Software Impacts 13 (2022) 100341

Contents lists available at ScienceDirect

Software Impacts

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

Original software publication

MMSS: A storytelling simulation software to mitigate misinformation on social media

Ahmed Abouzeid *, Ole-Christoffer Granmo

University of Agder, Norway

ARTICLE INFO

Keywords:
Intervention-based simulations
Hawkes process
Stochastic optimization
Misinformation mitigation
Learning automata

ABSTRACT

This paper proposes a modular python implementation of a storytelling simulation. The software evaluates misinformation mitigation strategies over social media and visualizes the investigated scenarios' potential outcomes. Our software integrates information diffusion and control models components. The control model mitigates users' exposure to misinformation with social fairness awareness, while the diffusion model predicts the outcome from the control model. During the interaction of both models, a graph coloring algorithm traces the interaction within specific time intervals. Then, it generates meta-data to construct visuals of predicted near-future states of the social network to help support decision-making and evaluate proposed mitigation strategies.

Code metadata

Current code version	v1.0
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2022-87
Permanent link to Reproducible Capsule	https://codeocean.com/capsule/6681283/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 3.8 or higher
If available Link to developer documentation/manual	https://github.com/Ahmed-Abouzeid/MMSS
Support email for questions	ahmed.abouzeid@uia.com

1. Introduction

Our Misinformation Mitigation Storytelling Simulation (MMSS) software works by receiving social network data. The data must consist of historical observations regarding the information dissemination behavior for the individual users. Such behavior is represented by the timestamps of each user for when a particular type of information was created. The MMSS expects such information to be labeled as either misinformation or regular content. Then the next task is for the information diffusion component to model and learn from these recorded observations. Typically, the information diffusion model task is to predict the future behavior of each user if either misinformation or regular content will be circulated during a specific time realization soon. However, the information diffusion model is controlled first before conducting any predictions, allowing for a misinformation mitigation strategy to alter the present for better consequences in the future. Therefore, each user is picked up and assigned an artificial intelligence-based agent to learn how such a user will influence and be influenced inside the network. The MMSS has a sampler component to shrink vast networks or target a specific community. Then, the picked user's behavior is examined by the controller agent within the sampled network. Sampled networks can be formed randomly or by targetting specific groups. When all users are assigned agents and

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 addresses: ahmed.abouzeid@uia.no (A. Abouzeid), ole.granmo@uia.no (O.-C. Granmo).

https://doi.org/10.1016/j.simpa.2022.100341
Received 6 June 2022; Received in revised form 20 June 2022; Accepted 22 June 2022

the latter finalized learning and converged, the controller component reports the results of its interventions to the data visualization layer to view intuitions about how effective a mitigation strategy could be. The MMSS helps investigate different mitigation strategies through system parameters. The experiments and empirical results from adopting our software are illustrated in [1]. Eventually, reported results are standard colored graphs meta-data that can be easily integrated with graph visualization software. Fig. 1 shows the main components and layers of the MMSS software.

2. The motivation for MMSS

Efficient Crisis Management Systems rely on visual analytics to filter and visualize relevant information extracted from social media platforms like Twitter and Facebook. The provided analytics from these tools equip emergency responders with different points of view to explore and better understand the situation and take a specific course of actions [2]. Recent works have proposed data visualization techniques for emergency operators. These tools took advantage of the large amount of data generated on the social networks every second. A non-exclusive list of applications is health monitoring [3], organized crime [4], hate speech [5], and gender bias [6]. On the other hand, simulation framework designs were introduced for crisis management as well. The latter has a distinctive advantage in informing emergency responders of potential risks or preferable actions to mitigate threats in the future. Furthermore, other efforts [7] presented a perennial simulation framework that targets crisis management simulation. Their framework incorporated concepts of dynamic data-driven systems, symbiotic simulation, and human-in-the-loop techniques [8]. Others worked on visualization-based techniques, [9] and proposed a mass transport system simulation to familiarize the operators on the way to handle emergencies by carrying out virtual drills. To the best of our knowledge, the literature lacks resolutions for simulation-based visualization techniques for the problem of misinformation mitigation on social media. Hence, the proposed software in this paper tries to fill the gap and provide a practical resolution to emergency responders to help support their decision-making when working on an infodemic crisis.

