



The fake news graph analyzer: An open-source software for characterizing spreaders in large diffusion graphs

Bodaghi, A., Oliveira, J., & Zhu, J. J. H. (2021). The fake news graph analyzer: An open-source software for characterizing spreaders in large diffusion graphs. *Software Impacts*, 10, 1-3.
<https://doi.org/10.1016/j.simpa.2021.100182>

[Link to publication record in Ulster University Research Portal](#)

Publication Status:

Published (in print/issue): 26/11/2021

DOI:

[10.1016/j.simpa.2021.100182](https://doi.org/10.1016/j.simpa.2021.100182)

Document Version

Publisher's PDF, also known as Version of record

Document Licence:

CC BY

General rights

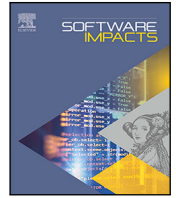
The copyright and moral rights to the output are retained by the output author(s), unless otherwise stated by the document licence.

Unless otherwise stated, users are permitted to download a copy of the output for personal study or non-commercial research and are permitted to freely distribute the URL of the output. They are not permitted to alter, reproduce, distribute or make any commercial use of the output without obtaining the permission of the author(s).

If the document is licenced under Creative Commons, the rights of users of the documents can be found at <https://creativecommons.org/share-your-work/licenses/>.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk



Original software publication

The fake news graph analyzer: An open-source software for characterizing spreaders in large diffusion graphs

Amirhosein Bodaghi ^{a,*}, Jonice Oliveira ^b, Jonathan J.H. Zhu ^a^a Department of Media and Communication, City University of Hong Kong, Hong Kong^b Department of Computing Science, Centre of Mathematical and Natural Sciences – CCMN, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

ARTICLE INFO

Keywords:

Fake news
Social network
Spreading process
Graph analysis
Twitter

ABSTRACT

In the study of fake news spreading, it is essential to know how different types of spreaders differ in terms of their characteristics, interconnections, and cascading flow. The fake news graph analyzer (FNGA) is an open-source software that provides the required computations for such extended analyses on large graphs. Moreover, FNGA generates data for graph visualizations. Also, FNGA is designed to consider the spreading of both fake and true news simultaneously in the graph, leading to a variety of confrontational patterns. FNGA facilitates future research on fake news and the diffusion of any contagion within a graph of entities.

Code metadata

Current code version	v1
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2021-156
Permanent link to Reproducible Capsule	https://codeocean.com/capsule/7614966/tree/v1
Legal Code License	MIT License
Code versioning system used	GIT
Software code languages, tools, and services used	Python
Compilation requirements, operating environments & dependencies	Python packages: <i>numpy</i> , <i>scipy</i> , <i>snap-stanford</i> , <i>xrld</i>
If available Link to developer documentation/manual	Not provided
Support email for questions	abodaghi@cityu.edu.hk

1. Introduction

Following the 2016 US Presidential Election, fake news has gained attention in academic scholarship [1]. The fast-growing literature on fake news has studied this phenomenon in various aspects in which investigations on user-based features play a significant role. The analysis based on user characteristics has roots in a variety of research contexts, from sports [2] to social media [3]. However, when it comes to spreading fake news or viral contagions, the underlying graph of users and the characteristics driven from it would play a major role in the analysis. Particularly when it comes to modeling the spreading process of fake news on social media, considering the graph-based characteristics of users appear to be a necessity [4,5].

When an external source, particularly a junk news website, publishes a fake news post for the first time, some Twitter users fetch and tweet it [6]. Then retweeters come along and diffuse the fake news. The exact process occurs for spreading the truth, which encounters that fake news. Therefore, the play's four main characteristics are the tweeter of fake news, retweeter of fake news, tweeter of truth, and retweeter of truth. However, some tweeters earn lots of retweets, making them Super Spreader, while many other tweeters of the same content find no success to have their tweets retweeted by anyone and end up as Unwelcome Spreaders. A fake news graph of diffusion is constituted of ego networks of all these different types of spreaders. Each ego network itself is made of all followers and followings of a single spreader as its nodes while connected to the spreader by directed edges.

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: abodaghi@cityu.edu.hk (A. Bodaghi).

