Extent prediction of the information and influence propagation in online social networks

Raúl M. Ortiz-Gaona[1,2] · Marcos Postigo-Boix[1] · José L. Melús-Moreno[1]

Abstract
We present a new mathematical model that predicts the number of users informed and influenced by messages that are propagated in an online social network. Our model is based on a new way of quantifying the tie-strength, which in turn considers the affinity and relevance between nodes. We could verify that the messages to inform and influence, as well as their importance, produce different propagation behaviors in an online social network. We carried out laboratory tests with our model and with the baseline models *Linear Threshold* and *Independent Cascade*, which are currently used in many scientific works. The results were evaluated by comparing them with empirical data. The tests show conclusively that the predictions of our model are notably more accurate and precise than the predictions of the baseline models. Our model can contribute to the development of models that maximize the propagation of messages; to predict the spread of viruses in computer networks, mobile telephony and online social networks.

Keywords Influence diffusion · Information diffusion · Influence threshold · Information threshold · Online social networks · Social tie-strength

This work was carried out by the financing of the Ecuadorian government of President Rafael Correa Delgado and supported in part by the Spanish Government (MINECO) and Fondo Europeo de Desarrollo Regional (FEDER) by means of the ADVICE (TEC2015-71329) project.

1. INTRODUCTION

Online social networks (OSNs) are increasingly used by many sectors of society for the purpose of informing and influencing people. These sectors are governments, social movements, providers of goods and services, etc. This situation is generating large volumes of information in Internet (Jin et al. 2013), and is playing an important role in society (Abbas 2013; Yun and Gloor 2015). They need to predict the extent, amplitude or distance of the spread of information and influence they transmit through these networks. One way to achieve this is to study this phenomenon through *Social Network Analysis* (SNA), which is currently used by many researchers (Kosorukoff and Passmore 2011; Yang et al. 2016; Scott 2017; Ito et al. 2018; Bandeli and Agarwal 2018), and it is the one that we have used in this research.
SNA is a research method that brings together a set of concepts and metrics, such as influence, centrality metrics, *threshold*, *social tie strength* and *homophily*, ideas that comes from sociology and anthropology. SNA uses mathematical tools, such as graph theory, probability and statistics, and computer science tools. SNA mathematically represents a social network as a graph, $G = (V,E)$, where V is a finite set of nodes and $E \subseteq VxV$ is a set of edges that connect pairs of nodes. Nodes represent individuals and edges represent relationships between individuals. Network analysis focuses on the association that exists between nodes rather than on attributes of those nodes. Within this perspective, *Linear Threshold Model* (*LTM*) and *Independent Cascade Model* (*ICM*) (Kempe et al. 2003) are two stochastic baseline models that predict the dissemination of information and influence through social networks. A model derived of *ICM* is *Weighted Cascade Model* (*WCM*) (Kempe et al. 2003). *LTM* and *ICM* are studied and used in many recent and important scientific works to optimize the propagation of information in social networks, some of them are the following (Arendt and Blaha 2015; Worrell et al. 2017; Bulumulla et al. 2018; Cui et al. 2018; da Silva et al. 2018; Fischetti et al. 2018; N. et al. 2018). For this reason, we have evaluated our proposal by comparing it with the *LTM*, *ICM*, *WCM* baseline models and with empirical results. Each of these three predictor models has its own criteria on which depends the extent of propagation.

✉ Raúl M. Ortiz-Gaona
raul.ortiz.gaona@gmail.com
Marcos Postigo-Boix
marcos.postigo@entel.upc.edu
José L. Melús-Moreno
jlmelus@entel.upc.edu

[1] Department of Telematics Engineering at the Polytechnic University of Catalonia (UPC), C/ Jordi Girona 1-3, Building C3, Campus Nord, Barcelona, E-08034, Spain.
[2] Faculty of Engineering at the Universidad de Cuenca, Cuenca, CP:010203, Ecuador

The Cambridge Dictionary defines the term "*inform*" as follows: "to tell someone about particular facts"; and the term "*influence*" is defined as "to affect or change how someone or something develops, behaves, or thinks". "*Inform*" and "*influence*" not only have different definitions but also they imply different processes among the people.
After studying the baseline models, *WCM* and related works, in order to fill the gaps of these models, we have developed from the ground the *Lucy Model* (*LM*). Our model improves both the representation of the message propagation process and the prediction of its extent in online social networks. This work does not pretend to establish a method to maximize the propagation.
LM has eight characteristics: it is predictive, computer numerical (nonanalytic), iterative, nonbased on historical data, stochastic, graph-oriented, time-independent, and customizable.
We identified three factors that are not considered in baseline models and their derivate models, which are: (1) Importance of the messages, (2) Existence of different classes of messages, and (3) We take the idea that the nodes are informed or influenced after a threshold is exceeded, but the threshold must be adapted to each node, according with its relevance or affinity, and with the importance and class of message. Not taking into account these factors causes the models to give results far from reality.
The different classes of messages that we consider are three: (1) Messages to inform, (2) Messages to influence which appeal to feelings and emotions, and (3) Messages to influence which appeal to interests and personal conveniences.
On the other hand, we consider that relevance of people and affinity between them affect the strength of interpersonal relationships called *tie strength*, consideration that is not made by existing models that quantify this strength. Based on this consideration, we propose a new model that represents and quantifies tie strength in online social networks. The relevance of people is quantified with centrality metrics (Newman 2010). Affinity is quantified with a metric that we also propose and that we will explain later. Affinity is an element not considered by the baseline models mentioned.
The method we have used to evaluate our proposal has the following steps. (1) We have created a simulation environment, implementing on computer the four predictive models. (2) We have tracked the social network of an anonymous *Facebook* user. (3) We have carried out simulations with the four predictive models, on the social network traced already indicated. (4) To obtain empirical data, we have used the *Facebook* platform to post messages in the social network of the anonymous user that was previously indicated, and we have observed how far they spread.
These tests show conclusively that the predictions of our proposal (*LM*) are notably more accurate and precise than the predictions of the baseline models and *WCM*.
In order to study the behavior of *Lucy Model*, we carry out additional laboratory tests, modifying its parameters and using a synthetic (artificial) *Power-Law* network (Newman 2010) and two social networks: *YouTube* (https://snap.stanford.edu/data/com-Youtube.hmtl) and the social network *Facebook* of a specific user.
LM may be suitable for use in different areas of human activity, for example*:* for predict the extent of the messages propagated in OSN by governments, humanitarian action, political parties, social movements, suppliers of goods and services, electoral and commercial campaigns, etc. Also, our model can be used as a basis to develop new models that maximize the propagation of messages (Wang et al. 2018), that prognostic the propagation of viruses both in computer networks (Piqueira and Araujo 2009), in online social networks (Fan and Yeung 2011; Luo et al. 2016), and in mobile telephony networks. Our model can also help predict the loss of clients in mobile telephony (Phadke et al. 2013).
This work is part of the topic *social cyber security*. (Carley et al. 2018), which is a new scientific area that has emerged to characterize and predict changes in both human (Luceri et al. 2019), social and political behavior (Robertson et al. 2019), governance (Paniagua et al. 2019) and national security (Mareswara Rao and Rajashekara Rao 2019). *LM* can be applied to prognostic the magnitude of the spread of misleading messages that actually lead to crimes (Zainudin et al. 2011) such as hate speech (Mathew et al. 2019), terrorism (Ishengoma 2013), pornography (Benevenuto et al. 2008), bullying (Kao et al. 2019), sexual harassment (Nova et al. 2019), prostitution and human trafficking (Ahmed et al. 2017); *LM* can also help predict the magnitude of a social outbreak such as the "Arab Spring" (Steinert-Threlkeld et al. 2015), fake news (Vishwakarma et al. 2019), disinformation (Bandeli and Agarwal 2018a), privacy violation (Kayes and Iamnitchi 2017) and frauds (Apte et al. 2019).
This paper is structured as follows: Section 2 describes the related works. Section 3 explains *Lucy Model* with its modules representing the propagation of nonhomophilic influence, homophilic influence, and propagation of information. Section 4 presents the validation method of *Lucy Model*. Section 5 shows validation results and analysis of *Lucy Model*. Section 6 contains the conclusions.

2. RELATED WORK

Currently, there are two types of predictive models of influence propagation: those focused on graphs, and those based on data from previous propagations. These late predictive models are used to predict the spread of diseases (Guille et al. 2013), and the propagation is represented by differential equations. The first work on the spread of diseases was published by Daniel Bernoulli in 1766 (Dietz and Heesterbeek 2002).
In this Section, we describe twelve models of propagation of influence, which have a direct relationship with *Lucy Model*. First we describe the two baseline models: *LTM* and *ICM*, which are fundamental in the creation of other models, some of which we will also describe. All these models are focused on graphs. In the following, to an informed or influenced node, we will call interchangeably "*active node*" or "*activated node*"

2.1 Baseline Models

2.1.1 Linear Threshold Model

This model is based on a threshold of influence for each node. A node *v* is influenced by each neighbor *w* according to a weight $b_{v,w}$ such as $\sum_{w\ neighbor\ of\ v} b_{v,w} \leq 1$. The process is as follows: each node *v* chooses a threshold Θ_v uniformly distributed in the interval [0, 1]. The threshold represents the weighted fraction of neighbors of *v* that must be active for *v* to become active. Given a random choice of thresholds, and an initial set of active nodes A_o (with all other nodes inactive), the diffusion process develops over time in discrete steps: all nodes that remained active in step *t - 1* are active in step *t*, and any node *v* is activated if the total weight of its active neighbors is at least Θ_v:

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

In this way, Θ_v represents the different latent tendencies of nodes to adopt innovation when their neighbors do; the fact that these are randomly selected is intended to model the lack of knowledge of those values. As it is observed, *LTM* is oriented to inactive nodes because it looks for the nodes that satisfy the activation condition.
LTM carries out successive iterations by visiting in each of them all the nodes of the graph, looking for inactive nodes whose weighted sum of tie strength with his active neighboring nodes is equal to or greater than his threshold. If at the end of the current iteration there is at least one new node activated, a new iteration is started; otherwise, the process ends. The authors of *LTM* do not specify if a node that was not influenced has a new opportunity of being influenced if it increases the number of active neighbors. We implemented *LTM* without considering this possibility because the resulting number of activated nodes was too large.
As an example, in Fig. 1 we show the inactive node *v* and its neighbors (*wi*), of which *w1*, *w4* and *w5* are active. The weights $b_{v,wi}$ of the edges of *v* with their active neighbors are represented by lines of different thickness. If the sum of these weights is at least equal to the threshold Θ_v, then *v* will be activated.

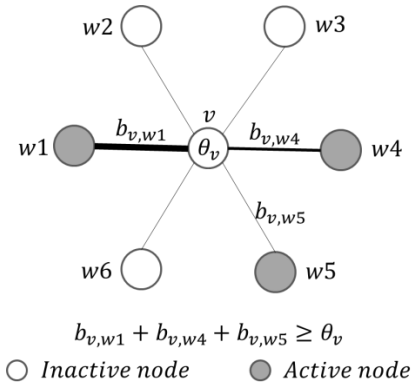

Fig. 1 *Linear Threshold Model: Activation of node v*

Its authors state that this model is suitable for a type of processes called *complex contagion*, in which it is more likely to activate a node if it has a certain amount of active neighbors, such is the case of the propagation of influence.

