T.R.N.C

TURKISH REPUBLIC OF NORTHERN CYPRUS

NEAR EAST UNIVERSITY

INSTITUTE OF HEALTH SCIENCES

THE BIBLIOMETRIC ANALYSIS OF BAYESIAN NETWORK METHOD IN WEB OF SCIENCE DATA FROM 1990 TO 2020

EWURU DEBORAH AMAKA

Master of Science in Biostatistics

NICOSIA, 2020

T.R.N.C

NEAR EAST UNIVERSITY

INSTITUTEOF HEALTH SCIENCES

THE BIBLIOMETRIC ANALYSIS OF BAYESIAN NETWORK METHOD IN WEB OF SCIENCE DATA FROM 1990 TO 2020

Master of Science in Biostatistics

Supervisor:

ASSISTANT PROF. DR. OZGUR TOSUN

NICOSIA, 2020

Thesis submitted to the Institute of Health Sciences of Near East University in partial fulfillment of the requirement for the degree of Master of Science in Biostatistics.

Thesis Committee;

Chair of the committee: Prof. Dr. İlker Etikan

Near East University

Sig: .....

Advisor: Assist. Prof. Dr. Özgür Tosun

Near East University

Sig: .....

Member: Assoc. Prof. Dr. Hülya Olmuş

Gazi University

Sig: .....

Approved by: Prof. Dr. K. Hüsnü Can Başer

Director of Health Science Institute

Near East University

Sig: .....

DECLARATION

I hereby declare that all information in this thesis has been obtained based on my research results pertaining to Bayesian Network and presented in accordance with academic rules and ethical conduct. I also declare that, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Deborah Amaka, Ewuru

Signature:

Date: 29th September, 2020

ACKNOWLEDGEMENT

First and foremost, I want to give thanks to the Almighty God for the successful completion of this research and also for the gift of life and the wisdom he bestowed upon me throughout this course.

I would like to express my heartfelt gratitude to my great supervisor, Assistant Prof. Dr. Ozgur Tosun for his immense guidance, corrections and support during this research work. I also want to thank our able Head of Department (HOD), Prof. Dr. Ilker Etikan for his motivation and encouragement throughout this course.

My sincere gratitude goes to my parent, family and friends for their constant prayer and support morally and economically.

Lastly, I acknowledge my very good friend, Opara Dennis Ifeanyi for the words of encouragement and his immense contribution throughout the course of this study.

DEDICATION

This work is dedicated to Almighty God for grace and gift of life. To my loving parent, siblings and friends, I love you all.

ABSTRACT

Introduction: Bibliometrics is the scientific field of study that quantitatively analyses all bibliographic data statistically. This method is often used to detect the growth of some particular fields. Bayesian Networks have been used to achieve great results in solving real world problems in different fields like medicine, biomedicine and health care and Artificial Intelligence.

Objective: The aim of this study was to determine the literature growth, research impact and the most productive journal s that have published articles concerning Bayesian Networks.

Method: This study applied a combination of two methods which are Bibliometric analysis and Visualization. The dataset used in this study was extracted, analyzed and downloaded from Web of Science. Co-authorship analysis, co-citation analysis, keyword co-occurrences analysis and science mapping were the bibliometric approaches assessed regarding our research topic.

Results: A total of 1739 publications and 46,167 sum of citations from the year 1990 to 2020 was produced by Web of Science from 1990 to 2020.Environmental Modelling & Software journal had the highest number of publications with 53 records. USA produced the greatest percentage of publications with 551 records and 31.69%. The key author was Bartels PH with 28 publications.

Conclusion: There was a rapid increase in the yearly research activities pertaining to Bayesian Network research from the period of 1990 to 2020 both in the number of publications and citations. The Bayesian Network research publications produced h-index of 83 which showed great productivity and impact.

Keywords: Bibliometric analysis, Bayesian network, Web of Science, VOS-viewer,Visualization.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

DAG: Directed Acyclic Graph

D-dimer: Domain dimer

JPD: Joint Probability Distribution

KSA: The Kingdom of Saudi Arabia

MERS-COV: Middle East Respiratory Syndrome Coronavirus

PGM: Probabilistic Graphical Model

SCI: Science Citation Index

SCIE: Science Citation Index Expanded

USA: United States of America

CHAPTER 1
INTRODUCTION

Background of study

Bibliometric analysis is an essential statistical method employed in several research fields to measure the impact of countries, authors, institutions, organizations, publications and journal articles in different subject areas.This method has also enabled analyst to identify new and emanating research areas,journals where one can publish his or her articles and possible funding agencies by providing them with evidence that shows the research they are promoting has the capacity to influencethe society with advances and innovations (Bornmann, 2014 and Brueton etal, 2014).Scientometrics is the sub-specialty of bibliometrics that is involved in evaluating and assessing scientific literatures.

Bibliometrics is used for research assessment in studying, evaluating, and examining scientific outputs. The importance of bibliometric studies is so numerous, it uncovers the space that require to be closed in the literature works and vital studies with the aim of supporting and promoting research proposals. It produces papers that have been regarded to give appropriate and dependable sources of results. They are frequently used in finding out recent and relevant research problems using software tools, and also in explaining conclusions or findings on job offers,promotions, funds and research policies (Campbell et al, 2010and Gläseret et al,2015).Citation analysis is often used in bibliometric study to discover relationships between intellectual authors or their research works in different subject field and it is formulated on building networks between documents or citation graphs (Schaer, 2013*).*

Bibliometric studies are conducted using citation databases and indexes. Citation databases are groups of published referenced journal articles, books and conference proceedings moved into online systems. Each record contains the followingitems; title or topic, abstract, author,keywords, full text and publication information, these items serve as different fields used in extracting details from the database using any of the items. Examples of citation databases include Web of Science, Scopus, Pubmed, CrossRef, Google Scholar,

Dimensions and many more, but the two most commonly used databases are Web of Science and Scopus.

Visualization methods and science mapping are used in analysis of data in scientometric and bibliometric studies. Both techniques have been integrated into the different invented software tools such as VOS-Viewer,CiteSpace,CitNet explorer and many more. They assist in visualizing the impact of scientific knowledge by displaying the strong structural relationship of scientific research domains (Pradhan,2016)

Bayesian Network is a type of Probabilistic Graphical Model(PGM) that shows if the relationship between random variables involves uncertainty. This method builds models manually from the knowledge of an expert in the underlying domain or automatically from large datasets by using the suitable software, or a combination of the two processes (Michal, 2014). These models or networks can be used in anomaly detection, prediction, automatically extract insight, prediction and decision automation.It can also be usedin several other tasks such as; medical diagnosis, modeling and explaining domains,promoting decision making under uncertainty, discovering the most possible configurations of variables, calculating conditional probability distributions and identifying outstanding strategies for solving problems in domains with unpredictability. It is a representation of Joint Probability Distribution(JPD) (Jirka, 2005).

Systematic Review

This is a high-level comprehensive summary of primary researchor scholarly literatureon a specific research question that systematically discovers, chooses, assesses, and integrates all high-quality research evidence related to that question so as to answer it. In a systematic review, the details from published and unpublished studies are reviewed and written by a group of experienced professionals. Systematic reviews present its evidence from a large set of research and different author’s perspectives. This study produces reliable and accurate conclusions. The importance of any systematic review is dependent on the quality,quantity, and heterogeneity of the involved studies. The commonly searched databases for systemic review are, Web of Science, PubMed (Medline, Embase, etc. Not all systematic reviews involve meta-analysis. Systematic reviews and meta-analyses are

found at the apexof the “Evidence Pyramid”. They are both regarded as the highest-quality evidence on a research topic and uses PICO (Population, Intervention, Comparison, Outcome) guide in formulating a research question. In addition, their study design reduces bias and also creates more valid findings.

Meta-Analysis

This type of quantitative, formal, and epidemiological study is usually a part of a systematic review. This study is typically but not necessarily built on randomized, controlled clinical trials. Meta-analysis is a statistical procedure that integrates the data or results from reviewed individual research studies to make conclusions about their findings and produce a single summary estimate of the effect. This study is carried out when multiple scientific research trying to answer a particular research question are expected to have some degree of error in their reported measurements (effect size or treatment effect). If there is a common effect size between studies, meta-analysis assists in discovering it. The results of the meta-analysis are usually displayed in a forest plot or blobbogram and can be combined using different approaches like Inverse- variance method, Mantel-Haenszel method, the Peto method, etc. (Haidich, 2010). The conclusions gotten from the meta-analysis are statistically greater than the analysis of any single study. There is an argument that the validity of meta-analysis produces unreliable or misleading results, this is because it combines data from different studies.

The difference between Bibliometric analysis with Systematic Review and Meta-Analysis

Bibliometric Analysis	Systematic Review	Meta-Analysis
This is the scientific field of study thatquantitatively and statistically analyses all bibliographic data to provide a summary of the body of knowledge for a given field of study.	This is the complete process of choosing, assessing, and integrating all high-quality research evidence.	This is the statistical approach that combines the data obtained from a systematic review.
Produces a summary of a specific topic.	Answers a certain clinical research question.	Uses statistical techniques to review the results of similar multiple research studies.
Increase Bias	Minimize Bias in all stages of the review process.	Increase Bias
Searches of one or more databases can be made	Comprehensive and include all relevant databases and individual journals are searched	Two or more databases can be searched.
Only published studies (peer-reviewed publications) are used.	Both publishedand unpublished studies are used	Unpublished studies are included in published studies to avoid publication bias

1.1Statement of the problem

It is very crucial to carry out Bibliometric studies on published scientific literatures with the purpose of finding out or discovering the significance, usage and relevance of these research articles. Many journal articles, books, editorials, conference proceedings and reviews on different topics or titles published by great researchers that have positively influenced the society by their applications are yet to be studied or reviewed. This study aims to analyze the journal articles on the application of Bayesian Networks in sciences.

Current research targets to measure the growth and impact of this topic to the society over the years from the appropriate information sources in Web of Science database

1.2Research questions

The study questions to be answered at the end of this study include:

- Is there any growing trend in Bayesian Network publications?
- Is there any impact of the applications of Bayesian Network research?
- Which scientific journals, disciplines, authors and countries are the most pertinent regarding the use of Bayesian Networks in research?

1.3 Aims and objectives

The aim of this study is to determine the literature growth and the research impact of the journal articles that have been published concerning Bayesian Networks. Bibliometric analysis and Visualization were applied to Bayesian Network journals in Web of Science database within the period of 1999 to 2020 to investigate its trends and productivity.

Objectives

The main objectives of this study are to identify the following the growing trend regarding Network,the impact of the usage of Bayesian Network in different fields and the most important scientific journal in the field of Bayesian Network.

1.4 Significance of the study

In this study, the researcher's target is to reveal the yearly changes in the usage of Bayesian Networks. This will help us to know if this method is becoming popular in research or not and also if there has been any increase in its usage. It will also discover the research area of Bayesian Network that has gotten greater awareness and attraction by analysts, organizations and Institutes. This study has the capability of making a positive impact in the world at large both in future research and clinically with respect to the importance of Bayesian Networks in different disciplines

1.5 Structure

The rest of this dissertation is arranged in chapters which include Chapter 2, talks about the literature review related to this study. Chapter 3 introduces the materials and methods used in this study such as the study design,research methodologies,data collection and analysisof data. The results in detail were explained in Chapter 4. Lastly, Chapter 5 summarizes the entire paper and discusses the significant results

CHAPTER 2
LITERATURE REVIEW

2.1 Bibliometric Analysis

Bibliometrics is the scientific field of study that quantitatively analyses all bibliographic data statistically and mathematically (Merigo et al, 2016). This method is often used to detect the growth of some particular field of study (Zeleznik, 2017 and Merigo, 2015). Bibliometric analysis of published journal articles, books, reviews, and conference proceedings have been carried out in different fields and topics. For example, a major Bibliometric study was done to analyze the research status, trends and hotspots in deep learning domain. The result revealed that the international academic communications in deep learning field are prosperingwhich are focused on three major regions: East Asia, North America and West Europe. The research hotspots included algorithm research and modeling (Mao,2018).

2.1.1 Applications of Bibliometric Analysis

Many bibliometric analyses have been done in recent studies. For instance, due to increase in the number of premature ejaculations over the years, bibliometric analysis was carried out to study global trend in the research results obtained in the field of premature ejaculation. The United State of America (USA) recorded the highest h-index, number of publications and also the frontline of research on premature ejaculation research (Hui, 2020).

Another bibliometric study was carried out to evaluate the characteristics of publications involving Middle East Respiratory Syndrome Coronavirus (MERS-COV) between 2012 and 2015 at global level using Scopus database.The USA was the largest contributor,with 319 articles published over 4 years,and also has the highest h-index of 32 superseded by the Kingdom of Saudi Arabia (KSA) with 113 articles and h-index of 36. Netherland produced the greatest proportions with International research collaboration with 72.7%.The MERS-COV associated publications were originated from 92 countries indicating the international spread of MERS-COV research.There is a rapid increase in research activities related to MERS-COV from 2012 to 2015 (Zyoud,2016).

Another study was carried out with the objective of investigating the author productivity and literature growth of block chain technology research from the period of 2008 to 2019. Results showed that the key author was Wang, F.Y with 38 different works by Scopus, the first publication with block chain technology was in 2010 and the peak time for block chain research was in 2019. China and USA are considered to be the most productive countries producing more research and Chinese organizations were leading in this new technology (Bukhari,2020).

2.1.2 Bibliometric indicators

Bibliometric indicators are becoming progressively significant because scientific findings are increasingly published in journals.These articles are being reviewed by professionals in the same field before they are published (Whitehouse, 2001). Their research findings are studied and cited by other scholars and it has been approved thatthe quantity of citations is an indication of its influence in scientific community (Rhen and Kronman, 2006).Bibliometric indicators have been of great help to organizations and researchers; they are being used in funding decisions, appointments, and promoting researchers; in detecting research accomplishment(Henderson and Breskey, 2003; Itagaki,2005); in analyzing research interests and predicting research resources that are wrongly allocated (Al-Shahi et al,2001);in identifying the trends of countries scientific influence and discovering geographical foundation of research (Rahman et al, 2005; Rahman and Fukui, 2002) and in enhancing research policies by surveying their research funding (Itagaki,2008; Mussurakis ,1994).

