

LA POPULARIDAD DE LAS ESPECIES. LAS DIEZ PRINCIPALES

Vidal Barrón* y Carlos Barrón†

Resumen: La base de datos Science Citation Index nos permite fácilmente obtener una medida de la popularidad científica de las especies animales, plantas y minerales. Utilizando en esta base de datos la estrategia de búsqueda en título, resumen y palabras clave hemos cuantificado cuántas veces ha sido incluido el nombre de una especie en los artículos publicados desde 1945 hasta nuestros días. La variación del número de citas con respecto al orden que ocupan en el ranking sigue una función altamente trascendente: $y = a x^{-b} e^{-cx}$. En este ranking, las 10 especies más citadas tienen todas un fuerte carácter antrópico.

Palabras clave: Science Citation Index, especies, funciones altamente trascendentes

Abstract: A measure of animal, plant and mineral species popularity can be easily obtained from the information available in the Science Citation Index. We used the title, abstract and keyword strategy in this index to find how many times the name of a species had been included in papers published since 1945 to the present. A good fitting of the variation of the number of citations with the rank was obtained with a linearizable and higher transcendental function $y = a x^{-b} e^{-cx}$. We also show the species «hit parade». The first ten species all have some anthropic character.

Keywords: Science Citation Index, species, higher transcendental functions.

Introducción

Más de un millón de animales y más de trescientas mil especies de plantas han sido, hasta la fecha, recopiladas usando distintas técnicas taxonómicas. La mayoría son sólo conocidas por su nombre científico y poco más. Alrededor de unas tres mil quinientas especies minerales han sido sistemáticamente estudiadas y descritas por los mineralogos, pero también, la mayoría sólo han tenido unos momentos de gloria muy efímeros. Dos cuestiones inmediatas se nos ocurren: ¿qué especies son las más populares? y ¿cuáles han recibido más titulares y chorros de tinta en la literatura científica?

Hoy es posible tener una estimación aproximada de la popularidad de las especies a partir de la información disponible en la base de datos Science Citation Index (SCI), que recopila el Instituto para la Información de la Ciencia, ISI, con sede en Filadelfia. La base de datos SCI recoge unas cinco mil novecientas revistas científicas (las mejores de todos los países) de cincuenta disciplinas científicas y dispone de referencias (más de 33 millones de citas) desde 1945 hasta nuestros días, incluyendo, también, resúmenes y palabras clave desde 1991. Aquí hemos usado la estrategia de búsqueda del título, resumen y pala-

* Departamento de Ciencias y Recursos Agrícolas y Forestales. Universidad de Córdoba. Correo-e: vidal@uco.es.

† I.E.S. Ángel de Saavedra, Córdoba, España.

Recibido: 3-11-04.

bras clave en el SCI para encontrar cuántas veces el nombre, en inglés, de una especie (en singular y plural) estaba incluido en los artículos científicos publicados desde 1945. Aunque esta búsqueda no es ni mucho menos exhaustiva, ni ofrece el número total de veces que una especie es nombrada en un artículo (puede haberse citado en el texto y no en el título, resumen o palabras clave), sí nos da una estimación relativa de la frecuencia con que ha sido estudiada una determinada especie. Utilizando este procedimiento hemos investigado más de trescientas especies de animales, plantas (incluyendo frutos) y minerales.

Distribución de las citas

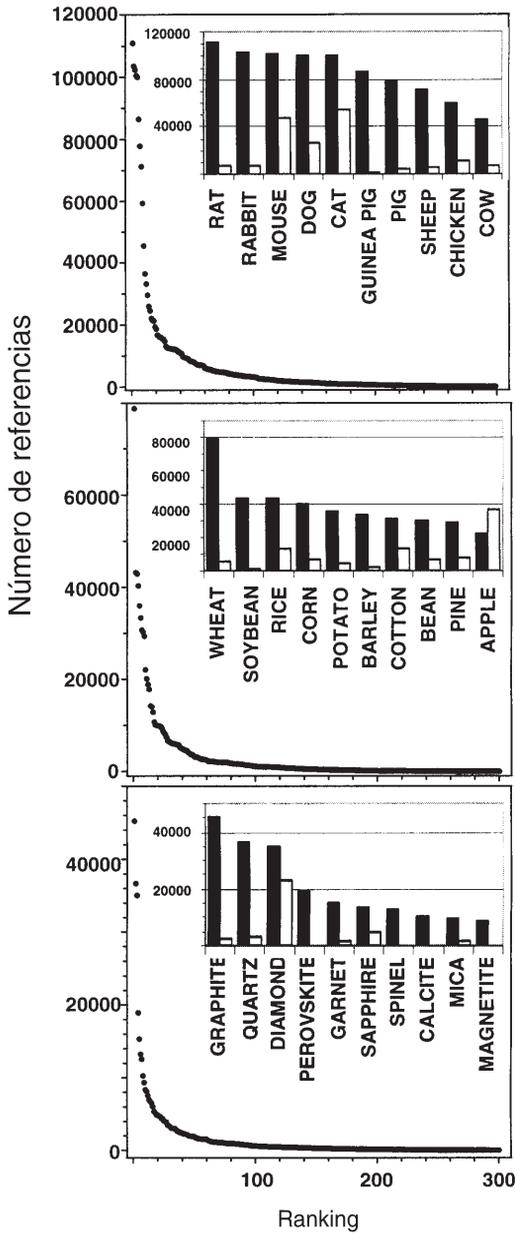
Como puede verse en la figura 1, la variación del número de citas de una especie (para los tres reinos) frente a la posición que ocupa en el ranking (el más citado es el número 1, el siguiente el número 2, y así, sucesivamente, hasta 300 en el ranking) sigue una tendencia hiperbólica antes que gaussiana. Este comportamiento no es sorprendente y puede ser interpretado a la luz del llamado «efecto Mateo» (1), por aquel pasaje del evangelio de San Mateo que dice: «Al que tiene se le dará... pero a quien no tiene, aun lo que tenga se le quitará». Este efecto ha sido estudiado también por otros autores denominándose «ventaja acumulativa» (2), esto es, unos pocos individuos monopolizan casi todo y la mayoría se reparten los restos, o en otras palabras (y otros campos) los grandes eventos son extremadamente raros mientras que los pequeños son altamente frecuentes.

