Part III: Misinformation mitigation

Xiuzhen Jenny Zhang (xiuzhen.zhang@rmit.edu.au)

RMIT University

RMIT UNIVERSITY

Contents

- Network-level misinformation mitigation
  - Social media information diffusion models
  - Heuristic-based mitigation models
  - Learning-based mitigation models
- Personalised misinformation mitigation
  - Personalised information recommendation
  - Veracity-aware information recommendation for mitigation

Misinformation spread, despite detection efforts

- Despite tremendous efforts for detection from fact checking services and automated systems, misinformation still going around.

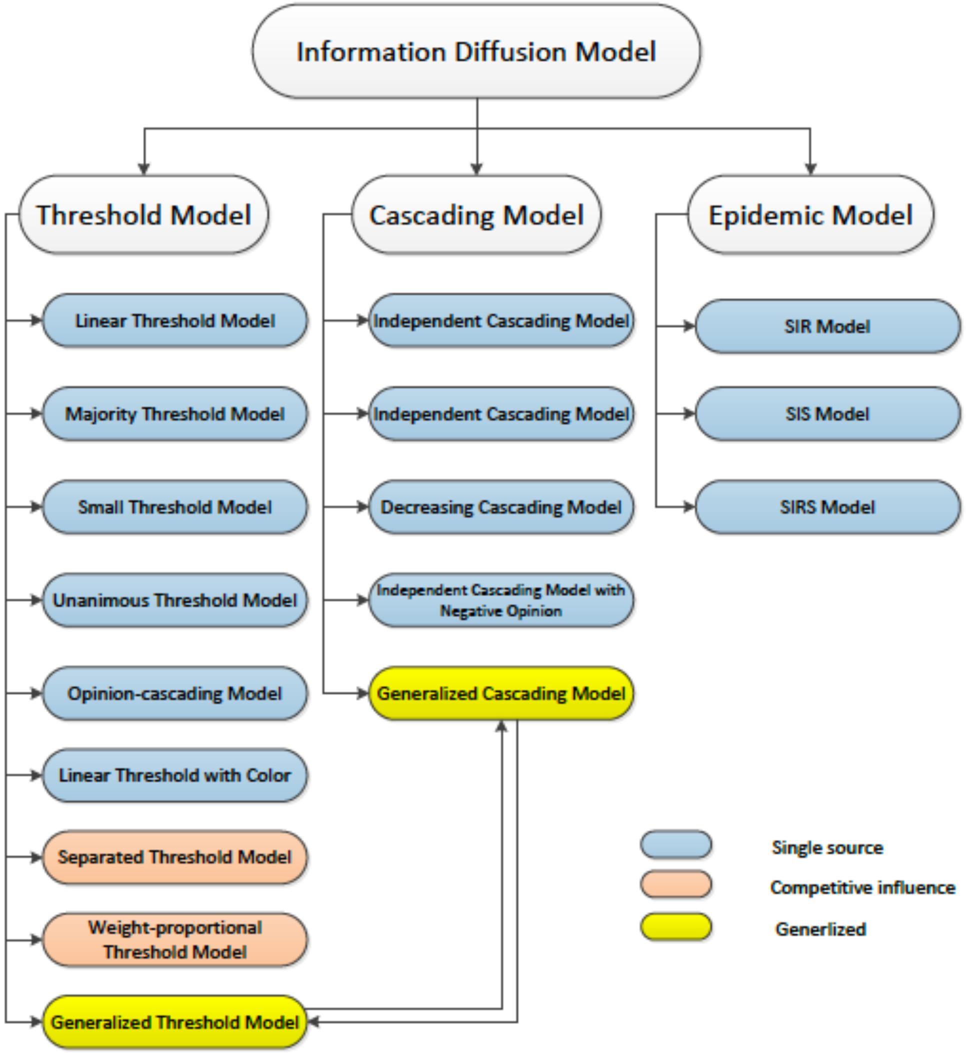
- So …

- *Mitigation* is critically important.

Information propagation models

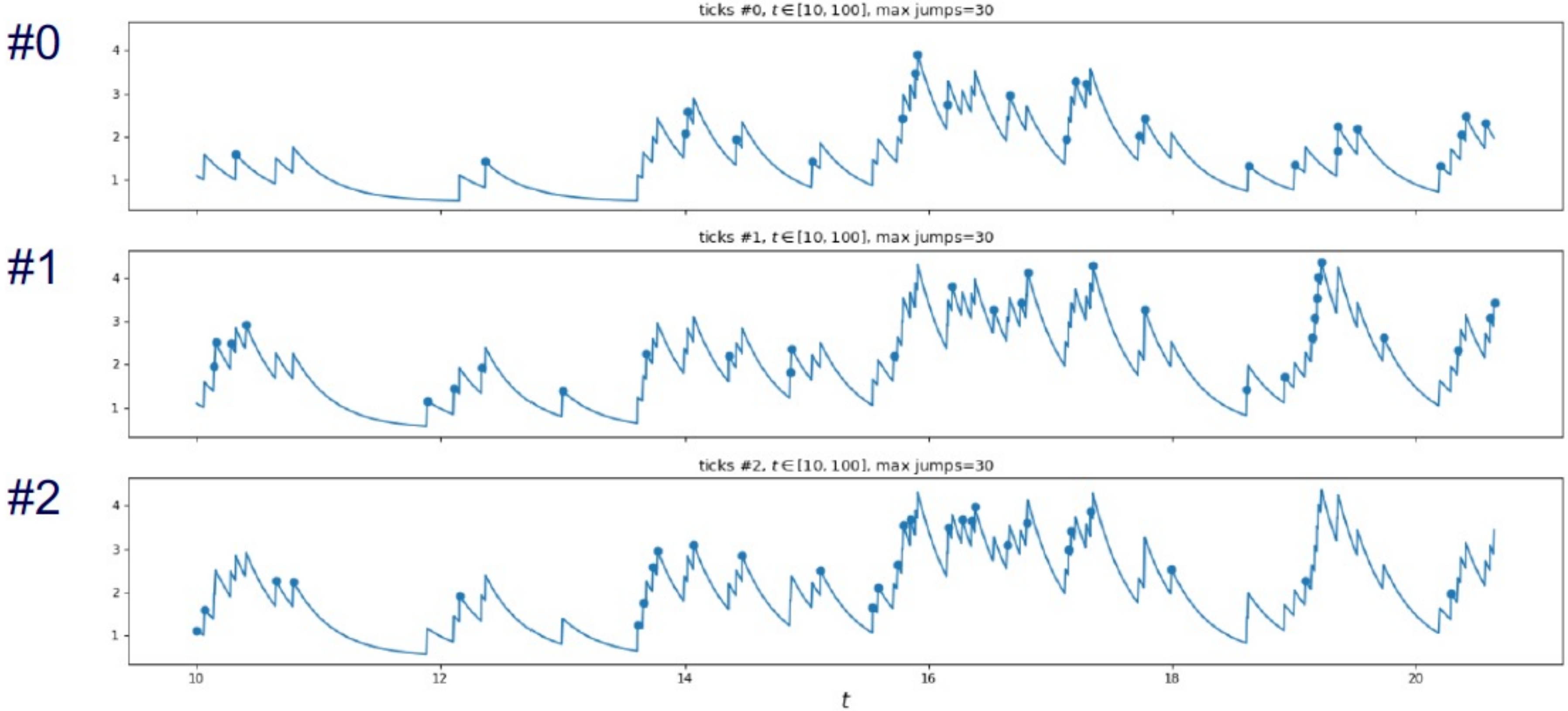
Information propagation models on social media	Multivariate Hawkes process (MHP)	Farajtabar et al. 2016; Farajtabar et al. 2017; Lacombe 2018; Shu, Bernard & Liu 2019; Goindani & Neville 2020a; Goindani & Neville 2020b
	Linear Threshold (LT) / Independent Cascade (IC)	Pham et al.; Saxena & Gera 2020a
	Information Aggregation Game	Aymanns et al. 2019
	Epidemic Models	Tan et al. 2019; Wen et al. 2015; Wen et al. 2014

