

Republic of Iraq

Ministry of Higher Education and Scientific Research

University of Babylon

College of Information Technology

Department of Software



Influential Nodes Detection of Social Networks Using Graph Theory

A Dissertation

*Submitted to the Council of the College of Information Technology,
University of Babylon, in Partial Fulfillment of the Requirements for the
Doctorate of Philosophy Degree in Information Technology/ Software*

By

Hussein Loia Hasan Mohammad

Supervised By

Prof. Dr. Salah A. Kadim Albermany

2022 A.D.

1443 A.H.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

فَإِنَّ مَعَ الْعُسْرِ يُسْرًا ﴿٥﴾ إِنَّ مَعَ الْعُسْرِ يُسْرًا ﴿٦﴾

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

سورة الشرح/آية(5-6)

Declaration

The work in this dissertation, **Influential Nodes Detection Using Graph Theory**, is original and no portion of the work referred to here has been submitted in support of an application for another degree or qualification of this or any other university or institution of learning.

Signature:

Name: Hussein Loia Hasan

Date: / /

Supervisor's Certification

I certify that the preparation of this dissertation "**Influential Nodes Detection Using Graph Theory**" has been prepared by "**Hussein Loia Hasan**" under my supervision at the University of Babylon, College of Information Technology, in partial fulfillment of the requirements for the degree of Doctorate of Philosophy in Information Technology/Software.

Signature: 

Name: **Prof. Dr. Salah Albermany**

Date: 1 / 6 / 2022

The Head of the Department Certification

I in view of the available recommendations, I forward the dissertation entitled "**Influential Nodes Detection Using Graph Theory**" for debate by the examination committee.

Signature:

Name: **Assistant Prof. Dr. Ahmed Saleem Abbas**

Date: / / 2022

(Head of Software Department)

Dedications

To my father...

my mother...

my wife...

my family...

I will never forget your sincere
love and support

Acknowledgments

First of all, I have to express all my thanks and gratitude to Almighty Allah, who gives me the ability to achieve such imperfect work and all what has been done with his blessing and support. I would like to thank my advisor Prof. Dr. Salah Albermany for his continuous support during my Ph.D.; He was always there to listen and to give advice.

Special thanks go to all the staff members of the Information Technology College for their faithful efforts to give us the utmost scientific topics and endless support in all directions aiming to bring into perfection their scientific followers, and for their unforgettable kind and wise management of our affairs during the research period.

I am immensely grateful to all my family members especially, my father, my mother, my wife, my children, my brother, and sisters, for their patience and encouragement, which have been my greatest impetus and enabled me to complete my research.

Finally, I apologize to those whose names might be missed. But, I am grateful to all of them for the help that I have received in various research and day-to-day life matters.

Hussein

Abstract

One of the critical topics associated with social interaction networks and the systems that manage these networks is the subject of who influences and controls movement, action, activity, trend, and interactions that occur between the parts of these networks. As an example, the movement and transmission of information, news, and others between the components of these networks. This dissertation focuses on finding influencers and controllers on the existing paths within these networks by creating and configuring a (Hybrid Method for Encircling Influential Vertices), abbreviated as HMEIV. The HMEIV algorithm consists of three essential parts.

The first part of the HMEIV algorithm is the creation of the (A Controlling Information Spread-Based Centrality), denoted by CISBC, which has the role of exploring the paths connecting the nodes. The knowing the roles, positions, and contributions of all the remaining nodes in the paths linking all pairs. The CISBC indicator has three gradient levels starting from the first level, which is the level of the individual paths between a pair of nodes and gives values to the nodes that affect and control the shortest paths between each pair of nodes. The second level in the indicator is the discovery of the nodes that control the paths of a particular node within the network. And the return of those who are those nodes that can control the transmission of information. The third level includes discovering who is controlling and affecting the nodes on the paths of the whole network.

The second part of the HMEIV algorithm is the creation of the (Interaction Weights Scale), denoted by IWS, This scale also consists of three stages corresponding to the stages in the CISBC indicator. The first stage explores and identifies the dominators by removing noise, analyzing the paths, finding out who the most active nodes are, and exchanging information on each pair of nodes that are not directly related. The second stage is to find out which

are the nodes that most affect a particular node through their activity in the exchange of data, The third stage is at the level of the entire network to form a clear picture of what the active nodes are in the exchange across the whole network.

The results obtained from CISBC and IWS, the implementation phase of the third part of the HMEIV algorithm begins. It is to find the trade-off between the values of the two indicators by excluding unwanted alternatives and sticking to the good alternatives and solutions. In other words, it is a filtering stage that is implemented by the Technique for Order of Preference by Similarity to Ideal Solution algorithm.

Five different methods were used in the evaluation stage, including comparing the results with the standard measures adopted in the networks (Closeness, Betweenness, Degree, Eigenvector, PageRank). The Person Correlation Coefficient demonstrates a linear association between the outcomes of the HMEIV and the most widely used network indicators. The method of calculating frequency of similar values also proves in nodes that share the same values where the HMEIV produces the best results. The Suspectable Infected model also measures the validity of the affected nodes and their effect on the remaining nodes. The last method tests the effective nodes resulting from the HMEIV algorithm to show the extent of their effect on the network's distance, average distance, and density when these nodes are removed and compare these results before and after removal. The summary of the findings obtained from the five assessment techniques gives satisfactory outcomes.

Publications Associated with this Dissertation

Some of the works presented in this dissertation have been published or accepted as listed below.

1. Hussein L. Hasan and Salah A. Albermany, "The Dominant Nodes Index in Shortest Paths (DNISP) as A Local Measure of Source Looking for the Target in Complex Networks", International Journal of Interactive Mobile Technologies (iJIM), 2020, vol 14(19) pp. 97-108.
2. Hussein L. Hasan and Salah A. Albermany, " A Weak Relation-Based Approach to Detect Influential Nodes in Complex Network" In AIP Conference Proceedings, 2022, vol. 2398(1), AIP Publishing LLC.

List of Contents

1. General Overview	0
1.1 Introduction.....	1
1.2 Problem statement.....	2
1.3 Dissertation Motivation	3
1.4 Challenges.....	4
1.5 Dissertation Aims.....	4
1.6 The Contributions of Dissertation.....	5
1.7 Related Works.....	5
1.8 Dissertation Layout.....	8
2. Theoretical Background	10
2.1 Introduction.....	11
2.2 Graph Concepts.....	11
<u>2.2.1</u> Size of Graph	12
<u>2.2.2</u> The Asymmetric and Symmetric Graphs	12
2.2.3 The Distance in Graph	14
2.2.4 The links in Undirected and Directed Graphs	16
<u>2.2.5</u> Path Description in the graph.....	17
2.2.6 The Sub-graph from a Graph.....	17
2.2.7 Types of Patterns in Graphs.....	17
2.3 Social Networks	18
2.4 Social Networks Types	19
2.4.1 Egocentric, Partial, and Full Networks.....	20
2.4.2 Networks of Unimodal, Affiliation, and Multimodal.....	22
2.4.3 Multiplex Networks	23
2.5 Metrics of Networks	23

2.5.1 Metrics of full network	25
2.5.2 Metrics of Nodes.....	27
2.5.3 Centrality Metrics Applications.....	31
2.6 Influencers in Networks	32
2.7 Interactions Inside Networks	34
2.8 Diffusion Models	35
2.9 Pearson Correlation Coefficient.....	37
2.10 TOPSIS Method.....	37
2.11 Breadth-first search (BFS).....	39
The Proposed System.....	40
3.1 Introduction.....	41
3.2 Methodology	41
3.3 Pre-Processing Stage.....	44
3.4 Components of HMEIV Algorithm	46
3.4.1 IWS Indicator.....	47
3.4.2 CISBC Indicator	53
3.4.3 Qualifying Stage	60
3.5 Reduce Computational Time	64
The Experimental Results	67
4.1 Introduction.....	68
4.2 Data Description	68
4.3 Apply HMEIV on sample network (Ds_1).....	69
4.3.1 Steps of indicator IWS.....	70
4.3.2 Steps of indicator CISBC	74
4.3.3 Qualifying stage.....	77
4.3.4 Discusses Results of Simple Network (Ds_1)	82
4.4 Results of Ds_2	83
4.4.1 Discusses of Results (Ds_2).....	87
4.5 Results of Ds_3	89
4.5.1 Discusses of Results (Ds_3).....	92

4.6 Results of Ds_4	93
4.6.1 Discusses of Results (Ds_4)	96
4.7 The Methods of Evaluation (Ds_2, Ds_3, Ds_4)	96
4.7.1 The results HMEIV compared with the centrality metrics.....	97
4.7.2 Employing the Pearson Correlation Coefficient method.....	98
4.7.3 Frequency of similar values.....	99
4.7.4 Susceptible Infected Model (SI)	101
4.7.5 Attacking the network.....	110
4.8 The Summary	112
Conclusions and Future Works	113
5.1 Conclusions in the following points	114
5.2 Future works	115
References	116

List of Tables

TABLE 3. 1 DESCRIPTION OF DATASETS	46
TABLE 3. 2 SAMPLE OF DATASET OF EXCHANGE DATA.....	47
TABLE 4. 1 GENERAL METRICS OF FOUR NETWORKS	69
TABLE 4. 2 THE RESULTS OF PAIR { 115, 1009} IN DS_2.....	84
TABLE 4. 3 THE RESULTS OF PAIR {409, 1122} IN DS_2.....	84
TABLE 4. 4 THE RESULTS OF NODE 454 IN DS_2	84
TABLE 4. 5 THE RESULTS OF NODE 982 IN DS_2	85
TABLE 4. 6 THE RESULTS OF WHOLE NETWORK (DS_2)	86
TABLE 4. 7 THE RESULTS OF PAIRS { 14, 915}, {348, 1103}, {162, 299} IN DS_3	90
TABLE 4. 8 THE RESULTS OF NODE 31 IN DS_3	90
TABLE 4. 9 THE RESULTS FOR NODE 687 IN DS_3.....	91
TABLE 4. 10 THE RESULTS FOR NODE 1022 IN DS_3.....	91
TABLE 4. 11 THE RESULT FOR WHOLE NETWORK TOP -15 (DS_3)	92
TABLE 4. 12 THE RESULT OF PAIRS {1112,1890}, {115,902}, {444,1118} IN DS_4	94
TABLE 4. 13 THE RESULTS OF NODE 123 IN DS_4	94
TABLE 4. 14 THE RESULTS OF NODE 1151 IN DS_4	95
TABLE 4. 15 THE RESULTS OF WHOLE NETWORK (DS_4)	95

TABLE 4. 16 THE TOP RANK 15 VERTICES OF THE METRIC STANDARDS WITH HMEIV_3 FOR DS_2.....	97
TABLE 4. 17 THE TOP RANK 15 VERTICES OF THE METRIC STANDARDS WITH HMEIV_3 FOR DS_3.....	98
TABLE 4. 18 THE TOP RANK 15 VERTICES OF THE METRIC STANDARDS WITH HMEIV_3 FOR DS_4.....	98
TABLE 4. 19 VALUES PERSON CORRELATION COEFFICIENT BETWEEN STANDARDS METRIC WITH HMEIV_3.....	99
TABLE 4. 20 DETAILS OF DISTANCE AFTER PROCEDURE DELETE NODES.....	111

List of Figures

FIGURE 2. 1 UNDIRECTED GRAPH WITH N=9(NODES, VERTICES), M=18 (EDGES, LINKS)	13
FIGURE 2. 2 UNDIRECTED WEIGHT GRAPH WITH N=21 (NODES, VERTICES), M=32(EDGES ,LINKS).....	14
FIGURE 2. 3 DIRECTED GRAPH UNREACHABLE TO NODE A (WEAK COMPONENTS)15	
FIGURE 2. 4 DIRECTED GRAPH (STRONG COMPONENTS).....	16
FIGURE 2. 5 CLIQUE UNDIRECTED GRAPH HAVING 20 NODES AND 190 EDGES	18
FIGURE 2. 6 SLICE OF NETWORK WITH 389 VERTICES AND 3567 EDGES	22
FIGURE 2. 7 SAMPLE NETWORK WITH NODE=13, EDGE= 18.....	26
FIGURE 2. 8 SAMPLE NETWORK SHOWS DEGREE CONTROLLING SIZE AND SHAPE NODES (NODES WITH MORE THAN FOUR DEGREES HAVE A TRIANGLE, WHILE THE REST IS A DARK CIRCLE, THE SHAPE SIZE CAN ALSO BE LARGER WITH THE LARGEST DEGREE)	29
FIGURE 2. 9 SAMPLE NETWORK SHOWS BETWEENNESS CONTROLLING SIZE AND SHAPE NODES(THE NODES WITH A HIGH BETWEENNESS ARE A DARK SQUARE, WHILE THE REST ARE A DARK CIRCLE: ALSO THE SIZE SHAPE CAN BE LARGER WITH THE LARGEST BETWEENNESS).....	30
Figure 3. 1 The general outline of HMEIV algorithm.....	43
FIGURE 3. 2 THE FIRST PHASE FOR (IWS_1).....	50
FIGURE 3. 3 THE SECOND PHASE (IWS_2).....	52
FIGURE 3. 4 THE THIRD PHASE (IWS_3)	53
FIGURE 3. 5 THE FIRST PHASE (CISBC_1).....	56

FIGURE 3. 6 THE SECOND PHASE (CISBC_2).....	58
FIGURE 3. 7 THE THIRD PHASE (CISBC_3)	59
FIGURE 3. 8 THE LEVELS OF IMPLEMENTATION OF THE QUALIFYING STAGE	63
FIGURE 3. 9 SIMPLE NETWORK WITH K-SHELL	65
FIGURE 4. 1 THE NETWORK TOPOLOGY FOR THE SAMPLE NETWORK CONSISTS OF 19 NODES WITH 32 EDGES.....	69
FIGURE 4. 2 RESULTS IWS_1 OF PAIR {Q, D}	71
FIGURE 4. 3 RESULTS IWS_1 OF PAIR {M, B}.....	71
FIGURE 4. 4 RESULTS IWS_2 OF N' PATHS	72
FIGURE 4. 5 RESULTS IWS_2 OF B' PATHS.....	73
FIGURE 4. 6 RESULTS IWS_3 OF WHOLE NETWORK	73
FIGURE 4. 7 RESULTS CISBC_1 OF PAIR {M,B}.....	74
FIGURE 4. 8 RESULTS CISBC_1 OF PAIR {Q, D}	75
FIGURE 4. 9 RESULTS CISBC_2 OF N' PATHS	76
FIGURE 4. 10 RESULTS CISBC_2 OF B' PATHS.....	76
FIGURE 4. 11 RESULTS CISBC_3 OF WHOLE NETWORK	77
FIGURE 4. 12 RESULTS HMEIV_1 OF PAIR {L, D}	78
FIGURE 4. 13 RESULTS HMEIV_1 OF PAIR {Q, D}	78
FIGURE 4. 14 RESULTS HMEIV_1 OF PAIR {M, B}.....	79
FIGURE 4. 15 RESULTS HMEIV_2 OF L' PATHS	79
FIGURE 4. 16 RESULTS HMEIV_2 OF D' PATHS.....	80
FIGURE 4. 17 RESULTS HMEIV_2 OF N' PATHS	80
FIGURE 4. 18 RESULTS HMEIV_2 OF B' PATHS.....	81
FIGURE 4. 19 RESULTS HMEIV_3 OF WHOLE NETWORK	81
FIGURE 4. 20 THE TOPOLOGY OF DS_2.....	83
FIGURE 4. 21 RESULTS HMEIV_2 OF NODE (454)	85
FIGURE 4. 22 RESULTS HMEIV_2 OF NODE (982)	86
FIGURE 4. 23 RESULTS HMEIV_3 OF WHOLE NETWORK	87
FIGURE 4. 24 THE TOPOLOGY OF DS_3.....	89
FIGURE 4. 25 THE TOPOLOGY OF DS_4.....	93
FIGURE 4. 26 THE FREQUENCY OF SIMILAR VALUES IN DS_2.....	100
FIGURE 4. 27 THE FREQUENCY OF SIMILAR VALUES IN DS_3.....	100
FIGURE 4. 28 THE FREQUENCY OF SIMILAR VALUES IN DS_4.....	101
FIGURE 4. 29 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & DEGREE WITHIN DS_2.....	102
FIGURE 4. 30 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & BETWEENNESS WITHIN DS_2.....	103
FIGURE 4. 31 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & CLOSENESS WITHIN DS_2.....	103
FIGURE 4. 32 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & EIGENVECTOR WITHIN DS_2.....	104

FIGURE 4. 33 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & PAGERANK WITHIN DS_2	104
FIGURE 4. 34 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & DEGREE WITHIN DS_3	105
FIGURE 4. 35 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & BETWEENNESS WITHIN DS_3	105
FIGURE 4. 36 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & CLOSENESS WITHIN DS_3	106
FIGURE 4. 37 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & EIGENVECTOR WITHIN DS_3	106
FIGURE 4. 38 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & PAGERANK WITHIN DS_3	107
FIGURE 4. 39 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & DEGREE WITHIN DS_4	107
FIGURE 4. 40 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & BETWEENNESS WITHIN DS_4	108
FIGURE 4. 41 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & CLOSENESS WITHIN DS_4	108
FIGURE 4. 42 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & EIGENVECTOR WITHIN DS_4	109
FIGURE 4. 43 THE INFECTED NODES THROUGH TOP- 25 NODES IN HMEIV & PAGERANK WITHIN DS_4	109
FIGURE 4. 44 THE SIMPLE NETWORK.....	110
FIGURE 4. 45 THE SIMPLE NETWORK AFTER DELETE NODE M	111

Table of Abbreviations

Abbreviation	Meaning
Ad _G	Average distance
Bc	Betweenness centrality
BFS	Breadth-first search
CbC	Community-based Centrality
DAG	Directed acyclic graph
Dc	Degree centrality
deg(v)	The number of arcs attached to the node v(indirected graph)
deg _i (v)	Arcs that enter to node v
deg _{max} (G)	maximal degree for graph G
deg _{min} (G)	minimal degree for graph G
deg _o (v)	Arcs that leave node v
Den _G	Graph density
Diag	Maximum distance
Dis(s, t)	Shortest distance between node s and node t
E	Edges or Links

E.c	Eigenvector centrality
E_{\max}	Maximum edges in clique graph
E_n	Empty graph
ESC	Evidential Semi-local Centrality
EVC	Existing Evidential Centrality
G	Graph
ICM	Independent Cascade Model
K_n	Full size graph
L_i	sets in indicator (CISBC)
LTM	Linear Threshold Model
LTR	Linear threshold rank
M	Edges number
MBI	Maximum betweenness improve
M_i	sets in indictors (IWS)
MRI	Maximum ranking improve
N	Nodes number
N_c	connected components numbers of graph
N_e -max	maximum edges in a connected component
N_v -max	maximum nodes in a connected component
OSN	Online Social Networks
RaC	Rank after clustering
S	Source node
SCC	Strongly connected component
SI	Susceptible Infected
SIR	Susceptible Infected Recovered
SIS	Susceptible Infected Susceptible
T	Target node
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
V	Vertex
W	Edge weight
WCC	Weakly connected component

CHAPTER ONE

General Overview

1.1 Introduction

Before the development of social networks and their creation in applications, the principle of social networks existed within our societies, so individuals have become part of social networks since ancient times. People live and work in a world that surrounds by connections and interactions. [1] [2].

People communicate through social networks that form and vary into language, exchange of information, cooperation, trade, kinship, citation, and conflict. [3] [4].

Existing platforms have created social networks in a relatively new form. But the idea of having exchanges and interactions within these networks was taken, which was primordial [5] [6].

The simple definition of a network is the gathering of things with their relationship to each other. Connected things in networks are called nodes, entities, or vertices, sometimes people. While connections between entities are called ties, links, edges, or relationships [7].

The transition of data within the network depends largely on the network topology and node activity within their communities [8].

In the networks, a node is affected by another node, for example, in social networks, an individual (node) takes the opinions or products that they have to buy through the node that has the most influence in transmitting data and controlling the paths of the rest of the nodes [9]. The effect can be within a narrower range, for example, when a specific node is sent to be a source to a specific node to be a target for this transmission [10].

The network analysts can see the world as a form of connected parts. Researchers in social networks say that relationships are the basic components of society; each group of relationships unites to arise distinct patterns of links between the nodes [9].

Social network analysis focuses on the relationships between nodes rather than what is contained within these nodes. In comparison, traditional social

sciences focus on research methods on individuals and their characteristics (age, gender, marital status, occupation) [9].

So that network scientists focus on connections, this difference in format focus from attribute information to relational information changes and affects how data is collected, analyzed, and represented [3].

Analyzing social networks is the crucial element in completing algorithms that focus more on users. This key dimension in links analyzing captures the interconnected structure and connective inside society.

Networks are always impressed with their visual quality and ability to generate patterns, each of us is entwined in various networks of friendship and business relationships. It is worth understanding the network, whose relationships and connections appear to be highly complex and erratic.

Therefore, it is practical to look at data and relationships in a new way to find important patterns that may be immersed in a great momentum of data [11].

It is a good idea to eliminate or isolate individual noise to extract the main groups of entities in a network that may be quite complex [12] [13].

1.2 Problem Statement

The subject of mutual exchange of data, such as news, rumors, thoughts, and information, which is exchanged from one section of the network to another, is a significant subject correlated with social network research. It can be determined that there is an open problem in networks which is to identify the influencers within the networks who have control over the paths and interactions that occur within the networks and thus have an impact in directing and adopting users to a certain opinion, decision, promoting a product or even spreading news and curbing rumors. There are many metrics (indicators, metrics) used, but these indicators may have overlooked some aspects and improving the performance of these indicators including:

- Indicator bring together the local and global characteristics of the network with the activities of nodes.
- Indicator gives us a close look at each part of the network (node or path)
- Indicator that provides differentiation to each node of the influencer and does not give comprehensiveness in the effect only.
- Indicator gives the probability of the paths for the spread of the data that it passes between the parts of the network, for example, the probability of the paths that data passes between a source and a target node.

1.3 Dissertation Motivation

Aspect of motivation is trying to explain the transmission of data between network nodes, especially if this network is complicated, why did the data decide to go this way and not follow another alternative path? Many researchers have attempted to identify the shortest path between nodes in complex networks, but what was inside the path was forgotten or ignored. For example, assuming that there is more than one short path within a complex network that connects the source node denoted by s and the target node denoted by t , the data will take the shortest path (although it is not necessary to take the shortest route in all cases). Thus, there are varying proportions of the control of these nodes in the spread of data from source node s to target node t ; a particular node may prevail in the contribution; this is attributed to several factors, including links or the location of this node. Most researchers ignore how data is transmitted or tracked between nodes. Yet, this study attempts to give a close-up picture of how paths are controlled by nodes located within the paths, showing how strategies are developed, for example, "going viral" immunization.

1.4 Challenges

It is hard to determine the influencers by selecting the suitable indication that combines the real user's activity and position within the network. Another step is finding the influences on a certain node. For example, trying to suppress news, rumors about this node or targeting it specifically for promotion or marketing. Another aspect of challenging extracting influencers across the entire network by calculating contributions of all nodes.

1.5 Dissertation Aims

This dissertation focused on forming an algorithm (Hybrid Method for Encircling Influential Vertices) denoted by HMEIV to find the nodes controlling or influencing the transmission of data, discouraging rumors, promotion, and marketing inside social networks. It consists of three main parts:

In the first part, where the indicator (A Controlling Information Spread-Based Centrality) denoted by CISBC, which depends on structural features of the network, where the metric uniquely derived through three phases:

CISBC_1: first phase within paths

CISBC_2: second phase for an individual node.

CISBC_3: third phase for whole networks.

The second part is specific to the exchange data that built them (Interaction Weight Scale), denoted by IWS. Also, the metric is uniquely derived through the three phases:

IWS_1: first phase within paths

IWS_2: second phase for an individual node.

IWS_3: third phase for whole networks.

In the third part, the HMEIV algorithm will select alternatives from both indicators with two different criteria through the qualifying stage.

1.6 The Contributions of Dissertation

1. Blending network properties with exchange properties for users to find influencers within a path between a source node and a target node.
2. Building a unique system for each network to determine influencers on a specific node. The effect of these influences is on the paths of a particular node.
3. Extracting influencers across the entire network by calculating contributions of all nodes.
4. Determining the best optimum paths for traffic between a target node and a source node among multiple shortest paths.

1.7 Related Works

There have been several studies conducted on the topic of the study, and they are as follows::

1. *Shang Q. et al, in 2021* suggested a method to extract influential nodes by measuring network efficiency by deleting a node each time and updating the weight. Based on the evaluation process on the methods of SI, correlation coefficient, and quality of the network, the method was promising and achieved good results [14].
2. *Zhang Bo, et al in 2020* suggested a model (rank after clustering (RaC)) which tries to reduce unnecessary computations in calculating the influencing nodes of the entire network. In the first stage, which depends on the network structure, the K-means algorithm is used, which collects the influential nodes. In the next step, these nodes are filtered by including the interactions between users for finding influential vertices depending on an individual's interactions and behavior among different individuals. Based on the management model, the individuals in a social network can be distributed into four

levels, i.e., ordinary individuals, active individuals, leaders of subject opinion, and leaders of a network [15].

3. *Liu F. et al, in 2020* combined two indicators that are eigenvector and degree, to extract the influential nodes. The model was tested on experimental data, and various methods evaluated the results; the researcher wanted to mix the multiple characteristics of both indicators to achieve better results. [16].
4. *Riquelme, F. et al. in 2019* proposed a "generalized Linear Threshold Rank (LTR)" measure to improve the original LTR, which takes into value the neighbors' distance for the initial activation set. They proved that the influential actors with neighbors at a larger distance are the best. They also showed no high correlation for the rankings that were gotten for the different neighborhood levels, which validates the measure generalization [17].
5. *Leung Carson K., et al., in 2019* studied the influence and suggested an algorithm for measuring influential social vertices by forecasting a user's activity to spread information. The effectiveness depends on the number of retweet counts. The algorithm base on both network topology and user activity [18].
6. *Shao Zengzhen, et al, in 2019* suggested centrality that depends on the second degree of neighbors to identify influential nodes in social networks. A DIL-centrality also determined the weight of the edge. The authors claimed that the results were satisfactory by using a network simulation and taking the average of these results. The researchers found that relying on the second degree was better than relying on the traditional degree [19].

7. *Yang, Pingle, Xin Liu, and Guiqiong Xuin, in 2018* employed the Topsis method in two phases to generate a hybrid model which extracts influential nodes in complex networks. The model depended on the Topsis method, where three of the standard metrics (Dc, Cc, Bc) were used as alternatives in the Topsis method. While calculating weights generated by the SIR model to estimate interaction, re-use the Topsis algorithm on the accumulated results of weights experiments. The algorithm gave good results in terms of extracting seeds for the influencers [20].
8. *Bergamini et al., in 2018* studied the problems of increasing the Betweenness metric of a given node "Maximum Betweenness Improve (MBI)" and of increasing the ranking of a given node "Maximum Ranking Improve (MRI)." They proposed a greedy algorithm for MBI and a new approach called "improvement of Betweenness (iBet)" to update the Betweenness score after the insertion of a weight increase or decrease for all vertices in the graph. They compared a suggested method with many baselines of centrality measures and a random method [21].
9. *Cao Linlin, et al., in 2018* suggested a model based on select Topic-Leader Rank. It summates the node weight (users' activity, links between two vertices, and specific topic), link weight, an individual interaction attribute, and content features to extract the influential vertices for a single topic [22].
10. *Ullah, Farman, and Sungchang Lee, in 2017* proposed a model for detecting influential vertices. They increase the effect on the nodes surrounding it and decrease the contagion time. The model combined specific interactions between nodes and topological networks [23].

11. *Liguo Fei et al, in 2017* proposed a new model that mixed the Topsis method and relative entropy, as it was applied to a set of datasets. The method tried to mix network metrics [24].
12. *Zhao Zhiying, et al, in 2015* linked methods for identifying communities within networks with finding influencers in those networks. They suggested a method called (Community-based Centrality CbC); the method is highly dependent on node connections within diverse communities. The researcher assumed that the node has strong and weak relationships, where strong relationships are within their original communities, while weak relationships are in sub or marginal communities [25].
13. *Li Qian, et at, in 2014* supposed the existence of a ground node based on which the weight correlation of the associated nodes depends on the degree of the ground node. This model is called the weighted LeaderRank algorithm, which developed the LeaderRank algorithm in 2011. As the possibility of the algorithm increased in terms of finding the influential nodes [26].
14. *Gao, Cai, et al, in 2013* developed a method (Evidential Semi-local Centrality ESC) that is updated to a previous method (Existing Evidential Centrality EVC) to extract the influential nodes that depend on DC centrality. The researcher claimed to get rid of a little local that was inherent to the index (EVC) by using the distribution node degree at network-level and integrating the results to form the new index (ESC) [27].

1.8 Dissertation Layout

After chapter one, which presents an overview of the entire dissertation, the rest of the dissertation is structured as follows:

- Chapter Two: Reviews an extensive description of Includes a definition of the graphs and the principles associated with it, social networks, centrality measures, and types of networks.

Chapter Three: Includes the steps for building the proposed algorithm model and the associated indicators, with figures that show the stages of work.

- Chapter Four: Includes the results implemented on the datasets. These results include a simple example with detailed steps and three main networks. It also includes five evaluation methods.

- Chapter Five: Reviews the conclusions reached by this dissertation and gives suggestions for future works.

CHAPTER TWO

Theoretical Background

2.1 Introduction

This chapter reviews several topics related to the dissertation. The definition of the graph and the most important considerations and types that the graph is interested in, moving to social networks, their analysis, and the most critical issues that are addressed within as an example metrics of networks, types of connection, many kinds of diffusion, methods, and algorithms associated with the dissertation.

2.2 Graph Concepts

The mathematical definition of relationship among a given group of actors or entities is the foundation for network analysis [28]. A graph combines the two sets; proteins, newspapers, streets names, cities, and terms are all examples of entities that can be distinguished [29]. Entities are groups that reflect subjects such as people, objects, and organizations. Every form of mechanism or contact that includes a subset of these entities, such as purchasing a device, sharing news, or email exchange between colleagues, is called a relationship between actors or entities. A connection may be expressed mathematically, and it is clearly described as a subset of the whole potential of actors. Entities, also called items, vertices, agents, or nodes, could describe various objects [30]. They often serve individuals or social systems, including workgroups, teams, states, organizations, nations, institutions, and countries. Relationships called connections, edges, links, and ties, are the network's copulative tissue, to show how to create and extract the graph from nodes and relationships [28] .

assuming a set of nodes; these nodes have a set of relationships among themselves.

So, the graph can be formed from the set of relationships and nodes linked by relationship by forming pairs of nodes linked to each other by a relationship. This collection is a (binary) connection that describes the actors' relationship mathematically. The expressed mathematically as follows: Let $V = \{\text{node}_1,$

node2, node3, node4, node5} denote nodes set, E reference to edge. So, $V \times V$ is the sequence of all feasible pairs of elements and $E = \{(node1, node3), (node2, node4), (node3, node5), (node1, node4), (node4, node5)\}$ is a subset of $V \times V$ [1].

$$E \subseteq V \times V. \quad (2.1)$$

G is outcome of combining V and E:

$$G = (V, E). \quad (2.2)$$

The set of edges are correlating with G is $E(G)$, and the set of nodes are correlating with G is $V(G)$

2.2.1 Size of Graph

Determining the size of a specific graph has order n that is within the range of 0 to $\binom{n}{2}$. Where $0 \leq m \leq \binom{n}{2}$, for a graph $G(n, m)$. Called the complete graph on a naming graph, which has full size $\binom{n}{2}$ denoted by K_n , while a graph with no edges is called an empty graph [31].

Usually known as an order of G for nodes number as:

$$n = |V| \quad (2.3)$$

And known as size G for edges number as:

$$m = |E| \quad (2.4)$$

2.2.2 The Asymmetric and Symmetric Graphs

This form of relationship is considered symmetric if it is such that node s is connected to node t, then t must be related to s. Unordered pairs of nodes, defined by $\{s,t\}$ reflect symmetric relationships: where $\{s,t\} = \{t,s\}$. The graph is called undirect, as shown in Figure 2.1. Graphs with a direction to the described relationship are considered directed graphs and are represented by ordered pairs (s,t) , where $(s,t) \neq (t,s)$ are called asymmetric [1].

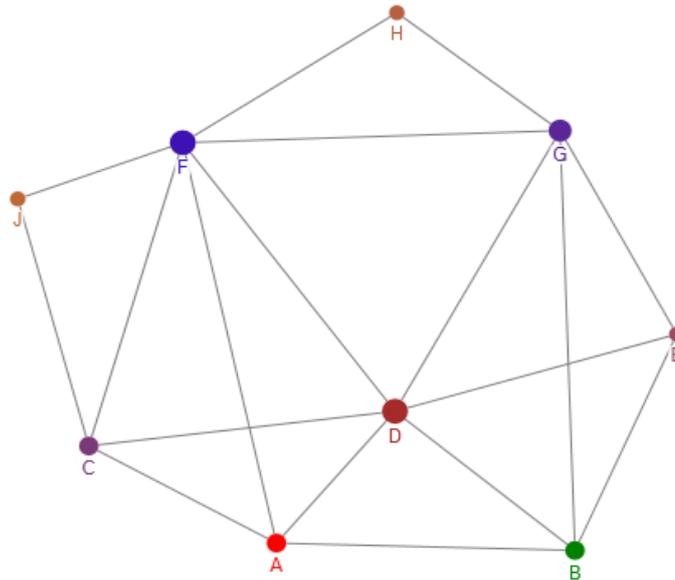


Figure 2. 1 Undirected graph with $n=9$ (nodes, vertices), $m=18$ (edges, links)

Both types of graphs can be weighted. A function of weight w can assign an actual value to links to represent a weight for links [8].

$$w: E \rightarrow \mathbb{R} \quad (2.5)$$

A graph with weights is

$$G = (V, E, w) \quad (2.6)$$

The relation contains vertices V , a number of links $E \subseteq V \times V$, and the expression of weight $\omega: E \rightarrow \mathbb{R}$. As shown in Figure. 2.2.

of complete nodes that make up the linked graph are called the connected components. The directed graph does not match the undirected graph inaccessibility to nodes, where the s node can reach the t node, but the opposite is not true. Where $\text{dis}(s,t)$ give some number of paths, but $\text{dis}(t,s)$ may give ∞ [31]. There are two cases in the directed graph, which are strong components and weak components. In the first case, some paths can be accessed to all nodes, as shown in Figure 2.4. If one node is not reached from another node, it will be weak components, shown in Figure 2.3 [8].

