

Bibliometría

Por: M.L.S. Alice Miranda Arguedas

Directora

Escuela de Bibliotecología,
Documentación e Información.

A principios de 1990, los documentalistas comienzan a interesarse en hacer aplicaciones matemáticas y estadísticas en las unidades bibliográficas. F. J. Coles y Nellie B. Eales en 1917 hicieron el primer estudio con un grupo de títulos de documentos cuyo análisis consideraba el país de origen ([White](#), p. 35). En 1923, E. Wyndham Hulme fue la primera persona en usar el término "estadísticas bibliográficas".

Y propuso la utilización de métodos estadísticos para tener parámetros que sirvan para conocer el proceso de la comunicación escrita y, la naturaleza y curso del desarrollo de una disciplina. Para lograr ese aspecto empezó contando un número de documentos y analizando varias facetas de la comunicación escrita empleada en ellos ([Ferrante](#), p. 201). En un documento escrito en 1969, Alan Pritchard propuso el término bibliometría para reemplazar el término "estadísticas bibliográficas" empleado por Hulme, argumentando que el, término es ambiguo, no muy descriptivo y que puede ser confundido con las estadísticas puras o estadísticas de bibliografías. El definió el término bibliometría como la aplicación de la matemática y métodos estadísticos a los libros y otros documentos (p. 348-349). Y desde ese momento se ha utilizado este término.

Muchas personas empiezan a interesarse en el estudio estadístico de los documentos, los patrones que se observan y se formulan leyes empíricas basadas en la observación. En 1926, Lotka desarrolla una ley relacionada con la distribución productiva de los autores. Zipf formuló, en 1933, una nueva ley relacionada con la frecuencia con que es usada una palabra dentro de un texto. En 1934, Bradford desarrolla una nueva ley para descubrir los patrones que se observan en la distribución de documentos (especialmente publicaciones periódicas) en una disciplina específica o un área problema ([White](#), p. 36). Las leyes citadas son las tres leyes básicas de la bibliometría. Hay casos especiales de distribuciones hiperbólicas donde un incremento geométrico provoca un descenso aritmético en la producción. Como las leyes han sido desarrolladas empíricamente, describen eventos pero no explican comportamientos. Las técnicas bibliométricas son usadas para analizar un tipo de literatura y así controlar ésta y entender el tema. A lo que B.C. Brookes dice ...parece ofrecer el único medio disponible en este momento para reducir la presente incertidumbre que nos indique qué cantidad de documentación científica, sistemas de información y servicios bibliotecológicos están funcionando de una manera ordenada que permita establecer criterios capaces de ser racional y económicamente planeados y organizados. ([Saracevie](#), p. 121).

Esta confusión ha aumentado en la literatura sobre bibliometría porque hay que considerar dos dimensiones en este campo: análisis cuantitativos y cualitativos de la literatura sobre una disciplina que involucro leyes empíricas para parámetros no cubiertos y, además, describir regularidades en la literatura ([Ferrante](#), p. 199). B. C. Brookes describe cinco objetivos del análisis cuantitativo de una literatura:

1. Diseñar sistemas y redes de información mas economicas.
2. Aumentar las cuotas de eficiencia de los procesos relacionados con la información.
3. Identificar y medir las deficiencias de las fuentes bibliográficas presentes.

4. Establecer predicciones sobre la tendencia publicitaria.
5. Descubrir y dilucidar las leyes empíricas que pueden proveer una base para el desarrollo de una teoría de la información científica ([White](#), p. 36).

Este tipo de análisis únicamente puede predecir y describir situaciones, pero no puede puntualizar causas.

El análisis cuantitativo de la literatura de una disciplina permite la utilización práctica de encuentros investigativos y provee comparaciones críticas. E. [White](#) señala cinco usos del análisis cualitativo en bibliotecas:

1. Identificar la esencia de la literatura.
2. Tener un número mayor de publicaciones en zonas de importante desprovisión.
3. Establecer un punto de equilibrio entre zonas de mayor y menor utilización de la información.
4. Establecer una lluvia de ideas como un estudio epidérmico.
5. Clasificar segmentos de una literatura a través de una interconexión de citas (P. 38).

[Saracevic y Perk](#) advierten que el análisis cualitativo es subjetivo y puede presentar un cuadro incompleto (p. 120). Sin embargo, [White](#) indica que aplicar leyes empíricas ha dado como resultado una combinación de dimensiones cualitativas y cuantitativas que han satisfecho necesidades investigativas en los dos aspectos anotados (p. 36). Parece indicar que cada dimensión del estudio bibliométrico ilumina algunas propiedades de una literatura.

