Characterizations of the generalized Wu- and Kosmulski-indices in Lotkaian systems

by

L. Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium
(*)
and
Universiteit Antwerpen (UA), Stadscampus, Venusstraat 35, B-2000 Antwerpen, Belgium
leo.egghe@uhasselt.be

___________________________________________________________________________

ABSTRACT

We define the generalized Wu- and Kosmulski-indices, allowing for general parameters of multiplication or exponentiation. We then present formulae for these generalized indices in a Lotkaian framework.

Next we characterise these indices in terms of their dependence on the quotient of the average number of items per source in the \( m \)-core divided by the overall average (\( m \) is any generalized Wu- or Kosmulski-index).

As a consequence of these results we show that the fraction of used items (used in the definition of \( m \)) in the \( m \)-core is independent of the parameter and equals one divided by the overall average.

(*) Permanent address

Key words and phrases: generalized Wu-index, generalized Kosmulski-index, characterization.
**Introduction**

The Hirsch-index (or $h$-index) is well known (Hirsch (2005)) and defined to be the largest rank $r = h$ such that all papers on ranks $1,\ldots, r$ received at least $r$ citations (here papers are ranked in decreasing order of the number of received citations).

Since the $h$-index can be applied to other source-item situations (other than source = paper and item = received citation) (see Egghe (2010) for a review on the $h$-index and other $h$-type-indices, up to (and including) 2008), we will henceforth use this more general source-item terminology.

There exist many papers describing advantages and disadvantages of the $h$-index (see again Egghe (2010) for a review). In this paper we discuss generalizations of two indices: the Wu-index (Wu (2010)) and the Kosmulski-index (Kosmulski (2006)).

The $w$-index of Wu (see Wu (2010)) is defined as the largest rank $r = w$ such that all sources on ranks $1,\ldots, r$ all have at least $10w$ items. With this index, Wu wants to focus on the sources with many items (or in Wu’s terminology: on the widely cited papers). Since the number $10$ is rather arbitrary, we replace it by the parameter $a \geq 1$, thereby generalizing the $w$-index to the $w_{a}$-index. This index was already introduced in van Eck and Waltman (2008). In the next section we will prove a formula for the $w_{a}$-index in the Lotkaian framework.

The $h^{(2)}$-index of Kosmulski (see Kosmulski (2006)) is defined as the largest rank $r = h^{(2)}$ such that all sources on ranks $1,\ldots, r$ all have at least $\left(h^{(2)}\right)^{2}$ items. With this index, Kosmulski wants to have an index similar to the $h$-index but requiring less ranked sources. Also here, the number $2$ is rather arbitrary. Therefore we replace it by $a \geq 1$, hence defining the generalized Kosmulski-index $h^{(a)}$. Since the notation $h^{(a)}$ is somewhat heavy (certainly in calculations) we will replace it by $h_{a}$. Also in the next section we will prove a formula for the $h_{a}$-index in the Lotkaian framework.
In Levitt and Thelwall (2007) (see also Egghe (2010)) one defines the “Hirsch $k$-frequency” $f(k)$ being the number of sources with at least $kh$ items ($h = h$-index). Hence, in our notation, $f\left( h^{a-1} \right)$ is the number of sources with at least $h^a$ items. So $f\left( h^{a-1} \right) = h_a$.

Reversely, for every $k > 1$, there exists an $a > 1$ such that $k = h^{a-1}$ (since $h > 1$), namely $a = \frac{\ln k}{\ln h} + 1$. This means that $h_a$ is equivalent with Levitt and Thelwall’s Hirsch $k$-frequencies.

The main disadvantage of the h-index is that it does not use the number of citations, above $h$, to the $h$ most cited papers. The (generalized) Wu- and Kosmulski-indices use more citations of the $w_a$ or $h_a$ most cited papers, but again, the citations above $aw_a$ (for $w_a$) and above $\left(h_a\right)^a$ are not used. Also, since for $a > 1$, $w_a$ and $h_a$ are smaller than $h$, these new indices use less cited papers which is considered in Wu (2010) and Kosmulski (2006) as an advantage, for calculatory reasons: a smaller list of papers in decreasing order of their received citations is needed.

