Republic of Iraq

Ministry of Higher Education and Scientific Research

University of Diyala

College of Science

Department of Computer Science

CITATION NETWORKS:
IRAQI UNIVERSITIES CASE STUDY

A thesis submitted to Council of College of Science, University of Diyala in partial fulfillment of the requirements for the degree of Master in Computer Science

By

Ahmed Jasim Mohammed

Supervisor by

Asst. Prof. Dr. Taha M. Hasan

Lecturer Dr. Basim Mahmood

July 2020 **IRAQ** **Dhul Hijjah1441**

بسم الله الرحمن الرحيم

﴿وَوُضِعَ الْكِتَابُ فَتَرَى الْمُجْرِمِينَ مُشْفِقِينَ مِمَّا فِيهِ وَيَقُولُونَ يَا وَيْلَتَنَا مَالِ هَٰذَا الْكِتَابِ لَا يُغَادِرُ صَغِيرَةً وَلَا كَبِيرَةً إِلَّا أَحْصَاهَا ۚ وَوَجَدُوا مَا عَمِلُوا حَاضِرًا ۗ وَلَا يَظْلِمُ رَبُّكَ أَحَدًا﴾

صدق الله العظيم

الكهف (٤٩)

Supervisors' Certification

We certify that this thesis entitled "Citation Networks: Iraqi Universities Case Study" was prepared by "Ahmed Jasim Mohammed" under our supervisions at the University of Diyala, Faculty of Science, Department of Computer Science as partial fulfillment of the requirements needed to award the degree of Master of Science in Computer Science.

(Supervisor)

Name: **Asst. Prof. Taha M. Hasan**

Signature:

Date: : 28 / 7 /2020

(Co-Supervisor)

Name: **Lecturer Dr. Basim Mahmood**

Signature:

Date: : 25 / 7 /2020

Approved by the University of Diyala, Faculty of Science, Department of Computer Science.

Signature:

Name: **Asst. Prof. Taha M. Hasan**

Date: 28 / 7 /2020

Head of Computer Science Department

Linguistic Certification

This is to certify that this thesis entitled "Citation Networks: Iraqi Universities Case Study" was prepared by "Ahmed Jasim Mohammed" at the University of Diyala, Faculty of Science, Department of Computer Science is reviewed linguistically. Its language was amended to meet the style of the English language.

Signature:

Name: Dr. Ghazwan Mohammed Jaafar AL-Hashimi

Date: 6 / 12 / 2020

Examination Committee Certification

We certify that we have read the thesis titled "Citation Networks: Iraqi Universities Case Study" and we have examined the student "Ahmed Jasim Mohammed" in its content and what is related to it, and in our opinion, it is adequate as a thesis for the degree of master in Computer Science.

Signature:

Name: Prof. Dr. Dhahir Abdulhade Abdullah **(Chairman)**

Date: 6/12/2020

Signature:

Name: Asst. Prof. Dr. Nada Hussein M. Ali **(Member)**

Date: 6/12/2020

Signature:

Name: Asst. Prof. Dr. Jumana Waleed Salih **(Member)**

Date: 6/12/2020

Signature:

Name: Asst. Prof. Dr. Taha Mohammed Hasan **(Supervisor)**

Date: 6/12/2020

Signature:

Name: Lecturer Dr. Basim Mahmood **(Co-Supervisor)**

Date: 9/12/2020

The thesis was ratified at the Council of College of Science, University of Diyala.

Signature:

Name: Prof. Dr. Tahseen H. Mubarak

Dean of College of Science University of Diyala

Date: / / 2020

DEDICATION

To the symbol of dedication and sincerity, my dear father

To whom was her prayer the secret of my success, my dear mother

To my honor and pride, my dear brothers

To whom is a piece of my heart, my affectionate sister

To my life partner and my life companion, my sweetheart my wife

To my soul, my life, and my happiness, my children, **Farah** and **Ameer**

To everyone who gave me advice and support ... my dear friends

Acknowledgment

*Praise be to **ALLAH**, Lord of the worlds, and prayers and peace be upon the master of the messengers, our Prophet **Mohammed** and his family and companions. I would like to express my thanks and appreciation to my supervisors, **Dr. Taha M. Hasan** and **Dr. Basim Mahmood** for their faithful guidance, valuable instructions, and constructive comments which have made the completion of this work possible.*

*Also, I would like to express my gratitude and my thanks to all the teaching staff who have taught me. Special thanks are extended to the members of the evaluation committee for discussing my thesis. Special thanks to all my friends for their help. Special thanks to **Mr. Khalil Al-Karkhi** who gave me everything he could. My final words go to my family. I want to thank my parents, my wife, and my children for their love and guidance which helped me in achieving my goals.*

Ahmed Jasim Mohammed

ABSTRACT

In the current technological era, scientific research is considered as one of the crucial factors for developing human life. The main sources for producing scientific research are worldwide universities, institutions, research centers, and scientific laboratories. Therefore, it is important to evaluate the performance of these institutions in terms of research production and quality. The main reason for this evaluation is to improve the performance of researchers and eventually reflect this improvement in scientific research status. Moreover, the productivity and quality of the researchers in a particular university can be measured based on two main indicators, namely, research citations and research publishing venues.

In this thesis, the current scientific status of the main Iraqi universities is deeply investigated. To this end, a Citation Network is generated among them. This kind of network can reflect the actual scientific research status of the main Iraqi universities. The approach that is used in this thesis is based on the concepts of complex networks. For the data collection, a special-purpose program is designed to crawling the Google scholar repository and retrieve all the required data. This crawler is designed to collect the published research articles based on the official educational domains of the Iraqi universities.

The first main contribution of this work is to generate a citation network of the Iraqi main universities and extract the main facts on scientific research activities. The second contribution is proposing a local rank for the main Iraqi universities based on network measurements and other academic indicators. Another aspect that is investigated in this work is the scientific collaboration among the Iraqi universities and with the worldwide universities. Furthermore, this thesis also shows the current status of the Iraqi universities compared to the world in terms of the Scopus repository. Based on the obtained results, this thesis provides recommendations and suggestions on how to improve the performance of

Iraqi universities in terms of scientific research and scientific collaboration among the universities.

The obtained results show an on-average performance of the scientific research in Iraqi universities according to network measurements such as the average clustering coefficient and the average path length. However, the University of Baghdad outperformed the other Iraqi universities in terms of the frequency of citations and the other network measurements. Also, the-top cited author was from the University of Baghdad in the field of Medicine with about 15566 citations (to the date of writing this thesis). However, the performance of scientific research in Iraq underperforms the neighbored countries such as Turkey, Iran, and KSA in terms of h-index, the number of the published papers, total citations, and the average citation per paper. The results also show that the collaboration among the Iraqi universities is based on the geographical area.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

List of Abbreviation

Abbreviation	Meaning
WoS	Web of Science
ECCN	Ego-Centered Citation Network
CTR	CiteTextRank
PTRA	Paper Time Ranking Algorithm
IUCN	Iraqi Universities Citation Network
l	Average Path Length
O	Diameter
D	Density
A_{CO}	Average Clustering Coefficient
cu	Communities
C_O	Clustering Coefficient
C_b	Betweenness Centrality
C_d	Degree Centrality
C_c	Closeness Centrality
RSS	Really Simple Syndication
SERP	Search Engine Result Pages
XML	Extensible Markup Language

GSC	Google Search Console
TV	Television
HTML	Hyper Text Markup Language
Geo Layout	Geographical Layout
IC	International Collaboration
C	Citations
RG score	ResearchGate Score
SI	Scopus Indicator

Chapter One

General Introduction

Chapter One
General Introduction

1.1 Introduction

In recent years, Iraqi universities have witnessed a great revolution in scientific research compared to the last decade. The reason behind this paradigm shift is that the ministry of higher education and scientific research in Iraq stepped forward towards the development of the Iraqi universities. One of the most important steps was encouraging researchers to publish their works in high indexed venues. The other reason was providing scholarship opportunities for the Iraqi scholars to perform their research abroad and bring some experience aiming at having colorful experiences at the local universities. Moreover, the collaboration opportunities with international institutions and universities enrich the local researchers with more experience in terms of the quality of the published researches. According to [1], the number and quality of the articles published in recent years have significantly increased. These facts lead to think more about how to increase the status of the Iraqi universities and obtain high international academic ranks. To this end, it is important to observe the patterns of the research activities by the researchers of the Iraqi universities. Then, an evaluation process should be performed aiming at having a deep look at the current situation of scientific research in Iraq. These processes help in developing and promoting the current patterns in a way that motivates to improve the whole scientific status of the Iraqi universities. In this regard, the citation network of a university or a group of universities can be used to investigate the citation and publishing patterns that are followed by the Iraqi researchers. In a citation network, two or more articles are considered to be connected if one of them is cited by the other article. In such a network, articles are represented as nodes and the links among them reflect the citations among them. Citation networks

are also used to measure the scientific status of a university or an individual researcher in a research community. They can also show all the past and the current collaboration activities performed by the authors. Furthermore, these networks can be utilized in identifying potential collaborations for future works. The proposed approach in this thesis is not based on traditional statistical analysis, instead, it is based on the concepts of *Complex Networks* in general and social networks in specific. The Complex networks' area is one of the modernist areas of research in computer science that appeared at the beginning of the 2000s. One of the main applications in complex networks is the field of Social Networks. It has emerged from sociology, statistics, and graph theory. Furthermore, using this field of study enables us to deeply investigate the relations among actors (authors). As one of the important types of social networks is Citation Networks, which can be used to measure network properties. Citation networks help to investigate and analyze the relationships among authors, research groups, and universities. The characteristics of this kind of network have been used earlier to understand and study the scientific collaboration among network actors [2]. In this thesis, the characteristics of the generated Iraqi citation network will be extracted at two levels: at the entire/global network and the author level. For the entire network, the *Giant Component* of the Iraqi citation network will be measured. At the author level, the centrality of the Iraqi authors/co-authors will be analyzed.

1.2 Literature Review

1.2.1 Analysis of Citation Networks

Researchers around the world have contributed to the field of citation networks in many different aspects.

J. Zhou et al. [3], 2019, investigated citation networks in finding the impact factor of publishing venues in similar disciplines based on the

average review cycle, the average number of references, and the yearly distribution of references. The results show that the yearly distribution of references of experimental disciplines (Nature Cell Biology/ Nature Chemical Biology) are mainly concentrated in the period 2000~2015 (up to 98.5% and 86.4% respectively), whereas in contrast, the percentages of references of journals in engineering and theoretical disciplines (IEEE Transactions on Automatic Control/ Linear Algebra and its Applications) before 2000 are 29.2% and even 51.7% respectively. However, this approach works well when the impact of the journal/conference is high but it has some issues when these venues do not have an impact factor, which is considered as a weak point.

Another study by **Y. Bu [4], 2020,** explored the citation networks using data from Web of Science (WoS) and Ego-Centered Citation Network (ECCN). This approach investigated three issues in these networks, namely, the structure of the network, the function of the network, and the bibliometric indicators. By using these factors, the scientific status of authors and institutions is evaluated. The method was based on the frequency of citations among different disciplines.

1.2.2 Analysis of the h-Index and Ranking

Citation networks can be a powerful tool for evaluating a particular research community. This evaluation can be useful for ranking authors or universities through their h-index.

A. P. Singh [5], 2011, proposed a new method based on citation networks for ranking the published research papers from different research fields in multiple conferences over the years. This method modifies the PageRank algorithm for ranking research papers by assigning an authoritative score to each paper. Depending on these scores, authors and conferences will have assigned formulated scores as well as they will be ranked. Additionally, the

approach has added another metric to the algorithm to rank papers taking into consideration the time factor for reducing the bias against recent papers that gained less time for being studied and consequently cited by the researchers as compared to the older papers. Moreover, in addition to paper scores, the algorithm included another feature that calculates the score of the year for each conference, and therefore, the researcher can find the best conferences in a specific year instead of the overall ranking of the conference. The results showed that the score of the top paper was (1.00000000) in both time-depended and time-independent domains. Also, the score of the top authors in time-dependent domain and time-independent domain were (0.34606894) and (0.37383376) respectively.

S. Das Gollapalli and C. Caragea [6], 2014, used citation networks to extract keyphrases from scientific papers. Keyphrases use *a small set of phrases* to give a brief description of a scientific paper. This approach used CiteTextRank (CTR) which is an efficient graph-based algorithm for ranking keyphrases using multiple sources of evidence such as the textual content of a paper, textually-similar neighbors, and neighbors in the interlinked paper in the citation network. The model obtained significant improvements over the state-of-the-art models. Specifically, on several datasets of research papers, CiteTextRank improves precision at rank 1 by as much as 9-20% over state-of-the-art baselines.

M. A. Hasson et al. [7], 2014, suggested an easily implemented new algorithm known as the Paper Time Ranking Algorithm (PTRA) for ranking scientific papers. This algorithm depends on three factors: paper age, citation index, and publication venue. Also, an assistance tool known as a web crawler is designed to crawl various databases of scientific papers for collecting the information needed by PTRA. Another crawler is designed for collecting the impact factors of the journals. The results

showed that PTRA depends on the paper age with a higher impact than the citation index and publication venue.

J. Li and P. Willett [8], 2015, proposed a new algorithm called ArticleRank which is a modified copy of the PageRank algorithm. This algorithm can be used in the analysis of citation data to calculate the number of citations. It distinguishes between papers that have the same number of citations, boosting cited papers by the papers with high impact. Additionally, ArticleRank can be used in the citation network analysis as an interesting alternative to *Times Cited* algorithm. However, it requires significant computation if it is applied to a large number of papers for many iterations.

X.-Y. Liu and B.-C. Chien [9], 2017, suggested a recommendation system for scientific research papers combined with cross-crawling-based tools for collecting papers and a *paper citation network analyzer*. The cross-crawling engine automatically collects relevant papers from various digital libraries. The *paper citation network analyzer* determines the degree of the papers based on both the citation relationship and the textual analysis of the collected papers. The results showed that the average number of authors per paper is 2.58, while the average number of papers per author is 1.94. Anyway, this approach makes paper collecting tasks more efficient than earlier paper recommendation systems.

O. Kinouchi et al. [10], 2019, used the K-index to calculate the number of citations in a complex network. A researcher has a K-index if and only if he or she is cited by K articles and each of these articles has a minimum number of citations equal to K. The K-index is examined on a list of researchers with higher citations, the list included twelve candidates of Physics Nobel Prizes for the year 2019 and above. Also, this study applied an improved ranking on the above list such that some candidates were

filtered out. In the final list, James Peeble who was the 2019 Physics Nobel laureate appeared to have a rank of 11 degrees. However, the list should be updated every year and should be compared to future years.

1.2.3 Analysis of Collaboration and Co-authorship Networks

Scientific collaboration is another aspect that is important to be investigated and analyzed. The collaboration among authors from the same discipline or different disciplines can be based on their co-authorship network. This kind of network formed authors as nodes and the edges among them are formed if they have co-authored a paper. This kind of analysis has been investigated in some works in the literature.

M. E. J. Newman [2], 2001, analyzed the collaboration networks of scientists in various disciplines such as medicine, biology, physics, and computer science. The study used author ascriptions from all papers published from 1995 to 1999. After the analysis, the scientific communities appeared like a small world, where the mean of the distance among the scientists over the intermediate collaborators' line differs *logarithmically* from the concerned community size. However, getting from any scientist to another from the same community needs no many steps. This study also discovered high clustering in the networks. This means that two scientists seem to be more collaborated if there was a third common collaborator selected randomly from the community. In this work, the distributions of collaborated scientists and their published articles have been tested. However, such distributions are suitable through a *power-law* and an *exponential cutoff*.

P. Ji and J. Jin [11], 2016, worked on two datasets of research papers: Co-authorship and Citation networks for statisticians. The study gathered the set of the published articles in the highest four statistics journals beginning with the year of 2003 till 2012’s half. After collecting and cleaning the

datasets, they have been analyzed in many various viewpoints such as *productivity*, *patterns and trends*, *centrality*, and *community structures*. The results highlighted *research habits*, *trends,* and *topological patterns* of statisticians.

1.3 Statement of Problem

According to the literature, there are several approaches to analyze research production and activities. However, there is no standard approach that can be adapted to analyze the scientific status of a particular author, discipline, or university. Moreover, the literature has a lack of providing approaches for measuring scientific collaboration in a particular research community. In addition to the aforementioned, there is a severe lack in the Iraqi literature that takes the Iraqi universities as their case study.

1.4 The Aim of Thesis

This thesis aims to generate an *Iraqi Universities Citation Network* (IUCN) for the Iraqi researchers based on their affiliations and disciplines. After generating the network, many facts will be revealed on the actual scientific situation in the research communities of the Iraqi universities. The main contributions of this thesis are:

1- Generate a citation network of the Iraqi main universities and extract the main facts on scientific research activities.
2- Propose a local rank for the main Iraqi universities based on network measurements and other academic indicators.
3- Investigate the scientific collaboration among the Iraqi universities and with the worldwide universities based on Scopus indicators. Also, investigate the scientific collaboration across disciplines in the Iraqi universities.

1.5 Work Challenges and Limitations

1- Simulation: it is very difficult to perform the complex computations.

2- Data: the collected data from Google Scholar contains only the authors who verified their accounts using the official domain. Some information is not collected because:

- Many authors did not have accounts on Google Scholar.
- Many articles have not yet published online.
- Many authors are passed away and their research are not available online.
- Duplicate accounts are not considered.
- Noise is removed from the data, which includes research articles but with missing information.

1.6 Organization of the Thesis

This thesis consists of five chapters; the following is a brief description of each one:

Chapter One: General Introduction

This chapter includes a general introduction to the topic of this thesis. It also contains the related works, the problem statement, aim, and challenges and work limitations of this thesis.

Chapter Two: Theoretical Background

It provides a comprehensive theoretical background on the topics that are considered in this thesis.

Chapter Three: The Proposed Approach

This chapter describes the data collection process. It also describes how the dataset of this thesis was generated and the main characteristics of the Iraqi universities' citation network.

Chapter Four: Network Analysis and Discussions

This chapter analyzes the citation network of the Iraqi universities and what are the main facts about them. Also, it provides a comprehensive and detailed discussion on the obtained results and the proposed rank.

Chapter Five: Conclusions and Future Works

This chapter contains conclusions and provides recommendations and suggestions on improving the scientific status of the Iraqi universities. Future works are also included in this chapter.

Chapter Two

Theoretical Background

Chapter Two

Theoretical Background

2.1 Introduction

This chapter describes the theoretical background of the topics that are taken in this thesis. It includes a summary of the Iraqi universities that are considered in this work. Google Scholar and its features are also described. Graph theory, Complex Networks, and Social Networks are described in sections (2.4), (2.5), and (2.6) respectively. Other subjects are described in this chapter such as Crawlers, Data Visualization, R language, and Gephi software.

2.2 Iraqi Universities

The first Iraqi university was established in 1957, which was the university of Baghdad. Then, the universities of Mustansiriyah, Basrah, and Mosul were established in 1963, 1964, and 1967 respectively. Appendix (1) presents a summary of the Iraqi universities that have been considered in this thesis. Scientific research in Iraqi universities was first triggered in the 70s. At that time, the main Iraqi universities were playing a leading role in scientific research in the Middle East. Most of the contributions were in the fields of Physics, Engineering, Medicine, Chemistry, and Mathematics [12]. Many Iraqi scientists around the world have significantly contributed to the world such as Eng. Zuha Hadeed, who was the icon of architectural engineering in the world.

2.3 Google Scholar

Google Scholar is one of the services of Google company that provides a simple way to broadly search for scholarly literature. It enables the researchers to seek across many fields of study: articles, dissertations, theses, review papers, books, and abstracts from different repositories such

as academic publishers, professional societies, online repositories, universities, and other web sources. Google Scholar enables us to search for related works over the world of scholarly research [http//: www.scholar.google.com].

2.3.1 Google Scholar Features

Google Scholar provides a lot of information related to research activities around the world [13].

- Search for all literature including any publication.
- Explore all the authors, co-authors, citations, related works, and publications.
- Provide an account with a library that can help researchers organize and store their searches.
- Makes us up-to-date with the recent works in any field of research.
- Enables authors and co-authors to see who is citing their works.

2.3.2 Ranking documents in Google Scholar

Google Scholar's method for viewing the publications varies according to the preferences of the authors. Usually, Google Scholar views the searched publications based on their relevance to the searched topic (e.g., keywords and other search settings). Users can also determine the year of the publications and the other indicators of interest. Users can also view their profiles and see their publications. The basic way of viewing one's publications is based on the citations of the publication; the highly cited publication is viewed first [13].

