

Caracterización del fenómeno de «cola larga» en los portales de boca a boca electrónico

María Olmedilla Fernández

Assistant professor de la SKEMA Business School
mariaolmedilla@hotmail.com

Sergio Luis Toral Marín

Catedrático de universidad de la
Escuela Técnica Superior de Ingeniería de la Universidad de Sevilla
storal@us.es

María del Rocío Martínez-Torres

Catedrática de universidad en la
Facultad de Ciencias Económicas y Empresariales de la Universidad de Sevilla
rmtorres@us.es

Extracto

El comercio *online* y los sistemas de recomendación tienen un efecto sobre la demanda de distintos tipos de productos. El objetivo del artículo es probar si internet promueve los productos más populares o *super-hit*, los productos menos populares o nicho, o ambos, así como analizar cuantitativamente la coexistencia de los efectos *super-hit* y de «cola larga». Analizando la curva de distribución de las acciones realizadas por los consumidores en internet sobre 28 categorías de producto, se proponen dos métodos: el método de ajuste de la ley de potencia de la distribución del número de productos (factor de la oferta) por número de comentarios *online* (factor de la demanda) y el método del codo demarcado por el ajuste de la ley de potencia para probar matemáticamente la presencia de ambos fenómenos. Los datos se extrajeron con un *crawler* programado con Python y con la librería de código abierto Scrapy. Los hallazgos revelan que el boca a boca electrónico promueve estos fenómenos según las diferentes categorías de productos, así como su coexistencia. Entre las implicaciones gerenciales destacan las nuevas perspectivas sobre los mercados potenciales que pueden abrirse por la expansión de la cola.

Palabras clave: boca a boca electrónico (*electronic word-of-mouth* [eWOM]); contenido generado por el usuario; «cola larga»; distribución de la ley de potencia; método del codo.

Fecha de entrada: 28-06-2019 / Fecha de revisión: 10-07-2019 / Fecha de aceptación: 15-07-2019

Cómo citar: Olmedilla Fernández, M.^a, Toral Marín, S. L. y Martínez-Torres, M.^a R. (2019). Caracterización del fenómeno de «cola larga» en los portales de boca a boca electrónico. *Tecnología, Ciencia y Educación*, 14, 97-125.



Characterization of the long tail in eWOM websites

María Olmedilla Fernández

Sergio Luis Toral Marín

María del Rocío Martínez-Torres

Abstract

Online commerce and recommender systems influence the demand for different products. The objective of this paper is to prove if the internet promote super-hits products, niche products or both, and to analyse the coexistence of super-hit effect and long tail phenomenon using a quantitative perspective. Through the analysis of the distribution curve of the actions developed by consumers in the internet in 28 different product categories, the power-law distribution of the number of products (supply factor) through the number of online reviews (demand factor), and the elbow criterion over the case of a power law distribution are proposed to mathematically prove the presence of both phenomena. Data were capture with a crawler designed using Python and an open source web crawler framework called Scrapy. Findings reveal that eWOM promotes both the super-hit and the long tail phenomenon depending on the different product categories, and that they can coexist. As a main managerial implication, the study provides new insights about potential markets that can be open as a result of the tail expansion.

Keywords: electronic word-of-mouth (eWOM); user generated content; long tail; power law fitting; elbow criterion.

Citation: Olmedilla Fernández, M.^a, Toral Marín, S. L. y Martínez-Torres, M.^a R. (2019). Characterization of the long tail in eWON websites. *Tecnología, Ciencia y Educación*, 14, 97-125.



Sumario

1. Introducción
 2. Enfoque teórico y marco conceptual
 - 2.1. Efecto *super-hit* versus fenómeno de «cola larga»
 - 2.2. Boca a boca electrónico
 - 2.3. Ajuste de las distribuciones de la ley de potencia *online*
 3. Motivación de la investigación y preguntas de investigación
 4. Metodología de investigación
 - 4.1. Caso de estudio y recolección de datos
 - 4.2. Técnicas cuantitativas
 - 4.2.1. Ajuste de la ley de potencia a los datos recopilados
 - 4.2.2. Encontrar el x_{min} óptimo según el método del codo
 5. Resultados
 6. Discusión e implicaciones
 - 6.1. Discusión
 - 6.2. Implicaciones para la gestión
 - 6.3. Limitaciones de la investigación
 - 6.4. Trabajo futuro
- Referencias bibliográficas

Nota: este trabajo se enmarca dentro de la Convocatoria de Ayudas a Proyectos de I+D+i 2017 de la Fundación Hergar (categoría: Investigación aplicada y tecnológica en Ciencias Sociales, Jurídicas y Humanidades).

1. Introducción

El efecto de internet y el aumento de la disponibilidad de información del consumidor redefinen un cambio en la distribución de las ventas *online* de los productos sobre los que se escriben comentarios, haciendo que productos que no son populares, también llamados «productos nicho», representen una parte importante de las ventas generales en los mercados *online* (Gu, Tang y Whinston, 2013). Dicha transformación en la distribución de las ventas del producto es lo que Anderson (2004) denominó «cola larga», cuyo fenómeno se puede explicar gracias a las diferentes tecnologías de búsqueda o sistemas de recomendación existentes en la actualidad, que conducen a la redistribución de la demanda, tanto de los productos nicho como de los productos más populares o *super-hit*.

La influencia de los sistemas de recomendación parece ser contradictoria al promover ambos tipos de productos (nicho y populares). En este sentido, estudios previos abren una brecha entre dos posturas diferentes. Por una parte, investigaciones a favor de la idea de la «cola larga» sostienen que estimular la oferta de productos accesibles a través de canales *online* fortalecerá las ventas de productos nicho (Elberse, 2008). Sin embargo, otras investigaciones abogan por la venta de productos populares con puntuaciones (*ratings*) altas (Standifird, 2001). En ambos casos, el rol de los canales *online* es el mismo: ayudar a los consumidores a encontrar los productos que están buscando, sin importar su popularidad. La única diferencia es que, en algunos casos, la cola se ensancha, mientras que, en otros, la cola se alarga (Lee, Lee y Shin, 2011).

Este trabajo proporciona una nueva perspectiva sobre la comprensión del efecto *super-hit* y el fenómeno de «cola larga» sobre los patrones de consumo de los usuarios *online* dentro de una comunidad de boca a boca electrónico, estudiando la presencia de los productos a lo largo de todo el espectro de distribución de categorías de productos.

El objetivo y la principal contribución de la investigación es determinar matemáticamente

El objetivo y la principal contribución de la investigación es determinar matemáticamente la presencia de productos *super-hit* y productos nicho, comparando dos metodologías que miden matemáticamente una distribución de «cola ancha» y una distribución de «cola larga»

Las metodologías utilizadas son el método de ajuste de la ley de potencia de la distribución del número de productos (factor de la oferta) por el número de comentarios *online* (factor de la demanda) y el método del codo demarcado por el ajuste de la ley de potencia

la presencia de productos *super-hit* y de productos nicho, comparando dos metodologías que miden matemáticamente una distribución de «cola ancha» y una distribución de «cola larga»: el método de ajuste de la ley de potencia de la distribución del número de productos (factor de la oferta) por el número de comentarios *online* (factor de la demanda), propuesto previamente por Clauset, Shalizi y Newman (2009) y desarrollado por Olmedilla, Martínez-Torres y Toral (2015), y el método del codo demarcado por el ajuste de la ley de potencia.

El resto del artículo está organizado como se describe a continuación. La siguiente sección analiza el trabajo relacionado y el marco conceptual mediante la realización de una revisión de la literatura y el planteamiento de nuestras preguntas de investigación. A continuación, se describe la metodología usada en la investigación, detallando el caso de estudio y la recolección de los datos, así como los dos métodos utilizados para ajustar la distribución de la ley de potencia. Los resultados se presentan posteriormente, para, finalmente, concluir el estudio con la discusión y sus implicaciones.

2. Enfoque teórico y marco conceptual

2.1. Efecto *super-hit* versus fenómeno de «cola larga»

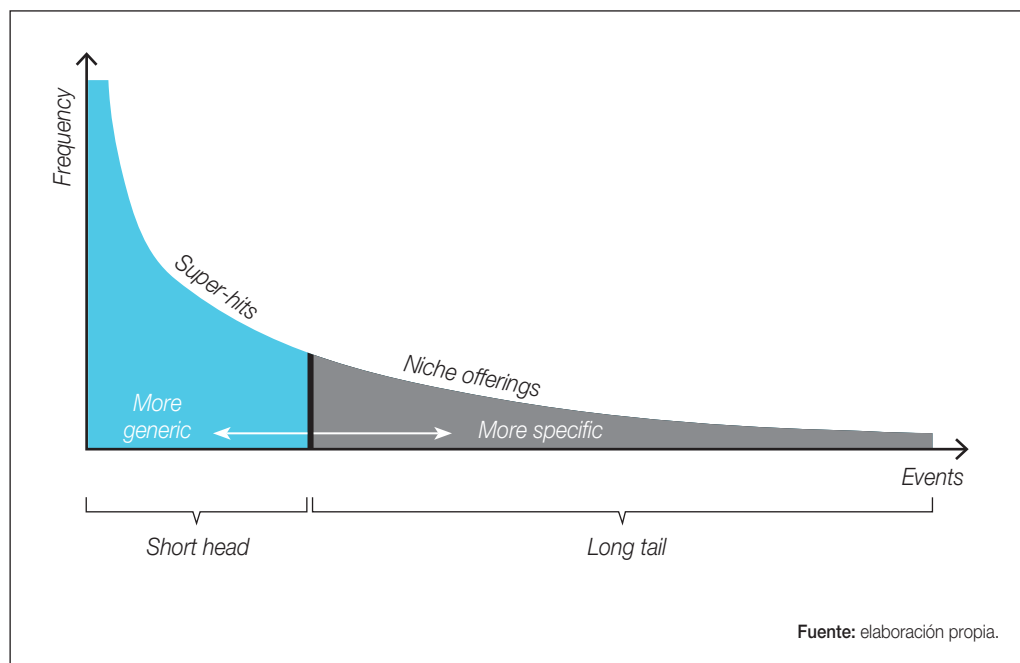
Hoy en día, gracias a internet, a las tecnologías digitales y a las redes, un producto se vuelve más apreciado a medida que lo tienen más usuarios, lo que significa que el efecto *super-hit* está afectando a más productos que antes. No obstante, internet también permite a los consumidores obtener información relacionada con productos menos comunes para un grupo de personas masivo y geográficamente disperso, lo que facilita este efecto de «cola larga» (Olmedilla, Martínez-Torres y Toral, 2015).

