Multi-objective Evolutionary Algorithms for Influence Maximization in Social Networks

Doina Bucur[1], Giovanni Iacca[2], Andrea Marcelli[3(✉)], Giovanni Squillero[3], and Alberto Tonda[4]

[1] University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
`d.bucur@utwente.nl`
[2] INCAS, Dr. Nassaulaan 9, 9401 HJ Assen, The Netherlands
`giovanni.iacca@gmail.com`
[3] DAUIN, Politecnico di Torino, Corso Duca Degli Abruzzi, 24, 10129 Torino, Italy
`{andrea.marcelli,giovanni.squillero}@polito.it`
[4] INRA, UMR 782 GMPA, Avenue Lucien Brétignières, 78850 Thiverval-Grignon, France
`alberto.tonda@grignon.inra.fr`

Abstract. As the pervasiveness of social networks increases, new NP-hard related problems become interesting for the optimization community. The objective of *influence maximization* is to contact the largest possible number of nodes in a network, starting from a small set of seed nodes, and assuming a model for information propagation. This problem is of utmost practical importance for applications ranging from social studies to marketing. The influence maximization problem is typically formulated assuming that the number of the seed nodes is a parameter. Differently, in this paper, we choose to formulate it in a multi-objective fashion, considering the minimization of the number of seed nodes among the goals, and we tackle it with an evolutionary approach. As a result, we are able to identify sets of seed nodes of different size that spread influence the best, providing factual data to trade-off costs with quality of the result. The methodology is tested on two real-world case studies, using two different influence propagation models, and compared against state-of-the-art heuristic algorithms. The results show that the proposed approach is almost always able to outperform the heuristics.

Keywords: Influence maximization · Social network · Multi-objective evolutionary algorithms

1 Introduction

Social networks (SNs) are graphs modeling a society, where edges represent the channels through which information, news, ideas, trends, or advertising flow dynamically in time. Nowadays SNs can effectively model the dynamic of an important part of people's digital lives: from interactions in established giants of

the field such as *Facebook* and *Google+*, or newcomers as *Twitter* and *Minds.com*; to contacts in specialized hubs dedicated to a specific target audience, such as *Academia.edu* and *ResearchGate*. As the number of users of such systems grows, so does the amount of behavioral data, and all classical network-related problems become computationally harder.

Interesting social phenomena can be studied by analyzing the underlying graph of the social network. A graph edge $a \rightarrow b$ could signify a likelihood that b will be exposed to a's opinions, or that a will support b in the election for the purpose of work-related promotions; given the initial set of network participants who broadcast information or cast a vote, the long-term opinions or the eventual outcome of the promotions may be easily foreseen. More in general, given an initial set of nodes in a SN, one can estimate the extent of their *influence* over the rest of the network. The dynamics by which the graph structure enables new information to spread varies with the nature of the SN: an edge $a \rightarrow b$ may model a fixed probability or a probability that varies according to other features of node b, such as its number of direct relationships with other nodes. Social sciences have studied a number of such probabilistic *propagation models* [13].

A common problem in SNs is the *influence maximization*: given the graph G, a discrete-time propagation model M, and a "budget" $k \geq 1$ of network nodes to be the initial "seeds", calculate the set of k nodes that will eventually produce the largest influence $\mathcal{I}$ upon the whole network. The problem was initially formulated in [19], and has been shown NP-hard for most propagation models [13].

Evolutionary algorithms (EAs) were found able to explore effectively the vast search space of all possible subsets of nodes [2], but the choice of the number of seeds was left to the user. In the present work, we instead propose to tackle the influence maximization problem using a Multi-Objective Evolutionary Algorithm (MOEA), trying to maximize the influence $\mathcal{I}$ and minimize the budget k concurrently. Including k as an explicit goal of the optimization provides users the necessary data to trade off between effort (the number of nodes that need to be influenced) and effect (the final influence over the whole network).

