Detecting misinformation in online social networks using cognitive psychology

KP Krishna Kumar and G Geethakumari

Computing and Information Sciences 2014

• ***Propaganda*** is defined as *information, especially of a biased or misleading nature*, used to promote a political cause or point of view.

• ***Misinformation*** is *false or inaccurate information*, especially that which is deliberately intended to deceive.

• ***Disinformation*** is *false information that is intended to mislead*, especially propaganda issued by a government organization to a rival power or the media.

The advent of Web 2.0 has resulted in faster dissemination of information using social media: social networks, blogs, emails, photo and video sharing platforms, bulletin boards etc.

The most important sources of misinformation are: media, governments and politicians, vested interests, **rumours**, and **works of fiction**.

*The **spread of misinformation*** is a result of a *cognitive process* by the receivers, based on an assessment of the truth value of information.
To make this assessment, people take into account four factors.

Information factors and 'analytic' factors

Consistency of message. Is the information compatible and consistent with the other things that you believe?

Coherency of message. Is the information internally coherent without contradictions to form a plausible story?

'Common knowledge' factors

Credibility of source. Is the information from a credible source?

General Acceptability. Do others believe this information?

Research in cognitive psychology has found that the ***acceptance of misinformation*** depends on *people's prior beliefs and opinions*: people accept information *without verification* if it conforms to their preexisting political, religious or social views.

> Poor truth discernment is linked to a ***lack of careful reasoning and relevant knowledge***, as well as to the use of *familiarity and source heuristics*.
>
> G Pennycook, DG Rand - Trends in cognitive sciences, 2021

Measuring credibility of tweets

Criteria	Metrics	Accuracy	Complexity
Consistency of message	Retweets, mentions	Retweets are better than mentions	No *Metrics are indirect measure*
Coherency of message	Questions, affirms, denial, no of words, pronouns, hashtags, URLs, exclamation marks, negative and positive sentiments, NLP techniques	Yes	***Computationally intensive***, requires ground truth and content analysis. *Metrics are indirect measure*
Credibility of Source	Tweets, retweets, mentions, indegree, user name, image, age, followers, followees	Retweets are better	No
General acceptability	Retweets	Good	No

Methodology

Consider only the retweets

Retweets are the easiest means by which tweets are propagated.
Further, this would also remove personal chats, opinions
and initial misinformation not considered credible.

Construct a retweet graph

We identified and segregated the retweets as per the source.

> This step would enable us to estimate
> the *tweets of the source which are being retweeted*
> and the *number of users who are retweeting the same*.
>
> We measure the *credibility of the source* using *Gini coefficient*.
> *Unevenness* amongst the users retweeting tweets of a source
> means the credibility of the source is poor.

Summary of the algorithm

- Identify the original sources of information (tweets) in the network.

- Rate the credibility of each source based on the acceptance of the tweets by the receivers, using a *retweet graph* .

- Segregate the possible sources of misinformation as non credible users and the corresponding tweets.

- Evaluate the general acceptance of tweets from credible users using *PageRank*.

The output of the previous steps is given to the user.

Based on his evaluation of the consistency and coherency of the message,
and the additional quantified inputs of the credibility of the source
and the general acceptability of the tweet, the user would be able
to make an informed decision on the authenticity of the tweet.

However:

We assume that for most of the news items spreading misinformation,
users are already suspicious of the news items.

There is a large disconnect between what people believe and what they will share on social media, and this is largely driven by ***inattention*** rather than by purposeful sharing of misinformation.

> G Pennycook, DG Rand - Trends in cognitive sciences, 2021

G. Pennycook, D. G. Rand The Psychology of Fake News

Trends in Cognitive Sciences, May 2021

Why Do People Fall for Fake News?

Truth 'discernment' is the extent to which misinformation is believed 'relative' to accurate content. It is typically calculated as belief in true news minus belief in false news. It captures the overall accuracy of one's beliefs.

Overall belief, or the extent to which news – regardless of its accuracy – is believed, is calculated as the average or sum of belief in true news and belief in false news. Factors that alter overall belief need not impact people's ability to tell truth from falsehood.

Reasoning

Across numerous recent studies, the evidence supports that people who are more *reflective* are less likely to believe false news content – and are better at discerning between truth and falsehood – regardless of whether the news is consistent or inconsistent with their partisanship.

Furthermore, experimentally manipulating participants' *level of deliberation* demonstrates a causal effect whereby deliberation reduces belief in false (but not true) news, regardless of partisan alignment (and has no effect on polarization).

It seems that people fail to discern truth from falsehood because *they do not stop to reflect sufficiently* on their prior knowledge (or *have insufficient or inaccurate prior knowledge*) – and not because their reasoning abilities are hijacked by political motivations.

Political Motivations

A popular narrative is that people engage in *'identity-protective cognition'* when faced with politically valenced content, and this leads them to be overly believing of content that is consistent with their partisan identity and overly skeptical of content that is inconsistent with their partisan identity.

People are somewhat better at discerning truth from falsehood when judging politically concordant news compared with politically discordant news.

Taken together, the evidence suggests that political identity and politically motivated reasoning are not the primary factors driving the inability to tell truth from falsehood in online news.

Heuristics

Feelings of familiarity likely contributes to increased belief in false claims.

Fake news is often geared toward provoking shock, fear, anger, or moral outrage. This is important because people who report experiencing more emotion (positive or negative) at the outset of the task are more likely to believe false (but not true) news.

Participants are more likely to believe information provided by people whom they view as being credible and a large literature from political science has robustly demonstrated the impact of elite messaging, in particular, on public opinion. Furthermore, social feedback provided by social media platforms (e.g., 'likes') also increases belief in news content, particularly for misinformation.

Believing versus Sharing Fake News

Participants who were asked about the accuracy of a set of headlines rated true headlines as much more accurate than false headlines.

Sharing intentions for false headlines were much higher than assessments of their truth, indicating that many people were apparently willing to share content that they could have identified as being inaccurate.

The confusion-based account posits that people genuinely (but mistakenly) believe that the false claims they share are probably true. Consistent with this proposal, of the false headlines that were shared, 33% were both believed and shared when participants were asked directly about accuracy.

The preference-based account is rooted in the idea that people place their preference for political identity (or related motives such as virtue signaling) above the truth, and thus share politically consistent false content on social media despite recognizing that it is probably not true. Of the false headlines that were shared, 16% of the headlines were shared despite being identified as inaccurate.

The inattention-based account argues that people have a strong preference to only share accurate content, but that the social media context distracts them from this preference. Consistent with this account, asking participants to rate the accuracy of each headline before deciding whether to share it decreased sharing of false headlines by 51% – suggesting that inattention to accuracy was responsible for roughly half of the misinformation sharing in the experiment.

Work on social media behavior often emphasizes the importance of the 'attention economy' where factors relating to engagement (likes, shares, comments, clicks,etc.) are selected for sharing of low-quality news content on Facebook.

What Can Be Done? Interventions To Fight Fake News

Fact-checking and inoculation approaches are fundamentally directed toward improving people's underlying knowledge or skills.

New Approaches for Fighting Misinformation

Recent research shows that a simple accuracy prompt – specifically, having participants rate the accuracy of a single politically neutral headline (ostensibly as part of a pretest) before making judgments about social media sharing – improves the extent to which people discern between true and false news content when deciding what to share online in survey experiments.

This approach has been successfully deployed in a large-scale field experiment on Twitter, in which messages asking users to rate the accuracy of a politically neutral news headline were sent to thousands of accounts who recently shared links to misinformation sites. This subtle prompt significantly increased the quality of the news they subsequently shared.