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°:……………………………………

DOMAIN: Mathematics and Computer Science
BRANCH: Computer Science
OPTION: SIGL

Dissertation submitted to obtain Master degree

By: MR. BOUKHALFA Radhwane

SUBJECT

Influence Analysis In Social Nets

Supported before the jury composed of:

………………………………….	University of M'sila	President
Mr. LOUCIF Hamza	University of M'sila	Supervisor
………………………………….	University of M'sila	Examiner
………………………………….	University of M'sila	Examiner

Academic year: 2018/2019

LIST OF FIGURES

LIST OF TABLES

General Introduction

In recent years, online social networks have become increasingly popular. Which defined as how users interact with each other. These interactions include: creating content (i.e., posting) by an individual and sharing it with its friends, re-sharing of the contents shared by an individuals' friends, submitting comments or liking friends' postings.

This dissertation focuses on the models of information propagation for anticipating number of infected nodes in a network, where we've used two modes Independent Cascade Model (ICM) & The Linear Threshold Model (LTM), we also aim to analyze the diffusion process on the same structure networks to understand how it can impact the flow of information and the influence of individuals in a social media. This analysis will be done on implicit network from page Facebook

This thesis is organized around four chapters organized as follows

Chapter 1

We present social networks in a global way

Chapter 2

Introduces some important definitions in social network analysis domain starting by graph theory basics for understanding network's structure. It also carries the network representations and visualization. Networks types and SNA major metrics are too

Chapter 3

Consecrated about information diffusion and influence definitions.

Chapter 4

Devote about implementation & explanation the results of the experiments, it carries information diffusion modeling, the tools, the used datasets and we also talked about analyze the structure and metrics of the datasets using Gephi

Finally, a **General Conclusion** recapitulates all the chapters besides the experiments steps and results.

CHAPTER 01

Social media Networks

Chapter 1: Social media Networks

1.1 Introduction

With the increasing growth in population, new communication methods were required to cross the geographical and temporal boundaries.

With the web and hard evolution, communications took a massive leap from being a difficult task to an instant, easy and accessible by the majority of people and therefore social networks were found as we know them nowadays. Thanks to that, the human interactions experience has been enhanced dramatically and new communicating aspect were found.

1.2 Definition Social media

Social Media is a technology that facilitates the sharing of information, thoughts, and ideas through the various networks between people. Social media is a mode of communication which is intended to create and strengthen relationships between people.

In present time most of the people on social media platforms. In social media platform we include: [1]

- **Facebook**
- **Instagram**
- **Twitter**
- **LinkedIn**
- **Pinterest**

1.3 History

Launched in 1995, classmates.com is the first online social network that allows Internet users to reconnect with former classmates. Each person can invite a former colleague to join him in his network.

Followed in 1997 by sixdegree.com then by Friendster.com in 2002 whose goal was to allow the members of this site to keep in touch with their friends and make new acquaintances.

In reality, the concept of social network was actually popularized by the arrival of the Myspace site in 2002, by the end of 2003, Friendster had 30 million members against a little over a million now. In which the user creates a personal space in his image and shares his passions with the people he has accepted as a "friend" for this site.

Mark Zuckerberg founded in 2004 Facebook which allows its users to enter personal information and interact with other users. Recent statistics[1] reveal that about 2,234 web users are monthly active in this fascinating platform.

The latest online social networks are more focused on specific topics such as the sport (eg the IPL Fan Club, a social network for cricket fans), or to professionals (Linkedln,Viadeo are examples) [2]

1.4 Characteristics of different types of social media [3]

Types	Characteristics
Online social networking	Online social networks are Web-based services that allow individuals and communities to connect with real-world friends and acquaintances online. Users interact with each other through status updates, comments, media sharing, messages, etc. (e.g., Facebook, Myspace, LinkedIn).
Blogging	A blog is a journal-like website for users, aka bloggers, to contribute textual and multimedia content, arranged in reverse chronological order. Blogs are generally maintained by an individual or by a community (e.g., Huffington Post, Business Insider, Engadget).

[1] https://www.blogdumoderateur.com/50-chiffres-medias-sociaux-2019/

Microblogging	Microblogs can be considered same a blogs but with limited content (e.g., Twitter, Tumblr, Plurk).
Wikis	A wiki is a collaborative editing environment that allow multiple users to develop Web pages (e.g., Wikipedia, Wikitravel, Wikihow).
Social news	Social news refers to the sharing and selection of news stories and articles by community of users (e.g., Digg, Slashdot, Reddit).
Media sharing	Media sharing is an umbrella term that refers to the sharing of variety of media on the Web including video, audio, and photo (e.g., YouTube, Flickr, UstreamTV).
Answers	These sites provide a platform for users seeking advice, guidance, or knowledge to ask questions. Other users from the community can answer these questions based on previous experiences, personal opinions, or relevent research. Answers are generally judged using ratings and comments (e.g., Yahoo! answers, WikiAnswers).

Table 1.1: Characteristics of different types of social media [3]

1.5 Function of social media

Commonly social media platforms are used for communication and sharing information but time has changed. In present time Social media platforms are being used not only for communication but also for lots of things like: [1]

- Use for Business growth
- Use for advertising and marketing of product and services
- Use for E-commerce
- Use for better customer relationship
- Use for link building

1.6 Summary

In the social life of the human being, social networks have always been present. They were waiting for a new way to grow. This way is internet and the new web technologies. Thus, it is possible to create networks of millions of people with anyone on the surface of the earth.

Behavior in a social network is defined as how users interact with each other. these interactions include: creating content (i.e., posting) by an individual and sharing it with its friends, re-sharing of the contents shared by an individuals' friends, submitting comments or liking friends' postings.

In the next chapter, we introduce some important definitions in social network analysis domain starting by graph theory basics for understanding network’s structure. It also carries the network representations and visualization. Networks types and SNA metrics are described in this chapter too.

CHAPTER 02

Social Networks Analysis (SNA)

Chapter 2: Social networks Analysis (SNA)

2.1 Introduction

In the precedent chapter, we have presented an introduction to social networks and its growth, and with that growth, data analysis is needed to obtain a better understanding about this data that circulates through those social networks.

In this chapter, we are going to briefly explore the concepts of graph theory to learn about social networks' structure. Then, we talk about network representation and visualization, and its different possible types. Afterwards, we cite the principal and most important social network analysis metrics. finally, we mention some reasons and the importance of studying the flow of information in social media networks. We want to mention that our aim through this chapter is to highlight the basics of SNA that help us to interpret social network structures based on graph theory concepts and social network analysis measures.

2.2 Definition Social network analysis

Mapping and measuring of relationships and flows between people, groups, organizations, computers or other information/knowledge processing entities, the nodes in the network are the people and groups, while the links show relationships or flows between the nodes [4]

Using network analysis, you can visualize complex sets of relationships as maps (i.e., graphs or sociograms) of connected symbols and calculate precise measures of the size, shape, and density of the network as a whole and the positions of each element within it.

Social network analysis helps to explore and visualize patterns found within collections of linked entities that include people. [5]

2.3 History

Social network analysis has its origins in social science, network science and graph theory fields. The beginning of graph theory has been led by the work of the mathematician Leonhard Euler by resolving the problem of Seven Bridges of Konigsberg in 1736. Years later, more work and results have been published by Polya (1935,1937), Paul Erdös and Alfréde Rényi in 1950's. Graph theory concepts and extensions have built social network analysis basics

In social science, the work of the first sociologist "Auguste Comte" in 1800's, and the work of the sociologist "Georg Simmel" in early 1900's, were the roots of social network analysis. They both defined a society as a group composed of relationships, where people can influence each other.

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

Modern social network analysis date from 1930's in the work of Jacob Moreno and his collaborators. They added the important algorithms and metrics of the modern social network analysis. In this period, the cost and lack of available network data sets and computing resources were the limiting factors against the widespread application of social network analysis especially in enterprises and organizations

In 1950's, Nadel wrote about social roles and the social structures that define them. Between 1960's and 1970's, a growing number of scholars worked to combine the different tracks and traditions of social network analysis, at the Harvard University department of social relations, university of California, university of Chicago, and Michigan State university, and many others .

The recent explosion of computer-mediated social relationships and the associated drop in the costs of creating network data sets have made network approaches increasingly practical. As more details about our interactions and associations are tracked and captured by mobile devices and social media services, network analysis becomes increasingly useful [9].

2.4 Social Network Analysis applications

Social network analysis has many strategic applications when people in an organization can analyze their position and the position of others. [5] Network analytic tools are used to represent the nodes (agents) and edges (relationships) in a network, and to analyze the network data. Like other software tools, data can be saved in external files. Additional information comparing the various data input formats used by network analysis software packages is available at NetWiki. Network analysis tools allow researchers to investigate large networks like the

Internet, disease transmission, etc. These tools provide mathematical functions that can be applied to the network model. [7]

Another significant application of SNA is measuring influence, diffusion and contagion. Contagions that flow through human-based networks can be bad (disease, gossip), good (ideas and information) or neutral (money and investments). [9]

2.5-Graph theory basic definitions

Graph theory has been useful in social network analysis for many reasons. Among these reasons are the following .First, graph theory provides a vocabulary which can be used to label and denote many social structural properties. This vocabulary also gives us a set of primitive concepts that allows us to refer quite precisely to these properties. Second, graph theory gives us mathematical operations and ideas with which many of these properties can be quantified and measured. Therefore, the following are some fundamental concepts and definitions surrounding graphs. [6]

2.5.1 A graph: is an ordered pair G = (V, E) a set V of vertices or nodes together with a set E of edges or lines [7] Figure (2.1) shows 2 simple graphs. We can represent a social network as a graph [8], so that vertices (also named nodes, agents, entities or actors) often represent people or social structures such as workgroups, teams, organizations, states or even countries. They can also represent content such as web pages (as we did in our work), keywords tags, or videos. They can even represent physical or virtual locations or events. The links that connect vertices (also called edges, ties, relationships or simply connections) are the building blocks of networks. They can represent many types of relationships like collaborations, kinship, and friendship [9]. Based on the orientation of the edges, we can define two types of graphs Directed and undirected.

2.5.2 Types of graphs

2.5.2.1 Undirected graph

A graph in which edges have no orientation, i.e., they are not ordered pairs, but sets {u, v} of vertices. Figure (2.1) (B) illustrates a undirected graph

2.5.2.2 Directed graph

A **directed graph** or **digraph** is an ordered pair $D = (V, A)$ with

- ✓ ***V*** a set whose elements are called **vertices** or **nodes**, and
- ✓ ***A*** a set of ordered pairs of vertices, called **arcs**, **directed edges**, or **arrows**.

An arc $a = (x, y)$ is considered to be directed **from** x **to** y; y is called the **head** and x is called the **tail** of the arc; y is said to be a **direct successor** of x, and x is said to be a **direct predecessor** of y.

