Library Review

Bibliometric and webometric methods for assessing research collaboration

Sharunizam Shari, Gaby Haddow, Paul Genoni,

Received 1 August 2012
Accepted 20 September 2012

Bibliometric and webometric methods for assessing research collaboration

Sharunizam Shari
Faculty of Information Management, Universiti Teknologi MARA Kedah, Merbok, Malaysia, and
Gaby Haddow and Paul Genoni
School of Media, Culture and Creative Arts, Curtin University, Perth, Australia

Abstract

Purpose – The purpose of this paper is to describe the methods and findings of a pilot study which applied bibliometrics and webometrics to examine collaboration in Malaysian biotechnology.

Design/methodology/approach – The research applied bibliometric and webometric methods to publications and web sites affiliated with Malaysian institutions. The bibliometric analysis focused on biotechnology-related journal articles indexed in Web of Knowledge. The webometric analysis examined the web sites of top biotechnology institutions generated in the bibliometric analysis. Collaboration behaviour was assessed in three ways: intra-institutional versus inter-institutional; national versus international collaboration; and by type of institution collaboration according to the triple helix model.

Findings – Findings of the pilot study, which applied bibliometric and webometric analyses to a limited sample, indicate that the methodologies will collect the desired data for a more extensive study.

Research limitations/implications – The quantitative research results describe the collaboration evident in publications and web sites, but not why it has happened in such a way.

Practical implications – The methodologies provide a framework for similar research exploring the impacts of collaboration in an e-research environment. The methodology is innovative and practical in terms of the combined use of bibliometric and webometric analyses.

Originality/value – This is one of few studies that has examined collaboration using both bibliometric and webometric methods, and elements of the methodology appear to be unique to the study. The methodologies will contribute to an emerging body of literature that explores the nature of research productivity and research collaboration.

Keywords Research work, Biotechnology, Bibliometrics, Webometrics, Research collaboration, Malaysia

Paper type Research paper

Introduction

As science research has become more complex and sophisticated, greater attention has been paid to the benefits to be gained from collaboration. Collaboration, on both national and international levels, has been identified in a series of studies (Abramo *et al.*, 2009; Adams *et al.*, 2005; Alcaide *et al.*, 2012; Anuradha and Urs, 2007; Bordons *et al.*, 1996; Finlay *et al.*, 2012; Gonzalez-Albo *et al.*, 2012; Seglen and Aksnes, 2000) as a crucial factor in bringing together the expertise that resides in different sectors and institutions in order to enhance research productivity. Assisted by developments in information technologies and e-research, research collaborations are increasingly multifaceted and large-scale. And whereas collaboration can be undertaken within one research

Library Review
Vol. 61 No. 8/9, 2012
pp. 592-607

0024-2535
DOI 10.1108/00242531211292097

sector (that is government, university or industry), it is increasingly being grounded in cross-sectoral partnerships; and while it can be undertaken on a national basis, it is increasingly being implemented on an international scale.

An ongoing large study, for which this pilot study was undertaken, is assessing the patterns of collaboration that are evidenced by the publishing output of Malaysian biotechnology researchers. The study is premised on the belief that research collaboration is critically important for a newly industrialized country such as Malaysia, as it attempts to lift its research and development productivity to match that of industrialized countries that are supported by fully developed research sectors. The biotechnology sector has been singled out because it has been identified and supported by the Malaysian government as an area of research focus, and it provides many of the key elements of a science discipline well suited to the development of broadly based collaborative research affiliations.

Government led research management in many countries has increasingly used the advent of e-research as a driver for broader and deeper collaboration. An issue facing researchers and research managers has been the need to develop methods for measuring the patterns and effectiveness of collaboration, particularly in a research environment where collaborative activity is apparent not only in traditional research outputs such as journal articles, but also through the patterns of networking and linking that can be assessed through the "non-traditional" means of the internet. The research that is the basis of this current paper has therefore applied social network analysis methods in order to assess Malaysian biotechnology research collaboration in both traditional and virtual environments. Specifically this entails using both bibliometric and webometric research methods to capture the widest possible coverage of Malaysian biotechnology research output in order to assess both the incidence and visibility of collaborations. The purpose of this paper is to explain these innovative research methods with the intention that they can be used or modified for similar studies of collaboration, and to report on the findings of a pilot study which tested the methods on a small sample of data.

Background to the study

Malaysia is a newly industrialized country in South East Asia. The Malaysian government has outlined several research foci as integral components of the national development objectives set out in the document *Vision 2020*. One of these research priorities is biotechnology. The government has announced that:

> Biotechnology is poised to drive the next wave of knowledge-based industries that will contribute to growth and wealth creation, new investment and employment opportunities as well as deliver social and environmental benefits (Economy Planning Unit, 2006, p. 157).

