Olympiads in Informatics – the Journal's First Six Years

SIMON

The University of Newcastle
Ourimbah Campus, PO Box 127, Ourimbah, NSW 2258, Australia
e-mail: simon@newcastle.edu.au

Abstract. This paper analyses the first six volumes of *Olympiads in Informatics*, examining the papers that have been published in these volumes and the authors who have written them. It finds that the journal is truly international, and that its authors appear to conform with the pattern expected of a discipline, notwithstanding that at this point it has only a short history and a small pool of publications. With regard to the papers themselves, the analysis finds that they cover a reasonable range of topics within the overall area of informatics olympiads, and that while they are predominantly reports, over the six years they show a pleasing increase in the number of analytical research papers.

Key words: olympiads in informatics, bibliometrics, paper classification.

1. Introduction

"*Olympiads in Informatics* is a refereed scholarly journal that provides an international forum for presenting research and development in the specific area of teaching and learning informatics through competition. The journal is focused on the research and practice of professionals who are working in this field." (Dagienė *et al.*, 2008).

After six annual volumes, it is timely to examine the journal and assess its contribution to research and development in the field. Who is contributing to the journal? Where do the authors come from? How many papers has each author written? To what extent are authors collaborating, forming a community of practice? What are the papers about? How many of the papers are factual reports, which would fall into the 'practice' category, and how many are more suited to the description 'research'?

This paper reports on a bibliometric study of the papers published in *Olympiads in Informatics* and of the authors of those papers. Bibliometric analysis of this sort is reasonably common in the library sciences, and has been applied to publications in numerous discipline areas including accounting (Chung *et al.*, 1992), crystallography (Behrens and Luksh, 2006), bioinformatics (Patra and Mishra, 2006), and computing education (Simon 2009a, 2009b). It has potential merit as the journal equivalent of genealogical research – it helps us to understand who we are ('we' in this case being the journal's authors) and where we have come from. It thus strengthens the notion of the informatics olympiad

community as a community and not just a set of authors who happen to present at the same conference and publish in the same journal.

The goal of the study, then, is to answer the question: what patterns can be discerned in the papers that have been published in the first six volumes of *Olympiads in Informatics*, and in the authors who have written them?

2. The Papers Being Considered

Any analysis of the papers in a particular publication must begin by clearly establishing which papers are to be considered and which are to be omitted from the analysis.

Volumes 1–4 of the journal included only a foreword or editorial and the peer-reviewed papers. A foreword is not generally considered to be a published paper, so the forewords were not included in the analysis.

Volumes 5 and 6 saw the introduction of a new section called 'Reviews, comments'. This section is described as being for "book reviews, comments on task solutions and other initiatives in connection with teaching informatics in schools". While it is not clear whether these contributions undergo the same sort of peer review as the papers, they are clearly treated in a different manner – for example, they are not listed individually in the table of contents – so it was decided to exclude them, restricting the analysis to the full papers that have clearly undergone peer review prior to acceptance in the journal.

By this criterion, there have been 101 full papers published in the six volumes of *Olympiads in Informatics*, and it is these 101 papers that will be analysed in this work.

3. The Authors

As with any other journal or conference, some papers have single authors while others have multiple authors. Likewise, there are authors who have contributed only one paper to the journal and authors who have contributed more. To deal with this many-to-many relationship we introduce an intermediate entity called the author contribution. This term should not be interpreted as suggesting that we are trying to measure *how much* each author contributed to each paper – this is something known only to the authors. Rather, the author contribution is a uniform measure of a single author's authorship of a single paper.

There are a number of different ways of counting an author's contribution to a paper (Larsen, 2008). In the system known as *complete counting*, each author of a paper is given a count of 1 for that paper, regardless of the number of authors. If a paper has four authors, each of the four will be given a count of 1, and the paper will register four contributions. In *complete-normalised counting*, on the other hand, the paper itself is given a count of 1, which is then divided equally among the authors. If a paper has four authors, each of the four will be given a count of 0.25.

