Software Impacts 9 (2021) 100107

Contents lists available at ScienceDirect

Software Impacts

journal homepage: www.journals.elsevier.com/software-impacts

Original software publication

An evolutionary framework for maximizing influence propagation in social networks

Giovanni Iacca [a,*], Kateryna Konotopska [a], Doina Bucur [b], Alberto Tonda [c]

[a] Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 9, 38123 Povo, Trento, Italy
[b] Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Zilverling, Hallenweg 19, 7522NH Enschede, The Netherlands
[c] INRA, UMR 782 GMPA, Avenue Lucien Brétignières, 78850 Thiverval-Grignon, France

ARTICLE INFO

Keywords:
Social network
Influence maximization
Evolutionary algorithm

ABSTRACT

Social networks are one the main sources of information transmission nowadays. However, not all nodes in social networks are equal: in fact, some nodes are more influential than others, i.e., their information tends to spread more. Finding the most influential nodes in a network – the so-called Influence Maximization problem – is an NP-hard problem with great social and economical implications. Here, we introduce a framework based on Evolutionary Algorithms that includes various graph-aware techniques (spread approximations, domain-specific operators, and node filtering) that facilitate the optimization process. The framework can be applied straightforwardly to various social network datasets, e.g., those in the SNAP repository.

Code metadata

Current code version	v1.0
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2021-60
Permanent link to reproducible capsule	https://codeocean.com/capsule/6362854/tree/v1
Legal Code License	Apache-2.0 License
Code versioning system used	git
Software code languages, tools, and services used	python, bash
Compilation requirements, operating environments & dependencies	python 3.6.8, inspyred 1.0.1, networkx 2.3, numpy 1.16.3, cython 0.29.13, node2vec 0.3.1, gensim 3.8.1, pandas 0.24.2
If available, link to developer documentation/manual	https://github.com/tsume82/Influence-Maximization
Support email for questions	giovanni.iacca@unitn.it

Software metadata

Current software version	v1.0
Permanent link to executables of this version	https://github.com/tsume82/Influence-Maximization
Permanent link to reproducible capsule	https://codeocean.com/capsule/6362854/tree/v1
Legal Software License	Apache-2.0 License
Computing platforms/Operating Systems	Linux, OS X, Microsoft Windows, Unix-like
Installation requirements & dependencies	python 3.6.8, inspyred 1.0.1, networkx 2.3, numpy 1.16.3, cython 0.29.13, node2vec 0.3.1, gensim 3.8.1, pandas
If available, link to user manual–if formally published include a reference to the publication in the reference list	https://github.com/tsume82/Influence-Maximization
Support email for questions	giovanni.iacca@unitn.it

Scope of the software

Social networks are used nowadays for various activities such as reading news, searching for jobs, watching movies, or shopping. One of the most important problems in social networks is the so-called

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding author.
E-mail address: giovanni.iacca@unitn.it (G. Iacca).

https://doi.org/10.1016/j.simpa.2021.100107
Received 1 June 2021; Received in revised form 5 July 2021; Accepted 11 July 2021

Influence Maximization (IM) problem, where one wants to find the most influential nodes in the network. This problem is typically addressed by modeling probabilistically how influence propagates over the network [1], starting from a group of *seed nodes*. Finding the optimal (i.e., most influential) set of seed nodes is an NP-hard [1] problem, for which various algorithms such as heuristics with provable guarantees [2–4] or metaheuristics [5–10] have been proposed in the recent literature. Both classes of methods provide good results, although their runtime tends to increase with the network size. On the other hand, IM is a crucial problem for a number of applications such as political campaigns and marketing, therefore its implications are relevant.

Here, we present an open-source `python` framework, based on Evolutionary Algorithms (EAs). In the EA, a candidate solution is modeled as a fixed-size sequence of seed node identifiers, with smart initializers, mutation and crossover operators generating candidate solutions to form a new population. The fitness of a candidate solution can be estimated by Monte-Carlo simulations of probabilistic spread models, but the framework also includes advanced graph-aware techniques (introduced in [11]) such as approximations of the spread models, domain-specific initialization and mutations of candidate solutions, and node filtering.

These techniques allow the EA to solve the IM problem in an efficient way. As we have shown in [11], the proposed framework is indeed able to find good sets of seed nodes in a comparably lower time than existing techniques, which makes it especially appealing in scenarios characterized by large network dimensionalities. As such, the proposed framework can be of interest in a number of research projects and practical applications.

The framework is publicly available on a GitHub repository[1] and contains the following main elements:

- the source code of the EA, which is based on the library `inspyred`[2];
- modules related to the graph-aware features introduced in [11], including an implementation of the UCB1 multi-armed bandit algorithm [12] used for dynamic mutation selection;
- scripts to prepare the experimental setup and run the execution pipeline (the latter loads the network data, runs the EA with the appropriate hyperparameters and enabled elements, collects outputs and logs);
- a `cython` wrapper of the influence propagation evaluation part of the pipeline, and the related script to compile it;
- an example of a `json` configuration file (containing the hyperparameters of the EA) that can be used to run the full pipeline;
- implementations of various heuristics from literature (e.g., high-degree [1] and CELF [2]), to provide baseline results for testing the graph-aware EA;
- the raw data of the numerical results presented in [11], obtained on various network data from the SNAP repository [13] (which can be downloaded automatically using a script included in the GitHub repository).

A more detailed description of the above elements can be found on the `README` file available on our GitHub repository.