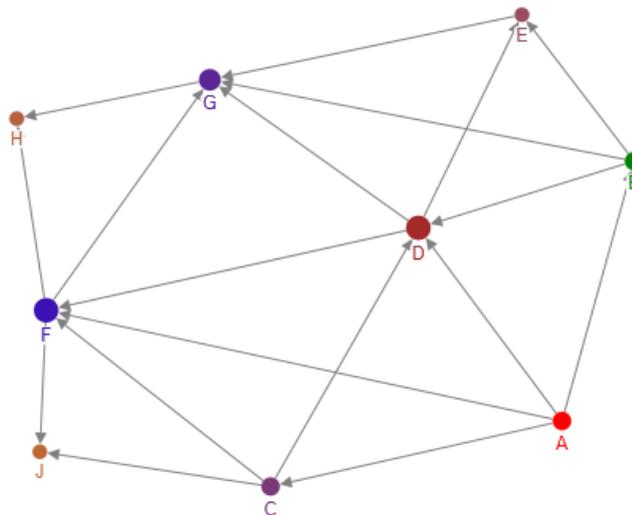


Figure 2. 3 Directed graph unreachable to node A (weak components)

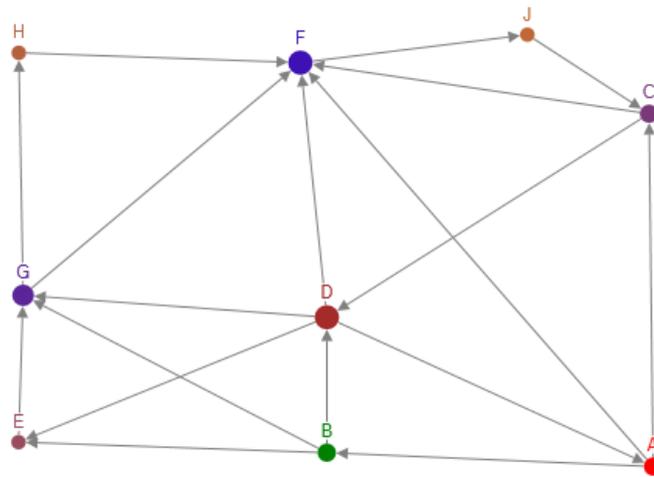


Figure 2. 4 Directed graph (Strong components)

Several graphs can be derived from the directed graph, such as the directed acyclic graph (DAG) that does not contain cycles.

2.2.4 The Links in Undirected and Directed Graphs

The node that has no connection to any part of the graph is called isolated node. The number of edges attached to node v represents the degree of the node $\text{deg}(v)$, representing the undirected graph. There is a minimal degree for graph G denoted by $\text{deg_min}(G)$ and a maximal degree denoted by $\text{deg_max}(G)$. Sometime it is $\text{deg_min}(G)=\text{deg_max}(G)$, thus the graph is called regular [31] [32].

In directed graphs, there is a distinction between the states indegree ($\text{deg}_i(v)$) for edges that enter node v and outdegree ($\text{deg}_o(v)$) for edges that leave node v . [1].

It can be defined as the set of nodes adjacent to node v as neighbors, where the v -node neighbors' set is referenced by $N(v)$, and the node itself may be among the neighbors [33].

2.2.5 Path Description in the Graph

The path P can be defined in graphs as a series of nodes $V(P) = \{v_0, v_1, v_2, v_3 \dots, v_L\}$ or use series of edges $E(P) = \{v_0 v_1, v_1 v_2, \dots, v_{L-1} v_L\}$, that relate to the node's pair $\{v_0, v_L\}$. In a direct network, it is not sufficient that only a pair of nodes may connect, such as (v_j, v_{j+1}) or (v_{j+1}, v_j) ; however, there must be accessible from the source node to the target node. Sometimes called v_0 is initial, and v_L is terminal in the path [34].

An independent term that can be found with paths, nodes, and edges within a graph extracts any connected subset from the parent graph.

2.2.6 The Sub-Graph from a Graph

There are two terms in sub-graph (proper-subgraph and induced-subgraph) where the proper subgraph checks if a single node or path is missing at least from the original graph. At the same time, an induced-subgraph is achieved when there are complete links among the nodes in the induced-subgraph equal to what is in the original graph [34].

2.2.7 Types of Patterns in Graphs

A special type of graph can be formed according to its relationships and links the elements within it. When it shares the same structure characteristics, a chain, clique, and cycle is an example of the graph. A chain graph ch_n consists of n nodes, indexed from 1 to n , node v_t , $t < n$ is linked to next higher index v_{t+1} . When the node v_n is connected to v_1 the graph is cycled with chain. In a clique graph, R_n is consists of n nodes each of which connects to all the others [35].

Figure.2.5 shows an undirected clique graph has 20 nodes and 190 edges. A clique graph can be a complete graph with equations 2.7 & 2.8. where E_{max} denotes the maximum number of edges:

$$E_{max} = \frac{n(n-1)}{2} \quad \text{undirected graphs} \quad (2.7)$$

$$E_{max} = n(n - 1) \text{ directed graphs} \quad (2.8)$$

Where n denoted the number of nodes.

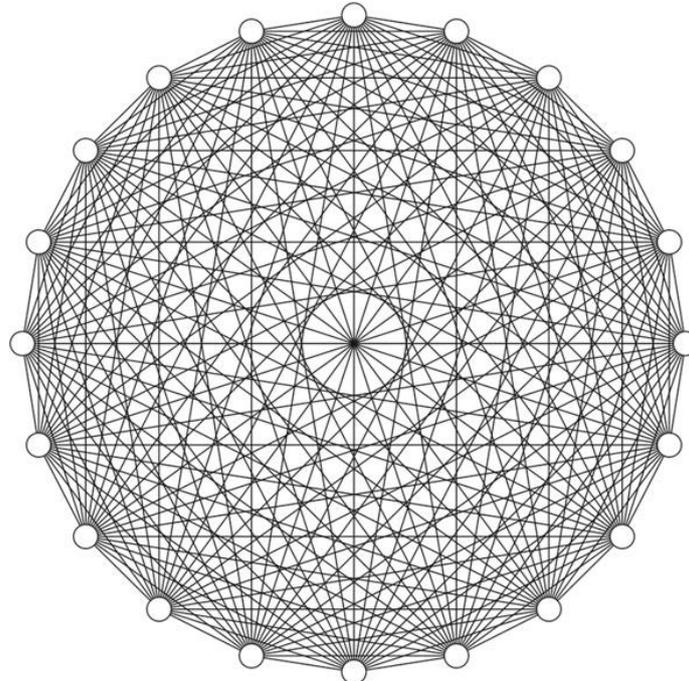


Figure 2. 5 Clique undirected graph having 20 nodes and 190 edges

2.3 Social Networks

Social networks are defined as a social structure made of a group of social entities, a set of links, and interaction among the entities [36]. On the other hand, the social network is a structure that depicts social relationships between entities and is used to describe a global map of inter-relationships between actors [9]. While social networks give outlets for studying social network theories, examining network dynamics, global and local trends, and patterns, and identifying essential entities, they also provide doors for studying social network theories [37].

The world is seen as a collection of linked components by network analysts, and ties are seen as the basic building blocks of social structure by those who research social networks. Every network of connections is joined together to

produce new and promising patterns of links between individuals, groups, and objects [38].

This field is constantly and rapidly updated, the dependence of individuals on it accelerates and presses strongly, so there is an urgent and useful need to analyze these networks and derive the interested directions [9].

Looking at science evidence differently, discovering basic trends that pull away at individual noise to isolate the most functional classes of individuals in the complicated structure at hand is undoubtedly one of the most rewarding experiences a scientist may have [36].

The network's methodology notes that people in similar positions often give the same action even with multiple activities. Network positions can be just as important as the characteristics of the individuals who inhabit them [39].

According to network analysis, reasons for an organization's success or failure may frequently be found in the structure form of interconnections that restrict or give possibilities for interaction among entities [37].

Several important concepts have been associated with network analysis, such as "word of mouth," "friend of a friend," "six degrees of separation." "triadic closure", "transitivity" and "centrality".

It is possible to have more links with a very popular node. Still, a low-popular node remains more influential because it occupies an important position, links multiple groups, or bridges the number of individuals among them [9].

2.4 Social Networks Types

It can be observed that range in social networks with each other in size from simple networks in their components such as nodes and links to larger networks representing a community. It also differs in the form of nodes contained within the connections between the nodes. We'll go through many of the differences that

network scientists have come up with to explain various social networks in this part. These discrepancies impact the metrics and maps that are created and how they are interpreted [36].

2.4.1 Egocentric, Partial, and Full Networks

The 'ego-net' or actor-centered, is a network that exclusively contains people who are linked to a certain ego. The person who is the center of attention is referred to as the "ego," and the others with whom they are associated are referred to as the "alters" [40].

In the analysis of social networks, the 'ego-net' or actor-centered can be described as one of the most abundant networks within any existing network, as these networks appear to form around a particular node [41].

Theoretically, all those connected within a private singular network revolve around a specific type of relationship (emotional closeness, exchange of information or opinions, economic cooperation, etc..). Relationships between alters themselves can be about the same type of relationships that brought them together in the ego network. Therefore, it is possible to deduce the type of relationship to some extent between the alters, even with the absence or lack of information about it [42] [43].

The ego-net can be within multiple degrees, starting from direct relationships with ego node to form the ego network and grading to include friends of friends of alters within a degree (1, 2, 3) [40] .

But it is noticeable that when the degree increases, the network is so wide that it is difficult to control the analysis of relationships. It is also possible to generate within this network sub-ego-net that needs to be analyzed.

In a "complete" or "full" network, only those individuals or entities that possess a particular interest or characteristic are present, as well as information

on the whole set of connections linking them together. In the nature of any complete network, every "ego" is treated equally, with none considered to be the "ego" of the network. On the other hand, analyses of complete networks will show that certain individuals have positioned in the network more advantageously than others [44]. A fully linked network is usually built and put into service such that a social networking platform is used operates as a focal point that connects many other individuals and systems. The Facebook network contains everyone who has signed up for the service and the relationships between them. Due to the inherent limitations of datasets, in many cases, it is neither practicable nor very enlightening to study the scale of entire platform networks based on a single dataset. Instead, analysts select a subset or parts of the full network to produce more selective sub-networks [45] [13]. Figure 2.6 contains part of a network that includes the connections to the college's 389 user accounts. This incomplete network is based on a list of already known other network types search-oriented instead of topic-based. In the process of collecting data, the set of users for whom other partial networks are built (for example, all conference participants) or time-based networks (for example, all persons and connections that happened during a particular time period) or characteristics-based networks (for example, persons who exhibit particular traits) (e.g., Top Ranks of 250 organizations, fan members of a sports team or Members of a local community) [3] [9].

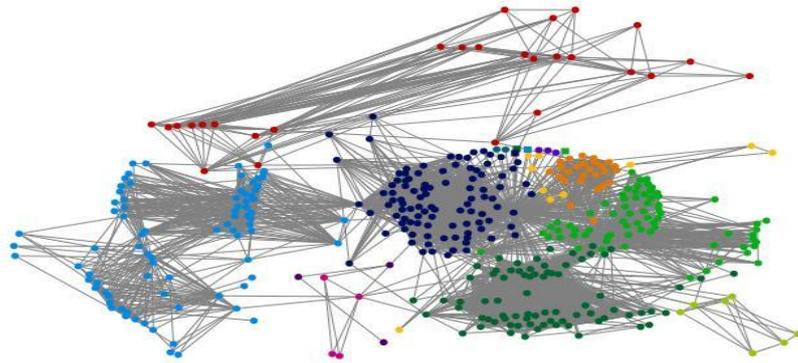


Figure 2. 6 Slice of network with 389 vertices and 3567 edges

2.4.2 Networks of Unimodal, Affiliation, and Multimodal

The primary networks that we utilize are termed unimodal networks since they have one kind of node (e.g., one mode) [44]. These connect people to people or films to films, but they do not cover the combination of people and films. However, a network may have several kinds of vertices, which may be interconnected with various modes of communication (e.g., multimodal) [46].

Compound networks consist of many interconnecting networks often occur in social media settings formed by connections between individuals, photographs, videos, organizations, messages, services, groups, places, and documents. Computational resources, which are usually needed to analyze networks, are sometimes unavailable for multimodal networks. Hence in many circumstances, these complex networks must be simplified to a unimodal network before significant analysis can be completed.

Many forms of recommender systems, which assist consumers in discovering new items of interest, need affiliation networks. When making a

network data structure searchable, it is feasible to provide results such as "persons linked to films are also connected to these films" and "if we connect to films, we may desire to connect to individuals." [47].

2.4.3 Multiplex Networks

. While individuals may expect to have several distinct connections (such as calling each other, emailing each other and being in the same group), most networks only have one sort of edge or connection. However, it is feasible to conceive of networks that support various forms of connections, referred to as multiplex networks [2]

The network may have represented every kind of link by utilizing various link types. In the single-type undirected edge multiplex network, the network data of a multiplex was converted into a network of uniplex that exhibited a single undirected edge if one or more of the four kinds of links were present. Unifying several kinds of edges using this method is rather frequent and mostly used to use network performance measurements, which are usually only available in uniplex networks [48] [49].

2.5 Metrics of Networks

The network element can be viewed at three different levels

- Dyadic
- Nodal
- Group

On this basis, the metrics or attributes pertaining to each component are addressed.

Relational data is the foundation of every social network study. In particular, network data that captures the social relationship between the two

components. (i.e., a dyad). Each of these types of relational data may be either undirected or directed; they could be categorical, valued, or binary; they is either positive or negative [50]; they count interactions or flow of information at a dyad [51]. Through the dyadic level, it can extract information about values, norms, and the behavior of transmissions between nodes.

Although dyadic social relationships serve as the fundamental parts of any network analysis, the analysis itself may take place at a variety of different levels. The three most often used levels of analysis are dyadic, nodal (or individual), and group [52] [53].

Network measures such as network complexity, density, and relatedness enable researchers to examine systematically and detect the patterns of connection inside social networks, establish a basis for comparing networks, follow alterations in the social networks throughout time, and evaluate people's and groups' relative positions inside a social network [54].

Initially, social network metrics concentrated on the most fundamental counts of connections, such as the number of friends or followers, and then eventually expanded into more complex calculation methods as the social network expanded and encompassed ideas of network density, transitivity, centrality, balance, and structural holes [55] [9].

While some metrics reflect the network as a whole, many others characterize subsets of the network. An example of this is the nodes count, which represents the network's node count, whereas the links count indicates the quantity of edges among them. The other entity network metric, called "density" quantifies the degree to which a collection of nodes is connected. This density metric calculates the proportion of links by assuming the maximum density of each node connected to the entire network.

Additional measurements are made at a single node in the social network; as an example, there are several measurements of "centrality." It finds ways to quantitatively assess how "essential" (of importance) a vertex is inside the network and use it to guide design decisions. Some pairs of people are not linked directly in several human networks, including highly linked ones. If a third member connects ("friend of a friend"), this member may be considered a "connector," "bridge," If the individual were missing, the gap might be seen as a "structural hole." [56].

2.5.1 Metrics of Full Network

Several metrics characterize and summarize the whole network. In certain circumstances, a single social network dataset includes sub-networks that are divided into numerous unconnected components. Some network measures function only on networks in which all the nodes are linked into one component. Others, on the other hand, maybe applied to complete networks, especially if they are divided into separate pieces [57] [58].

A- Network Density

The density of a network (Den_G) is calculated by dividing the observed connections by the total number of possible connections. Equation 2.9 calculate the density for the undirected graph [59] [44]:

$$Den_G = \frac{2m}{n(n-1)} \quad (2.9)$$

Where m is observed or real edges, n is number of nodes

It is a quantifiable method of capturing significant social concepts like membership, solidarity, and cohesiveness [60]. Centralization is a statistic that defines how much the network focuses on a single or a few significant nodes. Networks that are centralized feature numerous links branching off from a few

key nodes, while decentralized networks have several nodes with many interactions [61].

B- Maximum Distance (Diameter)

Considering the length of the distance between two nodes is the shortest path, then that is geodesic distance. The longest distance inside connected network, or the geodesic distance between two nodes that are furthest apart, is known as the maximum distance or diameter of the network [62].

Figure 2.7 shows the sample network from the previous network (Figure 2.6), where the Maximum distance value is 4. As an example, the distance that can be considered the shortest path between Jason and Carter is 4; the distance from Owen to Sam, Ayla, Sofia, and Carter is also 4. All other distances are smaller. For example, the distance between Max and Rose is 1.

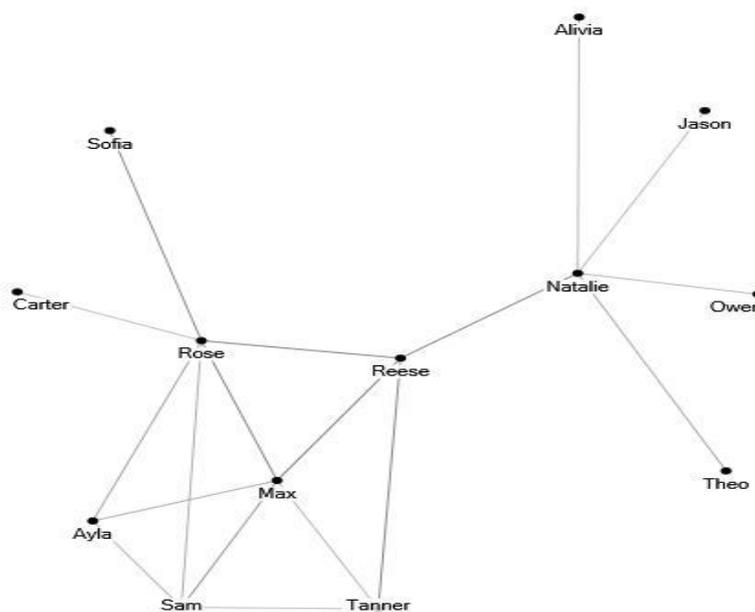


Figure 2. 7 Sample network with node=13, edge= 18

C- Distance Average

This is the average length of the distances between all pairs of nodes. It reveals whether network community members are close to one another. It is

among the most significant considerations for improving networks for performance and power [63]. The distance $Dis(s,t)$ between $\{s, t\}$, where $s, t \in V(G)$ can actually be considered as the shortest path between the nodes, where the average of distance (Ad_G) calculate by Equation (2.10) [64]:

$$Ad_G = \frac{\sum_{i \neq j} Dis(s_i, t_j)}{n(n-1)} \quad (2.10)$$

Where: $Dis(s,t)$ is shortest path between $\{s,t\}$, n is number of nodes in the network.

Returning to the previous figure 2.7, the average distance between members is 2.38. By applying Equation 2.10.

2.5.2 Metrics of Nodes

The sociologist Phillip Bonacich [65], established an improved definition of centrality, which took into account the varied worth a highly linked individual may possess compared to persons with a few exceptions and unusual relationships.

The collection of network metrics is analogous to the concept of latitude and longitude in geography, which are the coordinates that indicate every entity's location inside a network. The most important of these are the "centrality" measurements, which define how a certain vertex may be considered to be in the "center" of a network [66].

Networks scientists observed that simply having a large number of relationships, referred to as "degree centrality," was merely one method to be "in the core" of entities.

An individual who loses links may have more uncommon but possibly "influential" relationships than an individual with a greater number of connections.

In various ways, one link may be more significant than the other. Most are good since they link to well-connected individuals, while some are critical because they link to those who are not well-connected [67].

The centrality metrics listed below give tangible assessments of these ideas.

A. Centrality of Degree

There are many ways of comparing nodes' levels of centrality within a network. The most straightforward –centrality of degree – is to compare their respective numbers of ties. The centrality of a degree is means the number of connections that link it. It is a form of popularity metric, but it's a rudimentary one that does not distinguish both quantity and quality. In addition, for every one of the various node graph metrics, a frequency plot is constructed. When examining vast networks, the frequency plots are quite useful. In Fig 2.8, we use the visualization of the degree to influence the size and shape of the nodes. The mathematical expression for the centrality of degree in Equation [61] (2.11):

$$D. c(v_i) = \sum_{j=1}^{N_v} k_{ij} \quad (2.11)$$

Where K_{ij} denoted the link between node i and j , N_v number of nodes

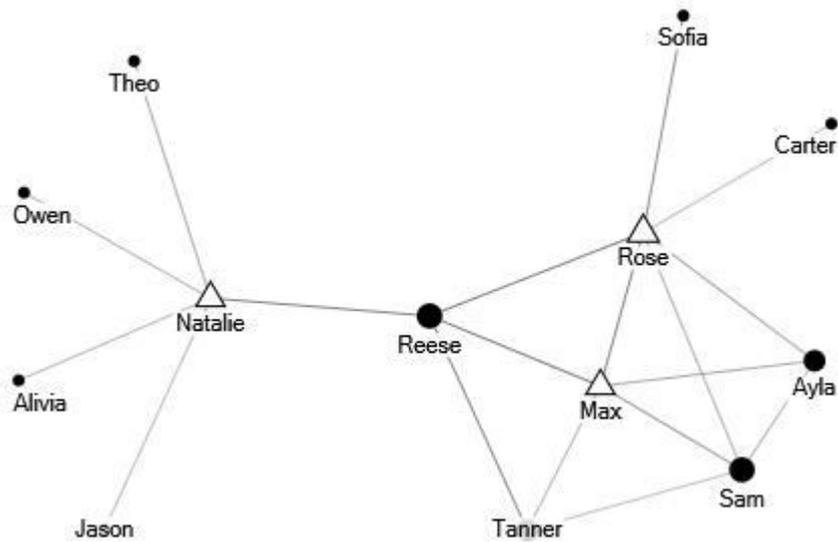


Figure 2. 8 Sample network shows degree controlling size and shape nodes (Nodes with more than four degrees have a triangle, while the rest is a dark circle, the shape size can also be larger with the largest degree)

B. Betweenness Centrality

The betweenness centrality of node v is the portion of geodesic paths between other nodes that v falls on. Mathematical expression for Betweenness centrality in Equation (2.12) [65]:

$$B. c(v) = \sum_{s \neq t \neq v} \frac{Dis_{st}(v)}{Dis_{st}} \quad (2.12)$$

Where: Dis_{st} denoted the total of distance between node s and t . $Dis_{st}(v)$ denoted the total distance between node s and t that pass through a node v . In Figure 2.9, using the visualization of the Betweenness is used to influence the size and shape of the nodes.

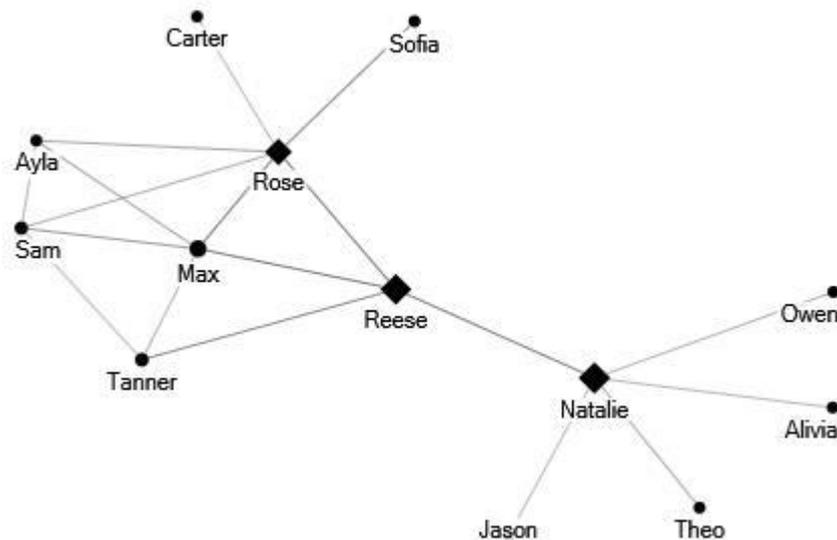


Figure 2. 9 Sample network shows Betweenness controlling size and shape nodes(The nodes with a high Betweenness are a dark square, while the rest are a dark circle: also the size shape can be larger with the largest betweenness)

C. Closeness Centrality

Closeness centrality is defined as the inverse of the sum of the shortest distances between each node and every other node in the network. Mathematically, it can be expressed by the Equation (2.13):

$$C.c(v) = \frac{1}{\sum_{s=1}^N Dis(v,s)} \quad (2.13)$$

Where: $Dis(v, s)$ denoted the distance between v & s . N denoted number of nodes [65] [2].

D. Eigenvector Centrality

With most nodes, it is possible to have more links. Still, a low-degree node remains more influential because it has important links with popular or influential

nodes within a network. This aspect is concerned with the indicator eigenvector centrality. Mathematical expression for eigenvector centrality in equation (2.14) [63] [68] :

$$E.c(v_i) = \frac{1}{\lambda} \sum_{j=1}^N A_{ij} v_j \quad (2.14)$$

Where λ is constant, A_{ij} is an element in the adjacency matrix.

E. PageRank Centrality

PageRank has three unique components, all of which contribute to determining the value of each node. (1) The count of nodes that are connected to the particular node. (2) the centrality of the interconnecting nodes based on the PageRank algorithm and (3) The inclination of connecting nodes to join one another. PageRank examines not just the quantity of neighbors but also the quality of neighbors [69].

PageRank is a common network scoring method in general and on the World Wide Web graph in particular. The literature commonly acknowledges that the PageRank for undirected networks is related to the degrees of the graph's vertices [70].

2.5.3 Centrality Metrics Applications

Social network analysis includes a description of the structure's characteristics and the connections that unite individuals and groups. These characteristics represent concepts in social theories, where patterns, models, and processes of change within these networks appear clearly .

Centrality metrics can be applied in various models to explain a network's characteristics. The following are the most important general outlines of the curricula that have been exploited in terms of metrics:

1. The centrality metrics were employed to describe the network characteristics and overall structure.
2. The centrality metrics are used to test the correlation with external criteria of the node or to test a model based on an analytical basis.
3. Directly metric-based operations can be configured to predict influential nodes within the network [1].

2.6 Influencers in Networks

The detection of influential nodes in complex networks may assist us in solving a variety of complicated but major issues in reality, such as game theory, disease propagation, the management of rumors, the finding of social center characters, and the location of major transportation hubs. Within many networks, nodes differ in the amount of power and influence they have[71-79]

Centralization can be considered one way to measure this effect, and nodes that have a position and value within the centralization metrics can have more control over the events within the network. These nodes are called "important" because they may control the paths of most of the network joints.

Network metrics range from simple such as degree centrality [80] [61], which measures the number of nodes connected to a given node. The local indicators are grading upwards to PageRank centrality [81], H-index employed in extracting influential nodes [82], to be associated with an indicator K-order. Global indicators have also been exploited early as an example closeness [61], Betweenness Centrality [83], and other very significant indicators.

Traditionally, node important ranking metric indices based on the network topology were calculated using locally and globally network attributes, network location, and randomized walks [29].

As a result, local and global network information must be considered together. Indeed, older, more conventional techniques of identification sometimes failed to take into consideration both global and local information about the network simultaneously, which resulted in a loss of useful information and ultimately influenced the final findings; Novel significant node identification approaches have emerged in recent years to address this issue and need additional investigation [14].

The limitations of these simple metrics prompted scientists to develop metrics that go beyond the flaws of these metrics, which are concerned with important theoretical aspects such as information flow, proximity, and control of paths.

The node's importance does not depend only on its location of the node within the network; an important aspect is the node's activity within the network and between its neighbors. Granovetter's [84] describes a strong relationship between two elements who can exchange information in any form for a long time together. At the same time, weak relationships are formed when there is a shortage of communication or exchange between individuals.

Social support is most visible in strong ties, as people who trust each other tend to form strong bonds. While providing new information is through weak relationships.

Differently, Burt [85] presented the idea of structural holes that indicate the lack of interconnectedness between the groups of communities within the networks, and he indicated that this often encourages researchers within the analysis of networks to fill the gaps and gaps to increase the interdependence between those communities. The basic building block of analysis that is supposed to start within networks is a dyad, and it is a pair of nodes where it is possible to examine the behaviors of those dyads.

2.7 Interactions Inside Networks

Connecting between nodes is the basis for the formation of communities within social networks. These communities contain individuals and corporations, which cannot be relied upon to perpetuate these communities and preserve them from scattering. Unless there is continuous interaction and strong relations between its members in terms of exchange of resources, cooperation, competition, support, and conflict. Therefore, we may find great homogeneity in many of these groups; the interactions between their members passed through advanced stages, which generated close ties [37].

Therefore, through increased interaction, we may find a large flow of information inside communities. A considerable scientific effort has been spent analyzing information flow, with most research examining which variables influence information dissemination.

As part of research presented within a thesis, Mark Granovetter [86] discussed interviews with people who wanted to change jobs to determine how they got new opportunities. It was found that many of these people got new jobs through contacts with acquaintances instead of people close to them. This result may be contrary to the expectation that the job comes through close people, not acquaintances. Granovetter found two perspectives to analyze the research: structural, concerned with the expansion of communications between parts of the network, and interaction between others.

In order to develop and comprehend networks based on multiple social media platforms, it is necessary to understand these relationships, or "links" which can be analyzed using tools, mathematics, and insights from social network analysis. Most social media networks' fundamental components may be expressly (explicitly) or indirectly linked (implicitly).

Individuals build explicit links on purpose and deliberately, while implicit relationships are deduced from the specifics of numerous digital paths. Friending

on social networking websites is the most frequent sort of explicit social media connection, in which both sides must consent to the connection before it can be executed. Examples of explicit linkages are, following another person on Twitter, hyperlinking a wiki page to another page, tagging two images or videos with the same tag, and adding someone to a text chat group.

Sociologist Matthew at Cornell University proposed a new concept called “link optimization” that describes two aspects of social relationships: latency and bandwidth, which means intimacy and frequency [44] [87] .

People today leave traces in various ways, forming an extensive network that connects each individual to the other people, places, and digital items surrounding them.

The number of messages, calls, photos, or videos they both send to one another might be used to gauge the quality of their communication relationship. Those weights have included crucial information regarding a tie's strength. [88]

2.8 Diffusion Models

Whenever a type of information flows from one person or group to others in a network, this process is referred to as information spread, information dissemination, information diffusion, or information propagation.

Considerable scientific effort has been devoted to analyzing information spread, with the majority of research focusing on the variables that influence information dissemination.. The process of spreading may be applied to diseases, and similar to the spread of information with the difference that there is no decision within the types of epidemic models [89]. Some features are crucial to comprehending the spreading phenomena. Trying to understand the transmission of information between network nodes, especially if the network is complicated. Why did the information decide to go this way and not follow another alternative path? There are generally three types of diffusion models, two of which depend

on the characteristics of the network directly, namely the Linear Threshold Model (LTM) [90], and the Independent Cascade Model (ICM) [91]. The third type is the epidemic models, which do not directly depend on the properties of the network [92]. The general characteristics of each type can be abbreviated as follows:

LTM - Every edge $\{s, t\}$ has a certain weight, and inactive nodes have a specific threshold. Node v will be activated if its overall weight exceeds or equals to edges weights. [93].

ICM – This model approaches a social network differently, with assume if the node is active, then node having just one chance to convert inactive neighbors to active (the active node can only send a data to inactive neighbors once) with a specific probability $p(s, t)$ [94].

Epidemic models- There are three types of epidemic models (SIR, SIS, SI). The susceptible infected recovered (SIR) model is a traditional infectious diseases model that may also explain the flow of information transfer abstract concepts [95]. The SIR model is partitioned into three groups: those who are susceptible, those who are infected, and those who have recovered. $S(t)$ signifies the number of nodes susceptible to the disease at time t , $I(t)$ means the number of infected at time t , and $R(t)$ denotes the number of recovered at time t . Every exposed node has one chance of getting infected by the surrounding infectors. This change is referred to as the infection rate. Furthermore, each infected node heals to become a recovered node with a probability of recovery rate at each time step. If and only if there is no infection, the preceding dissemination process is ended. While the SIS model differs in that it does not have a recovery period, the node can be infected more than once. And SI model does not have a recovery and the node can be infected once [96].

In this study, a Linear Threshold Model (LTM) was adopted, the value of threshold θ is normalized to be less than 1.

2.9 Pearson Correlation Coefficient

It is an indicator of the overall linear correlation between two variables. The most evaluation process relies on a correlation coefficient (denoted r), as it will give an impression of the extent to which the two values are related together. It is used to measure the correlation of the influential vertices extracted from standard measures with HMEIV_3. Where the scale measures the linear correlation between two variables, it will give a value between + 1 and -1, where a score of + 1 represents the highest value of correlation between the two variables, meaning in other words, there will be a match between the variables. In contrast, score 0 represents no correlation, and score -1 represents the full negative correlation between two variables [25].

2.10 TOPSIS Method

The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is a multi-criteria decision analysis method [97] .

It is a method of accumulating compensation that compares a group of alternates (which are the results of the two indicators at each phase) by giving weight that is equal to both criteria. It starts from the normalized result for each criterion and work inside the alternatives for the Euclidean distance calculation between the optimum alternative and the rest of the alternatives.