Para describir las tendencias mostradas en la figura 1 la ecuación más utilizada ha sido una función potencial del tipo: $y = ax^{-b}$. Esta ley ha sido utilizada durante mucho tiempo para describir distribuciones similares, tales como: la cantidad de ingresos de dinero en economía (3), la productividad científica (4), la frecuencia de aparición de una determinada palabra en textos literarios en inglés (5), el tamaño de las ciudades; las reservas petrolíferas; la magnitud de los terremotos; la intensidad de luz emitida por las galaxias (6), y más recientemente, la frecuencia de acceso a las páginas web (7). Si una ley potencial se ajustara perfectamente a los datos, entonces debería observarse una línea recta cuando los representamos en una escala logarítmica. Sin embargo, se observa en muchos casos una significativa curvatura. Esto ha requerido la aplicación de ecuaciones empíricas alternativas, tales como la ecuación fractal parabólica (8), la exponencial extendida (6) o la ley de Lavalette (9), entre otras, para asegurar un mejor ajuste a los datos.

Las curvas sobre la popularidad de las especies exhibían un muy buen ajuste a las anteriores funciones, con coeficientes de determinación (R^2) mayores de 0,97 (el valor 1 corresponde al ajuste perfecto). Sin embargo, obtuvimos un mejor ajuste ($R^2 > 0,99$; figura 2) cuando utilizamos una ecuación altamente trascendente del tipo: $y = a x^{-b} e^{-cx}$, donde a está relacionado con la frecuencia para el número uno en el ranking, b indica la caída de popularidad para las siguientes especies y c está relacionado con la extensión de la cola para los últimos del ranking. La ventaja de esta ecuación es que puede ser fácilmente linealizable (las funciones de Lavalette y la exponencial extendida requieren para su ajuste un procedimiento iterativo y no derivable).

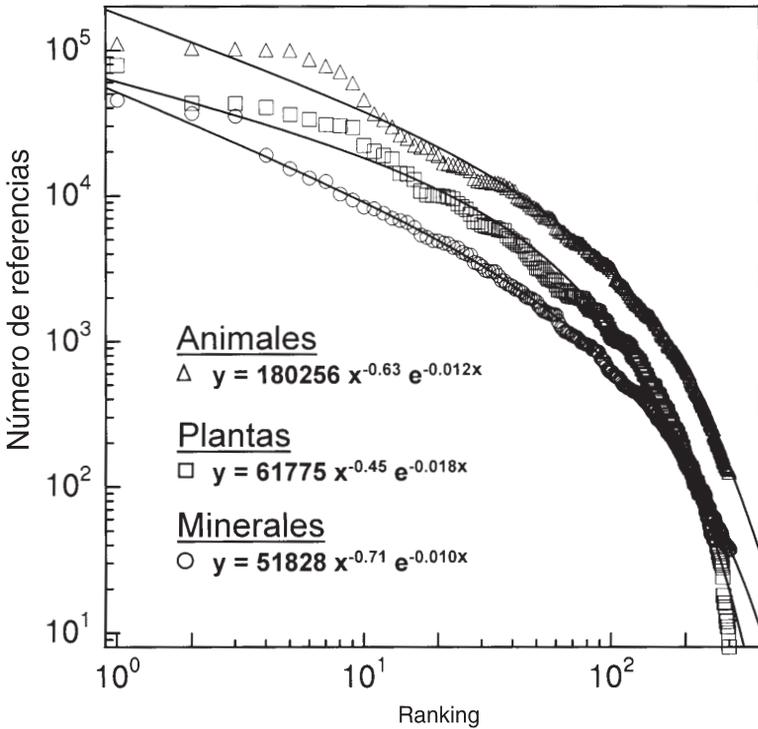
Por otro lado, la ecuación $y = a x^{-b} e^{-cx}$, puede considerarse como una función extendida de la ley potencial. A este respecto, la variación no gaussiana de la popularidad de las especies puede ser interpretada de forma similar a otras utilizadas en ciencias bibliométricas y cienciométricas (10) y puede ser considerada como una consecuencia de la «ley universal» propuesta por Stankov (11).

