Recommendation Systems and Sales Concentration:  
The Moderating Effects of Consumers’ Product Awareness  
and Acceptance to Recommendations

Ling-Ling Wu  
lwu@management.ntu.edu.tw  
Dept. of Information Management, National Taiwan University, Taipei, Taiwan

Yuh-Jzer Joung  
joung@ntu.edu.tw

Tzung-En Chiang  
betray1001@gmail.com

Abstract
This research investigates how recommendation systems affect sales concentration in the e-market. In the literature, there have been two opposite theories accounting for this issue: the winner-take-all theory and the long tail theory. The former predicts that the market will be homogenized by recommendation systems, whereas the later predicts that the market will be fragmented. We argue that both directions of influence are possible, depending on the types of products that recommendation systems favor. Furthermore, we argue that the effects of recommendation systems are moderated by consumers’ cognitive states before and after they encounter the systems, namely their awareness of products and their acceptance level to recommendations. We used the simulation method on the MovieLens dataset, which consists of 100,000 ratings from 943 users on 1682 movies. The simulation results support our hypotheses, showing that, relative to the condition of no recommendation, the content-based recommendation systems tend to decrease sales concentration, whereas the collaborative filtering tends to increase it. Moreover, the magnitudes of increase and decrease are moderated by consumers’ awareness type and acceptance level.

1. Introduction
Recommendation systems have been widely used in online stores. Such automated systems can reduce search costs for online consumers by providing help in locating the products they want from the seemingly endless list of alternatives. Moreover, recommendation systems can also reduce transaction costs for online sellers through efficiently targeting consumers with customized service. As such, an important question arises: how recommendation systems affect sales concentration, i.e. the extent to which sales of products are attributed to a small portion of products, given the lower search costs and transaction costs in the e-market? The answer of this question is very critical for decisions of product assortment, advertising expenses that support appropriate amount of product variety, and supply chain management that ensures availability of products requested by consumers. However, due to opposite views and mixed research findings, this important question still calls for further rigorous studies.

Interestingly, in the literature, there have been two opposite theories accounting for the effect of recommendation systems on sales concentration: the winner-take-all theory [1], [2] and the long tail theory [3], [4]. The winner-take-all theory predicts that the market will be homogenized due to reduced search costs in locating high quality products, which later become superstars [2]. This theory assumes that consumers have similar tastes of products. Therefore, recommender systems tend to recommend popular products and hence increase the sales concentration. This theory is supported by some research findings [5], [6]. The long tail theory, on the other hand, refutes the traditional wisdom of economics and argues that reduced search costs will lead to fragmentation of the market because recommenders make it much easier for the online consumers to find niche products that suit their special needs [7], [8]. Some empirical studies support this theory by showing that the sales of niche products do increase in the online market [7], [9], [10].

As we can see from the above discussion, these two theories have different assumptions of consumer taste variability, which leads to their distinctive arguments about how recommendation systems work. The winner-take-all theory assumes that people have similar tastes such that recommendation of popular products, such as what collaborative filtering systems tend to do, should satisfy consumers. On the other hand, the long tail theory assumes that people have unique tastes such that recommendation systems should recommend niche products, as content-based systems would do. That is, these two theories assume different mechanisms underlying recommenders due to their distinctive assumptions about variability of consumers’ tastes. Even though these two theories emphasize the role of recommenders and what recommenders “should” do, it still remains unknown what will happen to the market when different kinds of recommenders are put into service for various kinds of consumers. Therefore, the first purpose of this study is to investigate how different kinds of recommendation systems, such as
content-based systems and collaborative filtering systems, affect sales concentration.

Moreover, these two theories do not consider the cognitive states of individual consumers before and after they encounter the recommenders. Consumers with unique tastes might pay more attention to niche products whereas a big portion of consumers might only be aware of popular products before they encounter recommenders. Type of awareness is crucial because it relates to what kinds of products the consumer is interested in and because consumers compare recommended products against the products that they are aware of. Thus, we will assess this factor in our research. Moreover, when facing recommendation systems, consumers are susceptible to the influence of them to different degrees. Some consumers are inclined to accept recommendations derived from the systems, whereas some consumers are more likely to rely on the information they sought on their own. Therefore, this study will investigate the impacts of recommendation on sales concentration by taking consideration of these two crucial factors in consumers’ experience of using recommendation systems: awareness of products and acceptance of recommendations. By taking these factors into consideration, this study might provide some explanations for the contradictory divergence between the two theories discussed in the previous text.

In the following section, we will first discuss sales of concentration in E-commerce predicted by these two theories, and the related empirical results. Then, we will compare different types of recommendation systems and infer their effects on sales concentration. Lastly, the cognitive states of consumers before and after they encounter the systems, i.e. their awareness type and acceptance level, will be discussed to infer their interactive effects with recommendation systems.

