Examining the power-law distribution among eWOM communities: a characterisation approach of the Long Tail

M. Olmedilla[a], M. R. Martínez-Torres[a] and S. L. Toral[b]

[a]Facultad de Turismo y Finanzas, University of Seville, Seville, Spain; [b]E. S. Ingenieros, University of Seville, Seville, Spain

ABSTRACT

Nowadays electronic word-of-mouth (eWOM) communities symbolise a significant source of information that helps customers to make informed purchasing decisions. Through eWOM communities, a great audience of users is able to acquire knowledge from reviews concerning products and services that are less popular to the majority. The Long Tail effect is a manifestation of such redistribution of demand from popular products to niche products. In this paper, a new methodology that mathematically fits the relationship between the power-law distribution and the Long Tail from an eWOM community is developed. In addition, this paper defines a tool for finding niche products inaccessible through conventional channels. The results are consistent in showing that not all the categories fitting a power-law distribution are characterised by the Long Tail phenomenon, and conversely some of those having a Long Tail do not fit a power-law distribution.

KEYWORDS

eWOM; online user reviews; power-law distribution; Long Tail theory; product categorisation

1. Introduction

Continuous communication among people and ubiquitous online access are fundamental characteristics of online electronic word-of-mouth (eWOM) communities that are facilitating the distribution of a broad range of products and services. eWOM communities have emerged to influence customers directly and create interest with efficacy and flexibility in spite of geographic boundaries (Duan, Gu, and Whinston 2008). They provide rich and objective product information that is influencing customers' decision-making (Gu, Tang, and Whinston 2013; Kim and Gupta 2009; Zacharia, Moukas, and Maes 2000), due to the credibility, empathy and relevance they offer to customers as opposed to the information provided by marketer-designed websites (Bickart and Schindler 2001). Through eWOM, users can freely post their reviews about any product or service, and share those reviews with other users in order to better understand a product (Hennig-Thurau et al. 2004). Similarly, eWOM communities enable the interaction among users, as they can share their experiences and also comment on (or rate) other users' reviews (Arenas Márquez, Martínez-Torres, and Toral 2014). In that regard, eWOM systems provide an aggregated rating for each product given by the reviewers' scores, which shows a quick general impression of the product (Qiu, Pang, and Lim 2012).

Nevertheless, some concerns related to the use of an eWOM community have also been addressed in the literature. For instance, Cheng and Zhou (2010) suggested that 'consumers have to put more attention to evaluating the content of eWOM to determine whether it is credible or not'. Additionally, Shuang (2013) uses a scenario simulation experiment method to show that

CONTACT M. R. Martínez-Torres rmtorres@us.es

eWOM information quality and credibility are positive to information, which in turn depends on factors such as source's popularity or community status.

Although all the previous authors give a good understanding of the main idea behind an eWOM community, it is important to present an appropriate definition for the sake of clarity. Thus, an eWOM community can be described as a community of users who share freely their experiences, thoughts and opinions about a wide variety of products and services independently of any trade or business influence. Such users also receive the feedback of the rest of the community in the form of comments and usefulness score, which is key to rely on their shared opinions.

Many authors have claimed that eWOM communication among customers is influencing customer's purchasing decisions and even changing consumption habits (Hennig-Thurau et al. 2004; Khammash and Griffiths 2011). As a difference to conventional channels, the potential audience of shared opinions within eWOMs is enormous, facilitating the Long Tail effect (Anderson 2008; Arenas Márquez, Martínez-Torres, and Toral 2014; Lee, Lee, and Shin 2011). The Long Tail refers to an increase in the tail of the sales distribution of the less popular products, to wit, the sales of niche products that represent a significant portion of overall product sales (Brynjolfsson, Hu, and Smith 2006; Lee, Lee, and Shin 2011). Those products can collectively comprise a market share that competes or surpasses current bestsellers, but only if the distribution channel is large enough (Odić et al. 2013). Therefore, eWOMs can facilitate the findings of niche products and the Long Tail formation, as the web enables zero-cost dissemination of information about products.

A previous study of the Internet's Long Tail phenomenon made by Brynjolfsson, Hu, and Smith (2003) suggested that a Power Law can be used to describe the relationship between Amazon sales and Amazon sales rank and to estimate which niche books are not typically stocked in brick-and-mortar bookstore. Furthermore, according to a series of breakthroughs in network theory by authors such as Huberman (2003) among others, it is known that power-law distributions tend to arise in social systems where several people express their preferences among numerous choices. Also, some illustrations of such phenomena are the World Wide Web (Adamic and Huberman 2000), Internet TV (Cha et al. 2008), paper citations distribution (Redner 1998) or web pages visibility (Martínez-Torres and Díaz-Fernández 2013). Nevertheless, detection and characterisation of Power Laws remains still complex, especially due to the large fluctuations that arise in the tail of the distribution and by the difficulties of identifying the range over which power-law behaviour is supported (Clauset, Shalizi, and Newman 2009).

Following the previous study by Brynjolfsson, Hu, and Smith (2003) and Brynjolfsson, Hu, and Smith (2010b), who empirically examine the Long Tail within sales distributions in product markets such as Amazon, this paper applies and extends their methodology to mathematically measure the Long Tail in other environments, such as all the categories enclosed in product reviews of the eWOM community Ciao. More specifically, this study proposes an approach for detecting niche products within the Long Tail. The rest of the paper is organised as follows. The next section discusses the background of this research by conducting a review on the online eWOM communities, power-law Distribution and the Long Tail phenomenon. Then, the research framework is presented with several research questions. The design of the research using the power-law distribution approach is explained in the methodology section. The case study and data collection section details the data set gathered from the eWOM portal Ciao UK. Results are next presented in correspondence with each of the proposed research questions. The next section discusses the implications as well as limitations of this study and plans for future research. Finally, the last section concludes the study.

