ISSI Newsletter

Quaterly e-Newsletter of the International Society for Scientometrics and Informetrics
ISSN 1998-5460

#58 / Volume 15 Number 2
June 2019

CONTENTS

ISSI ELECTIONS 2019: **A REPORT ON THE PROCEDURES AND RESULTS**

BALÁZS SCHLEMMER
election assistant

ISSI holds elections in every two years, when those board members whose 4-year-long mandate expire, step down. (In 2019 it applied to Board members Birger Larsen, Grant Lewison and Nees Jan van Eck, as well as to President Cassidy Sugimoto). The election procedure that follows consists of two parts: nomination and voting. All members in good standing (i.e. those who have paid their membership fees until the beginning of the nomination turn) had the right to take part in the elections. Members having the right to vote (except for those board members who did not step down this year) became automatically eligible candidates as well. Similarly to earlier elections (and considering the members' global whereabouts, which range to 41 countries at the moment), the elections were carried out online. Anonimity was guaranteed throughout the procedures. The detailed election rules were sent to all members concerned in e-mail.

ROUND 1: NOMINATION

The Nomination Round took place between 01 and 21 April.

25% of the members in good standing took part in the first round and they nominated 20 candidates for president, together with no less than 59 candidates for board members.

It occurred twice that nomination forms were sent repeatedly and one form arrived after the submission deadline. According to the election rules, the invalid nomination forms were ignored when counting the results.

After the closure of the nominating round the nominees were asked if they were willing to accept the nominations. See Fig 1 for acceptances and rejections.

ROUND 2: VOTING

The Voting took place between 14 and 31 May.

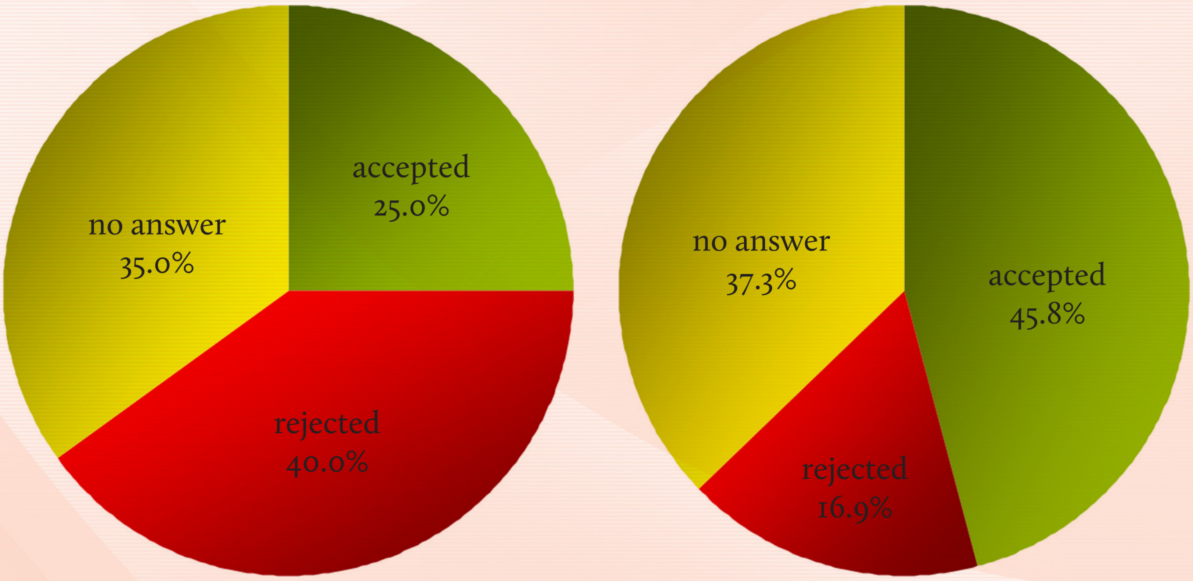

Figure 1 Acceptances and rejections of nominations for presidency (left) and board membership (right)

