1 - P(WB) - P(BW) \\ &= 1 - (1 - m) \left(1 - \frac{l}{1 - l} \right) \end{aligned}$$

Case 2. $l < \frac{1}{2}, h > \frac{1}{2} \Leftrightarrow (\frac{1}{2} - r)(1 - r)^{-1} < p < \frac{1}{2}(1 - r)^{-1}$

$$\begin{aligned}
P(WB) &= P(Z_{\tau+1} = 1)P(Z_t > 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = 1) \\
&= m[1 - P(1 \rightarrow 0)] \\
&= m \left(1 - \frac{1-h}{h} \right)
\end{aligned}$$

$$\begin{aligned}
P(BW) &= P(Z_{\tau+1} = -1)P(Z_t < 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = -1) \\
&= (1-m)[1 - P(-1 \rightarrow 0)] \\
&= (1-m) \left(1 - \frac{l}{1-l} \right)
\end{aligned}$$

$$\begin{aligned}
P(RTZ) &= 1 - P(WB) - P(BW) \\
&= 1 - m \left(1 - \frac{1-h}{h} \right) - (1-m) \left(1 - \frac{l}{1-l} \right)
\end{aligned}$$

Case 3. $l \geq \frac{1}{2}, h > \frac{1}{2} \Leftrightarrow p \geq \frac{1}{2}(1-r)^{-1}$

$$\begin{aligned}
P(WB) &= P(Z_{\tau+1} = 1)P(Z_t > 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = 1) \\
&= m[1 - P(1 \rightarrow 0)] \\
&= m \left(1 - \frac{1-h}{h} \right)
\end{aligned}$$

$$\begin{aligned}
P(BW) &= P(Z_{\tau+1} = -1)P(Z_t < 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = -1) \\
&= (1-m)[1 - P(-1 \rightarrow 0)] \\
&= (1-m)(1-1) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
P(RTZ) &= 1 - P(WB) - P(BW) \\
&= 1 - m \left(1 - \frac{1-h}{h} \right)
\end{aligned}$$

The above expressions show that for every case either $P(WB) > 0$, $P(BW) > 0$, or both are > 0 .

Recall that the parameter space is the unit square $\{(p, r) : 0 \leq p, r \leq 1\}$. The above cases cover the space $\{(p, r) : 0 \leq p \leq 1 \wedge 0 < r < 1\}$. To be exhaustive and cover the entire square, we point out the two trivial remaining cases. The results in these cases are clear, but the model does not apply to recommender systems unless $r \in (0, 1)$. When $r = 0$, this gives a Bernoulli process that converges to p by the weak

law of large numbers. When $r = 1$, the process immediately converges to the product purchased on the first occasion, which is w with the chance of a fair coin flip. Setting $r = 1$ means consumers, always accept the recommendation. The first purchase is determined by a Bernoulli trial; the system then recommends this product, since it has higher sales; and the consumer then accepts this recommendation since $r = 1$. This product now has even higher sales, so it continues to be recommended and purchased indefinitely. ■

The above lemma showed that it is possible for a product to obtain a majority share and never lose it. Next we show this must be the case: after sufficient time, some product is guaranteed to obtain a majority share and maintain it. Further, how likely it is for w versus b to obtain this majority, lasting share is shown.

Lemma 3 *Probability of obtaining a lasting majority share*

Case	$\lim_{t \rightarrow \infty} P(Z_t > 0)$	$\lim_{t \rightarrow \infty} P(Z_t < 0)$	$\lim_{t \rightarrow \infty} P(RTZ)$	Interpretation
1	0	1	0	b always wins
2	$1 - \gamma$	γ	0	either can win
3	1	0	0	w always wins

$$\text{where } \gamma = \frac{(1-m)(1-\frac{l}{1-l})}{m(1-\frac{1-h}{h})+(1-m)(1-\frac{l}{1-l})}$$

Proof.

