Letter to the Editor

Measures for Bibliometric Size, Impact,
and Concentration

Dear Sir,

Recently, Egghe (2014) suggested that the g-index (Egghe, 2006), normalized using the square root of the total number of citations C as $s = g/\lfloor \sqrt{C} \rfloor$ where $\lfloor * \rfloor$ denotes the floor function ($\lfloor x \rfloor$ is the largest integer smaller than or equal to x) is a "good" normalized measure of concentration. Rousseau (2014) in a thoughtful analysis showed that Egghe's s-measure is not an acceptable concentration measure and neither is the g-index or any other h-type index. Further, the new measure s no longer serves as a measure of impact.

It is meaningful to review all this in the light of the three-dimensional approach recently introduced by Prathap (2014).

Let c_k, $k = 1$ to P, represent the citation sequence of all P papers in a portfolio (Prathap, 2011). Note that the notation T is used instead of P for the total number of papers in Egghe (2014) and Rousseau (2014). Then $C = \Sigma c_k$, $k = 1$ to P is the total number of citations. We introduce a note of clarification here that k is used as the index of the citation sequence instead of i because of the historical legacy where the latter has served as the notation for impact. While P serves as a proxy for quantity or size of the academic effort in the portfolio, the impact $i = C/P$ is an empirical proxy for quality. P can be viewed as a performance indicator of the zeroth-order. Then, i and P become two orthogonal components of a three-dimensional performance evaluation protocol. $C = Pi$ can be considered a performance indicator of the first-order (Prathap, 2011). Prathap (2011) showed that it is possible to define second-order, energy-like terms $E = \Sigma c_i^2$ and $X = iC$. The product $X = iC = i^2 P$ becomes a higher-order measure. It is a robust second-order performance indicator (Prathap, 2011). Apart from X, the additional indicator defined by $E = \Sigma c_k^2$ also appears as a second-order indicator. The coexistence of X and E allows us to introduce a third attribute that is neither quantity nor quality. The simple ratio of X to E can be viewed as the third component of performance, namely, the consistency term $\eta = X/E$. Perfect consistency ($\eta = 1$, i.e., when $X = E$) is a case of absolutely uniform performance; that is, all papers in the set have the same number of citations, $c_k = c = i$. When the best work is concentrated in a very few papers of extraordinary impact, we have highly inconsistent performance; the inverse of consistency thus becomes a measure of concentration.

It is possible to show that η is related to other popular measures of concentration like Simpson's diversity index D (Simpson, 1949) or the Herfindahl–Hirschmann index (Herfindahl, 1950; Hirschman, 1964). The same logic that applies to bibliometric variation or skew also applies to ecological diversity, where species richness and abundance are governed by identical sequences (Jost, 2010) and Simpson's diversity index D is a measure of diversity and indirectly of the evenness ν of distribution of the species through the relation $D = P\nu$. A measure of market concentration, the Herfindahl–Hirschman Index (HHI) is used to describe the competitiveness of an industry by using the sum of the squares Σc_k^2 for $k = 1$, P, where the market share c_k is expressed as a fraction. The measure is essentially equivalent to the Simpson diversity index used in ecology in the manner described by $\eta = \nu = 1/(HHI \times P) = D/P$.

The h-index, as originally proposed (Hirsch, 2005), is a purely heuristic construction that is sensitive to the form of the citation distribution (described by consistency η), in addition to the normal bibliometric indicators which sense quantity (or size), namely, the number of papers P, and the quality (or impact i) as measured by the ratio C/P, where C is the total number of citations received by the P papers. The h-index, the g-index, and other h-type indices are actually heuristic constructions that try to condense P, i, and η into a single number that has the same dimensions as the number of papers P (Prathap, 2012).

References

Egghe, L. (2006). An improvement of the H-index: The G-index. ISSI Newsletter, 2(1), 8–9.

Egghe, L. (2014). A good normalized impact and concentration measure. Journal of the Association for Information Science and Technology, 65. doi:10.1002/asi.23168

Herfindahl, O.C. (1950). Concentration in the U.S. Steel Industry. Unpublished doctoral dissertation, New York: Columbia University.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102(46), 16569–16572.

Hirschman, A.O. (1964). The paternity of an index. American Economic Review (American Economic Association), 54(5), 761.

Jost, L. (2010). The relation between evenness and diversity. Diversity, 2, 207–232.

Prathap, G. (2011). The energy–exergy–entropy (or EEE) sequences in bibliometric assessment. Scientometrics, 87, 515–524.

Prathap, G. (2012). The thermodynamics-bibliometrics consilience and the meaning of h-type indices. Journal of the American Society for Information Science and Technology, 63, 430–430.

Prathap, G. (2014). Quantity, quality, and consistency as bibliometric indicators. Journal of the American Society for Information Science and Technology, 65, 214–214.

Rousseau, R. (2014). Egghe's g-index is not a proper concentration measure. Journal of the Association for Information Science and Technology. doi: 10.1002/asi.23276

Simpson, E.H. (1949). Measurement of diversity. Nature, 688, 163.

Gangan Prathap
CSIR National Institute for Interdisciplinary Science and Technology
Thiruvananthapuram,
India
695019
E-mail: gp@niist.res.in