In the Ninth Malaysia Plan (2006-2010), biotechnology was identified as providing a significant growth opportunity, as Malaysia becomes increasingly competitive amongst global economies. BiotechCorp, the leading agency responsible for the growth of the Malaysian biotechnology industry, has recognized the importance of creating greater synergy and collaboration between government research institutions, universities and the biotechnology industry (BiotechCorp, 2011).

In 2003, the Ministry of Science, Technology and Innovation of Malaysia (MOSTI) commissioned a bibliometric study to examine Malaysian Science and Technology (S&T) research outputs at the international level. In the study, all Malaysian S&T publications

(journal articles) published up to 2003 and indexed in international databases were examined. The objective of the study was to identify S&T research outputs by Malaysian scientists, measured by the number of published papers in the international literature. It also sought to identify areas of strength and weakness in certain fields; the contribution of individuals and institutions to relevant fields of research, and the extent of collaboration among institutions and countries. The findings provided an overview of the status of Malaysian S&T to 2003 and the second phase of the study commenced in 2008 to capture all publications published up to 2008.

The 2003 and 2008 studies focused on S&T research outputs in formal scholarly communication channels, however, increasingly scientists are using the web to communicate their work. Malaysian S&T research productivity in the web environment is therefore an important aspect of scholarly communication to consider in any assessment of current research collaboration.

The results of the large ongoing study will lead to a greatly enhanced understanding of the research process as it applies to the biotechnology sector in Malaysia, and the outcomes will provide an evidence-base for an improvement in the country's research productivity. The research draws upon the triple helix model of cross-sectoral collaboration to assess the extent and nature of the collaboration between government, universities and industry. Aided by the triple helix analysis the research findings will provide a basis for informed decision making within Malaysia as to whether particular parties within the biotechnology sector will benefit from enhanced collaboration, and whether the sector generally is sufficiently international in its existing collaborations. The data will also contribute to an understanding of the relationship between the research productivity recorded in traditional scholarly journals (the traditional focus of bibliometrics) and the internet (the emerging field of webometrics).

Another benefit of the study is in the development of methodologies for assessing research collaboration. It is one of the few studies to examine collaboration using both bibliometric and webometric methods, and elements of the methodology appear to be unique to this study.

Objectives

Focused on the issue of collaborative research as an aspect of national development, this pilot study assessed research collaboration in Malaysia. The study addressed the following research question:

RQ1. What patterns of collaboration are evidenced by the outputs of Malaysian biotechnology research?

The following objective has been developed to answer the *RQ1*:

(1) To determine the extent and types of collaboration demonstrated in Malaysian biotechnology research outputs (2005-2010) using bibliometric and webometric methods.

The "extent" of collaboration in the objective is defined as follows:

(1) *Number of authors in a paper*. Collaboration exists when at least two authors contribute to a paper. The number of authors contributing to a paper can be used to study whether collaboration has an impact on the citations received by that paper, which in turn indicates the visibility of the paper.

The "types" of collaboration in the objective are defined as:

(1) *Intra-institutional collaboration vs inter-institutional collaboration.* Collaboration can exist within a single institution or between two or more institutions. Studying the behaviour of intra- vs inter-institutional collaboration can determine whether it has an impact on the citation of papers, which in turn indicates the visibility of the institutions.

(2) *National vs international level collaboration.* National level collaboration involves collaboration between institutions from one country while international level collaboration involves collaboration between institutions from at least two countries. Studying this type of collaboration can determine whether national and/or international collaboration has an impact on the citation of papers, which in turn indicates the visibility of the research sector for the countries involved.

(3) *Types of institutional collaboration.* The triple helix model includes three types of institutions; government, university and industry. It is important to identify the key collaborating sectors in Malaysian biotechnology research to inform policy-making. This is examined by looking at the patterns of collaboration between government-university, government-industry, university-industry and government-university-industry.

Methodology

This study has adopted a quantitative research approach. Collaboration is measured in two ways:

(1) using bibliometric methods; and

(2) using webometric methods.

Data were collected, processed and analysed in two phases. First, a bibliometric analysis (Phase 1) was conducted using bibliographic information from biotechnology publications. The results from Phase 1 were then used to determine the most productive biotechnology contributors (by department or institution). The web site of each of these contributors were then analysed in the webometric analysis (Phase 2). The triple helix model was used as the framework to examine the types of collaboration that exist.

A number of collaboration studies using bibliometric methods employed the triple helix model as a framework. The triple helix has been described by its instigator, Etzkowitz (2002, p. 2) as "a spiral model of innovation that captures multiple reciprocal relationships at different points in the process of knowledge capitalization". The elements (or research sectors) incorporated in the triple helix are government, university and industry. Etzkowitz (2002, p. 2) describes the relationship between these three parties as "one of relatively equal, yet interdependent, institutional spheres, each of which can overlap and assume the role of the others". Several researchers have written extensively about the triple helix model and its implications for different research sectors. Irawati (2007) identified the importance of the triple helix for developing countries and discussed the essential stages required to establish a robust synergy between government, university and industry in the context of Indonesian research. Viale and Ghiglione (1998) analysed the characteristics of the triple helix model and the related effects on the innovation system of the European regional

socio-economic systems. Leydesdorf (2005) examined the concept and practice of interconnections between government, university and industry in knowledge-based innovation systems.