If a path leads from *x* to *y*, then *y* is said to be a **successor** of *x* and **reachable** from *x*, and *x* is said to be a **predecessor** of *y*. The arc (*y*, *x*) is called the arc (*x*, *y*) **inverted** Figure (2.1) (A) illustrates a directed graph [7]

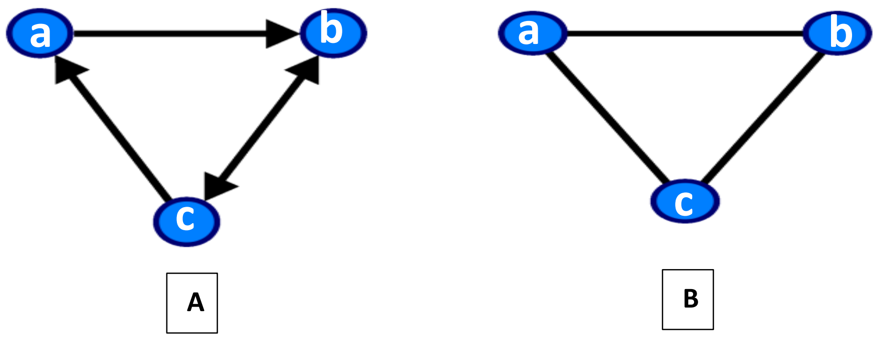


Figure (2.1): Directed and undirected graphs

2.5.3 A path: In graph theory, a **path** in a graph is a sequence of vertices such that from each of its vertices there is an edge to the next vertex in the sequence. A path may be infinite, but a finite path always has a first vertex, called its start vertex, and a last vertex, called its end vertex. Both of them are called end or terminal vertices of the path. The other vertices in the path are internal vertices. [7]

2.5.4 A cycle: is a path such that the start vertex and end vertex are the same. Note that the choice of the start vertex in a cycle is arbitrary. [7]

2.5.5 Diameter: The diameter diam(G) of a graph G is the maximum eccentricity over all nodes in a graph

2.5-6 Connectivity: a graph is connected if for every pair of nodes there is a path between them.

5-7 Bridge An edge is said to be a bridge if deleting it would cause its endpoints to lie in different components of a graph.

5-8 Structural hole in a network, we call a node a structural hole if it is connected to multiple local bridges. In figure (2.2), the node A is with her multiple local bridges

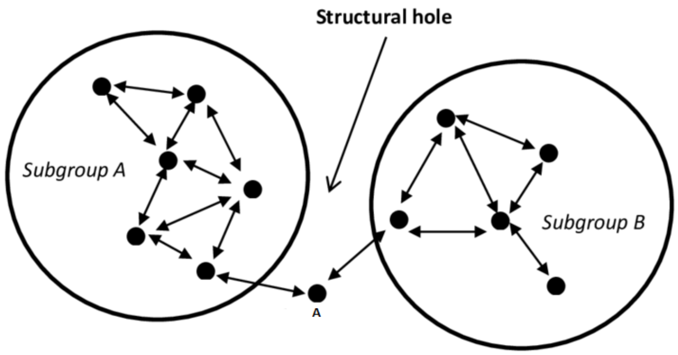


Figure (2.2): Bridges and structural hole

6 Representing and visualizing networks data

A: Representing

Graphs are very useful ways of presenting information about social networks, and it can be represented in many different ways. Representing the information in this ways also allows the application of mathematical and computer tools to summarize and find patterns

6-1 An adjacency matrix

It is a squared matrix in which $A_{ij} = 1$ if there is a link between i and j, and $A_{ij} = 0$ if there is no relationship between i and j. The adjacency is symmetric for undirected graph. In a case of a directed graph, rows represent the source nodes, and columns represent the end nodes of the edges. Table (2.1) shows the adjacency matrix of the graph (A) in figure (2.1). The value of the cell (A-A) is 0 which means there is no self-loop in the graph. Also, in the cell (A-B) the value is one, while in the cell (B-A) the value is 0; this means the link is from A to B, that proves that the graph is directed

	A	B	C
A	0	1	0
B	0	0	1
C	1	1	0

Table 2.1: Adjacency matrix

Column 1	Column 2
A	B
B	C
C	B
C	A

Table 2.2: Edge list representation

6-2 Edge list

One simple way to represent a graph is just a list, or array, of |V|. It is simply a list of all edges in the network as shown in the table (2.2) that represents the same graph (A) of the figure (2.1). Individuals in first column point to those in the second column. To describe the value of each edge or additional properties, more columns can be added. Self-loops are possible to be represented too in edge lists by repeating the same node in the both columns. Edge lists representation method is better than matrices in term of memory space, due to the fact that in the edge lists only existing relationships are represented.

6-3 Adjacency list

It is a collection of unordered lists used to represent a finite graph. Each list describes the set of neighbors of a vertex in the graph. This is one of several commonly used representations of graphs for use in computer programs. Easier to work with if network is large and parse, and quick in retrieving all neighbors for a node, in the table (2.3) that represents the Adjacency list of the graph in figure (2.3).

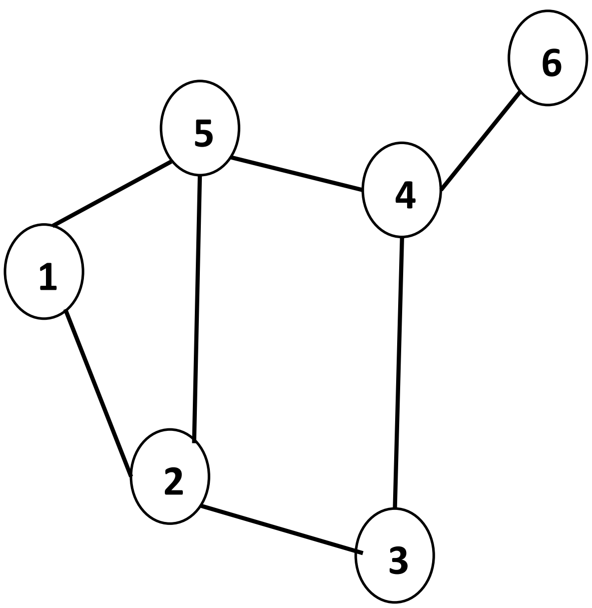


Figure (2.3): undirected graphs

node	neighbor
1:	2 5
2:	1 3 5
3:	2 4
4:	3 5 6
5:	1 2 4
6:	4

Table 2.3: Adjacency list

B: Visualization of networks

Network visualizations translate network data into a visual representation of some combination of the actors, relationships, clusters, and data attributes. Visualizing social networks is of immense help for social network researchers in understanding new ways to present and manage data and to effectively convert the data into meaningful information [10].

- ✓ Visualizations can help provide a fast and global understanding of the network
- ✓ Networks can be visualized in many different ways
- ✓ Diagrams of a network can be drawn by connecting nodes to each other using edges
- ✓ There are a variety of tools available for visualizing networks [11]

Recently, there are several different tools that facilitate quantitative or qualitative analysis of social networks, by describing the features of a network, either through numerical or visual representation. They generally consist of either packages based on graphical user interfaces (GUIs), or packages built for scripting/programming languages. GUI packages are easier to learn, while scripting tools are more powerful and extensible. Some are free and open sources, while others are commercial. Commonly used and well-documented tools are: Pajek, Gephi, NodeXL, UCINET/NetDraw, GUESS, NetMiner, InFlow, NetworkX library for Python and R statistical programming language package [9]

7 Types of networks

7.1 Egocentric networks

The egocentric network of vertex V in graph G is defined as the sub-graph of G induced by V an its neighbors It can be used to compute metrics over a local neighborhood, especially useful dealing with large networks As depicted in figure (2,7)(A) ,the egocentric network of 9 has nodes 3 ,6 and 8 .Similarly ,the ego net of 7 includes nod 5

7.2 Full networks

A full social network consists of all the network’s nodes and ties within a given boundary, such as the entire network of students attending a university or enrolled in a particular department or class. In principle, the whole population of the planet could be considered as a single social network, although it would be difficult to capture all these connections. [12]

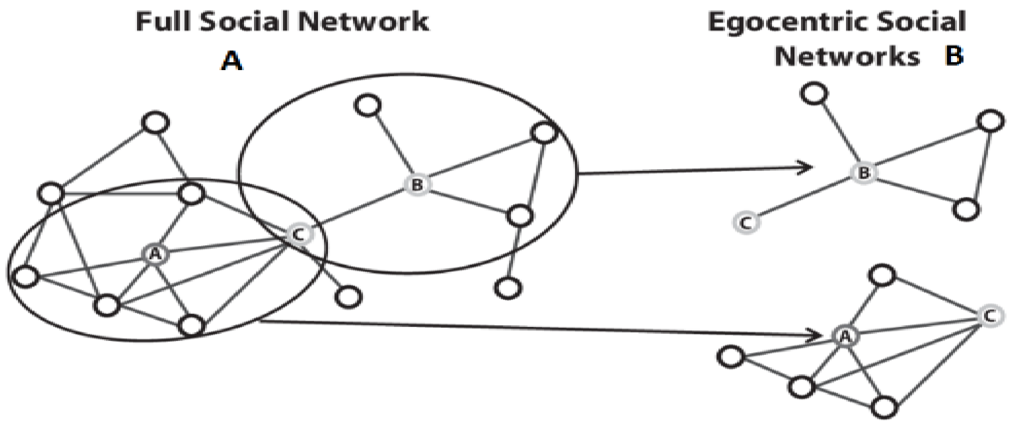


Figure (2.7): the egocentric and full social network

7.3 Unimodal networks

These are the standard networks that include one type (mode) of nodes. This means they connect a user to user, document to document, video to video, but not user to video, for example. Previous figures (2.7) (A and B) are Unimodal networks, where nodes are represented by the same color to describe that they are of the same type.

7.4 Multimodal networks

They are the networks that connect different type of nodes. For example, connecting a user to a blog, or a person to organization. In this type of networks, it is preferable to use various shapes and colors to separate different types of nodes. A common type of multimodal network is bimodal network which include exactly two types of nodes.

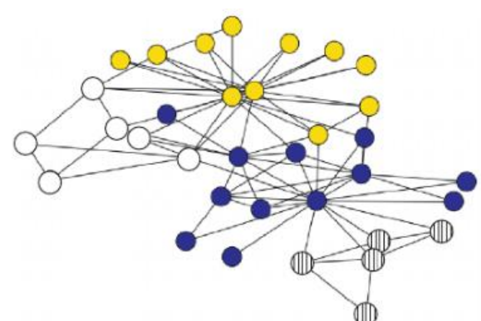


Figure (2.8): Multimodal networks

8 Social Network Analysis Metrics

8-1: Measures of centrality

Centrality is really a measure that tells us how influential or significant a node is within the overall network. This concept of significance will have different meanings depending on the type of network we are analyzing, so in some ways centrality indices are answers to the question "What characterizes an important node?"