There are three types of bibliometric indicators (Lundberg,2006):
1) Structural indicators evaluate the link between authors, areas of research and publication
2)Quantity indicators assess the performance of a specific author. This can be achieved by counting the number of published articles by a certain author or research group during a particular period of time, but this does not guarantee a good quality article. It is important to note that the number of publications is affected by a group’s size. This challenge can be resolved by counting the number of publications in top – ranked or highest quality journals with regards to their impact factor (Lundberg,2006)

3) Quality indicators analyze the standard or performance of a journal or researcher's works.More advanced bibliometric indicators have been initiated in assessing and comparing the quality of various journals, researcher groups or individual researchers. This type of bibliometric indicator is divided into two parts: Journal performance indicator and Researcher's performance indicators.

Eugene Garfield,one of the founders of bibliometric and Scientometric and also the father of citation studies invented the first bibliometric index tools (Ronald et al,2018). The bibliometric indicators currently used in studies for journals or researcher's performance include:Citation count, h-index, and article number.

2.1.3 Bibliometric Data Sources

Bibliometric indicators are built on details collected from citation indexes or databases. There are various citation databases used in bibliometric analysis such as Web of Science, Scopus, Google Scholar,Crossref, Microsoft Academic,PubMed, Dimension and many more. Current research will briefly discuss about Web of Science because the data used was collected from this database.

- Web of Science

Web of Science was formerly known as Web of Knowledge, it is a multidisciplinary resource that provides an extensive citation data. Web of Science is presently handled by Clarivate Analytics which was formerly owned by Thomas Reuters (Analytics, 2017). Science Citation Index (SCI) is the most frequently used citation index and it was invented by the Institute for Science Information (ISI) in the year 1963 (Neuberger and Counsell,2002).Journal Citation Reports, an annual publication was initially published with Science Citation Index but it is currently assembled from Science Citation Index Expanded(SCIE) and Social Science Citation Index(SSCI) and is formulated on citations. Journal Citation Reports has been incorporated into Web of Science and can be obtained from Web of Science-Core Collections. It enables access to academic journals from social sciences and natural sciences, as well as their impact factors (Garfield, 2007). The Web of Science is a multidisciplinary resource that consists of 12,000 high impact journals and 160,000 conference proceedings,as of 24th February 2017 (Reuters, 2017). It covers

disciplines such as Humanities, Arts, Sciences and Social Sciences. Web of Science Core Collection consist of nine(9) online databases which include:

- Science Citation Index Expanded (1900 - present)
- Social Sciences Citation Index (1900 - present)
- Arts & Humanities Citation Index (1975 - present)
- Conference Proceedings Citation Index - Science ((1900 - present)
- Conference Proceedings Citation Index - Social Sciences & Humanities - (1900 - present)
- Book Citation Index - Science - (2005 - present)
- Book Citation Index - Social Sciences & Humanities - (2005 - present)
- Emerging Sources Citation Index - (2015 - present)
- Current Chemical Reactions - (1985 - present)

It develops the most consistent,accurate and complete citation network to power both trusted evaluation and confident identification.Every article and cited references from each journal in Web of Science have been indexed.Web of knowledge had limitations such as estimating the performance indicators for journals by using only the citations of articles accepted by Thompson Reuter, this makes their data incomplete. In addition, its matching process is poorly automated by the computer which leads to misspelling of names,missing word in titles and this leads to some lostdata in the total number of publishing journals (Moed,2002).

Web of Science Core Collection Citation Report

This provides researchers with citation analysis for a set of search results in an easy-to-interpret way. These aggregate citation statistics are created to display the citation performance and publication trends of institutions,authors,journals,countries or result set on a particular topic.Web of Science Citation report shows graphical display of the total publications and the total number of times all records have been cited (sum of times cited) per year. It also presents the data for total number of publications from the results found field,h-index count, the average number of times a record has been cited for all years in the result set(average citations per item), sum of times cited with and without self-citations.

- Citation count

This is the number of times a publication was cited within a specified time. Citation count does not take into account the articles published in journals that are not covered by Web of Science. Average citations per year is calculated by dividing number of citations by the number of years within a period of time while the average citations per item is calculated by the total number of citations by the number of articles. (Durieux and Gevenois, 2010).

- Hirsh (h) Index

Jorge E Hirsh proposed this index in the year 2005 for the purpose of measuring the scientific impact and the productivity of individual researchers through the number of their publications and the number of citations these publications have received. Web of Science measures the h-index of researchers by using their publications and citation details gotten from different journals they index.The h-index count is built on the list of publications ranked in descending order by the times cited count the higher the h-index, the better the researcher. For example, the h-indexvalue of 20 simply means that 20 articles published by a particular researcher or author have been cited at least 20 times or more. It is very vital to note that the h-index value can never be greater than the number of published papers by the author. H-index has been identified to be useful both in scientific community (Batista et al, 2006, Braun et al,2005) and Infometrics literature (Van Raan, 2006; Egghe and Rousseau, 2006; and Bornmann,2005).

2.2 Bayesian Network

The concept of “Bayesian Network” was invented by Judea Pearl in 1985 (Pearl J,1985) and its application is built on Bayes’ theorem, Reverend Thomas Bayes’ discovery (McGrayne et al). The brief statement about Bayesian network that was made by the original texts of Probabilistic Reasoning in Expert Systems and Probabilistic Reasoning in Intelligent Systems assisted in recognizing Bayesian network as a field of study (Pearl, 1988 and Neapolitan,1989).Bayesian Network, also known as belief network, Bayes network, decision network or probabilistic directed acyclic graphical model; is a type of probabilistic graphical model that gives an easy way of employing Bayes theorem to solve difficult problems in different disciplines(Jason,2019).This type of model is known as the directed acyclic graph (DAG) that is well known in three major fields namely machine

learning, statistics and artificial intelligence societies. They allow the actual representation of joint probability distribution (JPD) of random variables (Pearl, 1988). Bayesian Network is a machine learning algorithm that uses Bayes theorem in detecting and explaining classification (Friedman et al,1997).

It is employed in denoting and evaluating cases or relationships between variables that has to do with uncertainty or unreliability. The details of Bayesian Network are represented in an acyclic graphic which comprises of nodes which indicates the variables and the directed arcs or edges indicatean influential relationship (the probability distribution and a conditional dependence). Bayesian Network models show the conditional independent and dependent relationships between random variables. This simply means that the absence of arc between two nodes explains that the two variables related to these nodes are independent of each other whereas the presence of an arc shows the two variables are dependent of each other. Bayesian Models can be prepared in two ways: either by experts or by learning from data. After that, they are used for inference to determine the probabilities for any usual events. Bayesian Networks incorporate principles from statistics, probability theory, and computer science together when carrying out its tasks (Garbolino and Taroni, 2002, Jason, 2019 and Ben Gal, 2007). This type of probabilistic graphical method can be applied with small and missing data. It has proven to give an excellent predictive accuracy result (Kontkanen et al, 1997).

2.2.1 Applications of Bayesian Network

It has really helped the society at large both medically and in research, though it also has its limitations. Bayesian Networks have been attracting a lot of attention and many researchers have used this model to achieve great results in solving real world problems in different fields like agriculture (Bressen et al,2009), molecular biology (Friedman et al,2004) (Sach et al,2005), genomics (Jansen et al,2003) (Sebastiani et al,2003) (Friedman et al,2000), medicine (Husmeier et al,2005) (Diez et al,1997), biomedicine and health care (Peter J.F et al,2004) and Artificial Intelligence (Guftodimos et al,2004).They have also been used to solve many problems in environmental modeling and management (Uusitalo et al,2007), in describing and assessing scientific evidence (Garbolino and Taroni,2002) and also in test analysis (Dong and Agogino,1997).

Bayesian Network analysis can be used in detecting post-stroke results from diagnosed acute ischemic stroke using its biomarkers such as" D-dimer" and it also gives a visual graph structure that shows the conditional probability of results following stroke (Eunjeong et al, 2018).Bayes Network is beneficial in classifying ovarian tumors from clinical data when text-based resources are converted into informative text-based priors (Peter Antal et al, 2004).

The relationship between bibliometric indices in computer science and artificial intelligence journals were evaluated using Bayesian network model, thesemodels were also used for probabilistic reasoning and to discover the probabilistic conditional (in)dependencies between the bibliometric indices (Ibanez,2011). Application of Bayesian Networks in risk modeling has merits over regression-based approaches; it has proven to be a flexible and good tool for precision medicine in assessing health economics and research outcomes (Paul et al, 2019).

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This section explains the different methods used in this research which include the study design, data collection, methodology and software tools,and analysis of data.

3.2 Study Design

This study was planned to be acombination of two methods which are Bibliometric analysis and Visualization.

Bibliometric Analysis: In this method, the literature (journal articles) on Bayesian Networks were reviewed and analyzed within a time span, between the year 1990 and 2020.

Visualization: This method showed the graphical representation of information and data using images, tables, networks and maps. It provided the best way to see and comprehend the literature growth, patterns, and outliers in data.

3.3 Bibliometric Data Collections

The dataset for this study was extracted and downloaded on 4thAugust 2020 from Web of Science (Clarivate Analytics), the major electronic database to access journals on a specified topic. The Web of Science Core Collection is the resource that contains over 21,000 high quality peer reviewed journals published globally was used to achieve high quality journal results. An in-depth search for journals using the topic Bayesian Network and Belief Network(TS =(Bayesian network AND belief network) were used as the search terms in advanced search covering the period of 1990 to 2020 and a total of 1,739 publications met the selection criteria. The search results were checked manually in order to eliminate irrelevant articles that are not related to the field of study before analyzing the data.

For the purpose of filtering the data, selection criteria that were used for this study had the inclusion criteria and the exclusion criteria:

3.3.1 Inclusion criteria were as follows:

- Document type: only journal articles.
- Year published or time span: only the years between 1990 and 2020 will be included in the study.
- Languages: English only.
- Fields: All subjects.
- Only the related literature in Science Citation Index (SCI) and Science Citation Index Expanded (SCIE) database was used.

3.3.2 Exclusion Criteria:

- Books, reviews, conference proceedings and the rest were excluded from the study to limit the research.
- The articles published before the year 1990 were excluded from the study.
- Other citation indexes aside the above mentioned two were excluded.

3.4 METHODOLOGY AND SOFTWARE TOOLS

Co-authorship analysis, co-citationanalysis and keyword co-occurrence analysis are the bibliometric approaches that were assessed regarding our topic. Network analysis was used to reveal the intellectual structures of Countries, Institutions and journals related to Bayesian Network research.

3.4.1 Co-authorship Analysis

This method uses only author papers as signal for multidisciplinary activities within a field (Newman 2001, 2004).This occurs when one author works together with the lead author and contributes earnestly to the manuscript of the research work.Co-authorship analysis evaluates the level of collaboration within a research program (Moral et al, 2009),the map preference topic areas(Gonzalez-Alcaide et al, 2013), and the technological and scientific development (Vasconcellos and Morel, 2012).

Co-authorship networks and Social Network Analysis (SNS) are progressively used to evaluate connections trends and to determine leading scientists and organizations(Brunade Paula et al,2016). It is a concrete and reliable method used in comprehending and

analyzing scientificcollaboration patterns and the result collected from these association develop what is called a “Collaboration Network” (Newman, 2001) like the one represented in Figure 1.It comprises of a node and a tie or link;a node illustrates an author of one or more papers while the link or the number of co-authored papers(between authors) represents the collaboration strength. This simply means that greater number of papers indicates more powerful levels of greater connection.

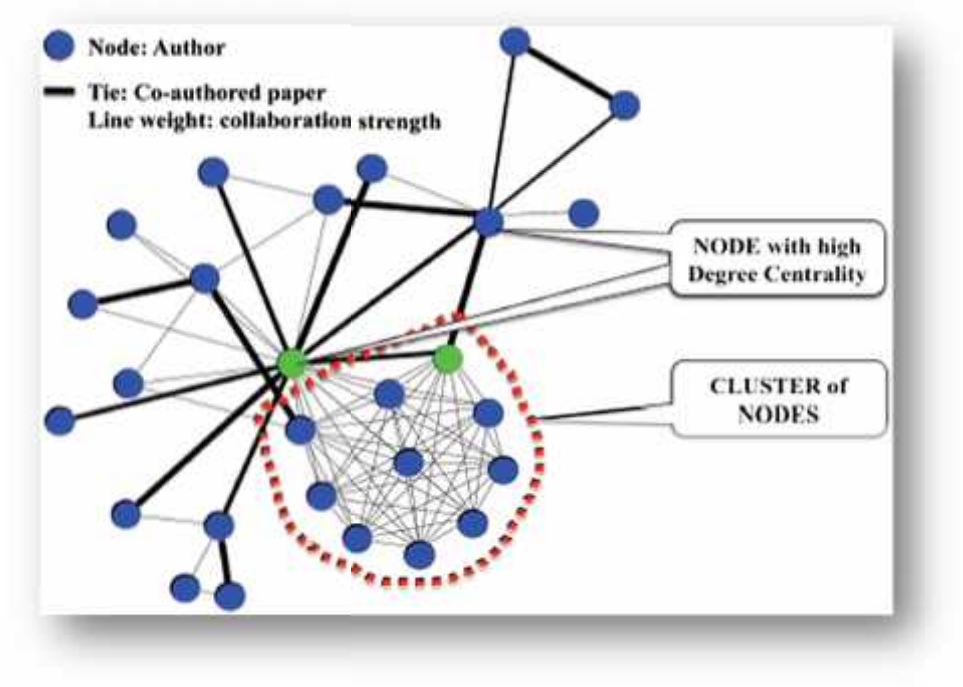

Figure 1. Co-authorship Network

A node with high centrality degree has a huge amount of links with other researchers whereas a cluster of nodes is a group of researchers that are interlinked and could show a semi-autonomous combination of a larger network (Reyes –Gonzalez, 2016). In co-authorship networks, nodes represent countries, authors, or organizations, which get linked when they share the authorship of a paper (Newman, 2004).