2. Literature review

2.1 Sales concentration in e-commerce

There is a much larger variety of products provided in the electronic market than those in the brick-and-mortar stores. For instance, large brick-and-mortar stores typically stock up to 1500 DVDs, while online retailers might have more than 100,000 titles. Despite the large amount of titles in online retailers, the transaction costs incurred for the seller and the search costs for the consumers are much lower due to the effective facilitation of online systems, such as recommendation systems. The recommendation systems can efficiently communicate selective products to the targeted consumers during the selling process. If the products are information goods, the distribution costs are even close to zero because they can be digitized and distributed online. Besides lowering down the transaction costs on the supply side, recommendation systems can reduce search costs on the demand side by actively recommending products or relevant product information to the consumers.

Reduction in transaction costs and search costs induces heated dispute on the change of demand. According to the long tail theory, new demands of niche products emerge and the market will shift from hit products to niche products as the transaction costs and search costs reduce [3], [7]. On the supply side, the online retailers have incentive to promote and recommend products in the long tail. As suggested by Goldstein & Goldstein [11], as a large amount of obscure products are made available online, the seller has incentive to promote the back catalog, which typically costs less than new releases. On the other hand, the winner-take-all theory argues that low search costs reduce information asymmetry between seller and buyer such that consumptions will converge to high quality products, which will eventually become superstars [1], [2]. Moreover, the manufacturers of hit products can “lock” their dominance by means such as providing complementary goods or extending their “installed base” with low prices [6].

Interestingly, the empirical results suggest a mixed picture that partially supports each theory. Some research results show that online consumers do explore the long tail more, suggesting that the diverse needs do exist. For instance, Elberse [10] found that a large share of DVD consumers, especially those with a higher consumption frequency and a narrower selection of DVD genres, regularly choose obscure products that are not likely available offline. On the other hand, she also found that even for consumers opt for the most obscure products, hit products constitute a significantly big share of their choices. The study of Elberse and Oberholzer-Gee [9] reported a slight shift toward the niche products whereas there was also a dramatic increase in the number of titles that did not sell at all. That is, the market is still dominated by a few superstars, even though best selling items do not reach previous sales levels. Moreover, the study of Fleder and Hosannagar [5] argues that recommendation systems may introduce new products to individual consumers, but they push similar consumers toward the same products such that sale concentration increases at aggregate. Even though the dominance of superstars supports the winner-take-all theory, significant price dispersion online [12], [13] refutes its prediction and suggests that search costs are still omnipresent in online environment. In conclusion, the empirical results suggest that both theories simplify the effects of e-market, and more constituents should be investigated to account for sales concentration in e-commerce.

Even though both theories emphasize the important role of recommendation systems in reducing transaction costs and search costs, the existence of different recommendation systems is usually overlooked because the underlying mechanism of
recommend systems have already been assumed. Nonetheless, the assumed systems are quite different in these two theories. As we know, different recommendation systems tend to recommend different kinds of products. For instance, some system may recommend products that have been popularly accepted by other consumers, while another system may recommend niche products specifically tailored to a consumer. It is quite possible that different kinds of recommendation systems might have different effects on sales concentration. Therefore, the factor of recommendation system types needs to be singled out in order to see its effects on sales concentration.

2.2 Recommendation systems and sales concentration

Online recommendation systems provide consumers better access to the products that fit their needs by making personalized suggestions based on their likes or dislikes [14]. We consider two classes of recommendation systems: collaborative filtering (CF) and content-based.

CF systems, the most prevalent recommendation systems in industry [15], use the preferences or ratings of other similar peers as the base for recommendations [16], [17], [18]. The peers could be similar to the matched consumer in terms of sharing the same demographic background, or the same interests on the products. For example, if Tom and John have bought a lot of identical albums, and Tom bought Michael Jackson’s new album while John hasn’t, the system will most likely recommend Michael Jackson’s new album to John. Ideally, such systems assume a sufficient number of ratings or purchase available to serve as the base of recommendations. However, most product items are rated or purchased by very few consumers, such that they will be less likely recommended [16]. Therefore, in actual practice, CF systems tend to recommend popular items, even to the consumers with special tastes.

Content-based systems recommend products based upon the information of the product itself, rather than on the preferences of other peer consumers [19], [20]. This could be achieved by calculating the similarities between product attributes and a consumer’s preference profile such that unique, personalized recommendation to each individual consumer is possible. For example, suppose a consumer likes action movies very much, moderately prefers comedy movies, and dislikes thriller movies. Content-based recommenders will recommend movies that are highly evaluated as action movies, but unfavorably evaluated as thriller movies.

