A Novel Influence Diffusion Model based on User Generated Content in Online Social Networks

Flora Amato, Antonio Bosco, Vincenzo Moscato, Antonio Picariello and Giancarlo Sperlí
University of Naples "Federico II", Via Claudio 21, 80125 Naples, Italy
CINI - ITEM National Lab, Complesso Universitario Monte Santangelo, 80125, Naples, Italy

Keywords: Multimedia Social Network, Influence Analysis, Big Data.

Abstract: Social Network Analysis has been introduced to study the properties of Online Social Networks for a wide range of real life applications. In this paper, we propose a novel methodology for solving the Influence Maximization problem, i.e. the problem of finding a small subset of actors in a social network that could maximize the spread of influence. In particular, we define a novel influence diffusion model that, learning recurrent user behaviours from past logs, estimates the probability that a given user can influence the other ones, basically exploiting user to content actions. A greedy maximization algorithm is then adopted to determine the final set of influentials in the network. Preliminary experimental results shows the goodness of the proposed approach, especially in terms of efficiency, and encourage future research in such direction.

1 INTRODUCTION

The growing popularity of Online Social Networks (OSN), and in particular, their huge amount of data lay the foundations for analyzing several sociological phenomena useful for a large number of applications.

From a technological point of view, OSNs are enabled as Internet applications that provide a set of common functionalities, such as: i) information sharing capabilities, ii) user generated content management and iii) support by means of several tools to different ways of communication and collaboration among users. Thus, OSN can be seen as a digital platform that "lives" on the Web and which content and services are delivered to final users through a variety of technological devices.

From a sociological perspective, OSNs are social structures constituted by a set actors (individuals or organizations), sets of dyadic ties, and other social interactions, often instantiated through the shared information, among actors themselves. In this case, mathematical models can be adopted to study the social network structure, the related generated content and its temporal evolution.

The most natural and simple way to model an OSN is to use a *directed graph* composed by a set of nodes/vertices – representing the individual actors of a social community – and a set of edges – representing the different kinds of relationships among actors, in many cases instantiated through the user generated content.

Social Network Analysis (SNA) (Scott, 2012) has been introduced to study the properties of OSNs with the aim of supporting a wide range of applications: information retrieval, influence analysis, recommendation, viral marketing, event recognition, expert finding, community detection, user profiling, security and social data privacy and so on.

Many OSNs are extremely rich in information, and they typically contain a tremendous amount of content and linkage data which can be jointly exploited for analysis. The linkage data is essentially the graph structure of the social network and the communications between entities; whereas the content data contains the text, images and other type of content in the network

The majority of SNA techniques mainly exploit "user to user" interactions, leveraging the graph theory as powerful tool to support the different kinds of analytics.

More recently, in according to a data-centric view of OSN, also "user to content"' relationships have been considered together with content features to provide more advanced forms of analysis.

Generally, the SNA techniques can be inspired by two different approaches (Aggarwal, 2011):

- *Linkage-based and Structural Analysis*: an analy-

Amato, F., Bosco, A., Moscato, V., Picariello, A. and Sperlí, G.
A Novel Influence Diffusion Model based on User Generated Content in Online Social Networks.
DOI: 10.5220/0006486703140320
In *Proceedings of the 6th International Conference on Data Science, Technology and Applications (DATA 2017)*, pages 314-320
ISBN: 978-989-758-255-4

sis of the linkage behavior of the network is performed in order to determine important nodes, communities, links, and evolving regions of the network.

- *Content-based Analysis*: Many social networks, such as Flick and Youtube, contain a tremendous amount of content (multimedia and tags) which can be leveraged to improve the analysis.

In this paper, we propose an hybrid SNA methodology to determine the most influential actors (*influentials*) wihin an OSN. In particular, we define a novel influence diffusion model that learns recurrent user behaviours from past logs to estimate probability that a given user can influence the other ones, basically exploiting user to content actions. Then, a greedy maximization algorithm is adopted to determine the final set of influentials.

The paper is organized as in the following. Section 2 introduces the theoretical background for our work. Section 3 presents a novel influence diffusion model for OSNs and some properties related to it, useful for defining the *influence graph*. Section 4 describes the adopted greedy algorithm for the influence maximization based on the influence graph. Finally several experiments are discussed in Section 5, while related works and conclusion are reported in Sections 6 and 7 respectively.