2.3.3 The h-Index in Google Scholar

The h-Index of an author or an educational institution can be defined as a value that indicates the influence of a researcher in the scientific research community. The h-Index was first suggested in 2005 by Jorge Hirsch [14], who is a physicist at California University. At that time, the established measure was raw citation counts. This means, to know the range of the productivity and influence of a particular author, it can be simply showing the number of times that other publications cited the papers of that author. Although this indicator is simple, authors around the world have discovered an issue in this indicator. For instance, if an author published one paper that was highly cited by other researchers around the world and then never published another publication again, it can easily be said that this author is successful. In this case, outliers will distortedly affect the comprehensive evaluation of the contribution of a researcher. The h-index value of an author/institution has been proposed for evaluating them in their area of research. Researchers use h-index routinely in various disciplines for evaluating themselves as well as evaluating other researchers in their field. Moreover, h-index has become a common requirement to evaluate the applicants for an academic position. It is used also in the evaluation of applicants for research grants. The calculation of a scholars' h-index has clear advantages through giving a degree of transparency about their influence in a field, which enables non-experts to evaluate the researchers' contributions in the field [15].

2.4 Graph Theory

Graph theory is a mathematical and computer science theory that studies graph features. A graph G can be represented as vertices (V) and edges (E) among them, such that G = (V, E). The edges of a graph can be directed when the relationship goes from a vertex to another one [16], as

shown in Figure (2.1). While the undirected edges mean the relationship is initiated in both directions (see also Figure 2.1). Graphs can be used to solve many scientific problems that can be formed as a graph. The applications of graph theory are very wide. Thus, it is needed to use computers in designing algorithms for graph theory in a way that helps to process every graph and get its features and information. Mathematically, a graph can be implemented by using an adjacency matrix [17].

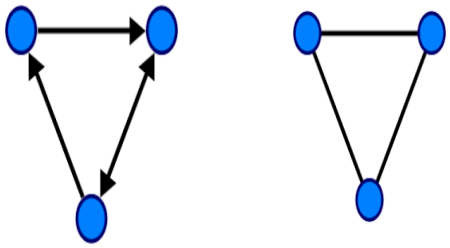

Figure 2.1 Directed and undirected graphs (The left part represents a directed graph and the right part represents an undirected graph) [16].

Graphs are characterized by many features; each has a certain purpose. The degree of an undirected graph node represents the number of edges connected to that node, and the average degrees of the nodes in a graph represent the degree of the whole graph [18]. The weight of an edge between two nodes represents the relationship strength between them [17]. The average shortest path, it is the shortest path to reach a certain node [17]. Diameter represents the longest distance in the graph (e.g., the longest distance between any pair of nodes) [18]. Graphs can have types based on their structure [19] such as Simple Graph, which is a graph that has no self-loop or multiple edges as shown in Figure (2.2).

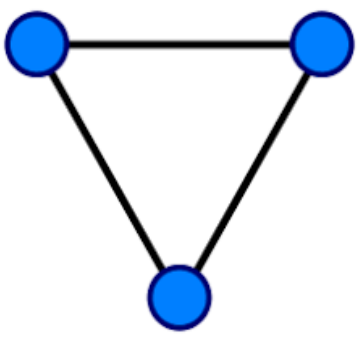

Figure 2.2 Simple Graph [19].

Multiple graphs, which is a graph that has no self-loop but it has at least multiple edges as depicted in Figure (2.3) [19].

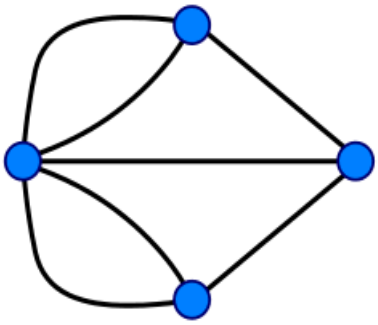

Figure 2.3 Multiple Graph [19].

Semi-graph, which is a graph that has self-loop and multiple-edges (see Figure 2.4) [19].

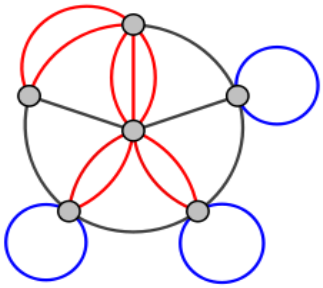

Figure 2.4 Semi-Graph [19].

A regular graph is a regular graph when the degrees of all nodes are equal in the graph as shown in Figure (2.5) [20].

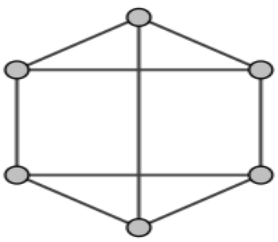

Figure 2.5 Regular Graph [20].

Complete graph, when every two nodes are connected as shown in Figure (2.6) [20].

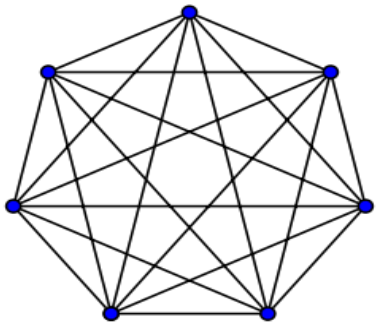

Figure (2.6) Complete Graph [20].

2.5 Complex Networks

In the theoretical principle of networks, a complex network can be defined as a graph having a considerable number of features that do not appear in ordinary networks like lattices or random graphs but mostly appear in graph models of real systems [21]. Studying complex networks has been started at the beginning of the 21st century and became an active field of scientific research. It is spired highly in general from the experimental study of real-world networks such as technological networks, biological networks, computer networks, climate networks, road networks, social networks, and brain networks [22]. Complex networks can use the same measurements that are used in graph theory.

2.6 Social Networks

Social networks have become one of the most important tools used for connecting people from around the world wide. They enable people to discuss different aspects regardless of their geographical locations, cultural background, religion, etc. Facebook and Twitter social networks represent the most popular networks of people. Social networks represent the structure of relationships amongst a group of effective individuals, organizations, or groups. The social network's study represents a

multidisciplinary discipline consisting of basic concepts that can be used in various disciplines such as biology [23], economics [24], geography [25], and others. The social networks analysis [26] is a combination of sociology, graph theory, statistics, and social psychology. The social network's study helps us to understand the relationships between people and the actors' influential behavior within the network. The World Wide Web includes many social networks used for different purposes. Some of these networks are used for academic purposes such as LinkedIn, ResearchGate, and Mendeley.

2.6.1 Analysis of Social Networks

Social networks' representation can be done by two common methods [16], as a graph, or as a matrix. In a graph representation, the graph G= (V, E), where G is a social network, V indicates the nodes in the network, and E indicates the relations between nodes. These relations may either be directed or undirected. More clearly, if there is a set of social actors {1, 2, 3, ..., n}, and a link (*i*,*j*) that links the actor *i* with the actor *j*, then it can be said that the relation between *i* and *j* is directed if and only if $(i, j) \neq (j, i)$, otherwise (i.e. $(i, j) = (j, i)$), the relation is undirected [16].

The second method for representing social networks is matrix representation. In matrix representation, an $(n \times n)$ matrix is used, where *n* is the number of individuals in the social network. Every cell in the matrix has a binary value of 0 or 1 that represents the relations between the actors. This value will be 1 if actor *i* is connected with actor *j*, otherwise, the value is 0. As mentioned previously, the relations between the nodes of a network are either directed or undirected. In the undirected connection, the distance between node *i* and node *j* represents the total number of connections that start from node *i*, pass through other nodes, and end with node *j* [27], [28].

There are a lot of features to characterize every node in a social network, one of them is a node degree, which represents the overall number of ties with other nodes. In directed networks, the number of incoming ties to a node within the network can be represented by an in-degree feature, whereas the number of outgoing ties from the node is represented by an out-degree [29]. A degree distribution [30] for a network is the probability distribution of degrees over the entire network. The importance of the degree distribution comes in the analysis of the real-world networks (e.g., the World Wide Web and on-line social networks). In such networks, some nodes may be as hubs of high connections, while the others have fewer connections. In this case, the degree distribution must follow a *power-law distribution*. These types of networks are known as *Scale-Free-Networks* [31].

In social networks, nodes can be clustered into groups or communities that share popularized features basing on popularized characteristics, like function and position [20]. As an example, citation networks [22] include communities by research topic; interaction protein-protein networks [32] may contain communities of protein that have similar functions. In such situations, the network will have a community structure.

Additionally, besides the above-mentioned characteristics, the relations in social networks have else characteristics such as:

- **Homophily**: Nodes incline for associating and connecting to similar ones [33], [34]. By McPherson's study et al. [29], Homophily is based on two major concepts, they are; Status Homophily and Value Homophily. In Status Homophily, individuals that have similarities in social status features are more probable for participating together fortuitously (e.g., religion, gender, age, or race). In Value Homophily, the participation is done between individuals that have similarities in

behavior and thinking regardless of the difference in social status. However, and according to the authors' study in [35], the impact of Value Homophily on Homophily is more significant than the impact of Status Homophily. Furthermore, Homophily is caused by many factors such as geography (e.g., people locations), social ties (e.g., strong and weak ties), organizational foci (e.g., school or work), isomorphic sources (e.g., people who occupy equivalent roles), and cognitive processes (e.g., people who have demographic similarity) [34].

- **Transitivity**: If a node, *i*, has a relation to a node, *j*, and node, *j*, has a relation to a node, *k*, then node, *i*, also has a relation to the node, *k*, (e.g., a friend of my friend is my friend)[36].
- **Reciprocity or Mutuality**: The inclination of actors for constituting reciprocal relations (double links) between each other (e.g., amongst friends)[37].

2.6.2 Network Measurements

In a social network, the importance of a node can be determined through the use of network measurements. There are two levels at which network measurements are applied; network level and node level.

2.6.2.1 Network Level Measurements:

a) *Average Path Length l*: For all the possible pairs of nodes in the network, it represents the average number of paths (steps) for all the shortest paths among the pairs. It shows the average shortest distance among the nodes and can be calculated using the following equation [38]:

$$l = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij} \qquad (2.1)$$

where d_{ij} is the distance between the nodes *i* and *j*, *n* is the number of nodes.

b) *Diameter O*: For a network, it is the longest path among all the shortest paths [16].

c) *Density D*: It represents the proportion of the number of network edges to the number of potential (possible) edges in that network. It can be calculated based on the following equation [16]:

$$D_G = \frac{2(E(G))}{N(N-1)} \qquad (2.2)$$

Where D_G is the degree of the graph G, E(G) is the actual number of edges in the graph, and N is the total number of nodes.

d) *Average Clustering Coefficient* A_{CO}: This measurement is also called the *global clustering coefficient*; it shows the average tendency of nodes to cluster with each other [20].

e) *Communities cu*: It represents a group of nodes that are densely connected [4].

f) *Degree Distribution*: A network's degree distribution is considered as a basic indicator of the nature of that network. For instance, if a network's degree distribution followed a power-law distribution, according to [27], and [28], this network is classified as a scale-free network. In such a network type, the newly attached nodes tend to cluster and collaborate with the highly connected nodes in the network. This phenomenon is called preferential-attachment.

2.6.2.2 Nodes Level Measurements:

a) *Clustering Coefficient* C_O: As mentioned in the previous section, this measurement shows the tendency level of nodes to cluster together in groups. The value of C_O for each node can be formulated as follows [20]:

$$Co_{(i)} = \frac{2|\{l_{jk}: n_j, n_k \in Ni, l_{ik} \in E\}|}{ki(ki-1)} \qquad (2.3)$$

where l_{jk} is a group between the nodes n_j and n_k. While N_i is the total number of nodes in a social network, and *ki* are the closest connected nodes in the network.

b) *Betweenness Centrality* C_b: It represents the overall number of the shortest paths passing over a specific node in the network. It shows the more probable information receiving nodes in the network. The C_b of a node *j* can be defined as follows [39], [40]:

$$C_b(j) = \sum_{i \neq j \neq k} \frac{\sigma_{ik}(j)}{\sigma_{ik}} \qquad (2.4)$$

where σ_{ik} is the shortest paths between the nodes *i* and *k*. $\sigma(j)$ is the number of paths that pass through node *j*.

c) *Degree Centrality* C_d: This measure represents the overall number of links to other nodes within a network. It detects only the number of connections for each node [20], [37].

d) *Closeness Centrality* C_c: It represents the reciprocal of the sum of all the shortest paths of a node to other network nodes. It also shows how close a node to other network nodes and formalized as follows [39], [41], [37]:

$$C_c(i) = \frac{N-1}{\Sigma d(ij)} \qquad (2.5)$$

where $d(ij)$ is the distance between the nodes *i* and *j*, *N* is the number of nodes.

To model social networks, capturing common characteristics of a real-world network is needed (e.g., friendship networks). As referred to previously, common characteristics of real-world networks are centrality,

transitivity, clustering, reciprocity, and homophily. When modeling social networks, these characteristics must be captured.

2.7 Crawlers

A crawler is computer software that performs an automatic document search on the Web [42]. Crawlers are mainly designed for repeated actions so that browsing is completely automated. Crawlers are used repeatedly by search engines for browsing the World Wide Web and building indexes. There are different types of crawlers, searching for various kinds of information such as Really Simple Syndication (RSS) feeds and e-mail addresses. The term crawler is inspired by the foremost internet search engine, the Web Crawler. Bot and Spider are other related words. Googlebot is the most famous web crawler. Now let's describe how a crawler program works. Principally, a crawler works as a librarian. It searches for the information on the Web, assigning them to certain categories, and then indexing them into catalogs, thus, it will be easy to retrieve and evaluate the crawled information. Unlike the librarian, who is performing his work self-sufficiently and assigns to his team other tasks, a crawler isn't independent. Before initiating such a crawl, it is necessary to establish all operations of the crawler. Thus, each order will be defined in advance. Then the crawler does an automatic execution of these instructions. Classically, the crawler's results will be added into a created index, which is accessible by an output software. The instructions determine the type of information that will be gathered by the crawler [43]. Figure (2.7) shows a graphic visualized the link relationships that the crawler uncovered.

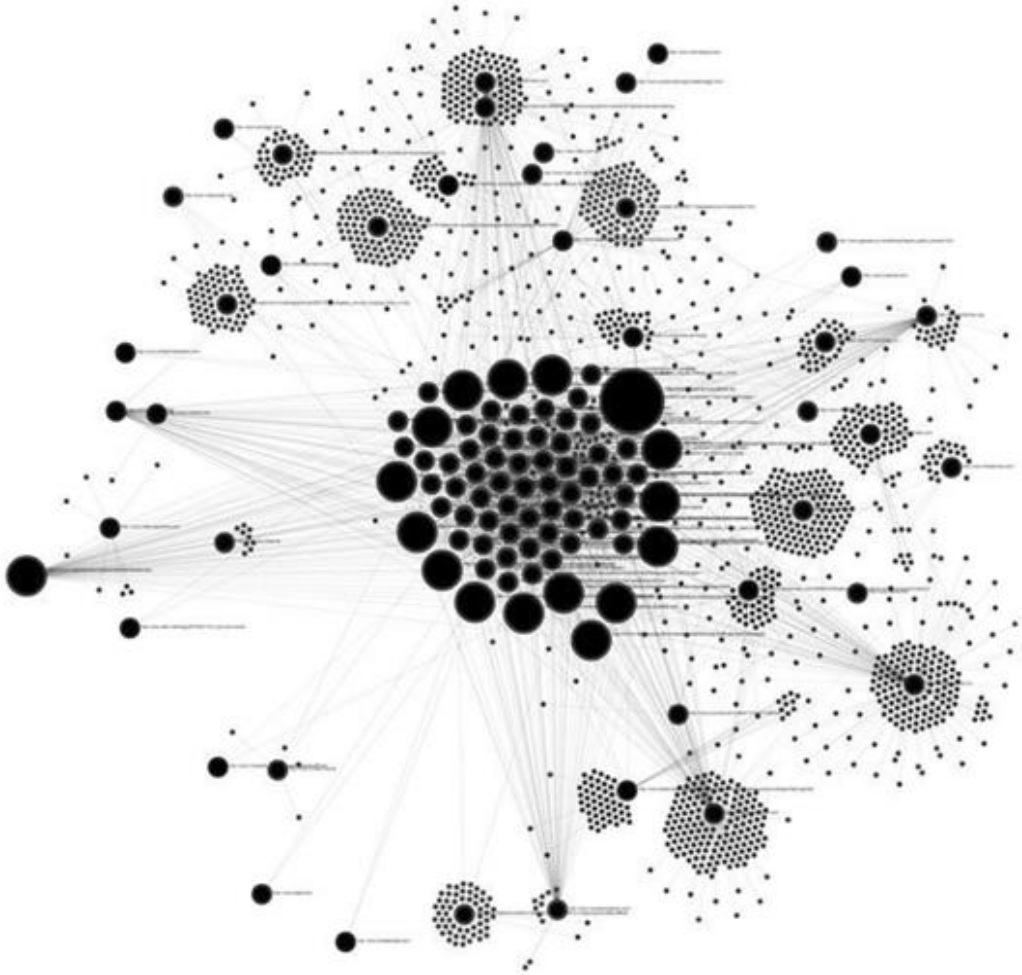

Figure 2.7 Crawler uncovered link relationships [43].

Often, crawlers visit sites without consent which causes resource consumption of such systems. When accessing large groups of pages, problems of schedule, load, and politeness will appear. Public sites that do not wish to be crawled have mechanisms to inform the crawling agent about that. For example, to include a file robots.txt, bots may do indexing for only some portions or nothing of the website. Because of the internet's significant number of pages, the largest crawlers cannot make complete indexes. For this reason, search engines grappled to present relevant search results before the year 2000 at the beginning of the creation of the World Wide Web. Presently, relevant results are presented at instant nearly. Figure (2.8) shows the crawler block diagram.

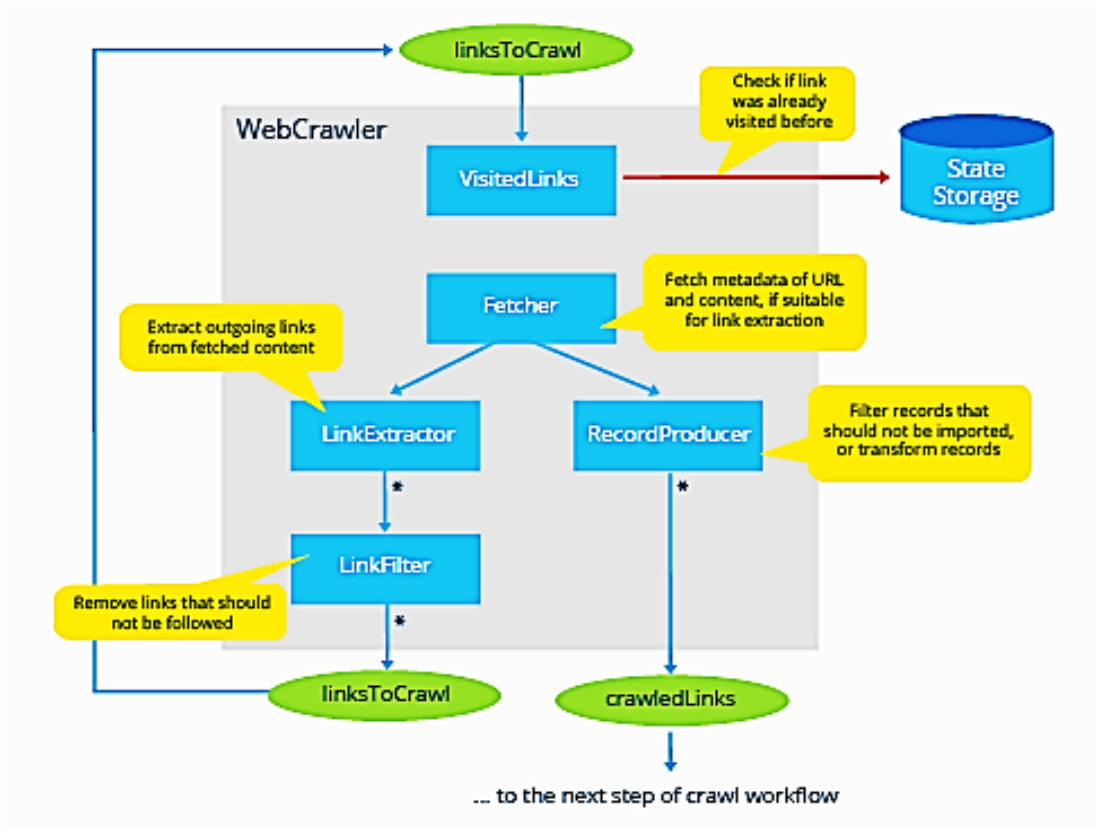

Figure 2.8 Crawler Block Diagram [https://en.ryte.com/wiki/Crawler]

A Web crawler behaves according to a collection of policies that determine its conduct [44]:

- A selection policy that determines the pages to be downloaded.
- A re-visit policy that determines the time for checking changes on pages.
- A politeness policy that determines the way of web sites' overloading avoidance.
- A parallelization policy that determines the way of distributed web crawlers' coordination.