The proposed approach is then applied to two real-world case studies: the *ego-Facebook* network, which describes social circles from Facebook, and *ca-GrQc*, which covers scientific collaborations between authors. The networks are taken from the Stanford large network dataset [15], a popular source of benchmarks for influence maximization algorithms, and have been tested against state-of-the-art heuristics using two different influence propagation models.

The rest of the paper is organized as follows: Sect. 2 formalizes the problem, discusses existing heuristics and approximation algorithms for this problem, and briefly presents the MOEAs related work; Sect. 3 describes our MOEA method; and Sect. 4 presents the experimental results on two case studies. Finally, Sect. 5 concludes this work.

2   Background and Related Work

To introduce the scope of the current work, the basics of influence propagation and influence maximization in a SN are briefly recalled, followed by a short review on the state of the art for both heuristics and EAs applied to this problem, and a brief introduction to MOEAs.

2.1   Models for Influence Propagation and Problem Formulation

Concrete models for message forwarding in a SN are the basic building blocks to be able to evaluate the effectiveness of a set of seed nodes, i.e., how much global influence a set of nodes will have, indirectly, over the network via peer-to-peer message propagation.

Since the propagation of a message from a network user to another is a discrete event, the propagation models are also time-discrete. As the receptiveness of users to incoming messages from the network differs, we experiment with two previously studied models from the "Cascade" family of propagation models [13], which views influence as being transmitted through the network in a tree-like fashion, where the seed nodes are the roots.

The pseudo-code common to the two Cascade models: Independent Cascade (IC) and Weighted Cascade (WC) is given in Algorithm 1. IC was first studied in the marketing domain, modeling the effects that word-of-mouth communication has upon macro-level marketing [11]. Each newly "activated" node n will succeed in activating each inactive neighbor m with a fixed probability p, which is a global property of the system, equal for all edges $n \rightarrow m$ in G. WC, on the other hand, assigns non-uniform probabilities: an edge $n \rightarrow m$ has probability $p(n \rightarrow m) = \frac{1}{in\text{-}degree(m)}$ of activating m when n is active. In both models, when node n has more than one neighbor, activation is sequenced in an arbitrary order.

In the classical problem of influence maximization, the goal is to optimize the seed set S given a budget $k = |S|$ so that its eventual influence over the whole network is maximal. The influence of a seed set S is measured as the size of the set A of active nodes, $\mathbb{E}[|A|]$, obtained by the propagation model.

For both propagation models, the problem is NP-hard [13]. Furthermore, an approximation hardness result is known: approximating the optimal solution by a factor better than $1 - \frac{1}{e}$ (with e the base of the natural logarithm, which means roughly 63% approximation) is also NP-hard [13].

Further complication stems from the propagation models being stochastic: the evaluation of the expected influence of a seed set S in polynomial time will not be exact; the problem of computing this for any S over Independent Cascade was proven #P-complete in [22]. Instead, an approximate estimation can be obtained empirically, by simulating the propagation process a given number of times.

2.2   Prior Heuristics and EAs for Influence Maximization

We compare the proposed approach against two state-of-art heuristics: *High degree* (HIGHDEG), and *Single discount* (SDISC), briefly described in the

---

Algorithm 1. The **Cascade** family of propagation models. G is the network graph, S the set of "seed" nodes, and $p(n \rightarrow m)$ the probability that information will reach across a graph edge $n \rightarrow m$.

1: **procedure** CASCADE(G, S, p)
2:     $A \leftarrow S$ ▷ A: the set of active nodes after the propagation ended
3:     $B \leftarrow S$ ▷ B: the set of nodes activated in the last time slot
4:     **while** B not empty **do**
5:         $C \leftarrow \emptyset$
6:         **for** each $n \in B$ **do**
7:             **for** each direct neighbor m of n, where $m \notin A$, **do**
8:                 with probability $p(n \rightarrow m)$, add m to C
9:             **end for**
10:        **end for**
11:        $B \leftarrow C$
12:        $A \leftarrow A \cup B$
13:    **end while**
14:    **return** the size of A
15: **end procedure**