The linear threshold model



Source: Zhang, H., Mishra, S., & Thai, M.T. (2014). Recent Advances in Information Diffusion and Influence Maximization of Complex Social Networks.

The multivariate Hawkes process



Source: https://x-datainitiative.github.io/tick/modules/hawkes.html

Modelling user influence and rumour propagation

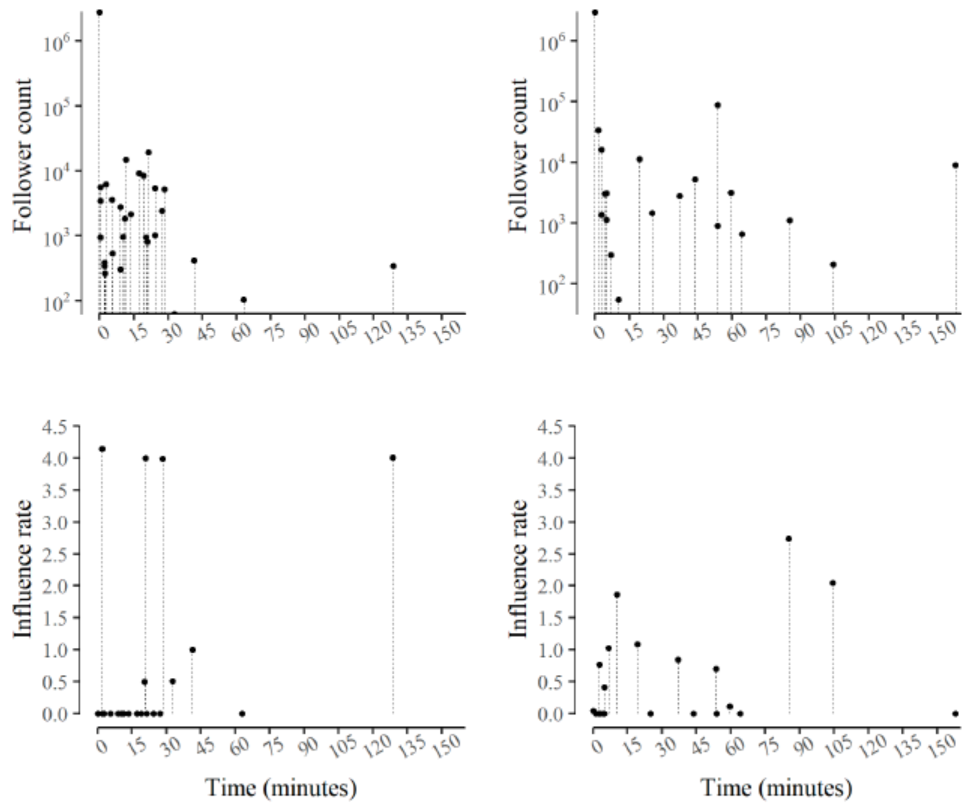
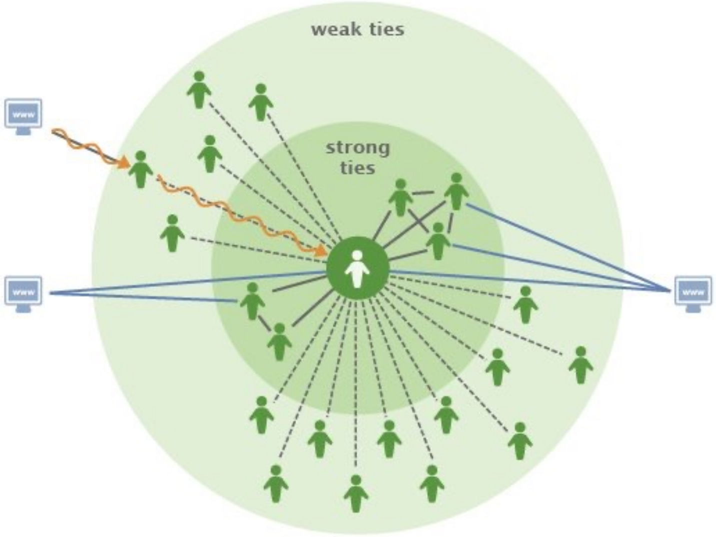


Figure 3.1: An example of the diffusion of a rumor (top and bottom left) and a non- rumor (top and bottom right) across time with different measures of user influence from twitter16 dataset. x-axis: time (in minutes); y-axis: user influence is either follower count or influence rate. The quantity, **Influence rate** is our proposed measurement of user influence in information propagation.

H. Ruda & X. Zhang et al. “Modelling User Influence and Rumor Propagation on Twitter using Hawkes Processes”. DSAA’2020.

Network-level mitigation

- Develop strategies to introduce true news to counteract the spread of fake news on social networks.
- Information diffusion models:
  - ✓ Independent Cascade (IC) and Linear Threshold (LT) models
  - ✓ Point process models
  - ✓ Reinforcement learning
- Maximize the propagation of true news on social networks.



Network-level mitigation

Social network misinformation mitigation

Debunker selection		Information diffusion
Heuristic approach	Learning-based approach	Learning-based approach
- (Wen et al. 2014) - (Tan et al. 2019) - (Saxena & Gera 2020a) - (Saxena & Gera 2020b)	- (Xu et al. 2022)	- (Farajtabar et al. 2016) - (Farajtabar et al. 2017) - (Goindani & Neville 2020a) - (Goindani & Neville 2020b)

