----PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
UNIVERSITY MOHAMED BOUDIAF - M'SILA

FACULTY: Mathematics and Computer Science

DEPARTMENT: Computer Science
N°:……………………………………

1985
جامعة محمد بوضياف - المسيلة
UNIVERSITY MOHAMED BOUDIAF - M'SILA

DOMAIN: Mathematics and Computer Science

BRANCH: Computer Science
OPTION: IDO

Dissertation submitted to obtain Master degree

By: MR. OUNICI Khaled
and
MR. MECHIKI Djaber

SUBJECT

Towards a new simplistic approach for influential web users identification in social networks

Supported before the jury composed of:

Mr. LOUNNAS Bilal…….	University of M'sila	President
Mr. LOUCIF Hemza…….	University of M'sila	Supervisor
Mr. TARAFI Abdallah ….	University of M'sila	Examiner

Academic year: 2020/2021

Table of contents:

LIST OF FIGURES

LIST OF TABLES

List of abbreviations:

SNA : social network analysis.

LTM : linear threshold model.

ICM : independent cascade model.

LIM : linear influence model.

ACM/IEEE : Institute of Electrical and Electronic Engineers/Association for Computi ng Machinery.

General Introduction

Behavior in social media is defined as how users interact. These interactions include: creating content by an individual (i.e. publishing) and sharing with its friends, re-sharing content shared by the individual's friends, submitting comments or liking friends' postings. Analyzing this behavior and predicting the next steps that users can take, and their impact on other users in the network is essential to bring various advantages to companies and non-profit organizations.

This article mainly studies the information dissemination model, which is used to predict the number of infected nodes in the network and the ability of individuals to disseminate information through the network. We try in this work to propose simple way, method or using metrics in order to identify influencer users in network. We also aim to use different models to analyze the diffusion process and focus on the structure of the network to understand how it affects the flow of information. This analysis will be done on Twitter user's network. The discussion of the network regards them as static structures. We take snapshots of nodes and edges at specific moments, and then analyze their structure and the spread of information processes.

In order to achieve the goal of this work, we organized the thesis as follows:

Chapter 2 introduces some important definitions in the field of social network analysis from the basis of graph theory to understand the structure of the network. It also comes with network representation and visualization. The network types and the main indicators of SNA are also described in this chapter.

Chapter 3 is devoted to the definition of information dissemination and influence. Likewise, it explains the importance and purpose of researching and analyzing the flow of information on social media. Some diffusion models from literature and related work are also provided.

Chapter 4 specifically introduces the realization and explanation of the experiment. First, we must clarify the phenomenon of information cascade. After that, the information diffusion modeling and process parameters are also described.

However, in this chapter, we will also describe the tools and data sets used to achieve our work goals. Furthermore, implementation and experimental results will be discussed in this chapter.

Finally, in addition to the experimental steps and results, the ***general conclusion*** summarizes all chapters.

Chapter 1:

Social Media And Social Network Analysis.

Introduction:

People's lives are based on social relationships that connect people, communities, and organizations in the form of networks. Early social networks were restricted by time and space, but due to the development and widespread use of the Internet, these restrictions have been eliminated. The popularity of the Internet, the amount of time people spend online, the highly interactive nature of Web 2.0 sites and their impact on our daily behavior indicate that the World Wide Web is actually a part of modern society. Early social networks were based on face-to-face interaction, while online social networks used information and communication technology tools to reduce the difficulty of interaction. Moreover, new and appropriate hardware and software have improved communication efficiency, allowing people to share information cheaply and reliably anytime, anywhere.

1.1. Social media

Social media describes online tools that people use to share content, personal data, insights, experiences, opinions, and the media itself, thereby facilitating conversations and online interactions between people. It is about converting monologues (one-to-many) into dialogue (many-to-many) [1] [2].

These tools contact individuals in different ways. The outstanding types of social media are: threaded conversations (email, Usenet newsgroups), WWW (homepage), collaborative creation (Wiki, shared documents), blogs and podcasts (LiveJournal, Twitter), social sharing (YouTube, Flickr), Digg), social networking services (Facebook, LinkedIn), online markets and products (eBay, Amazon), virtual worlds (Second Life, World of Warcraft), mobile-based services (location sharing and games)...[3]

Figure 1.1: Social media logos[1].

Social media can be seen as a resource of useful information, opinions and behaviors about different areas of interest. Currently[2], 500 hours of videos are uploaded to YouTube every minute, and the number of videos stored now exceeds

[1] https://www.perusal.in/events/learn-social-media/7
[2] https://www.omnicoreagency.com/youtube-statistics/

that of the entire 20th century. There are 500 million tweets every day[3], which means that if only a million tweets are worth it, you can still spend the entire day on Twitter. There are several times more web pages than people in the world. In addition, technological and economic systems have also become dependent on social media, which makes it very difficult to predict and reason about their behavior.

1.2. Social network analysis

Social network analysis is a wide application of network science in the research and application of human relations. It provides a systematic method to evaluate the work of social media based on scientific evidence. In recent years, social media and social network analysis methods (SNA) have aroused great interest in the social and behavioral science community. Psychologists, sociologists, anthropologists, economists and statisticians have made significant contributions, making it practically an interdisciplinary field of research. This growth coincides with the continuous development of methods used to collect and visualize network data in order to analyze the relationships between people, groups, organizations, and other knowledge processing entities on the Web [34].

The following sections are devoted to the history of social network analysis and its application in the real world. Later, we describe the purpose of this thesis and clarified its chapter organization.

1.2.1. Social Network Analysis history :

Social network analysis originated in the fields of social science, network science and graph theory. At the beginning of graph theory, the mathematician Leonhard Euler solved the Konigsberg Seven Bridges problem in 1736. Many years later, Polya (1935, 1937), Paul Erdös and Alfré de Rényi published more work and results. The concept and expansion of graph theory has established the foundation of social network analysis.

[3] https://www.internetlivestats.com/twitter-statistics/

In the field of social sciences, the work of the first sociologist "August Comte" in the 1800s and the work of the sociologist "George Simmel" in the early 1900s are

the roots of social network analysis. Both of them define society as a group composed of various relationships in which people can influence each other.

In the 20th century, anthropologists studied the kinship system and created a symbol system related to social network analysis, aiming to create a systematic language to record social relationships.

The work of Jacob Moreno and his collaborators dates back to 1930. They added important algorithms and indicators for modern social network analysis. During this period, cost and the lack of available network data sets and computing resources are the limiting factors for the widespread use of social network analysis (especially in enterprises and organizations).

In the 1950s, Nadel published articles on social roles and the social structure that defines social roles.

Between the 1960s and the 1970s, more and more scholars devoted themselves to combining the different trajectories and traditions of social network analysis in the Department of Social Relations of Harvard University, University of California, University of Chicago and Michigan State University.

The recent explosive growth of computer-mediated social relations and the consequent decline in the cost of creating network data have made network methods more and more practical. As mobile devices and social media services track and capture more detailed information about our interactions and associations, network analysis becomes more and more useful [35].

1.2.2. Social Network Analysis applications :

Social network analysis methods have been applied in a wide range of fields and have become more and more popular for individuals and businesses. Business applications are applying SNA to help manage business challenges, gain in-depth understanding of markets and communities, and establish stronger industry relationships [08]. Organizations and project management departments rely extensively on SNA, and its goal is to hire the best talents with the best network based on the idea that human resources directly affect important decisions and organizational results. Some organizations use SNA tools to learn how to improve collaboration and communication between departments to increase the effectiveness of the organization. Krebs and Holley proved that network maps can provide a clear snapshot of the business ecosystem at a specific point in time, which helps answer many questions in the community building process, such as: Is there a right connection? Who plays a leading role in the community ? Who is not, but who should it be? Do you share ideas and take action? Which companies can provide a better return on investment ?

Another important application of SNA is to measure influence, spread and infection. Infectious diseases spread through the human network may be bad (disease, gossip), good (thoughts and information) or neutral (money and investment) [36]. SNA method used to analyze the spread of different diseases. The feasibility and value of network analysis were checked to complement the routine tuberculosis (TB) contact investigation procedures[37]. Jafa et al. used SNA tools to investigate the spread of HIV in the state prison system from 1988 to 2005 [38].

In academia, Laszlo Barabasi and his colleagues analyzed the evolution of the social network of scientific cooperation by mapping an electronic database containing all relevant journals of mathematics and neuroscience over the eight-year period (1991-98) [39]. In their research, by applying social network analysis to the co-author network of past ACM, IEEE and ACM/IEEE joint digital library conferences, they studied the state of the digital library field after ten years of activities) [40].

Similarly, political scientists and complaint organizers also use SNA methods in political and power analysis. Scott and others studied how interpersonal social networks can help explain voting behavior. Krebs uses spacing and proximity indicators to compare the effects of physical location and location in the network to observe changes in power in the network over time.

After the tragic events of September 11, 2001, scientists focused on analyzing and modeling terrorist networks to prevent attacks and understand how to drive individuals into terrorist organizations.

Moreover, a successfully designed computer network depends on how its various elements are connected to each other to form an overall system with as few single points of failure as possible. To achieve this goal, social network analysis has been applied to the design of computer networks, and link changes are recommended.

Chapter 2:

Social Network Analysis Background

Introduction:

Social networks and social network analysis are called in multiple domains and contexts. Because of this difference, we will focus on we think some basic definitions in social network analysis are important for analyzing social networks media. We start with the basic concepts behind graph theory to understand the research of graph theory. Network structure. Then we talk about network representation and visualization and its Different possible types. After that, it's important to cite the main social network analysis index. Later, we mentioned some of the reasons and importance of studying information flow in social media networks. The purpose of this chapter is to understand the basics of SNA. Able to explain the structure of social networks based on the concept of graph theory and social interaction Network analysis measures.

2.1. Graph theory basic definitions

Graph theory is very important and useful in social network analysis because it provides a mathematical model of network structure. The social scientist John Barnes described graph theory as "the term jungle, in which any newcomer can plant trees." Therefore, the following are some basic concepts and definitions surrounding the graph [4] [19].

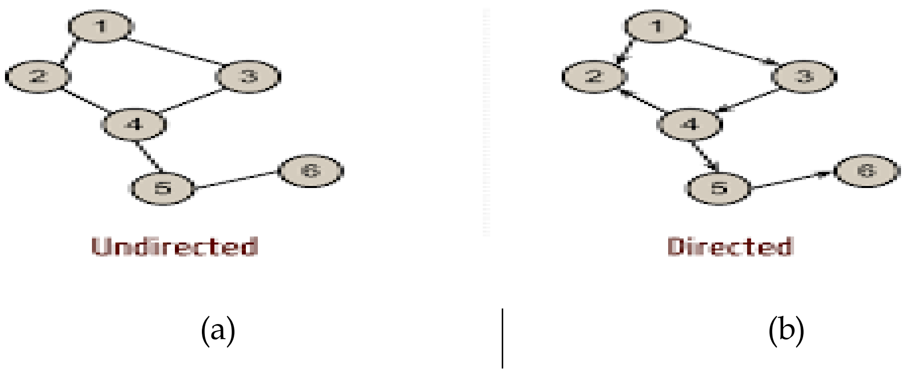


Figure (2.1): Directed and undirected graphs[4]

A graph is a way to specify the relationship between a collection of items. It consists of a group of objects (called vertices) are connected to some of these objects in pairs by links. Figure (2.1) shows two simple graphs. In social networks, vertices (also called nodes, Agents, entities or actors) usually represent people or social structures, such as Workgroup, team, organization, state and even country. They can also represent Content such as web pages, keyword tags, or videos. They can even represent the body or virtual places or events. Links connecting vertices (also called edges, ties, Relationship or simple connection) is the foundation of the network. They can Represents many types of relationships, such as collaboration, relatives, friendship, Investments, hyperlinks, transactions, shared attributes and information flow [5] [6].

[4] http://stoimen.com/2012/08/31/computer-algorithms-graphs-and-their-representation/

In social networks, there are two main types of links: corresponding strong links Friends are called strong ties, and weak links corresponding to acquaintances are called Weak relationship.

Directed graphs and undirected graphs: A directed graph is a graph whose nodes are located Connected by directed edges (arrows) that represent an asymmetric relationship. in Undirected graph, the edges indicate symmetrical relations, and the direction is not important. Figure (2.1)(b) shows a directed graph, and Figure (2.1)(a) shows an undirected graph [5] [6].

A path is just a sequence of nodes and a sequence of edges connecting these nodes, it has the characteristic that each successive pair in the sequence is connected by edges. A kind a simple path refers to a path that does not repeat nodes [5] [6].

The length of the path is the number of steps included from beginning to end, namely Refers to the number of edges in the sequence that contains it. The length is used to know whether two nodes are near or far in the graph (or network). The shortest path between two nodes is called distance [5] [6].

The diameter is the length of the longest distance of the graph [5] [6].

A cycle is a path with at least three edges, where the first and last nodes are identical. A loop containing only three nodes is called a "triple axis". Usually, the period is currently to allow redundancy [5] [6].

Connectivity: If there is a path between each node pair, it means that the graph is connected they [5] [6].

Weighted graph: If a number (called a weight) is assigned to each edge of the graph, it is called a weighted graph. These weights may represent cost, length or capacity; it depends on the question being studied. The sum of all weights is the weight of the chart [5] [6].

A component (in the disconnected diagram) is a group of connected nodes. Possible Treated as a connected graph in isolation. The concept of components reasoning about the network is also much smaller [5] [6].

Mega component: Contains a large part of the connected components of all components nodes in the figure are called giant components. It usually exists in large and complex environments internet [5] [6]..

Bridge: If deleting an edge causes its endpoints to be in different components of the graph, it is called a bridge. In other words, this is the only route between the two endpoint. If its endpoints do not share a common neighbor, it is called a "local bridge". Unlike bridges, local bridges are included in the loop. In Figure (2.2), the edges B-D and B-C is a bridge [5] [6].

Structural hole: This is the "blank space" between the two sets of nodes in the graph Otherwise, they cannot interact closely. In other words; a structural hole is a missing bridge. As shown in Figure (2.2), node B has multiple local bridges across a structural hole in the graph. Use structural loopholes in social networks to control People who have not been contacted [5] [6].

