Detection of influencers in social networks: A Survey

Ansam Ali AbdulAmeer
Dept. of computer science
College Science for Women
University of Babylon
Babylon, Iraq
ansam.abaas.gsci4@student.uobabylon.edu.iq
https://orcid.org/0000-0003-1073-5812

Muhammed Abaid Mahdi
Dept. of computer science
College Science for Women
University of Babylon
Babylon, Iraq
Wsci.muhammed.a@uobabylon.edu.iq
https://orcid.org/0000-0002-7820-4340

Mahdi Abed Salman
Dept. of computer science
College Science for Women
University of Babylon
Babylon, Iraq
Mahdi.salman@uobabylon.edu.iq
https://orcid.org/0000-0002-7805-6800

DOI: http://dx.doi.org/10.31642/JoKMC/2018/100103

Received Jun. 7, 2022. Accepted for publication Sept. 28, 2022

Abstract— *Social media influencers have the power to influence others. Identifying influencers in online social networks is essential for various applications in many domains such as advertisement, community health campaigns, administrative science and politics. Detecting influencers on online social networks is achieved in accordance with specific criteria such as the number of subscribers, the number of interactions with them, the extent of people's trust in them, etc. the present study encompasses different measures such as application, techniques, dataset, factors, and dataset. Besides, a table summarising and illustrating the main ideas and approaches is given.*

Keywords— Social networks, influencer identification, Twitter, Algorithms, Models, Followers, community

I. INTRODUCTION

People have grown more communicative due to the explosive expansion of social networking websites by distributing information about brands, services, and products. These social media platforms have developed into new informational resources for consumers and corporations as well. Social media platforms frequently disseminate information and promote products via viral marketing or product placement. By targeting specific individuals named " influential users," the success rate of this form of marketing can be boosted. [1]

Fig.1 demonstrates the whole marketing process. S chooses the users that have been chosen by the company to place advertisements among. T (S) indicates the users affected by S users. The organisation desires for the range of S to be as narrow as possible and the range of final T (S) to be as large as possible. [2].

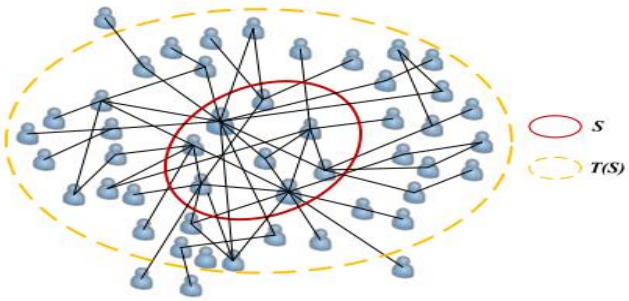

Fig. 1. A marketing demonstration of influence maximisation [2]

Several previous types of research have been done about these influencers, such as calculating their numbers, their influence, and how they have been exploited to solve some of the common problems in society and confront them on social media platforms. Among the most important problems come confronting rumours and finding solutions to repel them. [3]

The various methodologies, findings, solutions and outcomes that are used in the previous studies are included in this survey so as to better understand and identify the real social network influencers and their influence. Mathematical equations and algorithms are adopted to identify negative content in publications and rumours that are shared by these

influencers, and, thus, the propagation of rumours is confronted. [4][5].

The rest of this survey is organised as follows. Section II discusses the literature review, Section III discusses influencer detection methodologies, Section V discusses the comparative study, and Section VI presents the conclusions of the work.

II. LITERATURE REVIEW

Twitter users can create short blogs on any subject or area they want to convey to the community or information they want to disseminate. "Tweets" are short posts that do not exceed 140 characters. Text, links, quotes, and more may be found in these tweets. Each tweet is linked to several factors, such as the user's account and location, and includes a mention of another user's account preceded by a "@ sign" as well as hashtags for a unified phrase or word. This is done with the goal of publishing and republishing the phrase or word by most users as it appears in the "trend." The tweet or hashtag becomes a "Trend" when all of these factors are taken into account, and it then becomes a topic of discussion among the general public. [3]

Some of the most common concepts that will be used in this survey [3]:

- *Tweet*: The message or blog is written in the Twitter application and published.
- *Tweet Attributes*: Each tweet published on Twitter has its characteristics and is based on calculating these features, analysing and extracting several results that are used to solve research purposes or direct them to solve specific problems such as calculating the number of influencers or calculating the priority of tweets, or the hashtag account, which becomes a trend if it is used a lot, and so forth. [3]

  Among these attributes come the following:
  - Likes: It likes the tweet and its content, and a large number of likes for a specific tweet contributes to attracting more people to that community to read this tweet and find out why users like it.
  - Retweet: It means the re-posting of the tweet by several users who are interested in the same field, like the content, or may face the same problem. Influencers usually have a big role in publishing tweets for the reason that their followers republish them for their support or confidence in what this influencer offers. [3]
  - Friends or Followers: Any user can follow another user either as a friend or interested in the same field as the other user. The number of followers of a user is one of the most important metrics based on which it is calculated the impact of this user on his community or followers [6].

