MEASURING THE INFLUENCE OF ONLINE MISINFORMATION: A HIERARCHY OF SOCIAL MEDIA DATA

Dimitra (Mimie) Liotsiou, Philip N. Howard
University of Oxford, Oxford, UK

{mimie.liotsiou, philip.howard}@oii.ox.ac.uk

Keywords: misinformation, social media, social influence, causality, Big Data

Online Misinformation: From Attention to Influence

- Threat to democracy: misinformation and propaganda campaigns on social media around elections and key political events, around the world.
- Researchers have attempted to measure the reach of, engagement with, and attention paid to such campaigns on social media [3, 4, 7].
- But, still a challenge: measuring the influence (causal effect) of online misinformation on actual election behavior – requires more data than social media companies typically make externally available [2, 5].
- Objective: To help address this challenge, map out what kinds of data are needed to progressively move from measuring attention towards measuring influence.

Hierarchy of Social Media Data

We propose a hierarchical typology of social media data (Figure 1). It shows:

- The usefulness of each type of data for measuring the influence of online misinformation on actual voting behavior, from Level 0 (low) to Level 4 (high).
- How accessible social media platforms make each type of data to external researchers. Stars proportional to quantity and quality of data. Internal data covers all levels to some extent [1], while public API data only partially covers Levels 0 and 2 but not Levels 1, 3 and 4.
- To isolate the influence (causal effect) of online misinformation on election behavior, one needs data not only on the misinformation content (Level 0) and engagement with it (Level 2), which is in part available through platforms' public APIs. Rather, one also needs data on the exposure (Level 1), on the election behavior of interest (Level 3), and generally also on other relevant causes of the election behavior (Level 4) [5, 6], which is typically not made externally available.

Special Data: The Russian IRA & the 2016 US Elections

Special data can help move up the pyramid.

- We analysed the Russian Internet Research Agency’s (IRA) social media campaign targeting US voters around the 2016 US presidential election [4].
- Special data: Public and non-public data from social media platforms (Facebook, Instagram, Twitter, YouTube). Posts and ads, with aggregate and anonymised engagement numbers.
- Analysis contributed important findings:
  - Very active campaign (millions of tweets, tens of thousands of Facebook and Instagram posts), and continued and often rose after the election into 2017.
  - High engagement, reach (e.g. more than 189 million Instagram likes and comments in 2013-2018).
  - Strategically focused on key divisive issues (e.g. immigration, race), often targeting users based on interests in these issues, demographics, location.
  - Spikes in IRA activity around key political events.

Special data contributed crucial findings on the IRA’s activities and reach across platforms over time (richer data from Levels 0-2). So, continuing and expanding external researchers’ access to such previously internal data would have immense benefits for deepening our understanding of online misinformation campaigns.

Conclusions

- The platforms’ provision of special data enabled crucial findings on the IRA’s activities and reach across platforms over time (richer data from Levels 0-2).
- But, still a challenge: measuring the influence (causal effect) of online misinformation on actual election behavior – requires more data than social media companies typically make externally available [2, 5].
- Objective: To help address this challenge, map out what kinds of data are needed to progressively move from measuring attention towards measuring influence.

Figure 1: Hierarchical typology of social media data, in terms of their usefulness for measuring the influence of online misinformation on election behavior (0: low, 4: high), along with their accessibility (public, special, internal data). The number of stars is proportional to the quality and quantity of the data.