

Water Research Productivity, Trends and Collaborations in Malaysia between 1964 to 2012

Teong Han Chew¹, Mohd Shahir Shamsir^{1,*}, Zulkifli Yusop²

¹Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

²Institute of Environmental & Water Resource Management, Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

***Corresponding author email:**

shahir@utm.my

Received date: 2 Apr 2018

Published date: 31 Dec 2018

How to cite:

Chew, T., Shamsir, M., & Yusop, Z. (2018). Water Research Productivity, Trends and Collaborations in Malaysia between 1964 to 2012. *Journal of Research Management & Governance*, 1(1), 1-30. Retrieved from <https://ejournal.um.edu.my/index.php/JRMG/article/view/11325>

DOI:

<https://doi.org/10.22452/jrmg.vol1no1.1>

ABSTRACT

The research in water cycle management is an important endeavour in any nation's environmental management practices. Bibliometric examination the breadth and scope of research is one way to study past research focus. This will allow establishment of prior research trends, output and performance in order to reorient future resources, improve research collaborations and research return of investment. In this study, publications that are available on online from Web of Science and SCOPUS were examined using bibliometric analysis to visualize past research focus, institutional and author collaborations, clusters of peer research networks, lead institutions associated with the water cycle management in Malaysia from 1964 to 2012. Results showed that 80% of publications were indexed in Scopus with 5277 unique authors from 814 institutions. Significant 'alpha researcher' phenomenon was detected where strong principal investigators do not collaborate with each other, often only linked with 'bridging' individual that adopt a 'middle man' strategy, connecting two 'alpha researchers' together. Although UM is historically the initiator of many water research since 1972, UPM, USM and UKM are the current research leaders. Research universities UPM, USM, UKM and UTM possess good internal collaboration while UM only interact with external collaborators, possibly due to the highly competitive nature among themselves. Analysis of research topics showed that water pollution has been a perennial research interest since the 1990s while the geological focus has been on the river and coastal areas with studies on lake and islands receiving the least attention. Future awarding of research resources should take into account these historical insights in order to improve research allocations and institutional collaborations.

Keywords: water research; bibliometric analysis; network analysis; data visualization

1. INTRODUCTION

Water is essential to human; either to human body needs itself (Sawka and Cheuvron, 2005) or as support for human activities (Gleick, 1996). Although it is estimated that roughly 71% of earth is covered by water, only less than one percent of the water is accessible to human and therefore suitable for human activities (consumption, agriculture, development, etc.) whereas 97% of the water is saline with

North and South Poles constitute around two percent (Eric et al., 2001). Fortunately, this portion of “usable” water (or freshwater) is constantly and naturally recycled, a process known as hydrological cycle. This cycle is heavily dependent on climate, sources (of the water being extracted from) and of course human intervention to the nature. Some natural hydrological cycle may take as low as two weeks; others may take years (Taikan Oki and Kanae, 2006). Without the optimization of water withdrawal and usage, we will face water scarcity, an event that has already taken place in certain regions due to geographical factors. Worse, it is estimated that water resources issue would be even more critical, with the current trends of drastic climate change and the increase of human population growth, water demands will have higher priority than the global warming by year 2025 (Vörösmarty et al., 2000).

Such importance prompted research in a wide range, applicable to water including water cycle management, climate change (with its effect towards water supply) and even biodiversity. Understanding the trends of such research not only provide us with the correlation of current issue of interest and amount of research being conducted to study them, but also as an indicator to evolution of research (e.g. basic sciences research compared to applied sciences research). One of the methods of studying such phenomena is the bibliometric analysis.

The term bibliometric was made famous by Alan Pritchard, who defined it as the use of statistical method to analyze information regarding books and media (Pritchard, 1965). It covers a wide range of analysis ranging from the basic; number of publications over certain period of years and quality of the research (derived from the number of citations) to more complex studies such as identifying the state of growth in scientific publication (Larivi`ere et al., 2008) and formulations of a collaborative index (Liao and Yen, 2012). In addition to providing alternative perspective, in terms of research trend of a topic of interest, bibliometric analysis provides direct measurement to authors’ and institutions’ performance in scientific community (research outputs). Most importantly, such results from bibliometric analysis enable funding agencies to justify the research budget well spent.

Bibliometric analysis has been applied to various research related to water. Hagendijk and Smeenk (1989) reported their case study on Dutch freshwater ecology back in 1989 and strongly suggested that bibliometric (among others) helps in understanding the intellectual continuity of researchers with relevance to science policy. Zhang et al. (2010) adopted bibliometric study on global wetland research with a detailed analysis on the keywords used; hence the temporal trends of the research. Research on drinking water was also mapped (Fu et al., 2013; Hu et al., 2010) using bibliometric approach in which regional contributions were presented and popular journals were identified.

Bibliometric studies in water research are often topics specific (Hu et al., 2010; Fu et al., 2013; Hagendijk and Smeenk, 1989; Zhang et al., 2010) or journals specific (Wang et al., 2010, 2011) whereas water research is actually a wide research area covering basic sciences and applied sciences with major topics such as water resources management, alternative energy (hydroelectric), wastewater treatment and others. A more general bibliometric study on water research could provide even more detailed information regarding research trends and focus. Unfortunately, bibliometric studies are very much data dependent; the bigger and the more accurate the dataset is, the better the results reflect. Efforts have been done by various parties in order to index as many publications as possible. Thomson Reuters’s (formerly Institute for Scientific Information, ISI) collection of services (including the Journal Citation Reports, Web of Science, Web of Knowledge, etc.) and Elsevier’s SCOPUS, are known to be two of the most extensive academic publications indexing services. Datasets used for bibliometric analysis were mostly mined from ISI (Gleick, 1996; Falagas et al., 2006; Nazim and Ahmad, 2008; Hu et al., 2010; Rajendram et al., 2006; Francisco Mun~oz-Leiva et al., 2012; Zhang et al., 2010) while minority, were

from Scopus (Bajwa and Yaldram, 2013; Kumari, 2009). There is some literature which uses datasets from other established alternative sources as well, such as PubMed (Falagas et al., 2006; Vergidis et al., 2005) and Google Scholar (Sanni and Zainab, 2010).

Journals are normally ranked based on the number of publication and the number of citations within certain period of time. Although there are a number of different kinds of calculation to rank a journal, impact factor (Garfield, 1999) is arguably the most recognized which could be the reason majority of bibliometric studies were done on the ISI's datasets. For a more thorough representation, articles published in non ISI-indexed journals should also be considered during bibliometric studies. Although Archambault et al. (2009) indicated that bibliometric study is largely independent of databases, a study by Vieira and Gomes (2009) shows that Scopus, particularly, offered about around one third of additional publications (in addition to those available in ISI) with some of the highly cited publications included only in Scopus.

Data visualization techniques in bibliometric studies are often restricted to tables and simple xy/scatter/bar/line charts with few attempts on other representations; such as global geographical mapping of the publications (Fu et al., 2013), strategic diagram (Francisco Mun˜oz-Leiva et al., 2012), network diagram (Rosas et al., 2011; Francisco Mun˜oz-Leiva et al., 2012) and radar plot (Vieira and Gomes, 2010). Although sometimes, simple representations are good enough, the shortcomings of simpler visualizations are the limited parameters that can be visualized at one time. Moreover, numbers might not be able to highlight trends and patterns in one glance. Human is able, however, to quickly distinguish colors and sizes. Of course, we need to strike a balance between the complexity of data visualization and the easy data interpretation.

In this paper, we attempt a bibliometric analysis on the water research conducted in Malaysia using the data mined from ISI and Scopus. In addition to the basic statistical aspects of the bibliometric analysis, such as citations distribution and publications growth, authors' and institutions' networks were also analysed.

2. METHODS

The dataset used for this study covered the data from ISI's Web of Knowledge (all databases) and SCOPUS. Publications indexed in Scopus were mined using predefined search terms (Table 1). They were then organized using OutWit Hub and Zotero and exported into RIS (Research Information Systems) format. Data was merged in EndNote reference management software with the removal of duplicate items.

Additional data were also mined according to the list of (Malaysian) institutions indexed by Scopus, mostly institutions of higher learning. Data in ISI was obtained by using all publications in the Water Resource category, filtered by Country=Malaysia.

The two datasets were combined and duplications were removed with priority (of retaining the publications) given to publications indexed by ISI. Further filtering of unrelated articles was done in addition to filling in missing information of the published articles (e.g. abstract, authors, affiliations, keywords, etc.). These are done semi-automatically using scripts developed in-house with the information mined from OutWit Hub and Zotero, after which, resulting in the final dataset for our bibliometric study. Institutions' name and authors' name were standardized to the best of our ability to improve the data consistency. Additional information was added to the final dataset, specifically, the publications' category and geographical sources (of the water used in the particular publication). Several

Python scripts were written to assist in the unique author identification (and aliases), unique institution identification, data generation for analysis and chart plotting, among others.

Table 1: List of controlled search terms used to retrieve datasets from Scopus

Eutrophication Flood	Flood
Storm water Groundwater	Groundwater
Hydrology Lake	Lake
Reservoir Pond	Pond
Well (thermal/spring) Rainfall	Rainfall
River Estuary	Estuary
Sea Coast	Rainfall
Offshore Wastewater	Wastewater
Water and health issue Water conservation	Water conservation
Water pollution Water quality	Water Quality
Water analysis Water resources	Water Resources
Water issue Water supply	Water Supply
Wetlands Water	Water

3. RESULTS AND DISCUSSION

3.1. Basic statistics

A total of 2516 publications were finalized for our bibliometric study, with 489 from ISI and 2027 from Scopus. The publications span from year 1964 to 2012. There are 5277 unique authors from 814 institutions. Unique authors were identified automatically based on the author's name and his/her known affiliations. Authors were formatted as Author's Name [Affiliation1; Affiliation2]. There are, however, some authors with unknown institutions (either untraceable or affiliated to a generic street address), denoted as NULL, which is 495, including unknown authors from unknown affiliations (NULL [NULL]). Unknown institutions are also classified as NULL.

Although the total publications are 2516, the results shown in the subsequent sections might be "more" than that, namely the total publications in accordance to authors or in accordance to institutions. The fact is that we do not include weightage when assigning the number of publications to the authors or the institutions. For example, if a publication were jointed written by author 1 from institution A and author 2 from institution B, author 1 and 2 will get the number of publications incremented by one, respectively, as with institution A and B.

