Expert Systems With Applications 147 (2020) 113183

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Evolutionary multiobjective optimization to target social network influentials in viral marketing

Juan Francisco Robles[a,*], Manuel Chica[a,b], Oscar Cordon[a]

[a] Andalusian Research Institute DaSCI "Data Science and Computational Intelligence", University of Granada, Granada 18071, Spain
[b] School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia

ARTICLE INFO

Article history:
Received 5 October 2019
Revised 17 December 2019
Accepted 3 January 2020
Available online 10 January 2020

Keywords:
Viral marketing
Influence maximization
Influentials targeting
Social networks
Evolutionary multiobjective optimization
Agent-based modeling

ABSTRACT

Marketers have an important asset if they effectively target social networks' influentials. They can advertise products or services with free items or discounts to spread positive opinions to other consumers (i.e., word-of-mouth). However, main research on choosing the best influentials to target is single-objective and mainly focused on maximizing sales revenue. In this paper we propose a multiobjective approach to the influence maximization problem with the aim of increasing the revenue of viral marketing campaigns while reducing the costs. By using local social network metrics to locate influentials, we apply two evolutionary multiobjective optimization algorithms, NSGA-II and MOEA/D, a multiobjective adaptation of a single-objective genetic algorithm, and a greedy algorithm. Our proposal uses a realistic agent-based market framework to evaluate the fitness of the chromosomes by simulating the viral campaigns. The framework also generates, in a single run, a set of non-dominated solutions that allows marketers to consider multiple targeting options . The algorithms are evaluated on five network topologies and a real data-generated social network, showing that both MOEA/D and NSGA-II outperform the single-objective and the greedy approaches. More interestingly, we show a clear correlation between the algorithms' performance and the diffusion features of the social networks.

1. Introduction

On-line social networks such as Facebook or Instagram make people (potential consumers) more connected than ever before. With just a few actions, consumers can instantly communicate their products and brands' opinions. Social networks' influentials have thousands of friends and a word-of-mouth process can create a cascade of positive or negative information about a brand. This word-of-mouth process in social networks is crucial for marketers and advertisers (Haenlein & Libai, 2017; Leskovec, Adamic, & Huberman, 2007). In fact, people place more value in friends' recommendations than in those from traditional advertisement channels such as TV. Viral marketing (VM) consists of targeting certain consumers to encourage a faster product's adoption (Haenlein & Libai, 2017). The selection of these influentials is not a random but a complex optimization process that involves the analysis of the social network of consumers to trigger a large cascade of adoptions (known in the literature as influence maximization (IM) (Domingos & Richardson, 2001)) thus favoring a positive information diffusion.

Designing the best VM strategies before running them in the real market is possible through the use of artificial social networks (Watts & Dodds, 2007) and simulations using paradigms such as agent-based modeling (ABM) (Epstein, 2006). ABM can describe and simulate network-level interactions between consumers to reproduce word-of-mouth mechanisms that provide realism into marketing models and allow modelers to study emergent behaviors (Epstein, 2006; Janssen & Jager, 2003). These models enable the aggregation of individual-level interactions through the underlying artificial social network, creating high level outcomes and incorporating individual behavioral rules without higher level assumptions (Chica, Chiong, Kirley, & Ishibuchi, 2018).

In this contribution, we propose the use of evolutionary multiobjective optimization (EMO) (Coello, Lamont, & Van Veldhuizen, 2007) together with an ABM market simulation model to locate and target those influentials which are used as promoters of VM campaigns to encourage product adoption via IM through social networks. The ABM simulation model is used to evaluate the fitness of chromosomes. The market model incorporates realistic consumer behavior such as an awareness filtering process (Robles, Chica, & Cordón, 2016) and is inspired by the well-known Consumat marketing model (Janssen & Jager, 2003). Our novel definition of the multiobjective optimization problem jointly

* Corresponding author.
E-mail addresses: jfrobles@ugr.es (J.F. Robles), manuelchica@ugr.es (M. Chica), ocordon@decsai.ugr.es (O. Cordon).

https://doi.org/10.1016/j.eswa.2020.113183
0957-4174/

optimizes the sales revenue and costs of the targeting campaign using local social network measures (Newman, Barabási, & Watts, 2006) to make the problem more realistic. Sales revenue and campaigns' costs are natural conflicting objectives because increasing the total sales of a brand normally requires a higher budget for the marketing campaigns. Although there are single-objective optimization problem formulations for VM (Schlereth, Barrot, Skiera, & Takac, 2013; Stonedahl, Rand, & Wilensky, 2010), there is not any multiobjective approach to jointly optimize these two conflicting objectives up to our knowledge.

From the vast family of EMO algorithms, we use the NSGA-II (Deb, Pratap, Agarwal, & Meyarivan, 2002) and MOEA/D (Zhang & Li, 2007) as they are two well-known algorithms, successfully applied to a large number of multiobjective optimization problems. Our experimentation comprises different marketing scenarios by using five different artificial social network topologies representing the connections between consumers in the market: a real network based on the email communications between University scholars (Guimera, Danon, Diaz-Guilera, Giralt, & Arenas, 2003), a scale-free network (Barabási & Albert, 1999), a Watts-Strogatz small world network (Watts & Strogatz, 1998), an Erdös-Rényi random network, and a regular lattice network. Additionally, and to better validate the scalability of our approach, we use a real data-generated social network of 20,000 nodes, built from the users of a freemium app (Chica & Rand, 2017). We compare the quality of the non-dominated solutions found by the EMO algorithms with a single-objective genetic algorithm (GA) and a greedy algorithm. The design of our developed experimentation also allows us to analyze the influence of the social network topologies on the solutions and algorithms' performance. Finally, we explore the marketing policies obtained from the set of non-dominated solutions returned by the EMO algorithms.

Our approach enhances the current research on the IM problem by: (i) not only optimizing the global influence of adoptions upon the social network (as usual in IM, mainly based on the independent cascade model (Goldenberg, Libai, & Muller, 2001)) but the actual marketing campaigns' revenue and costs resulting from a marketing ABM, (ii) using awareness filters to better represent the consumer agents' purchase processes, resulting in a more realistic behavior, (iii) optimizing the number s of influential nodes to be used in the optimal solutions (i.e., not fixing the value of the number of seeds *a priori* but automatically estimating it to optimize the considered key performance indicators), (iv) returning not only a single but a non-dominated set of influential seeds in a single algorithm execution to provide the marketer with different decision alternatives, and (v) validating the proposed multiobjective influentials targeting optimization method on a real-world problem from a freemium app.

2. State of the art in viral marketing

IM on social networks was first presented as an optimization problem in Domingos and Richardson (2001). Later, Kempe, Kleinberg, and Tardos (2003) showed that IM is essentially an NP-hard optimization problem for most of the considered propagation models. These were the *linear threshold model*, the *weighted cascade model*, and the *independent cascade model*. The later was proposed to model the effects that micro-level word-of-mouth interactions have on macro-level marketing (Goldenberg et al., 2001) and established the roots of the relations between IM and VM.

In their seminal paper, Kempe et al. proposed a greedy hill-climbing algorithm to solve the IM problem, including an optimality approximation guarantee for the three kinds of cascade models. Nevertheless, the greedy heuristic showed some drawbacks: (a) its huge run time (due to the fact that the greedy selection function requires running the objective function to be optimized for every candidate node at each algorithm step), and (b) the need to have a global knowledge of the network. As a consequence, the authors also proposed several greedy heuristics based on social network measures such as degree and closeness centrality. Those algorithms showed a more usual greedy structure with selection functions based on quickly-to-compute indicators, different from those used in the objective function. In this case, s network nodes are selected as seeds in a decreasing order of degree or closeness centrality to generate a solution. This solution is finally evaluated in a Monte-Carlo simulation of the independent cascade model to get the influence spread value. When the considered measure is the degree of the node, the algorithm is called *high-degree heuristic*.

Several extensions of those first greedy algorithms were proposed later. In particular, Leskovec et al. (2007) introduced the *cost-effective lazy forward selection* which exploited the submodularity property to significantly reduce the run time of the greedy hill-climbing algorithm. The rationale is that the expansion of each node is computed *a priori* and only needs to be recomputed for a few nodes. Meanwhile, Chen, Wang, and Yang (2009) used the concept of degree-discount heuristics to optimize the high-degree heuristic. The greedy selection function considers the redundancy between likely influenced nodes and does not count those reached by the already selected seed nodes.

