How to Measure Influence in Social Networks?

Ana Carolina Ribeiro(✉) , Bruno Azevedo , Jorge Oliveira e Sá ,
and Ana Alice Baptista

Centro ALGORITMI, University of Minho, 4800-058 Guimarães, Portugal
id8730@alunos.uminho.pt, brunomiguelam@engagelab.org,
{jos,analice}@dsi.uminho.pt

Abstract. Today, social networks are a valued resource of social data that can be used to understand the interactions among people and communities. People can influence or be influenced by interactions, shared opinions and emotions. However, in the social network analysis, one of the main problems is to find the most influential people. This work aims to report on the results of literature review whose goal was to identify and analyse the metrics, algorithms and models used to measure the user influence on social networks. The search was carried out in three databases: Scopus, IEEEXplore, and ScienceDirect. We restricted published articles between the years 2014 until 2020, in English, and we used the following keywords: social networks analysis, influence, metrics, measurements, and algorithms. Backward process was applied to complement the search considering inclusion and exclusion criteria. As a result of this process, we obtained 25 articles: 12 in the initial search and 13 in the backward process. The literature review resulted in the collection of 21 influence metrics, 4 influence algorithms, and 8 models of influence analysis. We start by defining influence and presenting its properties and applications. We then proceed by describing, analysing and categorizing all that were found metrics, algorithms, and models to measure influence in social networks. Finally, we present a discussion on these metrics, algorithms, and models. This work helps researchers to quickly gain a broad perspective on metrics, algorithms, and models for influence in social networks and their relative potentialities and limitations.

Keywords: Influence metrics · Influence analysis · Social networks analysis

1 Introduction

Networks are one of the fundamental structures of our complex systems. In the evolution of our cultural information systems, networks are a ubiquitous way to represent the dynamics of economic and social systems [1–4].

The Web allowed simultaneously the exponential production and spreading of digital information. Users are "prosumers", meaning that they are simultaneous interchangeably producers and consumers of information [5]. Social networks exponentially increased the number of social actors that create a wide number of connections forming a vast structure of links between actors and other entities (e.g. documents, messages, posts, recommendations) [6].

The growing use of social networks has attracted many researchers, academics, and organizations to explore social network research topics, including the influence analysis [7]. Influence analysis and its spread on social networks have an important application value [8] by allowing to analyse and explain people's social behaviors. It also provide a theoretical basis for decision making [9]. However, there are still some challenges to work on [8]: there is no mathematical formula of influence; it is difficult to identify the parameters to measure the influence; and, the large amount of data generated by social networks, makes it difficult to analyse and, consequently, to determine the influence.

An influence analysis study covers the study of influence properties such as influence evaluation metrics and algorithms, influence maximization, and social data collection and big data analysis [10].

This paper falls within the scope of the project 6,849.32 New Scientific Journal Articles Everyday: Visualize or Perish! [11] and the main objective of this work is to identify and analyse the most relevant and metrics, algorithms and/or influence models currently available.

The articles' search and selection process was based on the recommendations given by [12] complemented by [13] and was following:

1. Search engines and databases: Scopus, IEEEXplore, and ScienceDirect;
2. Time constraints: January 2014 to January 2020;
3. Keywords: social networks analysis, influence, algorithms, metrics, and measurements;
4. Types of documents: reviews, journals and conference papers;
5. Languages: English;
6. Selection criteria: Iterative process where the titles, abstracts and parts of the articles were reviewed for inclusion/exclusion.

The search resulted in 12 articles. The backward process was applied to these articles, which resulted in an addition of 13 articles, totalling 25 articles.

The main contributions of this work are briefly summarized below:

1. A methodology sufficiently detailed to allow the analysis of this study by other reliable researchers and use this study as a basis for future research into the influence on social networks.
2. An overview of the most relevant and up-to-date metrics, algorithms and/or models in social networks: 21 metrics, 4 algorithms, and 8 models of influence analysis.

The remaining of this article is organized as follows: the section "Methodological Procedure" presents the methodology applied for the selection of articles; the section "Related Work" presents some the related works that analyse the algorithms, metrics, and models of influence; the section "A landscape of influence in social networks" aims to present the overview of the metrics, algorithms, and influence models, the section "Discussion" presents the discussion of the results obtained, and section "Conclusions and Future Work" presents the conclusions, limitations, and some future research directions.