Previous research has missed the effect of recommendation systems on sales concentration. As discussed above, CF systems tend to recommend popular products, while content-based systems are more likely to recommend obscure products to consumers if their tastes demand so. Moreover, past research has shown that consumers tend to be influenced by recommenders more than expected [21]. Therefore, it is predicted in this study that, given everything else being equal, CF systems tend to increase sales concentration whereas content-based systems tend to decrease sales concentration, as stated in Hypothesis 1.

Hypothesis 1: Relative to sales without recommendation systems, collaborative filtering systems increase sales concentration whereas content-based systems decrease sales concentration.

2.3 Seeking product information online: before and after encountering recommendation systems

As stated by Herbert Simon: “a wealth of information creates a poverty of attention”. Therefore, the attention of online consumers becomes particularly precious for electronic retailers because products that do not attract consumers’ attention will be excluded from consumption. Before recommendation systems have a chance to influence consumers, consumers might have already attended to some relevant product information and be aware of some items. Out of the huge number of product items, consumers selectively attend to a small set of them. Such selective awareness could be affected by their experiences with the products and their tastes. For instance, the existence of hit products suggests that quite a few consumers share the same interests and pay attentions to popular items. They could be aware of popular items due to exposure to words of mouth or advertisements in various media channels. On the other hand, consumers could be constantly seeking for variety and have a high degree of need for uniqueness [22], [23]. As a result, they could be attentive to niche products that fit their particular needs or tastes. For instance, consumers with higher consumption frequency and a narrower interest to DVD genres are more likely to explore into the niche products [10].

In this study, we will call the first kind of awareness as hit awareness, whereas the second, niche awareness. Nonetheless, this categorization does not imply that people are dichotomized as two distinct awareness types. Instead, there is a large amount of people entertain these two kinds of awareness simultaneously, which we will call hybrid awareness. As consumers usually choose products from those in their awareness, it is logically to predict that hit awareness leads to the highest sales concentration, niche awareness, the lowest, and hybrid awareness, an intermediate level of sales concentration, which is stated in Hypothesis 2:

Hypothesis 2: Sales concentration is the highest in the condition of hit awareness, followed by the
condition of hybrid awareness, and the lowest in the condition of niche awareness.

Note that recommendation systems will bring new products into the attentions of consumers. Therefore, these three awareness conditions should be affected differently by different kinds of recommendation systems. As postulated in Hypothesis 1, content-based recommenders generally decrease sales concentration as they tend to recommend products tailored to individual consumers’ preferences. However, the degree of decrease should be different in these three awareness conditions. The impact of content-based recommenders on the group of niche awareness should be the least because the consumers already look at special kinds of products that fit their needs, such that content-based recommenders would not recommend a fundamentally different set of products. On the other hand, the impact of content-based recommenders on the group of hit awareness would be the most conspicuous because they will direct consumers’ attention to a different set of products. Such a set of products could contain items that suit their needs but never have a chance to come into their attentions. Therefore, the decrease of sales concentration will be the largest for the group of hit awareness when content-based recommenders are used.

The logic is the same for the collaborative filtering systems. Such systems tend to recommend popular products, which, therefore, do not add many new products to the hit awareness. As a result, even though collaborative filtering systems typically increase sales concentration, they will increase the sales concentration in the niche awareness condition more than they do for those in the hit awareness condition. In another word, the impact of recommendation systems on sales concentration depends on the type of awareness that consumers initially have.

**Hypothesis 3:** There is an interaction effect between awareness types and recommendation systems on sales concentration.

After receiving recommendations from the systems, consumers would evaluate the recommended products, and decide whether they want to accept the recommendation or stick with what they have from their initial awareness. Generally speaking, consumers might have a higher degree of perceived risks towards the services provided online, compared to traditional channels [24], such that they could be suspicious of recommendation systems and treat them as a kind of advertisement. Nonetheless, if consumers feel that the recommender system is useful and justified, they are more likely to accept recommendations [25], [26]. As argued by Bodapati [27], besides purchase probability of recommended products, consumers’ reaction to the recommendation system should also be taken into consideration to assess the effectiveness of recommendation actions. Therefore, in order to accurately assess the impacts of recommendation systems, consumers’ susceptibility to the influence of the system will also be included in this research. If consumers are reluctant to accept the recommendations, then recommendation systems will not engender difference as predicted in Hypotheses 1 and 3. On the other hand, if consumers are highly susceptible to the influence of recommendation systems, then their impacts become significant in the sense that the interaction effect that we described in Hypothesis 3 will be even stronger. In this study, we assess the effects of three different acceptance levels (high, middle, low) to examine if this factor moderates the interaction effects between recommenders and awareness on sales concentration.

**Hypothesis 4:** The interaction effect between recommenders and awareness is the strongest in high acceptance level, next in the medium level, and lowest in the low level of acceptance to recommendation systems.