3.4.2 Co-citation Analysis

Co-citation analysis is a method that is used in signifying documents homogeneity between two documents. It evaluates the frequency with which two documents are cited simultaneously by other documents and they are said to be co-cited.. The more the two documents are cited by other documents, the stronger their co-citation strength and the

more probably they are semantically connected (Small, 1973; Gmür, 2003). For example, if document A and B, are co-cited by C, D and E, this means that A and B have a co-citation index or strength of three documents (Jeppe 2010; 2013).The co-citation score is known using citation indexes. Huge amount of co-cited documents is considered more related. Co-citation analysis examines the development and structure of scientific communities and areas. This type of methodology is built on the assumption that a citation is a justifiable and dependable indicator of scientific communication and evaluates the importance of articles or authors signals (Small, 1973, 1978).

3.4.3 Keyword Co-occurrence Analysis

Co-occurrence of keywords occurs when two or more keywords appear in the same article and it reveals the internal connection of knowledge between keywords in basic and technological research fields (Callon, 1991). It may also represent the evolution of multi-discipline research areas. To some extent high frequency keywords can study the research hotspots and research direction of some hot research topics. The greater the number of co-occurrence keyword, the stronger the connection between them which explains that the two keywords are linked to a particular research topic (Cambrosio, 1993). Node size represents h-index while the node color represents matrix-based clustering. H-index can extensively review the number of citations and the number of occurred keywords. The keywords are situated in the network center and have high degree centrality and betweenness centrality, which indicates that they are broadly applied in research.

3.4.4 Science Mapping

Science mapping is a vital quantitative method of bibliometrics used in visualizing the bibliographic data of scientific literatures. It is useful in creating large scale maps from large document spaces(Small,1997). This approach showed different fields situations and the development status related to Bayesian Networks. Itcan be applied in promoting the

organizational competitive intelligence(Alcara et al,2006) and communication management of networks for the health innovation system (Martins et al,2012), promoting the assessment of cross-disciplinary research programs(Yang and Heo,2014),building strategic public policy planning(Morel et al,2019) and in building up discovery management in public health systems (Vasconcellos and Morel, 2012).

Software Tools

The software tools that were selected for Science mapping are: VOS-viewer while Microsoft Excel was used for creating tables. VOS-viewer and Microsoft Excel was briefly explained below:

3.4.4.1 VOS-Viewer

This is a freely available computer software tool for evaluating and visualizing scientific literature used by industries, scientific publications, institutions, research funding agencies, organizations etc. VOS-viewer software was created by Van Eck N.J and Waltman L., it has been efficiently used in many projects done by the Center for Science and Technology Studies and it uses VOS mapping technique, where VOS symbolizes Visualization of Similarities.It gives the interactive visual exploration of bibliometric studies. VOS-viewer offers a distance-based visualization of bibliometric networks in an easy way for comprehension and interpretation, its features such as searching, zooming and scrolling enables the detailed examination of a map. Because of its main focus on visualization it is considered more suitable for visualizing larger networks from large Scientometric datasets.

It comprises of circles and lines where the circles denote words(authors, countries, journals, organizations) in the study while the line represents the existing relation between them.VOS-viewer does not display the edges between nodes by default in a bibliometric network; rather it displays only the nodes. The distance between two nodes approximately denotes the connection of the nodes. VOS-viewer also has a special feature of text mining but gives a less capability in analyzing bibliometric networks than other software tools (Van Eck and Waltman,2010,2014).

3.4.4.2 Microsoft Excel

This is a software program developed by Microsoft that enables users to arrange, format and calculate data with formulas utilizing a spreadsheet system. Microsoft Excel is used for data visualization; it shows data as histograms,line graphs and pie charts.

3.5 Analysis of Data

Web of Science was used to analyze the institutions, authors, countries and Journals related to Bayesian Networks. This is the development growth in the use of the method. VOS-Viewer was used to create the co-authorship, co-citation and keyword co-occurrence networks. The retrieved 1,739 scientific publications from Web of Science were entered into the software tools to construct the network visualization and metric calculations showing the total number of documents, citations and total link strength between authors,organizations,countries,keyword co-occurrence,cited reference,cited sources and cited authors. Microsoft Excel was used for displaying the downloaded text document files into tables for easy understanding and interpretation.

CHAPTER FOUR

RESULTS

This chapter presents the detailed results of this study carried out on the bibliometric data gotten on Bayesian Network. These data together with the analyzed results were extracted on 4th of August, 2020 from the Web of Science during the period of 1990to 2020.This part of the study shows the results in four sections. Section 4.1 presents the results of Web of

Science analyzed results. The co-authorship analysis and the co- citation analysis are presented in Section 4.2 and Section 4.3 respectively. Lastly, Section 4.4 introduces the keyword co-occurrence analysis on Bayesian Network.

4.1. The Current Status of Bayesian Network Research Study

This section of the study presents the Web of Science Citation Report including the h-index, the total number of document types retrieved from the search, annual publications and citations trends of Bayesian Network research, the distribution of Bayesian Network research publications in the view of journals (Source titles),countries/regions, organizations(institutes) , and research areas.

4.1.1. The Web of Science Citation Report and H-index

The data in Table 1 gives the summary of the Web of Science Core Collection Citation Report and h-index results.A total number of 1,739 articles were found from the query made during data collection from Web of Science and h-index of 83 indicated that 83 published papers on Bayesian Network research have been cited by other papers at least 83times.The average number a publication has been cited is 26.55,the total number of times all publications have been cited with and without self citations are 46,167 and 41,481 respectively while the total number of citations to any of the articles in the search result set with and without self citations are 34,324 and 33,116 respectively.

Table 1: *Web of Science Core Collection Citation Report and H-index Results*

Web of Science Citation Report Indicators	Number of Publications
Total Publications	1,739

H-index	83
Average Citations Per Items	26.55
Sum of times Cited	46,167
Without Self Citations	41,481
Citing Articles	34,324
Without Self Citations	33,116

4.1.2. Total Number of Document Types Retrieved from the Search.

The total number of document types retrieved from the search is revealed in Figure 2a. It shows a total number of 1737 articles, 170 proceeding papers and 8 early access. Figure 2b displays the percentage of 1739 publications of each retrieved document types.

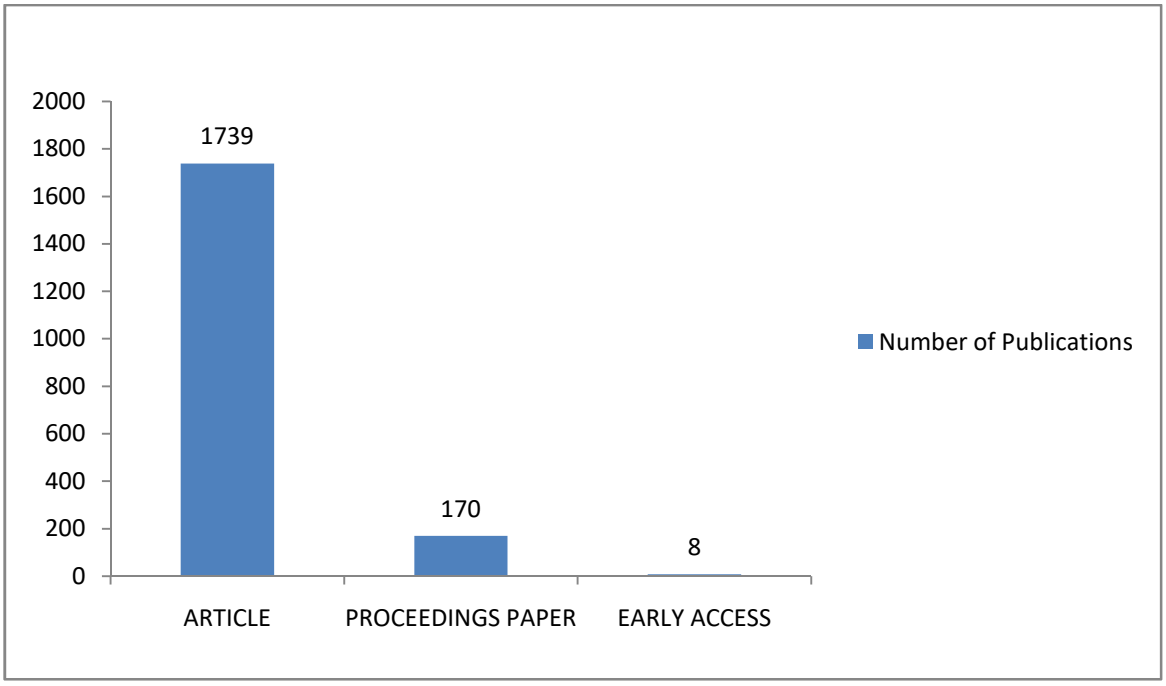

Figure 2a. Number of Publications Retrieved Document Type

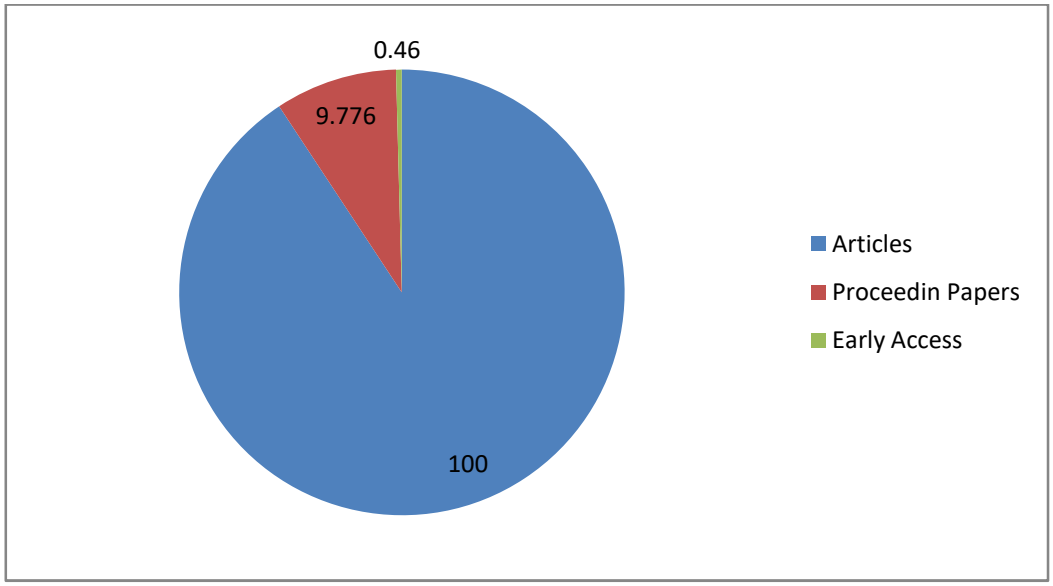

Figure 2b.Percentage of the total publications of Retrieved Document Types.

4.1.3. Annual Publication trends of Bayesian Network Research

Figure 3 plots the annual trends of Bayesian Network related publications from the period of 1990 to 2020 showing both the percentage and number of the total publications. The total number of 3 articles related to Bayesian Network Research was published in the year 1990 and a fluctuating increase in the number of publications in each year was observed. The year 2019 shows the peak (highest number) of the publications with 137 articles while the year 2020 as at 4thAugust gave 87 articles.

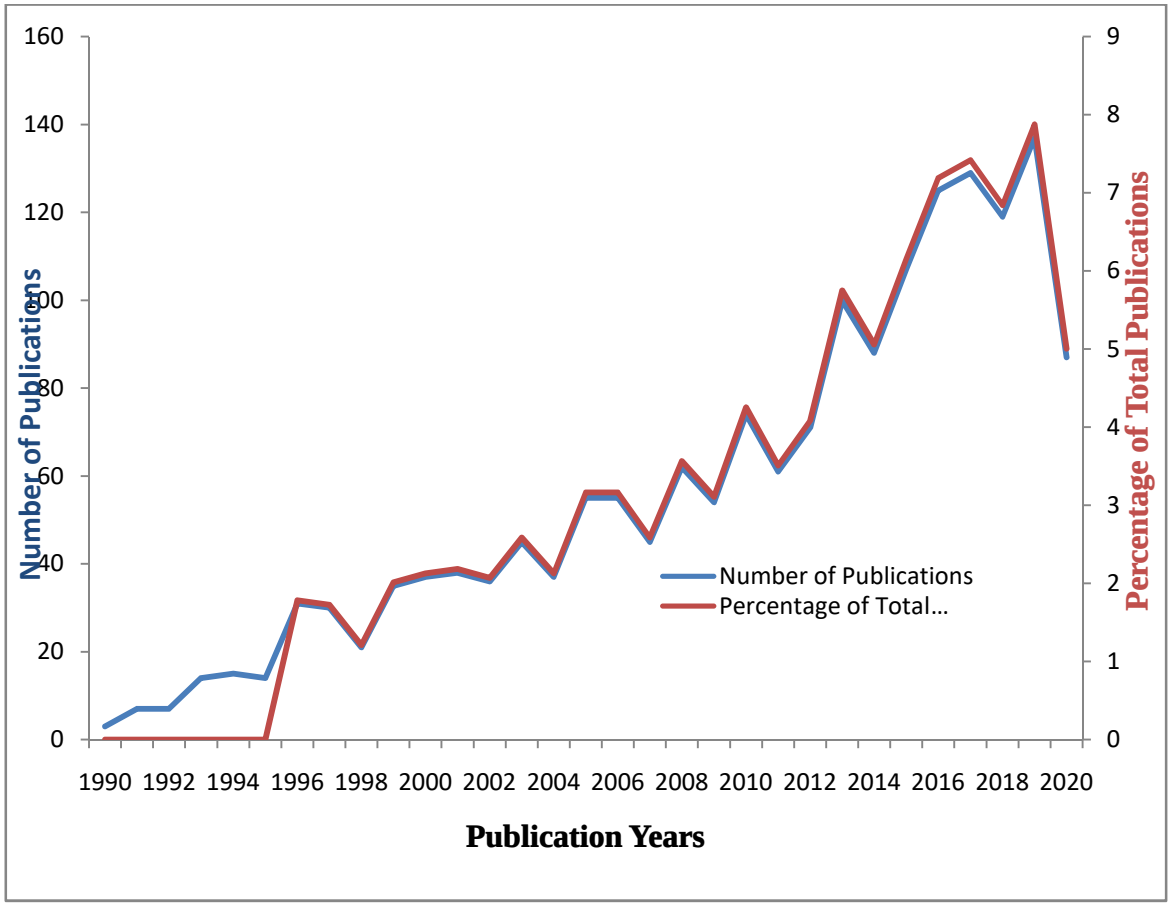

Figure 3. The annual Publication trends of Bayesian Network related publications from the period of 1990 to 2020.