En una curva de distribución que sigue una ley de potencia, como la que adquiere la curva de distribución de las ventas o las acciones realizadas por los consumidores en internet, se presenta un grupo pequeño de eventos (productos *super-hit*) que aparecen frecuentemente en la cabeza, seguidos por un gran conjunto de eventos (productos nicho) que aparecen con menos frecuencia en la cola (Koçaş y Akkan, 2016). Gráficamente esta curva de distribución de la ley de potencia representaría todos los eventos probables clasificados por su frecuencia (véase figura 1). Se observa que en ella los productos *super-hit* (a la izquierda de la figura 1, en color azul) son los que presentan mayor frecuencia, dando

Hoy en día, gracias a internet, a las tecnologías digitales y a las redes, un producto se vuelve más apreciado a medida que lo tienen más usuarios, lo que significa que el efecto *super-hit* está afectando a más productos que antes

lugar a la «cabeza» de la curva, mientras que los productos nicho (a la derecha de la figura 1, en gris) son más numerosos y menos frecuentes, constituyendo la cola de la curva. En este ejemplo, el corte en el eje x, marcado en color negro en la figura 1, se ha seleccionado para que las áreas de la cabeza y de la «cola larga» sean iguales, aunque, según Elberse (2008), el área gris de la curva puede llegar a ser más grande que el área azul a lo largo del tiempo.

Figura 1. Ejemplo de gráfico que representa la distribución de productos/eventos por la ley de potencia



La teoría de la «cola larga» defiende que una gran parte de los consumidores prefieren el contenido popular (*super-hit*), mientras que solo una minoría busca contenido de nicho (Goel, Broder, Gabrilovich y Pang, 2010). En este sentido, Anderson (2004) teoriza que la cola al final de la curva de distribución es prácticamente diferente a la cabeza y que los consumidores se están moviendo hacia los productos nicho en la cola, ya que esta satisface mejor los intereses que son menos comunes (Anderson, 2008). Además, en algunas áreas, como la música, las películas o los blogs, la cola es teóricamente ilimitada, ya que internet ofrece un espacio de almacenamiento infinito.

La teoría de la «cola larga» defiende que una gran parte de los consumidores prefieren el contenido popular (*super-hit*), mientras que solo una minoría busca contenido de nicho

2.2. Boca a boca electrónico

Los consumidores comparten cada vez más sus experiencias con un producto o servicio escribiendo comentarios *online*. La aparición y el impacto del contenido generado por los usuarios dentro de la web ha hecho que el boca a boca clásico avance hacia el boca a boca electrónico (Moe y Trusov, 2011), disminuyendo las barreras de entrada en los mercados de los productos (Martínez-Torres, 2014) y teniendo un fuerte impacto en los juicios sobre los productos al ser la información recibida más accesible. El boca a boca electrónico, o boca-oreja, es ampliamente reconocido como una fuente de información influyente en la toma de decisiones de los consumidores (Park, Lee y Han, 2007). Esta influencia depende de la calidad de la información suministrada, de la calidad de la web que sostiene las opiniones, del nivel de satisfacción del usuario con experiencias previas (Filiari, Algezauy y McLeay, 2015) y del nivel de confianza que aporte el sitio web (Casado-Aranda, Dimoka y Sánchez-Fernández, 2019).

Estudios existentes sobre el boca a boca electrónico han lidiado principalmente con información sobre productos como películas (Yeap, Ignatius y Ramayah, 2014), libros (Brynjolfsson, Hu y Smith, 2010; Chevalier y Mayzlin, 2006), viajes (Filiari *et al.*, 2015) o música (Morales-Arroyo y Pandey, 2010). Muchos han evidenciado su capacidad para encontrar productos nicho, facilitando el fenómeno de «cola larga» al disminuir el costo de la búsqueda de información sobre productos. Se ha evidenciado que este fenómeno depende de las características del producto (Manes y Tchetchik, 2018), de la reputación de la web y de la dirección de la información (Park y Lee, 2009).

2.3. Ajuste de las distribuciones de la ley de potencia *online*

No siempre los procesos medidos alcanzan su punto máximo alrededor de un valor típico; a veces divergen en un amplio rango (Newman, 2005). Así, cuando la varianza y/o la media no son finitas, como ocurre en eventos grandes, el teorema del límite central no predice distribuciones gaussianas, sino distribuciones que se asemejan a las leyes de potencia. Además, a diferencia de las distribuciones normales, las observaciones en las distribuciones de la ley de potencia están muy a la izquierda de la media (de ahí que también se llamen «distribuciones de cola pesada») [Bimpikis y Markakis, 2015]) y algunos valores atípicos representan una cantidad desigual de la producción de la distribución total, poniendo de manifiesto que los eventos en la cola de la distribución tienen más probabilidades de ocurrir (Crawford, Aguinis, Lichtenstein,

A diferencia de las distribuciones normales, las observaciones en las distribuciones de la ley de potencia están muy a la izquierda de la media y algunos valores atípicos representan una cantidad desigual de la producción de la distribución total, poniendo de manifiesto que los eventos en la cola de la distribución tienen más probabilidades de ocurrir

Davidsson y McKelvey, 2015). Por tanto, las colas extremas son importantes porque dan una idea de con qué frecuencia, en promedio, pueden ocurrir los eventos más grandes. En consecuencia, la distribución de la ley de potencia se considera una herramienta valiosa para medir estas incertidumbres de la información, ya que las distribuciones de Gauss no se pueden usar cuando ocurren a una cierta probabilidad.

Investigaciones anteriores y actuales han examinado si los fenómenos de internet, provocados por la interacción de los usuarios *online* que intercambian información, están mejor descritos por las distribuciones de la ley de potencia (Barabási y Albert, 1999; Mahanti, Carlsson, Mahanti, Arlitt y Williamson, 2013), demostrando que estas generalmente surgen en los sistemas sociales de internet donde los usuarios expresan sus preferencias entre muchas opciones (por ejemplo, YouTube, comunidades boca a boca electrónicas, Amazon, etc.). Por su parte, Brynjolfsson *et al.* (2003) argumentan que se puede usar una distribución de la ley de potencia para describir la relación entre el rango de ventas de un producto y la cantidad de ventas.

Numerosos estudios analizan el efecto de los comentarios *online* sobre las ventas de productos o las elecciones de los consumidores considerando también las características de los comentarios. Una alta proporción de opiniones *online* implica que hay una elevada participación de los consumidores, por lo que es más probable que un consumidor reciba información sobre un producto, lo que le podría suscitar la intención de comprarlo (Zhang, Zhao, Cheung y Lee, 2014). El volumen de comentarios no significa necesariamente que estos sean positivos. A este respecto, numerosos estudios han argumentado que los comentarios negativos en el boca a boca electrónico tienen un mayor impacto en la intención de compra que los comentarios positivos (Chen, Wang y Xie, 2011), ya que, en general, la información negativa es llamativa e incita a la curiosidad, y los comentarios controvertidos pueden provocar mucha discusión (Shao, 2012). Hay varios artículos relevantes que han verificado un impacto positivo de los comentarios *online* en las ventas teniendo en cuenta el volumen de comentarios, como los de Maslowska, Malthouse y Viswanathan (2017), Park *et al.* (2007), Zhou y Duan (2015), Chevalier y Mayzlin (2006) o Zhang, Ma y Cartwright (2013).

En línea con lo anterior, este trabajo se centra en la distribución de la ley de potencia, porque el efecto *super-hit* y el fenómeno de «cola larga» prevalecen en los patrones de consumo *online*, donde un número relativamente pequeño de productos muy populares representan la mayoría de las ventas (Brynjolfsson *et al.*, 2010) y la cola de la distribución incluye productos nicho (Anderson, 2008). Para este estudio se utiliza el número de comentarios *online* y no su contenido.