2. Research framework

Prior to Internet era, economic scale gave advantage to products oriented to a big amount of customers – such as books – instead of those aimed at niche markets (Anderson 2004). Nowadays, the inexpensive online medium and the reduced distribution costs have lowered the barriers of entrance (Kumar, Norris, and Sun 2009; Martínez-Torres 2014). Besides and as stated by Anderson (2008):

> increasingly, the mass market is turning into a mass of niche (…) and as the cost of reaching it falls – consumers finding niche products, and niche products finding consumers – it's suddenly becoming a cultural and economic force to be reckoned with.

eWOM has widened customers' choices for assembling impartial product information from other customers, giving them the opportunity to offer their own consumption-related recommendation (Jansen et al. 2009). In that respect, the distribution of product sales is changing due to the increment of product assortment available to consumers (Brynjolfsson, Hu, and Smith 2010a). Actually, the Long Tail phenomenon is a manifestation of such transformations (Anderson 2004).

The concept eWOM is starting to appear in the current literature and is also being assessed in the marketing environment with major repercussions on consumer behaviour (Khammash and Griffiths 2011; Phelps et al. 2004). However, this concept is not new. A good illustration of this was stated by Rogers and Cartano (1962), one of the first authors to foreshadow the so-called opinion leaders, who are defined as those connected to many people and thus influencing others' decisions. That means that people frequently look for an advice, information or an approval in order to strengthen their own opinions. Through classic WOM, customers exchanged their product-related experiences and opinions face to face, which has been proved to play a main role for customers' purchasing decisions (Bickart and Schindler 2001; Richins and Root-Shaffer 1988). Nevertheless, the emergence and impact of user-generated content through the Web has made classic WOM to shift towards eWOM (Hennig-Thurau et al. 2004).

Statements made by customers within eWOM can be exchanged through a variety of online communities such as discussion forums, electronic bulletin board systems, newsgroups, blogs, review sites and social networking sites (Goldsmith and Horowitz 2006). Furthermore, by using search engines, customers can seek out the opinions of strangers. This rarely occurs in classic WOM contexts where opinion providers are embedded in social networks, and well-known people might be more credible (Sun et al. 2006). Another important attribute of an eWOM is its speed, accessibility, one-to-many reach and its lack of face-to-face human pressure (Phelps et al. 2004).

Currently, literature on eWOM tends to deal with two thought streams, eWOM as an aspect of e-commerce (Alanah and Khazanchi 2008; Goldsmith 2008) or as interpersonal influence (Lis 2013; Zhang, Craciun, and Shin 2010). In that regard, Jalilvand, Esfahani, and Samiei (2011) present in their paper an analysis on how the advances of Information Technology and the emergence of online social network sites have changed the traditional limitations of WOM. This paper explains the challenges and opportunities related to the use of eWOM communities and highlights their role in the customer decision-making process. Cheng and Zhou (2010) also discuss in their paper the main eWOM characteristics and how eWOM differs from the traditional WOM. Likewise Brynjolfsson, Hu, and Smith (2006) point out that the eWOM Amazon.com contains a huge amount of product information created directly by customers in the form of reviews, which can drive sales to niche products on the Internet, enabling the Long Tail formation.

An early mention of the Long Tail in the context of the Internet was first outlined by Anderson (2004), who defined this new trend in the field of economics. He argued that niche products would become more prevalent since online retailers sell more products that are less popular than traditional retailers do. His view held that online recommendations networks based on user-generated content such as Amazon guide consumers into the Long Tail of personalisation by covering that percentage of the market of lower-selling niche products.

In the case of online shopping, having product information is essential since consumers purchase products without physical examination. Therefore and to avoid the risk of buying undesirable products, consumers gather all type of information presented on the Internet before making their buying decisions (Lee, Lee, and Shin 2011). In that respect, for online customers, eWOM communities are not only a primary source of product information, but they also assign value to informational content (Dwyer 2007). As consequence, several products that are not in the list of bestselling now become visible as they bring together users of similar interests, similar standards of judgment or geographic location (Elberse 2008). Besides, thanks to lower distribution costs and new ways of connecting demand and supply at a world scale, a shift in demand from the most popular products to niche products might occur facilitating again the Long Tail phenomenon (Peltier and Moreau 2012).

Nevertheless, there are two sets of studies suggesting opposite directions. Proponents of the Long Tail idea maintain that increasing the assortment variety of products offered through online channels will in turn intensify sales of lower-selling niche products (Anderson 2004; Brynjolfsson, Hu, and Simester 2011; Brynjolfsson, Hu, and Smith 2006; Elberse 2008), whereas others defend that these online channels promote the sales of popular products with high ratings (Standifird 2001). Paradoxically, unlike these studies, according to Lee, Lee, and Shin (2011), the roles of eWOM in both cases are the same; it helps consumers to find the product they are looking for. Whether consumers look for a less popular product, eWOM helps them and causes the Long Tail to appear. Otherwise, if customers look for a popular product, eWOM also helps them to find it and in this case the tail of the sales distribution is shortened.

From a mathematical point of view, the Long Tail is a manifestation of power-law relationships (Mahanti et al. 2013). Also, according to Anderson, customer demand across a product space takes the form of a Power Law (Anderson 2008). Such pattern has also been identified within web objects access frequencies (Breslau et al. 1999). In the same way, other widely reported examples are YouTube video popularity (Phillipa et al. 2007) and popularity of TV channels in Internet TV workloads (Cha et al. 2008), where they appear to follow a Zipf-like distribution, which belongs to the family of discrete power-law probability distributions (Jiang et al. 2013). On the Internet, power law seems to be the rule rather than the exception (Li et al. 2012).

Nonetheless, despite a growing research literature about the niche products across the Long Tail in Internet, there is a lack of papers focusing on analytical models within online recommender systems. Hence, the contribution of this paper has been to highlight such method in order to identify niche products within an eWOM community, providing a deeper understanding of the Long Tail phenomena.

It is important to emphasise that all the literature reviewed suggests a relevant association between eWOM communities and the Long Tail phenomenon. Firstly, eWOM communities allow customers to obtain information related to products from a massive, geographically dispersed group of people, who have experience with niche products. Secondly, such exchange of information facilitates a Long Tail effect, as more customers are able to access low-volume products.