4.1.4. Annual Citation trends of Bayesian Network Research

The bar graph (figure 4) below shows the yearly changes in the sum of cited publications related to Bayesian Network. It revealed a steady and rapid increase in the number of citations per year from 1990 to 2020. The year 1990 had a sum of citation of 3 which increased to 5302 citations in the year 2019 which happened to be the highest number of citation. The year 2020 recorded 3,101 number of citation as at 4th of August.

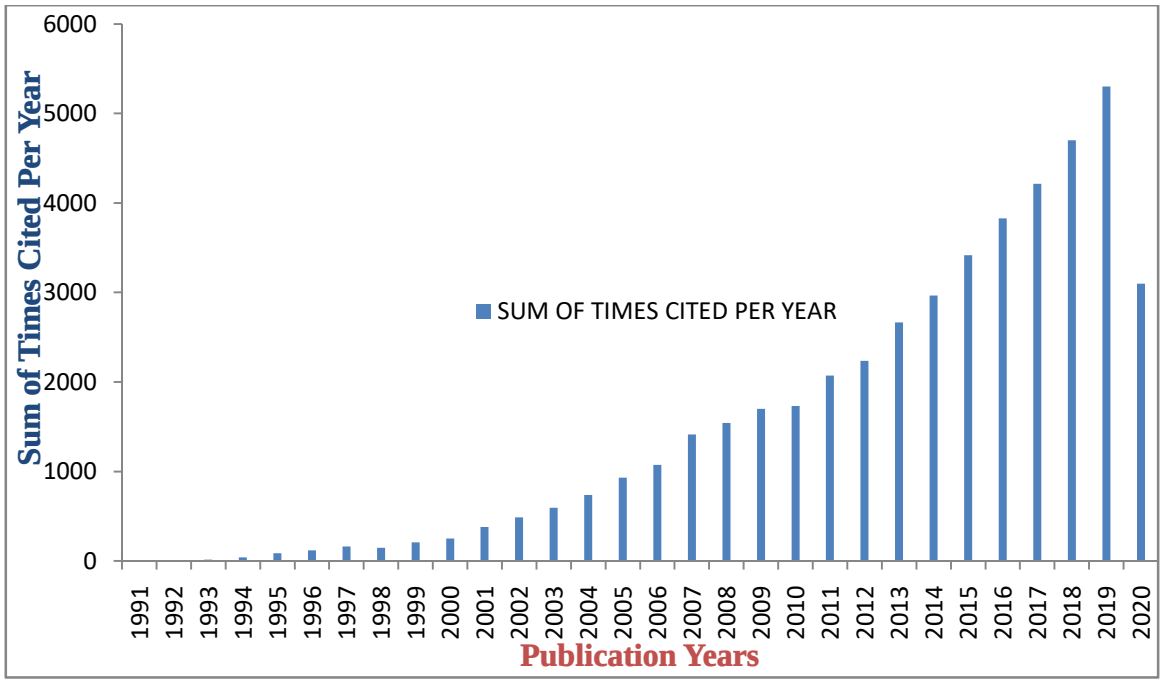

Figure 4.Annual Citation trends of Bayesian Network Research from the year 1990 to 2020.

4.1.5. The Distribution of Journals on Bayesian Network Research

The 1739total publications were published in 715 journals and only 25 journals with the highest number of publications(more productive journals) were used in the analyzes, which produced a total number of 518 publications (29.79%).The figure 5shows the top 10 journals that published the most Bayesian Network articles which gave a total number of 334 publications (19.08%). The Environmental Modelling &Software had the highest number of publications with a total of 53 publications (3.05%), followed by International Journal of Approximate Reasoning with 47publications (2.70%) and then Expert Systems with Applications with 39publications (2.24%).

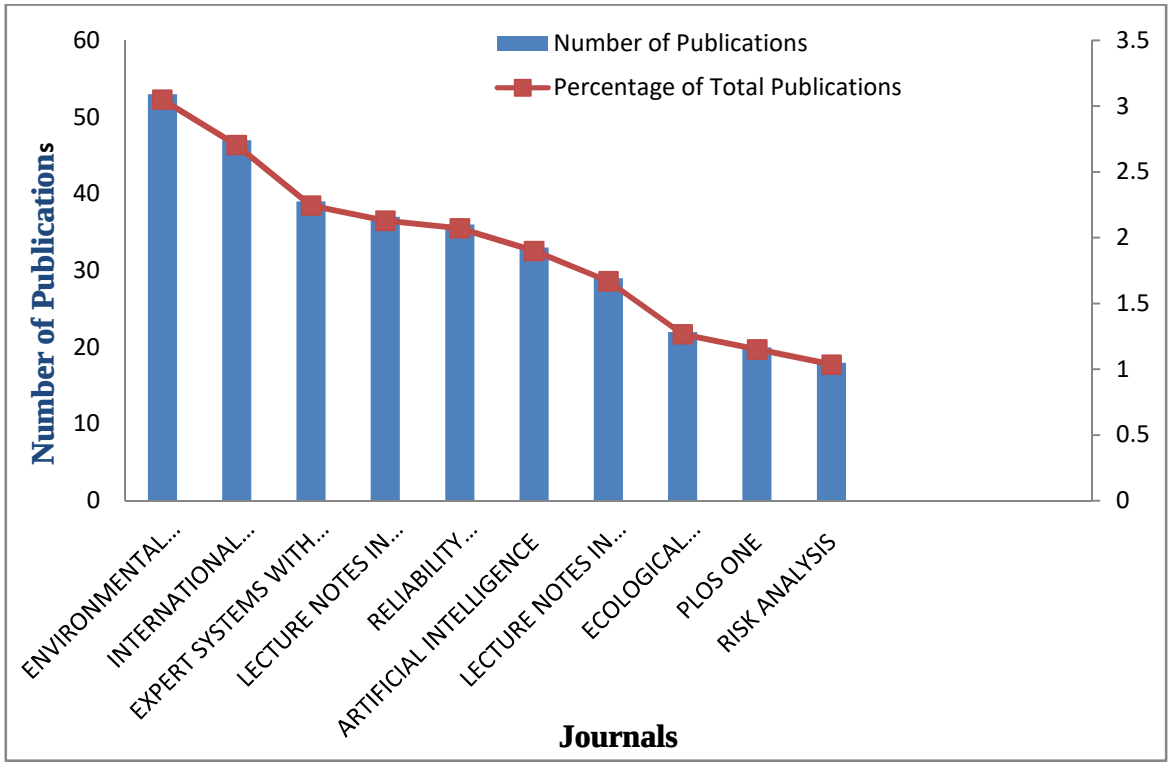

Figure 5.The top 10 journals with the most Bayesian Network publications.

4.1.6. The Distribution of Institutes on Bayesian Network Research

The top 25 journals accounted for 456 publications which gave 26.34% of all 1739 Bayesian Network publications. Figure 6 shows the top 10 Institutes with the most publications related to Bayesian Network research. University of Arizona has the highest number of publications with a total of 33 publications which explains 12.88% of all the publications in this study followed by University of Queensland with 33 publications (1.90%), Massachusetts Institute of Technology (MIT) and University of British Columbia have the same number of 23 publications and 1.32% of all publications. There are 3 Institutes from Australia,2 Institutes from USA and one Institute each from Canada, Italy, Spain, China,and Finland in the top 10 institutes.

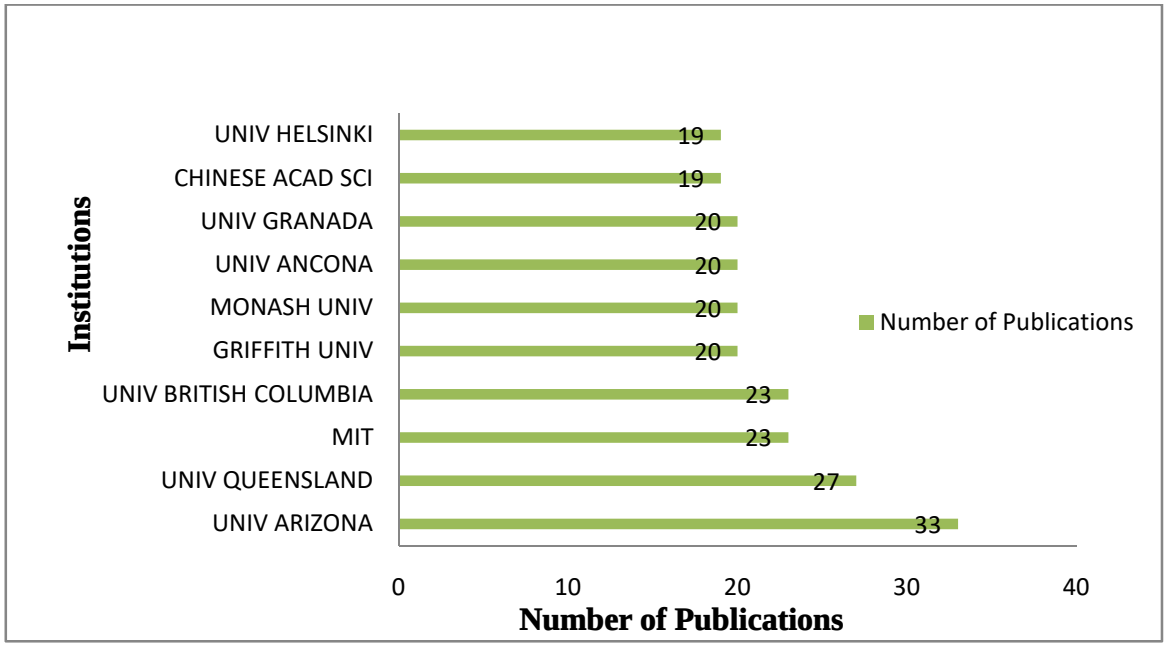

Figure 6. The top 10 Institutes with the most publications related to Bayesian Network research

4.1.7. The Distribution of Countries on Bayesian Network Research publications

The top 25 countries were used in the analysis and they produced a total of 2154 articles (123.87% of 1739) while the top 10 countries shown in Figure 7 recorded a total of 1668 articles(95.92% of 1739) on Bayesian Network research publications. USA is the most productive country with 551 articles(31.69% of 1739) on Bayesian Network related publications preceded by England and Peoples Republic of China with 206 articles(11.85%) each.

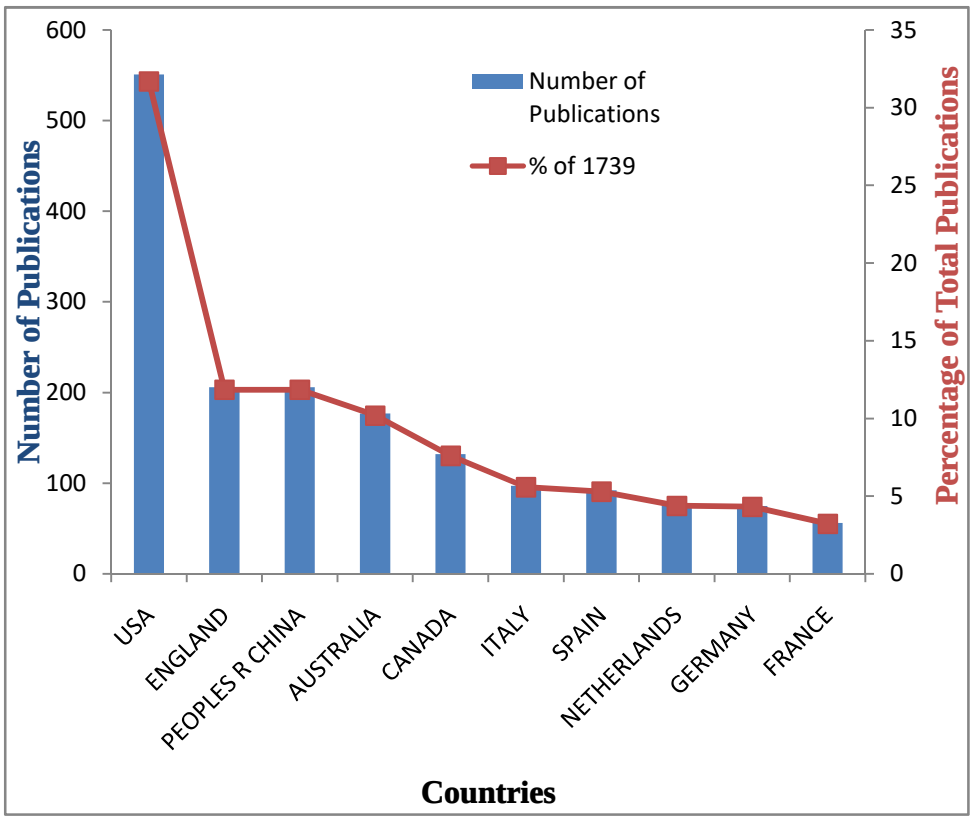

Figure 7. The top 10 Countries with the most publications related to Bayesian Network research.

4.1.7. The Distribution of Research Areas on Bayesian Network Research publications

The top 25 Research areas used in analyses accounted for 148.77% of 1739 publications(2587 articles) while the top 10 Research Areas gave 124.38% of 1739 publications(2163 articles) as seen in Figure 8.The Research Area with the highest number of publication is Computer Science with 721 publications (41.64%) preceded by Engineering having 575 articles (33.07%) and Environmental Sciences Ecology with 322 articles (18.52%).