2 THEORETICAL BACKGROUND

As prevously described, in a social network links among nodes usually represent a variety of relationships, from intimate friendships to common interests for a multimedia object (e.g., tweet, post, video, photo, etc.): they determine a "flow" of information and hence indicate a user's *influence* on the others, a concept that is crucial in sociology and viral marketing. As well known, studying influence patterns can help us better understand why certain trends or innovations are adopted faster than others and how we could help advertisers and marketers design more effective campaigns.

Traditional communication theories state that a minority of users, called *influentials*, excel in persuading others; a more modern view, in contrast, deemphasizes the role of influential and posits that the key factors determining influence are the different kinds of relationship among ordinary users and the readiness of a society to adopt an innovation (Watts and Dodds, 2007).

Moving from theory into practice, an *influence analysis* problem can be faced using two steps.

In the first one, a *diffusion model* is defined with the aim of describing the influence spread in the network; this phenomenon is usually modeled by a stochastic process where the activation of each node is based on its neighbours state. In the second step, a *maximization algorithm* is exploited to identify the set of nodes such that their activations maximize the diffusion or the propagation of influence.

The selection of the most influence nodes is an optimization problem that has been proven by Kempe et al.(Kempe et al., 2003) to be NP-Hard. In particular, chosen S as the initial active seed-set, Kempe et al. defined its influence $\sigma(S)$ as the total number of activated nodes in the network at the end of the diffusion process.

The influence function $\sigma(S)$ maps subsets of elements of a finite set to a non-negative number. The final goal is to find a k-element seed set S that maximizes $\sigma(S)$, which is a NP-Hard problem. To address this complexity, several greedy strategies exploiting a *non-negative*, *submodular* and *monotone* influence function have been proposed to obtain a solution that is no worse than $(1-1/e)$ of the optimal one.

In this paper, we model influence and the related spread in a novel way, as depicted in the following.

Let Giank and Boscus be two users of a OSN. Evidence suggests that user Giank influences Boscus if an action of Giank at time t causes one or more actions of Boscus at time $t' > t$. As an example considering user to content actions, if Giank publishes a certain photo at t his influence on Boscus can be proved by the fact that at time $t_1 > t$, Boscus puts a "like" on the picture, and that at time $t_2 > t_1 > t$ he publishes an audio that is semantically similar to the Giank's photo. In addition, the more the two multimedia objects are similar, the more user Giank has influenced user Boscus.

Assuming the same hypothesis of Kempe, we will define a maximization algorithm based on this novel influence model.

3 THE INFLUENCE MODEL

The basic assumption in our model is the existence of a finite set of *Actions* (A) representing all the possible "interactions" among the set of *Users* (U) and the set of *Objects* (O) in one ore more online social networks, which can be properly captured during user browsing sessions exploiting log information (Sang et al., 2015; Guo et al., 2016).

In particular, we denote with $A_u(t)$ the set of actions performed by a given user $u \in U$ in a specific instance of time t. Similarly, we indicate with $A_u(t, \Delta t)$ the set of actions of u within the interval $]t, \ldots, \Delta t]$.

In such a context, we can consider different examples of actions: users' reactions or comments to user generated content (e.g., post, pictures and so on), post visualization or rating, for giving few examples. Table 1 summarizes the available user-to-content relationships in the most diffused social networks.

In the following, we will describe the foundamental aspects of the proposed influence model.

Definition 3.1 (Log tuple). *A* log tuple *can be defined by the information $l = (a, u, o, \lambda_1, \cdots, \lambda_k)$, where $a \in A$, $u \in U$, $o \in O$ and $\lambda_1, \cdots, \lambda_k$, are particular attributes (e.g., timestamp, type of reaction, text and tags of a comment, etc.) used to describe an action.*

Definition 3.2 (Log). *A* Log *(L) is a finite sequence of log tuples.*

Intuitively, a log tuple corresponds to an observation of $l.a$ performed by the user $l.u$ on a given object $l.o$ along with the associated attributes of the observation $\lambda_1, \cdots, \lambda_k$. By convention, if action a_2 occurs after a_1 in a log, then the action a_2 occurred temporally after a_1.