Las tres leyes sobre bibliometría: Lotka, Zipf y Bradford, han sido examinadas y modificadas muchas veces y por muchas personas. [White](#) apunta un problema investigativo que existe con las leyes y es que son matemáticamente convincentes pero en la práctica son problemáticas (p. 35). Como las leyes son evaluadas en diferentes situaciones, luce como que los resultados observados no siempre se ajustan a la distribución esperada. Los investigadores han modificado las leyes para contar con diferentes variables que permitan analizar diferentes situaciones, pero todavía no existe una teoría general basada en la observación de las leyes empíricas que nos permitan indicar por qué diferentes aspectos de la literatura siguen patrones establecidos. [O'connor y Voos](#) sostienen que la diversidad de aplicaciones de los métodos prácticos de la bibliometría estarán limitados hasta que sea desarrollada una teoría única (p. 11).

Ley de Lotka

En 1926, Alfred J. Lotka concluyó un estudio sobre la productividad de los autores utilizando dos poblaciones de químicos. De los datos observados, formuló una ley empírica describiendo la productividad de los autores en una población dada: el número de químicos (N) que publican publicaciones químicas es proporcional a $1/n^2$ de esos químicos. Derek de Solla Price reescribió la ley así: la mitad de los documentos científicos son contribución de la raíz cuadrada del total del número de autores científicos (p. 36).

En una discusión sobre la ley de Lotka, [William Gray Potter](#) dijo que el artículo en el cual Lotka da a conocer su ley no fue publicado sino hasta 1941, su distribución no fue llamada ley de Lotka hasta 1949 y ningún intento fue hecho para evaluar la aplicabilidad de la ley de Lotka en otras disciplinas hasta 1973 (p. 21).

Ninguno de los trabajos hechos para replicar la ley de Lotka ha sido comparado al trabajo original, quizá por la época y la comunidad de autores involucrados ([Potter](#), p. 36). Potter enfatiza que la ley de Lotka no es una distribución estadística precisa, al contrario, es una generalización basada en dos muestras (p. 23). La ley parece que trabaja mejor con un amplio y representativo subconjunto del universo de la comunidad de autores; cuanto más pequeño y más definido el subconjunto de documentos, se observan mejor los datos derivados de esta ley ([Potter](#), p. 36).

Ley de Zipf

La segunda ley empírica desarrollada en el campo de la bibliometría es la ley de Zipf. George Kingsley Zipf formuló la ley en 1933 para describir la relación entre el rango de las palabras y la frecuencia dada en una porción de la literatura.

$rf = c$ donde r = rango de la palabra.

f = frecuencia de la palabra y c es una constante para la porción dada (usualmente 1/10 del total del tamaño de la porción de la literatura) ([Wyllys](#), p. 54).

La versión generalizada de la ley es $rB f = c$, donde B es la pendiente de la línea de los datos (puntos).

El estudio original mostraba una mejor relación entre los rangos centrales que los del final y el corpus debería ser al menos de 5000 palabras para que rf sea constante ([Wyllys](#), p. 55).

En estudios sobre lenguaje natural, rf muestra un consistente crecimiento ligero según aumente r , en lugar de permanecer constante.

Benoit Mandelbrot propuso una modificación de la ley de Zipf la cual incluye otra constante que tiene el más grande efecto cuando r es pequeña $(r + M)Bf = C$. Este refinamiento de la ley de Zipf, provee un mejor ajuste a los datos típicos, especialmente en las categorías bajas y palabras de alta frecuencia ([Wyllys](#), p. 59). La ley de Zipf es aplicable a un diverso rango de fenómenos, pero su uso en sistemas de información está muy limitado, porque en el presente no da más información que los conteos de frecuencia de cada palabra ([Wyllys](#), p. 63).