The distribution of publications and citations among the authors follows the hyperbolic trend (Figure 1). Large percentage of authors occupied the lower boundary of the graph for both publications and citations, with more than 3500 authors published only once and more than 1800 authors with zero citations, in which, while considered "unhealthy" trends, do follow the trends of some of the research field (Vieira and Gomes, 2010). An ideal trend for both the plot would be an exponential curve; with majority of the authors having high number of publications or citations. Publication and citation trends for institutions (Figure 2) show similar pattern. Table 2 details the distribution quantitatively. Both authors and institutions publications distributions have serious lower extremes (positive skew); with a low third quartile value.

Table 2: Statistics on the authors', their respective institutions' publications and citations distributions. Generally, all distributions are positively skewed; the most positively skewed being the authors' publications distribution (Q1=Median=Q3). Majority of authors and institutions have the lowest possible publications (1) or citations (0)

		Q1	Median	Q3	Mean	Mode
Authors	Publications	1.00	1.00	1.00	1.71	1.00
	Citations	0.00	1.00	6.00	7.87	0.00
Institutions	Publications	1.00	1.00	2.00	4.77	1.00
	Citations	0.00	2.00	13.00	24.50	0.00

3.2. Authors analysis

Top 20 authors in accordance to number of publications are highlighted in Figure 3. The angle of each arc is proportional to the number of publications (of the authors) with the radius proportional to citations (of the authors). Figure 4 provides a slightly different perspective on top authors, in this case, sorted according to citations, with the arc angle representing the citations and the arc radius representing the number of publications. Comparing both Figure 3 and 4, there is a shift in the authors' ranking, concluding that a high number of publications does not always translate to high citations (and vice versa); with only four authors belong to both top 20's. Low citations with high publications number could be due to "networking"; relying on random citations from other authors, assuming the publications is not impactful enough (less visibility). However, authors with high number of publications have the tendency to be cited more (12 out of 20 in Figure 3 have at least 100 citations). Some authors with high citations have a small number of publications (half of 20 in Figure 4 have less than five publications). These authors could be research officers, support staffs or postgraduates, rather than fulltime researchers which make sense since they would have not published as frequent. As far as publication sustainability is concerned, authors with frequent publications are generally preferred (at least with a high publications rate, the random probability of being cited is higher, as shown in Figure 6). Figure 6 shows that authors with more than six publications will definitely be cited although there are authors who published only once with high citations. Examples of these publications include review articles and publications with broad coverage (e.g. polycyclic aromatic hydrocarbons in rivers and estuaries in Malaysia). These kinds of publications often receive frequent citations due to the extensive information on the particular topic of research.

For a closer look at individual author's performance, we have come up with a "character's attributes" chart which illustrates the strength and weakness of that particular author (Figure 5). Each attribute has been normalized (with the maximum value of the respective attribute). Aziz, H.A. seems to be the most "balanced" authors in terms of all the attributes. Mokhtar, M. has the most number of collaborators but "owns" (being a corresponding author) only a small number of them in which he/she is likely to be the co-researcher in most of the publications. Yap, C.K., on the other hand, has the highest number of publications and the highest number of "ownerships". This could probably mean that he/she is the project leader (in the particular publications) and often receives funding for research (good credibility/track records). He/she, however, has a slightly weaker citations attribute. Being the most active author, Ismail, A., though has a reasonably good number of publications, does not have good citations and ownership attributes, hence being active does not guarantee the positive effects on the author's performance.

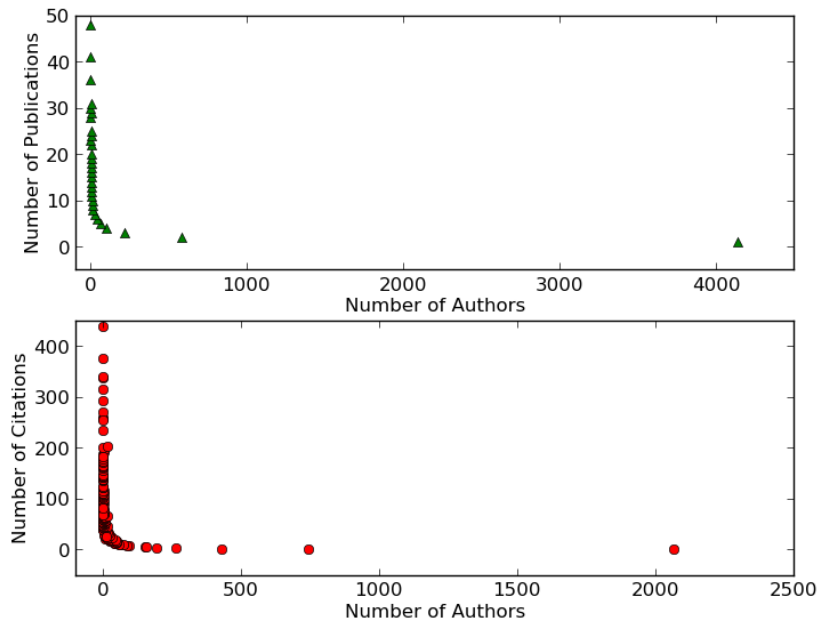


Figure 1: Publications (top) and citations (bottom) distribution for authors. More than 3500 authors published only once (top) and more than 1800 authors have zero citation (bottom). A relatively small number of authors occupy the bottom lower range in term of number of publications (top) and citations (bottom)

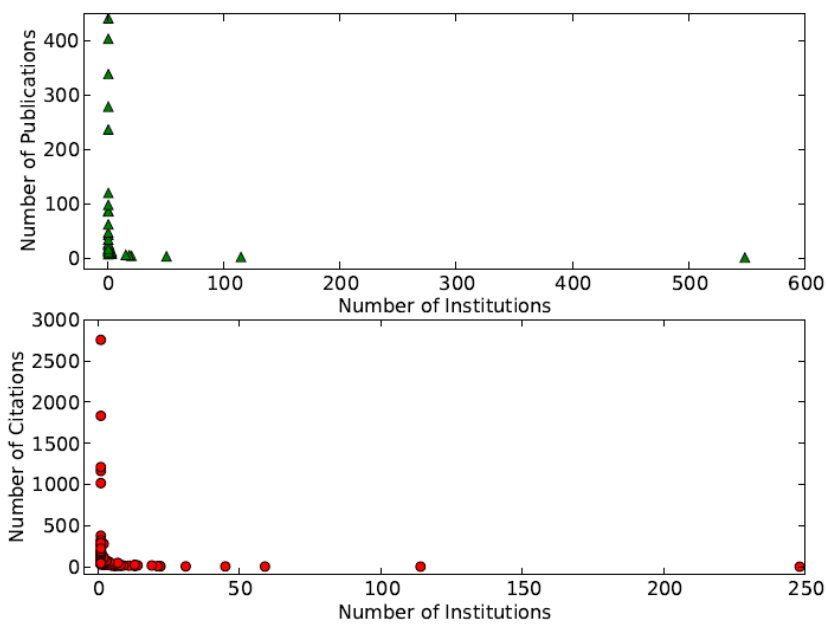


Figure 2: Publications (top) and citations (bottom) distribution for institutions. Around 550 institutions published only once (top) and slightly less than 250 institutions have zero citation (bottom). A relatively small number of institutions occupy the bottom lower range in term of number of publications (top) and citations (bottom)

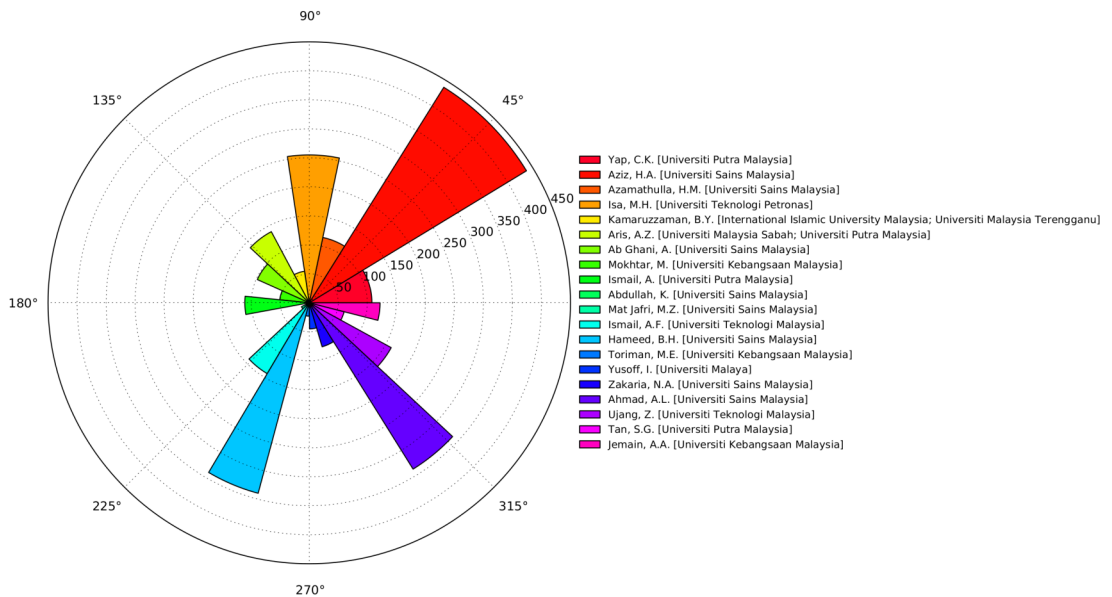


Figure 3: Top 20 authors sorted by number of publications; arc angle \propto number of publications and arc radius=number of citations

3.3. Institutional Analysis

Top institutions' performance, sorted by the number of publications is shown in Figure 7 (with Figure 8 as close-up). All of the institutions' names are abbreviated (Table 3). Only one of the top 20 institutions is a non-Malaysian institution (Kyoto University). Since the datasets are about water research in Malaysia, hence it is only logical that most of the top institutions are based in Malaysia. Generally, the number of authors is directly proportional to the number of publications (an almost linear plot in Figure 7 and 8), with few exceptions (clearly shown in Figure 8). Citations trend, on the other hand, is less dependent on both parameters. USM, being the institution with the second highest number of publications, tops the citations (Figure 4 shows that six authors with highest citations are in fact from USM). This situation is even more apparent in the dotted area in Figure 7 (see Figure 8 for close-up). FRIM possessed better citations count than most of the institutions in Figure 8 except for UMS and UTP.

