Long Tails and Superstars: The Impact of Collaborative Filtering across Online Platforms

By Henry Zhu Tang '14

ABSTRACT

Collaborative filtering (CF) is an ongoing development in the algorithms used for online recommendation systems that have become both a complement to and substitute for traditional search on online marketplaces. Most existing literature on the CF algorithm is understandably from an information systems standpoint, so this paper seeks to look at the economics behind this technological shift. It will examine namely how the distributions of online marketplaces have shifted and whether this shift favors niche products or large brand-name products. Are some marketplaces intrinsically more suited for collaborative filtering? What are the differences between an online mega-retailer such as Amazon and a subscription movie platform such as Netflix? After analyzing the Long Tail and Superstar effects in these marketplaces, the paper discusses broader implications for merchants, consumers, the platforms, and society as a whole.

I. Introduction

Collaborative filtering (CF) as an algorithm has been used to filter through large data sets since the early 1990s, but did not take off until within this past decade, due to the exponential growth of the Internet and data sets becoming increasingly more massive. Used for automated recommendations, CF "exploits similarities between the tastes of different users to recommend or advise against items. It relies on the fact that people's tastes are *not* randomly distributed" ("Strategy for Information Markets"). The algorithm treats two users as similar if they rated the same items similarly; two items are considered similar if they have received similar ratings from users who have rated both. The basic idea is to make a recommendation from the items liked by a user's similar users. As one would expect, collaborative filtering has become a useful tool to target-advertise items to customers (i.e. find items that a consumer is likely to like, yet was unlikely to find on his/her own because of the large selection).

We should note that there are actually two types of collaborative filtering to consider. Active (or explicit) collaborative filtering is essentially "word-of-mouth in the age of the information marketplace" ("Strategy for Information Markets"). Users are asked to actively participate in this kind of filtering by giving ratings or reviews, either with their accounts or anonymously. However, there can be a first-mover bias, in which the first person(s) to rank a product can skew the rest of the raters towards the first review(s). There may also be a "lone-mover" bias, in which a large majority of rankings are at extreme ends of the spectrum – either those very upset with the product or those actually associated with it. Passive collaborative filtering, on the other hand, requires no participation on the part of the consumer. Instead, information is collected on users as they navigate the site, and items are recommended to them by processing this information.

Many websites and applications have adopted some form of collaborative filtering in their recommendation systems, to be discussed throughout the rest of this paper. Section 2 will introduce a basic real-life application of CF, and its comparison with other algorithms. Section 3 will elaborate on how the improvements in recommendations can lead to two opposing effects: the Long Tail (rise of niche markets) and the Superstars ("winner-take-all"). Sections 4, 5, and 6 will then discuss specific applications of CF, compare its impact on two different platforms (Netflix and Amazon), and analyze the underlying significance of the Long Tail and Superstar effects. Finally, Section 7 will wrap up the paper with a summary of the different findings and final thoughts about how different players will be affected by CF in the future.

II. Collaborative Filtering and other algorithms

Before we dive into deeper applications of collaborative filtering in different marketplaces, we will first demonstrate how this new algorithm is a decided improvement upon other algorithms. A team of engineers and researchers at the Palo Alto Research Center designed and implemented a recommender system (codenamed Magitti) that considered a user's contextual data and their tastes in order to recommend certain leisure activities. Magitti first inferred which general category the user is most likely to be interested in (e.g. "eat", "shop", "read"), and then ranked the items in the chosen category by computing each item's utility to the user. In short, Magitti used six different models (including collaborative filtering), and the group then conducted a qualitative evaluation in which users assessed the usefulness and serendipity of recommendations obtained from these models. For their experiment, they recruited 16 random participants and asked them to rate a list of local restaurants they had already visited. The participants were then presented with five different lists of new restaurants (each list sorted in decreasing order of utility, though the participants were not made aware of this), and asked how interested they would be in dining at each restaurant. Each restaurant also included past user ratings and comments, and participants were asked to what degree the existing ratings and comments affected their decision. The results ultimately showed that collaborative filtering dominated in giving the most "useful" recommendations (see Figure 1).

Figure 1. The left table shows the average usefulness score for each algorithm. The right table shows average usefulness scores, excluding already visited locations. CF is clearly dominant in either case. (Ducheneaut et al.)

