

The Energy–Exergy–Entropy (or EEE) sequences in bibliometric assessment

Gangan Prathap

Received: 21 October 2010/Published online: 2 March 2011
© Akadémiai Kiadó, Budapest, Hungary 2011

Abstract Bibliometric research assessment has matured into a quantitative phase using more meaningful measures and analogies. In this paper, we propose a thermodynamic analogy and introduce what are called the energy, exergy and entropy terms associated with a bibliometric sequence. This can be displayed as time series (variation over time), or in event terms (variation as papers are published) and also in the form of phase diagrams (energy–exergy–entropy representations). It is exergy which is the most meaningful single number scalar indicator of a scientist’s performance while entropy then becomes a measure of the unevenness (disorder) of the publication portfolio.

Keywords Bibliometrics · $X = iC$ · Energy · Exergy · Entropy · $S = E - X$ · p -index

Introduction

Bibliometric research assessment is now in a quantitative phase where it is largely data and algorithm driven. Huge databases are available (Web of Science and Essential Science Indicators (a Thomson-Reuters product), Scopus, Google Scholar, SCImago (<http://www.scimagojr.com>), Eigenfactor.com, etc.) and powerful algorithmic and visualization techniques (eigenvalue, co-citation, factor analysis, etc.) have emerged.

In this paper, we define an energy like term which we shall call exergy, $X = iC$, where i is a measure of quality, expressed as the ratio of total citations C to total papers published P . The thermodynamic paradigm leads further to concepts of energy (E), and entropy (S). Also, it leads to time series and phase diagram representations which show the bibliometric progress of different scientists.

It is exergy which is a more meaningful, if not more accurate, single number scalar indicator of a scientist’s performance while entropy then becomes a measure of the unevenness (disorder) of the publication portfolio. Neither, P nor C (which are quantity measures), nor i (a quality measure), nor even the popular h -index, can serve this purpose.

G. Prathap (✉)

National Institute for Science Communication and Information Resources, New Delhi 110 012, India
e-mail: gp@emmacs.ernet.in

Being scalar quantities, E , X and S , can be added (or subtracted, as indeed $S = E - X$) sequentially and represented chronologically. These various terms can then be tracked as a time (or event) series showing the evolution of a scientist's career over time or over publication events.

The preliminaries towards a thermodynamic analogy

Here, we introduce the various terms and concepts. Consider that a person has published a single paper in a publication window (period over which papers are published). If the total number of papers is represented by P , then in this case, $P = 1$. Assume this paper, over a fixed citation window (period over which citations are counted) has collected c citations. The total number of citations during this citation window is designated by C , and in this case, $C = c$. We now define the energy e of the single paper as $e = C^2/P = (c^2/1) = c^2$. The basic or elementary unit of effort or energy is defined as the energy that a single paper gathering a single citation (measured over the citation window) possesses. The term $e = c^2$ can then be considered to have c^2 times the basic or elementary unit of effort or energy. It is the knowledge energy in a paper as measured over the citation window. We shall now take up a few structured exercises to enlarge upon this idea.

Exercise 1

Let us now assume that a second person has two papers, which have collected exactly c citations each. The total energy, i.e. the sum of the energies is $E = c^2 + c^2 = 2c^2$. But what is then the energies of the sum of the papers (we shall designate this by X for reasons which will become clear as we proceed)? If we invoke the C^2/P definition, this leads to $C = 2c$, $P = 2$ and $X = 2c^2$. Thus, in this thermodynamically "perfect" case, $E = X$.

