بسم الله الرحمن الرحيم

Tracing and Visualizing Knowledge Diffusion

By Sana Sikander

MCS143001

A thesis submitted to the

Department of Computer Science

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

Faculty of Computing
CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD
April 2017

C.U.S.T.

CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY ISLAMABAD

CERTIFICATE OF APPROVAL

Tracing and Visualizing Knowledge Diffusion

by

Sana Sikander

MCS143001

THESIS EXAMINING COMMITTEE

S No	Examiner	Name	Organization
(a)	External Examiner	Dr. Syed Nasir Mehmood Shah	KICSIT
(b)	Internal Examiner	Dr. Nayyer Masood	CUST, Islamabad
(c)	Supervisor	Dr. Muhammad Tanvir Afzal	CUST, Islamabad

Dr. Muhammad Tanvir Afzal

Thesis Supervisor

April, 2017

Dr. Nayyer Masood	Dr. Muhammad Abdul Qadir
Head	Dean
Department of Computer Science	Faculty of Computing
Dated : April, 2017	Dated : April, 2017

ACKNOWLEDGMENT

All praise to Almighty Allah, who gave me the understanding, courage and patience to complete my MS thesis.

Thanks to my parents, brother, well-wishers, who helped me in my most crucial times and it is due to their untiring efforts that I am at this position today.

Every mission has a brain behind that vivifies that theoretical raw idea. I lucky enough that a masterful intellect, in the mind of Dr. Muhammad Tanvir Afzal was with me. I have no words to thank the laborious and the tiring contribution of my supervisor.

I express my gratitude to my kind supervisor Dr. Muhammad Tanvir Afzal for providing me opportunity to learn and enhance my knowledge. He had been ready to help and guide me throughout my MS thesis in all possible way.

Lastly, I thank my fellow mates for the stimulating discussions, working together, and for all the work we have had during this time period. Also I thank my friends of the institution Capital University of Science & Technology, Islamabad.

Sana Sikander

DECLARATION

It is declared that this research work is original piece of my own work, except where otherwise acknowledged in the text and references. This work has not been submitted in any form for another degree or diploma at any university or other institution for tertiary education and shall not be submitted by me in future for obtaining any degree from this or any other university or Institution.

Sana Sikander
MCS-143001
April, 2017

ABSTRACT

Measuring knowledge diffusion is one of the popular area of the research. The rapid growth of scientific knowledge depends on the diffusion process in which knowledge transfers from one place to another. Measuring knowledge is vital to support number of important tasks such as: formulation of scientific policies, identification of producers and consumers etc. Citation is widely used for measuring the knowledge diffusion. For measuring knowledge diffusion, four different parameters are known in the literature. However, from critical review of literature, we have analyzed that is one other parameter, country's diffusion, should be added for measuring knowledge diffusion. Different visualization techniques used to visualize four parameters. Some techniques only visualized producer and consumer, others visualized knowledge diffusion and research activity level. Moreover, some techniques visualize multiple parameters. However, there exists no visualization that visualizes all important parameters collectively.

In this thesis, we proposed a new visualization technique that visualizes all five parameters collectively. For developing the visualization we have collected the data set of Geoscience from Microsoft academic search. That data set consists of 1100 publications and 110, 000 citations. After developing visualization we have performed a user study from forty users for evaluating proposed approach. Finally, we have a comparison between state of the art approaches and proposed approach.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

An essential phenomenon for the flow of useful ideas among members of a social system is characterized as knowledge diffusion. Knowledge Diffusion is actually a process through which the knowledge is deported among different social groups. "Diffusion is the process by which an innovation is communicated through certain channels over time, among the members of a social system" (Rogers, 1962). The measurement of knowledge diffusion is vital to support number of important tasks such as: formation of scientific policies and the identification of areas that need attention or improvement in research.

The focus of this thesis is to propose a visualization approach that could visualize the four important parameters of knowledge diffusion collectively. Furthermore, we have analyzed that there should be another parameter country's diffusion. Therefore, we propose a fifth parameter country's diffusions that represent the distinct edges to the country. We have calculated the value of proposed parameter (country's diffusion), and visualized all these five parameters in one visualization.

1.1 Background

Knowledge diffusion refers to the process of spreading knowledge. Newton says "science depends on the diffusion of knowledge". Measuring the knowledge diffusion is one of the most popular areas of research. Researcher use different data sources to trace the knowledge diffusion. These data sources are citations, patent citations and co-authorship network.

Citation analysis is usually utilized to appraise the global influence of any particular author, institution and scientific document. It is utilized as a measure to trace and discover the development of data diffusion. Configurations discovered by utilizing the citation analysis assist the researchers to develop scientific policies, trace academic leadership, ranking of authors, institutions, cities or journals. Periodicals count and citation rates are used by academic rankings to measure data production and consumption globally (Garfield, 1972). We assume that citations are the currency in the knowledge exchange. Countries that receive more citations, export their knowledge to others. Countries that cite other works, import knowledge from others (Zhang, 2013).

Information visualization helps to identify hidden patterns in the data and need to be carefully crafted for certain problems. It permits decision makers to see analytics presented visually, to allow them to grasp difficult ideas or identify new patterns. Geo visualization, a method of data visualization can be utilized for the purpose of visually representing knowledge diffusion.

A primary goal of Information visualization is to communicate information plainly and successfully via statistical graphics, plots and information graphics. Because of the true way the mind processes information, using charts or graphs to visualize huge amounts of complex data is simpler than poring over spreadsheets or reports. Information visualization is an instant, easy way to mention ideas in a universal manner. It makes complicated data more accessible, usable and understandable. Users might have particular analytical tasks, such as making comparisons and identify areas that require attention or improvement. Tables are being used where users will look up a specific measurement, while charts of various types are being used showing patterns or relationships in the data for one or even more variables.

From critical review of literature, we have found four important parameters (Producers, consumer, research activity level, and diffusion rate) of knowledge diffusion. Furthermore, we have to propose fifth parameter country's diffusion. We have visualized all these parameters collectively.

Producer countries are those, which are producing major part of research publications in various domains and these countries export knowledge more than they import. While, consumer countries are those which are consuming the knowledge and these countries import knowledge more than they export (Zhang, 2013). The identification of knowledge producers and consumers are helpful in many ways like formulating of scientific policies, sharing of useful information across the globe, and ranking of authors and publications. Mazloumian, et al Proposed a network based index with the aim to measure scientific knowledge flows in order to identify the knowledge producers and consumers globally (Mazloumian, et al., 2013).

Moreover, the number of published papers in a country reflects its level of activity in scientific research (Li, 2010). The research activity level of the country shows how much research has been done in that country. The countries that have more publications have a higher research activity level and the countries that have low publications have a lesser research activity level. Therefore, the countries with a huge number of published papers can be regarded as "research center" of its kind. Through this feature we can find about the countries that are higher in research and these are the main research center in the world.

Forth one is proposed parameter that represents the number of distinct edges coming towards a country and the name of proposed parameter is country's diffusion. The number of citations coming towards a country represents the diffusion rate of the country. For calculating the diffusion rate of countries the number of citations of each country has been counted (Katy and Penumarthy, 2006). In the literature volume of information flow is used instead of diffusion rate. However, we used the term diffusion rate as compared to the volume of information flow for better understanding of diffusion among countries. The goal of country's diffusion and diffusion rate is to identify how much knowledge diffusion takes place in the country.

There exists no visualization that visualizes the four features of knowledge diffusion collectively. The focus of this thesis is to present visualization, which visualize these five parameters (Producers, consumer, research activity level, country's diffusion and diffusion rate) at the same time so, as to provide the basis for further analysis and research. Such visualization provides convenience to see and evaluate the pattern and result of knowledge diffusion. Moreover, with the help of visualization it is easier for researchers to discover the hidden information in large amount of data. After developing visualization the evaluation process has been carried out. During the evaluation process, we have made benchmark and compared our approach with the benchmark. Finally, comparison between proposed approach and the state-of art approach have been performed. This comparison shows which approach performed well and satisfied the majority of users.

1.2 Problem Statement:

The critical analyses of literature have led us to the following two problems which have been focused in this thesis:

1) State-of-the-art approaches have proposed four parameters to measure the knowledge diffusion of countries such as identification of: knowledge producing countries, knowledge consuming countries, research activity level of countries, and country's diffusion rate, however, none of these parameters is able to gauge the number of distinct countries in which the knowledge of country 'X' has been diffused.

2) Most of the contemporary approaches are able to visualize one or two knowledge diffusion parameters, just few were able to visualize three parameters, however, none of the visualization approach was able to visualize all knowledge diffusion parameters.

1.3 Purpose

The purpose of this research work is to introduce a new visualizing technique to represent knowledge diffusion more accurately. The existing visualizations represent different parameters like Producers, consumer, research activity level and diffusion rate. Therefore, we will introduce new parameter country's diffusion and visualize all these five parameters collectively. The new visualization provides convenience to observe and analyze the trend and result of knowledge diffusion.

1.4 Scope

The scope of this thesis is to investigate about the knowledge producers and knowledge consumers in the field of *Geoscience*. Moreover, the identification of countries that have highest diffusion rates and the main research center in the world.

1.5 Definitions, Acronyms, and Abbreviations

MAS: Microsoft Academic Search

IJSE: International Journal of Science Education

JRST: Journal of Research in Science Teaching

ACM: Association for Computing Machinery

1.6 Applications of proposed solution

The results of this thesis are helpful for:

1. Formulation of scientific policies.
2. Identification of areas that need attention or improvement in research.
3. Identification of consumer and producer.
4. Awareness about academic leadership.

CHAPTER 2

LITERATURE REVIEW

The process of spreading knowledge is termed as knowledge diffusion. Scientific knowledge gets rapid growth with the help of knowledge diffusion (Chen, 2004). Measurement of knowledge diffusion is very important as it helps in many areas, for example: advancement of knowledge, enhancements of society needs, and creating the competition in the universal marketplace (Gardner, 2010).

This section will represent the critical review of literature. We have found that there are certain parameters to measure the knowledge diffusion. These parameters are sometime visualized as well. Therefore, we can categorize this section in two broad categories; one will depict parameters and the other will represent the visualizations. Let's start the discussion with parameters:

2.1 Parameters used to trace the knowledge diffusion

Researchers used various data sources and parameters for the measurement of knowledge diffusion. In a scientific community, these data sources include citations (Mazloumian, et al., 2013), patent citations and co-authorship network (Gans, 2013). Citations are used in order to measure knowledge diffusion in citation based approaches. Following this, in patent citation approach, the citations received by patent are used to measure the knowledge diffusion. While, a network of co-authors is used to measure knowledge diffusion in co-authorship networks.

The parameter used by different researchers to gauge the knowledge diffusion as: producers and consumers (Zhang Q. a., 2013), research activity level (Li, 2010) , country’s diffusion and diffusion rate (Katy and Penumarthy, 2006).

2.1.1 Citations

Citations are considered as an indicator to measure knowledge diffusion in a scientific community (Saeed, 2010). Citations are being given by the researchers who wish to link their work with the previous work done in the past. Fig2.1 represents the example of citation.

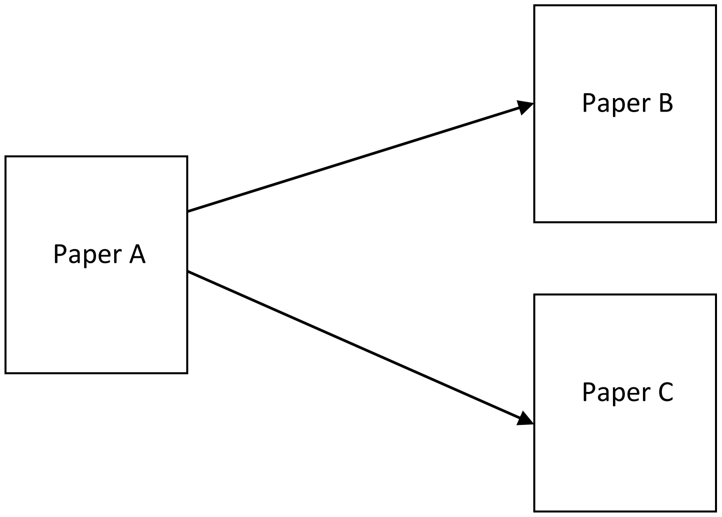


Figure 2.1: Example of paper citations

In this example paper A cites paper B and C, which indicate that the author of paper A has linked their paper with the work done by author B and C. While measuring diffusion of knowledge using citations it is acknowledged that the knowledge of paper B and C is diffused in paper A.

From many years, citations in the scientific publications are considered as a major pattern of how to reach the most relevant scientific research. Nonetheless the main focus of citation analysis is to quantify the impact of scientific research, during past decades. However, with time the process of measuring diffusion of knowledge by citation has geared up and ensured productive access to researchers of online journals and publications (Markl, 2009). Presently, there are various approaches that are being used for measuring knowledge diffusion. Some of them are given below:

Many researchers use citations for measuring the knowledge diffusion. Mazloumian, et al. Proposed a network based citation index with the aim to measure, flow of scientific knowledge in order to recognize knowledge producer and knowledge consumer globally (Mazloumian, et al., 2013). Trends that were identified shown that North America and Europe are major knowledge producers and Asia and Africa are the knowledge consumers.

Rowlands et al has introduced a new technique to measure published scientific knowledge (Rowlands, 2002). As per this technique journals are critical elements for measuring diffusion of knowledge. Rowland also focused on the citations of an article by some other journal, contrary to original journal of that article.

Liu et al has introduced another approach regarding knowledge diffusion based on citation. In this approach two form of diffusion has been studied (Liu, 2010). One is diffusion by citation while the other is diffusion by publications. Diffusion by citations refers to the published knowledge of researcher that is being cited by some field, while diffusion by publications refers to the community of researchers which have published their work in various fields. Afterward this approach is being proved by a case study in a Chinese university.

Another approach analyzes the flow of knowledge in Computer Science Citation Network. They trace citation networks by analyzing the associations between various citations and its impact on the other articles. To identify producers and consumers of knowledge dataset collected from Cite-Seer and ACM. They concluded that citing recent papers within the same scholarly's community get a larger quantity of citations normally (Shi, 2009).

Zhang et al measure researcher's publications, the process of knowledge diffusion plays an important role for measurement of knowledge diffusion in academic institutes (Zhang Q. a., 2013). The approaches of knowledge diffusion, which are based on citations, helpful in measuring the impact of academic institutions. A new approach for analyzing the dual role of top research institutions is introduced by (Katy and Penumarthy, 2006) in United State. He studied the knowledge diffusion among those institutions. Also, 20 year publications, data set are used to recognize the 500 research institutions which are often cited by the author. The Result of research reveals about the knowledge producer and knowledge consumer's institutes.

H- Index is another approach, regarding knowledge diffusion based on citations. This approach is introduced by (Hirsch, 2005) for estimating the importance of authors. The main theme is that, this approach ranks the researcher by combining the number of citations and number of publications. An H-index of X reveals that the researchers have published minimum X paper and each of them is being cited by at least X papers. This index categorizes the authors on the basis of citations which are being received by their publications. An author makes a top position in researcher list of his publications highly cited. Afterward, the H-index is also being assimilated in so called "citation reports" that are available online at the ISI's web of science. For a given set, the impact factor of scholar many decrease any time however, H-index cannot do so. Hence it can be considered as an accurate highlighter for lifetime achievement in the case of individual researchers.

Rousseau et al combine Gini "Evenness measure" the effect of publication and citations in order to measure diffusion of knowledge (Rousseau, 1998). Diffusion from publications and diffusion from citations are the two forms of knowledge diffusion that are being studied in this approach, which are used to measure the knowledge diffusion.

On the base of knowledge diffusion many critical decisions can be taken in the world. It can be measured between the countries (Hassan, 2013). In such scenario, authors of one country cite the work or publication of authors from some other country. In another case, while measuring the diffusion of knowledge from more then two countries, the country act as a source of knowledge production while the other countries which cites knowledge is considered as knowledge consumers countries.

From an overall literary survey of system, that is presently existed, which is used to measure diffusion of knowledge, it can be stated that citation are very crucial and are being used by scientist from many years. They are one of the critical sources for measurement of knowledge diffusion. According to some researchers, there are some limitations associated with citation count, which influence the validating of citations (Seglen, 1998). Another limitation is their time dependency. As citation of scientific papers generally increase over three to four years of its publications. However, still there is a significant portion of publications that has not been cited. But it does not mean that they lost their significance (Marx, 2001). Because of these problems, all the knowledge diffusion is not represented by citations. Hence, we can conclude that, due to such limitations, researchers use some other parameters for the measurement of knowledge diffusion.

2.1.2 Patent Citations

A set of rights given by a government to a patentee is termed as patent (Organization, 2004). They prohibit others from making offer for sale and importing the scientific inventions.

Scholarly paper's citation matches with patent citation. The main difference between the two is that in the prior one, only the author can cite other's work, while in later one both patent examiners and patent applicant cite other's work. The relevance of citations is determined by patent examiners, which took discussion on the basis of data provided by the patent applicant (Leydesdorff, 2007) .

In order to understand the mechanism of the co-author network, let us consider a simple example as shown in Figure 2.2. Patent A cited the three patents B, C and D. it is supposition that the author of patent A, citing patent B, knows about all of the information in patents B and using to build patent A.

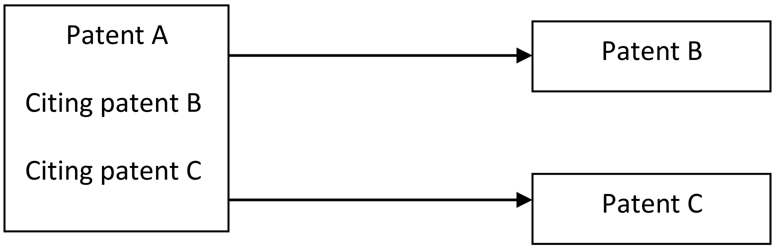


Figure 2.2 Example of Patent Citations

Many researchers assumed that the patent citations can help in measurement of knowledge diffusion. In order to trace the process of knowledge diffusion (Chen, 2004) has introduced an approach that combines network theory, network visualization and patent citation. Nonetheless, this approach still restricted to citation and patent citations for diffusion of knowledge.

A theoretical model, which is being introduced by Gans et al (Gans, 2013), represents that condition which support diffusion of knowledge by using patenting and publications. This model represents a relation between firm (patent) and researcher (publications) and keeps a bull's eye on the economic forces. These forces are helpful in the production of knowledge. Moreover, that model also thrashes light on previous techniques or method i.e. publication and patent of knowledge diffusion.

A comparison is made between Europe, Japan and US with the help of patent citations by (Bacchiocchi, 2006). In this comparison, a data of six centuries was used from Europe patent office to investigate the diversity across technological field and across countries. Final results introduced that the companies patent are less cited as compared to universities and public research. The nature of citations reveals that the knowledge of universities and public research circulated more rapidly as compared to the knowledge which is produced in companies.