  Or the tweet can be implicit attributes as:
  - Tweet content: Determines the type of what has been published or categorised for a particular direction. When calculating these classifications, it is possible to extract the extent of interest in a particular field or the extent of the spread of certain news by calculating the number of the spread of these tweets according to specific principles and processes [7].
  - The goal of the tweet: The goal of the tweet can specify a specific field. It can be a positive goal with the intent of supporting something, a person, or a specific segment, and it can be a negative goal with the intent of criticism or the intent of harming, such as spreading false rumours to harm a particular person or segment. These are problems that require solutions, and some are presented in advance, as will be clarified in this survey. [3][8]
  - Feelings: Tweets often include an expression of feelings for the author to express the user's mood or share positive or negative feelings. Also, this aspect is one of the aspects that are being worked on to be exploited in the analysis and to extract certain results, such as calculating the number of influencers in terms of bringing feelings, which is a very sensitive point in societies. [3]
- *Tweets priorities*: It means arranging tweets according to specific rules to prioritise some tweets over others. Researchers work on these priorities to extract specific results and direct them to solve problems or benefit from them in research. [5]
- *Influencers' priorities*: This means arranging users' accounts according to the most important account, the most active, and perhaps the most number of followers, according to several specific rules. These influencers are calculated and directed to a specific aspect or invested in solving specific problems or benefiting from them in a topic. [3][5]

The objective of this paper is to collect and categorise the various Twitter influence measures that have been published to date. These measures are extremely varied. Others are based on complicated mathematical models instead of simple measurements offered by the Twitter API. The PageRank algorithm has long been used to rank web pages on the Internet and generates several metrics. Others examine the release schedule, the substance of the communications, specific subjects and offer predictions. In addition, given measurements of activity and popularity, common procedures are used to connect metrics that emphasise the usage of retweets for influence estimations. The least frequent one is associated with favourites or likes. They merely learned about the several aspects that were investigated. It is essential to remember that none of the present measures uses all accessible metrics. On the one hand, few measures utilise favourites (or likes), and those that do nearly never combine them with another measurement. On the other hand, it is evident that adopting measurements such as F2, F4, F5, and F6 increases reaction time. This is because these metrics need a precomputing step to build implicit follow-up connections between actors. Notably, almost half of the presently accessible measurements of influence are based on the PageRank algorithm in some manner. This algorithm,which is initially meant to assess the relevance of Internet sites, has found a more permanent use as an effective tool for rating influential individuals in online social networks. Numerous efforts have been made to define topically sensitive metrics, which must require a content analysis of published tweets. Instead, there are few predicted effect metrics. Certain techniques for these measures are unavailable to other

researchers since they have not been made public [5]. Influence analysis is a crucial tool that supports these practical applications of social networks. This survey solicits the most recent work in this field to provide interested readers with a comprehensive and reliable starting point. An overview of social networks is offered first, along with definitions and classifications. Second, the current state of knowledge on social influence analysis was presented on multiple levels, including definition, properties, architecture, applications, and diffusion models. Finally, indicators of social impact are investigated. Finally, a rundown of the various approaches for assessing social effect in social networks offerring an overview of well-known methods for increasing influence [9].

This research explores the idea and purpose of opinion leaders, as well as studies that employed centrality and maximising methodologies to identify opinion leaders in social networks. In addition, a comparison of several kinds of centrality techniques was conducted. The idea of centrality tends to emphasise nodes that reflect influential users or opinion leaders. Using the same network [10], calculate the centrality of Betweenness (A), Closeness (B), Eigenvector centrality (C), Degree centrality (D), The Harmonic Centrality (E), and Katz centrality (F).

In the real world, the operation and integrity of a great majority of complex systems depend on a small number of important nodes or influences. These effects may be understood in a variety of ways, such as structurally important nodes that maintain network connectivity or dynamically significant units that have the capacity to have a disproportionate influence on certain dynamical processes. The identification of the most efficient influencer selection in a system has far-reaching effects across a vast array of industries and fields. This article provides cutting-edge solutions for a variety of applications and examines the most recent research advances in the detection of influencers from several perspectives. Using either continuous (for example, separate cascade models) or discontinuous phase transitions, investigate many techniques for identifying significant nodes capable of impacting the overall dynamics of the system (e.g. threshold models) [11].

III. DETECTION INFLUENCER WITH DIFFERENT TECHNIQUES

Finding influencers in a community is one of the main challenges to benefit from them in several areas, such as publishing important news or putting up advertisements to attract a larger number to interact with them, or perhaps to take their help in denying negative, false, harmful news or rumours and others. The main challenge lies in how to find these influencers, as some studies and research have proven that those who have the largest number of followers are not necessarily considered influencers, while another study has proven that the influencers are considered based on the number of retweets for them and other interactions depend on their tweets such as likes, comments, etc. [4][5][8].

In this survey, the most important studies conducted to extract influencers based on specific algorithms, formulas, or patterns, and perhaps surveys by users, will be clarified.

A. UIRank Technique

Jianqiang et al [12]. They propose an algorithm that is called "User Impact Rating Algorithm (UIRank)" to identify influencers among users based on interaction information flow and interaction relationships. This algorithm works through a graph in which it tracks influential users through the content of a Tweet, how the information is disseminated to the user, and their frequency of influence. The influence of the user network is determined by the person's position in the network of follower connections. When a user u tweets t, his follower's v will read it with probability p and retweet or comment on it with probability q, causing t to be redistributed among v's followers. The UIRank ranking formulae may be determined using random walk theory when, after reading t, v's supporters retweet and comment on it, causing it to spread again:

$$\text{IR(u)}=a\sum_{v\epsilon followers(u)}(IR\ (v)^{*}\pi_{uv}) + (1-a) \quad (1)$$

Where Followers (u) denotes user's collection that u follows, and a refers to decay factor.