Following this conceptual framework, the paper focuses on defining a tool for finding niche products within the Long Tail from eWOM communities. Hence, the outline emerging from this pertains to the following sets of research questions, with the first one primarily aiming to understand the power-law distribution among all the categories enclosed in product reviews:

Q_{1a}. Do all the distribution of categories from the reviews within eWOM fit a Power Law distribution?
Q_{1b}. Are all the categories fitting a Power Law distribution characterised by a Long Tail?

In order to further analyse consumer recommendations from eWOM, the second set of research questions focuses on more specific issues related to the reviews leading to niche products, including:

Q_{2a}. In which cases the formation of the Long Tail occurs among the distribution of all the categories enclosed in product reviews?
Q_{2b}. Are certain types of categories appearing in eWOM more likely associated to the Long Tail phenomenon?

3. Research methodology

Brynjolfsson, Hu, and Smith (2010a) made in their paper a review of the literature about the Long Tail. They found that the Long Tail could be defined and measured in three different ways: (1) Absolute Long Tail (2), Relative Long Tail and (3) exponent based, 'because the ordinal rank to cardinal sales relationship often follows a power-law distribution, the exponent provides an indication of the relative importance of the head versus the tail of the sales distribution'. Particularly, the third one has been taken into account in this paper.

According to this line of thought and to McKelvey and Andriani (2005), power-law distributions appear under two conditions, when tension increases and when the cost of connections decreases. Since in the global economy the costs of connections are quickly decreasing and IT infrastructures dramatically reduce the cost of information transmission, power-law distributions are becoming even more prevalent (Aarstad 2014).

The mathematics of power-law cumulative distributions imply a power-law form that has a probability $P(x)\,\mathrm{d}x$ of taking a value in the interval from x to $x + \mathrm{d}x$, where

$$P(x) = Cx^{-\alpha}, \tag{1}$$

with $\alpha > 0$. Actually, there must be some lowest value $x_{\min}$ at which the power law is obeyed, and the statistics are only considered for x values greater than this value. Also, the equation indicates that the probability of large events is very small and the probability of small events is high. In general, identifying power-law behaviour is difficult due to the large number of fluctuations that occur in the tail of the distribution (Mitzenmacher 2004).

In this paper, the method defined in Clauset, Shalizi, and Newman (2009) will be considered to decide whether or not the data set follows a power-law distribution:

(1) Estimate α using the maximum likelihood estimator for the α scaling exponent
(2) Find $x_{\min}$ using the goodness-of-fit value to estimate where the scaling region begins. The curve can follow a power law on the right or upper tail, so above a given threshold $x_{\min}$.
(3) Calculate the goodness of the model using the goodness-of-fit given by the Kolmogorov–Smirnov statistic equation

$$D = \max|S(x) - P(x)| \qquad x \geq x_{\min} \tag{2}$$

where $S(x)$ is the cumulative distribution function (CDF) of the data to be fitted with $x \geq x_{\min}$, and $P(x)$ is the CDF for the power-law model that best fits the data in the region $x \geq x_{\min}$. The estimation of $x_{\min}$ is actually the value of $x_{\min}$ that minimises the distance D. The distance D is calculated for the observed data set and the best-fit power-law distribution computed as described by the authors. A *p-value* can be calculated to determine whether the value of D is too high. Besides, the *p-value* quantifies if the data sets are consistent with a power-law distribution based on goodness-of-fit. That is to say, if the *p-value* is smaller than 1, the power-law model can be firmly discarded. Nevertheless, if the *p*-value is near 1, then the power law could be a plausible fit to the data.

4. Case study and data collection

To investigate the patterns of power-law distribution within an eWOM community, Ciao was chosen as the object of study. Ciao is a multi-million-strong eWOM community where registered users can critically review and rate products and services for the benefit of other consumers. Membership of Ciao is available free of charge to users in local-language versions. Ciao is an example of a mass eWOM community since there are numerous reviewers and much more readers. It is one of the largest eWOM in Europe, with more than 1.3 million members that have written more than 7 million reviews on 1.4 million of products. It is available free of charge, and registered users can write comments and score products using qualitative ratings that correspond to numerical values. Likewise, other users can score posted reviews.

The Ciao platform is organised through categories of products and services. There are 28 main categories, which in turn are subdivided in many other subcategories. Those 28 main categories are established by Ciao, while the subcategories are created by registered users whenever they post and share reviews.

Data collection within Ciao involves accessing the categories and subcategories in which registered users post their reviews. It is fundamental to gather data added by users in order to build the social network of the Long Tail, connecting those subcategories where a given user has posted his or her reviews. Ciao platform contains the webpage 'My Ciao' for each registered user where some statistics about his or her history past – contributions, review ratings, circle of trust or community score – are displayed. Such member webpage also includes those categories and subcategories where each particular user has posted his or her reviews. The principal limitation when trying to gather data is the absence of index about registered users. Consequently, data collection has been done as a two-stage procedure. First, members' webpages were collected. To that end, a crawler programme that follows the hyperlink structure of the Ciao website has been developed with R. The crawler browses the website, storing those webpages corresponding to members and discarding the rest of them. Second, the list of categories and subcategories for each member stored in the first step was extracted. For this purpose, the function *readLines*() from the R base package, which reads data from a URL, was used in order to access each member webpage. Nevertheless, since webpages are formatted in HTML code, and accessed data contain both the webpage content and the HTML tags, it was necessary to parse HTML file using the function *htmlParse*(), which generates an R structure representing the HTML tree. Finally and once online webpages were available as an R structure, significant data (categories and subcategories for each member) could be easily identified by using regular expressions that are also supported in R, for instance, in packages such as XML (Martínez-Torres 2015).

5. Analysis and results

Data were collected from the website www.ciao.co.uk, which is the Ciao website in UK. Currently, there are about 45,000 registered users in Ciao UK. Nevertheless, the website was only partially crawled by extracting a subset of 4.574 registered users, since extracting all of them would have taken an extraordinary amount of time. However, the sample corresponds to a 10.16% of the total sample, which, according to Pineda, de Alvarado, and de Canales (1994), is enough. A list of categories and subcategories of the posted reviews has been extracted for each registered user using the crawler. The gathered result is shown in Figure 1, which illustrates the distribution of posted reviews over the 28 main categories distinguished by Ciao.