Patent citation is also a useful measure of diffusion of knowledge across industries. A patent is the standard for the discovery of knowledge. In other words knowledge can be diffused more rapidly and widely. As more as patent is cited that increase the value and use of patent (Stolpe, 2002)

In another research, the patterns of patent citations between various countries are studied by (Jaffe, 1999). It measures the international diffusion of knowledge. A final result reveals that there is appreciable proof of geographical localization that becomes mild slowly as knowledge diffuse. Moreover, this study also found some phenomenal facts about countries. For instance, Japan localized but paying attention on latest developments.

Belenzon et al illuminates how geography, state policy and university influence diffusion of knowledge by using patent citations (Belenzon, 2010). The technique use to measure the knowledge diffusion in university patent and scientific publication. Result reveals that knowledge of university is localized. It also shows that the state policies university influences the knowledge torrent across borders. But measuring knowledge torrent through tags cannot be influence by geography, university and state policy.

Some researcher talked about the limitation of this approach is that, it cannot represent all the diffusion of knowledge. The researcher uses some other parameters like co-authorship network in which no need of paper’s citations. Therefore, it is concluded that this approach is also useful mean to measure the knowledge diffusion. However, there are certain limitations with this approach. Researchers use some other parameters for the measurement of knowledge diffusion.

2.1.3 Co-authorship Network

In scientific society, coauthor is referred to group of network of authors. These authors work together on different research ideas. This teamwork among researchers is called co-authorship (Newman, 2004). In another definition co-authors network is consider as a groups of authors who are working together. Additionally those authors who are working with large numbers of authors usually have large co-authors network. While, authors who are working with limited number of authors have small co-author network.

In order to understand the mechanism of co-authors network let us considered the simple example. Suppose author A having paper with B and C then author B and C are consider as a co-author of author A. Also, B and C are in network of author A. Fig 2.3 represent the illustration of example.

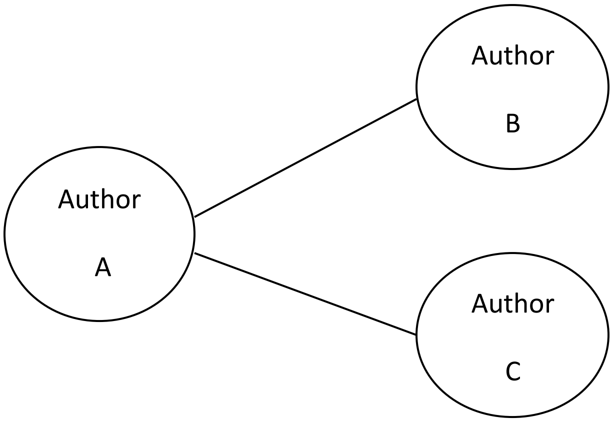


Figure 2.3: Co-author Network of Author A

From Fig 2.3 one can say that the co-author network of author A is 2 while co-author network for both B and C is 1.

Co-authorship is well known illustration of researcher’s relationship which is being existed in scientific research. It is mutual share of two or more researchers towards a new research. With the help of this greater numbers of scientific achievements can be made as compared to an individual researcher (Hudson, 1996).

Co-author network is an important approach to measure knowledge diffusion. A number of researchers use this approach to measure knowledge diffusion. Following are some previous approaches of co-author network for measuring knowledge diffusion.

Katz et al made an argument regarding production of new scientific knowledge (Katz, 1997). According to them co-author ship helps researchers to work collectively to attain goal of producing new scientific knowledge. This goal can be achieved with the help of diffusion of previous knowledge. Hence co-author network is one of the prominent approaches to measuring knowledge diffusion.

Many reasons for collaborations of scientists for production of new knowledge is spotted by (Sonnenwald, 2007). For instance, co-authorship helps researchers to produce quality new knowledge, to promote collaboration among countries, to support and growth of new scientific innovations and to gear up the transfer and diffusion of both implicit and published knowledge.

In regards to support researcher's collaboration, the main and important features are knowledge creation and diffusion. Also, scientific collaboration speed up the knowledge diffusion and creation process. Many researchers highlighted the positive after effect of scientific collaboration on diffusion and get spread around of knowledge one of the true way to analyzed researchers collaboration is co-author network (Lee, 2005).

The study of (Wagner, 2006) reveals that in growing countries big research network play significant role in understanding of educational, communal and environmental extent of growth and knowledge diffusion. Wusteman et al found in another study that the collaboration of researchers plays an important role in strengthening nationwide unity (Wusteman, 2009). It also plays role for the progress of research infrastructures by diffusion of knowledge through co-author network.

The calculation of almost complete or complete co-authorship becomes practically possible due to the wild spread of online bibliographies (Newman, 2004). They used date from three bibliographic databases and developed an interconnected group of scientists to study the patterns of knowledge diffusion.

As discussed above that the co-author network analysis is an important approach for measuring diffusion of knowledge. However, many researchers highlight the negative effect of co-authorship network. Such as, Australian academics, (Butler, 2003) study of publication, performance find that such policies enhancing internationally-oriented research collaboration. But such internationally-oriented research ought not to be conducted at the expense of nationally-oriented ones. Co-authorship network is an important mean for measurement of knowledge diffusion, but certain limitation also associated with it.

2.2 Visualizations

Some insights are given to us by visualization. These insights might be already known to us, while some of them might be completely new. We can identify a trend instantaneously with the help of visualization. A good visualization provides research data, to play with them in order to investigate some important cause-effect relationship. Visualization is very crucial for investigation and research work, as in journalism. Some visualizations are discussed below:

One of the approach that visualizes the knowledge diffusion based on a network base citation index proposed by (Mazloumian, et al., 2013). They have visualized producers and consumers as discussed in table 1. Their approach has effectively identify the producers and consumers and the dataset that analyzed by them, contain 80 million citations in 13 million paperwork published schedule of 2000 to 2009 retrieved from Thomson Reuters' Web of Science. They represented the Knowledge producers by Red dots and the knowledge consumers by Green dots in fig 1. Green pubs show that the amount of citations received is higher then expected and the red bars show that the amount of citations is smaller then expected. Trends that were identified show that North America and Europe are major knowledge producers and Asia and Africa were the knowledge consumers. However, this visualization does not represent the three important features like research activity level, country's diffusion and diffusion rate.

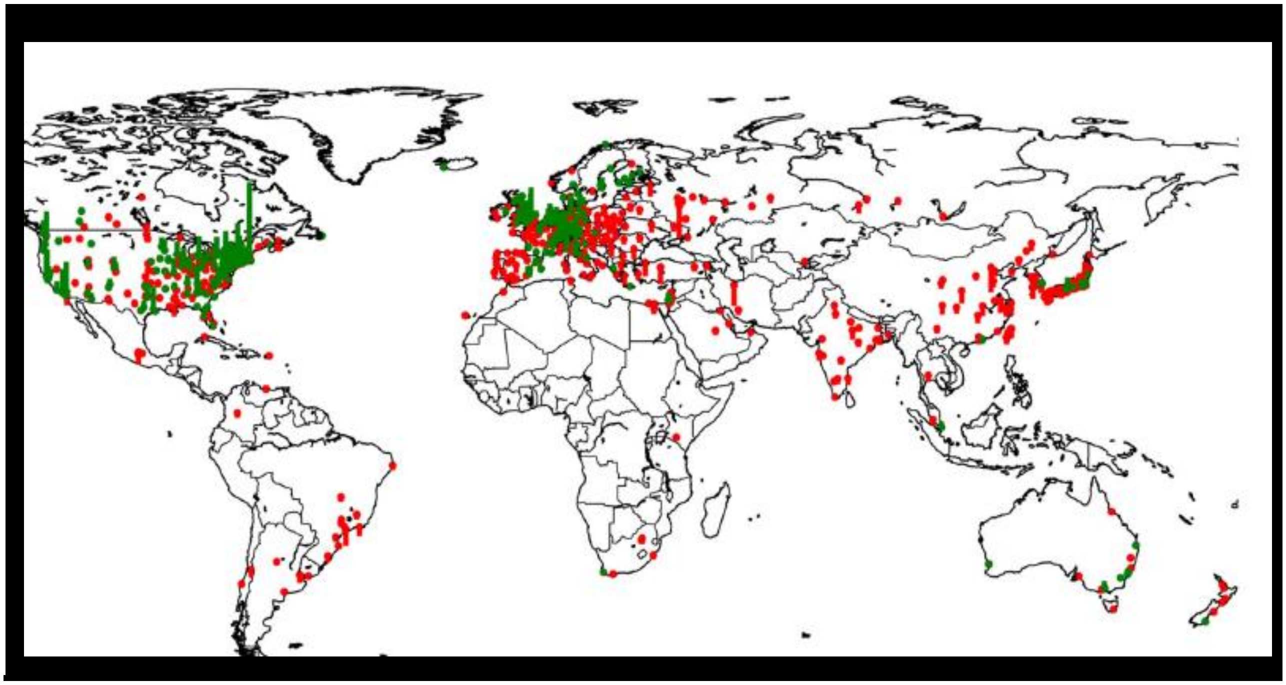


Figure 2.4: World map of the best knowledge producers and consumers.

Another approach present by Zhang et al (Zhang, 2013) use approach to visualize the scientific production and consumption in physics. They showed top locations for scholarly research in Physics represented in the fig 2. They analyzed the complete publication data source of North American Physical Society creating longitudinal (50 years) citation.

Algorithms that use by them were diffusion proxy and scientific production rank algorithms, to fully capture the spatial-temporal dynamics of Physics's knowledge in the world. Knowledge diffusion proxy recognizes the main places in the creation and utilization of Physics's knowledge. This visualization effectively identifies the producers and consumers as discussed in table 1. However, some other features like research activity level, country's diffusion and diffusion rate also absent.

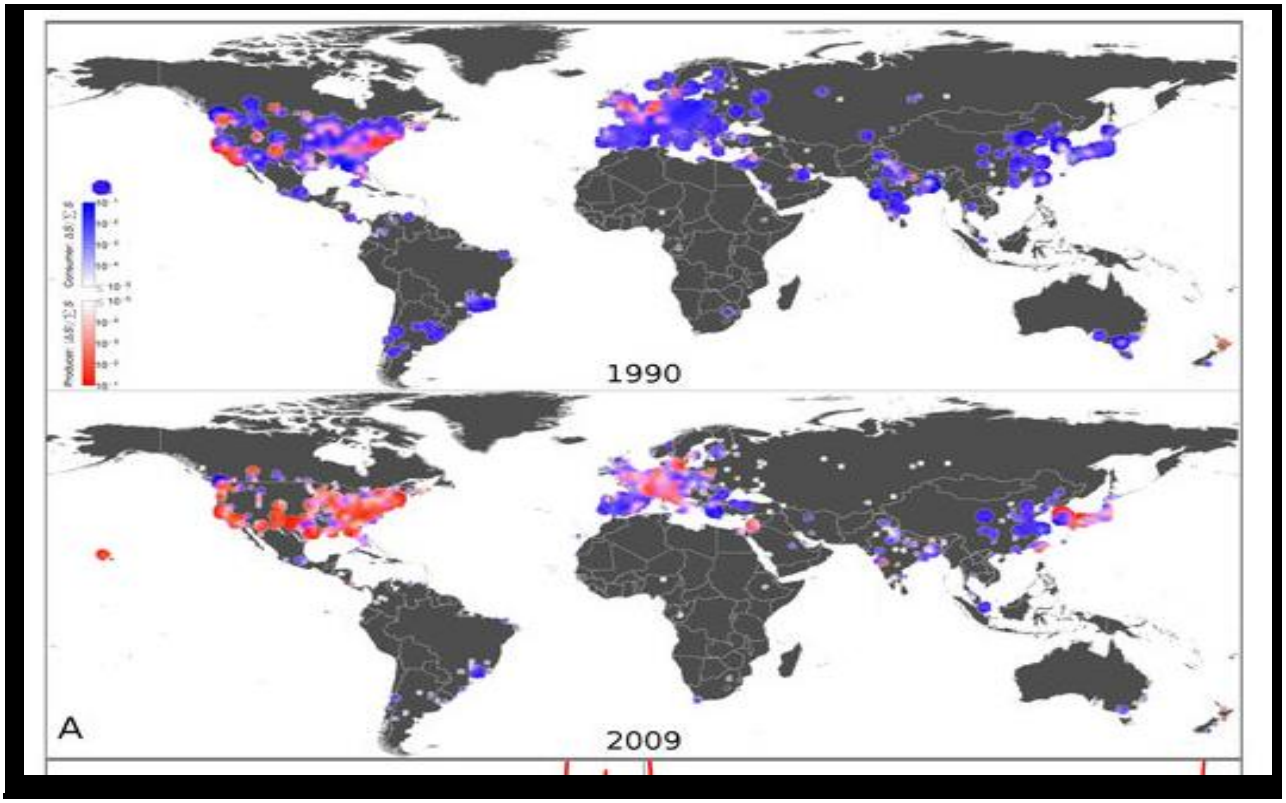


Figure 2.5: The globe map of knowledge source and sink.

Another approach that use geo-visualized technology to show visual knowledge diffusion of data mining papers published during four stages(1991-1995,1996-2000, 2001-2005, 2006-2008) in SCI (Li, 2010). Fifty key words of the highest frequency of data mining were analyzed to identify knowledge diffusion. Residential cities of the authors were collected by collecting contact addresses of authors in the original data and 788 cities were involved in the study. Factors and processes of knowledge diffusion were visualized using thematic maps in fig 3. Based on the results geo visualization of knowledge diffusion of data mining was given. This visualization does not contain information about the producer and consumers. However, diffusion rate and research activity level of different countries visualized as discuss in table 1.

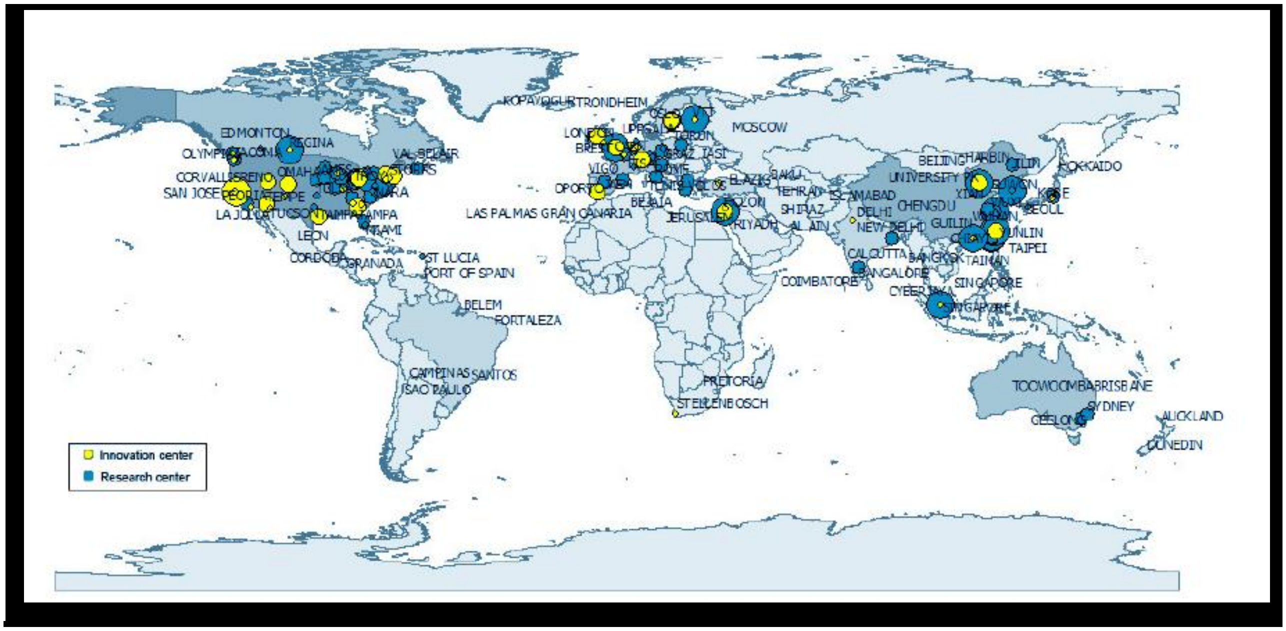


Figure 2.6: Analysis of research centers and innovation centers.

Another approach proposed by Chen et al (Chen, 2004) that combine network theory, network visualization and patent citation trace the process of knowledge diffusion. This approach improves the diffusion and transfer of knowledge. However, this approach is also limited to citation and patent for tracing the knowledge diffusion.

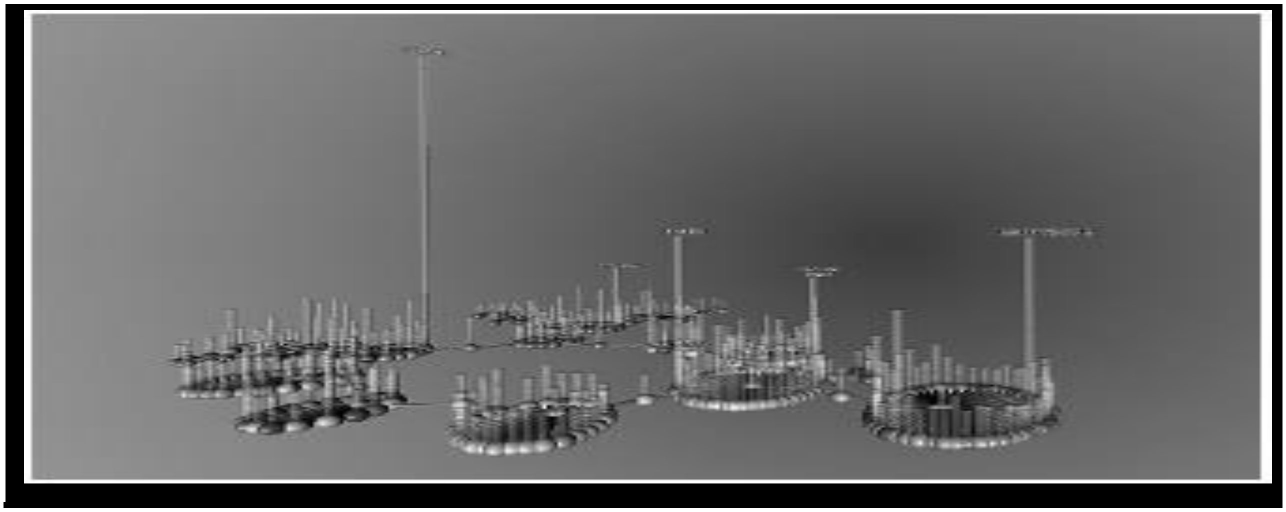


Figure 2.7: Visualization of a published paper.

Sorenson et al examined the importance of publication by comparing the rules of citations from future patents to three groups, those that mention scientific publications, those that mention commercial publications and those that mention neither (Sorenson, 2004). The forward citation that patients receive from future patents plays an important function in knowledge dispersion. Analyses strongly advocate that publication plays an important role for raising technological innovation rate. However, during the analysis, there is no information got about knowledge diffusion, consumer and producers.