The four most common centrality metrics are degree, Eigenvector, closeness and betweenness.

8.1.1Closeness centrality: "How near are you to everyone?"

The average of shortest distances to all other nodes in the graph. A node which has high closeness requires very little travel time to get to other nodes in the network, also estimates time to hear info, influence, and point of rapid diffusion

In the figure (2.4) node C has the highest closeness in the network, it has the smallest average distance to every node in the network. On the other hand, A and D are at the ends of the line, and so their average distance to the other nodes is at the maximum.

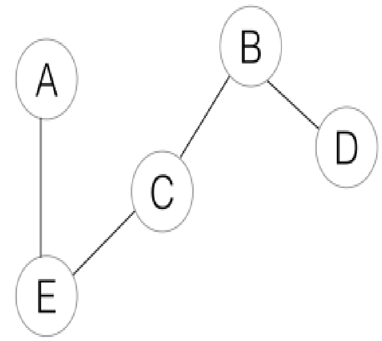


Figure (2.4): Closeness centrality of a node

8.1.2 Degree centrality: "How many friends do you have?"

The number of edges (links) connected to a node. It is a measure of how many relationships the node has. In the case of Facebook, degree is the number of friends that a user has.

In the figure (2.5) the node F is the node with the highest degree in the network.

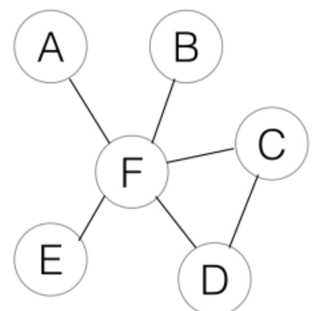


Figure (2.5): Degree centrality of a node

8.1.3 Betweenness centrality: "How often do you connect other people?"

Extent to which a particular node lies on the shortest path between other nodes. Betweenness centrality is a measure of how often a given vertex lies on the shortest path between two other vertices. This can be thought of as a kind of "bridge" score. [5]

In the figure (2.6) node C has the highest betweenness here, since the shortest paths from {A,E} to {B,D} need to go through C

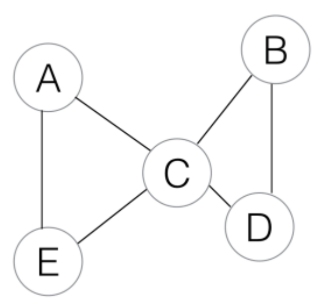


Figure (2.6): Betweenness centrality of a node

8.1.4 Eigenvector centrality

This is a measure of the importance of a node in a network. Contrary to the centrality of degree which gives a simple count of the number of connections of a vertex, a vector own centrality recognizes that not all connections are equal so that the connection to some vertices has more advantages than connecting to others. For example the private supervisor to the President. Other approaches focused on egocentric centrality, which determines the influence of a knot relative to its vicinity.

8-2: Network Density

Network density is a measure of the connectedness in a network. For valued data the sum of all tie strengths divided by all possible ties, number of ties, expressed as percentage of the number of ordered/unordered pairs

The problem with the measure of density is that it is sensitive to the number of network nodes; therefore, it cannot be used for comparisons across networks that vary significantly in size

8-3: Centralization

The difference between the numbers of links for each node divided by maximum possible sum of differences. A centralized network will have many of its links dispersed around one or a few nodes, while a decentralized network is one in which there is little variation between the numbers of links each node possesses

Summary

Social networks witnessed a gigantic increase during the past years. Analyzing's metrics and concepts gained popularity, collecting and analyzing data's obstacles were reduced by visualization tools.

Thanks to these tools, analyzing information helped experts like business managers and such to conclude more accurate and more efficient decisions in a shorter period of time.

CHAPTER 03

Information and Influence Diffusion Models

Chapter 3

Information and Influence Diffusion In Social Networks.

3.1: Introduction

There is no doubt that the main purpose behind inventing social networks is to facilitate the communication among web users. As a result, this digital society caused a wide presence of social influence on the attitudes and behaviors of the web users. People's opinions were altered because of social networks like what opinion they have about an artist, which products they buy or political views. This chapter talks about these influences and how important their study is.

3.2: Information diffusion

Diffusion is the process by which an innovation is communicated through certain channels over time among the members of a social system. It is a special type of communication concerned with the spread of messages that represent new ideas. Communication is a process in which participants create and share information with one another in order to reach a mutual understanding [13].

It also called "Diffusion of innovations" is a theory that seeks to explain how, why, and at what rate new ideas and technology spread. All the users of a social network are now potential producers and disseminators of information. Information diffusion can be simply defined as the process of spreading of information through the member of the social network. [9]

3.3: Influence in social networks

Sociologists defined social influence as a “change in an individual’s thoughts, feelings, attitudes, or behaviors that results from interaction with another individual or a group”. [23]

Social influence occurs when an individual changes his/her behavior after interacting with other individuals who tend to be similar

A social network plays a fundamental role as a medium for spread of information, ideas, and influence among its members. The actions of individuals can drive their friends to act in a similar way (ex: a user buys a product because of his/her friends has recently bought the same product)

The notion of influence is very present in its logic of tracking information given by others. With Twitter, we read the Tweets of people we follow the Retweeter to share them with your followers.

Accordingly, an **influencer** can be seen as that person who is followed by many people and has the power to make changes in a community

3.4: Diffusion models in social networks

Researchers study how information diffuses and explore different models of information diffusion, including the **L**inear **I**nfluence **M**odel, **I**ndependent **C**ascade **M**odel, **L**inear **T**hreshold **M**odel and **E**pidemic **M**odel. In these models, the network is represented by a directed graph. At the moment t = 0, a set of nodes initiators of the diffusion of a new idea are active. At a time t, if a given node adopts the new idea, it becomes active, otherwise it is remains inactive. We assume that a node can go inactive to active but not vice versa throughout the propagation process. An active node tries to activate these neighbors. The process continues until it has no more activations possible. [14]

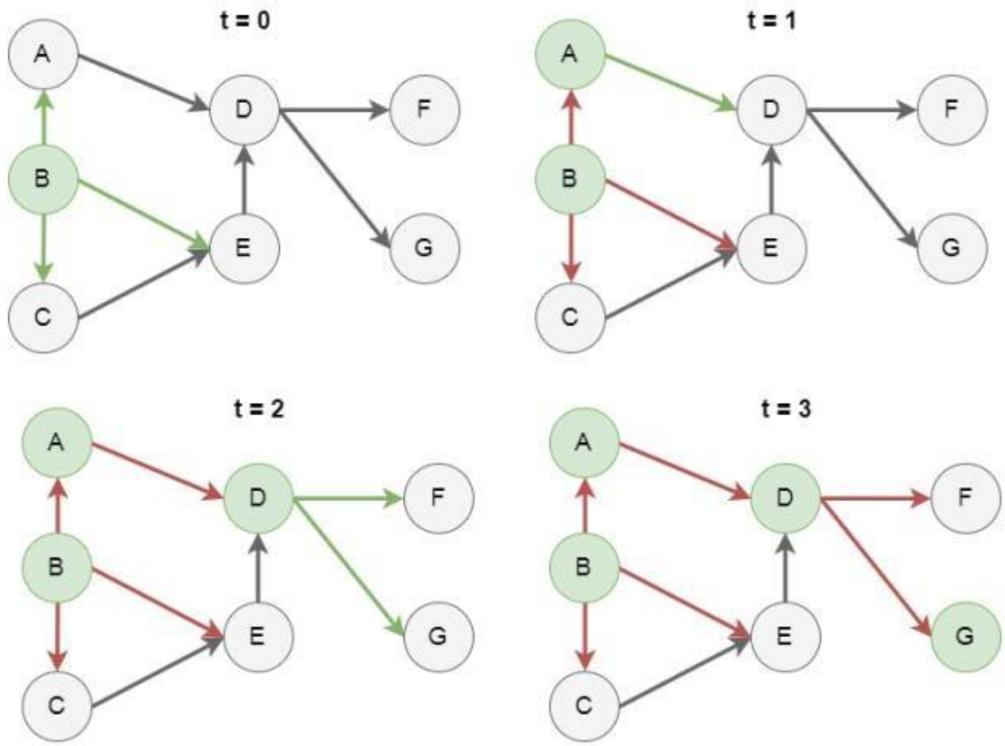
3.4.1: Independent Cascade Model (ICM) [15]

Independent Cascade Model is one of the most classical theoretical models of information diffusion where the information flows over the network through a cascade manner. Nodes can have two states, (i) Active: It means the node already influenced by the information in diffusion. (ii) Inactive: node unknowing of the information or not influenced

- When a node v becomes active, it is given a single chance to activate each currently inactive neighbor *w*
- The activation attempt succeeds with probability p_{vw}
- If v succeeds, then w will become active in step t+1; but whether v succeeds or not, it cannot make any other attempts to activate w in later tour.
- The process runs until no more activations are possible.

Example:

Figure 3.1 shows B is selected as the starting point node and activated at t = 0. Then it tries to activate its neighbors. A, C, and E have activation probabilities of P_{BA}, P_{BC}, and P_{BE} respectively. At t = 1, only A is activated by B; and B cannot activate any of its neighbors anymore. At t=2 A then returns to activate D in a similar way. At t=3 D attempts to activate its neighbors and activate G, the diffusion terminates because there doesn't remain any active node which can try to activate its neighbors.



3.4.2: The Linear Threshold Model (LTM) [16]

Is contains the same notion as the ICM, except that the single activation attempt rule is replaced by another activation mode. The diffusion process in discrete steps

- A node *v* is influenced by each neighbor *w* according to a *weight* $b_{v,w}$ such that

$$\sum_{v\ neighbor\ of\ w} b_{v,} \leq 1$$

- Each node has random threshold uniformly θ from interval [0,1].