As can be observed, the number of reviews is not evenly distributed over categories. The category Travel is the one gathering almost half of the total number reviews, whereas other categories only have a small number of reviews.

Table 1 describes in detail the power-law adjustment as well as the Long Tail characterisation for each of the 28 main categories contained in Ciao. The first column shows the number of subcategories created by the users who post reviews. The second and the third column are the $\propto$ and the $x_{\min}$ respective values of the fitted power-law distribution. The fourth column corresponds to the goodness–of-fit, given by the distance D and the *p-value*. The last column shows the length of the Long Tail, which has been calculated as the number of subcategories having a number of reviews below the $x_{\min}$, meaning the number of subcategories that are not part of the fitted power-law distribution.

For each category, were the calculated *p-value* is considerably lower than .1, the null hypothesis is rejected, which means that such category does not to follow a power-law distribution. But if the resulting *p-value* is greater than .1, that means that the null hypothesis cannot be rejected and hence the distribution is likely to follow a power-law distribution.

When examining all the categories gathered in Table 1, it is clear that some of them fit the power-law distribution but they do not exhibit a Long Tail. Similarly, there are other categories with highly right-skewed distributions that do not fit a power-law distribution, but they do exhibit a Long Tail. For instance, when observing the categories *Finance* and *Cars & Motorcycles* gathered in Table 1, it is clear that they do not fit a power-law distribution, since their *p-value* rejects the null hypothesis and there

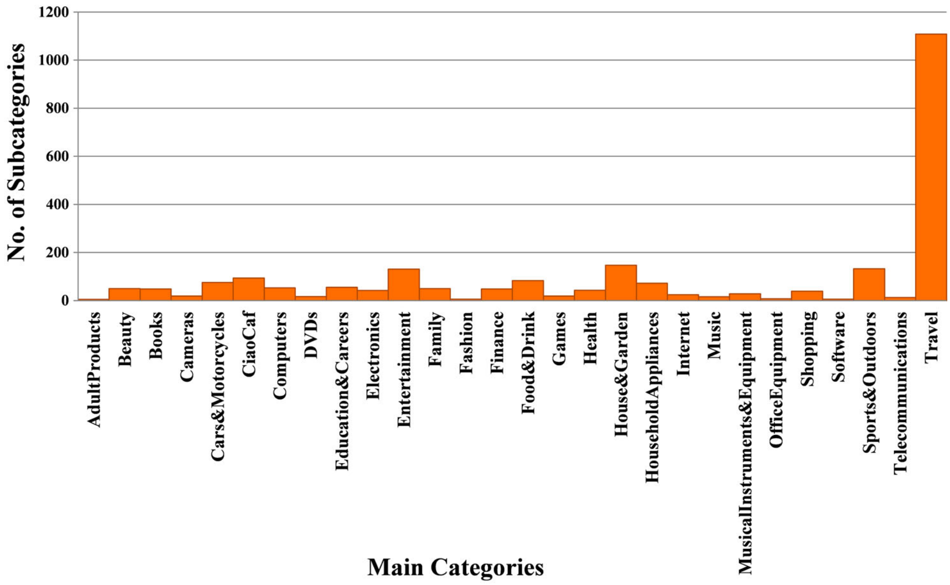

Figure 1. Distribution of posted reviews for the 28 Main Categories.

Table 1. Power law and long tail parameters of the 28 main categories of Ciao.

Category	Sub category	∝	x_{min}	D (p)	Length
Adult products	5	1.50	2	0.262 (.28)	0
Beauty	50	1.71	91	0.115 (.14)	11
Books	48	1.55	31	0.115 (.12)	12
Cameras	19	1.60	6	0.126 (.54)	2
Cars & motorcycles	75	2.13	13	0.125 (.03)	42
CiaoCaf	94	1.73	20	0.089 (.32)	60
Computers	53	2.36	55	0.082 (.96)	32
DVDs	17	3.50	2319	0.200 (.02)	12
Education & careers	55	1.69	2	0.112 (0.23)	21
Electronics	42	2.92	95	0.185 (.25)	33
Entertainment	131	1.97	10	0.055 (.89)	91
Family	50	1.68	37	0.139 (.01)	15
Fashion	6	1.50	10	0.288 (.52)	2
Finance	48	2.44	16	0.1159 (.00)	30
Food & drink	83	2.25	93	0.108 (.16)	39
Games	19	3.50	580	0.241 (.99)	12
Health	43	1.86	34	0.123 (.11)	12
House & garden	147	1.86	14	0.065 (.39)	58
Household appliances	72	3.36	170	0.102 (.89)	61
Internet	24	1.63	22	0.107 (.80)	6
Music	16	1.71	78	0.179 (.43)	5
Musical instruments & equipment	28	2.73	8	0.107 (.75)	19
Office equipment	7	1.54	9	0.194 (.70)	1
Shopping	39	2.12	24	0.147 (.11)	17
Software	6	1.50	6	0.170 (.92)	1
Sports & outdoors	132	2.88	17	0.060 (.86)	102
Telecommunications	13	1.50	4	0.113 (.85)	1
Travel	1108	2.04	6	0.028 (.30)	690

are not enough products highlighted as bestsellers. As revealed by Figures 2 and 3, it is possible to discern the number of products characterised by a Long Tail, which are below the x_{min} demarcated by a horizontal red line.