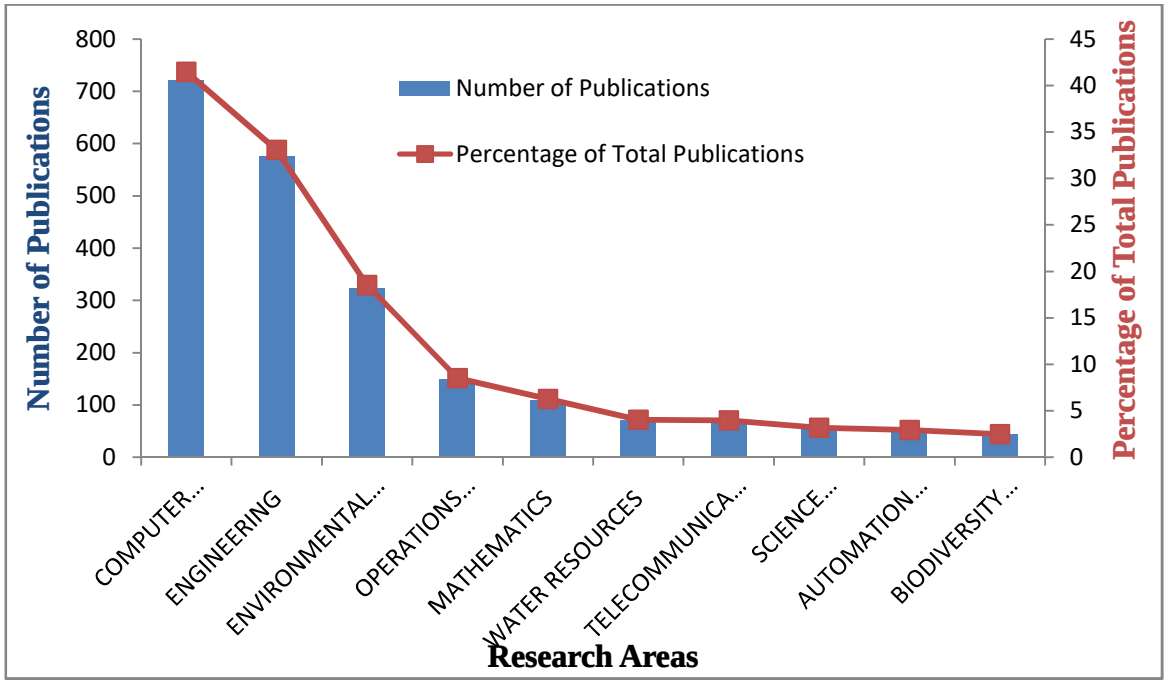

Figure8. Top 10 Research Areas Pertaining to Bayesian Network

4.1.8. The Distribution of Authors on Bayesian Network Research publications

The data in table 2 shows the top ten Authors with Bayesian Network related publications that covered 10.01% of 1739 publications which indicates 174 articles. Bartels P.H is the greatest author with 28articles followed by Thompson D. (26 articles) and Montironi R. (22 articles).

Table 2: Top Ten Authors with Bayesian Networks Related Publications

Authors	Number of Publications	Percentage(%) of 1739
BARTELS PH	28	1.61
THOMPSON D	26	1.495
MONTIRONI R	22	1.265
HAMILTON PW	18	1.035
KUIKKA S	16	0.92
TESFAMARIAM S	16	0.92
DE CAMPOS LM	13	0.748
MORAL S	13	0.748
MARCOT BG	12	0.69
FENTON N	10	0.575
Total	174	10.006

4.1.9. The Distribution of Web of Science Categories on Bayesian Network Research publications

Table 3 shows the top ten Web of Science Categories of Bayesian Network related articles where Computer Science Artificial Intelligence has the greatest number of 415 articles. The second position is Environmental Sciences with 243 articles and Engineering Electrical Electronic with 230 articles.

Table 3: *Top Ten Web of Science CategoriesRelated to Bayesian Network Research*

Web of Science Categories	Number of Publications	Percentage (%) of Total Publications
COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE	415	23.864
ENVIRONMENTAL SCIENCES	243	13.974
ENGINEERING ELECTRICAL ELECTRONIC	230	13.226
OPERATIONS RESEARCH MANAGEMENT SCIENCE	148	8.511
COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS	141	8.108
COMPUTER SCIENCE INFORMATION SYSTEMS	140	8.051
ECOLOGY	114	6.555
COMPUTER SCIENCE THEORY METHODS	113	6.498
ENGINEERING INDUSTRIAL	92	5.29
ENGINEERING ENVIRONMENTAL	90	5.175
Total	1726	99.252

4.1.10. The Distribution of Funding Agencies on Bayesian Network Research publications

The data in table 4 reveals that all the top ten funding agencies have similar number of publications pertaining to Bayesian Network but different values in their percentage. The National Natural Science Foundation of China is the greatest agency that funded the Bayesian Network Research and produced 126 articles (7.25% of 1739). The second funding agency is United States Department of Health Human Services with 126 articles (4.49% of 1739), and the third is National Institutes of Health NIH,USA having 216articles (4.31% of 1739).The top ten funding agencies published a total of 1260 articles explaining the 30.08% of 1739 total publications.

Table 4: *Top Ten Funding Agencies Related to Bayesian Network Research*

Funding Agencies	Number of Publications	Percentage (%) of Total Publications
NATIONAL NATURAL SCIENCE FOUNDATION OF CHINA	126	7.246
UNITED STATES DEPARTMENT OF HEALTH HUMAN SERVICES	126	4.485
NATIONAL INSTITUTES OF HEALTH NIH USA	126	4.313
NATIONAL SCIENCE FOUNDATION NSF	126	3.565
EUROPEAN UNION EU	126	2.473
NIH NATIONAL CANCER INSTITUTE NCI	126	2.128
NATURAL SCIENCES AND ENGINEERING RESEARCH COUNCIL OF CANADA	126	1.84
ENGINEERING PHYSICAL SCIENCES RESEARCH COUNCIL EPSRC	126	1.495
AUSTRALIAN RESEARCH COUNCIL	126	1.265
UNITED STATES DEPARTMENT OF DEFENSE	126	1.265
TOTAL	1260	30.075

4.2. The Co-Authorship Analysis on Bayesian Network Publications

This section presents the authors, organizations (institutions) and countries co-authorship network analysis.The co-authorship networks were created using the VOS-viewer software. The most productive or influential authors, organizations and countries having the greatest co-authorship connections with others are represented with big nodes including clusters which represent set of items(authors, counties and institutions) included in a

network map, their links describes the number of co-authorship links of a given author with other authors, the link strength explains the number of publications two items that have co-authored and their total link strength explains the total strength of the co-authorship links of a particular item with other items.The citations feature in co-authorship explains the total number of citations received by all articles published by a country, institution or author. The different colors in the map indicate variety of research fields. The links and the distance between two nodes indicate their level of co-operation and relationship.

4.2.1. Authors Co-authorship Network Analysis

The minimum number of documents of an author is 3 and out of 5045 authors only 222 authors met the thresholds. The most productive authors having the greatest connections are showed in figure 9 and it consists of13 Authors, 4 clusters, 50 links and 61.00total link strength. Bartels P.H has the greatest total link strength of 27.00,12 links,27 documents and 460 citations followed by Thompson, D. with 26.00 total link strength, 12 links, 26 documents and 455 citations and Hamilton, P.W.with19.00 total link strength,11 links,20 documents and 29citations.The link strength between Bartels, P.H. and Thompson,D. is 9.33,the link strength between Bartels, P.H. and Montironi,R. is 4.83.

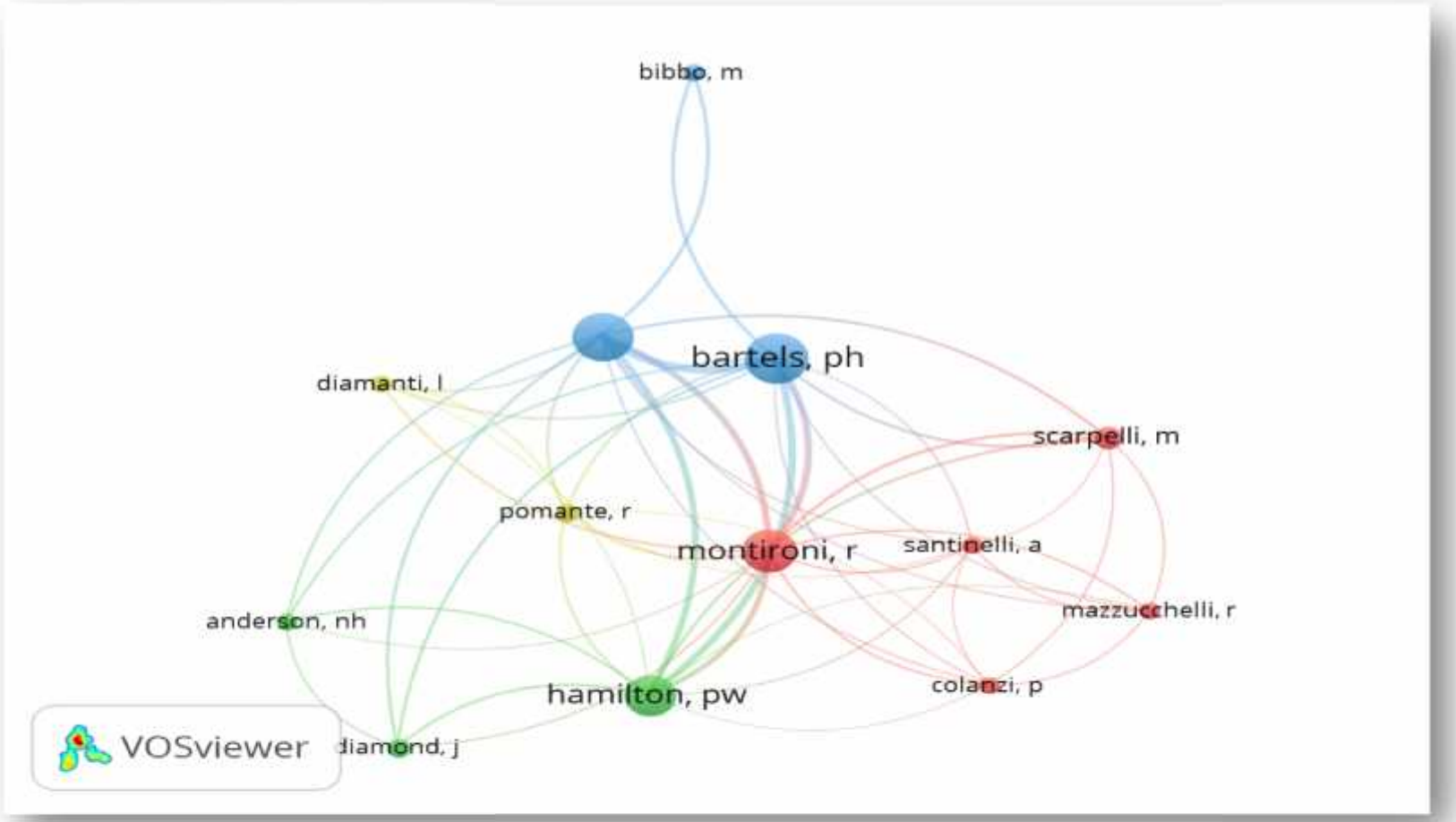

Figure 9. Authors Co-authorship Network Visualization of Bayesian Network Publications.

4.2.2. Countries Co-authorship Network Analysis

United States of America (USA) is the most productive country with 147.00 total link strength,38links,351 documents,15,037 citations preceded by England with 76.00 total link strength,135 documents,4187 citations and thirdly by Peoples Republic of China with 75.00 total link strength,25links,176 documents and 1869 citations. The link strength between USA and England 10.19 while between USA and Peoples Republic of China the link strength is 24.14.The minimum number of documents per country is 5 and out of 88 countries, 44 met the threshold. Figure 10 shows the overlay visualization of countries and their average citations in the task bar. It consists 44 countries having 9 clusters, 321 links and 432.50 total link strength.