Table 3: List of abbreviations used for institutions

Abbreviation	Full Name
UPM	Universiti Putra Malaysia
USM	Universiti Sains Malaysia
UKM	Universiti Kebangsaan Malaysia
UTM	Universiti Teknologi Malaysia
UM	Universiti Malaya
UiTM	Universiti Teknologi MARA
IUM	International Islamic University Malaysia
UMT	Universiti Malaysia Terengganu
UTP	Universiti Teknologi Petronas
UMS	Universiti Malaysia Sabah
UNIMAS	Universiti Malaysia Sarawak
MNA	Malaysia Nuclear Agency
UMP	Universiti Malaysia Pahang
FRIM	Forest Research Institute Malaysia
UTHM	Universiti Tun Hussein Onn Malaysia
UniMAP	Universiti Malaysia Perlis
Kyoto U	Kyoto University
UTAR	Universiti Tunku Abdul Rahman
MMU	Multimedia University
UniTEN	Universiti Tenaga Nasional
DID	Department of Irrigation & Drainage Malaysia

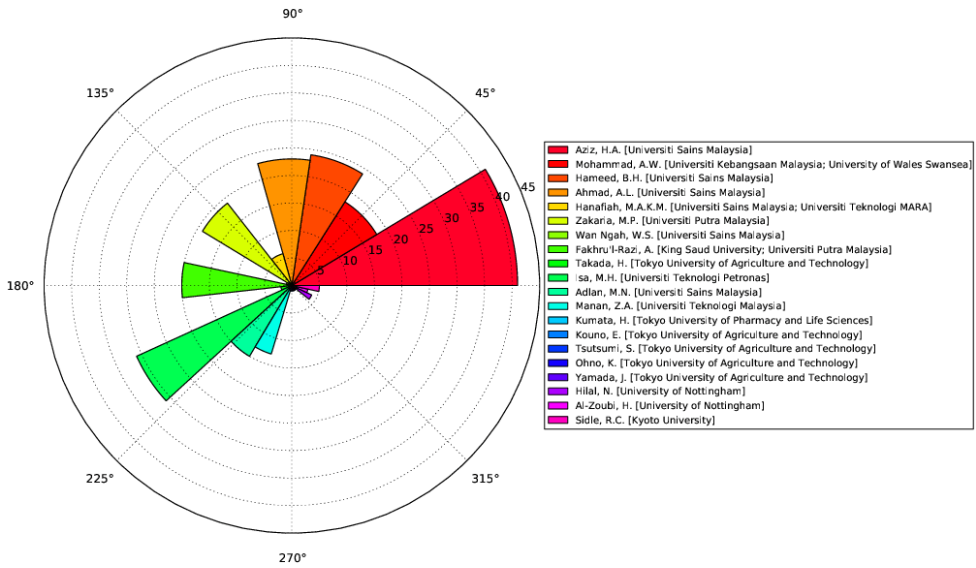


Figure 4: Top 20 authors sorted by number of citations; arc angle α number of citations and arc radius=number of publications

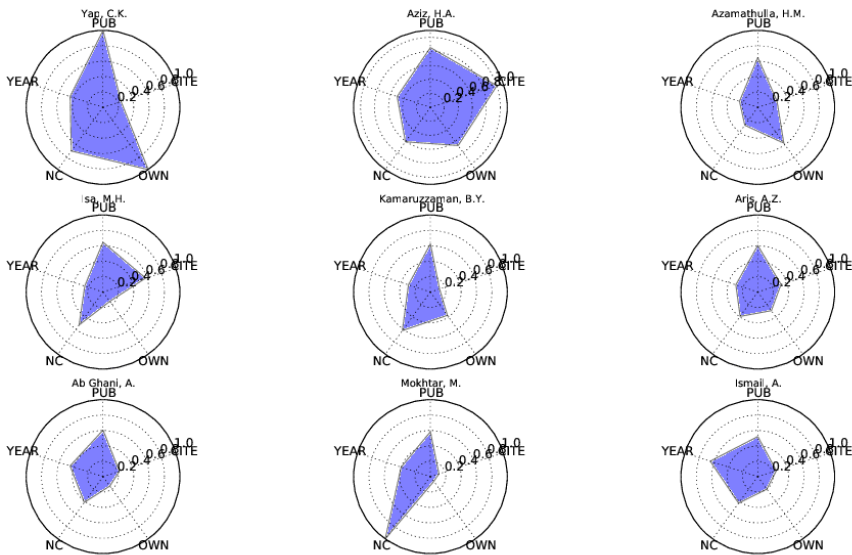


Figure 5: Performance chart of the top nine authors sorted by number of publications; PUB=number of publications, CITE=number of citations, OWN=number as corresponding author, NC=number of collaborators (authors) and YEAR=number of year active (at least one publication)

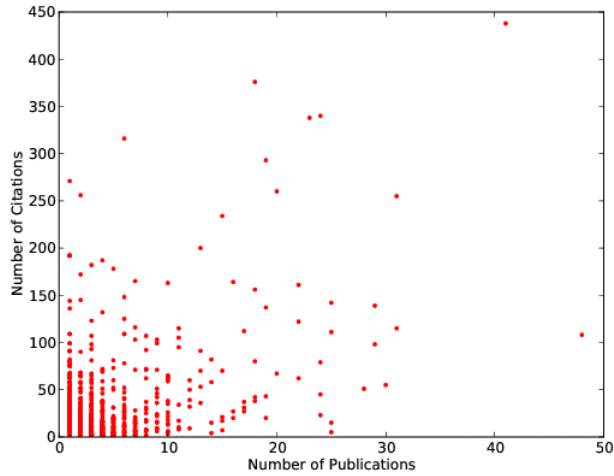


Figure 6: Authors' citations-publications distribution

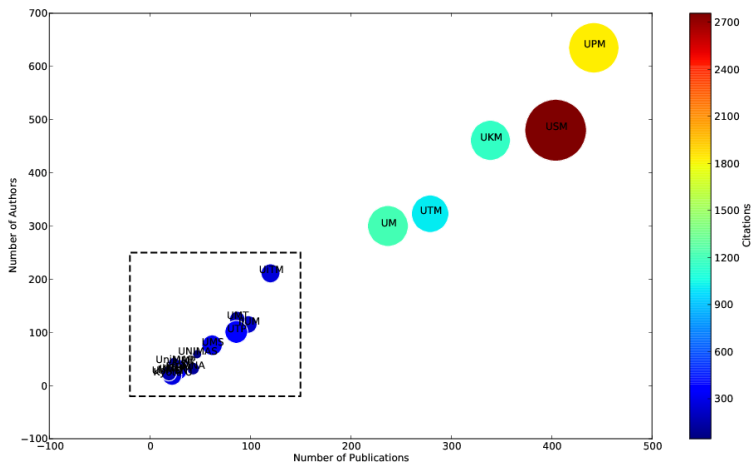


Figure 7: Top 20 institutions' trend on number of authors, number of publications and citations, sorted by the number of publications. The dotted area is enlarged as Figure 8. The size of the bubbles is \propto the number of citations (also highlighted as cool-warm color scheme). A clear split exists between two groups of institutions; those with more than 200 publications and those without

The top five institutions in Figure 7 are classified as research universities (RUs) in Malaysia, which have more research capacities in terms of number of researchers, equipment and budgets. The ranking could also be driven by funding received; unfortunately, we do not have sufficient data to prove that hypothesis. These RUs can be seen clearly leading the pack, with the nearest non-RU institution, IIUM is approximately 100 publications behind.

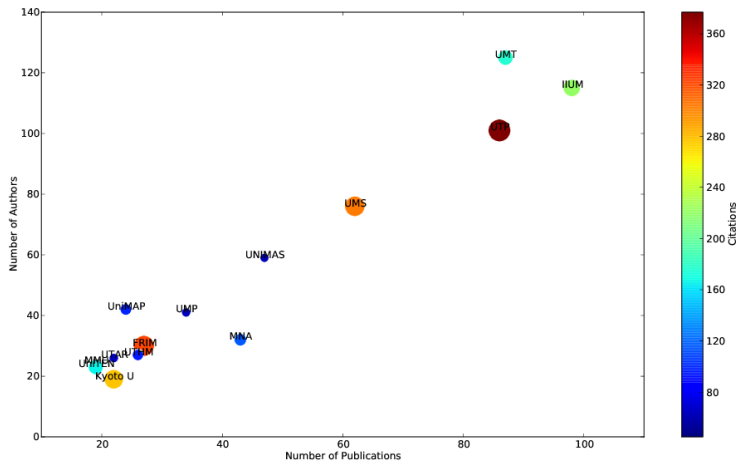


Figure 8: Close-up of the dotted area in Figure 7. While generally, the number of publications still correlates well with the number of authors, there are exceptions, such as UMT and UMP

Ranking could also be affected by the shift in the institutions' research focus. UTM, a traditionally technology-based institution is able to be ranked at number four in water research in Malaysia despite the fact that water research is a very diverse area of research. Famous for its law studies, IIUM is also gaining traction in water research in Malaysia. Non-public/private institutions can be seen populating the lower rank (Figure 8), such as MMU and UNITEN but not UTP. Although funding to these private institutions from the government is limited, research in UTP is partly collaborated with PETRONAS (an oil and gas company closely related to UTP), which actually does a lot of research regarding offshore and drilling with some of them related to water research in Malaysia. A rather surprise institution in the top 20 is the Malaysian Nuclear Agency (MNA). Though the main focus of MNA is nuclear research, a lot of projects involving heavy metals can be considered as water research related.

The yearly publications trend can be observed in Figure 9. UKM and UM can be seen as the pioneer in water research in Malaysia (as far as top 20 institutions with highest publications are concerned) with publications as early as 1972. They were joined by UTM, USM and UPM in year 1984-85. The rest of the institutions started to contribute to publications either in late 90s or early 2000s. The nature of starting involvement (year) of the institutions could be because of the establishment of the institutions (RUs are established earlier than the rest). As for the publication's growth, the top eight institutions generally recorded a yearly increment (denoted by the cool-warm color scheme in Figure 9).

Generally, as the number of new authors increases, the number of publications increases, following an exponential trend (Figure 10). However, the increase of number in new authors is more than the increase in publications; with the gap becoming obvious since year 2000. The increase in this gap could lead to a worry sign that too many authors are sharing the same publications. While some of the researches are indeed in a large scale, which more than five authors are significantly involved in, publications authorship should be limited to those who contribute considerably (authorship ethics). Pressure from the institutions and individual performance index/assessment (for salary increment/promotion) could be two of the reasons this trend is happening.

The top 20 institutions can be ranked similarly using the attributes radar plot (Figure 11). There is not much surprise here as the RUs have better attributes compared to the rest. UPM has the most "maximum" attributes with only citations being bested by USM. We do not see any significant or unique

trend for institutions unlike the authors' attributes radar plots. Figure 11, however, shows that there is a huge gap between the RUs and the rest.

