

# Anatomy of the Long Tail: Ordinary People with Extraordinary Tastes

Sharad Goel<sup>‡</sup>, Andrei Broder<sup>†</sup>, Evgeniy Gabrilovich<sup>†</sup>, Bo Pang<sup>†</sup>

<sup>‡</sup> Yahoo! Research, 111 West 40th Street, New York, NY 10018, USA

<sup>†</sup> Yahoo! Research, 4301 Great America Parkway, Santa Clara, CA 95054, USA

{goel, broder, gabr, bopang}@yahoo-inc.com

## ABSTRACT

The success of “infinite-inventory” retailers such as Amazon.com and Netflix has been ascribed to a “long tail” phenomenon. To wit, while the majority of their inventory is not in high demand, in aggregate these “worst sellers,” unavailable at limited-inventory competitors, generate a significant fraction of total revenue. The long tail phenomenon, however, is in principle consistent with two fundamentally different theories. The first, and more popular hypothesis, is that a majority of consumers consistently follow the crowds and only a minority have any interest in niche content; the second hypothesis is that everyone is a bit eccentric, consuming both popular and specialty products. Based on examining extensive data on user preferences for movies, music, Web search, and Web browsing, we find overwhelming support for the latter theory. However, the observed eccentricity is much less than what is predicted by a fully random model whereby every consumer makes his product choices independently and proportional to product popularity; so consumers do indeed exhibit at least some a priori propensity toward either the popular or the exotic.

Our findings thus suggest an additional factor in the success of infinite-inventory retailers, namely, that tail availability may boost head sales by offering consumers the convenience of “one-stop shopping” for both their mainstream and niche interests. This hypothesis is further supported by our theoretical analysis that presents a simple model in which shared inventory stores, such as Amazon Marketplace, gain a clear advantage by satisfying tail demand, helping to explain the emergence and increasing popularity of such retail arrangements. Hence, we believe that the return-on-investment (ROI) of niche products goes beyond direct revenue, extending to second-order gains associated with increased consumer satisfaction and repeat patronage. More generally, our findings call into question the conventional wisdom that specialty products only appeal to a minority of consumers.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WSDM'10, February 4–6, 2010, New York City, New York, USA.

Copyright 2010 ACM 978-1-60558-889-6/10/02 ...\$10.00.

## Categories and Subject Descriptors

J.4 [Computer Applications]: Social and Behavioral Sciences

## General Terms

Economics, Measurement

## Keywords

Long tail, infinite inventory

## 1. INTRODUCTION

The explosion of electronic commerce has opened the door to so-called “infinite-inventory” retailers, such as Amazon.com, Netflix, and the iTunes Music Store, which offer an order of magnitude more items than their brick-and-mortar counterparts [2]. The resulting *long tail markets* [1, 2] have been found to exhibit two near-universal properties: (1) the vast majority of products are “misses,” appealing to only a relatively small group of people; and (2) these “worst-sellers” in aggregate account for a sizable fraction of total consumption. For example, 30% of Amazon.com’s sales and 25% of Netflix’s sales are for items not available in the largest offline retail stores [2]. Based on these empirical observations, the success of online retailers has been largely attributed to the lucrative, and previously untapped, “tail markets.”

The long tail phenomenon, however, is in principle consistent with two fundamentally different hypotheses. The first, and generally accepted theory, is that a majority of consumers prefer popular offerings while only a minority seek niche content; the second hypothesis is that everyone is a bit eccentric, consuming both popular and specialty products. These two theories, importantly, predict substantively different tradeoffs between inventory size and user satisfaction. In the former case, a small inventory of popular items would satisfy most people nearly all of the time, while in the latter, such an inventory would frustrate most people at least some of the time. Thus, differentiating between the two is key to developing sound business strategies.

To distinguish between these possible alternatives, we examine extensive data on user preferences for movies, music, Web search, and Web browsing. In all of these domains, we find overwhelming evidence that nearly everyone is at least a bit eccentric. Our findings suggest an additional factor for the success of infinite-inventory retailers, that is, tail availability may boost head sales by offering consumers the convenience of “one-stop shopping” for both their mainstream

and niche interests. Hence, even small increases in direct revenue from niche products may be associated with much larger second-order gains due to increased overall consumer satisfaction and resulting repeat patronage. More generally, our work highlights the diversity of individual tastes, and calls into question the conventional view that niche products appeal only to a minority of consumers.

The remainder of our paper is organized as follows. In Section 2 we review related work. Section 3 describes our data and presents the main empirical findings. We propose and analyze a theoretical model of consumer behavior in Section 4, and discuss how small differences in inventories lead to “winner-take-all” effects. We conclude in Section 5 by summarizing and discussing our results.