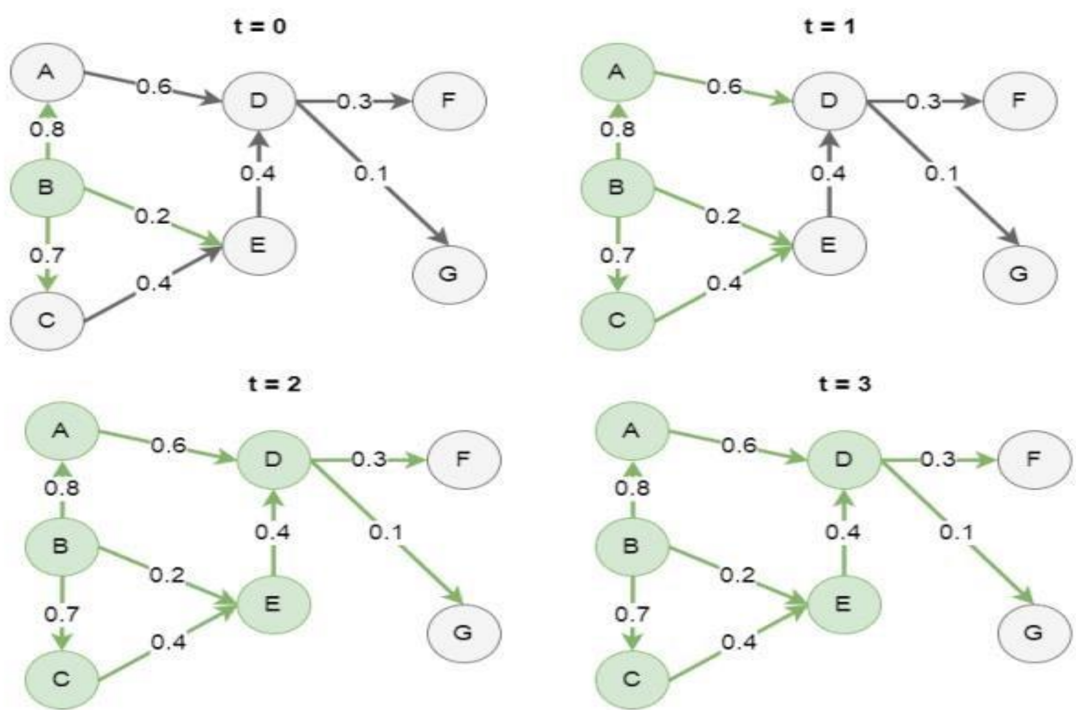
  Each node has a specific threshold θ to get activated.

- V will become active Depending on the following condition

$$\sum_{v\ active\ neighbors\ of\ w} b_{v,w} \geq \theta_v$$

Example:

Figure 3.2 shows all nodes have fixed thresholds of 0.5. At t = 0, B is selected as the starting point node and activated. At t = 1, A and C get activated because sum of incoming influence weights are bigger than their thresholds. In the next time step, D and E are activated in a similar way. The diffusion terminates after t = 3 because F and G cannot be activated.



3.4.3: The linear Influence Model (LIM) [17]

From this point is another way of find out about at spread through social networks, developed by Jaewon Yang and Jure Leskovec. 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.

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:

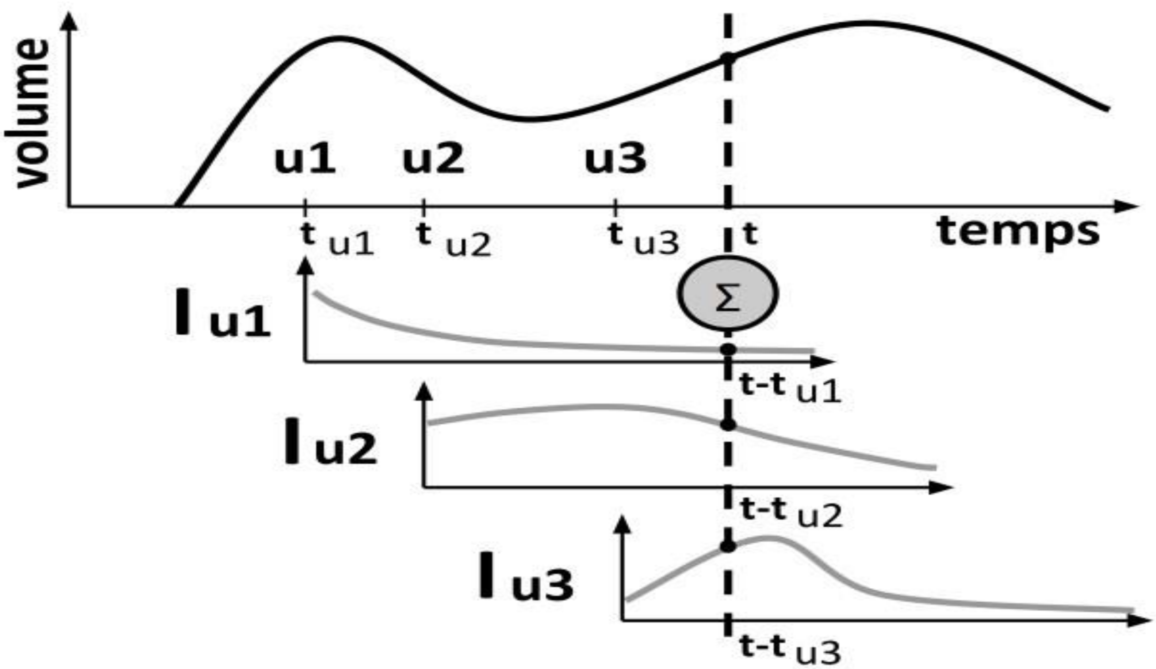
$$(t+1) = \sum_{u \in A(t)} I_u(t - t_u)$$

3.4.3.1: Modeling influence function

How to model ***$I_u(l)$***?

Influence function can be defined by an exponential law so $(l) = e^{-\lambda_u l}$ with a power law $I(l) = c_u l^{-\alpha_u}$, assuming all the nodes follow the same form

Figure 3.3: represents the volume V (t) over time, and t_u, t_v, and t_w denote the times when nodes, *u*, *v* and *w*, got infected. After the nodes got infected, they each influence additional $I_u(t - t_u)$, $I_v(t - t_v)$ and $I_w(t - t_w)$ infections at time t. So the volume V (t) at time t is the sum of the influences of the three nodes [17].



3.5: Influence Modeling

Social scientists have been exploring influence and homophily in social networks for some time. It is important to know whether the underlying social network is influence driven or homophily drive. For example, in the advertisement industry, if the social network is influence driven, then the influential users should be identified and incentivized to promote the product or services to the members of the social network. However, if the social network is homophily (similarity) driven, then some individual users should be directly targeted to promote sales. [3]

Social influence involves social correlations, which are divided into three categories as follows

3.5.1: Influence: where a user performs an action based on his friends' recent actions.

3.5.2: Homophily: a user chooses friends who share the same characteristics; this leads to perform the same actions.

3.5.3: Confounding factors: or external influence that affects individuals who are located near each other in the social network

3.6: Information Cascades

According to [18], "In an information cascade, people observe the choices of others and make their own decisions based on that observation while ignoring their personal knowledge or getting more information".

We find in information cascade three observations which are:

3.6.1: cascades can be wrong

That means do what you see them they do (*Following the Crowd*) for example suppose that you are went to a market and you found a big crowd around a new range of inexpensive clothes, but you bought it without thinking and when you got to your house, you didn't like it, which led to you didn't wear it. In this case, following the crowd was a wrong choice for you. [19]

3.6.2: cascades can be based on very little information

The people ignore their private information at the moment that a cascade starts, which means if you want to influence to the behavior of the population you must have precascade information about them, This means that if a cascade starts relatively quickly in a large population, most of the private information that is collectively available to the population is not being used. [19]

3.6.3: Cascades are fragile

Starting from "cascades can be based on little information" about makes them easy to start such as can easy to make them to stop

One manifestation of this is that people who receive a bit superior information, can overturn even enduring cascades**.** [19]

Summary

The number of situations in which people are influenced by others is endless. In order to find the reasons of why this happens, it is essential to understand and follow the propagation of information through networks. In this chapter, we address the influence and information diffusion process’s meaning.

Later we described some of the principals of 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 talked about Influence Modeling and Information Cascades.

In the next chapter which is devoted for the implementation and results of the experiments. It involves the clarification of information diffusion process and the process. We also present the tools that we employed and the datasets we have used to conduct the experimental part of this work.

CHAPTER 04

Implementation & Experiments' Results

Chapter 4: Implementation & Experiments' Results

4-1 Introduction

In the precedent chapter, we introduced general ideas behind the influence and modeling information diffusion. In addition, we have described the diffusion process with more details, then we have focused on the modeling of this process using Independent Cascade Model **(ICM)** and Linear Threshold Model **(LTM)** that we will use in our experiments.

In this chapter, we will learn on how to choose initial active set considering three criteria, which are the closeness, degree and random one for our comparison.

We will also explore the different tools and algorithms used to accomplish this work, accompanied by the obtained results.

4-2 Modeling Information Diffusion

There are several mathematical models that have been suggested to formally model the spread of information in the network, we introduced three famous models in previous sections: Linear Influence Model (LIM), Independent Cascade Model (ICM) and Linear Threshold Model (LTM).

We focused in Implementation on the following settings: given a network presented by a directed graph G = (V, E), each vertex v ∈ V can be active or inactive. Each node can switch to active from inactive, but does not switch in the other direction. The diffusion process, starts with an initial activated set of nodes, while time spreads out, continues until no more activation is possible. The influence of this initial set is the expectation number of active nodes by the end of the diffusion process running

4-2-1: Linear threshold model (LTM)

There are lot of way about the modeling of information diffusion. One basic way is based on the use of node-specific thresholds.

In this model, a node V is influenced by each neighbor W according to a weight b_{vw} .Each vertex V chooses a threshold uniformly at random from the interval [0, 1].

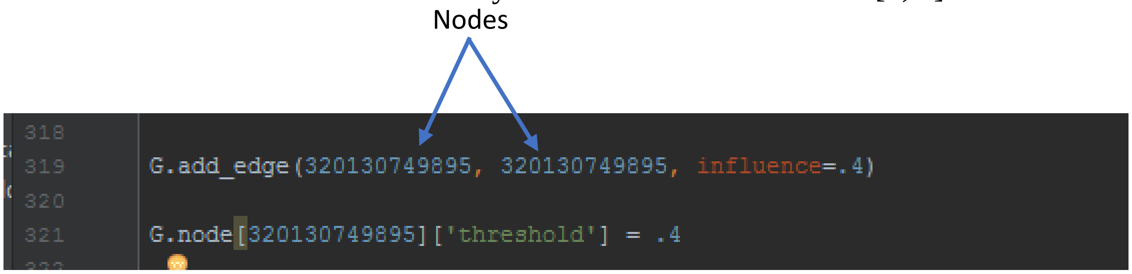


Figure 4.1: how to add edge and weight between two nodes
: how to add threshold to node

As explained before, the diffusion process is display in discrete steps: in step t, all nodes that were active in step t-1 remain active, and we activate any node V for which the total weight of its active neighbors is at least Theta (V)

$$\sum_{w \text{ active neighbor of } v} b_{v,w} \geq \theta_v$$

The LTM algorithm

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

4-2-2: Independent Cascade Model (ICM)

This model is other way for modeling the diffusion process. So that model starts with an initial set of active nodes, and the process unfolds in discrete steps according to the following randomized rule. When node v first becomes active in step s, it is given a single chance to activate each of its inactive neighbors w with a probability $\mathcal{P}_{VW}$ to succeed. If v succeeds to activate W in step s+1, then W will try to activate its inactive neighbors too (see chapter 03).