Algorithms	N ratings	Average	Variance
CF	110	0.75	0.13
PREFS	220	0.54	0.15
DISTANCE	200	0.43	0.15
ALL EQUAL	180	0.56	0.14
CUSTOM WEIGHTS	210	0.56	0.14

Algorithms	N ratings	Average	Variance
CF	46	0.57	0.14
PREFS	154	0.46	0.12
DISTANCE	139	0.37	0.13
ALL EQUAL	112	0.48	0.12
CUSTOM WEIGHTS	135	0.45	0.13

The Palo Alto researchers also saw that the participants found user comments more influential than ratings. Interestingly, when the participants were asked about the serendipity or "novelty factor" of the recommendations, the CF algorithm yielded the smallest number of novel items, suggesting it doe not do such a great job of introducing users to unfamiliar products, as one might expect it should. We will keep this in mind for the rest of the paper.

III. Long Tail and Superstar Effects

With CF playing a larger role in online markets, it is increasingly important to consider both Long Tail and Superstar effects and how CF has changed or shifted the sales distribution.

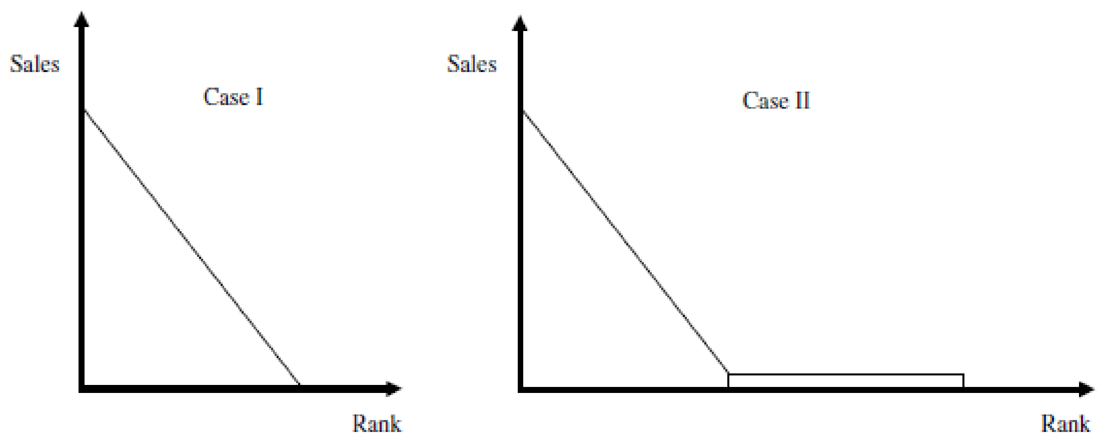


Figure 2. In case I, 100 products are available and the top 50% of products account for 75% of total sales (a variation on the Pareto principle). In case II, we add a "longer tail" of 100 niche products with minimal sales, so now 200 products are available and the top 50% of products account for 95% of total sales. (Brynjoflsson et al.)

Online recommendations can serve as an extension of the feedback effect/loop[17], which can easily breed "superstars" (products that dominate sales within marketplaces). The Long Tail consists of niche items that, though overshadowed by Superstars such as blockbuster movies or best-seller books, have demand from consumer groups outside of the mainstream. It's also likely that the marginal benefit of visible popularity information (the number of page views, visitors, clicks, etc.) is higher for niche products, which can contribute to a more prominent long tail.

With the continued growth of the Internet, it has become easier for retailers to profit by selling items too obscure for brick-and-mortar stores to carry, whose (physical) storage costs are often prohibitive. Jeff Bezos, founder and CEO of Amazon, observed that "there were more than three million book titles in print worldwide, whereas the largest physical

---

[17] Certain recommendations lead to higher sales for the recommended items, which in turn causes them to be recommended more.

superstores carry only about 175,000 titles" (Brynjolfsson et al.). Thus, he chose to focus on books when he first started Amazon, which has become known for supply chain strategies such as drop-shipping[18] that capitalize on demand for the Long Tail, as well as being behind the surging popularity of electronic books. Online retailers can therefore take advantage of the near-zero marginal costs in listing items as part of their inventory/selection, so that supply can and will always meet demand. For Amazon, even delivery fees (for items such as eBooks) can be virtually zero.