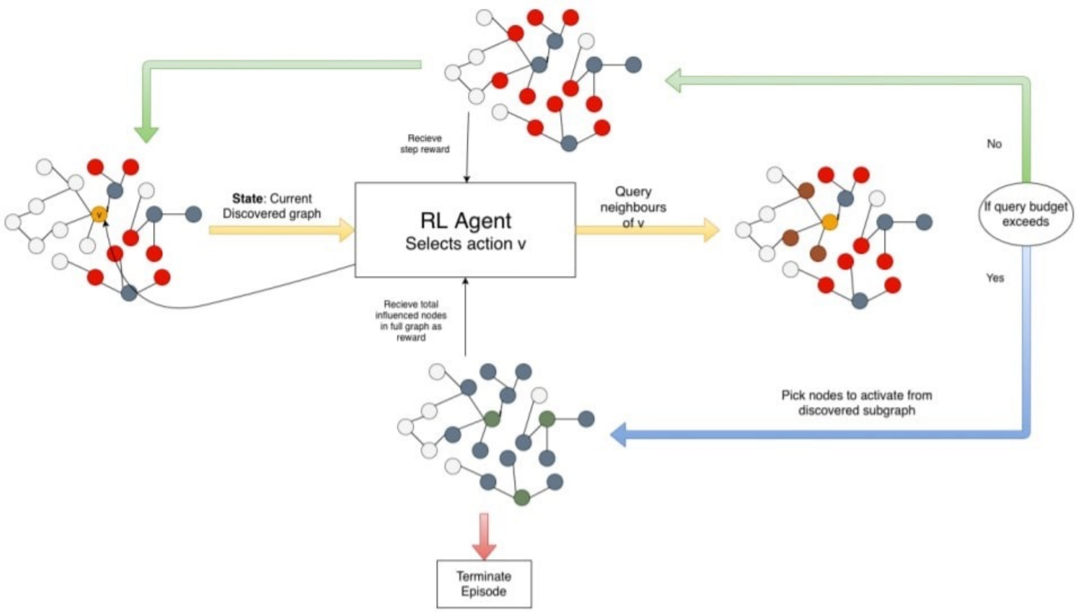
Selection of debunkers: Heuristic-based

Heuristically select influential users to block misinformation or propagate true information.

- Most influential users and/or community bridges to block misinformation and/or propagate true information based on the epidemic model (Wen et al, 2014).
- Select nodes to block misinformation based on the epidemic model (Tan, et al., 2019).
- Top-K debunkers to spread truth based on a dynamic opinion propagation model (Saxena et.al, 2020a, 2020b).

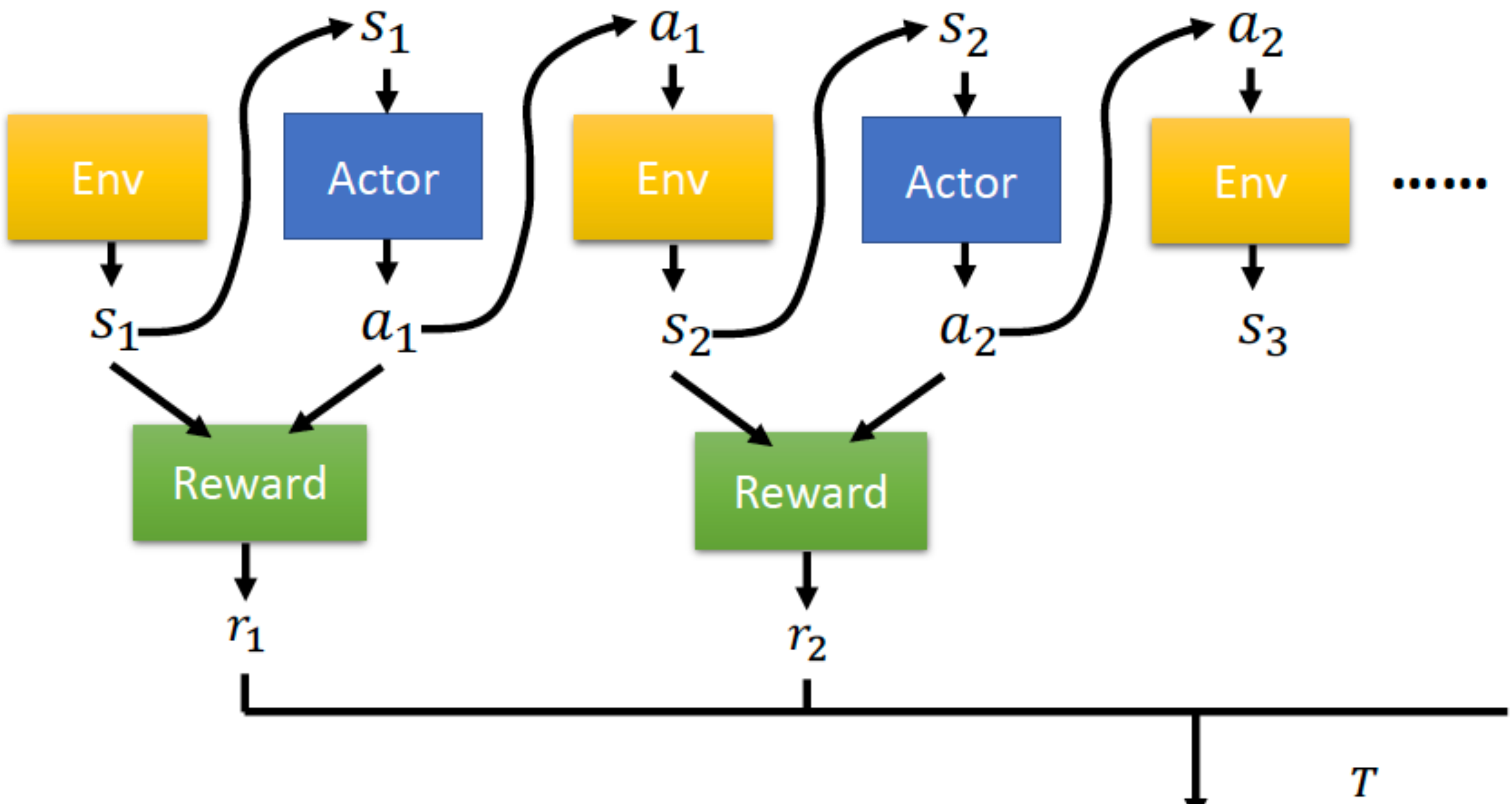
Selection of Cost-effective Debunkers for Multi-stage Fake News Mitigation

- A reinforcement learning problem: train a mitigation policy that optimizes debunker selection from social networks at multiple stages.



Xu, Xiaofei, Ke Deng, and Xiuzhen Zhang. "Identifying cost-effective debunkers for multi-stage fake news mitigation campaigns." *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 2022.