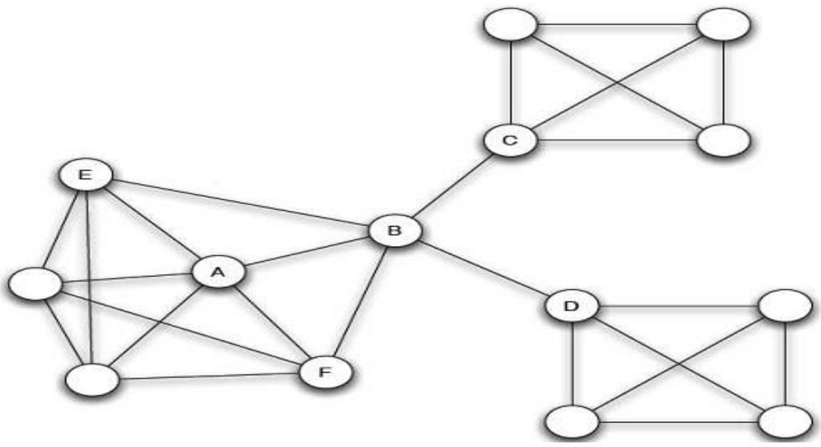


Figure (2.2): Bridges and structural hole[5]

2.2. Representing and visualizing networks data

Matrices are popular mathematical objects for processing graphics. Generally, when using a matrix to model a graph, the rows and columns of the matrix represent the nodes of the graph, and the values in the cells of the matrix represent the edges

[5] https://www.researchgate.net/figure/According-to-the-theory-of-structural-holes-node-As-bridging-role-in-the-social-network_fig4_221444745

(or the absence of edges) between the corresponding nodes. There are two main types of matrices used to represent social networks: adjacency matrix and incidence matrix [7] [8] .

The adjacency matrix is a square matrix, where rows and columns represent the same set of nodes. If there is no relationship between vertices v_i and v_j, the value is 0. If there is a relationship between two vertices, the value is 1 in the case of a non-weighted graph, or in the case of a weighted graph of the link between the pair of vertices. In the case of a directed graph, the rows represent the source nodes and the columns represent the end nodes of the edges. Table (2.1) shows the adjacency matrix of figure (2.1) (b). The value of cell (1-1) is 0, which means there is no self-loop in the figure. Similarly, in cell (12) the value is 1; in cell (21) the value is 0. This means that the link is from 1 to 2, which proves that the graph is directed.

The incidence matrix represents attribute data, where rows represent individuals, and columns represent individual characteristics, behaviors, or other. The principle of representing data is the same as explained in the adjacency matrix. The only difference is that the rows and columns do not represent the same person.

	1	**2**	**3**	**4**	**5**	**6**
1	0	1	1	0	0	0
2	0	0	0	0	0	0
3	0	0	0	1	0	0
4	0	1	0	0	1	0
5	0	0	0	0	0	1
6	0	0	0	0	0	0

Table (2.1): Adjacency matrix

Column 1	**Column 2**
1	**2**
1	**3**
3	**4**
4	**2**
4	**5**
5	**6**

Table (2.2): Edge list representation

Another form of network representation is called an "edge list." As shown in Table (2.2), it is just a list of all edges in the network, representing the same graph (b) in Figure (2.1). The people in the first column point to the people in the second column.

To describe the value or other attributes of each edge, more columns can be added. By repeating the same node in the two columns, self-loops can also be represented in the edge list. The edge list representation method is better than the matrix in terms of required storage space, because only existing relationships are represented in the edge list.

The characteristic of modern social network analysis is the use of visualization of complex networks. Unlike using a matrix or edge list to represent network data, viewing a network diagram can provide an instructive overview of the network structure Understanding and insights, and showing the main points about the network. Visualizing the network is a powerful point because of the ability to map attribute data to visualize vertex and edge attributes

Visualizing a large dense network diagram is not as easy as it seems (see Figure (2.3), which shows an example of a dense network diagram). Some obstacles (such as vertex occlusion and edge crossing) may limit the organization of the created network graph. Good network visualization must ensure four main principles:

1) Every vertex is visible.
2) The degree of each vertex is countable (that is, the number of connections starting or ending at that vertex).
3) Each edge can be tracked from source to destination.
4) Clusters and outliers can be identified.

To realize these principles, careful preparation, different layout and filtering techniques must be used. We will not solve these problems in this work.

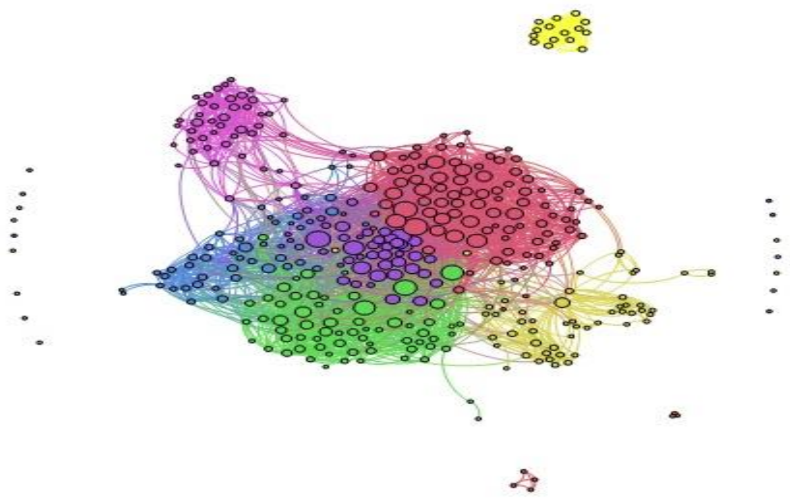


Figure (2.3): An example of a dense network graph of Facebook network counts [8]

Recently, there are several different tools that can facilitate quantitative or qualitative Describe the functions of the network through digital or visual representations to analyze social networks. They usually consist of a graphical user interface (GUI)-based software package or a software package built for scripting/programming languages. GUI software packages are easier to learn, while scripting tools are more powerful and extensible. Some are free and open source, while others are commercial. Commonly used and well-documented tools are: Pajek, Gephi, NodeXL, UCINET/NetDraw, GUESS, NetMiner, NetworkX library for Python and R statistical programming language package.

2.3. Types of networks

The size of the social network, the type of vertices, the nature of the edges, and how they are formed will vary. In this section, we introduced the basic types of networks and the differences that may affect the metrics and their interpretation.

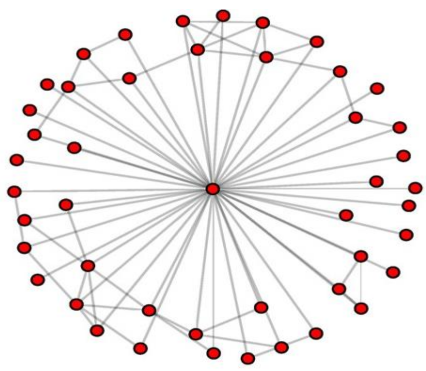


Figure (2.4): Egocentric network[6]

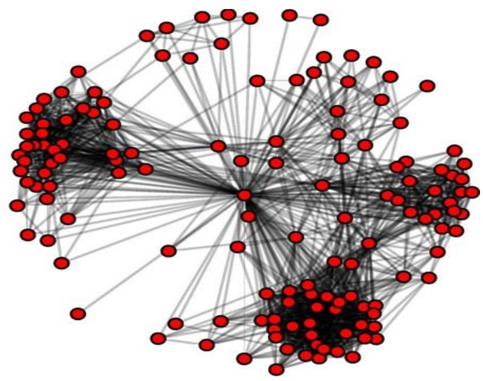


Figure (2.5): Full network[6]

1. **Egocentric networks**: The individual who becomes the focus of attention in the network is called the "self", and the person connected to him or her is called the "substitute" (see Figure (2.4)). A self-centered network refers to a specific network that treats individuals as a network. Connect to self. A good example of a personal-centric network is your personal Facebook network because you have established connections with all your friends [7] [8].
2. **Full networks**: It includes all interested individuals or entities and the connections between them, in which all conceit is treated equally. It is also called a complete network, and It is usually created when a single system acts as a hub in a connected person or group. For example, the Twitter network includes all users of the service and the connections between them. But in fact, it is more meaningful and feasible to create a partial network by selecting samples or slices of the entire network based on certain conditions. Figure (2.5) is a complete network, and the network in Figure (2.4) is a part of it, called a partial network [7] [8].
3. **Unimodal networks**: These are standard networks that contain one type (mode) of nodes. For example, this means they connect users to users, documents to documents, and videos to videos, instead of connecting users to videos. The previous figures (2.4) and (2.5) are unimodal networks, where nodes are represented by the same color to describe that they belong to the

[6] https://www.semanticscholar.org/paper/andom-errors-in-egocentric-networks-Almquist/98cc0b34f2d4a14b37863b47af30e222696d2b52

same type[7] [8].

4. **Multimodal networks**: They are networks that connect different types of nodes. For example, connecting users to blogs, or connecting people to organizations. In this type of network, it is best to use various shapes and colors to separate different types of nodes. A common type of multimodal network is the bimodal network, which happens to include two types of nodes. In Figure (2.6), the nodes represented by the black disks are users, the orange squares represent the forums they publish, and the blogs represented by the blue diamonds are blogs [7] [8].

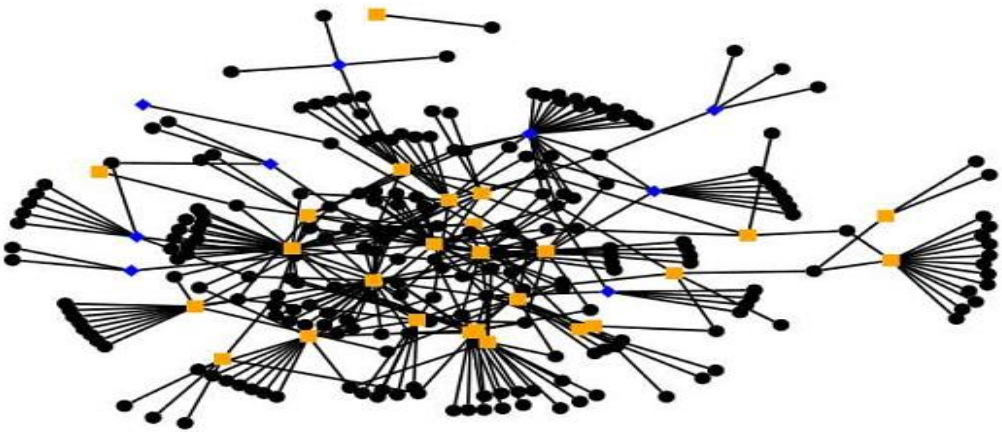


Figure (2.6): Multimodal network[7]

5. **Affiliation networks:** Affiliation network is a bimodal network, including individuals and some events, activities or content associated with them. A simple example is a network that connects users to the Wiki pages they edit. This kind of network can be transformed into a user-to-user network and a subordinate to subordinate network (in the example of a Wiki page, from article to article). Generally, this method can be used to associate all types of objects [7] [8].
6. **Multiplex networks:** They are networks that contain many types of connections. A good example of this type is the Twitter network, because it can include three types of directed edges: follow relationship, "reply"

[7] https://www.semanticscholar.org/paper/andom-errors-in-egocentric-networks-Almquist/98cc0b34f2d4a14b37863b47af30e222696d2b52

relationship, and "mention" relationship. In order to distinguish between various connection types, colors, different edge types or edge labels can be used. An example is given in the figure below (2.7), which visualizes users based on the Lostpedia Wiki user-to-user affiliation network, which is based on the number of unique pages that the user has edited. There are two types of edges: the edge that connects users based on the co-editing theory page (green) and the edge that connects users based on the co-editing of the article (maroon).

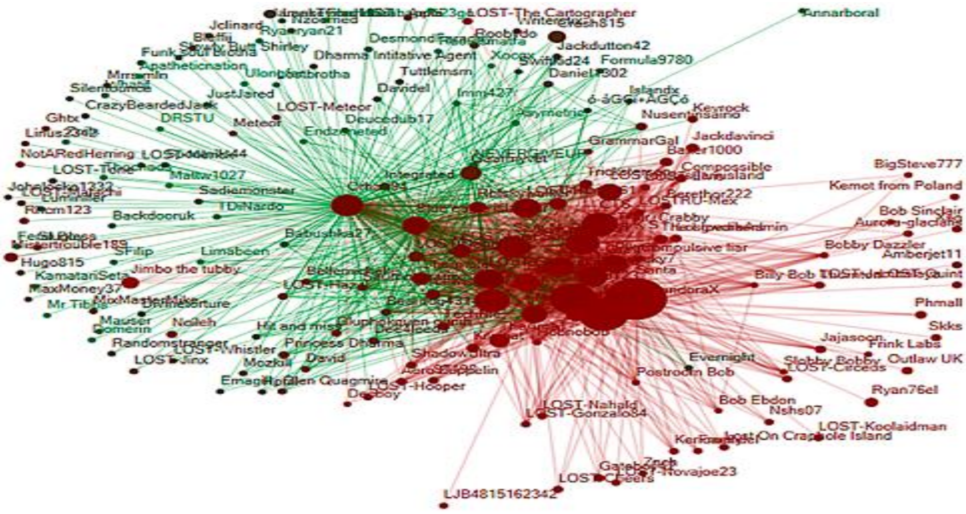


Figure (2.7): Multiplex network [8]

2.4. Social Network Analysis Metrics

Researchers in the field of social networks have collaborated to create theories and algorithms for calculating new quantitative measures of social networks. These metrics help to understand the global structure of the network, which allows us to: evaluate the effectiveness of the network when communicating, estimate network connection failures, predict the development of the network, understand the re-division of people and activities, compare networks and track changes in time . Some metrics describe the entire network, while others are specific to subgraphs of the network or components. Other metrics will be calculated for each vertex in the network.

Density: It is an aggregate network metric used to indicate the cohesion of the network (the degree of interconnection of vertices). Density refers to the number of relationships observed in the network divided by the total number of possible relationships. In Figure (2.8), the density of the undirected graph is 5/6 = 0.83, which means that the network is dense. In the second network (directed graph), the density is 5/12 = 0.41, which means that the network is sparse [7].

Figure (2.8): Examples of computing the density of a network[8]

Centralization: The overall concentration of social networks is calculated based on the local concentration of nodes. The calculation of concentration depends on the definition of local concentration that we consider. If we consider the centrality of local degrees, the overall centrality highlights the existence of high concerns in social networks. The global concentration measure based on local intermediate centrality provides an indication of the dependence of network connectivity and efficiency on its participants. Finally, global centrality based on local proximity centrality can measure the performance of communication networks, including traffic information. The global centralization of the network allows us to estimate the dependence of the network structure on its members [7].