### 3. Methodology

We model online purchase as a three-stage process to evaluate the four hypotheses: *product selection, recommendation, and decision making*. This process captures a typical consumer behavior when shopping in an e-commerce web site. First, a consumer selects some products into his shopping cart. Then, the system recommends some other products to the consumer. Finally, the consumer makes purchase decision by checking out items from the shopping cart or from the products recommended by the system. Below we describe in detail how we model each stage.

#### 3.1 Product selection

When a consumer shops online, he must have some knowledge on the products he is looking for. From these products, after some survey and comparison, he chooses some of them into shopping cart and proceeds to check out. For example, when a consumer buys/rents movies online, he is typically aware of some movies, either from browsing in the web site, or from other types of sources prior to visiting the web site. We refer to the set of products a consumer is aware of before making his selection as the *awareness pool*. The purpose of marketing, in essence, is to put some products into a consumer’s awareness pool; for otherwise, the consumer may never purchase the products. In general, there are two types of marketing: *mass* and *niche*. The former aims to promote a product popularly known, while the latter tails a product to a specific group of customers according to their needs [28]. Therefore, we model a consumer’s awareness pool based on two factors: a product’s “exposure” and its “similarity” to the consumer.

Formally, we use \( \text{expo}(x, M) \) to denote a product \( x \)'s exposure among the set \( M \) of products in
consideration. For similarity, we assume that each product $x$ has an attribute $x.Attr$, and each consumer $c$ has a preference $c.pref$ so that similarity $\text{sim}(c, x)$ between $x$ and $c$ can be calculated. Based on $\text{expo}(x, M)$ and $\text{sim}(c, x)$, we compute awareness level of a consumer $c$ to a product $x$ by

$$\text{aware}(c, x) = w_p \times \text{expo}(x, M) + w_r \times \text{sim}(c, x)$$ (1)

where $w_p$ and $w_r$ are weights on product exposure and consumer preference, respectively. A product’s awareness level then is used to determine the chance for it to be included in a user’s awareness pool; the higher the level, the greater the chance. When $w_p = 1$ and $w_r = 0$, a consumer’s awareness pool consists of solely highly exposed products, regardless of whether they match his preference or not. On the other hand, the case $w_p = 0$ and $w_r = 1$ let us model consumers who pay attention only to products that match their preferences. We refer to these two types of awareness hit and niche, respectively. In practice, however, awareness is often a mix of both. Therefore, in addition to the two extreme types of awareness (hit and niche), we will use nonzero $w_p$ and $w_r$ in (1) to calculate a consumer’s awareness level to a product. We refer to this type of awareness as hybrid. To make the hybrid case more realistic, we allow different consumers to have different $w_p$ and $w_r$. For example, a novice may have a higher weight on exposure, while an expert may have a higher weight on preference.

### 3.2 Recommendation

Before a consumer proceeds to check out the items he selected in the first stage, the web site will usually recommend some products to him. There are a number of recommendation strategies to adopt. We consider two popular ones: content-based filtering and collaborative filtering. In content-based filtering, the system recommends products that best fit a consumer’s preference. To model this, we again use the function $\text{sim}(c, x)$ to determine if a product $x$ fits a consumer $c$.

Collaborative filtering recommends products that have been purchased by other customers of the same interest. For this, we need a function to determine if two users are of the same interest. We assume that each user $c$ has a profile $c.profile$ to describe him. Depending on the design, $c.profile$ may or may not be equivalent to a user’s preference $c.pref$ discussed in the previous section. In general, a user’s profile may contain demographic data as well as personal interest and purchase history. We also assume a function $\text{sim}(c1, c2)$ for measuring the similarity between two users $c1$ and $c2$. Then, given a user $c$ and some threshold, we can compute the set $C$ of “neighbors” close to $c$. Collaborative filtering then recommends best selling products among $C$ to $c$.

### 3.3 Decision making

The final stage of the online purchase process is decision making. In this stage, a consumer needs to decide, from the products he has initially selected and the products recommended by the systems, the ones to purchase. We use acceptance level to characterize the chance a consumer leans towards accepting the system’s recommendation. The higher the level, the more likely he prefers recommended products to his initial selection. Several factors may affect a consumer’s acceptance level, including the quality of the recommendation, the consumer’s personal characteristics (e.g., whether or not he is easily persuaded), his experiences with the system and how well he trusts the system. Although understanding exactly how these factors affect acceptance level is crucial for building a more effective recommendation system, it is orthogonal to our research (and nontrivial!) Therefore, in the paper we simply assume that the level is given. Operationally, a system’s acceptance level can be measured by the percentage of recommendations accepted by the consumers.