---

following. In general, the heuristics in literature fall into two categories: (a) heuristics which provably obey the approximation guarantee but are too time-intensive, and (b) heuristics of much better time complexity, but with either no approximation guarantees or much weaker ones. From the latter category, the *High-degree* (or degree centrality) greedy heuristic simply adds nodes n to A in order of decreasing out-degree [13]. Chen et al. [5] refine greedy "degree-discount" heuristics based on *High-degree*, using the idea that if a node n is already active and also there exists an edge $m \rightarrow n$, then, when considering whether to add node m to A, this edge should not be counted towards the out-degree of m. This heuristic (known as *Single discount*) is applicable to all cascade models.

Many other heuristics are known. Among those without optimality guarantees, Jiang et al. [12] test Simulated Annealing under Independent Cascade, and find that it has a complexity advantage over greedy heuristics, and can also find narrowly better solutions. In [2], a first attempt to tackle the influence maximization problem by a classic (single-objective) Genetic Algorithm is made, with promising results obtained in comparison with existing heuristics.

Various studies [5,12,13] evaluated the previously described heuristics comparatively on a small number of large SNs, and generally find empirically that the two degree-based, inexpensive greedy heuristics *High-degree* and *Single discount* may not reach the approximation guarantee of 63% known to be possible, but in some cases they are only a few percentage points away from the target.

2.3   Multi-objective Evolutionary Algorithms

In many optimization problems, the quality of a solution is defined by its performance in relation to several, conflicting objectives. Such conflicting goals cannot be sensibly reduced to a single value using a weighted sum or another aggregate

function, but rather they must be considered independently from each other. Multi-Objective Evolutionary Algorithms (MOEAs) are a natural flavour of Evolutionary Algorithms specifically designed for tackling this kind of problems. As a result, differently from single-objective optimization where the output of an evolutionary algorithm is a unique optimal solution, the output of a MOEA is a *Pareto front*, that is, a set of *non-dominated* solutions, to choose from. In other words, a Pareto front contains a set of optimal trade-off solutions that are better than any other solution in the multi-dimensional objective space at hand.

MOEAs have been used with great success in a number of real-world applications, as surveyed for instance in [4,6,8]. In the SNs domain, so far MOEAs have been used mostly for community detection [16,18,23] and network clustering [14]. On the other hand, to the best of out knowledge no prior work exists on the application of MOEAs on the influence maximization problem, which is the main contribution of the present work.

3   Proposed Approach

In order to tackle influence maximization as a multi-objective problem, we propose to use a MOEA to optimize the conflicting goals of maximizing how much the network has been influenced by the seed nodes (i.e., the effect of the influence campaign), and of minimizing the number of such seed nodes (i.e., the cost of the campaign). As a result, the different solutions on the Pareto front provide users the necessary data to trade off between cost and effect. In what follows we present briefly the details of the MOEA used in our experimentation.

3.1   MOEA Engine

μGP (also called *MicroGP*) is a generic evolutionary algorithm, able to manage individuals encoded as multigraphs. Its original application was the creation of complex assembly-language programs for testing different microprocessors [20]. Afterward, it has been used on a wider range of problems, such as the creation of test programs for pre- and post-silicon validation; the design of Bayesian networks [21]; the analysis of the impact of network topology on routing protocols [3]; automatic software testing in mobile phones [10]; real-value parameter optimization; and even creation of *corewar* warriors [7]. μGP is freely available on SourceForge[1].

While handling complex and structured genomes is not needed for this case study, μGP is also able to self-adapt the activation probability of genetic operators, manage a genome of variable length, and perform multi-objective fitness evaluation, with a crowding distance assessment on the Pareto front similar to the state-of-the-art algorithm NSGA-II [9].

---

[1] http://ugp3.sourceforge.net/.