Because of the classical objective of crawlers which is creating indexes, they are considered as a basis for search engines' work. A crawler searches the content on the Web then it makes the results available to the user. A focused crawler, as an example, focuses on the current content-relevant

websites during the indexing operation. The web crawlers can also be used in other disciplines such as [45]:

- Price comparison portals look for information on particular products on the Web to compare prices or data precisely.
- In the data mining discipline, crawlers can gather publicly-available email or postal locations of organizations.
- Web analysis tools employ crawlers or spiders in collecting data for viewing pages, or incoming or outbound links.
- Crawlers provide information centers with data, for instance, news locales.

Unlike scrapers, crawlers only collect and prepare data. However, scraping is a black hat technique. The scraper’s objective is copying data in a content form from else sites and putting it in that way or a form (with a little modification) on one's website. The difference between the crawlers and the scrapers is that crawlers often deal with metadata that is invisible to a user at the first glim, whereas scrapers extract the perceptible content [45].

Web-crawlers like Googlebot rank websites in the Search Engine Result Pages (SERP) through crawling and indexing. They keep track of permanent links on websites and the World Wide Web. For every website, each crawler has a limited timeframe and budget. Through optimizing the structure of a website (such as navigation and file size), Googlebot's crawl budget can be better utilized by the operators of the website. Simultaneously, the budget increases by a variety of inbound links and a robustly visited website. Googlebot is controlled through the data of robots.txt and the Extensible Markup Language (XML) sitemap stored in the Google Search Console (GSC). In GSC, it can be texted about

Googlebot's possibility to reach and index all relevant areas of a website [46].

2.8 Data Visualization

Data visualization can be defined as a graphical representation of data and information. Through the use of visible objects such as maps, charts, and graphs, the tools of data visualization can provide accessible ways of seeing and understanding outliers, patterns, and trends in data. In a Big Data world, the importance of data visualization tools and techniques comes in analyzing huge quantities of information and making data-driven decisions [47].

Colors and patterns catch our eyes. By looking at graphics, for instance, we can directly recognize blue from red, circles from squares. In addition to our visual culture, which includes a variety of visualities such as advertisements, arts, movies, TV, and so on; the other type of visual art is a data visualization that triggers our interest and keeps our eyes on the message. On seeing a chart, our sight will immediately fall on outliers and trends. If we are capable of seeing something, then we will internalize it quickly. It is storytelling for a purpose. If you stare at a big spreadsheet of data and cannot see a trend, then you will know the great effectiveness of visualization [48].

2.9 R Language

R language is a programming media for computing statistics that allows carrying out statistical applications and setting statistical programs. On other hand, it is a free resource and has a free version of statistical

programming S plus (S+) which depends on S language. The inputs into the R language could be arrays, matrix, or time series. And, it could be a data graph. R language is a dependent language that means all its orders are dependent which allows many mediators. One of its characteristics is having a face language Hyper Text Markup Language (HTML) which helps to carry out programming instructions easily. Also, R language is considered a modern type, and it is increasingly used in scientific programming for statistical works and bioinformatics. It has been widely adopted by many universities and scientific research centers. Many articles in scientific magazines refer to it largely [49], [50]. There are many aspects of using the R language as follows [51]:

- Google uses R to calculate investment revenues in advertisement campaigns.
- R is used to perform calculations of complicated networks.
- Ford Company uses R to improve its car designs.
- Twitter uses R to watch user experiments.
- The National Weather Organization uses R for heavy floods for the cast.
- The Human Rights Committee uses R to analyze data concerning the war's effects.
- New York Times uses R to set indexes and press application of interaction data.

2.10 Gephi Software

Gephi can be defined as an open-source software package used in c and visualizing networks, written in Java on the platform of NetBeans. Gephi has been used in some research projects in academia, journalism,

and elsewhere. For example, it has been used to visualize the universal connectivity of the content of the New York Times, also in the network traffic examination for Twitter during social disorders along with traditional topics of network analysis. Gephi is widely used in digital humanities, literature, political sciences, etc., a society involving many of its developers. Gephi inspired the LinkedIn, InMaps, and was used for visualizing the network for truthy [52].

Chapter Three

The Proposed Approach

Chapter Three

The Proposed Approach

3.1 Introduction

This chapter describes the data collection process and network creation. It also provides a description of the methodology and the approaches involved in this thesis. Figure (3.1) shows the general workflow diagram of the proposed approach.

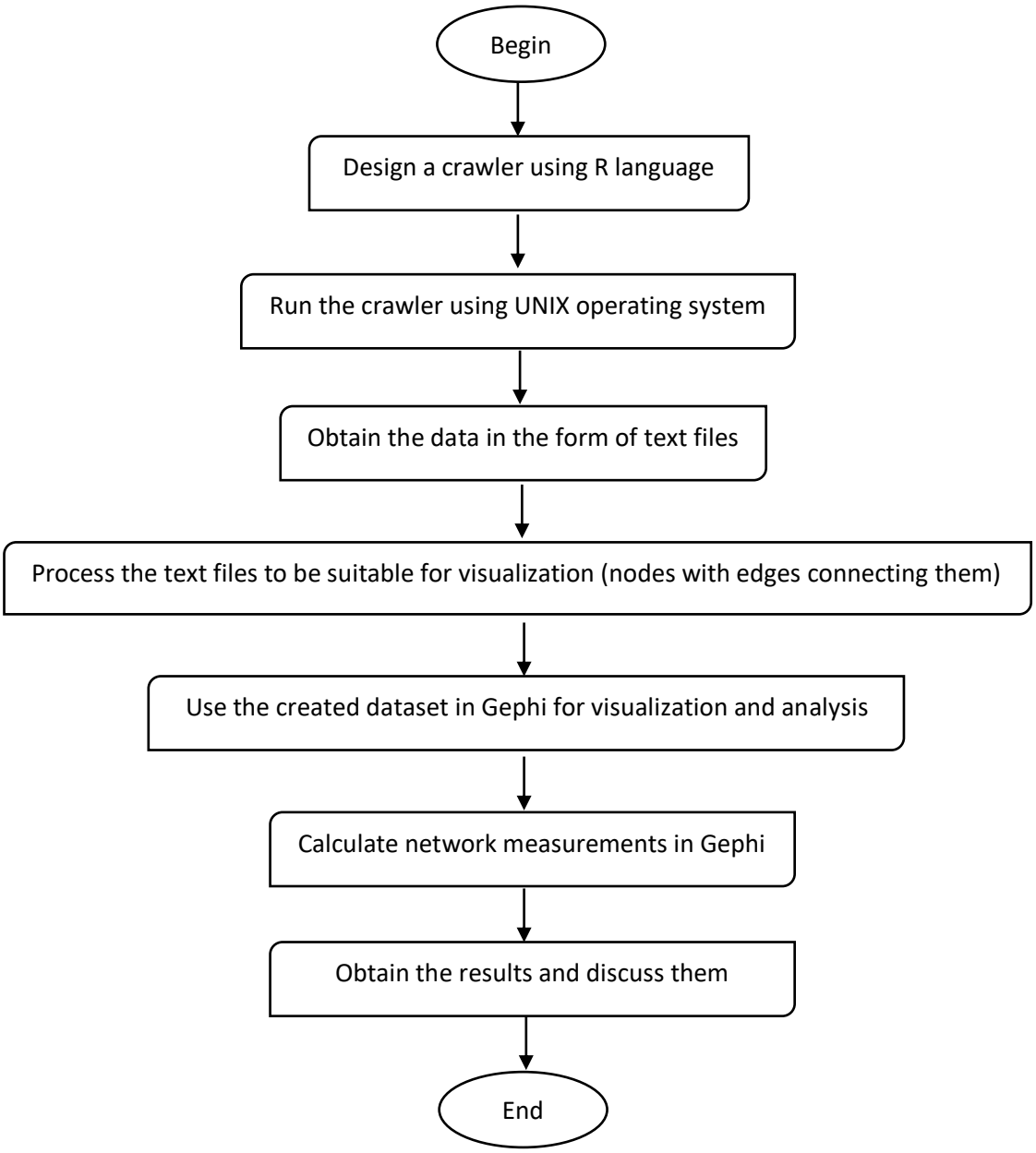

Figure (3.1) The general workflow diagram of the proposed approach

3.2 Data Collection

The data collection process was based on a particular dictionary that contains the main 22 Iraqi universities. This dictionary includes information on the Iraqi universities such as university name(s) and university domain name. Also, the collection process retrieved all the articles that have been published by authors who have been verified in Google Scholar using their official educational domain (e.g., *author@universitydomain.edu.iq*). A crawler program is designed to prepare that dictionary and collect researches' information that is contained in the Google scholar repository since the year 2000. The collected data includes information such as author identifier, author's affiliation, author total citations, author discipline, author total articles, and the affiliations of the author's co-authors. Then the crawler processes the data and formalizes. It aims at generating the dataset, which contains nodes (publications) and edges (citations). Algorithm (3.1) and Figure (3.1) show the details of the designed crawler program. In addition to the collected data from Google Scholar, some of the data have also been retrieved on the Iraqi universities from the Scopus repository aiming at using it in the analysis.

Algorithm (3.1) The crawler program

Input: Universities domain names
Output: Dataset
Begin Step1: Prepare Universities Dictionary;

Step2: Connect to Google Scholar using “Scholar” package in R;

Step3: If all links not visited, then Crawl next link in Google Scholar;

Else, Go to step 5;

Step4: If the link is not visited, then:

If publication found in the link, then:

retrieve University/Author information,

store the retrieved information in a file,

end if;

go to step3;

Step5: formalize the dataset;

End

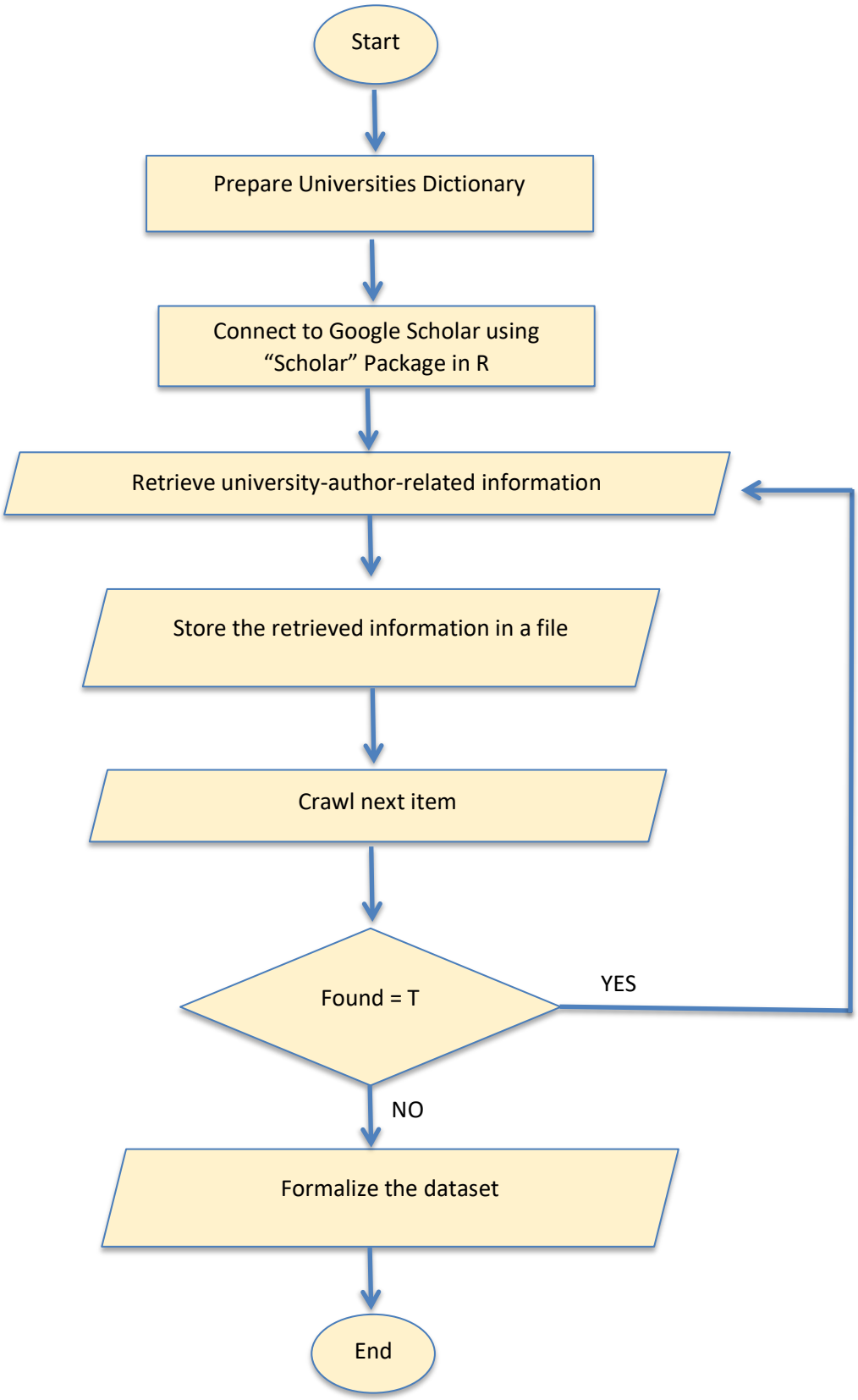

Figure 3.2 The flowchart of the crawler designed for data collection

3.3 Network Formation

As the final step in the data collection process, the program formalizes the dataset to be convenient for visualization. The dataset contains nodes and edges among them. The strategy followed in creating network objects (nodes and edges) is based on the following (see Figure 3.2):

a) Each publication that was authored by an Iraqi author(s) is formed as a node.
b) If the publication in (a) was cited by a publication authored by another Iraqi author, then an undirected edge is created between them.
c) Some of the publications might be authored by Iraqi and international authors, these cases are considered in this work since these kinds of publications have Iraqi authors.

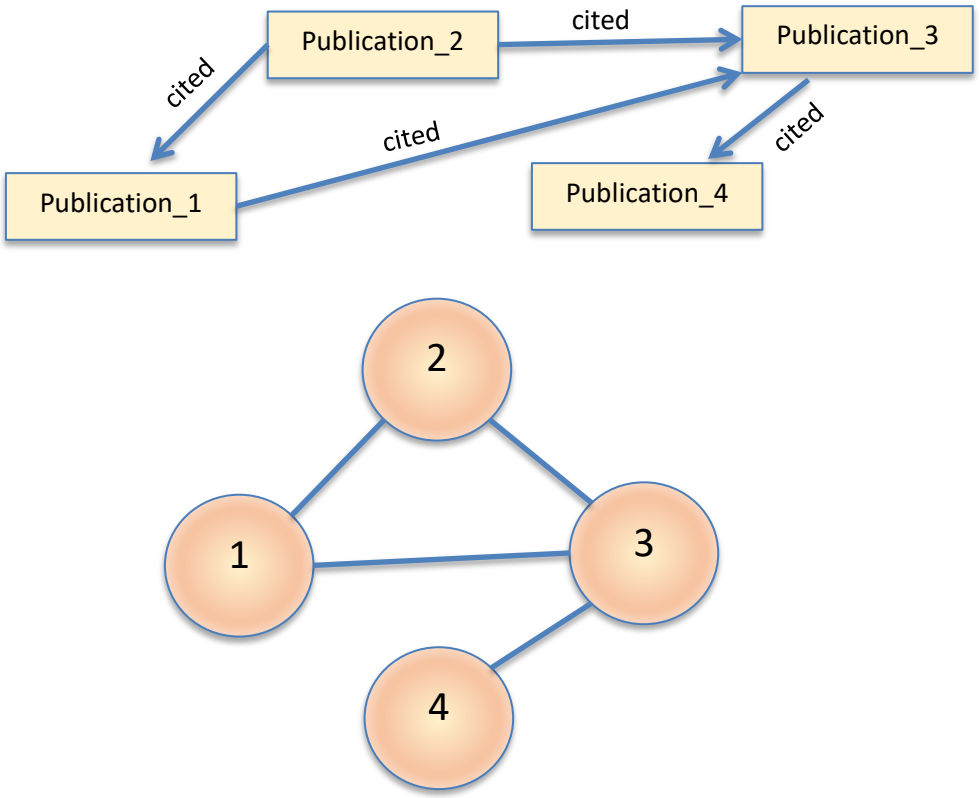

Figure 3.3 Network formation Strategy

The generated network contains 25,834 nodes (publications), and 167,267 edges among them. Finally, the generated network will be called the **I**raqi **U**niversities **C**itation **N**etwork (**IUCN**). It should be mentioned that although a citation is considered as one direction relation, the undirected type of edges is used in this work since the direction of the concept of this thesis is not of interest.

3.4 IUCN Measurements

The measurements that are used in this thesis are inspired by the complex networks. In Chapter Two, the network measurements have been described. However, this section specifies the measurements that are involved in the proposed approach at two levels as follows:

3.4.1 Network Level Measurements:

a) *Average Path Length l*: For all the possible pairs of publications on the IUCN network, it represents the average number of paths (steps) for all the shortest paths among the pairs. In IUCN, it shows the average shortest distance among the publications of the Iraqi authors.

b) *Diameter O*: In the IUCN network, it calculates the distance between the farthest publications that were authored by the Iraqi authors.

c) *Density D*: It represents the proportion of the number of network citations to the number of potential (possible) citations in that network. It also shows how dense the relations among the Iraqi authors in IUCN.

d) *Average Clustering Coefficient* A_{CO}: This measurement is also called the *global clustering coefficient*; it shows the average tendency of authors to cluster with each other. In other words, it reflects the average

tendency of Iraqi authors to collaborate and work as research groups [20].

e) *Communities cu*: This term can be referred to as the Iraqi authors in IUCN that are densely connected. This thesis benefits from Girvan-Newman [4] to extract IUCN clusters and reveal the collaborative groups of researches. This algorithm recognizes the relations that tie IUCN communities and eliminates these relations leaving the communities themselves. Girvan-Newman algorithm depends on the *betweenness centrality* in finding network communities.

f) *Degree Distribution*: According to what described in Chapter Two, and in this context of the IUCN network, fresh researchers have a strong tendency to collaborate with the most influential researchers.

3.4.2 Nodes Level Measurements:

a) *Clustering Coefficient* C_O: This measurement shows the tendency level of authors to work as groups and collaborate.

b) *Betweenness Centrality* C_b: This measurement shows how many times a publication cited in the shortest path of the IUCN publications pairs. It reflects the importance of a publication in being a bridge between the IUCN publications. This measurement can also be useful for evaluating the scientific collaboration level among network authors.

c) *Degree Centrality* C_d: Reveals the actual number of citations for a particular publication in IUCN.

d) *Closeness Centrality* C_c: It represents the reciprocal of the sum of all the shortest paths of a publication to other network publications. It also shows how close a publication/author to other network publications/authors.

The above measurements give a deep view of the relations of a particular individual to the other individuals in IUCN.

3.5 The Proposed Approach for Ranking the Iraqi Universities

This section describes the proposed approach for ranking the Iraqi universities. Figure (3.3) shows the structure of the proposed approach.