2.1.2 Independent Cascade Model

The influence propagation process starts with a set of active nodes A_o; the process is performed step by step as follows. When a node *v* becomes active at step *t* this has only one opportunity to activate each neighbor *w* with a probability $p_{v,w}$ which is a system parameter. Without explications its authors set $p_{v,w}$ at 1% and 10% in separate tests, the higher the value of $p_{v,w}$, the node *v* is more likely to influence the node *w*. If *w* has new neighbors recently activated, each of these tries to activate *w*. If *v* succeeds, then *w* becomes active in step *t + 1*. Otherwise, *v* will not make any other attempt to activate *w* in the following iterations. The process ends when further activation is not possible. *ICM* is oriented to the active nodes because it looks for these nodes to try to influence their inactive neighbors.
In Fig. 2 we show the inactive node *w* and an active neighbor *v*. We assume $p_{v,w} = 1\%$. Θ_w is a random number uniformly distributed in the interval [0, 1]. If $p_{v,w} \geq \Theta_w$, then *w* will be activated.

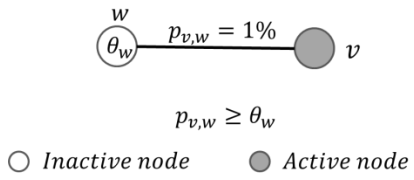

Fig. 2 *Independent Cascade Model: Activation of node w*

2.1.3 Baseline models revision

Considerations

The authors of the baseline models *LTM* and *ICM* (Kempe et al. 2003), consider nodes of two types, namely active and inactive. Active node is one that has adopted an innovation; inactive node is the opposite case. The authors say that influencing a node is to activate that node. The propagation of influence or information starts in a set of active nodes called seeds.

Over time, more and more neighbors of an inactive node (v) become active; at some point, this can cause v to become active, which in turn can cause that inactive neighbors of v also become active.

Modeling restrictions of baseline models

In the baseline models there are seven restrictions: (1) The sources of information and influence are within the social network, and nodes cannot be influenced by external sources. (2) No context is taken into account such as geographical region, time of the year, social environment, etc. (3) The topological structure of the graph remains static during the propagation process. (4) The activation thresholds of nodes remain constant through each message propagation process. (5) A node that has a certain state cannot return to an earlier state, so we say that these models are progressive. (6) No node can refuse to be seed. (7) No node can voluntarily be seed.

Properties

These models have two properties (Chen et al. 2013): monotonicity: adding nodes to a seed set does not reduce the final set of active nodes, and sub modularity: a decreasing marginal increment is obtained when nodes are added to a seed set. Nodes behave in a progressive way, that is, a node can pass from inactive to active but not from active to inactive.

2.2 Other models

Weighted Cascade Model (Kempe et al. 2003). This model is considered a special case of the *Independent Cascade Model*, where in each link $\{u,v\}$, node u activates node v with probability $1/d_v$, where d_v is the degree (number of links) of node v. *WCM* performs iterations by visiting the nodes of the graph in search of active nodes. Each active node has only one opportunity to activate its neighbors. If at the end of the current iteration there is at least one new node activated, a new iteration is started; otherwise, the process ends. As with *ICM*, this model is also oriented to active nodes.

Deterministic Linear Threshold Model (DLTM) (Swaminathan 2014). The difference of this model in respect to *LTM* is that the *DLTM* thresholds are assigned deterministically.

Deterministic Threshold Model (DTM)(Swaminathan 2014). In this model, each directed edge (u,v) is assigned a fixed weight 1. Each vertex v is assigned a fixed threshold Θ_v such as $\Theta_v \geq 1$ and is an integer. A vertex v changes its state S_v, from inactive to active ($0 \rightarrow 1$), if the sum of weights of coming edges from nodes that are in active state is greater or equal to Θ_v; otherwise vertex v remains inactive.

Triggering Model (Kempe et al. 2003). In this model, each node v chooses a random trigger set of neighbors T_v. If v is inactive at time t, but a neighbor on T_v has been activated, node v will also be activated at time $t + 1$.

Only-Listen-Once Model (Kempe et al. 2003). It is a special case of the *Triggering Model*. Here, each node v has a parameter p_v, so that the first neighbor of v that is activated, attempts to activate v with probability p_v, and all subsequent attempts to activate v fail. In other words, v only listens to the first neighbor that tries to activate it.

Nonprogressive Processes (Kempe et al. 2003). In a progressive process, nodes only go from inactivity to activity, but not vice versa. The nonprogressive case, in which nodes can switch in both directions, can in fact be reduced to the progressive case. This is achieved with thresholds of influence that vary over time, uniformly distributed in the interval [0, 1]. Node v remains active if $f_v(S) \geq \Theta_v^t$ where f is a function that determines the intensity with which the set of neighbors S tries to activate v.

Parallel Cascade (PC) model (Samadi et al. 2016). This is a model of influence propagation that uses Bayesian inference logic. Two opposing proposals are considered; each proposal competes to prevail in a group of people belonging to a social network. Each proposal is disseminated through the network starting in two sets of seeds S^+ and S^- respectively. Each node i receives the positive influence L_{it} and negative influence K_{it} at time t, which accumulate over time. To each node i is assigned a threshold of positive influence Θ_i^+ and a threshold of negative influence Θ_i^-. Node i is positively influenced if $L_{it} - K_{it} \geq \Theta_i^+$, or it is negatively influenced if $K_{it} - L_{it} \geq \Theta_i^-$; otherwise, node i remains uninfluenced. Each node accumulates influence that comes from its neighbors, regardless of whether it is influenced positively or negatively. Only an influenced node is able to propagate positive or negative influence to its neighbors.

Partial Parallel Cascade (PPC) model (Samadi et al. 2018). This model is a variant of the *PC* model that allows the partial activation of nodes. In each period of time $0 \leq t \leq T$ each node i delivers positive (X_{it}) or negative Y_{it} influence to its outgoing neighbors, where $0 \leq X_{it} \leq 1$ and $0 \leq Y_{it} \leq 1$ is the partial positive (negative) activation of node i at time t respectively. The amount of positive X_{it} (negative Y_{it}) influence is a function of $L_{it} - K_{it} \geq \Theta_i^+$ ($K_{it} - L_{it} \geq \Theta_i^-$). If $L_{it} - K_{it} \geq \Theta_i^+$ ($K_{it} - L_{it} \geq \Theta_i^-$), node i is considered totally influenced in a positive (negative) way. Θ_i^+ and Θ_i^- might or might not be equal.

Structural diversity model in social contagion (Ugander et al. 2012). Its authors found through empirical analysis that the probability of accepting a recommendation depends more subtly on the structure of the network, rather than on the number of friends that are influenced; that is, this acceptance depends on the different groups of neighbors instead of the number of neighbors. Each group or structural diversity represents a different social context to which the user belongs: family, coworkers, schoolmates, etc. In this model, only the first influenced friends can influence an uninfluenced friend.

Influence propagation model (Phadke et al. 2013). This model is based in the quantification of variable called *tie strength*. If the receiver is a close friend of the sender, reflected in greater tie strength in respect to other potential senders, then the

receiver is more likely to be influenced. The amount of influence retained by the receiver is relative to the social tie strength. As soon as the receiver receives some amount of influence, he sends the same quantity of influence to all his neighbors. The cumulative total influence is the sum of the influences received by a node through its active neighbors. This net influence is used as one of the predictors for the user to make a decision.

On the other hand, it is interesting the study that (Vishwakarma et al. 2019) perform to detect fake news in the form of images that are disseminated through social networks. The system uses technologies for text analysis and web scraping. Their study works on extracting features from the image and text. The algorithm applied uses various online resources to detect the credibility of the news (news channel like Fox News, CNN, and newspapers websites like The Washington Post), to detect the credibility of the news. The results of the algorithm are highly dependent upon the sources mentioned above, because the authors claim that whose trustworthiness is mandated by their good name in the market.

2.3 Social tie strength quantification

So far, we have described the work related to the propagation of information and influence. The number of nodes reached by a message that propagates through a social network depends directly on the social tie-strength between the nodes. *Lucy Model* also presents a way to quantify the tie-strength. Next, we will first describe some work related with the quantification of this strength.

Phadke et al. (2013) present a way to quantify tie strength among users of a mobile phone network, using the following mathematical expression: $w(x) = 1 - e^{-\frac{x}{\varepsilon^2}}$, where $x = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n$. $\alpha_1, \alpha_2, \ldots, \alpha_n$ and ε are assigned empirically using a training data set. $x_1, x_2, \ldots x_n$ are attributes expressed in values per unit, for example: number of calls placed between two users, the total duration of calls placed between two users, the proportion of neighbors that two users have in common with each other. $w(x)$ is restricted to the interval [0,1].

Girvan and Newman (2002) generalize Linton C. Freeman's betweenness centrality. Edge betweenness measures the number of shortest paths that pass through an edge to connect two nodes of a network.

Teixeira et al. (2013) define the spanning centrality of an edge *e* for an undirected and weighted graph as the relationship between the number of minimum expansion trees where *e* participates and the total number of minimum expansion trees.

Donald S. Sade designs the centrality of node k-path which is generalized by (De Meo et al. 2012) and apply to edges, which is defined as the number of paths, of at most, length k that connects with other edges.

(Postigo-Boix and Melús-Moreno 2018) propose to measure the distance between two nodes p and q: $distance_{pq} = \sqrt{w_1(p_1 - q_1)^2 + \cdots w_n(p_n - q_n)^2}$, $1, 2, \ldots n$ are attributes. w_i is the assigned weight to each dimension, which is used to normalize or to give more or less importance to each attribute. The authors define affinity as $affinity = 1 - distance_{pq}$. If affinity is a measure of closeness between two nodes, then affinity is a measure of tie strength between nodes.

3. LUCY MODEL

3.1 Introduction to the model

The *Lucy Model* is oriented to social networks modeled as an undirected graph $G = (V,E)$, where V is a finite set of vertices or nodes and $E \subseteq VxV$ is a set of edges that connect pairs of nodes. Nodes represent individuals and links represent relationships between individuals. The same considerations, restrictions and properties of baseline models (see Section 0) apply. As we will see later, *LM* is configurable so that it adapts to the different needs of each situation, allowing flexibility of use.

LM considers the propagation of three distinct classes of messages. They are (1) Messages to inform, (2) Messages to influence which appeal to feelings, (3) Messages to influence which appeal to personal. These messages provoke in individuals distinct activation thresholds, affecting the range of message propagation and the number of activated nodes. On the other hand, *LM* considers that the importance of a message and the intensity of relationship between individuals also affect the extent of its spread.

Baseline models consider that a node is activated without distinguishing if they have been informed or have been influenced. *Lucy Model* considers that a node is activated according to the class of message that is propagated in the network, as indicated in the previous paragraph. For this reason, *LM* is composed by three modules, one for each message class.

Messages are gradually propagated through the social network, so we modeled them as discrete time processes that are carried out step by step. Each of these steps represents one iteration for the process of each module. The process starts with a group of nodes called *seeds* that are previously activated. These nodes try to activate their neighbors, these to theirs, and so on, producing a cascading process. (Zuo et al. 2016) argue that the influence propagates a maximum of two hops away from a seed. There is a direct relationship between the number of seeds and the number of nodes that will be activated at the end of the process of propagating the message in the social network. Only active nodes can try to activate their neighbors. Our model is oriented to the active nodes because it looks for nodes from the network that are able to activate their neighbors.

