Improving Multi-Objective Evolutionary Influence Maximization in Social Networks

Doina Bucur Giovanni Iacca Andrea Marcelli Giovanni Squillero Alberto Tonda

Influence Maximization

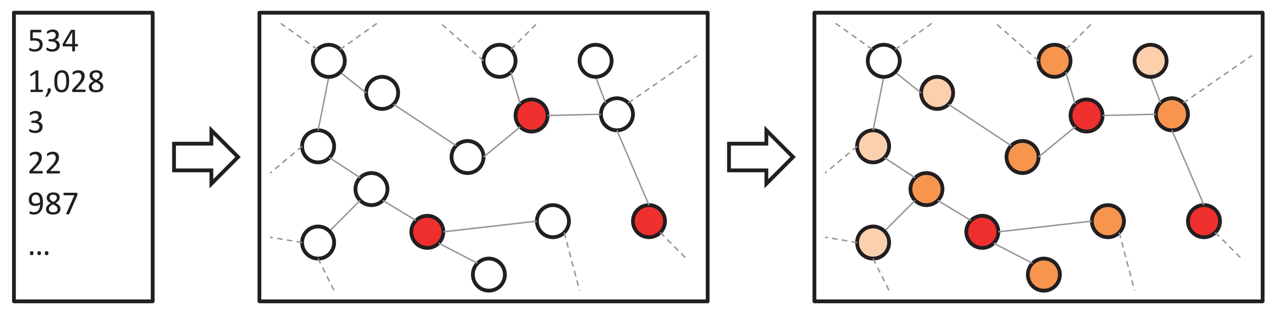
In the context of social networks, maximizing the influence means contacting the largest possible number of nodes starting from a set of *seed* nodes, while assuming a model for influence propagation.

The problem consists in finding the optimal set of seed nodes to maximize the final influence.

The MOEA approach

In [1] a multi-objective EA (MOEA) is used to maximize the influence while concurrently minimizing the size of the seed set, providing the necessary trade-off between *budget* and *effect*.

budget: the number of seed nodes
effect: the final influence over the whole network



Individual genotype (list of seed nodes) | Seed nodes in the target network | Nodes reached using an influence propagation model

[1] Bucur, Doina, et al. "Multi-objective Evolutionary Algorithms for Influence Maximization in Social Networks." European Conference on the Applications of Evolutionary Computation. Springer, Cham, 2017.

Propagation model

In the simplest SN model, a directed edge *a* -> *b* denotes that *b* is exposed to *a* and may be influenced by it. The rule used for determining whether the information is actually transmitted is called the propagation model.

The most popular belongs to the "Cascade" family which views influence as being transmitted through the network in a tree-like fashion, where the seed nodes are the roots. We use the *Independent Cascade* model.

Existing solution for influence maximization

Several heuristics allow to find good solutions to the influence maximization problem. The literature includes *High degree* (HIGHDEG), *Single discount* (SDISC), DISTANCE, *Generalized degree discount* (GDD), and *Cost-Effective Lazy Forward selection* (CELF).

The MOEA approach generates a Pareto front of candidate solutions, and shows [1] to outperform both HIGHDEG and SDISC. The main drawback is the computational time required to reach satisfying solutions: millions of individual evaluations are necessary, each one consisting of multiple runs of an influence spread model.

The proposed improvements considerably speed up the process, but also deliver higher-quality results

We show that the evolutionary algorithm is able to improve upon the solutions found by the heuristics, while also converging faster than an evolution algorithm started from scratch

Our contribution

We introduce an improvement over a previously proposed MOEA approach for influence maximization in social networks [1].

We show that seeding the initial population with the results of computationally cheap heuristics, the number of individual evaluations required to reach satisfying solutions drops dramatically.

Experimental Evaluation

	ego-Facebook	ca-GrQc	soc-ePinions1
Nodes	4,039	5,242	78,879
Edges	88,234	14,496	509,837
Type of graph	undirected	undirected	directed
Nodes in largest WCC	4,039	4,158	75,877
Nodes in largest SCC	4,039	4,158	32,223
Average clustering coefficient	0.6055	0.5296	0.1378
Diameter	8	17	14

Implemented in Python using the MOEA-NSGA-II algorithm from the *inspyred* library.