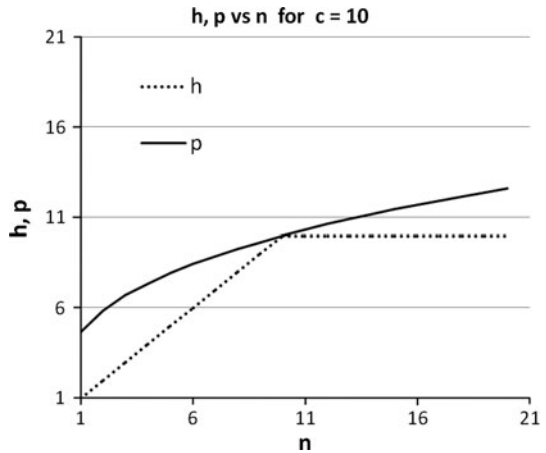
Exercise 2

We now examine a thermodynamically "imperfect" case where the two papers ($P = 2$) have collected c_1 and c_2 citations respectively. The individual energies of the papers are then $e_1 = (c_1)^2$ and $e_2 = (c_2)^2$. The total energy (defined as the sum of the individual energies) is, $E = (c_1)^2 + (c_2)^2$. We look next at the term we have called the total exergy X . To compute this, we need the sum of the citations, $C = c_1 + c_2$, and the sum of the papers ($P = 2$). The "energy" of the sums, i.e. the total exergy is, $X = (c_1 + c_2)^2/2$. We see at once that except in the case where $c_1 = c_2$, the total Energy will always be greater than the total Exergy. Let us define this discrepancy as the Entropy, which we designate by the term S . Then, $S = E - X$.

Exercise 3

Let us next consider yet another "perfect" case. An author has n papers during the publication window (this could be the lifetime of the scientist), all of which have collected c citations each during the citation window (again, this could be the lifetime of the scientist). Then the total number of papers published during the publication window is $P = n$. The energy of each paper is $e = c^2$. The total energy, i.e. the sum of the energies, is simply $E = nc^2$. The "energy" of the sum of the papers (which we have agreed to

Fig. 1 A portfolio, where the all papers gather 10 citations each. The h -index is initially limited by the number of papers published and when n crosses this threshold ($n = h = c$), h remains at the limiting value $h = c$. The p -index does not suffer from such a limitation



designate by the term X) called exergy, leads through the C^2/P definition, with $C = nc$ and $P = n$ to $X = nc^2$. Again, this is a thermodynamically “perfect” case with $E = X$, as the entropy S is zero. This is not unexpected. Knowing the number of citations of any one paper, that of all other papers is predictable, being exactly the same. Also, a p -index (Prathap 2010) could be defined as $p = X^{1/3}$, yielding $p = n^{1/3}c^{2/3}$. This is a formula that is first hinted at in Glanzel (2006) and developed further in Glanzel (2008) and Schubert and Glanzel (2007). However, the h -index (Hirsch 2005) depends on whether n is greater than or less than c . For all $n \leq c$, $h = n$ and for all $n > c$, $h = c$. Figure 1 shows a typical case, where the all papers gather 10 citations each. The h -index is initially limited by the number of papers published and when n crosses this threshold ($n = h = c$), h remains at the limiting value $h = c$. The p -index does not suffer from such a limitation. Note the significance of the curious conjunction at $n = h = p = c$. Here, we have the thermodynamically “perfect” portfolio of p papers published having p citations per paper so that $E = X = p^3$. The total energy is evenly distributed among all the states, i.e. p papers having p^2 units of bibliometric energy each.

A thermodynamical approach to bibliometric sequences

Consider what we shall call a portfolio of papers (this can belong to a scientist, or a group or any larger aggregation). This will be arranged as a bibliometric sequence. We take a very simple sequence, of 10 papers and their citations arranged in the chronological sequence as shown in Table 1. However, the usual bibliometric approach is to rearrange this in a ranked sequence so that the h -index can be easily found. This is displayed in Table 2 and by inspection one can see that $h = 6$.

We see from Table 2 that the 1st paper in this ranked list, that the author has published in 1976, has collected 14 citations. Standing alone, the energy of this single paper, which we shall designate by ΔE can be computed as $(\Delta c)^2$. Cumulatively these energies can be added up as shown in Table 3. Note that what we are computing are actual energy like terms. However, we must now contend with the quantity C^2/P which we noted had energy like properties and from which the p -index was derived. We shall call this the Exergy term, a sort of realizable energy, that modern thermodynamics has introduced (see below).