Degree centrality: Degree centrality is a simple count of the total number of connections linked to vertices. In other words, it measures the total number of edges connected to a particular vertex. Given the example in (a) in the figure(2.9), nodes 3 and 5 have the highest degree. For directional networks, there are two levels of measurement. The in-degree is the number of connections pointing inward from the vertex. Out degree is the number of connections starting from a vertex and pointing

[8] https://www.slideshare.net/gcheliotis/social-network-analysis-3273045

outwards to other vertices. Degree centrality regards the node with the highest degree as the most central, but it does not distinguish between quantity and quality.

Betweenness centrality:: The intermediate center point of a vertex is the fraction of the geodesic path that the vertex falls between other vertices. This metric focuses on the participant's ability to act as an intermediary between any two other participants in the network. The role on the geodesic path has a strategic position in the cohesion and information flow of the network, especially when the path is unique. Usually in social networks, vertices with high intermediateness will have great influence because they are located between other vertices. Like the degree centrality measure, the middle degree center regards the node with a higher middle degree as the most center. In the example given in Figure (2.9)(b); node 5 has the highest intermediateness [value calculated by NodeXL] [7] [8].

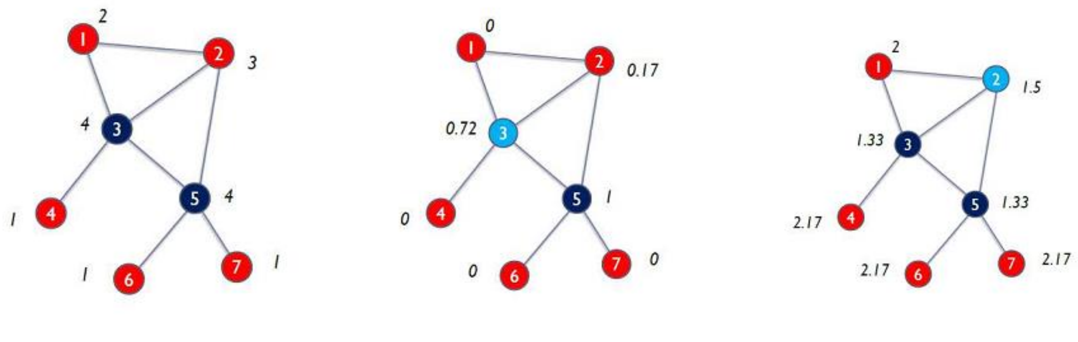


Figure (2.9): Centrality measures examples[9]

Closeness centrality: It is the average geodesic distance from a vertex to every other vertex. In the sense that the average network distance of the vertices is shorter than that of other vertices, the proximity of the vertices to the center position is low [7]. Tight centrality reveals the ability of a node to quickly connect with all other participants in the network. In a directed graph, the proximity centrality calculated on the outgoing and incoming edges respectively represents the ability of the character to reach or to reach in the entire network. In the example in Figure (2.9)(c),

[9] https://www.slideshare.net/gcheliotis/social-network-analysis-3273045

nodes 3 and 5 have the lowest tightness, which means they can be reached quickly [7].

Eigenvector centrality : This is a measure of the importance of nodes in the network. Unlike the degree centrality that only provides the number of vertex connections, eigenvector centrality recognizes that not all connections are equal, so connecting some vertices is more beneficial than connecting other vertices. An example is shown in Figure (2.10), node 3 has the highest value, where the output value comes from NodeXL[10] [7].

Clustering coefficient: The clustering coefficient of node n is defined as the probability that two randomly selected friends of n become friends of each other. In other words, it is the number of directional links that exist between node neighbors divided by the number of directional links that may exist between node neighbors [7]. The range is from 0 (when all friends of the node are not friends of each other) to 1 (when all friends of the node are friends of each other). The clustering coefficient of the graph is the average clustering coefficient of all its nodes. Specifically, the clustering coefficient is a 1.5 degree measure of the density of a self-centered network. When these connections are dense, the clustering coefficient is very high [7].

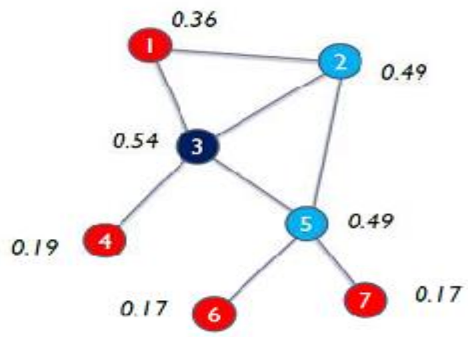


Figure (2.10): Eigenvector centrality example[11]

[10] https://en.wikipedia.org/wiki/NodeXL
[11] https://www.slideshare.net/gcheliotis/social-network-analysis-3273045

The centrality measures (degree centrality, intermediate centrality, closeness centrality) highlight the most important participants (nodes) and the strategic location of the network. The main problem is to define which node is more important than another in a different network.

2.5. Level of adjacent vertices in a network

The last thing we should explain about the network is the level of neighboring nodes in the network that we will use in the future. Level refers to the depth of the self-centered network. Let us give an example in Figure (2.11). Suppose that the self-specified in (a) is the purple disk at the center of the network, and the other light blue vertices represent changes directly connected to him (such as his friends). In this case, the depth level of the egocentric network is 1, because it only contains links to and from the ego. In (b), there is also a correlation between self-changes, indicating that these vertices are not only related to self, but also related to each other. The depth of the network is 1.5. The last network in (c) contains self-modifying blue changes that are not directly related to the center of the network. This is the second level of the egocentric network [7] [8].

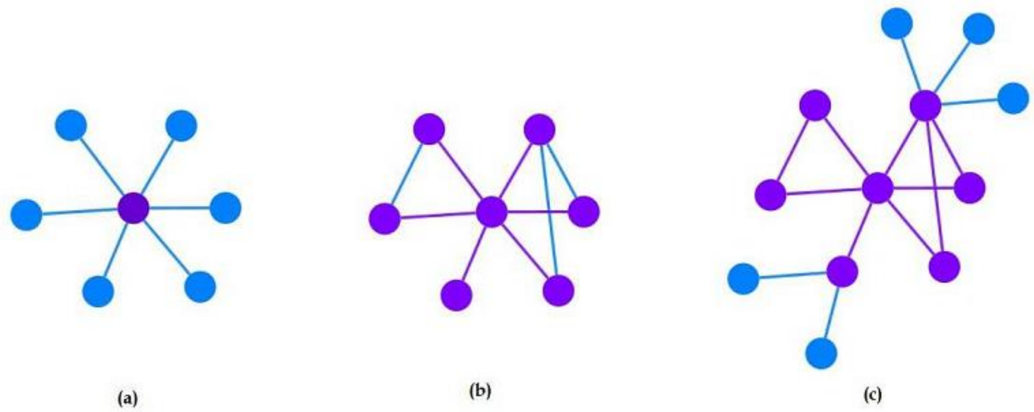


Figure (2.11): Egocentric network with a depth level equal to n.

(a) n = 1. (b) n = 1.5. (c) n = 2.

Summary

In recent years, social network analysis has increased dramatically. The concepts and metrics of network analysis have become popular. Powerful methods and visualization tools greatly reduce the difficulty of collecting and analyzing network data. Analyzing the information flow in social networks (more precisely, social media) can help organizations, business managers, and individuals make more effective decisions in hours instead of months.

Chapter 3:

Influence and Information diffusion Models

Introduction

Research on social networks has exploded over the last decade. To a large extent, this has been fueled by the spectacular growth of social media and online social networking sites, which continue growing at a very fast pace, as well as by the increasing availability of very large social network datasets for purposes of research [10].

This chapter is devoted to clarify the definitions of influence and information dissemination in social networks. It also addresses distinguished diffusion models in the literature and the importance of analyzing the flow of information.

3.1. Information diffusion

Information diffusion is a key aspect of interest for those who, for commercial reasons, want to get the right publicity to the right people. It is also of interest to analysts who want to track what is being transmitted through a given network and by whom [14]. One could say that, on the one hand, there is the legitimate diffusion of information (solicited, such as news bulletins a user subscribes to, or relevant targeted commercial information), and on the other hand, there is the undesirable diffusion of information (unsolicited, such as spam or ad hoc badly targeted commercial information and criminal activity under false pretenses). Another focus of research is how users propagate information from one to another (analogous to word of mouth); [14]

The analysis of information diffusion plays an important role in understanding and predicting information flows, and has been widely discussed in various disciplines including sociology, economics, physics and computer science. One of the applications is the analysis of collective action, in which a small group of individuals have been initially deprived of a right and ask for the return of this right. Their behaviors may then influence their friends or colleagues, triggering a cascade in the social network [15].

3.2. Influence in social networks

The notion of influence can help us in better understanding how we can help advertisers and marketing professionals design more effective campaigns. According to researches, a relatively small number of people are able to persuade, or influence, others in their society. By targeting these influential people, we will be able to achieve large-scale influence with a small marketing cost thereby maximizing our ability to sell products or services within the society.

This method has been mainly utilized by large companies to promote their products using high profile people, such as singers, actors, members of parliament and political leaders.

In other words, companies hire influencers as local opinion leaders with different expertise to promote their products or services [11].

This theory assumes interpersonal relationships among users and the receptiveness of a society to opinion change as key factors of influence. The technique of collaborative filtering is a direct outcome of this theory as a marketing strategy. These theories remained largely unproven prior to the emergence of social networks, such as Twitter and Facebook, due to lack of empirical data for validation. Furthermore, this type of study is not straightforward because we do not have specific criteria for quantification of influence in a society [11].

In Webster's dictionary the word influence is defined as "The power or capacity of a person or things in causing an effect in indirect or intangible ways". The study of influence dates back over a century. The role of influence and its effects is studied extensively in sociology, communication, marketing and political science and in understanding peer pressure, obedience and leadership. Three broad categories of social influence were introduced by Hernert Kelman [20]: (i) Compliance, which is defined as agreement among people while keeping their dissenting opinions private. (ii) Identification, in which people are being influenced by someone who is liked and respected, such as a famous celebrity. (iii) Internalization, in which people accept a belief or behaviour and agreement is made publicly and privately [20].

Morton Deutch and Harold Ferald [13] described two psychological needs; namely, the need to be right and the need to be liked, which are associated with informational and normative social influence respectively [12]. Accepting information from another person is called informational influence as evidence about reality. In normative influence people try to conform to the positive expectation of others. Normative influence leads to public compliance, whereas informational influence leads to private acceptance. Informational influence is applicable when people are uncertain because of social disagreement or stimuli are ambiguous. Some of the factors in being influential are considered to be: charisma, bully pulpit, peer pressure, psychological manipulation, psychotherapy, reputation, emotion, social trends and social structure.

In modern theories, influence is considered to be a subjective topic, which adds the dimension of topics and expertise to this domain. This arises as a consequence of influence having a root in trust, and trust is subjective [11].

In order to identify opinion leaders, Social Network Analysis (SNA) evaluates the location of actors in the network as well as analyzing their interactions [11]. The class of network analysis which concentrates on measuring centrality of users defining their roles in a network is called structural analysis.

3.3. Why study the flow of information in Social Media

In this short section, we outline a few of the many potential applications that could benefit from understanding the structure of networks and information propagation in social media.

- The first reason is "Trust": Trust is the foundation of interpersonal relationship and interaction as in social networks as in real world. Users in a network tend to trust other adjacent users more than other far random users [16].

The factors affecting trust between users could be classified into three dimensions: (i) similarity between users, (ii) familiarity between users, (iii) users' social reputation. The last two dimensions can be used to reflect asymmetry of trust between users [16].

A number of research systems have already been proposed to exploit this trust. Haifeng et.al [16] have used the social network to mitigate Sybil attacks in distributed systems, exploiting the fact that real people tend to have a diverse set of social relations. Scott et.al [16] presented Reliable Email (RE); a system that determines the social network distance between the sender and the receiver of an email to aid SPAM detection. A deeper understanding of the underlying topology of information flow in networks is an essential first step in the design and analysis of robust trust.

- The second reason to study the flow of information in media networks is "common interests": There is a recent rise of interest-based social networks (e.g., Pinterest and Goodreads), which connect users by relations based on shared interests. In these networks, links between users manifest from selecting common interests from a pool of available interests. For example, two users may establish a link on Pinterest because of both liking dog photos, or on Goodreads due to reading the same novel [16].

Users usually browse their friends' Facebook pages or YouTube channels because they are likely to find content that is of interest to them. Google for example use social networks to rank Internet search results relative to the interests of a user's social network using the content viewed and search results clicked on by members of a social network.

Clearly, understanding the structure of online social networks, as well as the processes that shape them, is important for these applications. For example, efficient algorithms are needed for inferring the actual degree of shared interest between two users, or the reliability of a user.

- The third reason is "Content exchange". The phenomenal popularity of social networking sites like YouTube, Flickr, and Twitter represents a shift in how content is published, located, and distributed on the internet. Understanding how content diffuses through these networks and becomes popular over time is not only of academic interest, but is increasingly important in commercial advertising, in political campaigning, and ultimately to society. Understanding how information flows in online social networks can aid designers of current social networking systems. It also has the potential to improve search algorithms [9].

By examining the content that users view or mark as a favorite, sites may be able to suggest other content that may be of interest to the user. New sites use social networks to predict music preferences, find potential job applications, and share content. By understanding the network structure and the properties of information flow, future systems will be designed [9].

3.4. Diffusion models in social networks

A rich body of this work has been devoted to the analysis of the propagation of information, influence, innovations, infections, practices and customs through networks. Can we build models to explain the way these propagations occur? How can we validate our models against any available real datasets consisting of a social network and propagation traces that occurred in the past? These are just some questions studied by researchers in this area [10].

Information propagation models find applications in viral marketing, outbreak detection, finding key blog posts to read in order to catch important stories, finding leaders or trendsetters, information feed ranking, etc. A number of algorithmic problems arising in these applications have been abstracted and studied extensively by researchers under the garb of influence maximization.

The literature offers four basic approaches to modeling influence propagation in social networks: linear influence model, cascade model, threshold model and epidemic model [10]

Before we discuss these models, we take a sight at the specific vocabulary in the context of information diffusion in a network. Given a network G = (V,E) with V is the set of vertices, and E the set of the edges existing in the network. A vertex $v \in V$ is said to be active if the information has reached the vertex and was accepted by it. "Contaminated vertex" is the synonym of active vertex, mostly used when the study involves some viruses instead of information propagation. If the information didn't reach the vertex or the vertex rejected it, then the vertex is said to be inactive or non-contaminated. A vertex can change its state during the process of information diffusion from inactive to active, but not the opposite, in order to simplify the models.