This approach presents promising results in finding influencers according to the above criteria. The references of cross-validation are drawn from four approaches supplied by scholars to validate the effectiveness of UIRank. The influential individual reference collection contains noteworthy nodes that have been examined using a variety of approaches. Each approach evaluated the reference set's precision, recall, and F1-Measure value, as the following evaluation formula explains:

$$\text{Precision}=\frac{|S\cap S_2|}{S}\ ,\quad \text{recall}=\frac{|S\cap S_2|}{S_2}\ ,$$

$$\text{F1-Measure}=\frac{2\times precision\times recall}{precision+recall} \quad (2)$$

B. Social Network Analysis (SNA)

In the same regard, the purpose of **[13]** is to identify social media influencers. In the present study, a fictitious dataset, Social Network Analysis (SNA) and weighting of SNA metrics are used as techniques (degree, closeness, betweenness), and then they are calculated via using the following formula:

1) *Degree Centrality (DC):* The number of edges linking vertices was calculated using DC. Each vertex has a different DC value than the other vertices. The DC measurement formula is as follows:

$$\text{DC}(i)=\sum_{j=1}^{n} a_{ij} \quad (1)$$

where n represents the entire number of vertices and aij represents the total number of connections between I and j vertices.

2) *Closeness Centrality (CC):* CC has discovered the shortest distance between all vertices and a vertex destination. The CC value influences the network's relationships; the bigger a vertex's CC value, the more influential it is. The formula for calculating CC is as follows:

$$CC(i)=\frac{1}{\sum_{j=1}^{n} d(i,j)} \quad (2)$$

where n is the total number of vertices and d(i,j) represents the shortest distance between vertices I and j.

3) *Betweenness Centrality (BC):* The shortest route between two vertices has been identified using BC. What follows is the formula for calculating BC:

$$BC(i)=\frac{\sum_{i\neq j\neq l} g_{ij}}{g_{il}} \quad (3)$$

where gjl is the shortest route between vertices j and l. While gjl(i) represents the number of paths from vertex j to vertex l that pass via vertex I.

This study finds a list of SNA measuring data which are then weighted. The results identify the top influencers on social media and their relationships with other accounts. The relationship has been visualised in a model to determine the spread's route. The SNA Measurement Results and Spreading Influencer Models are weighted differently. Each SNA calculation generates a unique set of influencers. The goal of weighing SNA data is to ascertain the primary influencer. Different SNA metrics must be equal using a range of values from 0 to 1. The formula for generating a range from 0 to 1 is as follows:

4) $$Z=\frac{x-min}{max-min} \quad (4)$$

where z is a new DC, CC, or BC value for each vertex. X denotes the DC, CC, or BC value at each vertex. The min and max values represent the lowest and maximum DC, CC, or BC values, respectively. Each new value of DC, CC, and BC has been weighted and added together to get the vertex's outcome. The weighting formula is as follows:

5) $$Y=\frac{W1.DC+W2.CC+W3.BC}{W1+W2+W3} \quad (5)$$

Y denotes the final weighted SNA value, while w1, w2, and w3 are weights with one, two and three values, respectively.

Weighting and modelling hierarchy findings enable the identification of influencers and the discovery of active accounts in the graph. The hierarchical model gives information on the different ways of an infection might spread. The findings of this identification may provide a potential answer to the problem of hoax influencer accounts on many social media platforms.

Associates Tridetti, Stéphane [14] presents, in his master's thesis, a study on the extraction of influencers and their use in the marketing campaigns of companies or brands for a specific target audience through their permanent activity in continuing to support this brand. Collecting data from Twitter [15] requires him to use a step-by-step algorithm to create a graph of who the influencers are and how they can exploit and use centrality measures (closeness [16] [17], Betweenness [18], Eigenvector [19]). Additionally, the study analyses the robustness and qualities of the three fundamental notions using a combination of conceptual and empirical experiments.

C. Association Learning Rules

Erlandsson et al. [20] suggest learning rules to discover influencers through relationships between users. Numerous metrics exist to aid in the understanding of acquired association rules. Erlandsson et al. [21] describes the crawler, provide the data used in this investigation and make available from an SQL database, as described in [22].

Support denotes the percentage of D covered by the item set. It is determined by dividing I by the total number of transactions (posts) in our dataset D. Alternately, it may be determined by dividing the frequency of A and B by the total number of items in D, as shown in "(1)":

$$\text{Support}(\{A,B\}) = \frac{\text{Support}(\{A,B\})}{|D|} \quad (1)$$

The second measure, **Confidence,** estimates the percentage of transactions having A and B that will also include C if the following condition holds true: {A, B} ⇒ C. As seen:

$$\text{confidence}(\{A, B\} \Rightarrow C) = \frac{\text{Support}(\{A,B,C\})}{\text{Support}(\{A,B\})} \quad (2)$$

The third measure is **lift** which refers to the degree of dependence between observed values.

$$\text{lift}(\{A, B\} \Rightarrow C) = \frac{\text{Support}(\{A,B,C\})}{\text{Support}(\{A,B\}) \times \text{Support}(\{C\})} \quad (3)$$

Finally, **conviction** indicates the ratio of projected support for A and B in the absence of C.

$$\text{conviction}(\{A, B\} \Rightarrow C) = \frac{1-\text{Support}(\{A,B\})}{1-\text{confidence}(\{A,B\} \Rightarrow C)} \quad (4)$$

The provided metrics explain taught rules in D whereby greater values for all four measures indicate that the learned rule has predictive importance. They report that the execution of many tests to validate the findings on social networks and contrasted the prominent people who are featured in the results based on their degree and page centrality. Their findings show that this strategy gives quicker outcomes than others.

