THE UNIVERSITY OF
MELBOURNE

Faculty of
Engineering and
Information Technology

News dispersal patterns across social media platforms

Presented by Jiachen Li, supervised by Prof. Richard Sinnott

Why News from Social Media?

- **People consume News from social media**
- **Journalists relying on social media to source (54.8%) and verify (48.8%) news [1]**
- **However, social media has posed a misinformation challenge**

% of U.S. adults who get news on a social networking site ...

62% get news on social media

Often 18%
Never 38%
Sometimes 26%
Hardly ever 18%
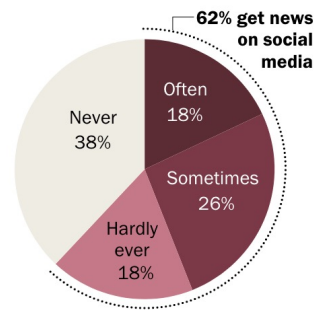
Source: Survey conducted Jan. 12-Feb. 8, 2016.
"News Use Across Social Media Platforms 2016"

PEW RESEARCH CENTER [2]

% of U.S. adults who get news from social media ...

Often	Sometimes	Rarely	Never	Don't get digital news
23%	30	18	21	7
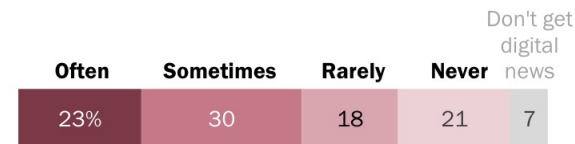
Note: This chart is not comparable to similar questions asked in the past due to question wording changes; see Appendix for more details.
Source: Survey of U.S. adults conducted Aug. 31-Sept. 7, 2020.
"News Use Across Social Media Platforms in 2020"

PEW RESEARCH CENTER [3]

Why not Rumour Detection？

- **Can achieve 0.96 accuracies [4]**
- **BUT has limited practicality**
  - **Only applicable to newsworthy posts**
  - **Other types of harmful information**
  - **Uninformative Binary Classification**

Pfizer purposely directed the evolution of covid-19 on monkeys in the laboratory environment to test the effectiveness of their product against potential mutation of covid-19 [5][6].

Pfizer was making money from the Pandemic by selling vaccine, and they have purposely directed the evolution of covid-19!

Related Works

- **Links-based epidemiological model by Adar and Adamic [7]**

- **Conserved content exact match model by Yang and Leskovec [8]**

- **MemeTracker by Leskovec et al. [9]**

- **NIFTY system by Suen et al. [10]**

- **Sentence-level grammatical relations model by Vakulenko et al. [11]**

- **Semantic similarity clustering model by Joshi and Sinnott [12]**
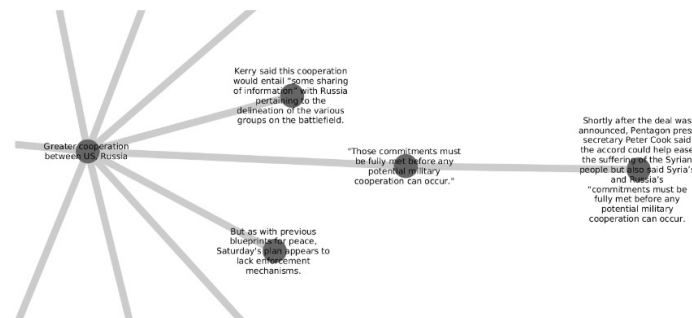
Method

- **Capture semantics with BERT [13]**
- **Specialised in social media post**
- **Multiply social media platforms**
  - **Twitter ?**
  - **Reddit**
  - **Mastodon**
- **Deploy with a cloud-based solution**

Why Cloud?

- **Performance**
- **Availability**
- **Accessibility**
- **Scalability**
- **Robustness**

Resource	Quanlity
Compute Resources - Virtual Cores	32 VCPUs
Compute Resources - Instances	6 servers
Compute Resources - RAM	Unlimited
Volume Storage	2000 GiB
Advanced Networking - Networks	3 Networks
Advanced Networking - Routers	3 Routers
Advanced Networking - Floating Ips	2 Floating IPs
Advanced Networking - Load Balancers	2 Load Balancers

System Design

- **Deploy on Melbourne Research Cloud**
- **Software as a Service**
- **ReSTful API**
- **Master-worker architecture**
- **Worker pool design**
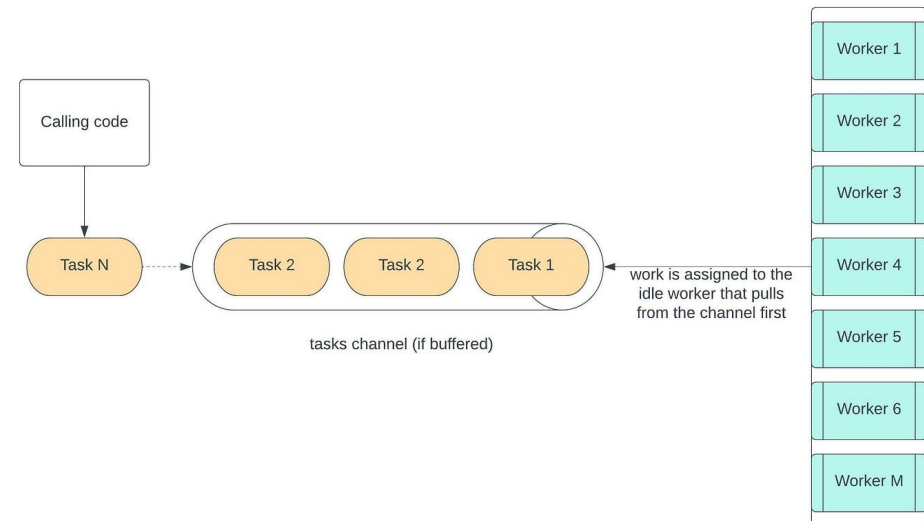
Future Improvement