The ideal alternative is considered the best among the alternatives at the criteria level [98].

Steps of TOPSIS

Step-1: Creating matrix that contains in each column single criteria, while items in the matrix include records in these columns

Step-2: Calculating normalized for values

$$\bar{X}_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^n X_{ij}^2}} \quad (2.15)$$

Where \bar{X}_{ij} is normalization value of item, and X_{ij} is original value of item within matrix. N number of items

Step-2 Calculate weighted for normalized values

$$V_{ij} = \bar{X}_{ij} \times W_j \quad (2.16)$$

Where \bar{X}_{ij} is normalization value, and W_j for criteria

Step-3: Find the solution best & solution worst from values

Step-4: Calculate the Euclidean distance relative to the solution best

$$S_i^+ = \left[\sum_{j=1}^m (V_{ij} - V_j^+)^2 \right]^{0.5} \quad (2.17)$$

Where S_i^+ is Euclidean distance for ideal best, V_{ij} is the value, and V_j^+ is ideal best.

Step-5 Calculate the Euclidean distance relative to the solution worst

$$S_i^- = \left[\sum_{j=1}^m (V_{ij} - V_j^-)^2 \right]^{0.5} \quad (2.18)$$

Where S_i^- is Euclidean distance for the solution worst, V_{ij} is the value, and V_j^- is solution worst.

Step-6 Calculate performance score

$$P_i = \frac{S_i^-}{S_i^+ + S_i^-} \quad (2.19)$$

Where S_i^- is Euclidean distance for the solution worst, S_i^+ is Euclidean distance for solution best, P_i is performance score.

2.11 Breadth-first search (BFS)

The Breadth-first search (BFS) algorithm is designed to search for a given vertex's tree structure data structure. The BFS algorithm starts at the root vertex and explores all remaining vertices by using queuing. Among the applications of its uses is to find the shortest paths between a specific vertex and the rest of the vertices [99] [100].

CHAPTER THREE

The Proposed System

3.1 Introduction

This chapter contains the steps for creating the algorithm HMEIV. It generally explains the structure of the HMEIV algorithm with all the detailed steps. It also shows pre-processing steps for the datasets used in the HMEIV - algorithm.

Two indicators IWS, CISBC are dealt with and work to create within the proposed HMEIV algorithm, which determines the influential and controlling nodes in transmitting news and rumors and curbing them within social networks.

3.2 Methodology

The first indicator is the (interaction weights scale) denoted by IWS; the first metric depends on the data exchange between nodes. The exchange and transfer of data is important in determining which nodes are actually active at the time the data is extracted. It will also divide the work into three levels:

IWS_1: Runs inside the paths between a pair of nodes (s,t) to extract the exchange contributions of the vertices.

IWS_2: Vertices exchange contributions aggregate for specific node's paths.

IWS_3: Vertices exchange contributions are aggregated for all nodes paths.

The second indicator is (a controlling information spread-based centrality) denoted by CISBC, which depends on the network's structure from fixed relationships. It extracts the important seeds for controlling most of the network paths. CISBC passes through three phases to get multiple visions from bottom to top (from single paths to the entire network) where:

CISBC_1: Runs within the paths of pair (s, t) to find the vertex's contribution, extracts its results, and passes it to the next phase.

CISBC_2: Aggregates the values of the contributions affecting a particular node's paths.

CISBC_3: Incorporating the results and contributions of the nodes for all paths to show the controlling and influencing nodes on the whole network.

After drawing conclusions from both indicators, the results are ready to be combined to extract the seeds that control the data. The algorithm used to derive new ranks for the two indicators is the TOPSIS algorithm. The stage responsible for choosing the best alternatives (IWS & CISBC) and releasing the final results of HMEIV is the qualifying stage, which includes three sequential phases:

Qualifying phase _1 denoted (QP_1): Choosing the best alternatives (IWS_1, CISBC_1) for the vertices affecting the shortest paths between (s,t). And launch the result of HMEIV_1, which is the final result.

Qualifying phase _2 denoted (QP_2): Provide the best results for the particular node's alternatives (IWS_2, CISBC_2). And release the result of HMEIV_2, which is the final result.

Qualifying phase _3 denoted (QP_3): Extracting the best values for the alternatives (IWS_3, CISBC_3) for the whole network.

And launch the result of HMEIV_3, which is the final result.

These phases of the qualifying stage are not connected to each other. The extracted vertices reflect the actual strength of the vertices controlling the paths in terms of network structure within the second indicator. Also, there was tracking and observation of the vertices' activity in terms of data exchange. The first indicator looked closely at all the nominated vertices within the second indicator, whether effective or inactive. Figure 3.1 shows the general outline of the HMEIV algorithm that the two indicators operate in, and their results are combined. The output of the HMEIV algorithm is ranked for results based on the result extracted from the algorithm.

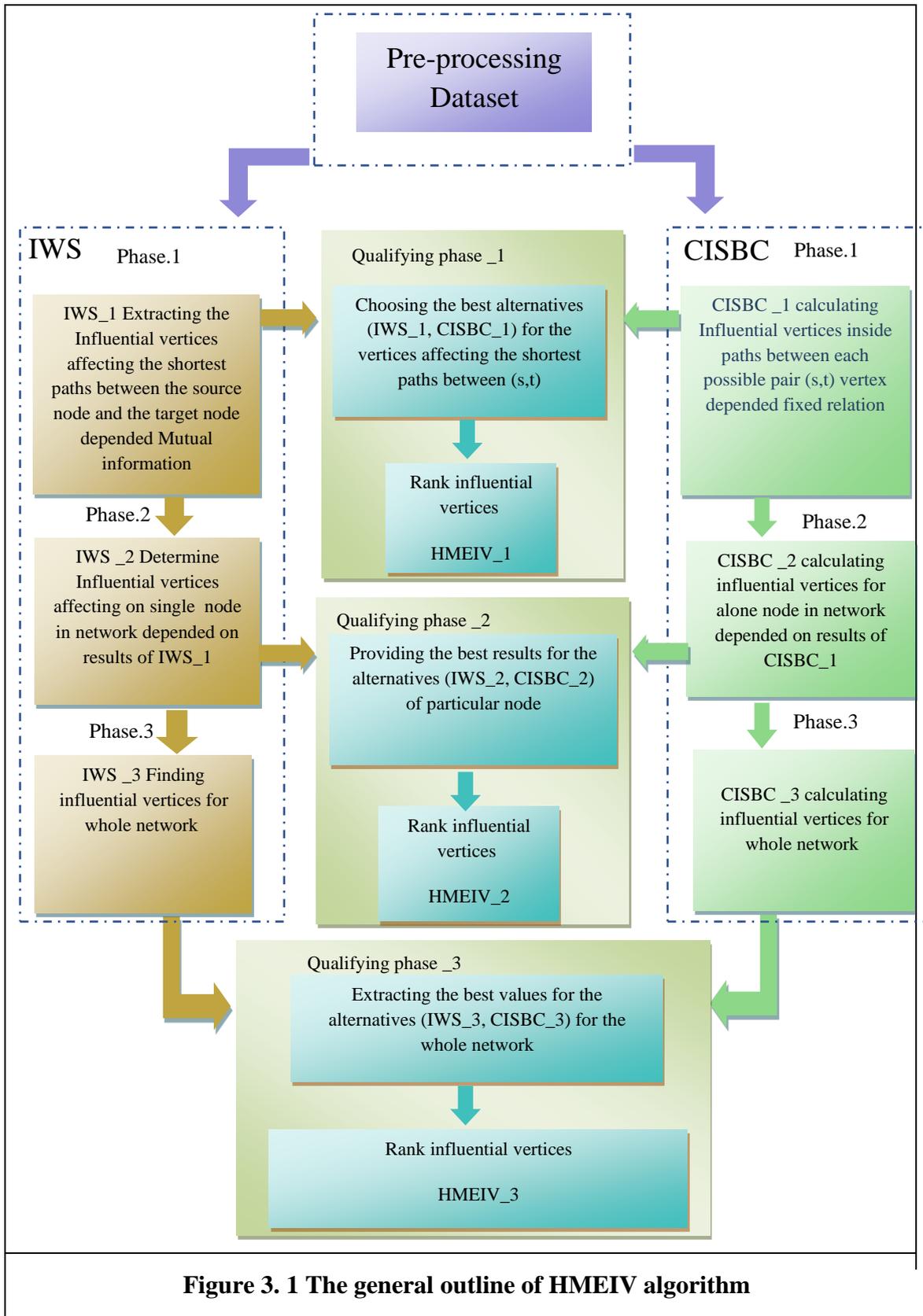


Figure 3. 1 The general outline of HMEIV algorithm

3.3 Pre-Processing Stage

The datasets are real data imported from [101] [102] [103] as a global edge list that expresses the exchange of emails between users. The first edge column is the source node, and the second edge column is the target node. Table 3.1 showing a description of datasets.

Table 3. 1 Description of datasets

Features	CDs_2	CDs_3	CDs_4
Nodes	36692	32430	265214
Edges	183831	54397	420045
Nodes in largest WCC	33696	32430	224832
Edges in largest WCC	180811	54397	395270
Nodes in largest SCC	33696	32430	34203
Edges in largest SCC	180811	54397	151930
Maximum Distance (Diameter)	11	9	14
Average Distance	4.025	4.082	4.899
Density of Network	0.0001365	0.0000517	0.0000597

The processing that was performed on the datasets are as follows:

- Removing isolated nodes from the networks that do not have any connection or edge with the network.
- Finding the frequency of the exchange between nodes in the network.
- Calculating the sum of the frequencies for each node with its surroundings.
- Sorting the nodes from highest rank to lowest rank.
- Cleaning the results from the nodes that are sent in one direction, which shows an imbalance between sending and receiving between the recipient and the sender. After eliminating values that do not demonstrate equilibrium in the trade, the values are ready to be examined using one of the measures of central tendency tests to exclude outliers.

- The seeds with the highest exchange are selected according to the specified sizes within the dataset to form the new network (Ds_2, Ds_3, Ds_4) which corresponds to (CDs_2, CDs_3, CDs_4). That will include the nodes with the highest exchange among them
- In each previous step there are a sequential encoding of the nodes

Followed by the process of finding the overall metric of networks, which have a role in the formation of the network's features. Also, among the features is to find the shortest paths, at this step, Breadth-First Search (BFS) is a very simple algorithm to detect the shortest paths relative to source node s . It applies to all nodes in graphs without any calculated for edge weights. The weights of the edges are calculated will be detailed in the index IWS.

The formation of data processing for the dataset will consist of three columns (Node.1, Node.2, Weight). As shown in the Table 3.2

Table 3. 2 **Sample of Dataset of exchange data**

Node.1	Node.2	Weight
12	3	7
14	24	4
73	5	6
101	8	6
90	7	8
259	55	7
4	1	8

The data set indicates in weight column the number of messages exchanged between individuals in a single connected network. It is possible that these

messages in other cases represent photos or even calls between members. This dataset is stored as a list of global edges. This list is for the source and target of the nodes, with the appropriate number of exchange values in messages, calls, and photos in the third column. Today we can find most relationships weighed. For example, friends are close or not, and this is measured through exchange between them. Cooperation relationships may be close and or loose, and they may also be frequent and rare.

3.3.3 Components of HMEIV Algorithm

The HMEIV algorithm relied equally on establishing two indicators (IWS, CISBC). The results were paired for both indicators to reflect the real results of the vertices within the network that participated and already controlled the paths. The results of HMEIV ranked in order of importance in the network according to the position, and the vertices linked to it. This is the direct effect of an index CISBC, and there is a condition for these vertices to remain in the forefront because they are of high exchange and overwhelming activity; this is what the IWS-Index achieves.

The parts of the algorithm (HMEIV) are:

HMEIV Steps

Input: Undirected graph (Network) $G=(V, E, w)$

Output: As the algorithm contains three levels, the results are divided into:

- Rank Influential vertices between (s, t) in phase HMEIV_1
- Rank Influential vertices for a particular node in phase HMEIV_2
- Rank Influential vertices for the whole network in phase HMEIV_3

Main Process: Three levels

- Calculate the influential vertices by the (IWS) indicator.
- Calculate the influential vertices by (CISBC) indicator.
- Performance of choosing the best alternatives (IWS & CISBC) by qualifying stage.

We will discuss in detail the operations within the HMEIV algorithm.

Inside of steps indicators, each letter L_i means a set i in CISBC & each letter M_i means a set i in IWS.

Also, two terms will be used within the algorithm's steps, although (node, vertex), the node will fall under the influence while the vertex is the one that influences it).

3.4.1 IWS Indicator

The exchange indicator (IWS) between nodes clearly shows the active vertices in the information delivery process. Identifying these vertices is important in finding the nodes seeds of influencers controlling the transmission of data and rumors. The indicator shows what is on the ground inside the network of operations, and it contains three phases:

Steps of IWS:

Input: undirected graph (Network) $G=(V,E, W)$

Output: As the indicator (IWS) contains three phases, so the results are divided into:

- Influential vertices between (s, t) in phase IWS_1
- Influential vertices for a particular node in phase IWS_2
- Influential vertices for the whole network in phase IWS_3

Pre-Processing: Perform a search for the shortest paths by algorithm BFs for each pair {s, t}. source vertex is s; target vertex is t. Exception case s adjacent t.

A. Main method: First phase (IWS_1):

1. Finding all possible connect of vertices pairs {a, b} inside the shortest pathways from source node s and destination node t, denoted by $M_{1,s}^t$ That contains pairs.
2. Extracting the weight of {a, b} $\in G$ by applying Equation (3.1):

$$\forall \{a, b\} \in M_{1,s}^t \quad W\{a, b\} = w, \quad \{a, b, w\} \in \mathbf{G} \quad (3.1)$$

Where: $W\{a, b\}$ represent weight assigned to pair {a, b}, w the value of $W(a, b)$ in graph G.

3. It is possible that there is more than one shortest path between {s,t}. Therefore it is necessary to unify the values by Distinguish and unify weights of pairs {a, b} within shortest paths between s & t by:

$$LP_s^t = \bigcup_{i=1}^k LP_{s,i}^t \quad (3.2)$$

$LP_{s,i}^t$: sequence weights of pair {a, b} in the shortest path i. k number of shortest paths between s & t.

4. Through this step we will invest the weight of the edges to extract the weight of the node. Where the weight of a vertex is created by summation of edges weight that connects. By applying Equation (3.3):

$$d_i(\alpha)_s^t = W(v, \alpha) + W(\alpha, u), \quad \forall W(v, \alpha), W(\alpha, u) \in LP_{s,i}^t \quad (3.3)$$

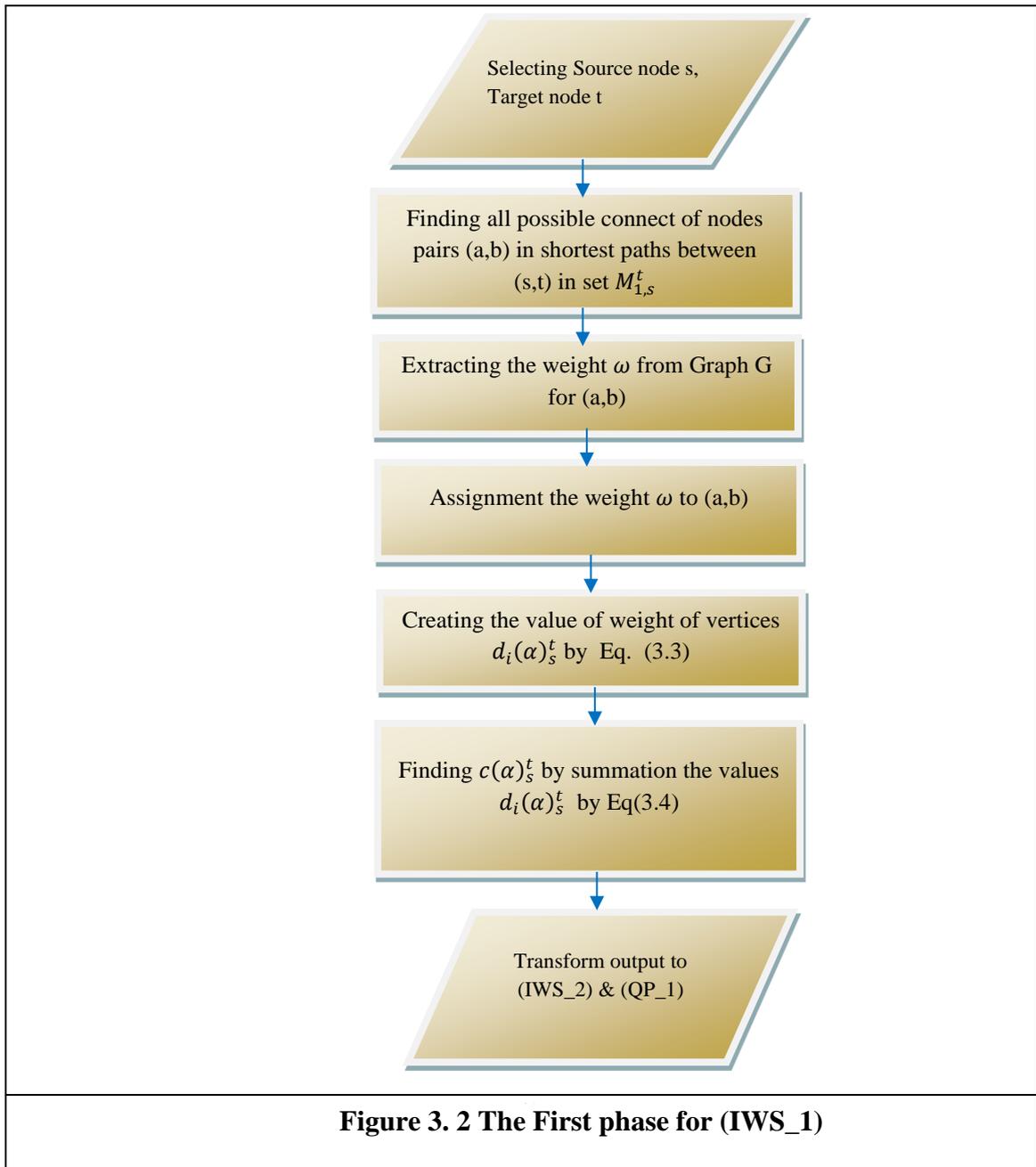
Where: $d_i(\alpha)_s^t$ represent a value of vertex inside of shortest path $LP_{s,i}^t$, $W(v, \alpha)$ represent weights that are assigned to pair {v, α }. $W(\alpha, u)$ Represent

weights that are assigned to pair $\{ \alpha, u \}$. The set of values of vertices denoted by $M_{2,s}^t$ that contains $d_i(\alpha)_s^t$ values.

5. There is a possibility of a node having existence within more than one path, the values of each unique node must be summed through summation the value $d_i(\alpha)_s^t$ in $M_{2,s}^t$. By applying Equation (3.4)

$$c(\alpha)_s^t = \sum_{i=1}^n d_i(\alpha)_s^t \quad , d_i(\alpha)_s^t \in M_{2,s}^t \quad (3.4)$$

Where: $c(\alpha)_s^t$ represents summation values of $d_i(\alpha)_s^t \in M_{2,s}^t$, $d_i(\alpha)_s^t$ represents a value of single vertex α in $M_{2,s}^t$. Figure 3.2 shows the First phase (IWS_1). Create a set $M_{3,s}^t$ that contains $c(\alpha)_s^t$.



B. Second phase (IWS_2):

1. In the first phase, the influencers were identified within a path of (s,t) and in order to complete the consideration and approach each node, we can test all the shortest paths related to a particular node. By completing all possible pairs between the source e (change the source in each time) and the target t in-phase (IWS_1) according to t in G.

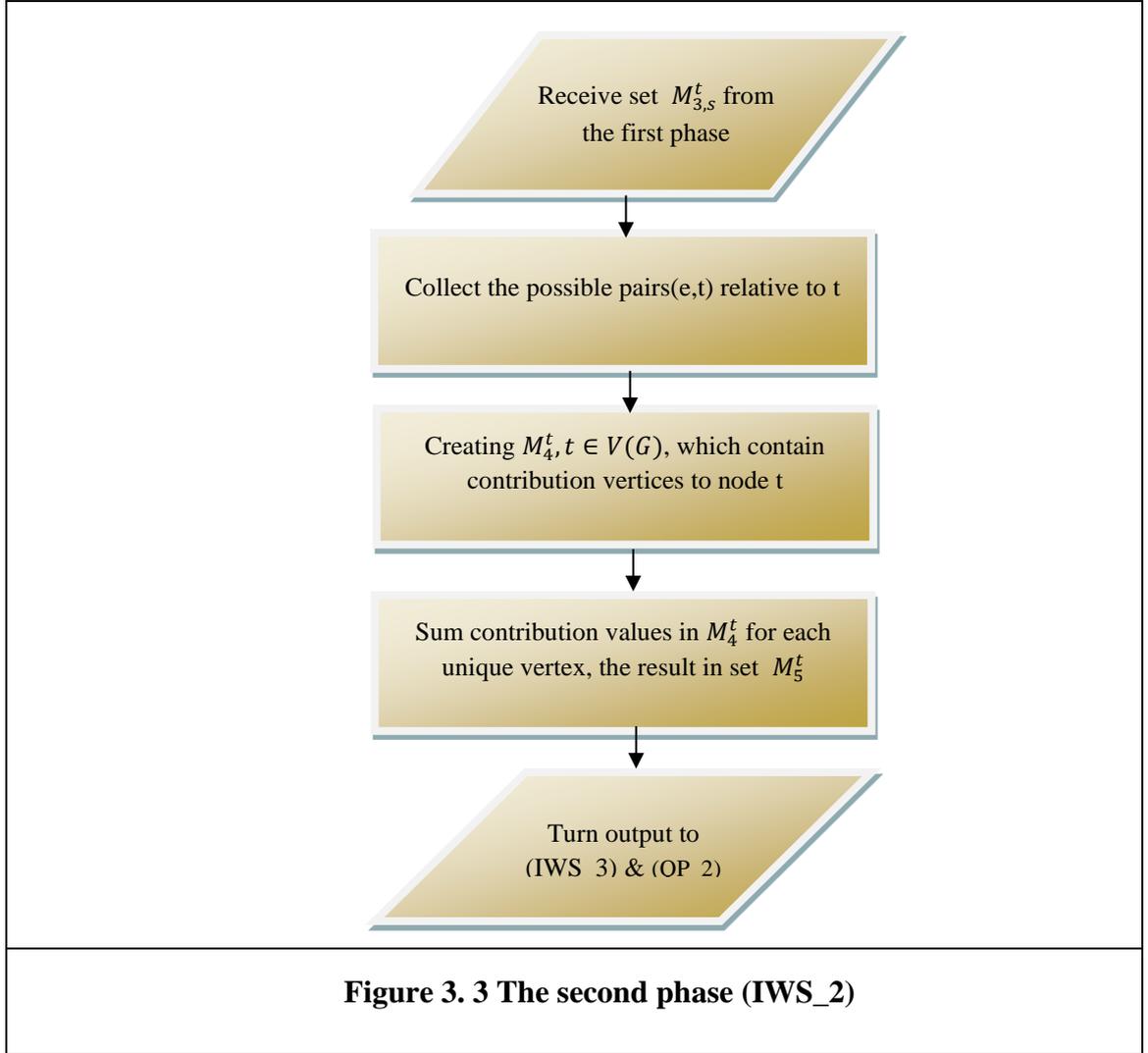
The set $M_4^t, t \in V(G)$ contains each possible $c(\alpha)_e^t$ $\forall \alpha, e, t \in V(G)$ $e \neq t$ the contribution values of the vertices relative to the node t.

2. By aggregating the values of the vertices that appear in the paths of node t, this process can reflect the amount of any vertice's contribution within the paths of node t. For each unique vertex α in $M_4^t, t \in V(G)$ summation the value $c(\alpha)_e^t$. By applying Equation (3.5):

$$r(\alpha)^t = \sum_{e, \alpha, t \in G} c(\alpha)_e^t, \quad c(\alpha)_e^t \in M_4^t \quad (3.5)$$

Creating $M_5^t, t \in V(G)$ which contain $r(\alpha)^t$.

Figure 3.3 shows the second phase (IWS_2)



C. Third phase (IWS_3):

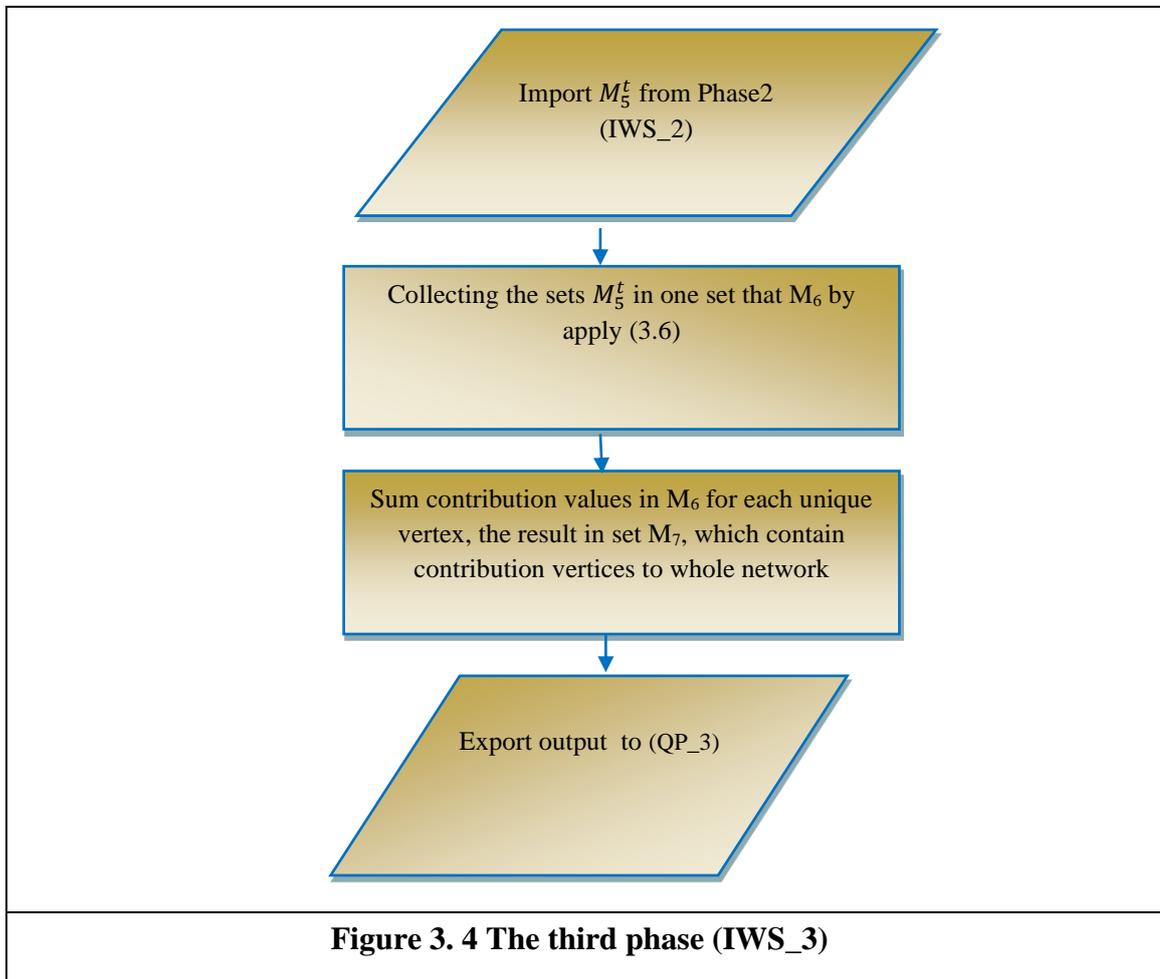
1. Continuity of results is produced by grouping all the M_5^t into one set by collecting M_5^t in one set that M_6 (which collect all Vertices)

$$M_6 = \bigcup_{\forall t \in V(G)} M_5^t \quad (3.6)$$

2. For each unique vertex α in M_6 summation the value $r(\alpha)^t$, create M_7 which contain $y(\alpha)$. By applying Equation (3.7)

$$y(\alpha) = \sum_{\alpha, t \in V(G)} r(\alpha)^t, \quad r(\alpha)^t \in \mathbf{M6} \quad (3.7)$$

Figure 3.4 shows the third phase (IWS_3)



3.4.2 CISBC Indicator

The indicator of (CISBC) between nodes clearly shows the important vertices in network topology; identifying these is important in finding seeds of influencers controlling the transmission of data. The indicator shows what is on the ground inside the network of operations, and it contains three phases:

Steps of CISBC:

Input: undirected graph (Network) $G = (V, E, W)$

Output: As the indicator (CISBC) contains three phases, so the results are divided into:

- Influential vertices between (s,t) in phase CISBC_1
- Influential vertices for a particular node in phase CISBC_2
- Influential vertices for the whole network in phase CISBC_3

Pre-Processing: Perform a search for the shortest paths by the algorithm BFs for each pair (s,t). source vertex is s; target vertex is t. Exception case s adjacent t.

Main method:

A. CISBC_1

1. Extracting all vertices in the shortest paths between s & t inset $L_{1,s}^t = \{v_1, v_2, \dots, v_k\}$ (may be more than one path).
2. Creating set $L_{2,s}^t$ which contain the values of appearance frequency of each vertex in the set $L_{1,s}^t$ inside of shortest paths between s & t that is S_v
3. The statement of the contribution of each vertex within the paths between (s,t) can be by calculating the effectiveness of appearing within these paths can be through Applying Equation (3.8) to find shares for each vertex v from credits in set $L_{3,s}^t$

$$f(v)_s^t = \begin{cases} 1, & S_v = 1, v \in L_{2,s}^t \\ 1 + ((S_v - 1)1/2), & S_v > 1, v \in L_{2,s}^t \end{cases} \quad (3.8)$$

Where $f(v)_s^t$ represent the amount of contribution from each vertex in the paths by frequency to appear in $L_{2,s}^t$, $S_v =$ amount of frequency of each unique vertex in $L_{2,s}^t$. Set $L_{3,s}^t$ is creating which contains $f(v)_s^t$

4. Applying Equation (3.9) to create a distribution for one credit on unique vertices in $L_{3,s}^t$.

$$Mc = \frac{1}{\sum_{j=1}^n f(v)_s^t} \quad f(v)_s^t \in L_{3,s}^t \quad (3.9)$$

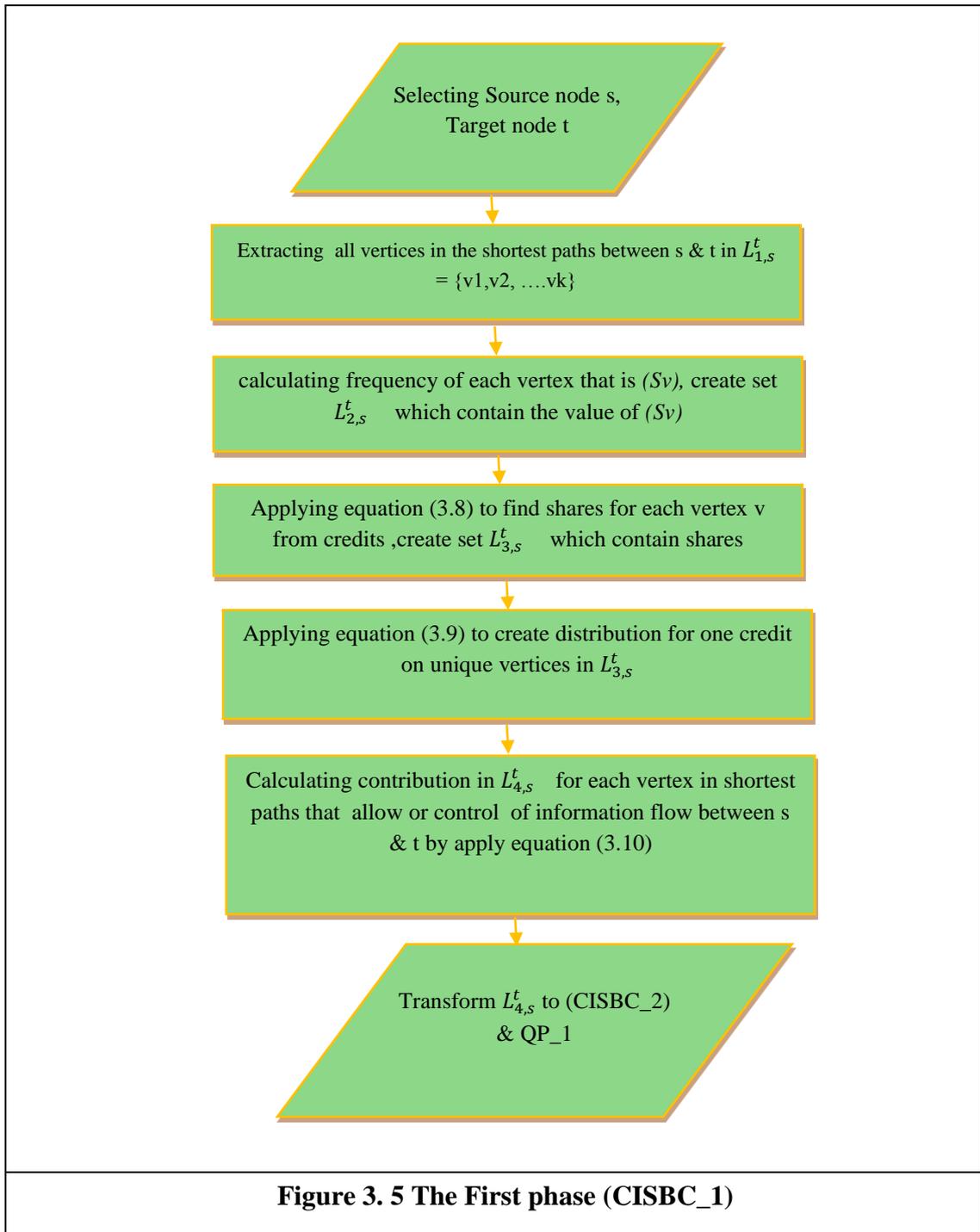
Where: Mc = represents a single credit, n = counter over the set vertices in $L_{2,s}^t$, $f(v)_s^t$ represent of v 's shares from the credits in $L_{3,s}^t$.