There must be an initial set of vertices activated to start the diffusion process targeted for initial activation. They are called 'initial adopters' of the information. The influence of this initial set of vertices is the expected number of active vertices in the end of the diffusion process [10].

We delve deep into the key problem of influence maximization, which selects key individuals to activate in order to influence a large fraction of a network. Influence maximization in classic diffusion models including both the independent cascade and the linear threshold models is computationally intractable, and we describe several approximation algorithms and scalable heuristics that have been proposed in the literature [10].

3.4.1. Epidemic Model

Epidemic models are used originally to study the spread of diseases among biological populations, including humans. Various epidemic models have been proposed and extensively studied over many years, starting from the early 20th century. Recently, researchers have also applied epidemic models to the diffusion of information and influence in social networks [10].

Each individual (or node) transitions between several possible states, which typically include state S (for susceptible), state I (for infected), and state R (for recovered or removed). A node in state S does not have the disease but is susceptible to getting the disease upon contact with an infected node. A node in state I has the disease and can transmit the disease to susceptible nodes upon contact, with infection rate , which is interpreted as the probability of successful transmission of the disease from an infected node to a susceptible node in a time unit [19].

A node in a network goes through three potential stages during the course of the epidemic: (1) Susceptible to infection from its neighbors, (2) Infectious if it has caught the disease and has some probability of infecting each of its susceptible neighbors, (3) Recovered or Removed from consideration (in case of death) after the full infectious period. Given a directed graph representing the contact network; so an edge pointing from node v to node w in the graph means that if v becomes infected at some point, the disease has the potential to spread directly to w. each node has the potential to go through the Susceptible-Infectious Removed cycle (S-I-R) [19].

The progress of the epidemic is controlled by the contact network structure and by two additional quantities: $\boldsymbol{p}$ (the probability of contagion) and $\boldsymbol{tI}$ (the length of the infection) [19] . Figure (3.1) shows an example of the SIR model in a small network through successive steps. In each step, shaded nodes with dark borders are in the I state and shaded nodes with thin borders are in the R state.Recovery may be followed by another susceptibility state, giving a S-I-S model, depending on the disease. However, it is questionable whether the S-I-R and S-I-S pattern applies to information to the same extent that it does to infectious disease, since information does not necessarily follow a rigid cycle like that of an infectious disease [19].

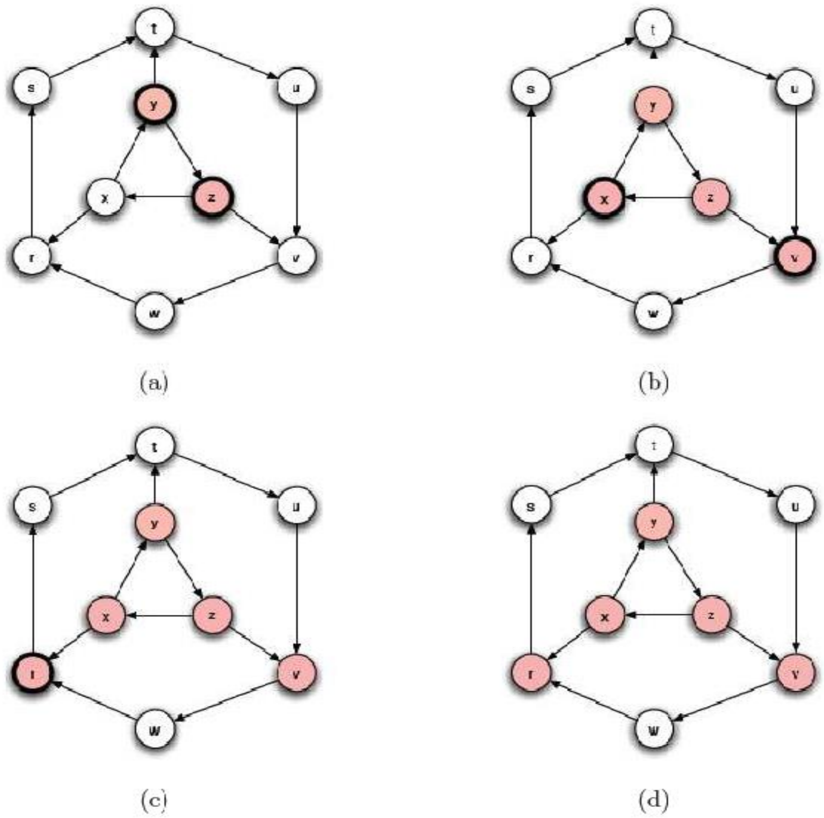


Figure (3.1): Illustration of the SIR Epidemic model[12]

3.4.2. Linear Threshold Model

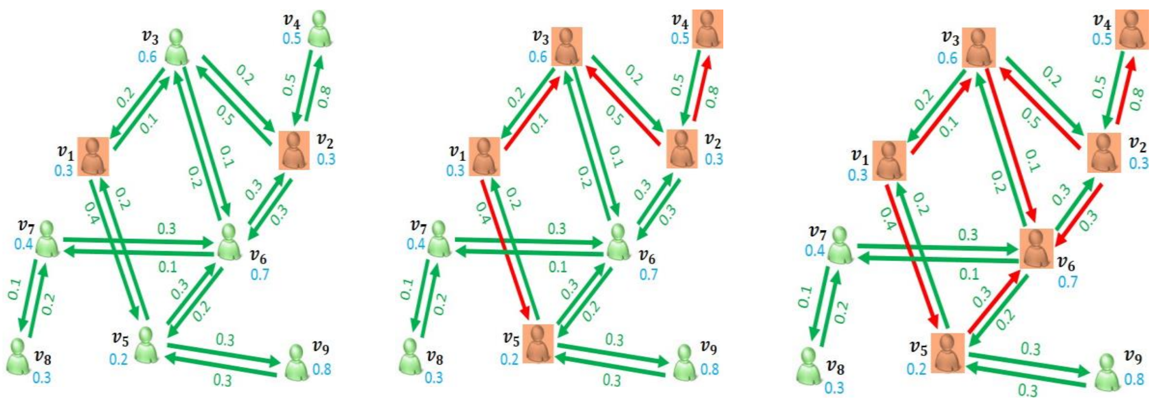
Social scientists have proposed threshold behaviors to model such diffusions (Granovetter, 1978, Schelling, 1978). When an aggregate function (e.g. count, or sum) of all of the positive signals received by a target exceeds a certain threshold, the target is activated. Centola and Macy (2007) refer to threshold behaviors in which an

[12] https://slideplayer.com/slide/3416033/

individual takes an action only after receiving influence from two or more sources as complex contagion [10].

The linear threshold model (LTM) is a stochastic diffusion model proposed by **Kempe et al**.(2003) to reflect this type of behavior. In the linear threshold model, every arc $(u,v) \in E$ is associated with an influence weight $w(u,v) \in [0,1]$ • (also denoted as w_{uv}), indicating the importance of u on influencing v. The weights are normalized such that for all v, the sum of weights of all incoming arcs of v is at most 1, i.e., $\sum_{u \in Nin(v)} W(u,v) \leq 1$ for all $v \in V$. For convenience, we set $w(u,v)=0$ for all $(u,v) \notin E$ [10] [17].

The linear threshold (LT) model takes the social graph G=(V,E) , the influence weights w(.) on all arcs, and the seed set S_0 as the input, and generates the active sets S_t for all $t \geq 1$ by the following randomized operation rule. Initially, each node $v \in V$ independently selects a threshold θ_v uniformly at random in the range [0,1] . At every time step $t \geq 1$, first set S_t to be $\mathbf{S_{t-1}}$; then for any inactive node $v \in V \setminus S_{t-1}$, if the total weight of the arcs from its active in-neighbors is at least θ_v, i.e., $\sum_{u \in S_{t-1} \cap N^{in}(v)} w(u,v) \geq \theta_v$, then add v into S_t (i.e., v is activated at time t) with the convention that $w(u,v)= 0$ for all $(u,v) \notin E$, the inequality can be equivalently written as $\sum_{u \in S_{t-1}} w(u,v) > \theta_v$ inequality in the main text provides more intuitive meaning, and we use these two type of expressions in the text interchangeably [10].



(a) $t = 0$ (b) $t = 1$ (c) $t = 2$

Figure 3.2: An example of the diffusion process of the linear threshold model.

Orange nodes in Figure (3.2) denote active nodes, and green nodes denote inactive nodes. Solid green arcs represent original arcs in the graph. Green numbers next to the arcs are the influence weights of the arcs. Blue numbers under the node labels are the randomly selected thresholds for the nodes. A set of red solid arcs pointing to the same node v means that the total weight of these arcs together causes the activation of v [10].

Next, sequence of steps for diffusion process of the linear threshold model (shown in figure 3.2)

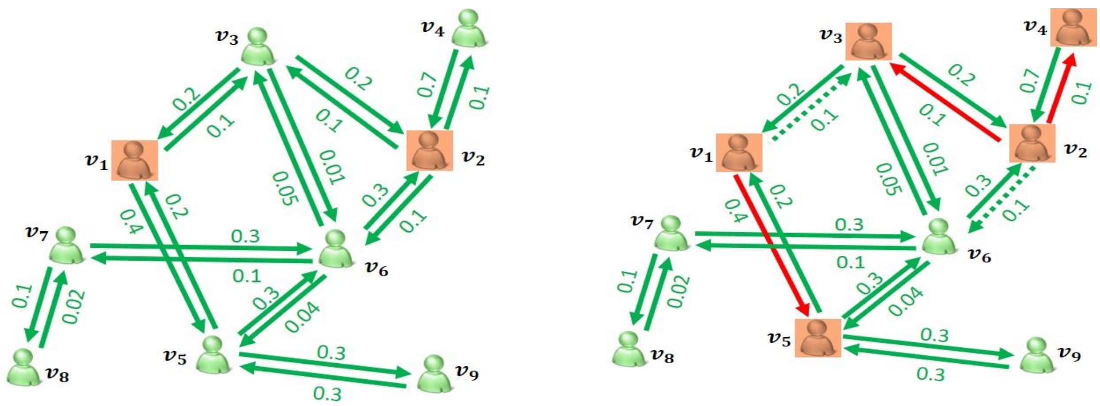
- Initially, the thresholds of all nodes are randomly chosen from [0,1] • (the blue numbers shown under the node labels), and nodes v_1 and v_2 are selected as seeds (Figure 3.2(a)).
- At step t = 1 (Figure 3.2(b)), nodes v_1 and v_2 jointly activate node v_3, because their total weight toward v_3 is 0.1 + 0.5 = 0.6, achieving the threshold of v_3. Node v_1 also activates v_5 and node v_2 activates v_4, because the corresponding arc weights exceed the thresholds of the target nodes. However, node v_6, although an out-neighbor of v_2, is not activated in this step because its threshold is 0.7 but the weight from its only active in-neighbor v_2 is 0.3.
- At step t = 2 (Figure 3.2(c)), three of v_6's in-neighbors, namely v_2, v_3, and v_5, are activated, and their total weight to v_6 is 0.7, achieving the threshold of v_3. Thus, v_6 is activated at this step. The diffusion stops at this point, since no other nodes can be activated by their active in-neighbors.

3.4.3. Independent Cascade Model :

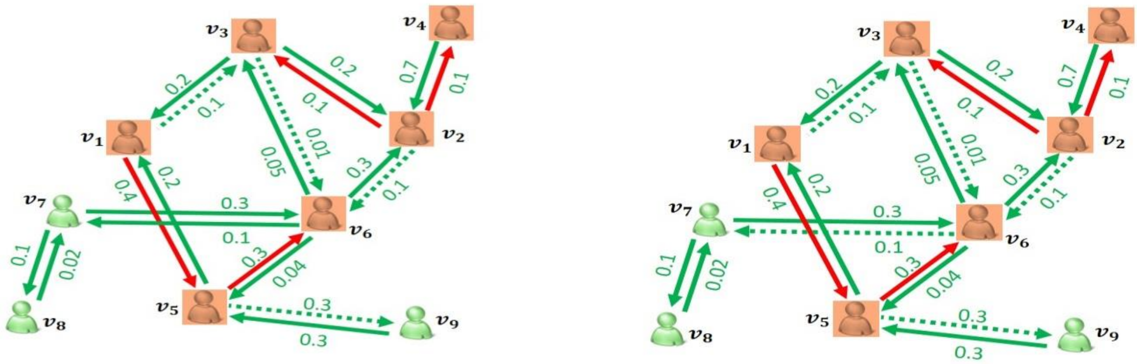
In the independent cascade model, every arc $(u, v) \in E$ has an associated *influence probability* $p(u,v) \in [0,1]$ (also denoted as $p_{u,v}$, corresponding to the extent to which node u influences node v. In the follow definition (and some later definitions), we will refer to a set S_{t-2} for t > 1, and thus as a convention we set $S_{-1} = \emptyset$. Also, for all $(u, v) \notin E$, we assume $p(u,v)$ =0. The technical definition of the model is given below [10].

The independent cascade model (ICM) takes the social graph G=(V,E) , the influence probability p(.) on all arcs, and the initial seed set S_0 as the input, and generates the active sets S_t for all $t \geq 1$ by the following randomized operation rule. At every time step $t \geq 1$, first set S_t to be S_{t-1} ; next for every inactive node $v \notin S_{t-1}$,for every node $u \in N^{in}(v) \cap (S_{t-1} \setminus S_{t-2}), u$ executes an activation attempt by performing a Bernoulli trial (flipping an independent coin) with success probability p(u,v); if successful we add v into S_t and say u activates v at time t. If multiple nodes activate v at time t, the end effect is the same — v is added to S_t [10].

Informally, after a node u is newly activated at time t-1, in the immediate next step t, u has a single chance of activating each of its inactive out-neighbors v with probability $p(u,v)$, and this activation is independent of any other activations. If u does not activate v at time t, it will not try to activate v in later steps. Once a node is activated, it stays active. If at some time t, no new nodes are activated, that is $S_t = S_{t-1}$, then the set of active nodes will no longer change, and the diffusion ends with the final active set S_t . We will see shortly that if we are only interested in the final active set, certain aspects of the model including single chance of activation and immediate activation are not essential [10].



(a) $t = 0$ (b) $t = 1$



(c) $t = 2$ (d) $t = 3$

Figure 3.3: An example of the diffusion process of the independent cascade model.

Orange nodes in Figure 3.3 denote active nodes, and green nodes denote inactive nodes. Solid green arcs represent original arcs in the graph. Green numbers next to the arcs are influence probabilities of arcs. A solid red arc from a node u to a node v means that u successfully activates v through this arc. A dotted green arc from a node u to a node v means that u fails to activate v through this arc.