A large number of research has been developed in the area from those first proposals. Different kinds of heuristic and metaheuristic algorithms have been considered to solve the IM problem such as simulated annealing (Jiang et al., 2011; Li et al., 2017), GAs (Bucur & Iacca, 2016; Zhang, Du, & Feldman, 2017), and memetic algorithms (Gong, Song, Duan, Ma, & Shen, 2016), among many others. They have allowed us to find both run time and performance advantages over Kempe et al.'s greedy heuristics. Other authors have addressed the IM problem by first defining a strategy space for seed selection —mainly based on the combination of some social network metrics and not on a single one as greedy heuristics— and then optimizing it using different metaheuristics such as GAs (Lahiri & Cebrian, 2010), swarm intelligence (ŞİMŞEK & Resul, 2018), and ant colony optimization (Salavati & Abdollahpouri, 2019). The latter algorithms explored the underlying structures of social networks and then used diffusion methods to maximize the influence upon them.

All the later proposals deal with the original and classical problem formulation where a fixed number of initial seeds must be selected with this value s being provided as input to the IM problem. Recently, Bucur, Iacca, Marcelli, Squillero, and Tonda (2018) dealt with a multiobjective IM problem variant where the number of seeds is also optimized. The goal is to look for a trade-off between effort (the number of seed nodes required) and effect (the influence spread obtained upon the whole network) using an EMO algorithm called MicroGP. A similar idea is considered in Yang and Liu (2018) where another multiobjective variant called *influence maximization-cost minimization* is introduced and solved using a multiobjective discrete particle swarm algorithm. In order to speed up the algorithm computation, authors considered a surrogate function to measure the influence spread (they approximate it by using the expected number of activated nodes with 2-steps in the social network). Meanwhile, the cost objective represents the overall cost of influencing the s seeds initially selected and is computed according to their individual influence, a function of each node's degree.

Notice that the influence propagation models considered in almost all the later proposals, as well as in the most of those existing in the specialized literature, are the independent cascade and linear threshold models. Alternatively, marketing scholars used ABM to describe other kinds of adoption processes and characterize the individuals having the greatest effect on global adoption (Schlereth et al., 2013). Hence, they dealt with VM, a specific variant of the IM problem focused on the important role that hubs or influentials play for new product adoption processes in social networks of consumers. In those cases, the objective function does not account for the influence spread but for the estimated revenue raising resulting from the marketing campaign. In this way, one can consider factors such as the opportunity cost related to the timing of the product adoption or purchase. Among the different approaches to the VM problem studied by other authors, we take as a base the one presented in Stonedahl et al. (2010), where authors proposed prescriptive policies for seeding campaigns handling with an affordable knowledge level to marketers. This proposal also relies on defining a weighted combination of local and global social network analysis measures to create a ranking-based seed selection function. The VM campaigns are generated with GAs by automatically learning the combination weights. In our proposed algorithm, we will only consider the local network metrics to make the approach realistic for a marketing environment.

Up to our knowledge, there is not any previous multiobjective evolutionary research dealing with conflicting objectives when finding and targeting influentials in VM campaigns. Normally, there is a simplification of the inherent two-objective problem (maximizing revenue and minimizing costs) into a single objective (see Stonedahl et al., 2010). In our case, we address the targeting problem from a multiobjective perspective by using EMO algorithms to jointly optimize the two key campaign objectives. As described in the introduction, our approach considers a realistic marketing ABM which for instance, includes an awareness mechanism. Our complete proposal computes those key performance indicators and estimates the number of seeds achieving optimal trade-offs between them. But the framework also allows marketers to have, in a single run of the algorithm, a non-dominated set of influential seed choices (i.e., different marketing campaign designs) and adapts the VM campaigns to the considered social network topology.

3. The agent-based market simulation model

3.1. Consumers and products of the artificial market

Our model is an extension of the one proposed in Robles et al. (2016), inspired by the well-known Consumat model (Janssen & Jager, 2003). We use an ABM with a population of I consumer agents, connected by an artificial social network, and J competing market products. The products of the set J differ in their defining features (e.g., price, quality, or design). We model these characteristics by a single variable $d_j \in [0, 1]$, $\forall j \in J$. Consumers have a personal preference $\pi_i \in [0, 1]$, $\forall i \in I$ about their needs in the market.

The simulation has T simulation steps and follows a synchronous update (Chica & Rand, 2017). The model has a parameter $p_b \in [0, 1]$ which defines, for each time-step $t \in \{1, 2, ..., T\}$, the probability for each agent i to start its decision-making process to choose and buy one product from J. An array $X_i = \{x_i^1, x_i^2, ..., x_i^T\}$ with $x_i^t \in \{0, 1, ..., |J|\}$ stores the selected products by agent i and 0 when the agent does not buy at the specific time-step.

The model also incorporates a consumer awareness behavior (Robles et al., 2016) for agents to have a limited knowledge about the set of available products in the market, based on the idea of bounded human rationality (Chica et al., 2018). Thanks to this awareness mechanism, the model restricts the agents' purchase decisions to those products they are aware of, which varies overtime. A binary variable $\theta_{ij}^t \in \{0, 1\}$ models if agent i is aware of product j ($\theta_{ij}^t = 1$) or is not ($\theta_{ij}^t = 0$) at time step t. Therefore, at each step, an agent i can only choose a product from a reduced set $J_i^t = \{j \subset J : \theta_{ij}^t = 1\}$. Values for θ_{ij}^0 are randomly initialized at the beginning of the simulation.

An agent can also lose their products' awareness during the simulation by a deactivation process to mimic reality. This process is modeled by a deactivation variable $\delta \in [0, 1]$, having the same value for all the agents. At every step, agent i can lose the awareness of product j (restricted to $j \neq x_i^t$) with a probability δ. High values of δ mean agents easily forget their products' awareness. The deactivation process occurs in reality, it is a natural process in the consumer's journey and this is why it is included in the model. The word-of-mouth process caused by viral marketing campaigns (to be introduced in the next sub-section) counteracts this deactivation process.

3.2. Peer influence by the social network

Agents are connected through an artificial social network where each node is a consumer and edges represent the connections of the agents with their direct contacts. Interactions between consumers occur during all the simulation steps and facilitate the information diffusion process among the agents (i.e., word-of-mouth) (Rogers, 2010). In some cases, not only the contacts of the network are important but the characteristics of these contacts (e.g., trust relationships) (Wang, Da, Li, & Liu, 2018). Some authors even distinguish among direct and indirect trust to calculate the reputation of each node of the social network and have integrated these mechanisms to find the most influentials nodes in IM formulations (Zhang et al., 2019). Another clear example is the work of Chica et al. (2018) where authors studied significant differences when promoting trust in social dilemmas when having diverse social network topologies. In our case, we aim to consider an additional trust layer in future research studies. Here, the social network influences three main components of the model, described in the following sub-sections.

3.2.1. Biased product utility

Each agent i has, for each product j, a biased product utility at each time-step, noted as $u_{ij}^t \in [0, 1]$. This biased utility takes into account the difference between the personal preference of the agent π_i and the product reality d_j; but also incorporates the last product choices of the direct network contacts of i ($\gamma_{ij}^{t-1} \in [0, 1]$):

$$u_{ij}^t = (1 - \beta)(1 - |d_j - \pi_i|) + \beta\gamma_{ij}^{t-1}, \tag{1}$$

where $\beta \in [0, 1]$ is a social peer influence parameter (global for all the agents). High values of β mean consumers are highly influenced by their contacts while low values mean the personal preferences when adopting a product j are more important (e.g., a market with more innovative and unconventional consumers). γ_{ij}^{t-1} is the ratio of direct contacts of agent i who chose product j in the previous time-step $t - 1$.

3.2.2. Uncertainty about the decision

The contacts in the social network also influence the uncertainty of an agent when making the buying decision. An agent will be less convinced about making its decision just based on the utility value u_{ij}^t if (s)he has a high number of direct contacts who consumed different products in their last buying decision (i.e., at time step $t - 1$). Agents' variable $\phi_i^t \in [0, 1]$ defines this uncertainty, and it is calculated based on the direct contacts in the social network:

$$\phi_i^t = \beta(1 - \tilde{\gamma}_i^{t-1}), \tag{2}$$

where $\tilde{\gamma}_i^{t-1} \in [0, 1]$ is the fraction of contacts of agent i buying a different product than the one bought by the agent in the previous time step. Social parameter β also modulates here the importance of the social network to generate uncertainty in the agents. If the market has a high social characteristic (high β values), an agent will have a high uncertainty when making their decisions if their contacts bought a different product. This uncertainty mechanism, in addition to the thresholds of the heuristics (see Section 3.3), presents similar features than previous ABM simulation models with Bayesian theory when managing uncertainty (Pope & Gimblett, 2015).