5. Exploiting and investing results of Equation (3.8) to form a clear picture of the contribution of each vertex by applying Equation (3.10) to calculate contribution for each vertex in shortest paths between source node s and target node t that allow or control information flow.

$$w(v)_s^t = Mc f(v)_s^t, \quad f(v)_s^t \in L_{3,s}^t \quad (3.10)$$

Where: $w(v)_s^t$ represent the value for each vertex of amount contribution in all paths between s & t which control or allow of information, Mc = amount of single credit, $f(v)_s^t$ shares of each vertex v in $L_{3,s}^t$. set $L_{4,s}^t$ is created which contains $w(v)_s^t$.

Figure 3.5 shows the first phase (CISBC_1).



B. CISBC_2:

1. Completing all possible pairs between the source e (change the source in each time) and the target t in-phase (CISBC_1) according to t in G . The set

$L_5^t, t \in V(G)$ contains each possible $w(v)_e^t \quad \forall v, e, t \in V(G), e \neq t$,

It represents the contribution values of the vertices relative to the node t.

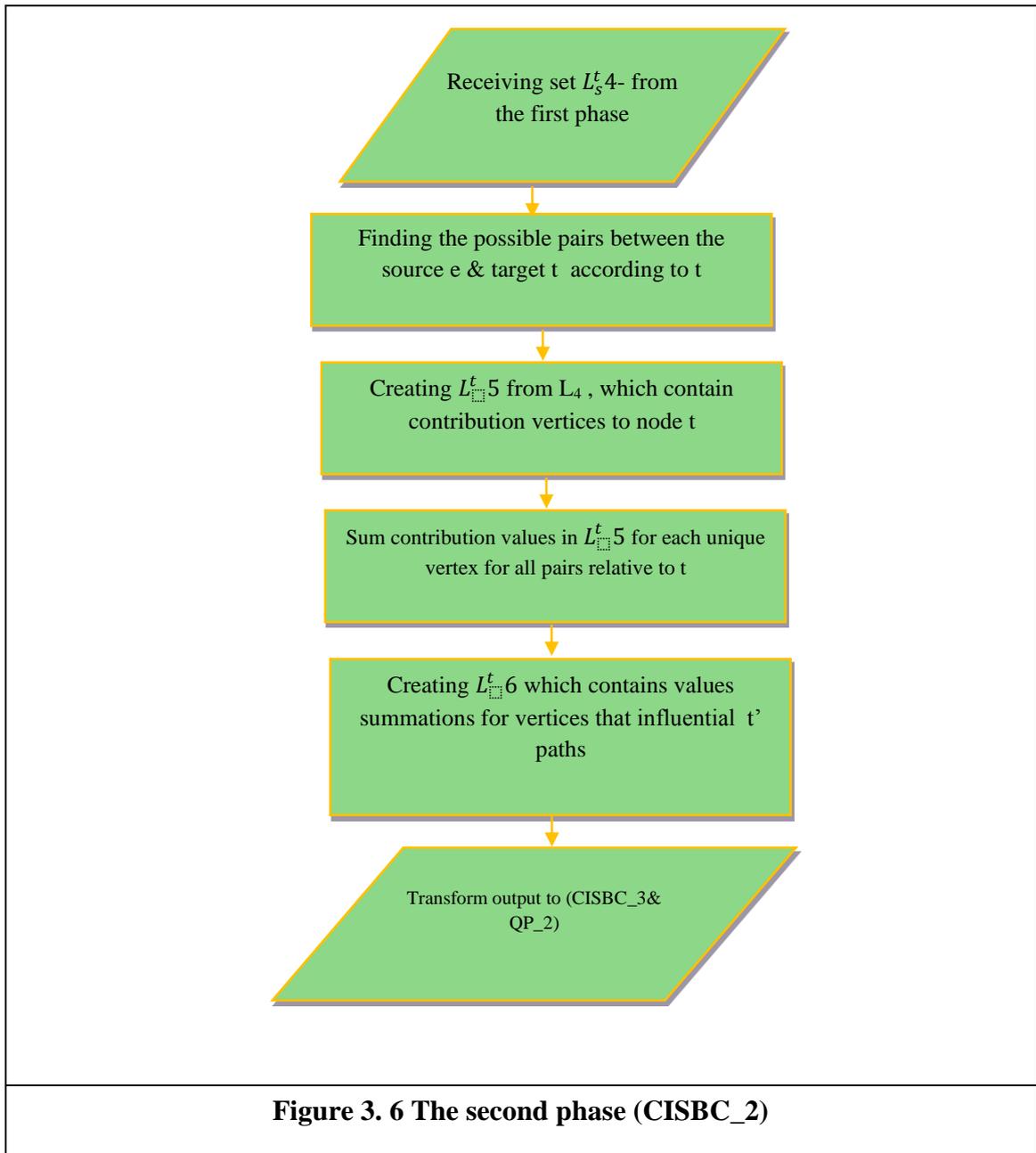
2. Finding the summations of $w(v)_e^t$ for each vertex by applying Equation (3.11).

$$h(v)^t = \sum_{e, v, t \in V(G)} w(v)_e^t, w(v)_e^t \in L_5^t \quad (3.11)$$

Where: $h(v)^t$ represents the summation of one type of vertex (v) in shortest paths between e & t for all pairs relative to t (change the source in each time). $w_e^t(v)$ represents the contribution of vertex (v) in one pair relative to t.

3. Creating L_6^t for each vertex v that contains summations $h(v)^t$ in which that contribution is values of the vertices relative to the t.

Figure 3.6 shows the second phase (CISBC_2)



C. CISBC_3:

1. Collecting the lists L_6^t in one list that L_7 .

$$L_7 = \bigcup_{\forall t \in V(G)} L_6^t \quad (3.12)$$

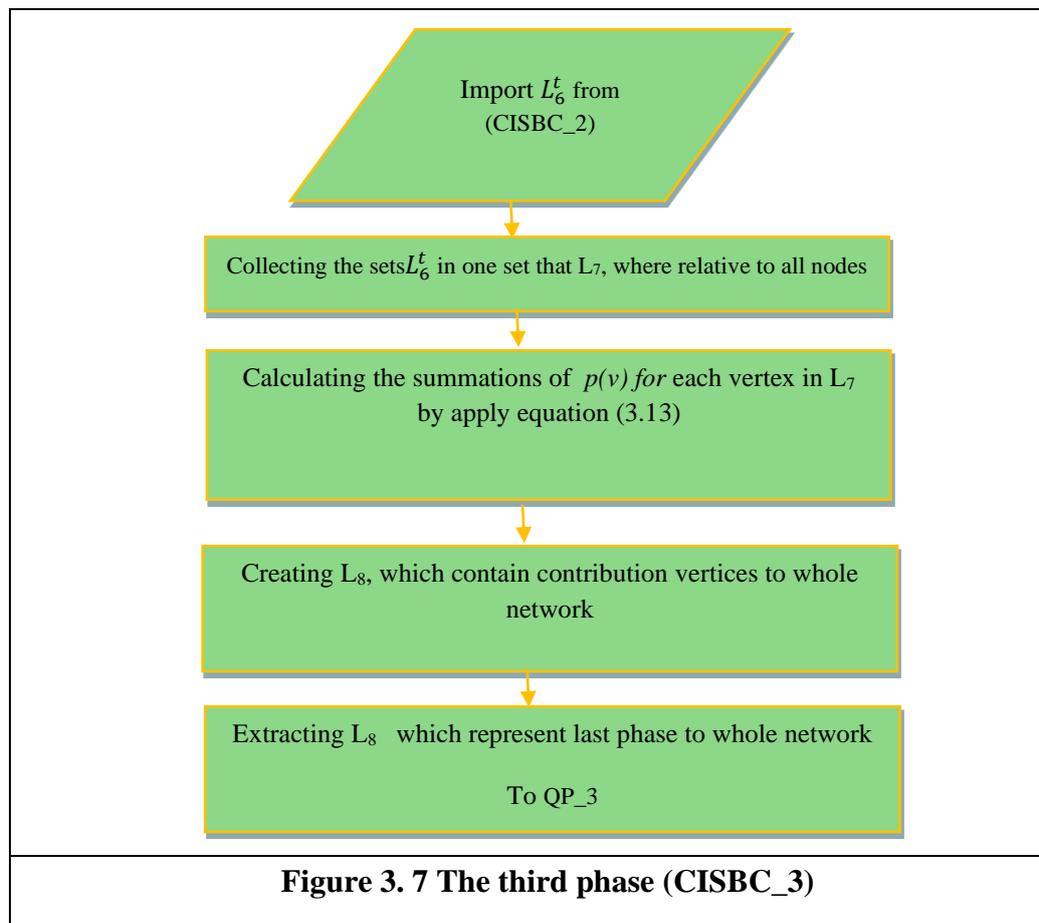
- Finding the summations of $h(v)^t$ for each vertex and creating L_8 by applying Equation (3.13)

$$p(v) = \sum_{v,t \in V(G)} h(v)^t, \quad h(v)^t \in L_7 \quad (3.13)$$

Where: $p(v)$ represents the summation of each vertex (v) in all shortest paths between all pair sources and targets in the whole network. $h(v)^t$ represent the contribution of (v) in all shortest paths relative t . Create set L_8 , is created which contain $p(v)$.

- Extracting values from L_8 that represent contribution values for each vertex in a complete graph (network) and export to QP_3.

Figure 3.7 shows the third phase (CISBC_3).



As shown in the two indicators steps, the algorithm assumes the presence of a source of data or rumor (s) within the network and the presence of a target (t); when the source s and the target t are related to a direct relationship, the algorithm excludes them, as it is not reasonable to have operators inside their paths because there is a previous direct relationship.

3.4.3 Qualifying Stage

In the last stage of the HMEIV -algorithm, after the results have been produced in both indicators and three stages in each indicator, the results pass to the qualifying stage, where the best alternatives outcomes are selected (IWS & CISBC). We have obtained result from both indicators with two different criteria. The first describes the vertices that exchange their surroundings and gives a close picture of those that have activity and those that are inactive. The second indicator describes the importance of the vertices according to their location. Of course, there was a path phase in which the highest control nodes were extracted between the specified nodes pair in both indicators.

In most cases, there was a great convergence in the results for the two indicators at the level of the path phase, and the difference is most likely in the order of the results that occupied the advanced positions.

In the second phase, at the level of a single node whose paths affect the rest of the vertices, the difference in the ordering of the results for the two indicators is greater than the difference in the arrangement of the results.

In the last phase of the two indicators, the results are important at the entire network level, and the results between the two indicators are also different in order, especially if the network is large in terms of the number of nodes and connections. A vertex can occupy first place in the first index depending on its activity, but we find it in fourth or fifth place in the second index depending on its location in the network. Therefore, in the qualifying stage, TOPSIS technology was employed to

choose the best alternatives and exclude the worst, as explained in the second chapter. Where the steps of the qualifiers will be:

Steps of Qualifying Stage

Input: The entry can be one of the three following:

- IWS_1, CISBC_1
- IWS_2, CISBC_2
- IWS_3, CISBC_3

Output: Also, the output can be one of the three following according to the order of the input:

- HMEIV_1: Rank list of Influential vertices between {s, t}, s is source node, t is target node within network.
- HMEIV_2: Rank list of Influential vertices on paths of particular node in the network.
- HMEIV_3: Rank list of Influential vertices of the whole network.

Main Method:

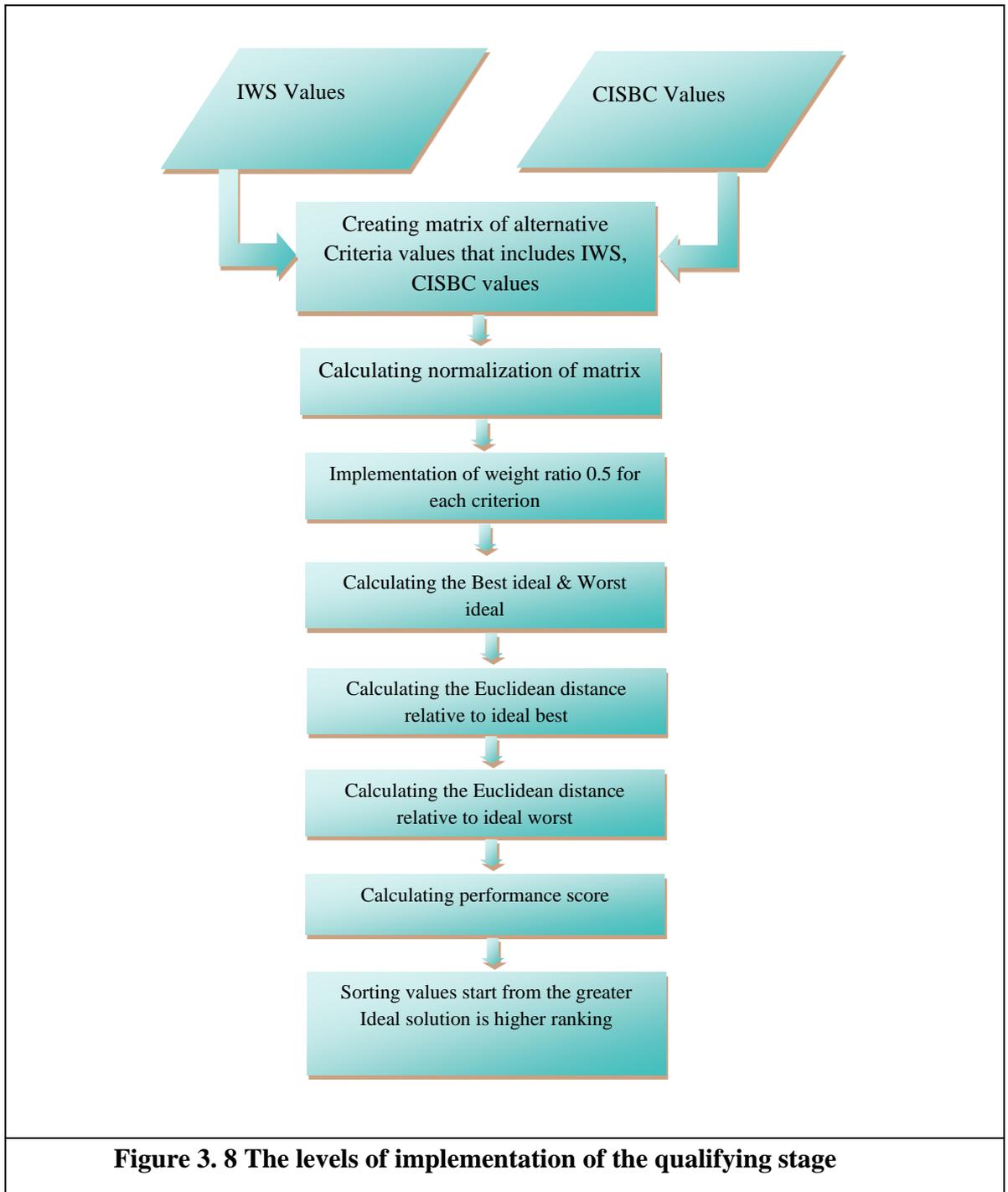
- Creating a matrix of alternative criteria values that includes IWS, CISBC values, which represents two criteria.
- Calculating normalization of matrix values by apply Equation (2.15)
- Implementing of weight ratio 0.5 for each criterion. by applying Equation (2.16)
- Finding the best ideal and worst solution.
- Calculate the Euclidean distance relative to best solution by applying Equation (2.17)
- Calculate the Euclidean distance relative to worst solution by applying Equation (2.18)
- Calculating performance score by applying Equation (2.19)

- Sorting the results to extract the final output.

The a qualifying stage is achieved by employing a method called TOPSIS, that compares a group of alternates (which are the results of the two indicators at each phase) by giving weightiness that is equal to both criteria. The initialization work includes is the normalized result for each criterion and working inside the alternatives to calculate the Euclidean distance between the ideal alternative and the rest of the alternatives. The ideal alternative is considered the best among the alternatives at the level of the criteria. TOPSIS allow for trade-offs within criteria, whereby a weak result can be nullified and excluded by the strong score of another criterion.

The final output of the qualifying stage is the result of HMEIV algorithm.

Figure 3.8 shows the levels of implementation of the qualifying stage and describes the three phases (Qualifying phase _1, Qualifying phase _2, Qualifying phase _3). The equations associated with the steps were explained in the second chapter of the TOPSIS method.



Through the main steps of the HMEIV algorithm and the results which obtained in the next chapter appears, it is clear that the proposed HMEIV algorithm not only specifies the influencing or dominant nodes within these paths dominate, but determines the exact amount of contribution of each vertex within these paths.

The work was done based on the derivation of new equations and new methods. Also, the steps for the indicators inside the HMEIV algorithm gave properties:

- The first phase gives a close-up of a single path from source to target. During this phase, all paths will be grouped based on which vertices affected these paths. This phase is the basis for knowing all the details about the vertices contributing to all the paths.
- The second phase is relative to one node to the rest of the node. It will discover what vertices are affecting the paths of a particular node; each node will have priority nodes that differ from the other nodes.
- The last phase is a global scale for the entire network, as this stage will show the influential and dominant vertices for the whole network with its nodes.

3.5 Reduce Computational Time

Sometimes, especially when we are dealing with huge data, we may need to consider the issue of how to reduce the processing time, so we have two methods that can be implemented both together or one of the following:

- **The parallel implementation**
- **Reduce traffic on frequent paths**

1. The parallel implementation:

It is possible to exploit this advantage when executing on paths associated with a particular node to extract the vertices that control these paths. This implementation can be executed completely independently and is not linked to the implementation or results of another stage.

Therefore, we can practically exploit any implementation on the cloud to execute in parallel for all nodes at once.

2. Reduce traffic on frequent paths:

By looking closely inside any network, we find that some paths within the network are repeated. So, it is practical to find nodes that have the same paths, so we will try to estimate the paths of the nodes at the ends, border or shell of network.

Through the simplest scale, which is the degree of centrality of the node, we find that the node with the smallest number of neighbors is the node with the least influence on the network, and we can find this node through the K-Shell method on the boundary. A simple network can be seen in the following Figure 3.9:

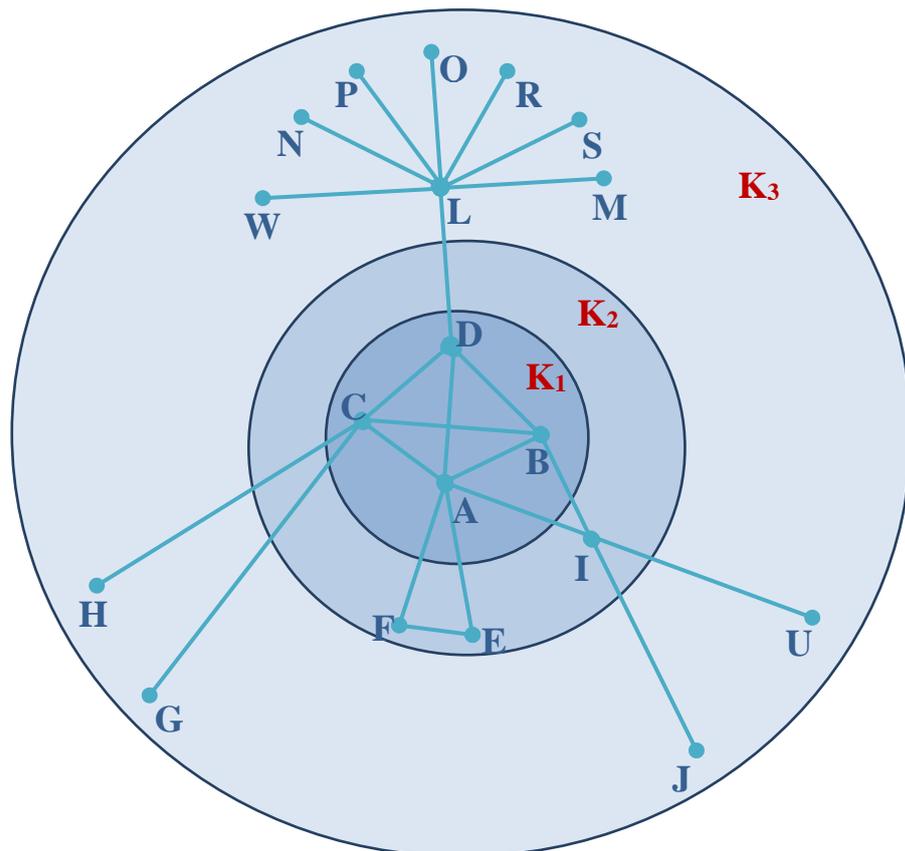


Figure 3. 9 Simple network with K-shell

Figure 3.9. shows the nodes (M, N, P, O, R, S, W) attached to the node L, where any shortest path from these nodes to the rest of the network must first pass through the node L. So, it can simply estimate that all the shortest paths of nodes (M, N, P, O, R, S, W) to the rest of the network are the same as the shortest paths of node L.

From here, the calculations can be performed once on the shared paths of these nodes. The part related to the L node remains, and its importance is calculated by the number of links from the terminal nodes.

CHAPTER FOUR

The Experimental Results

4.1 Introduction

This chapter will review the results of implementing the HMEIV-algorithm on several different datasets for networks. At each stage of the results for datasets there will be a discussion about it, so that there is an explanation for each part. The over metrics for these networks will be extracted, starting with an explanation of the topology network.

After applying the HMEIV algorithm to each network, the results will be evaluated using five experiments:

- I. Comparing the results of HMEIV with the standard metric.
- II. Employing the Pearson Correlation Coefficient which shows us the extent of the bond HMEIV with standard metrics.
- III. Using the frequency of similar values for standard metrics with HMEIV, which shows how the repeatability and differentiation of results can be shown for each metric and HMEIV.
- IV. Susceptible Infected model (SI), which was used as a model to measure the efficiency of diffusion applied to the metrics with HMEIV.
- V. Attacking the network, shows how the impact occurs on the network after deleting an influencer group, by diameter, average distance, and density of network.

4.2 Data Description

Four datasets are used and tested in this study; Ds_1 represents a sample where the algorithm was implemented in detailed steps. The rest of the datasets represent networks for exchanging data or emails between users. Table 4.1 shows the general metrics of the four networks.

Table 4. 1 General metrics of four networks

Features	Ds_1	Ds_2	Ds_3	Ds_4
Network Type	Undirected	Undirected	Undirected	Undirected
Nodes	19	1274	1141	2015
Edges	32	6462	5461	8560
Nodes in largest WCC	19	1274	1141	2015
Edges in largest WCC	32	6462	5461	8560
Nodes in largest SCC	19	1274	1141	2015
Edges in largest SCC	32	6462	5461	8560
Maximum Distance (Diameter)	7	9	10	7
Average Distance	2.903047	3.3172	3.6295	3.547341
Density of Network	0.1871345	0.0079689	0.0083952	0.0042171

4.3 Apply HMEIV on sample network (Ds_1)

Using a simple network (Ds_1) to show the algorithm results in detail, Figure. 4.1. shows the network as it consists of 19 nodes and 32 edges.

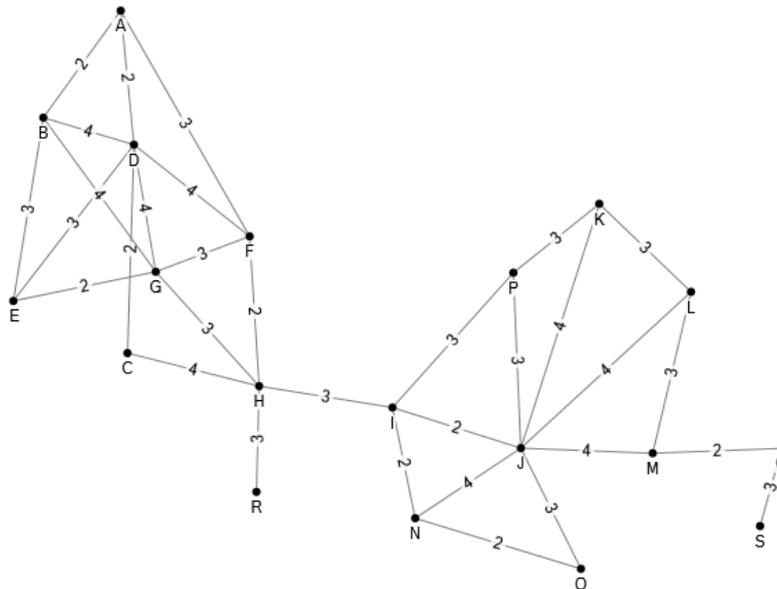


Figure 4. 1 The network topology for the sample network consists of 19 nodes with 32 edges

4.3.1 Steps of IWS Indictor

The work begins with identifying the possible pairs of source node s and target node t between nodes in the preprocessing dataset (exception case s adjacent t , it will not be included). We will also find that the shortest paths between them. An example of pair $\{L, D\}$ is taken. It has five edges for each of the three shorter paths. The shortest paths between nodes L & D are $\{L, J, I, H, C, D\}$, $\{L, J, I, H, F, D\}$, $\{L, J, I, H, G, D\}$. (Note: Each letter M_i means a set i)

A. IWS_1:

- Finding all possible connect pairs $\{a, b\}$ inside shortest paths between (L,D) in $M_{1,L}^D$.
 $M_{1,L}^D = \{(L,J), (J,I), (I,H), (H,C), (C,D), (L,J), (J,I), (I,H), (H,F), (F,D), (L,J), (J,I), (I,H), (H,G), (G,D)\}$
- Extracting the weight for $\{a, b\}$ from graph by applying Eq. (3.1):
- Distinguishing and unifying weights of pairs within shortest paths between L & D , by : $LP_{1,L}^D = \{(L,J) = 4, (J,I) = 2, (I,H) = 3, (H,C) = 4, (C,D) = 2\}$, $LP_{2,L}^D = \{(L,J)=4, (J,I)=2, (I,H)=3, (H,F)=2, (F,D)=4\}$, $LP_{3,L}^D = \{(L,J) = 4, (J,I) = 2, (I,H) = 3, (H,G) = 3, (G,D) = 4\}$
- The weight of the vertex is created by summation of edges weight that connect. By apply Equation (3.3)

$$M_{2,L}^D = \{ d_1(J)_L^D = 6, d_1(I)_L^D = 5, d_1(H)_L^D = 7, d_1(C)_L^D = 6, d_2(J)_L^D = 6, d_2(I)_L^D = 5, d_2(H)_L^D = 5, d_2(F)_L^D = 6, d_3(J)_L^D = 6, d_3(I)_L^D = 5, d_3(H)_L^D = 6, d_3(G)_L^D = 7 \}$$

- Summation the value $d_i(\alpha)_L^D$ in M_2 . By applying Equation (3.4)

$$M_{3,L}^D = \{ c(J)_L^D = 18, c(H)_L^D = 18, c(I)_L^D = 15, c(C)_L^D = 6, c(G)_L^D = 7, c(F)_L^D = 6 \} \quad (\text{vertices that are dominant in pair paths } (L,D)).$$

Figures 4.2, and 4.3 show the results for other pairs $\{M,B\}, \{Q,D\}$

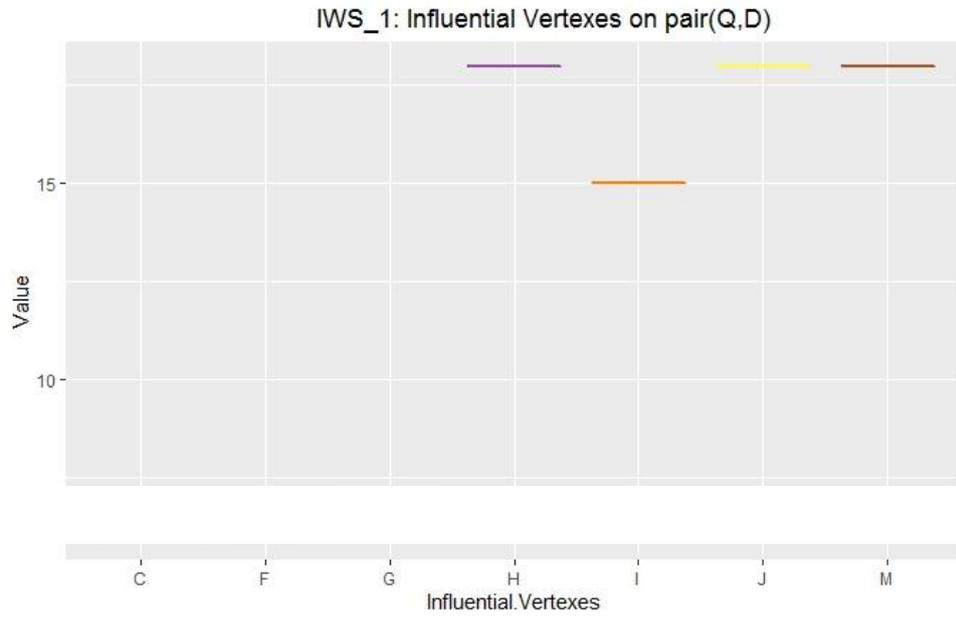


Figure 4. 2 Results IWS_1 of pair {Q, D}

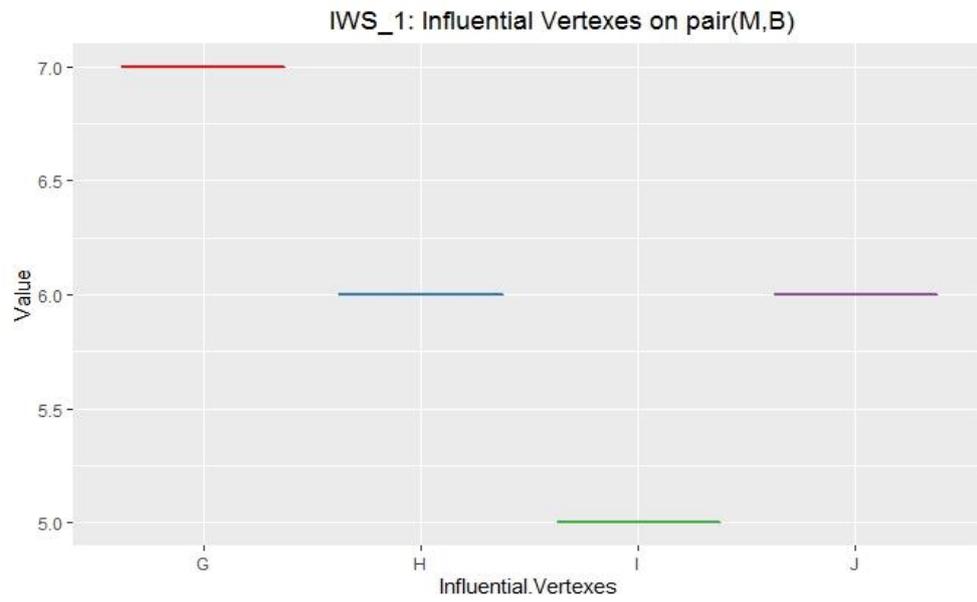


Figure 4. 3 Results IWS_1 of pair {M, B}

B. IWS_2:

- Completing all possible pairs between the source node { K, B, E, A, G, Q, I, M, J, L, C, N, S, O, H, P, F, R } (change the source in each time) and the target node D in-phase (IWS_1) according to node D in G. And create M_4^D Contains the contribution values of the vertices relative to the node D.
- For each unique vertex α in M_4^D summation the value $c(\alpha)_e^D$. by applying Equation (3.5)
- $M_5^D = \{r^D(H) = 234, r^D(I) = 171, r^D(J) = 105, r^D(G) = 98, r^D(C) = 84, r^D(F) = 84, r^D(M) = 36, r^D(P) = 18, r^D(Q) = 15, r^D(N) = 12\}$ (vertices that are dominant in path's node D)
- Also, values relative to node L.
- $M_5^L = \{r(J)^L = 94, r(H)^L = 59, r(I)^L = 55, r(G)^L = 19, r(F)^L = 11, r(M)^L = 10, r(C)^L = 6, r(K)^L = 6, r(Q)^L = 5\}$ (vertices that are dominant in path's node L)