Este trabajo se centra en la distribución de la ley de potencia, porque el efecto *super-hit* y el fenómeno de «cola larga» prevalecen en los patrones de consumo *online*, donde un número relativamente pequeño de productos muy populares representan la mayoría de las ventas y la cola de la distribución incluye productos nicho. Para este estudio se utiliza el número de comentarios *online* y no su contenido

3. Motivación de la investigación y preguntas de investigación

Investigaciones anteriores han descubierto el impacto de internet en el proceso de descubrimiento de productos por parte de los consumidores y han encontrado evidencias diferentes con respecto al cambio en la distribución de ventas para los distintos productos. Por una parte, Gu *et al.* (2013) examinan en Amazon el fenómeno de «cola larga» a través del efecto informativo en el contexto del boca a boca electrónico, sugiriendo sus resultados que «el boca a boca electrónico restringe la formación de la "cola larga"» (Gu *et al.*, p. 474). Por otra parte, Chevalier y Mayzlin (2006), utilizando también datos de Amazon, hallan el resultado contrario con relación a las ventas de libros; Zhu y Zhang (2010) defienden que los comentarios *online* influyen en la difusión y adopción de productos que son menos populares; y Hervás-Draney (2009) muestra que el boca a boca beneficia a los productos más populares y a los consumidores convencionales. Por lo tanto, la primera pregunta de investigación que nos planteamos es:

PI₁. ¿Promueve el efecto *super-hit* o el fenómeno de «cola larga» el boca a boca electrónico en todas las categorías de productos?

Asimismo, se han estudiado diferentes evidencias sobre cómo internet está afectando de manera diferente a la distribución de productos en la actualidad, evaluando la relación entre la «cola larga» y las ventas de productos. Hay investigadores que abogan por el boca a boca electrónico como facilitador del fenómeno de «cola larga» para ayudar a encontrar productos nicho (Dellarocas y Narayan, 2007), mientras otros investigadores describen la distribución de productos como muy desigual, con unos pocos valores extremos acumulados en la cabeza (Zhang *et al.*, 2013). Ambos fenómenos son ampliamente reconocidos en el artículo de Elberse y Oberholzer-Gee (2007), en el que se encuentran pruebas tanto del efecto de «cola larga» como del efecto *super-hit* en las ventas de videos domésticos.

Una distribución atractiva para los datos generados por este tipo de procesos acumulativos es la ley de potencia. Sin embargo, pocos trabajos se han centrado en si la distribución de la ley de potencia es un ajuste plausible a un conjunto de datos de concentración de diferentes surtidos de productos. Por lo tanto, sería conveniente poder modelar la distribución de las categorías de productos con la ley de potencia para discernir si hay eventos de alta frecuencia en la cabeza corta o eventos más frecuentes en la cola. Esto nos lleva a plantear la siguiente pregunta de investigación:

PI₂. ¿Es posible que el efecto *super-hit* coexista con el fenómeno de «cola larga»?

4. Metodología de investigación

4.1. Caso de estudio y recolección de datos

La comunidad elegida de boca a boca electrónico para la recopilación y el análisis de datos ha sido Ciao UK, por tratarse de una comunidad de boca a boca electrónico masiva, en la que los usuarios pueden opinar sobre cualquier producto o servicio, y pura, al contrario de Amazon, por ejemplo, que es una página de comercio electrónico que además posee una parte de comunidad boca a boca electrónico. Ciao UK, al igual que otras comunidades del estilo, como TripAdvisor, está formada exclusivamente por una red de usuarios que intercambian ideas en forma de comentarios sobre productos que podría haber anteriormente. Un punto muy importante a la hora de haber elegido Ciao UK es que, al tener 28 categorías de productos de muchos tipos y no estar enfocada en un solo sector, como, por ejemplo, TripAdvisor (turismo) o El Tenedor (restauración), consideramos que los datos extraídos tienen más riqueza y posibilidades de explotación.

Ciao UK es un sitio web con más de 1,3 millones de usuarios que han escrito más de 7 millones de comentarios sobre 1,4 millones de productos (Olmedilla *et al.* 2015). Tres secciones principales son las que constituyen esta comunidad de boca a boca electrónico: Comentarios, Compras y Mi Ciao, siendo esta última sección la que evalúa la información relevante sobre cada usuario registrado. Hay 28 categorías principales de productos establecidas por Ciao UK, subdivididas, a su vez, en muchas más subcategorías, por lo que los usuarios registrados publican y comparten sus comentarios dentro de una categoría específica y de una subcategoría relacionada con el tema del comentario.

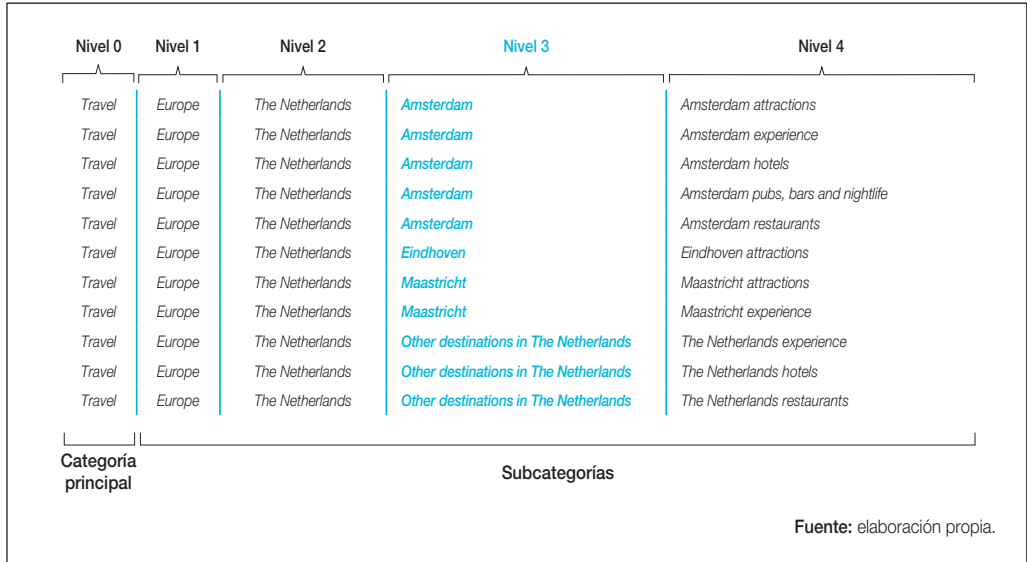
Casi todas las categorías principales tienen cuatro niveles de subcategorías. La figura 2 muestra como ejemplo los cuatro niveles de subcategorías para la categoría principal Viajes. En este caso, la subcategoría nivel 1 agrupa las revisiones por continente; la subcategoría nivel 2, por país; y la subcategoría nivel 3, por ciudad. El nivel 4 ofrece demasiado detalle para la oferta de destino de viaje.

Los usuarios acceden a Ciao UK de forma gratuita, teniendo la opción de registrarse, al crear una cuenta en la que proporcionan información no obligatoria sobre sí mismos (nombre, sexo, edad, país, etc.). Para escribir un comentario de al menos 120 palabras, dentro de la categoría y subcategoría adecuadas, los usuarios deben completar algunos campos, como el título del comentario, el nombre del producto, el cuerpo, que contiene la opinión del usuario, sus ventajas y desventajas, etc. Estos son los datos que se han recopilado

Ciao UK es un sitio web con más de 1,3 millones de usuarios que han escrito más de 7 millones de comentarios sobre 1,4 millones de productos. Tres secciones principales son las que constituyen esta comunidad de boca a boca electrónico: Comentarios, Compras y Mi Ciao

para este trabajo, formando un subconjunto de casi 45.000 usuarios, 106.000 comentarios aproximadamente de alrededor de 69.000 productos y cerca de 283.000 subcategorías.

Figura 2. Nivel de especificación de la categoría principal Viajes



Para la extracción de los datos de la web se ha aplicado el *crawler* desarrollado en Olmedilla, Martínez-Torres y Toral (2016), usando el lenguaje de programación Python y utilizando una librería de código abierto de extracción web llamada Scrapy. Para comenzar con el ciclo de extracción de todas las categorías, se ha definido un ítem Comentario, que contiene los campos Categoría Principal y Subcategoría que se quieren extraer. Luego, para extraer la información de cada categoría, se ha programado una clase llamada Spider. La clase Spider navega por la página web del producto. Mediante el uso de los selectores XPath para realizar extracciones de datos de la fuente HTML, se ha extraído toda la información sobre las categorías y subcategorías principales. Para ello, se ha llamado al método «response.xpath('text_to_gather').extract()» y se ha creado una lista con todos los elementos (vector) de categorías. A continuación, la información extraída se ha almacenado en una base de datos relacional diseñada en MySQL en los campos Categoría Principal y Subcategoría para la tabla Producto. El lenguaje de consulta SQL se puede utilizar para acceder y rastrear todos los datos almacenados en la base de datos.