Individuals

A candiate solution is a subset of the set of nodes in the original network. Individuals are unordered sequences of unique integer node identifiers, representing the seeds of influence in the network.

Genetic operators

We use three problem-specific mutations (add, remove, replace of a node in the set), and one crossover operator. The operators are always applied with uniform probability, while the parent individuals are selected through a tournament selection of size 2.

Fitness

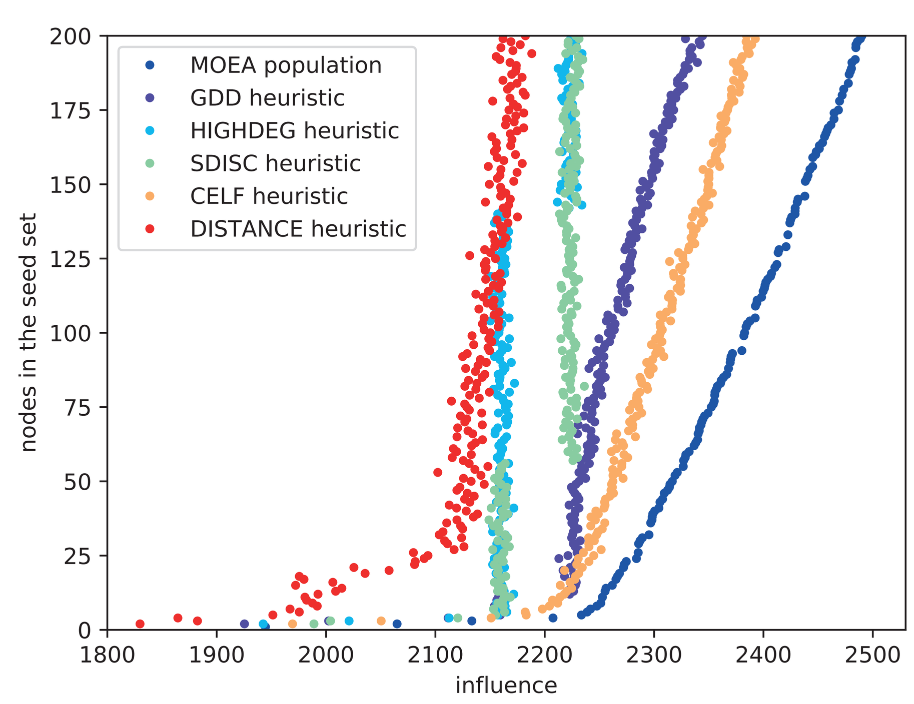
The fitness of a candidate solution is a probabilistic metric of the number of nodes that are likely to be reached, starting from a given set of seeds of influence - according to the IC model of influence of propagation.

MOEA configuration

$\mu = 2000$, $\lambda = 2000$, and tournament selection of size $\tau = 2$. Influence propagation model IC model with probability $p = 0.05$ and stop condition of 500 generations.