Next, sequence of steps for diffusion process of independent cascade model (shown in figure 3.3)

- Initially at t = 0, two seed nodes v_1 and v_2 are activated (Figure 3.3(a)).
- At step t = 1, v_1 successfully activates v_5 but fails to activate v_3, while v_2 successfully activates v_3 and v_4 but fails to activate v_6 (Figure 3.3(b)).
- At step t = 2, v_3 fails to activate v_6, while v_5 successfully activates v_6 but fails to activate v_9 (Figure 3.3(c)).
- At step t = 3, v_6 fails to activate v_7 (Figure 3.3(d)). At this point, the diffusion stops and nodes v_1 to v_6 are active nodes while nodes v_7, v_8 and v_9 are inactive nodes.

3.4.4. Linear Influence Model

Another way of looking at spread through social networks is the Linear Influence Model (LIM), developed by Jaewon Yang and Jure Leskovec. This model differs from the others in the sense that it assumes that the number of newly infected nodes depends on which other nodes got infected in the past. [18]

It is not like the precedent models LTM and ICM that require knowledge of the social graph and is modeled without any need for explicit knowledge of the network.

Formulate the Linear Influence Model (LIM) by starting with the assumption that the number of newly infected nodes depends on which other nodes got infected in the past. [18]

In this model, each node u has an influence function $I_u(l)$ (The Influence of the node u at time t) associated with it , $A_u(t)$ represents a set of nodes that mentioned u before time t and V(t) is the number of nodes that mention the info at time t. So a formulation of LIM is given as follows [18]:

$$(t+1) = \sum I_u(t - t_u) \qquad u \in A(t)$$

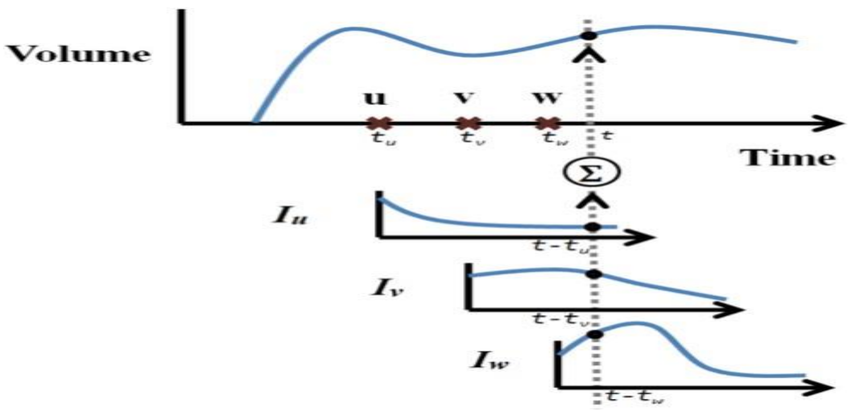


Figure 3.4: *"The LIM models the volume of diffusion over time as a sum of influences of nodes that got "infected" beforehand."* **[18]**

Summary

There is unlimited set of situations in which people are influenced by others. To find the reasons why this influence occurs, it is important to follow and understand the propagation of information through networks. In this chapter, we defined the influence in social network, and the meaning of information diffusion process. Then, we described the principals of some diffusion models that help to follow the propagation of information: the epidemic model, independent cascade model; linear threshold model and linear influence model. we also mentioned the reasons why analyzing information diffusion process is important in social media and we finished by giving some related work to information diffusion and influence measurement form the literature.

In the following chapter, we will explain with more details the models to be used in our experiments: the independent cascade model and the linear threshold model.

Chapter 4:

Implementation and Results Analysis

Introduction

Studying connectedness in social media networks can be done at two levels: the structure and behavior. At the first level, we study who is linked to whom, and at the behavior level we study the fact that each individual's action have implicit consequences for the outcomes of everyone in the network. It is frequent that we tend to do what most people around us do usually; especially those are directly connected to us. To be able to understand this collected behavior in networks, one needs to understand why influence happens and how it can propagate in networks.

In previous section, we introduced general ideas behind the influence and modeling information diffusion. In the following; we describe the diffusion process with more details, then we focus on modeling this process using Independent Cascade Model (ICM) and Linear Threshold Model (LTM) that we use in the experiments. Later, we describe the used tools and choices we made to achieve the principle objectives.

4.1. Information cascades

Many everyday situations involve sequential decision making, where one makes a decision based on some private information and the previous actions of others.

An Information cascade or informational cascade is a phenomenon described in behavioral economics and network theory in which a number of people make the same decision in a sequential fashion. It is similar to, but distinct from herd behavior.

An information cascade is generally accepted as a two-step process. For a cascade to begin an individual must encounter a scenario with a decision, typically a binary one. Second, outside factors can influence this decision (typically, through the observation of actions and their outcomes of other individuals in similar scenarios).

The two-step process of an informational cascade can be broken down into five basic components [33]:

- There is a decision to be made – for example; whether to adopt a new technology, wear a new style of clothing, eat in a new restaurant, or support a particular political position
- A limited action space exists (e.g. an adopt/reject decision)
- People make the decision sequentially, and each person can observe the choices made by those who acted earlier
- Each person has some information aside from their own that helps guide their decision
- A person can't directly observe the outside information that other people know, but he or she can make inferences about this information from what they do

Social perspectives of cascades, which suggest that agents may act irrationally (e.g., against what they think is optimal) when social pressures are great, exist as complements to the concept of information cascades.[21] More often the problem is that the concept of an information cascade is confused with ideas that do not match the two key conditions of the process, such as social proof, information diffusion,[22] and social influence. Indeed, the term information cascade has even been used to refer to such processes [23].

The information cascades can be the best explanation for many situations of influence in social media. Giving the example of sharing a popular video in YouTube or split a new rumor in Twitter …etc.

The main issue with information cascades is that they can be wrong. The main cause of this is that information cascades can be based on very little information: the actions of a few initial actors can determine all subsequent actions. This also explains why information cascades are fragile.

4.2. Modeling Information Diffusion

Information propagation, but what is most interesting about social media, about Twitter, is that it creates connections; networks that can be studied to understand how people interact or how news and opinions get spread, they are extracted using such tools like NodeXL, Gephi.

The aim of this work is to analyze the information diffusion process and predict the influence (represented by the rate of infected nodes at the end of the diffusion process) of an initial set of nodes in the Twitter network, several models are used for modeling the dissemination process: the Epidemic Model, Linear Influence Model

(LIM), Independent Cascade Model (ICM) and Linear Threshold Model (LTM). To achieve the goals of this work, we chose two basic diffusion models: the Linear Threshold Model and Independent Cascade Model.

Given a network G = (V, E) where V is the set of vertices, and E the set of existing edges in the network. A vertex $v \in$ V is said to be active if the information has reached the vertex and was accepted by it. If the information didn't reach the vertex or the vertex rejected it, then the vertex is said to be inactive. Each inactive vertex tends to become active, and it can switch from inactive to active, but it does not switch in the other direction. There must be an initial set of vertices activated to start the diffusion process targeted for initial activation. They are called 'initial adopters' of the information [24]. The influence of this initial set of vertices is the expected number of active vertices in the end of the diffusion process. Consequently, the cascading process will appear as follows: given an initial set of active vertices; while time spreads out, more of an inactive vertex v's neighbors become active which may cause this vertex to become active at some point. Then v may in turn trigger other vertices to which it is connected to adopt the same decision or action [24].

4.2.1. Linear threshold model (LTM)

Granovetter and Schelling were among the first to propose the threshold approach to capture influence [25]. In linear threshold model, a weight $b_{u,v}$ is used to measure the tendency of a node u to be influenced by each neighbor v such that $\sum_{v \text{ neighbor of } u} b_{u,v} \leq 1$. Starting with the initial set of active nodes A0, the influence propagation resumes as follows: each node u is assigned a threshold θ_u randomly from the interval [0, 1]; the threshold represents the weight fraction of u's neighbors that must adopt the behavior (be active) in order for u to become active and adopt the same behavior. At time stamp t , all nodes that were active in time t+1 remain active, and any node u for which the total weight of its active neighbors is at least θ_u gets activated; where [25]:

$$\sum_{v \text{ active neighbor of } u} b_{u,v} \geq \theta_u \qquad (1)$$

The thresholds θ_u represent the tendency of nodes to adopt the new behavior when their neighbors do. Figures 4.1a and 4.1b show an example of the process involved in linear threshold model.

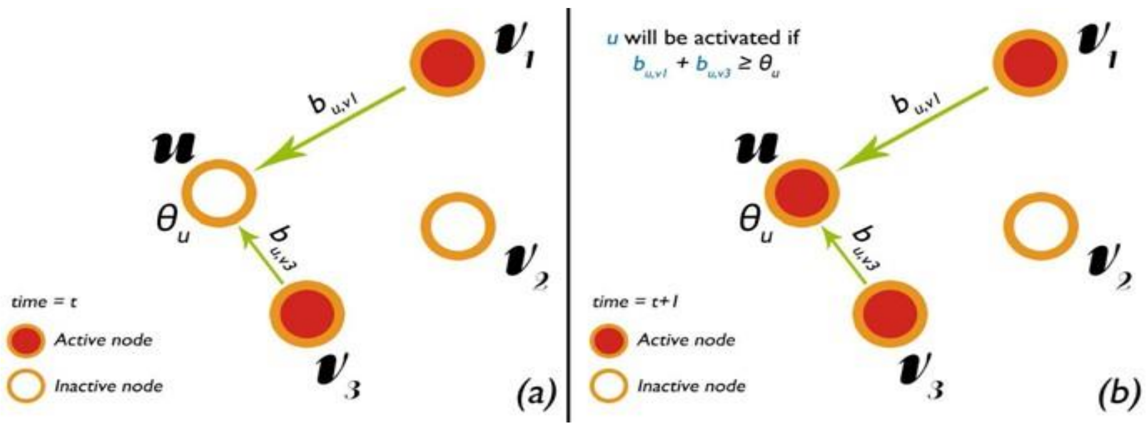


Figure 4.1. Linear Threshold Model process.

- At each time step, the active nodes can influence the inactive nodes
- Each node has an ==activation threshold==
- An inactive node becomes active if the sum of influence degrees ==exceeds== its threshold

Linear threshold Model Algorithm [26]:

1. $S_0 \neq \{\}$; *initial set of nodes selected for launching the diffusion.*
2. $i = 0$;
3. *thresholding* (V); *Attributing to each node* v *a threshold* θ_v *from the interval* **[0,1].**
4. While $i = 0$ *ou* $(S_{i-1} \neq A_i,\ i \geq 1)$ Do
5. $A_{i+1} = A_i$;
6. *inactif* $= V - A_i$; *The set of inactif nodes*
7. For $v \in$ *inactif* Do
8. If $\sum_{u \text{ an incoming neighbor of } v,\ \in A_i} P_{uv} \geq \theta_v$ Then
9. *Activate* (v);
10. $A_{i+1} = A_{i+1} \cup \{v\}$;
11. End if
12. End For
13. $i = i + 1$;

4.2.2. Independent Cascade Model (ICM)

An independent cascade model starts with an initial set of active nodes A0. This set of individuals should be chosen to generate the maximum influence during the cascade diffusion process. The process occurs in discrete steps as follows: when node u becomes active for the first time in time step t , its provided with one chance to activate each of its currently inactive neighbor v ; in that case u is called contagious, which means that it has the ability to affect other nodes as shown in Figure 4.2a. Node u succeeds to influence its neighbor v with a probability $p_{u,v}$ independent of past history. If u succeeds, then v will become active in time step t+1 as shown in Figure 4.2b; but whether or not u succeeds, it cannot make any further attempts to activate v in future rounds.18 The same process continues until us communicate with all neighbors for influence attempts and there are no more contagious nodes.[25]

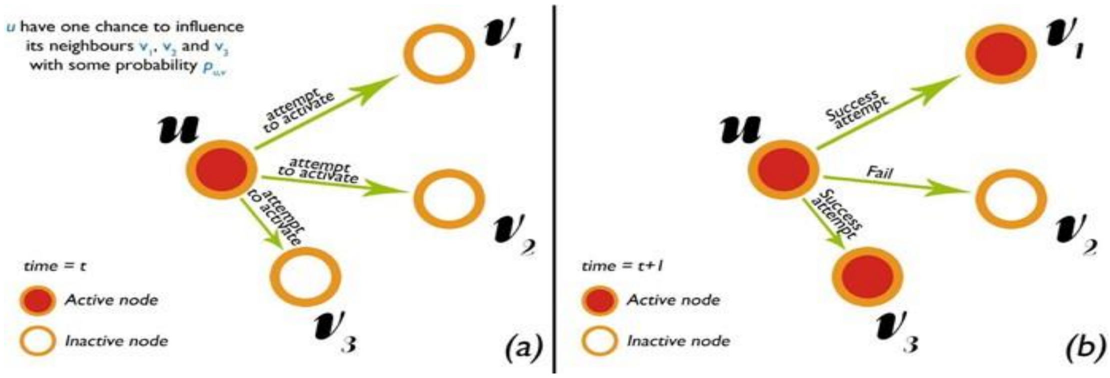


Figure 4.2. Independent Cascade Model Process.

- Nodes activated at time t, have a single chance, at time step t+1,to activate their neighbors
- Assume v is activated at time t, for any neighbor w of u, there is a probability $p_{u,v}$ that node w gets activated at time t+1

Independent cascade Model Algorithm [26]:

$S_0 \neq \{\}$;
$i = 0$;
While $S_i \neq \{\}$ **Do**
$i = i + 1$;
$A_i = \{\}$;
For $v \in A_{i-1}$ **Do**
For w a neighbor of v, $w \notin \cup_{j=0}^{i} A_j$ **Do**
r = random value $\in [0,1]$
If $r < P_v$, **Then**
Activate (w);
$A_i = A_i \cup \{w\}$;
End If
End For
End For
End While

4.3. How to choose initial active set A_0?

Another important step in the implementation of the chosen models is "choosing the initial active set" (the initial adopters). Many researchers focused on the problem of maximizing the influence or the spread of information in social networks by choosing the best initial adopters and the most influential ones [24]. Approximation algorithms and community-based greedy algorithms have been used for that purpose. This optimization problem is NP-hard.

The problem is posed as follows: "if you want to trigger a large cascade of adoptions of a new product or innovation, you need first to convince a subset of individuals to do so. In this case, which set of individuals (nodes) should you target initially? "

The initial active set is chosen based on 3 criteria: Betweenness centrality, Closeness centrality, and randomly. Next, every choice is described with details.