We consider that, nodes are found in one of the following three states: state 1: inactive node, state 2: active node enabled to activate, state 3: active node disabled to activate. If a node in state 1 is activated then the node changes to active state 2 and therefore acquires the ability to activate its neighbors. Each node in state 2 only has one chance to activate its neighbors. After a node in state 2 tries to activate its neighbors it goes to state 3. At the end of the process there will only be nodes in state 3 and eventually nodes in state 1. Depending on the situation, a node is activated either because it is influenced or because it is informed. For both situations in this paper we use the term *activated*. The process of changing the state of the nodes we will explain in detail later in this section and in Section 3.4.

Due to the considerations made in the previous paragraphs of this section, *Lucy Model* needs to know three elements: the topology of the network, the importance of the message, and the number and identity of the seeds.
A message that spreads through the network is processed according to the following three steps: (1) We search in the graph active nodes enabled to activate (nodes in state 2). (2) Each node found in state 2 sends the message to his inactive neighbors (nodes in state 1). These receive the message with probability w_{ji} (see next Section 3.2). (3) If the message was received, it is verified if the message exceeds the activation threshold of the node, if this is the case, the node goes to state 2. Step 1 is described in Algorithm 1 (Section 3.4); Steps 2 and 3 are described in Algorithm 2 (Section 3.5). Fig. 3 explains the three above-mentioned message processing.
Next, in this Section we describe: tie strength and probability of receiving messages; thresholds of activation and the probability of activating nodes; messages propagation process; and node activation process.

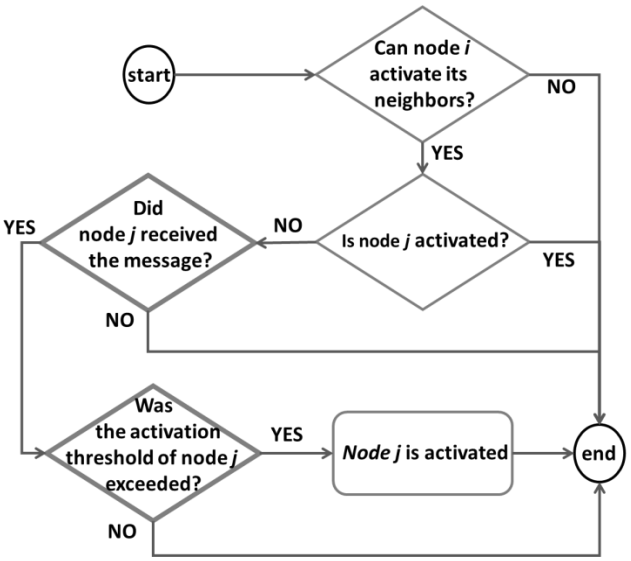

Fig. 3 *Node activation process*

3.2 Social tie strength and probability of receiving messages w_{ji}

The notion of tie strength in social networks is a property that characterizes the link between two nodes. Tie-strength can be quantified in three different ways: (1) In terms of the topology of the network, (2) In functional terms, that is, considering the number of messages that flow between nodes, and (3) Based on the attributes of the nodes: age, gender, race, etc. (Granovetter 1973) shows that there is a relationship between tie strength and the topological structure of the network. We quantify tie strength in terms of the topology of the network, which is described later.

By definition, tie strength is a "combination of the amount of time, the emotional intensity, the intimacy (mutual confiding) and reciprocal services which characterize the tie" (Granovetter 1973). Petróczi et al. (2006) asserts "The four indicators are actual components of tie-strength (closeness, duration and frequency, breadth of topics and mutual confiding), whereas contextual contingencies (neighbourhood, affiliation, socio-economic status, workplace and occupation prestige) are predictors. The first four contextual contingencies listed are features of affinity between people. Other contextual contingencies that characterize affinity between people are, for example: tastes, hobbies, passions, religion, family, age, gender, ethnicity, education, or people suffering the same tragedy. The last contextual contingency (occupational prestige) translates into relevance or importance of the individual within their social network. The phenomenon by which people tend to establish relationships for reasons of affinity is called *homophily* (McPherson et al. 2001; Mondani 2018), On the other hand, if people are interested in interacting with relevant people, these relationships are nonhomophilic.

In summary, the two great predictors of tie-strength are: affinity between individuals and relevance of these individuals. Then it follows that people establish relationships for two reasons: for affinity, for the interest of relating to relevant people, or for both reasons. *Lucy Model* proposes a mathematical expression to quantify tie strength for online social networks in function of those two great predictors.

Affinity is also a feeling. The affinity that node *i* feels with respect to node *j* does not necessarily equal the affinity that node *j* feels with respect to node *i*. On the other hand, the relevance of *i* in general is different from the relevance of *j*. Therefore there are two tie-strengths between *i* and *j*, one from *i* to *j* and one from *j* to *i*. The greater the affinity that *j* feels for *i*, and the greater the relevance that *j* considers that *i* has, then tie-strength of *j* to *i* (symbolically w_{ji}) is greater. This translates into a greater predisposition of *j* to accept messages from *i*.

The absolute affinity between nodes i y j is A_{ij}; R_i is the quantification of centrality of node i. A_{ij} and R_i are magnitudes that have different nature and scale. To be able to combine them in a mathematical expression to obtain w_{ji}, it is necessary to normalize these magnitudes, then:

$$w_{Aji} = normalized(A_{ij})$$

$$w_{Rji} = normalized(R_i)$$

Then:

$$w_{ji} = f(w_{Aji}, w_{Rji})$$

w_{Aji} is the affinity that node *j* feels with respect to node *i*. w_{Aji}, being normalized, is the probability that node *j* receives messages from node *i* considering the affinity between the two nodes. w_{Rji} is the relevance that *j* considers that *i* has. w_{Rji}, being normalized, is the probability that node *j* receives messages from node *i* considering the relevance of the node *i*. w_{ji} is the joint probability of two events w_{Aji} and w_{Rji} probabilistically independent of each other. Then:

$$w_{ji} = w_{Aji} * w_{Rji} \qquad (1)$$

Tie strength w_{ji} is normalized and represents the probability that node *j* will receives messages from node *i*. As will be seen later, in general $w_{ji} \neq w_{ij}$.

Eq. (1) is consistent with what is expressed by Mark S. Granovetter in the sense that the links within the communities are strong and the links between these groups are weak. In fact, the links that join communities have a small value w_{Aji}, and therefore w_{ji} is small. Within the communities the value of w_{Aji} is large, obtaining larger values of w_{ji}. Next, we find expressions to quantify A_{ij}, w_{Aji} and w_{Rji}.

3.2.1 Affinity A_{ij}

To quantify tie strength due to affinity w_{Aji}, we need to calculate the affinity between nodes A_{ij} based on the network topology. (Golbeck 2013) states that "*Those who have many mutual friends are likely to have stronger ties*". If this is so, then we can infer that these two friends are joined by strong feelings of affection, affinity or sympathy. Recalling what Mark S. Granovetter said that there is a relationship between tie strength and topological structure of the network, we will find a mathematical expression to compute w_{Aji} according to the number of common neighbors that each pair of related nodes has.

We observed that there are the following analogies between an electrical network and a social network: (1) The two structures are networks that can be represented by a graph. (2) Both networks can have sources and support a flow. In the first case, the sources are of energy and support the flow of electric current. In the second case, the sources are of messages and support the flow of the said messages. (3) In an electric circuit, the current flows from the points with greater electrical potential towards points with less potential. In a social network, a message flows from the individuals who have the message to the individuals who do not have it. (4) In an electrical circuit the conductors have the property to facilitate the flow of current, and it depends on the physical and electrical characteristics of the material. This property is the electric conductance *G*. In a social network, links have the property of facilitating the flow of messages, and it depends on the affinity A_{ij} that exists between individuals. To higher affinity higher tie strength between a pair of nodes. Then there is an analogy between G and A_{ij}.

Identifying the analogies between the two types of networks, we make the abstraction of treating a social network as if it were an electrical network. We then calculate the electrical conductance of the links, and assume that these measures are numerically equal to the affinities of the same links in the social network. Previously, for different purposes, similarities had already been established between two systems belonging to different areas of knowledge; for example, the ecosystem concept used in Biology and Life Sciences was adapted in an area of social sciences to model the business world with the name "Business Ecosystem" (Rong et al. 2015; Maturo et al. 2018). Another example is the modeling of computer virus propagation from disease propagation models in human populations (Piqueira and Araujo 2009; Uddin et al. 2015). Brandes and Fleischer (2005) did not determine the affinity between individuals, but establish variants of centrality measures, considering that the information is propagated efficiently as an electric current.

We proceed to calculate the equivalent electrical conductance G_{ji} between nodes j and *i* (Fig. 6). Let us consider the portion of a graph presented in Fig. 4 as if it were an electrical network. Nodes *1, 2, ... n* are connected to nodes *i* and *j*, and interconnected with each other. We assume that all links have an electrical resistance R = 1.

The calculation of G_{ji}can be done in two ways: by applying Kennelly's Theorem (Alexander and Sadiku 2017), or as follows. If a voltage is applied between nodes *i* and *j*, nodes *1, 2, ..., n* will be at the same electrical potential, therefore there will be no current between these nodes (Alexander and Sadiku 2017), then their links can be eliminated, as seen in the Fig. 5.

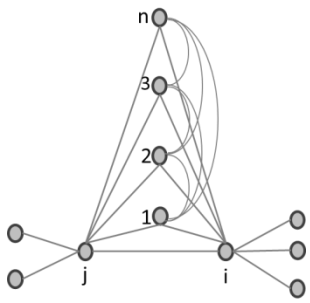

Fig. 4 *Graph of an electrical network*

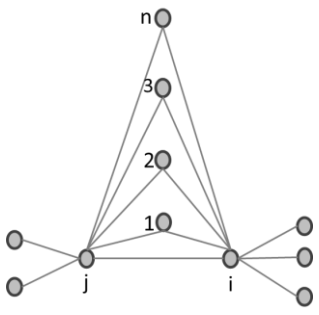

Fig. 5 *Equivalent electrical network*

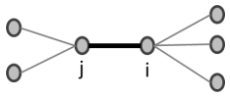

Fig. 6 G_{ij} *equivalent*

$R_{jni} = R_{jn} + R_{ni} = R + R = 2R$, similarly: $R_{j1i} = R_{j2i} = \cdots = R_{jni} = 2R$. The equivalent electrical conductance between nodes *j* and *i* (Fig. 6) is:

$$G_{ji} = \frac{1}{R_{ij}} + \underbrace{\frac{1}{2R} + \frac{1}{2R} + \cdots + \frac{1}{2R}}_{n} \quad (2)$$

$$G_{ji} = \frac{1}{R_{ij}} + \frac{n}{2R} \quad (3)$$

If R = 1, then:

$$G_{ji} = 1 + \frac{n}{2} \quad (4)$$

where n is the number of nodes connected to both nodes *i* and *j*.
By the existing analogy already indicated between an electric circuit and its corresponding social network, we conclude that the equivalent conductance $G_{ij\text{-}eq.}$ is numerically equal to the affinity A_{ij}.

$$A_{ij} = 1 + \frac{n_{ij}}{2} \quad (5)$$

where n_{ij} is the number of neighbors common to *i* and *j*. Observe that Eq. (5) tells us that the affinity between *i* and *j* is reciprocal, i.e., $A_{ij} = A_{ji}$.