2 Methodological Procedures

This section reports the methodological procedures applied and that were based on [12]. Figure 1 represents all stages of the process.
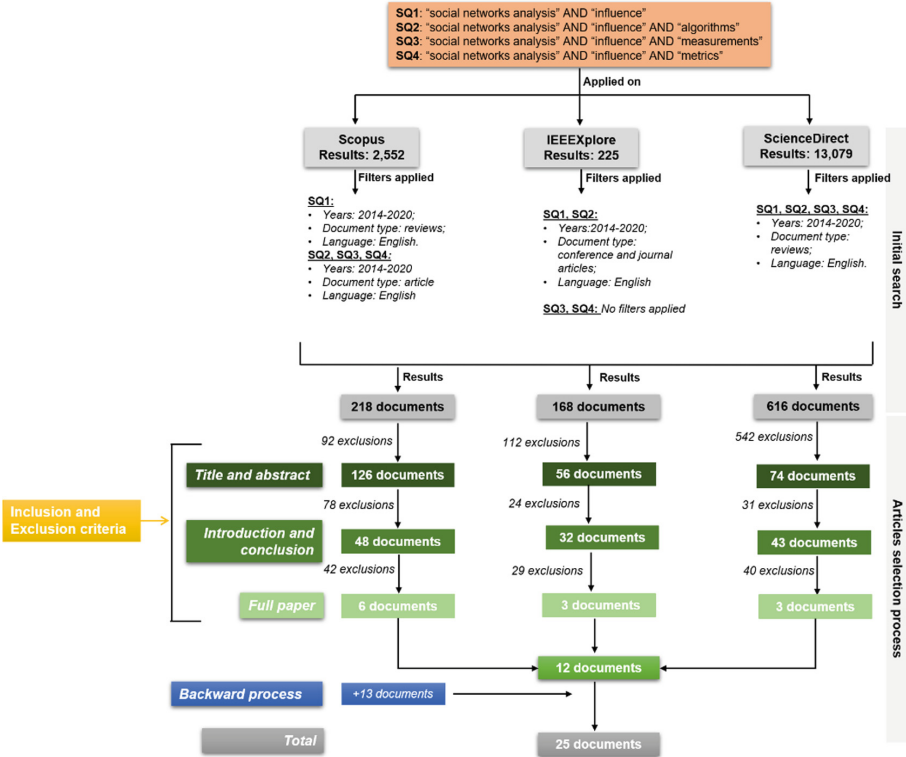

Fig. 1. Methodology of literature review.

2.1 Initial Search

The initial search starts with the selection of three databases: Scopus, IEEEXplore, and ScienceDirect. These databases have wide coverage of articles related to the topic and allow to filter the results: according to [14], in the social sciences, the coverage of Scopus is much higher than that of the Web of Science; The percentage of titles covered only by Scopus is above 60%, to which is added the almost 40% coverage overlap (Scopus and WoS), with WoS alone covering a very small percentage of titles; Sources indexed only by WoS are not necessarily disposable, however, it is safe to use only Scopus. IEEEXplore and Science Direct were used because they are widely used databases in the area of information systems, as a cross-check measure with Scopus results.

The keywords used were "social networks analysis", "influence", "algorithms", "measurements", and "metrics". In the initial search, we applied four search queries

(SQ) (Fig. 1) with the following results: Scopus 2,552 articles, IEEEXplore 225 articles, and ScienceDirect 13,079 articles, totalling 15,856 articles.

Considering these values, we used filters to get an acceptable number of results to analyse articles for all search queries: articles published from January 2014 to January 2020, conferences or journals or reviews, and written in English. However, according to the results obtained, it was necessary to adapt these filters for some search queries applied in some digital libraries, namely:

- For search query 1 on Scopus, we applied a different filter concerning the document type: we selected reviews because the values collected in the initial search were very high (2,121 articles). The reviews were selected because this type of articles describe, analyse, and discuss scientific knowledge already published.
- For search queries 3 and 4 applied on IEEEXplore, the values collected were low (6 and 9, respectively), and the application of filters was not necessary.

After applying the filters, all articles collected from Scopus, IEEEXplore, and ScienceDirect will be analysed in the next section.

2.2 Articles Selection Process and Results

After the articles collected in the previous phase, in this phase, all articles will be analysed according to the inclusion and exclusion criteria (Table 1), in parallel with the three phases described below

1. Title and abstract: Articles were selected if the title and abstract were aligned with the research objectives;
2. Introduction and conclusion: The introduction and conclusion of the articles accepted in phase 1 were analysed to proceed to a new selection;
3. Full article: The articles accepted in phase 2 were then fully read and subset was selected to be included in the review.

This process allowed the selection of 12 articles. We then applied a backward process were the references of the twelve previously selected articles were analysed. The implementation of the backward process resulted in the addition of 13 articles. The backward process, which allowed the identification of the most used metrics, algorithms and models to measure the influence on social networks, worked as a complemented the selection process, by allowing to gain a broader perspective on the topic. In total, 25 articles were collected and analysed.

Table 1. Inclusion and Exclusion criteria.

Inclusion criteria	Exclusion criteria
Articles about algorithms or metrics of social networks analysis	Articles not using metrics and/or influence algorithms (mention only metrics used in the social networks analysis, but not oriented to the analysis of influence)
Articles about the algorithms or metrics to computing influence on a social network	Articles focused on the influence that social networks have on people's lives, education, family life and, in general, on society
Articles about the algorithms or metrics to computing influence maximization on a social network	
Articles about the algorithms or metrics to computing influence diffusion on a social network	
Articles about the algorithms or metrics to computing influence applied on a social network	

3   Related Work

In this section, are reviewed the related works that analyse the algorithms, metrics, and models of influence.

The work reported in [15], presents a research on the latest generation of models, methods, and aspects of evaluation associated to influence analysis and provides a comprehensive analysis, helps to understand social behaviours, provides a theoretical basis to influence public opinion and reveal future directions of research and possible applications. The authors distinguish models in two types: microscopic (linear threshold, independent cascade, etc.) and macroscopic (epidemic models are the most common). The authors consider that, in the future, the microscopic models should concentrate on considering human interactions and different mechanisms during the information diffusion, while the macroscopic models consider the same probability of transmission and identical influential power for all users.

Differently, the authors of [8], present the state of the art on the influence analysis on social networks, presenting an overview of social networks, an explanation on the influence analysis at different levels, as a definition, properties, architecture and diffusion models, discuss the assessment metrics for influence and summarize the models for evaluating influence on social networks. In this work, the authors present some of the future trends in this topic that must be taken into account: the integration of cross-disciplinary knowledge due to the complexity of the topic; the development of an effective mechanism for influence analysis (hybrid approaches to improve the efficiency and effectiveness of

influence analysis) and an effective model for the efficiency and scalability of influence analysis.