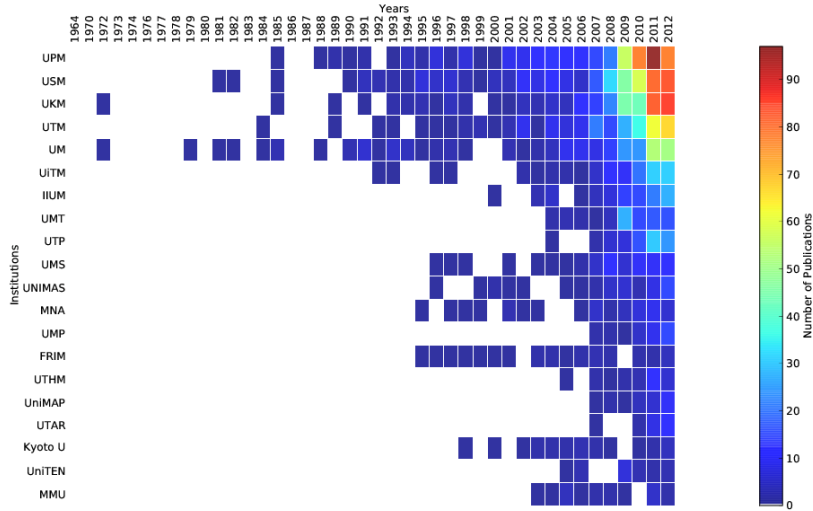


Figure 9: Yearly publications recorded for top 20 institutions; with white=no publications. All institutions seem to be constantly publish something about water research in Malaysia once they have started their first with the publication-less gap ranging from one to four years (from 1990 onwards)

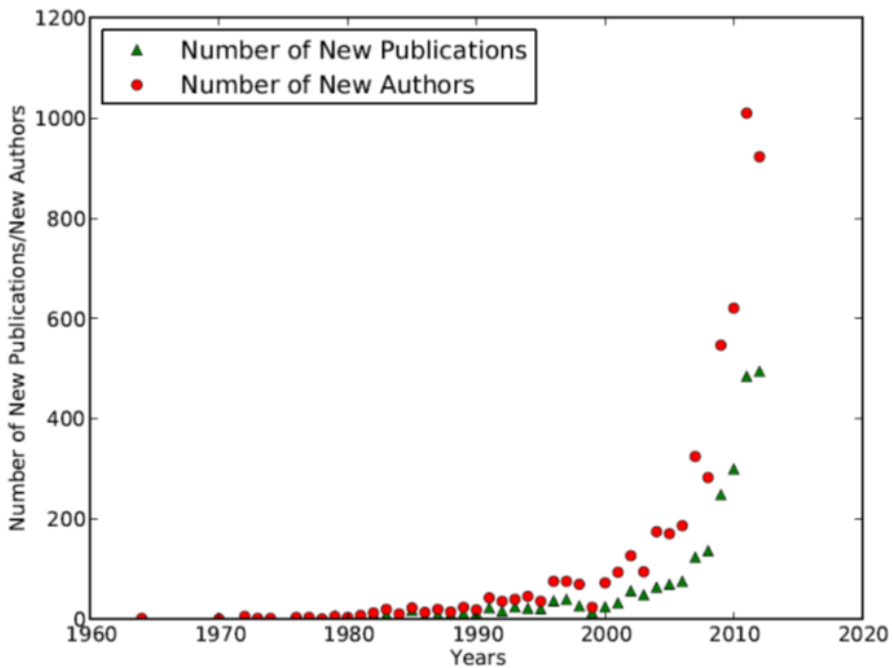


Figure 10: New authors and new publications by years

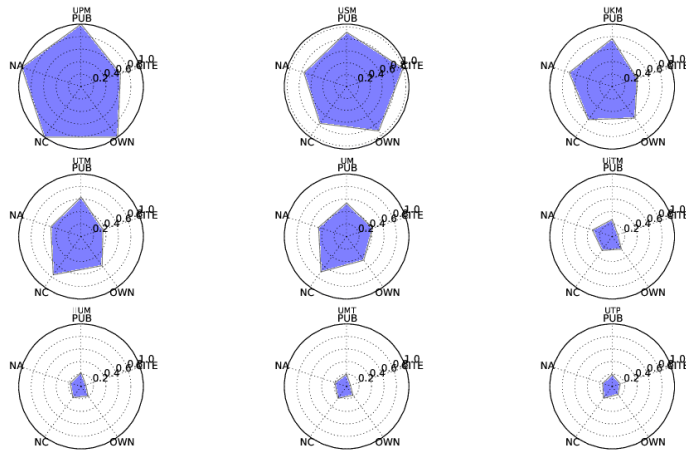


Figure 11: Performance chart of the top nine institutions sorted by number of publications; PUB=number of publications, CITE=number of citations, OWN=number as corresponding institution, NC=number of collaborators (institutions) and NA=number of authors

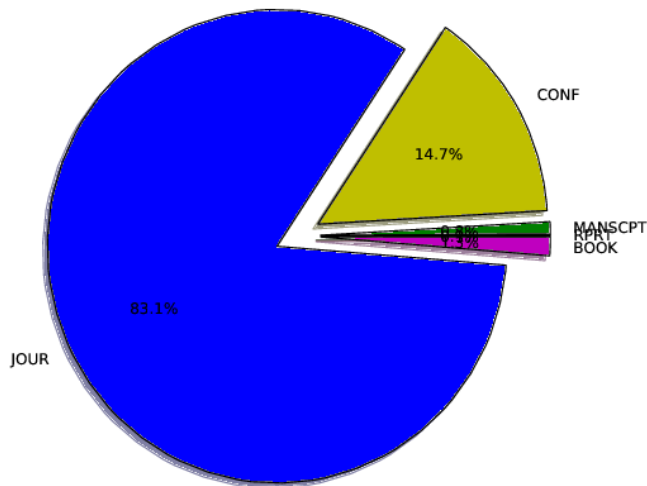


Figure 12: Distribution of publication types. Majority (83.1%) of the articles are JOUR followed by CONF at 14.7% with the rest scored less than 2% each

Five types of publications can be identified from the datasets, namely book (BOOK), report (RPRT), conference (CONF), manuscript (MANUSCPT) and journal (JOUR), with journal dominating at 83.1%, followed by conference at 14.7% (Figure 12).

Publications increase yearly with journal articles being the major contribution each year, which itself, is increasing as well (Figure 13). Conference articles are enjoying the increment, too, but at a slower pace. The publications are then categorized into seven categories, detailed out in Table 4, and ranked (fractional ranking) yearly as in Figure 14. Pollution is often ranked highest across year 1964–2012, being the preferred area of research. Pollution is getting more attention by the year because of nature degradation as a result of human activities. Energy and drainage categories are evidently the “coldest” area of research across the years. In our dataset, energy category often involves hydropower

(in Malaysia) and in order to conduct research, access to the dam is required which is fairly restricted. Drainage category, on the other hand, focuses on agriculture (according to our dataset); and it is only a small subset of drainage-based research. This can actually be expanded to drainage in terms of cityscapes, for example. With the annual recurring flood in Malaysia, drainage ought to be given a better attention.

We then took a look at the institutions' and authors' expertise for each category via simple ranking based on number of publications found in the respective categories (Table 5 and 6). Although there are only six unique institutions dominating the top three for each category, we can conclude that UKM is generally expert in water management and climate change types of studies whereas UPM excels in pollution, biodiversity and drainage types of researches. UTP surprisingly overtook RUs in energy category mainly because of its affiliation with PETRONAS.

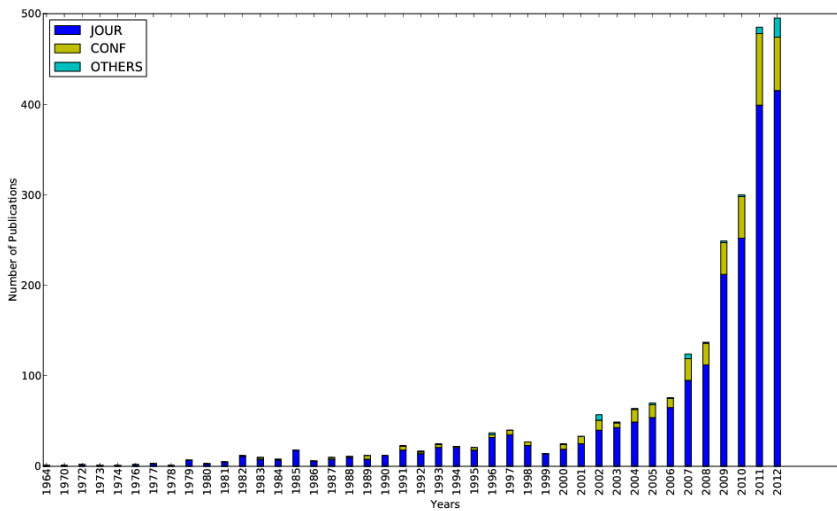


Figure 13: Yearly distribution of publication types. All publications types are generally increasing with JOUR at a faster pace compared to CONF

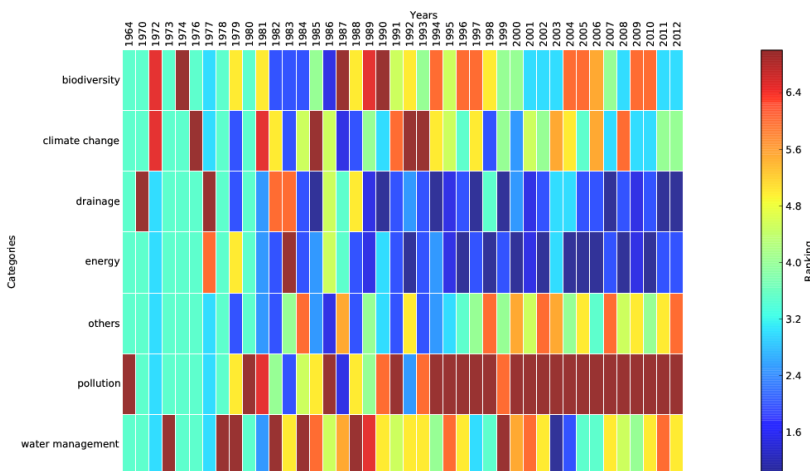


Figure 14: Yearly ranking of categories. Pollution is the “hottest” topic whereas energy and drainage are two of the lowest ranked categories

Table 5: Institutions' expertise based on categories in Table 4

Categories	Institutions		
	1 st	2 nd	3 rd
Water management	UKM	UPM	UTM
Climate change	UKM	UTM	USM
Pollution	UPM	USM	UKM
Biodiversity	UPM	UKM	USM
Drainage	UPM	UM	USM
Energy	UTP	UKM	UPM
Others	USM	UTM	UPM

Aris, A.Z. [UMS; UPM] is the only author that appeared twice as one of the top three in two categories: water management and biodiversity (Table 6). Comparing Table 5 and 6, we can identify authors with significant contribution to their institutions. For example, Mokhtar, M. [UKM] being the only author from UKM (Table 6) out of top three in water management could be the one that contribute significantly in water management that cause UKM to rank first in water management (Table 5). Only two categories do not tally in both tables: biodiversity and energy; however, this is subjected to argument that there is more than one author who scores similar rankings as the top three in Table 6. Arshad, A. [UPM] is ranked similar to Aris, A.Z. [UMS; UPM] and Ahmad, Z. [MNA; UKM]; meaning UPM actually has two authors instead of one in the top three. In energy, there are five authors with similar scores, which include the three in Table 6, plus one from UTP and one from UKM.