4.4. Finding one amazing Influencer

Though the objective sounds simple, however, when we started breaking it down further, we faced the following questions to answer [27]:

1. Should we find the person who could have fewer direct connections but its average outreach with all the other nodes on the network is the smallest? "K" below though doesn't have the highest direct connectivity but is on an average closest to all the other nodes in the universe. In graph theory, this is termed as **"Closeness Centrality"** [27]. Closeness Centrality (C_i) is measured by:

$$Ci = \frac{N}{\sum_{j=1}^{n} dij} \qquad (2)$$

Where, *d_{ij} = shortest distance between node I and node j, where I not equal to j* *N =total number of nodes*

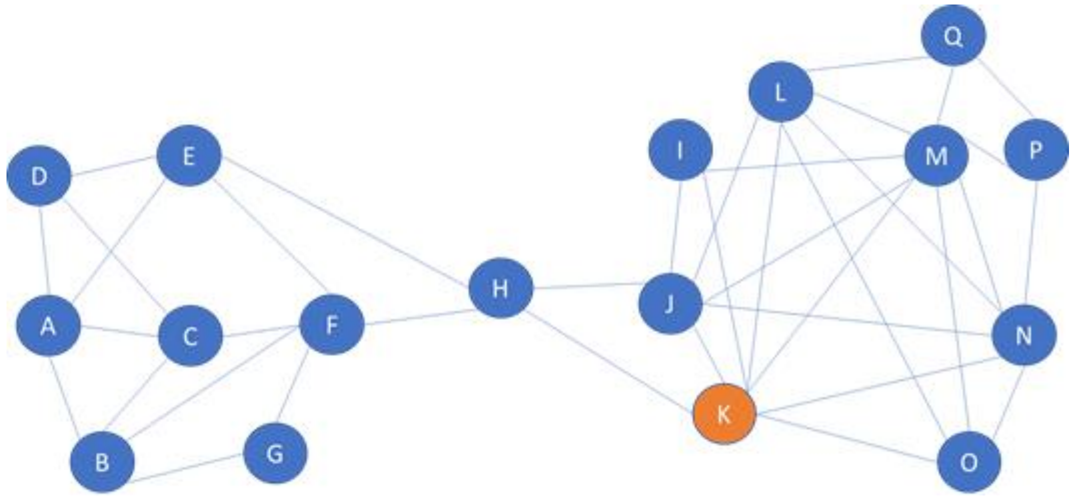


Figure 4.3 **Closeness Centrality**

2. Should we find the person who is in the center and act as a connection between various groups of nodes? “H” below is the center of the “universe”, which connects two big groups on either side. In graph theory, this is termed as **“ Betweenness Centrality ”** [27]**.** As one would expect most of the information would essentially pass through this node. Betweenness Centrality (b_i) can be defined as:

$$b_i = \frac{\#\boldsymbol{shortest\ path\ it\ crosses}}{\#\boldsymbol{total\ number\ of\ shortest\ path}}$$

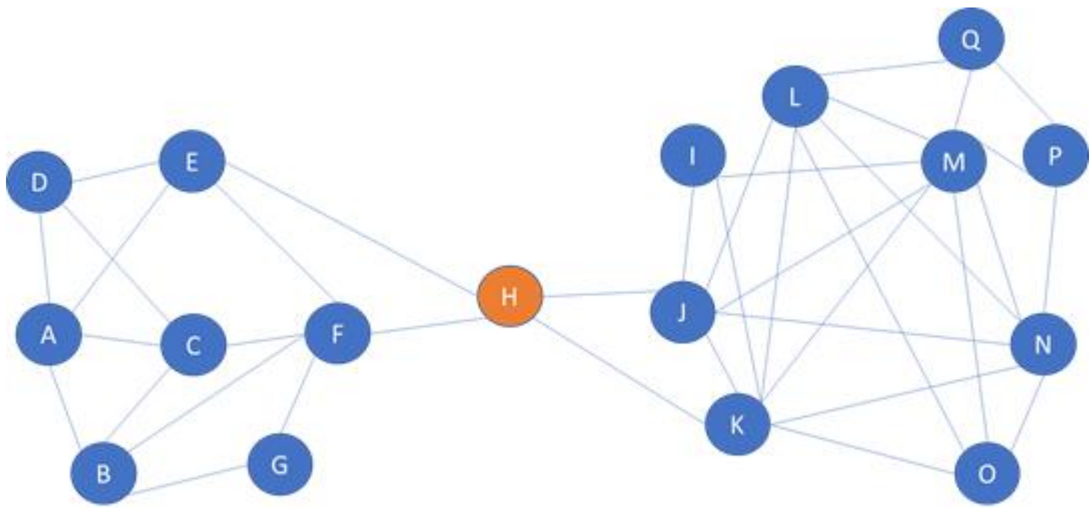


Figure 4.4 **Betweenness Centrality**

3. Random:

Finally, as a basic and intuitive choice, the same algorithms are tested by choosing the initial active set randomly with no specific condition. This choice will permit to understand and prove the importance of choosing the initial adopters for spreading the information (or influence) widely in a social media.

Regarding the number of initial active nodes, no general rules can be established this choice. From the literature, the size of the initial activity set is usually 30 to 50 nodes. It depends on the size, type and experimental field of the network. For example: in marketing; the cost of targeting a large set of initial adopters can be high because we must pay each initial adopter. On the other hand, such as spreading rumors on Twitter, without any cost, it is better to target a large number of initial adopters

4.5. Experiments tools

4.5.1. Gephi

Gephi is a tool for network analysis and visualization. It is developed by the Gephi Consortium which is a group of engineers and researchers in computer science. The tool is open source, freely available, and runs on Windows, Mac OS X, and Linux operation systems [28].

Gephi can be used to visualize networks with various layout algorithms and customizable nodes' or edges' properties such as colors, size, and labels. Gephi can be used to perform social network analysis (SNA) such as calculating the network diameter, shortest path, PageRank, modularity (for community detection), betweenness and closeness centralities, and clustering coefficient [28].

Gephi provides dynamic filtering and can be used to analyze dynamic networks (temporal graphs) to observe how a network evolves over time.

4.5.2. Python

Gephi doesn't contain any information diffusion model. For that reason we've used Python for programming the discussed models.

Python is a high-level programming language designed to be easy to read and simple to implement [29].

Features of Python programming language [29]:

a) **Readable:** Python is a very readable language.
b) **Easy to Learn:** Learning python is easy as this is an expressive and high level programming language, which means it is easy to understand the language and thus easy to learn.
c) **Cross platform:** Python is available and can run on various operating systems such as Mac, Windows, Linux, Unix etc. This makes it a cross platform and portable language.
d) **Open Source:** Python is a open source programming language.
e) **Large standard library:** Python comes with a large standard library that has some handy codes and functions which we can use while writing code in Python.
f) **Free:** Python is free to download and use. This means you can download it for free and use it in your application. See: Open Source Python License. Python is an example of a FLOSS (Free/Libre Open Source Software), which means you can freely distribute copies of this software, read its source code and modify it.

g) **Supports exception handling:** If you are new, you may wonder what is an exception? An exception is an event that can occur during program exception and can disrupt the normal flow of program. Python supports exception handling which means we can write less error prone code and can test various scenarios that can cause an exception later on.

h) **Advanced features:** Supports generators and list comprehensions. We will cover these features later.

i) **Automatic memory management:** Python supports automatic memory management which means the memory is cleared and freed automatically. You do not have to bother clearing the memory.

4.5.3. NetworkX

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks [30]

Software for complex networks:

- Data structures for graphs, digraphs, and multigraphs.
- Many standard graph algorithms.
- Network structure and analysis measures.
- Generators for classic graphs, random graphs, and synthetic networks.
- Nodes can be "anything" (e.g., text, images, and XML records).
- Edges can hold arbitrary data (e.g., weights, time-series).
- Open source 3-clause BSD license.
- Well tested with over 90% code coverage.
- Additional benefits from Python include fast prototyping, easy to teach, and multi-platform.

4.6. Datasets (twitter users):

4.6.1 Twitter

Twitter is an American microblogging and social networking service on which users post and interact with messages known as "tweets". Registered users can post, like, and retweet tweets, but unregistered users can only read them. Users access Twitter through its website interface or its mobile-device application software ("app"), though the service could also be accessed via SMS before April 2020 [31].

Twitter was created by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams in March 2006 and launched in July of that year [31].

Twitter is a fast-evolving, surprisingly powerful new way to exchange ideas and information, and to stay in touch with people, businesses, and organizations that you care about. It's a social network — a digital abstraction that represents who you know and who you're interested in (whether you know them personally or not) — that you can access from your computer or your mobile device anywhere that has an Internet connection.[30]

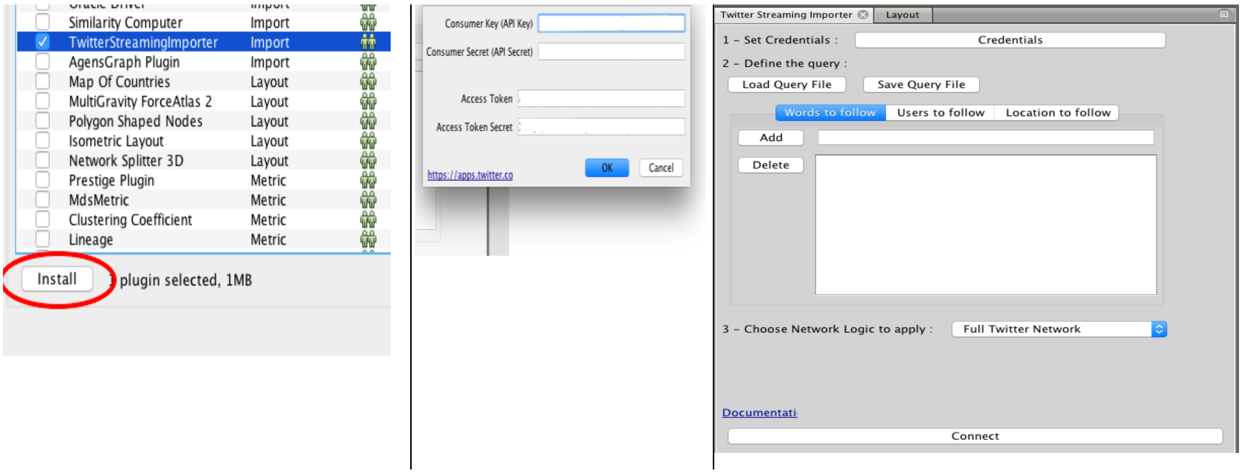
4.6.2 . Import data with Gephi:

Gephi has built-in "Plugins Center" which lists the available plugins from the Gephi plugin portal. This builtin plugins center extends the functionalities of Gephi, for example, the plugin "TwitterStreamingImporter" enables Gephi to collect data from Twitter Streaming API based on a keyword(s) or user screen name(s) then represent the collected data as a graph for further analysis. Three types of networks can be obtained using "TwitterStreamingImporter" plugin [28] :

- Full Twitter Network: a graph consisting of users, hashtags, tweets, media, URLs, and their connection.
- Twitter User Network: a network of users and the relations

between them.

- Twitter Hashtag Network: a network of co-occurring hashtags.



(1) (2) (3)

Figure 4.5. Importing Data with TwitterStreamingImporter

Then we need to enter our Twitter API credentials. If we don't have these already, we will need to go to https://apps.twitter.com/ and "create an app". After we create an app, go to "Keys and Tokens" to get your tokens.

Consumer_key = 'XXXXXXXXXXXXX'
consumer_secret = 'XXXXXXXXXXXXX'
access_token = 'XXXXXXXXXXXXX'
access_token_secret = 'XXXXXXXXXXXXX'