The study [16] is also a relevant work because it focuses on the problem of predicting influential users on social networks. In this work, the authors present a three-level hierarchy that classifies the measures of influence: models, types, and algorithms. The authors also compare, based on empirical analysis, in terms of performance, precision, and correlation the measures of influence using a data set from two different social networks to verify the feasibility of measuring the influence. The results of the study show that the prediction of influential users does not depend only on the measures of influence, but also on the nature of social networks.

In the article [17], the authors study the probability of an individual being an influencer. They grouped the influence measures in some categories: measures derived from the neighbourhood (that is, number of influencers, personal exposure of the network), diversity structural, temporal measures, cascade measures, and metadata. Also, they evaluated how these measures relate to the likelihood that a user will be influenced using actual data from a microblog. Subsequently, the authors evaluated the performance of these measures when used as a resource in a machine learning approach and compared performance in a variety of supervised machine learning approaches. Finally, they evaluated how the proportion of positive to negative samples in training and testing affects the results of predictions - still allowing the practical use of these concepts for applications of influence.

4   A Landscape of Influence in Social Networks

This section starts by presenting the concept of influence on social networks and, some influence analysis applications, and their main properties. Also presented are the various metrics, algorithms and models found in the literature for influence analysis. For each metric, we present its definition and, in some cases, the calculation formula.

4.1   Understanding Influence in Social Networks

In social sciences, the term *influence* is widely used: according to [18], influence is *"The power to change or affect someone or something: the power to cause changes without directly forcing them to happen"*; and [19], *"social influence occurs when an individual's thoughts, feelings or actions are affected by other people"* p. 184.

A social network can be represented as a graph *G = (V, E),* where *V* corresponds to the nodes (vertices) in the graph (users), and *E* corresponds to the edges that indicate the relationship between users [20, 21]. According to [20] the relationship (edges) connects the influencer and influenced node, i.e., who influences whom. The edges' weights correspond to the influence probabilities among the nodes.

Marketing is one of the areas were influence analysis is most frequent. These specialists select a set of influential users and try to influence them to adopt a new behavior, product or service; Later, they expect these users to recommend to others, for example, by spreading word-of-mouth in the social networks [22]. In sentiment analysis, text mining tools and natural language processing to allow extract subjective information from

data sets of social networks, for example, users' opinions and attitudes. This makes, it possible to analyse the influence of users [23]. Another interesting application is the influence analysis of academics in their communities. High impact researchers are not necessarily influential [24, 25].

According to [7], influence has the following properties: A user's influence can increase or decrease with new experiences or interactions – dynamic nature. These new experiences or interactions can be more important, and the old ones can become irrelevant over time, i.e., the user can stop being influential at any time; In a social network, information can be propagated from one user to another, allowing the development of chains of influence - propagative nature; Influence has no mathematical definition or measure. Its subjective nature leads to the personalization of the calculation of influence, where the biases and preferences of influencers have a direct impact on its calculation.

To measure the influence on social networks, several metrics, algorithms, and models are known. These are grouped in the following categories:

- **Influence diffusion models –** Influence diffusion models measure the influence of users through their ability to spread information [16].
- **Centrality measures –** Centrality measures classify users according to their position on the network. Centrality measures the central position and importance of a user in a social network [16].
- **Influence measures based on walks between pair of users –** These types of measures provide relative power or status of user in a network by accounting all length paths between pair of nodes [26].
- **Link topological ranking measures –** According to [8], most centrality metrics do not consider the variation of nodes in their calculation: these metrics consider that all nodes contribute equally to their calculation. However, different types of nodes execute an important role in social networks.
- **Types of influence maximization algorithm –** Maximizing influence is a problem widely studied by the community. Influence maximization algorithms should perform fast calculations, high accuracy, and low storage capacity [15].
- **Others –** This category includes measures used by social networks such as Twitter to measure the influence of users [27].

4.2 Metrics and Algorithms Overview

In the category of Influence diffusion models we found the following models: Linear threshold model (LT model), Independent cascade model (IC model), Heat diffusion model (HD model), and Epidemic models (Table 2).

To apply the LT model and IC model, it is necessary to perform the Monte Carlo simulation to determine the influence of a node for a given period. However, the Monte Carlo simulation is time-consuming and inadequate for large-scale social networks [15]. The IC model is used to find highly influential users, find the maximum influence, predict the development of cascades, and understand the diffusion structure in the networks [20, 28]. Similar to the IC model, the LT model is mainly used to maximize the influence of propagation on the network.

Epidemic models are used to find the source of the viral disease and to find the sources of rumours. The epidemic disease in the population is similar to the spread of rumours on a social network [8, 28]. However, these models ignore the topological characteristics of social networks [15].

Table 2. Influence diffusion models.