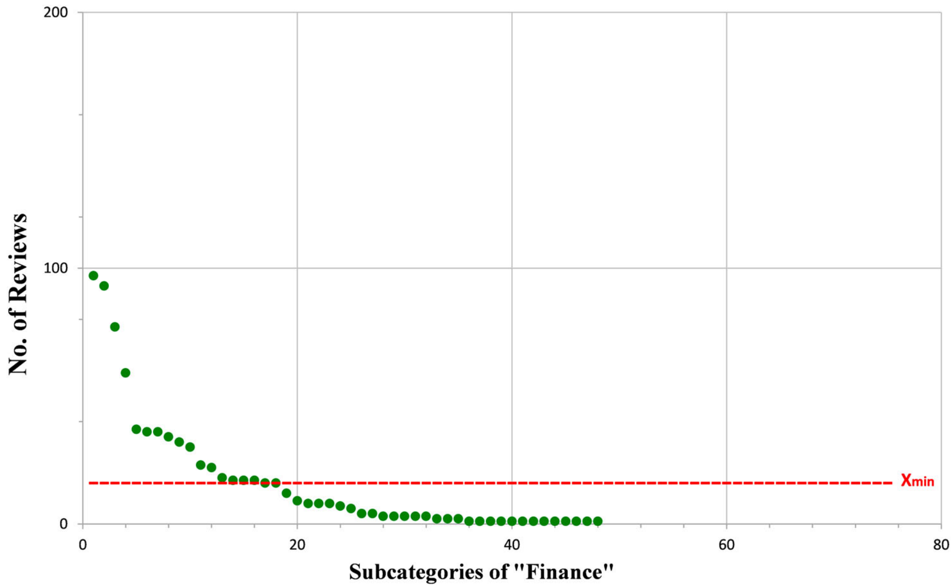

Figure 2. Distribution of reviews for the main category Finance.

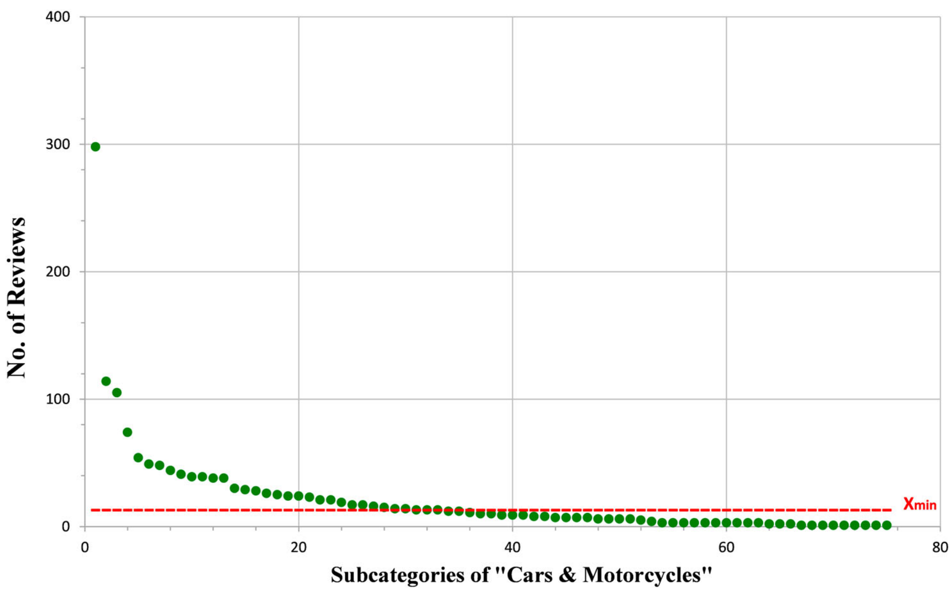

Figure 3. Distribution of reviews for the main category Cars & Motorcycles.

A good example of the Long Tail phenomenon is the category Travel, which is depicted in Figure 4. When observing its power-law adjustment in Table 1, it is obvious that it follows a power-law distribution exhibiting also a Long Tail.

There are some categories that, according to their $\propto$ exponent, are following a power-law distribution with a long peak filled with popular products. However, they are not characterised by a Long Tail, but by a short one, as plotted on Figure 5.

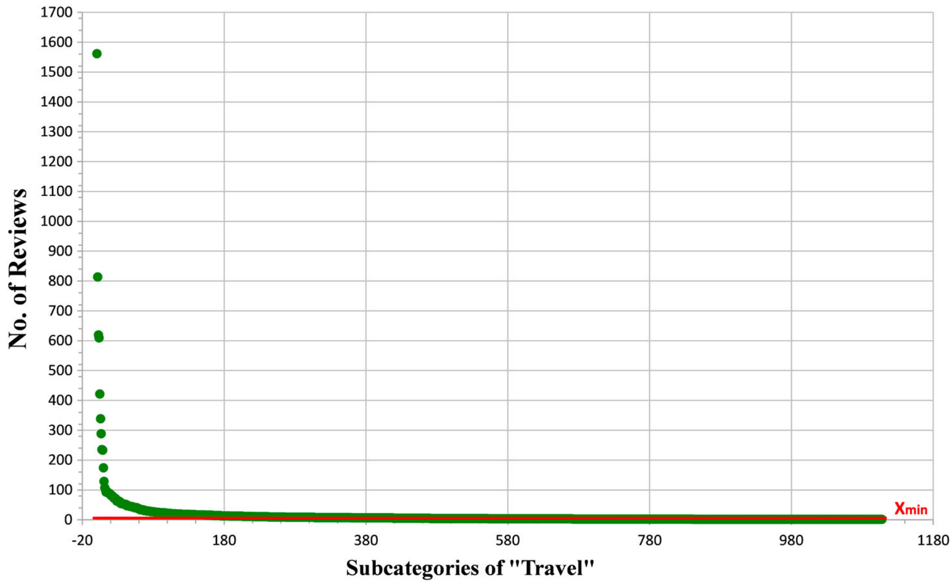

Figure 4. Distribution of reviews for the main category Travel.