Using information on past logs, we can introduce a *reaction operator* between two actions .

Definition 3.3 (Reaction Operator). *The* Reaction Operator *$reac^{\Delta t}(a_1, a_2)$ between two actions a_1 of user u_i and a_2 of user u_j – and both the actions are performed on the same object o (or on similar objects[1]) – returns the probability that a_2 occurs after a_1 within the interval Δt.*

It is simple to observe that the following property stands for the reaction operator:

$$
\begin{aligned}
reac^{\Delta t}(a_1, a_2) \geq \tau^1 \wedge reac^{\Delta t}(a_2, a_3) \geq \tau^2 \\
\rightarrow reac^{\Delta t}(a_1, a_3) \geq \tau^3 \\
\tau^3 = f(\tau^1, \tau^2)
\end{aligned}
$$

f being a function whose value is less then $min(\tau^1, \tau^2)$.

Example 3.1 (Log and reaction operator). *Consider a log, obtained from Flickr, whose associated sequence of actions is:*

$\langle$ *publishing, vinni, photo1,10/05/2017 13:30, 'sunset'* $\rangle$,
$\langle$ *like, flora, photo1,10/05/2017 13:31* $\rangle$,
$\langle$ *comment, flora, photo1,10/05/2017 13:32, 'wonderful'* $\rangle$,
$\langle$ *publishing, picus, photo2,10/05/2017 13:38, 'sunset'* $\rangle$,
$\langle$ *like, giank, photo2,10/05/2017 13:40* $\rangle$,
$\langle$ *like, boscus, photo2,10/05/2017 13:42,* $\rangle$,
$\langle$ *comment, vinni, photo2,10/05/2017 13:45,''you too..'* $\rangle$
$\langle$ *like, boscus, photo1,10/05/2017 13:47* $\rangle$

It is simple to observe that considering Δt =2 minutes the reaction operator returns a probability value of 1 *for the couple of actions (publishing,like). In turn, the probability value is* 0.5 *for the couple of actions (publishing,comment) and* 0.33 *for (like,like). If we consider a more wide temporal interval and assume the images published by picus and vinni very similar, the reaction operator will return a probaility value of* 0.5 *for the couple of actions (publishing,publshing) .*

Definition 3.4 (Influence Operator). *Let $u_1, u_2 \in U$ be respectively two users, we say that $u_1 \underset{\tau}{\rightarrow} u_2$, if each action $a_{u_1} \in A_{u_1}(t)$ of user u_1 at time t determines an action $a_{u_2} \in A_{u_2}(t, \Delta t)$ of user u_2 in the interval $]t, \ldots, \Delta t]$ within a log L:*

$$
u_1 \underset{\tau}{\rightarrow} u_2 \iff \forall t_i \in T \exists a_1 \in A_{u_1}(t_i), a_2 \in A_{u_2}(t_i, \Delta t) \in L : reac^{\Delta t}(a_1, a_2) \geq \tau
$$

$T = \{t_1, t_2, \ldots, t_m\}$ being a sequence of temporal instants such that $t_1 < t_2 < \ldots t_m$ and $\tau \in [0, 1]$ a probability value.

The *influence operator* estimates the influence that user u_1 exerts on u_2 within the time $t + \Delta t$.

Example 3.2 (Log and influence operator). *Considering the log of Example 3.1 as past log (it can be used for the learning of reaction operator) and as current log the following sequence of actions:*

$\langle$ *publishing, vinni, photo1,11/05/2017 19:30, 'california'* $\rangle$,
$\langle$ *like, flora, photo1,10/05/2017 19:31* $\rangle$,
$\langle$ *like, boscus, photo1,10/05/2017 19:32* $\rangle$
$\langle$ *comment, flora, photo1,10/05/2017 19:32, 'wow!!!!!!!'* $\rangle$,
$\langle$ *publishing, picus, photo2,10/05/2017 19:58, 'hollywood'* $\rangle$,
$\langle$ *like, giank, photo2,10/05/2017 19:59* $\rangle$,
$\langle$ *like, boscus, photo2,10/05/2017 19:59* $\rangle$

It is simple to observe that considering Δt =2 minutes we say that user vinni certainly influences flora and bosco and user picus certainly influences giank and boscus.