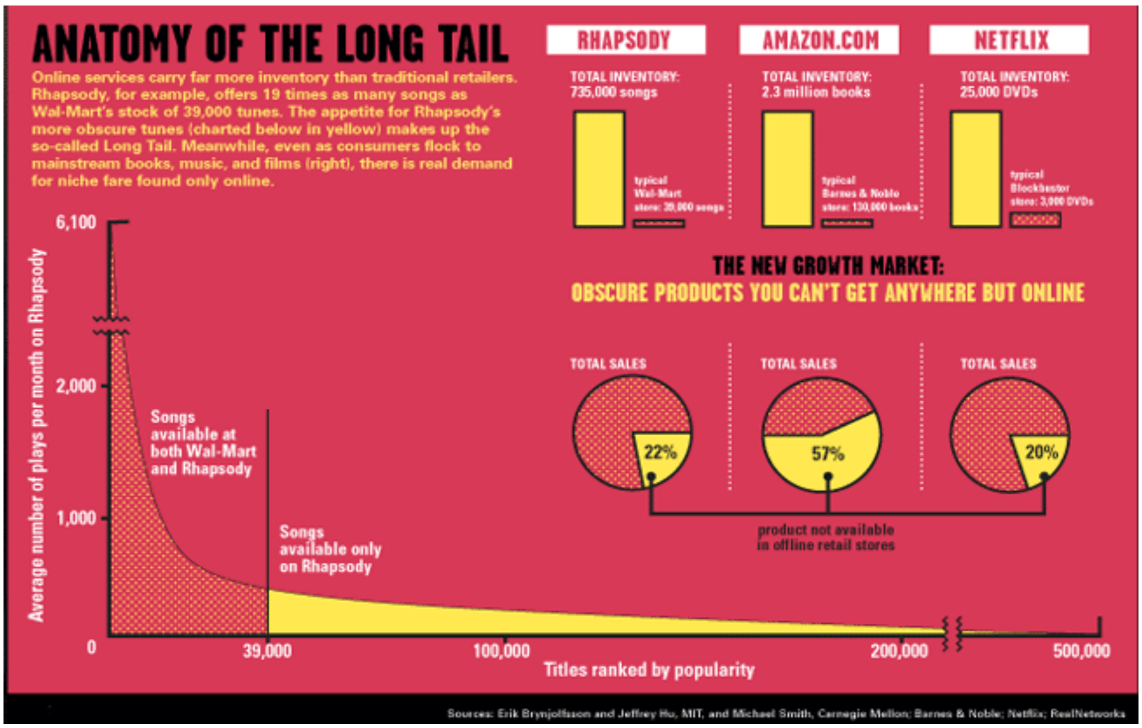


Figure 3. Diagram shows how many items are in the new Long Tail of traditional markets thanks to online alternatives such as Rhapsody, Amazon, and Netflix. (Hoppe)

There are two separate effects that can sustain either the Long Tail or the Superstars. There could be a "birds of a feather" (homophily) effect: As more people find niche items, they are more likely to show up in recommendation lists - other consumers would be drawn to them because they see that consumers similar to themselves have also used the items. Meanwhile, there is still a sense that the "rich get richer", if we assume that megahits continue to dominate search results and consumer demand. As Fleder and Hosanagar discuss, collaborative filters also have difficulty recommending products with

---

[18] Retailer does not keep goods in stock, but instead transfers customer orders and shipment details to a manufacturer or wholesaler. (Wikipedia)

limited historical data, even if they would actually be great matches. This will further reduce sales/product diversity.

Oestreicher-Singer and Sundararajan decide to calculate a Gini coefficient to measure the distributional inequality of revenues across categories, where 0 corresponds to perfect equality (all items in a category have the same revenue) and 1 corresponds to perfect inequality (one item has all the revenue while all other items have zero revenue). They find that based on a comparative analysis across over 200 categories of books on Amazon, more influential/effective recommendation networks are associated with a flatter revenue distribution (i.e. the relative revenue contribution of niche products increases). So, if the revenue fraction of the bottom 20% of products is about 2% for a category with a Gini coefficient of 0.8, it can increase to 4% when the Gini coefficient decreases to 0.6.