Influence diffusion models	Description
LT model	In this model, a new idea, or innovation is adopted by a user u, only when a certain number of users influence that user u [8] In a social network G = (V, E), the sum of the influence weights of all neighbouring nodes of node v_i corresponds to: $$\sum_{v_j \in N_{g_i} act} w_{ij} \leq 1,$$ where w_{ij} corresponds to influence weights between node v_i and its neighbour node v_j, and $N_{g_i act}$ corresponds to the neighbouring nodes activated by node v_i [15]
IC model	The IC model describes the procedure of influence propagation in a probabilistic way: a user can influence (activate) his neighbour with a certain probability [8, 16]. The IC model is represented as follows [20]: • The initial seed set creates the active sets S_t for all t $\geq$ 1 using the following rule: at each phase t $\geq$ 1, the first activation step is considered from the set S_t para S_{t-1}; then, for each inactive node u, an activation attempt is performed using the Bernoulli test with a probability of success $p(u, v)$
HD model	There is a similarity between the heat diffusion and the information spread on social networks: a user selecting information acts as a source of heat, which diffuses his influence on the social network [8, 15]
Epidemic models	Epidemic models correspond to models capable of studying the influence of a macroscopic perspective [8]. According to [29], epidemic models are classified into three categories: deterministic models, stochastic models, and space-time models • Deterministic models include the susceptible-infectious model (SI model), the susceptible-infectious-susceptible model (SIS model), and the susceptible-infectious-recovery model (SIR model) • The stochastic epidemic model includes the discrete-time, continuous-time Markov model, and the stochastic differential equation model • The space-time models introduce automated cell phones to model the spread of influence

In the category of Centrality measures we found the following metrics: degree centrality, closeness centrality, betweenness centrality and, eigenvector centrality (Table 3).

Centrality metrics measure a user position in a social network, and the most used tools are graph theory and network analysis [8]. These metrics are used to find the most central and influential node in the network. The centrality metrics for finding the centrality of the node depend on the structural properties of the network and make use of flows to analyse these characteristics [16, 26, 28].

Table 3. Centrality measures.

Centrality measures	Description
Degree centrality	In a social network G = (V, E), degree centrality metric correspond to the number of neighbours of a node, that is, the number of edges that a node has [30–33] It is usually calculated by dividing the degree of a node (k_i) by $N - 1$, restricting the value in the range of [0, 1]. The equation that defines it is as follows: $C_D(i) = \frac{k_i}{N-1}$
Closeness centrality	In a social network G = (V, E), closeness centrality corresponds to the average length of the shortest path from one node to all other nodes [30–33]. In the influence analysis, this metric measures the efficiency of each node to disseminate information on the network [8] The equation that defines it is as follows: $C_C(i) = \frac{N-1}{\sum_{j \neq i}^{N} d_{ij}}$, Where, N is the number of nodes in the network and d_{ij} is the distance between node i and node j
Betweenness centrality	In a social network G = (V, E), betweenness centrality describes the extent of nodes that need to be crossed to influence other nodes [31–33] The equation that defines it is as follows: $C_B(i) = \sum_{s \neq i \neq t \in V, s < t} \frac{\sigma_{st}(i)}{\sigma_{st}}$, Where $\sigma_{st}(i)$ corresponds to the number of shortest paths between nodes s and t through the node i, and σ_{st} corresponds to the number of shortest paths between nodes s and t
Eigenvector centrality	In a social network G = (V, E), eigenvector centrality provides the relative scores for all nodes, according to the nodes connected to the highest scores contribute more to the scores of the nodes than to the lowest scores [32]. Eigenvector centrality use the adjacency matrix, given by: $C_E(i) = \frac{1}{\lambda} \sum_j A_{ij} C_E(j)$, Where, A_{ij} corresponds i^{th} eigenvector of the adjacency matrix in the network

In the category of Influence measures based on walks between pair of users we found the following metrics: Katz centrality, Hubbel measure, and Bonacich Power Measure (Table 4).

Table 4. Influence measures based on walks between pair of users.

Influence measures based on walks between pair of users	Description
Katz centrality	Katz centrality allows not only direct links received by a user but also popularity or status of users sending links to him to be included in his score. Further, the status of each, who has link with these users in turn, should also be used for calculating scores in social network [26]. The equation that defines it is as follows: $$\vec{C}_{Katz} = \left(\left(I - \alpha A^T \right)^{-1} - I \right) \vec{I},$$ Where, I is the identity matrix, $\vec{I}$ is a vector of size n (n is the number of nodes) consisting of ones. A^T denotes the transposed matrix of A and $\left(I - \alpha A^T \right)^{-1}$ denotes matrix inversion of the term $\left(I - \alpha A^T \right)$ Through Katz measure, most influential node or individual positive tie network can be found who has connections with most of the other users and can influence or affect other users with his decisions or activities [26]. This measure is similar to PageRank algorithm and eigenvector centrality
Hubbel measure	Hubbel measure corresponds to the flow of influence through interpersonal links in social networks as input and output channels. The Hubbel measure has structural as well as functional significance. The structural significance of index is in identifying cliques and functional significance is in computation of status [26] This measure is similar to Katz centrality, the Katz measure uses an identity matrix (each node is connected to itself) while the Hubble measure does not

(*continued*)

Table 4. (*continued*)

Influence measures based on walks between pair of users	Description
Bonacich power measure	In social networks, the most central user is not always the most powerful one In order to distinguish between power and centrality, was proposed a set of measures given by $c(\alpha,\beta)$. The parameter β is used to reflect the degree and direction (positive or negative) in which individual user status depends upon status of other users in network [26] Bonacich power measure is useful in valued and signed graphs, negative ties and positive ties networks

The Katz centrality can be used to compute centrality in directed networks (citation networks, WWW, etc.); it can also be used estimate the relative status or influence of user in a social network [8, 20, 34]. Hubbel Measure and Bonacich Power Measure are measures similar to Katz centrality.

In the category of Link topology ranking measures, were found the following metrics: Hyperlink-Induced Topic Search (HITS) algorithm and PageRank Algorithm (Table 5).