3.2.3. Awareness diffusion

The model has an awareness diffusion probability $p_\theta \in [0, 1]$ that enables consumer agent i to talk about product j when (s)he is aware of it. To determine if a consumer talks about a product, a random number is drawn in [0,1]. If this number does not exceed the p_θ value, consumer agent i will spread its awareness about product j to its direct contacts and these contacts will also be aware of the product.

3.3. Heuristic rules for decision making

Agents can make decisions in different ways depending on their level of satisfaction and uncertainty. We consider four different heuristics, as done in Janssen and Jager (2003), to select a product from the set of products an agent i is aware of (i.e., sub-set $J_i^t \subseteq J$). Agents dynamically choose the heuristic by comparing their internal variables with respect to two global market thresholds. The first threshold is U and defines the minimum satisfaction level for the agents' utility. The second threshold is Φ and states agents' uncertainty tolerance level. These two thresholds are compared with the agent's values of their biased utility u_{ij}^t and uncertainty ϕ_i^t. The following four heuristics are chosen by each agent at each timestep t but when $t = 0$ (in that case there is not any previous step and the deliberation heuristic is the only one that applies):

- *Repetition*: Agent i repeats buying the same product with this heuristic because (s)he is satisfied and certain about its selection (i.e., $u_{ij}^t \geq U$ and $\phi_i^t \leq \Phi$). This is a repeat buying phenomenon related to the loyalty and satisfaction of a consumer to a brand (Ehrenberg, Uncles, & Goodhardt, 2004):

$$x_i^t = x_i^{t-1}. \tag{3}$$

- *Deliberation*: Agent i is not satisfied with its last purchase but is certain (i.e., $u_{ij}^t < U$ and $\phi_i^t \leq \Phi$). Therefore, this heuristic evaluates the utility u_{ij}^t for all the products the agent is aware of and generates a probabilistic decision using a logit function:

$$prob_i^t(j) = \frac{e^{u_{ij}^t}}{\sum_{k \in J_i^t} e^{u_{ik}^t}}, \forall j \in J_i^t. \tag{4}$$

- *Imitation*: Agent i is satisfied but uncertain about its choice (i.e., $u_{ij}^t \geq U$ and $\phi_i^t > \Phi$). A logit function is again used to calculate the probabilities for each product j the agent is aware of, taking into account the ratio of neighbors who bought it:

$$prob_i^t(j) = \frac{e^{2\gamma_{ij}^{t-1}}}{\sum_{k \in J_i^t} e^{2\gamma_{ik}^{t-1}}}, \forall j \in J_i^t. \tag{5}$$

- *Social comparison*: Agent i is neither satisfied nor certain about its choice (i.e., $u_{ij}^t < U$ and $\phi_i^t > \Phi$). A logit function similar to Equation 4 is used but by only evaluating the products bought by its direct contacts in their last buying decision ($J'^t_i \subset J_i^t$):

$$prob_i^t(j) = \frac{e^{u_{ij}^t}}{\sum_{k \in J'^t_i} e^{u_{ik}^t}}, \forall j \in J'^t_i. \tag{6}$$

4. Multiobjective viral marketing optimization problem

4.1. Decision variables

The goal of this problem is finding the optimal number of influentials for one of the brands of the market and their locations in the social network. We combine local social network metrics to evaluate all the consumers of the population and target those ones having the highest evaluation. We follow an adaptation of the approach defined in Stonedahl et al. (2010) to calculate the influential nodes in the social network (i.e., consumer agents). However, we do not take into account global social network metrics as it is unrealistic to have knowledge about the structure of the global network when launching the VM campaign. The three considered local measures are:

- Degree: Number of direct contacts of an agent i.
- 2-steps: Number of nodes reachable within two steps from agent i.
- Clustering coefficient (cc): It quantifies how much the neighbors of a node are interconnected among them (a local density measure) (Watts & Dodds, 2007). If a node is highly clustered (cc value close to 1) it will presumably generate a quicker adoption through a high diffusion speed in the social network. Values of cc close to 0 indicate a low clustered node in the social network and therefore lower diffusion speed.

We use a weighted combination of the latter three social network measures to obtain a value $q(i)$ for each agent i which indicates its influence in the network (higher values mean higher influence):

$$q(i) = \omega_d \frac{degree(i)}{|I| - 1} + \omega_{2s} \frac{\text{2-steps(i)}}{|I| - 1} + \omega_{cc}(1 - cc(i)), \tag{7}$$

where weights ω_d, ω_{2s}, and ω_{cc} are decision variables within [0,1]. Notice that the three measures are normalized in [0,1] to be properly combined to obtain $q(i)$.

We also incorporate the number of influentials to target as a fourth decision variable $s \in \{1, 2, ..., S_{\max}\}$, setting S_{max} as the maximum number influentials to target. Therefore, the set of decision variables is $\epsilon = \{\omega_d, \omega_{2s}, \omega_{cc}, s\}$. The process to identify the set S of influentials is to first evaluate all the agents by calculating their $q(i)$ values. Later, we sort them in descending order and pick the first s agents to be targeted in the campaign.

4.2. Objective functions

The ABM framework defined in Section 3 is used to simulate the market and evaluate the effectiveness of the VM campaigns after targeting s influentials with free samples of product $k \in J$ (i.e., the one promoted by the VM campaign). This effectiveness is measured by two objectives: (a) maximizing the number of sales of product k (i.e., the highest revenue), and (b) minimizing the costs of the campaign. Notice that influentials of set S cannot purchase products during the simulation as they obtain free product samples when they need to buy according to probability p_b. The cost of the campaign is associated to the cost of the free samples of the product provided to the selected influentials over time.

We use the net present value (NPV) to calculate both objectives. NPV is a measure of a company's revenue when consumers buy the product the moment in which the product has been commercialized rather than several months after product's commercialization. Mathematically, NPV is defined as $\sum_{t=1}^{\infty} |A^t| p \lambda^t$ where $|A^t|$ is the number of adopters or buyers of product k at time t, p is the profit for adoption of a product, and λ is the discount factor. This discount factor means that an earlier purchase is preferable because of the cumulative effect of several factors such as opportunity costs and the potential need of setting lower prices over time to be more competitive.

The first objective $f^0(\epsilon)$ is directly the NPV obtained accumulating the revenue of the purchases of product k at each time step t. The number of purchases obtained from the set of adopters $|A_k^t|$ is multiplied by discount factor λ related to the time of adoption as we set $p = 1$ for each sold product:

$$f^0(\epsilon) = NPV(A_k) = \sum_{t=1}^{T} |A_k^t| \lambda^t. \tag{8}$$

The second objective $f^1(\epsilon)$ is the cost of targeting the influentials of set S with free product k samples when they need to buy according to probability p_b. Note that this objective is clearly in conflict with $f^0(\epsilon)$ because increasing the number of influentials s consequently decreases the number of possible adopters of the product (A_k^t). We define $S_{x^t=k}$ as the set of influentials who are targeted at time step t. Campaign costs are computed by multiplying the number of provided samples by a manufacturing cost of $\frac{1}{10}$ and discount factor λ:

$$f^1(\epsilon) = Cost(S) = \frac{1}{10} \sum_{t=1}^{T} |S_{x^t=k}| \lambda^t. \tag{9}$$

5. Evolutionary multiobjective optimization methods for influentials selection

From the vast EMO family, we choose two of the best-known algorithms: NSGA-II (Deb et al., 2002) and MOEA/D (Zhang & Li, 2007). These two EMO algorithms follow two different approaches (i.e., Pareto-based selection and objective function decomposition). We also include a single-objective GA guided by a linear combination of the two objectives and a greedy approach as a baseline. We first describe the common design issues (Section 5.1) and later a brief description of the methods (Sections 5.2–5.4).

5.1. Common design of the methods

Each individual of the population has four genes which correspond to the set ϵ of four decision variables defined in Section 4.1. The first three genes are real-coded and represent the weights of the social network measures (ω_d, ω_{2s}, ω_{cc}). The fourth gene is the integer value s which states the number of influentials to target in the campaign. The three evolutionary algorithms follow a steady-state approach where the initial population is generated at random. The algorithms use a binary tournament selection method, a BLX-α crossover operator with a probability $p_c \in [0, 1]$, and a random mutation operator applied to each individual with a probability $p_m \in [0, 1]$.

5.2. NSGA-II Description

NSGA-II (Deb et al., 2002) is the most well-known EMO algorithm. In this algorithm, the offspring population Q is created, at each generation, from the parents population P (both composed of N chromosomes) by applying the considered selection mechanism and genetic operators. Afterwards, the two populations are joined to form a single intermediate population R with two times the population size ($2 \cdot N$). Then, the R population is partitioned into several fronts according to their dominance degrees by using a non-dominated sorting process.