3.2 Representation and Fitness of a Candidate Solution

For the problem at hand, a candidate solution is a set of nodes, of variable size, which is a subset of the set of nodes in the original network. Individuals are thus unordered sequences of integer node identifiers, representing the seeds of influence in the network. A visual representation of the problem encoding is reported in Fig. 1.
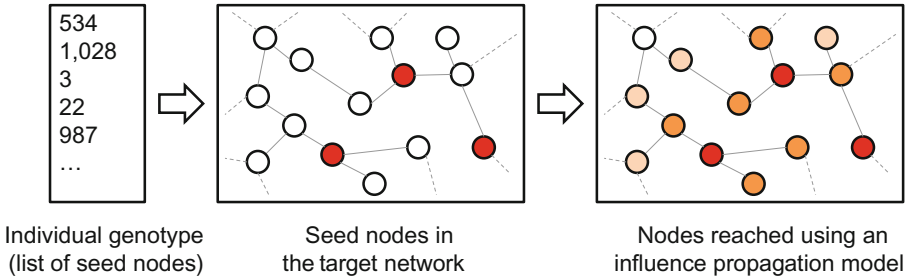

Fig. 1. Schema of the proposed encoding. Seed nodes are internally represented as a list of integers. The fitness value is the average number of other nodes that are reached, following a given model of probabilistic influence propagation. In the example, the only node not influenced by the set of seeds is the white one, in the top left corner of the rightmost frame.

The fitness value of a candidate solution is a probabilistic metric of the number of nodes that are likely to be reached, starting from a given set of seeds of influence — according to a given model of influence propagation (Independent Cascade or Weighted Cascade, described in Sect. 2). Given the stochastic nature of both propagation models, the fitness estimation is empirical, and itself a stochastic process: repeated simulations of the network propagation model yield an extent to which the network is reached, and the final fitness value is the average among these fitness samples.

4 Experimental Evaluation

In the experimental evaluation of the algorithm, μGP uses a $(\mu + \lambda)$ MOEA, with both μ (population size) = 100 and λ (number of operators applied at each generation) = 100, and the following genetic operators:

– Recombination operators: standard one- and two-point crossover.
– Three mutation operators, that add, remove, or change one node from a candidate seed set.

The activation probabilities of the operators are self-adapted during the evolutionary run, resorting to an improved dynamic multi-armed bandit selection [1].

All nodes in a target network can be added to a solution, and μGP is used with two different configurations, generating candidate seed sets of sizes between 1 and `max_nodes`. As the algorithm used to estimate influence propagation (either IC or WC) is stochastic, in order to get a reliable average each individual evaluation is repeated 100 times. The resulting fitness value is the average of the 100 repetitions.

The proposed approach is tested on two real-world case studies, taken from the Stanford large network dataset [15]. The first network, labeled **ego-Facebook**, describes social circles from Facebook; the second network, labeled **ca-GrQc**, covers scientific collaborations between authors from papers submitted to General Relativity and Quantum Cosmology category, from the e-print arXiv[2]. Their notable features are reported in Table 1.

Table 1. Main features of the case studies considered for the experimental evaluation of the proposed MOEA approach.

Name	ego-Facebook	ca-GrQc
Nodes	4,039	5,242
Edges	88,234	14,496
Type of graph	undirected	undirected
Nodes in largest WCC	4,039	4,158
Nodes in largest SCC	4,039	4,158
Average clustering coefficient	0.6055	0.5296
Diameter (longest shortest path)	8	17

4.1   ego-Facebook

The MOEA algorithm is compared with the *High degree* (HIGHDEG) and *Single discount* (SDISC) heuristics on the **ego-Facebook** case study, over both the IC propagation model, using two different probabilities of influence propagation, $p = 0.01$ or $p = 0.05$, and the WC model. The evolutionary algorithm's termination condition is set to 10,000 generations, and for each influence propagation model considered (IC with $p = 0.01$, IC with $p = 0.05$, WC), two different experiments are carried out, one with `max_nodes = 200` and one with `max_nodes = 400`.

