The Lost Tail

The myth of the Long Tail is possibly fading away as the digital book market grows and it's operated by few mega e-retailers

A white paper by Marcello Vena

June 2014

Introduction

The digital revolution of our age, especially in the business of creative goods, has mainly been fuelled by three-four key factors:

i) reduced transaction cost (distribution, logistics, payment, etc.);
ii) reduced search cost for the users (effort required to identify products, obtaining info on prices and key attributes, identify sellers, finalize transaction,..);
iii) reduced cost of producing and reproducing such creative goods (i.e. video, music, books).

There is a fourth one that is becoming very relevant, especially as the digital revolution matures. The world wide web is 21 years old now. As a matter of fact, creative goods are chosen by consumers basing also on subjective taste rather than only on objective quality criteria. This has led to the recent growth of advanced recommendation engines and/or collaborative filtering mechanisms to help consumer find the creative goods they like. (for example YouTube or Amazon's Recommendation engines, or Goodreads collaborative filtering/discoverability tool suite).

Thanks to the digital revolution, underpinned by the above said three-four factors, creative industries are more and more attracting armies of amateurs and outsiders who produce and disseminate their own creative IPs for various reasons and not necessarily only for profits. It's easy to see that all this has fuelled the growth of YouTube in online video, Amazon in books (and many other products – creative and non-creative -too), iTunes or Spotify in music and so on.

This the playground that led to the fascinating myth of the Long Tail. In a limitless world of digital goods, without inventory costs, powerful search engines, near-zero marginal cost of digital production and distribution, the niche products would get much more market relevance at expenses of the blockbusters. The old 80/20 Pareto rule would not apply anymore, and the worst-selling 80% of the creative goods in any market would be going to account for much more than the tiny old 20%.

Obviously the Pareto rule is just an empirical rule (not a theory) to convey the concept of a power-law curve. It just means a limited minority of action/goods/events accounts for the great the majority of results (85/15 or 90/10 carries the same concept with different numbers).

So the key point with the Long Tail is that the large majority of (low-selling) goods is going to massively expands beyond their old 20% (or whatever they had) small pound.

This is a myth (or a theory, if you want) that still fascinates many but it has not scientifically been proved yet. Not at least as the general theory for any digital creative market. The supply of creative content might be unlimited with ridiculously low costs, but content quality and consumer relevance are not constant. Creative goods aren't like industrially manufactured commodities. Furthermore, if content supply exceeds overall content demand there isn't much place for the consumption of additional supply. It's not necessarily a zero-sum game, but it a quite hard wall to break. A Long Tail on the supply-side of the creative market doesn't necessarily imply an equivalent Long Tail on the demand-side.

Digital Music

Let's take for example the digital music market, where more data are available and there is a very interesting and eye opening research led by Prof. Anita Elberse of Harvard Business School. She recently wrote a great book, any executive in the media industry should read: *Blockbusters: Hit Making, Risk-taking, and the big business of entertainment" , Published by Henry Holt and Co. in October 2013*.

The following data are from her latest book based on Nielsen's Soundscan. In 2011, 8 million different digital tracks were sold in the US. 94% sold less than 100 units. 102 tracks (0,001% of all sold tracks) sold over 1 million copies and accounted for 15% of total sales. A total 89,252 tracks (1,1% of all sold tracks) sold at least 1,000 copies and made 86% of all unit sales, see Figure 1. If we really need to believe in rules, then it's a 86/1 "rule", where the winners take all and leave next to nothing to the Long Tail. A very dry one, like a flat lost lake.

It's interesting to see the trend too, as the number of tracks selling next to nothing (just one copy) has been growing throughout the years. In 2007 it was 24%, in 2009 27% and in 2011 32% of the entire sold catalogue. So much for a growing Long Tail. In fact, the number of Long Tail failures has been growing. A Long Tail of songs becomes a Lost Tail on the market.

Digital Books

How about eBooks? It's hard say much as similar statistics aren't available yet. For instance, nobody knows yet how much self-published books really account in terms of catalogue and sales. A new approach is necessary to be able to capture, at least partially, some insights.