Once the non-dominated sorting process is completed, the new population is generated from the configurations of the non-dominated fronts. This new population is first built with the best non-dominated front, continues with the solutions of the second front, the third, and so on until the N available positions are occupied. Every member of each dominance group is inserted into the new population until there is not any free place in the whole front. For the case of the first front that cannot be fully allocated in the new population, a diversity preservation mechanism is applied. A crowding distance calculation is made for every individual in that last considered front and the most diverse individual are those included in the new population.

5.3. MOEA/D Description

MOEA/D (Zhang & Li, 2007) is an EMO algorithm based on explicitly decomposing a multiobjective optimization problem into m scalar optimization sub-problems. The algorithm solves these sub-problems simultaneously by evolving a population of solutions. At each generation, the population is comprised by the best solution found so far for each sub-problem. The neighborhood relations among these sub-problems are defined based on the distances between their aggregation coefficient vectors. Each sub-problem is optimized in MOEA/D by only using information from its neighboring sub-problems.

There are several approaches for converting the problem of approximating a Pareto front into a number of scalar optimization problems. We use the Tchebycheff approach (g^{te}) for our experiments due to the good results in terms of feasibility and efficiency obtained in Zhang and Li (2007). Mathematically, let $\lambda = (\lambda_1, \ldots, \lambda_m)^W$ be a weight vector and m be the number of sub-problems, i.e., $\lambda_i >= 0$ for all $i = 1, \ldots, m$ and $\sum_{i=1}^{m} \lambda_i = 1$. The Tchebycheff approach considers a scalar optimization problem in the form:

$$\begin{aligned} \min g^{te}(x|\lambda, z^*) = \min\nolimits_{1 \leqslant i \leqslant m} \{\lambda_i |f_i(\epsilon) - z_i^*|\} \\ \text{subject to}\, \epsilon \in \Omega, \end{aligned} \tag{10}$$

where $z^* = (z_1^*, \ldots, z_m^*)^W$ is the reference point and Ω is the decision variable space of the optimization problem (in our case, $\epsilon = \{\omega_d, \omega_{2s}, \omega_{cc}, s\}$ where the first three decision variables are in [0,1] and the fourth one is in {1, 2, ..., $S_{\max}$}). This point is initialized as the lowest value of the objective function $f_i(\epsilon)$ found in the initial population. For each Pareto optimal point ϵ^* there is a weight vector λ such that ϵ^* is the optimal solution of Eq. (10) if it is a Pareto optimal solution of the objective function. Therefore, we can obtain different Pareto optimal solutions by altering the weight vector.

5.4. Weighted single-objective genetic algorithm

We have implemented a single-objective approach to solve the multiobjective VM problem by linearly combining the two objectives, $f^0(\epsilon)$ and $f^1(\epsilon)$, in a single-objective fitness value. We use a parameter $\rho \in [0, 1]$ to weight both problem objectives by $f(\epsilon) = \rho f^0(\epsilon) + (1 - \rho) f^1(\epsilon)$.

An aggregated Pareto front approximation for this GA is obtained by joining the non-dominated set of solutions found by the algorithm with 11 values for the linear combination parameter ρ (from 0 to 1 with a step of 0.1). These values are able to cover a diverse solutions range and allow us to obtain a Pareto set approximation to be compared with those obtained by NSGA-II and MOEA/D, allowing a fair comparison.

5.5. Greedy algorithm

We have also included a greedy algorithm for the multiobjective VM problem. Our greedy heuristic can be seen as another extension of the high-degree heuristic (see Section 2), based on using different local social network metrics, both in isolation and in combination. The following four combinations are defined for weights ω_d, ω_{2s}, and ω_{cc}:

- [1, 0, 0]: Influentials are only sorted and targeted by their degree measure values.
- [0, 1, 0]: Influentials are only sorted and targeted by their 2-steps measure values.
- [0, 0, 1]: Influentials are only sorted and targeted by their clustering coefficient values.
- [0.33, 0.33, 0.33]: Influentials are equally sorted and targeted by considering their degree, 2-step, and clustering coefficient.

The algorithm generates sets of s influentials (from 1 to S_{max}) for each of the latter combinations. After generating all the solutions, an aggregated Pareto front is created with the non-dominated solutions from those generated by the algorithm, for each set of weights and number of targeted influentials s.

6. Experiments and analysis of results

Section 6.1 reports the experimental setup and reviews the EMO indicators considered. Sections 6.2 and 6.3 analyze the results and marketing insights from the EMO solutions.

6.1. Experimental setup

All the EMO algorithms were run 30 times with independent random seeds. Their population size is of $N = 200$ chromosomes, evolving for 100 generations. Crossover probability p_c is set to 0.6 (using $\alpha = 0.5$ for the BLX operator) and mutation probability p_m is set to 0.2. We define a neighborhood of 40 chromosomes for MOEA/D. We have used a grid search algorithm to find the best values for the latter EMO parameters in a preliminary experimentation by also taking into account different population sizes. This grid search was not applied to the parameters of the agent-based market simulation model and viral marketing conditions (e.g., maximum number of influentials to target).

The maximum number of influentials to target ($S_{\max}$) is $0.4 \cdot |I|$, with $|I|$ being the number of agents in the population. As done in Stonedahl et al. (2010), the discount factor of the objective functions is 10% ($\lambda = 0.9$). The optimization algorithms employ the agent-based market simulation model to evaluate the revenue and costs of the target VM campaigns (we average the 30 Monte-Carlo runs of the ABM's output for each individual evaluation). We set the parameters of the ABM simulation model as follows, with the goal of having a balanced scenario with respect to social, personal preferences, and heuristics. There are $|J| = 10$ products available in the market (including the one promoted by the VM campaign) and the simulations runs for $T = 365$ time-steps. Social influence parameter β is 0.5, thresholds U and Φ are both set to 0.5. Awareness deactivation δ is 0.2 and agents' probabilities to buy a product and talk about them are $p_b = 0.8$ and $p_\theta = 0.5$, respectively. Finally, the total number of agents $|I|$ is equal to the number of nodes of each artificial social network used in the experimentation.

Table 1 shows the features of the five artificial social network topologies and the case study. The first artificial social network topology, EMAIL, is a real email network of the Rovira i Virgili University in Spain (Guimera et al., 2003). The rest of the artificial networks were generated using similar features than those of the EMAIL network (i.e., an average degree of 10 and a density of 0.01). For the scale-free network (SF) we used the Barábasi-Albert preferential attachment algorithm (Barabási & Albert, 1999) with $m = 2$ (number of new connections for each added node). For the random network (RAND), we connected each node at random to other nodes in the network by maintaining the average degree of 10 ($p = 0.01$). For the small world network (SW), we used 0.2 as the rewiring probability in the Watts-Strogatz generation algorithm (Watts & Strogatz, 1998). We use 4 edges per node with periodic boundary conditions for the regular lattice network (REG). Finally, the data-generated network of the case study (APP) presents a heavily bi-modal degree distribution (Chica & Rand, 2017). It means there is a numerous group of users with one or two connections and another numerous group with a high number of connections (i.e., around 100).

We consider two unary and two binary multiobjective performance indicators (Coello et al., 2007; Zitzler, Thiele, Laumanns, Fonseca, & da Fonseca, 2003) to evaluate the performance of the algorithms. The first unary indicator is the cardinality of each Pareto set approximation (i.e., the number of solutions composing it). The second unary indicator is the hyper-volume ratio (*HVR*) (Coello et al., 2007) which jointly measures the distribution and convergence of a Pareto front approximation. In order to be able to calculate the *HVR* indicator we consider a pseudo-optimal Pareto front which is an approximation to the true Pareto front obtained by merging all the Pareto front approximations generated for each problem instance by each algorithm in every independent run. Additionally, we use two binary indicators (Zitzler et al., 2003), the multiplicative I_ϵ indicator and the set coverage indicator (C). Both of them are quality assessment methods for multiobjective optimization that avoid particular difficulties of unary and classical performance indicators.

6.2. Performance analysis on artificial social networks

Table 2 shows the values for the *HVR* and the cardinality unary performance indicators of all the algorithms. Table 3 reports the values for the C indicator and Table 4 for I_ϵ indicator. In all the tables, best values are in bold and average and standard deviation values for the 30 runs of the algorithms are in brackets.

MOEA/D obtains the highest *HVR* values in three of the five social network topologies (see Table 2). On the contrary, NSGA-II shows the best results in the other two topologies. The single-objective GA shows poor results compared to the EMO algorithms. Additionally, the greedy algorithm is the worst algorithm in all the scenarios. Results show that differences among MOEA/D, NSGA-II, and the single-objective GA are clear depending on the features of the social network topologies. For instance, MOEA/D performs between a 1.5% and 2% better than the other two algorithms when considering network topologies with high diffusion speed (EMAIL and SF). Meanwhile, the single-objective GA obtains the worst values for those networks. In the RAND network, where information diffusion speed is slow, MOEA/D is still the best algorithm but with a low difference with respect to NSGA-II (around 0.03%) and to GA

Table 1
Main features of the used social network topologies.