ego-Facebook/IC. Results of the evaluation for the considered case study, using the IC model of influence propagation are reported in Fig. 2 (for $p = 0.01$) and Fig. 3 (for $p = 0.05$). In the figures, the y axis shows k, the number of seed nodes in the solution, while the x axis shows the average number of nodes reached by the seeds. The final Pareto front found by the MOEA (in red, with

---

[2] https://arxiv.org/.

early candidate evaluations in blue) outperforms the heuristics when the number of nodes in the candidate solutions exceeds 25; the MOEA is able to find solutions that are more effective, fitness-wise, than the ones created by the heuristics for the same number of nodes in the seed set. For solutions featuring a number of nodes close to the upper bound (200 or 400, depending on the experiment), it is however noticeable how the evolutionary algorithm has issues populating the Pareto front. Remarkably, some of the candidate solutions explored are still better than those provided by the heuristics (see for example Fig. 2), even if they are Pareto-dominated by others. This might be due to our choice of maintaining a relatively small population ($\mu = 100$) and give the algorithm more generations to evolve.

Intuitively, when $p = 0.05$, the number of influenced nodes is higher: however, the difference in performance between the heuristics and the MOEA also grows, as even some of the initial randomly generated individuals for high k behave better than the solutions found by HIGHDEG and SDISC, see Fig. 3. It is noticeable that the MOEA has again issues populating the higher part of the Pareto front, especially when the maximum allowed number of seed nodes is 400.
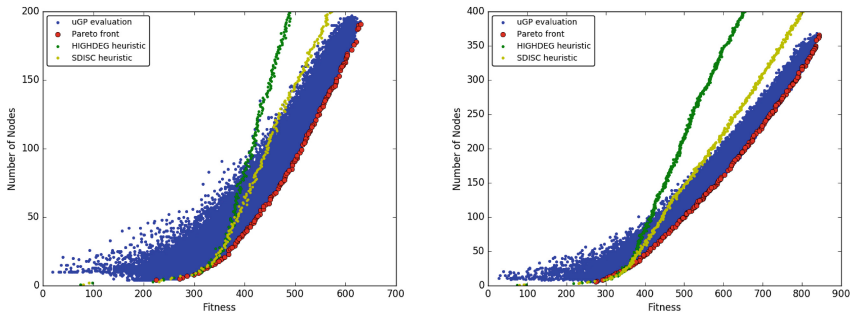

Fig. 2. Results obtained by the MOEA approach and the High Degree (HIGHDEG) and Single Discount (SDISC) heuristics on the ego-Facebook case study, using the Independent Cascade model for influence propagation, with $p = 0.01$ and an upper bound of 200 nodes (**left**) and 400 nodes (**right**) per candidate solution. (Color figure online)