The ICM algorithm

$S_0 \neq \{\}$;
$i = 0$;
While $S_i \neq \{\}$ **Do**
  $i = i + 1$;
  $A_i = \{\}$;
  For $v \in A_{i-1}$ **Do**
    For w a neighbor of v, $w \notin \bigcup_{j=0}^{i} A_j$ **Do**
      r = random value $\in [0,1]$
      If $r < P_{v,w}$ **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?

Many researchers focused on the problem of maximizing the influence or the spread of information in social networks by choosing the best initial and the most influential ones[20].

It is important step in my experiment of choosing the initial active set. In this work we focused only on the structure of the network.

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

4-3-1: Degree centrality

The degree centrality is a count of total number of connections linked to a node, there are two measures of degree. In-degree is the number of connections that point inbound at a node, Out-degree is the number of connections that originate a node and point outbound to other nodes.

In the figure (4.2) shows the number of degree between (5 to 92) Related to my dataset's work.

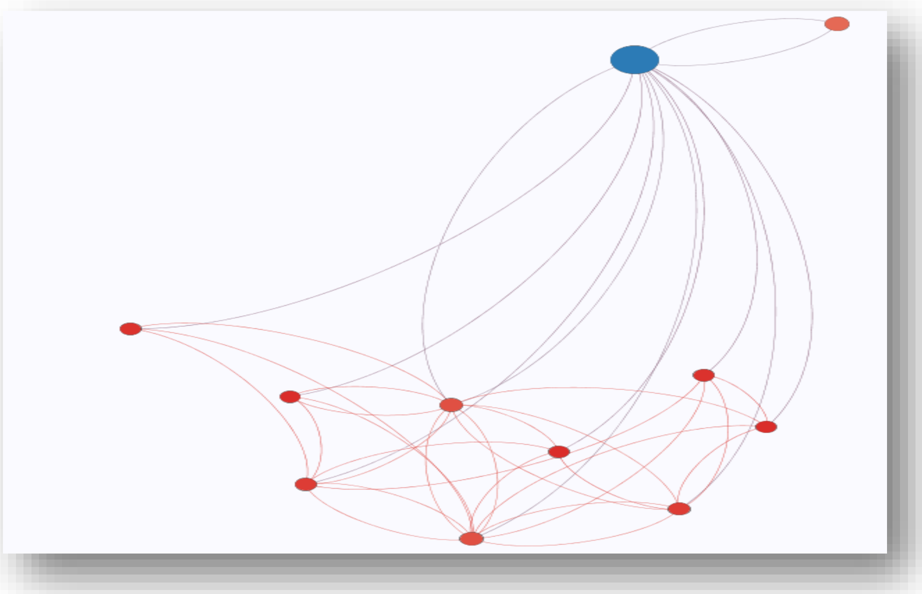


Figure 4.2 : Degree between 5 to 92

In the figure (4.3) shows the list of high degree that I used it in my experiment

Id	Label	Inte...	user...	cate...	post...	fan...	talking...	users...	link	In-D...	Out-...	Deg...
320130749895	msila28 صفحة ولاية المسيلة		wilay...	Com...	1.25	209998	40950	yes	https:/...	9	83	92
379613978719685	السلام		Al.sa...	Com...	0.0	3645	6	no	https:/...	2	15	17
129356317092664	... Matricule لي متركيل تناعك		Jimm...	Musici...	0.0	32815	11	yes	https:/...	5	7	12
186222058087214	J'aime Partager ORIGINAL		likesh...	Local ...	0.02	305032	4744	yes	https:/...	7	5	12
265483253468826	Copier Coller			Movie...	0.0	4976	1	yes	https:/...	6	2	8
117113448318421	...♥ قة مشاهيـــــــر و عظمـــة		mach...	Author	0.0	14397	7	yes	https:/...	4	4	8
232563643498516	سوق مك لي يحل في ليل ☐۞☐ ﻉ		so9m...	Movie	0.0	3352	4	yes	https:/...	1	5	6
153237994690552	تقاليد بلادي (الجزائر . 48 ولاية)			Public...	0.0	4344	5	yes	https:/...	1	5	6
206903016002643	Studio Elsaker استوديو الصقر		studi...	Movie...	0.02	3940	38	yes	https:/...	4	2	6
278649805560441	NessPlus.com		nessp...	Produ...	0.0	11321	5	no	https:/...	5	0	5
150555671647606	dz.com.فرغ قلبك			Sport...	0.0	10355	3	yes	https:/...	2	3	5
110455225690867	Mohamed Lakhdar-Hamina			Public...	0.0	1137	0	yes	https:/...	5	0	5
477005732359684	محمد رسول الله			Comic...	0.0	2949	1	yes	https:/...	1	3	4
148258735257726	Bladi213		bladi213	Com...	0.0	640	0	yes	https:/...	2	2	4
211700342199312	Ramadan m'sila رمضان المسيلة			Com...	0.0	737	1	yes	https:/...	2	2	4
742965095818568	استوديو الصقر - Mariage / الافراح		Studi...	Busin...	0.01	3415	36	yes	https:/...	2	1	3
324761227715373	Ismahane cuisine			Kitch...	0.0	1734	6	yes	https:/...	3	0	3
304284382949401	غزة اشواق واشواك		azou....	Com...	0.0	245	0	no	https:/...	1	2	3
198378655613	Le Buteur : Football Algérien			Maga...	2.24	244129	1737	no	https:/...	1	2	3
144325422279939	Tkarïij Algérien التقرعيج الجزايري		Tkarij...	Musici...	0.03	575479	939	yes	https:/...	3	0	3

4-3-2: Closeness centrality

Closeness centrality is a way of detecting nodes that are able to spread information very efficiently through a graph and it's the average of the shortest path lengths from the node to all other nodes in the network.

Nodes with a high closeness score have the shortest distances to all other nodes.

I chose closeness centrality for choose initial active set because it's used to research in the huge networks, where individuals with high closeness centrality are in a favourable position to control and get information and resources within .

In the figure (4.4) shows the list of high Closeness that I used it in my experiment

Id	Label	Int...	...	...	...	...	t...	u...	l...	In...	...	...	...	...	...	E...	Closeness Centrality
320130749895	msila28 صفحة ولاية المسيلة		...	...	...	2...	40...	yes	h...	9	83	92	...	8...	...	1.0	1.0
271750022866815	جمعية الحضنة للنشاطات الشبانية لبلدية المسيلة			...	...	2...	98	yes	h...	1	1	2	...	1.0	2.0	1.0	1.0
422095264505341	Djamel Mesbah		...	...	...	1...	6	yes	h...	1	1	2	...	1.0	2.0	1.0	1.0
208632939147751	La Gazette du Fennec		...	...	...	1...	7630	yes	h...	1	1	2	...	1.0	2.0	1.0	1.0
198378655613	Le Buteur : Football Algérien			...	...	2...	1737	no	h...	1	2	3	...	2.0	3.0	1.0	1.0
379613978719685	السلام		...	...	...	3...	6	no	h...	2	15	17	...	1...	...	2.0	0.549669
129356317092664	Matricule dz fr قولي متركيل تتاعك		...	...	...	3...	11	yes	h...	5	7	12	...	7.0	...	2.0	0.522013
477005732359684	محمد رسول الله			...	...	2...	1	yes	h...	1	3	4	...	3.0	4.0	2.0	0.509202
304284382949401	غزة اشواق واشواك		...	...	...	245	0	no	h...	1	2	3	...	2.0	3.0	2.0	0.506098
211700342199312	Ramadan m'sila رمضان المسيلة			...	...	737	1	yes	h...	2	2	4	...	2.0	4.0	2.0	0.506098
206903016002643	Studio Elsaker استوديو الصقر		...	...	...	3...	38	yes	h...	4	2	6	...	2.0	6.0	2.0	0.506098
106523136045172	مدرسة هواة السينما الجزائرية			...	...	3...	8	yes	h...	1	2	3	...	2.0	3.0	2.0	0.506098
324250214367731	صديقاتى اجمل هدية فى حياتى		...	...	...	2...	2	yes	h...	1	1	2	...	1.0	2.0	2.0	0.50303
273033719406933	مـــلتقى الأحـبة أولاد سيدي إبراهيم		...	...	...	857	39	no	h...	1	1	2	...	1.0	2.0	2.0	0.50303
148258735257726	Bladi213		...	...	...	640	0	yes	h...	2	2	4	...	2.0	4.0	3.0	0.359307
232563643498516	ღ □۞□ سوق مك لي يحل في ليل □۞□ ღ		...	...	...	3...	4	yes	h...	1	5	6	...	5.0	6.0	3.0	0.354701
186222058087214	J'aime Partager ORIGINAL		...	...	...	3...	4744	yes	h...	7	5	12	...	5.0	...	3.0	0.350211
117113448318421	ـقة مشاهيــــــــــر و عظمـــــة ♥♥♥ الجـــزائر ♥♥♥...		...	...	...	1...	7	yes	h...	4	4	8	...	4.0	8.0	3.0	0.348739
150555671647606	dz.com.فرغ قلبك			...	...	1...	3	yes	h...	2	3	5	...	3.0	5.0	3.0	0.34728
742965095818568	استوديو الصقر - Mariage / الافراح		...	...	...	3...	36	yes	h...	2	1	3	...	1.0	3.0	3.0	0.337398

Figure 4.4 : list of high Closeness

4-3-3: 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 set for spreading the information (or influence) widely in a social media.

When it comes to the number of initial active nodes, there is no general rule to make this choice. The size of the initial active set is usually between 30 to 50 nodes that's depends on the size of network, its type, and the domain of the experiments. In this work the size of initial active set is chosen between 1 to 20 nodes.