$$\begin{aligned}
& \lim_{t \rightarrow \infty} P(Z_t > 0) \\
&= \sum_{i=1}^{\infty} P(WB \text{ occurs after } i^{th} \text{ time } w = b) \\
&= \sum_{i=1}^{\infty} P(RTZ)^{i-1} P(WB | RTZ \text{ occurs } i-1 \text{ times}) \\
&= \sum_{i=1}^{\infty} P(RTZ)^{i-1} P(WB) \\
&= P(WB) \sum_{i=0}^{\infty} P(RTZ)^i \\
&= \frac{P(WB)}{1 - P(RTZ)} \\
&= \frac{P(WB)}{1 - (1 - P(WB) - P(BW))} \\
&= \frac{P(WB)}{P(WB) + P(BW)}
\end{aligned}$$

The analogous argument gives

$$\lim_{t \rightarrow \infty} P(Z_t < 0) = \frac{P(BW)}{P(WB) + P(BW)}$$

We can also confirm that

$$\begin{aligned} & \lim_{t \rightarrow \infty} P(RTZ) \\ &= \lim_{t \rightarrow \infty} \{1 - P(Z_t > 0) - P(Z_t < 0)\} \\ &= 1 - \frac{P(WB)}{P(WB) + P(BW)} - \frac{P(BW)}{P(WB) + P(BW)} \\ &= 0 \end{aligned}$$

Combining the above expressions with the results from the previous lemma gives

Case	$\lim_{t \rightarrow \infty} P(Z_t > 0)$	$\lim_{t \rightarrow \infty} P(Z_t < 0)$
1	$\frac{0}{0 + (1-m)(1-\frac{l}{1-l})} = 0$	$\frac{(1-m)(1-\frac{l}{1-l})}{0 + (1-m)(1-\frac{l}{1-l})} = 1$
2	$\frac{m(1-\frac{1-h}{h})}{m(1-\frac{1-h}{h}) + (1-m)(1-\frac{l}{1-l})} = 1 - \gamma$	$\frac{(1-m)(1-\frac{l}{1-l})}{m(1-\frac{1-h}{h}) + (1-m)(1-\frac{l}{1-l})} = \gamma$
3	$\frac{m(1-\frac{1-h}{h})}{m(1-\frac{1-h}{h}) + 0} = 1$	$\frac{0}{m(1-\frac{1-h}{h}) + 0} = 0$

■

The above result shows that in the limit (i) some product must develop and maintain a majority share and (ii) the chance of w versus b obtaining this lasting majority. We now determine what those limiting shares are.

Proposition 1 *Support of the limiting distribution: As $t \rightarrow \infty$, $\{X_t\}$ converges to either one or two values given by*

Case	Support point 1	Support point 2
1	l	n/a
2	l	h
3	h	n/a

Proof.

Case 1. By Lemma 3, $\lim_{t \rightarrow \infty} P(Z_t < 0) = 1$. Thus $\lim_{t \rightarrow \infty} X_t < \frac{1}{2}$ because by definition $Z_t < 0 \Leftrightarrow X_t < \frac{1}{2}$. Because $\lim_{t \rightarrow \infty} X_t < \frac{1}{2}$ and f is constant for $X_t < \frac{1}{2}$, $\lim_{t \rightarrow \infty} X_t = f(X_t | X_t < \frac{1}{2}) = l$.

In words, if w has the minority share, its chance of being bought is l ; since it was proved to remain the minority share for this case, its limiting share must also be l .

Case 2. By the previous result, $\lim_{t \rightarrow \infty} P(Z_t > 0) > 0$, $\lim_{t \rightarrow \infty} P(Z_t < 0) > 0$, and $\lim_{t \rightarrow \infty} P(RTZ) = 0$. Thus either $\lim_{t \rightarrow \infty} Z_t > 0$ or $\lim_{t \rightarrow \infty} Z_t < 0$. Suppose we have a sample path for which $\lim_{t \rightarrow \infty} Z_t < 0$. By the argument for Case 1 above, $\lim_{t \rightarrow \infty} \{X_t | Z_t < 0\} = l$. In contrast, suppose we have a sample path for which $\lim_{t \rightarrow \infty} Z_t > 0$. By the argument for Case 3 below, $\lim_{t \rightarrow \infty} \{X_t | Z_t > 0\} = h$.

Since once of these outcomes is guaranteed to occur (some product will have a lasting, majority share), then $\lim_{t \rightarrow \infty} X_t \rightarrow X$ where X is a random variable with support $\{l, h\}$.