ego-Facebook/WC. Results of the evaluation for the considered case study, using the WC model of influence propagation, are reported in Fig. 4 (the WC model does not use probability values p). The MOEA's results are very similar to the IC case, and while the evolutionary algorithm outperforms the heuristics on most values of k, it clearly has issues finding good solutions with a high number of nodes, especially when compared to the results of SDISC, which in this case study behaves particularly effectively.
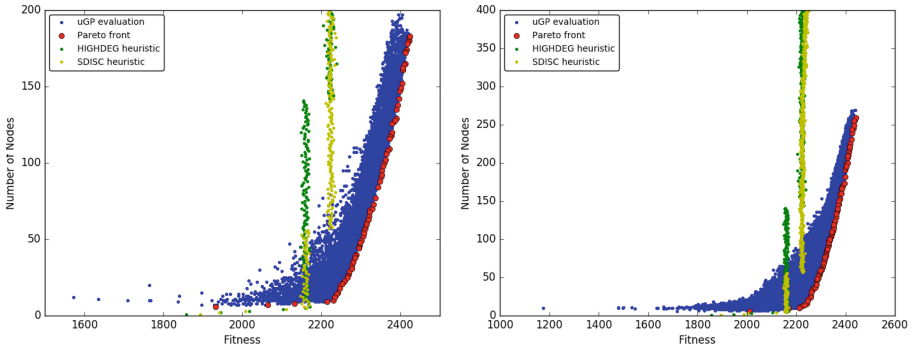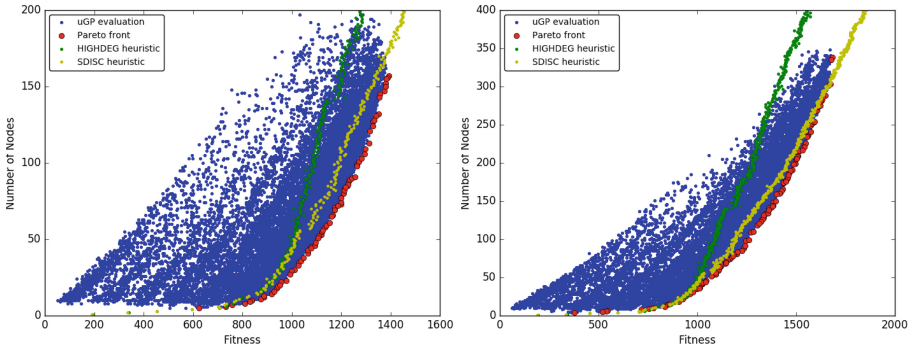
Fig. 3. Results obtained by the MOEA approach and the High Degree (HIGHDEG) and Single Discount (SDISC) heuristics on the ego-Facebook case study, using the Independent Cascade model for influence propagation, with $p = 0.05$ and an upper bound of 200 nodes (**left**) and 400 nodes (**right**) per candidate solution. (Color figure online)

Fig. 4. Results obtained by the MOEA approach and the High Degree (HIGHDEG) and Single Discount (SDISC) heuristics on the ego-Facebook case study, using the Weighted Cascade model for influence propagation, and an upper bound of 200 nodes (**left**) and 400 nodes (**right**) per candidate solution.

4.2   ca-GrQc

As **ca-GrQc** is a simpler case study, featuring less arcs than **ego-Facebook**, we choose to evolve solutions with upper bound of 400 candidate nodes, only (`max_nodes = 400`), and set a termination condition after 1,000 generations.

ca-GrQc/IC. Figure 5 shows results for the IC influence propagation model, using $p = 0.01$ and $p = 0.05$. Final outcomes for the two probability values are similar, even though the differences are more visible for $p = 0.05$; probably the lower value of p makes it harder for influence to spread on such a sparse network,
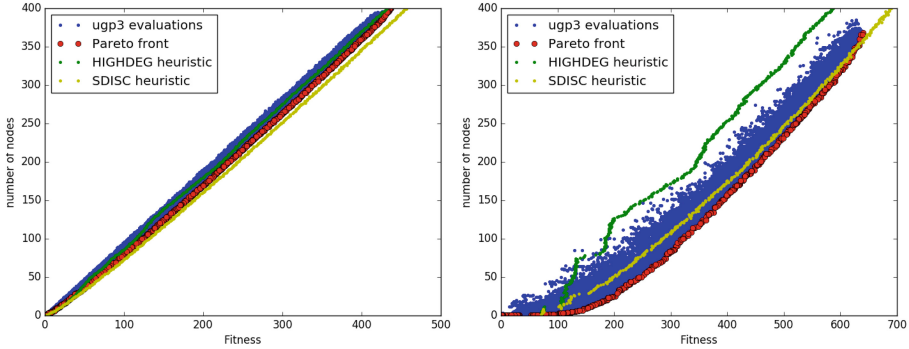
Fig. 5. Results obtained by the MOEA approach and the High Degree (HIGHDEG) and Single Discount (SDISC) heuristics on the ca-GrQc case study, using the Independent Cascade model for influence propagation, with an upper bound of 400 nodes per candidate solution, $p = 0.01$ (**left**) and $p = 0.05$ (**right**).

and as a result the fitness landscape is harder to climb, both for the heuristics and the evolutionary algorithm.