D. Massive Unsupervised. model Outlier Detection (MUOD)

Another use of influencers in FDA is shown in this study by Azcorra et al [23]. They categorised some user values into several different categories, each of which includes different types of influencers. By applying their 'Massive Unsupervised Model Outlier Detection (MUOD)" and testing it on social networks, they found that different groups of influencers were automatically identified and distinguished. They also found that there are features associated with each group of influencers, such as the ability to attract the largest number of likes, attract participation, or attract the largest number of followers. Consequently, OSNs may profile a collection of parameters that measure users' connection, activity, and other relevant characteristics:

Connectivity may be measured using the in-node degree, the out-node degree (which refers to friend or follower ties), the clustering coefficient, and other centrality measures.

Frequently, the activity parameters are separated into two primary categories. The user's actions are grouped into a single category. These are often published posts and "likes" (known as "plus ones" in Google+) given to other users' posts.

Likes and comments from other users and re-shares and re-posts constitute extra responses to a user's post (i.e., retweets in the case of Twitter). Each user of an OSN maintains a profile with information about themselves. Depending on the OSN, a user's profile may include, among other things, the user's name, location (such as the city in which she resides), profession, educational background, and gender. Since its inception in June 2011, Google+ has gathered an official total of 2.5 billion Google account holders who have enrolled for the network. As such, Google+ has become the most popular social networking site in terms of the user base, followed by Facebook and Sina Weibo (a Chinese-language OSN). These many platforms are applied to assess MUDO. [23].

E. Machine Learning Algorithms

There are numerous methods for detecting Influencers' significance in the marketing domain. Arora et al **.[24]** suggest a mechanism for calculating influencers on social networks such as Twitter, Instagram, and Facebook. In their work, they adop nested learning algorithms (ordinary least squares (OLS), K-NN regression (KNN), support vector regression (SVR), and Lasso Regression models).

1) Ordinary Least Squares (OLS)

MLR is based on OLS; the model is fitted so that the sum of squares of the observed and predicted values is as small as possible [25]. The MLR model is predicated on a number of assumptions (e.g., errors are normally distributed with zero mean and constant variance). The regression estimators are optimal if the assumptions are met. The optimality of the estimators is determined by their being unbiased (their expected and true values are identical), efficient (their variance is small in comparison to other estimators), and consistent (estimator bias and variance tend to approach zero as the sample size approaches infinity) [24].

$$Det_{Ceof}{}^2=\frac{SumSqTot}{SumSqReg}=1-\frac{SumSqEr}{SumSqTot} \quad (1)$$

where SumSqTot, SumSqReg, and SumSqEr are the sum of squares total, regression, and error values, respectively, as shown:

$$SumSqTot=\sum(Y-\overline{Y})^2$$

$$SumSqReg=\sum(Y`-\overline{\dot{Y}})^2$$

$$SumSqEr=\sum(Y-\grave{Y})^2$$

To have a perfect regression model, the SumSqEr should be zero and the $Det_{Ceof}{}^2$ should be one. On the other hand, if the regression model is a complete failure, SumSqEr and SumSqTot become equal, and the regression fails to explain any variance, resulting in a value of $Det_{Ceof}{}^2$ zero.

2) Support Vector Regression (SVR)

Support Vector Machines can be used to solve classification difficulties as well as regression problems. SVR solves the problem using a tiny selection of training points, which has huge computing benefits. The dataset is scaled to train the regression model with a linear kernel using "(2)":

$$F(x,w)=\sum_{j=1}^{k} w_j g_j(x)+b \quad (2)$$

where $g_j(x)$ signifies a collection of nonlinear transformations and b denotes the bias that can be discarded in the presence of zero-mean data.

3) K-NN regression

Calculating the average of the numerical target of the 'k' nearest neighbours is a straightforward implementation of KNN regression.

The regression variation of the KNN classification uses the same distance functions as the classification variant. The Euclidian (Euc_Dist) and Manhattan (Man_Dist) distances are denoted by "(3)" and" (4)", where x and y denote the two data points for which the distance is estimated.

$$Euc\_Dist=\sqrt{\sum_{i=1}^{k}(x_i-y_i)^2} \quad (3)$$

$$Man\_Dist=\sum_{i=1}^{k}|x_i-y_i|^2 \quad (4)$$

After obtaining the cumulative results, they are calculated to show the regression index of the influence. Their results prove that communication, admiration, feelings, attraction, and participation play a crucial role in determining influencers.

F. Categorical Influencer Detection(CID)

Using a deep learning methodology that includes VAE and word vectors to imitate the LDA (Latent Dirichlet Allocation) methods, a method for finding themes in microtext is devised. This study identifies category influencers by using microtext-based topic modelling that is based on the following factors [26]:

1) *Reach score:* This metric indicates how many people are reached by the influencer's social media posts. This score is based on empirical evidence:
(total followers þ total friends) /2,000,000 × 100
The value of 2,000,000 is decided based on actual data collected in the database at the time, as the majority of active prominent users who have more than 2,000,000 followers and friends.
2) *Resonance score:* This score indicates is initiated when an influencer in a topic is likely to generate conversation about that issue or not. This score is determined empirically in the following way:
resonance score = 0.4 × engagement_score + 0.6 × post_resonance_score
where:
engagement score = (avg_interactive_score/total followers)/0.27
The average interactive score is the ratio of the number of likes, shares, and comments received by this influencer's posts to the total number of posts published by this influencer in the last three months.
avg_interactive_score=(total_likes+total_shares+total _comments)/total_posts.
3) *Relevance score:* This score shows an influencer's importance to a topic. All of the influencer's posts and comments are gathered to accomplish this. Again, the microtext-based topic modelling method is used to infer the themes that emerged from the influencer's postings.

