# Long Tail or Superstar?

## An analysis of app adoption on the Android Market

**Nan Zhong**
Computer Science, ETH Zurich
Switzerland
zhongn@ethz.ch

**Florian Michahelles**
Information Management, ETH Zurich
Switzerland
fmichahelles@ethz.ch

### ABSTRACT

Many online markets are found with a long tail in sales distribution. With the analysis of a large data set of transactions in Android Market, this work first brings the examination of long tail to the mobile application market. The results suggest that, rather than being a "Long Tail" market where unpopular niche products aggregately contribute to substantial portion of sales, the Android Market is more a "Superstar" market strongly dominated by popular hit products. Hit apps are also found to have higher user consumption and satisfaction rate. Besides, we investigate the impact of price and finds that some expensive apps constitute unproportional large sales. Our findings reveal possible different market structure of mobile app market and point out challenges to app developers.

### Author Keywords

Mobile application market; long tail; sales distribution.

### ACM Classification Keywords

H.4.m. Information Systems Applications: Miscellaneous

### General Terms

Economics; Measurement.

### INTRODUCTION

A number of digital markets are found to be "Long Tail" markets where the aggregated sales of the huge amount of niches contribute a sizable fraction of the total revenue [2, 5, 4]. Nevertheless, some other markets are found to be "Superstar" markets where the blockbusters strongly dominate the revenue [7, 8].

Mobile app markets can be seen a long-tailed sales distribution. Among the tens of thousands of apps listed in Android Market, blockbusters such as Angry Bird have been downloaded millions of times, while numerous niches have only been downloaded dozens of times. Therefore, the examination of the long tail in mobile market could provide insights to developers in understanding the market structure and evaluating profitability of the long tail.

Despite research based on limited data [11], the general lack of data in sufficient size hinders research in mobile app market. Thus, this research, to our knowledge, is among the first to examine sales distribution of a mobile app market. In particular, we analyze a large data set of transactions in Android Market and examine the long tail in detail. We find evidence indicating that the Android Market whose downloads and sales are largely dominated by hit apps, is more a Superstar market than a Long Tail market. We also show that though most downloads of paid apps are from cheap apps, some expensive apps accounts for unproportional large revenue.

In the next section, we review related work, describe the dataset and methodology of research. Then results are presented and analyzed. Finally, we conclude our findings and summarize implications to business strategies.

### RELATED WORK

The term "Long Tail " was coined by Chris Anderson to describe how aggregated sales of niches products of online retailers can contribute to large portions of sales [2]. For example, 30% of Amazon's sales of books and 20% of Netflix revenue of movies come from titles unavailable in largest offline stores [2].

However, the value of long tail is in dispute in academia. There is evidence from video [7, 6, 12] and music markets [8] that online market sales concentrate further on hit products, therefore retailers should continue emphasizing the hit products.

Regarding the cause to the long tail, researchers have pointed out that 1 Low stocking and distribution costs that enable abundant supply; 2 Easy searching tools and smart recommender systems that allow users to access otherwise unnoticed niche products, are key factors [2, 5, 4, 3]. The mobile app market possess these factors and our work firstly examine the long tail of it.

### DATA

The data of this work has been provided by *42matters AG*[1] which captures installations, updates and removals of apps in real time and shares this information among its users [9]. Its central database receives records of transactions from Appaware clients running in users' Android phones, which is authorized on the terms of use when users install Appaware. A record contains user id, time, type of transaction (install, removal, and update), app name, app price, app rating, and etc. The dataset is part of those records and Table 1 shows some statistics of it. In general, this dataset consists of 208 thousand anonymous users' 84.1 million transactions from

[1] www.appaware.com


*LARGE 3.0*, September 21, 2012, San Francisco, CA, USA.

| | |
|---|---|
| Users | 208,187 |
| Paid Apps | 16,214 |
| Free Apps | 175,087 |
| Transactions | 17,609,041 |
| Paid App Sales | 1,887,175$ |
| Paid App Downloads | 530,168 |
| Free App Downloads | 6,079,398 |

**Table 1. Statistics of dataset.**

| | Paid Apps | Free Apps |
|---|---|---|
| $C$ | 71.5% | 76.0% |
| $U$ | 15.9% | 14.3% |
| $W$ | 12.6% | 9.7% |

**Table 2. Evaluation of dataset representativeness.**

March 2011 to November 2011. This dataset is one of the few sources that are statistically large enough for studies in sales distortion and user consumption patterns in mobile app markets.