Case 3. By Lemma 3, $\lim_{t \rightarrow \infty} P(Z_t > 0) = 1$. Thus $\lim_{t \rightarrow \infty} X_t > \frac{1}{2}$ because by definition $Z_t > 0 \Leftrightarrow X_t > \frac{1}{2}$. Because $\lim_{t \rightarrow \infty} X_t > \frac{1}{2}$ and f is constant for $X_t > \frac{1}{2}$, $\lim_{t \rightarrow \infty} X_t = f(X_t | X_t > \frac{1}{2}) = h$. The interpretation is analogous to Case 1 above. ■

Corollary 1 *In case 2, either product can obtain a lasting majority share: $\lim_{t \rightarrow \infty} P(X_t < \frac{1}{2}) > 0$ and $\lim_{t \rightarrow \infty} P(X_t > \frac{1}{2}) > 0$.*

Proof.

By Proposition 1, $\{X_t\} \rightarrow X$ where X is a random variable with support $\{l, h\}$. Now in Case 2, by definition $p < \frac{1}{2(1-r)}$. Thus $l \equiv p(1-r) < \frac{(1-r)}{2(1-r)} = \frac{1}{2}$. Similarly in Case 2, by definition $p > \frac{(\frac{1}{2}-r)}{(1-r)}$. Thus $h \equiv p(1-r) + r > \frac{(\frac{1}{2}-r)(1-r)}{(1-r)} + r = \frac{1}{2}$. This establishes that $l < \frac{1}{2}$ and $h > \frac{1}{2}$. Since the support is $\{l, h\}$, either product can obtain a lasting, majority share, regardless of p . ■

Proposition 2 *The distribution of $\lim_{t \rightarrow \infty} X_t$ is*

Case	$P(\lim_{t \rightarrow \infty} X_t = l)$	$P(\lim_{t \rightarrow \infty} X_t = h)$
1	1	0
2	γ	$1 - \gamma$
3	0	1

where again $\gamma = \frac{(1-m)(1-\frac{l}{1-l})}{m(1-\frac{1-h}{h})+(1-m)(1-\frac{l}{1-l})}$.

Proof.

The distribution's support was shown to be $\{l\}$, $\{h\}$, or $\{l, h\}$ depending on the case (from Proposition 1). The probabilities are determined as follows. For Cases 1 and 3, we know from Proposition 1 there is convergence to a single outcome; thus chance of converging to that particular outcome is 1. For Case 2, by Proposition 1 the process converges to one of two limiting outcomes, l and h . By Corollary 1, outcome l means b has majority share and outcome h means w has majority share. By Lemma 3, b versus w obtains the majority share with chance γ versus $1 - \gamma$, which means that l and h also occur with probabilities γ and $1 - \gamma$. ■

Thus far, we have derived the limiting distribution of $\{X_t\}$. With the limiting behavior of $\{X_t\}$ understood, we now ask whether that limit reflects more or less concentration.

The term “increased concentration” refers to shares that are less equal than they would be without recommendations. Formally, we define “increased concentration” to mean $\lim_{t \rightarrow \infty} X_t > p$ when $p > \frac{1}{2}$ and $\lim_{t \rightarrow \infty} X_t < p$ when $p < \frac{1}{2}$. When $p = \frac{1}{2}$, increased concentration occurs when $\lim_{t \rightarrow \infty} X_t \neq \frac{1}{2}$.

Proposition 3 *The relation of the limiting support points to concentration is*

Case	Support points	Effect on concentration relative to p
1	1	Increased concentration
2	2	Case 2A $p \in (\frac{1-r}{2-r}, \frac{1}{2-r})$. Increased concentration for both points. Case 2B $p \notin (\frac{1-r}{2-r}, \frac{1}{2-r})$. Increased concentration for one point, decreased for the other.
3	1	Increased concentration

As before, $p \in [0, 1]$ and $r \in (0, 1)$. (The case of $r = 0$ or 1 was discussed above.)

Proof.

Case 1.

