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