Each system has its merits and its drawbacks. Complete-normalised counting recognises that the work of writing a joint paper is shared among its co-authors, and so gives

each author less credit. But at the same time it could be seen as devaluing collaboration, by giving the authors of a two-author paper only half the credit that either would have earned by writing the paper alone. For this reason, this analysis will place more emphasis on complete counting, but will acknowledge the differences between it and complete-normalised counting.

Using the complete counting method, the 101 papers that have appeared in the first six volumes of the journal together comprise 213 author contributions; that is, there is an average of just over two authors per paper. The greatest number of papers (42) have single authors; nearly as many (35) have two authors; there are 14 three-author papers, six four-author papers, and one each of papers with five, nine, ten, and eleven authors. It can be argued that shared authorship is a good thing, as collaboration is one clear measure of engagement within the community (Simon, 2007). However, this should not be seen as diminishing the value of single-author papers.

There are 130 distinct authors who have contributed to papers published in the journal.

Lotka's law of author productivity (Nicholls, 1989) encapsulates the empirical observation that in a sufficiently large list of published papers within a discipline, 60% of authors will contribute to only one paper, 15% to two papers, 7% to three papers, and so on. More precisely, given a total pool of A authors, the number of authors A_n contributing to n papers will be CA/n^p, where C and p are constants that vary according to the discipline but are generally expected to be close to 60% and 2 respectively. Having validated Lotka's law on multiple diverse data sets, Nicholls suggests that values of 71% to 81% are more realistic for C than 60%.

Although this is not the most rigorous approach, C can be trivially estimated from the case where $n = 1$, that is, the proportion of authors who have contributed to just one paper. For *Olympiads in Informatics* this estimate gives a value of 68.5%, which sits comfortably between the generally quoted 60% and Nicholls's subsequent observations, suggesting that *Olympiads in Informatics* has a reasonable number of authors who have contributed two and more papers.

Figure 1 shows the observed numbers of authors contributing to specified numbers of papers, alongside the numbers predicted by Lotka's law. The power constant p has been set at 2.0 to give a good visual match with the observed counts of contributions. If this constant were any lower the curve would drop more quickly, indicating that very few authors come back to write further papers after their first.

The message from Figure 1 is that the papers in *Olympiads in Informatics* follow reasonably closely the expectations for a large list of publications within a single discipline. This is very positive, as the journal has been running for only six years, and it is not entirely obvious that informatics olympiads can be considered a discipline. It will be interesting to see whether the pattern continues as the number of issues increases.

Table 1 shows the same observed author contributions as in Figure 1, and lists the authors who have contributed to four or more papers in the six volumes of the journal.

Table 2 shows the journal's leading authors according to the complete-normalised counting system, which counts for each author a fraction of each paper according to how many authors the paper has.
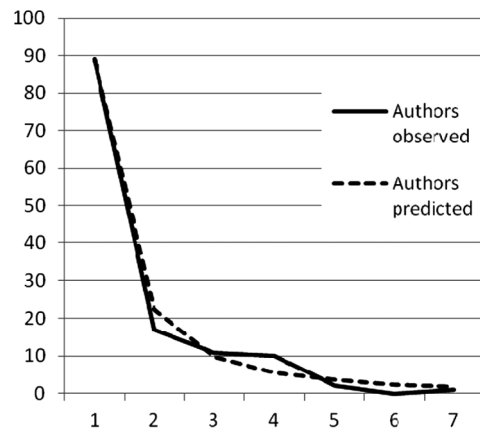
Fig. 1. Lotka's Law, with C=68.5% (from number of 1-paper authors) and p=2.0 (best fit by eye).