4) *Sentiment score:* quantifies the general opinion expressed about this influencer by other users on social media networks, whether good or bad.
   This score calculated as: (total_positive-total_negative)/ (total_positive +total_negative) × 100
5) *influence score:* this score is calculated as follows:
   Influence Score = 0.3 ×Reach score + 0.3 × Resonance score + 0.3 ×Relevance score + 0.1 × Sentiment score.

These experiments use two datasets: the benchmark dataset [27] [28] and the showcase dataset.

It is worth mentioning that the metric for evaluation used in this approach is the Normalized Point–wise Mutual Information (NPMI) [29], as the primary metric for evaluating the qualitative of the topic that the adopted model discovered in this experiment. As a result, it has been established that the vocabulary of found themes closely reflects human judgment.

G. Passion Point

The final approach in this study of influencer detection has been proposed by Huynh et al. [30]. Based on the connections between users and companies, they created a way for visualising social network influencers. This model, which we will call SNet, consists of the following components: **(U, T, R)**, where **U** represents a collection of social network members and **T** represents a collection of social network tags. **R** represents the total amount of users and tags. Each user may have a distinct collection of tags, and each tag may be associated with many users. The passion point is a collection of social network links that are based on a graph that illustrates the relationships between users' influence, speed of information propagation, favourite brand, and the sharing of comparable brand features. R is a collection of relationships in social networks. **R** is a collection of relationships in social networks. When a user u€ U is considered, the vertex represents the user's degree of influence.

$$\textbf{IU(u) = (Impress(u), Popularity(u))} \quad (1)$$

$$\text{Where} \quad \textbf{impress(u)} = \frac{\alpha.SI(u)+\beta.CI(u)+\gamma.Ir(u)}{\alpha+\beta+\gamma} \quad (2)$$

$SI(u)$, $CI(u)$, $Ir(u)$ are respectively computed. α, β, and γ: are weighted numbers.

$$\text{SI(u)}=\frac{\alpha_1.\#(\text{SU}_1(\text{u})) +\alpha_2 .\#(\text{SU}_2(\text{u})) + \alpha_3.\#(\text{SU}_3(\text{u}))}{\#(\text{u.ListFriends}) + \#(\text{u.ListFollowers})}$$

where SU(u): = $\cup_{t\in u.Listags}$ t.Sh: a set of users sharing u's tags.

$SU1(u) := \{v \mid v \in SU(u)$ and friend $(u, v)\}$: a group of users that share u's tags and who are friends with u.

$SU2(u) := \{v \mid v \in SU(u)$ and follower $(v, u)\}$: a collection of users who are sharing u's tags and who are also followers of u.

$SU3(u) := SU(u) \setminus (SU1(u) \cup SU2(u))$: set of users who share u's tags but are not related to u.

$$\alpha1, \alpha2, \alpha3: are\ weighted\ numbers, 0 < \alpha1 \leq \alpha2 \leq \alpha3 < 1.$$

A person who shares a tag signifies that they are interested in it. A colleague is more interested in the post than a follower, but an unconnected user is only interested if the content is really intriguing [31].

CI(u): influence of the user u's comment. It quantifies the effect that comments have on u's tags.

$$\text{CI(u)}=\frac{\beta_1.\#(\text{CU}_1(\text{u})) +\beta_2 .\#(\text{CU}_2(\text{u})) + \beta_3.\#(\text{CU}_3(\text{u}))}{\#(\text{u.ListFriends}) + \#(\text{u.ListFollowers})}$$

where CU(u): =$\cup_{t\in u.Listags}$ t.Com : A group of people who have commented on u's tags.

CU_1(u): = {v | v ∈ SU(u) and friend (u, v)}: set of users who have commented on u's tags and who are also u's pals.

CU_2(u): = {v | v ∈ SU(u) and follower (v, u)}: set of users who have commented on u's tags and are also followers of u.

$CU3(u) := CU(u) \setminus (CU1(u) \cup CU2(u))$: set of users who have commented on u's tags but are not related to u.

$\beta1, \beta2, \beta3$: are weighted numbers, $0 < \beta1 \leq \beta2 \leq \beta3 < 1$.

Where $Ir(u)$: interactor ratio for the tag of the user u.

$$\text{Ir(u)}=\frac{\gamma_1.\#(I_1(\text{u})) +\gamma_2 .\#(I_2(\text{u})) + \gamma_3.\#(I_3(\text{u}))}{\#(\text{u.ListFriends}) + \#(\text{u.ListFollowers})}$$

where $I(u) := \cup_{t\in u.Listags}$:interaction : A group of individuals who interact with u's tags.

$I1(u) := \{v \mid v \in I(u)$ and friend $(u, v)\}$: a set of users who engage with u's tags, and who are friends of u.

$I2(u) := \{v \mid v \in I(u)$ and follower $(v, u)\}$: a set of users who interact with u's tags and who are followers of u.

$I3(u) := I(u) \setminus (I1(u) \cup I2(u))$: a set of users who engage with u's tags but are not tied to $u.\gamma1, \gamma2, \gamma3$: are weighted numbers, $0 < \gamma1 \leq \gamma2 \leq \gamma3 < 1$

while Popularity [32] can compute as:

$$\textbf{Popularity(u)} = 1 - e^{-\lambda.\#(F)} \quad (3)$$

where F: = u. List Friends, ∪ u. List Followers, and λ: is a constant. And they used their results in influencer marketing and got positive results.