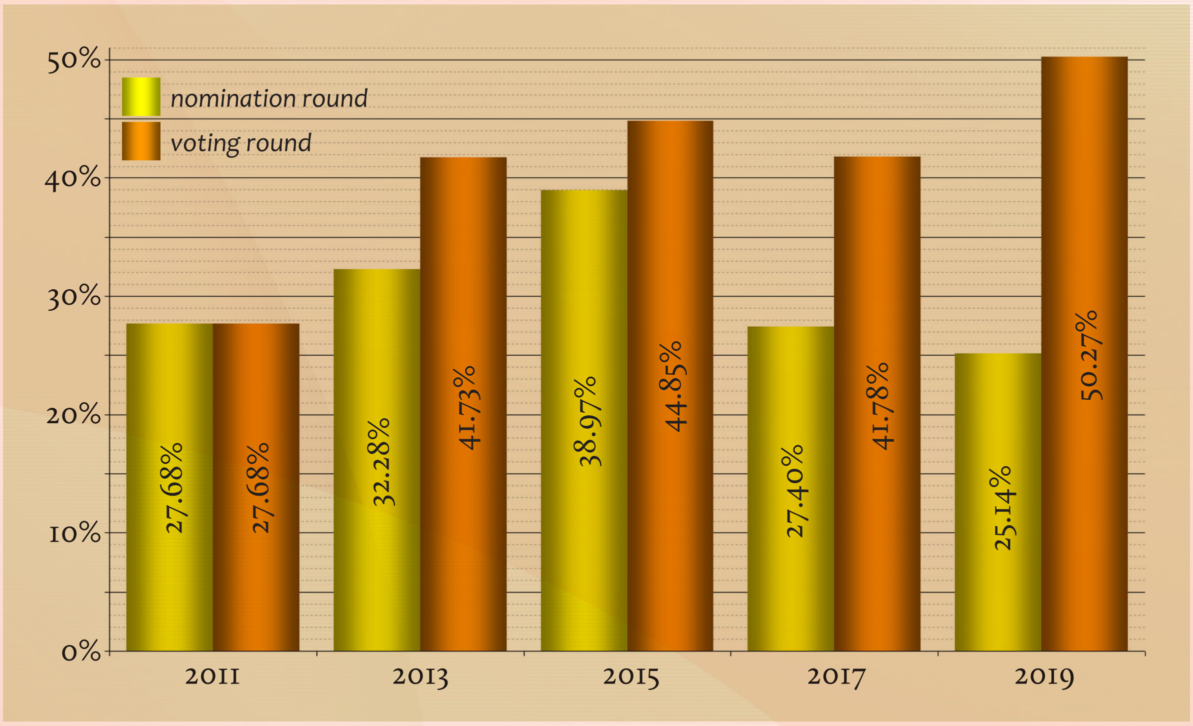

Figure 2 Nomination and voting turnout rates in the last five elections

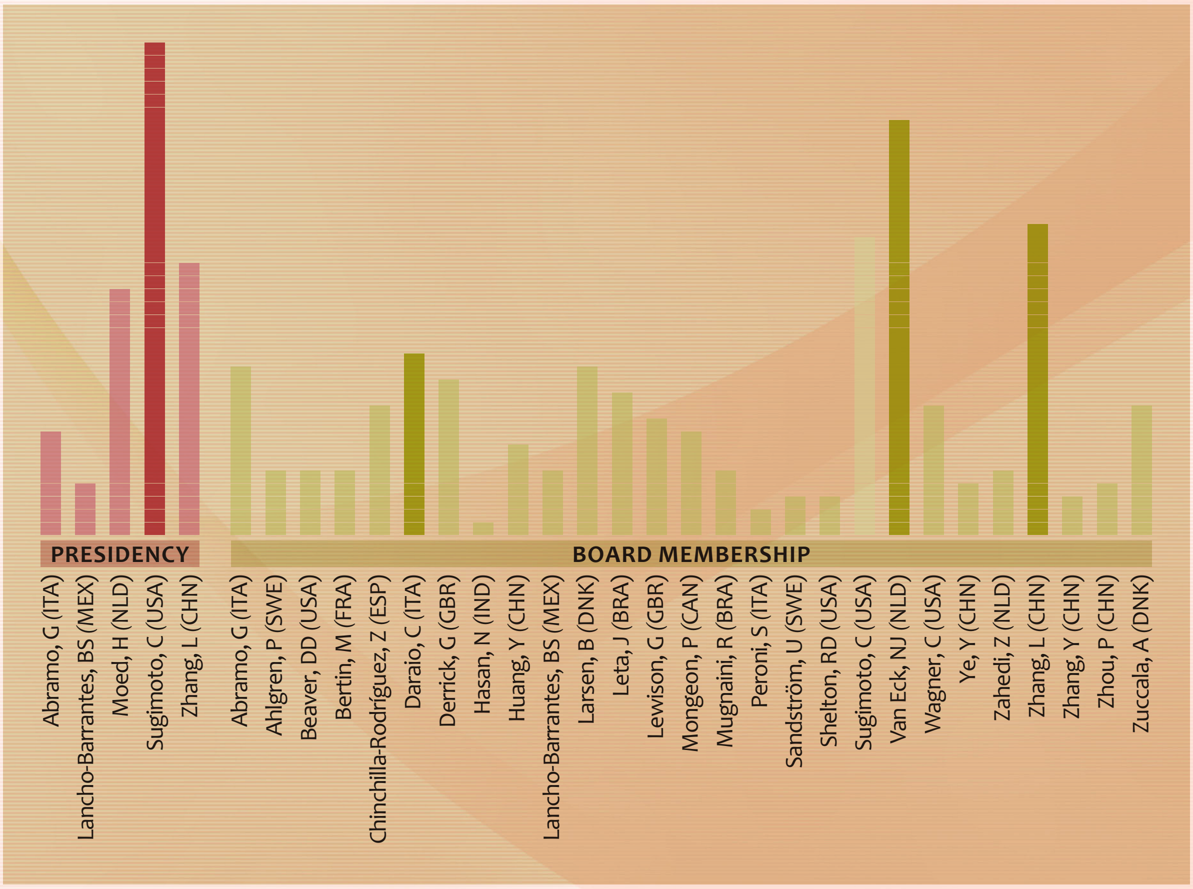

Figure 3 Results of the ISSI Elections in 2019

Exactly twice as many members took part in the second round as in the first one, which yielded the highest turnout (50.27%) in the last 10 years – counterbalancing nicely the last 10 years' lowest turnout in the nomination round (Fig 2).

It occurred twice that ballots were sent repeatedly. According to the election rules, only the last ballot sent by a particular voter was taken into account when counting the results. No invalid ballot has arrived.

RESULTS

After summing up the votes, the following results are hereby announced officially:

PRESIDENCY

Cassidy Sugimoto (USA) conveniently won by harvesting more than 40% of all votes. Runners-up were Lin Zhang (CHN) (23.3%), Henk Moed (NLD) (21.1%), Giovanni Abramo (ITA) (8.9%) and Bárbara S. Lancho-Barrantes (MEX) (4.4%).

BOARD MEMBERSHIP

Nees Jan van Eck (NLD) received 13% of all the votes, he was followed by Lin Zhang (CHN) (9.8%) and Cinzia Daraio (ITA) (5.7%). (Note: Cassidy Sugimoto's result (9.3%) was ignored, as she had been re-elected as president.) See Fig 3 for all the results.

THANKS!

Finally, on behalf of the Board of ISSI, I would like to say thanks to the Board members stepping down for the work they have done — as well as to the Board members who have just accepted their current nominations.

Last but not least, still on behalf of the whole Board, I would like to say thanks to all the members who participated in the 2019 Elections. See you in 2021 again!

ISSI LAUNCHES **QUANTITATIVE SCIENCE STUDIES**

VINCENT LARIVIÈRE

CASSIDY SUGIMOTO

LUDO WALTMAN

On January 10, the editorial board of the Elsevier-owned *Journal of Informetrics* (*JOI*) unanimously resigned. The resignation letter enumerated three main principles that were at odds with Elsevier ownership: (1) scholarly journals should be owned by the scholarly community rather than by commercial publishers, (2) journals should be open access under fair principles, and (3) publishers should make their citation data freely available. *JOI* had long served as a premier journal in the field of scientometrics and informetrics, populated with editors, reviewers, and authors from among the membership of the International Society of Scientometrics and Informetrics (ISSI). It was therefore a natural partnership for ISSI and the *JOI* editorial board to seek a new venue for the continued dissemination of high quality scientometric and informetric research.

Four days after the resignation of the *JOI* editorial board, ISSI launched a new journal, *Quantitative Science Studies* (*QSS*), constituted fully with the previous *JOI* editorial board. This journal is owned by ISSI and published jointly with the MIT Press. *QSS* is a journal run for and by the scientometric community. Funding for establishing and marketing the new journal has been provided in part by the MIT Libraries. APCs have been set at $600 per article for ISSI members and $800 per article for nonmembers. For the first three years of operation, however, APCs will be fully covered by the Technische Informationsbibliothek (TIB) – Leibniz Information Centre for Science and Technology (with support from the Communication, Information, Media Centre (KIM) of the University of Konstanz). The funds from TIB will be managed by the Fair Open Access Alliance (FOAA) to ensure that the journal is operating under fair open access principles. The MIT Press is also a full participant in the I4OC initiative, which promotes unrestricted availability of scholarly citation data. In these ways, the new partnerships are able to fully meet the criteria of the community: to provide a community owned journal, operating under fair open access principles, with fully available citation data.

News of the "flip" received international news coverage. There were dedicated reports in *Science, Nature, Inside Higher Ed, Times Higher Ed, Library Journal, Le Devoir, Science Guide, De Telegraaf,* and *Financieel Dagblad*. It was widely tweeted and continues to receive mentions major scientific newspapers (largely in relation to Plan S). It has been seen as a catalyst for other journals with multiple journals contacting MIT Press and FOAA to inquire about conducting their own flips from Elsevier and other large publishers. This is an important moment for our community: as experts in scholarly publishing and metrics, it is our obligation to model positive and responsible behaviors. This move to a community-owned and responsibly-managed journal is a critical signal to the wider community.

Shortly after the launch of QSS, the journal started to receive its first submissions. These submissions are currently under review and we are looking forward to a strong first issue in mid-2019. In addition, we have begun conversations with Clarivate and expect that *QSS* will soon be included in the Web of Science Emerging Sources Citation Index. We will continue conversations with other vendors to assure that *QSS* is widely indexed.