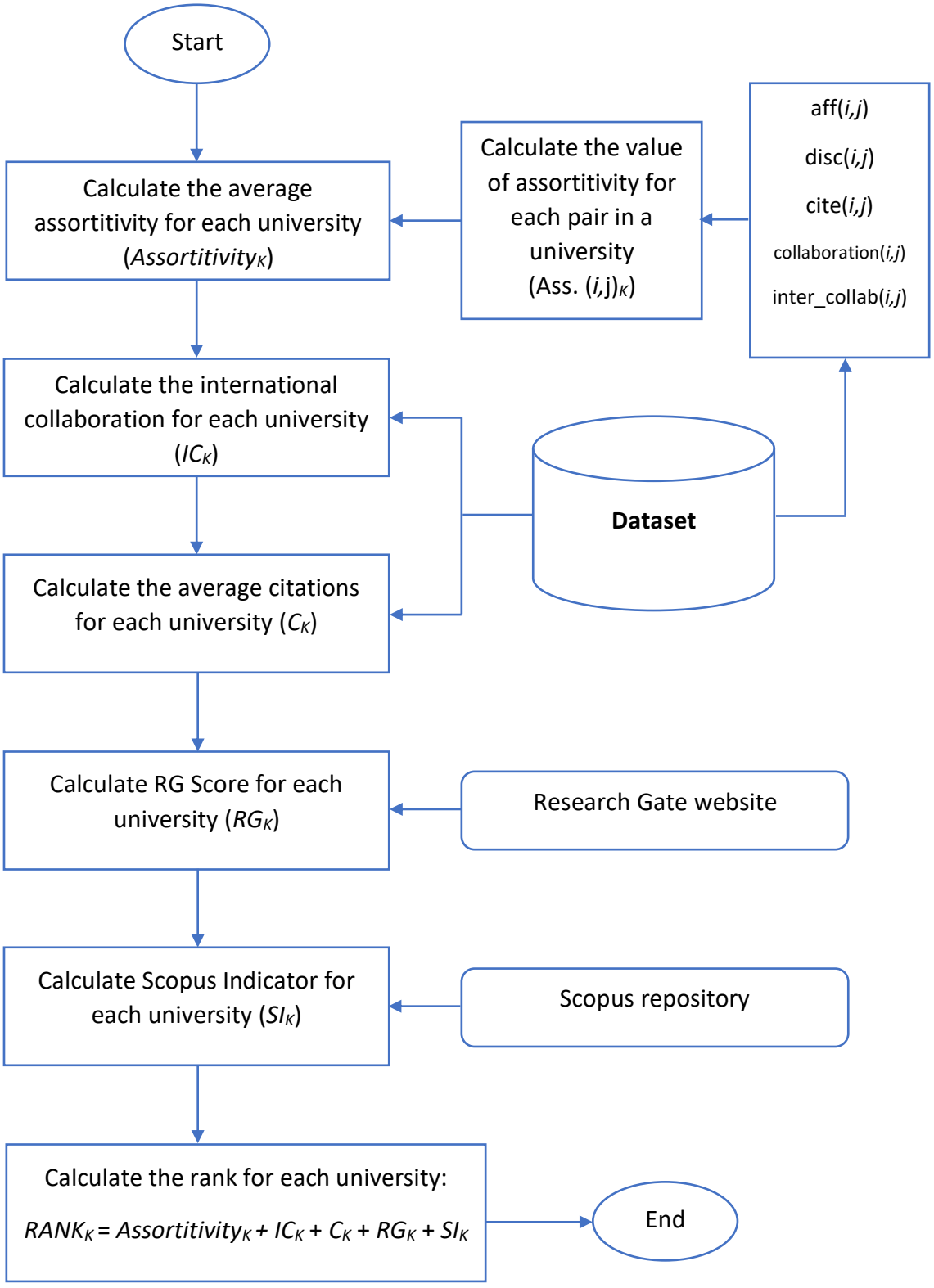

Figure 3.4 The general structure of the proposed rank

Algorithm (3.2) includes a pseudo code that describes the methodology used for calculating the proposed rank for the Iraqi universities. This methodology is based on the IUCN network measurements and other scientific indicators such as RG score and Scopus indicator.

Algorithm (3.2) The methodology of calculating the proposed rank

Input: Information from dataset
Output: Rank

Begin
Step1: k=0;
Step2: $k = k + 1$;
Step3: For i = 1 to N,
For j = 1 to N,
If *aff*(i)=*aff*(j), then *aff*(i,j)=1, else *aff*(i,j)=0;
If *disc*(i)=*disc*(j), then *disc*(i,j)=1, else *disc*(i,j)=0;
If *cite*(i)=*cite*(j), then *cite*(i,j)=1, else *cite*(i,j)=0;
While *co-author*(i,j)=true, l=l+1;
collaboration(i,j)=l;
If *co_author*(i,j,*international*)=*true*,
then *inter_collab*(i,j)=1, else *inter_collab*(i,j)=0;
Ass.(i,j)=*aff*(i,j)+*disc*(i,j)+*cite*(i,j)+*collaboration*(i,j)+*inter_collab*(i,j);
Assortitivity(k)=*Assortitivity*(k)+*Ass.*(i,j);

inter_collab(*i*,*j*)= *inter_collab*(*i*,*j*)+1;

cite(*i*,*j*)= *cite*(*i*,*j*)+1;

Step4: *Assortitivity*(*k*)= *Assortitivity*(*k*)/*N*;

Step5: *IC*(*k*)= *inter_collab*(*i*,*j*)/*N*;

Step6: *C*(*k*)= *cite*(*i*,*j*)/*N*;

Step7: For *i*=1 to *N*,

RG(*k*)= *RG*(*k*)+ *RG*(*i*);

Step8: *SI*(*k*)= *publications*(*k*)/*authors*(*k*);

Step9: *RANK*(*k*)= *Assortitivity*(*k*)+ *IC*(*k*)+ *C*(*k*)+ *RG*(*k*)+ *SI*(*k*);

Step10: If *k*<=22, then go to Step2;

End

The proposed approach is based on concepts inspired by sociology. These concepts can be as follows:

a) *Assortitivity* or sometimes called *Homophily*: This concept refers to the tendency for people to have (non-negative) ties with people who are similar to themselves in socially significant ways. In the context of this work, this concept means that authors tend to collaborate with others who have similar features such as affiliation and discipline. The value of assortitivity is calculated for each pair of authors in each university considered in this work separately aiming at having the average value of assortitivity for further use it in the proposed rank. The assortitivity is represented as a vector that includes the following information:

1. Affiliation:

$$aff(i,j) = \begin{cases} 1, & if\ the\ authors\ i\ and\ j\ work\ in\ the\ same\ university \\ 0, & otherwise \end{cases}$$

2. Discipline:

$$disc(i,j) = \begin{cases} 1, if\ the\ authors\ i\ and\ j\ have\ the\ same\ discipline \\ 0, otherwise \end{cases}$$

3. Citations:

$$cite(i,j) = \begin{cases} 1, if\ the\ authors\ i\ and\ j\ have\ cited\ at\ least\ one\ same\ publication \\ 0, \quad otherwise \end{cases}$$

4. Collaboration Frequency:

$$collaboration(i,j) = \begin{cases} the\ number\ of\ publications\ that\ the\ authors\ i\ and\ j\ worked\ together \\ 0, \quad otherwise \end{cases}$$

5. International Collaboration:

$$inter\_collab(i,j) = \begin{cases} 1, if\ both\ i\ and\ j\ co-authored\ a\ publication\ with\ international\ authors \\ 0, \quad otherwise \end{cases}$$

The formula of assortitivity for each pair of authors in IUCN for a university *k* can be as follows:

$$Ass.(i,j)_k = Aff(i,j) + disc(i,j) + cite(i,j) + collaboration(i,j) + inter\_colab(i,j) \quad \textbf{(3.1)}$$

After calculating the assortitivity for all the pairs in IUCN, the average assortitivity for each university (k) is calculated as follows:

$$\text{Assortitivity k} = \sum_{i,j=1}^{N_k} \text{Ass. (i, j)}_k / N_k \quad \textbf{(3.2)}$$

Where N_k is the number of authors in the university k.

b) *International Collaboration* (IC_k): This indicator is important to be included because it reflects how colorful the experience of a university in terms of scientific collaboration with worldwide universities. And can be formalized as follows:

$$IC_k = \sum_{i=1,j=1}^{N_k} inter\_collab(i,j)_k / N_k \qquad \textbf{(3.3)}$$

Where N_k is the number of authors in the university k.

c) *University Citations* (C_k): The proposed ranking includes the number of total citations for each university. This indicator is important since it reflects the impact of a university on the research communities around the world and can be formulated as follows:

$$C_k = \sum_{i=1,j=1}^{N_k} cite(i,j)_k / N_k \qquad \textbf{(3.4)}$$

Where N_k is the number of authors in the university k.

d) *RG Score* (RG_k): This score is issued by ResearchGate academic social network. Each university around the world has an RG Score based on the activities of its authors. These activities can be the number of published research articles, the number of projects, the number of interactions of the researchers, the amount of collaboration in terms of answering the questions of the other researchers, and the number of scientific contributions that were uploaded to the platform. This indicator is very important insofar as it contributes to reflecting the actual interactions of researchers with worldwide researchers. Moreover, this feature cannot be found on other platforms such as Google Scholar or other academic websites and can be calculated as follows:

$$RG_k = \sum RG_{ik} \qquad \textbf{(3.5)}$$

where *i* is an author in a university *k*.

e) *Scopus Indicator* (SI_k): This is also another important indicator that is added to the proposed university rank. It is calculated from the ratio of the number of documents that a particular university has in the Scopus repository to the number of authors registered in Scopus. This ratio reveals the impact of a particular university in worldwide research communities. Also, this indicator reflects how active a university compared to the other Iraqi universities. SI for a university k is calculated as follows:

$$SI_k = \frac{Number\ of\ publications\ of\ k}{Number\ of\ authors\ in\ k} \qquad \textbf{(3.6)}$$

Now, the formula of the proposed rank for the Iraqi universities is calculated as follows: (which is the collected values of the aforementioned indicators).

$$RANK_k = Assortitivity_k + IC_k + C_k + RG_k + SI_k \qquad \textbf{(3.7)}$$

Chapter Four

Results and Discussion

Chapter Four
Results and Discussion

4.1 Introduction

This chapter presents the results that were obtained from the generated network (IUCN). It also discusses all the obtained results and the actual scientific facts on scientific research in Iraqi universities.

4.2 Hardware and Software Requirements

The hardware requirements to perform the visualization process was; CPU Intel Core i7-9700TE, with a frequency of 3.8GHz, 8 cores, L2 Cache 8 × 256 KB, and L3 Cache 12M, and RAM of 64 GB. The Windows 10 operating system was used for the visualization. For the sake of data collection, a crawler program is designed using the R programming language. This crawler uses the Google Scholar service for crawling the articles that have been published by researchers from the Iraqi universities. The collected data was related to the scientific contributions that have been published since the year 2000 till the date of collecting the data (February 2020). The execution of the crawler has been done by using the Unix operating system. Network visualization was performed using the Gephi Visualization tool. Finally, Microsoft Office Word application was used for writing this thesis.

4.3 IUCN Network Visualization and Evaluation

The visualization of the IUCN network is shown in Figure (4.1). The network shown represents the giant component of the Iraqi Universities Citation Network. As mentioned, each node represents a publication by an Iraqi author, and the relations among them show the citations among the publications. It can be observed many clusters with different colors that reflect different disciplines within the Iraqi universities. These clusters

represent the research groups of the Iraqi universities. The biggest clusters belong to the biggest universities in Iraq.

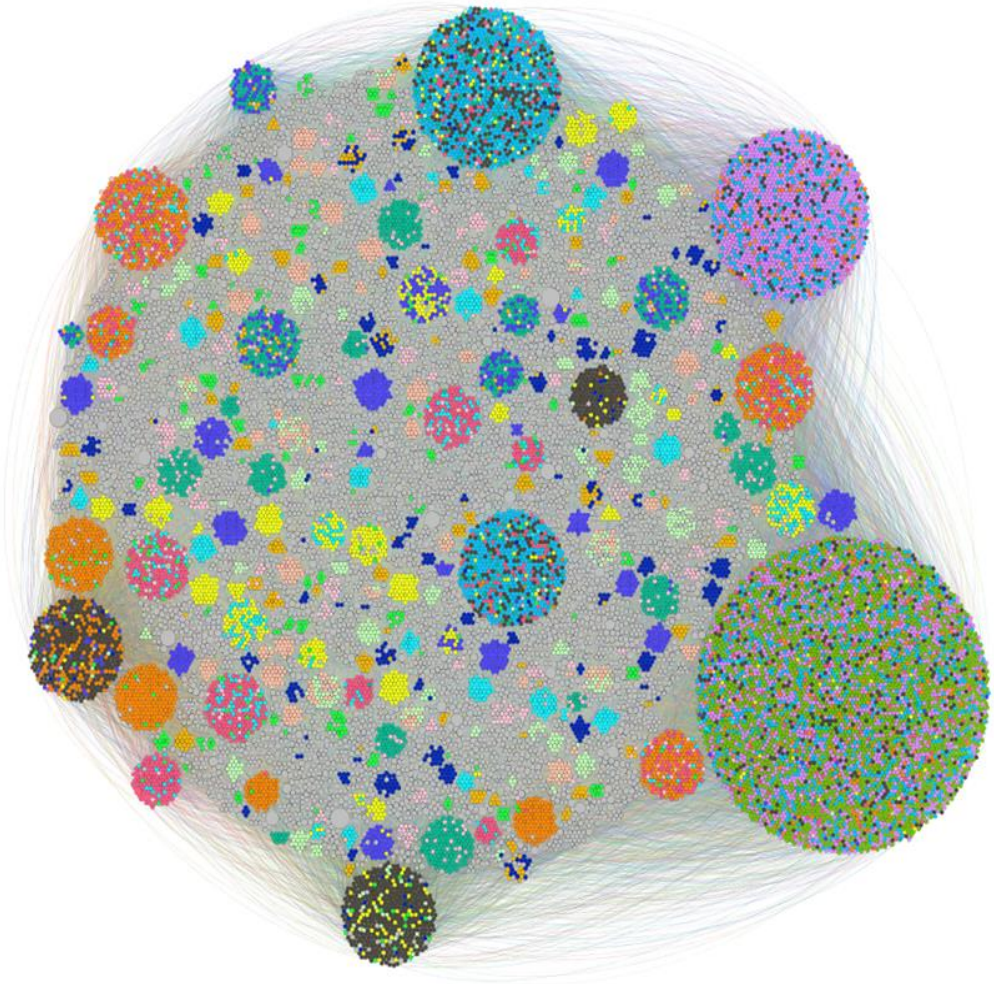

Figure 4.1 The visualization of the Iraqi Universities Citation Network (IUCN)

The main characteristics of the IUCN network can be summarized as follows (see Table 4.1):

1- *Average Degree* D_{IUCN}: IUCN network showed that the average number of publications of the Iraqi researchers was *3.671*, which is low compared to authors from international universities.

2- *Diameter O*: The distance between the farthest authors in IUCN is *15*, which is acceptable compared to similar networks. The diameter of a citation network is significantly affected by the level of scientific collaboration among the authors in that network.

3- *Density D*: the density level reflects an acceptable level of citations among the Iraqi authors and their publications.

4- *Girvan-Newman Modularity level*: This level reflects how strong the extracted communities when using the Girvan-Newman algorithm. In IUCN, this level was 0.752, which is considered a positive indicator of the scientific collaboration among the Iraqi authors.

5- *Average clustering coefficient* A_{CO}: It was 0.322 in the IUCN network; this means the tendency of the Iraqi authors to work in research groups is not strong, which means the probability of the current groups being expanded is not high.

6- *Average path length l*: The average distance between any given pair of authors in IUCN is 4.571. This value is acceptable taking into consideration the diameter of the network.

Table (4.1) The characteristics of the IUCN network

No. of nodes	No. of Edges	D_{IUCN}	*O*	*D*	*Modularity (Girvan-Newman)*	A_{CO}	*l*
25834	167267	3.671	15	0.01	0.752	0.322	4.571

Figure (4.2) shows the degree distribution of IUCN follows a power-law distribution. This feature is the most common feature when it comes to

citation networks. As mentioned in Chapter Two, the networks under this specific distribution are called *scale-free networks*. These networks have an important characteristic, which is *preferential attachment*. This means the fresh Iraqi authors prefer to collaborate and co-author articles with the highly connected authors (who published a high number of articles) in their research communities. This phenomenon is useful and contributes to enriching young researchers with experience from senior researchers. Moreover, the power-law distribution of IUCN indicates that there are few influential researchers compared to the total number of researchers in the Iraqi universities (see the left part of the distribution in Figure 4.2).

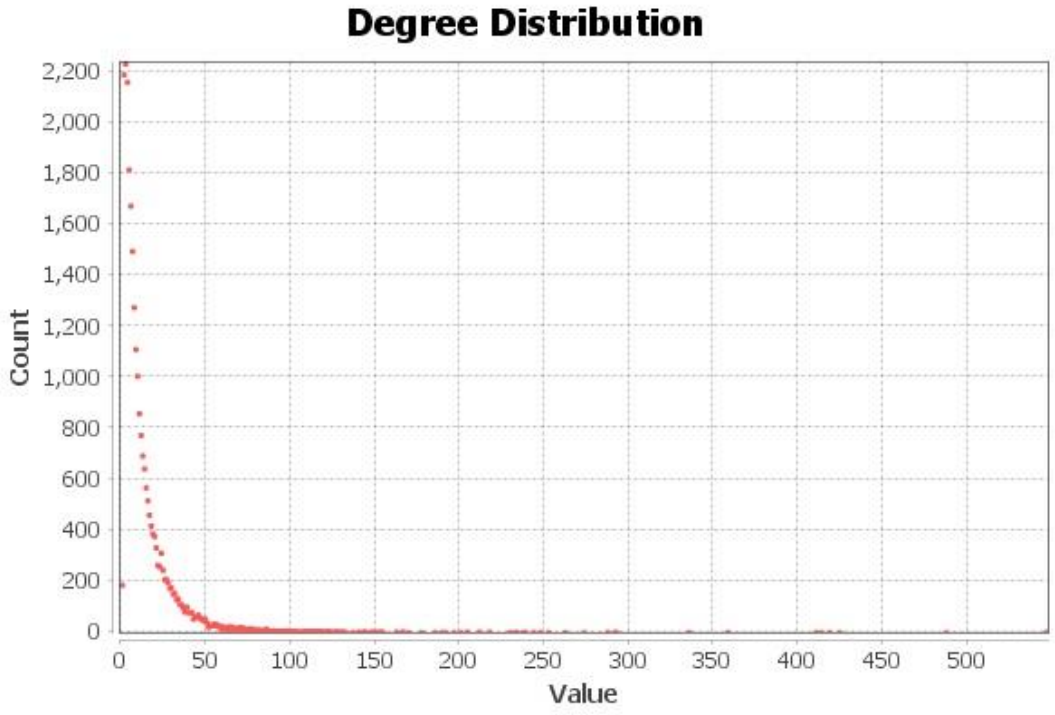

Figure 4.2 The degree distribution of the IUCN network (follows a power-law)

4.4 Resulted Numbers and Facts

Based on the collected dataset, main facts regarding scientific research in Iraqi universities have been extracted. Some of these facts were retrieved from ResearchGate academic social network. Table (4.2) presents the number of articles, RG score, number of authors verified in Google scholar, the actual number of authors, and Google Scholar citations for each

university considered in this thesis. It can be observed that the University of Baghdad outperformed the other Iraqi universities in terms of citations and RG score.

Table (4.2) Resulted Numbers and facts on scientific research in the Iraqi universities

University	articles	RG score	No. of authors	Actual Number of authors	Google Scholar Citations
University of Baghdad	8355	11,298.12	5070	6321	67883
University of Mosul	2805	4,593,54	3030	4180	40105
University of Basrah	2193	6,065.99	2861	2886	43971
University of Diyala	519	1,900.97	1244	1376	13741
University of Kufa	1062	3,540.40	2086	2144	26702
University of Al-Nahrain	1131	1,845.50	739	1061	4108
Tikrit University	692	1,699.24	1559	1978	17396
University of Technology	1211	4.118.39	1657	1573	45753
University of Al-Mustansiriyah	544	3,320.64	1977	3300	7233
University of Karbala	346	1,805.78	1097	1337	19222
University of Wasit	168	894.73	593	400	2669
University of Babylon	1359	3,986.94	2306	1998	49631
University of Thi-Qar	349	975.51	762	971	19326

University of Al-Muthana	165	925.56	615	535	237
University of Al-Anbar	755	2,308.17	1377	1674	4629
University of Kirkuk	428	738.51	399	813	19592
University of Al-Qadisiyah	393	2,965.07	1307	1432	21270
University of Samarra	172	137.42	92	338	675
Southern Technical University	51	245.93	218	535	719
Northern Technical University	98	780.88	675	836	3558
Middle technical University	261	1,235.71	792	1515	3186
Al-Furat Al-Awsat Technical University	201	1,347.53	571	989	1106
Total	23258		31027	38192	412712

Table (4.3) presents the scientific situation in Iraq compared to the neighbored countries. It is clear that most of these countries outperformed Iraq in terms of scientific research. This is in fact due to the unstable situation in Iraq and the side effect of the recent wars. In fact, a lot of work and effort is needed which may take several years to retrieve the Iraqi leading role in scientific research in the region and the world.