We did further articles analysis based on the source of water mentioned or used in the publications. The publications were classified based on the keywords (Table 7, with some regular expressions), title and abstract, with manual intervention if more than one water sources mentioned.

The result showed that four major geological categories captured the most attention, namely RIVER, COAST, WASTE and RAIN (Figure 15). Two least geological categories are LAKE and ISLAND. Around 14.3% of the publications did not mention the water sources used in the research.

Most publications categorized as LAKE are those involving large lakes as a result of dam constructions such as Kenyir, Chini and Titiwangsa. This correlates well with the result in Figure 12 (in which energy category research is one of the least published about). With RIVER as the top geological category, we can assume that it is because of the direct effect to human, especially when it comes to water consumption. COAST geological category could be attributed to the geographical nature of Malaysia which is surrounded by sea. Some of the publications in COAST also mention hydrocarbon or crude oils. WASTE focuses mainly on industrial by-products such as palm oil (agriculture) and dye (textile) wastewater. Membrane is also frequently mentioned in WASTE publications as a treatment method. Malaysia's climate (heavy rains and frequent floods) is obviously the reason RAIN category is one of the four top geological categories. A lot of publications in UNDEFINED geological category are either laboratory works (with generic water source) or policies and social sciences researches.

Table 6: Authors' expertise based on categories in Table 4. Note that there is more than one author with similar scores in the top three in categories with *

Categories	Authors		
	1 st	2 nd	3 rd
Water management	Mokhtar, M. [UKM]	Manan, Z.A [UTM]	Aris, A.Z [UMS; UPM]
Climate change	Jemain, A.A [UKM]	Deni,S.M [UITM]	Toriman, M.E [UKM]
Pollution	Yap, C.K [UPM]	Aziz, H.A. [USM]	Ismail, A. [UPM]
Biodiversity	Kamaruzzaman, B.Y. [IIUM, UMT]	Aris, A.Z [UMS; UPM]	Ahmad, Z. [MAN;UKM]
Drainage	Lee, T.S. [UPM]	Amin, M.S.M [UPM]	Rowshon, M.K. [UPM]
Energy	Choy, F.K. [TNB]	Cheong, B. [Schlumberger]	Daungkaw, S. [Schlumberger]
Others	Azamathulla, H.M [USM]	Marghany, M. [UTM]	Ismail, A.F. [UTM]

Table 7: Geological Classification

Classification	Words used for classifying
RIVER	stream, river, fresh (-)water, potable, ground(-)water, sg(.), sungai, drinking
RAIN	rain, rain(-)fall, storm(-)water, run(-) off, catchment,
LAKE	reservoir, lake, hydroelectric, tasik, dam
COAST	beach, sea, tsunami, estuary(y ies), coast, strait, sea(-)water, off(-)shore, kuala
ISLAND	island, pulau
WASTE	water(-)water, effluent, contaminated, polluted, leachate, sludge

Yearly trends (Figure 16) show that COAST, RIVER and WASTE are getting more attention in research in recent years compared to the rest of geological classification. In fact, during 1988–1993, COAST was ranked first. ISLAND mostly remained the “coldest” geological classification (although ranked first in 1978 since it was the only paper in 1978) whereas LAKE was ranked first twice: in 1973 and 1983. WASTE was not exactly the “hottest” research area until year 1999 onwards. This could be due to the industrial demands. With the availability of membrane treatment of waste water for the past ten years or so (Atkinson, 2006), research on membrane and waste water treatment is getting more attention.

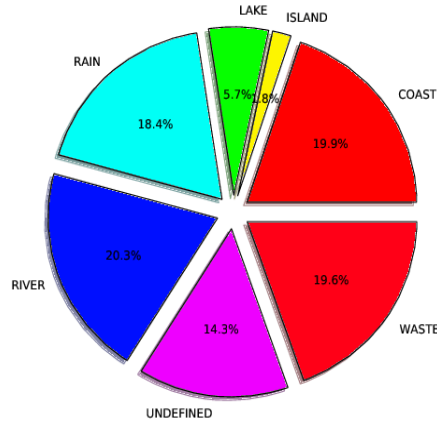


Figure 15: Distribution of publications based on geological classification (Table 7)

3.4. Authors' Collaboration Network Analysis

Gephi network visualization software (Bastian et al., 2009) was used to visualize and analyze our datasets in terms of authors networking and institutional networking. All the networks are undirected. There are 5277 authors (nodes) with 12227 interactions (edges) in our authors' network with 161 non-interacting authors and 514 authors who only interact once. The overview of the network is shown in Figure 17. Nodes' size is proportional to the nodes' degree and all the nodes are coloured according to "communities" detected via modularity analysis. The largest sub-network consists of 2465 nodes with 7397 interactions (Figure 18). We can now see a little bit more clearly on the communities. Only three out of top 50 authors, sorted by number of publications are not in the largest sub-network whereas all top 50 authors sorted by number of collaborators are in the largest sub-network. We then filtered the largest sub-network to observe any common 1st level "neighbours", with exception to Aziz-Isa. Aziz, H.A. [USM] and Azamathulla, H.M. [USM] do not share any collaborator although they are from the same institutions. (Figure 19). Only Aziz, H.A. [USM] (2nd) collaborates with Isa, M.H. [UTP] (4th) for 22 times directly. The other authors seemed to have their own "exclusive" collaborators and are not connected either directly or through any common 1st level "neighbours", with exception to Aziz-Isa.

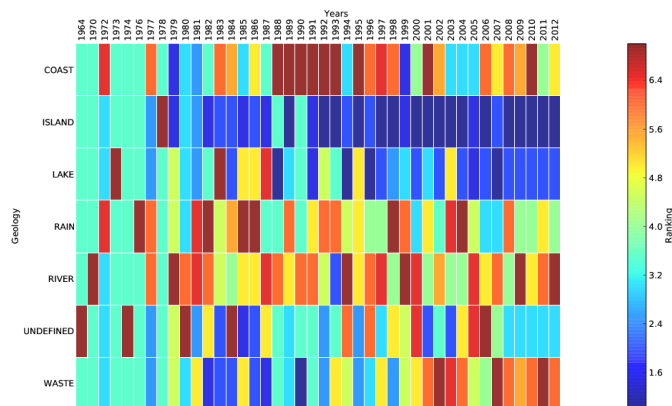


Figure 16: Yearly distribution of publications based on geological classifications. RIVER and WASTE are often ranked higher than the rest while ISLAND and LAKE are two of the least popular areas of research in terms of geological classification

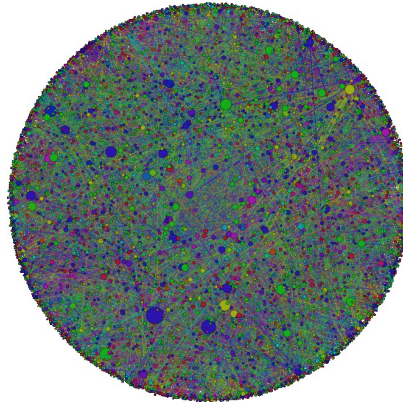


Figure 17: Overall authors' collaboration network via ARF layout. Color represents groups/communities detected by Gephi and nodes' size is proportional to the node's degree

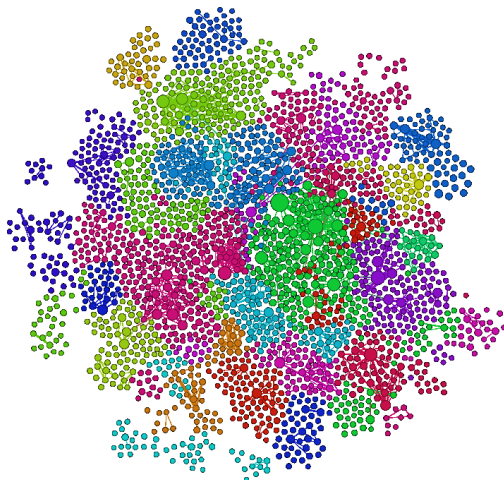


Figure 18: The largest authors' collaboration sub-network via OpenOrd layout. Colour represents groups/communities detected by Gephi and nodes' size is proportional to the node's degree

If we look at the network generated by the top five authors sorted by the number of collaborators, Mokhtar, M. [UKM] collaborates with Toriman, M.E. [UKM] ten times directly. Kamaruzzaman, B.Y. [IIUM; UMT] (5th) remains "isolated" from the other top four authors. Yusop, Z. [UTM] acts as the "middle-man" between Ujang, Z. [UTM] and Toriman, M.E. [UKM] whereas more than one common node connects between Yap, C.K. [UPM] and Mokhtar, M. [UKM], and, between Yap, C.K. [UPM] and Toriman, M.E. [UKM], notably Surif, S. [UKM]. Yusop, Z. and Surif, S. could be adopting the "middle-men" strategy to increase their publications. However, considering they both connect between high-degree nodes, they could play a bigger role in bringing big players together (e.g. Yap and Mokhtar or Ujang and Toriman), thus resulting in even bigger and more solid network. In fact, in the centrality analysis of the network, Yusop has the highest betweenness centrality score, which makes him/her the critical personnel in collaboration across different communities. Authors with high betweenness centrality normally have high inter-institutional collaborators/inter-institutions ratio (Table 8) but in the case of Mokhtar and Ujang, they rank relatively low even though they have high intra collaborators.

Another important observation is the eigenvector centrality of the network. Eigenvector centrality often gives a picture of the leader of highly connected communities (Figure 21). Authors such as

Mokhtar and Toriman are connected to several other high-profile authors in the cluster and thus, they are considered two of the most influential authors in their communities.