Ley de Bradford

La mayor parte de la literatura sobre leyes empíricas bibliométricas está relacionada con la ley de dispersión de Bradford. S. C. Bradford fue un bibliógrafo británico, quien estableció a inicios de 1930 que menos de la mitad de los documentos científicos útiles publicados están reunidos en las publicaciones periódicas sobre resúmenes. Esta afirmación fue más tarde verificada por Derek de Solla Price en 1965: en cualquier año, el 35% de todos los documentos existentes no son citados del todo y 49% son citados sólo una vez, por diferentes razones, tales como: inadecuada indización y resumen y

no disponibilidad. La deficiencia en el control bibliográfico en esa época permitió a Bradford examinar la extensión que verdaderamente se le da a los artículos en las publicaciones periódicas dedicadas a diversos temas. Sus observaciones le permitieron su formulación de la ley de dispersión. Si las revistas científicas están hechas de manera que disminuye la productividad del artículo en una disciplina determinada, pueden ser divididos en un núcleo de publicaciones periódicas más particularmente dedicadas a la materia y varios grupos o zonas que contengan el mismo número de artículos como el núcleo, cuando los números de publicaciones periódicas en el núcleo y zonas exitosas sean como $n : n^2$... (White, p. 37).

Desde que la ley fue publicada en 1948, se ha hecho una gran cantidad de trabajo tratando de verificarla y probar su estabilidad; no hay dos de estos estudios subsiguientes que interpreten la ley en los mismos términos matemáticos.

En su artículo "**La ambigüedad de la ley de Bradford**", Elizabeth Wilkinson señalaba que Bradford formuló su ley gráfica y verbalmente, y las dos formulaciones no son matemáticamente equivalentes. La formulación verbal expresaba la teoría de Bradford mientras que la formulación gráfica describía sus datos observados (p. 125). B. C. Vickey demostró que la fórmula algebraica de la ley de Bradford sólo predice la porción superior de la línea recta de las curvas observadas (White, p. 37). Las dos formulaciones de la ley no convergen donde es más grande el número de documentos aportados por un publicista. Brookes desarrolló la formulación gráfica y encontró que sus datos seguían una forma logarítmica lineal (Wilkinson, p. 125). La porción inicial de la curva representaba el núcleo en la porción lineal, las presiones restrictivas que afectan el núcleo no operaron por más tiempo (Salton, p. 176). Leimkuhler ha publicado evidencias para la formulación verbal de la ley de Bradford; otros no han hecho mucho por esta ley porque el núcleo es difícil de definir. El encontró un acuerdo relativamente cercano entre las distribuciones observadas y las teóricas (Wilkinson, p. 125). Además, Wilkinson señalaba que para que la distribución de Bradford fuera válida todos los documentos que son relevantes deben estar incluidos en los datos, y los criterios para reconocer un documento relevante deberían ser consistentes y sintió que la ambigüedad de las conclusiones relativas a las dos formulaciones de la ley de Bradford han sido responsables de gran confusión e incertidumbre respecto a ella (p. 126).

Leyes teóricas

Pocos trabajos han sido hechos para encontrar las causas de los patrones descritos en las tres leyes, previamente analizadas sobre bibliometría. Se han hecho mayores esfuerzos para relacionar las leyes entre sí y aplicarlas. Mandelbrot desarrolló una ley teórica basada en un modelo de palabras almacenadas en la mente del ser humano. Este modelo establece que $f(r)$, el número de ocurrencias del $r(s)$ (rango de la palabra, es proporcional a $1/(1 + Br)^y$, donde B y Y son constantes (Brookstein, p. 420).

Brookstein usó un modelo simple de población para desarrollar la ley que podría ser aplicada en cualquier época, texto, etc. (Brookstein, p. 420). Su distribución teórica tomó la forma siguiente: el número de palabras que ocurre n veces es proporcional a $1/n$ para la misma constante. Si α puede tomar cualquier valor, la fórmula de Brookstein es equivalente a la de Mandelbrot. Para $\alpha = 1$ ó $\alpha = 2$, estas distribuciones son equivalentes a la suma de las leyes de Bradford-Zipf-Lotka (Brookstein, p. 421). Es más, las derivaciones empíricas de las leyes de Bradford, Zipf y Lotka son casos especiales de las leyes teóricas desarrolladas por Mandelbrot y Brookstein.

Aplicaciones Prácticas

Como se anotó con anterioridad, mucha de la literatura sobre bibliometría está relacionada con las aplicaciones técnicas a situaciones reales. Según White el análisis de citas hechas en índices, producto de sistemas automatizados, podrían ser realmente los más exitosos para hacer uso de la bibliometría (p. 35). El concepto de citar una fuente es casi tan antiguo como la imprenta. La primera cita a pie de página fue encontrada en "A Dictionary of the Art of Printing" por William Savage en 1841 (White, p. 38). El primer "citation index" conocido fue el

Shepherd's Citation en 1873. Hoy día, el análisis y la indización de citas son los aspectos básicos para investigaciones sobre bibliometría.