Except for eigenvector centrality, most centrality metrics do not consider the variation of the nodes, which means that they consider that all nodes contribute equally to the measures [8]. However, the types of nodes execute an important role in social networks. The HITS algorithm aims to classify web pages based on links, while in PageRank all hyperlinked pages receive numerical weights, used to measure the importance of web pages [27].

The HITS algorithm is used to classify publications in citations networks by Citeseer (search engine). In the context of citation networks, it is natural to identify topical reviews as hubs, as they contain many references to influential articles in the literature [34].

In category of Influence maximization algorithms were found the following algorithms: Greedy-based algorithms and Heuristic-based algorithms (Table 6).

According to the literature, the greedy-based algorithms have higher accuracy compared to the heuristic-based algorithms. This is because greedy-based algorithms have high computational complexity and high execution time, decreasing their efficiency [15]. Concerning the heuristic-based algorithms, these algorithms were proposed to reduce the execution time of the solution and increase efficiency. Also, they present higher values of accuracy [8].

Table 5. Link topology ranking measures.

Link topology ranking measures	Description
HITS algorithm	The HITS algorithm is a popular classification method based on eigenvector to classify web pages [34]. In a network, this algorithm selects two scores to each node: score h – referred to as node hub-centrality score – is large for nodes that point to many authoritative nodes, and score a – referred to as node authority-centrality score – is large for nodes that are pointed by many hubs [8, 34] • Authority-centrality score: In the algorithm it is necessary update each node's authority score to be equal to the sum of the hub scores of each node that point to it. A node is given a high authority score by being linked from pages that are recognized as Hubs for information • Hub-centrality score: A hub is a web page serving as a large directories with no actual authoritative content that it points to. In the HITS algorithm, a directory points to many authorities, and an authority is a page with many incoming links from different hubs. In the algorithm, it is necessary update each node's hub score to be equal to the sum of the authority scores of each node that it points to The corresponding equations for node a and h are [34]: $a = \alpha A h$ $h = \beta A^T a,$ Where, α and β are parameters of the method
PageRank algorithm	In PageRank algorithm, all hyperlinked pages are given weights, which are used to measure the importance of web pages [35]. PageRank algorithm can be applied to social networks analysis since the relationships of nodes in social networks can be structured like links [36] The PageRank algorithm is defined by the following equation: $PR(r) = \frac{1-\lambda}{N} + \lambda \sum_{i=1}^{k} \frac{PR(r_i)}{K_{out}(r_i)},$ Where N represents the total number of nodes in the network, K_{out} is the out-degree of the node r, r_i denotes th in-degree of node r and λ is the damping factor

In the category of other influence metrics and algorithms were found the following metrics and algorithms: Popularity measures on Twitter (FollowerRank, Popularity, Popularity paradoxical discounted, Network Score, Acquaintance Score, Acquaintance-affinity score, Acquaintance-Affinity-Identification Score), Traditional measure used on Twitter (h-index), Measures based on Twitter metrics and PageRank (Retweet Impact, Mention Impact, Social Networking Potential, ThunkRank, UserRank), Topical influential users (Information diffusion), and Predicting influences (Activity and Willingness

Table 6. Influence maximization algorithms.

Influence maximization algorithms	Description
Greedy-based algorithms	The study of greedy algorithms is based on hill-climbing greedy algorithm, in which each option can provide the highest value of the impact of the node used to the local optimal solution to approximate the global optimal solution [8] Some examples of greedy-based algorithms are present below: 1) Target wise greedy algorithm based on the potential-based node-selection strategy. This algorithm does not have good results in an initial phase, but it can cover more nodes in a later phase of diffusion [37] 2) Community-based greedy algorithm was proposed to reduce the cost in terms of execution time. It is based on the IC model [38] 3) Upper bound-based lazy forward algorithm has been proposed to discover top-k influential nodes. This algorithm sets new limits to significantly reduce the number of Monte Carlo simulations, particularly in the initial phase [39]
Heuristic-based algorithms	According to the computational complexity of the greedy-based algorithms, several heuristic algorithms have been proposed to reduce the solution time and obtain more efficiency of the algorithm. These algorithms select nodes iteratively based on a specific heuristic, instead of computing the marginal gain of the nodes in each iteration. In contrast, its accuracy is relatively low [15] A proposed algorithm was Two-phase Heuristic Algorithm (TPH). This algorithm is composed of two phases: each node has its offline probability of a given product; therefore, the consideration of local-based maximization cannot focus only on the network topology, but also on the offline property of each node [40]

of users (AWI) model, Activeness, centrality, quality of post and reputation (ACQR) Framework, Time Network Influence Model, AuthorRanking) (Table 7).

These metrics were defined to try to combine metrics involving tweets, replies, tweets, and mentions to obtain information about a social network using a numerical value [27]. According to [41], the metrics of retweets are the best quantitative indicators for choosing to read a tweet over the other. Besides this, the most important indicators are qualitative, for example, the friendship between the reader and the author of the tweet.

Table 7. Other influence metrics and algorithms.