To show the representativeness of this dataset, we conduct an evaluation by comparing orderings of downloads in our data with those in Android Market. Intuitively, if an app $x$ has a higher ranking of downloads than $y$ in Android Market, then $x$ should also have more downloads recorded than $y$ in our data. With comparison of all possible combinations of two-app pairs, we could examine how much the dataset accords with Android Market available data.

In detail, for a given app, although the ground truth (precise number of downloads) is inaccessible, the range that how many downloads it has is listed in Android Market. These ranges are given by an ascending sequence of predefined consecutive intervals: $[1,5]$, $[5,10]$, $[10,50]$, $\cdots[10{,}000, 50{,}000]\cdots$. Every app $x$ fits in a range $r(x)$ and all the apps share the same sequence of ranges. We define $r(x) \succ r(y)$ if left bound of $r(x)$ is greater than or equal to right bound of $r(y)$. Let $A$ be the set of all apps in the dataset, and $d(x)$ number of downloads of an app $x \in A$. We calculate:

$$C = |\{(x,y) \mid x,y \in A, r(x) \succ r(y), d(x) \geq d(y)\}|/N$$
$$U = |\{(x,y) \mid x,y \in A, r(x) = r(y)\}|/2N$$
$$W = 1 - U - C$$

where $(x,y)$ is an ordered pair of apps. $N = |A|(|A|-1)/2$ is the total number of possible pairs. $C$ is the percentage of correct pairs, $U$ unclear pairs, i.e. the ones are in the same range, and $W$ wrong pairs. From Table 2 we could see that more than 70% of pairs have correct orderings in both paid and free apps. In short, the dataset preserves the ordering between apps in the Android Market fairly well.

## METHODOLOGY

We first partition the apps into two categories: paid and free, taking account of intrinsic difference between paid apps and free apps. Table 1 already indicates the difference: there are a lot more free apps than paid apps, and free apps also have much larger total number of downloads than paid ones. In addition to that, top downloaded free apps has around 10 times more downloads than top paid apps recored. Therefore it is reasonable to distinguish these two kinds of apps. Regarding the definition of free apps, an app is defined as free if it never charges users for downloading. This is because app prices could change over time, sometimes a paid app is free for promotion for a short period of time.

For paid apps, we unify local payment currency used in different countries by Android Market by converting all payments to US dollar using conversion rate given by Google Currency on May 1, 2012. We believe this has minor impact on the calculation of total sales.

Afterwards, we calculate downloads that contribute to sales of paid apps. Android Market has a return time for paid apps, within which a payment could be refunded if the purchased app is deleted. Since late December 2010, this return time has been set to 15 minutes. Spurious downloads are neglected accordingly.

Finishing these preparations, following [6, 5, 10], we conduct the analysis of sales distribution of paid apps and downloads distribution of free apps.

## RESULTS

### User Consumption

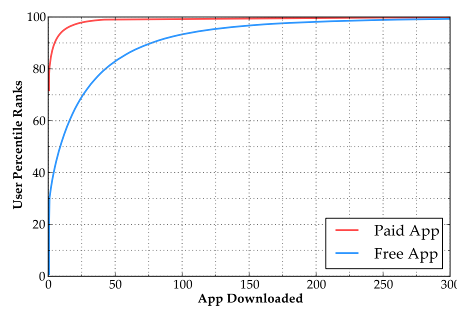



**Figure 1. User percentile of downloads.**

To begin with, Figure 1 depicts user consumption of apps. We find user consumption of paid apps is rather limited. For a given number of downloads $x$ in the horizontal axis, the corresponding $y$ value, i.e. percentile, is the percentage of users downloading less than or equal to $x$ apps. For example, 72% of users have not downloaded any paid apps and only 2% of users have not downloaded any free apps.[2] Most users ($90^{th}$ percentile) download less than 3 paid apps and 75 free apps. It may be caused by the fact that most apps in Android Market are free and as some business observers speculate, users in Android market are less willing to pay than in other mobile markets [1]. This strong distinction between paid and free apps supports our previous partition.