3.2.2 Normalization of affinity and relevance

The probability that node *j* receives a message from node *i* considering the affinity between the two nodes (w_{Aji}) is directly proportional to the affinity between *i* and *j* (A_{ij}) in relation to the sum of the affinities of *j* with all its neighbors (see Fig. 7), that is

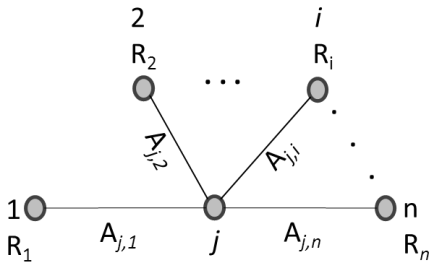

Fig. 7 *Affinity and relevance*

$$w_{Aji} = \frac{A_{ij}}{\sum^{n}_{\substack{k=1 \\ \{j,k\}\in E}} A_{jk}} \quad (6)$$

Similarly, the probability that node *j* receives a message from node *i* considering the relevance of the node *i* (w_{Rji}) is directly proportional to the relevance of *i* (R_i) in relation to the sum of the relevance of neighbors of *j* (see Fig. 7). R_i can be calculated by using some metrics of centrality, i.e., Degree, PageRank, Eigenvector, etc. (Bekiari and Hassanagas 2015). Then:

$$w_{Rji} = \frac{R_i}{\sum^{n}_{\substack{k=1 \\ \{j,k\}\in E}} R_k} \quad (7)$$

Equations (6) and (7) are normalized magnitudes whose values are in the interval [0, 1].

3.3 Activation thresholds and importance of a message

LM considers the propagation of three distinct classes of messages, each of them with a particular purpose. They are: Messages to influence which appeal to interests and personal conveniences; Messages to influence which appeal to feelings; and Messages to inform. The extent of this spread in a social network depends on three aspects: (1) the class of message; (2) the importance of the message, and; (3) the relevance of the sender of the message, or the affinity between sender and receiver according to the message class. These aspects awaken in each receiving node a threshold that determines the probability that it will be activated.

Nonhomophilic influence threshold
The first class of messages is intended to influence people by appealing to interests and personal conveniences. Examples of messages of this class are economic advice; health recommendations; travel recommendations; and safety recommendations

for machinery operators, vehicle and pedestrian drivers, etc. People analyze these messages in a rational and objective way, take precautions and seek the advice of other people. In this class of influence, not accepting the influence could mean risks. This class of messages awakens in the people a certain threshold of influence (U_{NHji}).
The greater the relevance of the individual (i) who sends a message or the greater the importance of said message (I_{pj}), the lower the threshold of influence of the individual j who receives the message, being easier to influence him. These considerations are fulfilled in the following mathematical equation (8).

$$U_{NHji} = 1 - \left(\alpha R_{Rij} + \beta I_{pj}\right) \qquad (8)$$

Where $R_{Rij}$is the relative relevance of i with respect to j; I_{pj} is the importance of message perceived by j; α and β are parameters that give weights to the variables.

Homophilic influence threshold
This class of message tries to influence people by appealing to feelings and emotions, where it counts the affinity, affection and sympathy of the people. Examples of messages of this class are a favor requested by someone dear; call to solidarity for someone dear; call to celebrate someone dear. People analyze these messages with their hearts. In this class of influence, the decision made does not mean risks. This class of messages awakens in the people a certain threshold of influence (U_{Hji}).
The greater the affinity that the receiver j feels for the sender (i) who sends a message or the greater the importance of said message (I_{pj}), the lower the threshold of influence of the individual j who receives the message, being easier to influence him. These considerations are fulfilled in the following mathematical equation (9)

$$U_{Hji} = 1 - \left(\alpha A_{Rji} + \beta I_{pj}\right) \qquad (9)$$

Where, A_{Rji} is the relative affinity that j feels for i.

Information threshold
The messages do not try to influence, they only try to inform. Examples of messages of this type are scandal, tragedy, political or sporting event. This class of messages awakens in the people a certain threshold of interest U_{Ij}. The receiving node (j) is activated if the news exceeds the threshold of interest. The threshold of interest U_{Ij} depends on the importance of the message I_{pj}. Greater interest in the news causes lower threshold U_{Ij} in node j. These considerations are reflected in the following mathematical equation (10)

$$U_{Ij} = 1 - I_{pj} \qquad (10)$$

3.3.1 Relative relevance and relative affinity

Each individual has a relevance (R_j) within his/her social environment. The relative relevance R_{ij}used in Eq. (8) is the quantification of how relevant the node i is in comparison with the node j. Both R_{ij} and I_{pj} have to be values within the range [0, 1] for U_{NHji} to be in this same range. For this reason the following expression that calculates the relative relevance of i $R_{ij} = \frac{R_i}{R_j}$ does not help us. An expression that meets the above requirements is the following one:

$$R_{Rij} = \frac{R_i}{R_i + R_j} \qquad (11)$$

There is not the possibility that $R_j = R_i = 0$

The values that can take R_{Rji} are in the interval [0, 1]. To quantify the relevance of nodes we use *PageRank* centrality metric, since it is widely used in Social Network Analysis (Bekiari and Hassanagas 2015; Yun and Gloor 2015).
In Eq. (9) we use A_{Rji} which is the relative value of A_{ji}. According to Fig. 7, we define A_{Rji} as follows:

$$A_{Rji} = \frac{A_{ji}}{Max_{\{j,k\}\in E}(A_{jk})} \qquad (12)$$

3.3.2 Importance of a message

In general, people do not perceive the importance of a message in the same way; some people give more importance and others give less importance to the same message. For example, if people, in general, feel that a message is important, one particular individual will seem to perceive it as moderately important, whereas another one will find it very important.
In the equations [(8)-(10)] the term *importance of the message* (I_{pj}) is present. The importance is perceived by a particular individual j. To deal with the variability in the perception of the importance of a given subject, we relate I_{pj} with *Is* in Eq. (13), where *Is* is a parameter of the system that represents the average importance of the subject. We defined a scale of five qualitative values with their corresponding quantitative values for *Is*; these are *very little important (0.2), little important (0.4), moderately important (0.6), important (0.8), very important (1.0)*.

$$I_{pj} = I_s + \text{random}(\epsilon) \qquad (13)$$

ϵ is a set of values that depends on the *Is* parameter, as seen in the Table 1. To calculate I_{pj}, a value of the set ϵ corresponding to *Is* is taken randomly and added to *Is*.

Table 1 *Relation between Is and the set ϵ*

Parameter *Is*	0.2	0.4	0.6	0.8	1.0
Set ϵ	{0.0, 0.2}	{-0.2, 0.0, 0.2}	{-0.2, 0.0, 0.2}	{-0.2, 0.0, 0.2}	{-0.2, 0.0}

3.4 Example of calculation of tie-strength w_{ji} and activation thresholds

In Table 2 we show an example of w_{ji} calculation for some links of the graph of Fig. 8. To make the example more explicit, the centrality metric R_i is the degree of the nodes.

Fig. 8 *Graph used as an example to explain the calculation of some variables*

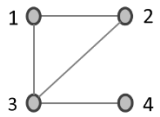

Table 2 *Calculation of w_{ji}*

j,i	n_{ji}	$A_{ji} = 1 + \frac{n_{ji}}{2}$	A_{jk}	$\sum A_{jk}$	$w_{Aji} = \frac{A_{ji}}{\sum A_{jk}}$	R_i	R_k	$\sum R_k$	$w_{Rji} = \frac{R_i}{\sum R_k}$	$w_{ji} = w_{Aji} * w_{Rji}$
1,2	1	1.5	A_{12}=1.5 A_{13}=1.5	3	0.5	2	R_2=2 R_3=2	4	0.5	0.25
1,3	1	1.5	A_{12}=1.5 A_{13}=1.5	3	0.5	3	R_2=2 R_3=2	4	0.75	0.375
3,1	1	1.5	A_{31}=1.5 A_{32}=1.5 A_{34}=1	4	3.75	2	R_1=2 R_2=2 R_4=1	5	0.4	0.15
3,4	0	1	A_{31}=1.5 A_{32}=1.5 A_{34}=1	4	0.25	1	R_1=2 R_2=2 R_4=1	5	0.2	0.05
4,3	0	1	A_{43}=1	1	1	3	R_3=3	3	1	1

In Table 3 we show an example of calculation of the three types of activation thresholds for some nodes of Table 1 with respect to a neighbor. The graph used is the one shown in Fig. 8. The parameters α= β= 0.5, and the perceived importance of the message I_{pj}= 0.2.

Table 3 *Calculation of activation thresholds*

j,i	R_j	R_i	R_{Rij}	A_{ji}	A_{jk}	max(A_{jk})	A_{Rji}	U_{NHji}=1-(αR_{Rij}+βI_{pj})	U_{Hji}=1-(αA_{Rji}+βI_{pj})	U_{Ij}= 1-I_{pj}
1,2	2	2	0.5	1.5	A_{12}=1.5 A_{13}=1.5	1.5	1	0.65	0.4	0.8
1,3	2	3	0.6	1.5	A_{12}=1.5 A_{13}=1.5	1.5	1	0.6	0.4	0.8
3,1	3	2	0.4	1.5	A_{31}=1.5 A_{32}=1.5 A_{34}=1	1.5	1	0.7	0.4	0.8
3,4	3	1	0.25	1	A_{31}=1.5 A_{32}=1.5 A_{34}=1	1.5	0.67	0.775	0.565	0.8
4,3	1	3	0.75	1	A_{43}=1	1	1	0.525	0.4	0.8

3.5 Search of active nodes

Lucy Model carries out successive iterations by visiting in each of them all nodes of the graph looking for nodes in "*active, enabled to activate*" state (state 2). These nodes have only one opportunity to try to activate their neighbors, after which these nodes become "*active, disabled to activate*" (state 3).

In Fig. 9 we show an example of the process of propagation of activation of nodes in a graph of six nodes with the first tour of the graph (iteration). Initially there are two nodes in state 2 (seeds): nodes 1 and 4. (a) The first node in state 2 is 1 that is able to activate node 2. (b) The nodes 1 and 2 go to state 3 and 2 respectively. The next node in state 2 is node 2 that is able to activate node 6. (c) The nodes 2 and 6 go to state 3 and 2 respectively. The next node in state 2 is node 4, which is not able

to activate any node. (d) Node 4 goes to state 3. The next node in state 2 is node 6 that is able to activate node 3. (e) The nodes 6 and 3 go to state 3 and 2 respectively. Then, the first iteration ends. Since the graph has a node in state 2, a new iteration is started. The next node in state 2 is node 3 that is not able to activate any node. (f) Node 3 goes to state 3. There are no more nodes in the graph in state 2, so at the end of this iteration the node activation process ends. This process is summarized in Algorithm 1 in the form of a pseudo code.