The process converges to $l \equiv p(1 - r) < p$. Since $p < \frac{1}{2}$ in Case 1, this implies increased concentration. To verify that $p < \frac{1}{2}$, start with Case 1's definition $p \leq (\frac{1}{2} - r)(1 - r)^{-1}$. Viewing p as a function of r , its derivative is $\frac{dp}{dr} = -(1 - r)^{-1} + (\frac{1}{2} - r)(1 - r)^{-2}$. The condition $r \in (0, 1)$ implies $\frac{dp}{dr} < 0$ on $(0, 1)$, and thus $p(r)$ is maximized as $r \rightarrow 0$. Since $p(0) = (\frac{1}{2} - 0)(1 - 0)^{-1} = \frac{1}{2}$, this bound shows $p < \frac{1}{2}$ on the interval $(0, 1)$.

Case 2.

First consider the case when $p < 0.5$. The two possible limits are $p(1 - r)$ and $p(1 - r) + r$. Note that $p(1 - r) < p$, and thus this outcome always involves increased concentration. Now, consider the other outcome $p(1 - r) + r$. Since Case 2's definition states $p > (\frac{1}{2} - r)(1 - r)^{-1}$, it follows that $p(1 - r) + r > \frac{1}{2}$. Clearly, this reverses the popularity order of the two products. However, it increases concentration only if $p(1 - r) + r > (1 - p)$. Simplifying this expression, concentration increases only if $p > (1 - r)(2 - r)^{-1}$. Similarly, for the case in which $p > 0.5$, concentration increases in both outcomes only if $p < (2 - r)^{-1}$. Combining results, we see that concentration always increases if $p \in (\frac{1-r}{2-r}, \frac{1}{2-r})$. Otherwise, concentration increases for one limiting outcome and decreases the other.

Case 3.

The process converges to $h \equiv p(1 - r) + r = p + r(1 - p) > p$. Since $p > \frac{1}{2}$ in Case 3, this implies increased concentration. To verify that $p > \frac{1}{2}$, start with Case 3's definition $p \geq \frac{1}{2}(1 - r)^{-1}$. Viewing p as a function of r , its derivative is $\frac{dp}{dr} = \frac{1}{2}(1 - r)^{-2}$. The condition $r \in (0, 1)$ implies $\frac{dp}{dr} > 0$ on $(0, 1)$, and thus $p(r)$ is minimized as $r \rightarrow 0$. Since $p(0) = \frac{1}{2}(1 - 0)^{-1} = \frac{1}{2}$, this bound shows that $p > \frac{1}{2}$ on the interval $(0, 1)$. ■

Part II: Alternative Simulation Settings

Note on the Simulation

The main simulation and sensitivity were programmed by the authors in Matlab. All code is available on request.

Overview of this Sensitivity Analysis

This section presents sensitivity analyses for the simulation with regard to the map distribution, awareness distribution, and the problem size (number of consumers versus products). Sensitivity to the salience parameter δ , recommender system employed, and variety seeking were presented in the main paper.

To aid the reader in understanding the parameter space, we have organized the online appendix by four cases of interest. Any results not covered by these cases are available from the authors.

The cases are defined by the distributions generating the consumer-product maps and the awareness states. Within each case, we also vary map parameters, such as the number of consumers (I), the number of products (J), and the recommender system itself.

Review of the Awareness Specification

For reference, we restate the awareness specification used in the paper so that it can be referred to below. Each consumer is assumed aware of a subset of the J products. Only items in this awareness set can be purchased. The initial awareness states for each consumer-product pair are sampled according to

$$P(c_i \text{ aware of } p_j) = \lambda e^{-distance_{0j}^2/\theta} + (1 - \lambda)e^{-distance_{ij}^2/k\theta}$$

Above, $distance_{0j}$ and $distance_{ij}$ are respectively the Euclidean distances from the origin to product p_j and from consumer c_i to product p_j . The constant $\lambda \in [0, 1]$. The higher is λ , the more users are aware of central, mainstream products (left term), and the higher is $1 - \lambda$, the more users are aware of products in their local neighborhood. The θ and $\kappa\theta$ terms are scaling parameters, determining how fast awareness decays with distance. Note, this does not mean users are aware of the same products. They are likely to overlap in their awareness of the central products but less so in the local ones.

Case 1. Normal Consumer-Product Maps. Awareness is central and local.