3.1 Properties and Theorem

We assume that the following properties stand for the influence operator:

Property 3.1 (Not Self Reflexivity). $u_1 \underset{\tau}{\nrightarrow} u_1$

Property 3.2 (Not Commutativity). $u_1 \underset{\tau}{\rightarrow} u_2 \nRightarrow u_2 \underset{\tau}{\rightarrow} u_1$

Property 3.3 (Not Transitivity). $u_1 \underset{\tau}{\rightarrow} u_2 \wedge u_2 \underset{\tau}{\rightarrow} u_3 \nRightarrow u_1 \underset{\tau}{\rightarrow} u_3$

[1]The evaluation of such condition needs the defiinition of a similarity function between two objects.

Table 1: User-to-Content relationships in Online Social Networks.

	Twitter	Facebook	Instagram	Google+	Last.FM	Flickr
Publishing	X	X	X	X	X	X
Tagging	X	X	X	X	X	X
Comment	X	X	X	X	X	X
Like	X	X	X	X		X
Resharing	X	X		X		
Favorites	X				X	X
Visualization		X	X		X	X

Theorem 3.1 (Influence Diffusion)**.** *Let u_1,u_2 and u_3 three users and L a given log,*

$$\begin{cases} u_1 \underset{\tau^1}{\rightarrow} u_2 \\ u_2 \underset{\tau^2}{\rightarrow} u_3 \end{cases} \Rightarrow u_1 \underset{\tau^3}{\rightarrow} u_3 \qquad (1)$$

$\tau^3 \leq min(\tau^1, \tau^2)$.

Proof. Let us consider the definition of influence operator and the property of reaction operator; we observe that, for the theorem hypothesis, there will always exist the two actions $a_{u_1} \in A_{u_1}(t)$ and $a_{u_3} \in A_{u_3}(t : \Delta t')$ and the reaction operator will always return for the couple (a_1, a_3) a probability $\tau^3 \leq min(\tau^1, \tau^2)$. □

3.2 Influence Graph

Definition 3.5 (Influence Graph)**.** *An* Influence Graph *is a labeled graph $G = (V, E, \tau)$ where:*

- *V is the set of nodes such that each $v \in V$ corresponds to a user $u \in U$;*
- *$E \subseteq V \times V$ is the set of edge (with no self-loops);*
- *$\tau : V \times V \rightarrow [0, 1]$ is a function that assigns to each edge $e = (v_i, v_j)$ a label, representing the probability that user u_i can influece user u_j.*

Example 3.3 (Example of Influence Graph)**.** *In the Figure 1, we show how from the log of Example 3.2 it is possible to derive an influence graph on the base of defined influence and reaction operators.*

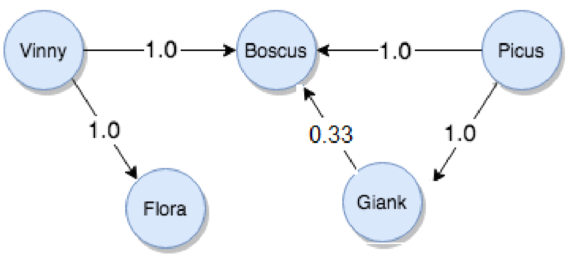


Figure 1: Influence Graph.

Definition 3.6 (Direct Influence)**.** *Let $G = (V, E, \tau)$ and u_i,u_j be respectively an influence graph and two users, we say that user u_i* directly influences *u_j if there exists an edge that connects u_i to u_j.*

Definition 3.7 (Indirect Influence)**.** *Let $G = (V, E, \tau)$ and u_i,u_j be respectively an influence graph and two users, we say that user u_i* indirectly influences *u_j if there exists a path $p = (e_1, \ldots, e_n)$, with $e_1, e_2, \ldots, e_n \in E$ that connects u_i to u_j.*

4 τ-greedy ALGORITHM

In this section we describe the proposed approach for influence maximization problem exploiting the influence graph.

To better explain our idea, we firstly analyze the chosen model to design the spread of influence over the network. We assume that each node of the graph could be either an *active* node or *inactive node*. Then, we choose a *Linear Threshold* (LT) model (Kempe et al., 2003) for describing the influence spread because of it represents in very effective way the typical behavior of a user that is led to adopt a new idea or technology as more of his friends become active.