### Long Tail vs. Superstar

Then we examine the sales distortion. In Figure 2 we use the Lorenz Curve and Gini Coefficient to study the concentration of consumption. Apps are ranked according to its popularity ascendingly. For paid apps, popularity is defined as value of sales, and free apps number of downloads. The Lorenz Curve

[2]Due to the design of Appaware, users not downloading any apps cannot be recorded.

depicts the cumulative percentage of popularity of the bottom $x$ percent most popular apps. The Gini Coefficient represents the deviation of the Lorenz Curve to the Line of Equality. A big Gini Coefficient indicates a Superstar market dominated by the hits, and a small Gini Coefficient shows a Long Tail market characterized by the long tail.

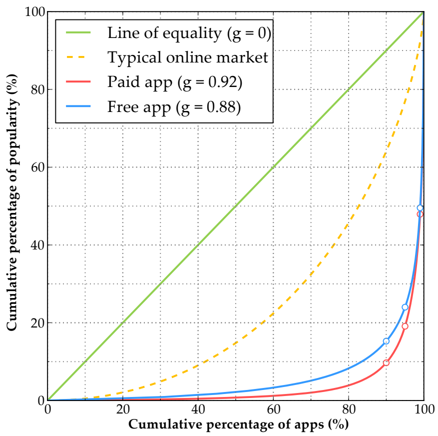
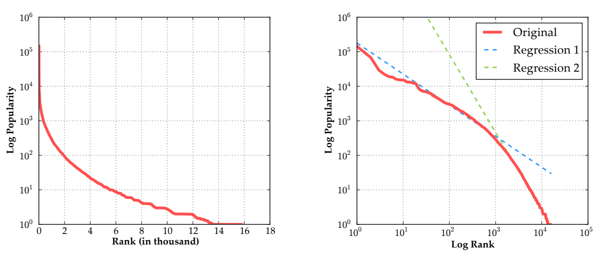



**Figure 2. Lorenz Curve and Gini Coefficient.**

We can see that the hits are dominating. For both sales of paid apps and downloads of free apps, top 1%, 5% and 10% most popular apps make up approximately 50%, 80% and 90% percent cumulative popularity. This dominance of hit products is even stronger than the well known Pareto Principle which claims that 20% most popular products possess 80% of popularity. These curves are also far different from Lorenz Curve of typical online market [5].




(a) Tall head and flat tail. (b) Piecewise power law.

**Figure 3. The long tail of sales and downloads in absolute terms.**

These patterns are depicted in absolute terms in Figure 3, where apps are ranked by popularity descendingly in $x$-axis, and its popularity value is in $y$-axis. Figure 3(a) takes logarithmic scale in $y$-axis, and shows that the popularity decreases sharply as the rank increases. Instead of having a long tail, the Android Market has a tall head and a flat tail. When logarithmic scale in both axises in Figure 3(b), curve does not posses a global power law which was found or assumed in similar studies in other online markets [5, 4]. Nevertheless, a piecewise power law is observed using two linear regressions segmented at $x = 10^3$, which divides the apps into two groups: the top 10% hits and the bottom 90% niches. The coefficient of determination ($R^2$) of Regression 1 and Regression 2 are 0.9997 and 0.9973 respectively implying the good fitting of the regression. The slope of the regression 2 is much deeper than that of regression 1. Namely, when compared to hits, popularity of niches drops much quicker. This suggests that for the niche apps, discovery is still an intractable task, especially in a market where most users download no more than 3 paid apps.