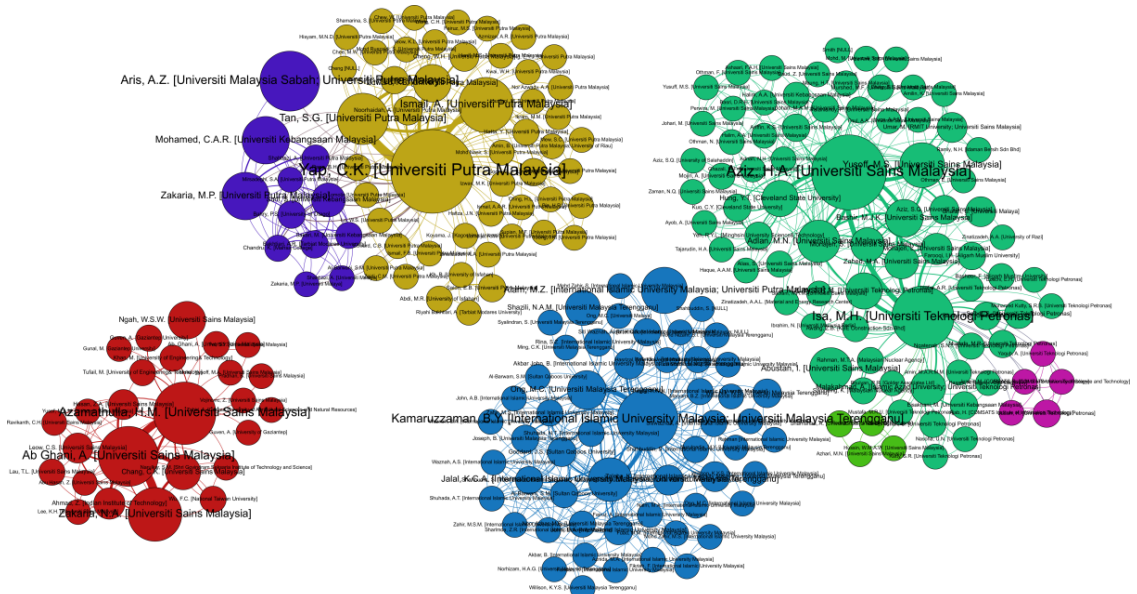


Figure 19: Top five authors' collaboration sub-network sorted by number of publications. Colour represents groups/communities detected by Gephi and nodes' size is proportional to the node's number of publications

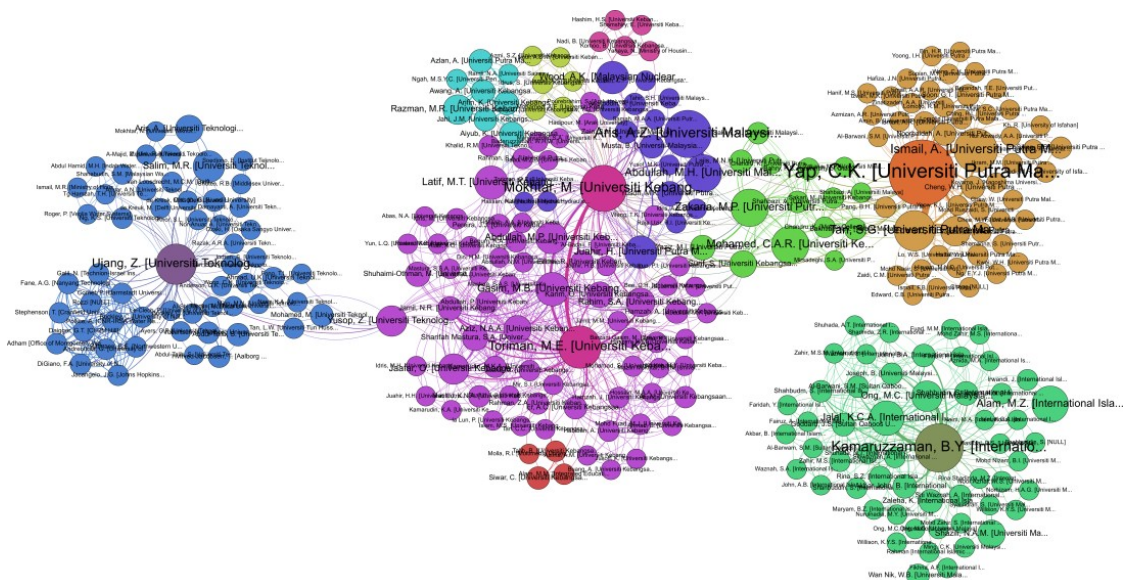


Figure 20: Top five authors' collaboration sub-network sorted by number of collaborators. Colour represents groups/communities detected by Gephi and nodes' size is proportional to the node's number of publications

Since most of the members in the cluster are from UKM, we can conclude that both authors are the leaders among UKM community. Table 9 lists the top five authors with highest eigenvector centrality in the overall network and in general, they have a large number of intra institutional collaborators

compared to inter institutional collaborators. Considering Mokhtar and Juahir both have a top five ranking of betweenness centrality and eigenvector centrality, both of them enjoy working among their local colleagues as well as having the ability to collaborate with those outside of their local communities. They would serve as the best example of striking a balance intra and inter partnership.

Surprisingly, Yap, who has the most number of publications and number of collaborators, does not appear in any top five centralities. It could be that, his network, although has the highest degree (connections), he does not connect to other nodes (Figure 22) often, especially with nodes from other communities, whereas Mokhtar's community network is denser compared to Yap's (Figure 22).

We also take a look at the authors' network in top institutions (Figure 23). These networks only map the connections between authors in the particular institution. Relevant nodes and edges were extracted and communities were detected by Gephi before being exported to visualization by Circos (Krzyszowski et al., 2009). The interactions between communities decreased as we move from Figure 23a to 23f although the number of communities is roughly similar. UPM, USM, UKM and UTM could be considered having good intra collaboration among their respective communities with at least 10% of the communities is interacting among each other. The interactions among communities are driven by the community size and community strength (number of publications) with interactions dominating the right portion of the graph. In the case of UM, although it has 51 local communities, the size of each community is small compared to the rest of the RUs. Surprisingly, the largest local community in UM, U63, does not interact outside of itself. A possible explanation to this could be the fact that researchers in UM are highly competitive among each other and only collaborate with members in the same community. Similar trend is observed in UITM; however, the situation is slightly different from UM. UITM is actually a network of universities with branches in almost every state in Malaysia and in our analysis, we group all branches together under UITM. The lack of interactions among UITM local communities is simply because the researchers (although affiliated to UITM) are based in different branches of UITM. For IIUM and UMT, interactions exist among communities with high publications number and low publications number although the interactions are still very much community size dependent.

Table 9: Top five authors with the highest eigenvector centrality. Rank is based on eigenvector centrality score, Intra=number of intra institution collaborators, Inter=number of inter institutions collaborators and Inter Inst=number of unique inter institutions

Rank	Author	Intra	Inter	Inter Inst
1	Mokhtar, M. [UKM]	59	28	11
2	Toriman, M.E. [UKM]	64	6	3
3	Aziz, N.A.A. [UKM]	40	3	2
4	Gasim, M.B. [UKM]	36	8	4
4	Juahir, H. [UPM]	20	32	12

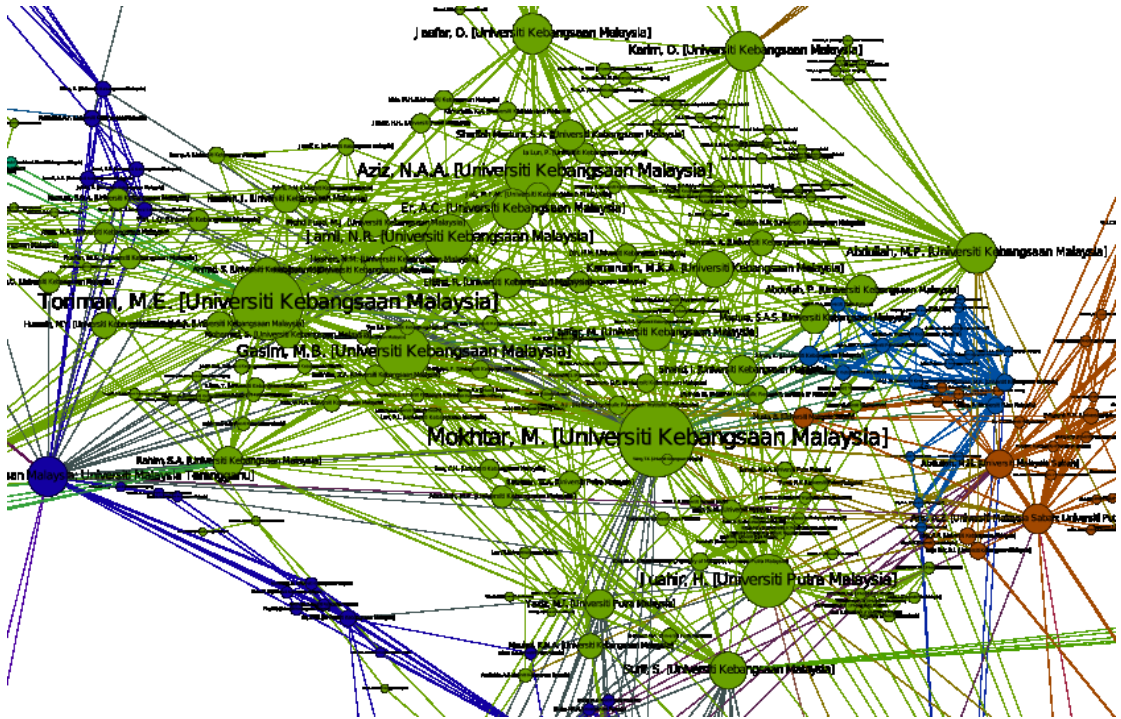


Figure 21: A highly connected clusters in authors' collaboration network showing authors with high eigenvector centrality (which is proportional to node size)