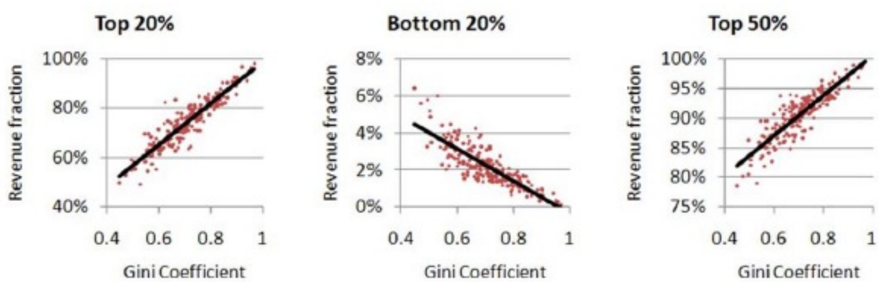


Figure 4. For the top 20% (or 50%) of titles, an increase in the Gini coefficient increases the share of total revenues it has in the market. For the bottom 20% of titles though, increasing the Gini coefficient decreases the shares of revenues that these titles have. (Oestreicher-Singer and Sundararajan)

They also find that categories with higher average demand are apparently less likely to have evenly distributed revenue, though a lot of this could be owed to the fact that the categories have higher demand due to the pull of a select few "mega-hits." The Long Tail and Superstar effects are important to pay attention to not only within the book marketplaces, but across all sorts of arenas that collaborative filtering has expanded to.

IV. Uses and Applications of Collaborative Filtering

Item type ↓	Commercial	Non Commercial
Music	iTunes, Last.fm, GenieLab.com	iRATEradio.com, mystrands.com
Movies	Netflix.com, blockbuster.com	movielens.umn.edu, filmaffinity.com
Books	StoryCode.com	gnooks.com
Dating	reciprodate.com	
Webpages		GiveALink.org, StumbleUpon.com
Aggregated	Amazon.com, half.ebay.com	

Figure 5. Different applications of collaborative filtering (Sahoo et al.)

Collaborative filtering has been applied across music (e.g. iTunes, Pandora, Last.fm, Spotify), movies (Netflix), retail (Amazon, eBay), and social media (Reddit, Quora), just to name a few categories. In fact, it has recently even moved into the food arena. Ness, a startup focused (for now) on matching consumers with restaurants, has been able to differentiate itself from a company like Yelp by using collaborative filtering in its recommendation system. It uses machine learning techniques (including CF) to calculate a Likeness Score for any restaurant based on a user's own ratings and his/her similarity to users who have been to the restaurant, among other factors. Ness will be looking to extend its offering to other lifestyle categories, including music, shopping, nightlife, and entertainment, all within the same app (which may very well challenge and even push out Yelp of the market).

We can argue that "early adopters" of collaborative filtering will have the decided advantage in platform wars; two "early adopter" platforms of particular interest are Amazon and Netflix, which have two of the most well-known and advanced recommendation systems that implement collaborative filtering. The two use a similar item-to-item CF algorithm from a technical standpoint, so we will control for this and concentrate instead on the different distributions of each marketplace because of the product offerings themselves.

V. Comparative Study of Amazon and Netflix

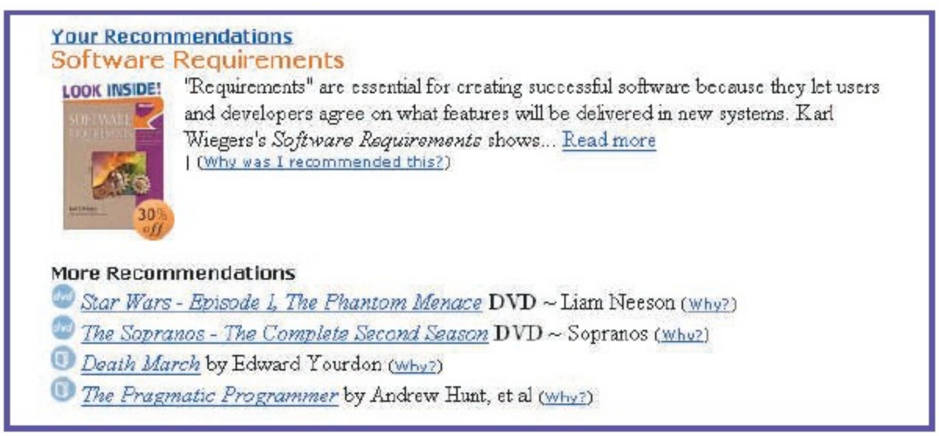
As described in an industry report by Amazon, the company's "algorithm's online computation scales independently of the number of customers and numbers of items in the product catalog" (Linden et al.). Rather than matching a user to similar customers, as traditional CF algorithms do, the item-to-item algorithm matches each of the user's purchased or rated items to similar items (building a table of correlated items), then combines those items into a real-time recommendation list ranked in order of best match. Thus, the algorithm (for Amazon or Netflix) is dependent only on how many items/titles in the user's history, a significant runtime improvement over going through a platform's entire library or set of users.