- **Short-video social media platforms**
- **TikTok has 150 million monthly active users in the US**
- **Speech-to-text + image-to-text**
- **However, developer APIs are US only now**

Research Timeline

Task	Semester 1, 2023												Winter Break		Semester 2, 2023											
	1	2	3	4	5	6	7	8	9	10	11	12			1	2	3	4	5	6	7	8	9	10	11	12
Phase 1: Secure computation reserouce and data source																										
Apply for developer API token from Twitter, Reddit and Mastodon		█																								
Apply for resource allocation from Melbourne Research Cloud	█																									
Apply for access token from Australian Digital Observatory		█																								
Phase 2: Preliminary knowledge Learning																										
Literture reading	█	█	█	█	█	█	█	█																		
Industry knowledge conversation		█	█	█	█	█	█																			
Phase 3: Cloud System Implementation																										
Architectural design							█	█	█																	
Database system deployment									█	█	█	█														
Gathering social media data from searching API													█	█	█	█	█	█	█	█						
Benchmark system with NLP algorithms and optimise correspingly															█	█	█	█	█	█	█	█	█	█	█	
Phase 4: NLP algorithm Implementation																										
Implement clustering algorithm										█	█															
Implement information extraction algorithm											█	█														
Implement entity alignment algorithm																										
Parallelise the algorithm													█	█	█	█										
Phase 5: Knowledge output																										
Research Proposal Writing						█	█	█																		
Thesis writing																	█	█	█	█	█	█	█	█	█	█

THE UNIVERSITY OF
MELBOURNE

Faculty of
Engineering and
Information Technology

Thank you for listening!

Special thanks to:
Luca Morandini
Qi Li
Jianzhong Qi

Reference

[1] X. Zhang and W. Li, "From Social Media with News: Journalists' Social Media Use for Sourcing and Verification," *Journal. Pract.*, vol. 14, no. 10, pp. 1193–1210, Nov. 2020, doi: 10.1080/17512786.2019.1689372.

[2] E. Shearer and A. Mitchell, "News Use Across Social Media Platforms in 2020," Pew Research Center, Report, Jan. 2021. Accessed: Apr. 22, 2023. [Online]. Available: https://www.pewresearch.org/journalism/wp-content/uploads/sites/8/2021/01/PJ_2021.01.12_News-and-Social-Media_FINAL.pdf

[3] J. Gottfried and E. Shearer, "News use across social media platforms 2016," Pew Research Center, Report, May 2016. Accessed: Apr. 22, 2023. [Online]. Available: https://www.pewresearch.org/journalism/wp-content/uploads/sites/8/2016/05/PJ_2016.05.26_social-media-and-news_FINAL-1.pdf

[4] Y. Fan, "Rumour Detection and Analysis on Twitter." arXiv, Apr. 04, 2023. Accessed: May 16, 2023. [Online]. Available: http://arxiv.org/abs/2304.01712

[5] *#DirectedEvolution TAKES THE WORLD BY STORM!!*, (Jan. 28, 2023). Accessed: Apr. 22, 2023. [Online Video]. Available: https://www.youtube.com/watch?v=BLE_3JElSNs

[6] "Pfizer Responds to Research Claims | Pfizer." https://www.pfizer.com/news/announcements/pfizer-responds-research-claims (accessed Apr. 23, 2023).

[7] E. Adar and L. A. Adamic, "Tracking Information Epidemics in Blogspace," in *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, Compiegne, France: IEEE, 2005, pp. 207–214. doi: 10.1109/WI.2005.151.

[8] J. Yang and J. Leskovec, "Modeling Information Diffusion in Implicit Networks," in *2010 IEEE International Conference on Data Mining*, Dec. 2010, pp. 599–608. doi: 10.1109/ICDM.2010.22.

[9] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, in KDD '09. New York, NY, USA: Association for Computing Machinery, 2009, pp. 497–506. doi: 10.1145/1557019.1557077.

[10] C. Suen, S. Huang, C. Eksombatchai, R. Sosic, and J. Leskovec, "NIFTY: a system for large scale information flow tracking and clustering," in *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro Brazil: ACM, May 2013, pp. 1237–1248. doi: 10.1145/2488388.2488496.

[11] S. Vakulenko, M. Gobel, A. Scharl, and L. Nixon, "Visualising the Propagation of News on the Web".

[12] A. Joshi and R. O. Sinnott, "Modelling Implicit Content Networks to Track Information Propagation Across Media Sources to Analyze News Events," in *2018 IEEE 14th International Conference on e-Science (e-Science)*, Oct. 2018, pp. 475–485. doi: 10.1109/eScience.2018.00136.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, May 24, 2019. doi: 10.48550/arXiv.1810.04805.