Table (4.3) The scientific research in Iraq compared to the neighbored countries

Country	H-Index	Total Number of Papers	Total Citations	Average Citation per cited paper
Turkey	443	1,279,318	14,039,162	10.97
Iran	329	581,253	5,509,736	9.48
KSA	361	422,538	5,169,638	12.23
Jordan	176	91,768	932,958	10.17
Iraq	109	82,978	385,306	4.571
Kuwait	162	54,494	635,924	11.67
Syria	113	16,020	217,694	13.59

Table (4.4) shows the performance of scientific research for each discipline in the Iraqi universities that has been verified under Scopus repositories. This table presents the h-index, number of papers, citations, and average citations for each paper published. It can be seen that the field of Medicine is the activist field of research in Iraq in 85 of the h-index. Also, the average citations per article in Medicine is 8.52, which is promising. The Engineering field also has an influential h-index of 60, which is the second activist discipline in Iraq. However, the average citations per paper are 3.81, which is not high compared to the h-index. The fields of humanities, economics, and business administrations did not gain acceptable h-index values. Therefore, these fields need more attention from their research communities.

Table (4.4) The performance of scientific research in the Iraqi universities for each discipline

Discipline	H-Index	Papers	Citations	Average Citation per paper
Medicine and Health Science	85	15598	132934	8.52
Engineering	60	24098	91900	3.81
Chemistry	55	8462	47742	5.64
Physics	52	11590	50950	4.40
Pharmacy	48	8072	22852	2.83
Computer Science	40	14324	34944	2.44
Agriculture	38	8592	23896	2.49
Veterinary	21	1600	5772	3.61
Humanities	16	1196	2984	2.49
Economics and Business administrations	12	406	1614	3.98
Note: The number of papers for each discipline interfered with each other since there is a collaboration among disciplines.				

4.5 Performance of Authors in Iraqi Universities

This section presents the most influential authors along with their citations and discipline. It starts with the University of Baghdad as shown in Table (4.5). According to the results in this table, the most dominated authors were from the Medicine field followed by a computer science author and other pure science authors. Moreover, the first-ranked author at the University of Baghdad gained the highest number of citations among all the Iraqi authors. However, there is a huge difference in citations between the first-ranked author and the second one.

Table (4.5) Top 10 highest citation authors at the University of Baghdad

Rank	Discipline	Citations
1st	Medicine	15566
2nd	Medicine	3000
3rd	Computer Science	1432
4th	Physics	1076
5th	Medicine	1011
6th	Medicine	870
7th	Chemistry	865
8th	Biology	829
9th	Mathematics	801
10th	Biomedical Engineering	772

Table (4.6) presents the top 10 authors in the second largest university in Iraq, which is the University of Mosul. It is clear that Chemistry discipline authors have dominated the list especially the rank of first who has gained 7002 citations among all the authors in the university. It can be observed that the Medicine discipline is not shown in the list of top 10 highest citation authors. However, the list is colorful of authors from different disciplines whose contributions gained high citations.

Table (4.6) Top 10 highest citation authors at the University of Mosul

Rank	Discipline	Citations
1st	Chemistry	7002
2nd	Chemistry	959
3rd	Statistics	611
4th	Electrical Engineering	567
5th	Chemistry	561

6th	Veterinary Medicine	540
7th	Administration and Economics	532
8th	Renewable Energy	415
9th	Computer Science	405
10th	Civil Engineering	387

The top 10 list of authors who have gained the highest citations in the third-largest university in Iraq is shown in Table (4.7). In the University of Basrah, the Physics discipline authors obtained the highest citations compared to the other disciplines. Also, the Marine Science discipline is on the list, which is a positive indicator of the future of this science. It should be mentioned that Marine science exists only at the University of Basrah.

Table (4.7) Top 10 highest citation authors at the University of Basrah

Rank	**Discipline**	**Citations**
1st	Physics	2008
2nd	Physics	1475
3rd	Mechanical Engineering	1453
4th	Physics	1085
5th	Marine Science	1076
6th	Biology	884
7th	Ecology	862
8th	Medicine	771
9th	Physics	720
10th	Agriculture	638

The performance of authors in terms of citations in the University of Diyala is depicted in Table (4.8). It can be seen that Chemistry discipline authors gained the highest citations among university authors with the absence of the Medicine field and health science. This university seems to be active in the Pure Sciences and Engineering fields of research.

Table (4.8) Top 10 highest citation authors at the University of Diyala

Rank	Discipline	Citations
1st	Chemistry	1463
2nd	Physics	847
3rd	Physical Chemistry	443
4th	Chemistry	342
5th	Civil Engineering	297
6th	General Science	294
7th	Civil Engineering	278
8th	Production and Manufacturing Engineering	268
9th	Architectural Engineering	239
10th	Chemistry	220

Table (4.9) shows the performance of authors in terms of citations at the University of Kufa. It is clear that three of the ten-highest citation authors are from the field of Medicine. Also, this university seems to be active in the fields of Engineering and Pure Sciences. However, the list includes colorful contributions in various disciplines.

Table (4.9) Top 10 highest citation authors at the University of Kufa

Rank	Discipline	Citations
1st	Medicine	3590
2nd	Signal Processing	2346
3rd	Renewable Energy	1450
4th	Mechanical Engineering	583
5th	Chemistry	443
6th	Medicine	436
7th	Mechanical Engineering	384
8th	Medicine	379
9th	Physics	371
10th	Pharmacy	363

At the University of Al-Nahrain, the performance of the authors in terms of citations is shown in Table (4.10). The Engineering and Chemistry disciplines' authors have obtained the highest citations among the university authors. The University of Al-Nahrain is very active in working on these two disciplines. The first ranked author is from Mechanical Engineering, which is interesting.

Table (4.10) Top 10 highest citation authors at the University of Al-Nahrain

Rank	Discipline	Citations
1st	Mechanical Engineering	1821
2nd	Medicine	292
3rd	Electronic Engineering	281
4th	Chemistry	273
5th	Biomedical Engineering	238

6th	Chemistry	197
7th	Chemistry	192
8th	Civil Engineering	192
9th	Chemistry	178
10th	Chemistry	133

The performance of authors in terms of citations at the University of Tikrit is depicted in Table (4.11). The most interesting observation is that an Arabic language author gained the highest rank out of all the disciplines in this university. This is, in fact, very promising since this discipline was the activist discipline in Iraq compared to all the Arab countries. However, Chemistry discipline authors gained the highest citations among the university authors.

Table (4.11) Top 10 highest citation authors at the University of Tikrit

Rank	**Discipline**	**Citations**
1st	Arabic Language	1931
2nd	Renewable Energy	1262
3rd	Electrical Engineering	1008
4th	Chemistry	930
5th	Medicine	883
6th	English Language	663
7th	Chemistry	597
8th	Chemistry	495
9th	Business and Administration	465
10th	Chemical Engineering	422

The performance of authors in terms of citations at the University of Technology is presented in Table (4.12). It can be seen that this university is specialized in technologies; therefore, the highest citations were gained from Science and Engineering discipline, which is expected. It should be mentioned that the University of Technology is considered one of the most influential universities in Iraq.

Table (4.12) Top 10 highest citation authors at the University of Technology

Rank	Discipline	Citations
1st	Renewable Energy	5606
2nd	Material Science	1274
3rd	Electronics Engineering	1123
4th	Biotechnology	1062
5th	Civil Engineering	1056
6th	Electrical Engineering	1045
7th	Environmental Science	1031
8th	Membrane Science	1030
9th	Laser and Optoelectronic Engineering	897
10th	Chemistry	882

The performance of authors in terms of citations at the University of Kerbala is depicted in Table (4.13). It is clear that Pharmacy discipline authors gained the highest citations among university authors with the absence of the Medicine field. The most interesting observation on this university is that the Nanotechnology discipline is on its top 10 list. This is also very promising since this discipline is considered as a new trend in scientific research these days. The other interesting thing is that the Teaching Methods discipline is also on this list, which is also promising

because this is one of the most important disciplines that are related to the recent techniques used in teaching.

Table (4.13) Top 10 highest citation authors at the University of Kerbala

Rank	Discipline	Citations
1st	Pharmacy	2005
2nd	Physics	891
3rd	Renewable Energy	619
4th	Chemistry	584
5th	Pharmacy	543
6th	Medical Science	443
7th	Nanotechnology	438
8th	Physics	401
9th	Mechanical Engineering	381
10th	Teaching Methods	308

The University of Babylon is considered as one of the most influential universities in Iraq. Therefore, its top 10 list includes colorful disciplines (Science and Engineering). Table (4.14) shows the list of top-cited authors along with their disciplines. As can be seen, there is no dominant discipline at this university, which means there is a strong competency among the disciplines, which is a positive indicator of the performance of this university in scientific research.

Table (4.14) Top 10 highest citation authors at the University of Babylon

Rank	Discipline	Citations
1st	Biotechnology	4678
2nd	Mechanical Engineering	1795

3rd	Physics	1584
4th	Environmental Science	1447
5th	Medicine	1034
6th	Civil Engineering	893
7th	Environmental Science	865
8th	Biology	831
9th	Computer Science	710
10th	Renewable Energy	606

The performance of authors in terms of citations at the University of Al-Muthanna is depicted in Table (4.15). Since this university is considered as a young university, its performance is significantly lower than the performance of the other Iraqi universities. The citation rate of the whole university needs more attention by the university officials. The highest author's citation is 81 in the Chemical Engineering discipline, and the second and third-ranked authors gained 51 and 50 respectively. The other citations were significantly decreased reaching two citations in the 10th rank.

Table (4.15) Top 10 highest citation authors at the University of Al-Muthana

Rank	**Discipline**	**Citations**
1st	Chemical Engineering	81
2nd	Construction Material	51
3rd	Physics	50
4th	Medicine	23
5th	Mathematics	13

6th	Chemistry	5
7th	Nursing	5
8th	Information technology	4
9th	Nursing	3
10th	Veterinary Medicine	2

The performance of authors in terms of citations at the University of Al-Anbar is depicted in Table (4.16). It is clear that Physics discipline authors gained the highest citations among university authors with the absence of the Medicine field and health science. This university seems to be active in the Pure Sciences and Engineering fields of research.

Table (4.16) Top 10 highest citation authors at the University of Al-Anbar

Rank	**Discipline**	**Citations**
1st	Physics	929
2nd	Structural and Material Engineering	751
3rd	Mechanical Engineering	742
4th	Physics	529
5th	Physics	464
6th	Mechanical Engineering	425
7th	Chemistry	407
8th	Networks and Communications	382
9th	Construction and Material Engineering	315
10th	Geophysics	275

The University of Samarra is one of the youngest universities in Iraq that was established in 2012. Therefore, it reflects a low performance in terms of the author's citations as depicted in Table (4.17). According to

Google scholar, there are only 7 authors who have been verified using their official educational domain. However, the list contains authors with excellent experience in Microbiology, Computer Science, and Mechanical Engineering disciplines.

Table (4.17) Top 7 highest citation authors at the University of Samarra

Rank	Discipline	Citations
1st	Microbiology	333
2nd	Computer Science	184
3rd	Mechanical Engineering	113
4th	Environmental Engineering	26
5th	Civil Engineering	9
6th	Transportation and Health	6
7th	Communication Engineering	4

The performance of authors in terms of citations at the University of Al-Mustansiriyah is depicted in Table (4.18). It can be observed that the Pharmacy discipline gained the highest citations among university authors. The list of this university is also fruitful of disciplines, which means all the disciplines are active and there exists a competency among them.

Table (4.18) Top 10 highest citations authors at the University of Al-Mustansiriyah

Rank	Discipline	Citations
1st	Pharmacy	705
2nd	Mathematics	522
3rd	Computer Engineering	349
4th	Electrical Engineering	312
5th	Dentistry	226

6th	Chemistry	206
7th	Pharmacy	184
8th	Physics	180
9th	Physics	165
10th	Medicine	149

The University of Wasit shows a medium level of citations as shown in Table (4.19). It is clear that the top 10 list of this university mainly includes Pure Science and engineering disciplines with the absence of the authors in the Medicine field and health science. This university seems to be active in the Pure Sciences and Engineering fields of research.

Table (4.19) Top 10 highest citation authors at the University of Wasit

Rank	**Discipline**	**Citations**
1st	Physics	583
2nd	Agriculture Science	552
3rd	Biology	497
4th	Physiotherapy	363
5th	Computer Science	291
6th	Civil Engineering	173
7th	Environmental Science	169
8th	Earthquake Engineering	156
9th	Mechanical Engineering	115
10th	Biotechnology	115

The performance of authors in terms of citations at the University of Thi-Qar is depicted in Table (4.20). It can be seen that the Pharmacy discipline dominates the citations of this university. Moreover, the 1st

ranked author at this university is one of the top 10 highest citation authors in the Iraqi universities.

Table (4.20) Top 10 highest citation authors at the University of Thi-Qar

Rank	Discipline	Citations
1st	Pharmacy	7549
2nd	Biomedical Engineering	1094
3rd	Pharmacy	602
4th	Mathematics	537
5th	Mechanical Engineering	531
6th	Medicine	405
7th	Arabic Language	330
8th	Petroleum Engineering	259
9th	Computer Science	241
10th	Computer Science	228

The performance of authors in terms of citations at the University of Kirkuk is depicted in Table (4.21). The most interesting fact on this university is the citation frequency of the 1st ranked author with 11676 in the field of Petroleum Engineering, which is on the top 10 list of the highly cited authors at the Iraqi universities.

Table (4.21) Top 10 highest citation authors at the University of Kirkuk

Rank	Discipline	Citations
1st	Petroleum Engineering	11676
2nd	Veterinary Medicine	1400
3rd	Physics	700
4th	Chemistry	482

5th	Biochemical	439
6th	Geochemistry	295
7th	Mechanical Engineering	281
8th	medicine	261
9th	Pharmacy	253
10th	Pharmacy	199

The University of Al-Qadisiyah is considered as one of the activist universities in Iraq compared to its established year in 2000. The performance of authors in terms of citations at this university is depicted in Table (4.22). It can be observed that the Medicine and Pharmacy disciplines dominate the list. The list also contains many different disciplines, which is a positive indicator of scientific research in this university.

Table (4.22) Top 10 highest citation authors at the University of Al-Qadisiyah

Rank	**Discipline**	**Citations**
1st	Medicine	5914
2nd	Pharmacy	3143
3rd	Biology	1098
4th	Ecology and Pollution	930
5th	Veterinary Medicine	823
6th	Statistics	563
7th	Electronics	383
8th	Chemistry	320
9th	Mathematics	283
10th	Material Engineering	269

The performance of authors in terms of citations at the Southern Technical University is shown in Table (4.23). It can be seen that the Computer Science discipline is at the head of the list in terms of the number of citations. However, the dominant discipline at this university is Engineering.

Table (4.23) Top 10 highest citation authors at the Southern Technical University

Rank	Discipline	Citations
1st	Computer Science	233
2nd	Thermal and Heat Engineering	76
3rd	Communications	49
4th	Renewable Energy	44
5th	Electronics and Communications	28
6th	Health and Medical Science	25
7th	Mechanical Engineering	18
8th	Control and Signal Processing	18
9th	Biology	16
10th	Thermal Mechanics	12

The performance of authors in terms of citations at the Northern Technical University is depicted in Table (4.24). It is clear that all the disciplines in the list belong to Engineering. Also, the list contains different disciplines with a little domination of Computer Engineering.

Table (4.24) Top 10 highest citation authors at the Northern Technical University

Rank	Discipline	Citations
1st	Construction Engineering	550

2nd	Chemical Engineering	399
3rd	Renewable Energy	289
4th	Computer Engineering	173
5th	Structural Engineering	159
6th	Power Engineering	137
7th	Mechanical Engineering	123
8th	Computer Engineering	112
9th	Computer Engineering	101
10th	Petroleum Engineering	99

The performance of authors in terms of citations at the Middle Technical University is depicted in Table (4.25). It can be seen that there is no dominant discipline in this list, which means all the disciplines work actively at this university.

Table (4.25) Top 10 highest citation authors at the Middle Technical University

Rank	**Discipline**	**Citations**
1st	Renewable Energy	1463
2nd	Environmental Technologies	355
3rd	Management Information Systems	273
4th	Environmental Technologies	247
5th	Physics	112
6th	Medical Technologies	96
7th	Renewable Energy	88
8th	Chemical Engineering	82
9th	Electrical Engineering	79
10th	Environmental Technologies	74

The performance of authors in terms of citations at Al-Furat Al-Awsat Technical University is depicted in Table (4.26). The list includes colorful disciplines and all work activities in this young university.

Table (4.26) Top 10 highest citations authors at Al-Furat Al-Awsat Technical University

Rank	Discipline	Citations
1st	Material Science	1341
2nd	Solar Energy	836
3rd	Material Science	693
4th	Chemical Engineering	517
5th	Health Community	278
6th	Parasitology	278
7th	Control Engineering	270
8th	Computer Engineering	236
9th	Mechanical Engineering	223
10th	Construction Engineering	154

Based on the results obtained in the aforementioned tables, the top 10 highest citation authors at the Iraqi universities are extracted. Table (4.27) shows the ranks along with their disciplines and the citation frequencies. Based on this table, it is clear that the Medicine discipline dominates the list. Moreover, the highest citations were obtained by an author from the University of Baghdad in Medicine with about 15566 citations to date. Also, the 10th rank in the list is an author from the University of Baghdad in the Medicine field. This means that this university is considered as a leading university in the field of Medicine, which is expected since the college of Medicine there is the best among all the Medicine colleges in Iraq. The second highly cited author is a

researcher from the University of Kirkuk in the field of Petroleum Engineering. The other ranks in the list are distributed to the other universities as shown in the table. It can be observed that 4 of the top 10 ranks are the authors' work in the Medicine discipline and 2 from the Pharmacy discipline. This means these two fields are the most influential compared to the other disciplines at the Iraqi universities.

Furthermore, according to these results, it can be observed that some of the biggest universities in Iraq are missing from the list such as the University of Basrah and the University of Al-Mustansiriyah. However, the performance of these universities still high compared to the other Iraqi universities (as will be seen later in this chapter).

Table (4.27) Top 10 highest citation authors by the Iraqi Universities

Rank	University	Discipline	Citations
1st	University of Baghdad	Medicine	15566
2nd	University of Kirkuk	Petroleum Engineering	11676
3rd	University of Thi-Qar	Pharmacy	7549
4th	University of Mosul	Chemistry	7002
5th	University of Al-Qadisiyah	Medicine	5914
6th	University of Technology	Renewable Energy	5606
7th	University of Babylon	Biotechnology	4678
8th	University of Kufa	Medicine	3590
9th	University of Al-Qadisiyah	Pharmacy	3143
10th	University of Baghdad	Medicine	3000

4.6 Iraqi Scientific Collaboration Network

4.6.1 Local Collaboration

This section describes the scientific collaboration among the Iraqi universities in terms of co-authoring research articles. Based on the

collected data from the Scopus repositories, Gephi software visualizes the collaboration network and shows its main characteristics. Figure (4.3) shows the visualization of the Iraqi Scientific Collaboration Network. The nodes in this network represent the universities considered in this thesis, while the edges represent the collaboration among the universities. It should be mentioned that there exists a relation between every pair of universities. However, the proposed approach considers the strongest level of collaborations among the Iraqi universities. In this context, the highest two collaborators are considered. For instance, the authors at the University of Diyala have scientific collaboration with authors from all the Iraqi universities, but the most frequent collaborators are considered (top-two universities). Figure (4.3) shows nodes with no edges, this means the top collaborators are international such as the University of Kirkuk. The nodes with one edge mean that one of its top-two collaborators is international. Nodes with three edges mean that two of the collaborators have the same level of collaborations. Moreover, the latitude and the longitude of the universities are used to project each to its coordinates on the map of Iraq using the Geo layout visualization. Based on the visualization, these communities can be described as follows:

- **North Community**: It is headed by the University of Mosul and followed by Northern Technical University, and the University of Tikrit.
- **South Community**: It is headed by the University of Basrah and followed by the Southern Technical University and Thi-Qar University.
- **Middle Community**: It is headed by the University of Baghdad and followed by the University of Al-Mustansiriyah, Al-Nahrain University, Middle Technical University, University of Diyala,

University of Technology, Al-Anbar University, Tikrit University (interfered), Samarra University, and the University of Wasit.

- **Middle South Community**: It is headed by the University of Babylon and followed by the University of Kerbala, University of Kufa, Al-Furat Al-Awsat Technical University, Al-Qadisiyah University, and Al-Muthanna University.
- **Blended Community:** This community combines the Middle community with the Middle South community.