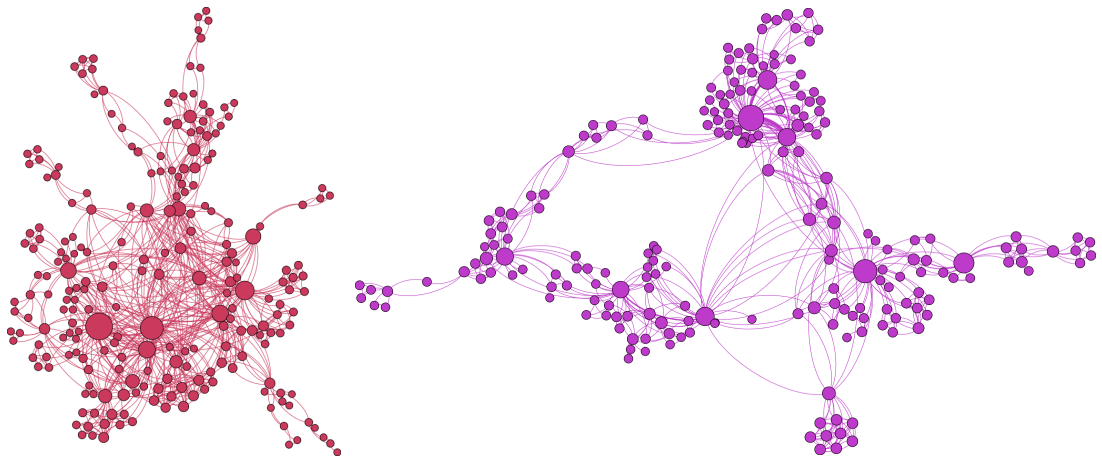
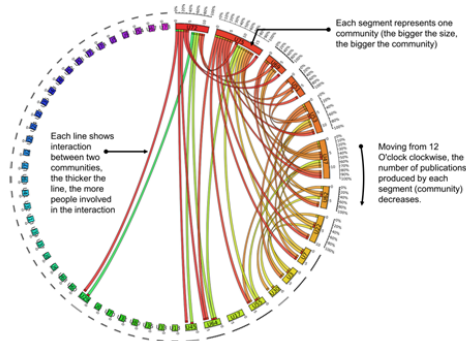
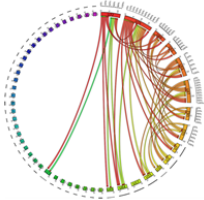


Figure 22: Comparison of two communities with different connections density with Mokhtar's community network in red (left) versus Yap's community network (right)

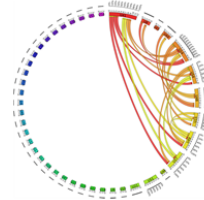
Last but not least, we observed the collaboration network among the top 20 authors (Figure 24 and Table 10). Five out of the top 20 authors do not interact among the top 20 authors; namely Kamaruzzaman (A5), Ismail (A10), Yusoff (A16), Jemain (A18) and Ujang (A20). The majority of the top 20 authors only collaborate with authors from the same institutions (true to the top 20 authors) and only four interact with authors from different institutions. The results showed that there is no significant inter institutional collaborations among top authors.



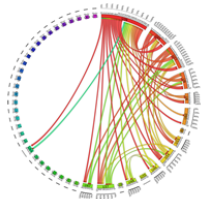
(*) Explanation of Diagram



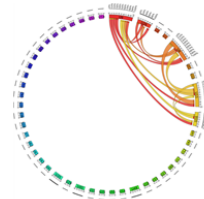
(a) UPM (50 communities with 15 interacting communities)



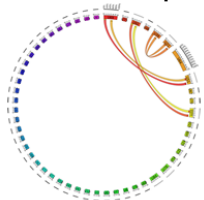
(b) USM (42 local communities with 9 interacting communities)



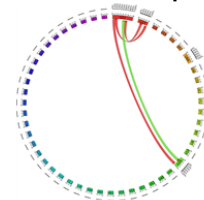
(c) UKM (45 local communities with 13 interacting communities)



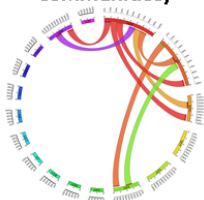
(d) UTM (46 local communities with 5 interacting communities)



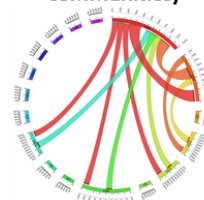
(e) UM (51 communities with 6 interacting communities)



(f) UITM (49 local communities with 3 interacting communities)



(g) IUM (17 local communities with 5 interacting communities)



(h) UMT (17 local communities with 6 interacting communities)

Figure 23 (*, a-h): Comparison of network of top institutional (local) communities. Outermost ring denotes each community detected by Gephi and only inter links (connections to different communities) are shown. Communities without any interactions (i.e. communities with only one member) are discarded. The ring segments are sorted by the cumulative number of publications (starting from 12 O'clock) and the ring size is \propto community size

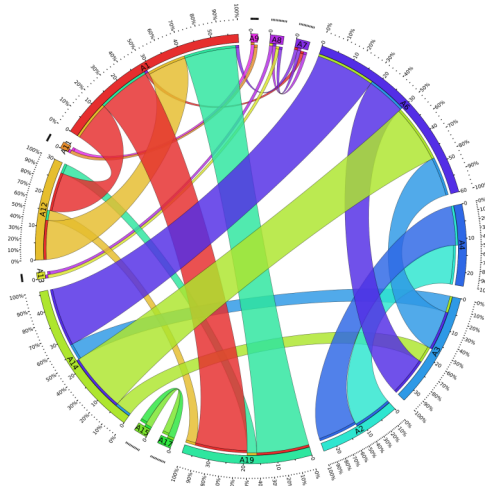


Figure 24: Top 20 author's collaboration network

Table 10: Top 20 authors with the corresponding code as in Figure 24

Code	Author
A1	Yap, C.K. [UPM]
A2	Aziz, H.A. [USM]
A3	Azamathulla, H.M. [USM]
A4	Isa, M.H. [UTP]
A5	Kamaruzzaman, B.Y. [IIUM; UMT]
A6	Ab Ghani, A. [USM]
A7	Aris, A.Z. [UMS; UPM]
A8	Mokhtar, M. [UKM]
A9	Abdullah, K. [USM]
A10	Ismail, A.F. [UTM]
A11	Mat Jafri, M.Z. [USM]
A12	Ismail, A. [UPM]
A13	Toriman, M.E. [UKM]
A14	Zakaria, N.A. [USM]
A15	Hameed, B.H. [USM]
A16	Yusoff, I. [UM]
A17	Ahmad, A.L. [USM]
A18	Jemain, A.A. [UKM]
A19	Tan, S.G. [UPM]
A20	Ujang, Z. [UTM]

3.5. Institutions' Collaboration Network Analysis

The overall institutional collaboration network is shown in Figure 25 with a total of 813 nodes (excluding NULL) and 1668 edges. The top five institutions (all RUs) scored the top betweenness centrality (highly connected to multiple communities) and eigenvector centrality (highly connected in a highly connected sub network). For the top ten institutions, UPM, UKM and UMS are grouped in the same community (detected by Gephi), USM and UTP are in another community, and, UTM–UITM and IIUM–UMT are in separate communities, respectively. Figure 27 shows the interactions among the top 20 institutions. Overall, the interactions are very much diverse but not all institutions in the top 20 collaborate with each other. The most diverse collaborations would be UKM (I3) and UTM (I4), with only no collaborations in two out of top 20 institutions. The least diverse RU would be UM (I5), which only interact with 13 out of the top 20 institutions. FRIM (I14) is the overall least diverse institutions with interactions only with four out of top 20 institutions. The highest frequency of collaboration is between UPM (I1) and UKM (I3) which is 39 times, followed by IIUM–UMT (27 times).

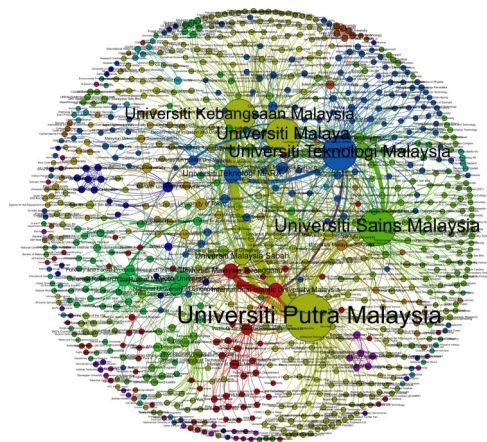


Figure 25: Institutional collaboration network

The institutional network can also be observed in terms of Malaysian and non-Malaysian institutions. There are 168 Malaysian institutions and 646 foreign institutions from our dataset. Table 12 shows the Malaysian-foreign institutional collaborators for the top 20 institutions

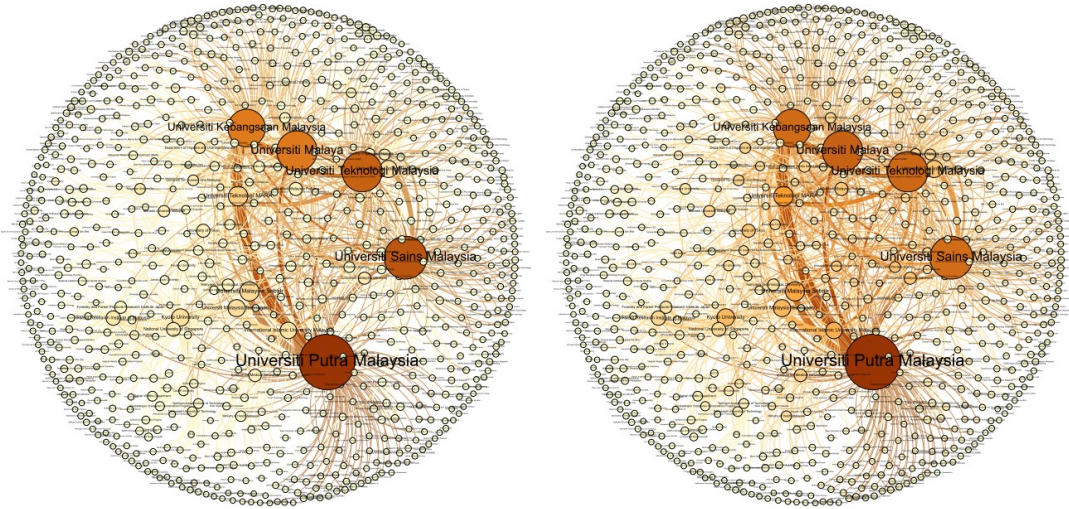


Figure 26: Institutional network with nodes (left) coloured by betweenness centrality (right) coloured by eigenvector centrality

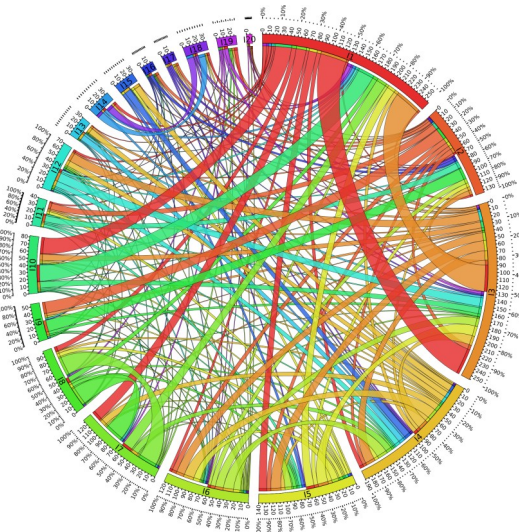


Figure 27: Top 20 Institutional collaboration network

RUs generally have more foreign collaborators than local ones, although UKM's ratio of FOR/MAL is relatively lower. IIUM, although it has smaller foreign collaborators, it manages to be one of the top ten institutions. The lack of foreign collaborators could be due to the institutions' reputation (in our case, reputation in water research). The established institutions (especially the RUs) have significant foreign collaborators compared to others. FRIM has a higher number of foreign collaborators compared to local collaborators since it is the de-facto guardian of Malaysian forest (foreign institutions will definitely need a "local" contact especially when the research is done in Malaysian forest). This could also mean that the foreign researchers are more interested in water research in Malaysian forest compared to local researchers. Kyoto University surprisingly has more foreign collaborators than Malaysian collaborators although the dataset is on water research in Malaysia.