Table 1 Chronological sequence of 10 papers and their citations

Date	P	Δc	C
1975	1	0	0
1975	2	2	2
1976	3	5	7
1976	4	13	20
1976	5	14	34
1977	6	0	34
1977	7	6	40
1977	8	6	46
1977	9	7	53
1977	10	9	62

Table 2 Ranked sequence of 10 papers and their citations

Date	P	Δc	C
1976	1	14	14
1976	2	13	27
1977	3	9	36
1977	4	7	43
1977	5	6	49
1977	6	6	55
1976	7	5	60
1975	8	2	62
1975	9	0	62
1977	10	0	62

Table 3 Energies of individual papers added up cumulatively

Date	P	Δc	C	ΔE	E
1976	1	14	14	196	196
1976	2	13	27	169	365
1977	3	9	36	81	446
1977	4	7	43	49	495
1977	5	6	49	36	531
1977	6	6	55	36	567
1976	7	5	60	25	592
1975	8	2	62	4	596
1975	9	0	62	0	596
1977	10	0	62	0	596

We see from Table 4 that exergies slowly diverge from the total energies! This can be explained by noting that the bibliometric sequences are distributions showing a significant dispersion and that the greater this variation, the greater is the entropy (Table 5).

Table 4 Exergies are not the same as Energies!

Date	P	Δc	C	ΔE	E	X
1976	1	14	14	196	196	196.00
1976	2	13	27	169	365	364.50
1977	3	9	36	81	446	432.00
1977	4	7	43	49	495	462.25
1977	5	6	49	36	531	480.20
1977	6	6	55	36	567	504.17
1976	7	5	60	25	592	514.29
1975	8	2	62	4	596	480.50
1975	9	0	62	0	596	427.11
1977	10	0	62	0	596	384.40

Table 5 Energy, exergy and entropy of scientist's output

Date	P	Δc	C	ΔE	E	X	S
1976	1	14	14	196	196	196.00	0.00
1976	2	13	27	169	365	364.50	0.50
1977	3	9	36	81	446	432.00	14.00
1977	4	7	43	49	495	462.25	32.75
1977	5	6	49	36	531	480.20	50.80
1977	6	6	55	36	567	504.17	62.83
1976	7	5	60	25	592	514.29	77.71
1975	8	2	62	4	596	480.50	115.50
1975	9	0	62	0	596	427.11	168.89
1977	10	0	62	0	596	384.40	211.60

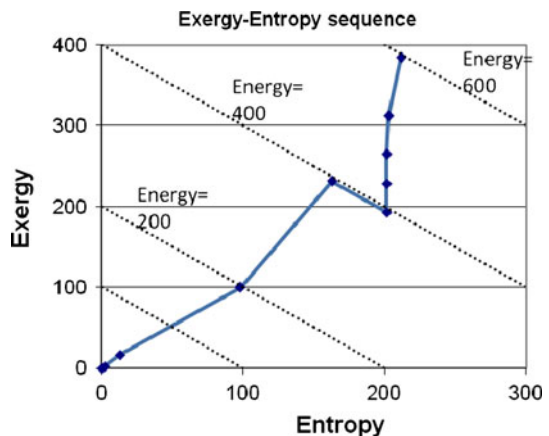
In a “perfect” sequence where every paper has the same number of citations, there is zero entropy, as in Exercise 3 above. Thus, we have a natural definition of entropy as the difference between energy and exergy ($S = E - X$). Indeed, once this thermodynamic paradigm is brought in, one need not use a ranked sequence but can actually use a chronologically ordered sequence as shown in Table 6. The Exergy–Entropy sequence of a scientist's output arranged chronologically, can be elegantly displayed in the form of the Energy–Exergy–Entropy diagram shown in Fig. 2.