### 4. Simulation design

We use simulation to collect sales information from the three-stage online purchase process, thereby to assess the effects of awareness type, recommendation, and acceptance level on sales concentration. Sales concentration is measured by the Gini coefficient [29]. For the simulation, we use a MovieLens dataset (http://www.grouplens.org/node/73) to model product exposure, product attributes, and user preferences. The dataset consists of 100,000 ratings, on the scale of 1 to 5, from 943 users on 1682 movies. Each user has rated at least 20 movies, and 100 ratings on average. Table 1 shows a typical rating record of a user. The record indicates that the user has rated three movies, with ratings 2, 4, and 5, respectively. Each movie’s genres are also indicated in the record. For example, movie 1 is of Action and Thriller. There are a total of 19 genres (unknown, Action, Adventure, Animation, Children’s, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western). For simplicity, we use only 4 genres in the example.

<table>
<thead>
<tr>
<th>Movie</th>
<th>Action</th>
<th>Comedy</th>
<th>Romance</th>
<th>Thriller</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movie 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Movie 2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Movie 3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>
4.1 Data translation

Based on the dataset, a movie $x$’s attribute can be modeled by the genres it belongs. We use $x.\text{Attr}$ to denote the attribute, and $x.\text{Attr}[i]$ (a Boolean value) to represent if $x$ is of genre $i$.

We model a user $c$’s preference as a vector $c.\text{pref}$, where $c.\text{pref}[i]$ indicates his preference level in a genre $i$. Let $x_1, \ldots, x_n$ be the movies he has rated, and $r_1, \ldots, r_n$ respectively be their ratings. Then, $c.\text{pref}[i]$ is calculated as follows:

$$c.\text{pref}[i] = \frac{\sum_{j=1}^{n} x_j.\text{Attr}[i] * r_j}{n}$$

For example, according to Table 1, the user has the following preference vector: $c.\text{pref} = \{2, 3, \frac{4}{5}, \frac{7}{8}\}$.

Based on a user’s preference, we can calculate the similarity between the user $c$ and a movie $x$ as follows:

$$\text{sim}'(c, x) = \frac{\sum_i c.\text{pref}[i] * x.\text{Attr}[i]}{\sum_i x.\text{Attr}[i]}$$

For example, the following shows $c$’s preference on the two new movies 4 and 5.

<table>
<thead>
<tr>
<th></th>
<th>Action</th>
<th>Comedy</th>
<th>Romance</th>
<th>Thriller</th>
<th>sim'(c, x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movie 4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>13/9</td>
</tr>
<tr>
<td>Movie 5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9/6</td>
</tr>
</tbody>
</table>

To integrate similarity and product exposure into awareness level, we further normalize similarity using min-max normalization as follows:

$$\text{sim}(c, x) = \frac{\text{sim}'(c, x) - \min \{\text{sim}'(c, x) | x \in M\}}{\max \{\text{sim}'(c, x) | x \in M\} - \min \{\text{sim}'(c, x) | x \in M\}}$$

where $M$ is the set of all movies.

For exposure, we use a movie’s rating frequency to represent its exposure, as intuitively, the more the number of users rating a movie, the more attention the movie has received from the public. Again, we use min-max normalization to normalize exposure. Let $x.\text{freq}$ denote a movie $x$’s rating frequency. Then, $x$’s exposure is defined as follows:

$$\text{expo}(x, M) = \frac{x.\text{freq} - \min \{z.\text{freq} | z \in M\}}{\max \{z.\text{freq} | z \in M\} - \min \{z.\text{freq} | z \in M\}}$$

4.2 Product selection

Next, we discuss how to model product selection, recommendation strategy and decision making into our simulations. The product selection stage allows a user to select, from his awareness pool, the one(s) that he would like to purchase. Due to human cognitive capacity, a user cannot be aware of all of the products available, but only a portion of them. To our knowledge, however, no empirical study has been conducted to determine the size of the product pool a user may be initially aware of. Therefore, for our experiment, we simply use the average number of movies a user has rated (100 in our dataset) to represent the size.

To determine the awareness pool, recall that there are three types of awareness: hit, niche, and hybrid. In hit awareness, a consumer pays attention to only highly exposed products. Therefore, we select movies to a consumer’s awareness pool based on their exposure defined in the previous section. To be able to model heterogeneity between users, instead of simply selecting the 100 most exposed (popular) movies, we make the process nondeterministic as follows. The probability of a movie $x$ to be included in the set is

$$\text{expo}(x, M)$$

Thus, the more popular a movie is, the higher chance it has to be in a user’s awareness pool.

In niche awareness, a consumer pays attention to only products that match his preference. Therefore, to build a consumer’s awareness pool, we select 100 movies that most match his preference. Like in the hit awareness case, we make the selection process nondeterministic by assigning a selection probability to each product, and the probability depends in linear proportion on the similarity between the product and the consumer.