In the figure (4.5) shows the list of randomly choose that I used it in my experiment

Id	Label	Int...	...	...	...	...	t...	u...	l...	In...	...	...	...	...	...	E...	Closeness Centrality
895585407127341	Solidarité M'sila			...	...	1...	0	no	h...	1	0	1	...	0.0	1.0	0.0	0.0
742965095818568	استوديو الصقر - Mariage / الافراح		...	...	...	3...	36	yes	h...	2	1	3	...	1.0	3.0	3.0	0.337398
633470946696033	المميز3			...	...	7...	7	yes	h...	2	0	2	...	0.0	2.0	0.0	0.0
632678806789996	السمعي البصري الجزائري			...	...	7...	647	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
607358172662546	Questions		...	...	...	8...	4	yes	h...	2	0	2	...	0.0	2.0	0.0	0.0
552652644920497	مركز الأيهم		...	...	...	2...	103	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
544576918940005	Khelwi		...	...	...	6...	20	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
497874413606647	Pôle Universitaire de M'sila			...	...	1...	3	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
496958313682153	Julj Studio		...	...	...	4...	3	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
490442451133145	EURL Chichi			...	...	1...	6	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
477005732359684	محمد رسول الله			...	...	2...	1	yes	h...	1	3	4	...	3.0	4.0	2.0	0.509202
445702862148148	CLEAR		...	...	...	9...	24	no	h...	1	0	1	...	0.0	1.0	0.0	0.0
422095264505341	Djamel Mesbah		...	...	...	1...	6	yes	h...	1	1	2	...	1.0	2.0	1.0	1.0
421900421346338	حوادث المرور في اجز ئر		...	...	...	2...	119	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
379613978719685	السلام		...	...	...	3...	6	no	h...	2	15	17	...	1...	...	2.0	0.549669
350260166090	أستغفر الله العظيم		...	...	...	8...	305	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
349960987378	Madjid Bougherra			...	...	5...	27	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
329621780469356	Zinet El Bouldane		...	...	...	193	0	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
327406860628094	صفحة تنقل أحاديث النبي صلى الله عليه وسلم		...	...	...	3...	4	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
324809660992803	أشوف فيك يوم			...	...	1...	7	yes	h...	2	0	2	...	0.0	2.0	0.0	0.0
324761227715373	Ismahane cuisine			...	...	1...	6	yes	h...	3	0	3	...	0.0	3.0	0.0	0.0
324250214367731	صديقاتي اجمل هدية في حياتي		...	...	...	2...	2	yes	h...	1	1	2	...	1.0	2.0	2.0	0.50303
320130749895	msila28 صفحة ولاية المسيلة		...	...	...	2...	40...	yes	h...	9	83	92	...	8...	...	1.0	1.0
304284382949401	غزة اشواق واشواك		...	...	...	245	0	no	h...	1	2	3	...	2.0	3.0	2.0	0.506098
287327228048824	Nadir Belhadj Officiel		...	...	...	2...	0	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0
280920165257144	Ness El Khir M'sila ناس الخير المسيلة		...	...	...	1...	51	no	h...	3	0	3	...	0.0	3.0	0.0	0.0
278649805560441	NessPlus.com		...	...	...	1...	5	no	h...	5	0	5	...	0.0	5.0	0.0	0.0
273033719406933	مـــلتقى الأحــبة أولاد سيدي إبراهيم		...	...	...	857	39	no	h...	1	1	2	...	1.0	2.0	2.0	0.50303
271750022866815	جمعية الحضنة للنشاطات الشبانية لبلدية المسيلة			...	...	2...	98	yes	h...	1	1	2	...	1.0	2.0	1.0	1.0
265483253468826	Copier Coller			...	...	4...	1	yes	h...	6	2	8	...	2.0	8.0	4.0	0.261006
252595808094949	كلـمـات المســاء		...	...	...	1...	3	yes	h...	1	0	1	...	0.0	1.0	0.0	0.0

Figure 4.5: the list of randomly choose

4-5: Experiments tools

4-5-1: Gephi [21]

Gephi is one of the most useful programs to explore and understand graphs that consist of nodes and links between them, or so-called graph, where the user can manipulation of structures, shapes and colors, to detect the properties of the network, and explore their hidden properties

It is a complementary tool to traditional statistics, as visual thinking with interactive interfaces is now recognized to facilitate reasoning

In my experiments, I've used "Gephi0.9.2" tool. Gephi is a free and an open-source software for network visualization and analysis.

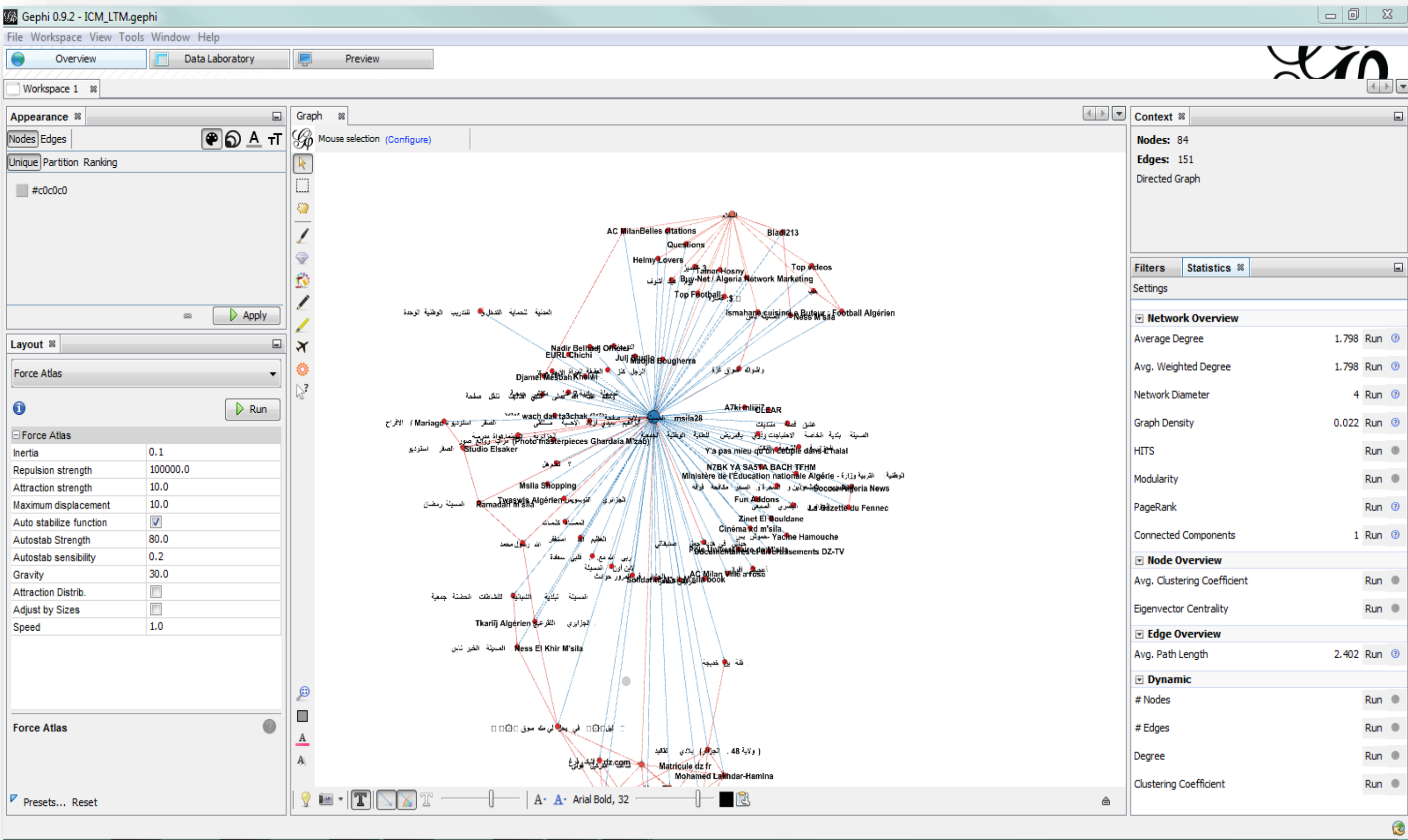


Figure 4.6: The complete Gephi workspace window

4-5-2: Python

Gephi doesn't contain any information diffusion model. For that reason I'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. It is open source, which means it is free to use, even for commercial applications. Python can run on Mac, Windows, and Unix systems. [22]

I've created a method to both Models ICM & LTM that contain three Parameters→**G**: network graph, **seeds**: list of initial nodes active, **steps**: type int is the number of steps to diffuse When steps <= 0, the model diffuses until no more nodes can be activated.

```
def independent_cascade(G, seeds, steps=0):

def linear_threshold(G, seeds, steps=0):
```

Figure 4.7: ICM &LTM by Python

4-5-3: Datasets

Facebook (Web page)

Facebook is an online social network that allows users to post images, photos, videos, files and documents, exchange messages, join and create groups and use different of applications.

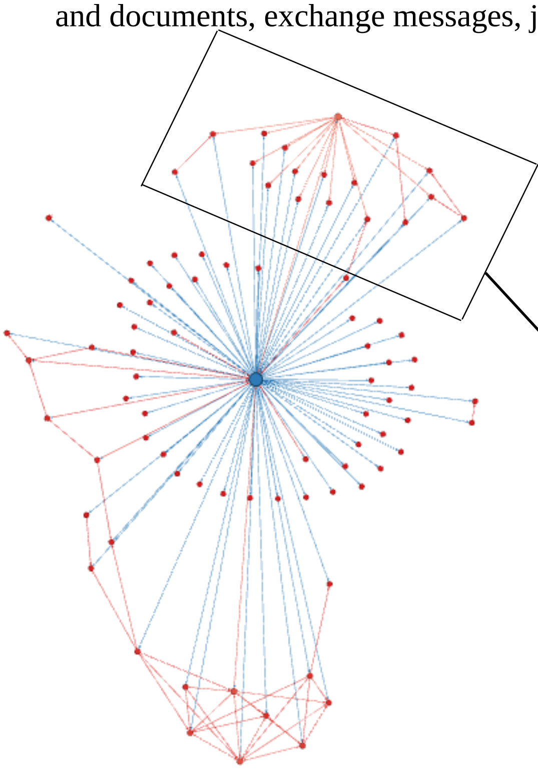


Figure 4.8: Graph of my datasets

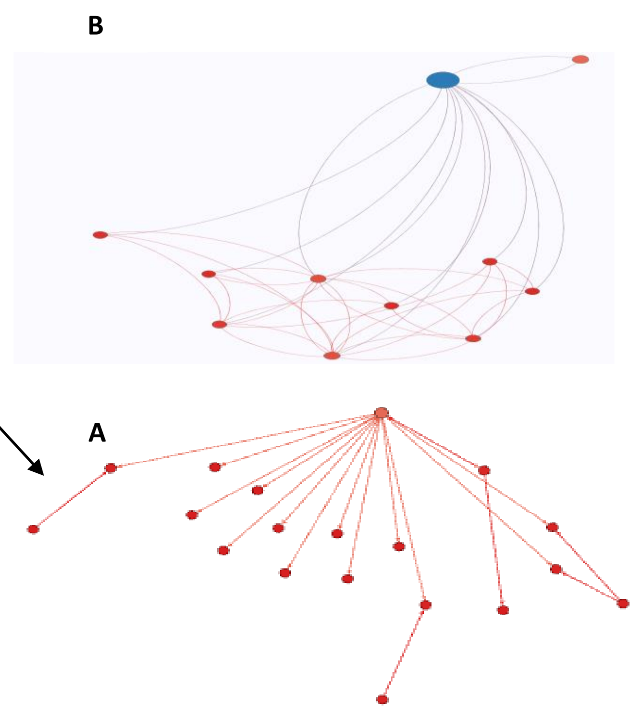


Figure 4.9: sub-graph of two different degrees

I tried many available Gephi layouts and I've chosen the automatic layout called Force Atlas which gave me the network illustrated in figure (4.8). The node"in bleu" is the high degree that's called "**msila28 صفحة ولاية المسيلة**". Gephi allows filter the nodes to create sub-graph by its degree, betweenness…etc. I show in figure (4.9) two sub-graphs which is illustrated their degrees between (2 to 21"graph A") and (5 to 92"graph B").