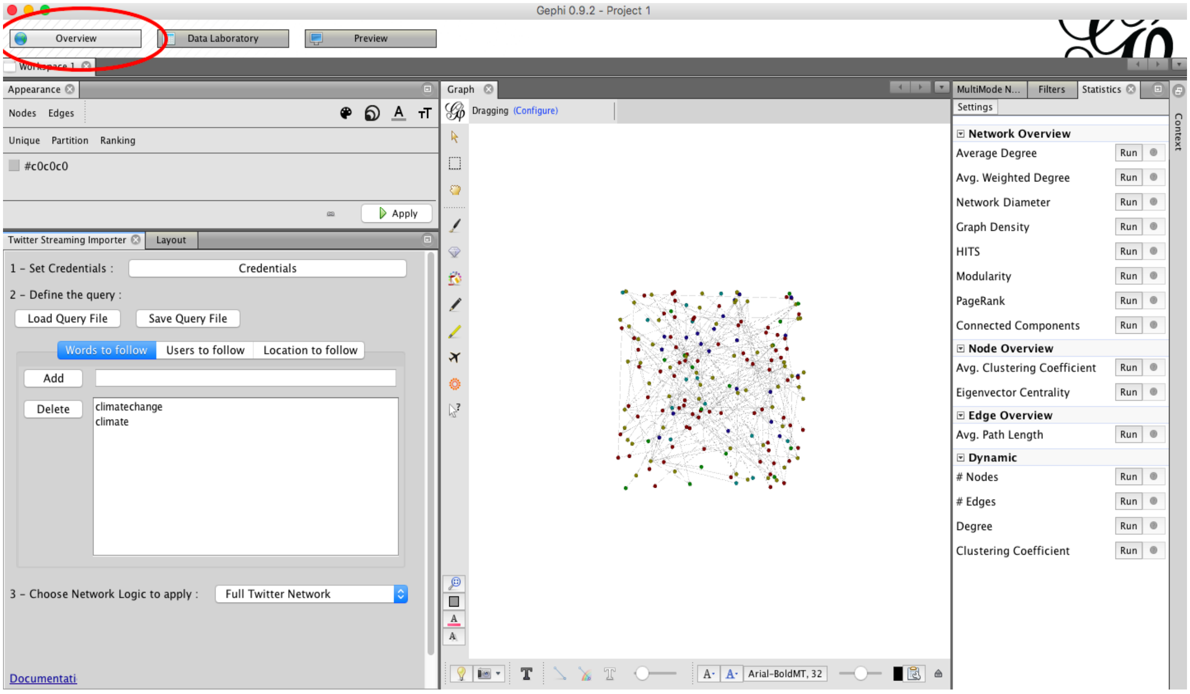


Figure 4.6: Gephi's Interface

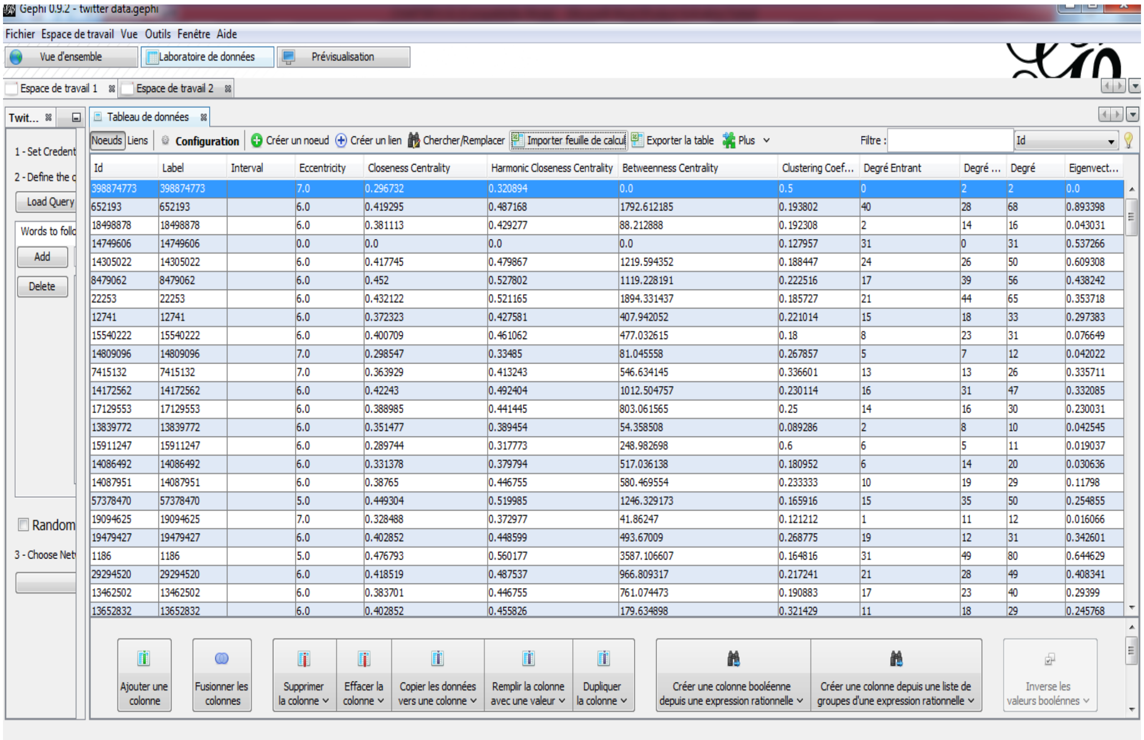
At the time of writing, the current version of Gephi (i.e., 0.9.2) can import data from CSV files, relational databases, and the majority of graph file formats such as DL …, and other types. The network graphs created with Gephi can be exported to .PDF, .PNG, and .SVG.

Once this is done running, we should have a csv with all of the edges of the network. We wrote this all to a csv just so that if it breaks while running I still have all the edges already scraped.

Now read the **csv** into a graph using **NetworkX** in **Python**.

Once the data has been converted to a graph, we can run some basic network analytics.

The network imported has 234 nodes and 2466 edges, graph is directed graph and has no self-loop and no duplicated edges



Id	Label	Interval	Eccentricity	Closeness Centrality	Harmonic Closeness Centrality	Betweenness Centrality	Clustering Coef...	Degré Entrant	Degré ...	Degré	Eigenvect...
398874773	398874773		7.0	0.296732	0.320894	0.0	0.5	0	2	2	0.0
652193	652193		6.0	0.419295	0.487168	1792.612185	0.193802	40	28	68	0.893398
18498878	18498878		6.0	0.381113	0.429277	88.212888	0.192308	2	14	16	0.043031
14749606	14749606		0.0	0.0	0.0	0.0	0.127957	31	0	31	0.537266
14305022	14305022		6.0	0.417745	0.479867	1219.594352	0.188447	24	26	50	0.609308
8479062	8479062		6.0	0.452	0.527802	1119.228191	0.222516	17	39	56	0.438242
22253	22253		6.0	0.432122	0.521165	1894.331437	0.185727	21	44	65	0.353718
12741	12741		6.0	0.372323	0.427581	407.942052	0.221014	15	18	33	0.297383
15540222	15540222		6.0	0.400709	0.461062	477.032615	0.18	8	23	31	0.076649
14809096	14809096		7.0	0.298547	0.33485	81.045558	0.267857	5	7	12	0.042022
7415132	7415132		7.0	0.363929	0.413243	546.634145	0.336601	13	13	26	0.335711
14172562	14172562		6.0	0.42243	0.492404	1012.504757	0.230114	16	31	47	0.332085
17129553	17129553		6.0	0.388985	0.441445	803.061565	0.25	14	16	30	0.230031
13839772	13839772		6.0	0.351477	0.389454	54.358508	0.089286	2	8	10	0.042545
15911247	15911247		6.0	0.289744	0.317773	248.982698	0.6	6	5	11	0.019037
14086492	14086492		6.0	0.331378	0.379794	517.036138	0.180952	6	14	20	0.030636
14087951	14087951		6.0	0.38765	0.446755	580.469554	0.233333	10	19	29	0.11798
57378470	57378470		5.0	0.449304	0.519985	1246.329173	0.165916	15	35	50	0.254855
19094625	19094625		7.0	0.328488	0.372977	41.86247	0.121212	1	11	12	0.016066
19479427	19479427		6.0	0.402852	0.448599	493.67009	0.268775	19	12	31	0.342601
1186	1186		5.0	0.476793	0.560177	3587.106607	0.164816	31	49	80	0.644629
29294520	29294520		6.0	0.418519	0.487537	966.809317	0.217241	21	28	49	0.408341
13462502	13462502		6.0	0.383701	0.446755	761.074473	0.190883	17	23	40	0.29399
13652832	13652832		6.0	0.402852	0.455826	179.634898	0.321429	11	18	29	0.245768

Figure 4.7: Nodes and Edges list with different metrics

The social network used in our case has structure, type, and size; we can resume his characteristics in Table (4.1) below:

	Twitter followers
Graph type	Directed
Egocentric network	Yes
Type of relationship	Explicit
Number of nodes (vertices)	234
Number of edges	2466
Graph density	0.045
Diameter	8
Average distance	2,957

Table 4.1: Network characteristics

The centrality metrics are summarized in table (4.2).

	Min	Max
In-degree	0	52
Out-degree	0	49
Closeness	0	0.75
Betweenness	0	4032.3974
Average degree	10.581	

Table 4.2: Graph Centrality Metrics

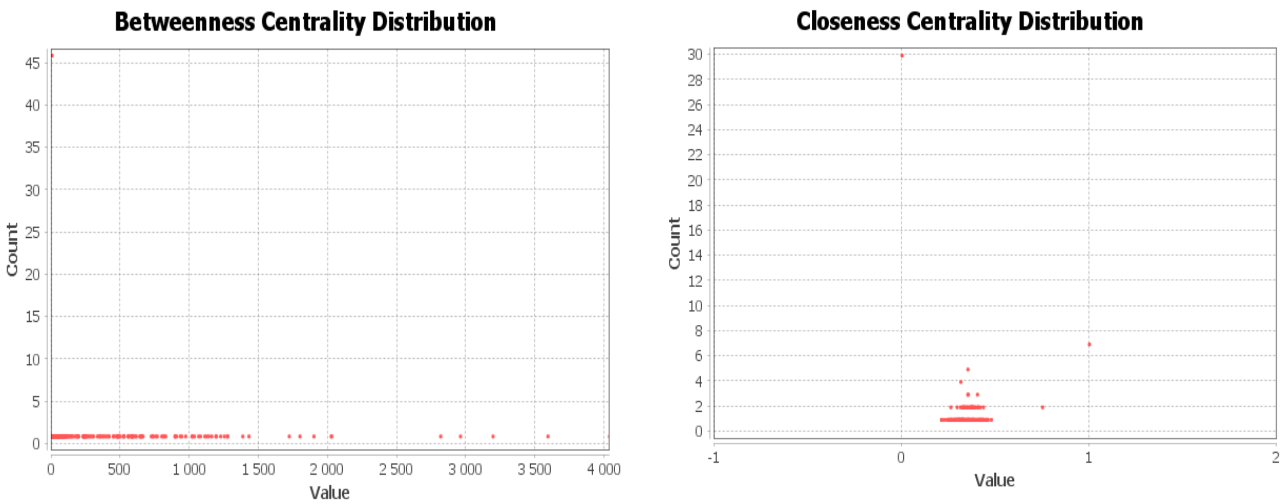


Figure 4.8: Network centrality metrics distribution plots

From centrality metrics distribution plots (Figure 4.8) we notice that a large portion of nodes have closeness centrality between 0 and 0.5; we can notice also that 80% of nodes have Betweenness centrality varies between 0 and 1500. For more insights and to provide more understandings about the used networks in our experiments, we visualize the graph with different metrics as parameters.

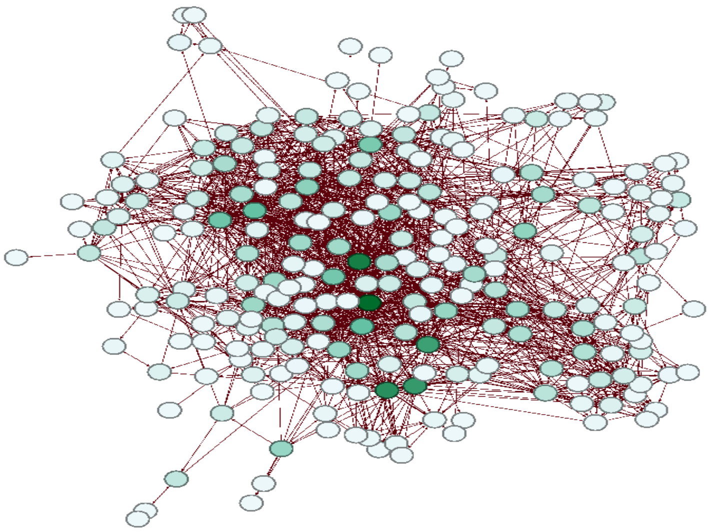


Figure 4.9: visualizing twitter network based on Betweenness centrality

From Figure 4.9; the Nodes with dark green color have high Betweenness centrality values and the clear green colors are low one

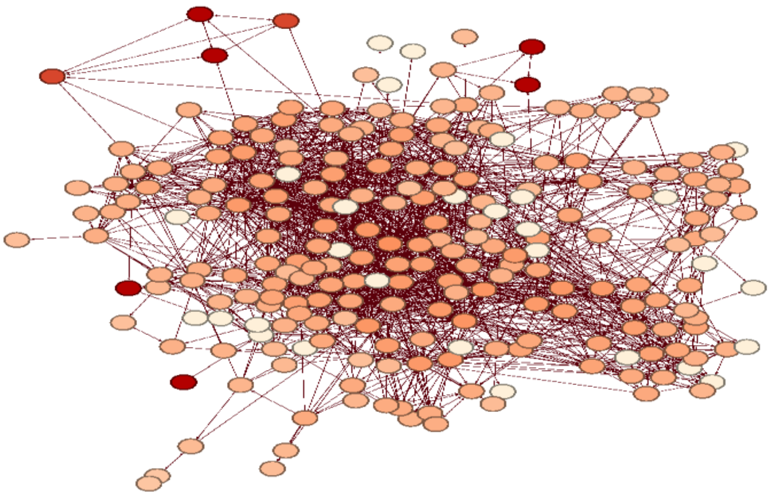


Figure 4.10: visualizing twitter network based on closeness centrality

Nodes (in figure 4.10) with dark brown color have high closeness centrality values and the clear brown colors are low one.

4.7. Simulation ICM /LTM models:

We have developed an application that simulate the execution of ICM and LTM model on our network with various metrics (Randomly, Betweenness and Closeness), by choosing the size of initial set0 and the appropriate metric the application plot the initial set0 in yellow and the activated nodes with green color .Figure 4.11 show application execution snapshot.

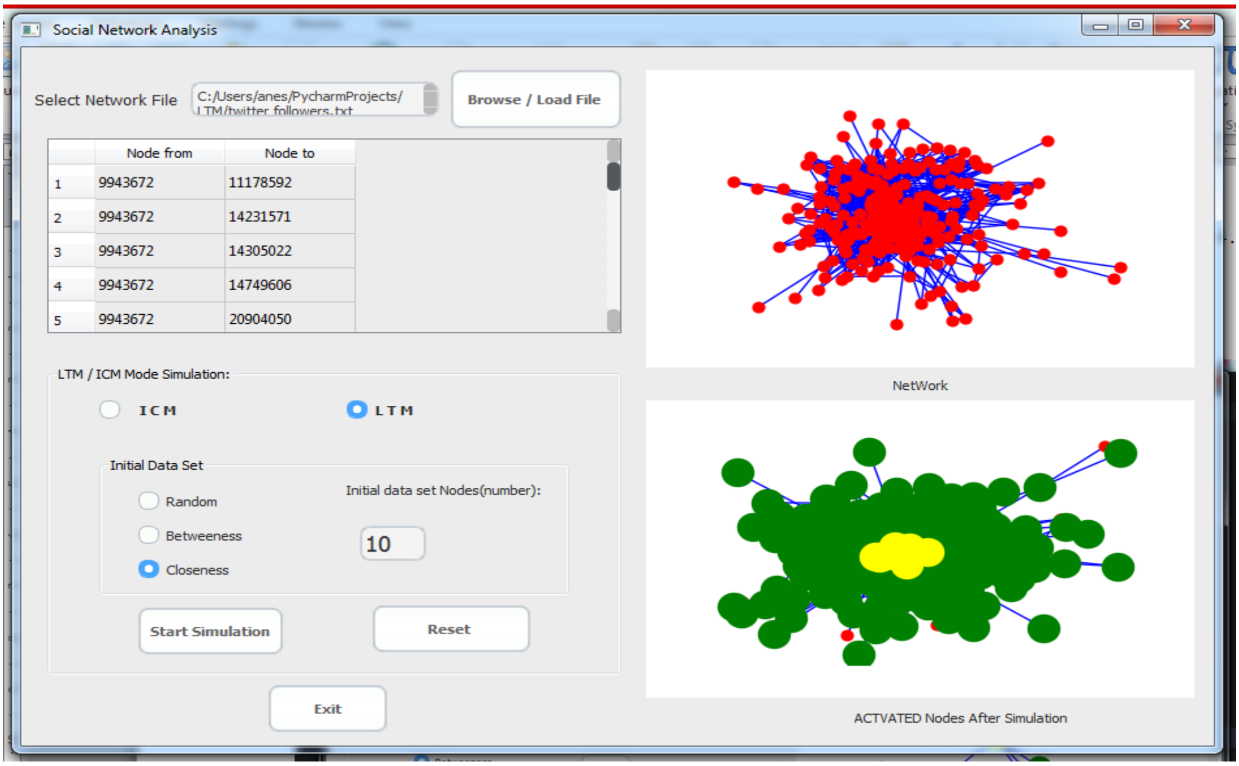


Figure 4.11: Application execution snapshot

4.8. Information diffusion results analysis

The selected diffusion models were implemented on the dataset (twitter followers) as explicit network. The results are presented in this section. For each algorithm (ICM/LTM), the propagation process was run 1000 times for every initial set $A0 \in \{5, 10, 15, 20\}$, then the average of the size active nodes after each run was computed. This average is considered as the influence of the initial set. The ratios from each model are summarized in table (4.3).