Hybrid awareness allows consumers to pay attention to both highly exposed products and products that match their preferences. Therefore, $w_p$ and $w_r$ are both nonzero in (1). Recall from Section 3.1 that we allow each consumer to have different $w_p$ and $w_r$. We distinguish two groups of consumers: heavy users and light users. For heavy users, we set $w_p=0.2$ and $w_r=0.8$, and for light users, we set $w_p=0.8$ and $w_r=0.2$. A consumer is a heavy user if he has rated more than the average number of movies (100 in the dataset); otherwise he is a light user. Again, to model heterogeneity between users, we assign a probability to each movie $x$ representing its chance to be included in a user’s awareness pool as follows:

$$\frac{\text{aware}(c, x)}{\sum \text{aware}(c, x)}$$

Thus, the higher the awareness level of a movie to a user, the higher chance it has to be in the user’s awareness pool.

To model product selection from an awareness pool, like Fleder and Hosanagar [5], we use the multinomial logit model to calculate the probability for a consumer $c_i$ to select a product $m_j$ as follows:

$$\text{P}(c_i \text{selects } m_j | c_i \text{ is aware of } m_j) = \frac{e^{u_{ij}}}{\sum_k e^{u_{ik}}}$$

where $u_{ij}$ is a function of $\text{aware}(c_i, m_j)$ such that

$$u_{ij} = \alpha \times \text{aware}(c_i, m_j)$$

We refer to the parameter $\alpha$ as consumer’s sensitivity to awareness. Introducing the parameter $\alpha$ allows us to fine-tune the baseline Gini coefficient of the simulation system. Specifically, without recommendation, the products selected by users represent their purchases. Thus, given
consumer’s sensitivity to awareness $\alpha$ , we can calculate the Gini coefficient of the sales. Conversely, if the Gini coefficient of the sales is known, we can determine the appropriate $\alpha$. In our simulation, we use some empirical studies on the Gini coefficient of online market to derive our $\alpha$. In the literatures, a Gini coefficient of 0.70 has been observed for online clothing retail [7], 0.724 for debut albums in music sales [30], and 0.75 for online book market [1]. We use the average 0.733 as the Gini coefficient for online sales. Based on this, we use hybrid awareness and no recommendation as the baseline to determine $\alpha$. We set $\alpha = 7$ as it results in a Gini coefficient of 0.735 in the baseline that is close to online sales that have been observed in the literature.

### 4.3 Recommendation strategy

In the product selection stage we let each user select one product from his awareness pool to his shopping cart. The recommendation system then recommends one product to him. For content-based filtering, we recommend the movie that best matches a user’s preference. To make the process nondeterministic, we consider a set of candidates, from which the recommended product is selected in a random style. The candidate set is determined as follows. Let $x_m$ be the product that has the highest $\text{sim}(c, x_m)$ to a user $c$. Then the candidate set consists of all products $y$ such that $\text{sim}(c, y) \geq 0.8 \times \text{sim}(c, x_m)$. Among the candidate products, the probability of a product $y$ to be recommended is in linear proportion to $\text{sim}(c, y)$.

Collaborative filtering recommends the most popular products that have been purchased by users of the same interest. A person $c$’s group of the same interest is determined as follows. First, we use Pearson’s correlation to measure similarity between two consumers

$$\text{sim}(c1, c2) = \frac{\sum c1.\text{pref}[i] \times c2.\text{pref}[i] - \frac{\sum c1.\text{pref}[i] \times \sum c2.\text{pref}[i]}{n}}{\sqrt{\sum c1.\text{pref}^2[i] - \left(\frac{\sum c1.\text{pref}[i]}{n}\right)^2 \times \sum c2.\text{pref}^2[i] - \left(\frac{\sum c2.\text{pref}[i]}{n}\right)^2}}$$

Let $c_m$ be the user with the highest similarity to a consumer $c$. Then, $c$’s group of the same interest is $\{c1 \in C | \text{sim}(c, c1) \geq 0.8 \times \text{sim}(c, c_m)\}$ where $C$ is the set of consumers. The most popular item in a user’s group of the same interest in a simulation round is recommended to the user in the next simulation round.

### 4.4 Decision making

In the decision making stage, a user needs to determine, between the product he has selected into the shopping cart and the product recommended by the system, the one to purchase. We model acceptance level $p$ as the probability that a user chooses the recommended product. (Thus, $1 - p$ is the probability for sticking to his initial selection.) Since the focus of our research is not to determine $p$, but rather to understand how $p$ affects sales concentration, we consider three given values $p = 0.2$, 0.5, and 0.8, representing low, medium, and high acceptance levels, respectively. Note that in our model we assume that a consumer must choose between the product he initially selected and the product recommended by the system. The model can be easily extended to allow a consumer to choose both by giving an additional probability to such event.