H. Soft Rumour Control

Another approach to detecting influencers is that of Mojgan et al. [33]. They, however, use influencer to spread anti rumours on social media, presenting a novel concept of social network-based soft rumour control. The premise of the suggested model is that as people's knowledge increases, they will be able to make more exact judgments concerning rumours. Anti-rumour communications are distributed to increase public awareness through the use of respected authority and trustworthiness. Influencers are calculated in this method as follows:

1) *Interest*: It is essential to consider his or her occupation and expertise if someone wants to assess a user as a trusted friend, on the one hand. On the other hand, obtaining the user's competence is difficult to undertake. It is typically challenging to build user expertise based on the user's accessible data, even more so when we want machines to do it automatically. As a result, we can hypothesise that persons who are interested in a certain topic are more likely to be experts in that field [33].

$$\mathrm{IN(f,q)}=\sum_{r_i \in R_f} \sum_{t_j \in q} \frac{tf(t_j,r_i).irf(t_j)}{|R_f|} \qquad (1)$$

In "(1)", the functions tf (tj, ri) and irf (tj) compute the term frequency for the given resource ri and the inverse term frequency for the whole resource collection, respectively. This calculation is run for each resource r in the resource collection Rf and each word in the rumour message q.

2) *Social intimacy and popularity*: Social intimacy quantifies the degree to which the requester user and his/her buddy share mutual relationships [34].
The social closeness between a requester user u and a friend is expressed by the formula below f_{ide}:

$$SI\ (\ ,f_i\ ) = \frac{(Sim(ol(u,A),ol(f_i,A))+Sim(il(u,A),il(f_i,A)}{2} \qquad (2)$$

where u and fi stand for the individual who made the request and a specific friend of his or hers, respectively, and u and fi stand for the requester and the friend, respectively. A denotes a group of user u's friends in which user u is included. Each element of ol (u, A) represents the number of out-degree friendships between the user u and each person in set A. Each element of il represents the quantity of in-degree friendships between all users in set A and user u. (u, A). These functions are not equal since the number of in-degree and out-degree connections between user, u, and their friends is not equal. Sim calculates the cosine similarity of two vectors, while social popularity reflects an individual's degree of reaction. The greater a user's popularity, the faster he or she answers to user requests. Prior interactions with a certain friend influence a person's popularity with that friend. We examine the number of retweets and answers to the requester's tweets to determine the user's popularity inside the buddy circle. The formula below expresses the social popularity (SP) relationship between a requester user u and a buddy of his/ her on Twitter:

$$SP\ (\ ,f_i\ ) = \frac{\sum_{k=1}^{n}\left(\frac{Retweet(t_{ik})}{maxRetweet(A)}\right)+\left(\frac{Reply(t_{ik})}{maxReply(A)}\right)}{2} \qquad (3)$$

n is the total number of tweets in the debate. The Retweet(t_{ik}) and Reply(t_{ik}) variables provide information on the number of retweets and replies, respectively, received by t_{ik}. The maximum number of retweets and responses for all users in set A, where A is the set of persons that user u considers to be his or her friends. Notably, leading components, which are comparable formulations based on the notion of social popularity, may be used to a variety of social networks.

3) *Choose dependable consultants:* Those with an interest in the context of the rumour, as well as a high degree of social connection and popularity with the requester user, are regarded as trustworthy consultants. Consequently, the following formula is used to evaluate a requester user u's trust in a friend fi for rumour message q consulting:

$$Trust\ (u, fi, q) = \alpha * IN\ (fi, q) + \beta * SI\ (u, fi) + \gamma * SP\ (u, fi) \qquad (4)$$

Where α, β, γ and are weights used to balance the relative value of interest (IN), social popularity (SP), and social intimacy (SI). Take note that the coefficients can be optimally chosen using real-world data training. They employed this method to determine appropriate coefficients, which will be provided in the evaluation section. As a result, the collection of trusted friends of a given user u is constructed as follows about the rumour query q:

$$Trusted(u, q) = \{fi \in A_u \mid trust(u, f_i\ , q) > \tau\}$$

Where A_u indicates the user u and r collection of friends and is a deemed trust threshold. As a result, when a user hears a rumour, he or she rates his or her friends differently depending on the rumour's content, taking into account his or her interactions with his or her friends and the neighborhood graph. This approach used three evaluation metrics as follows precision, recall, and F1 measure based on PHEME dataset [35].

I. Reliability and validity

While Ebuka et al. [36] explain the effect of social media influencers on the purchasing intentions of social media users, where focused on expanding the horizon of source credibility by applying the model experimentally to a diverse group of social media influencers for the first time. With a similar effect, this finding should be utilised in a formal industrial context to ensure that the stated objectives are met. The population of this study is composed of Anambra state's active social media users. Using convenience sampling for an infinite population, a sample size of 220 was determined. The construct reliability and discriminant validity [37] tests are conducted using smart- pls. SPSS version 24 is used to compute the data, and SEM using Smart-Pls is used to analyse it. The findings of this study indicate that trustworthiness, attractiveness, and influencer product pairing all have a favourable and significant effect on buy intention.