The HIGHDEG heuristic clearly lags behind both SDISC and the MOEA, being outperformed even by individuals produced in the first generations. The MOEA is able to best SDISC for the majority of values of k, with issues similar to the ego-Facebook case study: the multi-objective algorithm is unable to properly populate the highest part of the front ($k > 360$), and for small values of k, SDISC's solutions are better.
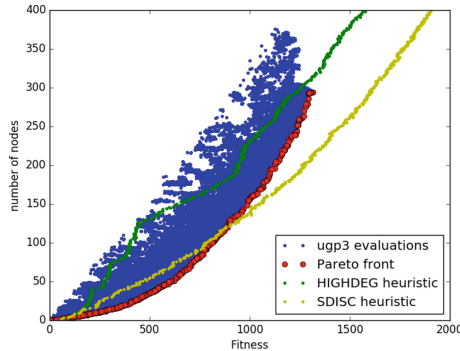

Fig. 6. Results obtained by the MOEA approach and the High Degree (HIGHDEG) and Single Discount (SDISC) heuristics on the ca-GrQc case study, using the Weighted Cascade model for influence propagation, with an upper bound of 400 nodes per candidate solution.

ca-GrQc/WC. Figure 6 shows results for the ca-GrQc case study using the WC influence model. This time, problems usually faced by the MOEA become more evident: the higher part of the Pareto front lags behind solutions found by the heuristics, already for $k > 150$; and while the MOEA clearly outperforms HIGHDEG, it cannot match the quality of seed sets found by SDISC. The general impression from this case study is that the termination condition came too early, and given more evaluations, the MOEA might have been able to populate other parts of the Pareto front: this is however counter-intuitive, as the same termination condition provided good results for the IC influence propagation model. Further analyses are required, in order to properly assess the source of these issues.

5   Conclusions

In this paper, we proposed a novel multi-objective evolutionary approach to influence maximization in social networks. A MOEA is tasked with finding the set of k seed nodes that, given a model of influence propagation, maximize the nodes reached in a target network. As minimizing the value of k is also given as an optimization objective, the MOEA is able to find a Pareto front of compromises between number of seed nodes in the set and global influence in the graph. The presented methodology is then tested on two real-world case studies, using two different influence propagation models. The MOEA is proven able to reliably outperform two state-of-the-art heuristics for intermediate values of k, facing mixed success for extremely high and low values of k.

While the results obtained are still preliminary, they nevertheless show promising outcomes. In particular, MOEAs prove to be able to overcome established heuristics for influence maximization for two real-world case studies. On the other hand, the heuristics are in general more efficient time-wise (although depending on the dataset and on the size of the seed nodes, as was shown in [2]), and still get a better performance on some corner cases. Further experimental evaluations on a wide range of social networks with different features are necessary, in order to assess the effective potential of the proposed approach; furthermore, the influence of the population size on the MOEA's performance is going to be studied. Future works will also focus on hybrid techniques, developing memetic algorithms [17] for influence maximization, which may be able to extract —and combine— the best qualities of EAs and heuristics.

References

1. Belluz, J., Gaudesi, M., Squillero, G., Tonda, A.: Operator selection using improved dynamic multi-armed bandit. In: Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation, pp. 1311–1317. ACM (2015)

2. Bucur, D., Iacca, G.: Influence maximization in social networks with genetic algorithms. In: Squillero, G., Burelli, P. (eds.) EvoApplications 2016. LNCS, vol. 9597, pp. 379–392. Springer, Heidelberg (2016). doi:10.1007/978-3-319-31204-0_25

3. Bucur, D., Iacca, G., Squillero, G., Tonda, A.: The impact of topology on energy consumption for collection tree protocols: an experimental assessment through evolutionary computation. Appl. Soft Comput. **16**, 210–222 (2014)