Table 1

Author contributions using complete counting, listing the authors with four or more papers

Papers	Author count	Authors
7	1	Pavel S Pankov
5	2	David Ginat, Marcin Kubica
4	10	Benjamin A Burton, Michal Forišek, Mathias Hiron, Emil Kelevedjiev, Vladimir M Kiryukhin, Krassimir Manev, Martin Mareš, Bruce Merry, Wolfgang Pohl, Tom Verhoeff
3	11	
2	17	
1	89	

Table 2

Leading authors using complete-normalised counting

CN Count	Authors
4.5	Pavel S. Pankov
4.0	Tom Verhoeff
3.5	Martin Mareš
3.0	Vladimir M. Kiryukhin
2.8	David Ginat, Bruce Merry
2.6	Benjamin A. Burton, Michal Forišek
2.2	Wolfgang Pohl
2.1	Willem van der Vegt
2.0	Marina S. Tsvetkova
1.9	Krassimir Manev
1.8	Emil Kelevedjiev

While Pavel S. Pankov is the clear leader by both counting systems – and also has the distinction of being the only author to have been published in all six volumes of the journal – the other rankings are quite different under the two counting systems. Tom Verhoef, for example, moves from equal fourth with complete counting to clear second with complete-normalised counting, because he was the sole author of all four of his papers. Conversely, David Ginat drops from equal second to equal fifth, and Marcin Kubica from equal second to fourteenth (not shown on Table 2), because their five papers were shared with various numbers of co-authors.

It is important to repeat that each of these counting systems has its merits and its drawbacks; and, of course, to point out that any author ranked in the top dozen by either system should be proud of that achievement.

4. The Countries

Olympiads in Informatics is a truly international journal. From the outset it has published papers from a wide range of countries. The first volume explicitly called for state-of-play reports from countries participating in the International Olympiad in Informatics, and accepted papers from Brazil, Bulgaria, Canada, China, Croatia, the Czech Republic, Germany, Italy, Kyrgyzstan, Lithuania, Macedonia, Mongolia, Poland, Portugal, the Russian Federation, Slovakia, and the USA. In subsequent issues these countries were joined by a further 23, making a total of 40 countries represented in the journal. As there are authors who have had more than one paper accepted over the years, so there are necessarily countries represented many times over (though there is one author who changed countries between publications). Table 3 shows the 40 countries represented in the 101 papers, with the number of papers that have come from each. A final row in the table notes the eight papers with authors from two or more countries; these papers are examples of the international collaboration that is almost certain to arise from a venture such as the International Olympiad in Informatics.

5. The Topics of Papers

What are the journal's papers about? Of course they are all related to the overriding theme of informatics olympiads, but within that theme they can be about quite different topics. For example, there will be papers about the national organisation of olympiads (Anido and Menderico, 2007), about the creation and choice of tasks (Burton and Hiron, 2008), and about the grading of tasks (Merry, 2010). Each of these topics could be further divided – for example, grading could be divided into automatic grading (Mareš, 2009) and manual grading (Pohl, 2008) – but this would lead to a proliferation of topics with very few papers in each, and so would make it more difficult to present a broad overview of what the papers are about.

Some papers clearly deal with more than one topic. In such cases, rather than trying to identify every topic, no matter how small its contribution, it was decided to classify each paper according to the most dominant of the topics it covers.

Volume 1 of the journal was based on the first Olympiads in Informatics Conference in 2007, which explicitly focused on organising olympiads at the national level. "Many of the issues at the national level differ from country to country. We have different educational systems and the availability and take-up of information technology varies, but even here there are as many similarities as differences. We also face many of the same problems: How do we pick our students? How do we train them? What is suitable material? etc" (Dagienè *et al.*, 2007). Therefore it is no surprise that almost all of the papers in that volume cover the topic of *organisation*.

The second conference, and therefore Volume 2, had a dual theme including task types. "Tasks are perennial issue for contests, their most visible aspect and, for many contestants, the primary reason for participation. We strive for quality, variety and suitability. We endeavour to make tasks interesting, understandable and accessible. They are used to test contestants and to train them, and perhaps even to capture the imagination of those outside the contest, be they family, sponsors or the media" (Dagienè *et al.*, 2008). Many of the papers in Volume 2 were therefore on the topic of *tasks*.

Overall, six distinct topics were identified across the six volumes of the journal, as summarised in Table 4. Each topic is briefly described below.