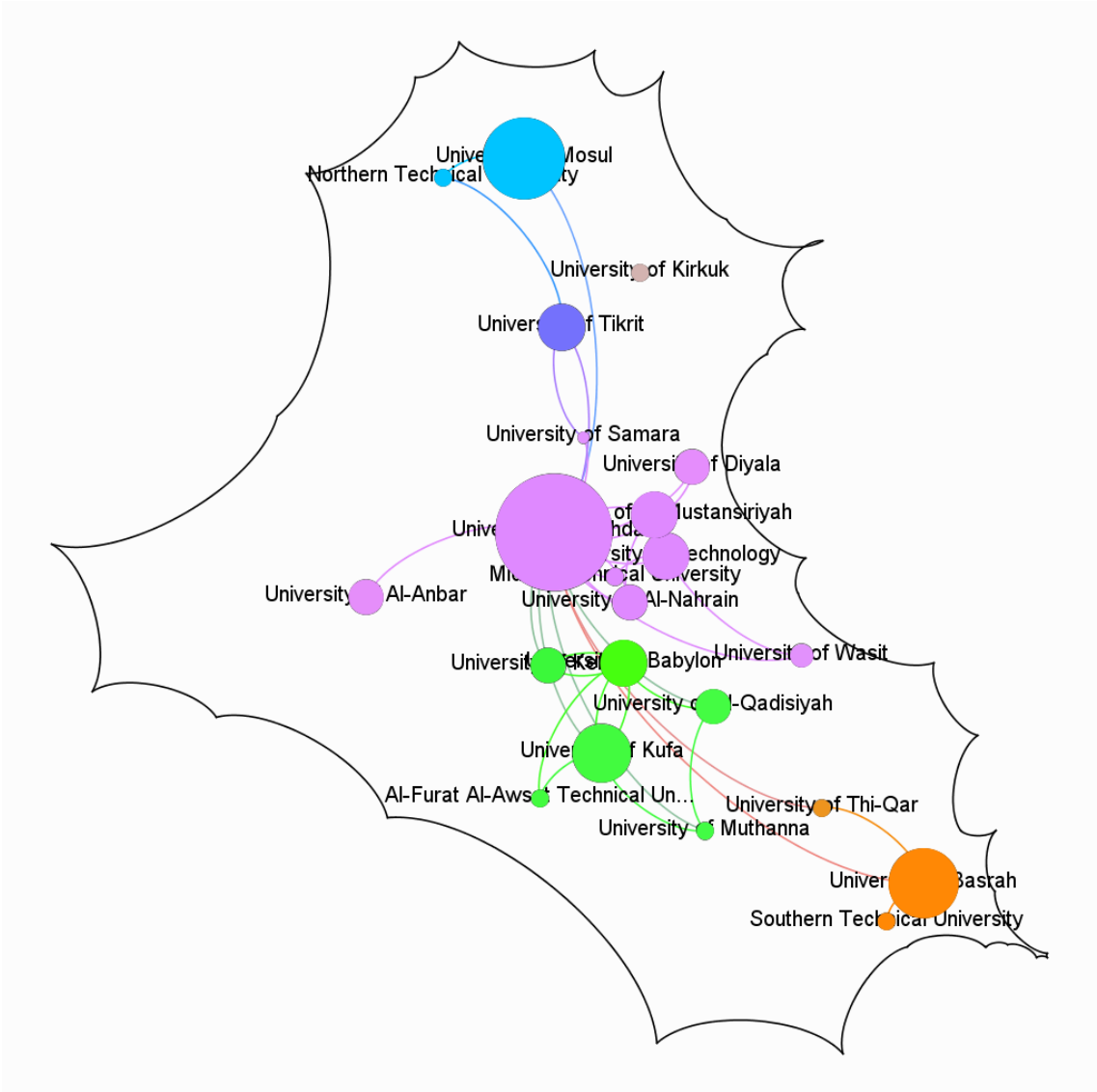

Figure 4.3 Visualization of the Iraqi Scientific Collaboration Network for the Iraqi Universities

According to the extracted communities, it is clear that the Middle community is the biggest and the most productive in Iraq. It should be mentioned that the University of Kirkuk does not belong to any of these communities according to the collected data from the Scopus repository. Therefore, it can be seen as an isolated node in the community. In fact, the University of Kirkuk has a strong collaboration with local universities when it comes to the local publishing in the Iraqi scientific journals.

Now, Table (4.28) presents the main characteristics of the network. The average degree of the generated network is 3.5, which is not strong taking into consideration the case that the degree from the University of Baghdad is 15 since it has collaborations with most of the Iraqi universities, and it is considered as the core center of the scientific research in Iraq. The diameter of the network is 4. It means to be between the farthest two universities in the network needs four edges. The density of the network is acceptable since only the strongest two collaborators are considered. The other interesting feature is that when using the Girvan-Newman clustering algorithm, five major communities are appeared with a weak modularity level of 0.376. This means the research communities in Iraq are clustered into five interfered communities.

Table (4.28) The characteristics of the Iraqi Scientific Collaboration Network

no. of nodes	no. of Edges	Average Degree	Diameter	Density	Modularity (Girvan-Newman)	Average Clustering coefficient	Average Path Length
22	39	3.5	4	0.169	5/0.376	0.543	2.133

According to the mentioned table, the average clustering coefficient is acceptable and there is an average tendency of the Iraqi universities to cluster together in one community. Finally, the average path length is 2.1,

which is also acceptable when it is needed to get from a university to another one.

4.6.2 International Collaboration

In the previous section, the collaboration among the Iraqi universities has been described. This section discusses international scientific collaboration at the Iraqi universities by using the same dataset that is used in the previous section. The visualization of the international collaboration is done based on the strongest levels (according to Scopus). Iraqi universities have scientific collaborations with many international universities around the world. This study, focuses on the strongest ones.

Figure (4.4) depicts international collaboration in scientific research. In the visualization, the latitudes and the longitudes of the countries have been determined using the Geo layout visualization and the Map of Countries in projecting the coordinates on the world map. Each node represents a country and the center node is Iraq since the looking is for collaborations with the Iraqi universities. The edges represent the existence of the collaboration. The weight of each edge reflects the level of collaboration. In this case, the weight is based on the frequency of collaborations. According to the results, the strongest level occurred with Malaysia. This is because a large number of the Iraqi researchers obtained their degrees from Malaysia and also there were a lot of postdocs positions were available in recent times. Moreover, the second strongest level was with UK universities, then with the United States. This is because of the scholarships of 2009-2014 that led to having many of the Iraqi researchers studying there. Iraqi universities also have strong collaborations with Turkey, Iran, China, Ukraine, and Jordan. The most successful collaborations were with Malaysia, the UK, the US, Turkey, and Iran. These countries have excellent records when it comes to scientific research.

Therefore, it is important to strengthen the level of collaboration with these countries aiming at strengthening scientific research in Iraq and obtaining better records in scientific research.

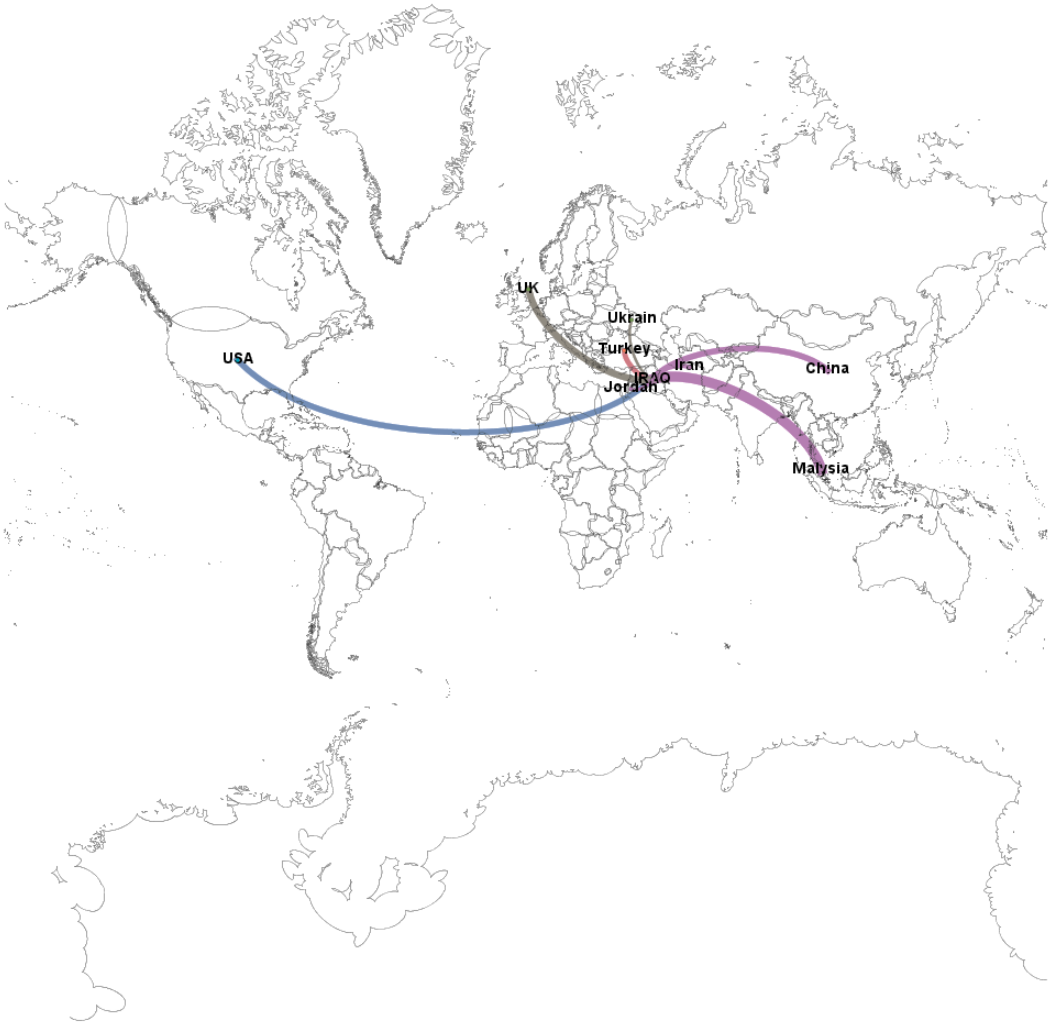

Figure 4.4 Visualization of the International Collaboration of the Iraqi Universities

4.7 The Iraqi Scientific Indicators Vs. World Indicators

4.7.1 World's Contributions

This section describes the position of scientific research in Iraq compared to the world for all the known-disciplines. Figure (4.5) shows the worlds' publications in all the disciplines. This figure is also used as a baseline of the Iraqi contributions and an indicator of the current status of

the scientific research in Iraq. Each node represents a publication in a discipline and different colors reflect different disciplines.

Figure 4.5 Visualization of the world contributions in all the disciplines [Scimago visualization tools]

4.7.2 Collaborations with International Authors

Figure (4.6) depicts the collaboration of the Iraqi authors with international authors in a particular discipline. Each node represents a contribution in a discipline, the color refers to the discipline of collaboration. Also, node size reflects the strength of collaboration (big nodes reflect strong collaboration). It should be mentioned that the background of the network (fuzzy) represents the world's contributions and the clear nodes in front are the nodes of the Iraqi contributions in this section. There are some interesting results shown in this graph. For instance, the fields of Neuroscience, Biochemistry, Genetics, and

Molecular Biology are represented as the biggest nodes in the network. This means that the level of collaboration in these fields is stronger than the other disciplines' levels of collaboration. Also, the fields of Mathematics and Computer Science have a strong level of collaboration with international authors. However, Figure (4.6) shows that the fields of Humanities and Arts have a lower level of collaboration. In this context, the authors in these disciplines need to focus more on working with the international authors aiming at filling this gap in the Iraqi scientific research. On the other hand, the authors of the Pure Sciences in the Iraqi universities work actively with international authors.

Figure 4.6 Visualization of the collaborations of the Iraqi authors with international authors in all the disciplines [Scimago visualization tools]

4.7.3 H-index Performance of the Iraqi Authors

As mentioned in Chapter Two, the h-index is an international strong indicator of the influence of a particular author/university. Figure (4.7) shows the performance of the Iraqi authors in terms of their h-index values. It can be seen that the disciplines of Biochemistry, Genetics, Molecular Biology, Immunology, and Microbiology have the highest h-index values (big nodes) compared to all the disciplines in Iraq. These results confirm what mentioned in Table (4.27). Moreover, the authors of Humanities disciplines show acceptable h-index values compared to their level of international collaboration. There is no doubt that if the international collaboration in these disciplines is improved, the h-index will be significantly improved.

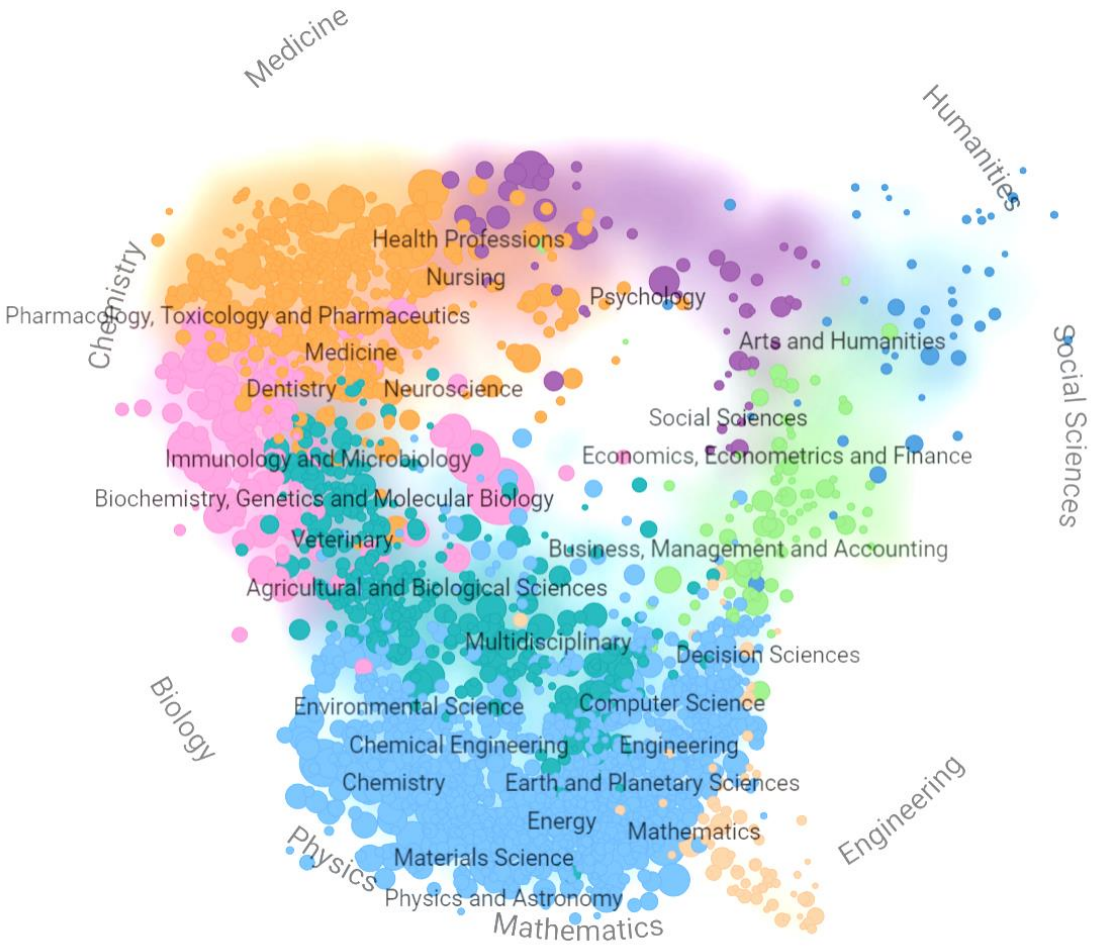

Figure 4.7 Visualization of the h-index of the Iraqi authors in all the disciplines [Scimago visualization tools]

4.7.4 Self-Citation Performance of the Iraqi Authors

Self-citation means an author cites his/her work. A high level of self-citation is considered as a negative indicator of a university/author. Figure (4.8) depicts some disciplines in the network show an extreme level of self-citation such as Biochemistry, Genetics, Molecular Biology, Immunology and Microbiology, Agricultural and Biological Science, Business Management, and Accounting. The Iraqi authors should be aware of this specific issue that can lead their h-index value to be decreased. This also can be a cause to drop a particular journal from a particular index repository such as Scopus and Clarivate. However, on many occasions, authors need to cite their work when it comes to using their previous methods for benchmarking or developing their previous work. Therefore, authors can cite their work only if it is necessarily needed.

Figure 4.8 Visualization of the self-citation of the Iraqi authors in all the disciplines [Scimago visualization tools]

4.7.5 Citation Performance of the Iraqi Authors

Figure (4.9) shows the performance of the Iraqi authors in terms of citations in all the disciplines. It is clear that the citations of the Iraqi authors are not high compared to the world's citations, although there exist some influential authors in Iraq.

Figure 4.9 Visualization of the citation performance of the Iraqi authors in all the disciplines [Scimago visualization tools]

4.8 Summary on the Collaboration of the Iraqi Universities

This section summarizes the Iraqi universities in terms of the top collaborators, number of documents in Scopus repositories, top subject area, and the top publishers. The universities are mentioned in the table based on their number of documents in Scopus. As can be seen in Table (4.29), the University of Baghdad is at the top of the list with 11784 documents and the top fields of research in the university is Medicine and Engineering. Most of the contributions of the university are published in the Plant Archives Journal and Iraqi Journal of Agricultural Sciences (top publishers). This information is mentioned for each university.

Table (4.29) Top collaborators and publishers for the Iraqi Universities

#	University	Top Collaborators	no. of Documents	Top Subject Areas	Top Publishers
1	University of Baghdad	- Al-Mustansiriya University - Al-Nahrain University	11784	Medicine Engineering	Plant Archives Iraqi Journal of Agricultural Sciences
2	University of Technology	- University of Baghdad - University of Al-Mustansiriyah	3972	Engineering Material Science	Iop Conference Series Materials Science and Engineering Aip Conference Proceedings
3	University of Babylon	- University of Kufa - University of Kerbala	3888	Engineering Medicine	Indian Journal of Public Health Research and Development Journal of Engineering and Applied Sciences
4	University of Al-Mustansiriyah	- University of Baghdad - University of Technology	3801	Engineering Medicine	Iop Conference Series Materials Science and Engineering Journal of Physics Conference Series
5	University of Basrah	- University of Baghdad - University Sains Malaysia	3518	Engineering Physics	Journal of Physics Conference Series Basrah Journal of Agricultural Sciences, Plant Archives

6	University of Mosul	- University Putra Malaysia - University of Baghdad	3441	Engineering Computer Science	Iraqi Journal of Veterinary Sciences Annals of The College of Medicine Mosul
7	University of Kufa	- University of Babylon - University of Baghdad	2692	Engineering Pharmacology, Toxicology, and Pharmaceutics	Plant Archives Indian Journal of Public Health Research and Development
8	Al-Nahrain University	- University of Baghdad - University of Al-Mustansiriyah	2087	Engineering Medicine	Iop Conference Series Materials Science and Engineering Arpn Journal of Engineering and Applied Sciences
9	University of Al-Anbar	- University of Baghdad - University Sains Malaysia	1607	Engineering Computer Science	Plant Archives Proceedings International Conference on Developments in Esystems Engineering Dese
10	University of Al-Qadisiyah	- University of Babylon - University of Baghdad	1594	Pharmacology, Toxicology, and Pharmaceutics Medicine	Journal of Physics Conference Series International Journal of Research in Pharmaceutical Sciences
11	Middle Technical University	- University of Baghdad - University of technology	1430	Engineering Material Science	Iop Conference Series Materials Science and Engineering Journal of Engineering and Applied Sciences
12	Tikrit University	- University of Baghdad - University Malaysia Pahang	1326	Engineering Material Science	Indian Journal of Public Health Research and Development Indian Journal of Forensic Medicine and Toxicology, Plant Archive
13	University of Diyala	- University of Baghdad - University of Al-Mustansiriyah	1215	Engineering Computer Science	Plant Archives Journal of Physics Conference Series
14	University of Kerbala	- University of Babylon - University of Baghdad	1097	Engineering Material Science	Iop Conference Series Materials Science and Engineering Indian Journal of Public Health Research and Development

15	Al-Furat Al-Awsat Technical University	- University of Babylon - University of Kufa	926	Engineering Computer Science	Iop Conference Series Materials Science and Engineering Plant Archives
16	University of Thi-Qar	- University of Baghdad - University of Basrah	881	Engineering Physics	Journal of Global Pharma Technology International Journal of Pharmaceutical Research
17	Al-Muthanna University	- University of Baghdad - University of Kufa - University of Al-Qadisiyah	669	Agricultural and Biological Sciences Engineering	Plant Archives Indian Journal of Public Health Research and Development
18	University of Kirkuk	- Manchester Metropolitan University - Erciyes Üniversitesi	611	Physics Material Science	Acta Crystallographica Section E Structure Reports Online Acta Crystallographica Section E Crystallographic Communications
19	University of Wasit	- University of Baghdad - University of Al-Mustansiriyah	590	Engineering Computer Science	Plant Archives Iop Conference Series Materials Science and Engineering
20	Northern Technical University	- University of Mosul - Tikrit University	581	Engineering Computer Science	Iop Conference Series Materials Science and Engineering Aip Conference Proceedings
21	Southern Technical University	- University of Basrah - Huazhong University of Science and Technology	314	Engineering Computer Science	Iop Conference Series Materials Science and Engineering Indian Journal of Public Health Research and Development
22	University of Samara	- Tikrit University - University of Baghdad	149	Engineering Agricultural and Biological Sciences	Iraqi Journal of Agricultural Sciences Journal of Engineering and Applied Sciences

According to the above table, it can observe the following:

- The top publishers of most of the Iraqi universities are similar. This means the Iraqi authors tend to publish their work in almost the same publishers frequently such as the Indian Journal of Public Health Research and Development. This phenomenon is not a good indicator of the Iraqi authors since there are hundreds of publishers around the world that can accept their work and eventually have colorful publishers and their works can reach more people around the world. Therefore, it is important to avoid publishing with the same publishers every time.
- The top field of research in all the Iraqi universities is in the Engineering field. In fact, this case needs to think about more seriously. It reflects that all the universities support the Engineering field rather than the other fields, which is a negative indicator of the trend of scientific research in Iraq.
- Some universities such as the University of Samarra have a little number of publications in Scopus and the high indexed repositories. This is a real issue that needs to be focused on by the university officials and the Ministry of Higher Education and Scientific Research in Iraq to provide more support to the authors of this university.