This case is identical to the main paper; we reproduce the results here to facilitate comparison. The distribution of consumer and product points on the map is standard bivariate normal. The awareness distribution has $\lambda = .75$, which means consumers are relatively more aware of mainstream goods than niche ones. This assumption is consistent with a market that has mass advertising, which makes consumers aware of (roughly) the same, central products.

A detailed discussion of the main results was presented in the main paper. Sensitivity to the problem size (I, J) did not appear in the main paper and is discussed next. The results are in Table A1 under Case 1.

To start, there are $(I, J) = (50, 50)$ consumers and products. When there are fewer consumers than products ($I = 25, J = 50$), $\overline{G_0}$ is higher than the original (50, 50) case. There are more products for the same number of consumers, and it results that there are more products with no or low sales. For example, if consumers always buy the closest product, then more products will have zero sales, yielding a higher Gini in the base case. Although $\overline{G_0}$ is higher, the change in Gini $\overline{G_i} - \overline{G_0}$ is still positive.

When there are more consumers than products ($I = 50, J = 25$) $\overline{G_0}$ is lower than the original (50, 50) case. There are more consumers for the same products, and so fewer products will have zero or low sales. In this case, $\overline{G_0}$ is lower, but the change in Gini $\overline{G_i} - \overline{G_0}$ is still positive as highlighted in Table A1.

When I and J are equal but lower ($I = 25, J = 25$), $\overline{G_0}$ is higher than the original (50, 50) case. With fewer data points, by chance some products are closer to more consumers; these products have higher sales and so increase the Gini. Conversely, when the map fills in with many more data points, the chance that some products have no or low sales decreases and so does the Gini. Again, though the base case Gini differs, the change in Gini $\overline{G_i} - \overline{G_0}$ is in the same direction.

Case 2. Normal Consumer-Product Maps. Central Awareness Only.

The distribution of consumer and product points on the map is again standard normal, as in the main paper. The awareness distribution is identical to the main paper except $\lambda = 1$. This creates a scenario in which users are more aware of central products than peripheral ones. This case provides a robustness check as to whether diversity will increase if users are only aware of central products and then discover the outer ones via the recommender.

As Table A1-Case 2 shows, concentration in the base case ($\overline{G_0}$) is higher than in the main paper. This arises because all users are focused on roughly the same, central products. In contrast, in Case 1 users were aware of their local neighborhoods as well, creating more diversity in the base case. Despite the differences in initial concentration, the change in Gini is in the same direction: it increases. The above

results hold under both r_1 and r_2 . However, the change in Gini is smaller under r_2 for the same reasons as in the main paper. Changing the balance between the number of consumers (I) and products (J) affects the initial concentration, but it does not affect the sign of $\overline{G_i} - \overline{G_0}$, as discussed in Case 1 above.

Case 3. Uniform Consumer-Product Maps. Awareness is central and local.

The distribution of consumer and product points on the map is now uniform on a square centered at the origin and with sides of length four. A box this size (± 2 in each direction from the origin) roughly captures data from a standard bivariate normal distribution. This helps change the distribution’s shape without changing its scale and facilitates comparisons among cases. The awareness distribution is identical to the main paper: $\lambda = .75$ again to create the idea that users are aware mainly of central products but also a few local ones.

As Table A1-Case 3 shows, $\overline{G_0}$ is higher than in the main paper. With the uniform map, products spread out more than in the normal map. Since awareness still has a large central component, few people are aware of the now more numerous peripheral products. With more low-selling peripheral products, the initial Gini is higher. However, despite the higher initial Gini, the change in Gini $\overline{G_i} - \overline{G_0}$ is in the same direction as the main results. Again, the effect under r_1 is greater than r_2 . Imbalances in I versus J have the same effects as described above.

Case 4. Pareto Consumer-Product Maps. Pareto Awareness.

We now test a heavy tailed distribution, the power law. The power law is synonymous with the Pareto distribution; the former typically refers to the PDF and the latter to the CDF (Adamic 2000).

The power law distribution has PDF

$$f(x) = Cx^{-\alpha} = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$

The restriction $x \geq x_{min} > 0$ is needed for the density to integrate to one. In addition, $\alpha > 2$ is required to have a finite mean.