The network has 151 edges and 84 nodes represented by the edges list (figure (4.10)). . There is no self-loop and no duplicated edge

Nodes Edges Configuration Add node Add edge Search/Replace Import Spreadsheet Export table More actions

Source	Target	Type
مدرسة هواة السينما الجزائرية - 106523136045172	Studio Elsaker استوديو الصقر - 206903016002643	Directed
مدرسة هواة السينما الجزائرية - 106523136045172	msila28 صفحة ولاية المسيلة - 320130749895	Directed
حقيـــقة مشاهيـــــر و عظمـــة ♥♥♥ الجـــزائر ♥♥♥ - 117113448318421	278649805560441 - NessPlus.com	Directed
حقيـــقة مشاهيـــــر و عظمـــة ♥♥♥ الجـــزائر ♥♥♥ - 117113448318421	186222058087214 - J'aime Partager ORIGINAL	Directed
حقيـــقة مشاهيـــــر و عظمـــة ♥♥♥ الجـــزائر ♥♥♥ - 117113448318421	110455225690867 - Mohamed Lakhdar-Hamina	Directed
حقيـــقة مشاهيـــــر و عظمـــة ♥♥♥ الجـــزائر ♥♥♥ - 117113448318421	Matricule dz fr قولي متركيل تتاعك - 129356317092664	Directed
Matricule dz fr قولي متركيل تتاعك - 129356317092664	278649805560441 - NessPlus.com	Directed
Matricule dz fr قولي متركيل تتاعك - 129356317092664	265483253468826 - Copier Coller	Directed
Matricule dz fr قولي متركيل تتاعك - 129356317092664	msila28 صفحة ولاية المسيلة - 320130749895	Directed
Matricule dz fr قولي متركيل تتاعك - 129356317092664	186222058087214 - J'aime Partager ORIGINAL	Directed
Matricule dz fr قولي متركيل تتاعك - 129356317092664	110455225690867 - Mohamed Lakhdar-Hamina	Directed
Matricule dz fr قولي متركيل تتاعك - 129356317092664	dz.com. فرغ قلبك - 150555671647606	Directed
Matricule dz fr قولي متركيل تتاعك - 129356317092664	حقيـــقة مشاهيـــــر و عظمـــة ♥♥♥ الجـــزائر ♥♥... - 11711344831ɛ	Directed
148258735257726 - Bladi213	السلام - 379613978719685	Directed
148258735257726 - Bladi213	Ness M'sila - ناس المسيلة - 181088711929732	Directed
الوحدة الوطنية للتدريب و التدخل للحماية المدنية - 149544255105896	الوحدة الوطنية للتدريب و التدخل للحماية المدنية - 149544255105896	Directed
dz.com. فرغ قلبك - 150555671647606	265483253468826 - Copier Coller	Directed
dz.com. فرغ قلبك - 150555671647606	186222058087214 - J'aime Partager ORIGINAL	Directed
dz.com. فرغ قلبك - 150555671647606	Matricule dz fr قولي متركيل تتاعك - 129356317092664	Directed
تقاليد بلادي (الجزائر . 48 ولاية) - 153237994690552	خديجة بن قنة - 240883675929542	Directed
تقاليد بلادي (الجزائر . 48 ولاية) - 153237994690552	265483253468826 - Copier Coller	Directed
تقاليد بلادي (الجزائر . 48 ولاية) - 153237994690552	186222058087214 - J'aime Partager ORIGINAL	Directed
تقاليد بلادي (الجزائر . 48 ولاية) - 153237994690552	110455225690867 - Mohamed Lakhdar-Hamina	Directed
تقاليد بلادي (الجزائر . 48 ولاية) - 153237994690552	حقيـــقة مشاهيـــــر و عظمـــة ♥♥♥ الجـــزائر ♥♥... - 11711344831ɛ	Directed
186222058087214 - J'aime Partager ORIGINAL	278649805560441 - NessPlus.com	Directed
186222058087214 - J'aime Partager ORIGINAL	حقيـــقة مشاهيـــــر و عظمـــة ♥♥♥ الجـــزائر ♥♥... - 11711344831ɛ	Directed
186222058087214 - J'aime Partager ORIGINAL	265483253468826 - Copier Coller	Directed
186222058087214 - J'aime Partager ORIGINAL	Matricule dz fr قولي متركيل تتاعك - 129356317092664	Directed

Figure 4.10: The edges list shows relationships among the nodes

4-6: Information diffusion results analysis

The social networks used in my experiments, namely Page Facebook that has size, type, and its structure. In table (4.1) I resume the networks structure characteristics.

	Page Facebook
Graph type	Directed
Egocentric network	Yes
Type of relationship	Implicit
Number of Nodes	84
Number of edges	151
Graph density	0.022
Connected component	1
Maximum number of nodes in connected component	84
Maximum number of edges in connected component	151
Diameter	4
Average Path length:	2.402268178785857

Table 4.1: The structure characteristics of Page Facebook

The centrality metrics are summarized in table (4.2) and their distribution plots are drawn on figure (4.11) for Facebook network. In the network (Facebook), one can notice that most nodes have a null value for out-degree, and their in-degree equals to 1. Those nodes represent the followers of the page's (M'sila) followers .Only few nodes, have significant values, where the highest in-degree value is only 9. Therefore, the Max in-degree and out-degree is 9 and 83 respectively.

Average Degree	**1.798**	
Out-degree	**Min**	**Max**
	0	**83**
In-degree	**1**	**9**
Closeness	**0.26**	**1**

Table 4.2: The centrality metrics

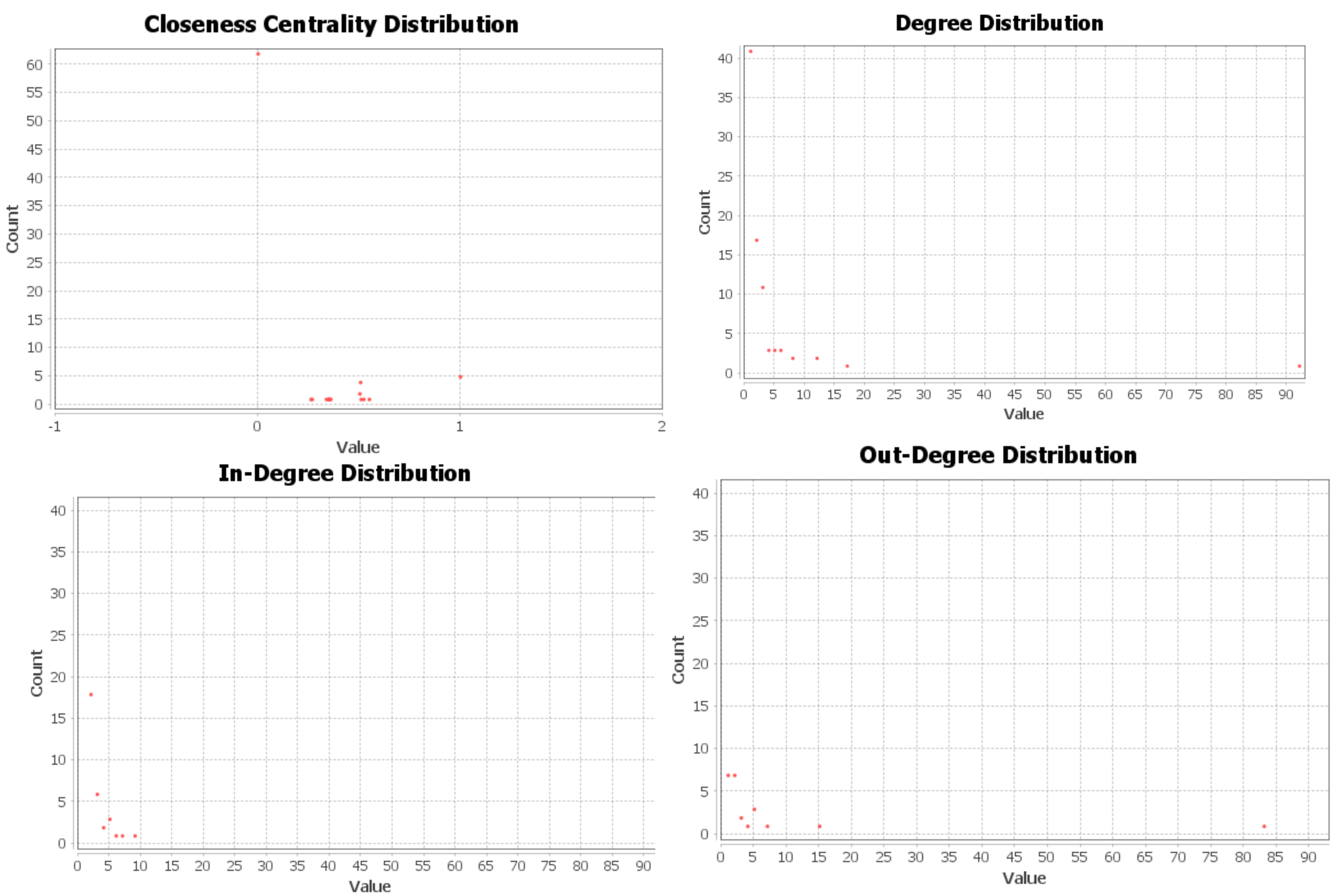


Figure 4.11: Page Facebook centrality metrics distribution plots

The results are presented in this section from dataset. For each algorithm ICM & LTM, the propagation process was run 1000 times for every initial set $A_0 \in \{1, 5, 10, 15, \text{and } 20\}$, then the average of the size of active nodes after each run was computed. This average is considered to be the influence of the initial set.