This study applied bibliometric methods using the triple helix model, following the examples of Leydesdorff and Sun (2009), Butcher and Jeffrey (2005) and Glanzel and Schlemmer (2007). Most bibliometric studies drawing upon the triple helix use the production of either publications or patents as the basis of analysis. In their examination of national and international dimensions of collaboration, Leydesdorff and Sun (2009) focused on Japanese science publications. Using journal publications as their samples, Butcher and Jeffrey (2005) explored trends in collaboration between industry and university in the field of membrane use for water treatment; and Glanzel and Schlemmer (2007) studied the different collaboration patterns in industry sectors from eight Eastern European countries. Klitjou *et al.* (2007) tracked research in the techno-science network of fuel cells and related hydrogen technology in Norway, and Lundberg *et al.* (2006) explored the adequacy of measuring university and industry collaboration through co-authorship and funding.

The triple helix model has also attracted interest from researchers using webometric methods to study collaboration. Leydesdorff and Curran (2000) mapped government-university-industry relationships on the internet in Brazil and The Netherlands and compared the findings for each country. They found that while the general patterns of collaboration are similar there is a difference in terms of the extent of international collaboration. Priego (2003) studied the relationship between the three sectors represented by the triple helix model by using the web pages of the National Research Council of Germany and Spain, focussing on the fields of biology and biomedicine. It was concluded that the triple helix model could provide qualitative indicators of scientific and technical research practice.

A detailed description of the methods developed for both Phases 1 and 2 are described below.

Phase 1: bibliometric analysis

The data required for bibliometric analysis in this research consists of bibliographic information of publications by Malaysian biotechnology researchers and the citations to these publications. It is important that these data are from authoritative and accurate sources and are available for the time period stated in the study's objective.

The Eighth Malaysia Plan (2001-2005) first identified biotechnology as a key area that would contribute to Malaysia's development as a knowledge-based economy and a biotechnology Research and Development Grant Scheme was introduced in 2001. On the basis that these events should have subsequently influenced biotechnology research, publications from the period 2005 to 2010 was selected as the sample for the bibliometric analysis.

Database and search strategies. There are many scientific databases which cover biotechnology research, such as Scopus, PubMed, Web of Knowledge, MedLine, and IEEE. Each of the databases has its own strengths and weaknesses that attract a specific researcher to use them. Of these databases, Web of Knowledge is one of the most widely used by researchers conducting bibliometric studies and was selected as the main source of data for this pilot study. All biotechnology research outputs published by Malaysian researchers from 2005 to 2010 indexed in the Web of Knowledge

were included in the primary sample. The Web of Knowledge database searched includes MedLine, Web of Science and Current Contents. Data were searched for and collected on March 23, 2012 and resulted in 784 articles related to "biotechnology and applied microbiology" subject area for Malaysia.

Several search strategies were tested before the following procedures were selected as the most effective to gain the best possible recall of biotechnology papers for Malaysia.

Using the basic search interface:

- Enter "Malaysia" as the keyword and limit the search by using the address field.
- Refine Search No. 1 by selecting "biotechnology and applied microbiology" subject area.
- Refine Search No. 2 by selecting "article" and "review" in the document type's option.
- Refine Search No. 3 by selecting "2010", "2009", "2008", "2007", "2006" and "2005" in the publication years option.

Summarized in the search string:

(1) Address = (Malaysia) Refined by: subject Areas = (Biotechnology Applied Microbiology) AND Document Type = (Article or Review) AND Publication Years = (2010 or 2009 or 2008 or 2007 or 2006 or 2005).

Where possible, the full text of the articles was obtained to identify ambiguous information. The bibliographic information of each article was gathered as in Table I.

For the pilot study being reported here, the 20 most cited papers located in the search were the subject of the analyses.

Processing and analysing. While the required bibliographic data are quite readily available from Web of Knowledge, there are problems with data accuracy that need to be addressed. The data therefore require extensive cleaning before analysis is conducted. The focus of the data cleaning is on ensuring correct attribution to both personal authors and institutions:

Title	Tamarind seed powder and palm kernel cake: two novel agro residues for the production of tannase under solid state fermentation by *Aspergillus niger* ATCC 16620
Authors	Abdul Hameed Sabu (Biotechnology Division, Regional Research Laboratory, Council of Scientific and Industrial Research (CSIR), India) Ashok Pandey (Biotechnology Division, National Institute of Interdisciplinary Science and Technology, India) M. Jaafar Daud (Malaysian Agricultural Research and Development Institute (MARDI), Malaysia) George Szakacs (Department of Agricultural Chemical Technology, Technical University of Budapest, Hungary)
Journal	*Bioresource Technology* Volume: 96 Issue: 11, Pages: 1223-1228
Year	2005
Language	English
Source	Web of Knowledge
Times cited	34