Hay 28 categorías principales de productos establecidas por Ciao UK, subdivididas, a su vez, en muchas más subcategorías, por lo que los usuarios registrados publican y comparten sus comentarios dentro de una categoría específica y de una subcategoría relacionada con el tema del comentario

4.2. Técnicas cuantitativas

4.2.1. Ajuste de la ley de potencia a los datos recopilados

El primer método, la distribución de la ley de potencia, utilizado por Olmedilla *et al.* (2015), se ha ampliado para decidir si el conjunto de datos sigue una distribución de la ley de potencia. Según Clauset *et al.* (2009), muchas de las cosas que los científicos miden tienen un tamaño o «escala» típica; un valor típico en torno al cual se centran las mediciones individuales. Por tanto, como resultado se produce una distribución gaussiana. Sin embargo, no todas las cosas medidas son gaussianas. Los eventos grandes son extremadamente raros dentro de las distribuciones gaussianas. Además, los eventos situados en la cola, que sería la parte de la curva de distribución alejada de la media y que caracterizaría grandes e infrecuentes eventos, tienen más probabilidades de ocurrir en una distribución de la ley de potencia (Clauset *et al.*, 2009). En consecuencia, cuando la probabilidad de alcanzar un valor máximo en torno a un valor típico de alguna cantidad varía en un amplio rango dinámico, dicha cantidad podría ser coherente con una distribución de la ley de potencia (Newman, 2005), que se define como:

$$P(x) = Cx^{-\alpha} \text{ para } x > x_{min} \quad (1)$$

Donde:

- $P(x)$ = Probabilidad (frecuencia) de que la variable tome el valor x .
- α = Exponente de la distribución.
- x = Variable a analizar.
- C = Constante que depende del tipo de evento.
- x_{min} = Valor mínimo de x sobre el cual comienza el comportamiento de la ley de potencia.

Dado que las leyes de potencia típicamente describen sistemas donde los eventos más grandes son menos que los eventos más pequeños, α se mantiene positivo. Esto confirma que la ley de potencia es una función que decrece monótonamente.

Las leyes de potencia típicamente describen sistemas donde los eventos más grandes son menos que los eventos más pequeños

Tomando logaritmos en ambos lados de (1), se observa que para una ley de potencia:

$$\ln(P) = \ln C - \alpha \ln x \quad (2)$$

en una gráfica con escala logarítmica, la relación entre $\ln(P)$ y $\ln x$ se describe mediante una línea recta cuya pendiente negativa es α .

En la práctica, identificar el comportamiento de la ley de potencia es difícil. En muchos casos, es conveniente utilizar la función de distribución acumulada (FDA) de una variable distribuida de la ley de potencia, que se denota como:

$$P(x) = \int_x^{\infty} p(x)dx = \left(\frac{x}{x_{min}}\right)^{-\alpha+1} \quad (3)$$

Esencialmente, como resultado de la aplicación de este método de distribución de la ley de potencia, se obtiene el valor de corte x_{min} . Este valor define el punto donde la distribución de la ley de potencia ya no es válida. Por lo tanto, es el punto más allá, donde no hay productos *super-hit* y, en consecuencia, comienza la cola de la distribución.

Se ha utilizado el lenguaje de programación R para evaluar la presencia de una ley de potencia en los datos mediante el uso de la biblioteca *poweRlaw*, que sigue el protocolo para ajustar un modelo de la ley de potencia como lo describen *Clauset et al.* (2009). La función *PLFIT* se utilizó para estimar el x_{min} y el α según la bondad de ajuste.

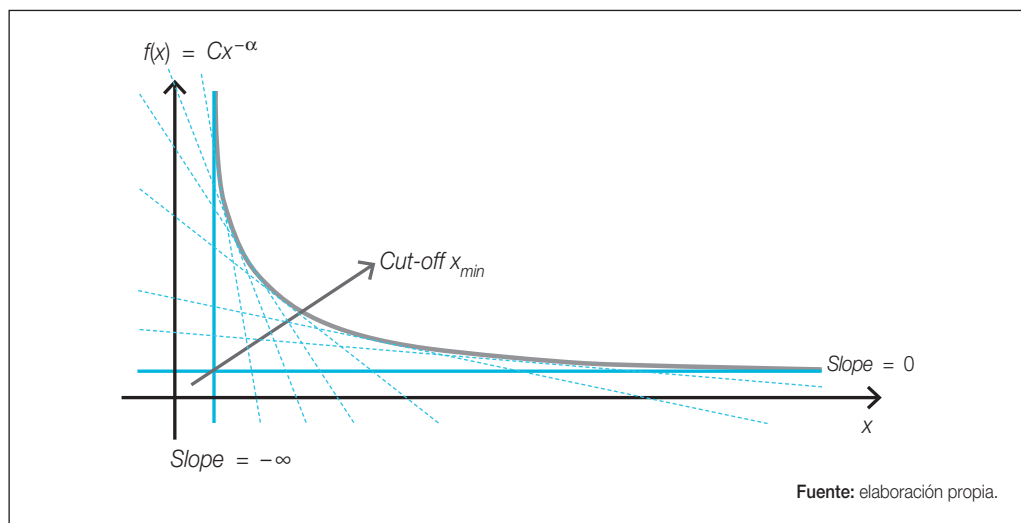
4.2.2. Encontrar el x_{min} óptimo según el método del codo

El segundo método utilizado se basa en el criterio del codo, que es un método visual alternativo para determinar el valor de corte x_{min} , es decir, para determinar qué número de *clusters* deben elegirse para el *clustering* de k-medias (Kodinariya y Makwana, 2013).

El método del codo se basa en la relación entre el porcentaje de varianza con respecto al número de *clusters* para encontrar el número óptimo de estos

El método del codo se basa en la relación entre el porcentaje de varianza con respecto al número de *clusters* para encontrar el número óptimo de estos. En esta relación, representada como un gráfico, se puede observar que las primeras posiciones revelan una pendiente alta debido a que, al tener un número bajo de *clusters*, es posible obtener un alto porcentaje de varianza entre ellos. Si el número de *clusters* continúa aumentando en cierto punto, la pendiente disminuye. Esto se debe a que el aumento del número de *clusters* podría no aumentar la varianza como antes, lo que significa que se ha superado el número óptimo de agrupaciones para la agrupación k-medias, dando un ángulo en la gráfica (Madhulatha, 2012). Extendiendo el criterio del codo al caso de una distribución de la ley de potencia, el problema consiste en encontrar el punto x_{min} , donde el «codo» puntiagudo es claramente visible en la gráfica (véase figura 3).

Figura 3. Aplicación del criterio de codo a la distribución de potencia



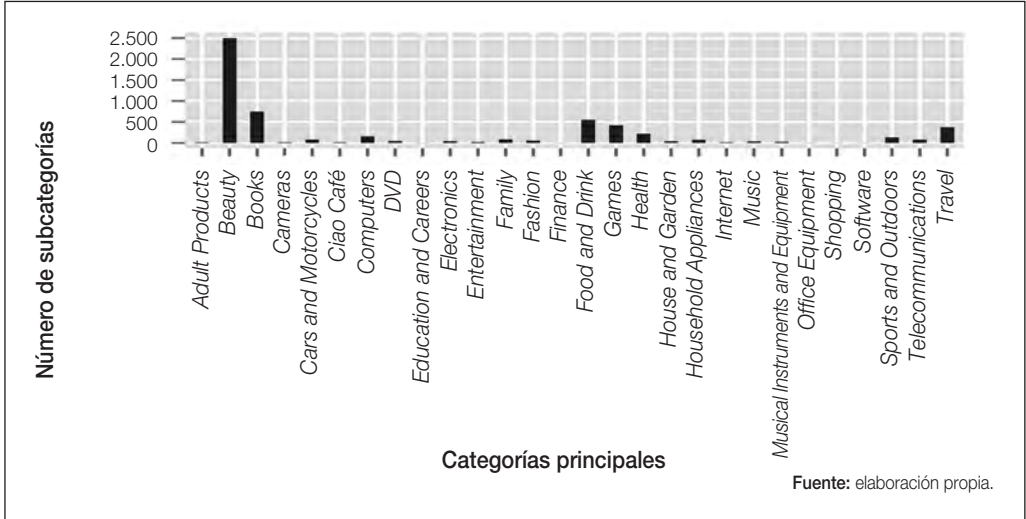
La figura 3 ilustra la selección del criterio del codo basándose en la elección del valor con una disminución brusca en la pendiente de la tangente. Dada la función $f(x) = Cx^{-\alpha}$, la línea de la tangente en el punto $(x_0, f(x_0))$ es la línea recta única que pasa por ese punto y tiene la misma pendiente que la gráfica en ese punto, que se define por el valor derivado $f'(x_0)$:

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \quad (4)$$

Las líneas continuas azules de la figura 3 representan la tangente inicial y final, correspondientes a los valores de pendiente de $-\infty$ y 0. Las líneas punteadas azules intermedias representan todos los valores intermedios posibles de la pendiente. x_{min} es el punto de la pendiente promedio entre las dos asíntotas del gráfico, dado por el valor -1 , que es el punto desde el cual la curva pasa de disminuir muy rápido a disminuir muy lento.

5. Resultados

Siguiendo la metodología propuesta, se ha ajustado una distribución de la ley de potencia a la curva de la distribución de subcategorías creadas por los usuarios que han publicado comentarios en las 28 categorías principales distinguidas por Ciao UK. El resultado recopilado se muestra en la figura 4, que ilustra dicha distribución de los comentarios en las categorías principales.