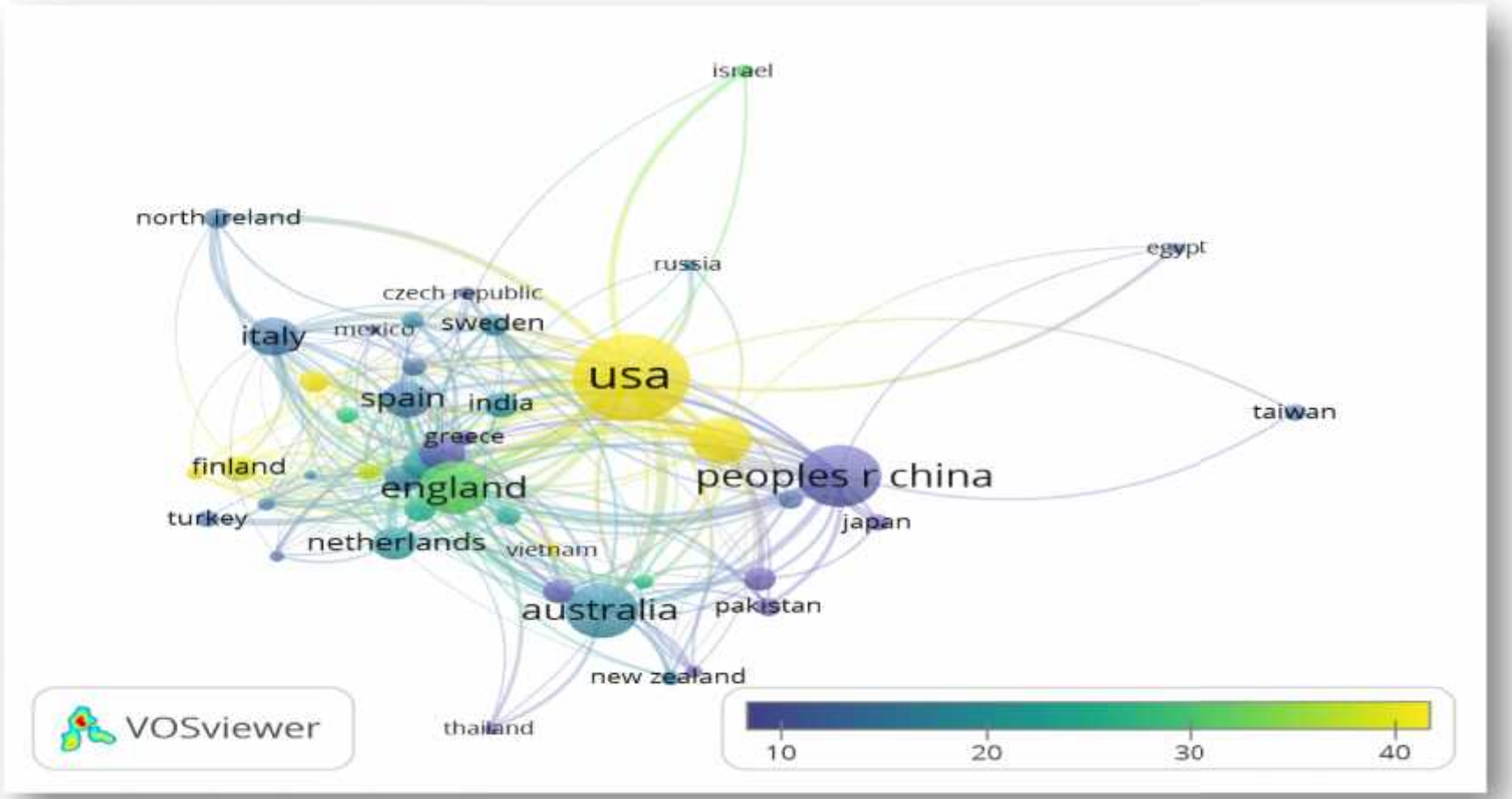

Figure 10: Countries Co-authorship Network Visualization of Bayesian Network Publications.

4.2.3. Organizations (Institutions) Co-authorship Network Analysis

Chinese Academy of Sciences (7links)and University of Arizona has the greatest total link strength of 12.00, University of Ancona and University of Queensland (8links) has a total link strength of 11.00 followed by Griffith University with 10.00 total link strength and 6 links .The minimum number of documents of an institution is 5 and out of 1462 institutions or only 113 institutions met the thresholds. The most productive institution having the greatest connections are showed in Figure 11 and it consists of 90 institutions, 16 clusters, 162 links and 167.50totallink strength.

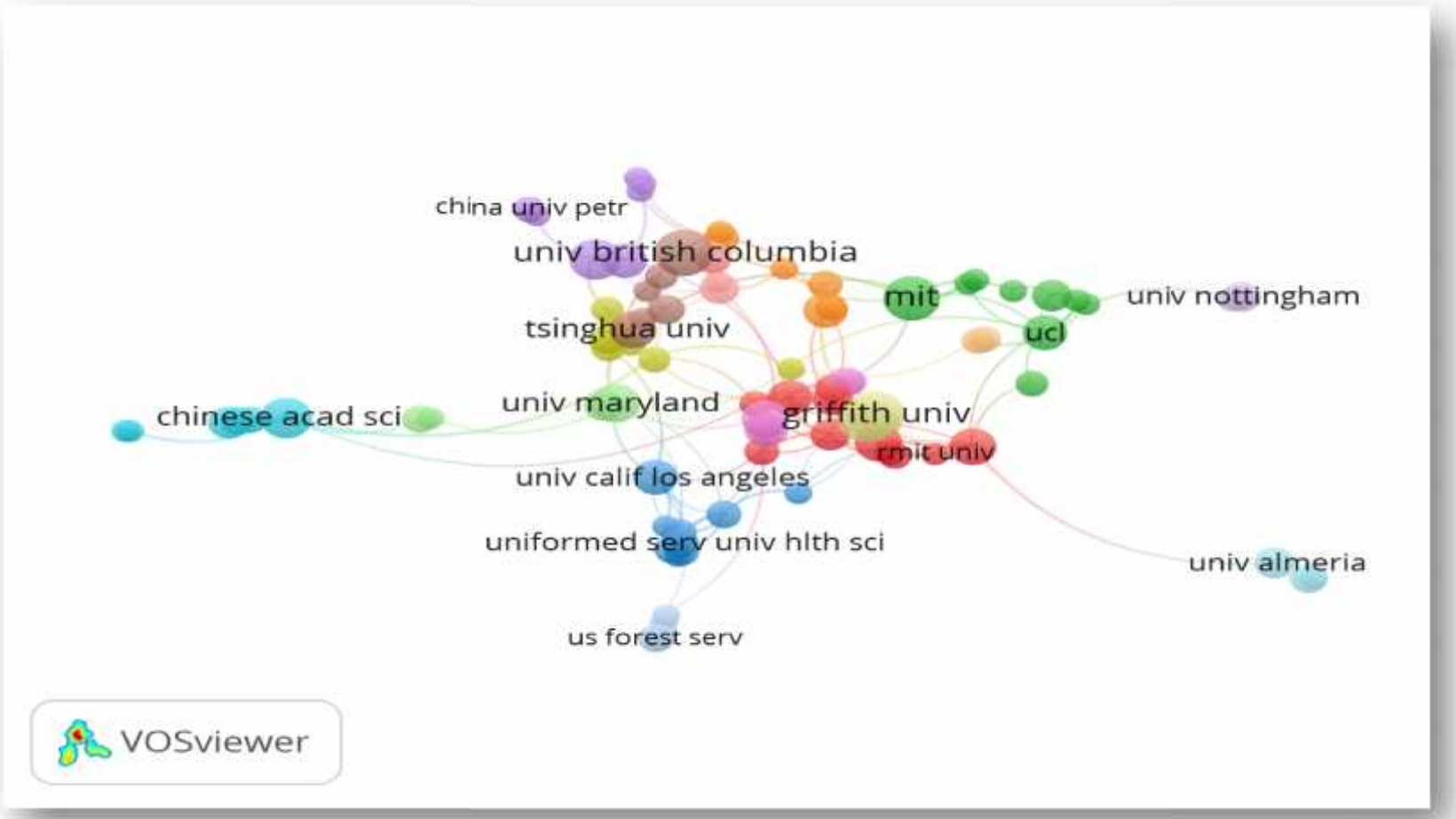

Figure 11: Institutions Co-authorship Network Visualization of Bayesian Network Publications.

4.2. The Co-Citation Analysis on Bayesian Network Publications

The number of citations made to cited authors, cited sources and cited references are shown in this part of the study.The co-citation networks were created using the VOS-viewer software.

4.2.1. Authors Co-Citation Network Analysis

Figure 12 shows the authors co-citation network analysis of Bayesian Network publications which includes 52 Authors,4 clusters, 982 links, and 2436.12 total link strength. The minimum number of citations of an author was increased from 20 to 50 and out of 29728 authors, 52 met the threshold. Table 5displays the most cited top 10 author.Pearl,J is the most cited author followed by Marcot, B.G. and Lauritzen, S.L.

Table 5

Top 10 most cited Authors

Authors	Number of Citations	Total link strength
Pearl,J	717	574.11
Marcot,B.G	354	297.43
Lauritzen,S.L	235	213.07
Jensen,F.V	224	209.66
Cooper,G.F	223	207.41
Heckerman,D	196	175.47
Uusitalo,I	164	158.3
Pollino, C.A	135	128.86
Aguilera,P.A	123	119
Borsuk,M. E	112	106.14
Total	2483	2189.45

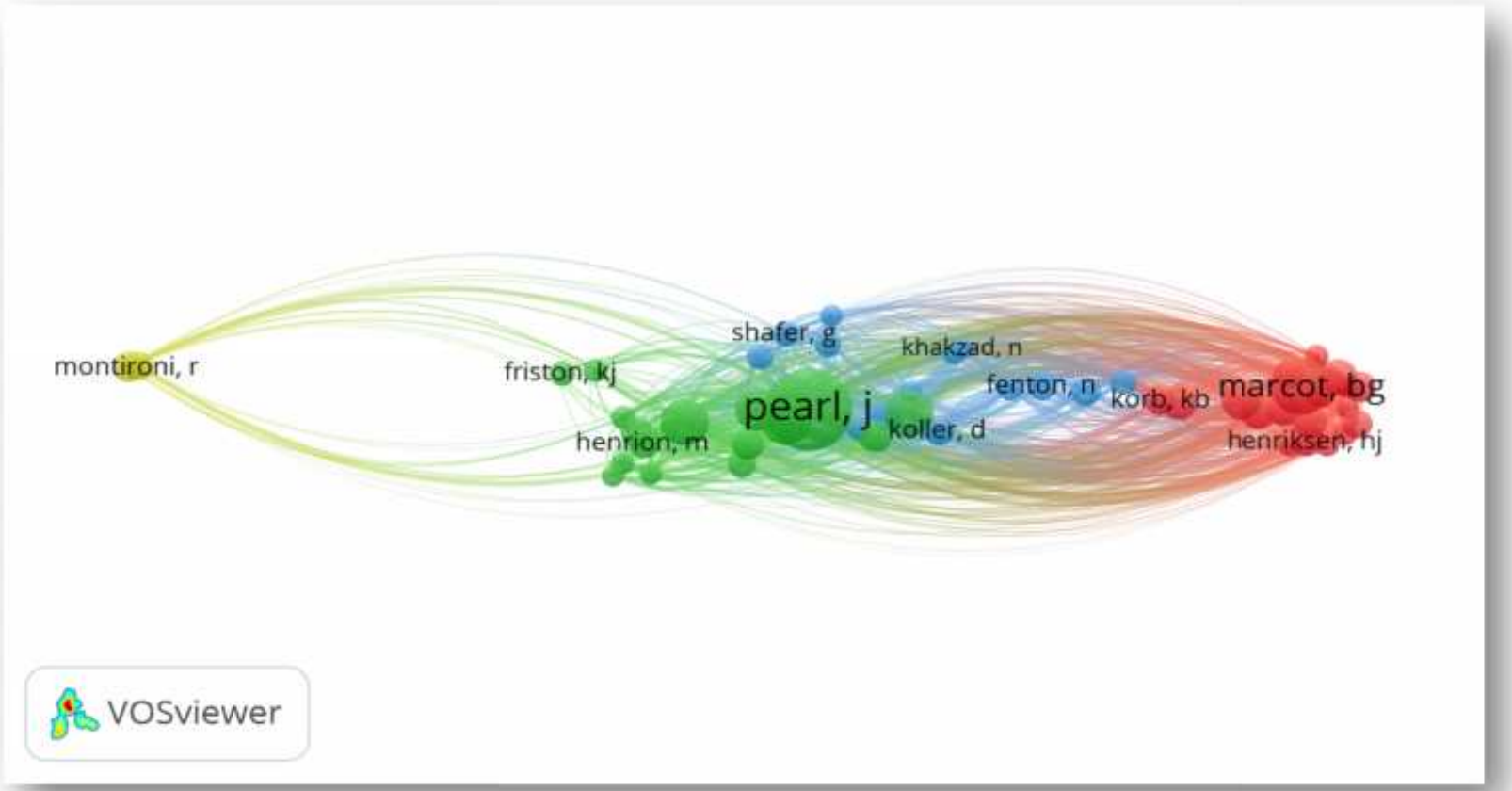

Figure 12: Author Co-Citation Network Visualization of Bayesian Network Publications.

4.2.2. The Journal (Source) Co-Citation Network Analysis

The minimum number of citations of a source is 20 at default but in this study it was increased to 50 and out of 17630 journals, 148 met the threshold. Figure 13 shows the 148 journals,4 clusters,7551 links and 8765.50 total link strength. The journal with the highest total link strength and citation is Environmental Modelling & Software followed by Ecological modeling and Reliability Engineering and system safety. The top 10 most cited journals can be seen in Table 6.