For this reason, we define a *reactive threshold* for each node describing the weighted percentage of its neighbors needed to led a specific user to adopt a given idea. This value corresponds to a lower bound for activating each node and is computed considering her/his behavior respect to the action performed by the community of users. More in details, given a specific interval of analysis $]t, \ldots, \Delta t]$, we compute for each user u the reactive threshold $\theta(u, \Delta t)$ on the basis of the ratio between the average number of actions of u in any interval $\hat{A}_u(t, \Delta t)$ and the maximum number of actions that it is possible to observe in the log within the same interval:

$$\theta(u, \Delta t) = 1 - \frac{\hat{A}_u(t, \Delta t)}{max_t(A(t, \Delta t))} \qquad (2)$$

Our idea is to leverage the greedy strategy proposed by Kempe et al.(Kempe et al., 2003), taking into account the τ value in the influence function together with the reactive threshold, that corresponds to the probability that a user is activated based on a specific weighted amount of its neighbor.

Thus, we provide the following definition of *influence function*.

Definition 4.1 (Influence function)**.** *Let* $G = (V, E)$ *be an influence graph,* S *a set of seed nodes and* $\delta(v)$ *a function that returns true if node* v *is actived, we define* influence function*:*

$$\sigma_\tau(S) = \sum_{v \in (V-S)} \sum_{u \in S} \tau_{uv} * \delta(v)$$

Our approach, called τ - *Greedy* (see Algorithm 1), provides in addition an a-priori pruning strategy based on τ values to reduce the space and time complexity of the problem. In particular, given a user defined threshold, we analyze the spread of influence through direct or indirect paths that allows to obtain an overall influence values greater than τ_s.

Algorithm 1: τ - Greedy Algorithm.

procedure τ - GREEDY ALGORITHM(G,τ_s)
    $S \leftarrow 0$
    for $i \leftarrow 1, k$ **do**
        $j \leftarrow argmax_{v \in V}(\sigma_\tau(S \cup v) - \sigma_\tau(S))$
        $S \leftarrow S \cup \{j\}$
    return S

5 EXPERIMENTS

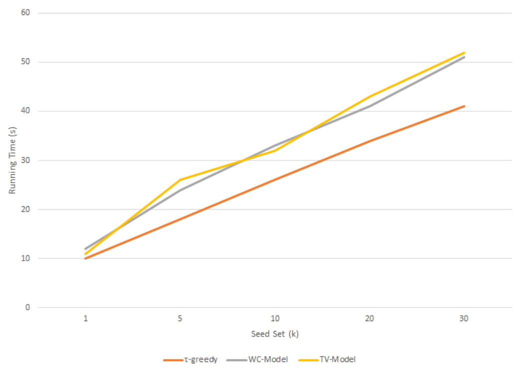
The experiments were performed using Databricks[2], a cloud-based big data processing environment based on Spark, using 5 computing nodes each one composed by 4 core and 30 GB Memory, on which are installed Spark 2.1.0 and Hadoop 2.7.3. Experiments for efficiency evaluation have been carried on the *Yahoo Flickr Creative Commons 100 Million Data* (YFCC100M)[3] dataset. Using the Flickr API[4] we extract the main social information and the different actions that compose a log. Table 2 provides the characterization of the dataset.

Table 2: Dataset characterization.

Log	Influence Graph	
Social elements	Nodes	Edges
17.493	1.264	4.686

Then, we performed a comparison of the execution times using the proposed Influence Graph, with respect to other influence diffusion models.

In particular, we computed the running times of τ-greedy algorithm on our proposed Influence Graph and compared them with the same approach based on another graph. We built the second graph by instantiating a direct edge between two node u_i and u_j if



(a) Running Times varying seed set

(b) Running Times varying τ_s

Figure 2: Efficiency of the τ-greedy algorithm using different influence diffusion models.

there is an action in the log from the first user to the second one (both performed on the same object within a given temporal interval), and we assign a weight to it using the two following methods:

- *Weighted Cascade Model*: in which the probability of in-coming edges for a vertex v is equal to $\frac{1}{d_{in}(v)}$, $d_{in}(v)$ being the in-degree of v.
- *Trivalency Model*: in which each edge weight is a random probability chosen from the three values: $0.1, 0.01, 0.001$.