	Betweenness	Closeness	Randomly
LTM	0.89	0.90	0.09
ICM	0.80	0.76	0.15

Table (4.3): Results for twitter follower's network

Figure 4.11 illustrates snapshot of execution independent cascade model by choosing 10 initial active nodes (in yellow) based on the high Betweenness centrality.by the end of the process, the initial adopters have influenced about 80% of nodes (infected nodes are green)

At the first view on the table (4.3) , we can notice that for both models (ICM and LTM) the results converge to be near when the initial set is chosen based on Closeness or Betweenness centrality. We can also remark that the influence ratio of random initial adopters is very low compared to other cases (choosing random initial set is generally not a good choice).

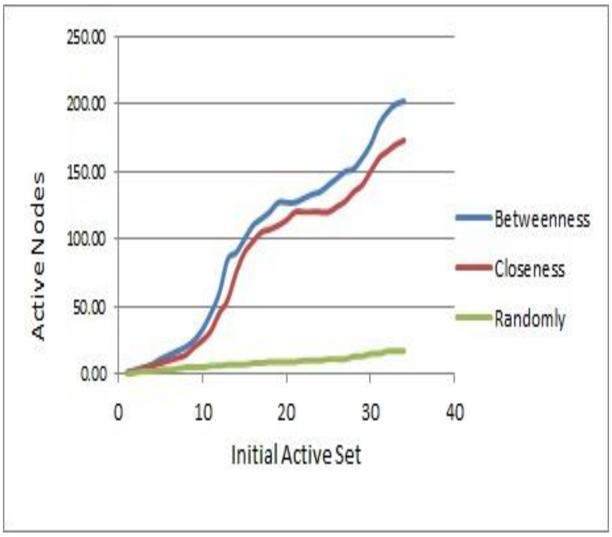


Figure 4.12: Results for ICM(twitter)

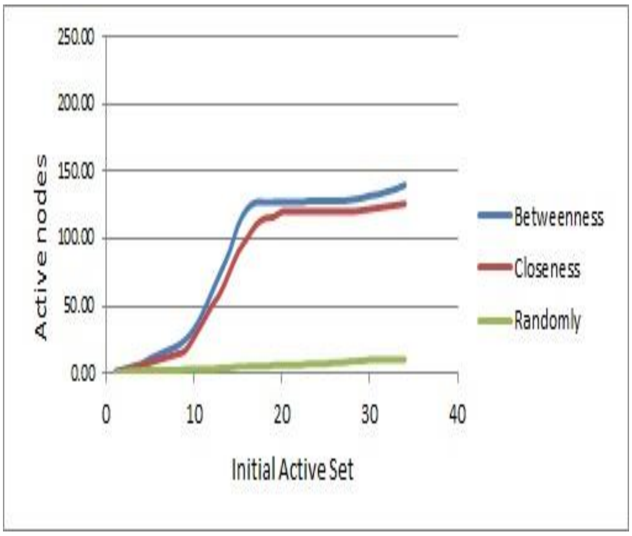


Figure 4.13:Results for LTM(twitter)

Figures 4.12 and 4.13 exhibit the performance of the algorithms in ICM and LTM models. The initial observation from these figures is that the curves become regular when the initial active set contains more than 10 nodes. The first 10 activated nodes (initial set with high Betweenness and closeness centrality) influence a large fraction of the network

Summary:

This chapter introduces the experimental results of the three main parts of our work. The first one describe different diffusion models used in our experiments, the various metrics are discussed also. The second is the simulation of diffusion models using the application. The last one is analyzing the results of the diffusion process.

General Conclusion

Due to the widespread availability of the Internet and the development of information and communication technology tools, human social life is closely related to the Internet and has become a supplement to social media, which makes the interaction between users easier, faster, and more reliable. Many behaviors and interactions occur through social media sites, which motivate organizations and researchers to develop new methods and tools to reason about users' behavior in order to predict their future behavior or experience new solutions. Social network analysis is a new research field that combines three main fields: network science, social science and computer science.

Analyzing the spread of information on social media and the influence of users on the Internet is one of the most important recent studies. Therefore, this article uses an information dissemination model to predict the number of infected nodes in social media networks and the ability of individuals to disseminate information through the network.

The prediction of information propagation on twitter users (followers) was done by implementing two models from the literature which are: the Linear Threshold Model (LTM), the Independent Cascade Model (ICM). However, the analysis and visualization of the structure Gephi , the implementation of the models was done with python and Networkx . To activate the diffusion process, an initial set that contains active nodes had to be chosen. Selecting these active nodes was based on three criteria Betweenness centrality, Closeness centrality and Randomly.

References:

[1] Kietzmann, Jan H.; Kristopher Hermkens (2011). Social media? Get serious! Understanding the functional building blocks of social media

[2] Obar, Jonathan A.; Wildman, Steve (2015). Social media definition and the governance challenge: An introduction to the special issue

[3] Investopedia, https://www.investopedia.com/terms/s/social-media.asp consulted at 20/04/2021

[4] John A. Barnes. Social Networks. Number 26 in Modules in Anthropology. Addison Wesley,1972

[5] Krishna Raj P. M., Ankith Mohan, K. G. Srinivasa, Practical Social Network Analysis with Python, Springer Edition

[6] Wikipedia, https://en.wikipedia.org/wiki/Graph_theory consulted at 03/05/2021

[7] Derek L. Hansen, B. Shneiderman, Itai Himelboim, Social Network Analysis Measuring, Mapping, and Modeling Collections of Connections,

[8] Derek L. Hansen, B. Shneiderman , Marc A Smith, Analyzing Social Media Networks with NodeXL, insights from a connected world,2010

[9] Alan E. Mislove, Online Social Networks: Measurement, Analysis, and Applications to Distributed Information Systems, thesis of doctorat, RICE UNIVERSITY

[10] Wei Chen, Laks V.S. Lakshmanan, and Carlos Castillo, Information and Influence Propagation in Social Networks, Morgan publishers (2014)

[11] Behnam Hajian, Tony White, Modelling Influence in a Social Network: Metrics and Evaluation SCS Technical Report: TR-11-09, Carleton University, Ottawa, Canada

[12] Wikipedia, https://en.wikipedia.org/wiki/Social_influence consulted at 02/05/2021

[13] *Deutsch, M. & Gerard, H. B. (1955).* "A study of normative and informational social influences upon individual judgment" Journal of Abnormal and Social Psychology

[14] David Nettleton, Commercial Data Mining processing, analysis and modeling for predictive analytics projects (2014)

[15] Xiaochen He, Haifeng Du, Marcus W. Feldman, Guangyu Li, Information diffusion in signed networks

[16] Guangyu Yin, Fan Jiang, Shaoyin Cheng, Xiang Li, Xing He, A Practical Trust Measurement for Adjacent Users in Social Networks, Proceedings of the 2012 Second International Conference on Cloud and Green ComputingNovember 2012 , Pages 360–367, Xiangtan, China

[17] David Kempe, Jon M. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In Proc. 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, pages 137–146, 2003

[18] Jaewon Yang, Jure Leskovec, Modeling Information Diffusion in Implicit Networks, Stanford University

[19] David Easley and Jon Kleinberg, Networks, Crowds, and Markets: Reasoning about a Highly Connected World, Cambridge University Press, 2010.

[20] Herbert C.Kelman, Compliance identification and internalization three processes of attitude change, Harcard University

[21] Schiller, R.J. (1995). "Conversation, Information and Herd Behavior". Rhetoric and Economic Behavior. 85 (3): 181–185

[22] Gruhl, Daniel; Guha, R.; Liben-Nowell, D.; Tomkins, A. (2004). Information diffusion through blogspace. WWW. pp. 491–501

[23] Sadikov, E.; Medina, M.; Leskovec, J.; Garcia-Molina, H. (2011). "Correcting for Missing Data in Information Cascades

[24] Samir Akrouf, Laifa Meriem, Belayadi Yahia, and Mouhoub Nasser Eddine, Social Network Analysis and Information Propagation: A Case Study Using Flickr and YouTube Networks, International Journal of Future Computer and Communication, Vol. 2, No. 3, June 2013

[25] Kanna AlFalahi, Yacine Atif, Ajith Abraham, Models of Influence in Online Social Networks, international journal of intelligent systems, VOL. 00, 1–23 (2013)

[26] H. Loucif, A. Boubetra, and S. Akrouf, S, “A Simplistic Model for Identifying Prominent Web Users in Directed Multiplex Social Networks: A CaseStudy using Twitter Networks,” New Review of Hypermedia and Multimedia Journal. Vol 22, Pages 287-302, 2016

[27] blog, https://prasunbiswas90.medium.com/mining-the-influencers-using-graph-neural-networks-gnn-cbd4d7965c2e consulted at 04/06/2021

[28] Samer Al-khateeb, Nitin Agarwal, Deviance in Social Media and Social Cyber Forensics, Uncovering Hidden Relations Using Open Source Information (OSINF), SpringerBriefs in Cybersecurity

[29] Python, https://www.python.org/doc/essays/blurb/ consulted at 01/06/2021

[30] Networkx , www.networkx.org, consulted at 20/05/2021

[31] Wikipedia, https://en.wikipedia.org/wiki/Twitter, consulted at 01/06/2021

[32] Laura Fitton, Anum Hussain, and Brittany Leaning, twitter for dummies, a Wiley Brand 3rd edition

[33] Wikipedia, https://en.wikipedia.org/wiki/Information_cascade consulted at 02/06/2021

[34] Ajith Abraham , Aboul-Ella Hassanien , Václav Snásel , “Computational Social Network Analysis” page 07 , https://www.springer.com/series/4198

[35] A dissertation : Hamza loucif , ,doctorat, University of Mohamed El Bachir El Ibrahimi -Bordj-Bou-Arreridj-faculty Mathematics and Computer Science Department of Computer Science ,2016/2017.

[36] A dissertation : BOUKHALFA Radhwane Influence Analysis In Social Nets , Master, University of M'sila faculty Mathematics and Computer Science Department of Computer Science ,2018/2019.

[37] Andre, M., Ijaz, K., Tillinghast, J. D., Krebs, V. E., Diem, L. A., Metchock, B., … McElroy, P. D. (2007). Transmission Network Analysis to Complement Routine Tuberculosis Contact Investigations. American Journal Journal of Public Health,97(3), 470–477. doi:10.2105/ajph.2005.071936 https://sci-hub.se/10.2105/AJPH.2005.071936

[38] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672174/

[39] A.L. Barabasi et al. / Physica , Evolution of the social network of scientific collaborations , Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA, 2002.

[40] https://dl.acm.org/doi/abs/10.5555/1103329.1710514

Abstract

Social influence is the science of influence, persuasion, and compliance. It is the process of influencing the behavior of one person by another, it is the process by which individuals adapt their opinion, revise their beliefs, or change their behavior as a result of social interactions with other people.

Nowadays, researchers in the Social Network Analysis (SNA) community are trying to figure out the factors that impact positively or negatively on the propagation of influence among the elements of a social graph. Our mission is to understand the diffusion process of influence and propose how it can be affected by the structural and behavioural aspects of social networks. Moreover, we will try to compare the proposed model with other models using real-world databases.

The results revealed the effectiveness of the LTM and ICM models in analyzing the diffusion process in social networks.

Key Word: Social Network Analysis (SNA), information diffusion, influence

الملخص

التأثير الاجتماعي هو علم التأثير والإقناع والامتثال. إنها عملية التأثير على سلوك شخص من قبل شخص آخر ، إنها العملية التي من خلالها يقوم الأفراد بتكييف رأيهم أو مراجعة معتقداتهم أو تغيير سلوكهم نتيجة للتفاعلات الاجتماعية مع أشخاص آخرين

في الوقت الحاضر ، يحاول الباحثون في مجتمع تحليل الشبكة الاجتماعية (SNA) معرفة العوامل التي تؤثر إيجابًا أو سلبًا على انتشار التأثير بين عناصر الرسم البياني الاجتماعي. مهمتنا هي فهم عملية انتشار التأثير واقتراح كيف يمكن أن تتأثر بالجوانب الهيكلية والسلوكية للشبكات الاجتماعية. علاوة على ذلك ، سنحاول مقارنة النموذج المقترح بالنماذج الأخرى باستخدام قواعد بيانات العالم الحقيقي.

كشفت النتائج عن فعالية نماذج LTM و ICM في تحليل عملية الانتشار في الشبكات الاجتماعية

الكلمات الاساسية : تحليل الشبكة الاجتماعية (SNA) ، نشر المعلومات ، التأثير

Résumé

L'influence sociale est la science de l'influence, de la persuasion et de la conformité. C'est le processus d'influencer le comportement d'une personne par une autre, c'est le processus par lequel les individus adaptent leurs opinions, révisent leurs croyances ou changent leur comportement à la suite d'interactions sociales avec d'autres personnes.

De nos jours, les chercheurs de la communauté Social Network Analysis (SNA) tentent de comprendre les facteurs qui ont un impact positif ou négatif sur la propagation de l'influence parmi les éléments d'un graphe social. Notre mission est de comprendre le processus de diffusion de l'influence et de proposer comment il peut être affecté par les aspects structurels et comportementaux des réseaux sociaux. De plus, nous essaierons de comparer le modèle proposé avec d'autres modèles en utilisant des bases de données du monde réel.

Les résultats ont révélé l'efficacité des modèles LTM et ICM dans l'analyse du processus de diffusion dans les réseaux sociaux

Mot clé: Analyse de réseau social (SNA), diffusion de l'information, influence