In the table (4.3) shows the nodes activated before 1000 runs (average)

Number of active nodes	Randomly		Degrees		Closeness	
	LTM	ICM	LTM	ICM	LTM	ICM
1	44	31.388	35	31	35	31
5	40	31.11	37	43	39	40
10	37	31.161	46	46	52	35
15	34	28.268	50	50	53	43.5
20	31	25.909	50	55	53	58.2

Table (4.3): The nodes activated before 1000 runs (average)

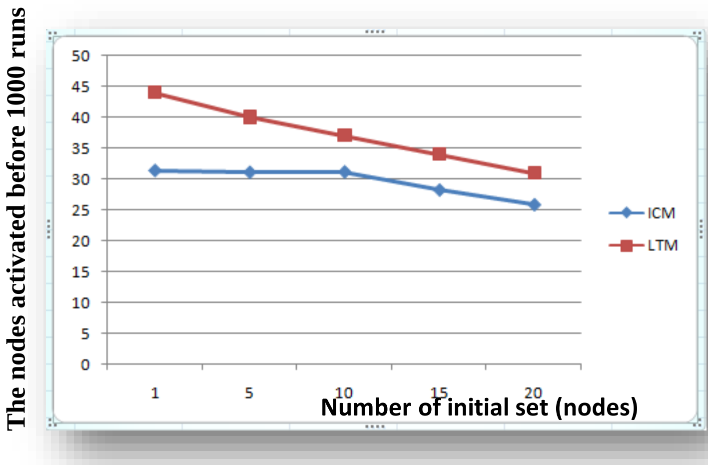


Figure 4.12: The Results of randomly for initial set

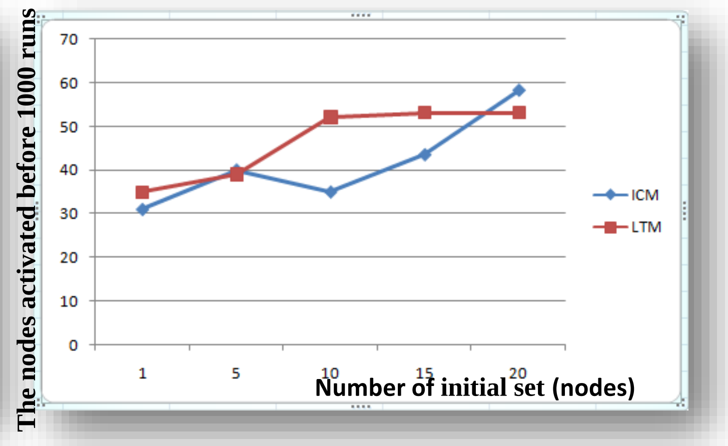


Figure 4.13: The Results of Closeness for initial set

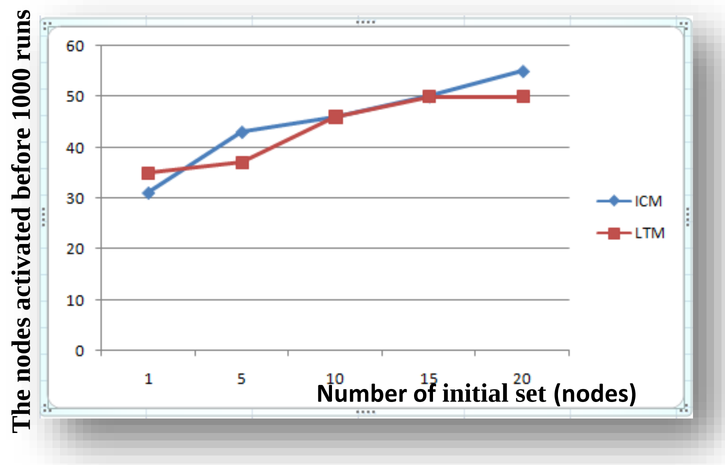


Figure 4.14: The Results of Degree for initial set

Each algorithm provides us analyzing diffusion process to get more insights and conclusions. We made three different analysis, in which they had different criteria to base the results on. We used both algorithms of ICM & LTM in order to review our analysis.

In the first phase, we made our analysis upon random criteria while increasing the number of active nodes. However increasing the number of active nodes didn't positively reflect the results due to the fact that the criteria we based the test on was randomly picked.

On the second phase, like the first phase. But we had a defined criteria which was the closeness. We obtained positive results using the ICM and LTM algorithms with minor differences, therefore, this criteria surely has an impact on determining the influencing node.

On the third phase, similar to the second, we used degree as criteria instead of the closeness, it also had a linear result however unlike the second phase, we obtained less nodes at the end of the study.

Summary

In this chapter contain two parts, the first one was about the implementation of the chosen models we used to analyze the propagation of the information in social media and how to choose initial active set which was important phase in. we talked about the too, the programming language and datasets We've used.

Whereas, the second part presented the results from the experiments of our work, it contain of analyze the structure and metrics of the datasets using Gephi. We made three different analysis, in which they had different criteria to base the results on. We also visualized some diffusion process instances to clarify the results and its analysis. We also visualized some diffusion process instances to clarify the results and its analysis.

General Conclusion

Social media, which has recently spread to social networking sites, is considered the latest development on the Internet. Social media has also provided many opportunities to be a trading place for some companies. Social media is a computer-based technology that facilitates the creation and exchange of information, ideas, professional skills, and other forms of expression through virtual networks.

The most important research recently on social media are Analyzing the information diffusion and the influence of users. Therefore, in this theses use the information diffusion models for get the number of infected nodes in social media networks and the capability of people to spread information through the networks. The used network was a page Facebook which is an implicit relationships between nodes.

There are a lot of information diffusion models, thus we used two most important models which are: the Linear Threshold Model (LTM) and the Independent Cascade Model (ICM). Whereas is the analysis was based on the structure (type, size, diameter, density, connected components). Both algorithms of ICM & LTM are applied three criteria which are: nodes with high overall degree, closeness centrality and random were chosen.

However, some experiments results were quite surprising and interesting, where we had observed from the results in the same network that the LTM model gives more nodes active than ICM due to the first one depends on a node which is influenced by each neighbor and random threshold whereas the second depends on the propagation probability of the edge from u to v (each edge has its own value). Ultimately, our experimental results show that the selection of initial set of nodes affects significantly the diffusion process.

References

Books

[5] D. L. H. B. Shneiderman and M. A. Smith, “analyzing social media networks with nodexl,” Insights from a Connected World, Elsevier, 2011.

[6] S. WASSERMAN and K. FAUST social network analysis methods and applications, 1994.

[7] S. WASSERMAN and K. FAUST Social Network Analysis Theory and Applications, Cambridge University Press, 03 Jan 2011.

[8] Charu C. Aggarwal, Social Network Data Analytics, IBM Thomas J. Watson Research Center, 2011.

[13] Everett M. Rogers DIFFUSION OF INNOVATIONS Third Edition, New York

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

An article

[3] Pritam Gundecha, Huan Liu,Mining Social Media: A Brief Introduction, Arizona State University, Tempe, Arizona, 2012.

[10] Neveen Ghali, Mrutyunjaya Panda, Aboul Ella Hassanien, Ajith Abraham and Vaclav Snasel, Social Networks Analysis: Tools, Measures and Visualization, Springer-Verlag London 2012.

[11] Angela Zoss, Adam Maltese, Stephen Uzzo, Network Visualization Literacy: Novel Approaches to Measurement and Instruction, Duke University, Indiana University, New York Hall of Science,2018

[12] B skiera, O hinz/M span, social media and academic performance: does the intensity of Facebook activity relate to Good Grades? 2015.

[16] Targeted and Budgeted Influence Maximization in Social Networks under Deterministic Linear Threshold Model

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

A dissertation or thesis

[2] HARZALLAH Messaouda KHABABA Amina, Analyse des réseaux sociaux (pour la détection des communautés), Master, University of Mohamed El Bachir El Ibrahimi -Bordj-Bou-Arreridj-faculty Mathematics and Computer Science Department of Computer Science ,2017/2018

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

[20] Laifa Meriem, Social Network Analysis and Information Propagation: A Case Study Using Flickr and YouTube Networks, Master, University of Mohamed El Bachir El Ibrahimi -Bordj-Bou-Arreridj-faculty Mathematics and Computer Science Department of Computer Science ,2011/2012

[14]K BENSALEM, S SOUALMI, Détection des rumeurs dans les réseaux Master, University of Mira _Bejaia

Faculty of science, 2016/2017

Web sites:

[1] Quora https://www.quora.com/What-is-the-function-of-social-media-platforms, Consulted at 15/03/2019

[4] Orgnet http://www.orgnet.com/, Consulted at 16/03/2019

[15] Sumankundu http://www.sumankundu.info/articles/detail/How-To-Code-Independent-Cascade-Model-of-Information-Diffusion . Consulted at 25/04/2019

[18] Investopedia https://www.investopedia.com/articles/investing/052715/guide-understanding-information-cascades.asp , Consulted at 20/05/2019

[21] Gephi https://gephi.org/about/, Consulted at 03/6/2019

[22] Techterm https://techterms.com/definition/python Consulted at 03/6/2019

[23] Opentextbc https://opentextbc.ca/socialpsychology/chapter/affect-behavior-and-cognition/ Consulted at 11/5/2019

الملخص

التأثير الاجتماعي هو العلم الذي يدرس كيفية تغيير أفكار الأشخاص أو مشاعرهم أو سلوكياتهم مصحوبًا بتحليل الشبكة الاجتماعية (SNA)، حيث أنها مهمة لتعيين وقياس العلاقات والتدفقات بين الأشخاص والمجموعات وما إلى ذلك. مشكلة هذا البحث هو التأثير و نشر المعلومات. الأساليب المستخدمة في هذا العمل هي نموذج العتبة الخطية ونموذج التتالي المستقل باستخدام مجموعة البيانات (الشبكة): صفحة فيسبوك ، تعتمد هذه الشبكة على هيكلها (الحجم ، النوع ، القطر ، الكثافة ، المكونات) و نوع العلاقات (ضمني). تم معالجة مقاييس الشبكات والتصور بواسطة برنامجGephi . في النهاية ، تظهر نتائجنا التجريبية أن اختيار مجموعة العقد الأولية يؤثر بشكل كبير على عملية الانتشار ، وأن نموذج (LTM) يعطي العقد أكثر نشاطًا من نموذج.(ICM)

الكلمة الأساسية: تحليل الشبكة الاجتماعية (SNA) ، نشر المعلومات ، التأثير

Abstract

Social influence is the science that study how to change people's thoughts, feelings or behaviors accompanied with Social Network Analysis (SNA) where it's important for Mapping and measuring of relationships and flows between people, groups etc. The problem of this research is influence and information diffusion. The methods are used in this work are Linear Threshold Model (LTM) and Independent Cascade Model (ICM) by use dataset (network): page facebook, this network depends on its structure (size, type, diameter, density, components) and the type of the relationships (explicit). Networks metrics and visualization were manipulated by Gephi as well. Ultimately, our experimental results show that the selection of initial set of nodes affects significantly the diffusion process, and (LTM) model gives more nodes active than (ICM) model.

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

Résumé

L'influence sociale c'est la science qui étudie la façon de changer les pensées, les sentiments ou les comportements d'une personne, accompagnée d'une analyse de réseau social (SNA), où il est important de cartographier et de mesurer les relations et les flux entre les personnes, les groupes, etc. Le problème de cette recherche est l'influence et la diffusion de l'information. Les méthodes utilisées dans ce travail sont modèle linéaire à seuil (LTM) et modèle en cascade indépendant (ICM) par usage base de données (réseau): page Facebook, ce réseau dépend de sa structure (taille, type, diamètre, densité, composants) et de la type des relations (implicite). Les métriques et la visualisation des réseaux ont également été manipulées par Gephi. En fin de compte, nos résultats expérimentaux montrent que la sélection de l'ensemble initial de nœuds affecte de manière significative le processus de diffusion, et le modèle (LTM) donne plus de nœuds actifs que le modèle (ICM).

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