4.9 Node-Level Network Measurements for the Iraqi Authors

This section shows the results of using node-level measurements in the IUCN network. These measurements are degree centrality, clustering coefficient, betweenness centrality, and closeness centrality.

- **Degree Distribution:** The degree distribution of the IUCN follows a power-law distribution. This means the majority of authors have a low degree (fewer publications). While a few authors in IUCN have a high number of publications. This feature is common in Citation networks

and Collaboration networks. The degree distribution of the network is shown in Figure (4.2) in Section (4.3).

- **Clustering Coefficient:** In the context of IUCN work, this measurement reflects the tendency of IUCN authors to collaborate with each other and cluster together in one research community. Figure (4.10) depicts the distribution of the clustering coefficient values of the Iraqi authors. It is clear that there exist acceptable tendencies of the Iraqi authors to collaborate with each other, which is a positive indicator of scientific research in Iraq.

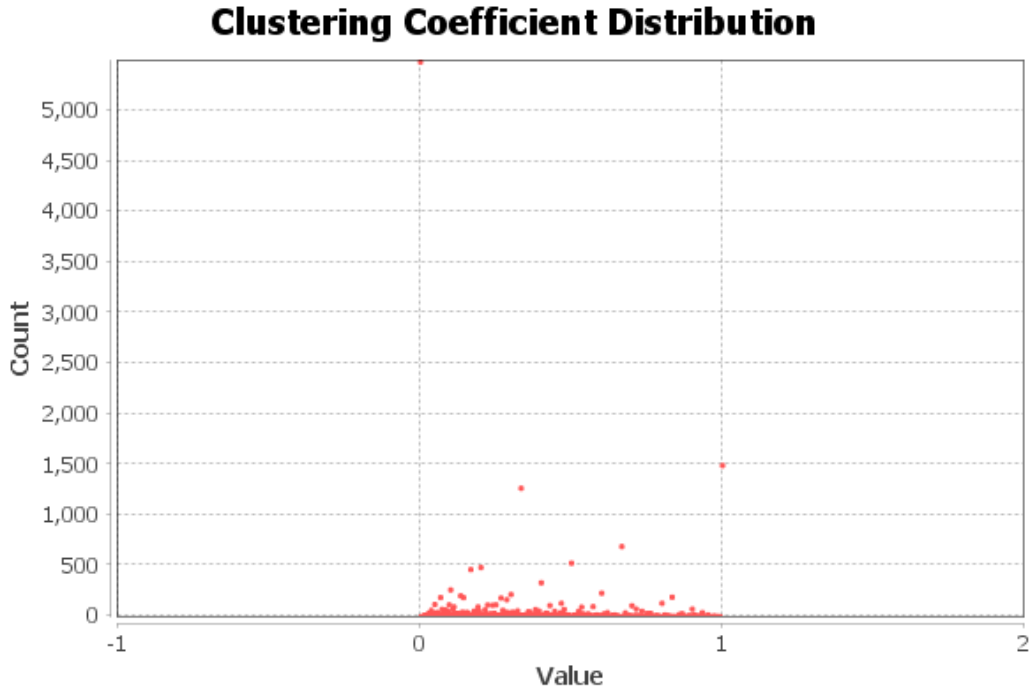

Figure 4.10 Distribution of the Clustering Coefficient of the Iraqi authors in IUCN

- **Betweenness Centrality:** This measurement reflects the importance of an author in connecting different communities. In other words, it shows the well-positioned authors that play a significant role in connecting the Iraqi research groups of the same university or different universities. Figure (4.11) shows the distribution of the betweenness centrality of IUCN. This figure shows that the majority of the Iraqi authors have

lower levels (less than 0.1) and very few influential authors have acceptable levels. Table (4.30) shows the top 10 highest betweenness centralities in IUCN.

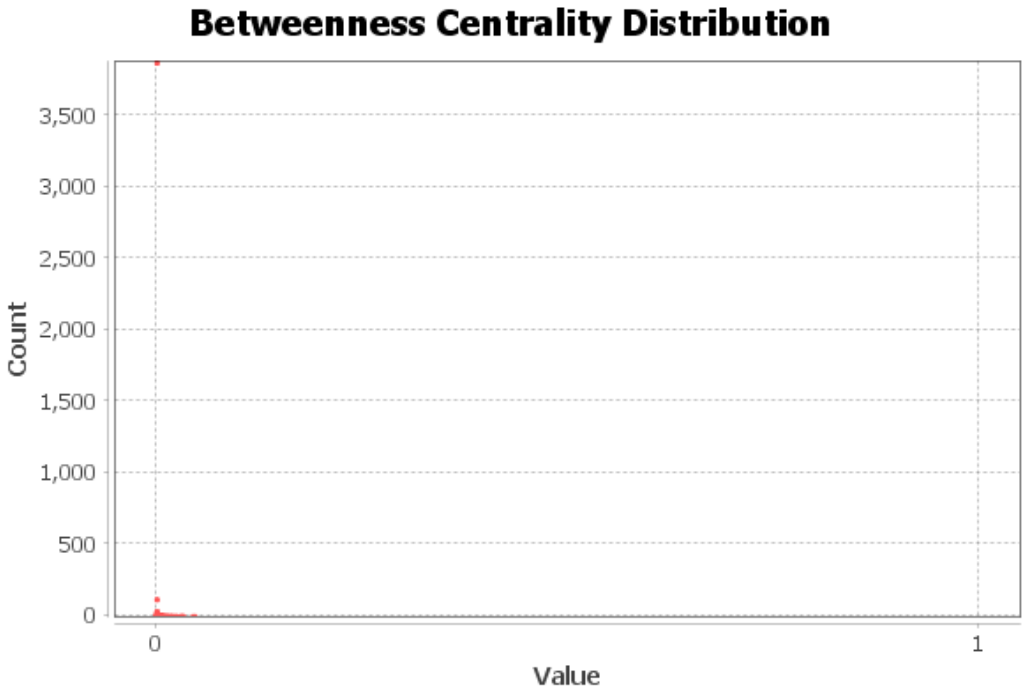

Figure 4.11 Distribution of Betweenness Centrality of the Iraqi authors in IUCN

Table (4.30) Top 10 highest betweenness centrality authors at the Iraqi Universities

Rank	University	Discipline	C_b	C_O
1st	University of Baghdad	Medicine	0.0459	0.0091
2nd	University of Babylon	Biotechnology	0.0437	0.0059
3rd	University of Mosul	Chemistry	0.0318	0.0098
4th	University of Baghdad	Medicine	0.0314	0.0094
5th	University of Al-Qadisiyah	Medicine	0.0305	0.0101
6th	University of Technology	Renewable Energy	0.0293	0.0077
7th	University of Kufa	Medicine	0.0286	0.0068
8th	University of Thi-Qar	Pharmacy	0.0240	0.0130
9th	University of Basrah	Physics	0.0226	0.0096
10th	Middle Technical University	Renewable Energy	0.0222	0.0089

According to the above table, the University of Baghdad has the most influential authors in terms of their positions in the IUCN network followed by authors from the universities of Babylon (Biotechnology), Mosul (Chemistry), Al-Qadisiyah (Medicine), Technology (Renewable Energy), Kufa (Medicine), Thi-Qar (Pharmacy), Basrah (Physics), and the 10th rank is for the Middle Technical University in the Renewable Energy, which is interesting.

- **Closeness Centrality:** Reflects how close the Iraqi authors from each other in the IUCN network. Figure (4.12) shows the distribution of the Iraqi authors in terms of closeness. It is clear that the majority of authors in IUCN are not close to each other and their closeness levels are below 0.3. While the minority of authors are very close to the other authors in the IUCN network and their closeness levels are 1, which is perfect.

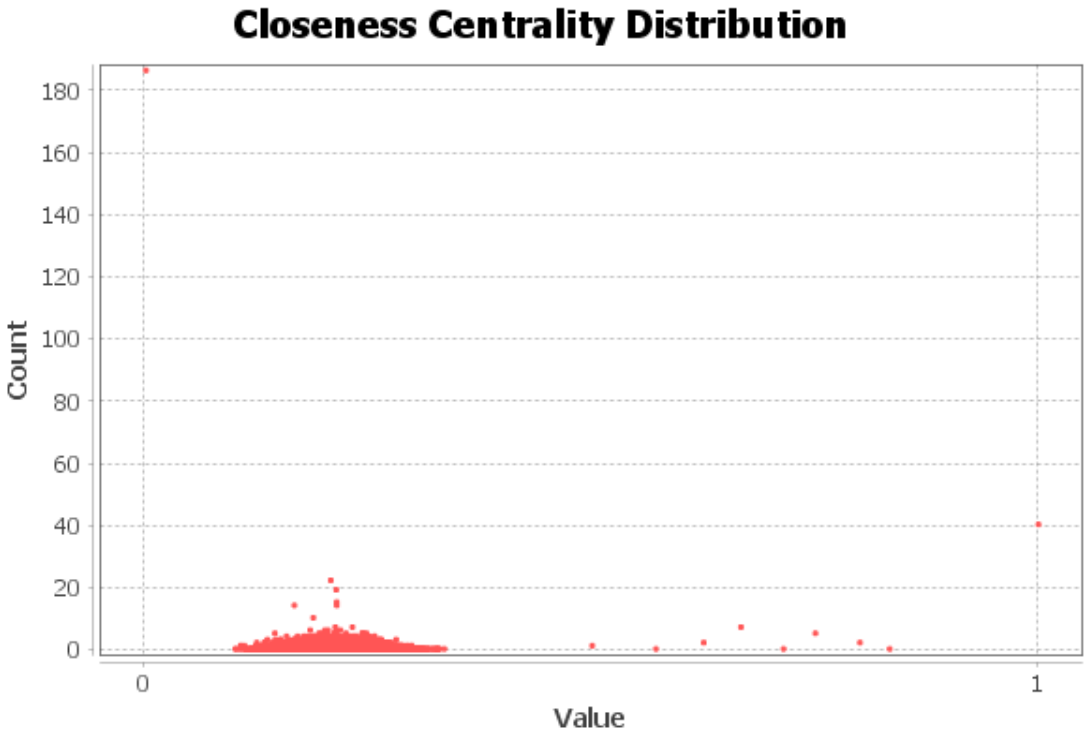

Figure 4.12 Distribution of Closeness Centrality of the Iraqi authors in IUCN

4.8 The Results of the Proposed Rank

This section describes the proposed approach in ranking the Iraqi universities according to what has been described in Chapter three. The proposed approach considers five indicators; the assortitivity level of a university, university international collaboration, university citations, university RG score, and Scopus indicator on the ratio of the number of publications to the number of authors. These indicators were described in detail in the previous chapter.

The first step in the proposed approach was to calculate the assortitivity for each pair of authors in IUCN. As mentioned, assortitivity means that authors tend to collaborate with others who have similar features. This indicator was important to be incorporated in the proposed rank since it digs deeply into the actual relations and the performance of each university in terms of scientific collaboration. To this end, the Equation (3.2) is used to calculate the assortitivity for each university. Figure (4.13) shows the distribution of assortitivity for all pairs of authors in the Iraqi universities. According to the figure, it can be seen that fewer pairs have high levels of assortitivity, while the majority have low levels.

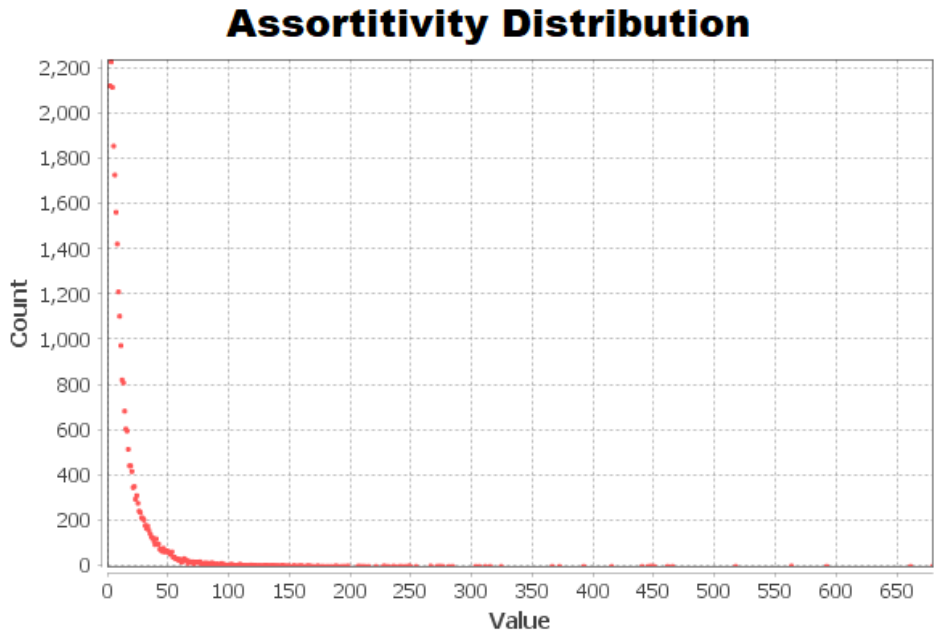

Figure 4.13 Distribution of the Assortitivity of the Iraqi authors in IUCN

In the proposed rank, the average assortitivity for each university is used. Also, other indicators have been included in the proposed rank aiming at having a rank for each university. Table (4.31) and Figure (4.14) show the Rank of each university based on the proposed approach.

The table also shows that only three universities (Baghdad, Al-Mustansiriyah, and Kufa) out of all the Iraqi universities have been listed in QS World University Rankings 2020. This is because QS ranking does not depend on scientific research only, but also other factors such as laboratories, physical space dedicated to each student, the quality of the graduated students, funds received, contents of the website, faculty members, etc. Webometrics is another organization that ranks universities based on their contents and the interactions on the website of a university. It can be seen that the University of Baghdad obtained the rank of first in 2020 and all the other universities considered in this work appeared in Webometrics with a specific rank. Also, the proposed rank has been benchmarked with the 4icu Rank. This rank uses indicators such as the university website size, university online resources, geographical classification, and contents of the website. Also, this rank considers all the Iraqi universities. The universities of Babylon, Baghdad, and Kufa obtained the highest ranks on the list.

Based on the aforementioned results, it can be observed that the proposed methodology reflects close results to what has been obtained from the mentioned benchmarking agencies. For instance, the University of Baghdad in all the ranks is considered as the best one in terms of the proposed approach and the benchmarking. However, the proposed approach takes into consideration only the scientific research aspects, while the other ranks consider other indicators that are related to universities' different aspects (e.g., website contents).

Table (4.31) The proposed rank of the Iraqi universities based on their performance in scientific research. The table also shows the ranks of the Iraqi universities based on three ranking agencies.

Proposed Rank	University	QS Ranking	Webometrics Ranking	4icu Ranking
1	University of Baghdad	1st	1st	2nd
2	University of Babylon	-	4th	1st
3	University of Basrah	-	7th	16th
4	University of Mosul	-	6th	8th
5	University of Technology	-	3rd	3rd
6	University of Kufa	3rd	5th	4th
7	University of Thi-Qar	-	17th	22nd
8	University of Al-Qadisiyah	-	9th	5th
9	University of Kerbala	-	15th	12th
10	University of Kirkuk	-	22nd	35th
11	University of Diyala	-	2nd	9th
12	University of Tikrit	-	12th	7th
13	University of Al-Anbar	-	10th	25th

14	University of Al-Mustansiriyah	2nd	8th	6th
15	Northern Technical University	-	38th	62th
16	University of Al-Nahrain	-	16th	21st
17	University of Wasit	-	19th	32nd
18	Middle Technical University	-	21st	37th
19	Al-Furat Al-Awsat Technical University	-	14th	18th
20	Southern Technical University	-	65th	45th
21	University of Al-Muthanna	-	36th	27th
22	University of Samarra	-	45th	52th

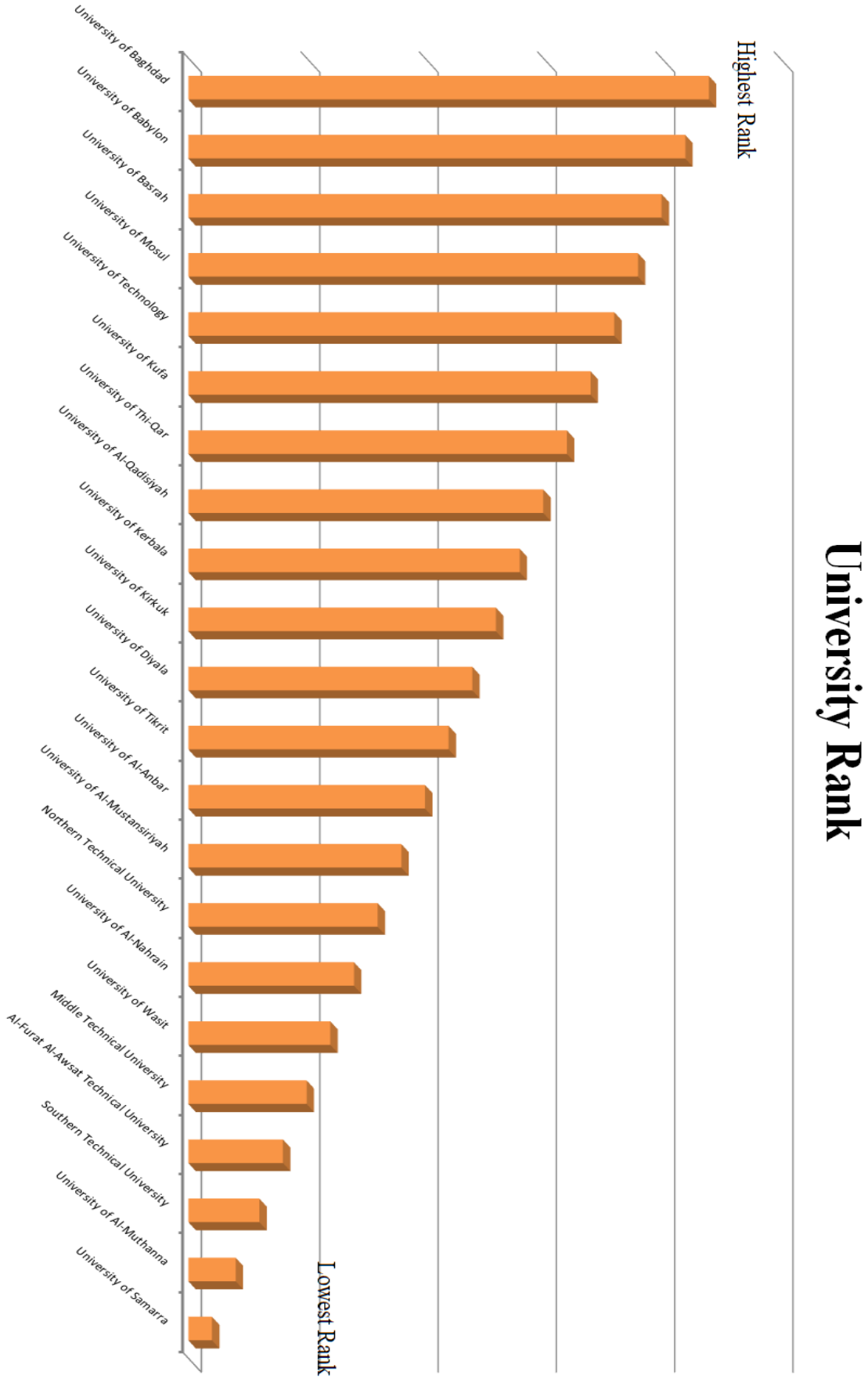

Figure 4.14 The rank of the Iraqi universities based on the proposed approach

4.11 Summary of Thesis Results

In this thesis, a comprehensive evaluation was performed based on the IUCN network measurements. The evaluation has been organized according to different aspects aiming at having a microscopic view of the current status of the Iraqi universities in terms of scientific research. These aspects can be summarized as follows:

- **Characteristics of the IUCN network:** The network reflects an on-average performance according to network-level measurements such as the average clustering coefficient and average path length. Also, the degree distribution of IUCN follows a power-law, which is common in this kind of network.
- **Numbers and Facts:** IUCN network showed that the University of Baghdad outperformed the other Iraqi universities in terms of the frequency of citations. Furthermore, the performance of the scientific research in Iraq underperforms the performance of the neighbored countries such as Turkey, Iran, and KSA in terms of h-index, the number of published contributions, total citations, and the average citation per contribution. In this regard, the Iraqi research communities have to spend more time and effort in order to compete with other countries. The results also showed that Medicine and Engineering disciplines are the activist fields of research in Iraq in terms of h-index, the number of published contributions, citations, and the average citation per contribution.
- **Performance of authors and universities:** IUCN network showed the most influential authors in Iraq in terms of the frequency of citations. In this regard, the top-cited author was from the University of Baghdad in the field of Medicine with about 15566 citations (to date). There were also other high influential authors from different disciplines (Engineering and Pure Sciences) and different universities in Iraq.