Shi et al analyzed the flow of knowledge in Computer Science Citation Network (Shi, 2009). They traced citation networks by analyzing the associations between various citations and its impact on the other articles. To identify producers and consumers of knowledge dataset collected from Cite-Seer and ACM. Information paths were identified in the domain of computer science, based on citation networks. Analysis was performed to identify the impact of different range of citation's options on the impact of article. They figured citing recent papers within the same scholarly community get a considerably larger amount of citations normally.

Azoulay et al measured the Diffusion of Scientific Knowledge across Time and Space (Azoulay, 2011). The dataset contains 9,483 top academic scientists. Analysis was performed, based on the citation's paths linked to individual articles, published prior to the scientist shifted to another institution. They figured article-to-article citations from the scientific community at the superstar's source location are less damaged by their departure. They further concluded the fact that at superstar's destination article-to-article citations increases after them, knowledge moves to industry might need more face-to-face contact than those to academics (Tsai, 2005) show worldwide geographic trends in publication of science education in three journals, the articles published by International Journal of Science Education (IJSE), Journal of Research in Science Teaching(JRST), and Science Education(SE) during a five year period, from 1998 to 2002. Analysis was performed on 802 research papers on the basis of authors' nationality, research issues and types. They conclude that researchers in a few major English speaking countries like US, UK, Australia, and Canada conduced to a significant area of the publication.

Hubbard et al show the geographical connection between the researchers in visualization of bibliographic data existing in the United States (Hubbard, 2012). Analysis was performed on bibliographic data which was taken from the Thomson Reuters Web of Science. They examined an insight with the help of visualization and geographical analysis to citation maps. Five aspects of the paper, i.e. citing publications, cited publications; co-author network and distance cited-citing publication network and hypothesis testing of average co-author's distances over time were used. Both cited and citing publications, involving visualization of bibliographic data was given on the basis of first author clustering.

Hu et al investigated the amount of knowledge circulation in East Asia and outside (Hu, 2008). East Asia had become a source of international knowledge diffusion and whether such diffusion is limited to the region. They discovered that intensive knowledge flow between East Asia leading innovators. Korea and Taiwan cite each other at least as they cite the US and Japan frequently. Such knowledge circulation has increased because the mid-1990s. They examined that Apart from Thailand, all the East Asian economies, Hong Kong, Singapore, China, and Malaysia, cite Korea and Taiwan at least as much as they cite the united states and Japan. The "G5" group, which include Britain, Canada, France, Italy and Germany, has been minimal, often cited way to obtain knowledge for East Asia

2.3 Comparison of different techniques

Table 2.1: Four different parameters that visualize in the different visualizations

1	**Research papers**	**Technique**	**Producers**	**Consumers**	**Diffusion rate**	**Research activity level**
2	Global Multilevel analysis of scientific food web (Mazloumian, et al., 2013)	Citations	Yes	Yes	No	No
3	Characterizing scientific production and consumption in physics. (Zhang Q. a., 2013)	Citations	Yes	Yes	No	No
4	Geovisulization of knowledge diffusion case study data mine (Li, 2010)	Publication	No	No	No	Yes
5	Tracing Knowledge diffusion (Chen, 2004)	Patent citation+ citations	No	No	Yes	No
6	Mapping the Diffusion of Information Among Major U.S. Research Institutions (Katy and Penumarthy, 2006)	Citations+ publication	Yes	Yes	Yes	No

In the Table 2.1, we have evaluated five previous visualizations. From the table 2.1 we can see that none of the previous approaches have visualized all of the knowledge diffusion parameters. There were total of four knowledge diffusion parameters found from literature. One can see that none of the approach has even visualized four parameters together. There were only two approaches out of five which have visualized three parameters. The majority of approaches have just visualized one or two parameters.

CHAPTER 3

PROPOSED METHODOLOGY

This chapter presents a comprehensive methodology to develop and evaluate the visualization approach that visualizes all of the features such as: producers, consumers, research activity level, diffusion rate and country's diffusion collectively.

The broad level architecture diagram for the proposed methodology has been explained in Figure 3.1.

3.1 Comprehensive dataset selection

For evaluating knowledge producers, consumers, diffusion rate, research activity level and country's diffusion. We need comprehensive dataset and such comprehensive dataset should contain publications, author profiles, paper citations, institutional affiliation, etc. Such data was available on Microsoft Academic Search (MAS). Within the MAS data collection, there are a number of different types of literature journal articles, and thesis. There are several categories in MAS, however, we have selected diversified domains of *Geosciences* as a comprehensive dataset. MAS classify *Geosciences* publications into 11 different categories like Atmospheric science, Geochemistry, Geodesy and Remote sensing, Geology, Geomorphology, Geophysics, Geotechnical engineering, Hydrology, Meteorology, Oceanography and Mineralogy. The details of the dataset are discussed in the chapter 4.

3.2 Determination of geographic location

The proposed visualization will map papers in their geographical locations, therefore, we need to find out the geographic locations of the published paper by identifying the geographical location of its first author. Each paper is geolocated using first author's institutional affiliation. In MAS only institutional affiliation is available against each author. However, this institutional affiliation cannot be visualized on geographical location unless that has been geolocated with respect to its country. The country of each author belonging to particular institute was extracted from Wikipedia infobox.

When we search institute's location on Wikipedia, the country of each institute can be acquired from location tag in Wikipedia infobox.

3.3 Proposed Context Diagram /System Architecture

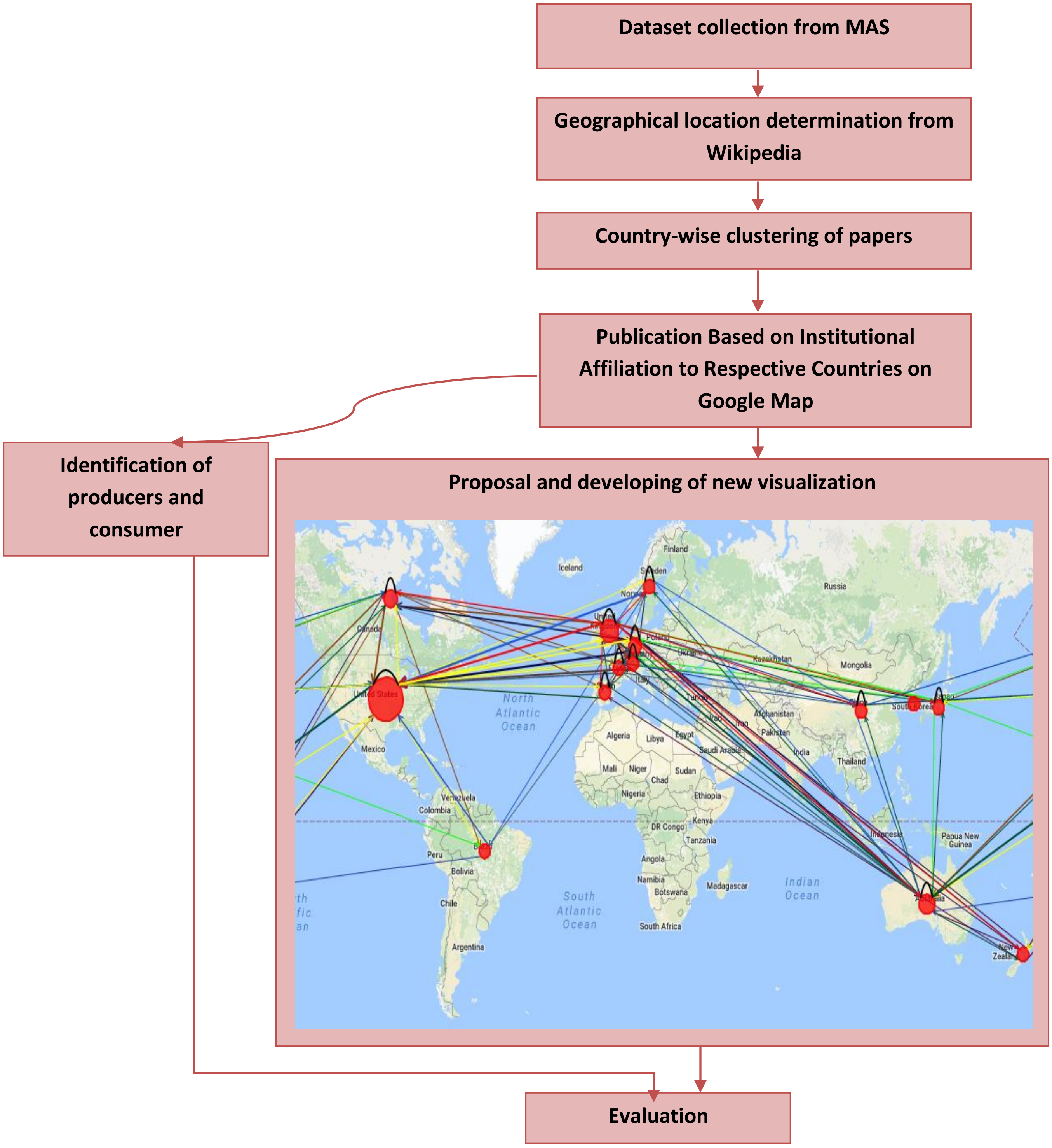


Figure 3.1: Architecture diagram of methodology

3.4 Country wise clustering of papers

After geolocating countries for each author, the papers belonging to the same country were grouped into the single cluster. We have different clusters of countries like UK, US, Australia, China, Japan etc.

3.5 Identification of producers and consumer

To measure the knowledge flow, different experiments were performed on the data set. Nodes that receive citations export their knowledge to the citing nodes. Form measuring the knowledge diffusion, we describe knowledge producing countries are those that export knowledge more than they import, and consuming countries are those that import knowledge more than they export (Zhang, 2013). We can measure it using the following equations as described by Zhang et al.:

$$\Sigma \mathrm{j}\ w_{ij} \text{ .... } (3.1)$$

” j” is the importer and ‘i” is the exporter in equation 3.1. Where “w_{ij}” present the amount of knowledge flow from papers written in “j” country to those papers which were written in “i”country.

$$\Sigma \mathrm{j}\ w_{ji} \text{... } (3.2)$$

“i” is the importer and “j” is the exporter in equation 3.2. Where “w_{ji}” present the amount of knowledge flow from papers written in ‘i” country to those papers which were written in “j”country.

$$\mathrm{S} = \Sigma \mathrm{ij}\ w_{ij} \text{... } (3.3)$$

The total amount of citations either from “i” country to “j” country or from “j” country to “i” country is represented in equation 3.3.

$$\mathrm{Si} = (\Sigma \mathrm{j}\ w_{ji} - \Sigma \mathrm{j}\ w_{ij})/\mathrm{S} \qquad (3.4)$$

To identify the "producer" or "consumer", the equation 3.4 computes the overall scores for country "i", If the value of S_i is positive then the country is producer, else the country is consumer.

3.6 Proposed visualization

In proposed visualization, we visualize which country is a knowledge producer or knowledge consumer? How much research activity is done in the country? Which country export knowledge and which country import knowledge? All these parameters visualized in one visualization. However, none of the previous approaches visualize these parameters collectively.

For visualizing these parameters, we need to represent the countries. For representing countries, we used nodes. For representing diffusion rate among countries, we used directed edges of different size. The size of directed edges is used to represent the diffusion rate. For example if a paper was published at institution A and which is cited by a paper that is published at institution B, then there will be an arrow going from B to A. The greater papers produced at A are cited by B, the higher the volume of diffusion rate.

These nodes and directed edges are displayed on Google map. The size of nodes, edges and the direction of edges represent different parameters as discussed below

3.6.1 Research activity level

Research activity level of different countries has been calculated on the basis of publication. The countries that have more publication have a higher research activity level and the countries that have low publications have a lesser research activity level (Li, 2010). The research activity level is represented by the size of the circle as represented in fig 3.2. The country with larger circle's size has higher research activity level. While, the country with smaller circle's size has lower research activity level.

Now we are going to show the research activity level of different countries.

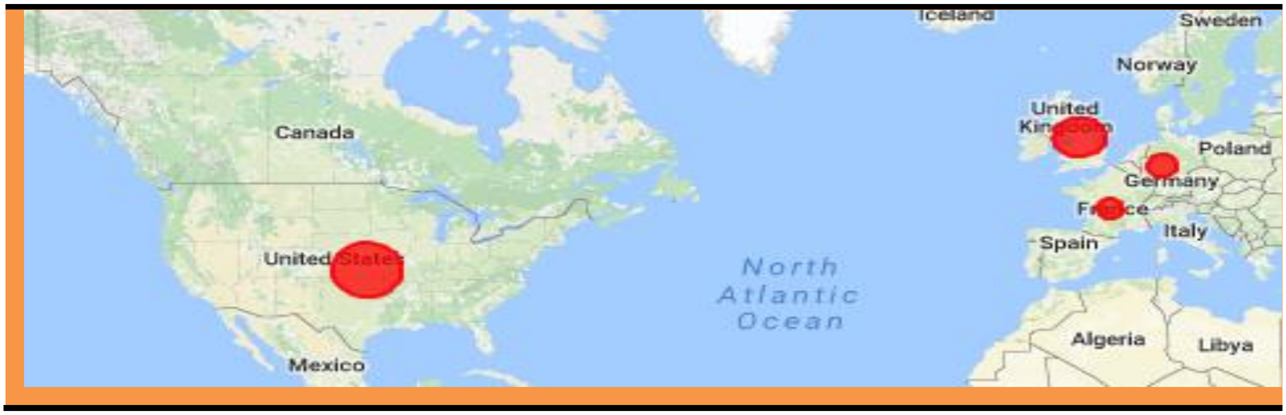


Figure 3.2: Research activity level of different countries.

Fig 3.2 shows the research activity level of four countries. The US has larger circle’s size, so its research activity level is high as compared to the other three countries.

We have faced a problem of adjusting the circle’s size while visualizing the research activity level. The problem is that the highest value of research activity level is six hundred and four, when the value of six hundred and four taken as a radius of the circle and when it displayed on the screen, then we can only view a portion of the circle even on a large LCD. Therefore, we use scaling function to avoid this problem.

Suppose 'Min' and 'Max' are the minimum and maximum values acquired from data. We want to scale 'Min' to 'a' and 'Max' to 'b'. Then for any number x, the function is as follows[1].

$$f(x) = (((b - a)\ (x - \min))/(\max - \min)) + a\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (3.5)$$

In our case the maximum value of research activity level is six hundred and four, while the minimum value is 6. We have scaled the values of research activity between 5 to 15.

Another formula is used for finding the meter per pixel because the circles drawn on maps have a property radius. Unit of radius is meters. Meters per pixel are not same at different latitudes due to the Maps projection system. We have computed the Meters per pixel by the following formula[2].

[1] http://stackoverflow.com/how-to-scale-down-a-range-of-numbers-with-a-known-min-and-max-value.
[2] http://wiki.openstreetmap.org/wiki/Zoom_level

$S = C * \cos(y)/2(z + 8)$.. (3.6)

C is the circumference of the earth, Z is the zoom level, Y is the latitude value of country for which the circle is being drawn represented in 3.6.

Now we can calculate the radius (in meters) by using the following formula

$\text{RadiusInMeters (latitude, scaledvalue)} = \text{scaledValue} * \text{MetersPerPixel (latitude)}$.... (3.7)

Latitude is the latitude value of the country for which the circle is being drawn.

Scaled value is the value computed by formula discussed in 3.5.

Meters per pixel is calculated in 3.6.

3.6.2 Country’s diffusion

The number of distinct edges coming to a country represents the country’s diffusion. For example, if seven distinct edges are coming to USA, this means that the knowledge of USA is diffusing into seven different countries.

Now we are going to show the country’s diffusion of US, Spain and Germany.

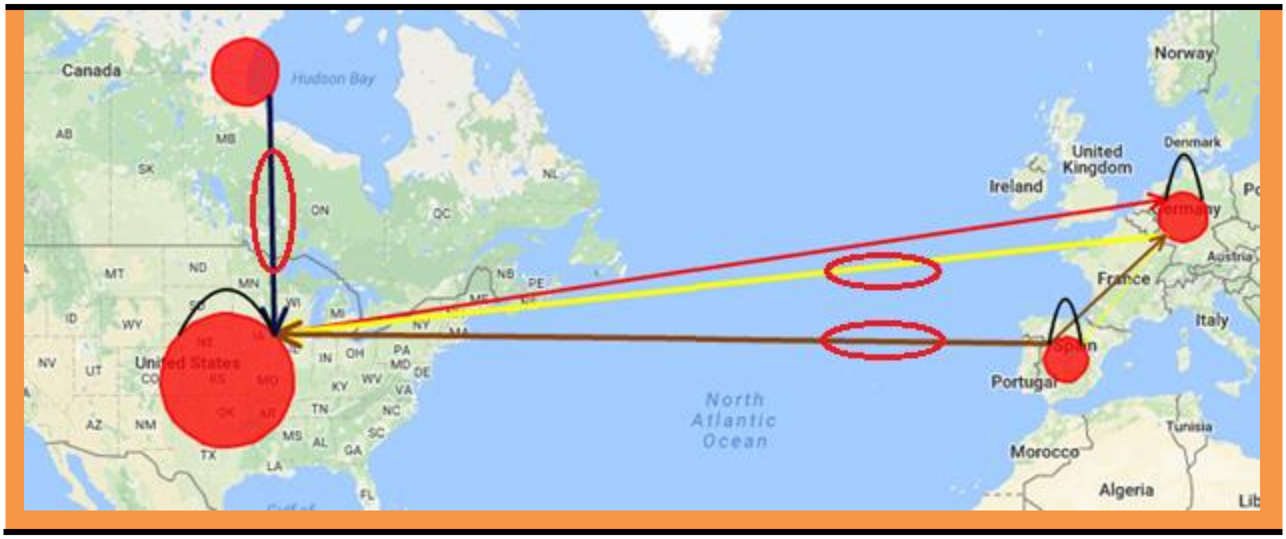


Figure 3.3: Country’s diffusion of US, Spain and Germany.

In fig 3.3 four distinct edges are coming to US (red oval shapes), while the distinct edges towards Canada and Spain is smaller as compared to US. Hence, we can identify from the figure 3.3 that the country’s diffusion of US is higher as compared to Germany and Spain.

3.6.3 Diffusion rate

The number of citations coming towards a country represents the diffusion rate. Countries with greater width of edges have a higher diffusion rate. While, countries with the smaller width of edges have a lesser diffusion rate.

Now we are going to show the diffusion rate of the US, Germany and Spain.