In Figure 2, we show the running times varying the number of seed set elements and using different values of τ_s. Note that in Figure 2(b) the TV-Model results are limited to the interval $[0.01, 0.1]$ due to the considered influence probability.

6 RELATED WORK

In the last decade, the huge amount of heterogeneous data that can be extracted from social networks (such

[2]https://www.databricks.com/
[3]https://webscope.sandbox.yahoo.com
[4]https://www.flickr.com/services/api

as Facebook, Twitter, Flickr and so on) led to a continuous growth of interest in the use of such networks for a large variety of applications.

Deciding whether to adopt an innovation (such as a political idea or product), individuals are frequently influenced, explicitly or implicitly, by their social contacts. Indeed, the way in which new practices spread through a population depends mainly on the fact that people influence each others behaviour. It is essential for companies to target “opinion leaders”, as influencing them will lead to a large cascade of further recommendations. This is the goal of each viral marketing and social advertisement campaigns, and corresponds in solving an *influence analysis* problem.

The influence analysis is composed by the following two aspects:

- *Influence Spread/Diffusion*: it consists of the analysis of the influence spread throughout the social network nodes;
- *Influence Maximization*: this means finding the seed set of users that maximize the total number of influenced users over the network.

6.1 Diffusion Models

A social network can be considered as a directed graph $G = (V,E)$, where V is the set of vertices/nodes and E is the set of edges. To design an influence algorithm we have to define a *diffusion model* that describes how the influence is propagated across the network nodes.

The *diffusion models* can be classified into different categories:

- *Stochastic diffusion models:* they exploit a randomized process for computing the influence spread. The most used models are the *Independent Cascade* (IC) and the *Linear Threshold* (LT) (Kempe et al., 2003) models. In the IC model, at each step an *active* node makes an attempt to activate its neighbours in according to a Bernoulli trial. The problem of this model is that a node can be influenced only by another one. The LT model is proposed to overcome this problem where each node could be activated if the sum of the influence weights of its active neighbours is greater than its activation threshold.
- *Epidemic diffusion models:* these approaches consider the influence diffusion as the spread of a disease among biological individuals of a population (Wallinga and Teunis, 2004; Rodriguez and Schölkopf, 2012). A node can be: *susceptible* (the node does not have the disease but upon a contact it will be infected), *infected* (when it has been infected and thus influenced) or *recovered*, (when it was *infected* in the past and now it cannot be infected). In particular, several models have been proposed, and they differ for the chance that a node has to return susceptible or not, after an infection.
- *Percolation theory:* the influence diffusion problem can be studied as a bond percolation from the nodes that belong to the seed-set (Li et al., 2015). This model does not stand for finding the expecting number of active nodes given a seed-set or determining the best k size seed-set that maximizes influence spread.

6.2 Influence Maximization

The *Influence maximization* problem try to identify the top-k nodes that allow to maximize the spread of influence in an OSN. Kempe et al.(Kempe et al., 2003) proved the NP-Hardness of this problem. The proposed approaches can be classified in four groups:

- *Stochastic approaches:* they leverage a randomized process for generating the active set of nodes S for each instance of time, from the initial seed sets. To address the complexity of problem, greedy algorithms can be used to find an approximation of the optimal solution with respect to the stochastic diffusion model. Kempe et al.(Kempe et al., 2003) developed a greedy algorithm that provides an approximation guarantee of $\sigma(S) \geq (1-1/e)\cdot\sigma(S^*)$, assuming that the influence function σ is a monotone and sub-modular function, where S^* is the seed-set that maximizes the value of σ and S is the initial seed set. The basic algorithms in this category are the Kempe et al. approach based on Monte-Carlo approximation, TIM/TIM^+ (Tang et al., 2015) and IMM (Tang et al., 2015). These algorithms use a martingale approach and are based on RIS algorithm (Borgs et al., 2014). The fastest one is IMM that outperforms TIM/TIM^+ and Kempe et al.’s approaches.
- *Biological Inspired approaches:* in this category there are the algorithms inspired by some biological phenomena. Two of main approaches are those inspired by the *bee waggle-dance* when they search food (Karaboga and Basturk, 2007), and by *ant colony optimization* (Yang et al., 2012; Yang et al., 2016).
- *Game Theory approaches:* they consider the influence maximization problem as model where each individual can make a selfish choice in terms of payoff in using a new product. One of the

main approach of this category is the *Multi-Armed Bandit* (MAB) theory (**?**): it leverages a classical probability model where a player has to choose one of a set of arms at each round. Each arm gives a reward to the player, based on some stochastic functions. The player has to choose the best set of arms to maximize the total reward.