Table 11: Top 20 institutions with the corresponding code as in Figure 24

Code	Institution
I1	Universiti Putra Malaysia
I2	Universiti Sains Malaysia
I3	Universiti Kebangsaan Malaysia
I4	Universiti Teknologi Malaysia
I5	Universiti Malaya
I6	Universiti Teknologi MARA
I7	International Islamic University Malaysia
I8	Universiti Malaysia Terengganu
I9	Universiti Teknologi Petronas
I10	Universiti Malaysia Sabah
I11	Universiti Malaysia Sarawak
I12	Malaysian Nuclear Agency
I13	Universiti Malaysia Pahang
I14	Forest Research Institute of Malaysia
I15	Universiti Tun Hussein Onn Malaysia
I16	Universiti Malaysia Perlis
I17	Universiti Tunku Abdul Rahman
I18	Kyoto University
I19	Universiti Tenaga Nasional
I20	Multimedia University

Table 12: Top 20 institutions with the number of Malaysian (MAL) and foreign (FOR) institutions (collaborators)

Institutions	FOR	MAL
Universiti Putra Malaysia	111	41
Universiti Sains Malaysia	88	25
Universiti Kebangsaan Malaysia	57	40
Universiti Teknologi Malaysia	70	35
Universiti Malaya	73	33
Universiti Teknologi MARA	22	20
International Islamic University Malaysia	12	21
Universiti Malaysia Terengganu	21	14
Universiti Teknologi Petronas	25	10
Universiti Malaysia Sabah	25	16
Universiti Malaysia Sarawak	7	13
Malaysian Nuclear Agency	4	12
Universiti Malaysia Pahang	14	9
Forest Research Institute of Malaysia	23	4
Universiti Tun Hussein Onn Malaysia	12	4
Universiti Malaysia Perlis	7	9
Universiti Tunku Abdul Rahman	12	5
Kyoto University	24	12
Multimedia University	3	6
Universiti Tenaga Nasional	1	6

4. CONCLUSIONS

In this study, historical analysis of indexed publications on water research was able to provide credible insights into past research focus, mapped collaborations, define research clusters, and identify leading institutions in Malaysia. The 814 institutions showed wide breadth of collaborations with UM being the historical lead in the research area. Significant research personas were also identified, showing research clusters created by 'alpha researchers'. Water pollution was highly focused as a research area, probably due to the pressing need and immediate impact of the applied research findings to the governance of the environment and society. The research bias on rivers and coastal areas could be due similar reasons, with accessibility to research sites being an added research incentive. Distribution of research resources should account for this research precedence in order to create a more effective and equitable research allocations. Dominant researchers identified from the publication strength present a challenge to the research community by creating research caucuses that might impeded research progress by creating personality bias or institutional research silos. Institutional diversity and transdisciplinary nature of researchers is paramount in ensuring breadth and depth of the research reach. Governing stakeholders would need to take into account all these factors when safeguarding a nation's continuous research momentum.

ACKNOWLEDGEMENTS

We would like to acknowledge Academy of Sciences Malaysia and Universiti Teknologi Malaysia for the funding of this project.

REFERENCES

- Archambault, E., Campbell, D., Gingras, Y. & Larivière, V. (2009). Comparing bibliometric statistics obtained from the web of science and scopus. *Journal of the American Society for Information Science and Technology*, 60, 1320– 1326.
- Atkinson, S. (2006). Research studies predict strong growth for mbr markets. *Membrane Technology* 8, 8–10.
- Bajwa, R.S. & Yaldrum, K. (2013). Bibliometrics analysis of biotechnology research in Pakistan. *Scientometrics*, 95, 529–540. doi:10.1007/s11192-012- 0839-x.
- Bastian, M., Heymann, S. & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks, in: International AAAI Conference on Weblogs and Social Media.
- Eric, H., Oelkers, E.H., Hering, J.G. & Zhu, C. (2001). Water: Is there a global crisis? *ELEMENTS*, 7, 157–162. doi:10.2113/gselements.7.3.157.
- Falagas, M.E., Papastamataki, P.A. & Bliziotis, I.A. (2006). A bibliometric analysis of research productivity in parasitology by different world regions during a 9-year period (19952003). *BMC Infectious Disease*, 6, 1–6. doi:10.1186/1471-2334-6-56.
- Francisco Muñoz-Leiva, F., María Isabel Viedma-del Jesús, M.I., Sánchez- Fernández, J. & López-Herrera, A.G. (2012). An application of co-word analysis and bibliometric maps for detecting the most highlighting themes in the consumer behaviour research from a longitudinal perspective. *Quality & Quantity*, 46, 1077–1095.
- Fu, H.Z., Wang, M.H. & Ho, Y.S. (2013). Mapping of drinking water research: A bibliometric analysis of research output during 1992-2011. *Science of the Total Environment*, 443, 757–765. doi:10.1016/j.scitotenv.2012.11.061.

- Garfield, E. (1999). Journal impact factor: a brief review. *Canadian Medical Association Journal*, 161, 979–980.
- Gleick, P. (1996). Basic water requirements for human activities: Meeting basic needs. *Water International*, 21, 83–92. doi:10.1080/02508069608686494.
- Hagendijk, R.P. & Smeenk, J.W. (1989). The analysis of national subfields: A case study of Dutch fresh-water ecology. *Scientometrics*, 15, 485–508. doi:10.1007/BF02017067.
- Hu, J., Ma, Y., Zhang, L., Gan, F. & Ho, Y.S. (2010). A historical re- view and bibliometric analysis of research on lead in drinking water field from 1991 to 2007. *Science of Total Environment*, 408, 1738–1744. doi:10.1016/j.scitotenv.2009.12.038.
- Krzywinski, M.I., Schein, J.E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. & Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*. doi:10.1101/gr.092759.109.
- Kumari, G.L. (2009). Synthetic organic chemistry research: Analysis by scientometric indicators. *Scientometrics*, 80, 559–570. doi:10.1007/s11192-007-1985-4.
- Larivè`ere, V., Archambault, E. & Gingra, Y. (2008). Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (19002004). *Journal of the American Society for Information Science and Technology*, 59, 288–296. doi:10.1002/asi.20744.
- Liao, C.H. & Yen, H.R. (2012). Quantifying the degree of research collaboration: A comparative study of collaborative measures. *Journal of Informetrics*, 6, 27–33. doi:10.1016/j.joi.2011.09.003.
- Nazim, M. & Ahmad, M. (2008). A bibliometric analysis on nanotechnology research. *Annals of Library and Information Studies*, 55, 292–299.
- Pritchard, A. (1965). Statistical bibliography or bibliometrics? *Journal of Documentation* 25, 348–349.
- Rajendram, R., Lewison, G. & Preedy, V.R. (2006). Worldwide alcohol-related research and the disease burden. *Alcohol & Alcoholism*, 41, 99–106.
- Rosas, S.R., Kagan, J.M., Schouten, J.T., Slack, P.A. & Trochim, W.M.K. (2011). Evaluating research and impact: A bibliometric analysis of research by the nih/niad hiv/aids clinical trials networks. *PLoS ONE*, 6, 1–12. doi:10.1371/journal.pone.0017428.
- Sanni, S.A. & Zainab, A.N. (2010). Google scholar as a source for citation and impact analysis for a non-ISI indexed medical journal. *Malaysian Journal of Library & Information Science*, 15, 35–51.
- Sawka, M. & Chevuron, S. (2005). Human water needs. *Nutrition Reviews*, 63, S30–S39.
- Taikani Oki, T. & Kanae, S. (2006). Global hydrological cycles and world water resources. *Science*, 313, 1068–1072. doi:10.1126/science.1128845.
- Vergidis, P.I., Karavasiou, A.I., Paraschakis, K., Bliziotis, I.A. & Falagas, M.E. (2005). Bibliometric analysis of global trends for research productivity in microbiology. *European Journal of Clinical Microbiology & Infectious Diseases*, 24, 342–345. doi:10.1007/s10096-005-1306-x.
- Vieira, E.S. & Gomes, J.A.N.F. (2009). A comparison of scopus and web of science for a typical university. *Scientometrics*, 81, 587–600.
- Vieira, E.S. & Gomes, J.A.N.F. (2010). Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics*, 4, 1–13. doi:10.1016/j.joi.2009.06.002.
- Vörösmarty, C.J., Green, P., Salisbury, J. & Lammers, R.B. (2000). Global water resources: Vulnerability from climate change and population growth. *Science*, 289, 284–288. doi:10.1126/science.289.5477.284.
- Wang, M.H., Li, J. & Ho, Y.S. (2011). Research articles published in water resources journals: A bibliometric analysis. *Desalination and Water Treatment*, 28, 353–365. doi:10.5004/dwt.2011.2412.

- Wang, M.H., Yu, T.C. & Ho, Y.S. (2010). A bibliometric analysis of the performance of water research. *Scientometrics*, *84*, 813–820. doi:10.1007/s11192-009-0112-0.
- Zhang, L., Wang, M.H., Hu, J. & Ho, Y.S. (2010). A review of published wetland research, 1991-2008: Ecological engineering and ecosystem restoration. *Ecological Engineering*, *36*, 973–980. doi:10.1016/j.ecoleng.2010.04.029.