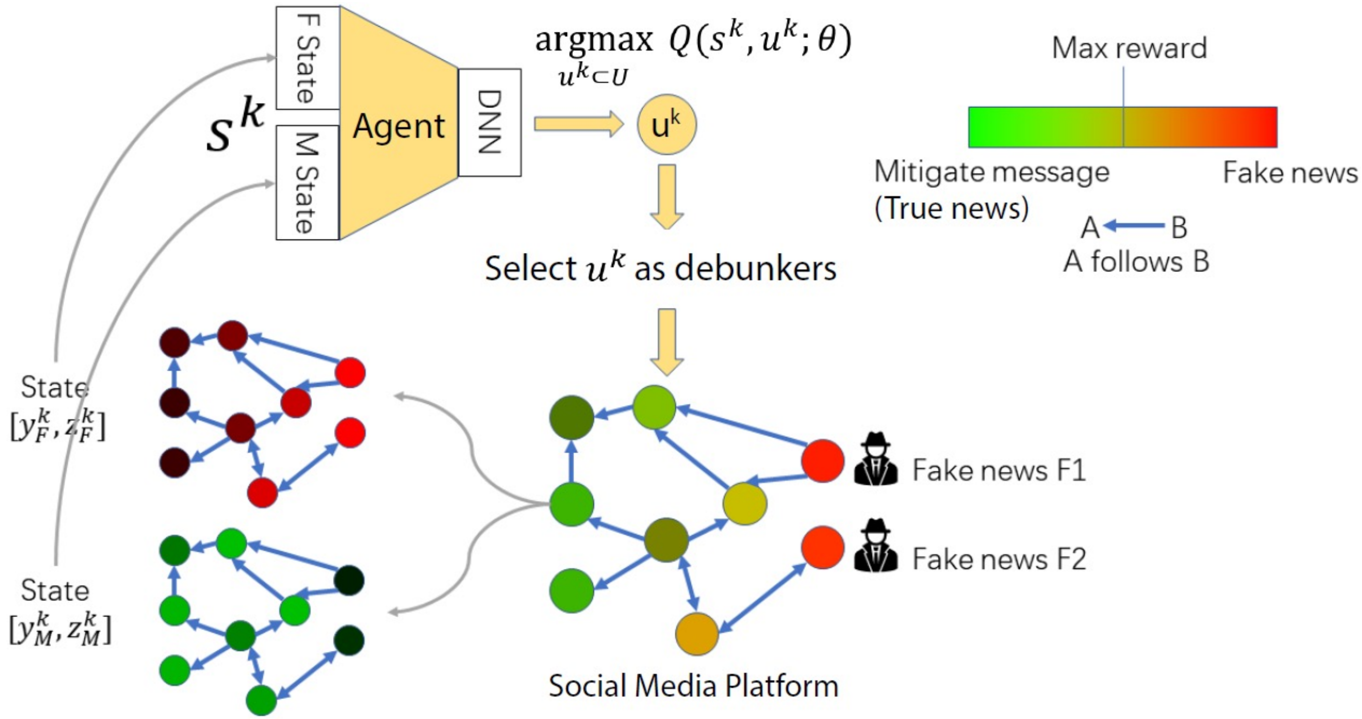
Preliminaries: reinforcement learning



Trajectory $\tau = \{s_1, a_1, s_2, a_2, \cdots, s_T, a_T\}$ $R(\tau) = \sum_{t=1}^{T} r_t$

Problem formulation

- Environment: Multivariate Hawkes Process.
- State: (intensity and number of posting true and fake news, #followers)
- Action: select users to propagate true news at higher intensity.



Problem formulation: reward function

- The objective of fake news mitigation is to maximise the correlation reward – more exposure to fake news correlates with more feed of true news.

$$r(s^k, u^k) = \frac{1}{n} \mathcal{M}^k(t_{k+1}; s^k, u^k)^\top \mathcal{F}^k(t_{k+1}; s^k, u^k)$$

Challenges

Mitigation overlap among debunkers

Large action space for selecting N users

Methodology: DQN-FSP

- Initialize with single-debunker mitigation.
- Use DQN with memory replay in action level.
- Use Future-State-Prediction (FSP) with memory replay in episode level.

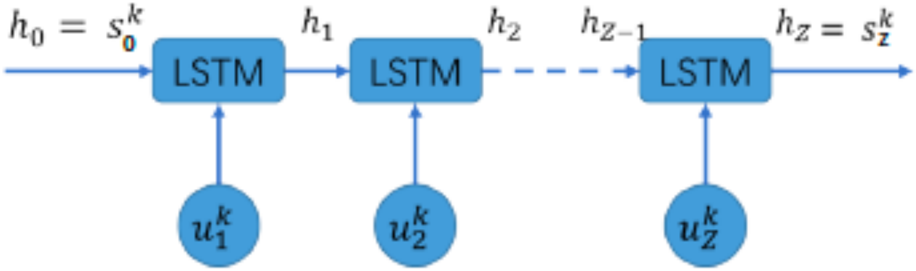


Figure 2: Future state prediction with the LSTM RNN model.

Methodology …

- Extend to multi-debunker mitigation.
- Stop training DQN.
- Use FSP with memory replay in episode level.

Experiment: datasets

- Synthetic: Randomly generate the environment parameters
  - Density test
  - Network size test
  - Average stage length test
  - Number of stages test
- Real world: Learn environment parameters using data
  - Dataset: PHEME
  - Density test
  - Average stage length test
  - Number of stages test

Experiment: baselines

- Random (RND)
- Max Influence (MAX-INF)
- Max Coverage (MAX-COV)
- Neural Network (NN)
- Deep Q-Network (DQN)
- Least-squares Temporal Difference (LTD, in analysis)

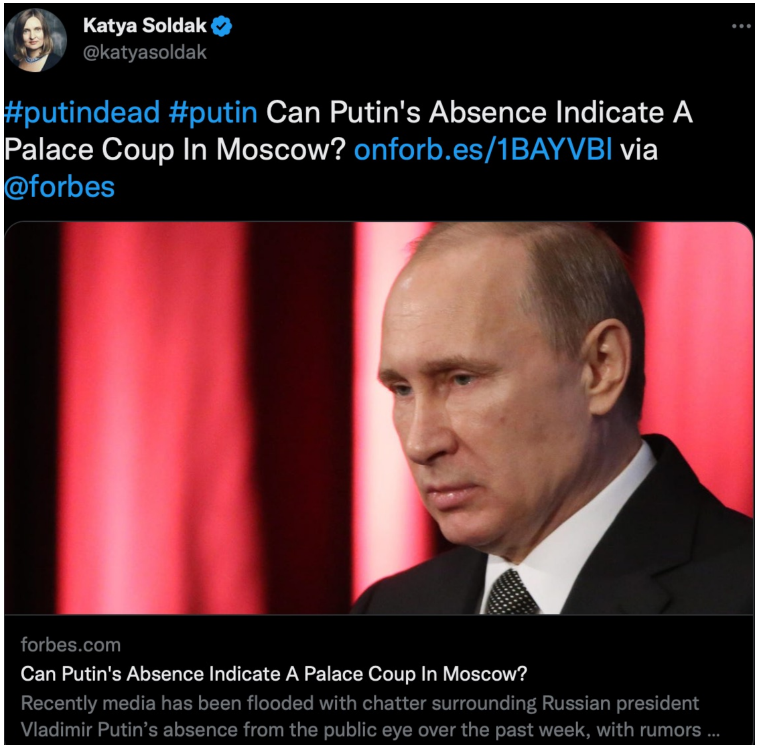
The PHEME Rumour dataset

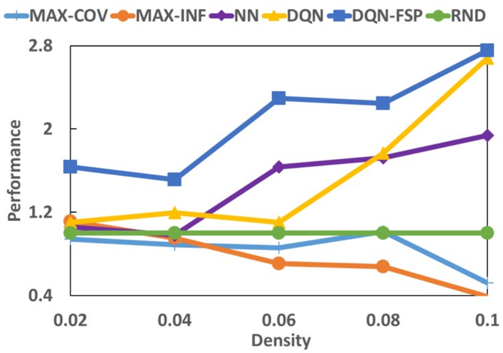
Table 1: Statistics of the real-world PHEME dataset