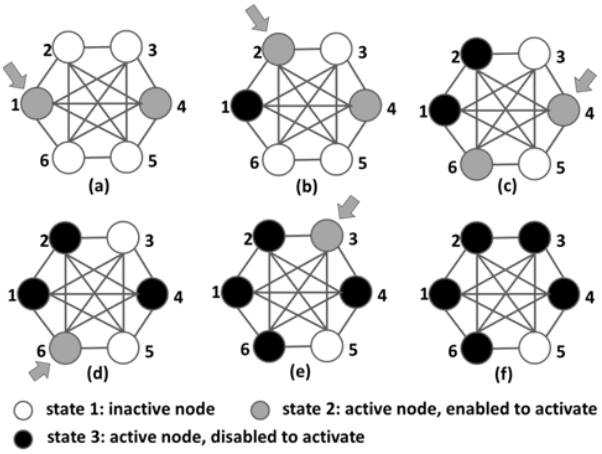

Fig. 9 *Activation propagation process*

Algorithm 1. Search active nodes enabled to activate

Input*: graph*
Output*: seeds (activated nodes)*
Begin
  Repeat
    newCycle ← 0
    for *i = 1* ***to*** *i = n* ***do***
      if *i is enabled to activate(state[i] == 2)* ***then***
        i is disabled to activate: state[i] ← 3
        for *j = 1* ***to*** *j = n* ***do***
          if *j is inactive (state[i] == 1)* ***then***
            Algorithm 2: Sending messages and activating nodes
          end if
        end for j
      end if
    end for i
  Until *newCycle == 0*
End.

3.6 Node activation process

Step 2 (Sending messages) and step 3 (activation of nodes) of our proposal *LM* are summarized in Algorithm 2.

Algorithm 2. Sending messages and activating nodes

Begin
  Calculate the affinity for each link: A_{ij}
  Calculate the normalized affinity: w_{Aji}
  Calculate the relevance of nodes: $R_i \leftarrow centrality[i]$
  Calculate the normalized relevance: w_{Rji}
  Calculate tie strength: $w_{ji} \leftarrow w_{Aji} * w_{Rji}$
  if *node j does not receive the message from node i* ***then end***
  Read the importance of the message Is
  Calculate the perceived relevance for each node: I_{pj}
  Calculate the relative relevance of sending node i: R_{Rij}
  Calculate the relative affinity of receiving node j: A_{Rji}
  if *the propagation is nonhomophilic influence* ***then***
    Calculate threshold of influence: U_{NHji}
  end if

```
    if the type of propagation is information then
        Calculate threshold of information:U_Ij
  end if
  if the type of propagation is hemophilic influence then
        Calculate threshold of influence:U_Hji
  end if
  if the random[0, 1] ≥ threshold then
        Node j is activated and enabled to activate: state ← 2
        At the end of the present iteration, a new one will begin:
        newCycle ← 1
  end if
end.
```

4. VALIDATION METHOD OF LUCY MODEL

The only criteria used by the baseline models (*LTM*, *ICM*), *WCM* and our model (*LM*) to predict the range of message propagation is the number of activated nodes. In this way, we have been able to determine the accuracy of the results of our model by comparing them with those of the other authors.
To demonstrate the validity of *Lucy Model*, we perform the following validation steps:

Designing Validation Objectives
The tests we have carried have three objectives: First, to demonstrate in a conclusive way that the predictions of *Lucy Model* are notably more accurate and more precise than the predictions of the baseline models and *WCM*. Second, to show the flexibility of Lucy Model when its parameters are modified to adapt to different situations. Third, to show the behavior of *Lucy Model* in three different types of networks: a synthetic *Power-Law* network (Newman 2010), a *Youtube* social network, and a tracked *Facebook* network.

4.1 Designing the validation

The tool we used to generate the synthetic network was *GenRndPowerLaw* from the *Stanford Network Analysis Project* (SNAP). The *Youtube* network was also obtained from SNAP.
Power-Law (Newman 2010) is a mathematical relation utilized by algorithms that generate synthetic networks. *Power-Law* networks are a class of random networks because the links are randomly added to a static set of nodes. The degree distribution obeys the power law $p(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$, where $p(k)$ is the probability of a node having degree k, $\zeta(\gamma)$ is Riemann's Zeta function, and $\gamma > 1$ is the exponent that takes values between 2 and 3 for social networks. When $2 \leq \gamma \leq 3$ occur, these networks are called *scale-free networks* because there is a high probability that nodes with high degree are connected to other nodes with a high degree (Li et al. 2005), characteristic that is typical of social networks.
We chose to use a *Power-Law* network because many existing social networks are said to be of this type, *e.g.*: friends networks (Hein et al. 2006), telephone call networks ($\gamma = 2.1$), email networks ($\gamma = 2.0$), World Wide Web ($\gamma = 2.1$) (Newman 2010), and online social networks (Duong-Ba, Thuan Hong 2014). In Table 4, we show the characteristics of the graphs that we are using.

Table 4 *Characteristics of the graphs used*

Graph	Node number	Edge number	Max degree	Mean grade
Facebook	968,810	2,742,523	4,804	5.66
YouTube	1,134,890	2,987,624	28,754	5.34
Power-Law exponent γ=2.1	1,000,000	3,382,218	143,309	6.76

Next, we elaborate the validation protocols to carry out their validation activities and present the results.

4.1.1 Protocol for empirical data extraction

To validate *Lucy Model*'s results, we extracted empirical data of messages propagation in the *Facebook* social network of an anonymous user, who has a personal profile. For this purpose, we spread messages through different seed users. Each published message included a hyperlink to a page of a web server; in this way, the *Facebook* user who received the message, by clicking on the hyperlink could access complete information related to that message. The number of people who accessed the web server we counted as activated people.
The seed users were twenty-three Discrete Mathematics students from the March-July 2018 period of the Faculty of Engineering of the Universidad de Cuenca.
Three classes of messages were designed and published: A nonhomophilic message, a homophilic message, and an informative message. Specifically, the messages were espionage in *WhatsApp*, abandonment of pets, and racism, respectively. Each seed published each class of message for a single time, scheduled as follows: Day 1: seed 1. Day 2: seed 2. Day 3: seed

3. Day 4: seed 4. Day 5: seed 5. Day 6: seeds 6, 7, 8, 9 and 10. Day 7: seeds 11, 12, 13, 14 and 15. Day 8: seeds 16, 17, 18, 19, 20, 21, 22 and 23. The participation of the seeds was in descending order according to their number of friends on *Facebook*. We program the publication of messages in this way to know the contribution of each seed or group of seeds in the total number of activated nodes. We had this same purpose when we carried out the laboratory tests with the *LM*, *LTM*, *ICM* and *WCM* models. On the other hand, we try to ensure that the publication of each message takes place in the shortest possible time so that the messages do not lose relevance and interest in the nodes to be activated. The publication of the three classes of messages took place between June 25 and July 4 of the year 2018.

4.1.2 Protocol for laboratory tests

The simulation to predict the extent of the propagation of messages with the respective models, was carried out in the tracked *Facebook* network of the same anonymous user used in the field experiments. We had to trace it up to three hops from the anonymous user because the messages propagate up to two hops of a seed (Afrasiabi Rad and Benyoucef 2012; Zuo et al. 2016). Tracking at various depth levels is not allowed by the official *Facebook* APIs, due to its security policies. For this reason we use *FBS script* (González Toral, Santiago 2018). This script uses *Facebook's* Front-end, is developed in the *Python* language, allows automatic tracking, and takes into account that *Facebook* could block the user's account. The process of tracking the network was slow since it was carried out in twenty-seven days from May 28 to June 23 of the year 2018.

The predictions made by the four models *LM*, *LTM*, *ICM* and *WCM* depend on probabilistic elements. In the case of *LM*, these elements are three: (1) Probability that the nodes receive the message [(Eq. (1)]. (2) Probability of exceeding the activation threshold of each receiver node [equations: (8)-(10)]; and (3) Estimating the importance of the message perceived by each receiving node [Eq. (13)].

We have used the same 23 seeds that we had used in the field experiments. The seeds were organized in the same way as in the field experiments, that is, in a descending manner by the number of neighbors or degree. We form the following sets of nodes: A = {1}, B = {1,2}, C = {1,2,3}, D = {1,2, ..., 4}, E = {1, 2, ..., 5}, F = {1,2, ... 10}, G = {1,2, ... 15}, and H = {1,2, ... 23}. We did it in this way to determine how much each seed or group of seeds contributes in the activation of the nodes.

We use the *Monte Carlo Method* (Bolthausen and Wüthrich 2013) to predict the number of activated nodes, executing thousand times each of the four models (*LM*, *LTM*, *ICM*, *WCM*), and for each group of seeds, starting with group A. In this way the results converge to an average value of extent, amplitude or distance of propagation of messages. In each execution we uniformly generate random numbers in the interval [0, 1] with the purpose of solving the thresholds for activation of nodes and perceived importance of the message (I_p).

Monte Carlo Method is based on Jakob Bernoulli's Theorem, also called Law of large numbers. The theorem states that if X_1, X_2, X_3, … is an infinite sequence of independent random variables having the same expected value μ and standard deviation σ, then $\overline{X}_n = (X_1 + \ldots + X_n)/n$ converges in probability to μ. That is, for any positive number ε we have

$$IC_{1-\alpha}(\mu) = \bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

This is a method of numeric stochastic research that allows obtaining approximate solutions to complex, deterministic and probabilistic problems, as is our case (Masayuki Yano et al. 2013).

In the case of *Lucy Model*, this process was done with each of the three classes of propagation of messages: informative, nonhomophilic, and homophilic. Both for the subject that the messages treat and for their design we consider that they have *very little importance* (*Is*). On the other hand, we have estimated that the weights (α and β) of the variables of the activation thresholds are equal. For these reasons, in *LM* the parameters were set with the following values: α = 0.5, β = 0.5 and *Is* = 0.2. In order that the *LTM* and *LM* results are comparable in similar conditions, we have decided that $b_{v,w}$ of *LTM* is equal to w_{ji} [Eq. (1)] of *LM*. In *ICM* we fix $p_{v,w}$ = 1%.

The message propagation models were implemented in computer simulators. The tools used were *C/C++* programming language, *SNAP* graph mining library, *NodeXL* network analyzer, spreadsheet, and PC core *i7* processor, 8GB, 64-bit *Ubuntu* operating system.

5. VALIDATION RESULTS AND ANALYSIS

5.1 Field experiments

Based on the protocol for empirical data extraction described in the previous section, we show in the Fig. 10 the results of the field experiments. We notice two characteristics:

1. There is a similar behavior in the propagation of the homophilic message and the nonhomophilic message.
2. The number of nodes activated with the informative message is lower than those reached by the other two classes of messages.

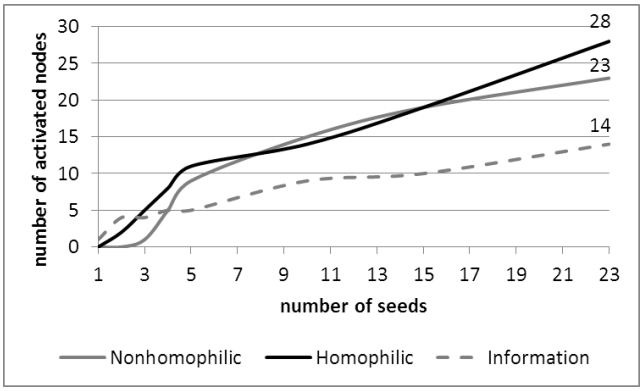

Fig. 10 *Tracked Facebook network: Empirical results*

5.2 Laboratory tests

Figures Fig**. 11**-Fig**. 13** show the grade of accuracy achieved by laboratory predictions made by *Lucy Model*, *LTM*, *ICM* and *WCM*, compared to empirical evidence. In the case of *Lucy Model*, these figures show the number of nodes activated as a result of the propagation of a nonhomophilic, homophilic and informative message, respectively. In these figures there is a gray band inside which are all the curves (B) of *Lucy Model*. Each curve corresponds to different values assigned to the parameters (α, β, *Is*) of the activation thresholds [Eq. (8) -Eq. (10)]. The curves that delimit the gray bands were obtained with α = 0 and β = 1. In this way, $U_{NHji} = U_{Hji} = U_I[j] = 1 - I_P[j]$. The curves marking the upper and lower borderline of these bands were obtained with Is = 0.2 and Is = 1, respectively.