	I	Edges	Avg. path length	Avg. degree	Clust. coeff.	Density
EMAIL	1133	5451	3.60	9.622	0.254	0.009
SF	1000	5007	3.04	10.014	0.056	0.01
RAND	1000	4980	3.27	9.852	0.008	0.01
SW	1000	5000	4.43	10	0.495	0.01
REG	1024	2048	16.02	4	0	0.04
APP	20000	481907	3.02	48.19	0.006	0.002

Table 2
Average and standard deviation values of *HVR* and cardinality (best results in bold).

	HVR				Cardinality			
	MOEA/D	NSGA-II	GA	Greedy	MOEA/D	NSGA-II	GA	Greedy
EMAIL	**0.96 (0.017)**	**0.947 (0.007)**	0.913 (0.01)	0.864 (0.008)	41.33 (3.32)	**65.47 (4.88)**	10 (0.83)	32 (3.1)
SF	**0.957 (0.017)**	0.938 (0.007)	0.928 (0.01)	0.862 (0.003)	35 (3.05)	**59.86 (6.21)**	9.03 (1.03)	32 (2.54)
RAND	**0.943 0.023**	0.94 (0.009)	0.937 0.01	0.857 (0.02)	41 (3.85)	**65.5 (6.25)**	9.56 (0.86)	29 (5.8)
SW	0.932 (0.028)	**0.934 (0.01)**	0.878 (0.01)	0.789 (0.023)	39 (3.76)	**63.6 (5.35)**	8.73 (1.01)	21 (2.7)
REG	0.865 (0.045)	**0.897 (0.022)**	0.687 (0.037)	0.653 (0.034)	28 (3.68)	**35 (3.74)**	6.1 (0.8)	15 (4.21)

Table 3
Average and standard deviation values of coverage *C* (best results in bold).

	C(MOEA/D,NSGA-II) / *C*(NSGA-II,MOEA/D)	*C*(MOEA/D,GA) / *C*(GA,MOEA/D)	*C*(MOEA/D,Greedy) / *C*(Greedy,MOEA/D)	*C*(NSGA-II,GA) / *C*(GA,NSGA-II)	*C*(NSGA-II,Greedy) / *C*(Greedy,NSGA-II)	*C*(GA,Greedy) / *C*(Greedy, GA)
EMAIL	**0.461 (0.019)** / 0.358 (0.0246)	**0.604 (0.0152)** / 0.068 (0.05)	**1.0 (0.0)** / 0.0 (0.0)	**0.526 (0.022)** / 0.082 (0.008)	**1.0 (0.0)** / 0.0 (0.0)	**1.0 (0.0)** / 0.0 (0.0)
SF	**0.4803 (0.0218)** / 0.333 (0.0212)	**0.439 (0.0252)** / 0.09 (0.075)	**1.0 (0.0)** / 0.0 (0.0)	**0.421 (0.0142)** / 0.145 (0.0076)	**1.0 (0.0)** / 0.0 (0.0)	**1.0 (0.0)** / 0.0 (0.0)
RAND	**0.4529 (0.026)** / 0.4072 (0.0277)	**0.448 (0.0121)** / 0.139 (0.065)	**1.0 (0.0)** / 0.0 (0.0)	**0.385 (0.0201)** / 0.14 (0.093)	**1.0 (0.0)** / 0.0 (0.0)	**1.0 (0.0)** / 0.0 (0.0)
SW	0.3261 (0.0238)/ **0.5537 (0.0263)** /	**0.3 (0.0151)** / 0.249 (0.0136)	**1.0 (0.0)** / 0.0 (0.0)	**0.414 (0.0116)** / 0.175 (0.059)	**1.0 (0.0)** / 0.0 (0.0)	**1.0 (0.0)** / 0.0 (0.0)
REG	0.2289 (0.0164) / **0.715 (0.0178)** /	**0.647 (0.0235)** / 0.113 (0.0132)	**1.0 (0.0)** / 0.0 (0.0)	**0.733 (0.0164)** / 0.045 (0.013)	**1.0 (0.0)** / 0.0 (0.0)	**1.0 (0.0)** / 0.0 (0.0)

Table 4
Average and standard deviation values of multiplicative I_ϵ (best results in bold).

	I_ϵ(MOEA/D,NSGA-II) / I_ϵ(NSGA-II,MOEA/D)	I_ϵ (MOEA/D,GA) / I_ϵ(GA,MOEA/D)	I_ϵ(MOEA/D,Greedy) / I_ϵ(Greedy,MOEA/D)	I_ϵ(NSGA-II,GA) / I_ϵ(GA,NSGA-II)	I_ϵ(NSGA-II,Greedy) / I_ϵ(Greedy,NSGA-II)	I_ϵ(GA,Greedy) / I_ϵ(Greedy, GA)
EMAIL	**1.023 (0.009)** / 1.04 (0.018)	**1.014 (0.007)** / 1.092 (0.023)	**0.928 (0.001)** / 1.22 (0.048)	**1.016 (0.009)** / 1.093 (0.023)	**0.953 (0.002)** / 1.21 (0.02)	**0.988 (0.002)** / 1.12 (0.037)
SF	**1.022 (0.006)** / 1.042 (0.017)	**1.019 (0.009)** / 1.086 (0.023)	**0.91 (0.001)** / 1.33 (0.028)	**1.025 (0.009)** / 1.09 (0.022)	**0.952 (0.025)** / 1.171 (0.035)	**0.973 (0.032)** / 1.22 (0.04)
RAND	**1.03 (0.018)** / 1.034 (0.018)	**1.024 (0.013)** / 1.108 (0.031)	**0.921 (0.004)** / 1.22 (0.03)	**1.023 (0.008)** / 1.103 (0.024)	**0.946 (0.002)** / 1.19 (0.018)	**0.983 (0.005)** / 1.12 (0.02)
SW	1.04 (0.019) / **1.031 (0.02)**	**1.028 (0.014)** / 1.094 (0.031)	**0.91 (0.001)** / 1.325 (0.032)	**1.022 (0.008)** / 1.09 (0.022)	**0.938 (0.003)** / 1.268 (0.003)	**0.964 (0.012)** / 1.25 (0.012)
REG	1.083 (0.026) / **1.03 (0.027)**	**1.023 (0.024)** / 1.205 (0.059)	**0.94 (0.002)** / 1.31 (0.03)	**1.01 (0.011)** / 1.205 (0.045)	**0.928 (0.001)** / 1.703 (0.002)	**0.993 (0.002)** / 1.19 (0.023)

(0.8%). Finally, NSGA-II obtains the best *HVR* values for SW and REG networks, the ones with the lowest diffusion speed.

NSGA-II is the EMO algorithm that returns the highest number of non-dominated solutions in the Pareto fronts (i.e., cardinality indicator). MOEA/D is the second one, and the greedy approach obtains a good number of non-dominated solutions but far from NSGA-II. The single-objective GA has the lowest cardinality value. *C* indicator values (shown in Table 3) draw the same conclusions than those observed when analyzing *HVR* values (Table 2). MOEA/D obtains the best *C* values in the EMAIL, SF, and RAND networks; but MOEA/D solutions are mainly covered by NSGA-II solutions for the SW and REG networks (low information diffusion speed). The single-objective GA shows low *C* values in comparison to pure EMO algorithms. Again, the greedy algorithm is the worst performing algorithm for all the tested social networks.

Table 4 shows the I_ϵ values. Again, the algorithms have the same behavior than with the previous indicators of Tables 2 and 3. MOEA/D Pareto front approximations need to be multiplied by lower values (with 1.03 being the highest one) to achieve the set of NSGA-II solutions for the first three scenarios (i.e., EMAIL, SF, and RAND). Meanwhile, NSGA-II needs a greater ϵ value to approximate the MOEA/D Pareto fronts. The opposite conclusion is derived for the case of SW and REG networks, where NSGA-II reports lower I_ϵ values. The greedy algorithm is again the worst performing algorithm.

We now focus our analysis on the attainment surfaces of the Pareto front approximations obtained by the algorithms. Figures 1, 2, and 3 show the aggregated Pareto front approximations for three of the five social network topologies (i.e., EMAIL, SF and REG). We do not show the remaining attainment surfaces because of the lack of space. Notice that previous indicators were calculated by using independent Pareto front approximations, obtained by the algorithms in the 30 runs. However, these attainment surfaces show the aggregated Pareto fronts of the different algorithms in the 30 runs for a better visualization.