Vincent Larivière served as the Interim Editor-in-Chief from founding until March 31. Ludo Waltman assumed the editorship on April 1, after fulfilling his contractual obligations to Elsevier. The *QSS* board is now managed by Waltman as EiC and Larivière and Stasa Milojevic as Associate Editors. The journal will be accepting submissions through the ScholarOne system soon. Until this is configured, submissions can be sent individually to Ludo Waltman at qss@issi-society.org. For more information on submissions, please visit:

https://www.mitpressjournals.org/journals/qss/sub

This is the official journal of ISSI. Therefore, we welcome all of you to the *QSS* family: as readers, authors, reviewers, and editors. We look forward to working with all of you to make our journal a success. For the latest updates, follow us on Twitter @QSS_ISSI.

INTERNATIONAL FORUM ON QUANTITATIVE STUDIES OF SCIENCE AND TECHNOLOGY HOSTED AT ZHEJIANG UNIVERSITY

INSTITUTE OF INFORMATION RESOURCES MANAGEMENT OF THE SCHOOL OF PUBLIC AFFAIRS, ZHEJIANG UNIVERSITY

14 APRIL, 2019

CONFERENCE REPORT BY **PING ZHOU**

With a purpose of providing a platform for Chinese scholars to communicate with international counterparts in scientometrics and informetrics, the Institute of Information Resources Management of the School of Public Affairs at Zhejiang University hosted the International Forum on Quantitative Studies of Science and Technology on 14 April, 2019. Invited speakers included Prof. Dr. Cassidy R. Sugimoto from Indian University Bloomington (USA), Prof. Dr. Ludo Waltman and Dr. Vincent Traag from Leiden University (The Netherlands), Prof. Dr. Vincent Larivière from the University of Montreal (Canada), Prof. Dr. Li Tang from Fudan University (China), Prof. Dr. Cong Cao from University of Nottingham Ningbo (China), Prof. Dr. Feishu from Hangzhou Dianzi University (China), and Prof. Dr. Ping Zhou from Zhejiang University (China).

The eight speakers covered four topics including the challenges of global metrics, bibliometric studies of China, open access, and altmetrics. Prof. Sugimoto opened the forum with a talk on global indicators in scientometrics. She expressed concern that measurements of science are largely static and do not account for the growing complexity involved in the making of science. She discussed the development of new indicators for science that examine scientific mobility, collaboration, specialization, and scientific leadership, and the implications for measuring the growth of national production in science. She emphasized the anachronistic lens of "brain drain" and sought to provide a more accurate lens through notions of "circulation." She also discussed data and policy implications at the national, global, and disciplinary levels.

In view of the use of scientometric indicators often being criticized because of the loss of important information when complex scientific activities are reduced to just a few numbers, Prof. Ludo Waltman calls for contextualized scientometrics. He presented an alternative approach to scientometrics—contextualized scientometrics—where indicators are complemented with contextual information in a systematic way, and this information can be presented using interactive visualizations. By providing contextual information, numerical indicators no longer need to be used in isolation. Instead, it is always possible to see what is behind the numbers.

China was a focus of four speakers. Dr. Traag studied the time trends of national self-citations, with a focus on Europe, the US and China, and found that across many scientific disciplines, the US is increasingly more likely to make such national self-citations, while China is making fewer such national self-citations. It therefore seems that the US is becoming more insular and that China is becoming more internationally involved. However, it turns out that these trends are mostly the result of the scientific rise of China and the scientific decline of the US. In the end, Dr. Traag discussed the relevance of these results for the assessment of national scientific impact.

Prof. Cao and his coauthors focused on returning scientists and China's science system. China has made investments in sending people abroad, in building domestic capability, and in prioritizing advancements in targeted areas of science and technology. Given the scale and scope of these investments, and the unprecedented rise of a huge nation into the global science system, and the implications of this development for international affairs, Prof. Cao presented the results of research in the growth of Chinese scientific personnel, their international engagement, the number of overseas Chinese scientists and returnees, and the contribution of these researchers to Chinese science. China's human resources development emerges at an unprecedented time, when the science system is global in scope and international in operation. This means that China cannot use the same model of participation

that has characterized the rise of other nations over the past 40 years. Prof. Cao presented an estimate of number of Chinese scientists in the U.S. and Europe, as well as estimates of the number of returnees from these countries based on a bibliometric approach. He also discussed the impact of mobility on the growth of Chinese science.

Prof. Tang and her co-authors focused on the relationship between international mobility and local connections, based on Chang Jiang Scholars of China. Utilizing novel curriculum vitae data of 1447 Chang Jiang Scholars, they examined the relationship between academic mobility and the speed of obtaining prestigious academic titles. Their results suggest that local connections accelerate the career development of Chinese scholars whilst international academic mobility has a negligible effect or even slows down the speed of late-phase career advancement. Returnee scholars tend to obtain national academic titles within a longer time period compared to their local counterparts. This penalty of international academic mobility also holds for returnees with only overseas PhD training experience and international research visits. Local scientists are more likely than their returnee peers with equivalent ties to have a quicker career trajectory. Policy implications were discussed at the end.

Prof. Zhou and her co-authors studied government funding, international collaboration, and their interactions in academic research, and found that different countries vary in arrangement of government funding sources: China and Brazil are centralized to a unique agency (i.e., NSFC and CNPq), whereas the rest four countries are relatively decentralized. The six focal agencies namely NSFC, NSF, DFG, NWO, NRF and CNPq are more efficient than non-focal agencies as a whole in raising citation impact, with the NSF of the US performing the best. Not every country gets benefit from international collaboration in terms of raising citation impact of publications: developing countries (i.e., China and Brazil in the current study) benefit more than developed countries. Collaborating with developed countries, especially the US, can be a first option for choosing foreign partners. With regard to synergy between funding and international collaboration, DFG performs best in supporting high-impact international collaboration.

Open access was the focus of Prof. Larivière's talk. He first provided an overview of research indicators and of their limitations, and discuss some of the adverse effects of the most well-known indicator, the Impact Factor. He then addressed the current transformations of scholarly communication, with an emphasis on the role of journals and of open access in this new digital ecosystem, and concluded with implications for journals, funders, universities and researchers.

Prof. Shu's talk was about altmetrics. He investigated the effect of diffusing scientific articles on Twitter on their citation impact. For a set of 1.3 million papers covered by the Web of Science and published in 2012, normalized citation rates are compared between tweeted and non-tweeted articles published in the same journal. The results indicate that tweeted papers, published in the same year and in the same journal, received around 30% higher citation rates than papers, which were not diffused on Twitter. This type of citation advantage appeared across disciplines and countries of authors, but varied in extent.

Except the speakers, the Forum had attracted 58 participants from universities, research institutions and enterprises (i.e., Clarivate and Scopus). Most participants were from China. Both the speakers and participants spoke highly of the organization of the Forum. It provided a precious chance for Chinese scholars especially young scholars to communicate directly with leading international young fellows, helped international scholars understand more of Chinese counterparts, and provided a platform for Chinese and international scholars to seek possibilities of collaboration. In general, the output of the Forum was more than expected.