Figures 4.4, 4.5 show the results for nodes N, B

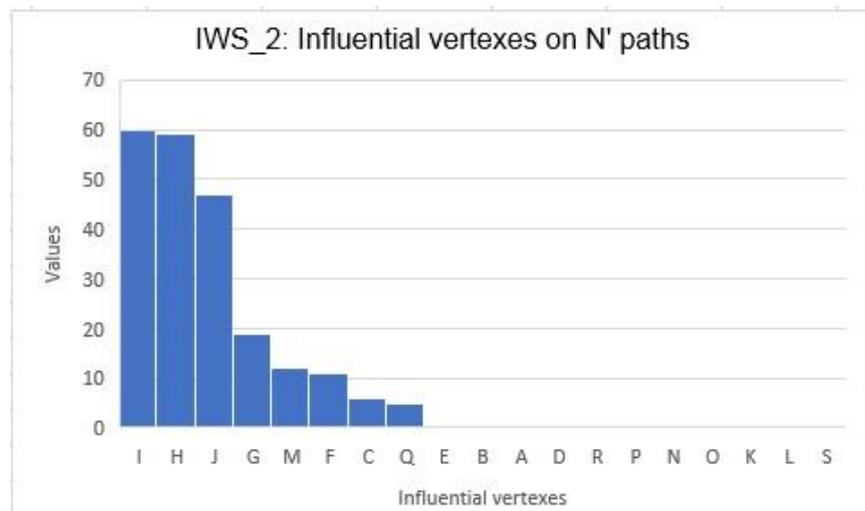


Figure 4. 4 Results IWS_2 of N' paths

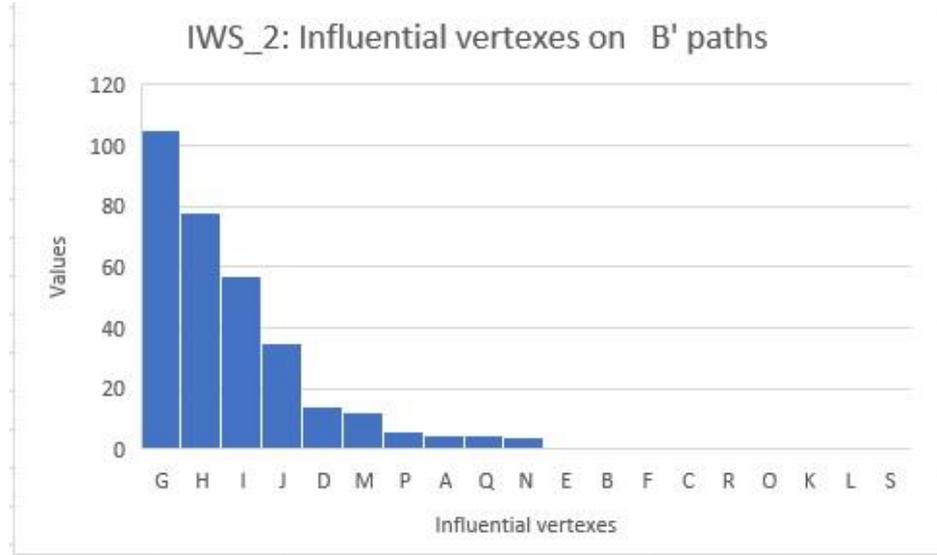


Figure 4. 5 Results IWS_2 of B' paths

C. IWS_3:

- i. Collect the lists M_5^t in one list that M_6 (which collects all vertices)
- ii. For each unique vertex α in M_6 , Eq (3.7) in M_7 is applied

$$M_7 = \{ y(H)=1548, y(I)=1264, y(J)=1120, y(G)=556, y(M)=448, y(F)=320, y(Q)=200, y(C)=168, y(P)=144, y(D)=106, y(N)=96, y(L)=36, y(B)=22, y(K)=12, y(A)=10 \}$$

Figure 4.6 shows the results IWS_3 for the whole network

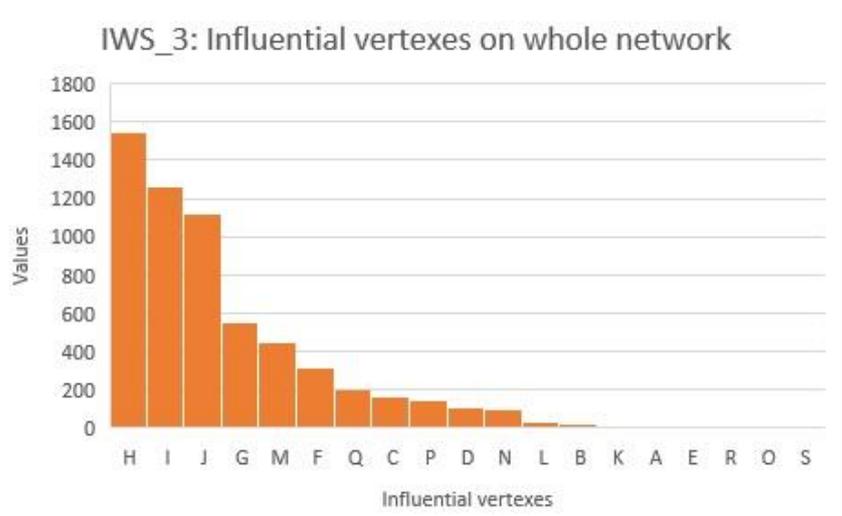


Figure 4. 6 Results IWS_3 of whole network

4.3.2 Steps of CISBC Indicator

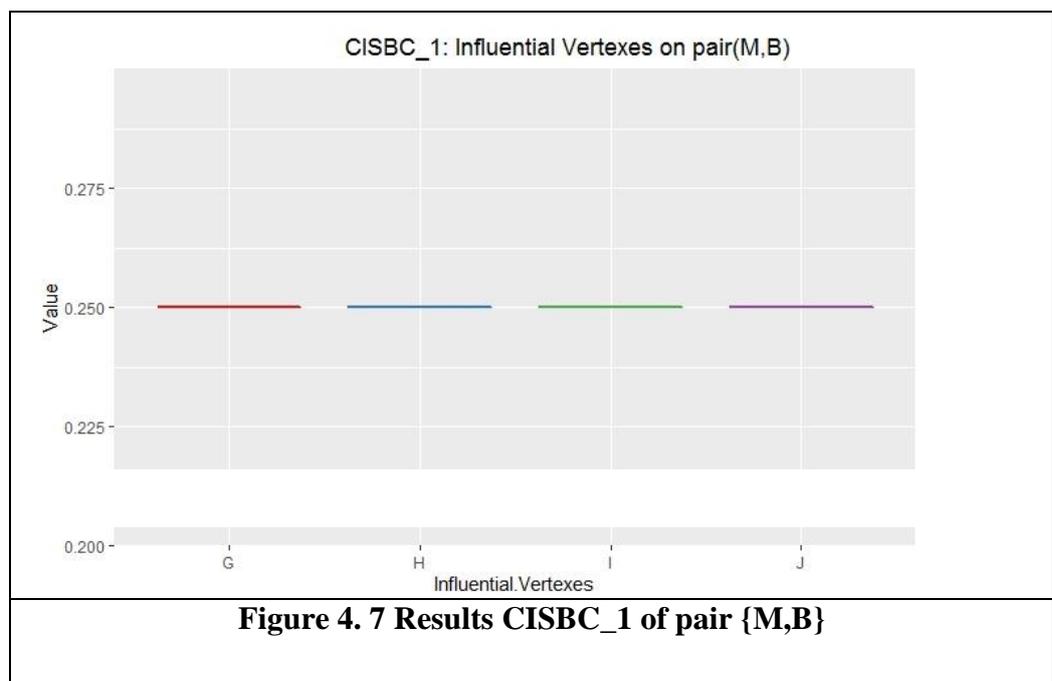
Continue with the same paths and nodes as in the first indicator.

The shortest paths between nodes L & D are {L, J, I, H, C, D}, {L, J, I, H, F, D}, {L, J, I, H, G, D}.

1. CISBC _1:

- It will be as the first step to unify paths and extract vertices between source node L & target node D into one set $L_{1,L}^D = \{J, J, J, I, I, I, H, H, H, C, F, G\}$
- Finding frequency of L1 by $L_{2,L}^D = \{J=3, I=3, H=3, C=1, F=1, G=1\}$.
- Applying Eq. (3.8). The result it: $L_{3,L}^D = \{f(J)_L^D = 2, f(I)_L^D = 2, f(H)_L^D = 2, f(C)_L^D = 1, f(F)_L^D = 1, f(G)_L^D = 1\}$.
- Applying Eq. (3.9). The result $Mc = 0.111$
- Applying Eq.(3.10). The result it by $L_{4,L}^D = \{w(J)_L^D = 0.222, w(I)_L^D = 0.222, w(H)_L^D = 0.222, w(C)_L^D = 0.111, w(F)_L^D = 0.111, w(G)_L^D = 0.111\}$.

There are different results for other pairs, figures 4.7 and 4.8 show the results for pairs {Q, D} and {M, B}.



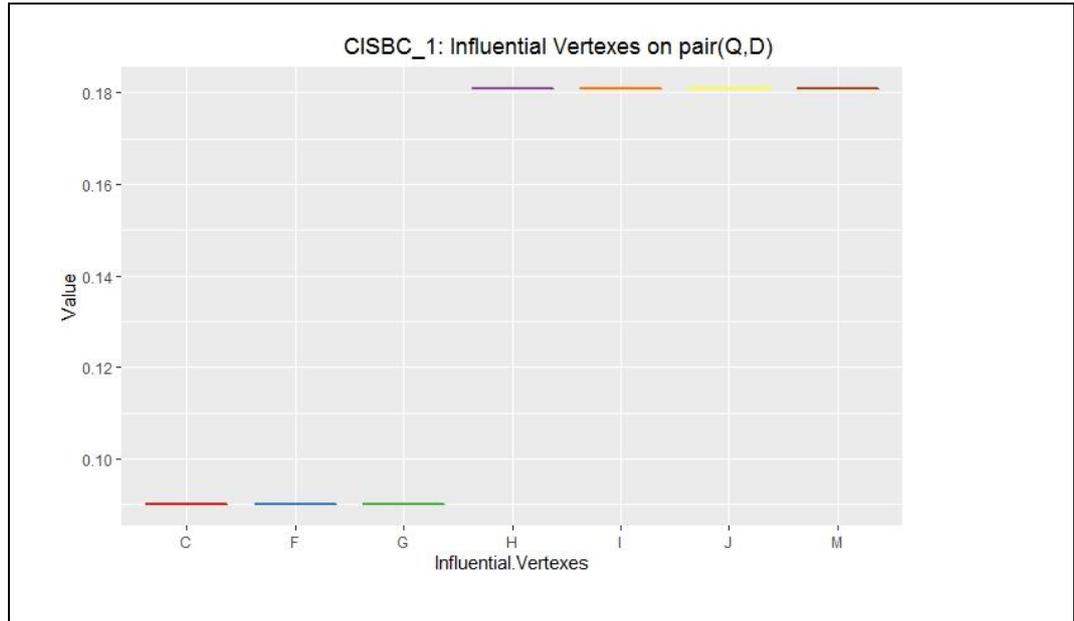


Figure 4. 8 Results CISBC_1 of pair {Q, D}

2. CISBC _2:

- Completing all possible pairs between the source node {K, B, E, A, G, Q, I, M, J, L, C, N, S, O, H, P, F, R} (change the source in each time) and the target node D in phase (CISBC_1) according to node D in G.
- The set L_5^D contains each possible $w(v)_e^D \quad \forall v, e, D \in V(G) \quad e \neq D$ the contribution values of the vertices relative to the node D.
- Finding the summations of $w(v)_e^D$ for each vertex inside the range of node D by apply Equation (3.11).
- The vertices control on the shortest path's node D set $L_6^D = \{h(H)^D = 2.852, h(I)^D = 2.052, h(F)^D = 1.724, h(G)^D = 1.724, h(C)^D = 1.724, h(J)^D = 1.018, h(M)^D = 0.332, h(Q)^D = 0.152, h(P)^D = 0.121, h(N)^D = 0.121\}$.
- While node L has the following vertices control its paths $L_6^L = \{h(J)^L = 6.292, h(I)^L = 2.792, h(H)^L = 2.292, h(M)^L = 1.5, h(G)^L = 0.611, h(K)^L = 0.5, h(Q)^L = 0.5, h(F)^L = 0.361, h(C)^L = 0.111\}$
- Create L_6^t for each vertex v which contains summations $h(v)^t$ that contribution values of the vertices are relative to the t.

There are different results for other nodes. figures 4.9 and 4.10 show the results for nodes N and B

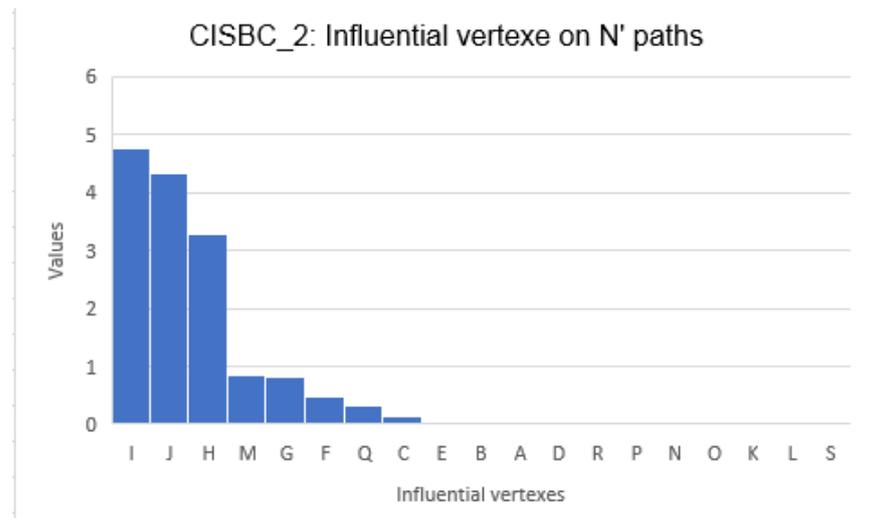


Figure 4. 9 Results CISBC_2 of N' paths

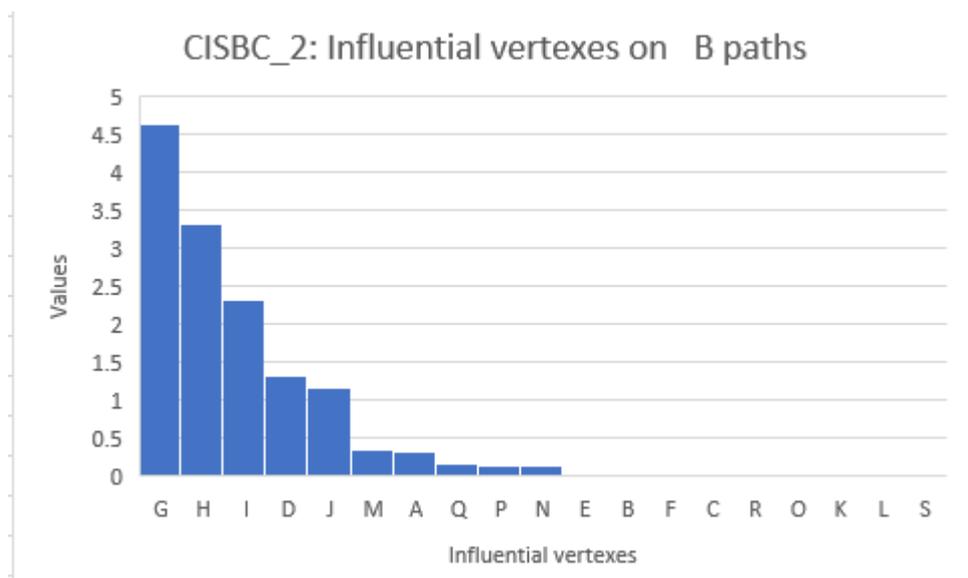


Figure 4. 10 Results CISBC_2 of B' paths

3. CISBC_3:

- Collecting the lists L_6^t in one list that L_7 .
- Applying Eq. (3.13)
- Where the influential vertices in the whole network are in $L_7 = \{ p(H) = 67.48, p(J) = 58.2, p(I) = 55.76, p(G) = 22.47,$

$p(M)=21.76, p(F)=12.81, p(D)= 11.32, p(Q)= 10.67, p(N)=4.35, p(P)= 4.35, p(C)= 3.49, p(L)=1.97, p(B)=1.66, p(K)=1, p(A)=0.66\}$.

Figure 4.11 shows the values of applying the phases CISBC_3 to the simple network (DS_1).

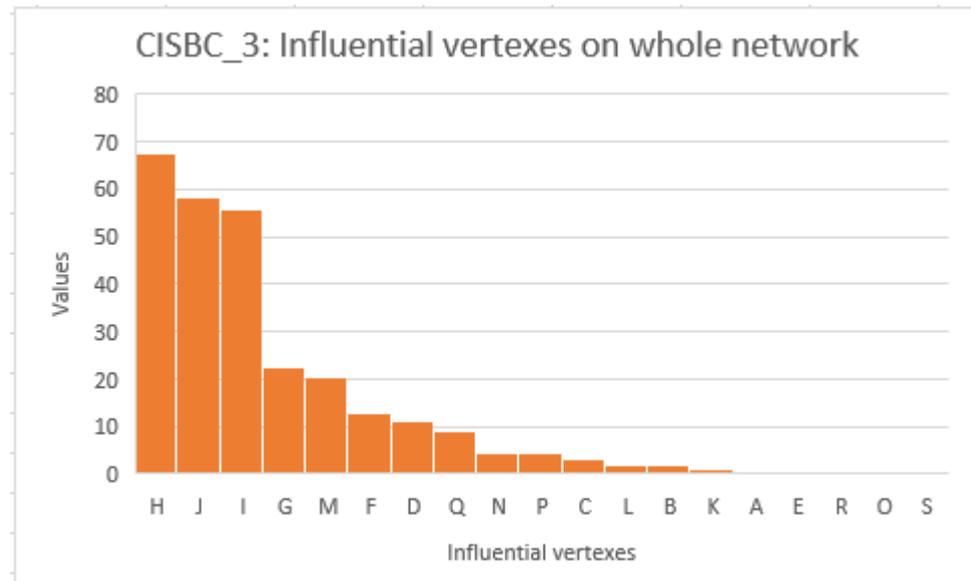


Figure 4. 11 Results CISBC_3 of whole network

4.3.3 Qualifying Stage

The last stage is applying the TOPSIS algorithm (steps are explained in the second and third chapters) on the result of (IWS) and (CISBC) on three phases to create final result (HMEIV_1, HMEIV_2, HMEIV_3).

At this stage the TOPSIS algorithm will choose the best alternatives and exclude the weaker alternatives.

a. Qualifying Phase _1

Completing the results of the paths $\{L,D\}$, $\{Q,D\}$ and $\{M,B\}$, through values (IWS_1,CISBC_1) to obtain HMIDV_1, where the following figures 4.12, 4.13, 4.14 show the results of HMEIV_1.

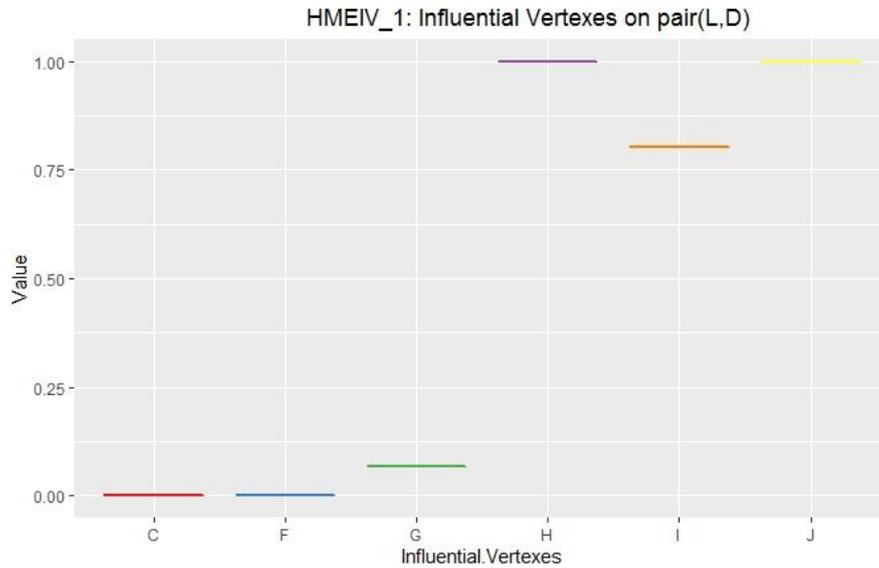


Figure 4. 12 Results HMEIV_1 of pair $\{L, D\}$

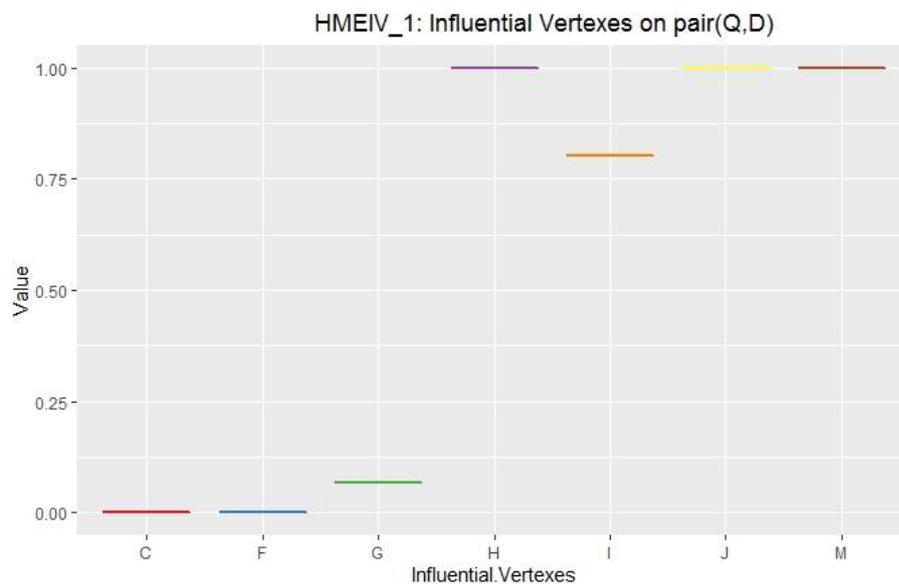


Figure 4. 13 Results HMEIV_1 of pair $\{Q, D\}$

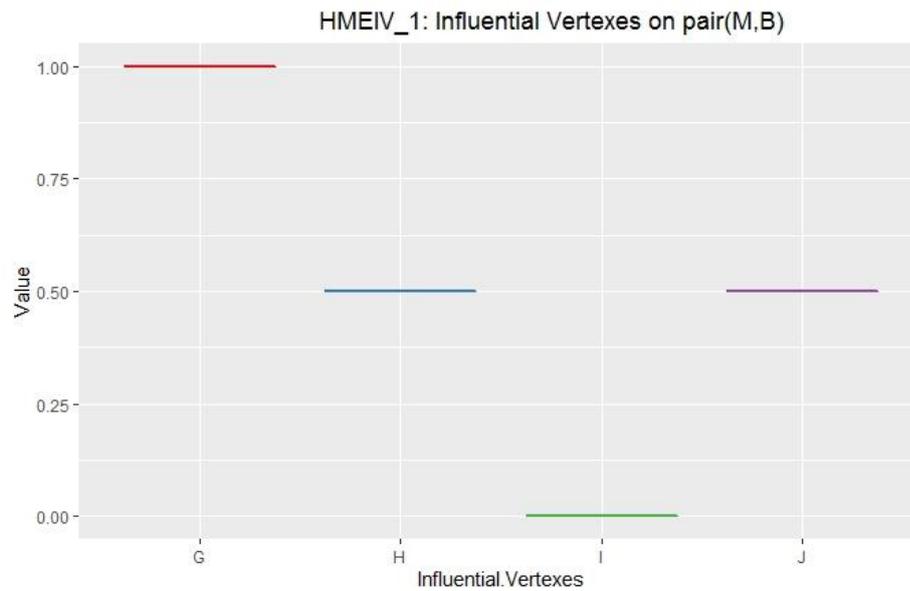


Figure 4. 14 Results HMEIV_1 of pair {M, B}

b. Qualifying Phase _2:

Completion of the same nodes in the results of the indicators (IWS_2) and (CISBC_2) to choose best alternatives that represent (HMEIV_2). Figures 4.15, 4.16, 4.17 and 4.18 show the influential vertices on nodes (L, D, N, B) as results of HMEIV_2.

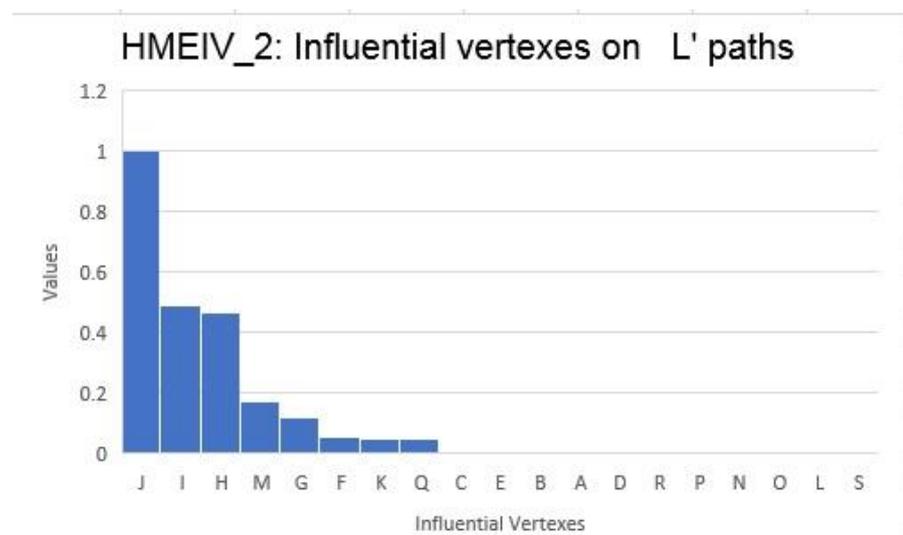


Figure 4. 15 Results HMEIV_2 of L' paths

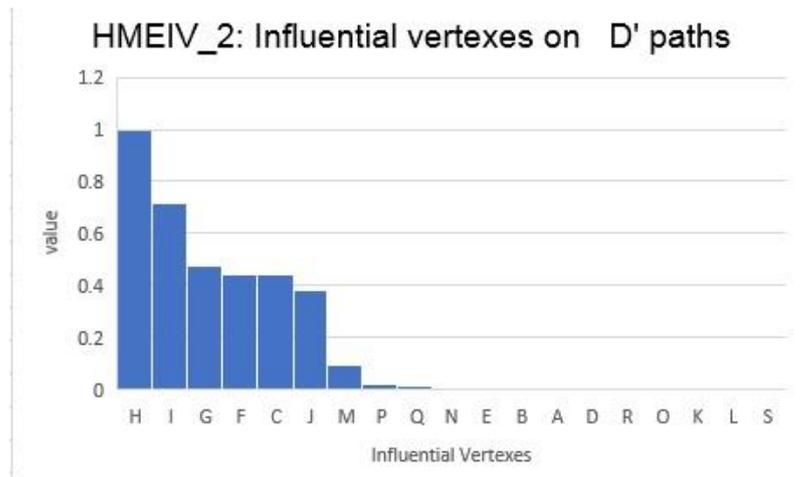


Figure 4. 16 Results HMEIV_2 of D' paths

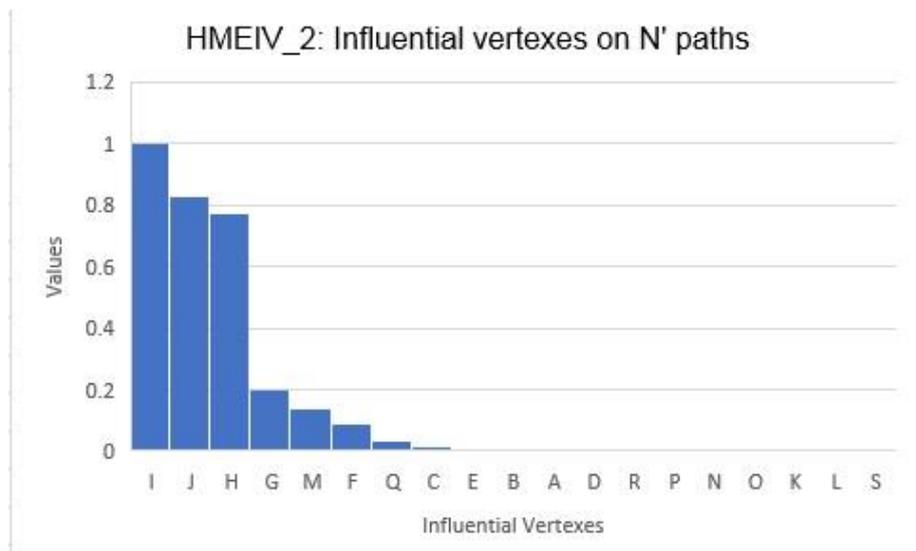


Figure 4. 17 Results HMEIV_2 of N' paths

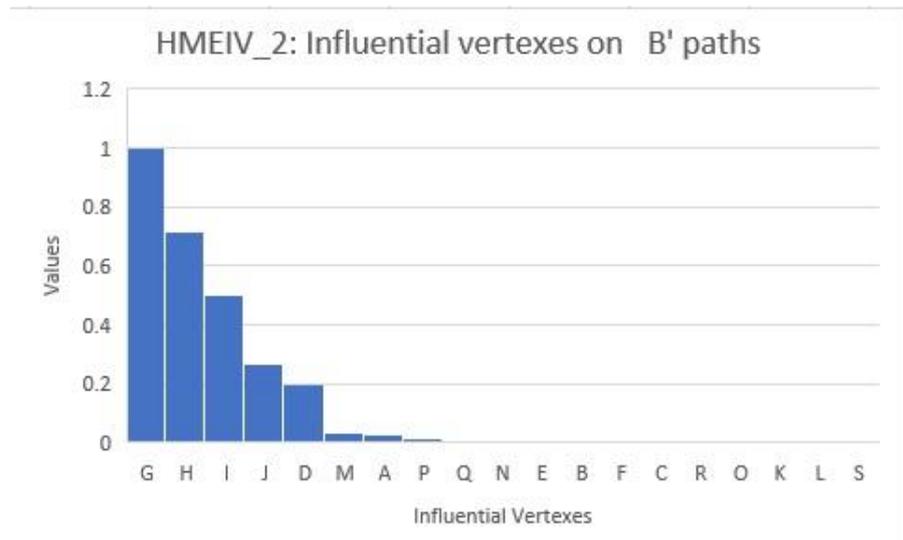


Figure 4. 18 Results HMEIV_2 of B' paths

c. Qualifying Phase _3:

Incorporating results and finding better alternatives between (IWS_3 and CISBC_3) that affect the entire network.

HMEIV_3 is represented in Figure 4.19

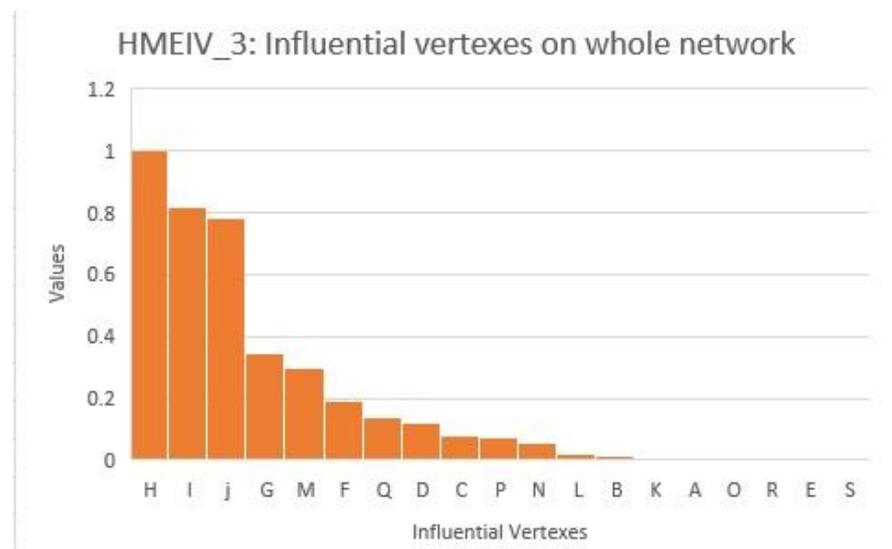


Figure 4. 19 Results HMEIV_3 of whole network

4.3.4 Discusses Results of Simple Network (Ds_1)

Execution of both indicators takes place in parallel. After applying the CISBC_1, that each pair {s, t} has a unique case. For example, a pair of nodes {M, B} in the figure 4.8 have equal values of vertices {G, H, I, J} at 0.25 that affect their paths, whereas a pair {Q, D} in figure 4.7 has high values of vertices {H, I, J, M} at 0.181, which the influential or dominant vertices within this network are identified in the data spread between the intended node pair.

While in indicator IWS_1 for the same pairs, there are different results because this indicator depends on the activity of the vertex in the process of exchanging data within the path. The trade-off between the two indices takes it upon itself in the qualifying phase_1 within the HMEIV algorithm, which it can show the exact amount of influence and contribution of each vertex within the paths that will influence in one way or another the process of selecting the shortest candidate path among the paths that the data will take.

The first phase is the basis for building and completing the remaining parts of the algorithm. In CISBC_2, IWS_2 involves collecting results for each node's paths. These results will be specified within the paths of each node. We will see each node separately with the results of the vertices contributing to its paths. This leads us to determine what vertices are affecting each node's paths; Therefore, this phase is unique in its results compared to the related work.

Each node has a different case. Therefore, as data are transmitted to a single node and according to the HMEIV_2, which combines the results of indicators, can provide a very close view of the influencers and controllers on the paths of each node, so that we will have a clear picture of which vertices have strong access paths to a particular node, through these powerful nodes that appear, the paths that data will take (to and from) this node can be predicted from the rest of the network. On the other hand, it can be detected the vertices have weak control at the level of one node. For example, in figure 4.17 (I) vertex control N's paths at 0.99, while

in figure 4.18 B's paths control by (G) vertex. Similarly, the HMEIV_2 finds the vertices that have the weakest control on the paths of a specific node.

As an example of a different scenario, a node (R), where is the (Q) vertex controls the least paths of (R) node, whereas for paths between a pair {R, Q} the vertices {H ,J, I, M} be controlling paths equally at 0.25. Figure. 4.18 reflects the result of HMEIV_3 for the whole network. There will be a precise measure of the nodes affecting the whole network paths and the with H-node showing the highest amount.

4.4 Results of Ds_2

Although DS_1 was a small network, it showed differences in results at the level of paths and at the level of individual nodes. So, what is the expectation if the network is much larger? We will observe the application of the HMEIV algorithm on Ds_2 and the differentiation of the results. Figure 4.20 shows a the topology of Ds_2.