MOEA/D provided the best solutions for the EMAIL network (Fig. 1) and SF network (Fig. 2). Only few NSGA-II solutions dominate the MOEA/D ones (e.g., those with revenue in [1,500] and

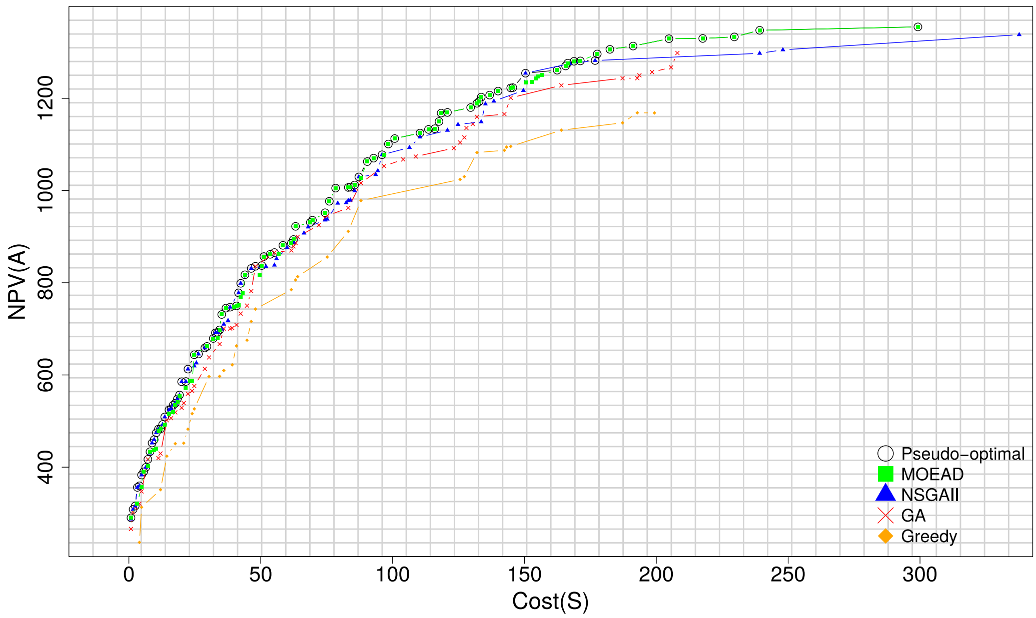

Fig. 1. Pareto front approximations obtained for the EMAIL network.

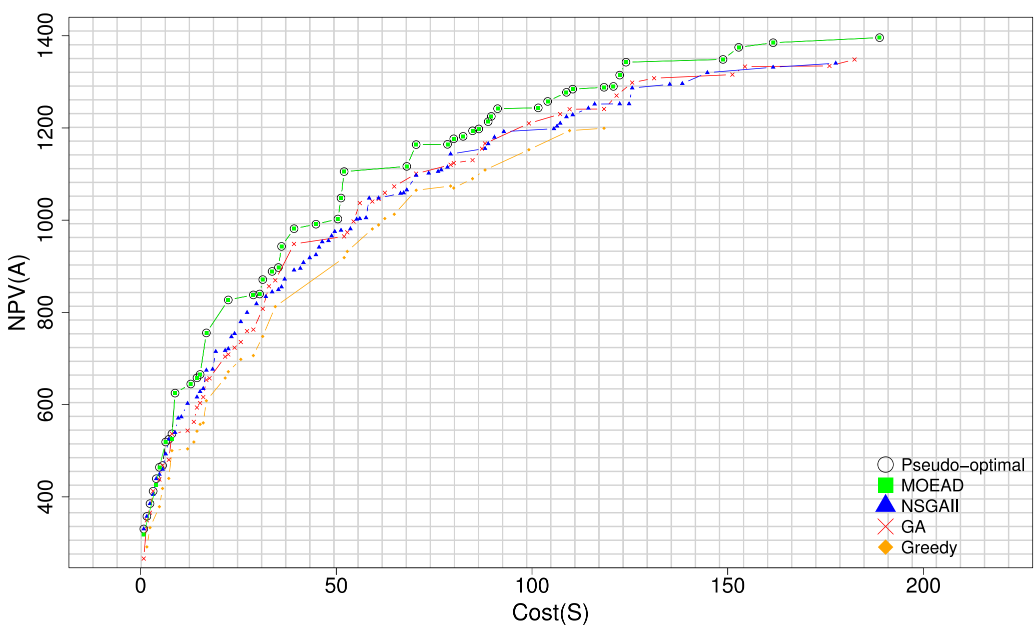

Fig. 2. Pareto front approximations obtained for the SF network.

costs in [1,20], located in the bottom left part of the Pareto front for the SF network). The performance of MOEA/D in these first two scenarios shows its potential when having network topologies with a power-law degree distribution. Finally, Fig. 3 shows the attainment surfaces for the REG network, the one with the lowest information diffusion speed. In this scenario, we see the opposite behavior: NSGA-II outperforms MOEA/D by generating a significantly higher number of pseudo-optimal Pareto front solutions. For all the cases, the performance of the single-objective and greedy algorithms is poor as the majority of their solutions are dominated by NSGA-II and MOEA/D.

6.3. Analysis of the returned marketing campaigns

We validate here the multiobjective VM proposal from a marketing perspective. Therefore, we analyze the quality of the marketing campaigns obtained by the three evolutionary algorithms. To ease the analysis we only consider the EMAIL and REG network scenarios as they represent the most different social network topologies.

Table 5 shows the composition of three different solutions selected from the Pareto front approximations for each scenario. Two of them are those with the best value for each individual objective (revenue raising and cost of the VM campaign). The third one presents a trade-off between these two objectives. The trade-off solution is chosen as follows. We compute 1000 random values for weight $\rho \in [0, 1]$ and the aggregated value of the weighted combination of the two objectives is taken for each solution (i.e., $f(\epsilon) = \rho NPV(A) + (1-\rho)Cost(S)$). The solution with the highest aggregated value $f(\epsilon)$ is then selected. Remind that the decision variables set ϵ is formed by three ω weights for the local network measures and the fourth variable is the number of influentials s.

Table 5
Decision variable values obtained by solutions of MOEA/D, NSGA-II and GA in EMAIL and REG social networks.

	MOEA/D				NSGA-II				GA				Greedy			
	ω_d	ω_{2s}	ω_{cc}	s	ω_d	ω_{2s}	ω_{cc}	s	ω_d	ω_{2s}	ω_{cc}	s	ω_d	ω_{2s}	ω_{cc}	s
EMAIL	Highest revenue solution ($f^0(\epsilon)$=1398.17, $f^1(\epsilon)$=191.25)				Highest revenue solution ($f^0(\epsilon)$=1350.2, $f^1(\epsilon)$=174.38)				Highest revenue solution ($f^0(\epsilon)$=1354.2, $f^1(\epsilon)$=182.25)				Highest revenue solution ($f^0(\epsilon)$=1180.0, $f^1(\epsilon)$=201.3)			
	0.58	0.9	0.23	170	0.48	0.98	0.29	155	0.34	0.78	0.22	162	1.0	0.0	0.0	198
	Best trade-off solution ($f^0(\epsilon)$=972.12, $f^1(\epsilon)$=76.3)				Best trade-off solution ($f^0(\epsilon)$=958.4, $f^1(\epsilon)$=79.88)				Best trade-off solution ($f^0(\epsilon)$=950.8, $f^1(\epsilon)$=84.6)				Best trade-off solution ($f^0(\epsilon)$=912.7, $f^1(\epsilon)$=92.2)			
	0.7	0.84	0.27	69	0.64	0.96	0.32	71	0.58	0.75	0.27	75	1.0	0.0	0.0	88
	Lowest cost solution ($f^0(\epsilon)$=204.92, $f^1(\epsilon)$=15.75)				Lowest cost solution ($f^0(\epsilon)$=238.8, $f^1(\epsilon)$=22.5)				Lowest cost solution ($f^0(\epsilon)$=70.03, $f^1(\epsilon)$=14.625)				Lowest cost solution ($f^0(\epsilon)$=52.12, $f^1(\epsilon)$=26.8)			
	0.91	0.87	0.4	14	0.87	0.85	0.49	20	0.86	0.72	0.37	13	1.0	0.0	0.0	24
REG	Highest revenue solution ($f^0(\epsilon)$=155.4, $f^1(\epsilon)$=398.25)				Highest revenue solution ($f^0(\epsilon)$=148.1, $f^1(\epsilon)$=282.38)				Highest revenue solution ($f^0(\epsilon)$=123.43, $f^1(\epsilon)$=148.5)				Highest revenue solution ($f^0(\epsilon)$=120.5, $f^1(\epsilon)$=152.3)			
	0.31	0.52	0.47	354	0.72	0.85	0.68	251	0.5	0.64	0.53	132	0.33	0.33	0.33	138
	Best trade-off solution ($f^0(\epsilon)$=112.4, $f^1(\epsilon)$=132.4)				Best trade-off solution ($f^0(\epsilon)$=122.38, $f^1(\epsilon)$=125.03)				Best trade-off solution ($f^0(\epsilon)$=108.6, $f^1(\epsilon)$=111.58)				Best trade-off solution ($f^0(\epsilon)$=89.31, $f^1(\epsilon)$=120.45)			
	0.34	0.56	0.32	117	0.47	0.63	0.0.25	111	0.74	0.58	0.49	99	0.33	0.33	0.33	114
	Lowest cost solution ($f^0(\epsilon)$=27.13, $f^1(\epsilon)$=1.13)				Lowest cost solution ($f^0(\epsilon)$=22.24, $f^1(\epsilon)$=1.13)				Lowest cost solution ($f^0(\epsilon)$=15.41, $f^1(\epsilon)$=1.13)				Lowest cost solution ($f^0(\epsilon)$=10.58, $f^1(\epsilon)$=1.13)			
	0.27	0.62	0.27	1	0.41	0.26	0.23	1	0.82	0.47	0.53	1	0.33	0.33	0.33	1