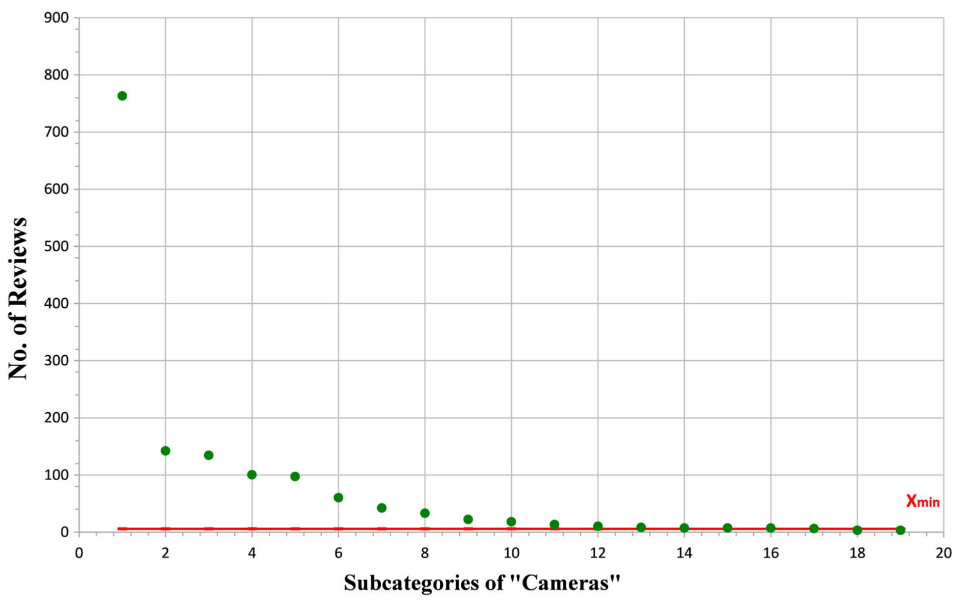

Figure 5. Distribution of reviews for the main category Cameras.

6. Discussion

Anchored in the results gathered form the analysis of the case study, this paper offers several useful insights into the Power Law and Long Tail domain for the eWOM communities. Particularly, while existing literature focuses attention on the niche products across the Long Tail in Internet (Anderson 2004; Brynjolfsson, Hu, and Simester 2011; Brynjolfsson, Hu, and Smith 2006; Elberse 2008; Lee, Lee, and Shin 2011), very little is known in terms of the eWOM communities or online recommender websites.

Specifically and responding to each of the above-described research questions, this paper first answers that not all the distribution of categories gathered from the users' reviews within Ciao UK fit a power-law distribution. Besides, results also confirm that some of the distributions of categories fit a power-law distribution only in the tail, while others concentrate most of the information in the body or head region. In such cases, to reveal the Power Law form of the distribution as well as the Long Tail, it is necessary to plot the chart of the distribution in order to visualise the characteristic form of the power-law distribution.

Furthermore, the findings reveal that categories with subjective and personal evaluation standards are the ones encouraging the Long Tail phenomenon, just as identified by Lee, Lee, and Shin (2011) on their conclusions. Actually, posted reviews in the above-displayed Travel category shows the users' subjective perceptions, desires and preferences about travelling. Otherwise, those categories of products sustained to a more impartial evaluation standards are the high-volume ones joined in the head region, which encourage the superstar phenomenon. Typically, this is the case of those categories where reviews can be done with objective evaluation criteria, such as Electronics, Telecommunications or Cameras.

6.1. *Research contributions*

Theoretical and practical implications of this paper contribute with a new approach on analysing the information gathered from eWOMs. Following from the conceptual framework, several explanations for the study of Long Tail in Internet can be traced to several authors. Brynjolfsson, Hu, and Simester (2011) draw attention to Internet's Long Tail by analysing data collected from a clothing retailer. They discover that consumers' usage of Internet search and discovery tools, for instance recommendation engines, are related with an increase in the share of niche products. Elberse (2008) analyses the Anderson's Long Tail idea and also studies how the tail of the sales distribution is getting longer and fatter by looking at Rhapsody Music or Quickflix among others. Elberse and Oberholzer-Gee (2007) show evidence that Internet retailing has shifted demand towards niche video products over time and argue that exploiting the tail could be unprofitable if many titles do not sell at all. Additionally there is also literature that mathematically examines the Power Low within the Long Tail. For instance, Clauset, Shalizi, and Newman (2009) present a statistical framework for discerning and quantifying power-law behaviour in empirical data occurring in the tail of the distribution where occur large but rare events.

Although those ideas continue to hold sway and this paper absorb all of them, the specific attention resides on defining a new methodology for finding niche products focusing in the Long Tail from eWOM communities. Thus, this paper explores cases characterised by the formation of the Long Tail among all the categories enclosed in product reviews of Ciao by examining the power-law distribution of such data. Likewise, this paper contributes a method for finding niche products within the Long Tail, inaccessible through conventional channels.

In that regard, from the managerial perspective, this paper also presents a challenge to businesses by suggesting that they should embrace the opportunity of finding first-hand information on niche products that exceed the geographic boundaries. What is more, methods such as the one proposed here permit managers to identify and find niche products, as well as to leverage advantages into dominating global blockbusters. Thus, they should look for those markets where those niche products stand to be of great interest for customers outside their immediate social network, since that could have a significant impact on business' sales.

6.2. *Limitations and future research*

One possible methodological limitation in the analysis of this paper would be the sample size. Although the one used is sufficient, taking into account that there are 45 thousand registered users in Ciao UK, a larger sample size might ensure a clearer representative distribution of the gathered data or discover new properties about it. To that end, other type of crawler programme should be developed where data collection would not require crawling the whole website and thus, would take less time.

Further research can extend the findings by characterising the Long Tail for each main category of the posted reviews in an eWOM portal, and defining a tool for extracting and representing all the niche products across the Long Tail. The goal would be to discover some common patterns among niche products through the creation of social network models, where nodes would represent types of products and edges would connect products that have received reviews from the same user. Additionally, it is important to highlight that it would be possible to extend this methodology to measure the Long Tail of other eWOM communities on the Internet.

