The Mechanisms of Misinformation

Ethan Grinberg, University of Illinois at Urbana-Champaign, ethanbg2@illinois.edu
Dr. Jacob T. Fisher, University of Illinois at Urbana-Champaign, jtfisher@illinois.edu

Data Availability

All code and data can be found at: https://github.com/ethanbg2/misinformation-cluster-analysis

Method

1. The list of misleading sources comes from the *iffy+* list. A list of sources deemed misleading by various fact-checking organizations. This list is used by the Hoaxy API misinformation tracking tool used in this project. [2–4]

2-3. For each misleading source indexed by the Hoaxy API, up to 100 articles per source were retrieved by the Hoaxy API. For each of these articles the Twitter Diffusion network was retrieved by the Hoaxy API.[2–4]

4. Before embedding the diffusion networks into vectors, networks with a number of tweets below the 15% quantile and above the 85% quantile were filtered out. This resulted in 569 diffusion networks being considered for analysis. These diffusion networks were converted to 96-dimensional vectors using Unsupervised Inductive Graph-Level Representation Learning via Graph-Graph Proximity.[5] This method requires a metric for graph proximity, in which graph edit distance was used. This was implemented using Python libraries Stellar Graph and Networx.[6,7]

5. A primary concern with K-means clustering is finding the correct number of clusters before clustering. To do this the "elbow" method heuristic was used and automatically decided by the Python Kneed library.[8] K-means was implemented using the Python Scikit-Learn library.[9]

References

1. Pierri F, Piccardi C, Ceri S. Topology comparison of Twitter diffusion networks effectively reveals misleading information. *Sci Rep*. 2020;10(1):1372. doi:10.1038/s41598-020-58166-5
2. Shao C, Ciampaglia GL, Flammini A, Menczer F. Hoaxy: A Platform for Tracking Online Misinformation. In: *Proceedings of the 25th International Conference Companion on World Wide Web*. WWW '16 Companion. International World Wide Web Conferences Steering Committee; 2016:745-750. doi:10.1145/2872518.2890098
3. Shao C, Ciampaglia GL, Varol O, Yang KC, Flammini A, Menczer F. The spread of low-credibility content by social bots. *Nat Commun*. 2018;9(1):4787. doi:10.1038/s41467-018-06930-7
4. Shao C, Hui PM, Wang L, et al. Anatomy of an online misinformation network. *PLOS ONE*. 2018;13(4):e0196087. doi:10.1371/journal.pone.0196087
5. Bai Y, Ding H, Qiao Y, et al. Unsupervised Inductive Graph-Level Representation Learning via Graph-Graph Proximity. *ArXiv190401098 Cs Stat*. Published online June 2, 2019. Accessed March 29, 2022. http://arxiv.org/abs/1904.01098
6. Data61 C. StellarGraph machine learning library. Published online 2018. https://github.com/stellargraph/stellargraph
7. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux G, Vaught T, Millman J, eds. *Proceedings of the 7th Python in Science Conference*. ; 2008:11-15.
8. Satopaa V, Albrecht J, Irwin D, Raghavan B. Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. In: *2011 31st International Conference on Distributed Computing Systems Workshops*. IEEE; 2011:166-171. doi:10.1109/ICDCSW.2011.20
9. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.