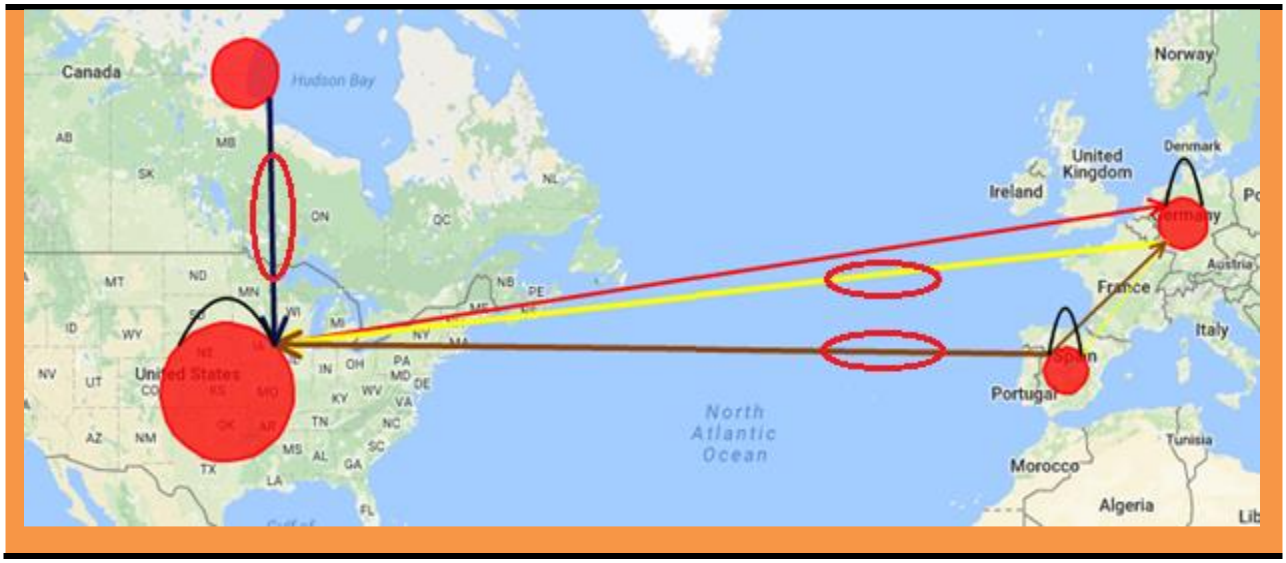


Figure 3.4: Diffusion rate of US, Spain and Germany.

The number of citations coming towards US is higher as represented by the greater width of incoming edges (red oval shape). Therefore, US have higher diffusion rate as compared to other three countries. While, the number of citations coming towards Germany and Spain have lower as compared to the US. Scaling function that is defined in equation 3.5 is again used to scale the values for the width of edges. In this case the maximum value of diffusion rate is thirteen thousand and four hundred, while the minimum value is 11. We have scaled the values of diffusion rate diffusion between 1 to 6.

3.6.4 Producer

Knowledge producing countries are those countries that export knowledge more than they import. These countries identified with the help of incoming and outgoing edges. The greatest width of incoming edges and smaller width of outgoing edges represents that the country is producer.

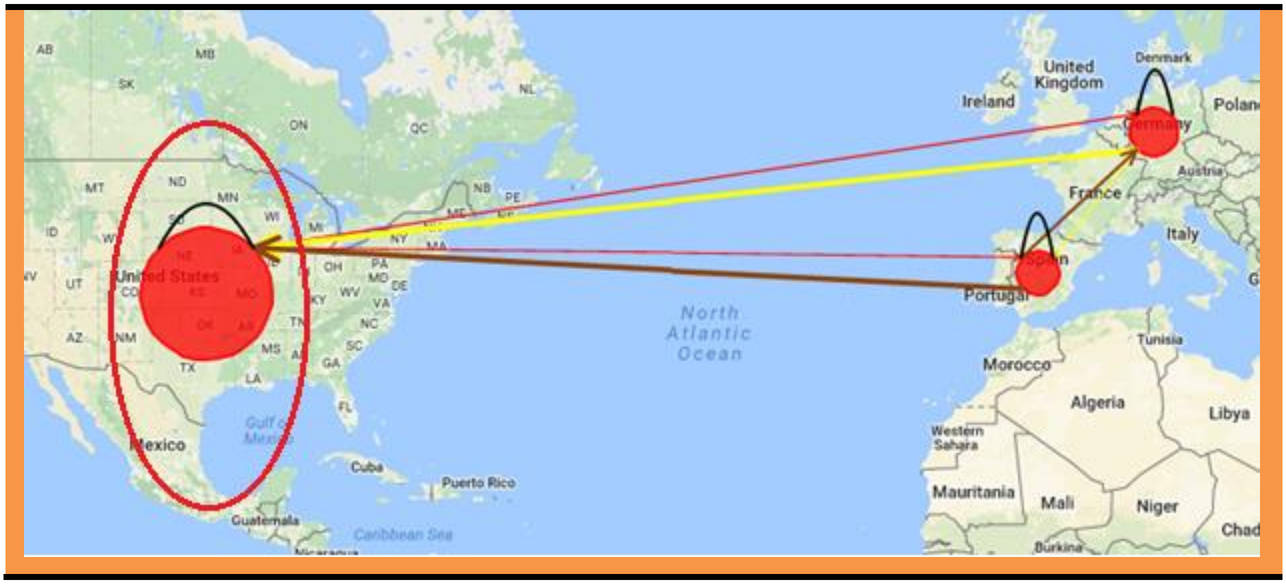


Figure 3.5: Knowledge producer.

In fig 3.5 US have incoming edge's width greater and outgoing edge's width smaller as compared to Germany and Spain, So US is a knowledge producer (red oval shape).

3.6.5 Consumer

Knowledge consuming countries are those that import knowledge more than they export. These countries identified with the help of incoming and outgoing edges. The smaller width of incoming edges and greater width of outgoing edges represents that the country is consumer.

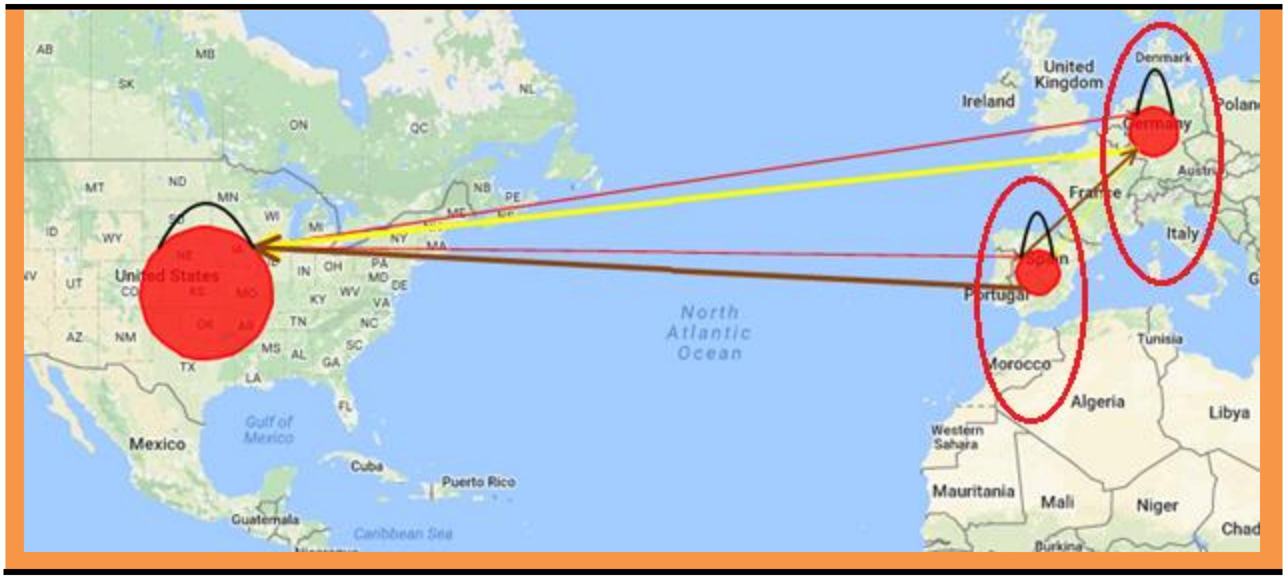


Figure 3.6: Knowledge consuming countries.

In fig 3.6 Germany and Spain both have incoming edge's width smaller and outgoing edge's width larger. Thus, it represents that the Germany and Spain are consumers (red oval shapes).

3.7 Evaluation

The evaluation of visualization has been done in the literature. User studies are the most commonly used evaluation methods in Information Visualization. Many authors have chosen users study for evaluating their visualization (Hearst A., 2016) (Carpendale S., 2008).

To evaluate the proposed visualization against the benchmark, we have also used the user study. Total of 40 users have been used in this focused study having diversified experience, expertise and background. After taking the feedback we find correlation between benchmark dataset and the user's feedback on the developed visualization. The detail of evaluation has been provided in the chapter 4.

After taking the feedback we find a correlation between actual results that we have stored and the user's feedback on new developed visualization.

Following is the correlation formula used for evaluation.

Spearman's Rank correlation coefficient is used to identify and test the strength of a relationship between variables. Excel can be used to calculate and graph Spearman's Rank correlation to discover if a relationship exists between the variables, and how strong this relationship is. The formula used to calculate Spearman's Rank[3] is shown below.

$R = (1 - 6\,\Sigma d2)/n3 - n$.. (3.9)

The value n is the number of sites at which we took measurements. The answer will always be between 1.0 (a perfect positive correlation) and -1.0 (a perfect negative correlation).

[3] http://www.statisticssolutions.com/correlation-pearson-kendall-spearman/

CHAPTER 4

EXPERIMENTS AND RESULTS

In the previous chapter, a comprehensive methodology was discussed. This chapter presents the results obtained from each step of the methodology which is being discussed in the previous chapter.

4.1 Dataset Collection

For evaluating knowledge producers, consumers, diffusion rate, research activity level and country's diffusion we need a comprehensive dataset. This dataset was obtained from MAS by using crawler. We have selected diversified domains of *Geoscience* to perform critical analysis. MAS classify *Geoscience* publications into 11 different categories like Atmospheric science, Geochemistry, Hydrology Geodesy and Remote sensing, Geology, Geomorphology, Geophysics, Geotechnical engineering, Meteorology, Oceanography and Mineralogy. In order to identify the trends of knowledge diffusion, top 100 papers were selected from all categories which are sorted citations wise and 100 citations of each paper were also extracted. Extracted dataset contains information about the publication title, first author, citations received to publication and the first author's institutional affiliation. While crawling dataset, we found that there are many duplicate records in the dataset. Unique records are filtered by using MS Excel 'Remove Duplicate' command. Many duplicates records were removed from the collected dataset. Moreover, records that do not contain the author's name were also removed from the dataset.

The dataset consist of domain, categories and papers. The statistics of dataset are shown in the Table 4.1.

Table 4.1 Total No. of categories and their statistics

Dataset Collection					
Domain	**Categories**	**Root papers**	**Total Citations**	**Duplicate citations**	**Final citation**
Geoscience	11	1100	1,10000	8225	101775

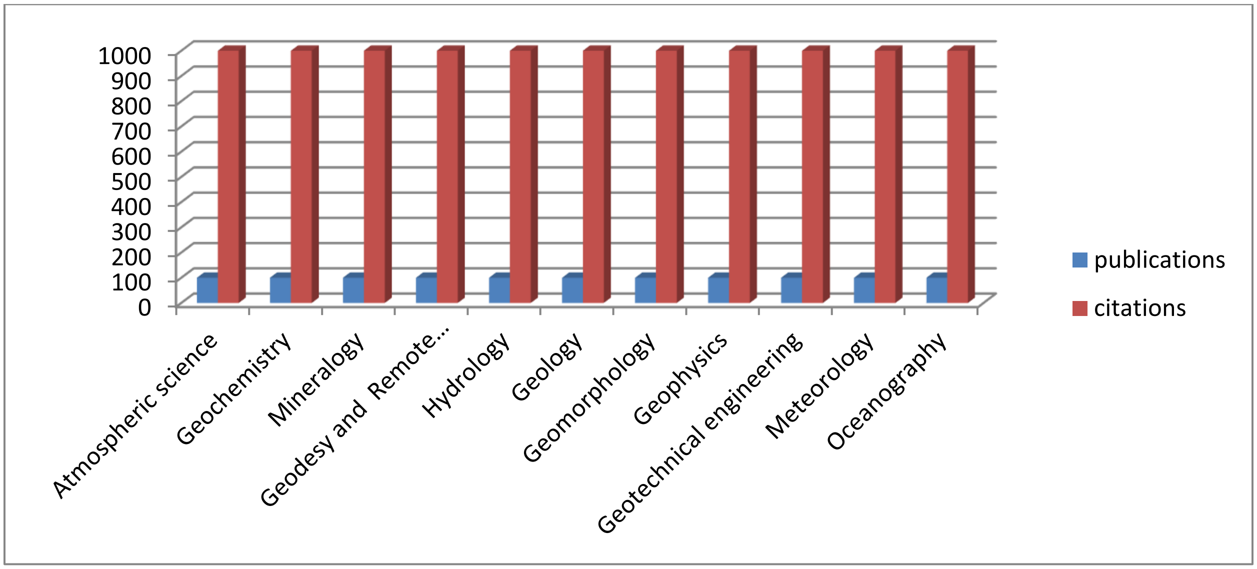


Figure 4.1: Statistics of Geoscience categories.

This dataset contains all of the categories, hundred publications of each category and thousand citations of each category of *Geoscience*. Therefore, this is a comprehensive dataset for the experiments.

4.2 Determination of geographic location

The proposed visualization will map papers in their geographical locations, therefore, we need to find out the geographic locations of the published paper by identifying the geographical location of its first author. When we search institute’s location on Wikipedia, the country of each institute can be acquired from location tag in Wikipedia infobox.

There are some cases when country is not retrieved from Wikipedia, location of the author is extracted from manual inspection of author's website. Some institutions have multiple branches in different countries in such scenario head quarter’ location is taken into consideration for further experiments. There were total 330 unique institutions found in the collected dataset.

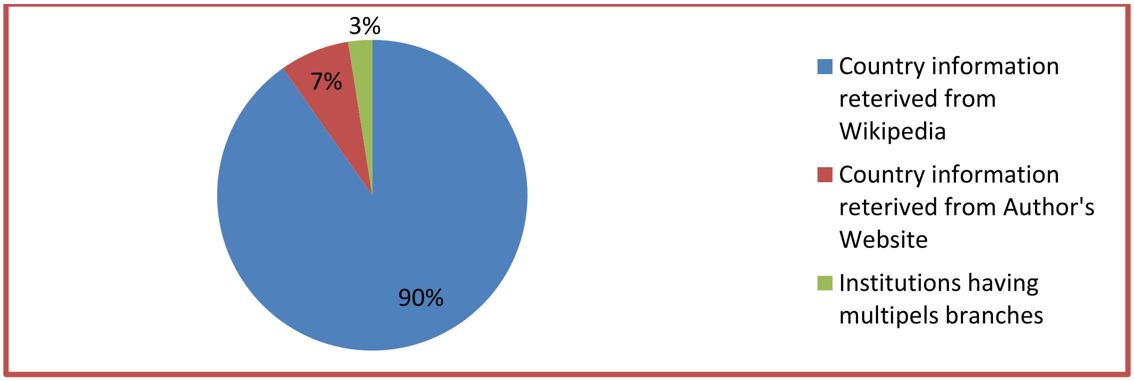


Figure 4.2: Extracted countries from institutional affiliation

4.3 Country wise clustering of papers

In this step we have collected the geographical location of each author in the dataset and the papers were classified on the basis of author's country. The authors who are associated with the same country were categorized into the single group. We have found clusters of 24 different countries like US, UK, Canada and Germany etc.

4.4 Identification of producers and consumer

To measure the knowledge flow, knowledge producing countries are those that export knowledge more than they import, and consuming countries are those that import knowledge more than they export (Zhang, 2013). For instance, if a paper written in node *i* cites one paper written in node *j*, there is a link from *i* to *j*, i.e. *j* receives a citation from *i* and *i* sends a citation to *j*.

After applying the formula on the dataset that was discussed in chapter 3, we found the knowledge producing and knowledge consuming countries. In chapter three, we have discussed that if the value of i is positive, then the country is a producer else the country is consumer. In our dataset, we have 24 different countries. After applying the formula we have found only two countries have positive values that are producers and the remaining 22 countries have negative values that are consumers.

Results of different countries are represented with line chart given below:

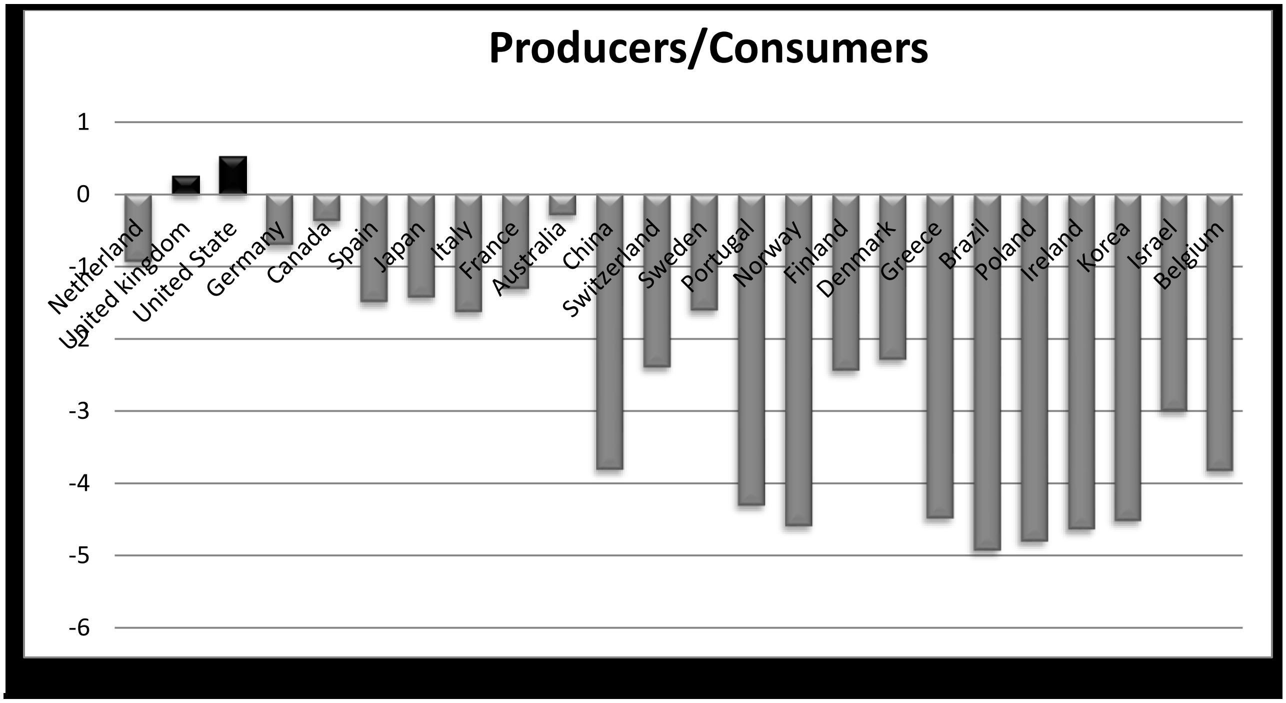


Figure 4.3: Knowledge producing and consuming countries

In the Figure 4.3 only two countries United Kingdom and United States have positive values and the bars represents these two countries are above zero, so these are knowledge producing countries. However, another 22 countries have negative values and the bars represent these 22 countries are below zero, so these are knowledge consuming countries. After the analysis of the Figure 4.3. We can easily identify that the United State is the greatest knowledge producer in the field of *Geoscience*, while Brazil is greatest knowledge consumer in the field of *Geoscience*.