7. Conclusions

This paper analyses whether there is a Long Tail characterisation in an eWOM community. To that end, the previous sections explore a methodology that allows to find if there is a power-law distribution within the data set gathered from 28 main categories of the portal Ciao UK and to specify in which cases occur the Long Tail. This paper presents two important new findings. First, not all the categories that fit a power-law distribution are characterised by a Long Tail. Otherwise, some of those having a Long Tail do not fit a Power Law. Second, it is also evidenced that most of the cases with subjective evaluation standards are those encouraging the Long Tail phenomenon, while those with more objective evaluation standards encourage the superstar phenomenon. These findings suggest that, in general, eWOM can generate a Long Tail effect, where a large number of small-volume vendors coexist with a few high-volume ones.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Consejería de Economía, Innovación, Ciencia y Empleo under the Research Project with reference P12-SEJ-328, and by the Programa Estatal de Investigación, Desarrollo e Innovación Orientada a los Retos de la Sociedad under the Research Project with reference ECO2013-43856-R.

Notes on contributors

M. Olmedilla is a doctoral student at the Business Administration and Marketing Department, University of Seville. Her research involves the study of eWOM communities and social network analysis.

María del Rocío Martínez-Torres is an Associate Professor in Management and Business Administration at Business Administration and Marketing Department, University of Seville. Her main research interests include *Intellectual Capital and Knowledge Management*, *Social Network Analysis*, *Open Innovation and Virtual Communities*. She has co-authored articles in many leading academic and professional journals, including: *Information and Management*; *IEEE Transactions on Education*; *Computers & Education*; and *Behaviour and Information Technology*.

Sergio Toral is an Associate Professor in Digital Electronic Systems at the Department of Electronic Engineering, University of Seville. His main research interests include Open Source Software projects, Open Innovation and Social Network Analysis.

ORCID

M. Martínez-Torres http://orcid.org/0000-0002-1640-0020
S. L. Toral http://orcid.org/0000-0003-2612-0388

References

Aarstad, J. 2014. "Possible Suboptimal Diffusions of Technological Innovations in Clustered Scale-free Networks." *Technology Analysis & Strategic Management* 26 (3): 267–277.

Adamic, L. A., and B. A. Huberman. 2000. "Power-law Distribution of the World Wide Web." *Science* 287 (5461): 2115–2115.

Alanah, D., and D. Khazanchi. 2008. "An Empirical Study of Online Word of Mouth as a Predictor for Multi-product Category e-Commerce Sales." *Electronic Markets* 18 (2): 130–141.

Anderson, C. 2004. "The Long Tail." *Wired Magazine* 12.10.

Anderson, C. 2008. *Long Tail: Why the Future of Business is Selling Less of More*. New York: Hyperion Books.

Arenas Márquez, F. J., M. R. Martínez-Torres, and S. L. Toral. 2014. "Electronic Word-of-mouth Communities from the Perspective of Social Network Analysis." *Technology Analysis & Strategic Management* 26 (8): 927–942.

Bickart, B., and R. M. Schindler. 2001. "Internet Forums as Influential Sources of Consumer Information." *Journal of Interactive Marketing* 15 (3): 31–40.

Breslau, L., P. Cao, L. Fan, G. Phillips, and S. Shenker. 1999. "Web caching and Zipf-like distributions: Evidence and implications." INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings, 126–134. New York: IEEE.

Brynjolfsson, E., Y. Hu, and D. Simester. 2011. "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales." *Management Science* 57 (8): 1373–1386.

Brynjolfsson, E., Y. Hu, and M. D. Smith. 2003. "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers." *Management Science* 49 (11): 1580–1596.

Brynjolfsson, Erik, Yu Jeffrey Hu, and Michael D. Smith. 2006. "From Niches to Riches: The Anatomy of the Long Tail." *MIT Sloan Management Review* 47: 67–71.

Brynjolfsson, E., Y. Hu, and M. D. Smith. 2010a. "Research Commentary-long Tails vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns." *Information Systems Research* 21 (4): 736–747.

Brynjolfsson, E., Y. J. Hu, and M. D. Smith. 2010b. "The Longer Tail: The Changing Shape of Amazon's Sales Distribution Curve." *Available at SSRN 1679991*.

Cha, M., P. Rodriguez, J. Crowcroft, S. Moon, and X. Amatriain. 2008. "Watching Television over an IP Network." Proceedings of the 8th ACM SIGCOMM conference on Internet measurement, 71–84. New York: ACM.

Cheng, X., and M. Zhou. 2010. "Study on Effect of eWOM: A Literature Review and Suggestions for Future Research." *International Conference on Management and Service Science* 1–4.

Clauset, A., C. R. Shalizi, and M. E. Newman. 2009. "Power-law Distributions in Empirical Data." *SIAM Review* 51 (4): 661–703.

Duan, W., B. Gu, and A. B. Whinston. 2008. "Do Online Reviews Matter? An Empirical Investigation of Panel Data." *Decision Support Systems* 45 (4): 1007–1016.

Dwyer, P. 2007. "Measuring the Value of Electronic Word of Mouth and its Impact in Consumer Communities." *Journal of Interactive Marketing* 21 (2): 63–79.

Elberse, A. 2008. "Should You Invest in the Long Tail?" *Harvard Business Review* 86 (7/8): 88–96.

Elberse, A., and F. Oberholzer-Gee. 2007. "Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales." *Hardvard Business School* (Division of Research, Harvard Business School) Working Paper Series 7–15.

Goldsmith, R. E. 2008. "Electronic Word-of-Mouth." In *Electronic Commerce: Concepts, Methodologies, Tools, and Applications*, edited by A. Becker, 2143–2149. Hershey, PA: Information Science Reference.

Goldsmith, R. E., and D. Horowitz. 2006. "Measuring Motivations for Online Opinion Seeking." *Journal of Interactive Advertising* 6 (2): 2–14.

Gu, B., Q. Tang, and A. B. Whinston. 2013. "The Influence of Online Word-of-mouth on Long Tail Formation." *Decision Support Systems* 56: 474–481.

Hennig-Thurau, T., K. P. Gwinner, G. Walsh, and D. D. Gremler. 2004. "Electronic Word-of-mouth via Consumer-opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?" *Journal of Interactive Marketing* 18 (1): 38–52.

Huberman, B. A. 2003. *The Laws of the Web: Patterns in the Ecology of Information*. Cambridge, MA: MIT Press.