<https://doi.org/10.1016/j.simpa.2021.100182>

Received 12 November 2021; Accepted 19 November 2021

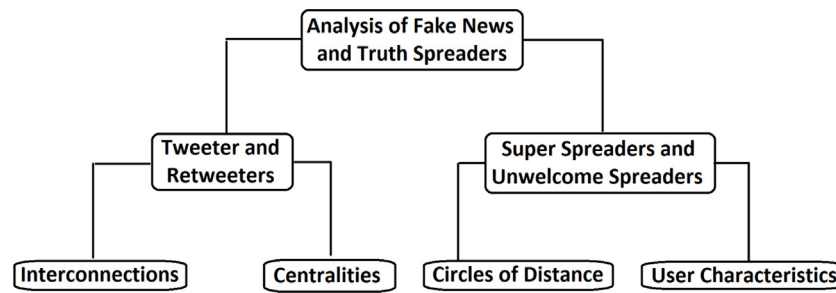


Fig. 1. The structure of the analysis.

2. Description

The FNGA^{1,2} is written in python and uses libraries such as Snap-Stanford [7], Numpy, and Scipy. The inputs for each fake news are: 1- Dataset of Diffusion (DD), an Excel file containing all the tweets related to the fake news. For each tweet that occupies one row in the file, various information is provided for the user who published it and the tweet itself. This information for the user includes ID, #tweets, #followings, #followers, data and time of the account creation, language, and description sentence. While the information for the tweet include date and time of publishing, #favorite, type of tweet, which can be a retweet, quote, reply, or original, the tweet and user IDs of the referred tweet in case the tweet be a retweet, quote, or reply, frequency of tweet occurrence in the dataset, and finally the last feature is the nature of tweet which can be fake news (marked as 'r'), truth (marked as 'a'), or just questioning about the fake news (marked as 'q'). 2- Dataset of Graph (DG), a Graph file containing the nodes and edges of the fake news graph. Indeed, this file could be achieved by applying the Snap.py a Python interface of SNAP, which confers an agile and light environment for big graph analysis [7], to run a process of compression and conversion upon the raw list-type data of followers and followings for each spreader of the fake news, to attain the agile graph dataset. In this way, for each dataset, we acquired a directed network in which each node represents a user, and a link from user i to user j is established if i follows j . 3- The ID of users in DG, which is brought into a Jason file.

After reading the inputs, six analysis steps will be done on each fake news by the software. Step1: Extract the information of retweeters of the superspreaders' tweets at different circles of distance around the super spreaders. Indeed, first a set of categories based on the distance to the super spreader are defined, then for members of each category, the following characteristics are calculated; 1- ratio of followings to followers (FwFr), 2- ratio of mutual followings to followings in general (MFw), 3- number of followings who had spread the most popular tweet before, as the antecedent spreaders of the given member (NA), 4- the number of passed seconds from the time the super spreader had spread the most popular tweet until the time the given member retweeted this tweet (TDS), 5- the number of passed seconds from the time the first antecedent spreader had spread the most popular tweet until the time the given member retweeted this tweet (TDFA), 6- number of published tweets and retweets per year from the given member (NTPY), which is calculated based on the overall activities since the creation of the account, 7- number of antecedent spreaders for the opposite contagion i.e. users who have been a following of the given member and had spread the truth/fake news before the given member spread fake news/truth (NAO). The software at this step calculates the mean and standard deviance of the above characteristics for users in each category, and the correlations of those characteristics with the TDFA at each category are computed.