As United States and United Kindom have emerged as knowledge producing countries. We have performed another experiment in which we have considered them as outlier and have recomputed the dataset of remaining countries to identify knowledge producing and knowledge consuming countries. After removing the producers (United State and the United Kingdom) above graph converted into the graph shown in Figure 4.4.

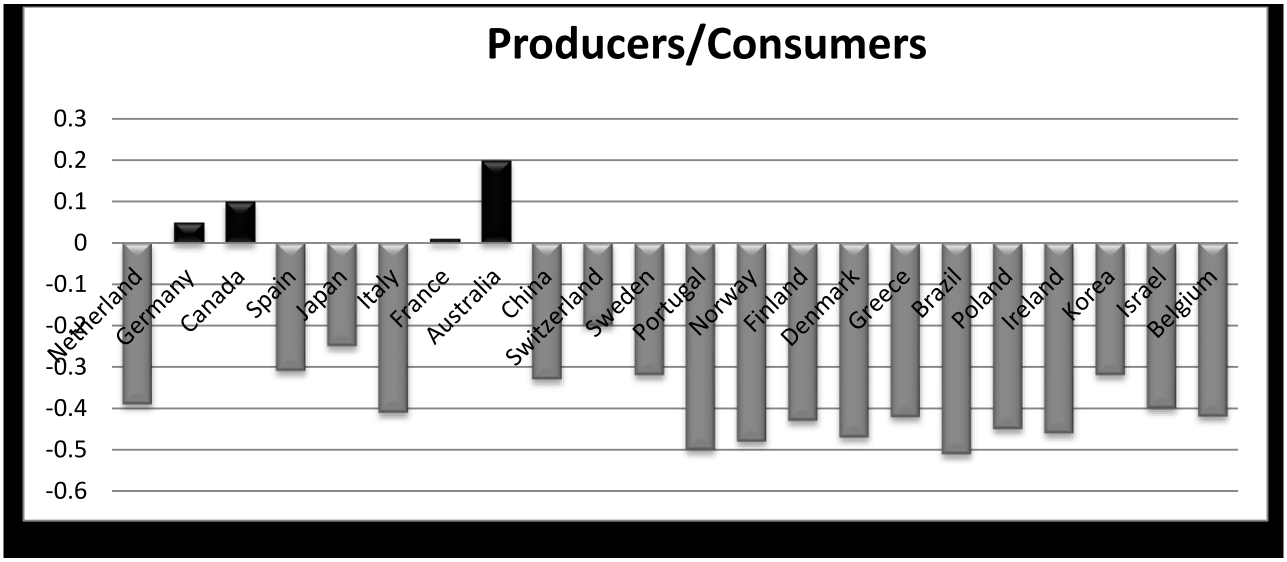


Figure 4.4: Knowledge producing and consuming countries.

From the analysis of figure 4.4 we have identified that four countries (Australia, Canada, Germany and France) are producers while remaining all the other countries are consumers. We can also say that these four countries import higher amount of knowledge from US and UK. When we remove US and UK, the import of these four countries become less then its export and these countries turned as producing countries.

4.5 Country's research activity level

Research activity level of different countries has been calculated on the basis of publication. The countries that have more publication have a higher research activity level and the countries that have low publications have a lesser research activity level (Li, 2010). Our dataset consist of 1100 publications of different countries. The countries that have higher research activity level are the main research centers in the world. For calculating the research activity level of each country we have counted the publications of that country from 1100 publications.

The research activity level of different countries is represented by the following graph.

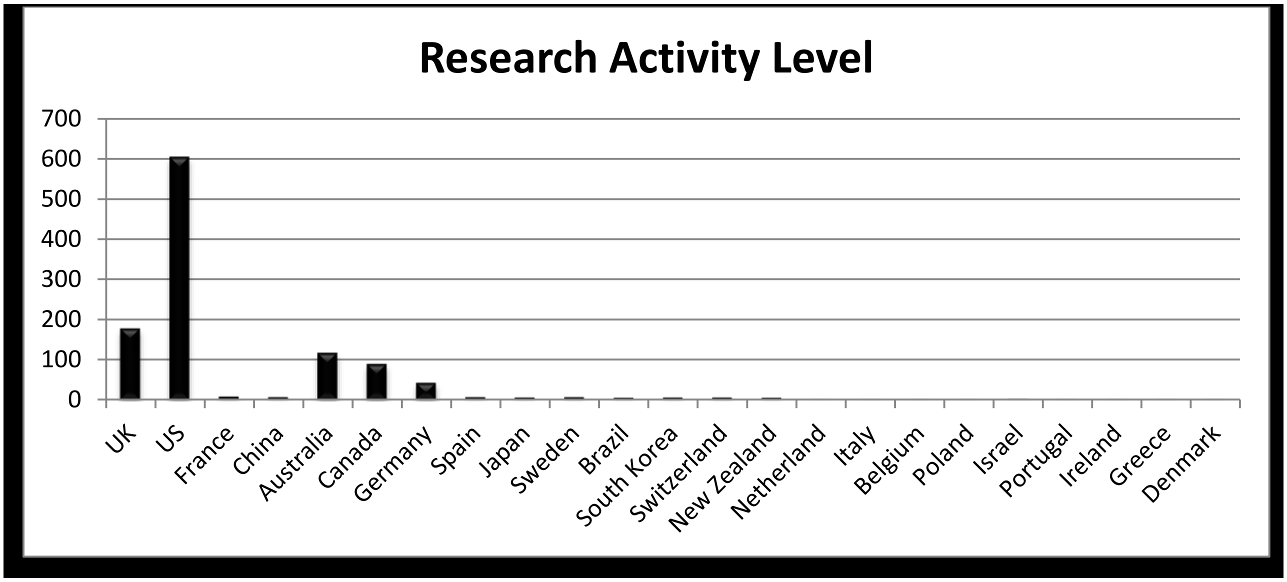


Figure 4.5: Research activity levels of different countries.

In the Figure 4.5 we have analyzed that the United State has the highest research activity level among all countries in the *Geoscience* dataset from *MAS*. Countries like UK, Australia, Canada, Germany and France have a lesser research activity level as compared to the US. All the remaining countries have very low research activity level and the bars represent these countries lie near to zero.

4.6 Country's Diffusion

The number of distinct edges coming to a country represents country's diffusion. For example, if seven distinct edges are coming to USA, this means that the knowledge of USA is diffusing into seven different countries.

The country's diffusion of different countries is represented by the following graph.

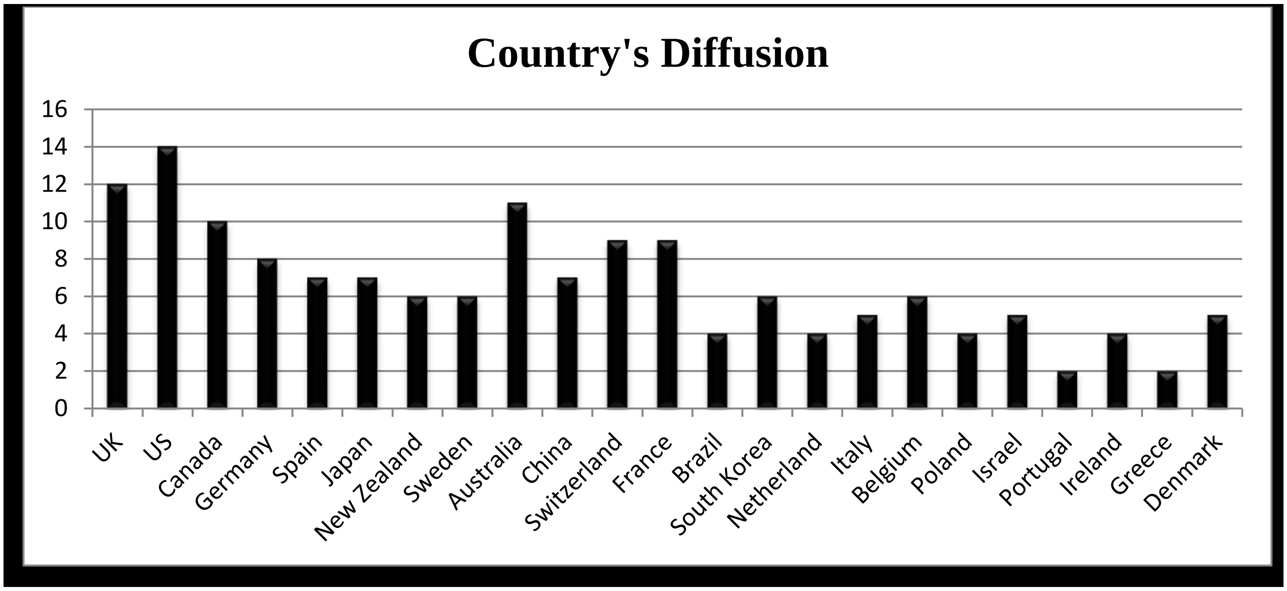


Figure 4.6: Country’s diffusion of different countries

From the Figure 4.6 we can analyze that the United State have the higher country’s diffusion among all countries in the *Geoscience* dataset from *MAS*. All other countries have a low country’s diffusion as compared to US.

4.7 Diffusion Rate

The number of citations coming towards a country represents the diffusion rate of the country. For calculating the diffusion rate of countries the country’s diffusion of each country has been counted. The Diffusion rate of different countries is represented by the following graph.

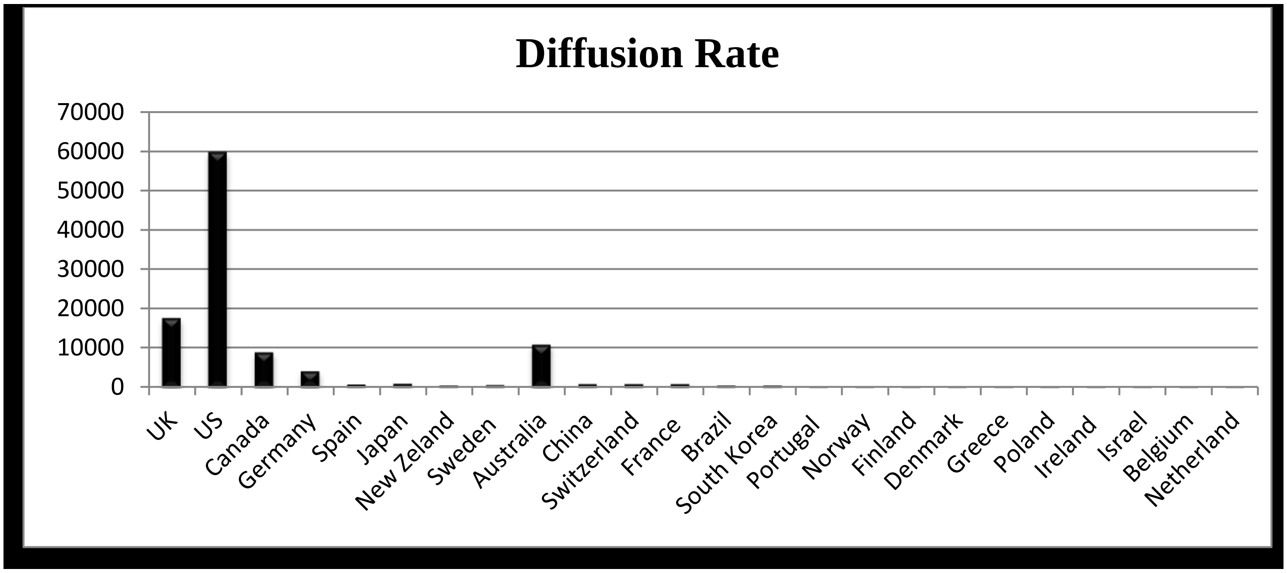


Figure 4.7: Diffusion rates of different countries

From the Figure 4.7, we can analyze that the United State has the highest diffusion rate among all countries in the *Geoscience* dataset from *MAS*. UK have lower diffusion rate as compared to the US. All other countries have a very low diffusion rate as compared to US.

4.8 Visualization

After the identification of knowledge producer, consumers, research activity level, country's diffusion and diffusion rate. We have visualized these five parameters on to the Google map using the Google Visualization Developer tool, so that one can have an overview of knowledge producer, consumers, research activity level, country's diffusion and diffusion rate at once by looking at the map in the Figure 4.8.

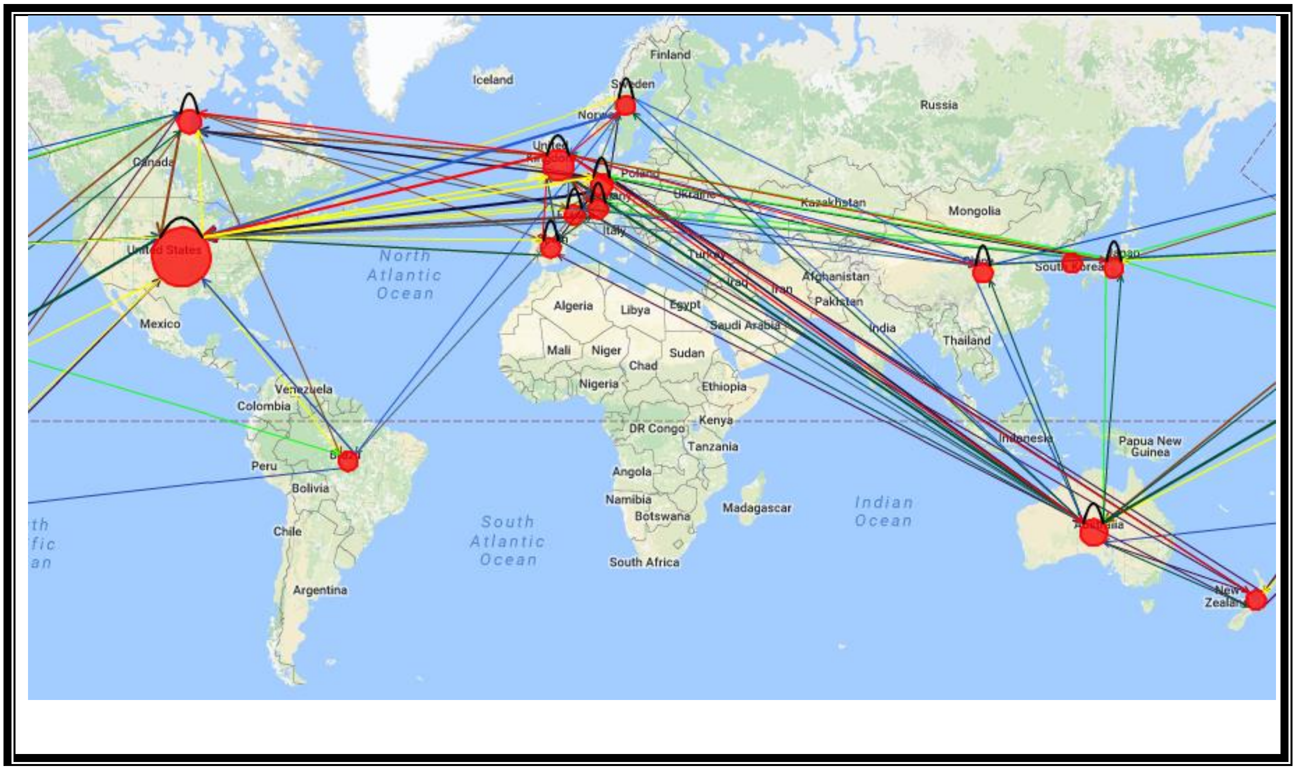


Figure 4.8: Knowledge producers, consumer, research activity level and diffusion rate of countries.

4.9 Evaluation

Every system needs to be evaluated. We have performed an in depth evaluation of our visualization. For evaluating the visualization, we have made a comparison between benchmark and proposed approach.

In this section, we will explain how the benchmark was made and how proposed approach evaluated with benchmark?

4.9.1 Benchmark

We make a benchmark for four parameters producers, consumers, research activity level, country's diffusion and diffusion rate by using well known formulas.

For making the benchmark of producers and consumers, we have used a well known that is described by Zhang et al (Zhang, 2013).

$$\Sigma\text{j}\ w_{ij} \qquad (4.1)$$

" j" is the exporter and "i" is the importer in equation 3.1. Where "w_{ij}" present the amount of knowledge flow from papers written in "j" country to those papers which were written in "i"country.

$$\Sigma\text{j}\ w_{ji} \qquad (4.2)$$

"i" is the exporter and " j" is the importer in equation 3.2. Where "wji" present the amount of knowledge flow from papers written in "i" country to those papers which were written in "j"country.

$$\text{S} = \Sigma\text{ij}\ w_{ij} \qquad (4.3)$$

The total amount of citations either from "i" country to "j" country or from "j" country to "i" country is represented in equation 3.3.

$$\text{Si} = (\Sigma\text{j}\ w_{ji} - \Sigma\text{j}\ w_{ij}\ )/\text{S} \qquad (4.4)$$

To identify the "producer" or "consumer", the equation 3.4 computes the overall scores for country "i", if the value of S_i is positive then the country is producer, else the country is consumer.

We show the results of top five countries in Table 4.2.

Table 4.2 Benchmark of Producer and Consumer

Countries	Values	Identification of producer and consumer
United States	0.53	Producer
Germany	-0.72	Consumer
United Kingdom	0.25	Producer
Canada	-0.31	Consumer
Australia	-0.29	Consumer

The results of five countries represented in the Table 4.2. The results revealed that only two countries US and UK have positive values these are producers. While, other three have negative values and these are consumers. These results used as a benchmark to evaluate proposed approach.

For making the benchmark of research activity level, we used publications. The research activity level of a country is calculated on the basis of publication. The countries that have more publication have a higher research activity level and the countries that have low publications have a lesser research activity level. The research activity level of top five countries is shown in Table 4.3.

Table 4.3 Benchmark of research activity level

Countries	Research activity level
United States	604
Germany	43
United Kingdom	178
Canada	80
Australia	118

The results of top five countries represented in the Table 4.3. Results revealed that US has highest research activity level as compared to the other countries. These results used as a benchmark to evaluate proposed approach.

For making the benchmark of country's diffusion, the number of distinct edges coming to a country represents country's diffusion. For example, if seven distinct edges are coming to US, this means that the knowledge of USA is diffusing into seven different countries. The country's diffusion of top five countries is shown in Table 4.4.

Table 4.4 Benchmark of country's diffusion

Countries	**Country's diffusion**
United States	14
Germany	08
United Kingdom	12
Canada	10
Australia	11

The results of top five countries represented in the Table 4.4. Results revealed that US and UK have higher country's diffusion as compared to the other countries. These results used as a benchmark to evaluate proposed approach.

For making the benchmark of diffusion rate, total numbers of citations are counted for each country. These citations represent the diffusion rate of the country. The countries that have more citation have higher diffusion rate and the countries that have less citations have a lower diffusion rate. The diffusion rate of top five countries is shown in Table 4.5.

Table 4.5 Benchmark of Diffusion rate.