In the third section we prove characterizations of the generalized Wu- and Kosmulski-indices. The results are as follows. Let us define the $r$-core as the set of sources on the first $r$ ranks. Denote by $\mu_r$ the average number of items in these $r$ sources and by $\mu$ the overall average number of items per source.

We prove that

$$ r = \frac{\mu_r}{a\mu} \tag{1} $$

if and only if $r = w_a$. Similarly, we prove

$$ r = \left( \frac{\mu_r}{\mu} \right)^{\frac{1}{a}} \tag{2} $$

if and only if $r = h_a$. Since for $a = 1$ we have $w_a = h_a = h$, the $h$-index, the above results reprove the characterization of the $h$-index, proved in Jin, Liang, Rousseau and Egghe (2007):

$$ r = \frac{\mu_r}{\mu} \tag{3} $$
if and only if \( r = h \) (note that \( \mu_h \) is denoted \( A \) in Jin et al. (2007) and called the \( A \)-index).

As a corollary of results (1) and (2) we prove that the fractions of used items of the items in the \( w_a \)- or \( h_a \)-core are independent of \( a \) ("used" means: used in the definition of the \( w_a \)-index and \( h_a \)-index: for the \( w_a \)-index we use \( aw_a \) items in the first \( w_a \) sources and for the \( h_a \)-index we use \( (h_a)^a \) items in the first \( h_a \) sources).

These fractions of used items are not only independent of \( a \); they are even equal for the \( w_a \)- and \( h_a \)-index, being \( \frac{1}{\mu} \).

The paper then closes with some conclusions and open problems.

We close this introductory section by repeating some results of Lotkaian informetrics that we need here. They can also be found in Egghe (2005) and Egghe and Rousseau (2006) (in which also a proof is given).

We consider the size-frequency function \( f \), where

\[
f(j) = \frac{C}{j^\alpha}
\]

\((C > 0, \ j \geq 1, \ \alpha > 1)\). Here \( f \) is a decreasing power law (the law of Lotka) being the density of the sources with item-density \( j \).

The total number \( T \) of sources equals (if \( \alpha > 1 \))

\[
T = \int_1^\infty f(j) \, dj = \frac{C}{\alpha - 1}
\]

and the total number of items equals (if \( \alpha > 2 \))

\[
A = \int_1^\infty j f(j) \, dj = \frac{C}{\alpha - 2}
\]
From this it is clear that $\mu$, the average number of items per source, equals (if $\alpha > 2$)

$$\mu = \frac{A}{T} = \frac{\alpha - 1}{\alpha - 2}$$

(7)

Further, Lotka’s law is equivalent with Zipf’s law:

$$g(r) = \frac{B}{r^\beta}$$

(8)

($B, \beta > 0, 0 < r \leq T$) and we have

$$B = \left( \frac{C}{\alpha - 1} \right)^{\frac{1}{\alpha - 1}} = T^{\frac{1}{\alpha - 1}}$$

(9)

and

$$\beta = \frac{1}{\alpha - 1}$$

(10)

Here $g(r)$ is the density of the items in source density $r$.

In terms of the function $g(r)$, $\mu_r$ (defined above) can be expressed as

$$\mu_r = \frac{1}{r} \int_0^r g(r')dr'$$

(11)

The generalized Wu-index and Kosmulski-index

**Definition 1:** The generalized Wu-index, denoted $w_a$, for $a \geq 1$ is the largest rank $r = w_a$ such that all sources on ranks $1, \ldots, r$ all have at least $aw_a$ items.
Note that \( w_a = h \), the \( h \)-index, for \( a = 1 \). We have the following formula for \( w_a \) in the Lotkaian framework.