Figura 4. Distribución de comentarios publicados en las 28 categorías principales


Todas las distribuciones obtenidas muestran que hay algunas subcategorías muy populares, mientras que otras solo están respaldadas por una pequeña cantidad de comentarios. Al aplicar los dos métodos descritos en la sección anterior, nuestro objetivo es determinar la presencia del efecto *super-hit* y la longitud de la cola para cada una de las 28 categorías principales de Ciao UK.

El cuadro 1 describe en detalle los parámetros de «cola larga» de las 28 categorías principales de acuerdo con el ajuste de la ley de potencia. La segunda y la tercera columnas muestran los valores x_{min} y α del ajuste de la distribución de la ley de potencia, respectivamente. La cuarta columna corresponde a la bondad del ajuste, el p-valor. La quinta columna muestra la longitud de la cola, calculada como el número de subcategorías que tienen un número de comentarios por debajo del umbral x_{min} , es decir, el número de subcategorías que no forman parte de la distribución de la ley de potencia ajustada. Las columnas sexta y séptima representan las áreas de la cabeza y la cola de la distribución, respectivamente, calculadas también usando el valor x_{min} .

Los valores de x_{min} distinguen entre aquellas subcategorías que forman parte de la cabeza de la distribución de la ley de potencia (productos *super-hit*, por encima de x_{min}) y aquellos que pertenecen a la cola (productos nicho, por debajo de x_{min}). El exponente α determina la forma de la distribución de la ley de potencia. El ajuste tiene un p-valor asociado, de modo que en aquellos casos en que el p-valor calculado sea considerablemente inferior a 0,05, se rechaza la hipótesis nula, lo que significa que dicha categoría no sigue una distribución de la ley de potencia y, por lo tanto, no existe ningún efecto *super-hit*. Pero si el p-valor

resultante es mayor que 0,05, la hipótesis nula no puede ser rechazada y, por lo tanto, la distribución es probable que siga una distribución de la ley de potencia, mostrando entonces un efecto de *super-hit*. Finalmente, la columna de porcentaje del área principal indica el área contenida a la izquierda del punto de corte de x_{min} , que incluye los productos dominantes o los más vendidos. La columna de porcentaje del área de la cola incorpora la cola, es decir, la región a la derecha del punto de corte de x_{min} .

El cuadro 1 describe en detalle los parámetros de «cola larga» de las 28 categorías principales de acuerdo con el ajuste de la ley de potencia

Cuadro 1. Parámetros de «cola larga» de las 28 categorías principales de Ciao UK según el método de la ley de potencia

Principales categorías	x_{min}	α	p-valor	Longitud de la cola (%)	Área de la cabeza (%)	Área de la cola (%)
<i>Adult Products</i>	3	2,97	0,602	68,75	46,03	53,97
<i>Beauty</i>	3	2,02	0,000	65,91	74,32	25,68
<i>Books</i>	7	2,20	0,011	76,03	72,02	27,98
<i>Cameras</i>	4	1,63	0,231	10,53	98,04	1,96
<i>Cars and Motorcycles</i>	3	1,84	0,047	30,00	89,88	10,12
<i>Ciao Café</i>	18	1,62	0,085	20,00	98,27	1,73
<i>Computers</i>	5	1,74	0,002	55,48	91,29	8,71
<i>DVD</i>	5	1,78	0,435	52,17	91,13	8,87
<i>Education and Careers</i>	8	1,53	0,098	0,00	100	0
<i>Electronics</i>	20	2,21	0,683	55,56	88,74	11,26
<i>Entertainment</i>	18	2,72	0,940	56,52	80,92	19,08
<i>Family</i>	57	3,50	0,520	84,88	56,41	43,59
<i>Fashion</i>	2	2,09	0,065	35,71	77,92	22,08



Principales categorías	x_{min}	α	p-valor	Longitud de la cola (%)	Área de la cabeza (%)	Área de la cola (%)
<i>Finance</i>	118	3,24	0,590	50,00	82,78	17,22
<i>Food and Drink</i>	2	1,81	0,086	47,55	87,76	12,24
<i>Games</i>	9	2,13	0,171	80,52	73,59	26,41
<i>Health</i>	17	2,39	0,064	86,70	67,32	32,68
<i>House and Garden</i>	19	2,01	0,683	35,90	94,82	5,18
<i>Household Appliances</i>	71	3,50	0,703	81,33	52,62	47,38
<i>Internet</i>	13	1,50	0,048	11,76	99,45	0,55
<i>Music</i>	8	3,50	0,179	71,05	60,18	39,82
<i>Musical Instruments and Equipment</i>	8	2,70	0,711	83,33	45,28	54,72
<i>Office Equipment</i>	4	1,52	0,161	0,00	100	0
<i>Shopping</i>	11	1,50	0,458	0,00	100	0
<i>Software</i>	7	1,52	0,538	20,00	98,13	1,87
<i>Sports and Outdoors</i>	5	2,21	0,036	68,46	72,56	27,44
<i>Telecommunications</i>	71	2,22	0,020	90,00	84,10	15,90
<i>Travel</i>	3	1,98	0,157	53,23	83,12	16,88

Fuente: elaboración propia

El cuadro 2 describe los resultados del mismo análisis para las 28 categorías principales, pero utilizando el método del codo. La primera columna muestra los valores de x_{min} , que corresponden al punto de inflexión, es decir, el punto desde el cual la función pasa de disminuir muy rápido a disminuir muy lentamente. En consecuencia, la función se recorre de derecha a izquierda, de x_{max} a -1 , hasta el primer punto/valor que cumple la condición de búsqueda. Este es el punto de la pendiente promedio entre las dos asíntotas de la gráfica. La tercera

columna muestra la longitud de la cola, que se halló utilizando el punto de inflexión calculado. Finalmente, la cuarta y quinta columnas representan las áreas de la cabeza y la cola de la distribución, respectivamente, calculadas utilizando el punto de inflexión obtenido.

El cuadro 2 describe los resultados del mismo análisis para las 28 categorías principales, pero utilizando el método del codo

Cuadro 2. Parámetros de «cola larga» de las 28 categorías principales de Ciao UK según el método del codo

Principales categorías	x_{min}	Longitud de la cola (%)	Área de la cabeza (%)	Área de la cola (%)
<i>Adult Products</i>	2	56,25	63,49	36,51
<i>Beauty</i>	2	50,50	80,60	19,40
<i>Books</i>	5	69,51	77,23	22,77
<i>Cameras</i>	4	10,53	98,04	1,96
<i>Cars and Motorcycles</i>	2	18,75	95,63	4,37
<i>Ciao Café</i>	8	13,33	99,19	0,81
<i>Computers</i>	4	50,32	92,52	7,48
<i>DVD</i>	4	50,00	93,51	6,49
<i>Education and Careers</i>	8	0,00	100,00	0,00
<i>Electronics</i>	24	62,22	81,47	18,53
<i>Entertainment</i>	23	70,83	62,31	37,69
<i>Family</i>	65	87,21	46,34	53,66
<i>Fashion</i>	2	35,71	77,92	22,08
<i>Finance</i>	149	66,67	54,04	45,96
<i>Food and Drink</i>	2	210,31	87,76	12,24



Principales categorías	X_{min}	Longitud de la cola (%)	Área de la cabeza (%)	Área de la cola (%)
<i>Games</i>	11	83,61	69,11	30,89
<i>Health</i>	23	90,37	59,76	40,24
<i>House and Garden</i>	10	30,77	95,64	4,36
<i>Household Appliances</i>	82	85,33	43,82	56,18
<i>Internet</i>	10	5,88	99,76	0,24
<i>Music</i>	5	68,42	68,86	31,14
<i>Musical Instruments and Equipment</i>	9	86,67	38,58	61,42
<i>Office Equipment</i>	8	16,67	97,11	2,89
<i>Shopping</i>	107	33,33	85,56	14,44
<i>Software</i>	7	20,00	98,13	1,87
<i>Sports and Outdoors</i>	7	77,69	63,88	36,12
<i>Telecommunications</i>	83	91,25	79,21	20,79
<i>Travel</i>	2	36,83	88,47	11,53

Fuente: elaboración propia

Ambos métodos se han comparado en el cuadro 3 para probar su coincidencia o no, identificando la presencia de la «cola larga» y considerando los criterios de área y longitud. El resultado de esta comparación se muestra en la última columna del cuadro 3. Un valor verdadero significa que ambos métodos están de acuerdo con la presencia o ausencia de la «cola larga», un valor falso significa que ambos métodos están en desacuerdo y un valor incierto

Un valor verdadero significa que ambos métodos están de acuerdo con la presencia o ausencia de la «cola larga», un valor falso significa que ambos métodos están en desacuerdo y un valor incierto significa que ambos métodos generan resultados contradictorios

significa que ambos métodos generan resultados contradictorios. La toma de decisiones se ha llevado a cabo aplicando al área debajo de la cola o longitud de la cola la relación de unión 80:20. Es decir, siempre que el área o la longitud de la cola estén por debajo del 20 % del valor máximo, se considera que no hay «cola larga». De lo contrario, se acepta la presencia de la «cola larga».