3. Methodological foundation

3.1. The diffusion model

The information diffusion model is responsible for predicting the behavior of individual users on the social network. The model is based on a Multivariate Hawkes process (MHP) as practiced by [10,11], and [12]. An MHP is a multivariate stochastic process [13] which models the occurrence of temporal or spatio-temporal asynchronous events by capturing their mutual dependencies. The behavior of each individual on the network is modeled through two Hawkes processes (HP), one for the misinformation dissemination the user is involved with and the other for the regular content dissemination of the same user. The associated user HPs generate estimated random counts for both information kinds, given some behavior observation in the past. For instance, predicting future re/tweeted events given the observed mutual dependencies with other users, where observed dependencies are estimated from the given social network historical and labeled data. These final counts indicate the intensity of each behavior style at a specific time realization. Mathematically, an HP can be defined with its conditional intensity function λ. The conditional intensity function has two significant elements: base intensity μ and exponential decay function g over an adjacency matrix A which represents the estimated mutual dependencies. The formal explanation of the conditional intensity function is given as the below equation:

$$\lambda_i(t_r|H^{t_r}) := \mu_i + \sum_{t_s<t_r} g(t_r - t_s). \tag{1}$$

Where μ is the base intensity that represents some external motivation to propagate some content. On the other hand, g is some kernel function over the observed history H^{t_r} associated with user i from the discrete time realization t_s prior to time t_r. g is concerned with the history of some influence matrix $A_{i.}$, where $A_{ij} = 1$ if there is an influence indicating that user i follows user j or quotes (with agreement) content from j, and $A_{ij} = 0$ if not. We used an exponential decay kernel function $g = A_{i.}e^{-wt}$ as practiced by [12], where w is the decay factor which represents the rate for how the influence is reduced over time. For all users, the base intensity vector μ and the influence matrix A can be estimated using maximum likelihood as proposed in [14]. To predict all users' behaviors for each content type, an MHP is created, given that different intensity rates are generated at different discrete-time realizations. Hence, we model each user behavior at each time realization as an estimated number of events (misinformation or regular). We utilized the *tick* library [15] for our implementation of the Hawkes-based information diffusion model.

3.2. The control model

On the other hand, controlling the diffusion model means introducing additional information to the network to alter the diffusion's future outcomes. The misinformation mitigation strategies could be implemented by introducing specific information to the network. For instance, when an individual is exposed to a certain amount of misinformation and then being exposed to its correction. Such imposed correction can be viewed as mitigating the impact of manipulating content. Therefore, the base intensity μ in the regular content HP is the element being under control. Therefore, For each user, an incentivization value x is added to the external motivation of the HP. The incentivization values are bounded by a predetermined mitigation budget C, representing time limitations or other incentivization constraints. Hence, let x_i be the incentivization amount decided for user i and $x_i \leq C$, the modified HP for the mitigation purposes can be redefined by the below equation:

$$\lambda_i(t_r|H^{t_r}) := x_i + \mu_i + \sum_{t_s<t_r} g(t_r - t_s). \tag{2}$$

To intelligently apply incentivizations to the network, we need an intervention model that can learn from the observed social network dynamics. Therefore, we utilized Learning Automata (LA) [16] for its capabilities, easy implementation, and light-weight computation. We believe easier implementations and lower computation costs are important and needed for the practicality of our proposed MMSS. For each individual user, we assign an LA controller to intervene with the simulated user from the information diffusion model. Therefore, We built a network of user-assigned LAs [11] for learning optimum incentivization value needed for each assigned user in an optimum mitigation strategy.

3.3. Results visualization

The data visualization layer takes advantage of our proposed graph coloring algorithm. The latter mainly produces meta-data for colored and sized graph nodes. The outputted information provides a detailed story that includes temporal changes in the predicted consequences over the network. For instance, when users start to follow, tweet, and retweet specific information types with other users, that includes starting and ending times. The traced temporal information stored for nodes and edges is important to provide dynamic transitioned graphs overtime frame by frame. Also, the colored graph meta-data contains different nodes flags to help distinguish users' nodes from content nodes. For a fully detailed illustration about the dynamic graph transitions generated for the different told stories by the MMSS, please visit the following link. https://www.youtube.com/watch?v=Lqmp4PdWCp4.