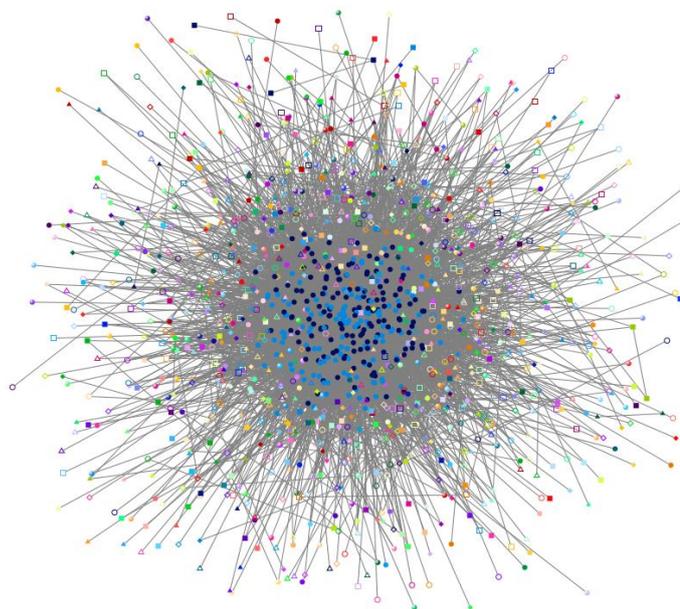


Figure 4. 20 The topology of Ds_2

- For the path phase, Tables 4.2 and 4.3 contain some results for selected pairs.

Node pair	Influential Vertex	CISBC_1	IWS_1	HMEIV_1
{115,1009}	194	0.148148148	23	0.999999846
	444	0.111111111	23	0.751436534
	84	0.111111111	22	0.738614524
	97	0.111111111	18	0.632679113
	788	0.074074074	14	0.348825159
	145	0.074074074	12	0.270049589
	974	0.074074074	10	0.184877144
	203	0.074074074	9	0.14028169
	508	0.074074074	7	0.04773122

Node pair	Influential Vertex	CISBC_1	IWS_1	HMEIV_1
{409,1122}	793	0.4	31	0.999999871
	514	0.2	13	0.181000272
	52	0.2	11	0.110108808
	1239	0.2	8	1.05E-07
{33,788}	1009	0.5	12	0.999999361
	21	0.5	10	1.47E-07

- For individual nodes, Tables 4.4 and 4.5, with Figures 4.21, 4.22 contain some results for selected nodes.

individual nodes	Influential Vertex	CISBC_2	IWS_2	HMEIV_2
454	1228	199.9794119	67524	0.999999999
	77	122.2056272	59798	0.705517798
	814	126.4835179	46193	0.653245511
	828	79.91174237	41054	0.48582211
	1172	53.70806686	28916	0.338619953
	956	69.07138095	20943	0.331112716
	117	51.22446733	17196	0.255508569
	910	41.92549565	12708	0.201008865
	5	39.01031718	13170	0.195025766
	21	1.841963758	20444	0.180424388
	553	10.05396829	12374	0.121052201
	1091	20.09899398	8742	0.113151748

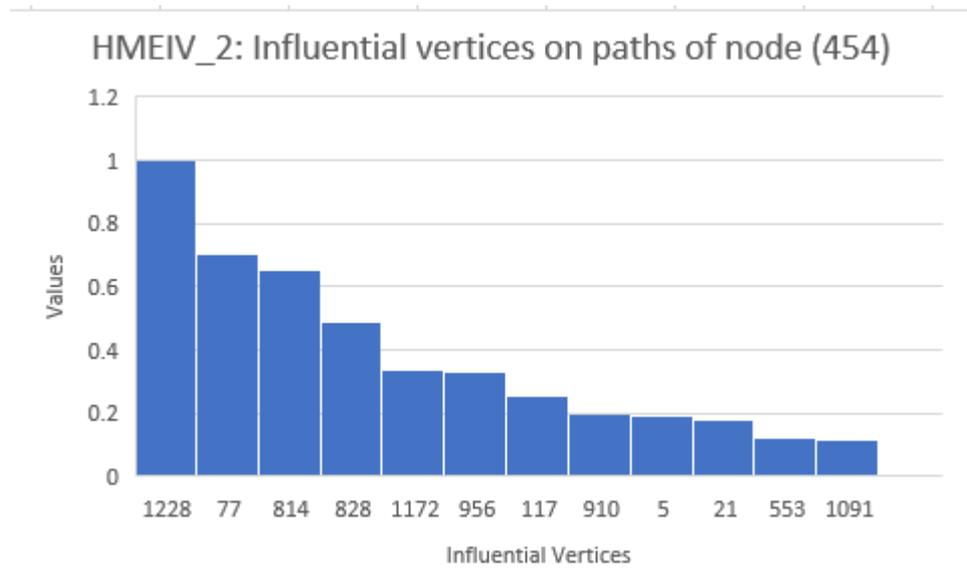


Figure 4. 21 Results HMEIV_2 of node (454)

individual nodes	Influential Vertex	CISBC_2	IWS_2	HMEIV_2
982	194	94.46416033	8134	0.999999999
	145	89.21428401	6345	0.865734857
	1210	74.40632818	7936	0.833446049
	828	56.85950774	7369	0.689837284
	956	52.72316095	5824	0.610932998
	97	44.20537564	4326	0.490315842
	85	29.86294714	6340	0.482505798
	365	42.78978759	4179	0.474288563
	1060	27.71207422	4734	0.402496791
	681	25.27572618	4394	0.372610391
	1220	23.42687422	3798	0.333093252
	203	29.53098079	2858	0.326297962

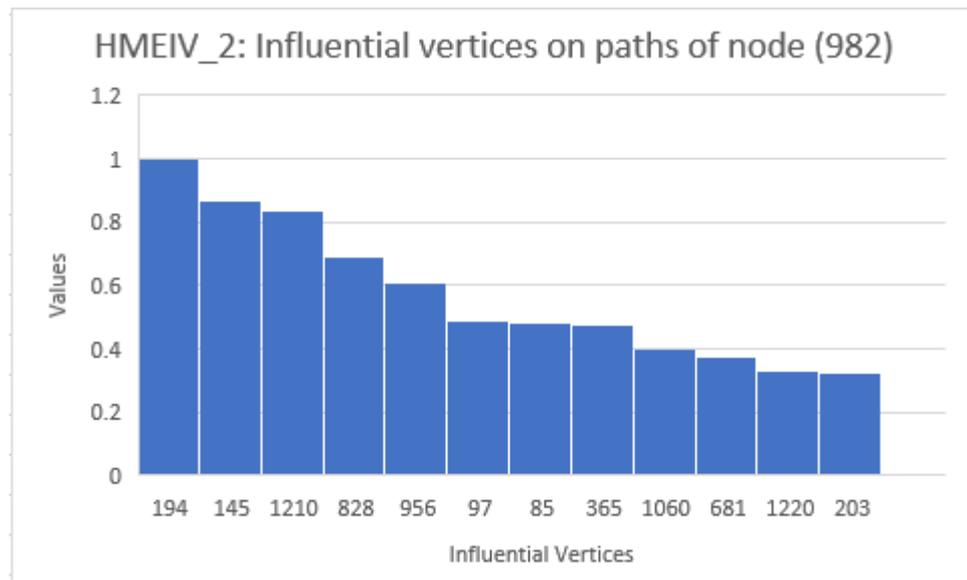


Figure 4. 22 Results HMEIV_2 of node (982)

- For the whole network

Table 4.6. & Figure 4.23 show the results of the whole network.

Table 4. 6 The results of whole network (Ds_2)			
Influential Vertex	CISBC_3	IWS_3	HMEIV_3
194	45557	8956142	0.999999998
145	38544	7212176	0.830610218
828	29892	7622926	0.716301271
7	28458	8109512	0.706470267
334	28654	7789870	0.701854018
430	27792	8301602	0.700553159
44	25808	8635968	0.678006775
1210	26182	6857346	0.637795987
131	20718	7431224	0.578130811
52	19984	6617188	0.542189799
1221	19238	6059242	0.512234914
679	18321	4898884	0.454674221
21	19766	4283824	0.449791806
553	19385	4415396	0.449770203
399	16791	5231798	0.447698231

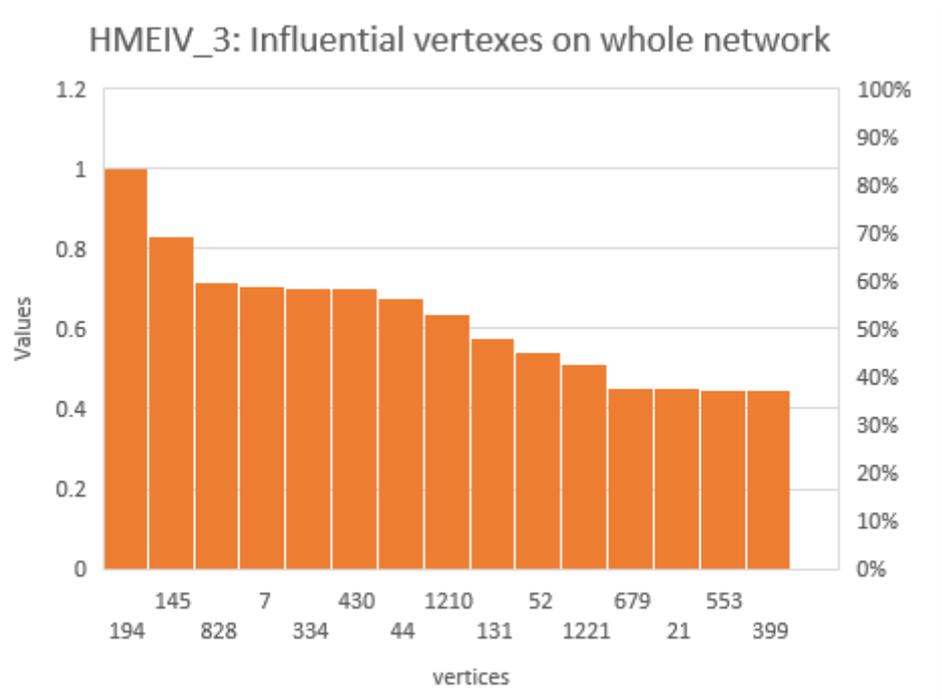


Figure 4. 23 Results HMEIV_3 of Whole network

4.4.1 Discusses of Results (Ds_2)

Working inside the HMEIV algorithm is a systematic approach, where work begins with eradicating illiteracy from paths between pairs of nodes and knowledge of their details. Where the algorithm, with its parts, moves from one state to another to give in a state a visualization of the spatial in which it is located. Various scenarios can be found in each part, so we can see a unique case for each pair $\{s, t\}$. Within the Ds_2, there is more variety than the simple example applied previously. For example, a pair of nodes $\{115,1009\}$ in Table 4.2, where we find the result in HMEIV_1 after the qualifying process for the two indicators, the top ranks are occupied by influential vertices $\{194,444,84\}$, it is followed by six other vertices that influence the path. In pair $\{409,1122\}$ in Table. 4.3 have a different scenario where share three vertices $\{514,52,1239\}$ at 0.2 and top rank $\{793\}$ at 0.4 within indicator CISBC_1, where the exchange value within the indicator IWS_1 has a differential position, to be reflected in the final values of HMEIV_1 and the results are $\{0.999, 0.181, 0.1101, 1.05E-07\}$, where this pair has three

paths $\{409,514,793,1122\}$, $\{409,52,793,1122\}$, $\{409,1239,793,1122\}$. Through the results of the HMEIV_1, we can give a high probability of data passing on the first path.

The last path is $\{33,788\}$, where two influence vertices $\{1009,21\}$ share at (0.5) within CISBC_1, to settle the IWS_1 at values $\{12, 10\}$. The path's construction results are the basis for all subsequent stages in the algorithm HMEIV, where the influential or dominant vertices within this network are identified in the data spread between the intended node pair. While in indicator IWS_1 for the same pairs, there are different results because this indicator depends on the activity of the vertex in the process of exchanging data within the path. The trade-off between the two indices takes it upon themselves in the qualifying phase of HMEIV_1 algorithm, which can show the exact amount of influence and contribution of each vertex within the paths that will influence in one way or another the process of suppressing the rumor or increasing the spread of the data.

In CISBC_2, IWS_2 collects results for each node's paths by applying Eq. (3.6) and (3.12) These results will be specified within the paths of each node. We will see each node separately with the results of the vertices contributing to its paths. This leads us to determine what vertices are affecting each node's paths; Therefore, this phase is unique in its results compared to the related work. Each node has a different case. Therefore, as data are transmitted from a single node and according to the HMEIV_2, which combines the results of indicators, can provide a very close view of the influencers and controllers on the paths of each node, so that we will have a clear picture of which vertices have strong access paths to a particular node, and also, on the other hand, it can be clear which detect the vertices have weak control at the level of one node. For example, in Table. 4.4 Vertex (1228) control node (454)'s paths at 0.99 in first rank, while vertex(77) is ranked second although it has a lower value than the vertex (814) within CISBC_2. Because it has a high value in the index (IWS_2), it ranked second.

In Table 4.5. node (982)'s paths controlled by vertex (145) is ranked second, although it has a lower value than the vertex (1210) within IWS_2. Because it has a high value in the index (CISBC_2), it ranked second.

Similarly, the HMEIV_2 finds the vertices that have the weakest control on paths of a specific node. As an example of a different scenario, a node (982) in Table 4.5. it is clear that the vertex (203) controls the least node paths (982), Table 4.6. reflects the result of HMEIV_3 with the indicators (CISBC_3, IWS_3) for the whole network. There will be a clear measure of the nodes affecting the entire network paths, and the node (194) showing the highest amount.

4.5 Results of Ds_3

To reinforce the results given by the algorithm in Ds_1, Ds_2. Here we used the different dataset in terms of a number of nodes, relationships, and data exchange. We will choose some results to display as it is difficult to display all of the results due to the magnitude of the results.

Figure 4.24 shows topology of Ds_3

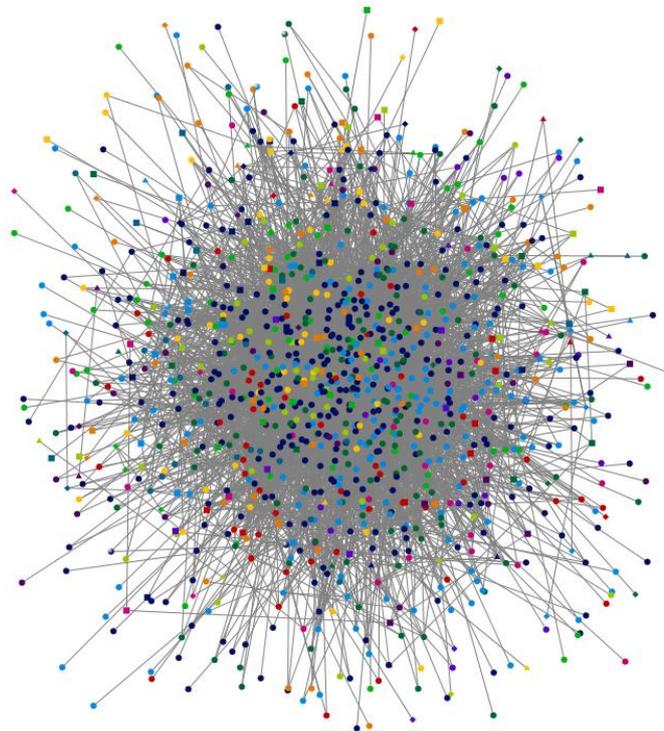


Figure 4. 24 The topology of Ds_3

- For the path phase, Table 4.7 contains some results for selected pairs.

Node pair	Influential Vertex	CISBC_1	IWS_1	HMEIV_1
{14,915}	322	0.4	30	0.999999883
	163	0.2	13	0.156107338
	164	0.2	10	0.039914615
	168	0.2	9	1.19E-07
{348,1103}	186	0.230769231	31	1
	768	0.230769231	28	0.864540264
	38	0.230769231	25	0.733816698
	19	0.153846154	15	0.179656471
	27	0.153846154	11	2.43E-07
{162,299}	341	0.428571429	20	0.9801
	76	0.285714286	13	0.174442442
	196	0.285714286	11	1.58E-07

- For individual nodes, Tables 4.8,4.9 and 4.10 contain some results for selected nodes.

Individual nodes	Influential Vertex	CISBC_2	IWS_2	HMEIV_2
31	341	77	8887	0.999999999
	332	63	7837	0.84926309
	1	64	5540	7.22E-01
	54	51	6907	0.714499792
	241	54	5291	0.648639482
	339	40	3831	0.481315113
	45	39	3078	0.432658714
	329	25	3406	0.356992792
	331	22	3668	0.354026001
	330	23	2696	0.305890931
	334	25	2147	0.287787385
	104	22	2435	0.280607738

Individual nodes	Influential Vertex	CISBC_2	IWS_2	HMEIV_2
687	434	65	7696	0.999999998
	367	61	6631	0.894028218
	532	55	7082	0.872839435
	351	57	6471	0.858421936
	543	44	6311	0.742594338
	588	56	4585	0.710686066
	698	38	3400	0.51662122
	199	29	3498	0.454647181
	122	25	3520	0.418358899
	250	22	3339	0.384637845
	692	23	2934	0.372186673
	306	22	2163	0.314240964

Individual nodes	Influential Vertex	CISBC_2	IWS_2	HMEIV_2
1022	460	267	53811	0.999999999
	483	116	19042	0.396295794
	52	66	14764	0.260924637
	564	33	11654	0.174972733
	429	48	8293	0.16815162
	82	38	8168	0.148457445
	81	36	7577	0.138235521
	341	11	9548	0.125265136
	131	20	3675	0.072290289
	57	15	4567	0.07187559
	238	15	3488	0.060590023
	64	14	3153	0.056941493

- For the whole network

Table 4.11. shows the result for the whole network

Influential Vertex	CISBC_3	IWS_3	HMEIV_3
109	19361	4974148	0.999999972
341	19291	4710828	0.961308205
28	17097	3555630	0.780214649
53	15314	3568664	0.749349787
49	15055	3457632	0.730930465
378	14207	3571882	0.725208435
43	13591	3375052	0.689139212
78	13754	3157492	0.668339175
16	13679	2958478	0.644030005
580	13034	2606226	0.590264681
242	12455	2597512	0.576551671
21	11546	2584302	0.554399155
196	10183	2681056	0.533000708
429	11272	2385684	0.526339514
141	9682	2725242	0.525890785

4.5.1 Discuses of Results (Ds_3)

In the same way, the results of a dataset Ds_3 can be tracked. Still, we will discuss how the IWS index affects changing the results of the algorithm through that the indicator gives large values to some nodes because of their activity in sharing data. Still, these nodes do not have large values within the CISBC index, This is what we find in the influential paths within the second stage of the algorithm, and it appears clearly as in Table 4.8 , where vertex (332) take the second rank in influential of path's node (31). However, the vertex (332) has only 63 within the CISBC index, and there is a vertex that has much higher values than it (1). Also, vertex (532) in Table 4.9 has the third rank in influential on path's node (687). It also has few values within the CISBC Index compared to the values that follow. The reason is the large and significant superiority of the two nodes that occupied the first positions in the IWS Index.

The same is the case with Table 4.10, where the vertex (564) that affect the node (1022) occupies the fourth rank, knowing that it has a small and far value within the CISBC index, but what happened is that it obtained high values within the IWS index; also, the vertex (341) had an eighth position with a small value of CISBC.

At the level of the entire network in Table 4.11, we find that there is a differentiation that occurs by overcoming the values of the CISBC index over the values of the IWS index, through what we notice in the vertex (28) that occupied the third-ranked, as the values it has in the IWS index are less compared to the vertices (53, 378), because the vertex (28) has a very high position and values within the CISBC index.

4.6 Results of Ds_4

The last database is Ds_4, where the large number of nodes and links shows the magnitude of this data. Figure 4.25 shows topology of Ds_4.

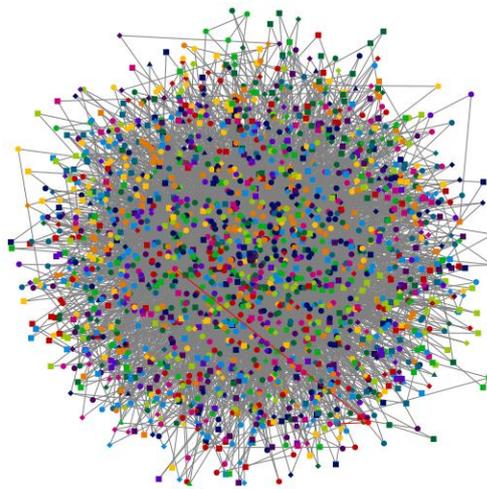


Figure 4. 25 The topology of Ds_4

For path phase Table 4.12 contains some results for selected pairs.

Node pair	Influential Vertex	CISBC_1	IWS_1	HMEIV_1
{1112,1890}	111	0.1764706	25	0.999999851
	1176	0.1176471	17	0.398198542
	135	0.1176471	15	0.2913381
	1271	0.1176471	13	0.178004755
	1706	0.1176471	12	0.119558778
	1	0.1176471	11	0.060175997
	48	0.1176471	10	8.53E-08
	731	0.1176471	10	8.53E-08
{115,902}	22	0.4	46	0.999999899
	472	0.2	16	0.076946621
	217	0.2	15	0.051512132
	949	0.2	13	1.14E-07
{63,1941}	217	0.272727273	30	0.999999905
	1186	0.181818182	16	0.270873691
	473	0.181818182	14	0.182548256
	20	0.181818182	13	0.137547508
	425	0.181818182	10	9.29E-08

- For individual nodes Tables 4.13 and 4.14 contain some results for selected nodes.

Individual node	Influential Vertex	CISBC_2	IWS_2	HMEIV_2
123	470	135.5802412	25888	0.968786763
	357	134.244055	24178	0.92645535
	587	99.17199505	27138	0.815864953
	1058	106.4046837	21723	0.791994894
	984	104.9928815	19915	0.754787602
	666	102.3831534	19700	0.741138893
	992	79.18172217	19122	0.638271144
	897	63.14526967	12314	0.460029052
	510	66.37457744	9159	0.420118302
	960	57.43454213	10021	0.398357906
	124	52.29381545	11093	0.39644467
	774	57.38314743	9268	0.385650339

individual node	Influential Vertex	CISBC_2	IWS_2	HMEIV_2
1151	44	158.4561177	25325	0.982800866
	135	162.2505035	21991	0.913929031
	682	122.674851	19612	0.764240965
	425	124.323897	18272	0.745223124
	34	112.7835836	19269	0.723735059
	1058	68.72343437	13761	0.478196906
	1253	56.87404965	14273	0.449504872
	797	43.65794225	10598	0.340737668
	181	41.03163622	8551	0.293024908
	605	46.03744482	7345	0.28646363
	54	46.30511137	6203	0.267353694
	513	33.41437087	4765	0.19785331

- For whole network

Table 4.15 shows the result for whole network

Influential Vertex	CISBC_3	IWS_3	HMEIV_3
21	78144	33708160	0.971598162
189	81531	27394126	0.869879323
44	70173	27575946	0.837226261
425	72877	22349994	0.754163606
1199	55232	24816354	0.707373839
34	63916	20916068	0.692217487
949	55283	22184714	0.667534273
52	49469	24158294	0.662181285
113	43698	21602208	0.589919238
135	46445	14631332	0.498589457
328	39098	16720864	0.488188896
242	35833	17307204	0.47836411
682	38043	15360606	0.460879181
1	40421	12774780	0.435432395
41	40755	12622896	0.435161001

4.6.1 Discusses of Results (Ds_4)

The dataset Ds_4 gives a large diversity because it has a larger number of nodes and relationships between them. With the same approach as in the previous discussion, the effect can be traced within individual paths, as well as for individual nodes and for the entire network, but we will only discuss how the IWS index affects changing the results of the algorithm through that the indicator gives large values to some nodes because of their activity in sharing data, but these nodes do not have large values within the CISBC index, This is what we find in the influential paths within the second stage of the algorithm, and it appears clearly as in Table 4.13 , where vertex (587) take the third rank in influential of path's node (123), although the vertex (587) has only 99.17 within the CISBC index. Table 4.14 vertex(44) have high value within IWS index , it has first rank inside of list influential on path's node (1151). It also has the least value within the CISBC Index compared to the value that follow.

At the level of the entire network, we find that there is a differentiation that occurs by overcoming the values of the CISBC index over the values of the IWS index, through what we notice in the vertex (189) that occupied the second place, as the values it has in the IWS index are less compared to the vertex (44) in Table 4.15.

4.7 The Methods of Evaluation (Ds_2, Ds_3, Ds_4)

Previously in the second chapter, the importance of metrics in extracting network characteristics and applications is mentioned. Katharina A. [1] clarified several points related to applying standard or centrality metrics. These metrics can be applied in various models to explain a network's characteristics. The following are the most important general outlines of the curricula that have been exploited metrics:

1. The Centrality metrics are employed to describe the network characteristics and overall structure.

2. The Centrality metrics are used to test the correlation with external criteria of the node or to test a model based on an analytical basis.
3. Directly metric-based operations can be configured to predict influential nodes within the network.

4.7.1 The Results HMEIV Compared with the Centrality Metrics

Comparing the fifteen values at the top resulting from the algorithm HMEIV_3 with the top fifteen values resulting from the standard metric (Degree centrality, Closeness centrality, Betweenness centrality, Eigenvector centrality, PageRank centrality).

Tables 4.16, 4.17 and 4.18 show a clear appearance of most of the values of the five metrics in the results of the HMEIV_3. This generates the conclusion that the HMEIV_3 was able to a large extent to reconcile the five measures. So, its results reflected the clear diversity and comprehensiveness of a local and global scale on the network.

Rank	HMEIV_3		Degree		Closeness		Betweenness		Eigenvector		PageRank	
	ID	Value	ID	Value	ID	Value	ID	Value	ID	Value	ID	Value
1	194	1	194	112	194	0.000341	194	50614.84	194	0.00835	194	9.5958
2	145	0.83061	145	103	1210	0.000333	145	42959.04	1210	0.00770	145	9.1848
3	828	0.716301	1210	93	145	0.000332	131	34498.04	145	0.00715	131	7.7809
4	7	0.70647	52	88	7	0.000331	52	31244.97	7	0.00697	553	7.6515
5	334	0.701854	131	87	52	0.000323	1210	31082.97	52	0.00625	1210	7.4867
6	430	0.700553	21	83	131	0.000321	21	30832.82	131	0.00590	52	7.4226
7	44	0.678007	553	82	679	0.000319	553	30042.08	553	0.00584	21	7.2712
8	1210	0.637796	7	81	21	0.000314	7	29426.15	679	0.00570	7	6.5712
9	131	0.578131	430	76	553	0.000314	828	23502.55	828	0.00510	430	6.4418
10	52	0.54219	828	73	535	0.000314	334	23159.61	21	0.00508	334	6.3948
11	1221	0.512235	253	72	44	0.000312	122	22900.31	956	0.00499	828	6.1733
12	679	0.454674	334	71	430	0.000312	44	22282.13	253	0.00495	253	6.0422
13	21	0.449792	679	69	956	0.000312	430	21124.92	535	0.00482	122	5.9477
14	553	0.44977	44	65	1221	0.000312	679	20756.14	1221	0.00480	679	5.8634
15	399	0.447698	535	64	6	0.000311	535	20559.22	430	0.00469	44	5.6924

Table 4. 17 The Top rank 15 vertices of the metric standards with HMEIV_3 for Ds_3.

Rank	HMEIV_3		Degree		Closeness		Betweenness		Eigenvector		PageRank	
	ID	Value	ID	Value	ID	Value	ID	Value	ID	Value	ID	Value
1	109	0.999999	109	71	341	0.000334	341	25753.40	109	0.01146	109	5.7798
2	341	0.961308	341	52	28	0.000333	109	23713.57	16	0.00827	28	4.5028
3	28	0.780214	28	51	109	0.00033	28	21653.91	196	0.00772	341	4.3968
4	53	0.749349	43	51	43	0.000329	580	20254.31	204	0.00675	53	4.3742
5	49	0.730930	16	51	78	0.000327	78	19666.79	43	0.00658	24	4.1698
6	378	0.725208	53	49	53	0.000327	242	17844.09	49	0.00640	43	4.1415
7	43	0.689139	196	47	242	0.000326	141	17663.46	56	0.00632	16	4.0999
8	78	0.668339	242	45	52	0.000326	53	17043.84	116	0.00625	242	4.0507
9	16	0.644030	78	43	141	0.000323	360	17007.01	341	0.00618	360	3.9761
10	580	0.590264	21	43	378	0.000321	43	16797.08	53	0.00603	21	3.9503
11	242	0.576551	24	43	54	0.00032	378	15964.75	128	0.00593	196	3.9210
12	21	0.554399	49	40	3	0.00032	429	15306.82	28	0.00583	78	3.7876
13	196	0.533000	141	40	137	0.000319	52	14823.77	21	0.00552	141	3.7528
14	429	0.526339	354	40	134	0.000318	196	12815.03	206	0.00544	580	3.6090
15	141	0.525890	360	40	39	0.000318	564	12779.65	210	0.00538	14	3.5555

Table 4. 18 The Top rank 15 vertices of the metric standards with HMEIV_3 for Ds_4.

Rank	HMEIV_3		Degree		Closeness		Betweenness		Eigenvector		PageRank	
	ID	Value	ID	Value	ID	Value	ID	Value	ID	Value	ID	Value
1	21	0.97159	189	115	189	0.000201	189	98323.93	189	0.00716	21	11.374
2	189	0.86987	21	111	1	0.000201	21	96614.45	1199	0.00684	189	10.962
3	44	0.83722	44	105	135	0.000199	135	88287.47	1	0.00632	44	10.410
4	425	0.75416	135	105	1199	0.000198	44	85306.51	135	0.00615	135	10.308
5	1199	0.707374	1199	97	44	0.000197	1	77657.47	44	0.00612	1	9.0439
6	34	0.692217	1	95	21	0.000194	128	68725.33	21	0.00556	1199	8.9391
7	949	0.667534	128	90	128	0.000193	1199	66290.63	128	0.00530	128	8.8677
8	52	0.662181	545	84	34	0.000193	545	62336.39	545	0.00522	545	8.4668
9	113	0.589919	113	83	425	0.000192	425	56185.71	682	0.00496	113	8.0323
10	135	0.498589	425	80	682	0.000192	113	53369.34	34	0.00480	425	7.7696
11	328	0.488189	34	74	545	0.000191	41	52455.67	113	0.00467	328	7.3204
12	242	0.478364	328	74	531	0.00019	328	48918.58	949	0.00453	34	7.0156
13	682	0.460879	242	74	120	0.000189	34	48392.93	242	0.00444	242	7.0002
14	1	0.435432	682	72	949	0.000189	682	46743.40	531	0.00437	41	6.9487
15	41	0.435161	531	66	113	0.000188	120	46652.11	425	0.00430	682	6.9465

4.7.2 Employing the Pearson Correlation Coefficient method

The second method relies on a correlation coefficient (denoted r), as it will give an impression about the extent to which the two values are related together. It is used here to measure the correlation of the influential vertices extracted from standard measures with HMEIV_3. Where the scale will measure the linear correlation between two variables, it will give a value between + 1 and -1, where a score of + 1 represents the highest value of correlation between the two

variables, while a score of 0 represent no correlation and a score of -1 represents the full negative correlation between two variables. There are gradients that will appear in between +1 and -1, where will depend on that there is a strong high correlation between the two variables in the case $|r| > 0.8$, strong correlation in the case $|r|$ between (0.59, 0.79), moderate correlation in the case $|r|$ between (0.39, 0.59). Table 4.19 shows values of the Pearson correlation coefficient between five metrics (Degree centrality, Closeness centrality, Betweenness centrality, Eigenvector centrality, PageRank centrality) and HMEIV_3.

Dataset	Degree	Closeness	Betweenness	Eigenvector	PageRank
Ds_2	0.904946	0.618204	0.961789	0.858701	0.917166
Ds_3	0.887774	0.668554	0.971075	0.771343	0.90083
Ds_4	0.922628	0.621146	0.962906	0.876023	0.92859

4.7.3 Frequency of Similar Values

In the direct way of using centrality methods to extract the affected nodes, a case can occur where the same value of the effect goes to more than one node so that the rank for those nodes is equal. This leads to the fact that it is not possible to distinguish between those nodes.

In general, in the case of any procedure, if several nodes share the same value, this leads to confusion in performance.

Therefore, the frequency measurement method for the values produced by these scales is an efficient tool for measuring the efficiency and reliability of the scale so that the lower frequency means the higher efficiency of the scale.

Figures 4.23 and 4.24 show the frequency of each of the five scales with the HMEIV algorithm for the top-rank of 1000 nodes in each scale for Ds_2, Ds_3. While figure 4.25 show the frequency of the five scales with the HMEIV for the top -rank of 2000 nodes in each scale for Ds_4.

So Figures 4.26, 4.27 and 4.28 show how the HMEIV- algorithm got the least number among all the measures.