**Proposition 1:** In the notation of the introductory section, we have, if \( \alpha > 1 \),

\[
w_a = \frac{1}{T^\alpha} a^{\frac{1}{\alpha}} = h a^{\frac{1}{\alpha}}
\]  
(12)

**Proof:** The proof is an extension of the one in Egghe and Rousseau (2006) where we proved

\[
h = \frac{1}{T^\alpha}
\]
(13)

, showing already that (11) and (12) are equivalent. Now, if \( \alpha > 1 \)

\[
\int_n^\infty f(j) dj = \frac{C}{\alpha - 1} n^{1-\alpha} = T n^{1-\alpha}
\]
(14)

is the total number of sources with item density larger than or equal to \( n \). Now replacing \( n \) by \( an \) yields the definition of the \( w_a \)-index: \( n = w_a \) for

\[
T( an )^{1-\alpha} = n
\]

Hence

\[
T a^{1-\alpha} = w_a^a
\]

, yielding (11). □

Note that \( w_a \leq h \) since \( a \geq 1 \) and \( \alpha > 1 \). If \( a > 1 \) then \( w_a < h \). In Wu (2010) one uses \( a = 10 \) and one finds \( h \approx 4w_{10} \). According to our model (12) this leads to

\[
\frac{1}{h 10^{\frac{1}{\alpha - 1}}} = \frac{h}{4}
\]

or
\[ \alpha = \frac{1}{1 - \log_{10} 4} = 2.5129416 \]

For the “classical” value \( \alpha = 2 \) we have

\[ w_{10} = h10^{-\frac{1}{2}} \]

or

\[ h = \sqrt{10}w_{10} = 3.1622777w_0 \]

In general we have for \( \alpha = 2 \)

\[ w_a = \frac{h}{\sqrt{a}} \]

**Definition 2:** The generalized Komulski-index, denoted \( h_a \), for \( a \geq 1 \) is the largest rank \( r = h_a \) such that all sources on ranks 1,...,\( r \) all have at least \( (h_a)^a \) items.

Note that \( h_a = h \) for \( a = 1 \). We have the following formula for \( h_a \) in the Lotkaian framework.

**Proposition 2:** In the notation of the introductory section, we have, if \( \alpha > 1 \),

\[ h_a = T^{\frac{1}{1-a(1-\alpha)}} \]  

(15)

**Proof:** Using again (14) (with \( n \) replaced by \( n^a \)) in the previous proof, we have that \( n = h_a \) if

\[ T\left(n^a\right)^{1-\alpha} = n \]  

(16)

Hence

\[Th_a^{\alpha(1-a)} = h_a \]
yielding (15). □

Note that, in (15), \( h_a = h = T^{\frac{1}{a}} \) for \( a = 1 \).

### Characterizations of the generalized Wu-index \( w_a \) and the generalized Kosmulski-index \( h_a \)

We have the following characterization of the generalized Wu-indices \( w_a \).

**Proposition 3:** Let \( \alpha > 2 \). The following assertions are equivalent:

(i) \( r = \frac{\mu_r}{a \mu} \)

(ii) \( r = w_a \)

with \( \mu_r \) and \( \mu \) as in the introductory section.

**Proof:** (ii) \( \Rightarrow \) (i)

We have to show that

\[
 w_a = \frac{\mu_{w_a}}{a \mu}
\]  

(17)

By definition of \( \mu_{w_a} \) we have (see (11))

\[
 \mu_{w_a} = \frac{1}{w_a} \int_{0}^{w_a} g(r) dr
\]  

(18)

where \( g(r) \) is Zipf's law (8).
Since $\alpha > 2$ we have that $0 < \beta < 1$ and hence

$$
\mu_{wa} = \frac{1}{w_a} B \frac{1}{1 - \beta} w_a^{1-\beta}
$$

$$
\mu_{w_a} = \frac{1}{w_a} \frac{\alpha - 1}{\alpha - 2} T^{\alpha - 1} w_a^{\frac{\alpha - 2}{\alpha - 1}}
$$

$$
\mu_{w_a} = \frac{1}{w_a} \mu T^{\alpha - 1} w_a^{\frac{\alpha - 2}{\alpha - 1}} \tag{19}
$$

by (7), (9) and (10). In order to prove (17) we have to show, by (19), that

$$
aw_a = \frac{1}{w_a} T^{\frac{1}{\alpha - 1} \frac{\alpha - 2}{\alpha - 1}}
$$

or

$$
a = T^{\frac{1}{\alpha - 1} \frac{\alpha - 1}{\alpha}}
$$

But (11) yields

$$
- \frac{\alpha}{\alpha - 1} T^{\frac{1}{\alpha - 1} \frac{\alpha - 1}{\alpha}} = \left( \frac{1}{T^{\alpha} a^{\frac{1}{\alpha}}} \right)^{\frac{\alpha}{\alpha - 1}} T^{\frac{1}{\alpha - 1}} = a
$$