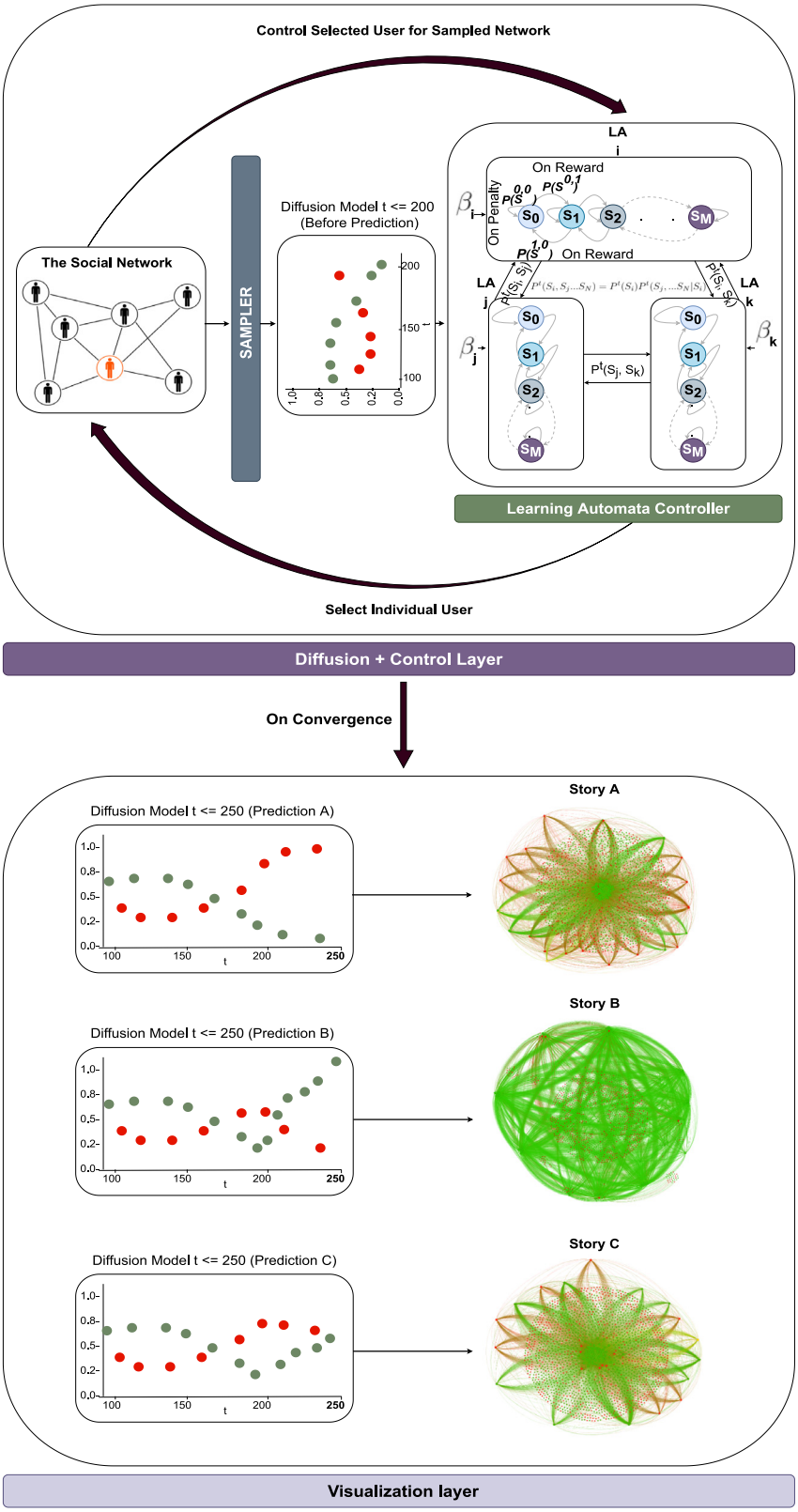

Fig. 1. MMSS architecture and underlying components. Red nodes indicate misinformation, while green nodes indicate regular content. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. MMSS impact and target domain

As discussed earlier, there is a critical need for simulation-based visualization techniques for the problem of misinformation mitigation on social media. The proposed MMSS software provides a resolution and fills such gap. Moreover, the provided mitigation strategy parameters can help investigating different strategies with different consequences. The latter is a big advantage when working on emergencies and crisis related to political manipulation for instance. Hence, the proposed MMSS software can be utilized by emergency responders when an infodemic is causing social disorder or an extreme level of political polarization based on some political misinformation. Therefore, mitigating the misleading content on social media could be approached by using the MMSS to evaluate and adopt learned misinformation mitigation strategies, where the outcome of the MMSS would be the different consequences of the different evaluated and simulated mitigation strategies. The latter outcome can also be represented by a dynamic social network state transitions on temporal basis, generated by the graph coloring component the MMSS has. The mentioned impact was scientifically evaluated on both real and synthetic data and the work [1] was accepted under the track of AI for social impact at the AAAI22 venue.

5. Adoption of MMSS in misinformation research

The development of MMSS considered multiple aspects to provide practicality and robustness as a misinformation crisis management system. Some of these offered practicalities are mitigation strategy parameters, traced temporal meta-data while simulating the mitigation strategy, and an easy and decentralized implementation inspired by the capabilities of the Learning Automata network. However, the problem of social media misinformation is intricate and interconnected with other issues like hate speech, political polarization, and the echo chamber effect inherited by social media platforms' technology. Therefore, we believe the MMSS's modular implementation could be the first step towards a fully integrated pipeline for a storytelling simulation for misinformation/hate speech/polarization mitigation on social media. The mitigation incentives could be viewed as motivational causes social media providers can apply to specific users at specific time intervals. After evaluating different mitigation strategies and supporting mitigation decisions, how intensive each user's motivation could be is decided from the MMSS learning. Moreover, the modular implementation of the MMSS software allows for reusing a subset of its components while replacing the others for the sake of scientific evaluation. For instance, researchers can adopt different controller models to make an analogy between different controllers within the MMSS pipeline. Similarly, information diffusion model can be replaced with another diffusion methodology.