The consumer and product points are sampled from this density with each point’s coordinate an i.i.d. draw. This distribution creates a roughly L-shaped map. Most points are in the bottom left of the map. There is some spread northward and some eastward, creating an L shape. After generating the maps, they are mean centered. Centering does not affect the shape or inter-point distances; it shifts the origin so that awareness can be defined relative to $(0, 0)$ rather than some other point that would change with each map generated. Figure A1 shows an example.

For the distribution’s parameters, we set $\alpha = 2.2$. This value is chosen to match estimates discussed in the main paper for media and retail markets. Two parameters remain, x_{min} for the power law distribution and k for the main simulation. x_{min} affects the map’s scale. It determines whether points fall, for example, mostly in the 0-1 range versus the 0-1,000 range. k , as in the main text (§5.2), helps transform distance to similarity; k determines whether a given distance on the map is considered large or small by the consumer. Rather than keep both free, we fix $k = 10$ as in the main paper and adjust x_{min} to obtain empirically observed levels of the Gini. Setting $x_{min} = 0.05$ gives a base-case Gini of 0.70–0.75. These Gini values match the estimates from prior work as discussed in the main paper.

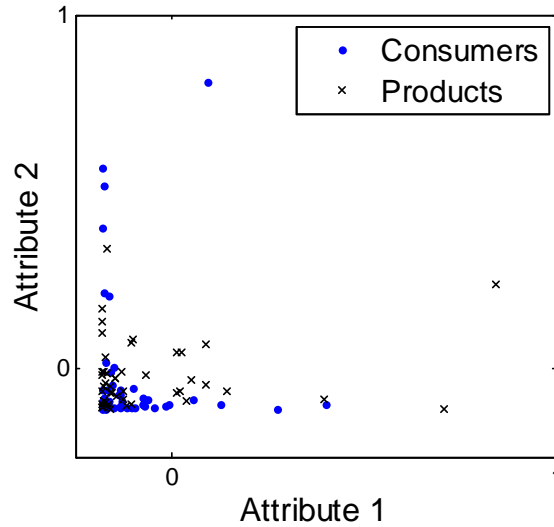


Figure A1. Map of product and consumer points with Pareto distribution

Awareness is defined analogous to the previous cases. The general form of the awareness specification is

$$P(c_i \text{ aware of } p_j) = \lambda f(\text{distance}_{0j}^2; \alpha) + (1 - \lambda) f(\text{distance}_{ij}^2; \kappa\alpha)$$

with f some function and α and $\kappa\alpha$ scaling parameters. The constant $\lambda \in [0, 1]$. In the main paper, f has the decay of the normal distribution. Here, we again test the power law $f = Cx^{-\alpha}$, which has a heavier tailed decay.

For the expression above to be a probability, $f = Cx^{-\alpha}$ must be between 0 and 1. To bound this by 1, we assume the user is aware of a product if it is less than x_{min} distance away. Letting $x_{min} = 0.05$ again, this means $f(x_{min} = 0.05) \equiv 1$. Since $f(x_{min}) = Cx_{min}^{-\alpha} = 1$, this implies $C = x_{min}^{\alpha}$. Thus the awareness decays as $f = x_{min}^{\alpha} x^{-\alpha}$.

For awareness decay relative to the origin, we again use the $\alpha = 2.2$ estimate. For awareness decay relative to the consumer’s neighborhood, we use $\kappa\alpha = 3\alpha$. This is consistent with the main paper’s goal of making the local awareness neighborhood smaller than the center one. (When the local neighborhood is larger than the center one, consumers know almost every product that might interest them, and the recommender serves no purpose.)

As Table A1-Case 4 shows, despite the different map, the results are similar. The Gini increases under r_1 and r_2 ; imbalances in I versus J have the same effects as described above; and r_2 has a smaller effect than r_1 . However, the effects under both r_1 and r_2 are smaller compared with the main paper. In the Pareto maps, there is a mass of points in the bottom-left and a few points much farther away. For the large mass, the recommender has the same effects as in Case 1. For the latter, peripheral points, they are too spread out for the recommender to be effective. The system does recommend products that a user’s peers purchased, but those products are too far away for the recommendations to be accepted. As a result, recommendations effectively influence only a subset of the population and the change in Gini is less.