Table I.
Example of bibliometric data

- *Personal authors*. Problems with the authors' names include: authors with same exact name; authors with the same surname and initials; authors who have changed their family name after marriage; inconsistency of name used, i.e. some authors use their first name while others used their last name; and data input (typographical) errors. To address these problems further research is often required to distinguish between authors, or to ensure that a single author is correctly identified when publishing under more than one name. This is done by obtaining the full text copy of an article. In many cases accessing the author's web page is necessary to compare their list of publications against the articles obtained from the Web of Knowledge. In this study, authors with the same name and using the same initial(s) were distinguished by putting an identifying numeral after their name. Authors who had changed their family name were identified using the most recent name. To standardise the format of the name used, the full name of the authors were identified and recorded.
- *Institutional names*. Problems that arise with the institutional name are: change of the name of the institution, for example Universiti Pertanian Malaysia was changed to Universiti Putra Malaysia; the use of acronyms or abbreviated form, for example, some authors refer to UPM rather than Universiti Putra Malaysia in their affiliation; different acronyms or abbreviated forms for the same institution, for example some authors may use UiTM or ITM to refer to Universiti Teknologi MARA; and data input (typographical) errors. In addition to verifying or obtaining information from the full text copy of the article, the web sites of the institutions were used to ensure the correct name, and the most recent name was recorded. To standardise the format of institutional names, the full names were identified and recorded.
- *Assigning an institution to an author with multiple affiliations*. Some authors may be affiliated with more than one institution. For example, a PhD student studying at Curtin University but attached as a staff member to Universiti Teknologi MARA, or a medical doctor affiliated with both a university and a government hospital. Most authors attach the primary affiliation to a research output, but may also indicate that they have a secondary affiliation with another institution. In these cases the primary institution was used to assign the affiliation to the respective author.

For the triple helix analysis it is important to correctly classify and code each institution into the three categories of:

(1) government;
(2) university; and
(3) industry.

Institutions that do not belong to any of these categories were categorised as "others" and were not included in the analysis.

Counting collaboration. An important component of the methodology used in collaboration studies is in determining how collaboration is to be counted. There have been extensive debates and arguments regarding the preferred method for counting publications with more than one author (Hagen, 2010; Ioannidis, 2008; Long and McGinnis, 1982;

Narin, 1976), and to date, no one measurement has been agreed upon as the "correct" method. However, since publications represent the productivity of authors or institutions, the most common measurement used to assess collaboration relies on calculating this productivity. The two most popular methods used in counting publication productivity are adjusted or fractional counting, and unadjusted or full counting.

Adjusted counting occurs when each author or institution is given a fractional credit value of 1/n; where n represents the number of authors or institutions. Therefore, if there are five authors contributing to an article, each author will get a credit value of 0.2. If all the five authors come from different institutions, each institution is then given 0.2 credit value. If the five authors come from two institutions (for example, two authors from institution A and three authors from institution B), then institution A is given 0.4 credit value and institution B is given 0.6 credit value.

Unadjusted counting is a more straight forward measure whereby each author or institution is given full credit or a credit value of one. Therefore, if there are five authors contributing to an article, each author is given one credit value. If all the five authors come from five different institutions, then each institution is given one credit value. If the five authors come from two institutions (two authors from institution A and three authors from institution B), then each institution is given one credit value.

Long and McGinnis (1982, p. 386) have concluded that "there appear to be no major differences between adjusted and unadjusted measure". Since the objective of this research was to study collaboration behaviour (what happened), not to measure the contribution of each scientist to the collaboration (giving credit), it is appropriate to use the unadjusted counting. In this study, productivity refers to the number of papers that are published by institution, type of institution, and country affiliated with the authors, and is the metric used to indicate the extent of collaboration. When publication productivity (which refers to the number of papers) is being counted, it is logical to use whole numbers and interpretation of the data should also be more comprehensible and approachable to readers.

For the types of collaboration counts, the two-way collaboration counting technique was adopted to measure the institution, category of institution, and country collaboration. For example, in relation to data in Table I, there are four authors, each with an institutional affiliation. For the purpose of counting collaboration, this study only considers the institution (omitting department). Since two of the authors are from the same institution (Council of Scientific and Industrial Research (CSIR), India), there are three institutions are involved in the collaboration counting, namely CSIR, India (government), Malaysian Agricultural Research and Development Institute (MARDI), Malaysia (government), and Technical University of Budapest (TUB), Hungary (university).

By using the two-way collaboration counting, the first collaboration is between CSIR, India and MARDI, Malaysia. The second collaboration is between CSIR, India and TUB, Hungary. The third collaboration is between MARDI, Malaysia and TUB, Hungary. The collaboration between countries is also assessed using this same method.