Knowing that their algorithms are technically similar, we consider different incentive structures for each platform that can cause the algorithms to be implemented differently.

Netflix has transitioned to a prepaid subscription service from a pay-per-use pricing model, which has changed both its content acquisition costs and consumer preferences. It has encountered further trouble with licensing, leading to a fairly limited selection and very broad genres for its titles. As a result, matches can often feel more "out there" or nonsensical to the user. This is a problem that a music platform such as iTunes has been able to avoid, due to a much wider library and very specific niches within the music industry nowadays (e.g. independent artists of all sorts have become increasingly mainstream within the populace, due to lower search and promotion costs). Also, rather than having everyone drawn towards the latest blockbusters (which tends to be the trend), Netflix wants to steer users towards older, less known movies that are less expensive in terms of acquisition costs. Its recommendation system is likely geared to favoring niche titles in some way, so that Netflix can induce demand for the older and lesser known movies it's already paid for. Movies are experience goods and cultural products though, so it is hard for less popular titles from a different time period to suddenly catch on or have the same appeal today. Meanwhile, Amazon, as an aggregate retailer, is perfectly happy with having consumers all buy the same item in any category, and when it ranks items in a recommendation list, the consumer will likely only need the "best" recommended item (on Netflix, users likely go through multiple titles in a list, starting with the best).

The user feedback is also different on both platforms. Netflix uses an explicit ratings system, "encouraging [and constantly reminding] its users to rate the movies that they have watched both outside and within Netflix to improve its recommendations for them, so users have direct incentives to provide truthful and complete ratings" (Tan and Netessine). Netflix also uses pure ratings data rather than review data, where giving a rating is much less costly to a user than writing a review. Thus, Amazon reviews (which are seen as completely optional and less intrusive) must be less frequent. However, each review may have much more of a marginal impact on consumers' decisions that a rating (from 1 to 5 stars) would have, and Amazon may be faced with more of the "lone-mover" bias that was previously discussed. Amazon, while it does ask users to review items, relies more on passive CF by collecting information on user activity as soon as they enter the site.

By seeing user purchase and rating activity, it has been able to establish a Long Tail that serves as a core component of its business. Getting back to Oestreicher-Singer and Sundararajan, the two discovered that books on Amazon have a flatter revenue distribution (a decrease in the Gini coefficient) thanks to more influential recommendation networks. Recommendations within categories have a higher impact on flattening revenue (as traffic stays largely within the category) than do recommendations across categories (since they are more likely to terminate at the more popular product choices and not get to the Long Tail)



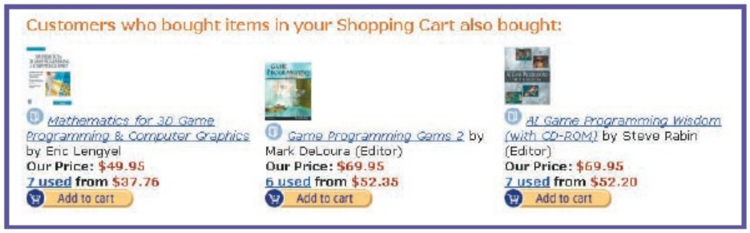


Figure 6. Example of passive collaborative filtering. Amazon collects info on all users as they navigate the site. Upon entering the site, users see a graphic of "recently viewed products" and underneath this, a widget displaying other products the user might be interested in, based on this history. Amazon will also show what "Customers who bought items [you did] also bought", encouraging users to consider purchasing products they otherwise would not have searched. (Linden et al.)