### Natural Monopoly and Double Jeopardy

The dominance of hit apps is further shown in the two phenomena of sales distribution: *natural monopoly* and *double jeopardy* [6, 10].

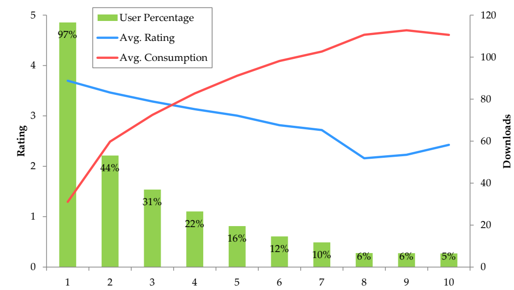



**Figure 4. Natural Monopoly and Double Jeopardy**

Natural monopoly claims that not only does popular products attract disproportionate share of customers, but also these customers purchase more popular products than unpopular ones. We find evidence supporting this theory. In Figure 4 and Figure 5, apps are sectioned into ten deciles where the most popular 10% apps are at left most and least popular 10% right most. The green bars in Figure 4 represent the percentage of users downloading at least one app in this decile[3]. Almost every user download the most popular apps while very few users download the least popular ones. Additionally, the red line shows the average number of apps downloaded by users downloading at least one app of a decile. It tells that, consumers of niche apps download more than those of hit apps. When we drill down these downloads in Figure 5, in which the top 10% apps are tittled as Head and bottom 90% as Tail. Light users in 1stdecile , i.e. those who download most popular apps, have larger portion of apps downloaded from most popular apps.

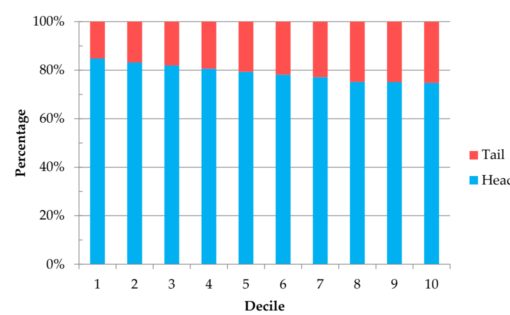



**Figure 5. Distribution of downloads in tail and head.**

Double jeopardy describes that the unpopular products have both less consumers and lower satisfaction rate, therefore in a "double jeopardy". This is shown in Figure 4 by the descending bar chart and blue line, which represent number of consumers and their average rating of apps.

[3]Users not downloading any paid apps are not shown in this chart.

To sum up, the majority of users download hit apps and the few minority users download niche apps; all users consume much more hit apps than niche apps; and hit apps have higher user ratings than niche apps. This accords with the natural monopoly and double jeopardy observations, which clearly demonstrate the superiority of hit apps.

### Price Distribution

At last, we analyze the distribution of sales and downloads of paid apps versus prices. In Figure 6, the height of a bar is the percentage of total apps in a section of prices, and corresponding percentages of total sales/downloads of all apps in this section are represented by the red and blue lines. Most apps are rather cheap, actually the average price of all paid apps is 2.6$. Interestingly, among cheap apps which are below 3$, the usual 1$ apps have less aggregated downloads and sales than apps whos prices are ranging from 1$ to 3$. However, counter intuitively, a few expensive apps acquire unproportional large revenue, whose price are dozens of times higher than cheap apps, thus a few downloads results in big revenue. These apps are usually professional apps, such as navigation, which may have different market position than games and daily apps.

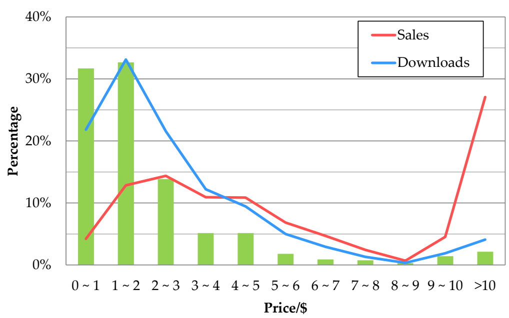



**Figure 6. Distribution of sales and downloads of paid apps.**

## DISCUSSIONS

We found that the Android Market is a Superstar market largely dominated by hit apps. Among the limited number of apps downloaded or purchased by most users, hit apps make up the vast majority and achieve better user rating.