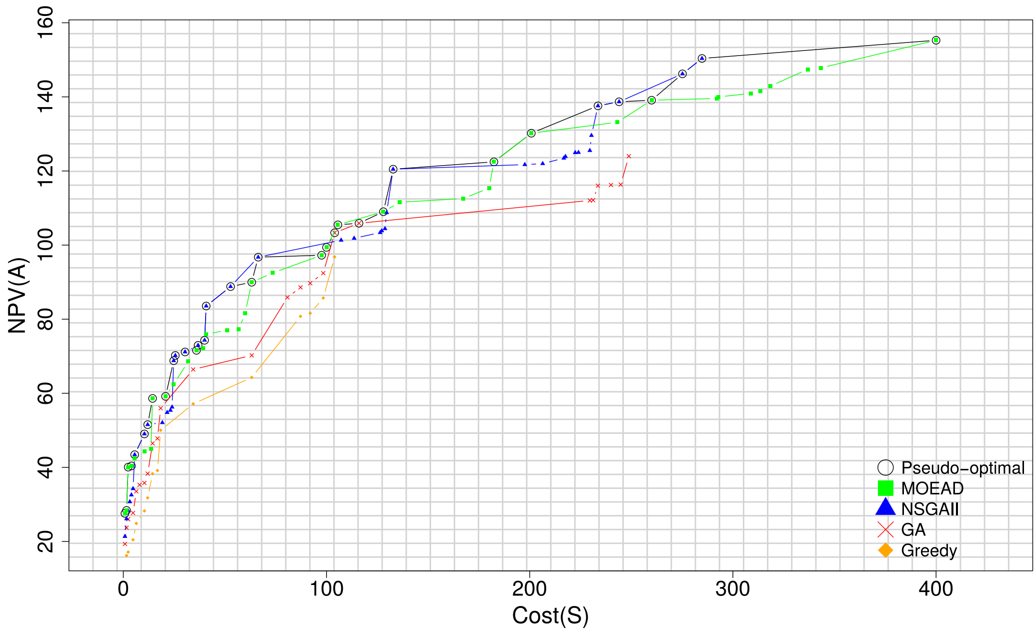

Fig. 3. Pareto front approximations obtained for the REG network.

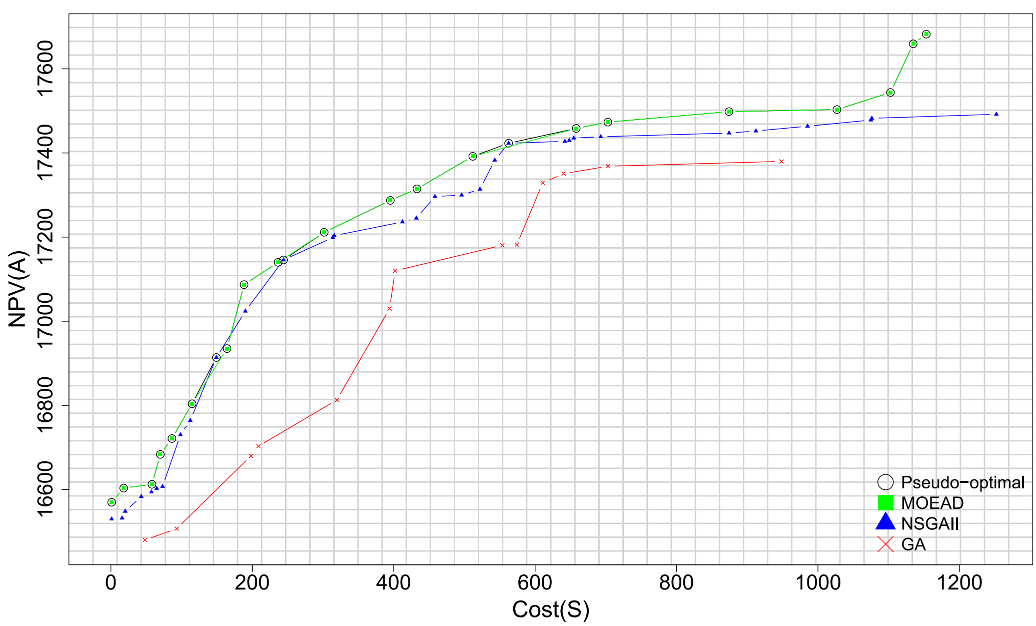

Fig. 4. Pareto front approximations obtained for the APP network.

We can find interesting differences in the solutions of Table 5. First, we can see that the algorithms return solutions with different ω weights for the local network measures. These weights are adapted to the characteristics of the social network topology to achieve a higher diffusion of the campaign. For example, in the EMAIL network, we see that the weights of measures related to the degree of the target nodes have a higher value to use the "hubs" of the network as influentials. As expected, solutions will have a higher or lower number of influentials depending on the revenue and cost of the campaign, but also depending on the network topology (e.g., a large number of influentials is required to get good revenue in a REG network with low diffusion speed).

We see two interesting differences when comparing the highest revenue and lowest cost solutions for the EMAIL network. The first is related to the number of influentials s: it is higher for the solutions with the highest revenue to enable a quick spread of the product. The second difference is that, for the highest revenue solution, the two-steps degree is weighted after the degree or the clustering coefficient. This is a direct consequence of the power law-distribution of the EMAIL network topology: by using a combination of a low number of targeted consumers with a high degree and a high 2-steps value, the VM campaign targets those consumers which are "hubs" in the social network. In contrast, the composition of the solutions for the REG network has less variability than those for the EMAIL network. Solutions for REG show similar ω_d, ω_{2s}, and ω_{cc} values, both for the highest revenue and lowest cost solutions. In this case, the main difference between these solutions resides in the number of targeted consumers s: the algorithm needs to select a high number of nodes s to obtain a high revenue.

Table 6
Average and standard deviation values of multiobjective performance indicators for the real case study.

		MOEA/D	NSGA-II	GA
HVR		**0.999 (0.0008)**	0.997 (0.0002)	0.94 (0.0031)
Cardinality		22.6 (3.51)	**34.2 (5.18)**	12.5 (1.2)
I_ϵ	MOEA/D		**1.002 (0.002)**	**1.005 (0.001)**
	NSGA-II	1.008 (0.001)		**1.003 (0.002)**
	GA	43.08 (0)	42.7 (0.02)	
C	MOEA/D		**0.742 (0.121)**	**1.0 (0)**
	NSGA-II	0 (0.001)		**1.0 (0.001)**
	GA	0 (0)	0.05 (0.001)	

6.4. Case study on a data-generated social network

We have extended the set of social networks by adding a real case study. This case study allows us to analyze if MOEA/D, NSGA-II and single-objective GA behave similarly than with the five artificial social networks. We use the same indicators to measure the algorithms' performance (Table 6) and the attainment surfaces with the set of non-dominated solutions (Fig. 4). The greedy algorithm is not included because of its poor performance in the previous experiments.

Results from Table 6 shows the superiority of MOEA/D over NSGA-II and GA. MOEA/D obtains the best values for all the multiobjective performance indicators except for cardinality, where NSGA-II is the best performing algorithm. For this real-world network, MOEA/D has higher I_ϵ and C values than NSGA-II and the single-objective GA (for instance, we see that MOEA/D covers around the 75% of the solutions obtained by NSGA-II). Fig. 4 shows the attainment surfaces. We observe MOEA/D has the highest number of solutions belonging to the pseudo-optimal Pareto front. Only three NSGA-II solutions belong to the pseudo-optimal Pareto front. The single-objective GA has a poor performance and is far from the performance obtained by MOEA/D and NSGA-II.