Before proceeding to further applications of the idea, it will be worthwhile to review the thermodynamic understanding of the terms energy, exergy and entropy. Sciubba and Wall (2007) document that ‘at a scientific meeting in 1953, Zoran Rant suggested that the term exergy should be used to denote “technical working capacity”—energy literally means “internal work” (from the Greek en [$\epsilon\nu$] and ergon [$\epsilon\rho\gamma\omicron\nu$]), and the prefix ex [$\epsilon\zeta$] implies instead an “external” quantity.’

An excellent explication is available in Dincer and Cengel (2001). The definition of the terms entropy and exergy and their properties emerged from attempts to quantify the quality or “work potential” of energy in the light of the second law of thermodynamics. This was based on the realisation that it was not enough to state the principle of conservation of energy through the first law of thermodynamics to have a full understanding of

Table 6 Energy, exergy and entropy of scientist's output in chronological sequence

Date	P	Δc	C	ΔE	E	X	S
1975	1	0	0	0	0	0.00	0.00
1975	2	2	2	4	4	2.00	2.00
1976	3	5	7	25	29	16.33	12.67
1976	4	13	20	169	198	100.00	98.00
1976	5	14	34	196	394	231.20	162.80
1977	6	0	34	0	394	192.67	201.33
1977	7	6	40	36	430	228.57	201.43
1977	8	6	46	36	466	264.50	201.50
1977	9	7	53	49	515	312.11	202.89
1977	10	9	62	81	596	384.40	211.60

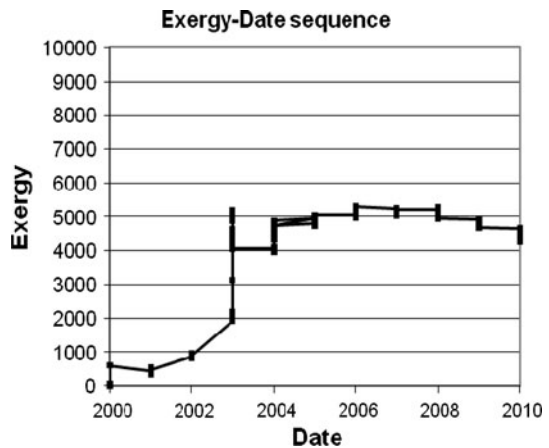
Fig. 2 The Exergy–Entropy sequence of a scientist's output arranged chronologically can be shown in an Energy–Exergy–Entropy diagram

such processes. According to the first law of thermodynamics, energy has quantity (and by implication 100% quality). Empirical evidence has shown (and so far never contradicted) that thermodynamical processes always occur in the direction of decreasing quality of conversion of energy (Dincer and Cengel 2001). Entropy and exergy were introduced as concepts to reconcile this. Indeed these definitions go beyond thermodynamics and can be applied to any process where quantity and quality need to be stated and measured (e.g. statistics, information theory, ecology, etc.). The bibliometric process is one such. In a bibliometric sequence, we have both quantity (the number of papers published) and quality (impact = ratio of citations and papers). Thus, it is not surprising that the concepts of energy, exergy and entropy can be stated in terms of P and C in the manner shown in this paper.

Next, we shall take up by way of fuller illustration, the portfolio of an active scientist and show the time-series, event-series and phase diagram representations of his bibliometric progress. The scientist has published 111 papers in the period 2000–2010 (Table 7 shows an abbreviated abstract of the complete bibliometric sequence).