In this online appendix, we have tested the sensitivity of our main results under different map distributions, different recommenders, and different numbers of consumers and products. This analysis shows the results to be qualitatively similar to those discussed in the main paper.

TABLE A1. Sensitivity for the four cases ($\delta = 5$; $n = 1000$ simulations each)

Case	I	J	Recomm- ender	$\overline{G_0}$	$\overline{G_i}$	$\overline{G_i - G_0}$
1. Normal Maps. Awareness in Center and Local	50	50	1	0.72 (0.05)	0.81 (0.03)	0.09 (0.03)
	25	50	1	0.80 (0.04)	0.87 (0.03)	0.07 (0.03)
	50	25	1	0.70 (0.07)	0.79 (0.06)	0.08 (0.04)
	25	25	1	0.77 (0.06)	0.85 (0.05)	0.08 (0.04)
	50	50	2	0.72 (0.05)	0.74 (0.05)	0.02 (0.02)
	25	50	2	0.80 (0.04)	0.80 (0.04)	<.01 (0.02)
	50	25	2	0.71 (0.07)	0.75 (0.07)	0.04 (0.03)
	25	25	2	0.77 (0.06)	0.79 (0.07)	0.02 (0.04)
2. Normal Maps. Awareness in Center Only	50	50	1	0.75 (0.05)	0.84 (0.03)	0.09 (0.03)
	25	50	1	0.80 (0.04)	0.89 (0.02)	0.09 (0.03)
	50	25	1	0.75 (0.07)	0.81 (0.05)	0.06 (0.03)
	25	25	1	0.78 (0.06)	0.86 (0.04)	0.08 (0.04)
	50	50	2	0.75 (0.05)	0.77 (0.04)	0.02 (0.02)
	25	50	2	0.80 (0.04)	0.81 (0.04)	<0.01 (0.02)
	50	25	2	0.75 (0.07)	0.76 (0.07)	0.01 (0.03)
	25	25	2	0.78 (0.06)	0.79 (0.06)	0.01 (0.03)
3. Uniform Maps. Awareness in Center and Local	50	50	1	0.77 (0.05)	0.84 (0.04)	0.07 (0.02)
	25	50	1	0.85 (0.04)	0.90 (0.03)	0.05 (0.02)
	50	25	1	0.76 (0.07)	0.85 (0.05)	0.09 (0.04)
	25	25	1	0.83 (0.06)	0.89 (0.04)	0.06 (0.03)
	50	50	2	0.77 (0.05)	0.81 (0.05)	0.04 (0.03)
	25	50	2	0.85 (0.04)	0.87 (0.04)	0.02 (0.02)
	50	25	2	0.75 (0.07)	0.83 (0.06)	0.07 (0.04)
	25	25	2	0.83 (0.06)	0.87 (0.06)	0.04 (0.03)
4. Pareto Maps. Awareness in Center and Local	50	50	1	0.75 (0.04)	0.81 (0.04)	0.06 (0.02)
	25	50	1	0.83 (0.04)	0.86 (0.03)	0.03 (0.02)
	50	25	1	0.72 (0.07)	0.77 (0.05)	0.06 (0.03)
	25	25	1	0.80 (0.06)	0.83 (0.05)	0.04 (0.03)
	50	50	2	0.75 (0.05)	0.76 (0.05)	0.01 (0.02)
	25	50	2	0.83 (0.04)	0.83 (0.04)	<0.001 (0.01) *
	50	25	2	0.72 (0.06)	0.74 (0.06)	0.01 (0.03)
	25	25	2	0.80 (0.06)	0.80 (0.06)	<0.01 (0.02)

All comparisons of the change in Gini are significantly different from zero ($p < 0.05$) except those marked * (t-test of paired differences for zero mean difference).

References

- [1] Adamic, L. A. 2000. Zipf, Power-law, Pareto: a ranking tutorial. *HP Research*.
<http://www.hpl.hp.com/research/idl/papers/ranking/>. Accessed April 6, 2008.
- [2] Durrett, R. 2005. *Probability: Theory and Examples*. Belmont, CA: Thompson Brooks–Cole.
- [3] Hill, B. M., D. Lane, and W. Sudderth. 1980. A strong law for some generalized urn processes. *Annals of Probability* 8(2):214-226