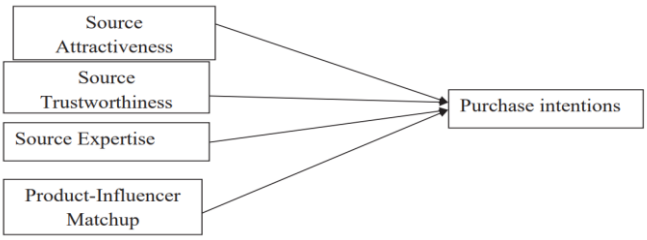

Fig. 2. The proposed scheme of Purchase intention [36]

IV. COMPARATIVE STUDY

As shown in Table 1, a summary of the most important studies related to different techniques for finding influencers in social networks, with the attributes and characteristics of each study and what it focused on and targeted, as well as the direction of research and dataset.

Table1. A brief summary of the most important studies in this field

Reference	Application	Technique	Data set	factor	Evaluation metric
Jianqiang et al.2016 [12]	Micro-blog marketing	UIRank	Sina Weibo API	Interaction information flow and interaction relationships	Precision, Recall, F1 measure
Tridetti et al.2016[14]	Fashion topics	step-by-step	Twitter	Centrality measures	Conceptual and empirical studies
Erlandsson, et al. 2016 [20]	———	Association Rule Learning	Facebook	Relationships between users	Degree and page centrality
Azcorra et al. 2017 [23]	———	MUOD model	Facebook, Twitter, or Google+	Connectivity, Activity parameters, and profiling data	SI (susceptible-infected)
Andre et al. 2018[13]	Eradicating hoax accounts	SNA and weighting of SNA measurements	Facebook hoax	Degree, Closeness, and Betweenness Centrality	Spread pattern
Arora et al.2019 [24]	Marketing and brand management	underlying machine learning	Facebook, Twitter and Instagram	Regression index of the influence	MAE, MSE, and RMSE.
Thanh et al.2019[26]	Marketing campaigns.	CID	Benchmark dataset, showcase dataset	Reach, Resonance, Relevance, Sentiment score	NPMI
Huynh et al.2020[30]	Marketing the brand/product/news	passion point	Facebook and Twitter	Popularity, Impress	Positive Feedback from the Customers
Mojgan et al. 2021 [33]	Anti-rumor spreader	Soft rumor control	Phem e dataset	Interest, social intimacy, and popularity	Precision, Recall, F1 measure
Ebuka et al. 2021[36]	Purchase intentions	Reliability and validity	active social media users	Attractiveness, Trustworthiness, Expertise, Product-Influencer Matchup	Discriminant Analysis, validity

V. DISCUSSION

It may be useful to emphasize, many of these approaches are useful from this influencer in marketing, anti- rumour, etc. However, using influencer identification theory in biological and technical systems is now very rare. Implementing recent advancements in influencer detection in various real-world systems helps demonstrate and disseminate these strategies more effectively, which is advantageous given that some of the more complex processes in statistical physics are difficult to comprehend and require specialised knowledge. In addition, the bulk of modern procedures are developed in ideal conditions. The current methods that depend on supervised techniques have drawbacks [24]. They require extensive manual study of the problem and data in order to define attributes. Because their use is entirely dependent on the definition: if the definition is incorrect or unsuitable in a specific context, the consequences will be similarly incorrect or wrong, but it works on active users. From another angle, the primary disadvantage of association rule learning is that rules for the largest pages in the dataset cannot be extracted [12,33].

VI. CONCLUSIONS

At present, Social networks are more widely used. They are characterised as fast-spreading and influential in public life regarding what is being published. Controlling social networks is one of the critical challenges, as it is not based on a control centre and is not subject to rules and provisions. Accordingly, there must be appropriate solutions to control the spread of news and what is circulated in it. One of the most important solutions is to extract influential users from among all users, where they have some importance in controlling what is published or denying some of the news. Hence, several factors, such as include feelings, likes, similar user opinions, reposts, and so on, are needed.

In this survey, most types of research are taken on extracting influencers from social network users. They are classified according to different measures according to their influence on society, popularity, account activity, and impact on the information spread. The results can be seen briefly through all these researches in Table 1. After discovering these influencers, they can be directed to solve some problems or issues raised in a society, according to certain criteria through experience.

REFERENCES

[1] S. Arrami, W. Oueslati, and J. Akaichi, "Detection of Opinion Leaders in Social Networks: A Survey," vol. 76, pp. 362-370, 2018.

[2] Z. Jianqiang, G. Xiaolin, and T. Feng, "A New Method of Identifying Influential Users in the Micro-Blog Networks," *IEEE Access*, vol. 5, pp. 3008-3015, 2017.

[3] E. Guzman, R. Alkadhi, and N. Seyff, "A Needle in a Haystack: What Do Twitter Users Say about Software?," *2016 IEEE 24th International Requirements Engineering Conference (RE)*, pp. 96-105, 2016.

[4] Romero, D. M., Galuba, W., Asur, S., & Huberman, B. A. (2011, September). Influence and passivity in social media. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 18-33). Springer, Berlin, Heidelberg.

[5] Riquelme, F. and P. González-Cantergiani, *Measuring user influence on Twitter: A survey.* Information Processing & Management, 2016. **52**(5): p. 949-975.

[6] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: quantifying influence on twitter," in *WSDM '11*, 2011.

[7] E. Guzman, R. Alkadhi, and N. Seyff, "A Needle in a Haystack: What Do Twitter Users Say about Software?," *2016 IEEE 24th International Requirements Engineering Conference (RE)*, p. 96-105.

[8] Agarwal, N., Liu, H., Tang, L., & Yu, P. S. (2008, February). Identifying the influential bloggers in a community. In Proceedings of the 2008 international conference on web search and data mining (pp. 207-218).