- *organisation*
  This is a broad topic covering many aspects of the organisation of olympiads, typically at the national level. It includes levels of competition, the logistics of getting students to suitable venues and of training them, costs and budgets, timing, and many other aspects of how an olympiad is organised.
- *tasks*
  This topic deals with many aspects of the tasks used in olympiads: task selection, examples of tasks, new styles of task, and other related matters.
- *grading*
  Papers with this topic describe or propose ways to grade the students' submissions in an olympiad.

Table 3

The countries from which the papers have come, with a count of papers from each country

Albania	1	Finland	2	Lithuania	1	Singapore	1
Australia	3	France	2	Macedonia	1	Slovakia	3
Belgium	2	Germany	2	Mexico	1	Slovenia	1
Brazil	1	Ghana	1	Mongolia	1	Spain	1
Bulgaria	9	Indonesia	1	New Zealand	1	Switzerland	1
Canada	2	Israel	6	Poland	4	Syria	1
China	2	Italy	4	Portugal	3	Thailand	1
Croatia	1	Japan	2	Romania	1	The Netherlands	6
Czech Republic	3	Kyrgyzstan	8	Russian Federation	5	United Kingdom	3
Estonia	1	Latvia	1	Serbia	1	United States	2
International	8						

Table 4

The six identified topics and their distribution across the six volumes

Volume	Organisation	Tasks	Grading	Preparation	Infrastructure	Impact
1	13	3	1			
2	5	8	2	1		
3	3	6	3	1	1	
4	7	2	3	3		
5	7	5	2	3	1	1
6	7	5	3		4	1
Total	42	29	14	8	6	2

- *preparation*
  These papers deal with the preparation of students for olympiads, that is, preparing students prior to an olympiad. Papers about the training of students once they are in the olympiad are included in the *organisation* topic.
- *infrastructure*
  Papers with this topic describe or propose infrastructure to be used in running olympiads. They might, for example, discuss the way that wireless networks were used for submission of the competitors' work (Imajo, 2011).
- *impact*
  Only two papers were indentified within this topic, but it was not possible to ascribe those papers to any of the other five topics. These are papers that clearly deal with the impact of the olympiads in broader areas of education (Audrito *et al.*, 2012) and recruitment (Jakacki *et al.*, 2011).

6. The Natures of Papers

Independently of their topics, the papers have been divided into three categories specifying more about the nature of the work they present.

- *report*
  Report papers describe the past and/or present situation, or sometimes propose future situations. They are essentially capturing and reporting on existing factual knowledge.
- *software*
  Software papers are also reports, but of a particular kind. Rather than describing existing knowledge, they present and describe software that has been designed and constructed for a specific purpose related to the olympiads. For example, a *software* paper whose topic is *grading* will describe software that has been created to assist with the grading of students' submissions (Maggiolo and Mascellani, 2012).

Table 5

The three identified natures and their distribution across the six volumes

Volume	Report	Software	Analysis
1	16	1	0
2	12	1	3
3	7	4	3
4	12	2	1
5	11	0	8
6	6	7	7
Total	64	15	22

- *analysis*
  Analysis goes well beyond reporting, gathering data and analysing it to produce hitherto unknown results. The data might be pre-existing (Forišek, 2009) or gathered expressly for the analysis (Merry, 2010).

Table 5 shows the breakdown of papers by their natures. While there is of course great value in reporting existing facts, it is good to see the recent growth in the number of analysis papers in the journal, as these papers are more readily perceived as research (Simon, 2007), and make a clear contribution to the journal's expressed aim of publishing high-quality research.

7. Correlating Nature and Topic

Table 6 groups the papers according to both their topic and their nature. It is interesting to see that a clear majority of the analysis papers focus on the olympiad tasks, with a few focusing on organisation and a few on grading. It is not surprising that much of the software presented in the journal focuses on grading and on infrastructure. And while the bulk of the reports focus on olympiad organisation and on tasks, overall the reports cover the full range of topics.