For the collaboration of institutional category used in the triple helix analysis, there are only two categories involved in the article, which are government and university. Therefore, the collaboration for institutional category for this article is government-university collaboration (two-way-collaboration counting technique). If, however, CSIR, India was a private organisation (industry), then the collaboration would be government-university, government-industry, and university-industry.

Phase 2: webometric analysis

Webometric research methods have the potential to gather an overwhelming amount of data. For this reason it was necessary to restrict the number of institutions selected for the webometric component of the pilot study. The top four contributors with the highest number of papers found in the bibliometric analysis (Phase 1) were analysed.

The method of "link analysis" was employed for the webometric phase of the research, where only links in selected web sites pointing to another web sites were assessed. Thus, the links information included in web sites of the top four biotechnology contributors comprised the data for the webometric analysis. Thelwall has outlined nine steps to an information science approach to link analysis, seven of which apply to aspects of methodology. These seven steps are as follows (2004, p. 3):

(1) formulate an appropriate research question;

(2) conduct a pilot study;

(3) identify web pages or sites that are appropriate to address a research question;

(4) collect link data;

(5) apply data cleaning techniques to the links;

(6) partially validate the link count results through correlation tests; and

(7) partially validate the interpretation of results through a link classification exercise.

The first two of these seven steps were included in developing the webometric phase of this pilot study. Within step 2, the data was also examined in light of Thelwall's recommendations for undertaking steps 3 and 4. The discussion below outlines how issues relating these steps were resolved.

Formulate an appropriate research question. The research question formulated for the project is the same question that drives the implementation of a webometric analysis: that is, what patterns of collaboration are evidenced by the outputs of Malaysian biotechnology research?

The primary reason for employing webometric analysis in addressing this question is to compare the evidence for collaboration in non-traditional forms with the bibliometric analysis, given that not all collaboration will be reflected in research outputs such as journal articles. Therefore, for the webometric analysis it is of interest to look at how the webometric data for Malaysian biotechnology research collaboration compares to the bibliometric data.

Conduct a pilot study. A pilot study is necessary to examine the feasibility and validity of data. In this webometric analysis, feasibility is a means of assessing whether there will be sufficient links to produce interesting and useful results, while validity assesses whether the types of link found are consistent with the research goal.

Two questions were asked in the pilot study:

(1) Are there sufficient links to give useful results in a full-scale investigation?

(2) Are the types of links appropriate to address the research question?

In order to check for feasibility, the top four contributors were identified and the counts of links performed manually. If there were too few links from the sample web sites, two assumptions can be made:

(1) there is a need to raise the number of sample; and/or
(2) the study may not be feasible.

Thelwall (2004) suggested that 40 link counts are sufficient to give a good indication of the overall results.

Validity was tested by randomly choosing 40 links from the sampled web sites and conducting a link classification exercise in order to create an appropriate schema as recommended by Thelwall (2004, p. 37). Since the aim of this webometric analysis was to study collaboration, two classifications used for this schema:

(1) links reflecting meaningful collaboration; and
(2) links reflecting anything else.

A meaningful link was determined by the extent to which the link reflected genuine collaborative relationships, such as a link to research partners. Other links, for example internal links to general information about a department and its research activities were not classified as links that reflected meaningful collaboration.

Identify web pages or sites that are appropriate to address a research question. This stage is crucial as it determines that the right sample is being studied to address the research question. As biotechnology is the field of interest in this research, it was important to identify a valid representation of biotechnology institutions for Malaysia. The sample selected for the pilot study was a subset of the data gathered in the large ongoing study. For the webometric analysis, this was the four most productive departments identified from the bibliometric analysis of the top 20 most cited publications.

For a large institutions such as a university, a further classification was conducted where appropriate, to narrow the sample of web pages to the level of unit, department, faculty or school. In this study, this classification was necessary as it was not intending to examine university level links which reflect complex organisations that are typically multidisciplinary (Thelwall, 2004).

Collect link data. Data on web site links can be obtained in two ways:

(1) observation; and
(2) computer-assisted measurement (Park, 2003).

Observation requires the researcher to navigate individual web sites and record all the links information (where the links point to, such as individuals and organisations). This approach is time-consuming as a researcher must "surf" web sites and the many web pages within each site carefully. As Park has noted, there are advantages in using the observation method relating to the comprehensiveness of data collected. A disadvantage when dealing with large number of sites is that it can incur "a high labour cost and the possibility of coding error" (2003, p. 57).

Computer-assisted measurement, on the other hand, provides a more efficient method of gathering the links information, especially when dealing with a large number of web sites. There are two methods that can be used for computer-assisted measurement:

(1) using the advanced search and application feature of search engines such as Yahoo and Google; and
(2) using a web crawler, a program designed to automatically collect link data from the web sites.

Search engines such as Yahoo, Google and Alta Vista have, in the past, provided a platform to conduct link analysis such as listing all links to a specific web site. These services are very useful in counting the occurrences of links and pages (Holmberg, 2009, p. 118). However, most search engines have now discontinued these services since they have the potential to be of considerable commercial value.