7. Main conclusions and limitations of the study

This study proposed a novel multiobjective optimization problem formulation to maximize the revenue while minimizing the costs of a VM campaign that targets influentials with free product samples. We applied two EMO algorithms, NSGA-II and MOEA/D, a single-objective GA and a greedy algorithm which evaluate the fitness of the VM campaigns using an artificial market framework based on ABM. The ABM simulation closely represents consumer agents' reality when choosing among existing products and spreading consumers' peer influence in the social network.

Our experimentation comprised social networks from high information diffusion speed (e.g., SF networks) to low information diffusion speed (e.g., RAND and REG networks) and applied statistical tests to confirm the results. The validation of our EMO approach was enhanced by a real case study with a social network of 20,000 nodes and a bi-modal degree distribution.

This study shows EMO algorithms can be applied to a VM problem to obtain sets of non-dominated solutions with different trade-offs between high revenue and low costs in a single run. Therefore, using an EMO algorithm is an appropriate way to help marketers by having a set of solutions that optimizes the selection of influentials. Another important highlights of our work are:

- Both EMO algorithms behave similarly, with MOEA/D showing the best performance. The single-objective GA had low performance indicators' values and returned sets with a low number of non-dominated solutions. The traditional greedy algorithm obtained the worst performance results.
- We found that the EMO algorithms showed different performance depending on the characteristics of the social network. The properties of the social networks affected the best performing EMO algorithm as well as the type of non-dominated solutions obtained by the algorithms to optimize the VM campaign. For instance, MOEA/D obtained more extended fronts and had the best performance for SF and EMAIL networks. In contrast, NSGA-II outperformed MOEA/D for social networks having slow diffusion speed such as REG.
- We found differences among the features of the Pareto front approximations. For instance, Pareto fronts obtained by NSGA-II had, on average, a higher number of solutions than those obtained by MOEA/D.

One of the limitations comes from the single type of seeding we apply here. There are other approaches such as sequential seeding (Jankowski et al., 2017) which can be applied and compared to. Also, the measures to rank the network nodes are basically local and others can be incorporated to enhance the actual decision variables. Future research can be focused on solving these limitations. As already mentioned, trust properties can also be incorporated into the social network to enrich social peer influence when defining the VM problem.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Credit authorship contribution statement

Juan Francisco Robles: Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Visualization. **Manuel Chica:** Writing - review & editing, Supervision, Project administration. **Oscar Cordon:** Writing - review & editing, Conceptualization, Methodology, Supervision, Project administration, Funding acquisition.

Acknowledgments

This work is supported by the Agencia Española de Investigación, Spanish Ministry of Science, Innovation and Universities, and ERDF under grant EXASOCO (PGC2018-101216-B-I00). M. Chica is supported by the Ramón y Cajal program (RYC-2016-19800).

References

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.
Bucur, D., & Iacca, G. (2016). Influence maximization in social networks with genetic algorithms. In *European conference on the applications of evolutionary computation* (pp. 379–392). Springer.
Bucur, D., Iacca, G., Marcelli, A., Squillero, G., & Tonda, G. (2018). Multi-objective evolutionary algorithms for influence maximization in social networks. In *European conference on the applications of evolutionary computation*. In *Vol. 10199 of Lecture notes in computer science* (pp. 221–233). Cham, pp. 221–233: Springer.
Chen, W., Wang, Y., & Yang, S. (2009). Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 199–208). ACM.
Chica, M., Chiong, R., Kirley, M., & Ishibuchi, H. (2018). A networked N-player trust game and its evolutionary dynamics. *IEEE Transactions on Evolutionary Computation, 22*(6), 866–878.
Chica, M., & Rand, W. (2017). Building agent-based decision support systems for word-of-mouth programs. a freemium application. *Journal of Marketing Research, 54*, 752–767. October

Coello, C. A., Lamont, G. B., & Van Veldhuizen, D. A. (2007). *Evolutionary algorithms for solving multi-objective problems* (2nd ed.). Springer.

ŞİMŞEK, A., & Resul, K. (2018). Using swarm intelligence algorithms to detect influential individuals for influence maximization in social networks. *Expert Systems with Applications, 114*, 224–236.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation, 6*(2), 182–197.

Domingos, P., & Richardson, M. (2001). Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, pp. 57–66.

Ehrenberg, A. S., Uncles, M. D., & Goodhardt, G. J. (2004). Understanding brand performance measures: using Dirichlet benchmarks. *Journal of Business Research, 57*(12), 1307–1325.

Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.

Goldenberg, J., Libai, B., & Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters, 12*(3), 211–223.

Gong, M., Song, C., Duan, C., Ma, L., & Shen, B. (2016). An efficient memetic algorithm for influence maximization in social networks. *IEEE Computational Intelligence Magazine, 11*(3), 22–33.

Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F., & Arenas, A. (2003). Self-similar community structure in a network of human interactions. *Physical Review E, 68*(6), 065103.

Haenlein, M., & Libai, B. (2017). Seeding, referral, and recommendation: Creating profitable word-of-mouth programs. *California Management Review, 59*(2), 68–91.

Jankowski, J., Bródka, P., Kazienko, P., Szymanski, B. K., Michalski, R., & Kajdanowicz, T. (2017). Balancing speed and coverage by sequential seeding in complex networks. *Scientific Reports, 7*(1), 891.

Janssen, M. A., & Jager, W. (2003). Simulating market dynamics: Interactions between consumer psychology and social networks. *Artificial Life, 9*(4), 343–356.

Jiang, Q., Song, G., Cong, G., Wang, Y., Si, W., & Xie, K. (2011). Simulated annealing based influence maximization in social networks. In *Proceedings of the twenty–fifth AAAI conference on artificial intelligence* (pp. 127–132). AAAI'11.

Kempe, D., Kleinberg, J., & Tardos, E. (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* (pp. 137–146). ACM.

Lahiri, M., & Cebrian, M. (2010). The genetic algorithm as a general diffusion model for social networks. In *Proceedings of the twenty-fourth AAAI conference on artificial intelligence.* (pp. 494–499).

Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web (TWEB), 1*(1), 5.

Li, D., Wang, C., Zhang, S., Zhou, G., Chu, D., & Wu, C. (2017). Positive influence maximization in signed social networks based on simulated annealing. *Neurocomputing, 260*, 69–78.

Newman, M., Barabási, A.-L., & Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton University Press.

Pope, A. J., & Gimblett, R. (2015). Linking Bayesian and agent-based models to simulate complex social-ecological systems in semi-arid regions. *Frontiers in Environmental Science, 3*, 55.

Robles, J. F., Chica, M., & Cordón, O. (2016). Incorporating awareness and genetic-based viral marketing strategies to a consumer behavior model. In *Proceedings of the 2016 IEEE congress on evolutionary computation (CEC)* (pp. 5178–5185). July

Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.

Salavati, C., & Abdollahpouri, A. (2019). Identifying influential nodes based on ant colony optimization to maximize profit in social networks. *Swarm and Evolutionary Computation, 51*, 100614.

Schlereth, C., Barrot, C., Skiera, B., & Takac, C. (2013). Optimal product-sampling strategies in social networks: How many and whom to target? *International Journal of Electronic Commerce, 18*(1), 45–72.

Stonedahl, F., Rand, W., & Wilensky, U. (2010). Evolving viral marketing strategies. In *Proceedings of the 12th annual conference on genetic and evolutionary computation. GECCO '10* (pp. 1195–1202). New York, NY, USA: ACM.

Wang, N., Da, J., Li, J., & Liu, Y. (2018). Influence maximization with trust relationship in social networks. In *2018 14th International conference on mobile ad-hoc and sensor networks (MSN)* (pp. 61–67).

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

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.

Yang, J., & Liu, J. (2018). Influence maximization-cost minimization in social networks based on a multiobjective discrete particle swarm optimization algorithm. *IEEE Access, 6*, 2320–2329.

Zhang, B., Zhang, L., Mu, C., Zhao, Q., Song, Q., & Hong, X. (2019). A most influential node group discovery method for influence maximization in social networks: A trust-based perspective. *Data & Knowledge Engineering, 121*, 71–87.

Zhang, K., Du, H., & Feldman, M. W. (2017). Maximizing influence in a social network: Improved results using a genetic algorithm. *Physica A: Statistical Mechanics and Its Applications, 478*, 20–30.

Zhang, Q., & Li, H. (2007). MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation, 11*(6), 712–731.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., & da Fonseca, V. G. (2003). Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation, 7*(2), 117–132.