Table 6: Top 10 Most Cited Journal Pertaining To Bayesian Network Research

Journals(Sources)	Number Of Citations	Total Link Strength
Environmental Modelling & Software	1440	1112.49
Ecological Modelling	656	597.57
Reliability Engineering And System Safety	753	520.18
Artificial Intelligence	650	509.9
Probabilistic Reason	454	440.62
Canada Journal Of Forest Research	380	363.34
Science	352	333.49
Expert System With Applications	383	329.06
Journal Of Environmental Management	311	297.96
Risk Analysis	297	270.44
Total	5676	4775.05

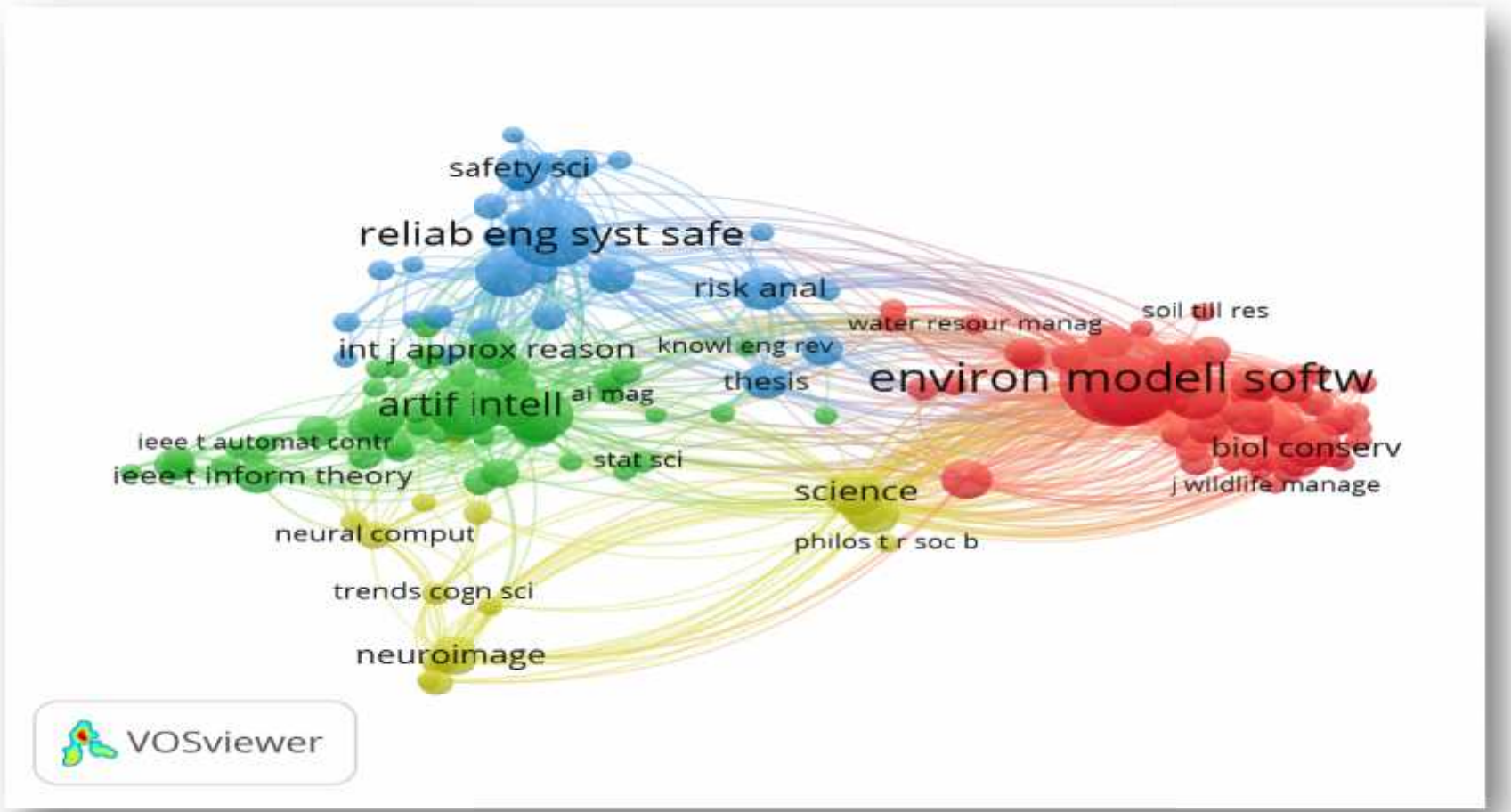

Figure 13: Journal Co-Citation Network Visualization of Bayesian Network Publications.

4.2.3. The References (Articles) Co-Citation Network Analysis

The minimum number of citation of a cited reference was 20 and out of 43727 cited reference only 77 met the threshold. The reference co-authorship network of Bayesian Network Research is showed in Figure14.It consists of 77 articles,5 clusters,1877 links and 1805.50 total link strength. The most cited reference with the biggest node is Pearl Judea 1988, the title of his article is Probabilistic Reasoning in Intelligent Systems: Networks of plausible inference followed by Marcot Bruce G.and the title of his paper isGuidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation.Table 7lists the top 10 most co-cited articles related to Bayesian Network Research.

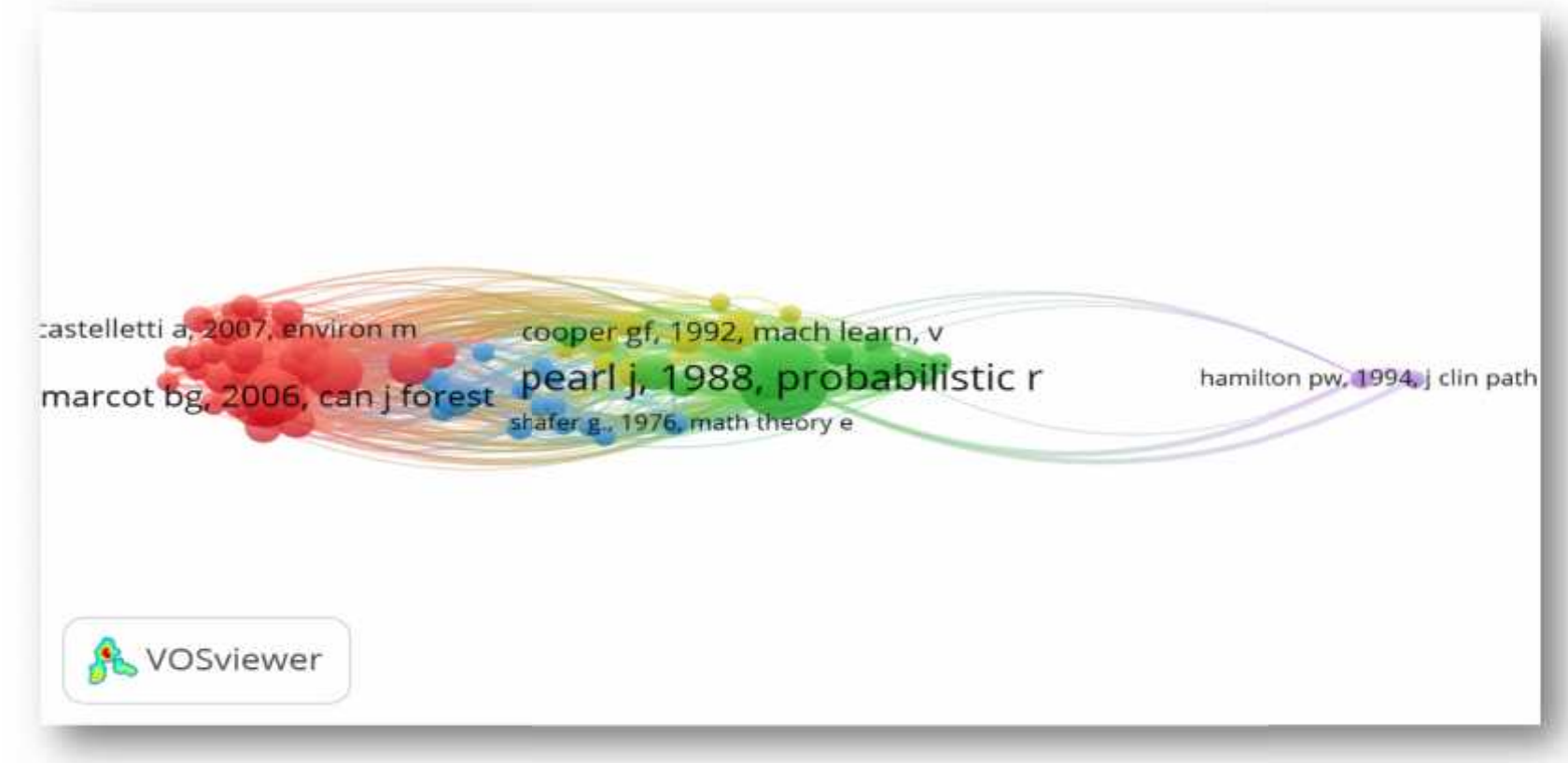

Figure 14: Reference Co-Citation Network Visualization of Bayesian Network Publications.

Table 7: Top 10 Most Cited Articles Related To Bayesian Network Research

Position in the citation ranking	Title of the article	Authors	Year of Publication	Journal	Number of Citations	Total Link Strength
1	Probabilistic Reasoning in Intelligent Systems: Networks of plausible inference	Pearl Judea	1988	Probabilistic Reason(Morgan Kaufmann Publisher)	334	304.00
2	Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation	Marcot Bruce G. , J Douglas Steventon, , Glenn D Sutherland, and , Robert K McCann	2006	The Canadian Journal of Forest Research	172	171.00
3	Advantages and challenges of Bayesian Networks in Environmental modelling	Uusitalo Laura	2007	Ecological Modelling	145	142.00
4	Local Computations with Probabilities on Graphical Structures and their applications to expert systems	Lauritzen S.L., Spiegelhalter, D.J	1988	Royal Statistical Society	135	130.00

5	Bayesian Networks in Environmental modelling	Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R., &Salmerón, A.	2011	Environmental Modelling & Software	102	101.00
6	Bayesian belief networks: Applications in ecology and natural resource management	McCann Robert K., Bruce G. Marcot and Rick Ellis	2006	Canadian Journal of Forest Research	101	99.00
7	Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment	Pollino Carmel A.,Owen Woodberry, Ann Nicholson, Kevin Korb Barry Hart.T.	2007	Environmental Modelling &Software	91	91.00
8	The computational complexity of probabilistic inference using Bayesian Belief Networks	Cooper G.F	1990	Artificial Intelligence	91	88.00
9	Bayesian Artificial Intelligence	KorbKevin .B& Nicholson,A.	2010	CRC Computer Science &Data Analysis	91	82.00
10	Bayesian Networks and Decision Graphs	Jensen Finn Verner &Thomas Nielson D.	2007	Springer	84	81.00

4.4.The keyword Co-occurrence Network Analysis

The minimum of occurrences of author keyword of 5 set at default in VOS-viewer software was used and a total of 3692 keywords were gotten but 108 met the threshold. The total result of 108 keywords,15 clusters,529links,and 521 total link strength gotten is shown with two different visualizations in figure 15 and figure 16. The most occurred keyword is Bayesian networks with 162 occurrences and total link strength of 105 followed by Bayesian belief network with 151 occurrence and total strength of 67.The Table 8 below shows that Bayesian network and Bayesian belief network appeared twice each of the first four top rows which indicate more occurrences of these two keywords.

Table 8: Top 10 Most Occurred Keywords in Bayesian Network Publications

Rank Number	Keyword	Occurrence	Total link strength
1	Bayesian networks	162	105
2	Bayesian belief network	151	67
3	Bayesian belief networks	105	60
4	Bayesian network	114	59
5	Uncertainty	46	40
6	Belief networks	33	27
7	Risk assessment	31	24
8	Ecosystem services	25	21
9	Risk analysis	25	21
10	Probabilistic reasoning	20	20
	Total	712	444

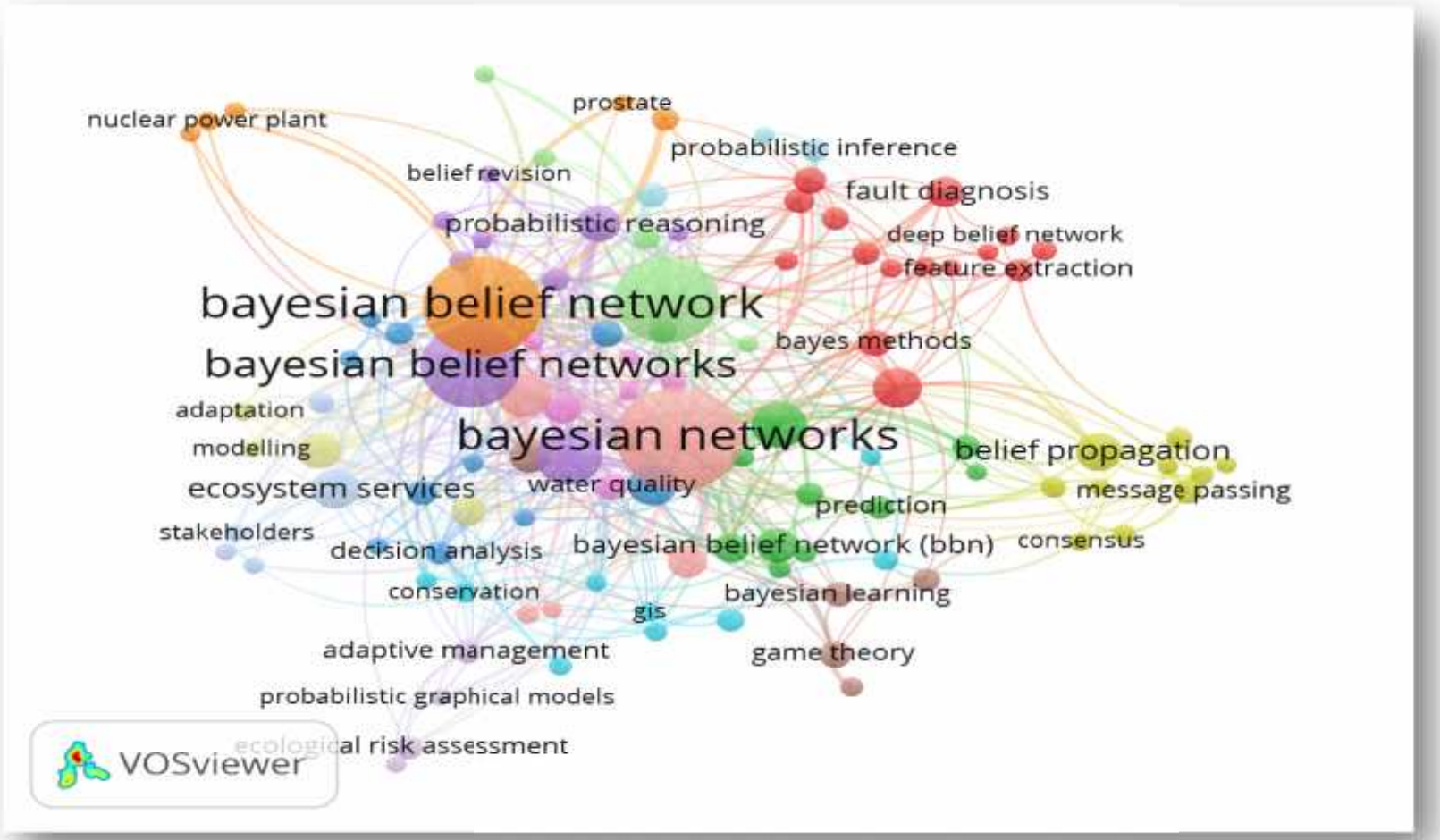

Figure 15: Keyword Co-Occurrence Network Visualization of Bayesian Network Publications