A web crawler is a program that has the ability to search and gather every link and the content of any given web site. To date, there is only one web crawler program available which is free for non-commercial use, Web Analyst (formerly known as LexiURL Searcher), created by Mike Thelwall of the University of Wolverhampton. The platform for data searching for the Web Analyst is the Bing and Yahoo search engines. Web Analyst has the capacity to detect links in a web site and produce relevant reports. However, there were some issues that result in the Web Analyst being unsuitable for this study, including the reduced access to search engine data (discussed above), and changes to the searching and retrieval functions of Web Analyst.

Due to these reasons, the observation approach to collecting the webometric data was used for this pilot study. Although time-consuming, the number of web sites involved was relatively small. In the large ongoing study, a content analysis coding guide will be fully developed for determining meaningful links for inclusion.

In the pilot study, Thelwall's steps 5-7 were not conducted. These steps relate to data cleaning, validation of link counts, and validation of interpretation. A series of procedures are being developed for the ongoing large study in order to undertake these steps in link analysis using systematic and reliable methods.

Findings – pilot study

The pilot study examined the 20 most cited papers extracted from Web of Knowledge, with citations to the papers ranging from 44 to 154. The characteristics of the papers were as follows:

- All papers were collaborations with the number of authors ranging from two to seven.
- There were 63 authors, from 21 departments in 15 institutions.
- The 15 institutions consisted of 12 universities and three government institutions.
- There were five countries other than Malaysia involved in the collaborations.

Table II reports the rank of departments by the number of papers they produced. Three departments produced four papers; one department produced two papers; and 17 departments produced only one paper.

Table II.
Departments ranked by the number of papers

Rank	Department, institution	Papers
1	Department of Chemical Engineering, Faculty of Engineering, Universiti Malaya	4
	School of Chemical Engineering, Universiti Sains Malaysia	4
	School of Chemical Sciences, Universiti Sains Malaysia	4
2	School of Civil Engineering, Universiti Sains Malaysia	2
3	17 other departments from 14 institutions	1

Table III ranks institutions according to the number of papers they produced. From the top 20 most cited papers, ten were produced by Universiti Sains Malaysia, followed by the oldest university in Malaysia, Universiti Malaya with four papers, and Universiti Putra Malaysia with two papers. These three institutions are well-established research universities. The remainder of the institutions produced one paper each. While all the papers were produced by Malaysian institutions, Iran, Japan, Singapore, Taiwan and United Arab Emirates were associated as collaborators with one paper each. University affiliated authors had collaborated on 19 papers and authors affiliated with government institutions had collaborated on two papers.

When collaboration is examined in terms of the number of contributing authors, there appears to be higher visibility, measured by citations per paper (CPP), to publications with two authors. Table IV presents this data, showing the CPP, calculated as a mean of total citations and number of papers.

The majority of the publications were collaborations within the same department of an institution. Only three papers were collaborations between different departments from the same institution and five papers involved inter-institution collaboration, as shown in Table V with CPP indicating visibility relating to type of collaboration.

Table VI presents the national and international collaboration results for the 20 publications, with CPP calculations. Only six papers were collaborations between different institutions, of which two were national collaborations and four were international collaborations.

In terms of the triple helix analysis, it was found that only one paper was a collaboration between categories of institution; that was government-university (Table VII). While there

Rank	Institution	Papers
1	Universiti Sains Malaysia, Malaysia	10
2	Universiti Malaya, Malaysia	4
3	Universiti Putra Malaysia, Malaysia	2
4	12 other institutions from six countries	1

Table III. Institutions ranked by the number of papers

No. of authors	No. of papers	Total citations	CPP
2	4	344	86.00
3	9	572	63.56
4	2	169	84.50
5	4	193	48.25
7	1	60	60.00
Total	20	1,338	66.90

Table IV. Number of contributing authors and citations per paper

Type of collaboration	No. of papers	Total citations	CPP
Within department	12	845	70.42
Intra-institutional	3	177	59.00
Inter-institutional	5	316	63.20
Total	20	1,338	66.90

Table V. Intra/inter-institution collaboration

is only one collaboration between different categories of institution, it was found that six papers were collaborations within the same category; university-university (five papers) and government-government (one paper).

The findings of this pilot study suggest that bibliometric analyses applied to examine the extent and type of collaboration in Malaysian biotechnology research are appropriate to address the research question. The results of the pilot study, which was based on limited data, do not permit any firm conclusions to be drawn. The ongoing study will involve a very large data set and this will enable the testing of associations between different types of collaboration and visibility.