THE GOLDEN SECTION IN SCIENCE – AN INTERDISCIPLINARY CASE

WOLFGANG GLÄNZEL
ECOOM and Dept. MSI, KU Leuven, Belgium

SARAH HEEFFER
ECOOM and Dept. MSI, KU Leuven, Belgium

Abstract
Scientometrics and qualitative linguistics have traditionally a mutually fruitful relationship. The commonly used Zipf distribution applied to word-frequency ranking in linguistics and publication activity and citation impact in scientometrics might serve just as an example. Also the Waring model has been successfully applied in both fields. More recently, the h-index originally introduced by J.E. Hirsch to quantify an author's scientific research output has been adopted by Popescu and Altmann and applied along with some derivatives to the linguistics of text corpora in different languages. In our present study we reassume the Hirsch-Popescu model for further application in scientometrics. We also add some new parameters to this model and show that similar rules hold in both disciplines although we conclude that the golden ratio as defined by Popescu and Altmann is – in scientometrics – probably located outside the h-point.

1. INTRODUCTION

1.1. SCIENTOMETRICS AND GLOTTOMETRICS

Since text-mining techniques and lexical analyses have found their way into the scientometrics agenda, quantitative linguistics has become an increasingly important component of our field. At a first sight, their seems to be some similarities indeed, but in practice there are essential differences as well and that not only in their goal settings. Similarities and differences manifest in both conceptual and methodological issues. To begin with, scientometrics applies linguistics, that is, properties and characteristics of the use of natural language, mostly of English language, in order to detect similarities (or dissimilarities) between text of parts or the full body of different documents, where the individual properties of the language used in a single documents or by a single author are of lesser importance. This approach, that can be considered a "non-parameter" approach, results from the needs in our field, namely linking or clustering documents, or matching those based on their content or their language. The purpose ranges from topic detection and classification to issues related to intellectual property. However, little research has been done regarding a more "parametric" approach, namely the characterization of individual documents or the vocabulary of individual scientists as has been done in the classical field of quantitative linguistics based on quantitative laws and the quantification and analysis of language phenomena. Another consequence of these differences is that the first-mentioned approach unfolds its particular strength on the large scale, in a macro environment, when it comes to the analysis of large document spaces, while the second one displays it full advantages on the micro/nano scale as already been pointed to by Glänzel et al. (2017). This has also led to some first applications of "glottometrics" to our field (Glänzel et al., 2016; 2017).

In turn, scientometrics has fertilised the field of quantitative linguistics as well. Quantitative linguists have recently detected the usefulness of Hirsch's h-index in their field (Popescu and Altmann, 2006; 2007), notably in the context of characterising the vocabulary and language used by writers in their work, above all, in poems (Popescu et al., 2012; Pan et al., 2015).

Furthermore, both approaches use similar models for citation and word frequency distributions. The Zipf distribution and, more generally, Pareto-type distributions are the most commonly used models. The Waring distribution, one of the models successfully applied to publication activity, citation impact and other cumulative advantage processes and so-called success-breads-success phenomena in scientometrics, has been transferred to quantitative linguistics as well. Already Herdan (1964) has used the Waring distribution although he had some reservations about the goodness of fit at higher frequencies. Telcs et al. (1985) have applied this model, on the basis on a characterisation theorem, to some word frequency statistics of Macaulay's essay on Bacon and Pushkin's story *The Captain's Daughter*. About three decades later, Glänzel et al. (2016; 2017) analysed word frequency samples of András Schubert's scientometric oeuvre on the basis of this model.

Despite all similarity between the sciento- and the glottometric interpretation, there are some fundamental differences we need to discuss before we will have a closer look at some *geometric properties of the h-index* in both approaches. In scientometrics we deal with publications and citations, where we rank publications in descending order according to their citation they have received, we do the same with words and their use (the frequency of their occurrence) in a text for linguistic analysis. We just mention in passing that this can be done in scientific text as well. However, we have already pointed to one important difference regarding the notion of the sample. In the scientometric (publication–citation) model, the publication output of an author, author group or any research unit under study can be considered a sample the size of which can be increased by adding new publications. In linguistics, by contrast, the sample is formed by the vocabulary, increasing the length of the text will not automatically increase the sample size, at least not proportionally. In particular, the vocabulary is usually growing slower than the document's length and will finally not diverge, whereas the document length is theoretically unlimited (Glänzel et al., 2017). Another difference results from the nature of unused items. In scientometrics the number of uncited publications is known since they form part of the sample. In glottometrics the sample, i.e., the vocabulary does not comprise unused words. In this case we can distinguish between two types of unused words, those that are not used by the author/writer in the actual text but nevertheless used in other documents and those that are never used by the author/writer. Both can be estimated from the observed vocabulary using appropriate frequency-distribution models (cf. Telcs et al., 1985; Kornai, 2002). In order to synchronise methodology of both models, we can, however, simply restrict the scientometrics model to *cited* publications. If we apply this in the context of the h-index, this will not result in any loss of generality concerning the results since adding uncited items to the set under study will not have any effect on the h-index except for the trivial case if $h = 0$. In the following we will adopt the glottometric model and apply to the *geometry* of the scientometrics h-index. Before we do so, we still have to have a look at some important properties found in poetry, music and more generally in the arts.

1.2. THE GOLDEN SECTION AND FIBONACCI NUMBERS

The golden section or golden ratio stands for an ideal geometric relationship and has already been studied in ancient mathematics since Euclid or maybe even earlier. The golden ratio was and still is employed by artists, architects and composers to proportion their work. This can be deliberately designed but also done unconsciously as the golden ratio expresses a fundamental aspect of human aesthetics. Some manifestations of the application of this principle are quite obvious like those in Leonardo da Vinci's work (e.g., the *Vitruvian Man* and *Mona Lisa*) although there is no evidence that da Vinci has consciously used this principle in his work (cf. Livio, 2002). It was shown that composers have employed the golden ratio in different contexts. Larson (1978) reports the presence of the golden section in early western music, Lendvai (1966) detected its characteristic use in Bartók's musical themes and Putz (1995) found the golden section in the structure of Mozart's piano sonatas, namely in the relationship of *exposition* and *development & recapitulation*, where in contrast to Bartók, subjects of classical music usually follow rather symmetric patterns. The popularity of the golden section and its unconscious and even intuitive implementation in creative work is probably a consequence of its frequent occurrence in various manifestations in our immediate environment. The transformation of structural patterns in nature such as the arrangement of seeds in plants, crystals, human and animal anatomy, and others into visual art and architecture seems to be quite obvious, its application to more abstract artefacts like music or other human activities is less plausible. The fact that the role of the golden ratio is not restricted to aesthetic perception and its application to artwork alone is shown by its relationship with Fibonacci sequences through which its role broadens to other phenomena such as financial market, game theory and, more generally, stochastic processes (e.g., Glover et al., 2013).