Thus, developers should focus on hit apps to achieve a spot in the relatively small screen of smart phones which physically constraint user choices. Our results also suggest developers to employ more flexible pricing policy. Also, we do not find any pattern of affection of discount promotion in the data.

Our findings suggest that, mobile app market may follow a different market structure than other online markets. First of all, in a highly connected world full of social networks and social apps, mobile market could be influenced by the tyranny of network effect which let users tend to choose the same app. Studies investigating the impact of social features on mobile app market would be beneficial. A second consideration is the diversity of users' tastes. Do users really have diverse needs in choosing most apps? Unlike books or music whose perception is highly subjective, a user's need for an app, e.g. a navigation app, tends to be more objective. However, for different categories of apps, e.g. games, the perception may be subjective as well. Diversity of consumer needs of different categories of apps is another point of research. All these open research problems could help researchers and developers in understanding the underlaying mechanism of mobile app market.

Developers or market operators may have the chance to change the market structure by providing a smart recommender system which better help consumers reach the niches. This has been proven to be beneficial in other online markets [8]. Currently, the recommendation is more based on current popularity of apps which contributes to the dominance of hit apps. How could recommender systems better enable users to explore the growing long tail where thousands of new apps are added to everyday? Is this able to change the market structure?

Finally, we want to mention two limitations of our work. Firstly, our data-set is limited to AppAware users who would yet have to be proven to be representative for the total Android user population. Secondly, we had to neglect the impact of in-app purchase and revenue of add's, which has been important sources of revenue to developers, too, besides the price of the app.

## ACKNOWLEDGMENT

We thank 42matters AG for providing data to foster research.

## REFERENCES


1. Iphone owners more willing to pay for apps than android phone owners, 5 2012. `http://www.cellular-news.com/story/39327.php`.
2. Anderson, C. *The long tail: Why the future of business is selling less of more*. Hyperion, 2006.
3. Brynjolfsson, E., Hu, Y. J., and Simester, D. Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales. *Management Science 57*, 8 (June 2011), 1373–1386.
4. Brynjolfsson, E., Hu, Y. J., and Smith, M. D. From Niches to Riches : The Anatomy of the Long Tail. *MIT Sloan Management Review 47*, 4 (2006), 67–71.
5. Brynjolfsson, E., Smith, M. D., and Hu, Y. J. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science 49*, 11 (2003), 1580–1596.
6. Elberse, A. Should you invest in the long tail? *Harvard Business Review 86*, 07/08 (July 2008), 88–96.
7. Elberse, A., and Oberholzer-Gee, F. Superstars and underdogs: An examination of the long tail phenomenon in video sales. *Marketing Science Institute 4* (2007), 49–72.
8. Fleder, D. M., and Hosanagar, K. Blockbuster Culture' s Next Rise or Fall : The Impact of Recommender Systems on Sales Diversity. *Management Science 55*, 5 (2009), 697–712.
9. Girardello, A., and Michahelles, F. AppAware: Which mobile Applications Are Hot? In *Mobile HCI* (2010), 431–434.
10. Goel, S., Broder, A., Gabrilovich, E., and Pang, B. Anatomy of the Long Tail : Ordinary People with Extraordinary Tastes. In *WSDM* (2010), 201–210.
11. Lee, G., and Raghu, T. S. Product Portfolio and Mobile Apps Success: Evidence from App Store Market. In *AMCIS* (2011), 444–454.
12. Tan, T. F., and Netessine, S. Is Tom Cruise Threatened? An Empirical Study of the Impact of Product Variety on Demand concentration. In *ICIS* (2011), 1–18.