(i) $\Rightarrow$ (ii)

The function

$$
r \rightarrow \frac{\mu}{a \mu} - r
$$

is strictly decreasing in $r$ (by definition of $\mu_r$ or just calculate the derivative of the function $\mu_r$)

$$
r \rightarrow \frac{1}{r} \int_0^{r'} g(r') dr'
$$

which is
$$\mu'_r = \frac{rg'(r) - \int_0^r g'(r') dr'}{r^3} < 0$$

since \( g'(r) \) strictly decreases).

Hence, since (i) supposes that there exists an \( r \) such that

$$\frac{\mu_r}{a\mu} - r = 0$$

then this \( r \) must be unique. But (17) implies that there is a solution in \( r = w'_a \). Hence \( r = w'_a \). □

We have the following characterization of the generalized Kosmulski-indices \( h_a \).

**Proposition 4:** Let \( \alpha > 2 \). The following assertions are equivalent:

(i) \( r = \left( \frac{\mu_r}{\mu} \right)^{\frac{1}{\alpha}} \)

(ii) \( r = h_a \)

with \( \mu_r \) and \( \mu \) as in the introductory section.

**Proof:** (ii) ⇒ (i)

We have to show that

$$h_a = \left( \frac{\mu_{h_a}}{\mu} \right)^{\frac{1}{\alpha}} \quad (20)$$

By definition of \( \mu_{h_a} \) we have (see (11))

$$\mu_{h_a} = \frac{1}{h_a} \int_0^{h_a} g(r) dr \quad (21)$$
As in the proof of Proposition 3 we have (see formula (19))

\[ \mu_{h_a} = \frac{1}{h_a} \mu T^{\alpha-1} h^{\alpha-1} \]  

(22)

In order to prove (20) we have to show, by (22), that

\[ \frac{1}{h_a} T^{\alpha-1} h^{\alpha-1} = h_a^a \]

or that

\[ \frac{a-2}{h_a^{\alpha-1}} \frac{1}{a-1} = T \]

(23)

But (15) yields

\[ \frac{a-2}{h_a^{\alpha-1}} \frac{1}{a-1} = T^{1-a[1-a]} \frac{(a-2-a(a-1)-(a-1))}{a-1} \]

\[ = T^{1-a[1-a]} \frac{-1+a-a^2}{a-1} \]

\[ = T^{1-a[1-a]} \frac{1}{a-1} \]

proving (23), hence (20).

(i) \( \Rightarrow \) (ii)

This proof follows the lines of the proof of (i) \( \Rightarrow \) (ii) in Proposition 3. \( \square \)

Both Propositions 3 and 4, for \( \alpha = 1 \), yield a new proof of the result proved in Jin et al. (2007) and described in Proposition 5 below (just take \( \alpha = 1 \) in Proposition 3 or Proposition 4).

**Proposition 5:** Let \( \alpha > 2 \). The following assertions are equivalent:

(i) \( r = \frac{\mu_i}{\mu} \)
(ii) \( r = h \)

The results in Proposition 3 and 4 imply that, for every \( a \geq 1 \), if \( \alpha > 2 \)

\[
w_a = \frac{\mu_{w_a}}{a \mu} \tag{24}
\]

and

\[
h_a = \left( \frac{\mu_{h_a}}{\mu} \right)^{\frac{1}{a}} \tag{25}
\]

We have the following consequences. Since

\[
\mu_{w_a} = \frac{1}{w_a} \int_{0}^{w_a} g(r)dr
\]

we have that (24) implies that

\[
\frac{aw_a^2}{\int_{0}^{w_a} g(r)dr} = \frac{1}{\mu} \tag{26}
\]

This formula can be interpreted as follows. The left-hand side of (26) is the fraction of used items (in the definition of the \( w_a \)-index) in the \( w_a \)-core. Indeed, \( aw_a^2 \) equals the minimum number \( aw_a \) of items in sources in the \( w_a \)-core times the \( w_a \) sources in the \( w_a \)-core and \( \int_{0}^{w_a} g(r)dr \) is the total number of items in the \( w_a \)-core. Then formula (26) says that this fraction is independent of \( a \) and equals \( \frac{1}{\mu} \).