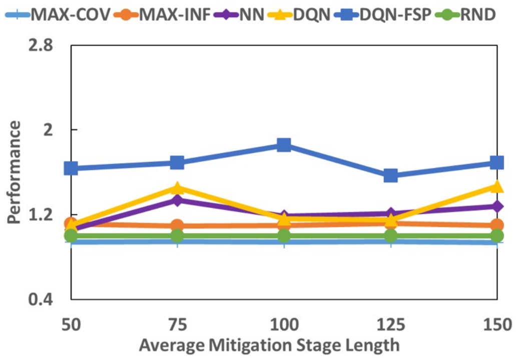
Topic	#Users	Fake tweets	True tweets
Gurlitt (GUR)	98	70	159
Prince Toronto (PRI)	322	483	489
Putin Missing (PUT)	352	251	468

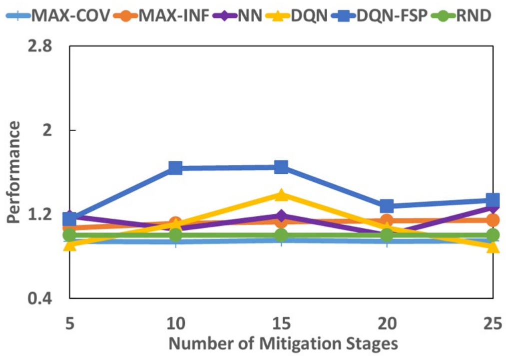
Zubiaga, Arkaitz, et al. "Analysing how people orient to and spread rumours in social media by looking at conversational threads." *PloS one* 11.3 (2016): e0150989.

Results: the PHEME "Putin Missing" story









Network-level mitigation: summary

- Generally overall objective is to block misinformation and/or maximise true information propagation on social networks.
- Effectiveness for individual users not considered.
- Need to consider the diverse topics and events.

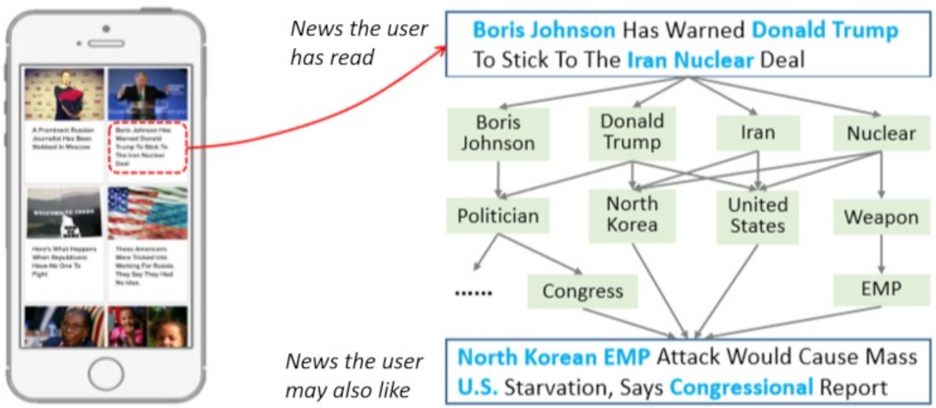
So, …

How to design practical intervention strategies to deter individual users from sharing misinformation?

Recommend personalised corrective true news to individual users.

News recommender systems

News recommender systems have been playing an increasingly important role in influencing and even changing users’ reading behaviours.



An example of conventional news recommender system

Conventional news recommender systems

Can we just employ the existing conventional news recommender systems for fake news mitigation?

No!

Veracity-aware news recommendation for fake news mitigation

- News recommendation for fake news mitigation is based on uniquely characterized data and has a different goal.
- Data characteristics:
  - ✓ Event information (e.g., US election, Covid-19)
  - ✓ Veracity information (i.e., true or fake)

- Goal:
  - ✓ Personalized (the recommended news should be relevant to the events the user recently focused on)
  - ✓ True (the recommended news should be true)

Wang, S., Xu, X., Zhang, X., Wang, Y. and Song, W., 2022, April. Veracity-aware and Event-driven Personalized News Recommendation for Fake News Mitigation In *Proceedings of the Web Conference.*

Challenges for our task

- How to recommend relevant news?
- How to only recommend true news when the veracity of candidate news is unknown?
- How to model the transition over latent events while avoiding the interference from news veracity related information (e.g., news content style)?

Existing news recommender systems

- Only focus on the relevance between news, namely users' personalized preference
- Oblivious to the veracity of news
- Event information is less studied

- So they cannot be used for fake news mitigation

Problem formulation: The input data

- A user-news interaction dataset:
  - A collection of user-news interaction (e.g., click) sequences D:

$$\mathcal{D} = \{\mathcal{S}_1, \cdots, \mathcal{S}_u, \cdots, \mathcal{S}_{|\mathcal{U}|}\}$$

  - Each sequence S consists of t pieces of news which were interacted by a given user u:

$$\mathcal{S}_u = \{v_1, \cdots, v_t\} (v \in \mathcal{V})$$

  - $U = \{u_1, u_2, \dots, u_{|U|}\}$ is a user set, $V = \{v_1, v_2, \dots, v_{|V|}\}$ is a news set.

- A news meta information table N records the title and the abstract of each piece of news.

Problem formulation: The problem

- For each user u, given the news set $C_u = \{v_1, \cdots, v_{t-1}\}$ interacted by u, we build a model M to first predict the veracity of each candidate news and then generate a list of true news R_u which interest the user to the most:

$$R_u = M(C_u)$$

User_2	**Context news (CN)**	CN_1: selena gomez brings a and a bikini to australia u but not justin bieber	CN_2:justin bieber selena gomez their time apart is driving him crazy	CN_3:justin bieber and selena gomez may have broken up for good this time	CN_4:justin bieber s ex baskin champion wows in a bikini amid his engagement to hailey baldwin	?

Green color indicates true news.
Red color indicates fake news.

A running example of news recommendation for fake news mitigation

Research questions

- How to recommend relevant news?
- How to only recommend true news when the veracity of candidate news is unknown?
- How to model the transition over latent events while avoiding the interference from news veracity related information (e.g., news content style)?

The Rec4Mit model

- Rec4Mit model contains three main modules:
  - (1) Event-veracity disentangle module, to divide the event information and veracity information of each news into two separate latent vectors: event embedding and veracity embedding,
  - (2) Event detection and transition module, to detect the possible events associated with each news, and the sequential transitions over events by taking the event embedding as the input,
  - (3) Next-news prediction module, predict the next news according to both the event information and veracity information.