### 5. Simulation results

There are 3 awareness types, 3 recommendation conditions (2 recommendation strategies and no recommendation), and 3 acceptance levels. Since acceptance level does not apply to the case of “no recommendation”, in total, there are 21 conditions to examine. For each condition, we conducted 500 rounds of the three-stage purchase process, where in each round every consumer (with a total of 943 consumers in the dataset) makes a purchase. In the end we calculated the Gini coefficient of the amount of inequality in the sales distribution of the products. The Gini coefficients of all these 21 conditions are illustrated in Figure 1, 2, and 3, and the exact numbers are listed in Table 2.

Before we compare the numbers in these 21 conditions in depth for the interaction effect, let’s compare the Gini coefficients of the two recommendation systems (content-based and CF) against that of the base line, namely, the no recommendation condition. In the case of hybrid awareness and medium acceptance level, the Gini coefficient of the content-based recommendation is 0.601, which is lower than the base line .733, which is lower than the CF system .849. As predicted in Hypothesis 1, sales concentration is the lowest in the content-based recommendation condition, followed by the no commendation condition, and highest in the CF condition. Such result pattern also prevails in the hit and niche awareness conditions, across all three acceptance levels. In average, the Gini coefficient of no recommendation condition in all cases is .731, which is higher than that in the content-based recommendation condition .622, and lower than that in the collaborative filtering recommendation .824. Hypothesis 1 is strongly supported by the results. The effect of awareness type is also strong. When no recommendation system is used, the Gini coefficient of the hit awareness is .829, higher than that of the hybrid
awareness, .733, followed by that of the niche awareness, .630. However, the effect of awareness seems to be attenuated when recommendation systems are used. In average, the Gini coefficient of the hit awareness is .782, which is higher than that of the hybrid awareness .737, followed by the niche awareness .665. Nonetheless, the effect of awareness type is more conspicuous for collaborative filtering systems (Hit: .90; Hybrid: .845; Niche: .727) than that of the content based systems (Hit: .623; Hybrid: .629; Niche: .614). That is, only the content-based systems attenuate the effect the awareness type. Such results suggest that the content-based systems help consumers of all three awareness types find products suitable to their needs, such that the difference of these three awareness types is overridden. In conclusion, Hypothesis 2 is supported as well.

### Table 2. Gini coefficients of Various Conditions

<table>
<thead>
<tr>
<th></th>
<th>Hit Awareness</th>
<th>Hybrid</th>
<th>Niche Awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td>No recommend.</td>
<td>0.829</td>
<td>0.733</td>
<td>0.630</td>
</tr>
<tr>
<td>High Acceptance Rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Based</td>
<td>0.574</td>
<td>0.598</td>
<td>0.607</td>
</tr>
<tr>
<td>Collaborative Filtering</td>
<td>0.938</td>
<td>0.907</td>
<td>0.800</td>
</tr>
<tr>
<td>Medium Acceptance Rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Based</td>
<td>0.596</td>
<td>0.601</td>
<td>0.612</td>
</tr>
<tr>
<td>Collaborative Filtering</td>
<td>0.902</td>
<td>0.849</td>
<td>0.719</td>
</tr>
<tr>
<td>Low Acceptance Rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Based</td>
<td>0.700</td>
<td>0.688</td>
<td>0.623</td>
</tr>
<tr>
<td>Collaborative Filtering</td>
<td>0.860</td>
<td>0.780</td>
<td>0.661</td>
</tr>
</tbody>
</table>

In order to test the interaction effect proposed in Hypothesis 3, we first look at the numbers in group of medium acceptance to temporarily exclude the effect of acceptance. As illustrated in Figure 1, compared to the results of the no recommendation, the content-based recommender does decrease the Gini coefficients. Interestingly, it decreases the Gini coefficients of the hit awareness and hybrid awareness (from .829 to .596 vs. from .733 to .601), but has less effect for the niche awareness (from .630 to .612). On the contrary, the CF system does increase the Gini coefficients in all three awareness types, but the increase in the hit awareness is not as much as the increase in the other two awareness types (about 14-16% increase in the niche and hybrid awareness vs. 9% increase in the hit awareness). Consistent with the prediction of Hypothesis 3, the effects of the recommendation systems on sales concentration depend on the initial awareness of consumers. Moreover, this interaction effect is even stronger for the high acceptance level group, as illustrated in Figure 2, but was cancelled out in the low acceptance level condition, as demonstrated in Figure 3. That is, Hypothesis 4 is also strongly supported by our simulation results.

In conclusion, all of the four hypotheses proposed in this research are strongly supported by the results of simulation. Recommendation systems can increase or decrease sales concentration, depending on what kind of products that they recommend. When recommenders incline to recommend popular products, such as collaborative filtering systems, sales concentration is increased, whereas if recommenders provide recommendation according to every consumer’s unique preference, and recommend niche products, sales concentration will decrease. That is, the phenomenon of winner-take-all and long tail could exist simultaneously in the e-market, given the service of both types of recommendation systems.