For the webometric analysis, the top four departments found in the bibliometric analysis were identified as shown in Table II. Each of the department's web sites was studied. The findings were:

- The Department of Chemical Engineering, Faculty of Engineering, Universiti Malaya does not have its own dedicated web sites. It was incorporated in the Faculty of Engineering web sites. There was links to respective researchers but only their CV available. No particular links was found that point to other departments or institutions, except to the Faculty of Engineering web sites and Universiti Malaya in general.
- The rest of the top contributors were from the same university: School of Chemical Engineering, School of Chemical Sciences and Scholl of Civil Engineering of Universiti Sains Malaysia. All three have a very comprehensive web sites. There were many links related to research however they were mainly descriptive. No institutions.

It was found that while there were links in the studied web sites, the links do not reflect meaningful collaboration. For this reason, the webometric analysis for this pilot study indicated that more data are needed to draw a valid conclusion on the feasibility of the study.

On the basis of these results of the pilot study it can be deduced that the bibliometric analysis can be used to observe the collaboration in Malaysian biotechnology research, while more data need to be analysed for the webometric analysis.

Table VI.
National/international collaboration

Type of collaboration	No. of papers	Total citations	CPP
Within institution	14	967	69.07
National	2	105	52.50
International	4	266	66.50
Total	20	1,338	66.90

Table VII.
Triple helix collaboration

Type of collaboration		No. of papers	Total citations	CPP
Different categories	Government-university	1	48	48
	Government-industry	0	0	0
	University-industry	0	0	0
Same category	University-university	5	311	62.20
	Government-government	1	60	60.00

Conclusion

The advent of e-research has facilitated collaborative research, and the near ubiquitous use of scholarly web sites by both individuals and institutions has made such collaborations increasingly transparent. The findings from bibliometric and webometric studies can illustrate a range of phenomena related to transformations in research and scholarly communication, including patterns of growth in a field's literature, characteristics of a field's research and communication practices, and citation rates. The challenge for researchers interested in research collaboration and associated phenomena is to develop or adopt both bibliometric and webometric methods that are relevant to the needs of a particular project, and responsive to the e-research context. The methods described above have been developed for a specific ongoing large project based on biotechnology research collaboration in a developing country. While they are likely to be transferrable to other science fields, the scholarly communication behaviour of humanities and social science researchers may not lend itself to the same forms of analysis, particularly the bibliometric methods.

References

Abramo, G., D'Angelo, C.A. and Di Costa, F. (2009), "Research collaboration and productivity: is there correlation?", *Higher Education*, Vol. 57 No. 2, pp. 155-71.

Adams, J.D., Black, G.C., Clemmons, J.R. and Stephen, P.E. (2005), "Scientific teams and institutional collaborations: evidence from US universities, 1981-1999", *Research Policy*, Vol. 34 No. 3, pp. 259-85.

Alcaide, G.G., Zurian, J.C.V. and Benavent, R.A. (2012), "Analysis of the internationalization process of Spanish research in science and technology (1980-2007)", *Revista Espanola De Documentacion Cientifica*, Vol. 35 No. 1, pp. 94-118.

Anuradha, K.T. and Urs, S.R. (2007), "Bibliometric indicators of Indian research collaboration patterns: a correspondence analysis", *Scientometrics*, Vol. 71 No. 2, pp. 179-89.

BiotechCorp (2011), BioMalaysia 2011 Conference & Exhibition and the 2011 Pacific Rim Summit Opens with Strategic Announcements to Spur Malaysia's Biotechnology Industry, available at: www.biotechcorp.com.my/media/biomalaysia-announces-new-collaborations/ (accessed 25 November).

Bordons, M., Gomez, I., Fernandez, M.T., Zulueta, M.A. and Mendez, A. (1996), "Local, domestic and international scientific collaboration in biomedical research", *Scientometrics*, Vol. 37 No. 2, pp. 279-95.

Butcher, J. and Jeffrey, P. (2005), "The use of bibliometric indicators to explore industry-academia collaboration trends over time in the field of membrane use for water treatment", *Technovation*, Vol. 25, pp. 1273-80.

Economy Planning Unit (2006), *Ninth Malaysia Plan 2006-2010*, Prime Minister's Department, Putrajaya.

Etzkowitz, H. (2002), "The triple helix of university-industry-government implications for policy and evaluation", Working Paper 2002-11, Science Policy Institute, Stockholm.

Finlay, S.C., Ni, C.Q. and Sugimoto, C.R. (2012), "New methods for an old debate: utilizing reader response to investigate the relationship between collaboration and quality in academic journal articles", *Library & Information Science Research*, Vol. 34 No. 2, pp. 131-7.

Glanzel, W. and Schlemmer, B. (2007), "National research profiles in a changing Europe (1983-2003): an exploratory study of sectoral characteristics in the triple helix", *Scientometrics*, Vol. 70 No. 2, pp. 267-75.

Gonzalez-Albo, B., Moreno, L., Morillo, F. and Bordons, M. (2012), "Bibliometric indicators for the analysis of the research performance of a multidisciplinary institution: the CSIC", *Revista Espanola De Documentacion Cientifica*, Vol. 35 No. 1, pp. 9-37.

