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

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

Lemma 1 *One-away return probabilities*

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

$$P(i \to i-1) = \begin{cases} \theta^{-1}(1-\theta) & , \ \theta > \frac{1}{2} \\ 1 & , \ \theta \leq \frac{1}{2} \end{cases}$$

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

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

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 ammenable 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 immediatley 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. $l < \frac{1}{2}, h \le \frac{1}{2} \Leftrightarrow p \le (\frac{1}{2} - r)(1-r)^{-1}$

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

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

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

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

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

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

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

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

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

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

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

The above expressions show that for every case either 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	$\lim_{t\to\infty} P(Z_t > 0)$	$\lim_{t\to\infty} P(Z_t < 0)$	$\lim_{t\to\infty} P(RTZ)$	Interpretation
1	0	1	0	b always wins
2	$1-\gamma$	γ	0	either can win
3	1	0	0	w always wins

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

Proof.

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

The analogous argument gives

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

We can also confirm that

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

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

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

■

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

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

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

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

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

Proof.

By Proposition 1, $\{X_t\} \to X$ 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* $\lim\limits_{t\to\infty} X_t$ *is*

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

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

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 know 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 $\lim_{t\to\infty} X_t > p$ when $p > \frac{1}{2}$ and $\lim_{t\to\infty} X_t < p$ when $p < \frac{1}{2}$. When $p = \frac{1}{2}$, increased concentration occurs when $\lim_{t\to\infty} X_t \neq \frac{1}{2}$.

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 $p \in \left(\frac{1-r}{2-r}, \frac{1}{2-r}\right)$. Increased concentration for both points.
		Case 2B $p \notin \left(\frac{1-r}{2-r}, \frac{1}{2-r}\right)$. 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 convergres 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 $r \to 0$. 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 $p \in \left(\frac{1-r}{2-r}, \frac{1}{2-r}\right)$. Otherwise, concentration increases for one limiting outcome and decreases the other.

Case 3.

The process convergres 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 $r \to 0$. Since $p(0) = \frac{1}{2}(1-0)^{-1} = \frac{1}{2}$, this bound shows that $p > \frac{1}{2}$ on the interval $(0,1)$. $\blacksquare$

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 $\overline{G_i - G_0}$ 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 $\overline{G_i - G_0}$ 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 $\overline{G_i - G_0}$ 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 r2 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 $\overline{G_i - G_0}$, 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 $\overline{G_i - G_0}$ 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.

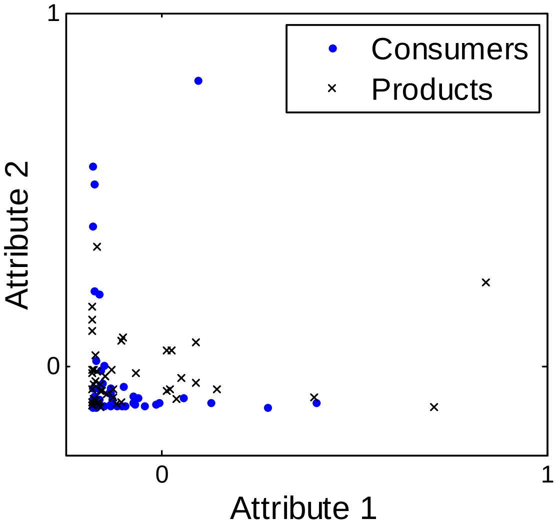

Figure A1. Map of product and consumer points with Pareto distribution

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

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

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 $f(x_{min}) = Cx^{-\alpha}_{min} = 1$, this implies $C = x^{\alpha}_{min}$. Thus the awareness decays as $f = x^{\alpha}_{min} x^{-\alpha}$.

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	Recommender	$\overline{G_0}$	$\overline{G_i}$	$\overline{G_i - G_0}$
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