Countries	**Diffusion rate**
United States	58102
Germany	4132
United Kingdom	17734
Canada	6945
Australia	10955

The results of top five countries represented in the Table 4.5. Results revealed that US and UK have higher diffusion rate as compared to the other countries. These results used as a benchmark to evaluate proposed approach.

We used all these results of top five countries that are discussed above as a benchmark for evaluating proposed Visualization. For evaluating visualization we performed a user study that is discussed below.

4.9.2 User study

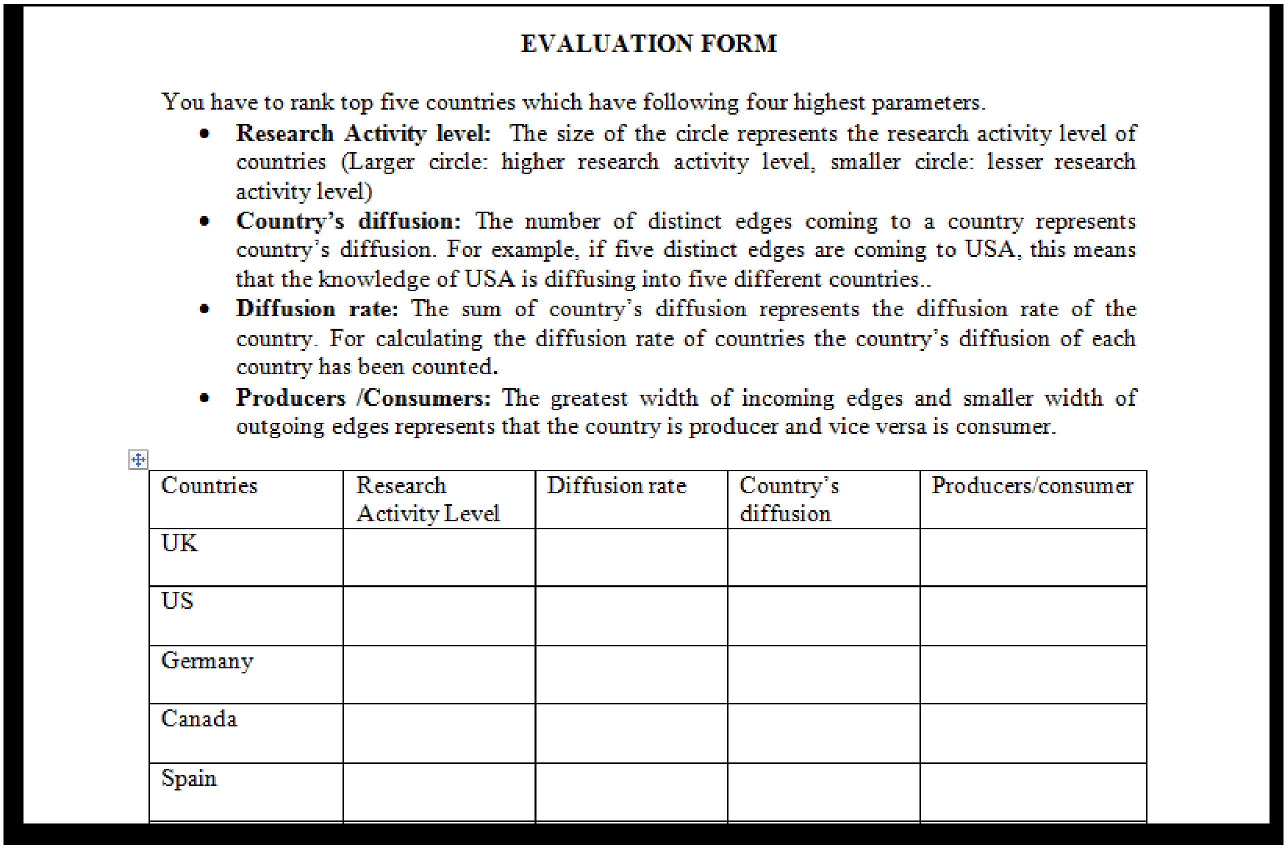
Visualization techniques are evaluated using a dedicated user study (Carpendale S., 2008).This is due to the fact that visualizations are developed to provide ease for the users. Therefore, visualizations should be evaluated by users. For evaluating our visualization, we performed a user study. For user study, we wanted to take those users that have ideas about the parameters like producers, consumer e.t.c. We also wanted to take some users novice users, for evaluation, that have no idea about visualization and its parameters. We explained the parameters in front of these users and they ranked countries based on all five mentioned parameters. Users belonged to four different cities of Pakistan as shown in the Table 4.6.

Table 4.6 statistics of evaluators

Evaluator	**No. of Teachers/ Students**	**Qualification**	**Geographic location**
Faculty members	12	Masters	Muzaffarabad, Mirpur
Students	28	MPhil / Masters	Muzaffarabad Mirpur, Rawalpindi, Islamabad

From 40 users, some users belong to the computer science. These users have knowledge about the four parameters producers, consumers, research activity level, diffusion rate and country's diffusion. However, others are novice users and they have no idea about the visualization.

For the evaluation process, firstly we briefly explain the visualization in front of users that discussed in table 4.6. Visualization that visualizes four parameters producers, consumers, research activity level, diffusion rate and country's diffusion. Every user is provided with an evaluation form and visualization on the laptop. A snapshot from evaluation form is presented in appendix A has been mentioned below.



EVALUATION FORM

You have to rank top five countries which have following four highest parameters.

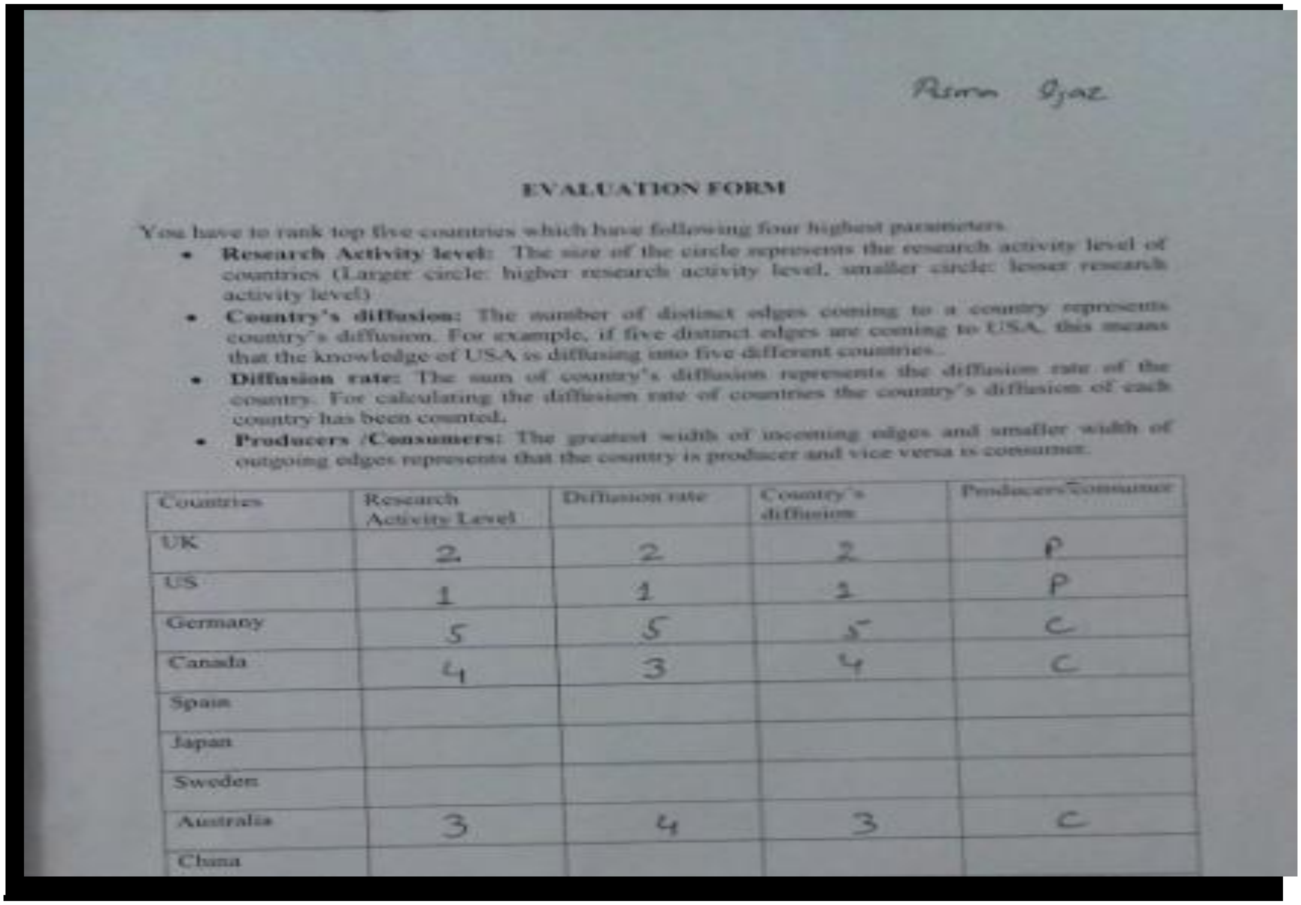
- **Research Activity level:** The size of the circle represents the research activity level of countries (Larger circle: higher research activity level, smaller circle: lesser research activity level)
- **Country's diffusion:** The number of distinct edges coming to a country represents country's diffusion. For example, if five distinct edges are coming to USA, this means that the knowledge of USA is diffusing into five different countries..
- **Diffusion rate:** The sum of country's diffusion represents the diffusion rate of the country. For calculating the diffusion rate of countries the country's diffusion of each country has been counted.
- **Producers /Consumers:** The greatest width of incoming edges and smaller width of outgoing edges represents that the country is producer and vice versa is consumer.

Countries	Research Activity Level	Diffusion rate	Country's diffusion	Producers/consumer
UK				
US				
Germany				
Canada				
Spain				

Figure 4.9 Evaluation form

There were total of 24 countries, however, some of the countries had very low number of publications. Therefore a threshold of 5 publications was applied, all countries having publications less than 5 were removed from the evaluation. This resulted in the exclusion of 9 countries, therefore, we left with 14 countries which were evaluated from 40 users for all five tasks. It was required from the user to rank the top five countries against four parameters.

A snapshot from evaluation form evaluated by user is presented in the Figure 4.10.



Asma Ijaz

EVALUATION FORM

You have to rank top five countries which have following four highest parameters.

- **Research Activity level:** The size of the circle represents the research activity level of countries (Larger circle: higher research activity level, smaller circle: lesser research activity level)
- **Country's diffusion:** The number of distinct edges coming to a country represents country's diffusion. For example, if five distinct edges are coming to USA, this means that the knowledge of USA is diffusing into five different countries.
- **Diffusion rate:** The sum of country's diffusion represents the diffusion rate of the country. For calculating the diffusion rate of countries the country's diffusion of each country has been counted.
- **Producers /Consumers:** The greatest width of incoming edges and smaller width of outgoing edges represents that the country is producer and vice versa is consumer.

Countries	Research Activity Level	Diffusion rate	Country's diffusion	Producers/Consumer
UK	2	2	2	P
US	1	1	1	P
Germany	5	5	5	C
Canada	4	3	4	C
Spain				
Japan				
Sweden				
Australia	3	4	3	C
China				

Figure 4.10 Filled evaluation form

All users filled the evaluation form as represented in the Figure 4.9. Now we have to analyze that, is there any sort of agreement between 40 users. If the agreement between users is more than 80%, it means that this is a good agreement. If the agreement between users is less than 80%, it means another user study should be performed for evaluation or the results need to be evaluated very deeply. For finding this agreement we have calculated Spearman correlation between the 40 users. The Spearman correlation[4] was used because we have ranked lists. The average values that we achieved from correlation between users are: research activity level is 0.93, country's diffusion is 0.88, diffusion rate is 0.84 and producers/Consumer is 0.75. we have analyze that if the value of co-relation is 0.8 or above then these values represent the high positive correlation. The closer the value is to +1, the stronger the relationship.

[4] http://www.statisticssolutions.com/correlation-pearson-kendall-spearman/

4.9.3 Comparison of proposed approach with benchmark

(a) Research activity level

The top 5 countries that have highest research activity level are US, UK, Australia, Canada and Germany as per benchmark. Following graph show the correlation between different users' results obtained from the visualization and the benchmark obtained by state-of-the-art formulae.

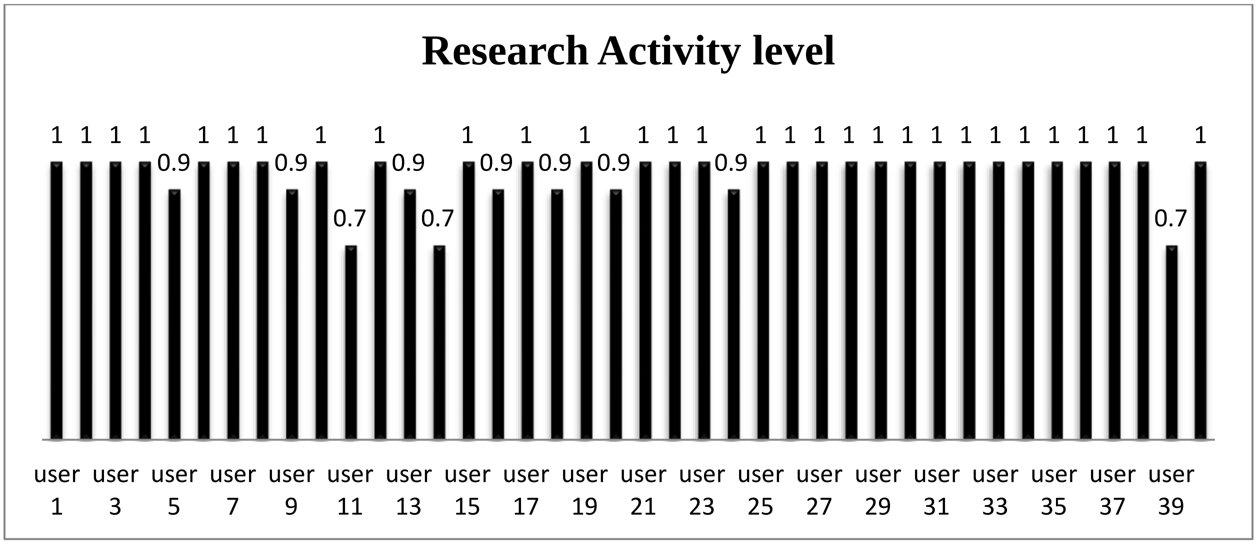


Figure 4.11 Co-relation of research activity level between users' results and benchmark.

In the Figure 4.11 demonstrates the correlation between benchmark and users' ranked list. We have found an important finding that is; all users selected top five countries similar to the countries that are in the benchmark irrespective of their ranks. Therefore, if we just compare the top 5 countries of benchmark and all of the users, we have achieved 100% results. However, we were interested in the ranked results that either the countries listed at specific rank in the benchmark have been placed by the users on the same specific ranks or not? Figure 4.11 demonstrates such results – belonging to correlation of ranking lists. The x-axis represents user's correlation with benchmark, while the y-axis represents the values of correlations. Overall the results remained quite encouraging, 30 out of 40 users fully agreed with the benchmark, 7 agreed as 0.9 while just three have 0.7 agreements with the benchmark.

(b) Country's Diffusion

The top 5 countries that have highest diffusion rates are US, UK, Australia, Canada and Germany as per benchmarks. Following graph show co-relation of country's diffusion between different users' ranked list and benchmark.

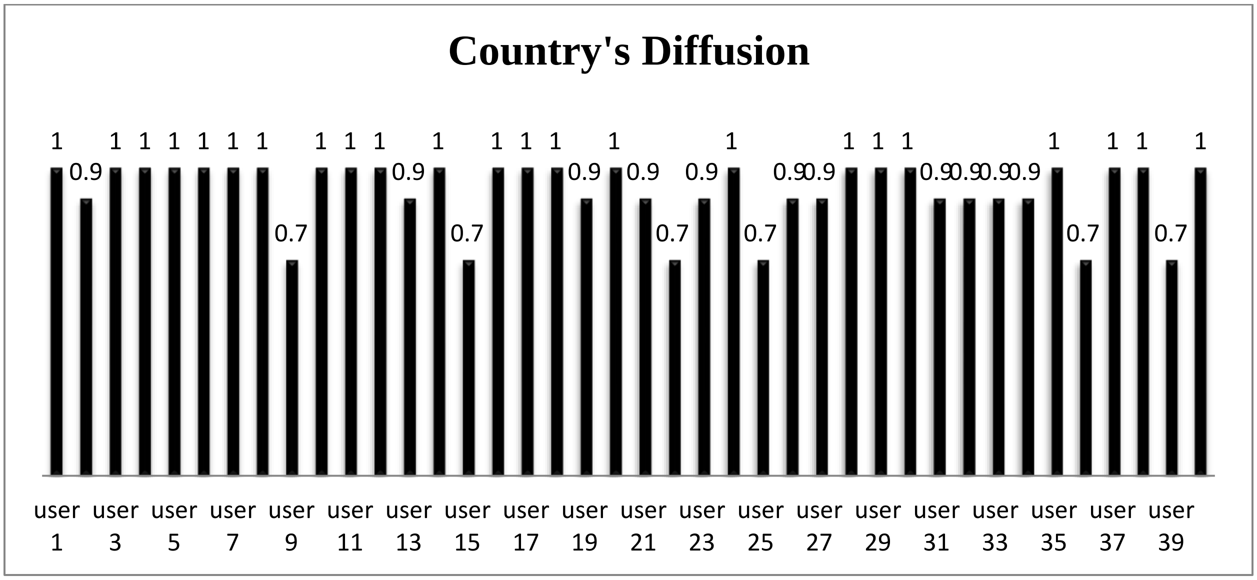


Figure 4.12 correlation of country's diffusion between users' results and the benchmark.

In the Figure 4.12 demonstrates the correlation between benchmark and users' ranked list. We have found an important finding that is; all users selected top five countries similar to the countries that are in the benchmark irrespective of their ranks. Therefore, if we just compare the top 5 countries of benchmark and all of the users, we have achieved 100% results. However, we were interested in the ranked results that either the countries listed at specific rank in the benchmark have been placed by the users on the same specific ranks or not? Figure 4.12 demonstrate such results – belonging to correlation of ranking lists. The x-axis represents user's correlation with benchmark, while y-axis represents the values of correlations. Overall the results remained quite encouraging, 23 out of 40 users fully agreed with the benchmark, 11 agreed as 0.9, while just six have 0.7 agreements with benchmark.

(c) Diffusion Rate

The top 5 countries that have highest diffusion rates are US, UK, Australia, Canada and Germany as per benchmarks. Following graph show co-relation of diffusion rate between different users' ranked list and benchmark.