Jalilvand, M. R., S. S. Esfahani, and N. Samiei. 2011. "Electronic Word-of-mouth: Challenges and Opportunities." *Procedia Computer Science* 3: 42–46.

Jansen, B. J., M. Zhang, K. Sobel, and A. Chowdury. 2009. "Twitter power: Tweets as Electronic Word of Mouth." *Journal of the American Society for Information Science and Technology* 60 (11): 2169–2188.

Jiang, Q., C. H. Tan, C. W. Phang, J. Sutanto, and K. K. Wei. 2013. "Understanding Chinese Online users and their Visits to Websites: Application of Zipf's Law." *International Journal of Information Management* 33 (5): 752–763.

Khammash, M., and G. H. Griffiths. 2011. "Arrivederci CIAO.com, Buongiorno Bing. com'—Electronic word-of-mouth (eWOM), Antecedences and Consequences." *International Journal of Information Management* 31 (1): 82–87.

Kim, H. W., and S. Gupta. 2009. "A Comparison of Purchase Decision Calculus between Potential and Repeat Customers of an Online Store." *Decision Support Systems* 47 (4): 477–487.

Kumar, C., J. B. Norris, and Y. Sun. 2009. "Location and Time do Matter: A Long Tail Study of Website Requests." *Decision Support Systems* 47 (4): 500–507.

Lee, J., J.-N. Lee, and H. Shin. 2011. "The Long Tail or the Short Tail: The Category-specific Impact of eWOM on Sales Distribution." *Decision Support Systems* 51 (3): 466–479.

Li, X., Y. Xu, Y. Zhang, and J. Shi. 2012. "Long Tail Distribution in the Web Usage of a Chinese Learning Website." Proc. 2012 International Symposium on Information Science and Engineering (ISISE), 64–67. Shanghai: IEEE.

Lis, B. 2013. "In eWOM We Trust." *Business & Information Systems Engineering* 5 (3): 129–140.

Mahanti, A., N. Carlsson, M. Arlitt, and C. Williamson. 2013. "A Tale of the Tails: Power-laws in Internet Measurements." *IEEE Network* 27 (1): 59–64.

Martínez-Torres, M. R. 2014. "Analysis of Open Innovation Communities from the Perspective of Social Network Analysis." *Technology Analysis & Strategic Management* 26 (4): 435–451.

Martínez-Torres, M. R. 2015. "Content Analysis of Open Innovation Communities using Latent Semantic Indexing." *Technology Analysis & Strategic Management,* 27 (7): 859–875.

Martínez-Torres, M. R., and M. C. Díaz-Fernández. 2013. "A Study of Global and Local Visibility as Web Indicators of Research Production." *Research Evaluation* 22 (3): 157–168.

McKelvey, B., and P. Andriani. 2005. "Why Gaussian Statistics are Mostly Wrong for Strategic Organization." *Strategic Organization* 3 (2): 219–228.

Mitzenmacher, M. 2004. "A Brief History of Generative Models for Power Law and Lognormal Distributions." *Internet Mathematics* 1 (2): 226–251.

Odić, A., M. Tkalčič, J. F. Tasič, and A. Košir. 2013. "Predicting and Detecting the Relevant Contextual Information in a Movie-recommender System." *Interacting with Computers* 25 (1): 74–90.

Peltier, S., and F. Moreau. 2012. "Internet and the 'Long Tail versus Superstar Effect' Debate: Evidence from the French Book Market." *Applied Economics Letters* 19 (8): 711–715.

Phelps, J. E., R. Lewis, L. Mobilio, D. Perry, & N. Raman. 2004. "Viral Marketing or Electronic Word-of-mouth Advertising: Examining Consumer Responses and Motivations to Pass along Email." *Journal of Advertising Research* 44 (4): 333–348.

Phillipa, G., M. Arlitt, Z. Li, and A. Mahanti. 2007. "YouTube Traffic Characterization: A View from the Edge." *Proceedings of the 7th ACM SIGCOMM conference on Internet Measurement.* New York: ACM, 15–28.

Pineda, E. B., E. L. de Alvarado, and F. H. de Canales. 1994. "Metodología de la investigación: Manual para el desarrollo de personal de salud." *OPS*, 1994.

Qiu, L., J. Pang, and K. H. Lim. 2012. "Effects of Conflicting Aggregated Rating on eWOM Review Credibility and Diagnosticity: The Moderating Role of Review Valence." *Decision Support Systems* 54 (1): 631–643.

Redner, S. 1998. "How Popular is Your Paper? An Empirical Study of the Citation Distribution." *The European Physical Journal B-Condensed Matter and Complex Systems* 4 (2): 131–134.

Richins, M. L., and T. Root-Shaffer. 1998. "The Role of Evolvement and Opinion Leadership in Consumer Word-of-mouth: An Implicit Model Made Explicit." *Advances in Consumer Research* 15 (1): 32–36.

Rogers, E. M., and D. G. Cartano. 1962. "Living Research Methods of Measuring Opinion Leadership." *Public Opinion Quarterly* 26 (3): 435–441.

Shuang, Y. 2013. "Effects of Information Quality and Source Credibility on EWOM Adoption in Context of Virtual Community." Proc. International Conference on Management Science and Engineering, 194–200. Harbin: IEEE.

Standifird, S. S. 2001. "Reputation and Ecommerce: eBay Auction and the Asymmetrical Impact of Positive and Negative Ratings." *Journal of Management* 27 (3): 279–295.

Sun, T., S. Youn, G. Wu, and M. Kuntaraporn. 2006. "Online Word-of-Mouth (or Mouse): An Exploration of Its Antecedents and Consequences." *Journal of Computer-Mediated Communication* 11 (4): 1104–1127.

Zacharia, G., A. Moukas, and P. Maes. 2000. "Collaborative Reputation Mechanisms for Electronic Marketplaces." *Decision Support Systems* 29 (4): 371–388.

Zhang, J. Q., G. Craciun, and D. Shin. 2010. "When Does Electronic Word-of-mouth Matter? A Study of Consumer Product Reviews." *Journal of Business Research* 63 (12): 1336–1341.