Cuadro 3. Validación de la presencia de «cola larga» a través de la consistencia de reglas de decisión

Principales categorías	Áreas		Longitud de la cola		Comparación
	Ley de potencia	Codo	Ley de potencia	Codo	Áreas versus cola
<i>Adult Products</i>	Fuerte/Sí	Sí	Sí	Sí	Verdadero
<i>Beauty</i>	Sí	Sí	Sí	Sí	Verdadero
<i>Books</i>	Sí	Sí	Fuerte/Sí	Sí	Verdadero
<i>Cameras</i>	No	No	No	No	Verdadero
<i>Cars and Motorcycles</i>	No	No	Sí	Sí	Falso
<i>Ciao Café</i>	No	No	Sí	No	Incierto
<i>Computers</i>	No	No	Sí	Sí	Falso
<i>DVD</i>	No	No	Sí	Sí	Falso
<i>Education and Careers</i>	No	No	No	No	Verdadero
<i>Electronics</i>	No	Sí	Sí	Sí	Incierto
<i>Entertainment</i>	Sí	Sí	Sí	Sí	Verdadero
<i>Family</i>	Sí	Fuerte/Sí	Fuerte/Sí	Fuerte/Sí	Verdadero
<i>Fashion</i>	Sí	Sí	Sí	Sí	Verdadero
<i>Finance</i>	Sí	Sí	Sí	Sí	Verdadero



Principales categorías	Áreas		Longitud de la cola		Comparación
	Ley de potencia	Codo	Ley de potencia	Codo	Áreas versus cola
<i>Food and Drink</i>	Sí	Sí	Sí	Sí	Verdadero
<i>Games</i>	Sí	Sí	Fuerte/Sí	Fuerte/Sí	Verdadero
<i>Health</i>	Sí	Sí	Fuerte/Sí	Fuerte/Sí	Verdadero
<i>House and Garden</i>	No	No	Sí	Sí	Falso
<i>Household Appliances</i>	Sí	Fuerte/Sí	Fuerte/Sí	Fuerte/Sí	Verdadero
<i>Internet</i>	No	No	No	No	Falso
<i>Music</i>	Sí	Sí	Sí	Sí	Verdadero
<i>Musical Instruments and Equipment</i>	Fuerte/Sí	Fuerte/Sí	Fuerte/Sí	Fuerte/Sí	Verdadero
<i>Office Equipment</i>	No	No	No	No	Verdadero
<i>Shopping</i>	No	Sí	No	Sí	Incierto
<i>Software</i>	No	No	Sí	Sí	Falso
<i>Sports and Outdoors</i>	Sí	Sí	Sí	Fuerte/Sí	Verdadero
<i>Telecommunications</i>	Sí	Sí	Sí	Fuerte/Sí	Verdadero
<i>Travel</i>	Sí	No	Sí	Sí	Incierto

Fuente: elaboración propia

Las figuras 5.1 y 5.2 representan algunas de las distribuciones para las categorías principales mostradas en el cuadro 3. Las gráficas muestran una distribución de la ley de potencia ajustada a la curva de la distribución de subcategorías de productos. En consecuencia, el eje y corresponde al volumen de comentarios (número de comentarios escritos por los usuarios) y el eje x representa todas las subcategorías de productos de cada una de las 28 categorías principales de productos distinguidas por Ciao UK.

En ambas figuras, el valor de x_{min} está demarcado por una línea gris horizontal que representa su valor calculado por el método del codo y por una línea azul horizontal que representa su valor calculado por el método de la ley de potencia. De acuerdo con el cuadro 3, observamos que 15 de las 28 categorías principales (*Adult Products* [Productos para Adultos], *Beauty* [Belleza], *Books* [Libros], *Entertainment* [Entretenimiento], *Family* [Familia], *Fashion* [Moda], *Finance* [Finanzas], *Food and Drinks* [Alimentos y Bebidas], *Games* [Juegos], *Health* [Salud], *Household Appliances* [Electrodomésticos], *Music* [Música], *Musical Instruments and Equipment* [Instrumentos Musicales y Equipos], *Sports and Outdoors* [Deportes y Aire Libre] y *Telecommunications* [Telecomunicaciones]) exhiben claramente un comportamiento de «cola larga», ya que ambos criterios se cumplen utilizando los dos métodos propuestos.

Figura 5.1. Distribución de los comentarios para algunas de las principales categorías en Ciao UK

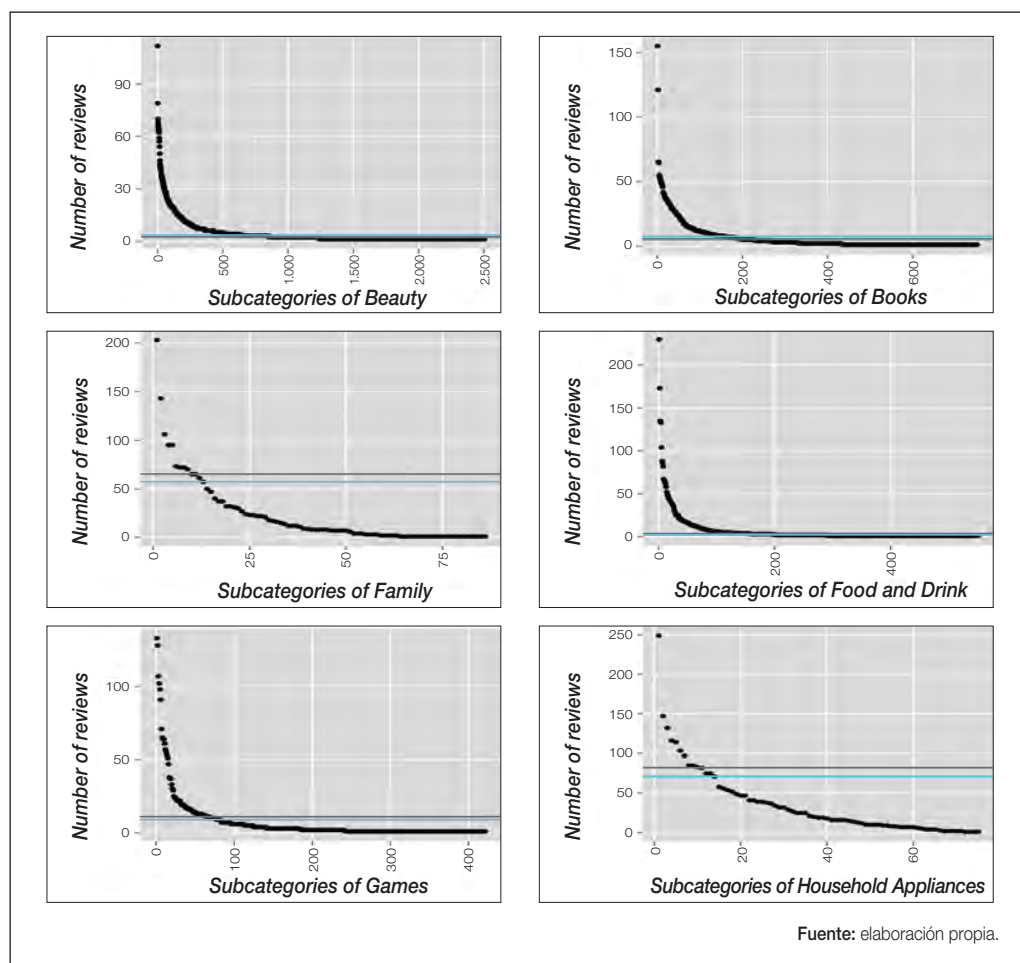
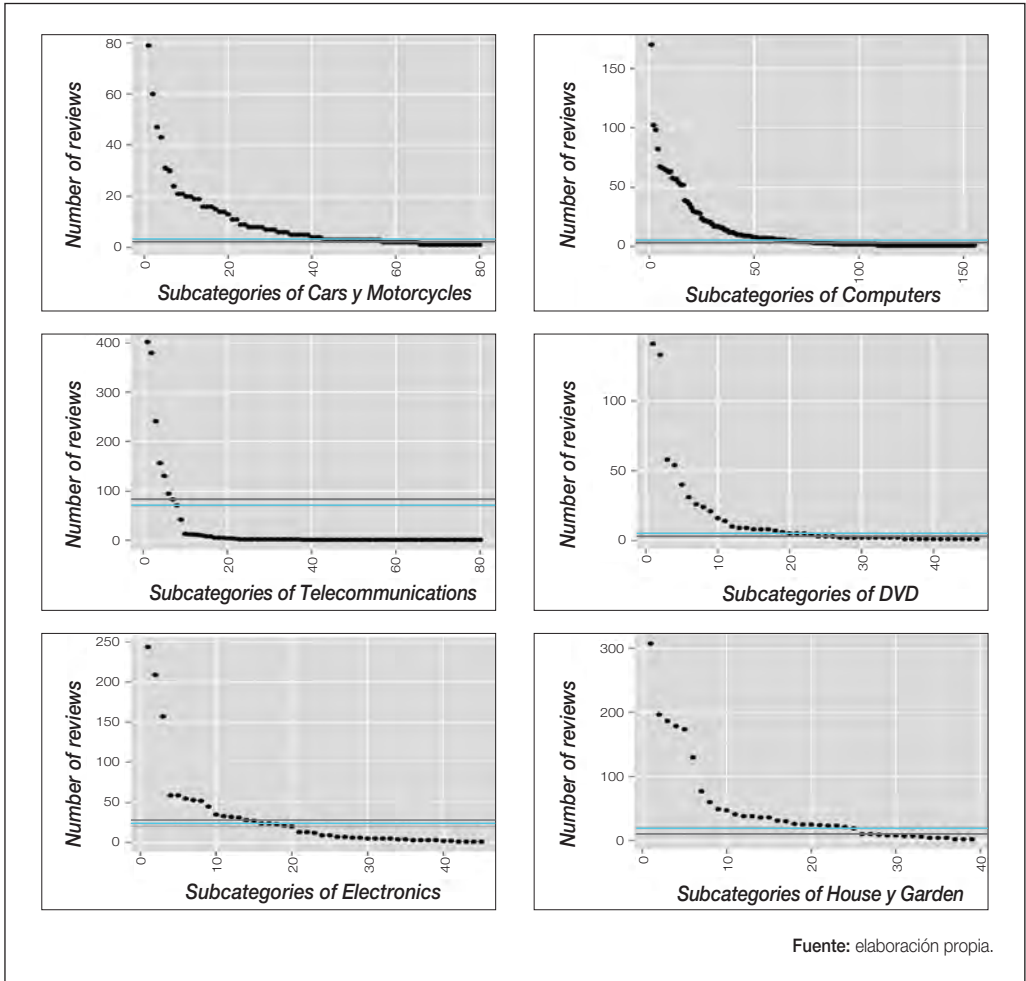