Others	Description
Popularity measures on Twitter	**1) FollowerRank** – Corresponds to the standardized version of the in-degree measure. $FollowerRank(i) = \frac{Number\ of\ followers}{Number\ of\ followers+Number\ of\ followees}$ **2) Popularity** – This measure was developed to mitigate differences in followers between users $Popularity(i) = 1 - e^{\lambda.Number\ of\ followers}$, Where, λ it is a constant that, by default, is equal to 1 **3) Popularity paradoxical discounted** – Corresponds to the number of reciprocal actors of a user, that is, the number of followers who are also followed $Pradoxixal\ discounted(i) =$ $$\begin{cases} \frac{N^{\underline{o}}\ of\ followers}{N^{\underline{o}}\ of\ followees} & if\ N^{\underline{o}}\ of\ followers > N^{\underline{o}}\ of\ followees \\ \frac{N^{\underline{o}}\ of\ followers-reciprocal(i)}{N^{\underline{o}}\ of\ followees-reciprocal(i)} & otherwise \end{cases}$$ Measuring the value of reciprocal value (i) considerably increases computational costs **4) Network Score (NS)** – Corresponds to a measure of popularity, based on the user's active non-reciprocal followers $NS(i) = \log(N^{\underline{o}}\ topically\ active\ followers + 1)$ $\qquad - \log(N^{\underline{o}}\ topically\ active\ followees + 1)$ **5) Acquaintance Score A(i)** – Measures how well-know user i is. Let n be the number of considered user accounts, it is defined as: $A(i) = \frac{N^{\underline{o}}\ of\ followers+UMA+URA+UPA}{n}$ Where, UMA = number of users mentioning the author, URA = number of users who have retweeted author's tweets, and UPA = number of users who have replied author's tweets **6) Acquaintance-Affinity Score AA(j)** – Measures how dear user j is, by considering how well know are those who want him $AA(j) = \sum\limits_{i \in E_{RP}} A(i).\frac{\#replies\ of\ i\ to\ j}{\#replies\ of\ i} + \sum\limits_{i \in E_M} A(i).\frac{\#mentions\ of\ i\ to\ j}{\#mentions\ of\ i} +$ $\sum\limits_{i \in E_{RT}} A(i).\frac{\#retweets\ of\ i\ to\ j}{\#retweets\ of\ i}$ Where ERP, EM, and ERT are the set of users who reply, mention and retweet the tweets of j, respectively **7) Acquaintance-Affinity-Identification Score AAI(j)** - Measures how identifiable user j is, by considering how dear those who identify him $AAI(j) = \sum_{i \in F_r} \frac{AA(i)}{\#followees\ of\ i}$, Where, Fr is the set of followers of j. The AAI Score is well correlated with the number of followers and was used to identify celebrities in the "real world", i.e., outside the Twitter network

(continued)

Table 7. (*continued*)

Others	Description
Traditional measure used on Twitter	**h-index** – In the context of Twitter, it can be defined as the maximum value h such that h tweets of the user have been replied, retweeted, or liked, at least h times
Measures based on Twitter metrics and PageRank	**1) Retweet Impact (RI)** – Estimates the impact of the user tweets, in terms of the mentions received by other users $RI(i) = RT2 * \log(RT3)$, Where RT2 = Number of original tweets posted by the author and retweeted by other users, RT3 = Number of users who have retweeted author's tweets **2) Mention Impact (MI)** – Estimates the impact of the user tweets, in terms of the mentions received by other users $MI(i) = M3 * \log(M4) - M1 * (M2)$, Where M1 = Number of mentions to other users by the author, M2 = Number of users mentioned by the author, M3 = Number of mentions to the author by other users, M4 = Number of users mentioning the author **3) Social Networking Potential (SNP)** – Measure considers all kind of actions on Twitter, except the favorites or likes $SNP(i) = \frac{Ir(i)+RMr(i)}{2}$, Where the Interactor Ratio, Ir(i), and the Retweet and Mention Ratio, RMr(i), are defined as: $Ir(i) = \frac{RT3+M4}{F1}$ and $RMr(i) = \frac{\#tweets\ of\ i\ retweetes+\#tweets\ of\ i\ replied}{\#tweets\ of\ i}$, Where RT3 = Number of users who have retweeted author's tweets, M4 = Number of users mentioning the author, F1 = Number of followers **4) ThunkRank** – Direct adaptation of PageRank algorithm into the context of Twitter $TunkRank(i) = \sum_{j \in followers(i)} \frac{1+p.TunkRank(j)}{\#followees\ of\ j}$, Where $0 \leq p \leq 1$ is the probability that a tweet is retweeted. This probability is assumed to be equal for all users. In the literature, normally use p = 0.5, but in fact this value should vary from case to case **5) UserRank** – A variation of ThunkTank, defined to measure the influence of a user according to the relevance of his tweets $UserRank(i) = \sum_{j \in followers(i)} \frac{1+\frac{\#followers\ of\ i}{\#tweets\ of\ i}*UserRank(j)}{\#followers\ of\ j}$
Topical influential users	**Information diffusion** – Estimates the possible influence of the users's tweets among his followers who are non-followees $ID(i) = \log(N^{\underline{o}}\ of\ followers\ tweeting\ on\ topic\ after\ the\ author + 1)$ $- (N^{\underline{o}}\ of\ followees\ tweeting\ on\ topic\ before\ the\ author + 1)$ The "+1" in the logarithms avoids divisions by zero. This measure only considers follow-up relashionships, but it is independent of the number of followers and followees on the network

(*continued*)

Table 7. (*continued*)

Others	Description
Predicting influences	**1) Activity and Willingness of users (AWI) model** – AWI model is a user interaction model that considers the activity and willingness of users to retweet through time, in order to measure the influence among pairs of users. This model also predicts retweet ratios and influential users **2) Activeness, centrality, quality of post and reputation (ACQR) Framework** – This framework uses data mining to detect activity (original tweets, retweets and replies), centrality, and user reputation (mechanism to distinguish between real users and spammers). It also considers the quality of tweets through the number of replies and retweets, and the reputation of users that reply and retweet. ACQR framework was used to identify and predict the influential users in a relatively small network that was restricted to a specific topic **3) Time Network Influence Model** – Uses a probabilistic generative model to make an offline estimation of the influence power between users. This model considers the time intervals between messages, follow-up relationships, and the relationships of similarity in the content of the tweets **4) AuthorRanking** – Uses the style of the tweets (words, hastags, websites, references to other accounts) and user behavior (profile information, following ratios, number of tweets, and main user activity, previously determined by a text classification task)

5   Discussion

The growing development of social networks has also allowed the production of large amounts of information that can tell who the most influential users are. To try to solve this problem were developed algorithms, metrics, and models to compute the influence of a user on social networks [8]. For this reason, this work presents an extended set of several algorithms, metrics, and models and their applicability found in the literature.