Table 7 Energy, exergy and entropy of an active scientist's output in chronological sequence

Date	P	Δc	C	Energy E	Exergy X	Entropy S
2000	1	1	1	1	1.00	0.00
2000	2	12	13	145	84.50	60.50
2000	3	30	43	1045	616.33	428.67
2001	4	0	43	1045	462.25	582.75
2001	5	0	43	1045	369.80	675.20
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2010	107	0	686	16202	4398.09	11803.91
2010	108	0	686	16202	4357.37	11844.63
2010	109	0	686	16202	4317.39	11884.61
2010	110	0	686	16202	4278.15	11923.85
2011	111	0	686	16202	4239.60	11962.40

Fig. 3 The time-series variation of exergy (X) during the window 2000–2010 over which the scientist has published 111 papers

The time series representation

Figures 3 and 4 show how exergy and entropy vary with time over the period 2000–2010 during which the scientist has published 111 papers. Note the steep rise in activity in the early years—this is indicative of the fact that the early papers have collected more citations with time. The flattening of the curve shows either a decline in the scientist's activity and/or that more recent papers will need some time to receive their full complement of citations.

The event series representation

Figure 5 shows how exergy (a similar exercise can be carried out for entropy and energy) varies with event (marked by publication of each item in the portfolio) from the first event to the 111th event. Again, the steep rise in activity seen in the early years is indicative of the fact that the early papers have collected more citations over time. The flattening of the

Fig. 4 The time-series variation of entropy (S) during the window 2000–2010 over which the scientist has published 111 papers

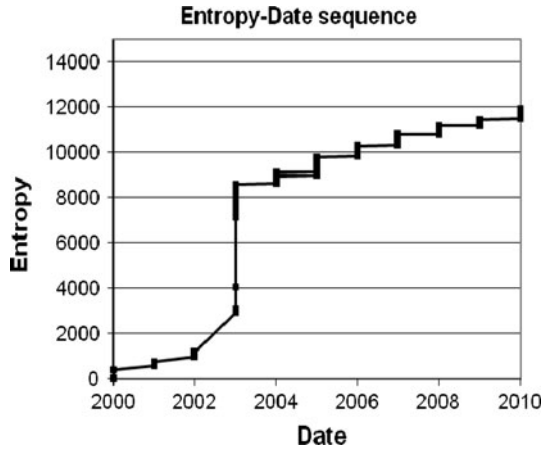
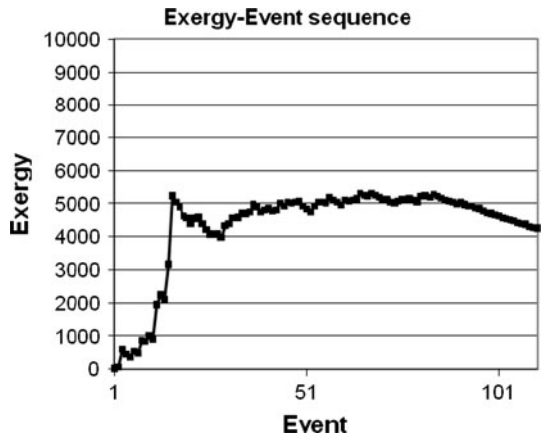


Fig. 5 The event-series variation of exergy (X) during the window 2000–2010 over which the scientist has published 111 papers



curve shows either a decline in the scientist's activity and/or that more recent papers will need some time to receive their full complement of citations.

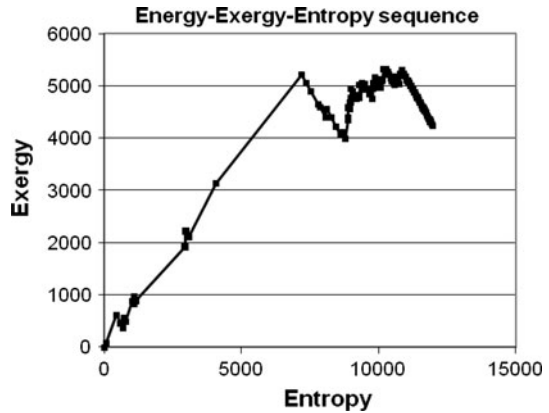
The phase diagram representation

Since energy = exergy + entropy ($E = X + S$), the bibliometric progress can also be represented as a phase diagram as shown in Fig. 6. As in Fig. 2 earlier, we see that energy (which is indestructible) and entropy never diminish and can only increase—in consonance with the classical laws of thermodynamics.