- **Scientific Collaboration:** In terms of local scientific collaboration, the collaboration network of the Iraqi universities showed five different research communities in Iraq, namely, the north community, the south community, the middle community, the middle south, and the blended community. This phenomenon reflects the fact that the geographical area of a university plays a significant role in establishing a scientific collaboration among universities. This, in fact, is a negative indicator since we live in a cyber world in which the geographical barriers are not related when it comes to scientific collaboration. On the other hand, the international scientific collaboration of the Iraqi universities showed a strong level of collaboration with some countries such as Malaysia, the UK, Turkey, Iran, the USA, Ukraine, and Jordan. The results also showed that the fields of Neuroscience, Biochemistry, genetics, and Molecular Biology fields have gained the highest levels of collaboration for the Iraqi authors with internationals. Moreover, the h-index of the aforementioned fields gained the highest values compared with the other disciplines in Iraq. However, these disciplines have a negative aspect in terms of a high level of self-citation.
- **International Publishing:** The Iraqi research communities showed a negative indicator in terms of publishing their contributions. They almost use the same publishers that they published with before. This leads to limit the readers to a specific audience, which is not desired since the Iraqi contributions should be disseminated widely and accessed by as many as international readers.
- **Network Measurements:** The clustering coefficient in the IUCN network showed an acceptable level and a tendency of the Iraqi authors to collaborate and create research groups in on-average. The betweenness centrality showed fewer influential authors who connect the research communities. This specific feature should be improved by

encouraging authors to collaborate with the most influential ones in their research communities or outside their research communities and eventually improve the quality of the published contributions. This is confirmed when observing the closeness centrality of the authors.

- **Ranking universities:** This study proposed a novel approach for ranking the Iraqi universities based on five indicators; assortitivity of university authors, international collaboration, citations, RG score, and Scopus indicator, which is the ratio of the number of publications to the number of authors in that university. The proposed rank put the University of Baghdad at the top of the list, followed by the University of Babylon, then the University of Basrah, then the other universities. Also, the rank has been benchmarked against the ranks of three ranking agencies namely, QS, Webometrics, and 4icu. The results showed close behavior since the University of Baghdad outperformed the other Iraqi universities in all the approaches used. However, the proposed rank is based on scientific research indicators only, while the other ranks have relatively used other different factors.

Chapter Five

Conclusions and Future Work

Chapter Five

Conclusions and Future Work

5.1 Conclusions

This thesis has presented a deep analysis of scientific research in the main Iraqi universities. The data was collected mainly from Google Scholar and Scopus repositories. The collected data was related to the scientific contributions that have been published since the year 2000 till the date of collecting the data (February 2020). Moreover, this thesis has considered the main twenty-two Iraqi universities as the case study. The approach that was used in this thesis was based on the concepts of Complex Networks. Therefore, the analysis approach was performed using network measurements at both node level and network level. The data collection process was performed using a crawler program that was designed for the purpose of this thesis. The output of the crawler was the dataset of this work. Based on the collected dataset, a network has been created in which the contributions (e.g., articles, papers, books, reviews) are represented as nodes and if there is a citation between two contributions, an edge is created between them and so on. The generated network represents the Giant Component of what has been called the Iraqi Universities Citation Network (IUCN).

Finally, we believe that this is the first kind of academic works that deeply investigates the scientific research in the Iraqi universities using data mining and network science approaches, which is of interest to both; the Iraqi universities and the ministry of higher education and scientific research in Iraq.

5.2 Future Work

In future work, the plan is to investigate IUCN using other network measurements such as Bridging Centrality and Eigenvector Centrality aiming at having a different perspective on the Iraqi authors in terms of how to strengthen the scientific collaboration among the Iraqi authors and the best strategies when it comes to initiating research groups. Another plan is to deeply investigate the impact of moving influential authors from their source universities to a target university and how this strategy can contribute to improving the quality of scientific research in the targeted university. This needs to use sophisticated network measurements and mobility patterns. However, this kind of work requires very fast computers in terms of hardware to perform the simulations smoothly.

References

REFERENCES

[1] S. M. Al-Najjar and M. K. Jawad, “Total Quality Management Practices and Impediments in the Arab Countries with Special Reference to Iraq,” *J. Law Soc.*, vol. 6, no. 1, pp. 86–96, 2019.

[2] M. E. J. Newman, “The structure of scientific collaboration networks,” *Proc. Natl. Acad. Sci.*, vol. 98, no. 2, pp. 404–409, 2001.

[3] J. Zhou, N. Cai, Z.-Y. Tan, and M. J. Khan, “Analysis of effects to journal impact factors based on citation networks generated via social computing,” *IEEE Access*, vol. 7, pp. 19775–19781, 2019.

[4] Y. Bu, “Understanding the Citation-based Impact of Scientific Publications Through Ego-centered Citation Networks.” Indiana University, 2020.

[5] A. P. Singh, K. Shubhankar, and V. Pudi, “An efficient algorithm for ranking research papers based on citation network,” in *2011 3rd Conference on data mining and optimization (DMO)*, 2011, pp. 88–95.

[6] S. Das Gollapalli and C. Caragea, “Extracting Keyphrases from Research Papers Using Citation Networks.,” in *AAAI*, 2014, vol. 14, pp. 1629–1635.

[7] M. A. Hasson, S. F. Lu, and B. A. Hassoon, “Scientific Research Paper Ranking Algorithm PTRA: A Tradeoff between Time and Citation Network,” in *Applied Mechanics and Materials*, 2014, vol. 551, pp. 603–611.

[8] J. Li and P. Willett, “ArticleRank: a PageRank-based alternative to numbers of citations for analysing citation networks,” 2009.

[9] X.-Y. Liu and B.-C. Chien, “Applying citation network analysis on recommendation of research paper collection,” in *Proceedings of the 4th multidisciplinary international social networks conference*, 2017, pp. 1–6.

[10] O. Kinouchi, A. J. Holanda, and G. C. Cardoso, “Citation network centrality: a scientific awards predictor?,” *arXiv Prepr. arXiv1910.02369*, 2019.

[11] P. Ji and J. Jin, “Coauthorship and citation networks for statisticians,” *Ann. Appl. Stat.*, vol. 10, no. 4, pp. 1779–1812, 2016.

[12] F. Mazzocchi, “Scientific research across and beyond disciplines: Challenges and opportunities of interdisciplinarity,” *EMBO Rep.*, vol. 20, no. 6, p. e47682, 2019.

[13] A. Martin-Martin, E. Orduña-Malea, A.-W. Harzing, and E. D. López-Cózar, “Can we use Google Scholar to identify highly-cited documents?,” *J. Informetr.*, vol. 11, no. 1, pp. 152–163, 2017.

[14] J. E. Hirsch, “An index to quantify an individual’s scientific research output,” *Proc. Natl. Acad. Sci.*, vol. 102, no. 46, pp. 16569–16572, 2005.

[15] J. E. Hirsch and G. Buela-Casal, “The meaning of the h-index,” *Int. J. Clin. Heal. Psychol.*, vol. 14, no. 2, pp. 161–164, 2014.
[16] L. C. Freeman, “Social networks and the structure experiment,” *Res. methods Soc. Netw. Anal.*, pp. 11–40, 1989.
[17] C. J. R. Machado, A. M. A. Maciel, R. L. Rodrigues, and R. Menezes, “An approach for thematic relevance analysis applied to textual contributions in discussion forums,” *Int. J. Distance Educ. Technol.*, vol. 17, no. 3, pp. 37–51, 2019.
[18] O. Frank, “Transitivity in stochastic graphs and digraphs,” *J. Math. Sociol.*, vol. 7, no. 2, pp. 199–213, 1980.
[19] S. Liu, D.-G. Zhang, X. Liu, T. Zhang, J. Gao, and Y. Cui, “Dynamic analysis for the average shortest path length of mobile ad hoc networks under random failure scenarios,” *IEEE Access*, vol. 7, pp. 21343–21358, 2019.
[20] S. Fortunato, “Community detection in graphs,” *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, 2010.
[21] C. Santana, E. Keedwell, and R. Menezes, “An approach to assess swarm intelligence algorithms based on complex networks,” in *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, 2020, pp. 31–39.
[22] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proc. Natl. Acad. Sci. USA*, vol. 99, no. cond-mat/0112110, pp. 8271–8276, 2001.
[23] A.-L. Barabasi and Z. N. Oltvai, “Network biology: understanding the cell’s functional organization,” *Nat. Rev. Genet.*, vol. 5, no. 2, pp. 101–113, 2004.
[24] M. O. Jackson and A. Watts, “The evolution of social and economic networks,” *J. Econ. Theory*, vol. 106, no. 2, pp. 265–295, 2002.
[25] X. Zhang, “Model design of wireless sensor network based on scale-free network theory,” in *2009 5th International Conference on Wireless Communications, Networking and Mobile Computing*, 2009, pp. 1–4.
[26] S. B. Andrews and D. Knoke, *Networks in and around organizations*. Jai Press, 1999.
[27] A.-L. Barabási and E. Bonabeau, “Scale-free networks,” *Sci. Am.*, vol. 288, no. 5, pp. 60–69, 2003.
[28] A.-L. Barabási, “Scale-free networks: a decade and beyond,” *Science (80-.).*, vol. 325, no. 5939, pp. 412–413, 2009.
[29] M. McPherson, L. Smith-Lovin, and J. M. Cook, “Birds of a feather: Homophily in social networks,” *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, 2001.
[30] A. Clauset, C. R. Shalizi, and M. E. J. Newman, “Power-law distributions in empirical data,” *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, 2009.

[31] C. Castellano, S. Fortunato, and V. Loreto, “Statistical physics of social dynamics,” *Rev. Mod. Phys.*, vol. 81, no. 2, p. 591, 2009.
[32] J.-F. Rual *et al.*, “Towards a proteome-scale map of the human protein–protein interaction network,” *Nature*, vol. 437, no. 7062, pp. 1173–1178, 2005.
[33] H. Ibarra, “Homophily and differential returns: Sex differences in network structure and access in an advertising firm,” *Adm. Sci. Q.*, pp. 422–447, 1992.
[34] P. V Marsden, “Homogeneity in confiding relations,” *Soc. Networks*, vol. 10, no. 1, pp. 57–76, 1988.
[35] H. Bisgin, N. Agarwal, and X. Xu, “A study of homophily on social media,” *World Wide Web*, vol. 15, no. 2, pp. 213–232, 2012.
[36] M. Ou, P. Cui, J. Pei, Z. Zhang, and W. Zhu, “Asymmetric transitivity preserving graph embedding,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 1105–1114.
[37] D. Garlaschelli and M. I. Loffredo, “Patterns of link reciprocity in directed networks,” *Phys. Rev. Lett.*, vol. 93, no. 26, p. 268701, 2004.
[38] J. Lewis, “How Transdisciplinary is Design? An Analysis Using Citation Networks,” *Des. Issues*, vol. 36, no. 1, pp. 30–44, 2020.
[39] T. Agryzkov, L. Tortosa, and J. F. Vicent, “A variant of the current flow betweenness centrality and its application in urban networks,” *Appl. Math. Comput.*, vol. 347, pp. 600–615, 2019.
[40] N. Eltantawy and J. B. Wiest, “The Arab spring| Social media in the Egyptian revolution: reconsidering resource mobilization theory,” *Int. J. Commun.*, vol. 5, p. 18, 2011.
[41] B. Wei and Y. Deng, “A cluster-growing dimension of complex networks: From the view of node closeness centrality,” *Phys. A Stat. Mech. its Appl.*, vol. 522, pp. 80–87, 2019.
[42] M. Thelwall, “A web crawler design for data mining,” *J. Inf. Sci.*, vol. 27, no. 5, pp. 319–325, 2001.
[43] V. Shkapenyuk and T. Suel, “Design and implementation of a high-performance distributed web crawler,” in *Proceedings 18th International Conference on Data Engineering*, 2002, pp. 357–368.
[44] A. Heydon and M. Najork, “Mercator: A scalable, extensible web crawler,” *World Wide Web*, vol. 2, no. 4, pp. 219–229, 1999.
[45] A. Menshchikov, A. Komarova, Y. Gatchin, A. Korobeynikov, and N. Tishukova, “A study of different web-crawler behaviour,” in *2017 20th Conference of Open Innovations Association (FRUCT)*, 2017, pp. 268–274.
[46] N. Algiryage, G. Dias, and S. Jayasena, “Distinguishing real web crawlers from fakes: Googlebot example,” in *2018 Moratuwa Engineering Research Conference (MERCon)*, 2018, pp. 13–18.
[47] N. Bikakis, “Big data visualization tools,” *arXiv Prepr.*

arXiv1801.08336, 2018.
[48] S. D. H. Evergreen, *Effective data visualization: The right chart for the right data*. Sage Publications, 2019.
[49] E. Kaya, M. Agca, F. Adiguzel, and M. Cetin, "Spatial data analysis with R programming for environment," *Hum. Ecol. risk Assess. An Int. J.*, vol. 25, no. 6, pp. 1521–1530, 2019.
[50] S. U. S. Chebolu, "A General Summarization Matrix for Scalable Machine Learning Model Computation in the R Language." 2019.
[51] M. L. Rizzo, *Statistical computing with R*. CRC Press, 2019.
[52] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks.," *Icwsm*, vol. 8, no. 2009, pp. 361–362, 2009.

APPENDICES

Appendix (1) A summary of the main Iraqi Universities (According to the websites of the Iraqi Universities)

No	University	City	Governorate	Year founded	No. of Faculty members	No. of Students
1	University of Baghdad	Baghdad	Baghdad	1957	6321	70615
2	Mustansiriyah University	Baghdad	Baghdad	1963	3300	37757
3	University of Basrah	Basra	Basra	1964	2886	45000
4	University of Mosul	Mosul	Ninawa	1967	4281	47000
5	University of Technology	Baghdad	Baghdad	1975	1573	10111
6	University of Kufa	Kufa	Najaf	1987	2144	27006
7	Al-Nahrain University	Baghdad	Baghdad	1987	1061	11531
8	Tikrit University	Tikrit	Salah Al-Din	1987	1978	54869
9	University of Anbar	Ramadi	Anbar	1987	1674	22719
10	University of Al-Qadisiyah	Qadisiya	Diwaniya	1987	1432	21303
11	University of Babylon	Hillah	Babylon	1991	1998	25462
12	University of Diyala	Baquba	Diyala	1999	1376	22450
13	University of Thi-Qar	Nasiriya	Thi-Qar	2000	1179	20634
14	University of Kerbala	Karbala	Karbala	2002	1337	20072
15	University of Wasit	Kut	Wasit	2003	400	24000
16	University of Kirkuk	Kirkuk	Kirkuk	2003	813	27500
17	Al-Muthanna University	Sumbawa	Muthanna	2007	535	13266
18	University of Samarra	Samarra	Salah Al-Din	2012	338	8205
19	Southern Technical University	Basra	Basra	2014	535	17370
20	Northern Technical University	Mosul	Ninawa	2014	836	17555
21	Middle Technical University	Baghdad	Baghdad	2014	1515	40758
22	Al-Furat Al-Awsat Technical University	Kufa	Najaf	2014	989	22702

ملخص

في العصر التكنولوجي الحالي، يعتبر البحث العلمي أحد أهم عوامل التطور في حياتنا. إن المصادر الرئيسية لإنتاج البحث العلمي هي الجامعات والمؤسسات ومراكز البحث والمختبرات العلمية في جميع أنحاء العالم. لذلك، فإن من المهم تقييم أداء هذه المؤسسات من حيث إنتاجية وجودة البحث العلمي. السبب الرئيسي لهذا التقييم هو تحسين أداء الباحثين وعكس هذا التحسن على حالة البحث العلمي. علاوة على ذلك، يمكن قياس إنتاجية وجودة الباحثين في جامعة معينة بناءً على مؤشرين رئيسيين، وهما الاستشهادات البحثية وأماكن نشر الأبحاث.

في هذه الرسالة، تم التحقيق بعمق في الوضع العلمي الحالي للجامعات العراقية الرئيسية. من أجل تحقيق هذه الغاية، تم إنشاء شبكة اقتباس من الجامعات العراقية. هذا النوع من الشبكات له القدرة على عكس حالة البحث العلمي الفعلي للجامعات العراقية الرئيسية. يعتمد المنهج المستخدم في هذه الرسالة على مفاهيم الشبكات المعقدة. من أجل جمع البيانات، صمم برنامج ذو غرض خاص للزحف إلى مستودع الباحث العلمي من جوجل واسترداد جميع البيانات المطلوبة. تم تصميم هذا الزاحف لجمع المقالات البحثية المنشورة بناءً على المجالات التعليمية الرسمية للجامعات العراقية.

تتمثل المساهمة الرئيسية الأولى لهذا العمل في توليد شبكة الاقتباس من الجامعات العراقية الرئيسية واستخلاص الحقائق الرئيسية حول أنشطة البحث العلمي. أما المساهمة الثانية فتقترح تصنيفاً محلياً للجامعات العراقية الرئيسية بناءً على قياسات الشبكة والمؤشرات الأكاديمية الأخرى. جانب آخر تم بحثه في هذا العمل وهو التعاون العلمي بين الجامعات العراقية والجامعات العالمية. علاوةً على ذلك، توضح هذه الرسالة أيضاً الوضع الحالي للجامعات العراقية مقارنة بالعالم من حيث مستودع سكوبس. بناءً على النتائج التي تم الحصول عليها تقدم هذه الرسالة توصيات واقتراحات حول كيفية تحسين أداء الجامعات العراقية من حيث البحث العلمي والتعاون العلمي بين الجامعات.

تظهر النتائج التي تم الحصول عليها أداءً متوسطاً للبحث العلمي في الجامعات العراقية حسب قياسات الشبكة مثل متوسط معامل التجميع ومتوسط طول المسار. لكن جامعة بغداد تفوقت على باقي الجامعات العراقية من حيث تكرار الاستشهادات. كما أن المؤلف الأعلى اقتباسًا كان من جامعة بغداد في مجال الطب بحوالي ١٥٥٦٦ استشهادًا (حتى تاريخ كتابة هذه الرسالة). ومع ذلك، فإن أداء البحث العلمي في العراق يقل عن أداء البحث العلمي في الدول المجاورة مثل تركيا وإيران والمملكة العربية السعودية من حيث مؤشر h وعدد الأوراق المنشورة ومجموع الاقتباسات ومتوسط الاقتباس لكل ورقة. كما أظهرت النتائج أن التعاون بين الجامعات العراقية يعتمد على المنطقة الجغرافية.

جمهورية العراق

وزارة التعليم العالي والبحث العلمي

جامعة ديالى

كلية العلوم

قسم علوم الحاسوب

شبكات الاقتباس العلمي: دراسة حالة الجامعات العراقية

رسالة مقدمة إلى مجلس كلية العلوم / جامعة ديالى

وهي جزء من متطلبات نيل درجة الماجستير في علوم الحاسوب

من قبل

أحمد جاسم محمد

بإشراف

أ.م.د طه محمد حسن

م.د باسم محمد محمود الخشاب

ذو الحجة ١٤٤١ هـ العراق تموز ٢٠٢٠ م