If we look at (25) we see that
\[
\frac{h^{e+1}}{\int_0^{h_a} g(r)dr} = \frac{1}{\mu} \tag{27}
\]

using that

\[
\mu_{h_a} = \frac{1}{h_a} \int_0^{h_a} g(r)dr \tag{28}
\]

Again, the left-hand side of (27) is the fraction of used items (in the definition of the \( h_h \)-index) in the \( h_a \)-core. Indeed, \( h^{e+1}_a \) equals the minimum number \( h^n_a \) of items in sources in the \( h_a \)-core times the \( h_a \) sources in the \( h_a \)-core and \( \int_0^{h_a} g(r)dr \) is the total number of items in the \( h_a \)-core. Then formula (27) says that this fraction is independent of \( a \) and, again, equals \( \frac{1}{\mu} \).

We can conclude that the fractions of used items in the \( w_a \)- and \( h_a \)-core are not only independent of \( a \) but are equal for the generalized \( w_a \)- and \( h_a \)-indices, namely \( \frac{1}{\mu} \).

Note that this fraction, when taking the limit for \( \alpha \rightarrow +\infty \) is 1 using (7). So the larger \( \alpha \), the larger the fraction of the used items in the \( w_a \)- and \( h_a \)-core (in the limit being 1).

Note that, for \( \alpha \rightarrow +\infty \), \( w_a \rightarrow \frac{1}{a} \) (as follows from (11), since \( T \) and \( a \) are constant) and \( h_a \rightarrow 1 \), as follows from (15).

Results (26) and (27) contradict the (intuitive) feeling that the \( w_a \)- and \( h_a \)-indices use (relative) more items from the more productive sources, when \( a \) increases: the used fractions are the same for all \( w_a \)- and \( h_a \)-indices! In absolute terms, they even use less items for \( a \) increasing. This follows from (26) and (27) and since the denominators of the left-hand sides in (26) and (27) decrease for \( a \) increasing (since \( w_a \) and \( h_a \) decrease for \( a \) increasing - a logical fact which also follows from (12) and (15) and the fact that \( \alpha > 1 \)).
Concluding remarks

The Wu-index and Kosmulski-index have been generalized using a general parameter $a$. We then proved formulae for these indices in a Lotkaian framework. Then these measures are characterized using the quotient of the average number of items per source in a certain $r$-core ($r = \text{rank}$) (denoted $\mu_r$) and the overall average number of items per source.

As a corollary (for $a = 1$) we reproved the result of Jin et al. (2007), characterizing the $h$-index as the unique index $h$ such that

$$h = \frac{\mu_h}{\mu}$$

(29)

where $\mu_h$ is the average number of items per source in the $h$-core and where $\mu$ is the overall average number of items per source in the system (still supposing a Lotkaian framework).

In this connection we can make the following remarks.

It is clear that in any system (Lotkaian or not, continuous or discrete), $\frac{\mu_r}{\mu}$ decreases in $r$ and that

$$\lim_{r \to T} \frac{\mu_r}{\mu} = 1.$$ 

In the discrete case we define $m$ as the largest rank $r = m$ such that

$$\frac{\mu_m}{\mu} \geq m$$

(30)

This rank $r = m$ exists due to the above argument and since $r \geq 1$.

$m$ is a new impact measure and equals $h$ in the Lotkaian framework with Lotka exponent $\alpha > 2$.

Final Remark: for the $g$-index (see Egghe (2006)) result (26) (or (27)) is not true since, for the $g$-index, we use all items in all sources in the $g$-core (except, in the discrete case, possibly a few items in the source on rank $r = g$).
**References**


M. Kosmulski (2006). A new Hirsch-type index saves time and works equally well as the original h-index. ISSI Newsletter 2(3), 4-6.