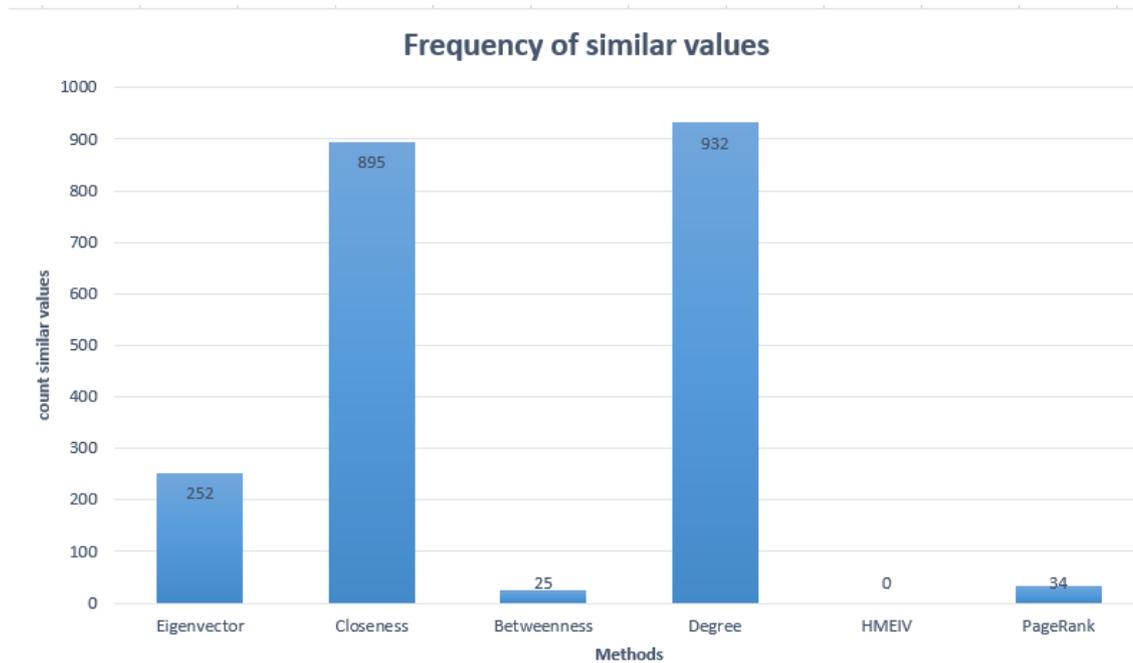


Figure 4. 26 The Frequency of similar values in Ds_2

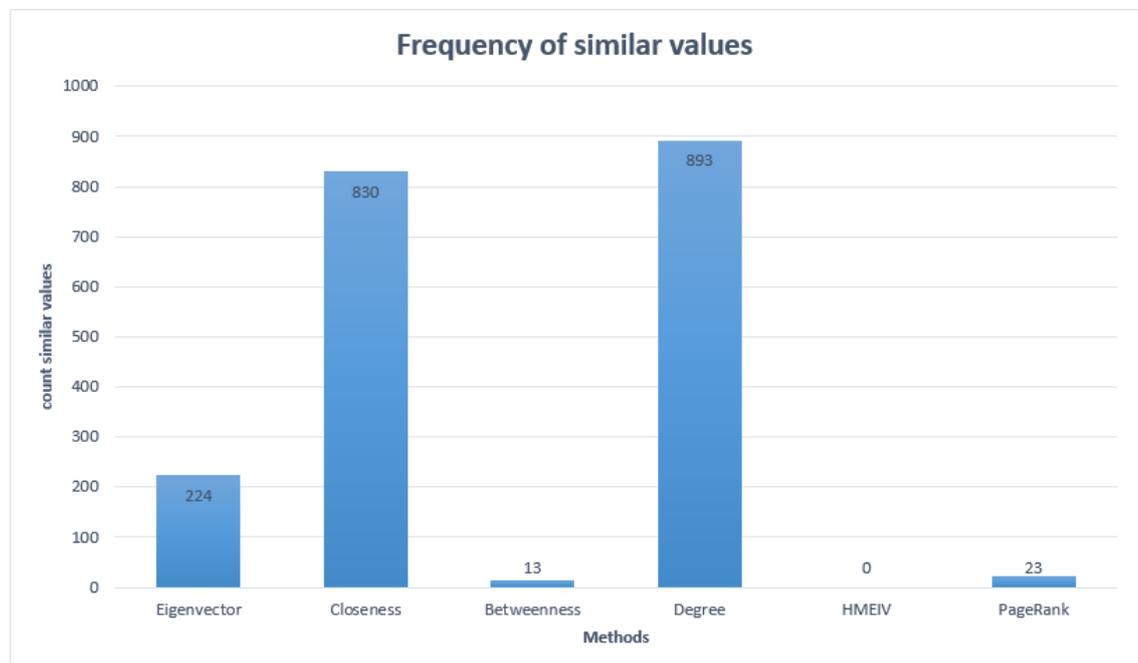


Figure 4. 27 The Frequency of similar values in Ds_3

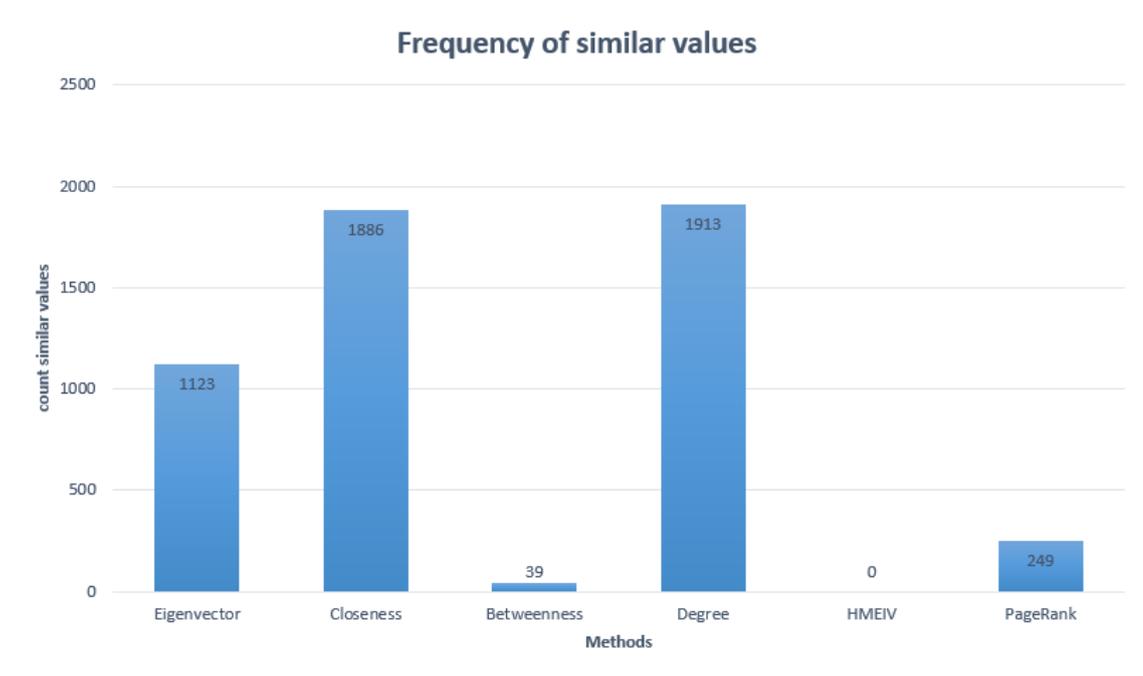


Figure 4. 28 The Frequency of similar values in Ds_4

4.7.4 Susceptible Infected Model (SI)

It is common in networks if the node is of great importance and occupies an advanced position in the network in terms of influence. This means that the ability of that node to transmit the infection to the rest of the network is also high. In this experiment, the top twenty-five nodes were taken in each of the indexes (HMEIV, Closeness centrality, Betweenness centrality, Degree centrality, Eigenvector centrality, PageRank centrality). Where each time, one node is the source of the infect, covering 100 cycles. The next step is to take the average of infected nodes over a period of 100 cycles for sources in each index. With the infection rate remaining constant for all cycles.

The output of each standard index was taken and compared with the result of the algorithm HMEIV.

It can be inferred by the parameter $F(t)$ inside of Susceptible Infected Model, which specifies the average of infected nodes at time t .

The 100-cycles resulting from each index were divided into four sections, each section containing 25-cycles and distinct in one color for each section.

Figures 4.29, 4.30, 4.31, 4.32, 4.33, 4.34, 4.35, 4.36, 4.37, 4.38, 4.39, 4.40, 4.41, 4.42 and 4.43 showing how the average of infected nodes of HMEIV at every $F(t)$ is accelerating more than other indicators.

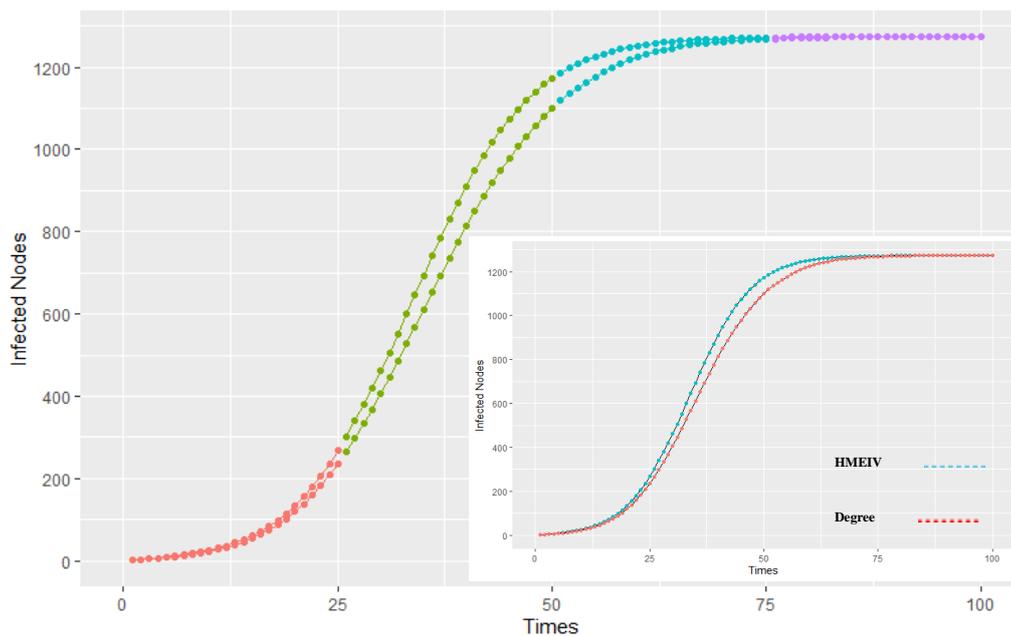


Figure 4. 29 The infected nodes through Top- 25 nodes in HMEIV & Degree within Ds_2

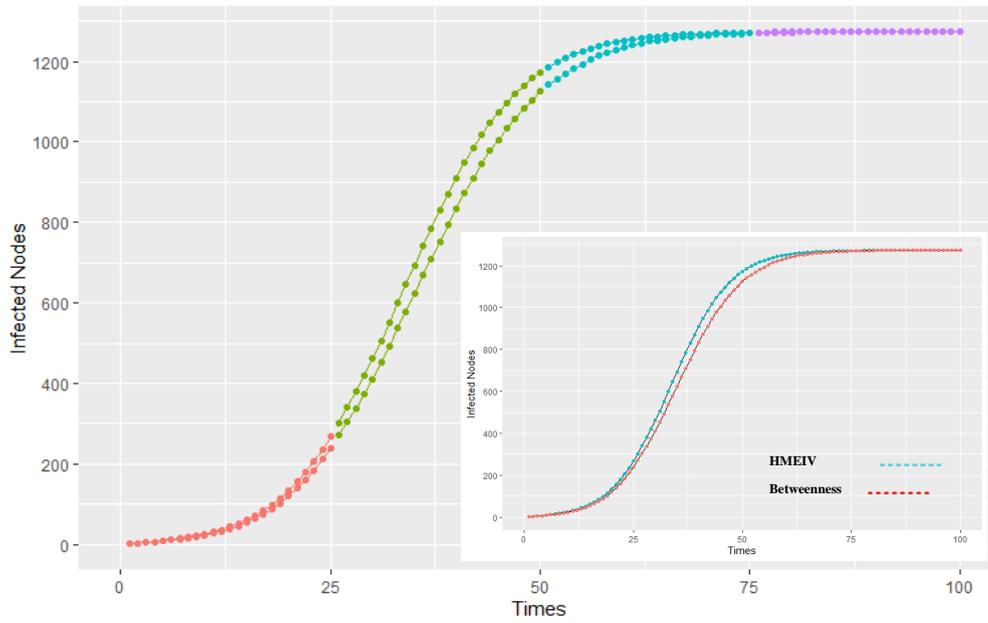


Figure 4. 30 The infected nodes through Top- 25 nodes in HMEIV & Betweenness within Ds_2

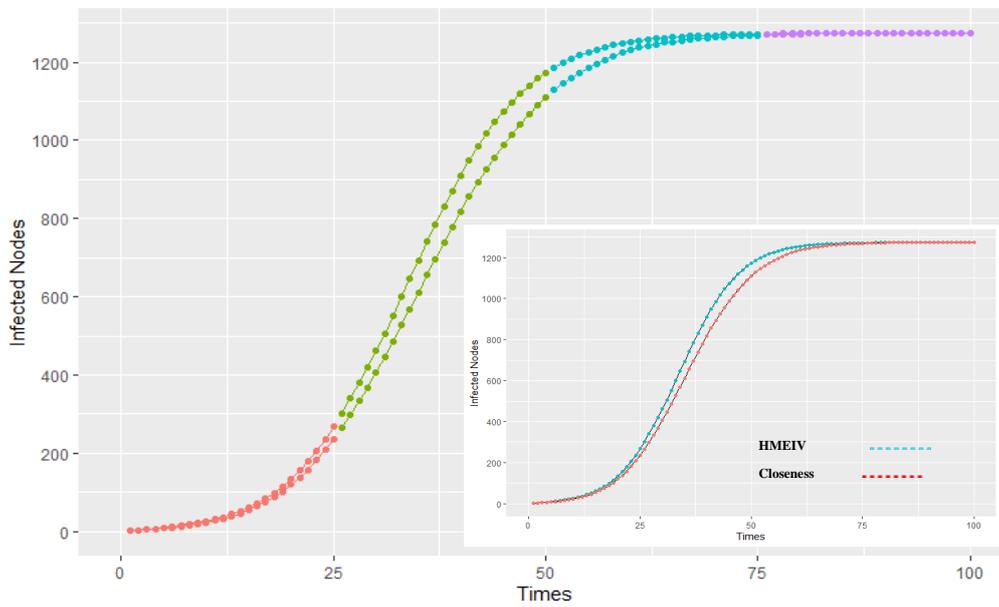


Figure 4. 31 The infected nodes through Top- 25 nodes in HMEIV & Closeness within Ds_2

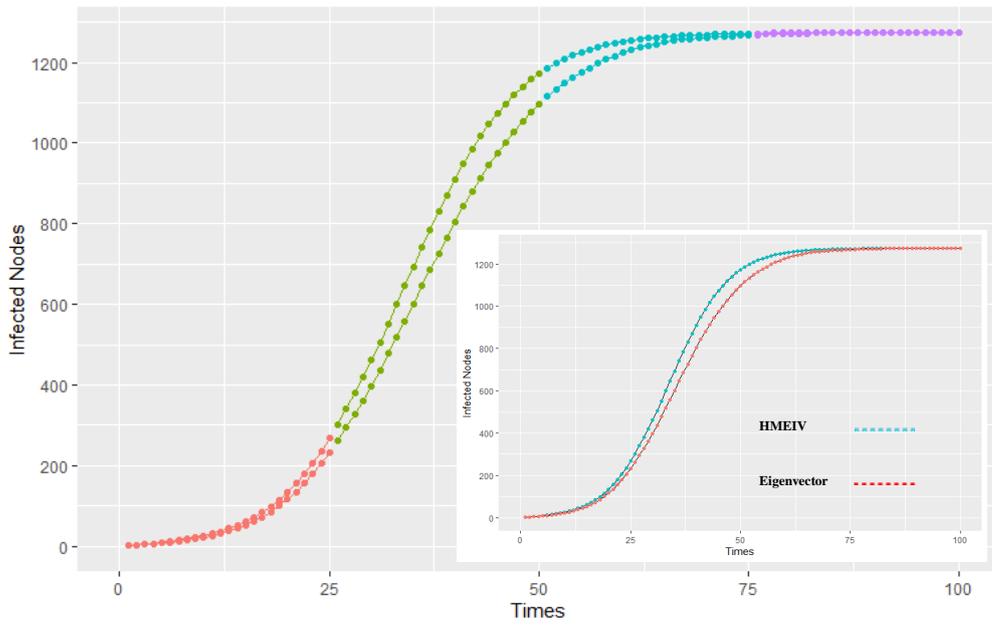


Figure 4. 32 The infected nodes through Top- 25 nodes in HMEIV & Eigenvector within Ds_2

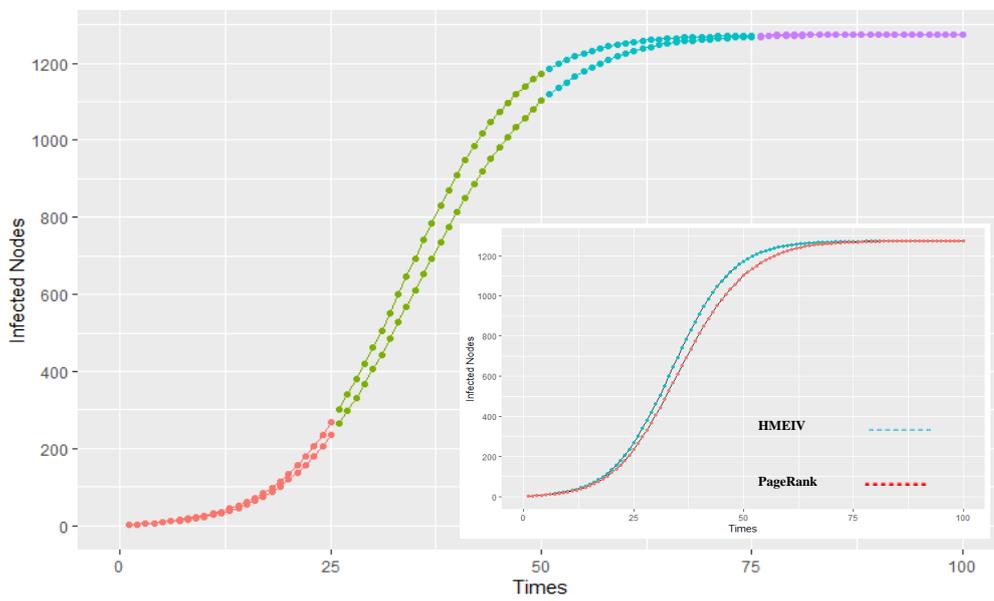


Figure 4. 33 The infected nodes through Top- 25 nodes in HMEIV & PageRank within Ds_2

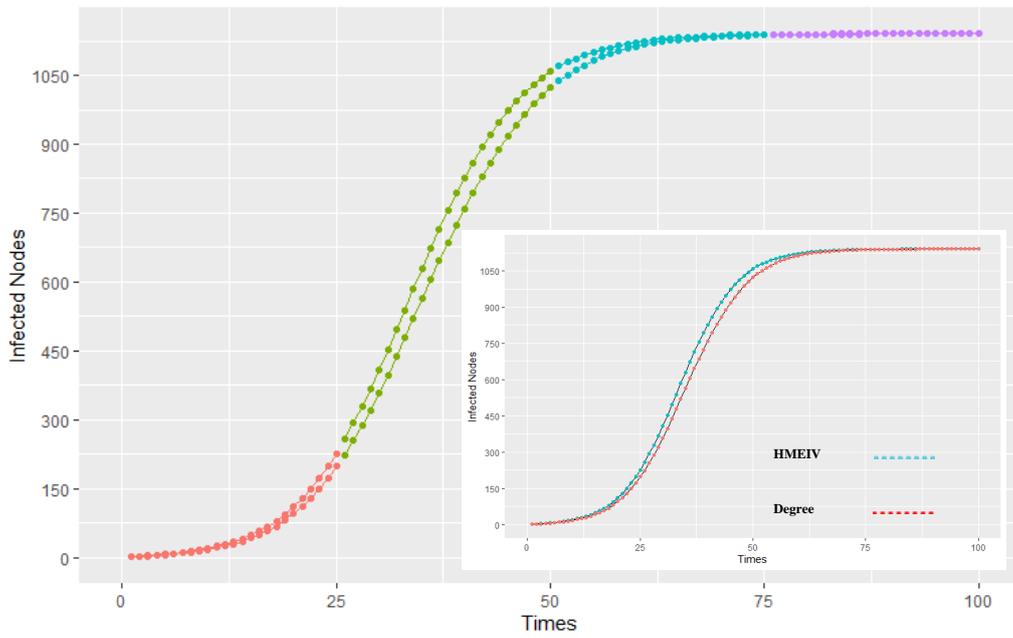


Figure 4. 34 The infected nodes through Top- 25 nodes in HMEIV & Degree within Ds_3

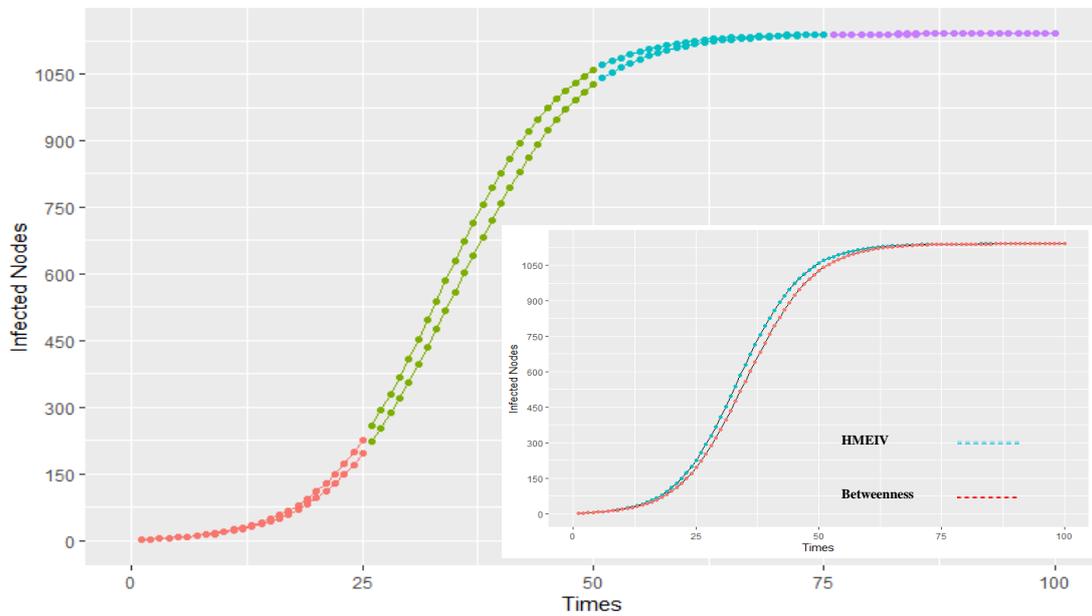


Figure 4. 35 The infected nodes through Top- 25 nodes in HMEIV & Betweenness within Ds_3

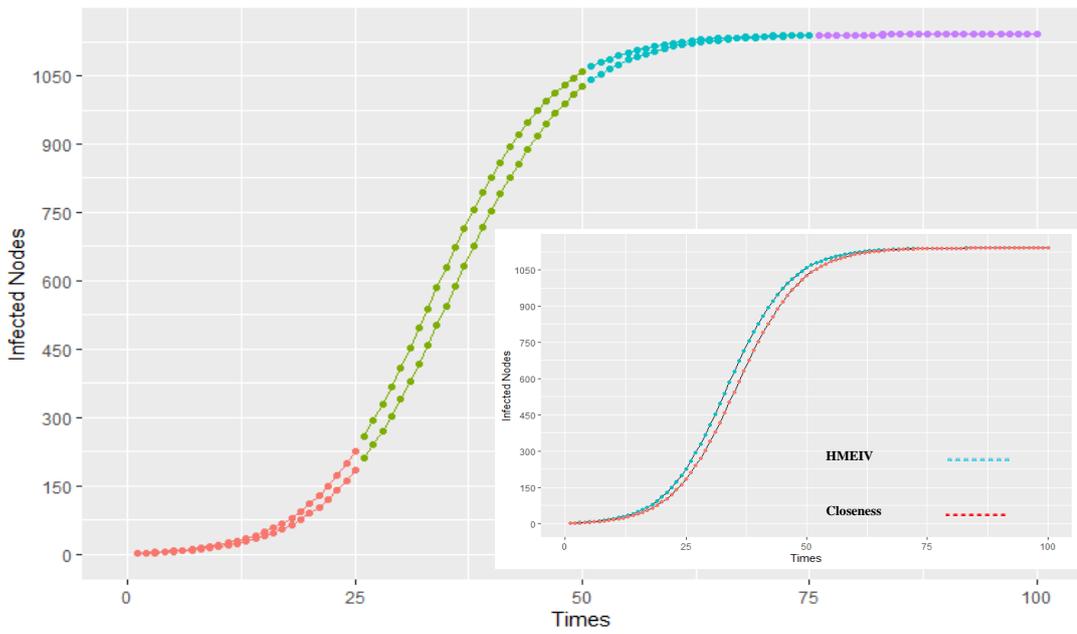


Figure 4. 36 The infected nodes through Top- 25 nodes in HMEIV & Closeness within Ds_3

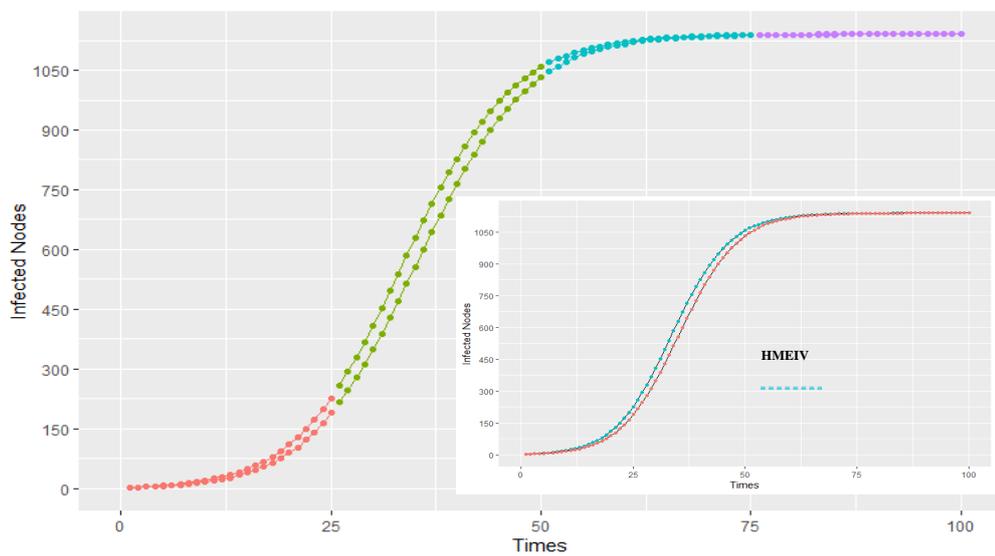


Figure 4. 37 The infected nodes through Top- 25 nodes in HMEIV & EigenVector within Ds_3

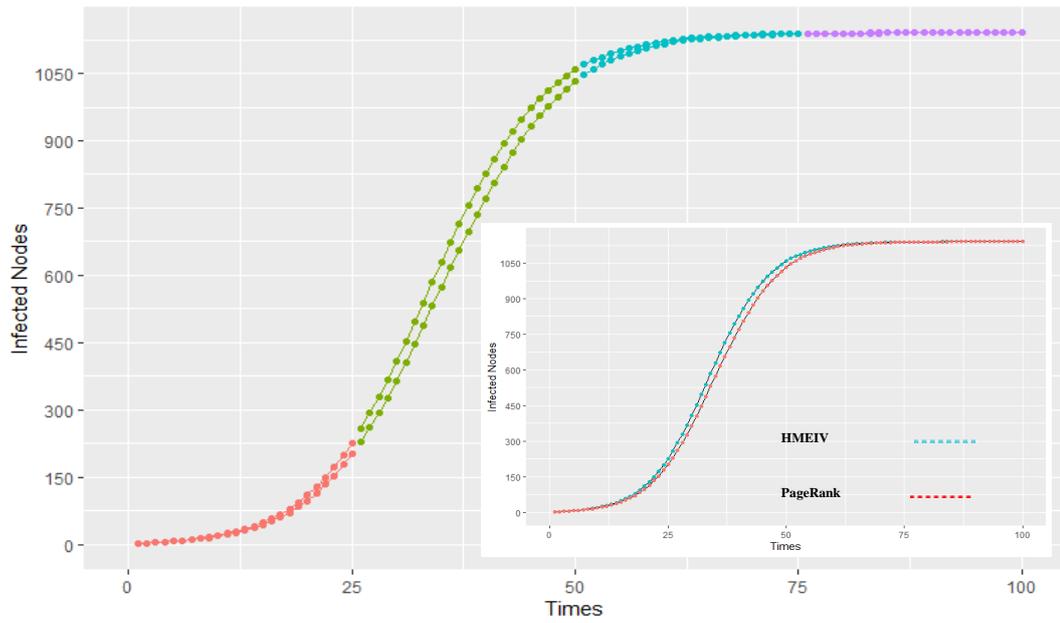


Figure 4. 38 The infected nodes through Top- 25 nodes in HMEIV & PageRank within Ds_3

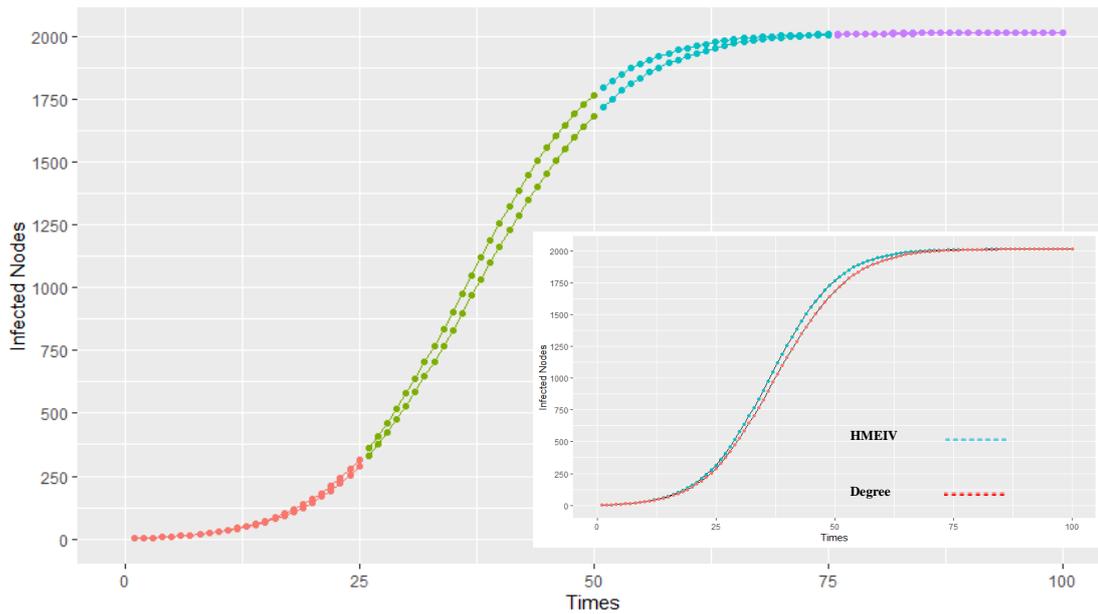


Figure 4. 39 The infected nodes through Top- 25 nodes in HMEIV & Degree within Ds_4

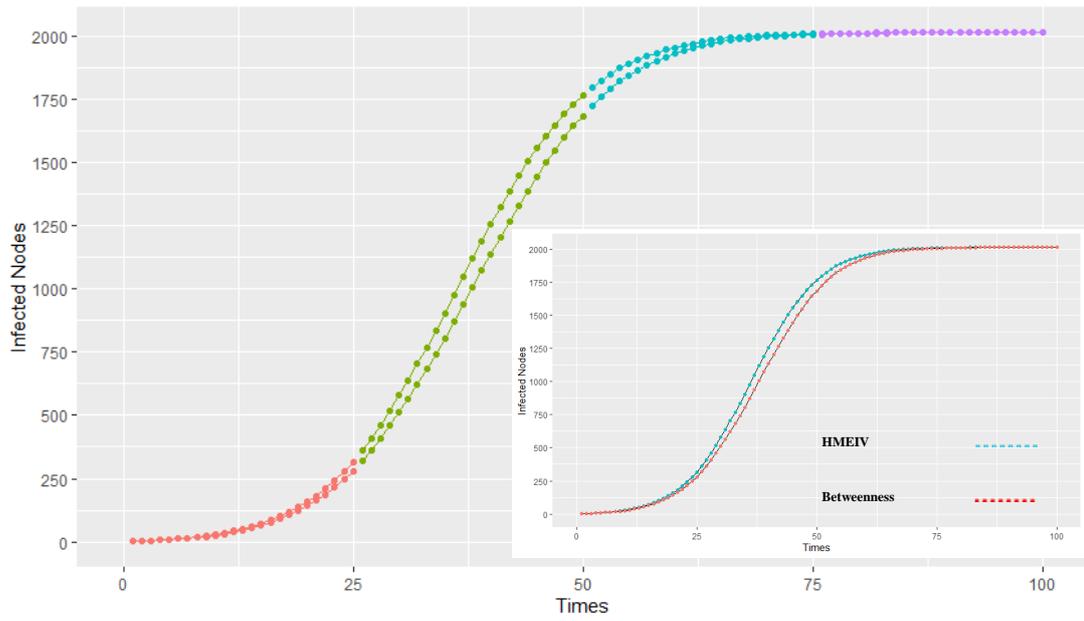


Figure 4. 40 The infected nodes through Top- 25 nodes in HMEIV & Betweenness within Ds_4