Figura 5.2. Distribución de los comentarios para algunas de las principales categorías en Ciao UK



Con respecto al efecto *super-hit*, los p-valor del cuadro 1 revelan que 7 de las 28 categorías del conjunto de datos (*Beauty* [Belleza], *Books* [Libros], *Cars and Motorcycles* [Automóviles y Motocicletas], *Computers* [Computadoras], *Internet* [Internet], *Sports and Outdoor* [Deportes y Aire Libre] y *Telecommunications* [Telecomunicaciones]) no obedecen a una distribución de la ley de potencia (véase figura 5.2). Sus p-valor son lo suficientemente pequeños para que el modelo de la ley de potencia pueda ser descartado firmemente, según su bondad de ajuste. Por otra parte, y de acuerdo con sus p-valor, las 21 categorías restantes pueden ser consistentes con una distribución de la ley de potencia. En consecuencia, y respondiendo a PI_1 , el boca a boca electrónico promueve tanto el efecto de *super-hit* como el fenómeno de «cola larga», aunque el efecto *super-hit* es más frecuente en las 28 categorías consideradas.

Entre los 21 casos con el efecto *super-hit*, hay 6 casos sin «cola larga» (DVD, *Education and Careers* [Educación y Carreras], *Cameras* [Cámaras], *House and Garden* [Casa y Jardín], *Office Equipment* [Material de Oficina] y *Software*), 11 con «cola larga» (*Adult Products* [Productos para Adultos], *Entertainment* [Entretenimiento], *Family* [Familia], *Finance* [Finanzas], *Food and Drinks* [Alimentos y Bebidas], *Games* [Juegos], *Health* [Salud], *Household Appliances* [Electrodomésticos], *Music* [Música], *Musical Instruments and Equipment* [Instrumentos y Equipos Musicales] y *Fashion* [Moda]) y, finalmente, 4 de ellos que no está claro si exhiben una «cola larga» o no (*Ciao Café* [Ciao Café], *Shopping* [Compras], *Travel* [Viajes] y *Electronics* [Electrónica]). Estos hallazgos responden a la PI_2 , al mostrar que es posible que el efecto de *super-hit* coexista con el fenómeno de la «cola larga» dependiendo de las categorías de productos.

6. Discusión e implicaciones

6.1. Discusión

Este artículo contribuye a la creciente investigación sobre cómo internet influye en el efecto *super-hit* y la formación de «cola larga» a través de sus datos, y, más específicamente, del contenido generado por el usuario. Los resultados sugieren, como establecen Park y Lee (2009), que es importante considerar el impacto de las comunidades de boca a boca electrónico en el proceso de descubrimiento de productos por parte de los consumidores, ya que muchos productos solo se venden *online* y sus consumidores son más propensos a utilizar los comentarios *online* como fuente principal de información. La interacción de los usuarios que intercambian información sobre los productos dentro de las comunidades de boca a boca electrónico crea una dinámica compleja en la demanda.

La literatura existente relacionada con la «cola larga» ha demostrado que el boca a boca electrónico tiene un impacto positivo en las ventas (Elberse y Oberholzer-Gee, 2007; Hervas-Draney, 2009). Además, otros estudios de boca a boca electrónico han encontrado evidencia del efecto *super-hit*, apoyando que este boca a boca promueve las ventas de productos populares, por lo que la parte principal de la distribución de ventas se vuelve más gruesa (Standifird, 2001). La premisa de nuestro estudio es aclarar estas dos escuelas de pensamiento. En este sentido, se ha encontrado evidencia de que el boca a boca electrónico es un promotor tanto de uno como de otro dependiendo de las categorías de productos. Sin embargo, y siguiendo estudios previos (Standifird, 2001), la presencia del efecto *super-hit* está más extendida que la «cola larga», lo que indica que el fenómeno de «cola larga» está estrechamente relacionado con ciertas categorías de productos donde las revisiones difunden nuevos conocimientos sobre artículos o subcategorías desconocidas para la mayoría de los usuarios. Además, también encontramos evidencia de que el fenómeno de «cola larga» está estrechamente relacionado con ciertas categorías de productos. Los hallazgos están respaldados por dos métodos cuantitativos (Clauset *et al.*, 2009; Kodinariya y Makwana, 2013) que proporcionan evidencia de forma matemática sobre el efecto *super-hit* y el fenómeno de «cola larga».

En línea con los estudios de Dellarocas y Narayan (2007) y Elberse y Oberholzer-Gee (2007), demostramos que el fenómeno de «cola larga» es capaz de coexistir con el efecto *super-hit* dependiendo de la categoría de producto. No es sorprendente que internet haga que a los consumidores les resulte más fácil recopilar información a bajo costo y les permita confiar más en la experiencia de otros usuarios cuando evalúan productos (Filieri *et al.*, 2015). Una vez más, la gran disponibilidad de comentarios *online* sobre muchos tipos de productos dentro de las comunidades de boca a boca electrónico es el factor explicativo de este fenómeno, dado que este canal *online* permite a los clientes encontrar más productos nicho que se ajusten a sus preferencias o promover la popularidad de los productos, llevando a una situación de «el rico se enriquece más», favoreciendo los productos que ya son muy populares en ciertas categorías de productos y también a una distribución más igualitaria que facilita la formación de la «cola larga» entre otras categorías de productos. La coexistencia de productos *super-hit* y productos nicho también significa que el boca a boca electrónico probablemente ha expandido una cola corta original. Casos como la música, el entretenimiento o la moda son ejemplos de categorías de productos donde los mercados *online* ofrecen la difusión de nuevas opciones desconocidas de otra forma para los clientes. Además, los comentarios *online* proporcionan no solo la descripción de nuevas opciones, sino también la calidad de la experiencia. La mayoría de los comentarios *online* se centran más en los sentimientos, en las percepciones y en la satisfacción que en las características objetivas de los productos. Por lo tanto, los canales *online* ofrecen información más completa que la que se puede encontrar en catálogos o publicidad.

6.2. Implicaciones para la gestión

Hoy en día, cualquier marca o empresa puede llegar sin costo alguno a muchos consumidores potenciales a través de internet. Esto hace que sea rentable invertir tiempo y esfuerzo para crear productos que puedan interesar incluso a un pequeño segmento de ellos. Por lo tanto, una estrategia para aumentar la variedad e información de productos nicho podría permitir a los consumidores descubrir aquellos que, de otra manera, no estarían disponibles.

Por ello, un minorista debe comprender el impacto del boca a boca electrónico en la distribución de diferentes productos para ofrecer la variedad precisa y administrar el inventario. En este sentido, nuestros resultados muestran qué productos de la distribución de categorías

Una estrategia para aumentar la variedad e información de productos nicho podría permitir a los consumidores descubrir aquellos que, de otra manera, no estarían disponibles

Un minorista debe comprender el impacto del boca a boca electrónico en la distribución de diferentes productos para ofrecer la variedad precisa y administrar el inventario

de productos forman la «cola larga» o son *super-hits*, lo que permite especificar los casos en detalle. Por tanto, es posible que los minoristas *online* deban prestar más atención si su objetivo es capturar el mercado masivo, adquiriendo y administrando en este caso clientes mediante el uso de los productos más populares o *super-hit* y centrándose también en la comercialización de esos productos de gran éxito. Por el contrario, si su objetivo es utilizar estrategias de «cola larga» para vender una gama más amplia de productos en cantidades más pequeñas, deberían centrarse en ampliar su surtido con más productos nicho.

Asimismo, las empresas pueden aprovechar los resultados para desarrollar mejores estrategias de marketing en publicidad y en promoción de productos que los consumidores tienen más probabilidades de comprar en el futuro.