Moreover, the effects of recommendation systems are moderated by consumers’ awareness of products and acceptance of recommenders. If recommenders recommend products distinctive from those already existent in consumers’ awareness, then the effect of recommenders stands out, in terms of either increasing or decreasing sales concentration. On the other hand, if recommenders recommend products similar to those in consumers’ awareness, the effect of recommenders levels out. Recommending hit products to consumers who are aware of hit products is not likely to impose much impact on them; whereas recommending niche products might make some sort of difference, namely, decreasing sales concentration. The same logic applies to recommenders that favor niche products. That is, the extent of recommenders’ increasing or decreasing sales concentration is dependent upon their congruency to consumers’ awareness type. In addition, the effect of recommenders also depends on the acceptance level of consumers. If consumers do not accept their recommendation, recommending hit or niche products does not make as much difference in sales concentration. It is only when consumers accept their recommendation that the effects of recommendation systems described above appear. In a word, consumers’ cognitive states before and after encountering recommendation systems moderate the effects of the systems on increasing or decreasing sales concentration.
6. Discussions, limitations and future research

6.1 Discussions

How recommendation systems affect e-commerce has been an important question. Such question can be further translated into two questions based on the perspectives of individual consumers and the market. From the perspective of individual consumers, the question is how recommendation systems affect their information seeking process and final product selection? From the perspective of the whole market, the question is how recommendation systems influence sales concentration? However, these two questions are closely intertwined because sales concentration is generated from the collective results of the product selection decisions of individual consumers. A comprehensive account of recommendation systems should take care of these two perspectives simultaneously. This research examines the effects of recommendation systems on the macro level, by way of simulating the information process of individual consumers when using recommendation systems.

In particular, this research suggests that the seemingly contradictory phenomenon of winner-take-all and long tail could co-exist in the e-market due to various recommendation strategies adopted in the systems. Therefore, instead of asking how recommendation systems affect sales concentration, the valid question to ask should be how sales concentration is affected by different kind of recommendation systems? Besides recognizing the heterogeneity of recommendation systems, this research also emphasizes the heterogeneity of consumers, especially the way they process the information that they search on their own and enter their awareness vs. the information that is provided by the recommendation systems. The results of this research suggest that these cognitive factors of individual consumers moderate the impacts recommendation systems.

For managerial implications, our results indicate that online retailers that use CF recommendation systems can more optimally utilize their advertising and supply chain expenses by focusing on selected popular products, as sales concentration tends to increase. In contrast, an entirely different strategy would have to be adopted if a content-based recommendation system is used, as the sales concentration tends to decrease. Moreover, to embrace the effect of recommendation, consumers’ awareness type and their acceptance level to recommendations must be considered. Awareness type may be obtained by analyzing consumers’ profiles, purchase history and browsing behaviors in the web site. Acceptance level may be increased by knowing consumers’ needs and increasing their satisfaction to the recommendation system. Still, more research is needed to understand these two moderating factors.

6.2 Limitations and future research

There are three major limitations in this research. First of all, we use MovieLens’ dataset as the base for simulation. Even though the data was obtained from real consumers, we have no idea if this dataset can be
representative of consumers of online movies. Therefore, more datasets need to be investigated for the hypotheses. Moreover, in this research, we use only one product, i.e., movie, which is relatively low-priced, and is consumed with higher frequencies. Therefore, it calls for caution when the results are applied to other kinds of products. Second, we use product exposure and preference as the indicators of consumers’ awareness. However, there are other valid indicators of consumer’s awareness, such as price, brand name and manufacturers, which we did not include in this research. Therefore, a more comprehensive set of indicators of awareness should be used in the future research. Third, we use the method of simulation to assess responses of consumers, which is lack of external validity. Therefore, other kinds of method that collect first-hand data from consumers, such as experiment or survey, should be used to further verify the hypotheses in this research.

Finally, in this research we argued that the effects of recommendation systems on sales concentration are moderated by consumers’ awareness types and their acceptance levels to recommendations. We did not look at the decision making strategies or attribute evaluation process. For example, Aksoy et al. [21 found that a recommender that thinks in a similar fashion as the consumer in terms of attribute weights or decision strategies would be more useful. This suggests that a consumer’s acceptance level to a recommendation system might be affected by the degree of alignment between the thinking process of user and the recommender. In this research, we did not look at the reasons why consumers accept the recommendations. It will be interesting to see how these factors interact with each other and ultimately affect sales concentration.

Acknowledgment We thank the anonymous reviewers for their valuable comments and suggestions.

7. References