In fact, any publisher with a large digital catalogue (1,000+ eBooks) and sizeable sales (1M+ yearly volume) can make the following quick-and-dirty data analysis on its own sales. It can find some basic insights on whether the Long Tail theory applies or not.

It just takes to count what % of titles makes 80% of sales over the year. For example in our case the percentage (%) of titles that accounts for 80% of total sales went down from 23% to 15% from 2011 through 2014 ytd (first 23 weeks), while overall sales grew more than 10x over the years. As the digital market grows, our bestsellers take a bigger share of our market and not a smaller one.

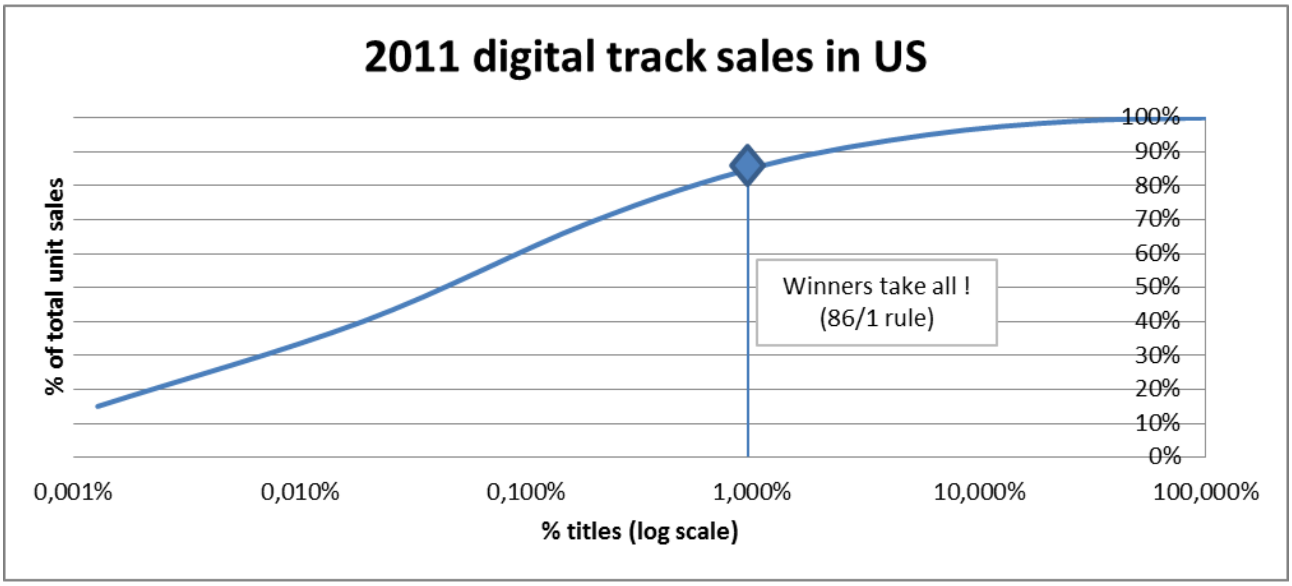

Figure 1 –Cumulative volume sales distribution by % of published digital music tracks (titles) in US.

In other words, the head has been growing faster than the tail, although the number of tail titles grew dramatically over time (3x-4x).

One very interesting aspect to notice is the concentration of the retail market. As the market share of the global players grows, the sales impact of small and independent retailers diminishes dramatically. A less variegated e-retail market can hardly be capable to better promote a huge portfolio of titles, on the contrary: the more the e-retail market is concentrated in few hands, the less a chance for the tail titles to stand out. This might further contribute to dry out the Long Tail and to foster a starving Tail of hopeless titles: *The Lost Tail*.

We've found a strong correlation between the decreasing percentage of titles that make 80% of sales and the increasing concentration of ebook retail market in Italy. Even if correlation doesn't necessarily mean causation, it's a noteworthy finding. More detailed analysis in the future will help to verify causality.