Impact overview

The EA algorithm provided with the framework achieved state-of-the-art results (both in terms of improved influence and reduced computing time) on various network data from the SNAP repository, compared to some of the existing heuristics. Furthermore, the framework has been used already to address a number of research questions concerning the influence maximization problem, most notably the effect of the various graph-aware elements, discussed in [11], on the optimization performance. There is no similar open-source software or toolbox that we are aware of.

Nevertheless, the framework can be easily extended to pursue new research questions. For instance, one possibility could be to include other objectives into the problem formulation (e.g., minimizing the time to propagation). Other options include (but are not limited to): analyzing dynamic social networks; associating features to the nodes in the network, then finding influential sets of seed nodes that also maximize the feature-wise diversity of the influenced nodes; designing and testing novel algorithmic elements to be included in the EA. In all these cases, the EA included in the framework, as well as at least some of its graph-aware mechanisms, can be reused.

Currently the framework is being used in several student projects and research activities aimed at extending the results obtained in [11]. So far, the framework has been already used in a number of scholarly publications. In particular, the original version of the framework, which did not include any graph-aware mechanism, was used firstly in [14], with a single-objective formulation of the IM problem, and later in [15–17], with a multi-objective formulation aimed at maximizing influence while minimizing the number of seed nodes. Eventually, the framework has been consolidated as presented here, and extended with the graph-aware mechanisms introduced in [11].

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.

References

[1] David Kempe, Jon Kleinberg, Éva Tardos, Maximizing the spread of influence through a social network, in: International Conference on Knowledge Discovery and Data Mining, KDD, ACM, New York, NY, USA, 2003, pp. 137–146.

[2] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Christos Faloutsos, Jeanne VanBriesen, Natalie Glance, Cost-effective outbreak detection in networks, in: International Conference on Knowledge Discovery and Data Mining, KDD, ACM, New York, NY, USA, 2007, pp. 420–429.

[3] Youze Tang, Yanchen Shi, Xiaokui Xiao, Influence maximization in near-linear time: A martingale approach, in: International Conference on Management of Data, SIGMOD, ACM, New York, NY, USA, 2015, pp. 1539–1554.

[4] H.T. Nguyen, M.T. Thai, T.N. Dinh, A billion-scale approximation algorithm for maximizing benefit in viral marketing, IEEE/ACM Trans. Netw. 25 (4) (2017) 2419–2429.

[5] Michal Weskida, Radosław Michalski, Finding influentials in social networks using evolutionary algorithm, J. Comput. Sci. 31 (2018) 77–85.

[6] Abel García-Nájera, Saúl Zapotecas-Martínez, Roberto Bernal-Jaquez, Selection schemes analysis in genetic algorithms for the maximum influence problem, in: Mexican International Conference on Artificial Intelligence, MICAI, Springer, Cham, 2020, pp. 211–222.

[7] Rodrigo Rodrigues, Arthur Silva, Viní cius Vieira, Carolina Xavier, Optimization of the choice of individuals to be immunized through the genetic algorithm in the SIR model, in: International Conference on Computational Science and Its Applications, ICCSA, Springer, Cham, 2018, pp. 62–75.

[8] Arthur Rodrigues da Silva, Rodrigo Ferreira Rodrigues, Vinícius da Fonseca Vieira, Carolina Ribeiro Xavier, Influence maximization in network by genetic algorithm on linear threshold model, in: International Conference on Computational Science and Its Applications, ICCSA, Springer, Cham, 2018, pp. 96–109.

[9] Krzysztof Michalak, Informed mutation operator using machine learning for optimization in epidemics prevention, in: Genetic and Evolutionary Computation Conference, GECCO, ACM, New York, NY, USA, 2018, pp. 1294–1301.

[10] Jian-bin Guo, Fu-zan Chen, Min-qiang Li, A multi-objective optimization approach for influence maximization in social networks, in: International Conference on Industrial Engineering and Engineering Management, IEEM, Springer, Singapore, 2019, pp. 706–715.

[11] Kateryna Konotopska, Giovanni Iacca, Graph-aware evolutionary algorithms for influence maximization, 2021, arXiv preprint arXiv:2104.14909.

[12] Peter Auer, Nicolo Cesa-Bianchi, Paul Fischer, Finite-time analysis of the multiarmed bandit problem, Mach. Learn. 47 (2-3) (2002) 235–256.

[13] Jure Leskovec, Andrej Krevl, SNAP Datasets: Stanford large network dataset collection, 2014, http://snap.stanford.edu/data.

[1] https://github.com/tsume82/Influence-Maximization.

[2] https://github.com/aarongarrett/inspyred.

[14] Doina Bucur, Giovanni Iacca, Influence maximization in social networks with genetic algorithms, in: European Conference on the Applications of Evolutionary Computation, in: LNCS, vol. 9597, Springer, Berlin, Heidelberg, 2016, pp. 379–392.

[15] Doina Bucur, Giovanni Iacca, Andrea Marcelli, Giovanni Squillero, Alberto Tonda, Multi-objective evolutionary algorithms for influence maximization in social networks, in: European Conference on the Applications of Evolutionary Computation, in: LNCS, vol. 10199, Springer, Cham, 2017, pp. 221–233.

[16] Doina Bucur, Giovanni Iacca, Andrea Marcelli, Giovanni Squillero, Alberto Tonda, Evaluating surrogate models for multi-objective influence maximization in social networks, in: Genetic and Evolutionary Computation Conference Companion, GECCO, ACM, New York, NY, USA, 2018, pp. 1258–1265.

[17] Doina Bucur, Giovanni Iacca, Andrea Marcelli, Giovanni Squillero, Alberto Tonda, Improving multi-objective evolutionary influence maximization in social networks, in: International Conference on the Applications of Evolutionary Computation, Springer, Cham, 2018, pp. 117–124.