Hagen, N.T. (2010), "Harmonic publication and citation counting: sharing authorship credit equitably-not equally, geometrically or arithmetically", *Scientometrics*, Vol. 84 No. 3, pp. 785-93.

Holmberg, K. (2009), *Webometric Network Analysis: Mapping Cooperation and Geopolitical Connections between Local Government Administration on the Web*, Åbo Akademi University Press, Turku.

Ioannidis, J.P.A. (2008), "Measuring co-authorship and networking-adjusted scientific impact", *PLoS One*, Vol. 3 No. 7, p. e2778, available at: www.ncbi.nlm.nih.gov/pmc/articles/PMC2464713/pdf/pone.0002778.pdf (accessed 1 February 2012).

Irawati, D. (2007), "Understanding the triple helix model from the perspective of the developing country: a demand or a challenge for Indonesian case study?", MPRA Paper No. 5829, available at: http://mpra.ub.uni-muenchen.de/5829 (accessed 13 February 2010).

Klitjou, A., Nygaard, S. and Meyer, M. (2007), "Tracking techno-science networks: a case study of fuel cells and related hydrogen technology R&D in Norway", *Scientometrics*, Vol. 70 No. 2, pp. 491-518.

Leydesdorf, L. (2005), "The triple helix model and the study of knowledge-based innovation systems", *International Journal of Contemporary Sociology*, Vol. 42 No. 1, pp. 12-27.

Leydesdorff, L. and Curran, M. (2000), "Mapping university-industry-government relations on the internet: the construction of indicators for a knowledge-based economy", *Cybermetrics*, Vol. 4 No. 1, pp. 1-17.

Leydesdorff, L. and Sun, Y. (2009), "National and international dimensions of the triple helix in Japan: university-industry-government versus international coauthorship relations", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 4, pp. 778-88.

Long, J.S. and McGinnis, R. (1982), "On adjusting productivity measures for multiple authorship", *Scientometrics*, Vol. 4, pp. 379-87.

Lundberg, J., Tomson, G., Lundkvist, I., Skar, J. and Brommels, M. (2006), "Collaboration uncovered: exploring the adequacy of measuring university-industry collaboration through co-authorship and funding", *Scientometrics*, Vol. 69 No. 3, pp. 575-89.

Narin, F. (1976), "Evaluative bibliometrics: the use of publication and citation analysis in the evaluation of scientific activity", Working Paper Project No. 704R, Computer Horizons, Inc., Parsippany, NJ, 31 March.

Park, H.W. (2003), "Hyperlink network analysis: a new method for the study of social structure on the web", *Connections*, Vol. 25 No. 1, pp. 49-61.

Priego, J.L.O. (2003), "A vector space model as a methodological approach to the triple helix dimensionality: a comparative study of biology and biomedicine centres of two European National Research Councils from a webometric view", *Scientometrics*, Vol. 58 No. 2, pp. 429-43.

Seglen, P.O. and Aksnes, D.W. (2000), "Scientific productivity and group size: a bibliometric analysis of Norwegian microbiological research", *Scientometric*, Vol. 49 No. 1, pp. 125-43.

Thelwall, M. (2004), *Link Analysis: An Information Science Approach*, Academic Press, San Diego, CA.

Viale, R. and Ghiglione, B. (1998), *The Triple Helix Model: A Tool for the Study of European Regional Socio-economic Systems*, The IPTS Report, Vol. 20, IPTS, Seville.

Further reading

Pyle, D. (1999), *Data Preparation for Data Mining*, Academic Press, San Diego, CA.

About the authors

Sharunizam Shari is a Senior Lecturer with the Faculty of Information Management at Universiti Teknologi MARA Kedah. He is currently pursuing his doctoral degree at the Department of Information Studies, Curtin University under the supervision of Dr Gaby Haddow and Associate Professor Paul Genoni. Sharunizam Shari is the corresponding author and can be contacted at: s.shari@postgrad.curtin.edu.au

Dr Gaby Haddow is a Senior Lecturer with the Department of Information Studies at Curtin University. Her research has focused on research communication and bibliometrics in research assessment.

Associate Professor Paul Genoni is the Head of the Department of Information Studies at Curtin University. He has published widely on matters related to collection management, research assessment, and e-research.

This article has been cited by:

1. Filistéa Naudé. 2016. Country Trends and Scholarly Collaboration in the ICT4D Research Community 2000-2013: A Single Journal Study. *The Electronic Journal of Information Systems in Developing Countries* **72**:1, 1-18. [Crossref]
2. Filistéa Naudé. 2016. Electronic Journal of Information Systems in Developing Countries (2000 - 2013): A Bibliometric Study. *The Electronic Journal of Information Systems in Developing Countries* **72**:1, 1-23. [Crossref]
3. Jane G. Payumo, Taurean C. Sutton. 2015. A bibliometric assessment of ASEAN collaboration in plant biotechnology. *Scientometrics* **103**:3, 1043-1059. [Crossref]