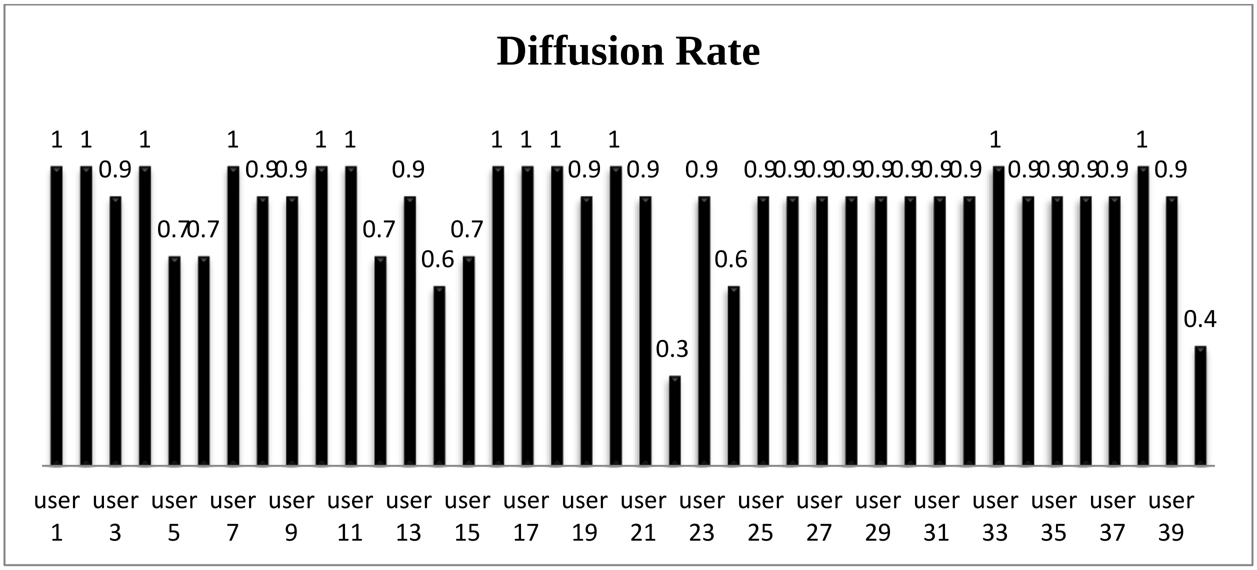


Figure 4.13 Co-relation of diffusion rate between users' results and the benchmark.

In the Figure 4.13 demonstrates the correlation between benchmark and users' ranked list. We have found an important finding that is; all users selected top five countries similar to the countries that are in the benchmark irrespective of their ranks. Therefore, if we just compare the top 5 countries of benchmark and all of the users, we have achieved 100% results. However, we were interested in the ranked results that either the countries listed at specific rank in the benchmark have been placed by the users on the same specific ranks or not? Figure 4.13 demonstrates such results – belonging to correlation of ranking lists. The x-axis represents user's correlation with benchmark, while y-axis represents the values of correlations. Overall the results remained quite encouraging, 12 out of 40 users fully agreed with the benchmark, 20 agreed as 0.9, 4 agreed as 0.7, 2 agreed as 0.6, 1 agreed as 0.4 while 1 has 0.3 agreements with benchmark.

(d) Producer/Consumer

Producing countries export knowledge more than import while consuming countries import knowledge more than export. According to the benchmark from the top 5 countries US and UK are producers. While, Australia, Canada and Germany are consumers.

Following graph show co-relation between different users ranked list and benchmark of producer and consumer.

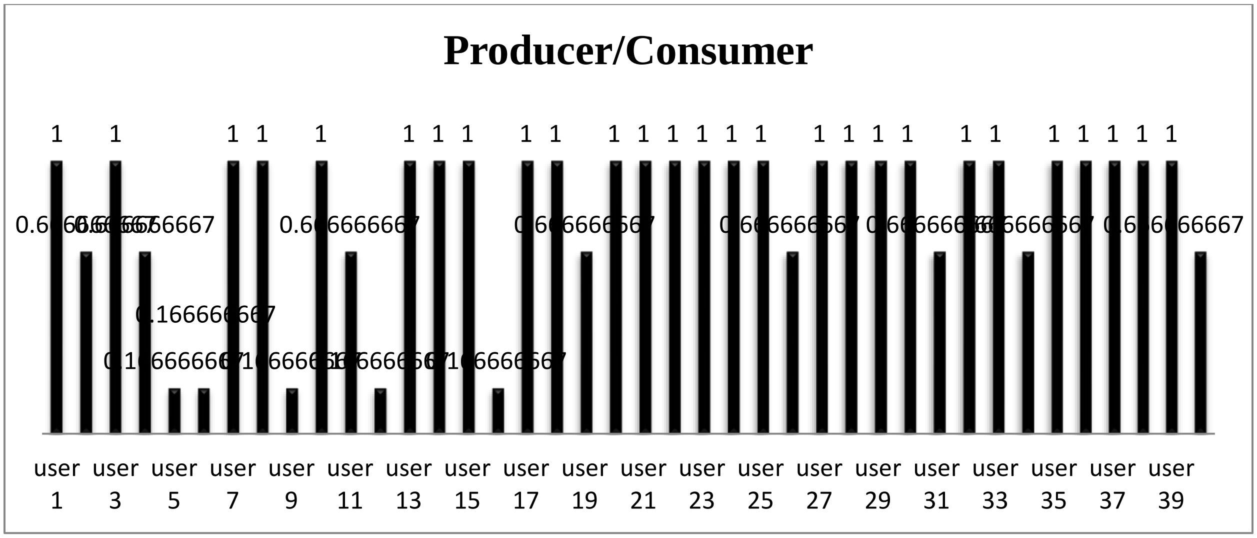


Figure 4.14 Co-relation of consumer/producer between different users' results and benchmark.

In the Figure 4.14 demonstrates the correlation between benchmark and users' ranked list. We have found an important finding that is; all users selected top five countries similar to the countries that are in the benchmark irrespective of their ranks. Therefore, if we just compare the top 5 countries of benchmark and all of the users, we have achieved 100% results. However, we were interested in the ranked results that either the countries listed at specific rank in the benchmark have been placed by the users on the same specific ranks or not? Figure 4.14 demonstrate such results – belonging to correlation of ranking lists. The x-axis represents user's correlation with benchmark, while y-axis represents the values of correlations. Overall the results remained quite encouraging, 27 out of 40 users fully agreed with the benchmark, 8 agreed as 0.67 while 5 has 0.16 agreements with benchmark.

4.9.4 Comparison of state of art approaches

Now we evaluated state-of-the-art approaches. Our first intention was to evaluate the state-of-the-art approaches on the same dataset and with same environment; however, the contemporary approaches have not provided their software code openly available. Therefore, we had to design other ways to evaluate the state-of-the-art visualizations :(1) First thing was to validate previous approaches and the proposed on all knowledge diffusion parameters (producer, consumer, research activity level, country's diffusion and diffusion rate). We used the visualizations that were available in their research papers for evaluation. Those visualizations were also shown to same users used in the above user study. The results of the first comparison have been discussed in the Table 4.7. In this Table, 5 previous visualizations have been evaluated, the details were also discussed in the chapter 2. In this Table, the last row is the new entry depicting the proposed visualization. It is obvious from this Table that none of the previous approaches have visualized all of the knowledge diffusion parameters. There were total of four knowledge diffusion parameters found from literature. One can see that none of the approach has even visualized four parameters together. There were only two approaches out of five which have visualized two parameters. The majority of approaches have just visualized one or two parameters. However, the proposed approach was able to manage all five parameters in a single visualization approach. This was another reason for not being able to use all previous approaches on the same scale for evaluation.

Table 4.7 Five different parameters visualizes in the visualizations.

1	Research papers	Technique	Producers	Consumers	Country's diffusion	Diffusion rate	Research activity level
2	Global Multilevel analysis of scientific food web	Citations	Yes	Yes	No	No	No
3	Characterizing scientific production and consumption in physics	Citations	Yes	Yes	No	No	No
4	Geovisulization of knowledge diffusion case study data mine	Publication	No	No	No	No	Yes
5	Tracing Knowledge diffusion	Citations	No	No	No	Yes	No
6	Mapping the Diffusion of Information Among Major U.S. Research Institutions	Citations+ publication	Yes	Yes	No	Yes	No
7	Tracing and Visualizing Knowledge Diffusion.	Citations+ Publications	YES	YES	YES	YES	YES

The second strategy of evaluation was qualitative comparison in which existing visualizations is presented to the users and to enquire their subjective opinions in comparison to the proposed visualization. The previous visualizations that presented to the users and feedback that has been taken from users is given below:

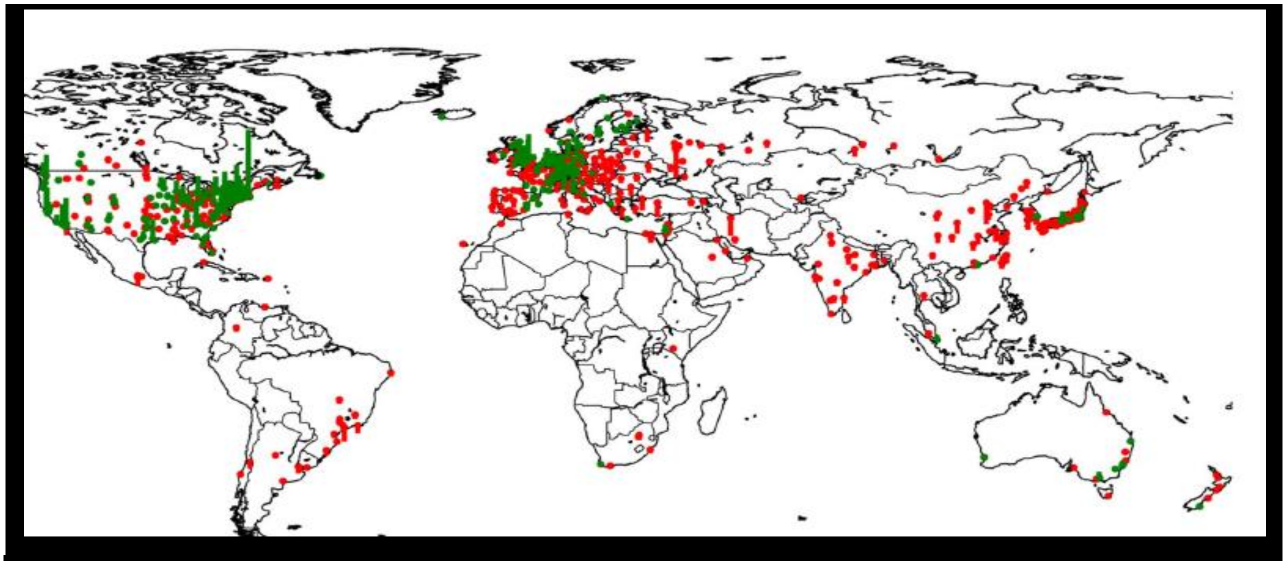


Figure 4.15 Visualization of knowledge producing and knowledge consuming countries.

We have taken feedback from 40 users against the visualization that has been represented in the Figure 4.15. The notable feedback that obtained from users is mentioned below:

“In this visualization we can identify only two parameter producers and consumer from the color of node. Other three parameters country diffusion, diffusion rate and research activity level are missing in the visualization.”

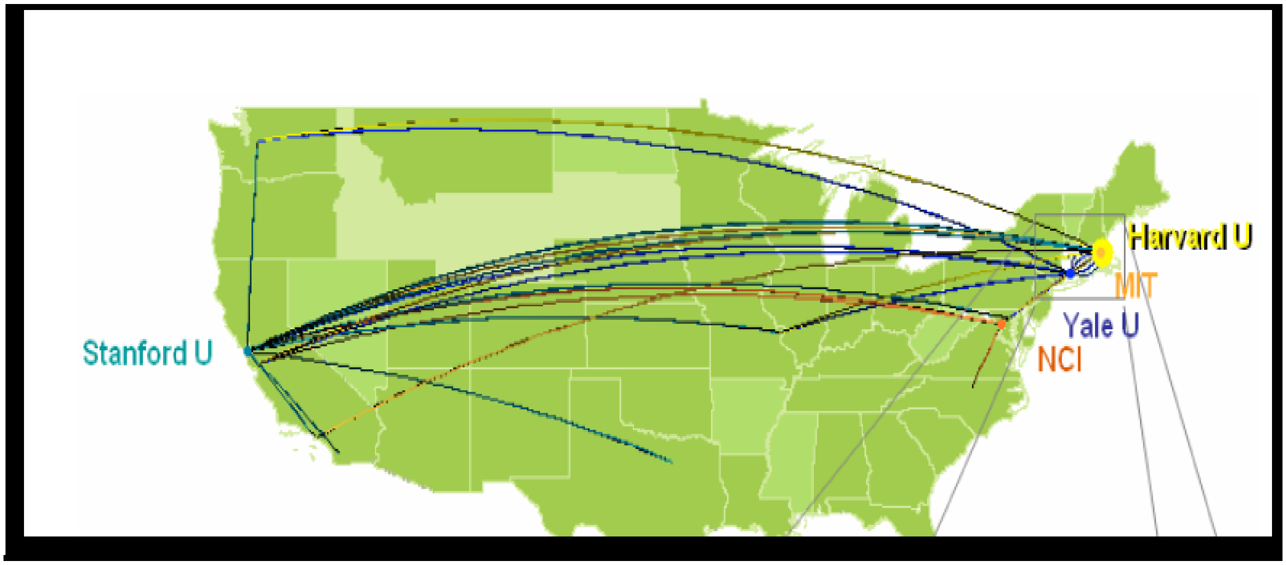


Figure 4.16: Visualization of diffusion among different institutes.

We have taken feedback from users against the visualization that represented in the Figure 4.16. The notable feedback obtained from users is mentioned below:

"In this visualization we can only identify the consumer and producer from the size of node. The knowledge diffusion occurs between the countries. However, we cannot identify how much knowledge is diffused. Moreover, the direction of knowledge flow cannot be identified clearly."

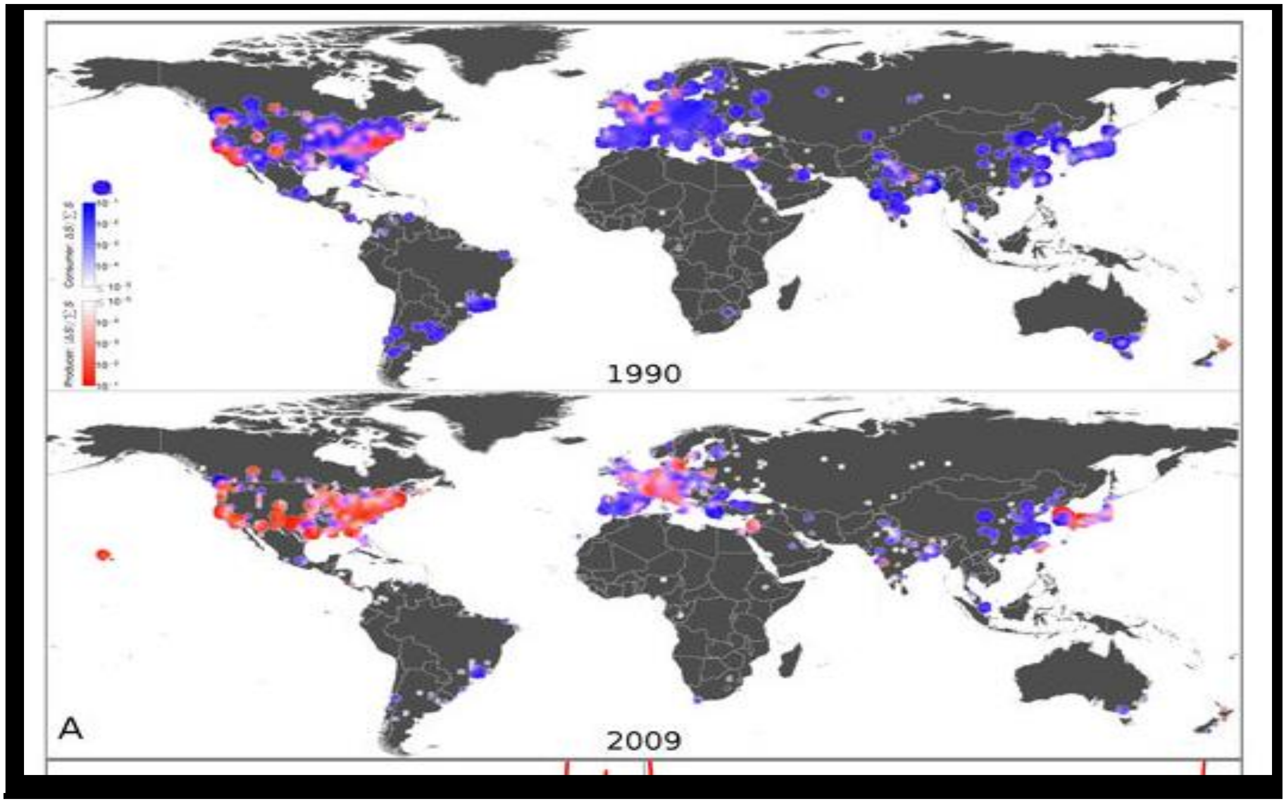


Figure 4.17: Visualization of knowledge source and sink.

We have taken feedback from users against the visualization that represented in the Figure 4.17. The notable feedback obtained from users is mentioned below:

“In this visualization, we can only recognize the consumer and producer that are represented with red and blue color. We cannot identify how much knowledge is diffused between countries. Moreover country’s diffusion and research activity level are also missing.”

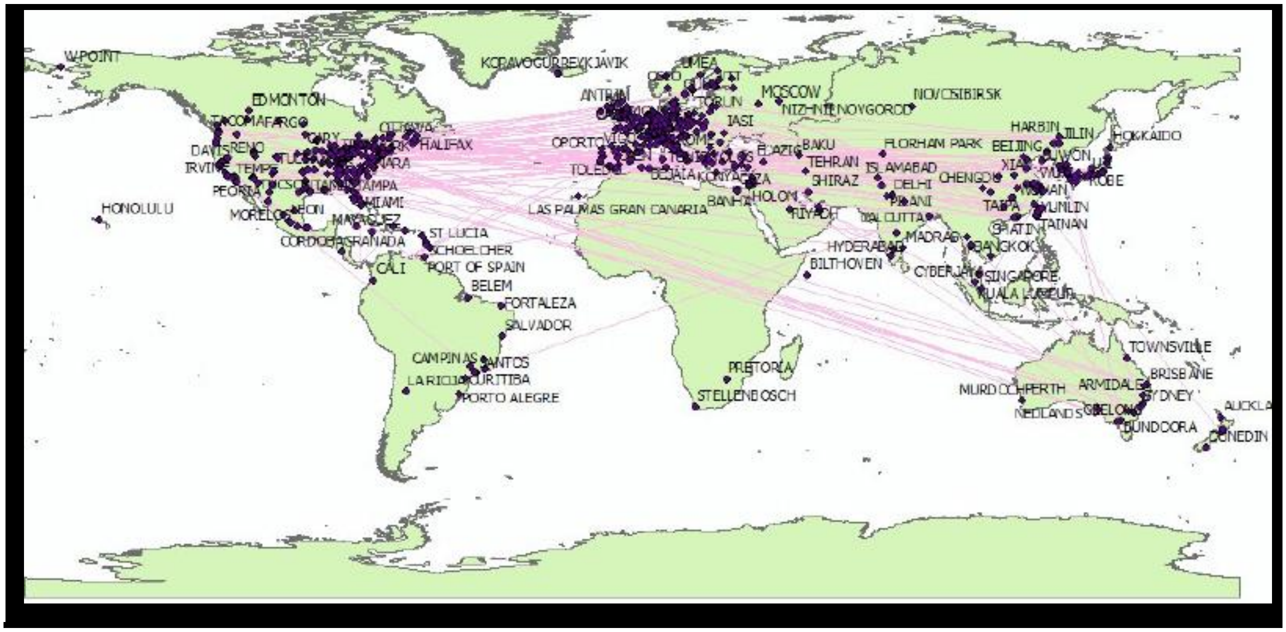


Figure 4.18: Visualization of diffusion among different cities.