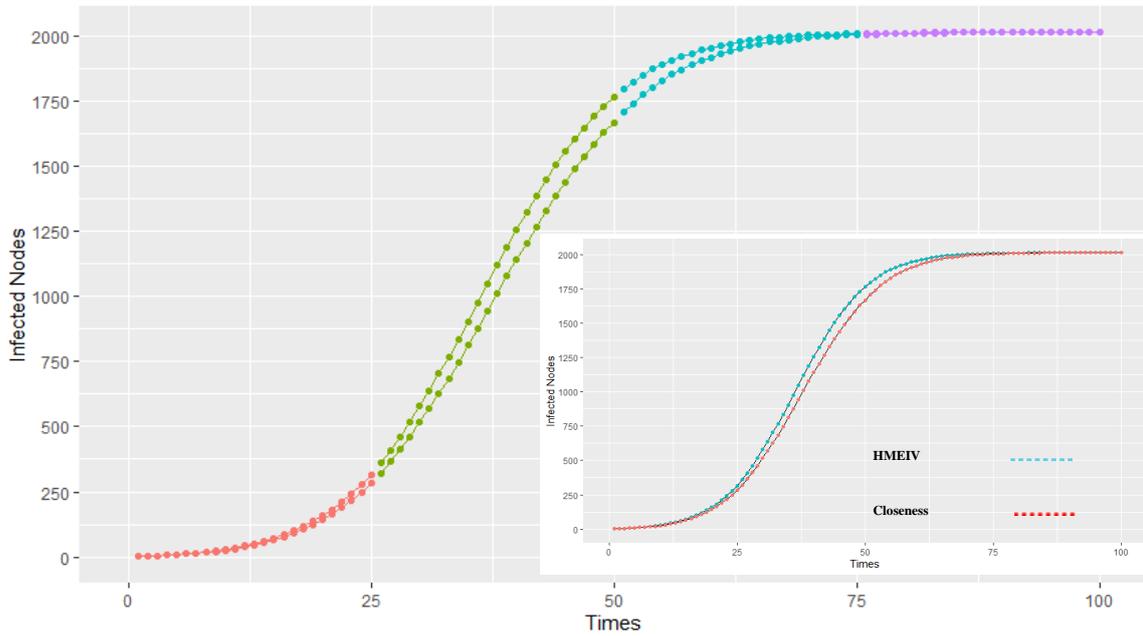


Figure 4. 41 The infected nodes through Top- 25 nodes in HMEIV & Closeness within Ds_4

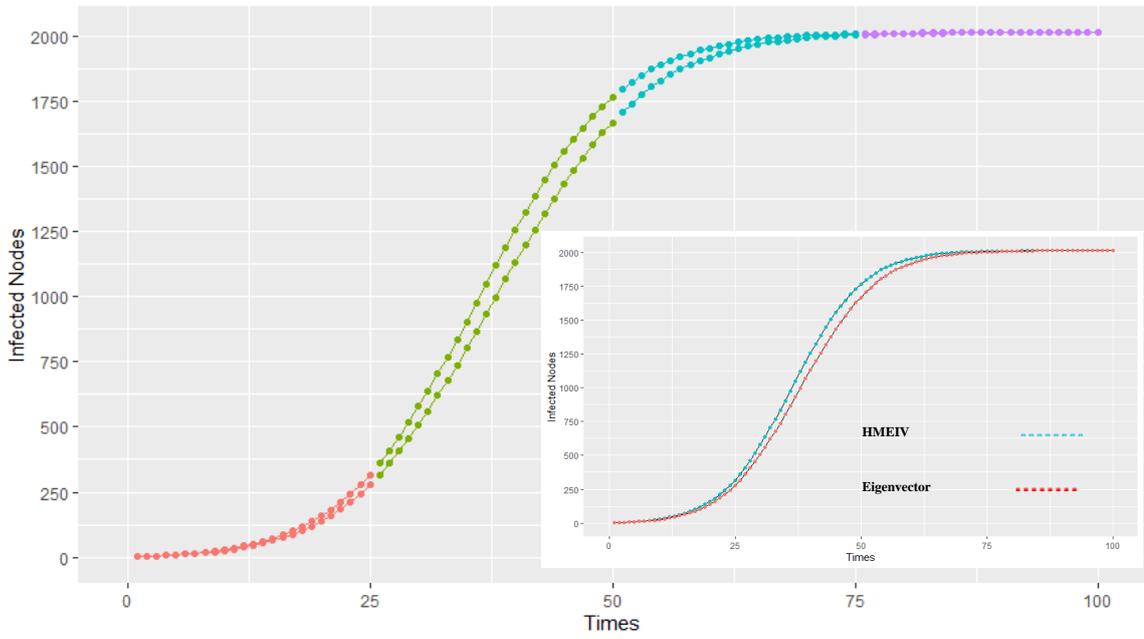


Figure 4. 42 The infected nodes through Top- 25 nodes in HMEIV & EigenVector within Ds_4

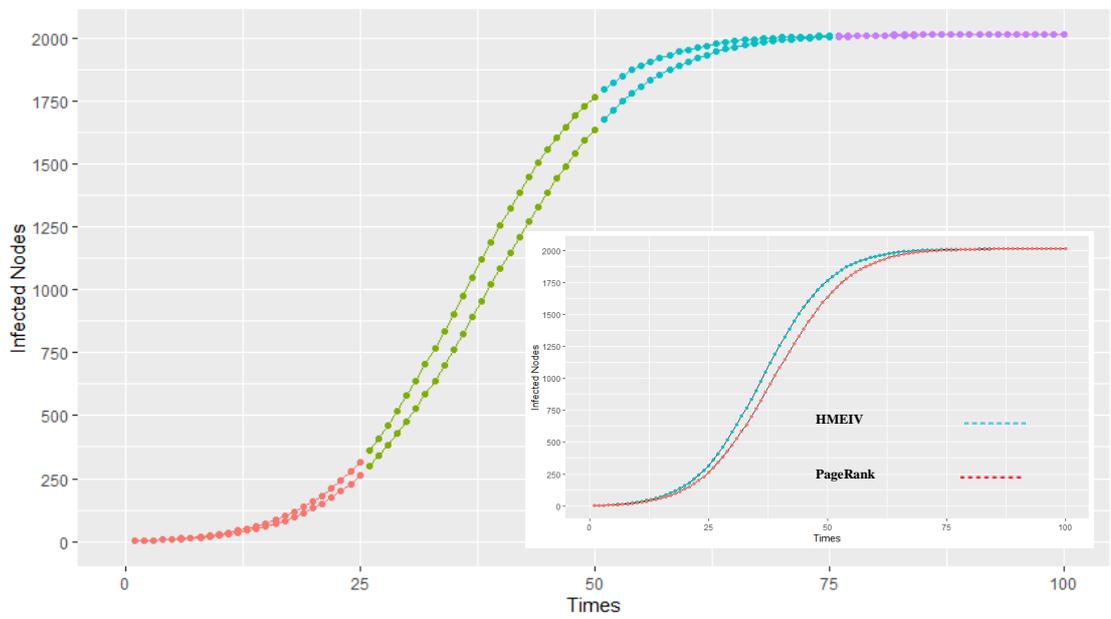


Figure 4. 43 The infected nodes through Top- 25 nodes in HMEIV & PageRank within Ds_4

4.7.5 Attacking the Network

By attacking some nodes, the effectiveness of these nodes can be demonstrated. In other words, the effectiveness of influencing nodes can be tested to measure the extent of their influence on the change in the distance between nodes in the network, the average distance, and the density of the network. This is achieved by deleting these influencing nodes to demonstrate their effect on distance. The following example can be seen for a simple network of nodes. Whereas the average distance to the previous network is 3.2575, the maximum distance (diameter) is 8. But when deleting an important vertex like M, the network will look like Figures 4.44 and 4.45.

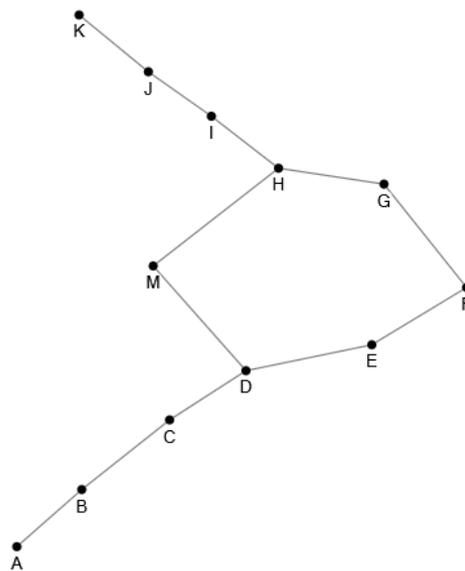


Figure 4. 44 The simple network

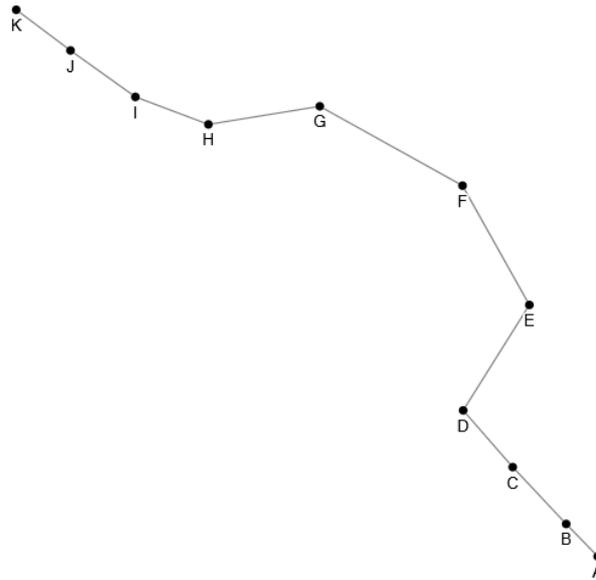


Figure 4. 45 The simple network after delete node M

Deleting this node led to three important things, increase the maximum distance between nodes to 10, the average distance to 4. and decrease the density of simple network.

We will apply this procedure to networks (Ds_2, Ds_3, Ds_4), So that we will delete the nodes that occupy the top three ranks as shown in Table 4.20.

Table 4. 20 Details of distance after procedure Delete nodes						
Dataset	Original Diameter	Diameter after deleting	Original Average Distance	Average Distance After deleting	Original Density	Density after deleting
Ds_2	9	11	3.3172	3.388157	0.007968914	0.007649752
Ds_3	10	11	3.629533	3.676102	0.008395221	0.0081737
Ds_4	7	8	3.547341	3.611769	0.0042171	0.004070151

The importance of these deleted nodes comes from the fact that they control the shorter paths between the complexes; in addition to that, they play a significant role in the process of exchanging information, so the process of deleting it leads to an increase in the distances between the vertices, and thus the average distance

also increases. Thus, the process of controlling news or information is more dispersed between other nodes.

4.8 The Summary

The proposed HMEIV- algorithm implemented on several datasets can lead us to many details about knowing the influencer within the path, about a particular node, or about the entire network, which provides a better and more accurate understanding of all the interactions within the network.

The methods used in evaluating the HMEIV algorithm depended on benchmark metrics. So, all the methods gave good results to support the HMEIV algorithm.

Chapter Five

Conclusions and Future Works

5.1 Conclusions in the following points

The issue of the mutual exchange of data, such as news, rumors, ideas, and information, which is shared from one sector of the network to another is an important subject that is associated with social network study. Highlights the role of influencers within this network in addressing these issues. Identifying the influencers within networks who have control over the paths and interactions that occur within the networks and thus have an impact on directing and adopting users to a certain opinion, decision, promoting a product, or even spreading news and preventing rumors is an open problem in networks. Through the issues this study focused on forming an HMEIV algorithm to find the nodes controlling or influencing the transmission of data, discouraging rumors, promotion, and marketing inside social networks. By applying the HMEIV on a variety of datasets, the results can be summarized as follows:

1. The initial step is to acquire the outputs of the paths connecting two nodes in order to locate the influencer in these paths in a certain approach.
2. The second step obtaining outputs relating to a specific node's paths. Identifying the influencer in these paths in a certain manner.
3. The last step is to get the outputs of the paths for the complete network's nodes and locate the controller in these paths.
4. At the initial and second levels, the results were able to accurately describe the path influencers.
5. The influential nodes drawn from the results have characteristics that bring together the local and global characteristics of the network with the activities of nodes.
6. Evaluation of results A variety of quantitative and qualitative comparisons between HMEIV results and standard indicators

7. The efficiency of the influential vertices that were extracted using the HMEIV was evaluated by testing their influence on the cohesiveness, length of distance, and average distance within the network if removed.

5.2 Future works

1. The HMEIV algorithm can be invested in dynamic networks that apply the algorithm at different time intervals to find the influencers within each period.
2. The HMEIV algorithm can be developed on drug-drug networks, as the drug are connected and affected by each other, so after some modifications to the algorithm can be applied to such networks. Where the nodes represent the drugs, while in networks the relationships between the nodes represent the side effects. The role of the algorithm is to detect side effects between a pair of drugs.
3. The HMEIV algorithm can also be invested in networks of protein interactions, where the nodes represent the proteins while the edges represent the direct relationship between them. Where the role of the algorithm is to find the affecting proteins between the pairs of interacting proteins.

References

- [1] Katharina. A. Zweig, NETWORK ANALYSIS LITERACY, Verlag Gmbh Austria: Springer, 2016.
- [2] M. Kaya, J. Kawash, S. Khoury and M. Y. Day, Social Network Based Big Data Analysis and Applications., Springer., 2018.
- [3] R. a. J. M. e. Light, The Oxford Handbook of Social Networks, Oxford : Oxford University Press, 2020.
- [4] A. S. A. a. J. M. Zareie, "Influential node ranking in social networks based on neighborhood diversity," *Future Generation Computer Systems*, vol. 1, no. 94, pp. pp.120-129, 2019.
- [5] S. K. RAVINDRAN and V. GARG, Mastering social media mining with R, Packt Publishing Ltd, 2015.
- [6] J. a. P. J. C. Scott, The SAGE handbook of social network analysis, SAGE publications, 2011.
- [7] W. M. A. & B. V. (. De Nooy, Exploratory social network analysis with Pajek: Revised and expanded edition for updated software, Cambridge university press, 2018.
- [8] R. K. PM, A. Mohan and K. Srinivasa, Practical Social Network Analysis with Python., Springer International Publishing, 2018.
- [9] D. Hansen, M. A. S. B. Shneiderman and I. Himelboim, Analyzing Social Media Networks with NodeXL: Insights from a Connected World, Morgan Kaufmann, 2019.
- [10] A. a. S. A. Zareie, " A hierarchical approach for influential node ranking in complex social networks," *Expert Systems with Applications*, vol. 1, no. 93, pp. 200-211, 2018.
- [11] P. a. B. C. Praveen, " Big Data Clustering: Applying Conventional Data Mining Techniques in Big Data Environment," *Innovations in Computer Science and Engineering*, vol. 74, no. 1, pp. 509-516, 2019.
- [12] S. K. a. A. P. Pal, Pattern Recognition and Big Data, World Scientific, 2016.
- [13] R. MISSAOUI and I. . SARR, Social network analysis-Community detection and evolution, Springer, 2015.
- [14] Z. B. L. H. D. Y. Shang Q, " Identifying influential nodes: A new method based on network efficiency of edge weight updating," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 3, no. 31, p. 033120, 2021.

- [15] B. e. a. Zhang, "An Opinion-Leader Mining Method in Social Networks With a Phased-Clustering Perspective," *IEEE Access* 8, pp. 31539-31550, 2020.
- [16] W. Z. D. Y. .. A. 6. Liu F and 193:., "GMM: A generalized mechanics model for identifying the importance of nodes in complex networks.," *Knowledge-Based Systems*, no. 193, p. 105464, 2020.
- [17] G.-C. P. M. X. S. M. Riquelme F, " The neighborhood role in the linear threshold rank on social networks," *Physica A: Statistical Mechanics and its Applications*, no. 528, p. 121430, 2019.
- [18] C. K. e. a. Leung, "Personalized Deep Inf: enhanced social influence prediction with deep learning and transfer learning.," in *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, 2019..
- [19] L. S. Z. Y. L. Y. Shao Z, "Identifying influential nodes in complex networks based on Neighbours and edges," *Peer-to-Peer Networking and Applications*, vol. 6, no. 12, pp. 1528-1537., 2019.
- [20] L. X. X. G. Yang P, "A dynamic weighted TOPSIS method for identifying influential nodes in complex networks," *Modern Physics Letters B*, vol. 19, no. 32, p. 1850216, 2018.
- [21] E. e. a. BERGAMINI, "Improving the betweenness centrality of a node by adding links.," *Journal of Experimental Algorithmics (JEA)*, vol. 12, no. 1, pp. 1-32, 2018.
- [22] L. e. a. Cao, "An algorithm of opinion leaders mining based on signed network.," *AIP Conference Proceedings*, vol. 1955, no. 1, p. 040044, 2018.
- [23] F. a. S. L.] Ullah, "Identification of influential nodes based on temporal-aware modeling of multi-hop neighbor interactions for influence spread maximization," *Physica A: Statistical Mechanics and its Applications*, vol. 1, no. 486, pp. 968-985, 2017.
- [24] L. FEI, H. MO and Y. DENG, "A new method to identify influential nodes based on combining of existing centrality measures," *Modern Physics Letters B*, vol. 31, no. 26, p. 1750243, 2017.
- [25] Z. e. a. Zhao, "A community-based approach to identifying influential spreaders," *Entropy* , vol. 17, no. 4, pp. 2228-2252, 2015.
- [26] Q. Li, T. Zhou, L. Lu and D. Chen, "Identifying influential spreaders by weighted leaderrank," *Physica A*, vol. 404, pp. 47-55, 2014.
- [27] W. D. H. Y. M. S. D. Y. Gao C, "A modified evidential methodology of identifying influential nodes in weighted networks," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 21, pp. 5490-5500, 2013.
- [28] S. H. Bornholdt S, *Handbook of graphs and networks, From Genome to the Internet*, Willey-VCH , 2001.
- [29] B. U., *Network analysis: methodological foundations*, Springer Science & Business Media, 2005 .

- [30] B. Bollobás, *Modern graph theory*, Springer Science & Business Media Vol .184, 2013.
- [31] B. Bollobás, *Modern graph theory*, Springer, Heidelberg Vol.100, 2001.
- [32] U. S. R. a. A. B. Murty, *Graph Theory (graduate texts in mathematics 244)*, springer, 2008.
- [33] N. a. G. R. Hartsfield, *Pearls in graph theory: a comprehensive introduction*, Courier Corporation , 2013.
- [34] J. L. J. Y. a. P. Z. e. Gross, *Handbook of graph theory*, CRC press, 2013.
- [35] A. Bickle, *Fundamentals of Graph Theory*, American Mathematical Soc Vol 43, 2020.
- [36] C. Prell, *Social network analysis: History, theory and methodology*, Sage, 2012.
- [37] R. a. J. R. Alhajj, *Encyclopedia of social network analysis and mining*, Springer, 2014.
- [38] S. F. B. K. a. L. Z. Yang, *Social network analysis: Methods and examples*, Sage Publications, 2016.
- [39] S. a. K. F. Wasserman, *Social network analysis: Methods and applications*, springer, 1994.
- [40] B. L. B. A. P. a. S. P. B. Perry, *Egocentric network analysis: Foundations, methods, and models*, Cambridge : Cambridge university press Vol.44, 2018.
- [41] N. B. E. E. G. E. M. G. K. J. & T. M. Crossley, *Social network analysis for ego-nets: Social network analysis for actor-centred networks*, Sage, 2015.
- [42] M. V. A. M. C. a. A. P. La Gala, "Ego-net digger: a new way to study ego networks in online social networks," in *In Proceedings of the First ACM International on Hot Topics on Interdisciplinary Social Net*, 2012.
- [43] C. M. P. A. P. F. Arnaboldi V, "Analysis of ego network structure in online social networks," in *In 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, 2012.
- [44] M. NEWMAN, *Networks*, Oxford university press, 2018.
- [45] S. Fortunato, "Community detection in graphs," *Physics Reports*, 2010.
- [46] C. Kadushin, *Understanding social network analysis: Theories, concepts & findings*, Oxford University Press, 2012.
- [47] A. CC., *Recommender systems*, Springer International Publishing, 2016.
- [48] S. & H. B. Dominguez, *Mixed methods social networks research: Design and applications*, New York: Cambridge University Press., 2014.
- [49] N. P. A. S. & L. D. Eagle, "Inferring friendship network structure by using mobile phone data," *Proceedings of the National Academy of Science*, vol. 106, no. 36, p. 15274–15278, 2009.

- [50] N. M. L. G. J. & A. F. Harrigan, "Negative ties and signed graphs research: Stimulating research on dissociative forces in social networks," *Social Networks*, vol. 1, no. 60, pp. 1-10., 2020.
- [51] S. P. M. A. B. D. & L. G. Borgatti, "Network analysis in the social sciences," *Science*, vol. 5916, no. 323, p. 892–895, 2009.
- [52] P. V. Marsden, "Network data and measurement," *Annual Review of Sociology*, vol. 16, no. 1, p. 435–463, 1990.
- [53] N. S. W. S. & F. K. Contractor, "Testing multitheoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example," *Academy of Management Review*, vol. 31, no. 3, p. 681–703, 2006.
- [54] M. S. & M. C. Mizruchi, "Egocentric, sociocentric, or dyadic? Identifying the appropriate level of analysis in the study of organizational networks," *Social Networks*, vol. 28, no. 3, p. 187–208, 2006.
- [55] D. J. & B. S. P. Brass, A practical guide to multilevel social network research, the handbook of multilevel theory, measurement, and analysis, American Psychological Association, 2019.
- [56] L. J. B. M. e. Fu X, Social network analysis: interdisciplinary approaches and case studies, CRC Press, 2017.
- [57] E. R. Barnes, "An algorithm for partitioning the nodes of a graph," *SIAM Journal on Algebraic Discrete Methods*, vol. 3, no. 4, p. 541–550, 1982.
- [58] M. W. O. L. & V. N. K. Mahoney, "A local spectral method for graphs: With applications to improving graph partitions and exploring data graphs locally," *Journal of Machine Learning Research*, vol. 13, no. 1, p. 2339–2365, 2012.
- [59] S. R. "J. F. Peters, "Proximal three-way decisions: theory and applications in social networks," *Knowledge-Based Syst., Elsevier*, vol. 91, no. 1, p. 4–15, 2016.
- [60] J. a. B. J. Podolny, "Resources and relationships," *American Sociological Review*, vol. 62, no. 1, p. 673–693, 1997.
- [61] L. Freeman, "Centrality in social networks: conceptual clarification," *Social Network*, vol. 1, no. 12, pp. 35-41, 1979.
- [62] F. B. S. P. & E. M. G. Agneessens, "Geodesic based centrality: Unifying the local and the global," *Social Networks*, vol. 49, no. 1, p. 12–26, 2017.
- [63] M. E. Newman, "The mathematics of networks," *The new palgrave encyclopedia of economics*, vol. 2008, no. 2, pp. 1-12, 2008.
- [64] B. Bollobás, Random Graphs, New York: Academic Press, 2nd, 2001.
- [65] P. Bonacich, "Power and centrality: A family of measures," *American Journal of Sociology*, vol. 92, no. 5, p. 1170–1182, 1987.

- [66] P. Bonacich, "New Worlds of Social Network Research," *Contemporary Sociology*, vol. 37, no. 5, p. 426, 2008.
- [67] P. G. O. a. P. M. Bonacich, "Family coalitions: A new approach and method," *Social Psychology Quarterly*, vol. 1, no. 1, pp. 42-50, 1985.
- [68] B. P, "Some unique properties of eigenvector centralities," *Social Network*, vol. 29, no. 1, p. 555–564, 2007.
- [69] T. a. Y. Deng, "Identification of influencers in complex networks by local information dimensionality," *Inform. Sci*, vol. 512, no. 1, p. 549–562, 2020.
- [70] V. Grolmusz, "A note on the pagerank of undirected graphs," *Information Processing Letters*, vol. 115, no. 6, pp. 633-634, 2015.
- [71] T. Zhou, "Comment on: 'Epidemic spreading on heterogeneous networks with identical infectivity,'" *Phys. Lett.*, vol. 372, no. 10, p. 1725–1726, 2008.
- [72] M. A. Z.-X. W. L. W. C. B. Z. Wang, "Coupled disease behavior dynamics on complex networks: A review," *Phys. Life Rev.* 15, vol. 1, no. 1, p. 1–29, 2015.
- [73] Z. J. M. W. R. S. L. I. Y. M. Y. a. K. J. Wang, "Onymity promotes cooperation in social dilemma experiments," *Science advances*, vol. 3, no. 3, p. 1601444, 2017.
- [74] D. W. L. W. Z. a. X. G. Zhao, "Virus propagation and patch distribution in multiplex networks: modeling, analysis, and optimal allocation," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 7, pp. 1755-1767, 2018.
- [75] A. S. Klovdahl, "Social networks and the spread of infectious diseases: The AIDS example," *Social Science Medicine*, vol. 21, no. 1, p. 1203–1216, 1985.
- [76] M. L. K. G. S. H. F. L. L. M. H. E. S. a. H. A. M. Kitsak, "Identification of influential spreaders in complex networks," *Nature physics*, vol. 6, no. 11, pp. 888-893., 2010.
- [77] J. Z. R. W. P. Z. Z. F. L. & L. X. .. Tang, "A discrete shuffled frog-leaping algorithm to identify influential nodes for influence maximization in social networks," *Knowledge-Based Systems*, vol. 187, no. 1, p. 104833, 2020.
- [78] X. J. M. W. Z. L. H. S. L. P. B. .. & B. S. Li, "Punishment diminishes the benefits of network reciprocity in social dilemma experiments," *Proceedings of the National Academy of Sciences*, vol. 115, no. 1, pp. 30-35, 2018.
- [79] C. B. S. B. A. d. P. M. M. P. N. P. M. S. D. Z. Z. Wang, "Statistical physics of vaccination," *Phys. Rep.* 664, 2016.
- [80] .. J. P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *Math. Sociol*, vol. 2, no. 1, p. 113–120, 1972.
- [81] S. B. R. a. T. L. Page, "The pagerank citation ranking: Bringing order to the web," Technical Report (Stanford InfoLab), 1999.

- [82] Z. T. Z. Q. S. H. Lü L, "The H-index of a network node and its relation to degree and coreness," *Nature communications*, vol. 7, no. 1, pp. 1-7, 2016.
- [83] M. E. Newman, "A measure of betweenness centrality based on random walks," *Social networks*, vol. 27, no. 1, pp. 39-54, 2005.
- [84] M. Granovetter, "The strength of weak ties," *American Journal of Sociology*, vol. 78, no. 1, p. 1360–1380, 1973.
- [85] R. S. Burt, *Structural holes*, Harvard university press, 1992.
- [86] M. Granovetter, *Getting a job: A study of contacts and careers*, Cambridge: Harvard University Press, 1974.
- [87] R. a. R. J. Alhajj, *Encyclopedia of social network analysis and mining*, Springer, 2020.
- [88] A. H. J. A. McCulloh I, *Social network analysis with applications*, John Wiley & Sons, 2013.
- [89] C. S. Namatame A, *Agent-based modeling and network dynamics.*, Oxford University Press, 2016.
- [90] M. Granovetter, "Threshold models of collective behavior," *American journal of sociology*, vol. 83, no. 6, pp. 1420-1443, 1978.
- [91] J. B. L. a. E. M. Goldenberg, "Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata," *Academy of Marketing Science Review*, vol. 9, no. 3, pp. 1-18, 2001.
- [92] H. H. F. C. Z. D. Guille A, "Information diffusion in online social networks: A survey," *ACM Sigmod Record*, vol. 42, no. 2, pp. 17-28, 2013.
- [93] A. Y. F. a. J. O. Borodin, "Threshold models for competitive influence in social networks," in *In International workshop on internet and network economics*, Berlin, Heidelberg, 2010.
- [94] T. C. N. S. M. W. a. A. V. Z. Carnes, "Maximizing influence in a competitive social network: a follower's perspective," in *In Proceedings of the ninth international conference on Electronic commerce*, 2007.
- [95] C.-Y. L. C.-L. W. T.-H. & S. C.-T. Huang, "A computer virus spreading model based on resource limitations and interaction costs," *Journal of Systems and Software*, vol. 86, no. 3, pp. 801-808, 2013.
- [96] M. (. Morris, *Network epidemiology: A handbook for survey design and data collection*, Oxford, UK: Oxford University Press, 2004.
- [97] C. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications*, New York: Springer-Verlag, 1981.

- [98] C. Hwang, Y. Lai and T. Liu, "A new approach for multiple objective decision making," *Computers and Operational Research*, vol. 20, no. 8, p. 889–899, 1993.
- [99] T. H. e. a. Cormen, *Introduction to Algorithms*, MIT Press and McGraw-Hill, 2009.
- [100] S. S. Skiena, *The algorithm design manual*, Springer International Publishing, 2020.
- [101] J. Leskovec, "<https://snap.stanford.edu/data/email-EuAll.html>," European research institution, 2010. [Online]. Available: <https://snap.stanford.edu/>. [Accessed 2021].
- [102] R. A. R. a. N. K. Ahmed, "<https://networkrepository.com/email-EU.php>," network repository, 2015. [Online]. Available: <https://networkrepository.com>. [Accessed 2021].
- [103] J. Leskovec, "<https://snap.stanford.edu/data/email-Enron.html>," Federal Energy Regulatory Commission, 2010. [Online]. Available: <https://snap.stanford.edu/>. [Accessed 2021].

المستخلص

من الموضوعات التي لها أهمية ولها ارتباط مع الشبكات التفاعلية الاجتماعية والانظمة التي تدير هذه الشبكات وما تحتويه من علاقات بين مكوناتها هو موضوع تحديد من هم المؤثرين والمسيطرين على الحركة والانتقال والتفاعلات الحاصلة بين اجزاء هذه الشبكات كمثال على ذلك حركة وانتقال المعلومات والايخبار وغيرها بين اجزاء هذه الشبكات ، لذلك اهتمت هذه الاطروحة في كيفية ايجاد المؤثرين والمسيطرين على المسارات الموجودة داخل هذه الشبكات من خلال انشاء وتكوين (الخوارزمية الهجينة لتطويق العقد المؤثرة) والتي اختصارها HMEIV.

حيث تتكون هذه الخوارزمية من ثلاث اجزاء مهمة هي انشاء (المؤشر المركزي القائم على سيطرة انتشار المعلومات) والذي له اختصار CISBC ، حيث ان هذا المؤشر كان له الدور في استكشاف الطرق الرابطة بين العقد الغير متصلة بشكل مباشر من خلال معرفة أدوار مواقع ومراكز كل العقد الباقية الموجودة في المسارات الرابطة بين كل الازواج ، حيث ان هذا المؤشر له ثلاثة مستويات متدرجة تبدأ من المستوى الاول الذي هو عبارة عن مستوى المسارات المنفردة بين زوج من العقد واعطاء القيم للعقد التي تؤثر وتسيطر على المسارات الاقصر بين كل زوج من العقد ، اما المستوى الثاني في المؤشر هو اكتشاف العقد التي تسيطر على مسارات عقدة معينة داخل الشبكة ومعرفة من هي تلك العقد التي بإمكانها ان تسيطر على نقل المعلومة لها ، المستوى الثالث يتضمن اكتشاف من هي العقد المسيطرة والمؤثرة على مسارات الشبكة باكملها.

الجزء الثاني في خوارزمية HMEIV هي انشاء (مقياس تفاعل الازان) والذي له اختصار IWS حيث هذا المقياس يتكون ايضا من ثلاث مراحل تقابل المراحل الموجودة في مؤشر CISBC، المرحلة الاولى هي مرحلة استكشاف وتحديد المسيطرين من خلال ازالة الضوضاء وتحليل المسارات ومعرفة من هي العقد الاكثر نشاطا وتبادل للمعلومات على كل زوج من العقد غير المرتبطة بشكل مباشر ، والمستوى الثاني هو معرفة من هي العقد التي تؤثر بشكل كبير على عقدة معينة من خلال نشاطها في تبادل البيانات ، المستوى الثالث يكون على مستوى الشبكة بأكملها لتكوين صورة واضحة حول ما هي العقد النشطة في التبادل على مستوى الشبكة .

بعد ان تم الحصول على النتائج من كلا المؤشرين CISBC , IWS تبدأ مرحلة تنفيذ الجزء الثالث من خوارزمية HMEIV التي هي عبارة عن ايجاد المفاضلة بين قيم المؤشرين من خلال استبعاد البدائل السيئة والتمسك بالبدائل والحلول الجيدة اي بعبارة أخرى هي عبارة عن مرحلة تصفيات تتم تنفيذها بواسطة خوارزمية TOPSIS .

مرحلة التقييم استخدمت فيها خمس طرق متنوعة حيث تضمنت مقارنة النتائج مع

المقاييس القياسية المعتمدة في الشبكات (Closeness ,Betweenness , Person Correlation Degree,Eigenvector, PageRank) ، حيث تم اعتماد Coefficient لبيان مقدار الترابط بشكل خطي بين نتائج HMEIV وبين المقاييس القياسية المعتمدة في الشبكات و اعتماد طريقة Frequency of similar values في اثبات ان الخوارزمية اعطت افضل النتائج في النسب القليلة للعقد التي تشترك بنفس القيم ، كذلك استخدم موديل Suspectable Infected لقياس مدى صلاحية العقد المؤثرة وتأثيرها على العقد الباقية ، والطريقة الاخيرة هي اختبار العقد المؤثرة الناتجة من خوارزمية HMEIV وبيان مدى تأثيرها على المسافة ومعدل المسافة والكثافة للشبكة عند إزالة هذه العقد ومقارنة هذه النتائج قبل وبعد الازالة .



جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل / كلية تكنولوجيا المعلومات
قسم البرمجيات

كشف العقد المؤثرة في شبكات التواصل الاجتماعي باستخدام نظرية المخططات

أطروحة مقدمة

الى مجلس كلية تكنولوجيا المعلومات في جامعة بابل وهي جزء من متطلبات نيل درجة الدكتوراه
فلسفة في تكنولوجيا المعلومات / برمجيات

من قبل

حسين لؤي حسن محمد

اشراف

الاستاذ الدكتور صلاح عبد الهادي كاظم البيرماني

1443 هـ

2022 م