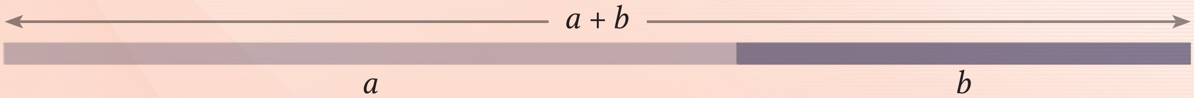

Figure 1 Visualisation of the golden section

The best way to understand the variety of applications of the golden ratio is to consider the definition of the golden ratio and its Fibonacci relationship. First, we recall the classical definition of the golden ratio. A line is subdivided into two parts (see Figure 1). If the ratio of the larger and the smaller part (a/b) equals the ratio of the total and the larger part ($(a+b)/a$), then these ratios are called golden ratio and denoted by ϕ. Consequently, we then have $\phi = 1 + 1 / \phi$. It is straightforward that ϕ is the (positive) solution of the quadratic equation $\phi^2 - \phi - 1 = 0$, particularly $\phi = ½(1 + \sqrt{5}) \approx 1.618$.

The strong relationship with the Fibonacci sequence reveals itself only at closer investigation. The Fibonacci sequence $\{f_n\}_{n \geq 0}$ is formed by the following recursion: $f_n = n$, if $n \leq 1$, and $f_n = f_{n-1} + f_{n-2}$, for all $n > 1$.

Hence, we have $f_n / f_{n-1} = 1 + f_{n-2} / f_{n-1}$ and, as n tends to infinity, we obtain $\phi = 1 + 1/\phi$ for $\phi := \lim_{n\to\infty} f_n / f_{n-1}$, where ϕ is the golden ratio. Furthermore, ϕ can be used to formulate a closed-form expression for the Fibonacci numbers. It is not difficult to derive the following formula:

$$f_n = 1/\sqrt{5} \cdot [\phi^n - (-1/\phi)^n]$$

The quintessence of the Fibonacci property, namely that a future value of something is directly determined by the present and the immediate past one, has certainly a strong effect on its widespread occurrence and applicability in so many fields and contexts.

2. THE GOLDEN RATIO AND THE H-INDEX

After the short discourse on the Golden Ratio and its appearance in even unexpected fields and contexts, we would like to shed light on its link with scientometrics, in general, and the h-index, in particular. At the first sight, any relationship between this ratio and a scientometric indicator seems to be beyond the imaginable. However, the h-index is not just a single number, it covers, in fact, a much more complex reality. For instance, the particular properties of the *h-core*, which is uniquely defined by the h-index, reveal much more relevant information than the index itself. Reaching out beyond citations, the characteristic properties of the h-index and h-core in the context of *networks* has been analysed, for instance, by Schubert et al. (2009) and Glänzel (2012). Moreover, we find a complete geometry behind the h-index, if we intend to obtain a more complete picture. This has been actually detected by our colleagues in quantitative linguistics. We have already pointed to some similarities and differences between the 'scientometric' and the 'glottometric' approach in Section 1.1. Most notably, we know the number of uncited papers but, by contrast, we do not have sufficient information about a writer's unused vocabulary. To avoid adjusting the glottometric model to cover also uncited papers, we decided to synchronise the model by dropping the uncited papers. Usually we would not do so because many of those will become cited as well earlier or later, but in the calculation of the actual h-index these papers do not have any effect.

In order to translate the glottometric approach to scientometrics, we draw an author's h-graph based on Figure 1 in Popescu et al. (2012), which can be considered an extended version of the h-graph provided, for instance, by Scopus. In Figure 2, we have plotted all citations of a selected author with 223 cited papers till June 2019 according to Clarivate Analytics Web of Science Core Collection (WoS). The total amount of citations is immaterial, the maximum number of citations (m) amounted to 359, the h-index was exactly 50. The graph plots the rank (r) of papers versus its frequency (f). We obtain three characteristic points (r, r), namely, (1, m) representing the paper with the highest rank, (h, h) the point representing the h-index and, finally, (v, 1) the point, which stands for the paper ranking last, where we assume that the number of citations will equal 1. If the number of citations received by the least cited but not uncited paper is larger, a marginal modification of the model could adjust for this. However, in most cases $f_v = 1$ will be observed, indeed. The notation "v" for the least rank has been adopted from the glottometric model, where it stands for the *vocabulary*, that is, the total number of words used. These three points define several angles and triangles.

The blue line running through the points (1, 1) and (h, h) in Figure 2 bisects the positive quadrant of the coordinate system and defines several angles and line segments. The

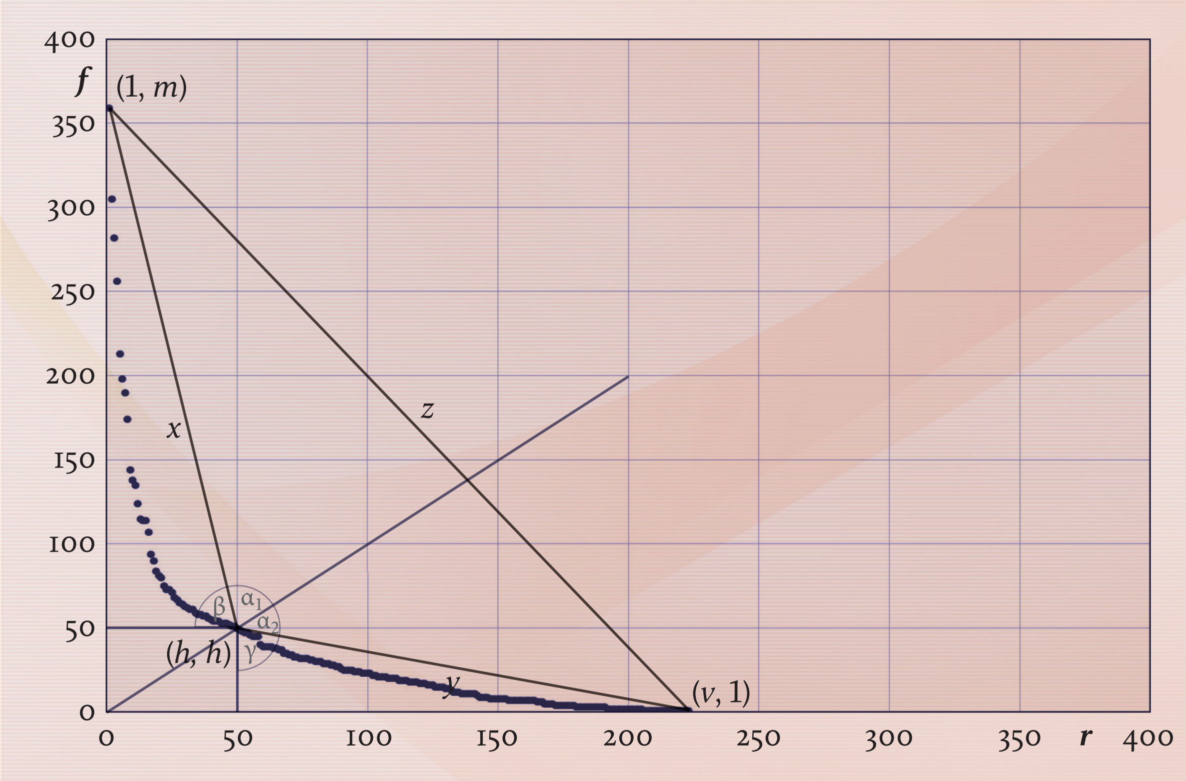

Figure 2 The characteristic triangle of the rank frequency distribution according to Popescu and Altmann (2007).