Las empresas pueden aprovechar los resultados para desarrollar mejores estrategias de marketing en publicidad y promoción de productos que los consumidores tendrán más probabilidades de comprar en el futuro

6.3. Limitaciones de la investigación

La principal limitación podría ser que la metodología se ha implementado en una sola comunidad de boca a boca electrónico. Sin embargo, sería posible extender los dos métodos (ajuste de la ley de potencia y método del codo) para ajustar la distribución de productos en otras comunidades de boca a boca electrónico. Otra limitación metodológica podría ser la muestra del conjunto de datos, debido a que Ciao UK no representa a toda la población, sino a un subconjunto particular de usuarios *online*. Además, los usuarios de Ciao UK pueden tener varios perfiles, o incluso los perfiles de los usuarios pueden ser empleados por varias personas. Si bien esta limitación no influye en los resultados, debe entenderse que se aplican a un determinado conjunto de usuarios *online*. Una limitación adicional es el hecho de que un usuario no tiene que proporcionar ninguna prueba de compra del producto para que se le permita escribir un comentario. Sin embargo, al escribir comentarios, los usuarios consiguen beneficios de estatus y económicos, y, para obtenerlos, pueden no estar interesados en ser deshonestos. Finalmente, otra limitación es la ausencia de datos de ventas, ya que este documento se basa en el volumen de comentarios *online*, que es la mejor información disponible dentro de una comunidad de boca a boca electrónico. A pesar de ello, este estudio se puede contextualizar en que analiza la diversidad de los productos que supuestamente han sido comprados antes de que el usuario escriba un comentario *online*.

6.4. Trabajo futuro

Investigaciones futuras pueden extender los hallazgos caracterizando cuáles son los productos nicho de la «cola larga» y cuáles los productos *super-hit*. El objetivo sería descubrir

algunos patrones comunes entre los productos nicho mediante la creación de modelos de redes sociales o reglas de asociación y descubrir grupos de productos nicho que pueden representar un objetivo rentable para los minoristas. Además, los productos pueden clasificarse como productos de búsqueda y experiencia, por lo que, como trabajo futuro, se podría analizar qué tipo de producto es más propenso a exhibir los dos comportamientos estudiados. También sería interesante poder probar la validez de los modelos presentados en otras plataformas de boca a boca electrónico con información específica disponible de ventas (por ejemplo, Amazon).

Referencias bibliográficas

- Anderson, C. (2004). The long tail. *Wired Magazine*, 12 (10), 170-177.
- Anderson, C. (2008). *Long Tail: Why the Future of Business is Selling Less of More*. New York: Hyperion Books.
- Barabási, A. L. y Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509-512.
- Bimpikis, K. y Markakis, M. G. (2015). Inventory pooling under heavy-tailed demand. *Management Science*, 62(6), 1.800-1.813.
- Brynjolfsson, E., Hu, Y. y Smith, M. (2003). Consumer surplus in the digital economy: estimating the value of increased product variety at online booksellers. *Management Science*, 49(11), 1.580-1.596.
- Brynjolfsson, E., Hu, Y. y Smith, M. (2010). Research commentary-long tails vs. supers-tars: the effect of information technology on product variety and sales concentration patterns. *Information Systems Research*, 21(4), 736-747.
- Casado-Aranda, L. A., Dimoka, A. y Sánchez-Fernández, J. (2019). Consumer proces-sing of online trust signals: a neuroimaging study. *Journal of Interactive Marketing*, 47, 150-180.
- Chen, Y., Wang, Q. y Xie, J. (2011). Online social interactions: a natural experiment on word of mouth versus observational learning. *Journal of marketing research*, 48(2), 238-254.
- Chevalier, J. y Mayzlin, D. (2006). The effect of word of mouth on sales: online book reviews. *Journal of Marketing Research*, 43(3), 345-354.
- Clauset, A., Shalizi, C. y Newman, M. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4), 661-703.
- Crawford, G. C., Aguinis, H., Lichtenstein, B., Davidsson, P. y McKelvey, B. (2015). Power law distributions in entrepreneurship: implications for theory and research. *Journal of Business Venturing*, 30(5), 696-713.
- Dellarocas, C. y Narayan, R. (2007). Tall heads vs. long tails: do consumer reviews increase the informational inequality between hit and niche products? *School of Business Research Paper*, 06-056.

- Elberse, A. (2008). Should you invest in the long tail? *Harvard Business Review*, 86 (7/8), 88-96.
- Elberse, A. y Oberholzer-Gee, F. (2007). Superstars and underdogs: an examination of the long tail phenomenon in video sales. *Harvard Business School*, 07-015.
- Filieri, R., Algezauzi, S. y McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, 51, 174-185.
- Goel, S., Broder, A., Gabrilovich, E. y Pang, B. (2010). Anatomy of the long tail: ordinary people with extraordinary tastes. *Proceedings of the Third ACM International Conference on Web Search and Data Mining* (pp. 201-210). New York.
- Gu, B., Tang, Q. y Whinston, A. (2013). The influence of online word-of-mouth on long tail formation. *Decision Support Systems*, 56, 474-481.
- Hervas-Drane, A. (2009). Word of mouth and taste matching: a theory of the long tail. *NET Institute Working Paper*, 07-041.
- Koçaş, C. y Akkan, C. (2016). A system for pricing the sales distribution from blockbusters to the long tail. *Decision Support Systems*, 89, 56-65.
- Kodinariya, T. y Makwana, P. (2013). Review on determining number of cluster in k-means clustering. *International Journal*, 1(6), 90-95.
- Lee, J., Lee, J. N. y Shin, H. (2011). The long tail or the short tail: the category-specific impact of eWOM on sales distribution. *Decision Support Systems*, 51(3), 466-479.
- Madhulatha, T. (2012). An overview on clustering methods. *Journal of Engineering IOSR*, 2(4), 719-725.
- Mahanti, A., Carlsson, N., Mahanti, A., Arlitt, M. y Williamson, C. (2013). A tale of the tails: power-laws in internet measurements. *IEEE Network*, 27(1), 59-64.
- Manes, E. y Tchetchik, A. (2018). The role of electronic word of mouth in reducing information asymmetry: an empirical investigation of online hotel booking. *Journal of Business Research*, 85, 185-196.
- 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.
- Maslowska, E., Malthouse, E. C. y Viswanathan, V. (2017). Do customer reviews drive purchase decisions? The moderating roles of review exposure and price. *Decision Support Systems*, 98, 1-9.
- Moe, W. W. y Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, 48(3), 444-456.
- Morales-Arroyo, M. y Pandey, T. (2010). Identification of critical eWOM dimensions for music albums. *IEEE International Conference on Management of Innovation and Technology* (pp. 1.230-1.235). Singapore.
- Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323-351.
- Olmedilla, M.^a, Martínez-Torres, M.^aR. y Toral, S. L. (2015). Examining the power-law distribution among eWOM communities: a characterisation approach of the long tail. *Technology Analysis & Strategic Management*, 28(5), 601-613.
- Olmedilla, M.^a, Martínez-Torres, M.^aR. y Toral, S. L. (2016). Harvesting big data in social science: a methodological approach for collecting online user-generated content. *Computer Standards & Interfaces*, 46, 79-87.

- Park, C. y Lee, T. M. (2009). Information direction, website reputation and eWOM effect: a moderating role of product type. *Journal of Business Research*, 62(1), 61-67.
- Park, D. H., Lee, J. y Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: the moderating role of involvement. *International Journal of Electronic Commerce*, 11(4), 125-148.
- Shao, K. (2012). The effects of controversial reviews on product sales performance: the mediating role of the volume of word of mouth. *International Journal of Marketing Studies*, 4(4), 32-38.
- Standifird, S. (2001). Reputation and e-commerce: eBay auction and the asymmetrical impact of positive and negative ratings. *Journal of Management*, 27(3), 279-295.
- Yeap, J., Ignatius, J. y Ramayah, T. (2014). Determining consumers' most preferred eWOM platform for movie reviews: a fuzzy analytic hierarchy process approach. *Computers in Human Behavior*, 31, 250-258.
- Zhang, L., Ma, B. y Cartwright, D. K. (2013). The impact of online user reviews on cameras sales. *European Journal of Marketing*, 47(7), 1.115-1.128.
- Zhang, K. Z., Zhao, S. J., Cheung, C. M. y Lee, M. K. (2014). Examining the influence of online reviews on consumers' decision-making: a heuristic-systematic model. *Decision Support Systems*, 67, 78-89.
- Zhou, W. y Duan, W. (2015). An empirical study of how third-party websites influence the feedback mechanism between online word-of-mouth and retail sales. *Decision Support Systems*, 76, 14-23.
- Zhu, F. y Zhang, X. (2010). Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2), 133-148.



Máster en Enseñanza del Español como Lengua Extranjera

Este máster oficial [60 créditos ECTS] tiene una duración normal de 12 meses.

Dirigido a: Titulados universitarios en las distintas áreas de Lenguas Modernas, Filología, Lingüística, Traducción, Ciencias de la Educación, Humanidades, Pedagogía y Ciencias de la Comunicación y profesionales de las lenguas que deseen especializarse en la enseñanza del español como lengua extranjera.

Objetivos: Formar profesionales cualificados en el ámbito de la enseñanza del español como lengua extranjera, dotándolos de un perfil práctico y competitivo que los capacite para realizar adecuadamente y con garantía de éxito su labor profesional, facilitándoles su integración en un mercado laboral de enorme proyección internacional.

Inicio en **octubre** y **febrero** de cada año

www.udima.es | 918 561 699