An interesting concentration figure is the Herfindahl-Hirschman Index (HHI), a commonly accepted measure of market concentration. The US DoJ uses HHI for evaluating mergers. The Index is calculated by squaring the market share of each company competing in any market (i.e. ebook retail market) and then adding up the resulting numbers. The HHI number can range from almost 0 to 10,000. The closer a market to a monopoly (or monopsony), the higher the market's concentration. If there were only one company with 100% market share, the HHI would be 10,000. Or if there were N competing firms with same market share, the HHI would be 1/N tending to 0 is N if relatively large. For your reference the US DoJ considers a market with HHI less than 1,500 to be a competitive marketplace. Between 1,500 and 2,500 a moderately concentrated market and above 2,500 a highly concentrated market.

Technically speaking the HHI applies to markets, not to single companies. However nobody forbids any Publisher to use it as a proxy to measure the effective concentration of the e-retail market for its own ebooks. In fact, any title is typically protected by copyright and constitutes a market on itself.

Any Publisher can easily calculate the HHI of its ebook retail market based on its own sales and how they split across the main e-retailers (we call it Pub HHI to distinguish it from the market HHI that includes the titles from the entire market). For example, using the data Hachette recently revealed, its Pub HHI would be between 4,134 and 4,194 in US and between 6,253 and 6,278 in UK (the uncertainties are due the 8% of others sales in US and 5% of other sales in UK, which should be accounted store by store. However in this case they don't noticeably affect the final Pub HHI results).

We have been able to correlate the percentage (%) of our titles that makes 80% of sales in any given year with our Pub HHI for the same year. For 2014 we took the first 23 weeks of year. We weren't interested into the actual hard figures but rather to see if there is a qualitative trend. And yes there it is, as shown in Figure 2.

The trend is quite clear: the higher Pub HHI, the lower the % of titles necessary to make 80% of sales. In other words **a growing retail market concentration is correlated to a growing weight of bestsellers on overall publisher's sales** and therefore to a lower weight of low sellers.

We still are in blockbuster business, even more than ever before. The myth of the Long Tail is possibly fading away due to the lack of factual evidence. Big digital retail ecosystems, who were supposed to enable and benefit from this theory, might ironically be instrumental for its ultimate falsification (in the epistemological sense). *Long* and *Lost* might (partially) be the same Tail, after all.