One of the main problems of some metrics, algorithms, and models detected in the literature is the scalability-efficiency capacity [8, 15]. Also, with the continued increase of social networks, most existing methods find the problem of efficiency in runtime, and it becomes difficult to implement them in a large-scale context.

The literature argues that the application of the LT model and the IC model is time-consuming and unsuitable for large-scale networks [20, 28]. Also, greedy-based algorithms present high computational complexity and high execution time, decreasing their efficiency [15]. Other algorithms such as heuristic-based algorithms have been developed to reduce these execution times and, consequently, increase their efficiency [8].

The diversity of metrics, algorithms, and models of influence analysis is due to the need to solve several types of problems: influence maximization [15], the influence diffusion [16], the distinction of the importance of the various nodes in a social network, among others. Centrality measures are the best known and most used in the social networks analysis, but to be used in the analysis of the most influential node, they are dependent on the properties of the networks [15]. The metrics that fall into the Others category are very interesting: the investigators used quantitative measures such as tweets, retweets or mentions to obtain a numerical value and thus be able to classify the user as influential or not [27].

It is important to consider the objectives of the problem and the type of data in hands in order to be able to apply the most appropriate set of metrics to obtain the greatest possible precision of the influence.

6   Conclusions and Future Work

As mentioned at the beginning of the article, influence analysis is one of the biggest problems in social networks analysis. Therefore, the main objective of this literature review was to identify and analyse the most relevant metrics, algorithms, or models to measure the influence on social networks. Also, methodological limitations were recognized and should be refined in future work, namely:

- The article selection process for literature review was performed by only one researcher. This may affect the results because articles were select according to perspective of a single researcher. Recommendation: This phase should be conducted in parallel with other researchers to reduce error and bias in article selection. The usage of social networks (Twitter, Facebook, etc.) may also support the research allowing identify the perspective of other researchers and get new research outputs faster;
- Since Scopus only used reviews, several important studies may have been missed. As future work, a meta-analysis of the reviews must be made. Thus, it will be possible to complement the work with a review of what was produced after the last review analysed.
- Only 3 databases were used – Scopus, IEEEXplore, and ScienceDirect. Recommendation: although these databases have high coverage of scientific articles, other sources (SpringerLink, Web of Science, scientific journals, and social networks) may complement the research.
- The keywords used in search queries can be improved, including new keywords, changing their order and combination to cover more works. For example, "social network", "social networks influence analysis", "models", "social media", "social media platforms", etc.

In this article, was reported a study of influence and respective the metrics, algorithms, and models used for its analysis their challenges and opportunities. Through this search and the analysis of the articles, it was possible to collect 21 metrics, 4 types of algorithms, and 8 models of influence analysis.

The metrics, algorithms, and models of influence found in the literature allowed us to obtain a broad view of this topic: the LT model and the IC model are the most

time-consuming and inappropriate models for large-scale networks; the greedy-based algorithms are considered very complex and time-consuming to implement; and the centrality measures are the most well-known measures and the measures based on indicators such as tweets, retweets, and mentions should be deepened to understand how they can contribute when used in conjunction with other types of metrics. Also, as the metrics of Twitter were analysed, metrics from other social networks (for example, Facebook) should be analysed and compare for existing differences; if they can be adapted to other social networks, since it depends on the organization of the social network and the types and numbers of resources it has.

However, it is necessary to consider that, in addition to these metrics, algorithms, and models, other measures should be studied due to their potential in the influence analysis.

Several challenges and opportunities may stimulate, in the future, new theoretical and practical perspectives. This article may serve as a basis for researchers interested in measuring the influence on social networks as they can gain a broad perspective on the topic.

References

1. Newman, M., Watts, D.J., Barabási, A.-L.: The Structure and Dynamics of Networks. Princeton University Press, Princeton (2006)
2. Wright, A.: Glut: Mastering Information Through the Ages. Cornell University Press, Ithaca (2008)
3. Castellano, C., Fortunato, S., Loreto, V.: Statistical physics of social dynamics. Rev. Mod. Phys. **81**(2), 591–646 (2009)
4. Hidalgo, C.A.: Disconnected, fragmented, or united? A trans-disciplinary review of network science. Appl. Netw. Sci. **1**(1), 6 (2016)
5. Wurman, R.S.: Information Anxiety 2, 2nd edn. QUE (2001)
6. Hansen, D., Shneiderman, B., Smith, M.: Analyzing Social Media Networks with NodeXL: Insights from a Connected World, 1st edn. Morgan Kaufmann (2010)
7. Peng, S., Wang, G., Xie, D.: Social influence analysis in social networking big data: opportunities and challenges. IEEE Netw. **31**(1), 11–17 (2017)
8. Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., Jia, W.: Influence analysis in social networks: a survey. J. Netw. Comput. Appl. **106**(January), 17–32 (2018)
9. Yu, S., Liu, M., Dou, W., Liu, X., Zhou, S.: Networking for big data: a survey. IEEE Commun. Surv. Tutor. **19**(1), 531–549 (2017)
10. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 11, pp. 137–146 (2003)
11. Azevedo, B.M., Oliveira e Sá, J., Baptista, A.A., Branco, P.: Information visualization: conceptualizing new paths for filtering and navigate in scientific knowledge objects. In: 2017 24o Encontro Português de Computação Gráfica e Interação (EPCGI), pp. 1–8 (2017)
12. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a review. MIS Q. **26**(2), 13 (2002)

13. Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., Linkman, S.: Systematic literature reviews in software engineering - a systematic literature review. Inf. Softw. Technol. **51**(1), 7–15 (2009)
14. Mongeon, P., Paul-Hus, A.: The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics **106**(1), 213–228 (2016)
15. Li, K., Zhang, L., Huang, H.: Social influence analysis: models, methods, and evaluation. Engineering **4**(1), 40–46 (2018)
16. Almgren, K., Lee, J.: An empirical comparison of influence measurements for social network analysis. Soc. Netw. Anal. Min. **6**(52), 1–18 (2016)
17. Kumar, N., Guo, R., Aleali, A., Shakarian, P.: An empirical evaluation of social influence metrics (2016)
18. Merriam, W.: Definition of influence. In: Definition of influence (2011)
19. Li, H., Cui, J.-T., Ma, J.-F.: Social influence study in online networks: a three-level review. J. Comput. Sci. Technol. **30**(1), 184–199 (2015)
20. More, J.S., Lingam, C.: A gradient-based methodology for optimizing time for influence diffusion in social networks. Soc. Netw. Anal. Min. **9**(1), 5 (2019)
21. Jalayer, M., Azheian, M., Agha Mohammad Ali Kermani, M.: A hybrid algorithm based on community detection and multi attribute decision making for influence maximization. Comput. Ind. Eng. **120**, 234–250 (2018)
22. Liqing, Q., Jinfeng, Y., Xin, F., Wei, J., Wenwen, G.: Analysis of Influence Maximization in large-Scale Social Networks. IEEE Access **7**(4), 42052–42062 (2019)
23. Li, D., Shuai, X., Sun, G., Tang, J., Ding, Y., Luo, Z.: Mining topic-level opinion influence in microblog. In: ACM International Conference Proceeding Series, pp. 1562–1566 (2012)
24. Li, N., Gillet, D.: Identifying influential scholars in academic social media platforms. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 608–614 (2013)
25. Kong, X., Shi, Y., Yu, S., Liu, J., Xia, F.: Academic social networks: modeling, analysis, mining and applications. J. Netw. Comput. Appl. **132**, 86–103 (2019)
26. Kaur, M., Singh, S.: Analyzing negative ties in social networks: a survey. Egypt. Inform. J. **17**(1), 21–43 (2016)
27. Riquelme, F., González-Cantergiani, P.: Measuring user influence on Twitter: a survey. Inf. Process. Manag. **52**(5), 949–975 (2016)
28. Shelke, S., Attar, V.: Source detection of rumor in social network – a review. Online Soc. Netw. Media **9**, 30–42 (2019)
29. Peng, S., Yu, S., Yang, A.: Smartphone malware and its propagation modeling: a survey. IEEE Commun. Surv. Tutor. **16**(2), 925–941 (2014)
30. Russell Bernard, H.: The development of social network analysis: a study in the sociology of science. Soc. Netw. **27**(4), 377–384 (2005)
31. Borgatti, S.P.: Centrality and network flow. Soc. Netw. **27**(1), 55–71 (2005)
32. Frantz, T.L., Cataldo, M., Carley, K.M.: Robustness of centrality measures under uncertainty: examining the role of network topology. Comput. Math. Organ. Theory **15**(4), 303–328 (2009)
33. Kosorukoff, A.: Theory. In: Social Network Analysis - Theory and Applications, pp. 1–4 (2011)
34. Liao, H., Mariani, M.S., Medo, M., Zhang, Y.C., Zhou, M.Y.: Ranking in evolving complex networks. Phys. Rep. **689**, 1–54 (2017)
35. Brin, S., Page, L.: Reprint of: the anatomy of a large-scale hypertextual web search engine. Comput. Netw. **56**(18), 3825–3833 (2012)
36. Liu, Q., et al.: An influence propagation view of PageRank. ACM Trans. Knowl. Discov. Data **11**(3), 2–28 (2017)

37. Wang, Y., Feng, X.: A potential-based node selection strategy for influence maximization in a social network. In: Proceedings of the 5th International Conference on Advanced Data Mining and Applications (ADMA 2009), pp. 350–361 (2009)
38. Wang, Y., Cong, G., Song, G., Xie, K.: Community-based greedy algorithm for mining top-k influential nodes in mobile social networks categories and subject descriptors. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1039–1048 (2010)
39. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., Vanbriesen, J.: Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 420–429 (2007)
40. Zhou, T., Cao, J., Liu, B., Xu, S., Zhu, Z., Luo, J.: Location-based influence maximization in social networks. In: CIKM 2015, no. 93, pp. 1211–1220 (2015)
41. Chorley, M.J., Colombo, G.B., Allen, S.M., Whitaker, R.M.: Human content filtering in Twitter: the influence of metadata. Int. J. Hum Comput. Stud. **74**, 32–40 (2015)