Meanwhile, Tan and Netessine examine the Long Tail effect on Netflix and find that the demand for hits rises while the demand for niches falls, in relative terms. One explanation is that new movie titles appear much faster than consumers discover them: "Consumers over time indeed watch more niche movies in absolute terms, but the rate at which [they] shift demand from the hits to the niches is considerably less than the growth rate of product variety." From the experiment they run, while the average popularity ranking of movies rated by consumers goes down sevenfold (indicating the presence of more obscure movies), after normalizing products ratings by dividing by current product variety[19] and then redefining a measure of popularity, they discover that consumers watch the hits more and more over time. The average consumer watched, on average, movies in the 11th percentile of product variety at the start of Tan and Netessine's study, while the average consumer watched movies in the 5th percentile at the end of the study. This shows that new titles are appearing more quickly than people can (or want to) discover them, so the question is now whether Netflix should continue acquiring or recommending these relatively unknown titles.

It is interesting to remember that Netflix, a platform that wants consumers to embrace niche titles, is having trouble doing so, while Amazon, a platform that would be fine with all of its consumers choosing the "greatest hits" and "best-sellers", succeeds because of its sales diversity. Next we will further discuss the broader implications behind the Long Tail and Superstar effects.

[19] The total number of different movies rated in a time period

VI. Analysis of Findings (Implications)

What do the Long Tail and Superstar effects mean from a macro perspective? There are a number of different externalities that can result, including information externalities (knowing how others acted under similar circumstances saves me the effort of evaluating the options myself), coercive externalities (heightened peer pressure to do as others do), and probably most importantly, market externalities. Market externalities occur when a particular option is chosen by more and more people, and becomes more valuable to all who chose it as a result. This can make it difficult for new collaborative filtering sites to challenge an established site such as Amazon, which already has a critical mass of users, inventory, and preference data. This suggests that current online retailers are safeguarded from the downfall that traditional brick-and-mortar stores fell victim to (notable examples are Blockbuster and Borders, which were slow to adapt in the information age against Netflix and Amazon). Not only do online retailers have more cost-effective operations and supply-chain methods, but they can maintain their market share through the large number of partnerships they already have with third-party merchants/sellers that a "new entrant" platform simply would not have yet.

Even if a Long Tail does continue to develop, there does not seem to be an end to the greatest hits anytime soon. Producers may decide to decrease the number of "superstars" but increase the marketing budget for those which remain, leading to increased sales concentration and "Super Superstars." Tan and Netessine found, using their Netflix data, that the demand for the top 0.1% of movies increases five times as fast the demand for the top 10%, indicating that demand for the "hits of the hits" still is very strong. The fact that consumers gravitate away from obscure titles (which lie in metaphorical dust on the online shelves) to the newest popular releases is further evidence. Thus, the (studios behind the) Tom Cruises and (the publishers for the) James Pattersons of the world can still rest easy.

And once again, we return to the question of inequality. In terms of social welfare implications, consumer surplus gains and producer surplus gains may not be evenly distributed across different types of consumers and firms. To be discussed as a concluding thought, even if everyone is "benefiting", society could still be negatively affected and operating at a suboptimal point.

VII. Concluding Remarks

We have seen that collaborative filtering is indeed a very powerful algorithm that is changing the way we, as consumers, shop, listen to music, watch videos, eat, search for information, etc.

It has the potential to enhance consumer utility, by presenting alternatives that more closely match individual preferences. However, increased choice ("overchoice") may also create an overload, leading to poorer choices. We also must be careful about defining hits and niches now in the Internet era, as opposed to a brick-and-mortar world (where product variety was relatively stable and hits/niches could be defined in absolute terms). In the Internet era, as product variety is skyrocketing, the customer base may be expanding across the distribution, but some products may be left unnoticed. On the merchant/seller side, a downside of increased product variety is that a greater number of products will naturally take longer now to be discovered by consumers after making it onto the marketplace. For future consideration, it would be interesting to forecast the time for products in the Long Tail to accumulate demand (i.e. the average time between entering the market and the first sale, and then between successive sales). It may sooner or later become undesirable for the platforms to list the new titles/items, or for the merchants to sell their niche products on the platforms.