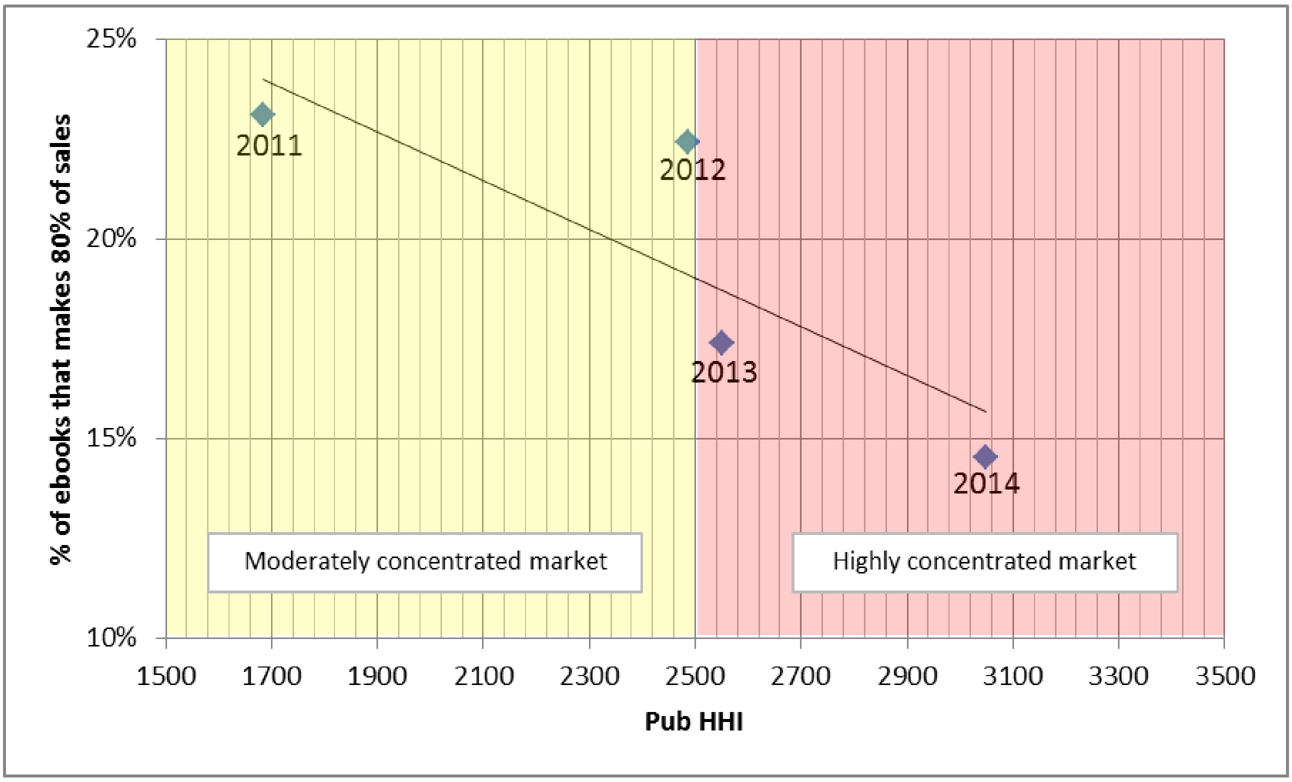

Figure 2 – Head sales % vs e-retail market concentration (Pub HHI)

Last remarks

In this white paper we just wanted to focus on the key ideas and insights. Not so much on technicalities and measuring methods. They are tricky but not very helpful to understand the general trends on digital book sales. Of course actual figures might vary depending on accounting methods, different assumptions, markets, portfolio of own titles and competing titles, publishers' growth rate vs market's growth rate and so on.

Publishers might find interesting to correlate the concentration of their e-retail market with the percentage of titles of own catalogue that generates a certain significant percentage (75%-85%) of their overall sales over the years. All in all, it's *big data* analysis in action with a quick and easy replicable methodology.

It is very well possible that the weight of Long Tail titles grows or erratically moves over the years for a certain publisher, for example if its sales growth underperforms the market growth due to ineffective or insufficient marketing effort on its own top sellers. Do not forget that the weight of Long Tail titles depends also on how good the bestsellers perform. Thus, such analysis might bring additional unexpected market insights.

Finally it's hard to overstate that the findings of this paper are preliminary and might be related to other factors as well. Correlation doesn't mean causation. Stronger evidences could be provided by replicating this analysis at country level. It might be hard to execute it as not all data might be available for analysis.

As Karl Popper once put it: *"Science must begin with myths, and with the criticism of myths"*. There definitely is space to develop new thoughts, challenge without prejudices new and old theories so as to get a better and better understanding of the digital business environment while it continuously changes.

Disclaimer

The opinions expressed in this article are Marcello Vena's own at personal level and might not reflect the view of his employer. Nothing expressed here shall be deemed, either directly or indirectly, as an official statement or endorsement by RCS Libri SPA.

About the author

Marcello Vena has founded and managed the digital trade book business of RCS Libri (fully owned by RCS Mediagroup), one of the largest European book publishers and owner of Rizzoli, Bompiani and Fabbri Editori. He has 17 years of digital innovation experience across three continents (Europe, US & Asia), encompassing the roles of designer, entrepreneur, venture capitalist, management consultant and general manager. Renowned speaker at international events on publishing (New York, London, Berlin, Frankfurt, Munich, Paris, Milan) and member of the Digital Book World 2015 Conference Council. He holds an MBA with Dean's List Distinction from INSEAD and a MS Electronic Engineering. He speaks all five main languages of Western Europe. Information accurate as of June 23rd 2014. www.*about.me/marcellovena*