three characteristic points also form the vertices of a triangle, the sides of which are denoted by x, y and z (cf. Figure 1). Its internal angle divided by the blue line is denoted by α, its two parts by α_1 and α_2 with $\alpha = \alpha_1 + \alpha_2$. Two other external angles (β and γ) are defined by the square with the vertices (1, 1), (1, h), (h, h) and (h, 1), and the edges x and y. These three angles will play an important part in the geometry of the h-index. Popescu and Altmann (2007) called the α angle metaphorically "writer's view" expressing the balance between autosemantic and synsemantic, in scientometrics terms one could say, between significant and supporting work. The "view angle" ranges between $\pi/2$ and π. The most striking observations they made is that this angle converges to the Golden Ratio with growing length of the texts (n). For some mixed texts with n ranging between 30,000 (Goethe: Faust 1) and 790,000 (Old and New Testament), they found values ranging between $\alpha = 1.588$ (Joyce: Ulysses) and $\alpha = 1.656$ (Goethe: Faust 1). Joyce had also the largest vocabulary ($v = 29{,}457$) in the selection. The English translation of Homer and parts of the Bible were very close to the value $\phi \approx 1.618$ (cf. Popescu and Altmann, 2007). Popescu et al. (2012) and Tuzzi et al. (2010) have found similar trends for Slovak poetry and for Italian texts. Furthermore, Tuzzi et al. found the following relationship for the Italian case: $\alpha \approx 1.618 + 8.710/\sqrt{n} = \phi + 8.710/\sqrt{n}$ with $r^2 = 0.833$, where the expression on the r.h.s. depending on the text length n indicates the convergence speed. By contrast, Pan et al. (2015) could not fully confirm this observation for Chinese poetry: They have found that the golden ratio of the selected Chinese poetry is located outside the h-point. Nevertheless, these interesting findings are largely independent of the language used since only part of the analysed work was in English translation. Popescu et al. (2012) noticed in the context of the convergence to the golden ratio that "the writer does not strive for it and does not even know that it exists or, if so, how to attain it."

The chart in Figure 2 provides the basis of even more interesting geometrical properties. Before we study those, we briefly discuss the basics, which we have already used above, more in detail. The characteristic triangle in Figure 2 directly results in the following solution (cf. Popescu and Altmann, 2007).

$$\cos\alpha = -(z^2 - x^2 - y^2)/2xy,$$

where $x^2 = (h-1)^2 + (m-h)^2$ and
$y^2 = (h-1)^2 + (v-h)^2$

(1)

$$\cos\alpha = -\frac{(h-1)\left[(m-h)+(v-h)\right]}{\sqrt{\left[(h-1)^2+(m-h)^2\right]\left[(h-1)^2+(v-h)^2\right]}}$$

or

(1*)

$$\cos\alpha = -\frac{\frac{m-h}{h-1}+\frac{v-h}{h-1}}{\sqrt{\left(1+\left[\frac{m-h}{h-1}\right]^2\right)\left(1+\left[\frac{v-h}{h-1}\right]^2\right)}}, \text{ if } h>1.$$

In addition to these results by Popescu and Altmann, we will also give a solution for α_1, α_2 and the external triangles β and γ for the case that $h > 1$. The following properties are then straightforward:

$$\tan\beta = (m-h)/(h-1) \text{ and}$$
$$\tan\gamma = (v-h)/(h-1).$$

This results in the following solution.

$$\tan\alpha_1 = \tan(3/4\pi - \beta) =$$
$$= -(1+\tan\beta)/(1-\tan\beta) =$$
$$= (m-1)/[(m-h)-(h-1)]$$

$$\tan\alpha_2 = \tan(3/4\pi - \gamma) =$$
$$= -(1+\tan\gamma)/(1-\tan\gamma) =$$
$$= (v-1)/[(v-h)-(h-1)]$$

(2) $$\cos\alpha = \cos(\alpha_1+\alpha_2) =$$

$$= -\frac{\tan\beta+\tan\gamma}{\sqrt{(1+\tan^2\beta)(1+\tan^2\gamma)}} =$$

$$= -(\sin\beta\cos\gamma + \sin\gamma\cos\beta)$$

Obviously, the following equivalence relations hold: $\alpha_1 = \alpha_2 \Leftrightarrow \beta = \gamma \Leftrightarrow m = v$. For $h = 1$, we obtain the trivial cases $\beta = \gamma = 0$ and $\alpha_1 = \alpha_2 = \pi/4$.

Note that none of the above relations depends on the text length n or, in the scientometric equivalent, on the total number of citations received.

Coming back to the scientometric example now, we need only to use the three values $h = 50$, $v = 223$ and $m = 359$ to determine all relevant angles. The view angle of $\alpha = 2.004$ we obtain from these three values is larger than the expected value of 1.618, while the equivalent of the text length (i.e., the number of citations) cannot be considered small (n = 7,821). Of course, the question arises whether this case is representative or, if just in the case of Chinese poetry, also here the golden section is located somewhere outside the h-point. Quite a number of samples out of the 176 texts from 20 different languages discussed in Table 1 in Popescu and Altmann (2007) show similar deviations. Furthermore, we have $\alpha_1 = 0.943$ and $\alpha_2 = 1.061$, that is, the characteristic triangle is not isosceles ($x > y$) and thus cannot be a golden triangle either, which would otherwise be an isosceles triangle with the largest internal angle $\alpha = 1.885 \cong 108°$.

Popescu and Altmann (2007) analysed a further geometric property of the characteristic triangle. This regards the *area* of the characteristic triangle with respect to maximum possible area defined by the points (1, 1), $(m-1)$ and $(v-1)$. Unlike Popescu et al. (2012), who used an analytical approach, we give an elementary solution (see Figure 3). In particular, we have $A_{max} = A_0 + A_\beta + A_\gamma + A_h$ with

$$A_0 = (h-1)^2$$

$$A_\beta = (h-1)(m-h)/2$$

$$A_\gamma = (h-1)(v-h)/2$$

$$A_{max} = (m-1)(v-1)/2$$

Hence, we obtain

(3)

$$A = \frac{A_h}{A_{max}} = 1 - \frac{h-1}{m-1} - \frac{h-1}{v-1}, \text{ if } m \text{ and } n > 1.$$

According to Popescu et al. A is related to the use of the given vocabulary. A ranges between 0, if h is located on side z and the