In these figures we clearly observed that the band of *Lucy Model* is much closer to the result of the empirical experiments, compared to the predictions of the baseline and *WCM* model.

The lower and upper borders of the gray bands of the three next figures mark the minimum and maximum number of activated nodes predicted for each message class. The black curve that is within the each gray band indicate the number of nodes predicted for *Is=0.2* and *α=0.5*.

The laboratory tests that we carried out with *Lucy Model* and the field experiments, resulted in a very small number of activated nodes compared to the almost one million nodes of the tracked *Facebook* network and the millions of nodes of the *Facebook* platform. This was due to the fact that, in both tests, we had available a set of seeds with very little relevance with respect to other nodes in the network.

In Fig. 11, Fig. 12 and Fig. 13, L*M* has with seed 8 the greatest growth in the number of activated nodes. This is because the neighbors have stronger tie strength with the seed 8 compared to the previous seeds. Therefore, the probability that the neighbors receive the message and they are activated is greater.

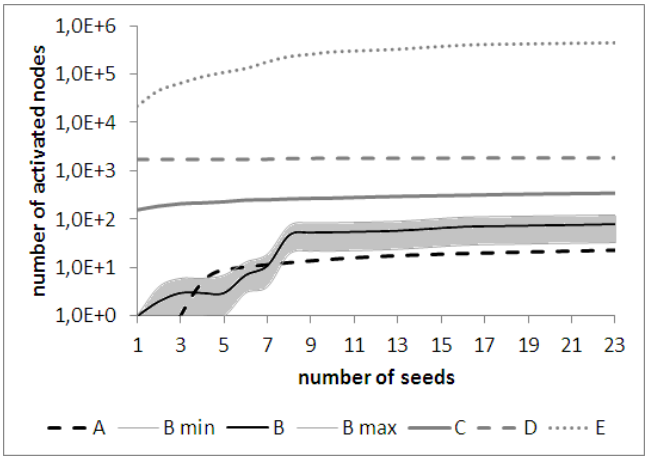

Fig. 11 *Accuracy: Propagation of a nonhomophilic message. A: empirical data. B min: LM min. B: LM (Is=0.2, α=0.5). B max: LM max. C: ICM (pv,w=1%). D: LTM. E: WCM*

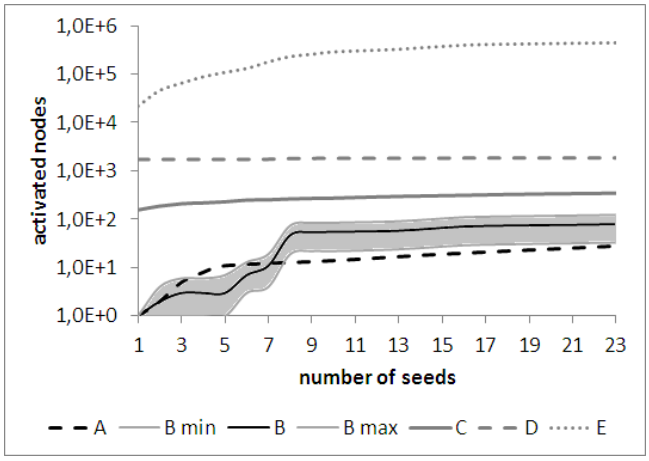

***Fig. 12**. Accuracy: Propagation of a homophilic message. A: empirical data. B min: LM min. B: LM (Is=0.2, α=0.5). B max: LM max.. C: ICM (pv,w=1%). D: LTM. E: WCM.*

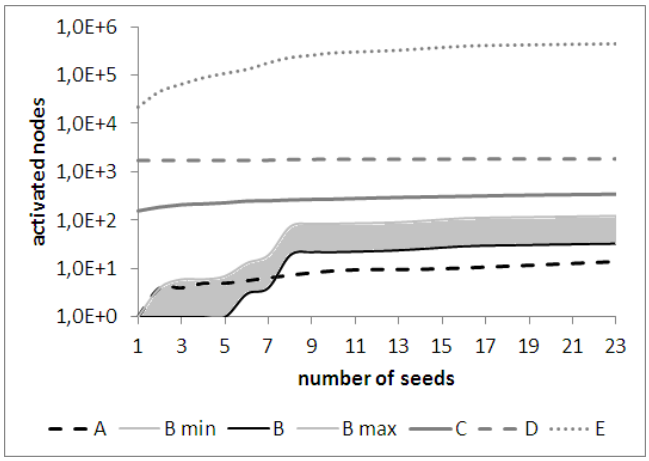

***Fig. 13** Accuracy: Propagation of an informative message. A: empirical data. B min: LM min. B: LM (Is=0.2). B max: LM max. C: ICM (pv,w=1%). D: LTM. E: WCM*

The Figs. Fig**. 11**-Fig**. 13** also show the flexibility of Lucy Model modules to adapt to the different levels of importance of the message to be published, as well as the different weights that the user of the model wishes to give to the parameters of the activation thresholds of nodes.
LM activates a much smaller number of nodes in relation to the other models because we consider the probability (w_{ji}) that the nodes receive the message, while the other models do not. If in *LM* we only consider the probability of influence U_{NHji} and without considering w_{ji}, the results are as follows (see Fig. 14).

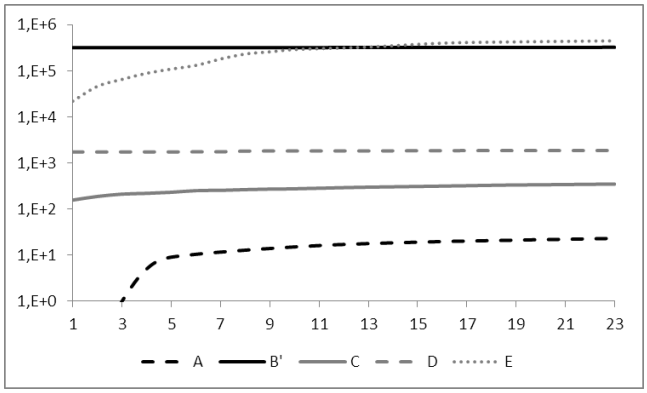

***Fig. 14** Accuracy: Propagation of a nonhomophilic message. A: empirical data. B': LM (Is=0.2, α=0.5). C: ICM (pv,w=1%). D: LTM. E: WCM*

In figures 15-17 we show *Lucy Model* behavior in different types of networks: the tracked *Facebook* network, a *Youtube* network, and a synthetic *Power-Law* network. The laboratory tests allowed us to make some findings.

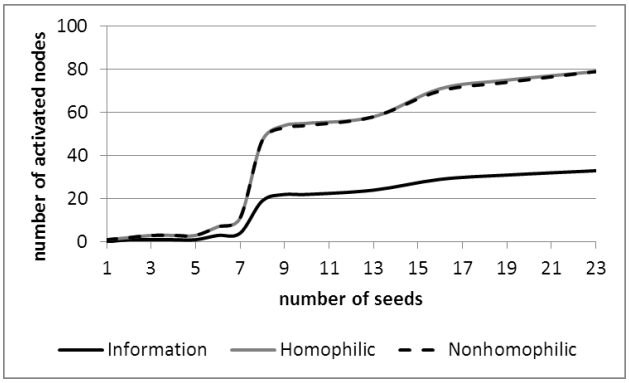

Fig. 15 *Lucy Model. Tracked Facebook network. Three classes of messages. Is = 0.2, α = 0.5*

In figures Fig**. 16** and Fig**. 17** since we used other networks, we had to use other seeds. We chose the 23 nodes with the highest value of PageRank centrality as seeds and ordered them in a descendent way. With these seeds, the nodes activation is much greater as we see in the respective figures. In three *Lucy Model* modules we use the same parameters: *Is* = 0.2, α = 0.5.

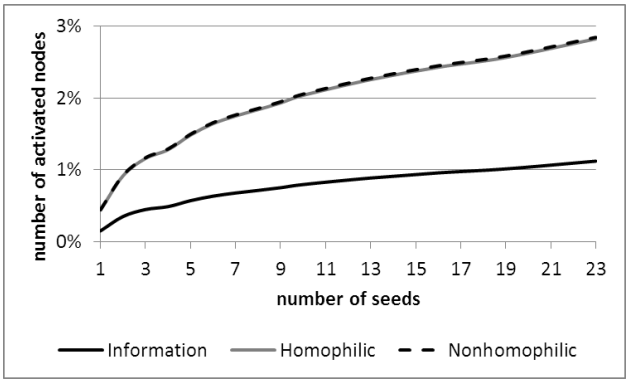

Fig. 16 *Lucy Model. Youtube network. Three classes of messages. Is = 0.2, α = 0.5*

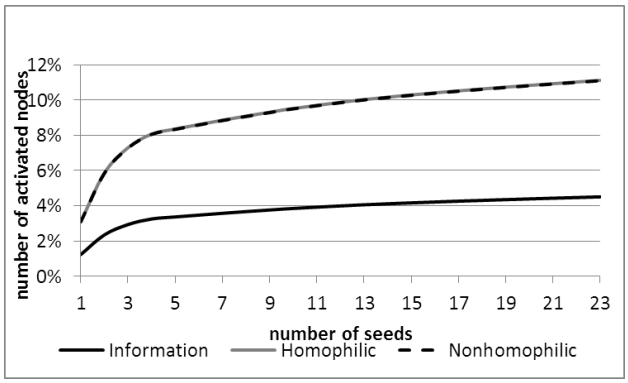

Fig. 17 *Lucy Model. Synthetic network Power-Law. Three classes of messages. Is = 0.2, α = 0.5*

Firstly we find two characteristics:

1. The informative message reaches the smallest propagation with respect to the other types of messages. This result is the same obtained in the field experiments (see Fig. 10).
2. The nonhomophilic message and the homophilic message had almost the same number of activated nodes. Indeed, the average relative difference in the number of activated nodes with these two types of messages in the *Facebook*, *YouTube* and *Power-Law* networks is 0.46%, 0.85% and 0.19% respectively. We obtained similar results in the empirical experiments.

Regarding the second characteristic, the activation of nodes depends on tie-strength and the activation threshold. Tie-strength between nodes is the same regardless of the class of message that propagates through the network. The activation of nodes depends on their respective thresholds (8) and (9). These thresholds are almost identical ($U_{NHji} \approx U_{Hji}$), differing in that (8) depends on the relevance of nodes, while (9) depends on the affinity between nodes. Therefore, the number of activated nodes in these cases is very similar as reflected in Figs. **Fig. 15** to **Fig. 17**.

Then:

$$A_{Rji} \approx R_{Rij} \qquad (14)$$

Then, the relative affinity that the receiver *j* feels for the transmitter *i* is numerically equal to the relative relevance of the transmitter *i* in relation to the receiver *j*. Therefore, the propagation of both nonhomophilic and homophilic messages can be predicted using either the activation threshold (8) or the (9).