- *Genetic algorithms based approaches:* It is possible to find a solution to influence maximization problem in feasible time using genetic algorithm. Promising approaches are the *evolutionary algorithms* (Bucur and Iacca, 2016). Here, individuals evolve during time and can be in several states: *selection*,*reproduction*,*mutation* and *recombination*. This is inspired by the natural selection of the species and can overcome the drawback of the searching for a local maximum of the greedy approach. Such approach is non deterministic but it has results even better compared to deterministic algorithms like greedy ones.

Differently from the other works, in this paper we have provided novel stochastic diffusion model and we have exploited a simple greedy algorithm to maximize the influence spread in the network.

7 DISCUSSION AND CONCLUSIONS

In this paper, we defined a novel influence diffusion model that learns recurrent user behaviours from past logs to estimate probability that a given user can influence the other ones, basically exploiting user to content actions. Then, a greedy maximization algorithm is adopted to determine the final set of influentials. We reported some preliminary experimental results that show the goodness of the proposed approach.

Future work will be devoted to improve the diffusion models defining other properties that can allow to optimize the calculus of the influentials. In addition, we are planning to extend the proposed experimentation considering other influence analysis algorithms and big data coming from heterogeneous networks.

REFERENCES

Aggarwal, C. C. (2011). An introduction to social network data analytics. *Social network data analytics*, pages 1–15.

Borgs, C., Brautbar, M., Chayes, J., and Lucier, B. (2014). Maximizing social influence in nearly optimal time. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 946–957. SIAM.

Bucur, D. and Iacca, G. (2016). Influence maximization in social networks with genetic algorithms. In *European Conference on the Applications of Evolutionary Computation*, pages 379–392. Springer.

Guo, G., Zhang, J., and Yorke-Smith, N. (2016). A novel recommendation model regularized with user trust and item ratings. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1607–1620.

Karaboga, D. and Basturk, B. (2007). Artificial bee colony (abc) optimization algorithm for solving constrained optimization problems. In *International Fuzzy Systems Association World Congress*, pages 789–798. Springer.

Kempe, D., Kleinberg, J., and 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*, pages 137–146. ACM.

Li, D., Zhang, Q., Zio, E., Havlin, S., and Kang, R. (2015). Network reliability analysis based on percolation theory. *Reliability Engineering & System Safety*, 142:556–562.

Rodriguez, M. G. and Schölkopf, B. (2012). Influence maximization in continuous time diffusion networks. *arXiv preprint arXiv:1205.1682*.

Sang, J., Deng, Z., Lu, D., and Xu, C. (2015). Cross-osn user modeling by homogeneous behavior quantification and local social regularization. *IEEE Transactions on Multimedia*, 17(12):2259–2270.

Scott, J. (2012). *Social network analysis*. Sage.

Tang, Y., Shi, Y., and Xiao, X. (2015). Influence maximization in near-linear time: A martingale approach. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, SIGMOD '15, pages 1539–1554, New York, NY, USA. ACM.

Wallinga, J. and Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology*, 160(6):509–516.

Watts, D. J. and Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4):441–458.

Yang, Q., Chen, W.-N., Yu, Z., Gu, T., Li, Y., Zhang, H., and Zhang, J. (2016). Adaptive multimodal continuous ant colony optimization. *IEEE Transactions on Evolutionary Computation*.

Yang, W.-S., Weng, S.-X., Guestrin, C., Faloutsos, C., VanBriesen, J., and Glance, N. (2012). Application of the ant colony optimization algorithm to the influence-maximization problem. *International Journal of Swarm Intelligence and Evolutionary Computation*, 1(1).