Figura 1
Distribución del número de referencias para 300 animales, plantas (incluido frutos)
y minerales



Nota: Se incluyen las 10 principales a partir del SCI (barras oscuras) y a partir del motor de búsqueda en internet Google (barras claras; × 10³).

Figura 2
Gráfico logarítmico de la distribución del número de referencias frente a la posición que ocupan en el ranking



Nota: Las tres ecuaciones se ajustan a los datos con coeficientes de determinación mayores de 0,99.

Las 10 principales

Finalmente, vamos a echar un vistazo al «hit parade», a «los 10 principales» de las especies. ¿Qué especies están más a menudo en las mentes de los científicos? En la figura 1 se muestran los 10 más citados para cada reino. Sorprende, tal vez, que no hayamos incluido la especie humana en la lista de los animales. De hecho, se han encontrado más de trescientas cuarenta mil citas (hombre, mujer, humanos), esto es, casi tres veces el número del siguiente, la rata, el primero que se muestra en el ranking (en menor extensión, esto también parece ocurrir para el caso del trigo en el ranking de las plantas). Esta anomalía en la distribución (se sale mucho del ajuste a la ecuación) es bien conocido como el «efecto rey»: cuando un nuevo rey es coronado, rápidamente elimina a sus rivales para asegurarse su reinado y quedarse con todas sus riquezas. Este efecto ha sido ya estudiado, por ejemplo, cuando se ha comparado el tamaño de ciudades de un país (París se sale de las previsiones cuando se compara con el número de habitantes de otras ciudades de Francia) (8). Pero el «efecto rey» va aún más lejos. Así, las primeras 10 especies, tienen todas un carácter antrópico, bien se utilizan como animales de laboratorio o se trata de animales y plantas agroganaderas.

Los mineralógos son, quizá, los más románticos (les encantan las piedras preciosas) y desinteresados. ¿Cómo si no explicar el cuarto lugar en el ranking para la perovskita? La mayoría de las referencias están dedicadas al $MgSiO_3$, un mineral con estructura de perovskita pero del cual no se tienen muestras naturales aunque se cree que es el más abundante en la Tierra, si bien se encuentra entre 670 y 2.900 km de profundidad de la superficie terrestre.

¿Comparte la gente las mismas preferencias que los científicos? Los datos obtenidos usando una búsqueda análoga para las mismas especies en el motor de búsqueda Google, muestra que no existe un buen acuerdo (figura 1). De nuevo los humanos, por supuesto, pero luego los gatos, la manzana y claramente los diamantes son en este caso los reyes de la web.

¿Hay algo más que añadir después de los miles de artículos dedicados a estas especies? ¿Se imagina que una prestigiosa editorial científica le encargue que lleve a cabo una extensa revisión sobre cualquiera de las más famosas especies?

Referencias

1. MERTON, R. The Matthew effect in science. *Science*, 1968, vol. 159, 56-63.
2. DE SOLLA PRICE. D. J. A general theory of bibliometric and other cumulative advantage process. *Journal of American Society for Information Science*, 1976, vol. 27, 292-306.
3. PARETO, V. Cours d'Economie Politique. Paris; Rouge et Cie, 1897.
4. LOTKA, A. J. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 1926, vol. 16, 317-323.
5. ZIPF, G.K. Human Behavior and the Principle of Least Effort. Cambridge; Addison Wesley Publishing, 1949.
6. LAHERRERE, J.; SORNETTE, D. Stretched exponential distributions in nature and economy: «fat tails» with characteristic scales. *European Physical Journal B*, 1998, vol. 2, 525-539.
7. LEVENE, M.; FENNER, T.; LOIZOU, G.; WHEELDON, R. A stochastic model for the evolution of the web. *Computer Networks*, 2002, vol. 39, 277-287.
8. LAHERRERE, J. C. Distributions de type «fractal parabolique» dans la nature. *Comptes Rendus de l'Académie des Sciences II*, 1996, vol. 322, 535-541.
9. POPESCU, I.I. On a Zipf's law extension to impact factors. *Glottometrics*, 2003, vol. 6, 83-93.
10. IVANCHEVA, L.E. The Non-gaussian nature of bibliometric and scientometric distributions: A new approach to interpretation. *Journal of the American Society for Information Science and Technology*, 2001, vol. 52, 1100-1105.
11. STANKOV, G. The Universal Law. A general Theory of Physics and Space Science Munich, New European Academy Press, 1999.