Rec4Mit: The architecture

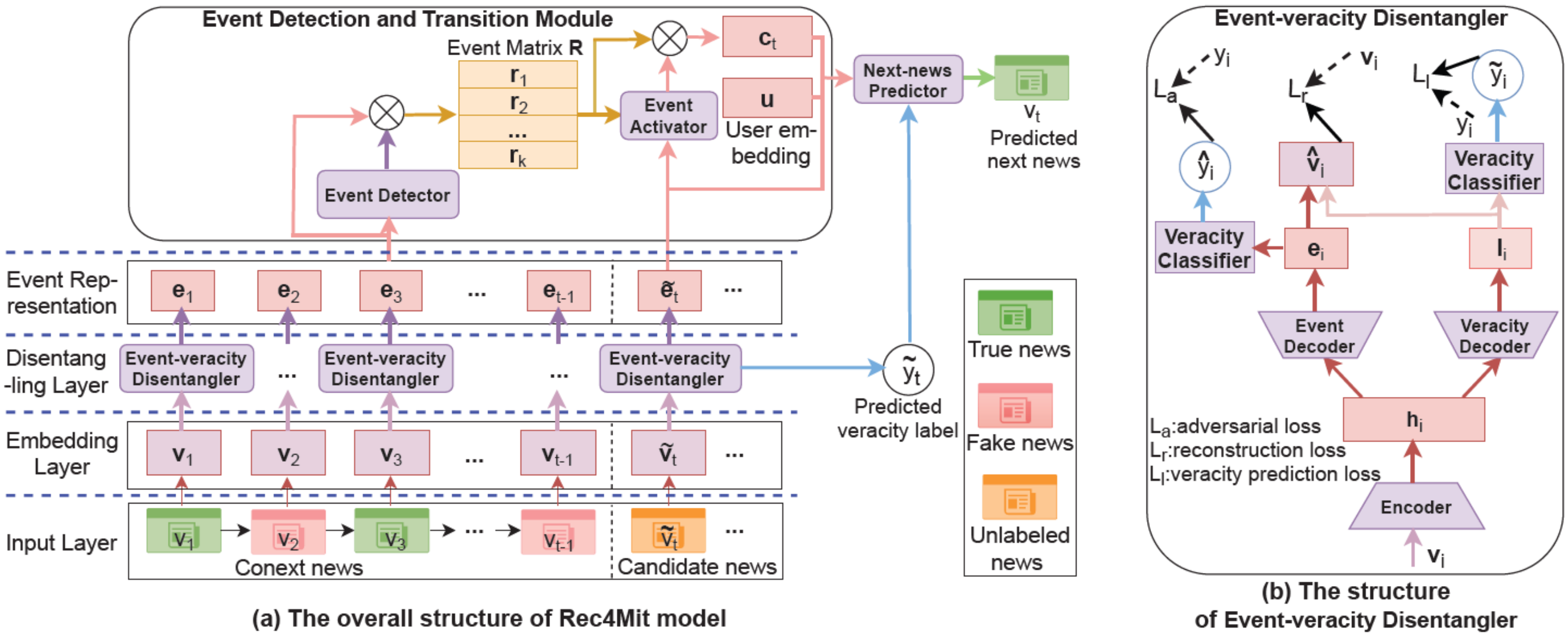


Figure 1: (a) Rec4Mit is built on three main components: Event-veracity Disentangler, Event Detection and Transition Module, and Next-news Predictor; (b) The Event-veracity Disentangler is built on the Encoder, Event Decoder, Veracity Decoder and Veracity Classifier.

Experiments: datasets

The FakeNewsNet data was used for the experiments, which contains two datasets: PlolitiFact and GossipCop.

Table 1: The characteristics of experimental datasets

Statistics	PolitiFact	GossipCop
#Users	37,873	22,540
#True news	306	6,792
#Fake news	310	2,737
#User-news interactions	150,350	646,154
#Training instance	38,062	108,802
#Test instance	4,701	13,601
#Validation instance	4,701	13,601

Gossip Cop

Type of site	Celebrity News Pop Culture
Owner	Gossip Cop Media
Created by	Michael Lewittes (Co-Founder)
URL	http://www.gossipcop.com/
Launched	July 29, 2009
Current status	Defunct

Shu, Kai, et al. "FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media." *Big data* 8.3 (2020): 171-188.

Baseline methods

We select nine representative and/or state-of-the-art approaches as baseline approaches. They are based on various models:

- ✓ KNN based approaches: SKNN
- ✓ Memory network based approaches: CSRM
- ✓ RNN based approaches: CSRM, LSTUR
- ✓ CNN based approaches: FIM, FedNewsRec
- ✓ GNN based approaches: SR-GNN
- ✓ Attention model based approaches: SASRec, DAN, NRMS

Evaluation metrics

We evaluate the recommendation results from two perspectives:

- Relevance: Prediction accuracy: weather the recommended news can well match the users' reading preference:
  - Recall
  - mean reciprocal rank (MRR)
  - normalized discounted cumulative gain (NDCG)

- Veracity: the ratio of true news (RT) in the recommendation list: weather only true news was recommended to users:

$$RT@K = \frac{\#True\ news}{K} * 100\%.$$

Experimental result 1: comparison with baselines

Relevance: our proposed Rec4Mit model can achieve the highest recommendation accuracy.

Table 2: Comparison of prediction accuracy with baselines on two datasets, *the improvement is significant at $p < 0.05$.

	PolitiFact						GossipCop					
	REC@5	REC@20	MRR@5	MRR@20	NDCG@5	NDCG@20	REC@5	REC@20	MRR@5	MRR@20	NDCG@5	NDCG@20
SKNN	0.2176	0.6088	0.1171	0.1553	0.1414	0.2524	0.1697	0.6252	0.0394	0.0911	0.0703	0.2074
CSRM	0.3752	0.6773	0.2629	0.2923	0.2906	0.3763	0.4764	0.6387	0.3496	0.3661	0.3813	0.4281
SR-GNN	0.3678	0.6741	0.2562	0.2865	0.2837	0.3711	0.4920	0.6239	0.3933	0.4067	0.4180	0.4560
SASRec	0.2962	0.6608	0.1582	0.1933	0.1924	0.2954	0.2419	0.4655	0.1009	0.1244	0.1358	0.2010
DAN	0.1874	0.7405	0.0784	0.1338	0.1049	0.2637	0.3174	0.4541	0.3157	0.3257	0.3161	0.3512
NRMS	0.4752	0.8260	0.3103	0.3449	0.3511	0.4511	0.6354	0.8239	0.4505	0.4702	0.4966	0.5516
LSTUR	0.4827	0.8111	0.3166	0.3491	0.3577	0.4515	0.6950	0.8817	0.4955	0.5156	0.5454	0.6005
FedNewsRec	0.3584	0.7949	0.1940	0.2377	0.2344	0.3596	0.2267	0.4892	0.1248	0.1498	0.1499	0.2237
FIM	0.3711	0.7042	0.1930	0.2250	0.2371	0.3311	0.3521	0.5911	0.2340	0.2570	0.2631	0.3312
Rec4Mit	**0.5561***	**0.8868***	**0.3462***	**0.3808***	**0.3979***	**0.4944***	**0.7552***	**0.9543***	**0.4984***	**0.5205***	**0.5625***	**0.6220***
Improvement[1]	15.21%	7.36%	9.35%	9.08%	11.24%	9.50%	8.66%	8.23%	0.59%	0.95%	3.14%	3.58%

[1] The improvement over the best-performing baseline methods whose performance is underlined.