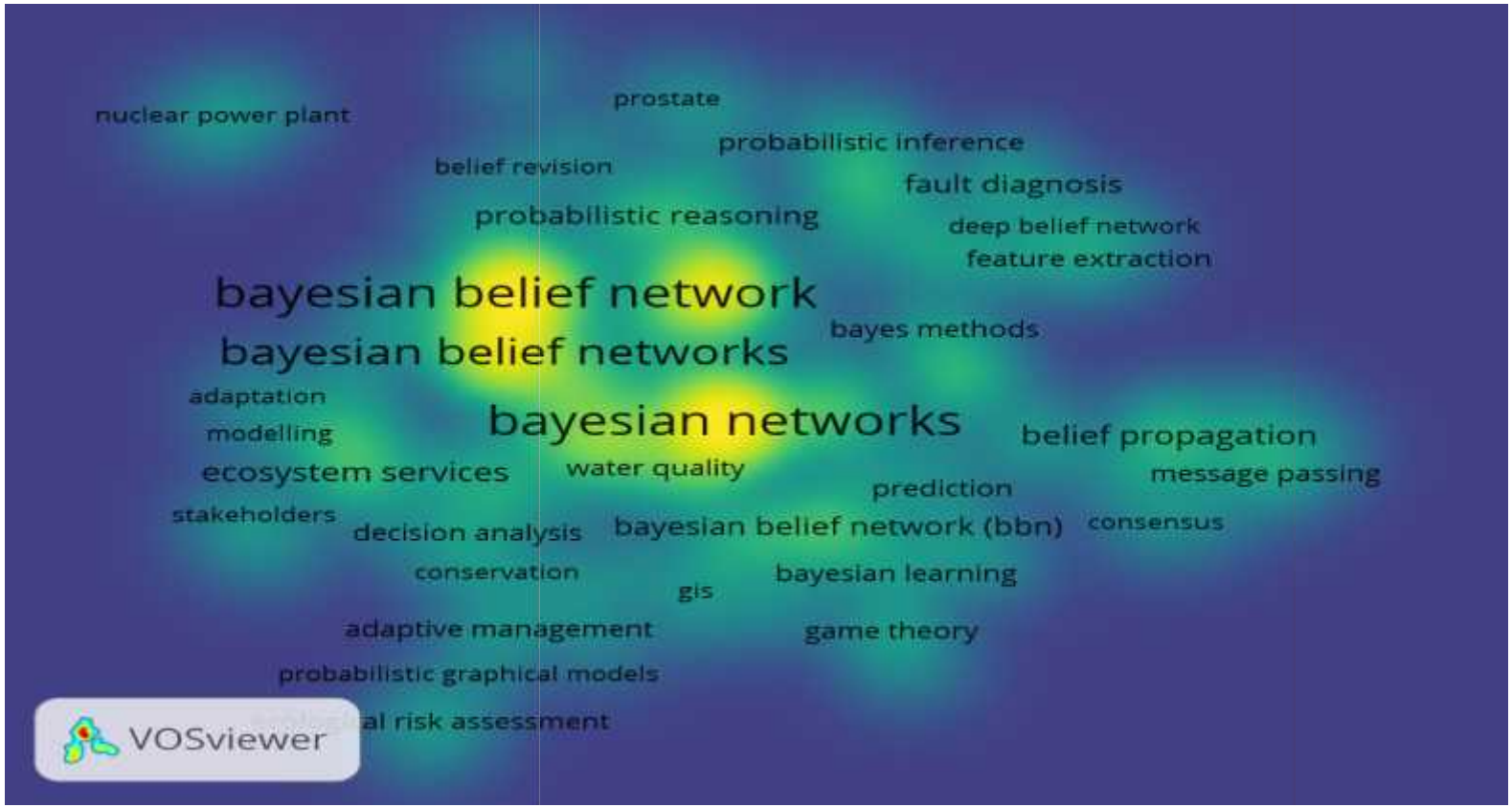

Figure 16. Keyword Co-Occurrence Density Visualization of Bayesian Network Publications.

CHAPTER 5

DISCUSSIONS AND CONCLUSIONS

5.1. Discussion

Bibliometric studies are very beneficial in evaluating the productivity and impact of hot research topics like Bayesian Network in different research fields, countries, institutions and journals. It revealed that the rapid increase and explosion of interest with in Bayesian Network articles was as a result of a strong scientific collaboration among authors, countries and institutions. In addition, increase in the number of citations by other researchers explains how educative and important Bayesian Network articles are to the academic domain.

Many funding agencies also support the Bayesian Network research with resource allocations.USA became the most publishing country in this study because of its effective engagement in research activities involving the application of Bayesian Network method giving rise to high number of publications and citations from researchers and institutions.

Bayesian Network method uses a graphical modeling, and this requires a bit more of computer knowledge. According to the results, this explains why Bayesian Network method is used mostly by Computer science and Engineering research area, the major field was Computer science artificial intelligence and also Environmental modeling & software was the best productive journal. This study showed that a few applications of Bayesian Network method are used in Medicine because statistical analysis approach and computer science is not that common.

There is huge data accumulation in health sciences but there is lack of proper application of this method to research studies due to little or no Computer science and Statistical method background. Bayesian Networks are used in prediction, solving decision problems and in detecting and assessing relationships between symptoms of diseases. This is why more researchers from Medicine department are encouraged to apply this method to these real healthcare data.

Bayesian Network has been applied in predicting pregnancy after psychiatric interventions in infertile couple (Abbas et al, 2016), epidemiological systems analysis (Lewis et al,2012)

and in building a clinical Bayesian Network based on data from the electronic medical report for medical Ontology (Ying and Kai,2018).

Jason Brownlee indicated in their research (Jason, 2019),Bayesian Networks are used in solving complex problems in different disciplines and their models can be learned from discrete or continuous data and by experts.Zelenik and Merigo showed that (Zeleznik, 2017 and Merigo,2015),bibliometric analysis must be done to analyze the research status,trends and hotspots of a topic. Friedman proposed that (Friedman et al, 1997). Bayesian network is a machine learning algorithm in detecting and explain classification. Also, Huchang Liao showed that (Huchang et al,2018) bibliometric analysis and visualization can be applied in medical big data research.

Lastly, bibliometric analysis is really important in analyzing and understanding the previous literature work on special subjects or research topics. In this sense, I really believe that bibliometric analysis, systematic reviews and meta-analysis studies must be part of a whole.

Generally, some analysts are not yet aware of the benefits of bibliometric analysis. They only do systematic reviews and meta-analysis studies without supporting it with bibliometrics.

5.2. Conclusion

There was a rapid increase in the yearly research activities pertaining to Bayesian Network research from the period of 1990 to 2020 both in the number of publications and citations. It produced a total of 1739 publications and 46,167 sum of citations from the year 1990 to 2020.The Bayesian Network research publications of 30years produced a h-index of 83 which explained the researchers that worked on this topic have published 83 journal articles (measures productivity) with at least 83 citations (measures impact) each.

This study showed that Bayesian Network research had so much attention and attraction from authors, journals, countries, institutions, research areas and funding agencies. The Environmental Modelling & Software was the most important scientific journal with 53 publications (3.05% of 1739) according to Web of Science and 1440 number of citations from VOS-viewer software.

The United States of America (USA) was discovered to be at the frontline of Bayesian Network research and also the most productive country with 551 publications (31.69% of 1739) followed by England and Peoples Republic of China with 206 publications (11.85%) each. Pearl Judea, 1988 had the highest number of co-citations (the most cited reference). The most frequent used keyword was Bayesian Network. Bartels P.H had the highest number of co-authored citations and publications.

5.3 Recommendations

Though compelling results were gotten through the bibliometric analysis and Visualizations on Bayesian Network research publications, this study had some limitations. The documents were only downloaded from Science Citation Index (SCI) and Science Citation Index Expanded (SCIE) databases through Web of Science and all the articles were written in English language. Further studies carried out on this research topic can include articles included in other citation indexes and languages in order to reduce the underestimation of researchers who use other languages. Statistical tests and other science mapping visualization software tools can be added to the study.

REFERENCES

Al-Shahi, R., Will, R. and Warlow, C. (2001). Amount of research interest in rare andcommon neurological conditions: bibliometric study. BMJ, 323 (7327): 1461 – 1462.

Bruna de Paula, F., Ricardo, B.S., Marcus, V. and Fabio, Z. (2016). Co-authorship network analysis in health research method and potential use.Health Res Policy Syst., 14: 34. .

Bukhari, D. (2020). Blockchain Technology: A Bibliometric Analysis. In: Stephanidis C., Antona M. (eds) Communications in Computer and Information Science, vol 1226. Springer,Cham.

Diez, F.J., Mira, J.,Iturralde, E. and Zubillaga,S. D. (1997).A Bayesian expert system for echocardiography. *Artificial Intelligence in Medicine*, 10(1),59-73.

Eungeong, P.,Chang, H. and Hyo Suk, N. (2018). A Bayesian Network Model for Predicting Post-Stroke outcomes with Available risk factors.*Front.Neurol*.

Friedman, N.,Geiger, D. and Goldszmidt, M. (1997). Bayesian network classifiers.*MachLearn*, 29:131-63.

Friedman, N.,Linial, M.,Nachman, I. and Pe'er, D. (2000). Using Bayesian network to analyze expression data.*Journal of Computational Biology*, 7,601-620

Friedman, N. (2004). Inferring cellular networks using probabilistic graphical models.*Science*, 303,799- 805.

Garbolino, P. and Taroni, F. (2002).Evaluation of Scientific Evidence Using Bayesian Networks.*ForensicSci Int.*,125(2-3):149-155.

Gyftodimos, E. and Flachi, P.A.(2004).Hierarchical Bayesian Networks:an approach to classification and learning for structured data. In Methods and Applications of Artificial Intelligence:Proceedings of the Third Hellenic Conference on AI(SETN 2004). *Lecture Notes in Artificial Intelligence*, 3025,291-300,Springer.

Henderson, S.O. and Brestky, P. (2003).Predictors of academic productivity in emergency medicine.*AcadEmerg Med*, 10 (9):1009 –1011**.**

Hui ,J., Wang, L., Liu, R., Yang,C., Zhang, H., He, S., Chen, Z. and Wei, A. (2020). A bibliometric analysisof international publication trends *in* premature ejaculation research (2008-2018). .*Int J Import Res*.

Husmeier, D., Dybowski, R. and Roberts, S.(eds) . (2005). Probabilistic modeling in bioinformatics and medical informatics and medical informatics.Springer,London. ISBN 9781852337780,.

Itagaki, M.W. (2008). Impact of the National Institutes of Health on radiology research.*Radiology*, 247 (1): 213 – 219.

Itagaki, M.W. and Pile-Spellman, J. (2005).Factors associated with academic radiology research productivity.*Radiology*, 237 (3): 774 – 780 .

Jansen R, Haiyuan, Y.,Dov, G., Yuval, K. and Nevan, J. (2003). A Bayesian network approach for predicting protein-protein interactions from genomic data.*Science*,302,49-453.

Kontkanen, P.,Myllymaki, P.,Silander, T. and Tirri, H. (1997).Comparing Predictive inference methods for discrete domains.*Proceedings of the Sixth International Workshop in Artificial Intelligence and Statistics*.311-318.Ft.Laurderdale,USA.

Mao, M.,Li, Z.,Zhao Z. and Zeng L. (2018).Bibliometric Analysis of the Deep Learning Research Status with the Data from web of science.In:TanY.,ShiY.,Tang Q.(eds)Data Mining and Big Data.DMBD 018. *Lecture Notes in Computer Science*, vol 10943.Springer,Charm.

Merigo, J.M.(2016). Academic research in innovation;A country analysis.*Scientometrics*,108,559-593.

Merigo,J.M.,Mas-Tur,A.,Roig-Tierno, N. and Ribeiro-Soriano D.A.(2015):A bibliometric overview of the Journal of Business Research between 1973 and 2014.*J.Bus.Res*,68,2645-2653.

Paul, A., Devo, B., Justine, J. S.,Alind, G., Brenner, D.R., Marek, J. D. (2019). Bayesian Networks for Risk Prediction Using Real-World Data:A tool for Precision Medicine.Value Health,22(4):439-445.

Pearl, J.(1985).”Bayesian Networks:A Model of Self-Activated Memory for Evidential Reasoning”(UCLA Technical Report CSD-850017)In*Proceedings of the 7th Conference of the Cognitive Science Society*,University of California,Irvine,CA.:329-334.Retrieved on 2009-05-01.

Pearl,J. (1988). Probabilistic Reasoning in Intelligent Systems,Morgan Kaufmann,San Francisco,CA.

Peter, A.,Geert, F.,Timmerman, D.,Moreau,Y. and De Moor, B. (2004). Using literature and data to learn Bayesian networks as clinical models of ovarian tumors.*Artificial Intelligence in Medicine*, 30(3), 257-281.111.

Peter,J.L., Vander Gaag, L. and Abu-Hanna, A. (2004).Bayesian networks in biomedicine and health care.*ArtifIntell Med*, 30(3):201-214.

Reyes-Gonzalez, L., Gonzalez-Brambila, C.N. andVeloso, F. (2016). Using co-authorship and citation analysis to identify research groups: a new way to assess performance. *Scientometrics* ,108**,** 1171–1191 (2016).

Schaer, P. (2013). "Applied Informetrics for Digital Libraries: An Overview of Foundations, Problems and Current Approaches". *Historical Social Research. **38** (3): 267–281*.

Sebastiani, P.,Yu, Y.-H.,Ramoni, M.F. (2003).Bayesian machine learning and its potential applications to the genomic study of oral oncology.*Advances in Dental Research*, 17,104-108.

Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics*, 38(2), 275–293.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecol Model*, 203:312-8.doi:10.1016/j.ecolmodel.2006.11.033.

Van Eck, N.J. and Waltman, L. (2010). Software survey: VOS-viewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.

Van Eck, N.J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (eds.), Measuring scholarly impact: Methods and practice. Springer,Cham.

Zeleznik, D.,Vosner,H.B. and Kokol, P.A. (2017).A bibliometric analysis of the journal of Advanced Nursing,1976-2015.*J.Adv.Nurs*.,73,207-2419.

Zyoud, S.H. (2016). Global research trends of Middle East respiratory syndrome coronavirus: a bibliometric analysis. *BMC Infect Dis*, 16**,** 255.