Los índices de citas ilustran el uso e impacto de las publicaciones periódicas, colecciones organizadas que muestran la frecuencia de documentos citados y pueden ser usados para plantear estudios históricos y sociológicos (White, p. 39).

La autora [Emile White](#) resume los conceptos básicos de análisis de citas en:

1. **Cita directa:** relación entre los documentos y el investigador que los usa.
2. **Acoplamiento bibliográfico:** la relación de dos documentos en virtud de su descenso a un tercer documento.
3. **Co-cita:** dos citas que son citadas juntas. (p. 39).

El tesoro de ERIC define el acoplamiento bibliográfico como la separación de una porción de literatura en pequeños grupos a través de correlaciones de conjuntos similares de referencia o citas bibliográficas. ([Ferrante](#), p. 202). Estos subgrupos forman redes de publicaciones relacionadas. Esta relación en el desarrollo de colecciones bibliotecológicas, es una relación fija, donde la co-cita puede cambiar dependiendo de los autores citados. La ley de Bradford puede ser usada para tales aplicaciones prácticas como estadísticas de circulación, composición de usuarios, la relación publicador-monografía para el desarrollo de una materia determinada. El número de citas puede ser usado para evaluar autores individuales y unidades académicas.

En 1973, [Saracevie y Perk](#) analizaron la literatura en bibliotecología utilizando la ley de Bradford, usando citas de la acumulación de 1967 del [Library Literature](#), ellos realizaron una serie de pruebas estadísticas sobre la selección de materiales de diversas fuentes. Ni el total ni la selección de la literatura siguió la distribución de Bradford; la distribución observada era demasiado fuerte en la cima de la curva y se desvanecía al final. Los autores concluyeron que la literatura sobre bibliotecología tiene un comportamiento fuertemente inesperado, está autocontenida y no es interactiva. (p. 121).

Conclusiones

La bibliometría está basada en tres leyes que se derivan como resultado de observaciones empíricas. Todas estas leyes son distribuciones que describen situaciones complejas, las distribuciones persisten en toda clase de circunstancia probada, y son casi equivalentes. Se han hecho muchos estudios sobre el uso de las leyes en aplicaciones prácticas en bibliotecología y ciencias de la información, aunque se ha dedicado atención en explicar las causas fundamentales de los modelos. Los investigadores todavía están examinando las leyes en nuevas situaciones; todos ellos están de acuerdo con que las distribuciones están presentes y son válidas en muchos y casi todos los casos, pero no saben por qué existen y trabajan.

Literatura Consultada

Bookstein, Abraham. The bibliometrics distributions / Abraham Bookstein. -- P. 416-423. -- [En](#) Library Quarterly. -- Vol. 46(Oct. 1976)

Buckland, Michael K. Library services in theory and context / Michael K. Buckland. -- New York : Pergamon Press, 1983.

Ferrante, Barbara Kopelock. Bibliometrics : access in the library literatura / B. K. Ferrante. -- P. 199-204. -- [En](#) Collection management. -- Vol. 2(Fall 1978)

O'connor, Daniel O. Empirical laws, theory construction, and bibliometrics / Daniel O. O'connor and Henry Voos. -- P. 9-20. -- [En](#) Library Trends. -- Vol. 30(Summer 1981)

Potter, William Gray. Lotkals law revisited / William Gray Potter. -- P.21-39. -- En Library Trends. -- Vol. 30(Summer 1981)

Pritchard, Alan. Statistical bibliography of bibliometries? / Alan Pritchard. -- P. 348-349- En Journal of Documentation. -- Vol. 25(Dec. 1969)

Saracevie, Tefko. Ascertaining activities in a subject area through bibliometric analysis / Tefko Saracevie and Lawrence J. Perk. -- P. 120-134. -- En Journal of the American Society for Information Science. -- Vol. 24(Mar.-Apr. 1973)

White, Emilie C. Bibliometries : from curiosity to convention / Emilie C. White. -- P. 35-42. -- En Special Libraries. -- Vol. 76(Winter 1985)

Wilkinson, Elizabeth A. The ambiguity of Bradford's law / E. A. Wilkinson. -- P. 122-130. -- En Journal of Documentation. -- Vol. 34(June 1972)

Witting, Glenn R. Statistical bibliography-A historical footnote / Glenn R. Witting. -- P. 240-241. - En Journal of Documentation. -- Vol. 34(Sept. 1978)

Wyllys, Ronald. Zipf's law / Ronald Wyllys. -- P. 53-64. -- En Library Trends. -- Vol. 30(Summer 1981)