Experimental result 1: comparison with baselines

Veracity: our proposed Rec4Mit model can achieve the highest ratio of true news in the recommendation list.

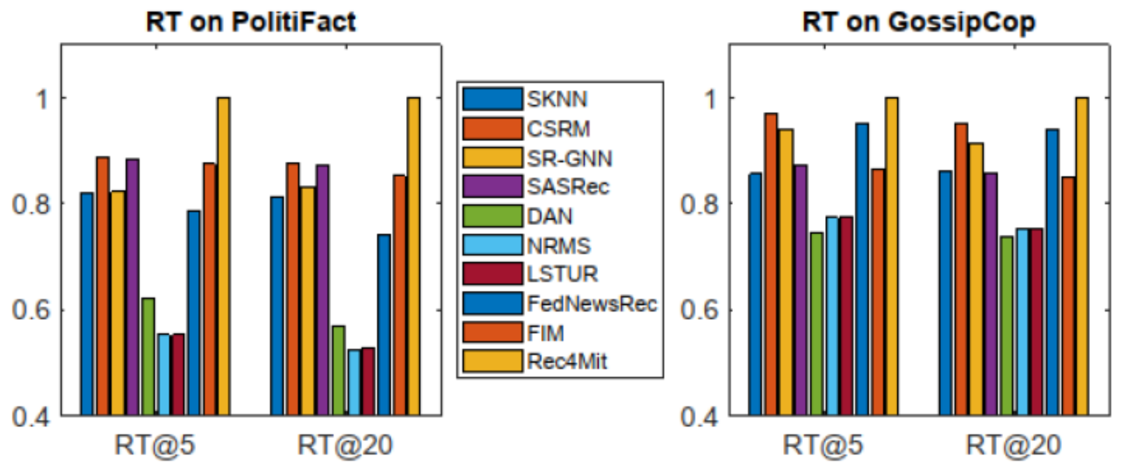


Figure 2: The ratio of true news (RT) in recommendation lists.

Experimental result 2: ablation analysis

Three simplified versions of Rec4Mit are designed to test th
and effectiveness of different core modules in Rec4Mit:

- Rec4Mit-Disen: which removes the event-veracity disentangle module
- Rec4Mit-Event: which remove the event detector inside the disentangle module
- Rec4Mit-Label: which removes the veracity classifier inside the disentangle module

Table 3: Comparison of Rec4Mit with its Variants on two real-world datasets, *the improvement is significant at $p < 0.05$.

	PolitiFact						GossipCop					
	REC@5	REC@20	MRR@5	MRR@20	NDCG@5	NDCG@20	REC@5	REC@20	MRR@5	MRR@20	NDCG@5	NDCG@20
Rec4Mit	**0.5561***	**0.8868***	**0.3462***	**0.3808***	**0.3979***	**0.4944***	**0.7552***	**0.9543***	**0.4984***	**0.5205***	**0.5625***	**0.6220***
Rec4Mit-Disen	0.5527	0.8721	0.3377	0.3702	0.3816	0.4919	0.6932	0.9249	0.4418	0.4678	0.5053	0.5737
Rec4Mit-Event	0.5375	0.8783	0.3314	0.3687	0.3823	0.4855	0.6991	0.9349	0.4523	0.4778	0.5138	0.5838
Rec4Mit-Label	0.5502	0.8837	0.3427	0.3786	0.3938	0.4920	0.6673	0.8546	0.4288	0.4495	0.4883	0.5442

A case study: recommendation results

Table 4: Recommendation Lists for 5 Users Sampled from GossipCop Dataset.

User$_1$	Context news (CN)	CN$_1$:[1]jennifer lawrence says u mother u led to darren split	CN$_2$:[2]where is travis scott why kylie jenner s boyfriend avoids the spotlight	CN$_3$:jennifer lawrence says u mother u led to darren split	CN$_4$:chris pratt[3]files for divorce from anna faris	
	Recommended news (RN)	RN$_1$:tori kelly is engaged to basketball player boyfriend u e	RN$_2$:steven innovative co creator of u nypd blue u u hill street blues u dies at	RN$_3$**(ground truth)**:[4] chris pratt and anna faris finalize divorce one year after separating reports	RN$_4$:rita ora kisses cardi b in the new video for controversial track u girls u	RN$_5$:harvey weinstein timeline how the scandal unfolded
User$_2$	Context news (CN)	CN$_1$: selena gomez brings a and a bikini to australia u but not justin bieber	CN$_2$:justin bieber selena gomez their time apart is driving him crazy	CN$_3$:justin bieber and selena gomez may have broken up for good this time	CN$_4$:justin bieber s ex baskin champion wows in a bikini amid his engagement to hailey baldwin	
	Recommended news (RN)	RN$_1$**(ground truth)**:selena gomez u s mom responds to justin bieber relationship rumors	RN$_2$:taylor swift s stalker sentenced to year probation and gps monitoring	RN$_3$:celebrities with tattooed eyebrows including helen mirren rooney michelle	RN$_4$:prince harry and harry styles reunite	RN$_5$:kristen bell hosts sag awards in series of gowns see the stunning looks
User$_3$	Context news (CN)	CN$_1$: brad pitt he had a blast playing with kids during secret cambodian family reunion	CN$_2$:kim kardashian responds to claims she was attacked in los angeles such weird rumors	CN$_3$:pepsi pulls controversial kendall jenner ad after outcry	CN$_4$:girls cast spoofs golden girls on jimmy kimmel live	
	Recommended news (RN)	RN$_1$**(ground truth)**:brad pitt u s red carpet surprise at u lost city of z u premiere	RN$_2$:the fast food guide	RN$_3$:kesha s mother drops against dr luke	RN$_4$:jason aldean and wife brittany kerr revealed the gender of their baby in the cutest way	RN$_5$:video justin timberlake announces opening act for man of the woods tour u z
User$_4$	Context news (CN)	CN$_1$:selena gomez demi lovato bond over boys possible duet more during epic reunion	CN$_2$:real reason behind justin bieber and selena gomez u s breakup has finally been revealed	CN$_3$:poor joe jonas is trying desperately to look like ex gigi hadid s new boyfriend zayn malik	CN$_4$:katie holmes pushing jamie foxx to go more public with their relationship u why he u s u hesitant u	
	Recommended news (RN)	RN$_1$:first look at ryan murphy s new fox series	RN$_2$:video justin timberlake announces opening act for man of the woods tour u z	RN$_3$:jennifer aniston	RN$_4$:best royal wedding gowns of all time	RN$_5$**(ground truth)**: justin s wife, his character may have relationship issues

[1] Green color indicates true news.
[2] Red color indicates fake news.
[3] ____ is used to highlight the identical/related events/topics from the context news and the recommended news of each user, e.g., "chris prat divorce" appears both in the fourth context news and the third recommended news of User$_1$.
[4] "**ground truth**" means the corresponding recommended news has been really read by the user in the test set.