Table 6

Papers according to topic and nature

	Organisation	Tasks	Grading	Preparation	Infrastructure	Impact
Report	39	14	3	6	1	1
Software		1	8	1	5	
Analysis	3	14	3	1		1

8. Conclusion

This analysis help to form a picture of the authors and papers of *Olympiads in Informatics*, a picture that will help members of the olympiads community to better know and understand that community.

Even though it has been going for only six years, and has published only 101 papers, the journal demonstrates a pattern of authorship that looks remarkably like a discipline when examined in the light of Lotka's Law.

The papers published in the journal are predominantly reports – not least because the first conference and the first volume explicitly called for such reports – but there is a clearly perceptible increase in the number of analytical papers being published.

This analysis supports the view of *Olympiads in Informatics* as a scholarly journal that provides an international forum for presenting research and development in the teaching and learning of informatics through competition.

References

Anido, R.O., Menderico, R.M. (2007). Brazilian olympiad in informatics. *Olympiads in Informatics*, 1, 5–14.

Audrito, G., Demo, G.B., Giovannetti, E. (2012). The role of contests in changing informatics education: a local view. O*lympiads in Informatics*, 6, 3–20.

Behrens, H., Luksh, P. (2006). A bibliometric study in crystallography. *Acta Crystallographica*, B62, 993–1001.

Burton, B.A., Hiron, M. (2008). Creating informatics olympiad tasks: exploring the black art. *Olympiads in Informatics*, 2, 16–36.

Chung, K.H., Pak, H.S., Cox, R.A.K. (1992). Patterns of research output in the accounting literature: a study of the bibliometric distributions. *Abacus*, 28(2), 168–185.

Dagienė, V., Cepeda, A., Forster, R., Manev, K. (2007). Foreword. *Olympiads in Informatics*, 1, 3–4.

Dagienė, V., Cepeda, A., Forster, R., Manev, K., Vasiga, T. (2008). Foreword. *Olympiads in Informatics*, 2, 3–4.

Forišek, M. (2009). Using item response theory to rate (not only) programmers. *Olympiads in Informatics*, 3, 3–16.

Imajo, K. (2011). Contest environment using wireless networks: a case study from Japan. *Olympiads in Informatics*, 5, 26–31.

Jakacki, G., Kubica, M., Waleń, T. (2011). Codility. Application of olympiad-style code assessment to pre-hire screening of programmers. *Olympiads in Informatics*, 5, 32–43.

Larsen, P.O. (2008). The state of the art in publication counting. *Scientometrics*, 77(2), 235–251.

Maggiolo, S., Mascellani, G. (2012). Introducing CMS: a contest management system. *Olympiads in Informatics*, 6, 86–99.

Mareš, M. (2009). Moe – design of a modular grading system. *Olympiads in Informatics*, 3, 60–66.

Merry, B. (2010). Performance analysis of sandboxes for reactive tasks. *Olympiads in Informatics*, 4, 87–94.

Nicholls, P.T. (1989). Bibliometric modeling processes and the empirical validity of Lotka's law. *Journal of the American Society for Information Science*, 40(6), 379–385.

Patra, S.K., Mishra, S. (2006). A bibliometric study of bioinformatics literature. *Scientometrics*, 67(3), 477–489.

Pohl, W. (2008). Manual grading in an informatics contest. *Olympiads in Informatics*, 2, 122–130.

Simon (2007). A classification of recent Australasian computing education publications. *Computer Science Education*, 17(3), 155–169.

Simon (2009a). Informatics in Education and Koli Calling: a comparative analysis. *Informatics in Education*, 8(1), 101–114.

Simon (2009b). Ten years of the Australasian Computing Education Conference. *Australian Computer Science Communications*, 31(5), 157–163.

Simon is a senior lecturer in information technology at the University of Newcastle in Australia. Several years ago he devised a scheme for classifying publications in computing education, and together with a small team he has published a number of papers in that area. He also conducts research in programming education and in academic integrity.