In the Eq. (11): $R_{Rji} = \frac{R[j]}{R[i]+R[j]}$

Analogously $R_{Rij} = \frac{R[i]}{R[i]+R[j]}$

Then:

$$R_{Rji} + R_{Rij} = 1 \qquad (15)$$

To explain the first and second characteristics, we replaced the values of the parameters in the equations of the activation thresholds [(8)-(10)], and using equations (14)-(15), we verified that: $U_{Ij} > U_{Hji} \approx U_{NHji}$
On the other hand, and analogous to Eq. (14):

$$A_{Rij} \approx R_{Rji} \qquad (16)$$

From equations [(14)-(16)]:

$$A_{Rji} + A_{Rij} \approx 1 \qquad (17)$$

Also, from the equations (14) and (17):

$$A_{Rij} + R_{Rij} \approx 1 \qquad (18)$$

Equation (14) empirically obtained, and the deduced equations (15), (17) and (18), are very interesting because they relate relative affinity and relative relevance between nodes.
Regarding performance, our model is mathematically expressed with matrices (algorithms 1 and 2), but its implementation in software is inefficient both in processing time and in memory utilization. For this reason, we do not use adjacency matrices but linked lists of adjacency. For example, the tracked *Facebook* network has 968.810 nodes and 2.742.523 edges, with an average degree of only 5,66. To represent this network with an adjacency matrix, we need a matrix of 968.810^2 cells. To represent the same network with linked adjacency lists, we need only $968.810 * 5.66$ memory locations. The amount of memory required with adjacency matrices is 171.167,84 times larger than the amount of memory required with linked lists of adjacency. The space of memory used is in the order of $O(n * \langle k \rangle)$ where *n* is the number of nodes and $\langle k \rangle$ is the average grade of the graph. Clearly $n \gg \langle k \rangle$. The simulators *LM*, *LTM*, *ICM* and *WCM* were constructed with both adjacency matrix and linked lists. In Table 5 we compare the performance of the models with the two data structures. In Table 6 we present the execution times of the four models implemented with linked lists. The performance of the models with the YouTube and Power-Law graphs were very similar to the results presented in Table 5 and Table 6.

Table 5 *Performance of simulators implemented with tracked Facebook network and different data structures*

LM, LTM, ICM, WCM			
Data structure	Required memory positions	Relative memory size	Execution time
Adjacency matrix	n^2	$\frac{n}{\langle k \rangle} = 171.167,84$	It was not possible to run the simulator because the size of the matrix overflowed the computer memory
linked lists	$n * \langle k \rangle$	1	Less than 20 minutes

Table 6 *Execution times (mm:ss) of simulators, implemented with tracked Facebook network and linked lists*

LM:			*LTM*	*ICM*	*WCM*
Nonhomophilic	*Homophilic*	*Informative*			
13:14	18:39	15:20	6:22	6:32	19:27

In Table 7 we show the accuracy of the results expressed through the root mean square error $\left(RMSE = \sqrt{\frac{\sum_1^n (Predicted_i - Actual_i)^2}{n}}\right)$ of the predictions of *LM*, *ICM*, *LTM* and *WCM* regarding empirical experiments. The error of *LM* was calculated with the curve that marks the upper borderline of the gray bands, which is the one that produces the maximum error.

Table 7 Accuracy: Root Mean Square Error

Message class	*LM*	*LTM*	*ICM*	*WCM*
Nonhomophilic	53	238	1,759	241,352
Homophilic	52	236	1,757	241,350
Informative	58	241	1,761	241,357

Table 8 shows the precision in terms of standard deviation of the predictions of *LM*, *LTM*, *ICM* and WCM shown in Figs. Fig. 11, Fig. 12 and Fig. 13. We found that *Lucy Model* obtains significantly more precise results compared to the other models.

Table 8 Precision: Standard Deviation

Lucy Model			*LTM*	*ICM*	*WCM*
Nonhomophilic	Homophilic	Informative			
12.47	15.69	6.99	861	81	356,667

5.3 Findings

A message of homophilic influence and a message of nonhomophilic influence are of a different nature, and therefore their respective activation thresholds are a function of different variables. But, we found that their node activation curves are practically the same in laboratory tests, and similar in field experiments, That is, the two classes of messages produce the same results. These results led us to the conclusion that it is not necessary to define two thresholds of activation of nodes but only one, unifying the two variables (affinity and relevance) into a single variable.

6. CONCLUSIONS

Benefits derived from Lucy Model

The most conclusive result of our research is that the predictions made by *Lucy Model* are notably more accurate and precise than those made by the baseline models *Linear Threshold Model* and *Independent Cascade Model*, and by *Weighted Cascade Model*.

The contribution of our work is to identify new elements that are present in the processes of propagation of information and influence in real life, and that have not been taken into account before. These elements were incorporated into *LM*, allowing the model to represent these processes in a finer and closer way to reality. These are (1) three classes of message propagation in social networks: homophilic influence, nonhomophilic influence, and informative. (2) The concept of affinity between nodes and the metric to calculate it. (3) Concept of tie strength defined as the probability that a node receives a message, and the mathematical expression to quantify it. (4) Concept of activation threshold defined as the probability of activation a node. Each node to influence (*j*) has a different activation threshold for each message propagation class and for each influencing node (*i*). (5) Importance of a message for people in general and the perceived importance for each individual in particular. (6) *LM* is configurable to allow flexibility of use to adapt to different levels of importance of a message, as well as to assign different weights to some variables that intervene in the model. It is important to indicate that we design *LM* before knowing experimental data.

On the other hand, the results of node activation of *Lucy Model*, like *LTM*, *ICM*, and *WCM*, have the characteristics of monotonicity and submodularity, concepts that we defined in Section 0.

Regarding the performance of our model, the use of adjacency linked lists instead of adjacency matrices has allowed us to use very small processing time and memory space, and to perform experiments with much larger graphs in the order of millions of nodes at much higher speeds, demonstrating the effectiveness of *LM*.

In relation with the utility of *LM*, some sectors of society (governments, social movements, sellers, etc.) that use OSNs need to know the extent of the spread of information and the influence they transmit through these networks. *LM* will facilitate telecommunication engineers with the task of performing different types of analysis about the characteristics of the information and influence propagation through online social networks, satisfying these sectors´ need. Our proposal can be used to help develop new models that maximize the propagation of messages; to predict the spread of viruses in both computer networks and mobile phone networks that could be propagated even through online social networks. Our model can also help predict the loss of clients in mobile telephony.

Limitation of Lucy Model implementation

The *LM* implementation allows to work with graphs of up to five million nodes., in a computer PC core i7, memory 8 GB and *Ubuntu* (64 bit) operating system. For larger graphs we require hardware with more capacity.

Difficulty of validation

In order to validate the results of this kind of models, it is necessary to trace the corresponding social network, a task that is very cumbersome and long, since it could last several weeks and even months. This is due to the security measures imposed

by the companies that own the online social networks such as *Facebook*. To overcome this difficulty, synthetic networks are used.

Future work

LM considers only homophilic or only nonhomophilic processes of propagation of influence. It would be important to design a model that considers both types of influence at the same time. On the other hand, our model only considers the propagation of influence in a particular sense; it would be very interesting to model the propagation of influence in at least two senses or opposite influences. We suggest studying the propagation of nonhomophilic influence in an online social network, where accepting influence carries risks, for example: acquiring a new technology or investing in a new business. Another future work could be to predict not only how many nodes but which nodes will be activated. We recommend studying the propagation processes with networks that evolve their topology over time using *LM* model. Another future work could be to determine the time expressed in hours or days that takes the activation of a certain percentage of nodes in a network. On the other hand, it may be interesting to study the propagation characteristics of negative feelings, bad news and fake news. Finally, we recommend studying the maximization of activation with *LM*.

ACKNOWLEDGMENT

R.M.O-G. thanks Dr. Joan Serrat Fernández, Dr. Xavier Muñoz and Dr. Josep Fàbrega, professors at the University Polytechnic of Catalonia. Similarly, Dr. Esteban Samaniego Alvarado and PhD student Vladimiro Tobar Solano, professors at Universidad de Cuenca, and the PhD student Lucía Mendez Tapia for the accomplishment of this work.

References

Abbas SMA (2013) An agent-based model of the development of friendship links within Facebook. Computational and Mathematical Organization Theory 19:232–252. https://doi.org/10.1007/s10588-013-9156-z

Afrasiabi Rad A, Benyoucef M (2012) Measuring propagation in online social networks: the case of youtube. Journal of Information Systems Applied Research 5:26

Ahmed S, Kabir A, Sharmin S, Jafrin S (2017) Cyber-crimes Against Womenfolk on Social Networks: Bangladesh Context. IJCA 174:9–15. https://doi.org/10.5120/ijca2017915407

Alexander CK, Sadiku MNO (2017) Fundamentals of electric circuits, Sixth edition. McGraw-hill Education, New York, NY

Apte M, Palshikar GK, Baskaran S (2019) Frauds in Online Social Networks: A Review. In: Özyer T, Bakshi S, Alhajj R (eds) Social Networks and Surveillance for Society. Springer International Publishing, Cham, pp 1–18

Arendt DL, Blaha LM (2015) Opinions, influence, and zealotry: a computational study on stubbornness. Comput Math Organ Theory 21:184–209. https://doi.org/10.1007/s10588-015-9181-1

Babcock M, Cox RAV, Kumar S (2019) Diffusion of pro- and anti-false information tweets: the Black Panther movie case. Comput Math Organ Theory 25:72–84. https://doi.org/10.1007/s10588-018-09286-x

Bandeli KK, Agarwal N (2018a) Analyzing the role of media orchestration in conducting disinformation campaigns on blogs. Computational and Mathematical Organization Theory. https://doi.org/10.1007/s10588-018-09288-9

Bandeli KK, Agarwal N (2018b) Analyzing the role of media orchestration in conducting disinformation campaigns on blogs. Comput Math Organ Theory s10588-018-09288–9. https://doi.org/10.1007/s10588-018-09288-9

Bekiari A, Hassanagas N (2015) Verbal Aggressiveness Exploration through Complete Social Network Analysis: Using Physical Education Students' Class as an Illustration. Int'l J Soc Sci Stud 3:30

Benevenuto F, Rodrigues T, Almeida V, et al (2008) Identifying video spammers in online social networks. In: Proceedings of the 4th international workshop on Adversarial information retrieval on the web - AIRWeb '08. ACM Press, Beijing, China, p 45

Bolthausen E, Wüthrich MV (2013) Bernoulli's Law Of Large Numbers. ASTIN Bulletin: The Journal of the IAA 43:73–79

Brandes U, Fleischer D (2005) Centrality measures based on current flow. In: Annual symposium on theoretical aspects of computer science. Springer, pp 533–544

Broniatowski DA, Reyna VF (2019) To illuminate and motivate: a fuzzy-trace model of the spread of information online. Comput Math Organ Theory. https://doi.org/10.1007/s10588-019-09297-2

Bulumulla C, Chan J, Padgham L (2018) Enhancing Diffusion Models by Embedding Cognitive Reasoning. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, pp 744–749

Carley KM, Cervone G, Agarwal N, Liu H (2018) Social Cyber-Security. In: Thomson R, Dancy C, Hyder A, Bisgin H (eds) Social, Cultural, and Behavioral Modeling. Springer International Publishing, pp 389–394

Chen W, Lakshmanan LVS, Castillo C (2013) Information and Influence Propagation in Social Networks. Synthesis Lectures on Data Management 5:1–177. https://doi.org/10.2200/S00527ED1V01Y201308DTM037