Concluding remarks

In this paper, we have proposed a thermodynamic analogy that can assess more meaningfully the bibliometric progress of a scientist. The energy e of the single paper is defined

Fig. 6 Since energy = exergy + entropy ($E = X + S$), the bibliometric progress can also be represented as a phase diagram as shown above



as $e = C^2/P = (c^2/1) = c^2$. The basic or elementary unit of effort or energy is defined as the energy that a single paper gathering a single citation possesses. The term $e = c^2$ can then be considered to have c^2 times the basic or elementary unit of effort or energy. It is the knowledge energy in a paper as measured over the citation window. Note that for a single paper, the unit of exergy is the same as the unit of energy, i.e. $x = e$. These distinctions were highlighted using a few structured exercises.

The definition of the terms entropy and exergy and its properties emerged from attempts to quantify the quality or “work potential” of energy in the light of the second law of thermodynamics (Dincer and Cengel 2001). Energy has quantity (and by implication 100% quality), but thermodynamical processes always occur in the direction of decreasing quality of conversion of energy (Dincer and Cengel 2001). In a bibliometric sequence, where we have both quantity (the number of papers published) and quality (impact = ratio of citations and papers), it is not surprising that the concepts of energy, exergy and entropy can be invoked in terms of P and C in the manner shown in this paper.

From this, for any portfolio of papers belonging to a bibliometric sequence, energy, exergy and entropy terms can be defined. A p -index, which has the same units as papers and citations, can be derived from the exergy term. The performance index (p -index) can effectively combine size and quality of scientific papers, mocking what the h -index could do, and emerges from the exergy term, $X = iC$, where i is a measure of quality, expressed as the ratio of citations C to papers published P .

Here, it is important to emphasise the practicality of these analogies. The analogy is not only an expression of “*consilience*” between the areas of thermodynamics and bibliometrics but goes beyond that in having a very practical value. The exergy X is a more meaningful, if not more accurate, single number scalar measure of a scientist’s performance. Neither P nor C (which are quantity measures), nor i (a quality measure), nor even the popular h -index, serves this purpose. Also, the entropy S becomes then a measure of the internal unevenness (disorder) of a scientist’s publication portfolio. These various terms can then be tracked as a time (or event) series showing the evolution of a scientist’s career over time or over publication events. This is possible because E , X and S , being scalar quantities, can be added sequentially and represented chronologically.

It is also instructive to appreciate that the formula for entropy given here is different from that given by Shannon for information theory. In Shannon’s definition, the logarithmic term comes from the assumption of an a priori distribution in the conventional probability interpretation. Here, no such assumption need be made. We are no longer

talking of probabilities of events. The internal bibliometric sequence (c_i , $i = 1$ to P) is known. The challenge is to represent this externally by a single number (based on C and P). Consequently, there is a loss of information. The definition of energy (from full knowledge of the internal structure), exergy (based on the externally available information) and entropy now comes from first principles and definitions, once the sequence is known.

Acknowledgments The author is grateful to anonymous referees for their valuable comments which have helped to improve the paper.

References

- Dincer, I., & Cengel, Y. A. (2001). Energy, entropy and exergy concepts and their roles in thermal engineering. *Entropy*, *3*, 116–149.
- Glanzel, W. (2006). On the h -index—A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, *67*, 315–321.
- Glanzel, W. (2008) On some new bibliometric applications of statistics related to the h -index. OR 0801, KU Leuven.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the USA*, *102*, 16569–16572.
- Prathap, G. (2010). Is there a place for a mock h -index? *Scientometrics*, *84*, 153–165.
- Schubert, A., & Glanzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics*, *1*, 179.
- Sciubba, E., & Wall, G. (2007). A brief commented history of exergy from the beginnings to 2004. *International Journal of Thermodynamics*, *10*, 1–26.