Personalised mitigation: summary

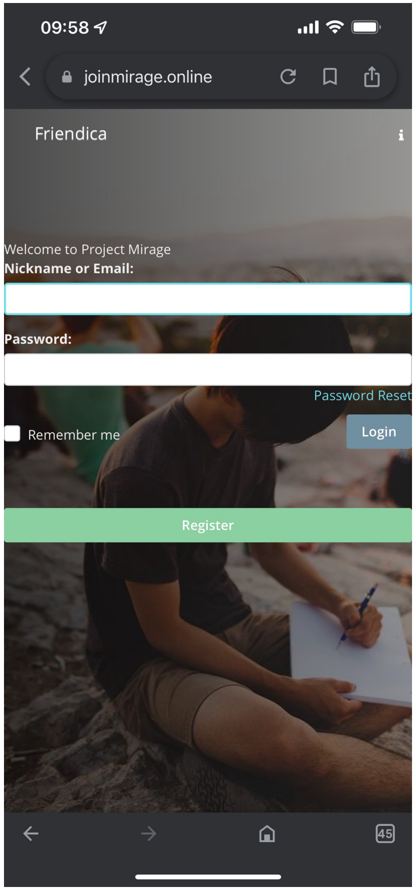
- Aim for both personalized and corrective true news for fake news mitigation.
- Veracity-aware personalised news recommendation is effective for personalised mitigation.
- Modelling the dependencies and transitions between events in news sequence and classifying the veracity of candidate news is crucial for effective personalised mitigation.

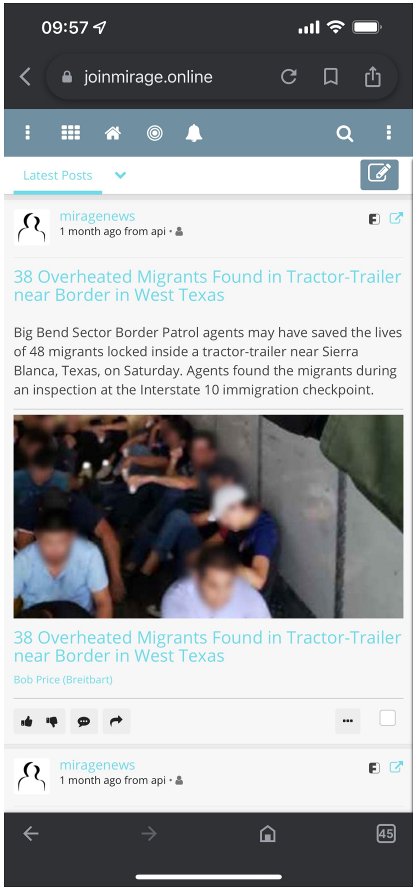
What's next?

Online evaluation of mitigation strategy for changing user information behaviour.

An ad hoc social network for misinformation research

Call for volunteers for Mirage: https://joinmirage.online/





Xu X., Deng, K., and Zhang, X. 2022. Mirage: An ad-hoc social network for research on social media misinformation mitigation.

Acknowledgements

Australian Government

Australian Research Council

References: social media information diffusion models

- H.R. Nie and X. Zhang et al. 2020. Modelling user influence and rumour propagation on social networks. DSAA’2020.
- Pham, C.V., Phu, Q.V., Hoang, H.X., Pei, J. and Thai, M.T., 2019. Minimum budget for misinformation blocking in online social networks. Journal of Combinatorial Optimization, 38(4), pp.1101-1127.
- Saxena, A., Hsu, W., Lee, M.L., Leong Chieu, H., Ng, L. and Teow, L.N., 2020, April. Mitigating misinformation in online social network with top-k debunkers and evolving user opinions. In Companion proceedings of the web conference 2020 (pp. 363-370).
- Tan, Z., Wu, D., Gao, T., You, I. and Sharma, V., 2019. AIM: Activation increment minimization strategy for preventing bad information diffusion in OSNs. Future Generation Computer Systems, 94, pp.293-301.
- Wen, S., Haghighi, M.S., Chen, C., Xiang, Y., Zhou, W. and Jia, W., 2014. A sword with two edges: Propagation studies on both positive and negative information in online social networks. IEEE Transactions on Computers, 64(3), pp.640-653.
- Zhang, H., Mishra, S., & Thai, M.T. 2014. Recent Advances in Information Diffusion and Influence Maximization of Complex Social Networks.

References: heuristic-based debunker selection

- Pham, C.V., Phu, Q.V., Hoang, H.X., Pei, J. and Thai, M.T., 2019. Minimum budget for misinformation blocking in online social networks. Journal of Combinatorial Optimization, 38(4), pp.1101-1127.
- Saxena, A., Hsu, W., Lee, M.L., Leong Chieu, H., Ng, L. and Teow, L.N., 2020, April. Mitigating misinformation in online social network with top-k debunkers and evolving user opinions. In Companion proceedings of the web conference 2020 (pp. 363-370).
- Saxena, A., Saxena, H. and Gera, R., 2020, December. k-truthscore: Fake news mitigation in the presence of strong user bias. In International Conference on Computational Data and Social Networks (pp. 113-126). Springer, Cham.
- Wen, S., Jiang, J., Xiang, Y., Yu, S., Zhou, W. and Jia, W., 2014. To shut them up or to clarify: Restraining the spread of rumors in online social networks. IEEE Transactions on Parallel and Distributed Systems, 25(12), pp.3306-3316.

References: learning-based network-level mitigation

- Farajtabar, M., Ye, X., Harati, S., Song, L. and Zha, H., 2016. Multistage campaigning in social networks. Advances in Neural Information Processing Systems, 29.
- Farajtabar, M., Yang, J., Ye, X., Xu, H., Trivedi, R., Khalil, E., Li, S., Song, L. and Zha, H., 2017, July. Fake news mitigation via point process based intervention. In International conference on machine learning (pp. 1097-1106). PMLR.
- Goindani, M. and Neville, J., 2020, May. Cluster-Based Social Reinforcement Learning. In Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (pp. 1852-1854).
- Goindani, M. and Neville, J., 2020, August. Social reinforcement learning to combat fake news spread. In Uncertainty in Artificial Intelligence (pp. 1006-1016). PMLR.
- Shu, K., Bernard, H.R. and Liu, H., 2019. Studying fake news via network analysis: detection and mitigation. In Emerging research challenges and opportunities in computational social network analysis and mining (pp. 43-65). Springer, Cham.
- Xu, X., Deng, K. and Zhang, X., 2022, February. Identifying cost-effective debunkers for multi-stage fake news mitigation campaigns. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining* (pp. 1206-1214).

Personalised mitigation by recommendation

- Wang, S., Xu, X., Zhang, X., Wang, Y. and Song, W., 2022, April. Veracity-aware and Event-driven Personalized News Recommendation for Fake News Mitigation. In *Proceedings of the Web Conference.*

Contact: xiuzhen.zhang@rmit.edu.au

http://www.xiuzhenzhang.org/