Step2: Extract the information of unwelcome spreaders, including calculating a set of characteristics such as FwFr, MFw, and NTPY. Step3: Generating the input files for graph visualization by Cytoscape. Step4: Extracting the information of interconnections between tweeters and retweeters of fake news and truth. First, the software divides users into four groups; Fake News Tweeters (FT), Fake News Retweeters (FR), Truth Tweeters (TT), and Truth Retweeters (TR). The members of these groups extract a series of information, including 1- The fraction of followings (for the given member) who are members of the other groups. 2- The fraction of whole members of each group that are the followings of the given member. 3- The fraction of followers (for the given member) who are members of the other groups. 4- The fraction of whole members of each group that are the followers of the given member. Eventually, we compute the mean and standard deviation of these variables for members of each category. Step5: Extracting the general information of each fake news graph, including the number of nodes and edges of different types, triangles, triads, connected components, and length of the effective diameter. Step6: Extracting the information of centralities for tweeters and retweeters of fake news and truth. These centralities are calculated at the singular level (for single nodes) and plural level (for all nodes of a category). The former includes centralities such as Page Rank, Closeness, Hubs, and Authorities score, while the latter encompasses Modularity and the ratio of intra- to inter-links for each of the groups i.e. FT, FR, TT, and TR. Each of these six steps provides a unique output file (a .text file) in which the results are presented.

3. Impact

The FNGA enables an extensive range of analysis over fake news graphs driven from social media. A summary of these analyses can be seen in Fig. 1.

The analyses conducted by this software in recent research [8] proved its ability to reveal new findings towards a better understanding of fake news diffusion on social media. Furthermore, findings of different categories' centralities and interconnections of spreaders depict a practical perspective that can guide practitioners to build better strategies against fake news spreading on social media. Moreover, the investigations on collective behaviors such as the wave-like forms of diffusion and their associations with micro behaviors are still in their infancy [9–11]. In fact, research on the impact of cyberpsychology [12] at the aggregate level remains open to yield more profound insights into the diffusion process of fake news in social media. This software may pave the way for such research by providing an extensive range of analyses.

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.

¹ <https://github.com/AmirhoseinBodaghi/P9>.

² <https://codeocean.com/capsule/7614966/tree/v1>.

Acknowledgments

The study was partially funded by City University of Hong Kong Centre for Communication Research (No. 9360120) and Hong Kong Institute of Data Science (No. 9360163).

References

- [1] E. Tandoc, J. Jenkins, S. Craft, Fake news as a critical incident in journalism, *Journalism Pract.* 13 (6) (2019) 673–689.
- [2] A. Bodaghi, A novel pervasive computing method to enhance efficiency of walking activity, *Health Technol.* 6 (2016) 269–276.
- [3] X. Zhang, A. Ghorbani, An overview of online fake news: Characterization, detection, and discussion, *Inf. Process. Manage.* 57 (2) (2019) 102025.
- [4] A. Bodaghi, S. Goliaei, A novel model for rumor spreading on social networks with considering the influence of dissenting opinions, *Adv. Complex Syst.* 21 (2018) 1850011, <http://dx.doi.org/10.1142/S021952591850011X>.
- [5] A. Bodaghi, S. Goliaei, M. Salehi, The number of followings as an influential factor in rumor spreading, *Appl. Math. Comput.* 357 (2019) 167–184.
- [6] A. Bondielli, F. Marcelloni, A survey on fake news and rumor detection techniques, *Inform. Sci.* 497 (2019) 38–55.
- [7] J. Leskovec, R. Soric, Snap: A general-purpose network analysis and graph-mining library, *ACM Trans. Intell. Syst. Technol. (TIST)* 8 (1) (2016) 1.
- [8] A. Bodaghi, J. Oliveira, The theater of fake news spreading, who plays which role? A study on real graphs of spreading on Twitter, *Expert Syst. Appl.* (2021) 116110, <http://dx.doi.org/10.1016/j.eswa.2021.116110>.
- [9] A. Bodaghi, J. Oliveira, The characteristics of rumor spreaders on Twitter: A quantitative analysis on real data, *Comput. Commun.* 160 (2020) 674–687.
- [10] P. Meel, D.K. Vishwakarma, A temporal ensembling based semi-supervised ConvNet for the detection of fake news articles, *Expert Syst. Appl.* 177 (1) (2021) 115002.
- [11] A. Choudhary, A. Arora, Linguistic feature based learning model for fake news detection and classification, *Expert Syst. Appl.* 169 (1) (2021) 11417.
- [12] Amirhosein Bodaghi, Jonice Oliveira, A longitudinal analysis on Instagram characteristics of Olympic champions, *Social Network Anal. Mining* 12 (1) (2022) <http://dx.doi.org/10.1007/s13278-021-00838-9>.