We have taken feedback from users against the visualization that represented in the Figure 4.18. The notable feedback obtained from users is mentioned below:

“In this visualization, we can identify the knowledge diffusion take place between the countries. However, we cannot identify the direction of knowledge flow. Moreover, producing and consuming countries cannot identify with the visualization that represented in the Figure 4.18.”

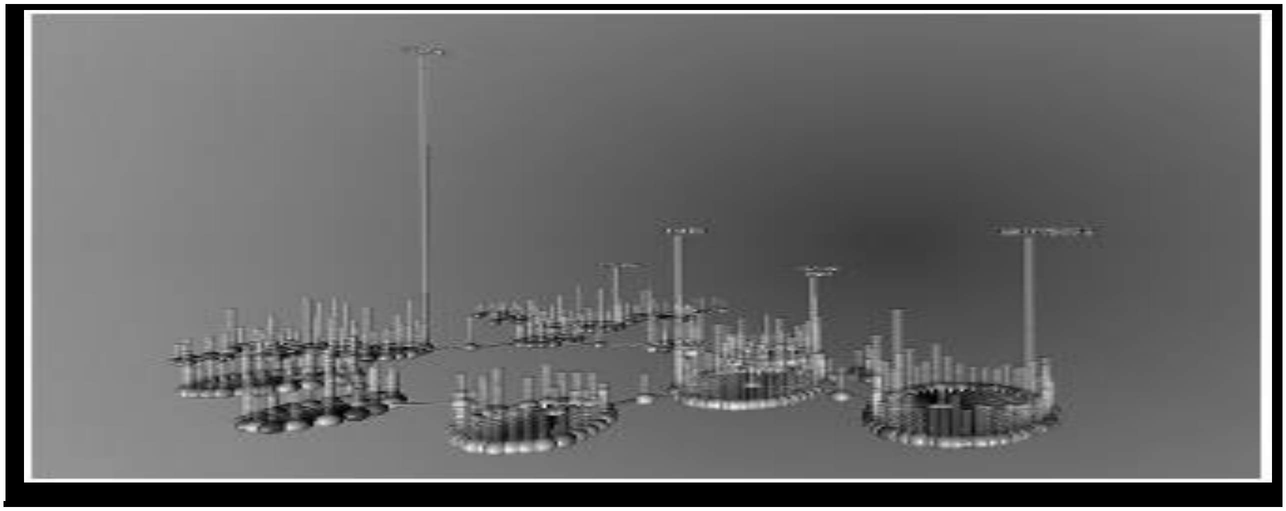


Fig 4.19 Visualization of diffusion among published papers.

We have taken feedback from users against the visualization that represented in the Figure 4.19. The notable feedback obtained from users is mentioned below:

“In this visualization, we can identify the knowledge diffusion take place between the countries with the link. However, we cannot identify the knowledge producer, knowledge consumer country’s diffusion and research activity level in the country.”

The third strategy of evaluation was quantitative comparison in which existing visualizations is presented to the users and they identify how many parameters are there in every visualization. The result of this comparison shows in the table 4.8.

Table 4.8 Users's evaluation against existing approaches and proposed approach.

1	**Visualization**	**Users's evaluation against existing approaches and proposed approach**	
2	Global Multilevel analysis of scientific food web	40 users 2 parvameters visualize	
3	Characterizing scientific production and consumption in physics	29 users 3 parameters visualize	11 users 2 parameters visualize
4	Geo-visualization of knowledge diffusion case study data mine	16 users 2 parameters visualize	24 users 1 parameters visualize
5	Tracing Knowledge diffusion	18 users 2 parameters visualize	22 users 1 parameters visualize
6	Mapping the Diffusion of Information Among Major U.S. Research Institutions	40 users 2 parameters visualize	
7	Tracing and Visualizing Knowledge Diffusion.	40 users 5 parameters visualize	

We proposed an innovative visualization in this chapter. All five parameters visualized in the proposed visualization. All of four parameters (such as: producer, consumer, research activity level, and diffusion rate) were acquired using standard formulae from literature. In this chapter, we made a benchmark of top five countries for all five parameters. We evaluated proposed approach on the benchmark dataset. The correlation results that obtained from the evaluation were research activity level is 0.98, country's diffusion is 0.91, diffusion rate is 0.84 and producers/Consumer is 0.81.

After evaluating proposed approach we have evaluated the state of the art approaches from 40 users. We have done the evaluation process of the previous approaches in two steps. In the first step, we have shown that in previous approaches which parameters were visualized "yes" or "no" according to the users' understanding. In the 2nd step, we have taken feedback from 40 users against visualization and the colourful quotations from users have been mentioned in this chapter. The overall impression of the user was that the state-of-the-art approaches do not visualize all five parameters collectively. However, some authors try to visualize these parameters, but they cannot visualize them effectively as visualized in the proposed approach. To identify knowledge producers, consumers, research activity level, country diffusion, and diffusion rate was very easy with the help of the proposed approach in comparison to the standardized formulae known in the literature.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Knowledge diffusion is a process through which the scientific knowledge is diffused among different social groups. Measuring knowledge diffusion helps researchers to understand the usage of knowledge and generate the facts for the impact of research on the economic and scientific development.

Citation is widely used for measuring the knowledge diffusion. Different researchers contributed in this field and have proposed different measures to measure the knowledge diffusion for example: the identification of knowledge producers and consumers, research activity level, diffusion rate. This thesis after careful and in depth analysis of state-of-the-art research has identified that diffusion rate only let us know the total amount of knowledge diffusion by country 'X', however, the number of countries which get benefits from the knowledge produced by the country 'X' are not used as an indicator of knowledge diffusion. This may happen that the papers written by country 'X' gets acknowledged by just small number of countries say only by 2 countries, however, producing an impact of 'A'. On contrary, there could be large number of independent countries say 15 which might get the benefits from the papers written by country 'Y', however, making a total impact of 'B' where 'B' is smaller than 'A'. On ranking based on diffusion rate, the country 'X' would be placed higher in ranking as compared to country 'Y' which is okay, however, the versatility of country 'Y' for getting attraction by 15 independent countries as compared to diffusion rate to just 2 countries for country 'X' would remain hidden. Therefore, this thesis proposes another measure called country's diffusion for measuring the versatility of knowledge consumption by independent other countries for a specific country.

Furthermore, this thesis identified many state-of-the-art visualization approaches and concluded that majority of them just visualize only one or two parameters at one place. There are only two approaches which could visualize three parameters in their proposed visualization. Therefore, an innovative visualization was developed and evaluated which is not only able to visualize all of the known four parameters, but also able to visualize the newly proposed fifth parameter of 'country diffusion'.

For developing and evaluating the visualization, we have collected the data set of Geoscience from Microsoft Academic Search (MAS). That data set consists of 1100 publications having over 110, 000 citations. After developing visualization, we performed user study. The detailed results of user study are explained. During the development of visualization, we have to face a problem, because the publication and citations are in thousands and hundreds of thousands, so we have to scale these values for visualizing them on Google map.

After the development of visualization, we have made benchmark from well-known formulae of producer, consumer, and research activity level and diffusion rate. For the evaluation process we performed extensive user study. For selecting users, we have made sure to include users from different groups ranging from expert to novice users based on the knowledge of visualization and domain knowledge. This user study was performed from 40 users and we have evaluated the visualization from these users against five research tasks. During the evaluation process, we have calculated the spearman correlation between forty users to analyze inter rater agreement between 40 users. The average values that we achieved from inter rater agreement between users are as follows: research activity level is 0.93, country's diffusion is 0.88, diffusion rate is 0.82 and producers /Consumer are 0.75. We have evaluated the user's results with benchmarks. An important finding is that: all users selected top five countries exactly the same to the countries that were present in top five positions in the benchmark irrespective of their ranks. Therefore, if we just compare the top 5 countries of benchmark and of users, we have achieved 100% correct results. However, we were interested to evaluate the corresponding ranking positions between the users and benchmark dataset. In case of research activity level 30 out of 40 users fully agreed with the benchmark, 7 agreed as 90%, while just three have 70% agreements with benchmarks. The results of country' diffusion are 23 out of 40 users fully agreed with the benchmark, 11 agreed as 90%, while just six have 70%, agreement with benchmark. The results of diffusion rate are 12 out of 40 users fully agreed with the benchmark, 20 agreed as 90%, 4 agreed as 70%, 2 agreed as 60%, 1 agreed as 40%, while 1 has 30% agreements with benchmark. While, the results of producer and consumer are, 27 out of 40 users fully agreed with the benchmark, 8 agreed as 67% while 5 has 16% agreements with benchmark. The average agreements for all five tasks by 40 users remained 84%. After evaluating benchmark with proposed approach, we have evaluated the state of the art approaches in two ways. The first thing was to validate previous approaches and the proposed approach on all knowledge diffusion parameters (producer, consumer, research activity level, country’s diffusion and diffusion rate). The results of this manual evaluation remained inline as per the critical analysis of the literature, most of the approaches just visualized only one or two parameters, there were only two visualizations which visualized three parameters, whereas the proposed approach was able to visualize all five parameters in one visualization. The second type of evaluation with the state-of-the-art was to evaluate their visualizations available in their research papers by the 40 users who evaluated the proposed approach. Users have given similar comments as highlighted in the above manual evaluation. The users were of the point of view that the state-of-the-art visualizations can just only be used to visualize a part of the dimensions of knowledge diffusion; however, it’s not possible to visualize all knowledge diffusion parameters comprehensively in any of the contemporary visualizations which were possible in the proposed visualization. Some of the colorful quotations of the users have been mentioned in the chapter 4.

Our contributions in this thesis are follows:

(1) We proposed one of the important parameters, i.e. country’s diffusion. This parameter represents the amount of knowledge diffusion spectrum of one country into other independent countries.

(2) We proposed an innovative visualization that visualizes the five important parameters of knowledge diffusion collectively as compared to the state-of-the-art approaches which were only able to visualize one or two parameters most of the time.

(3) We have performed in depth user’s study and the results have shown the potential of the proposed approach in comparison with the state-of-the- art approaches.

5.2 Limitations

There are certain limitations associated with this research. For example, we have collected the top hundred papers from all the categories of Geoscience and top hundred citations of each paper were also extracted. It is not necessary that the producers and consumer will always remain the same if more comprehensive dataset is added. However, in this research, our aim was not to identify the knowledge producers and consumers. However, our aim was to propose a visualization technique which could visualize the five important parameters of knowledge diffusion collectively.

5.3 Future work

The future possibilities to extend this work could be to apply it for variety of datasets like mathematics, computer science, chemistry and physics, etc. This will further demonstrate the potential of the proposed visualization when there might be more producers as compared to consumers in domains other than *Geoscience.*

Moreover, the proposed visualization can be developed to make interactive features. When the dataset is more comprehensive, the interactive visualization may be more helpful for better understanding the trends and patterns of knowledge diffusion.

Bibliography

Azoulay, P., Zivin, J. S. G., & Sampat, B. N. (2011). The diffusion of scientific knowledge across and space. Evidence from professional transitions for the superstars of medicine. (No. w16683) *In: "National Bureau of Economic Research*",*(*January 4).

Börner, K., Penumarthy, S., Meiss, M., & Ke, W. (2006). Mapping the diffusion of scholarly knowledge among major US research institutions. In: "*Scientometrics". (*July 26, USA), *68*(3), (pp. 415-426).

Bacchiocchi, E., & Montobbio, F. (2009). Knowledge diffusion from university and public research. A comparison between US, Japan and Europe using patent citations. In: "*The Journal of Technology Transfer",34*(2), (pp. 169-181).

Belenzon, S., & Schankerman, M. A. (2010). Spreading the word: geography, policy and university knowledge diffusion. In: "*LSE STICERD", Research Paper No. EI50*.

Butler, L. (2003). Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. In: "*Research policy", 32*(1), (pp. 143-155).

Chen, Zifeng, and Jiancheng Guan (2016). "The core-peripheral structure of international knowledge flows: evidence from patent citation data." *R&D Management* 46.1 (pp. 62-79).

Carpendale, S. (2008). Evaluating information visualizations. *Information Visualization*, In: "*Springer Berlin Heidelberg",(July,* Waikoloa, HI, *USA),* (pp. 19-45).

Gans, J. S., Murray, F. E., & Stern, S. (2013). Contracting over the disclosure of scientific knowledge.Intellectual property and academic publication, In: *"National Bureau of Economic Research",* (No. w19560).

Gardner, Philip L., Ann Y. Fong, and Roshena L. Huang. (2010). "Measuring the impact of knowledge transfer from public research organizations: a comparison of metrics used around the world. In: " *International Journal of Learning and Intellectual Capital",* (7) (pp. 318-327).

Hudson, J. (1996). Trends in multi-authored papers in economics. In: "*The Journal of Economic Perspectives"*, (June 9) *10*(3), (pp. 153-158).

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. In: "*Proceedings of the National academy of Sciences of the United States of America",*(Nov. 15, San Diego, USA) 102(46), 16569-16572.

Hassan, S. U., & Haddawy, P. (2013). Measuring international knowledge flows and scholarly impact of scientific research. In: "*Scientometrics", 94*(1), (pp.163-179).

Hubbard, D. E. (2012). *Geovisualization of knowledge diffusion: Visualization of bibliographic data 1995-200, (* (Doctoral dissertation).

Ho, J. C., Saw, E. C., Lu, L. Y., & Liu, J. S. (2014). Technological barriers and research trends in fuel cell technologies: A citation network analysis. Technological Forecasting and Social Change, In: *Scientometrics* 82, (pp. 66-79).

Jaffe, A. B., & Trajtenberg, M. (1999). International knowledge flows: evidence from patent citations. In: "*Economics of Innovation and New Technology",* (October 18, USA), *8*(1-2), (pp. 105-136).

Katz, J. S., & Martin, B. R. (1997). What is research collaboration?. In: "*Research policy", 26*(1), (pp. 1-18).

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. In: "*Social studies of science",* (June 2, Maryland, USA), *35*(5), (pp. 673-702).

Leydesdorff, L., & Zhou, P. (2007). Nanotechnology as a field of science: Its delineation in terms of journals and patents. In: "*Scientometrics", 70*(3), (pp. 693-713).

Liu, Y., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. In: "*Journal of the American Society for Information Science and Technology",* (September 8), *61*(2), (pp. 340-351).

Marx, W., Schier, H., & Wanitschek, M. (2001). Citation analysis using online databases: feasibilities and shortcomings. In: "*Scientometrics", 52*(1), (pp. 59-82).

Mazloumian, A., Helbing, D., Lozano, S., Light, R. P., & Börner, K. (2013). Global multi-level analysis of the "Scientific Food Web". (January 30), "*Scientific reports", 3.*

Ma, F. C., Lyu, P. H., Yao, Q., Yao, L., & Zhang, S. J. (2014). Publication trends and knowledge maps of global translational medicine research. In: “*Scientometrics”*, *98*(1), (pp.221-246).

Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the national academy of sciences*,(April 6, UK), *101*(1), 5200-5205.

Plaisant, C. (2004, May). The challenge of information visualization evaluation. In: “*Proceedings of the working conference on Advanced visual interfaces”, (*June 3, Maryland, USA), (pp. 109-116).

Rowlands, I. (2002, April). Journal diffusion factors: a new approach to measuring research influence. *Aslib Proceedings, (*April 3, Germany), 54(2), (pp. 77-84)

Rousseau, R., Van Hecke, P., Nijssen, D., & Bogaert, J. (1999). The relationship between diversity profiles, evenness and species richness based on partial ordering. In: “*Environmental and Ecological Statistics”*, *6*(2), (pp. 211-223).

Sorenson, O., & Fleming, L. (2004). Science and the diffusion of knowledge. In: “*Research policy”*, *33*(10), (pp. 1615-1634).

Sonnenwald, D. H., Lassi, M., Olson, N., Ponti, M., & Axelsson, A. S. (2009). Exploring new ways of working using virtual research environments in library and information science. In: “*Library Hi Tech”, (*Feb. , 15 USA), *27*(2), (pp. 191-204).

Stolpe, M. (2002). Determinants of knowledge diffusion as evidenced in patent data: the case of liquid crystal display technology. In: “*Research Policy”*,*31*(7), (pp. 1181-1198).

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. In: “*BMJ: British Medical Journal”*, *314*(7079), 498.

Saeed, A. U., Afzal, M. T., Latif, A., & Tochtermann, K. (2010). Disseminating Knowledge through Tags: Recommending Tags for Scientific Resources. In: “*Journal of IT in Asia”,3.*

Tsai, C. C., & Lydia Wen, M. (2005). Research and trends in science education from 1998 to 2002: A content analysis of publication in selected journals. In: "*International journal of science education"*, (April, 23, USA), *27*(1), (pp. 3-14).

Wagner, C. S. (2006). International collaboration in science and technology: Promises and pitfalls. In: *Science and technology policy for development,* Anthem press, 2006.

Zhang, L. (2012). A tapered diffusion impact indicator: A preliminary exploration on the journal level. In: "*Malaysian Journal of Library & Information Science"*, (Decmber 4), *17*(3), (pp. 67-72).

Zhang, Q., Perra, N., Gonçalves, B., Ciulla, F., & Vespignani, A. (2013). Characterizing scientific production and consumption in Physics. "*Scientific reports"*(Feb. 26, USA), *3*.

Appendix A

EVALUATION FORM

You have to rank top five countries which have following four highest parameters.

- **Research Activity level:** The size of the circle represents the research activity level of countries (Larger circle: higher research activity level, smaller circle: lesser research activity level)
- **Country's diffusion:** The number of distinct edges coming to a country represents country's diffusion. For example, if five distinct edges are coming to USA, this means that the knowledge of USA is diffusing into five different countries..
- **Diffusion rate:** The total number of citations toward a country represents the diffusion rate of the country. For calculating the diffusion rate of countries the country's diffusion of each country has been counted**.**
- **Producers /Consumers:** The greatest width of incoming edges and smaller width of outgoing edges represents that the country is producer and vice versa is consumer.

Countries	Research Activity Level	Diffusion rate	Country's diffusion	Producers/consumer
UK				
US				
Germany				
Canada				
Spain				
Japan				
Sweden				
Australia				
China				
Switzerland				
France				
South Korea				
New Zealand				
Brazil				