6. Limitations and future work

Modeling human behavior is entirely different and more complicated than modeling an artificial system. Hence, formal verification of such social models should be studied along with interpretable Artificial Intelligence consideration in the utilized methods. For instance, interpretable information diffusion models are highly needed to avoid mistaken imitation of users' behaviors which is considered a drawback of misinformation classification methods. Additionally, the proposed MMSS depends on labeled historical data from the social network, which makes the MMSS still dependent on such classification methods. Although, the mitigation approach reduces the risk of these classifiers' drawbacks since the former works by incentivization, making a wrong decided incentivization value less harmful. However, integrating an interpretable misinformation classifier with MMSS in future work is also highly recommended for both interpretability and dependability characteristics.

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.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.simpa.2022.100341.

References

[1] Ahmed Abouzeid, Ole-Christoffer Granmo, Christian Webersik, Morten Goodwin, Socially fair mitigation of misinformation on social networks via constraint stochastic optimization, 2022, arXiv preprint arXiv:2203.12537.
[2] Teresa Onorati, Paloma Díaz, Belen Carrion, From social networks to emergency operation centers: A semantic visualization approach, Future Gener. Comput. Syst. 95 (2019) 829–840.
[3] Rongheng Lin, Zezhou Ye, Hao Wang, Budan Wu, Chronic diseases and health monitoring big data: A survey, IEEE Rev. Biomed. Eng. 11 (2018) 275–288.
[4] Simon Andrews, Ben Brewster, Tony Day, Organised crime and social media: a system for detecting, corroborating and visualising weak signals of organised crime online, Secur. Inf. 7 (1) (2018) 1–21.
[5] Arthur TE Capozzi, Mirko Lai, Valerio Basile, Fabio Poletto, Manuela Sanguinetti, Cristina Bosco, Viviana Patti, Giancarlo Ruffo, Cataldo Musto, Marco Polignano, et al., Computational linguistics against hate: Hate speech detection and visualization on social media in the" contro L'odio" project, in: 6th Italian Conference on Computational Linguistics, CLiC-It 2019, Vol. 2481, CEUR-WS, 2019, pp. 1–6.
[6] Prashanth Rao, Maite Taboada, Gender bias in the news: A scalable topic modelling and visualization framework, Front. Artif. Intell. 4 (2021).
[7] Seth N. Hetu, Samarth Gupta, Vinh-An Vu, Gary Tan, A simulation framework for crisis management: Design and use, Simul. Model. Pract. Theory 85 (2018) 15–32.
[8] Fabio Massimo Zanzotto, Human-in-the-loop artificial intelligence, J. Artificial Intelligence Res. 64 (2019) 243–252.
[9] P.K. Kwok, Bill K.P. Chan, Henry Y.K. Lau, A virtual collaborative simulation-based training system, in: Proceedings of the 10th International Conference on Computer Modeling and Simulation, 2018, pp. 258–264.
[10] Mahak Goindani, Jennifer Neville, Social reinforcement learning to combat fake news spread, in: Uncertainty in Artificial Intelligence, PMLR, 2020, pp. 1006–1016.
[11] Ahmed Abouzeid, Ole-Christoffer Granmo, Christian Webersik, Morten Goodwin, Learning automata-based misinformation mitigation via Hawkes processes, Inf. Syst. Front. (2021) 1–20.
[12] Mehrdad Farajtabar, Jiachen Yang, Xiaojing Ye, Huan Xu, Rakshit Trivedi, Elias Khalil, Shuang Li, Le Song, Hongyuan Zha, Fake news mitigation via point process based intervention, in: International Conference on Machine Learning, PMLR, 2017, pp. 1097–1106.
[13] Yuanda Chen, Thinning algorithms for simulating point processes, Florida State University, Tallahassee, FL, 2016.
[14] Tohru Ozaki, Maximum likelihood estimation of Hawkes' self-exciting point processes, Ann. Inst. Statist. Math. 31 (1) (1979) 145–155.
[15] Emmanuel Bacry, Martin Bompaire, Philip Deegan, Stéphane Gaïffas, Søren Poulsen, Tick: a Python library for statistical learning, with an emphasis on Hawkes processes and time-dependent models, J. Mach. Learn. Res. 18 (1) (2017) 7937–7941.
[16] Kumpati S. Narendra, Mandayam A.L. Thathachar, Learning automata-a survey, IEEE Trans. Syst. Man Cybern. (4) (1974) 323–334.