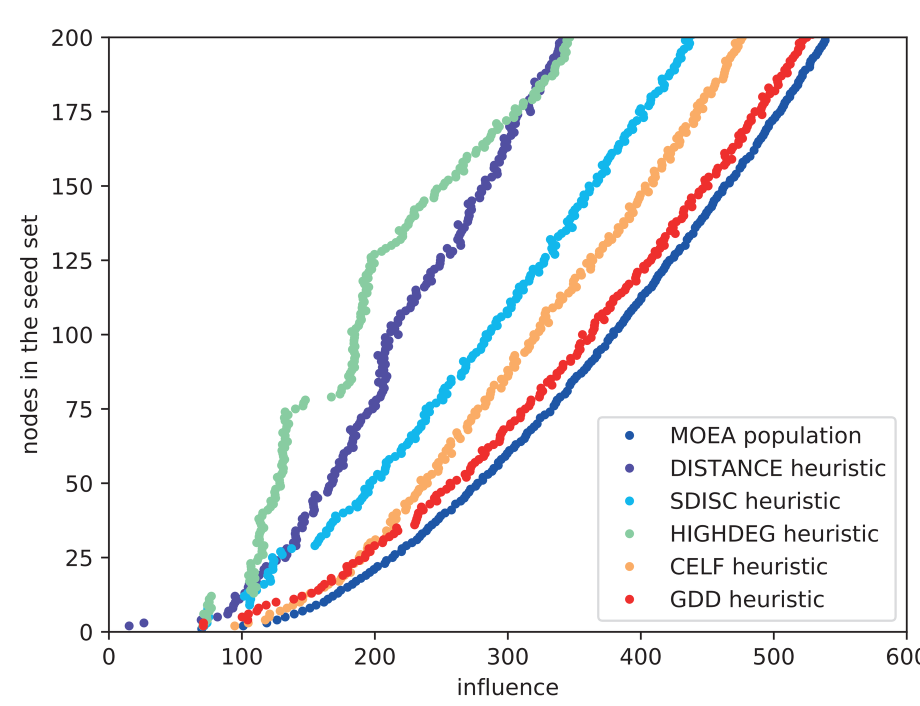
ego-Facebook

The MOEA is seeded with the results of the GDD heuristic, and the evolution is able to outperform the effective CELF heuristic too.



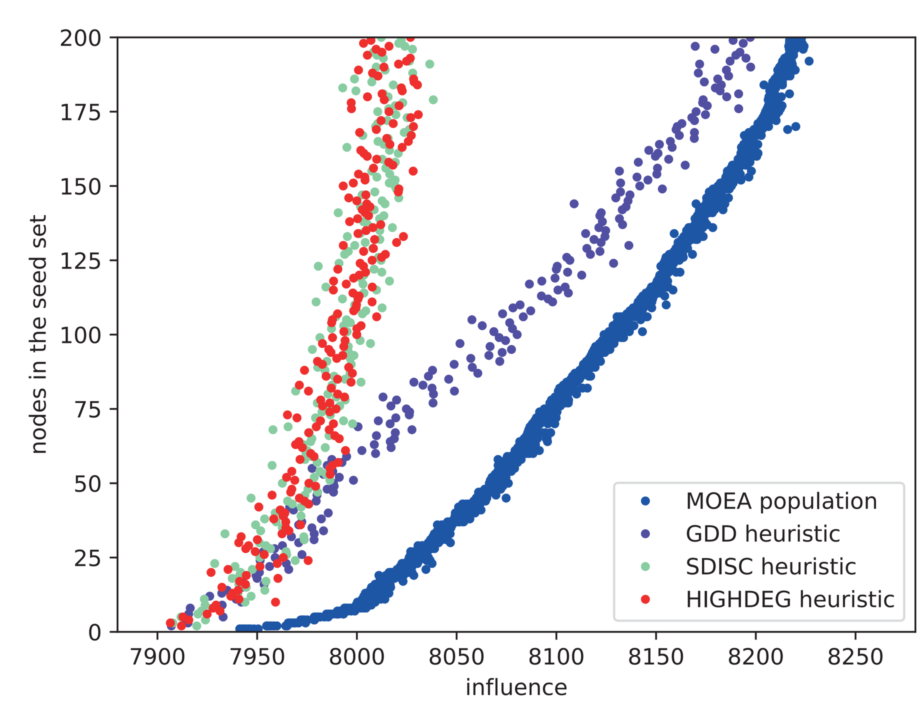
ca-GrQc

The MOEA is seeded with the results of the most performing GDD heuristic. Finally the evolution is able to find the best results.



soc-ePinions1

The MOEA is seeded with the results of the GDD heuristic. DISTANCE and CELF could not run on this network due to their excessive time complexity.



UNIVERSITY OF TWENTE.
RWTH AACHEN UNIVERSITY
POLITECNICO DI TORINO 1859 1906
INRA SCIENCE & IMPACT