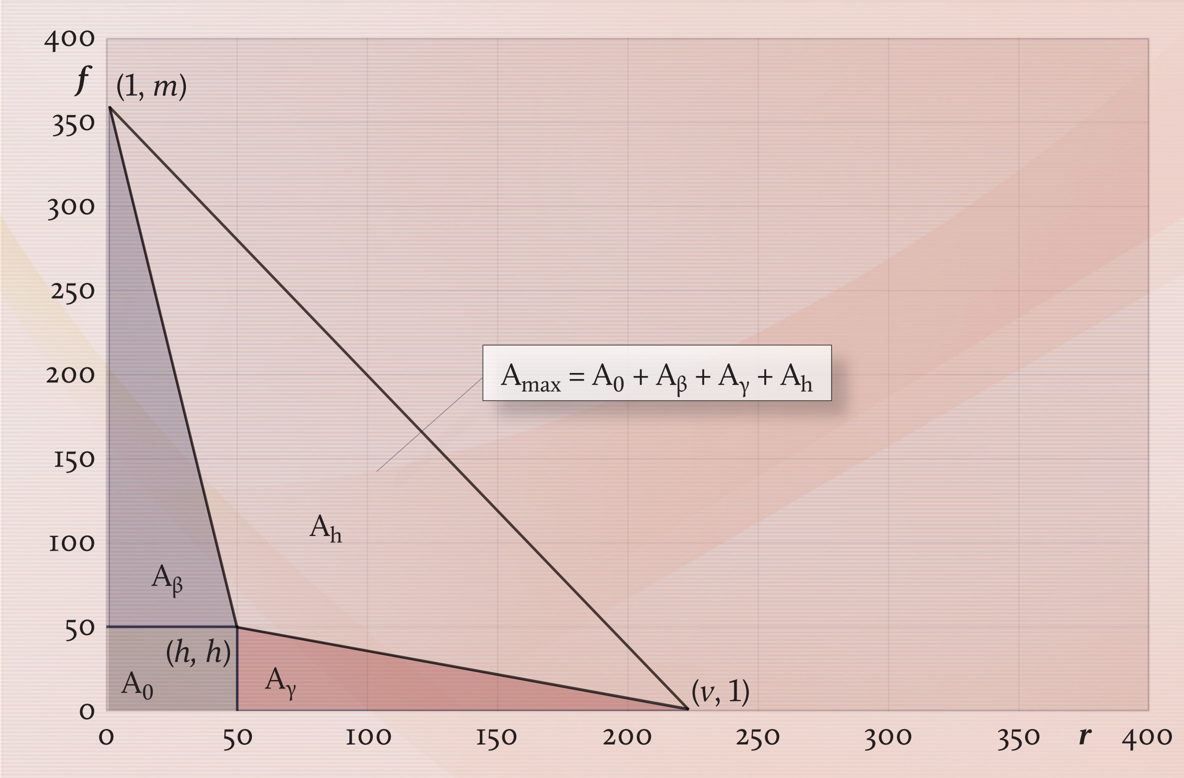

Figure 3 The characteristic triangle and the elementary calculation of its area

characteristic triangle is degenerate with $\alpha = \pi$ and 1 (if $\alpha = \pi/2$, i.e., $h = 1$). In the first case, we have $h = (mv - 1)/(m + v - 2)$, provided $v > 1$ and $m > 1$. Note that this is a theoretical case since h has to be an integer. According to Popescu et al., a small A value is associated with a frequent use of a large number of words, while a greater A value indicates that more words have very small frequency and only few words are used often. Translating this to scientometric terms, greater A values stand for polarised citation patterns with few highly cited papers contrasted by many less cited papers, while a smaller value of A would indicate a more balanced distribution of citations. Again, the actual value of A depends only on v and m, while the text length n (and thus the total number of citations received) is not relevant. In their sample collection, Popescu et al. have not found any correlation between A and other measures of vocabulary richness but they considered it a result of a kind of self-organisation as it proved to be strongly correlated with the writer's view (α). They conclude that a writer, who still has strong control of the word-frequency structure in the beginning of writing, loses this with increasing text, and the frequency structure, due to self-organisation, strives to the golden section. The question arises of a similar property characterises citations, in other words, are citations received at the beginning of the reception rather the expression of kind of individual expression of citers and do they follow later on, when a sufficient number of citations has been accumulated, more general, self-organising patterns? This would bring us back to the well-known cumulative advantage or success-breads-success phenomenon. In any case, Popescu et al. found the linear regression line $A = 1.826 - 0.571 \cdot \alpha$ (with $r^2 = 0.997$) for 54 poems by the Slovak poet Eva Bachletová. From the above-mentioned properties of A and α we obtain a deterministic relationship for the extremes $\alpha = \pi/2$ and $\alpha = \pi$, and $A = 1$ and $A = 0$, respectively. This does not yet imply any *linear* relationship between these two parameters derived from Eqs (2) and (3). A theoretical linear regression would rather be unexpected since both A and α depend of three free parameters. Now, if there exists any linear regression line, it should run through the point (π/2, 1) and another point close to (π, 0) since h needs to be an integer, otherwise α and A will not take the values π and 0. Hence, we obtain the theoretical function $A = 2.00 - 2/\pi \cdot \alpha$ (with $2/\pi \approx 0.6366$. While the low-end of α reflects a quite frequent case, where the observations meet the theoretical situation and imply $A = 1$, the high-end of α will rather be the exception to the rule and we thus do not consider these expected values.

The intention of using the above example from citation analysis was to illustrate the possible equivalence between the geometry of the h-index in scientometrics and quantitative linguistics and to show a possible starting point for the application of the glottometric model to our field since this could serve a real example of cross-fertilisation between two disciplines both of which are studying quantitative aspects of different but related phenomena. However, before we could draw any conclusion in that regard, we need to analyse at least a number of samples of citations distributions from different subject fields. This will be done in the following section.

3. THE EXPERIMENT

Analogously to the above mentioned glottometric papers, we decided to select real-live examples. These are taken from the Clarivate Analytics Web of Science Core Collections (WoS) databse. In particular, we have chosen 25 individual authors from about 20 different fields in the sciences and social sciences. The aim was not to collect a random sample but to find specific cases with sometimes almost extreme combinations of cited publication output, maximum citation rate and h-index. Table 1 gives all relevant indicators for the selected authors, including the number of cited papers (v), de maximum citation rate (m), the h-index (h) and the total number of citations received (n). In addition, we have calculated the angles β, γ, α_1, α_2, α as well as the relative area of the characteristic triangle A. Data have been retrieved from the WoS on 01–14 June 2019.

Case #1 is the sample already used to introduce the model in the previous section. The "minimum" case (# 24) is represented by a publication output comprising three cited publications with the maximum citation rate of 3 and an h-index of 1. This case implies the unique solution $\alpha = \pi/2$, $\alpha_1 = \alpha_2 = \pi/4$, $\beta = \gamma = 0$ and $A = 1$. By contrast, case # 25 with more than 700 cited papers but an already extreme maximum citation rate of nearly 70, 000 citations and an h-index of 69 yields a quite similar α value. By contrast, the values of β and γ are because of the extreme citation rates rather close to the value π/2. Cases #25 and #4 (both chemistry) are quite similar with regard to α and A although h, m and v differ. #18 (Economics), #8 (Biochemistry & Molecular Biology), #14 (Psychology) and #23 (Behavioural sciences) have the largest views but the two components (α_1 and α_2) are different, which is line with the large differences in their h, m and v values. This is contrasted by the "small-view" cases #24 (Education), #4 (Chemistry), #20 (Materials science) and #25 (Inorganic chemistry), again with different h, m and v triplets. For the cases #7, #9, #20, #22 and #24, the blue line running through the points (1, 1) and (h, h) in Figure 1 practically bisects the angle α. #14 (Psychology), #17 (Sociology), and #18 (Economics) reflect asymmetric division of α. Finally, the relatively small A values (< 0.6) of cases #5 (Mathematics), #8 (Biochemistry & Molecular Biology), #10 (Environmental), #14 (Psychology), #17 (Sociology), #18 (Economics), #21 (Mathematics) and #23 (Behavioural sciences) indicates frequent citations of many publications although in some of these cases (#14 and #18) outlier maximum citation rates could be observed. This is contrasted by #4, #20 and #24 with $A > 0.9$, which reflect a more polarised situation and where #24 has, however, to be considered a rather singular case.