Cui L, Hu H, Yu S, et al (2018) DDSE: A novel evolutionary algorithm based on degree-descending search strategy for influence maximization in social networks. Journal of Network and Computer Applications 103:119–130

da Silva AR, Rodrigues RF, da Fonseca Vieira V, Xavier CR (2018) Influence Maximization in Network by Genetic Algorithm on Linear Threshold Model. In: Gervasi O, Murgante B, Misra S, et al. (eds) Computational Science and Its Applications – ICCSA 2018. Springer International Publishing, pp 96–109

De Meo P, Ferrara E, Fiumara G, Ricciardello A (2012) A novel measure of edge centrality in social networks. Knowledge-based systems 30:136–150

Dietz K, Heesterbeek JAP (2002) Daniel Bernoulli's epidemiological model revisited. Mathematical biosciences 180:1–21

Duong-Ba, Thuan Hong (2014) Resource allocation optimization in large scale distributed systems. Dissertation, Oregon State University

Fan W, Yeung KH (2011) Online social networks—Paradise of computer viruses. Physica A: Statistical Mechanics and its Applications 390:189–197. https://doi.org/10.1016/j.physa.2010.09.034

Fischetti M, Kahr M, Leitner M, et al (2018) Least cost influence propagation in (social) networks. Mathematical Programming 170:293–325

Fonseca A, Louçã J (2018) Explaining the emergence of online popularity through a model of information diffusion. Comput Math Organ Theory 24:169–187. https://doi.org/10.1007/s10588-017-9253-5

Frantz TL, Carley KM (2017) Reporting a network's most-central actor with a confidence level. Comput Math Organ Theory 23:301–312. https://doi.org/10.1007/s10588-016-9229-x

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. PNAS 99:7821–7826. https://doi.org/10.1073/pnas.122653799

Golbeck J (2013) Analyzing the Social Web. Newnes

González Toral, Santiago (2018) Running scrapper. In: Gist. https://gist.github.com/santteegt/2e70fe88b67ce52842ec451bd53ac4d2. Accessed 20 Mar 2019

Granovetter MS (1973) The Strength of Weak Ties. American Journal of Sociology 78:1360–1380

Guille A, Hacid H, Favre C, Zighed DA (2013) Information diffusion in online social networks: A survey. ACM Sigmod Record 42:17–28

Hein O, Schwind M, König W (2006) Scale-free networks. Wirtschaftsinformatik 48:267–275

Ishengoma FR (2013) Online Social Networks and Terrorism 2.0 in Developing Countries. 12

Ito S, Vymětal D, Šperka R, Halaška M (2018) Process mining of a multi-agent business simulator. Comput Math Organ Theory 24:500–531. https://doi.org/10.1007/s10588-018-9268-6

Jin L, Chen Y, Wang T, et al (2013) Understanding user behavior in online social networks: A survey. IEEE Communications Magazine 51:144–150

Kao H-T, Yan S, Huang D, et al (2019) Understanding Cyberbullying on Instagram and Ask.Fm via Social Role Detection. In: Companion Proceedings of The 2019 World Wide Web Conference. ACM, New York, NY, USA, pp 183–188

Kayes I, Iamnitchi A (2017) Privacy and security in online social networks: A survey. Online Social Networks and Media 3–4:1–21. https://doi.org/10.1016/j.osnem.2017.09.001

Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 137–146

Kosorukoff A, Passmore DL (2011) Social Network Analysis: Theory and Applications. Passmore, D. L

Lee C, Sung C, Ma H, Huang J (2019) IDR: Positive Influence Maximization and Negative Influence Minimization Under Competitive Linear Threshold Model. In: 2019 20th IEEE International Conference on Mobile Data Management (MDM). pp 501–506

Li L, Alderson D, Doyle JC, Willinger W (2005) Towards a theory of scale-free graphs: Definition, properties, and implications. Internet Mathematics 2:431–523

Luceri L, Braun T, Giordano S (2019) Analyzing and inferring human real-life behavior through online social networks with social influence deep learning. Appl Netw Sci 4:34. https://doi.org/10.1007/s41109-019-0134-3

Luo W, Tay WP, Leng M (2016) Infection Spreading and Source Identification: A Hide and Seek Game. IEEE Transactions on Signal Processing 64:4228–4243. https://doi.org/10.1109/TSP.2016.2558168

Mareswara Rao P, Rajashekara Rao K (2019) Extended Security Model over Data Communication in Online Social Networks. In: Satapathy SC, Joshi A (eds) Information and Communication Technology for Intelligent Systems. Springer Singapore, pp 239–249

Masayuki Yano, James Douglass Penn, George Konidaris, Anthony T Patera (2013) Math, Numerics, & Programming (for Mechanical Engineers) - Buscar con Google. https://ocw.mit.edu/ans7870/2/2.086/S13/MIT2_086S13_Textbook.pdf. Accessed 20 Mar 2019

Mathew B, Dutt R, Goyal P, Mukherjee A (2019) Spread of Hate Speech in Online Social Media. In: Proceedings of the 10th ACM Conference on Web Science - WebSci '19. ACM Press, Boston, Massachusetts, USA, pp 173–182

Maturo F, Migliori S, Paolone F (2018) Measuring and monitoring diversity in organizations through functional instruments with an application to ethnic workforce diversity of the U.S. Federal Agencies. Comput Math Organ Theory. https://doi.org/10.1007/s10588-018-9267-7

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a Feather: Homophily in Social Networks. Annu Rev Sociol 27:415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Michalski R (2015) Linear threshold model in temporal networks — Seed selection for social influence. In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). pp 922–923

Mondani H (2018) The underlying geometry of organizational dynamics: similarity-based social space and labor flow network communities. Comput Math Organ Theory 24:378–400. https://doi.org/10.1007/s10588-017-9260-6

Mungovan D, Howley E, Duggan J (2011) The influence of random interactions and decision heuristics on norm evolution in social networks. Comput Math Organ Theory 17:152–178. https://doi.org/10.1007/s10588-011-9085-7
N. S, B. A, Bhattacharya S (2018) Influence maximization in large social networks: Heuristics, models and parameters. Future Generation Computer Systems 89:777–790. https://doi.org/10.1016/j.future.2018.07.015
Nan N, Zmud R, Yetgin E (2014) A complex adaptive systems perspective of innovation diffusion: an integrated theory and validated virtual laboratory. Comput Math Organ Theory 20:52–88. https://doi.org/10.1007/s10588-013-9159-9
Newman M (2010) Networks: An Introduction. Oxford University Press
Nova FF, Rifat MDR, Saha P, et al (2019) Online sexual harassment over anonymous social media in Bangladesh. In: Proceedings of the Tenth International Conference on Information and Communication Technologies and Development - ICTDX '19. ACM Press, Ahmedabad, India, pp 1–12
Paniagua J, Rivelles R, Sapena J (2019) Social Determinants of Success: Social Media, Corporate Governance and Revenue. Sustainability 11:5164. https://doi.org/10.3390/su11195164
Petróczi A, Bazsó F, Nepusz T (2006) Measuring tie-strength in virtual social networks
Phadke C, Uzunalioglu H, Mendiratta VB, et al (2013) Prediction of subscriber churn using social network analysis. Bell Labs Technical Journal 17:63–76
Piqueira JRC, Araujo VO (2009) A modified epidemiological model for computer viruses. Applied Mathematics and Computation 213:355–360. https://doi.org/10.1016/j.amc.2009.03.023
Postigo-Boix M, Melús-Moreno JL (2018) A social model based on customers' profiles for analyzing the churning process in the mobile market of data plans. Physica A: Statistical Mechanics and its Applications 496:571–592
Roberts SG, Dunbar RI (2011) Communication in social networks: Effects of kinship, network size, and emotional closeness. Personal Relationships 18:439–452
Robertson C, Fernandez L, Shillair R (2019) The Political Outcomes of Unfriending: Social Network Curation, Network Agreeability, and Political Participation. Social Science Research Network, Rochester, NY
Rong K, Hu G, Lin Y, et al (2015) Understanding business ecosystem using a 6C framework in Internet-of-Things-based sectors. International Journal of Production Economics 159:41–55
Samadi M, Nagi R, Semenov A, Nikolaev A (2018) Seed activation scheduling for influence maximization in social networks. Omega 77:96–114. https://doi.org/10.1016/j.omega.2017.06.002
Samadi M, Nikolaev A, Nagi R (2016) A subjective evidence model for influence maximization in social networks. Omega 59:263–278. https://doi.org/10.1016/j.omega.2015.06.014
Scott J (2017) Social Network Analysis, Fourth edition. SAGE Publications Ltd, Thousand Oaks, CA
Steinert-Threlkeld ZC, Mocanu D, Vespignani A, Fowler J (2015) Online social networks and offline protest. EPJ Data Sci 4:1–9. https://doi.org/10.1140/epjds/s13688-015-0056-y
Swaminathan A (2014) An Algorithm for Influence Maximization and Target Set Selection for the Deterministic Linear Threshold Model. Virginia Polytechnic Institute and State University
Talukder A, Alam MGR, Tran NH, Hong CS (2018) A cost optimized reverse influence maximization in social networks. In: NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium. pp 1–9
Teixeira AS, Monteiro PT, Carriço JA, et al (2013) Spanning edge betweenness. In: Workshop on mining and learning with graphs. pp 27–31
Uddin S, Khan A, Hossain L, et al (2015) A topological framework to explore longitudinal social networks. Comput Math Organ Theory 21:48–68. https://doi.org/10.1007/s10588-014-9176-3
Ugander J, Backstrom L, Marlow C, Kleinberg J (2012) Structural diversity in social contagion. Proceedings of the National Academy of Sciences 109:5962–5966. https://doi.org/10.1073/pnas.1116502109
Vishwakarma DK, Varshney D, Yadav A (2019) Detection and veracity analysis of fake news via scrapping and authenticating the web search. Cognitive Systems Research 58:217–229. https://doi.org/10.1016/j.cogsys.2019.07.004
Wang F, Jiang W, Li X, Wang G (2018) Maximizing positive influence spread in online social networks via fluid dynamics. Future Generation Computer Systems 86:1491–1502. https://doi.org/10.1016/j.future.2017.05.050
Weng X, Liu Z, Li Z (2016) An Efficient Influence Maximization Algorithm Considering Both Positive and Negative Relationships. In: 2016 IEEE Trustcom/BigDataSE/ISPA. pp 1931–1936
Worrell JC, Rumschlag J, Betzel RF, et al (2017) Optimized connectome architecture for sensory-motor integration. Network Neuroscience 1:415–430
Yang S, Keller FB, Zheng L (2016) Social Network Analysis: Methods and Examples. SAGE Publications
Yun Q, Gloor PA (2015) The web mirrors value in the real world: comparing a firm's valuation with its web network position. Computational and Mathematical Organization Theory 21:356–379. https://doi.org/10.1007/s10588-015-9189-6
Zainudin NM, Merabti M, Llewellyn-Jones D (2011) Online social networks as supporting evidence: A digital forensic investigation model and its application design. In: 2011 International Conference on Research and Innovation in Information Systems. pp 1–6
Zuo X, Blackburn J, Kourtellis N, et al (2016) The power of indirect ties. Computer Communications 73:188–199
Snap.py - SNAP for Python. https://snap.stanford.edu/snappy/. Accessed 20 Mar 2019b
SNAP: Network datasets: Youtube social network. http://snap.stanford.edu/data/com-Youtube.html. Accessed 20 Mar 2019a