## 2. RELATED WORK

The “long tail” view was coined and popularized by Chris Anderson [1, 2] to describe consumers’ demand for niche products in an age of infinite-inventory retailers. In particular, Anderson finds that a substantial fraction of revenue is generated from specialty items not available in traditional brick-and-mortar stores, and argues that the “future of business is selling less of more” [2]. The economics of long tail markets have been further analyzed by Brynjolfsson et al. [4, 5, 6], who provide a theoretical framework and empirical detail. They consider drivers that increase the collective share of niche products both on the supply-side (e.g., lower stocking and distribution costs) and the demand-side (e.g., improved recommendation and search tools). On the other hand, Elberse et al. [9, 10] have suggested that tail inventory is overrated. Noting that the number of DVD titles in the top 10% of weekly sales dropped by more than 50% from 2000 to 2005, they conclude the importance of best sellers has been growing, not diminishing, over time. And Tan et al. [17], after adjusting for increasing product variety, likewise find that demand for hits has been rising.

In contrast to past work, which primarily considers the volume of tail sales, we focus on consumer satisfaction and the resulting second-order effects of tail inventory. By focusing on the consumer, we shed light on—and largely refute—the perception that niche content appeals only to a minority of consumers. In part, this misconception may be traced to what Levine describes as the “emergence of a cultural hierarchy” in early twentieth century America that established a stark divide between “lowbrow” and “highbrow” entertainment [12]. Looking primarily at high-status individuals, Peterson et al. suggest a relatively recent “historical shift from highbrow snob to omnivore is taking place” [14, 15]. Although we do not explicitly address the cultural status of consumers’ choices, our results are consistent with this view of “omnivorous” individuals.

Elberse [9], writing in Harvard Business Review, reaches qualitative conclusions similar to some of our observations. Specifically, she posits in part that: (1) “a large number of customers occasionally select obscure offerings;” and (2) “customers with a large capacity for content venture into the tail.” We provide extensive empirical evidence to support and refine these statements, and analyze, both empirically and theoretically, the consequences of these results on business strategies. Elberse further argues that consumers appreciate obscure movies less than popular movies, and thus advises retailers to “resist the temptation to direct customers to the tail.” While we find—in agreement with Elberse—

that popular movies receive the highest user ratings, the opposite appears to be true with music: The highest average ratings on Yahoo! Music are given to the most obscure songs. Furthermore, even in the case of movies, we find that typical users regularly give high marks to tail inventory (cf. Section 3.3.2).

Many authors have examined Web search query distributions. Spink et al. [11, 16] studied query logs of the Excite search engine, and analyzed basic properties of this query stream. Later, Downey et al. [7, 8] juxtaposed rare and common queries with rare and common information goals, and described distinctions in user behavior observed for queries and goals of differing rarity. We believe, however, that the long tail phenomenon previously has not been explicitly addressed in the context of Web search.

## 3. EMPIRICAL ANALYSIS

### 3.1 Data Description

Our empirical results are based on an analysis of user behavior across five large datasets: (1) ratings on the movie rental service Netflix; (2) ratings on the music service provider Yahoo! Music; (3) queries on Yahoo! Search; (4) clicked search results on Yahoo! Search; and (5) Web browsing activity collected by the Nielsen Company. Summary statistics for these datasets are given in Table 1.

We examined nearly 100 million Netflix movie ratings from over 400,000 users, and over 700 million Yahoo! Music ratings from over two million users. Netflix ratings were collected between November 1999 and December 2005, and Yahoo! Music ratings were collected between 2002 and 2006. As we are primarily concerned with user-centric statistics, we excluded users for whom we have limited data. In particular, the Netflix dataset was trimmed to include only users who had rated at least 10 movies, and the Yahoo! Music dataset was trimmed to include only users that rated at least 20 songs.<sup>1</sup> Additionally, the music dataset was comprised only of songs that received at least 20 ratings. In both datasets, users rated items (i.e., movies and songs, respectively) on a five point scale, and the primary incentive for users to rate items was to receive personalized recommendations. In neither case was there a requirement that users have purchased or intend to purchase the items they rate. Although these rating records are distinct from purchase histories, we believe they provide a reasonable indication of user interests.

For Web search related data, we analyzed one month of Yahoo! search logs (September, 2008). Simple transformations (e.g., mild stemming) were applied to collate equivalent queries. Furthermore, URLs for clicked search results were truncated to only include domains; for example, [http://en.wikipedia.org/wiki/Long\\_tail](http://en.wikipedia.org/wiki/Long_tail) was normalized to [en.wikipedia.org](http://en.wikipedia.org). Infrequent users—those who issued fewer than 10 queries, or clicked on fewer than 10 URLs in the month-long dataset—were excluded from our analysis. In total, we considered approximately 2.6 billion queries and 2.5 billion click events across nearly 60 million users.

Analysis of Web browsing behavior was based on complete activity logs for the approximately 100,000 users in the Nielsen MegaPanel for the month of March, 2009. Users in

<sup>1</sup>Trimming retains nearly all users in these datasets.

