Table 1 Computation of A and "view" in the work of academic authors in the mirror of the WoS (data sourced from Clarivate Analytics Web of Science Core Collections)

#	h	m	v	n	β	γ	α_1	α_2	α	A	Subject field
1	50	359	223	7821	1.414	1.295	0.943	1.061	2.004	0.642	Social Sciences
2	55	1063	176	25594	1.517	1.151	0.839	1.205	2.044	0.641	General intern. medicine
3	58	2181	325	13063	1.544	1.360	0.812	0.996	1.808	0.798	Social Sciences
4	91	7117	1247	52379	1.558	1.493	0.798	0.863	1.661	0.915	Chemistry
5	11	33	82	479	1.144	1.431	1.212	0.925	2.137	0.564	Mathematics
6	69	318	2570	25951	1.304	1.544	1.052	0.813	1.865	0.759	Chemistry
7	32	150	180	3697	1.314	1.364	1.042	0.992	2.034	0.619	Physics
8	28	191	76	2323	1.407	1.058	0.950	1.298	2.247	0.498	Biochem. & Molec. Biol.
9	21	176	137	1620	1.442	1.400	0.914	0.956	1.870	0.739	Cardiology
10	36	389	108	5208	1.472	1.118	0.884	1.238	2.122	0.583	Environmental sciences
11	58	3788	251	19110	1.556	1.284	0.801	1.073	1.873	0.757	Physics
12	103	2528	396	39045	1.529	1.236	0.827	1.120	1.948	0.701	Oncology
13	33	2365	101	9186	1.557	1.131	0.799	1.225	2.024	0.666	Agriculture
14	39	11953	87	36900	1.568	0.901	0.789	1.455	2.244	0.555	Psychology
15	49	5004	183	23634	1.561	1.227	0.795	1.129	1.924	0.727	Psychology
16	59	6642	234	25664	1.562	1.251	0.794	1.105	1.900	0.742	Engineering
17	25	3007	60	8632	1.563	0.970	0.793	1.386	2.180	0.585	Sociology
18	48	14808	78	46345	1.568	0.568	0.789	1.788	2.577	0.386	Economics
19	12	286	41	960	1.531	1.208	0.826	1.148	1.973	0.686	Architecture
20	122	2249	4251	84947	1.514	1.541	0.842	0.815	1.657	0.918	Materials Science
21	17	54	102	982	1.163	1.385	1.194	0.971	2.165	0.540	Mathematics
22	18	98	92	1241	1.361	1.345	0.995	1.011	2.006	0.638	Economics
23	19	68	58	824	1.219	1.138	1.137	1.218	2.355	0.416	Behavioural sciences
24	1	3	3	4	0.000	0.000	0.785	0.785	1.571	1.000	Education
25	69	69833	713	126023	1.570	1.466	0.786	0.891	1.677	0.904	Inorganic chemistry

Figure 4 finally shows the empirical relationship between α and A, where we found a linear regression with the linear trendline $A = 1.970 - 0.647 \cdot \alpha$, which is indeed close to the theoretical equation ($A = 2.000 - 0.637 \cdot \alpha$). The correlation coefficient ($r^2 = 0.966$) is slightly lower than in the glottometric case reported by Popescu et al. (2012), but still reflects a strong correlation.

The subject field has certainly an effect on publication activity, citation impact and h-index, but, using glottometric parlance, the "view" and the extent of (citation) exploitation of the given publication output are largely independent of the subject. In this regards we see a clear similarity with quantitative linguistics, however, the golden ratio in science, as reflected by citations, seems not to be located in the h-point.

4. CONCLUSION

In our study we have reassumed the geometric model according to Popescu and Altmann (2006; 2007), which, in turn, is based on the well-known h-graph. The model proceeds from the *h-point* (in linguistics also called *Hirsch-Popescu-point*) and the characteristic triangle created by the h-point, the frequency of rank 1 and the greatest rank. Its internal angle at the h-point is called the *writer's view*. In particular, Popescu and Altmann analysed the geometric properties of this triangle in order to answer several questions related to the richness and exploitation of vocabularies. We have reassumed their model for application to citation analysis and introduced several new parameters originated from the geometric properties of the characteristic triangle. The adoption of the "writer's view", the relative area of the characteristic triangle and the new parameters α_1 and α_2 help obtain a more differentiated picture of the citation exploitation of an author's publication output. Our study could furthermore confirm similarities between the glottometric and scientometric model regarding the area of the characteristic triangle and the view parameter. Nevertheless, similarly to the findings by Pan et al. (2015) concerning Chinese poetry, we could not confirm, in our sample, that the golden ratio is located at the h-point. Our results could, however, be considered a true example of cross-fertilisation between the two disciplines both of which are studying quantitative aspects of different but nevertheless similar and related phenomena.

REFERENCES

Glänzel, W. (2012), The role of core documents in bibliometric network analysis and their relation with h-type indices. *Scientometrics*, 93(1), 113–123.

Glänzel, W., Debackere, K., Thijs, B., Heeffer, S. (2016), *A Researcher in the Mirror of his Models*. In: Glänzel, W., Schlemmer, B. (eds) András Schubert – A World of Models and Metrics A festschrift for his 70th birthday. Leuven, Belgium, 2016, pp. 25–39.

Glänzel, W., Heeffer, S., Thijs, B. (2017), Lexical analysis of scientific publications for nano-level scientometrics. *Scientometrics*, 111(3),1573–1589.

Glover, K., Hulley, H., Peskir, G. (2013), Three-dimensional Brownian motion and the golden ratio rule. *Annals of Applied Probability*, 23(3), 895–922.

Herdan, G. (1964), *Quantitative linguistics*. London, Butterworths.

Kornai, A (2002), How many words are there? *Glottometrics,* 4, 61–86.

Lendvai, E. (1966). Bartók und der Goldene Schnitt. *Österreichische Musikzeitschrift*, 21(11), 607–615.

Larson, Paul, The Golden Section in the Earliest Notated Western Music. The Fibonacci Quarterly 16 (1978), Nr. 6, 513–515.

Pan, X., Hui Qiu, H., Liu, H. (2015), Golden section in Chinese Contemporary Poetry. *Glottometrics*, 32, 55–62

Popescu, I.-I., Altmann, G. (2006), Some aspects of word frequencies. *Glottometrics*, 13, 23–46.

Popescu, I.-I., Altmann, G. (2007). Writer's view of text generation. *Glottometrics*, 15, 71–81.

Popescu, I.-I., Čech, R., Altmann, G. (2012), Some Geometric Properties of Slovak Poetry. *Journal of Quantitative Linguistics*, 19(2), 121–131

Schubert, A., Korn, A., Telcs, A. (2009). Hirsch-type indices for characterizing networks. *Scientometrics*, 78(2), 375–38

Telcs, A., Glänzel, W., Schubert, A. (1985), Characterization and statistical test using truncated expectations for a class of skew distributions. *Mathematical Social Sciences*, 10, 169–178.

Tuzzi, A., Popescu, I.-I., Altmann, G. (2010). *Quantitative Analysis of Italian texts*. Lüdenscheid: RAM.