Hotelling's law[20], as well as the fact that consumers find utility in discussing the experience of reading popular books or seeing popular movies with each other, could place a limit or upper bound on the degree of diversity consumers would ever prefer. Platforms have changed their pricing too in response to demand for the big hits. One notable example is Apple, which continues to charge $0.99 for many of its songs, but now sells new hit songs for $1.29 while older, more obscure songs can actually be found for $0.69.

---

[20] Essentially, minimum product differentiation is good, in the perspective of producers.

An online platform such as Netflix may also be inhibited, not by an ineffective algorithm, but by the intrinsic nature of its product. Netflix actually carried out a $1 million contest to see if any team of engineers could design a collaborative filtering algorithm more accurate than the one it currently was implementing. While there was a winner, Netflix eventually decided to pass on it, indicating the added costs of implementation outweighed any incremental revenues. We could infer, therefore, that the algorithm was not the root of the problem, but rather the movie industry itself and consumers' fixation on the newest, trending blockbusters.

Marketing and news have also become more targeted, thanks to CF. However, we are led to wonder whether selective consumption of products and information would lead to either a "global village of well-informed citizens, or to fractured communications between balkanized groups of consumers" (Brynjolfsson et al.), especially if the Long Tail effect increases. As it becomes easier for consumers to filter content based on their own tastes, interests, viewpoints, etc., it is very possible that citizens who are fed an exclusive diet of news catered to them would be less able to understand and debate with individuals holding alternative viewpoints. This is a strange outcome in which individuals could all report higher satisfaction, yet most would agree that there is a negative impact on society as a whole (if ignorance ≠ bliss).

Lastly, it may be time to distinguish between a "greatest hits" and a "winner-take-all" effect. The "superstars of the superstars" could become even more popular, but it is hard to imagine that a true "winner-take-all" mentality can still apply when barriers to entry online are increasingly difficult to establish/hold (for merchants, anyway. It remains to be seen whether mega platforms such as Amazon or eBay can remain the go-to marketplaces for popular and/or niche products).

Works Cited

Brynjoflsson, Erik, et al. “Long Tails vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns.” *Information Systems Research* 21.4 (2010): 736-747. Web. 1 March 2013.

Ducheneaut, Nicolas, et al. “Collaborative Filtering Is Not Enough? Experiments with a Mixed-Model Recommender for Leisure Activities.” *Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization (UMAP)*. June 22-26, Trento. Rome: Springer Berlin Heidelberg, 2009. 295-306.

Fleder, Daniel, and Kartik Hosanagar. “Blockbuster Culture’s Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity.” Working Paper, University of Pennsylvania, Philadelphia, PA.

Hoppe, Bruce. “The Long Tail.” *Webwhompers*. Webwhompers, 2009. Web. 10 March 2013.

Linden, Greg, et al. “Amazon.com Recommendations.” *IEEE Internet Computing* 7.1 (2003): 76-79. Web. 18 Feb 2013.

“Ness Launches Personal Search Engine, Starting with Restaurants.” Ness Computing Press Release, 25 Aug 2011. Web. 20 Feb 2013.

Oestreicher-Singer, Gal, and Arun Sundararajan. “Recommendation Networks and the Long Tail of Electronic Commerce.” Working Paper, New York University, New York, NY.

Peltier, Stéphanie, and François Moreau. “Internet and the Long Tail vs. Superstar effect debate.” *Applied Economic Letters* 19.8 (2012): 711-715. Web. 10 March 2013.

Sahoo, Nachiketa, et al. "Collaborative Filtering with Multi-component Rating for Recommender Systems." Working Paper, Carnegie Mellon University, Pittsburgh, PA.

“Strategy for Information Markets/Collaborative Filtering.” *Wikibooks*. Wikimedia Foundation, Inc. 4 May 2012. Web. 22 Feb 2013.

Tan, Tom, and Serguei Netessine. “Is Tom Cruise Threatened? Using Netflix Prize Data to Examine the Long Tail of Electronic Commerce.” Working Paper, University of Pennsylvania, Philadelphia, PA.