4. Bucur, D., Iacca, G., Squillero, G., Tonda, A.: The tradeoffs between data delivery ratio and energy costs in wireless sensor networks: a multi-objective evolutionary framework for protocol analysis. In: Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation, pp. 1071–1078. ACM (2014)

5. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2009, NY, USA, pp. 199–208. ACM, New York (2009)

6. Coello, C.C., Lamont, G.B., van Veldhuizen, D.A.: Evolutionary Algorithms for Solving Multi-objective Problems. Springer, Heidelberg (2002)

7. Corno, F., Sanchez, E., Squillero, G.: Evolving assembly programs: how games help microprocessor validation. IEEE Trans. Evol. Computat. **9**(6), 695–706. http://dx.doi.org/10.1109/TEVC.2005.856207

8. Deb, K.: Multi-Objective Optimization Using Evolutionary Algorithms, vol. 16. Wiley, Hoboken (2001)

9. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)

10. Gandini, S., Ruzzarin, W., Sanchez, E., Squillero, G., Tonda, A.: A framework for automated detection of power-related software errors in industrial verification processes. J. Electron. Test. **26**(6), 689–697 (2010)

11. Goldenberg, J., Libai, B., Muller, E.: Talk of the network: a complex systems look at the underlying process of word-of-mouth. Mark. Lett. **12**(3), 211–223 (2001)

12. Jiang, Q., Song, G., Cong, G., Wang, Y., Si, W., Xie, K.: Simulated annealing based influence maximization in social networks. In: Burgard, W., Roth, D. (eds.) AAAI. AAAI Press (2011)

13. Kempe, D., Kleinberg, J.: Éva Tardos: maximizing the spread of influence through a social network. Theor. Comput. **11**(4), 105–147 (2015)

14. Kim, K., McKay, R.B., Moon, B.R.: Multiobjective evolutionary algorithms for dynamic social network clustering. In: Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation, GECCO 2010, NY, USA, pp. 1179–1186. ACM, New York (2010)

15. Leskovec, J., Krevl, A.: SNAP Datasets: Stanford Large Network Dataset Collection (2016). http://snap.stanford.edu/data

16. Liu, C., Liu, J., Jiang, Z.: A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks. IEEE Trans. Cybern. **44**(12), 2274–2287 (2014)

17. Neri, F., Cotta, C., Moscato, P.: Handbook of Memetic Algorithms. Springer, Heidelberg (2012)

18. Pizzuti, C.: A multiobjective genetic algorithm to find communities in complex networks. IEEE Trans. Evol. Comput. **16**(3), 418–430 (2012)
19. Richardson, M., Agrawal, R., Domingos, P.: Trust management for the semantic web. In: Fensel, D., Sycara, K., Mylopoulos, J. (eds.) ISWC 2003. LNCS, vol. 2870, pp. 351–368. Springer, Heidelberg (2003). doi:10.1007/978-3-540-39718-2_23
20. Squillero, G.: MicroGP - an evolutionary assembly program generator. Genet. Program. Evol. Mach. **6**(3), 247–263 (2005)
21. Tonda, A.P., Lutton, E., Reuillon, R., Squillero, G., Wuillemin, P.-H.: Bayesian network structure learning from limited datasets through graph evolution. In: Moraglio, A., Silva, S., Krawiec, K., Machado, P., Cotta, C. (eds.) EuroGP 2012. LNCS, vol. 7244, pp. 254–265. Springer, Heidelberg (2012). doi:10.1007/978-3-642-29139-5_22
22. Wang, C., Chen, W., Wang, Y.: Scalable influence maximization for independent cascade model in large-scale social networks. Data Mining Knowl. Discov. **25**(3), 545–576 (2012)
23. Zeng, Y., Liu, J.: Community detection from signed social networks using a multiobjective evolutionary algorithm. In: Handa, H., Ishibuchi, H., Ong, Y.-S., Tan, K.C. (eds.) Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems, Volume 1. PALO, vol. 1, pp. 259–270. Springer, Heidelberg (2015). doi:10.1007/978-3-319-13359-1_21