[9] Peng, S., et al., *Influence analysis in social networks: A survey.* Journal of Network and Computer Applications, 2018. **106**: p. 17-32.

[10] Arrami, S., Oueslati, W., & Akaichi, J. (2018, May). Detection of opinion leaders in social networks: a survey. In International conference on intelligent interactive multimedia systems and services (pp. 362-370). Springer, Cham.

[11] S. Pei, J. Wang, F. Morone, and H. A. Makse, "Influencer identification in dynamical complex systems," *Journal of Complex Networks*, vol. 8, p. cnz029, 2020.

[12] Jianqiang, Z., G. Xiaolin, and T. Feng, *A New Method of Identifying Influential Users in the Micro-Blog Networks.* IEEE Access, 2017. **5**: p. 3008-3015.

[13] Pudjajana, A.M., et al., *Identification of Influencers in Social Media using Social Network Analysis (SNA).* 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2018: p. 400-404.

[14] Tridetti, S. (2016). Social network analysis: detection of influencers in fashion topics on Twitter. (Unpublished master's thesis). University of Liège, Liège, Belgium. Retrieved from https://matheo.uliege.be/handle/2268.2/1348

[15] Gentry J. (2011). Package 'twitteR'. Available on line in http://cran.r-project.org/web/packages/twitteR/twitteR.pdf

[16] Bavelas, A. (1950). Communication patterns in task-oriented groups. Journal of the acoustical society of America.

[17] Beauchamp, M.A., *An improved index of centrality.* Behavioral science, 1965. **10**(2): p. 161-163.

[18] Brandes, U., *A faster algorithm for betweenness centrality.* The Journal of Mathematical Sociology, 2001. **25**(2): p. 163-177.

[19] Bonacich, P., *Some unique properties of eigenvector centrality.* Social networks, 2007. **29**(4): p. 555-564.

[20] Erlandsson, F., et al., Finding Influential Users in Social Media Using Association Rule Learning. ArXiv, 2016. **abs/1604.08075**.

[21] Erlandsson, F., et al., *Crawling Online Social Networks.* 2015 Second European Network Intelligence Conference, 2015: p. 9-16.

[22] Nia, R.; Erlandsson, F.; Bhattacharyya, P.; Rahman, M.R.; Johnson, H.; Wu, S.F. Sin: A platform to make interactions in social networks accessible. In Proceedings of the 2012 International Conference on Social Informatics (SocialInformatics), Washington, DC, USA, 14–16 December 2012; pp. 205–214.

[23] Azcorra, A., et al., *Unsupervised Scalable Statistical Method for Identifying Influential Users in Online Social Networks.* Sci Rep, 2018. **8**(1): p. 6955.

[24] Dwivedi, Y., Arora, A., Bansal, S., Kandpal, C., & Aswani, R. (2019). Measuring social media influencer index- insights from facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services, 49*, pp. 86-101. doi:10.1016/j.jretconser.2019.03.012

[25] D. F. Andrews."A Robust Method for Multiple Linear Regression", Technometrics, Vol. 16, 1974, pp. 523-531.

[26] Quan, T.-T., D.-T. Mai, and T.-D. Tran, *CID: Categorical Influencer Detection on microtext-based social media.* Online Information Review, 2020. **44**(5): p. 1027-1055.

[27] Vitale, D., Ferragina, P. and Scaiella, U. (2012), "Classification of short texts by deploying topical annotations", Proceeding of European Conference on Information Retrieval, Springer, Berlin, Heidelberg, pp. 376-387.

[28] Phan, X., Nguyen, C., Le, D., Nguyen, L., Horiguchi, S. and Ha, Q. (2011), "A hidden topic-based framework toward building applications with short web documents", Proceeding of IEEE Transactions on Knowledge and Data Engineering, Vol. 23 No. 7, pp. 961-976. doi: 10.1109/ TKDE.2010.27.

[29] Bouma, G. (2009), "Normalized (pointwise) mutual information in collocation extraction", Proceedings of the Biennial GSCL Conference 2009, Germany, pp. 31-40.

[30] Huynh, T., et al., Detecting the Influencer on Social Networks Using Passion Point and Measures of Information Propagation †. Sustainability, 2020. **12**(7): p. 3064.

[31] Zimmerman, J.; Ng, D. Social Media Marketing All-in-One, 4th ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2017.

[32] Aleahmad, A., Karisani, P., Rahgozar, M. & Oroumchian, F. 2016. 'OLFinder: Finding opinion leaders in online social networks', Journal of Information Science, vol. 42, no. 5, pp. 659-674.

[33] Askarizade, M. and B. Tork Ladani, *Soft rumor control in social networks: Modeling and analysis.* Engineering Applications of Artificial Intelligence, 2021. 100: p. 104198.

[34] Li, Y.-M. and C.-W. Chen, A synthetical approach for blog recommendation: Combining trust, social relation, and semantic analysis. Expert Systems with Applications, 2009. **36**(3): p. 6536-6547.

[35] dataset, P., *PHEME dataset of rumours and non-rumours. figshare. Dataset.*, A.W.S.H. Zubiaga, Geraldine; Liakata, Maria; Procter, Rob, Editor. 2016.

[36] Ezenwafor, E.C., Olise, C.M. & Ebizie, P. I. (2021). Social media influencer and purchase intention amongst social media users in developing African economy. Quest Journal of Management and Social Sciences, 3(2), pp. 217-228.

[37] Hair, E., et al., Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. Early Childhood Research Quarterly, 2006. **21**(4): p. 431-454.