Characterizing and Predicting Social Correction on Twitter

Yingchen Ma
Georgia Institute of Technology
Atlanta, Georgia, USA
yma473@gatech.edu

Bing He
Georgia Institute of Technology
Atlanta, Georgia, USA
bhe46@gatech.edu

Nathan Subrahmanian
Brandeis University
Waltham, Massachusetts, USA
nsubrahmanian@brandeis.edu

Srijan Kumar
Georgia Institute of Technology
Atlanta, Georgia, USA
srijan@gatech.edu

ABSTRACT

Online misinformation has been a serious threat to public health and society. Social media users are known to reply to misinformation posts with counter-misinformation messages, which have been shown to be effective in curbing the spread of misinformation. This is called social correction. However, the characteristics of tweets that attract social correction versus those that do not remain unknown. To close the gap, we focus on answering the following two research questions: (1) "Given a tweet, will it be countered by other users?", and (2) "If yes, what will be the magnitude of countering it?". This exploration will help develop mechanisms to guide users' misinformation correction efforts and to measure disparity across users who get corrected. In this work, we first create a novel dataset with 690,047 pairs of misinformation tweets and counter-misinformation replies. Then, stratified analysis of tweet linguistic and engagement features as well as tweet posters' user attributes are conducted to illustrate the factors that are significant in determining whether a tweet will get countered. Finally, predictive classifiers are created to predict the likelihood of a misinformation tweet to get countered and the degree to which that tweet will be countered. The code and data is accessible on https://github.com/claws-lab/social-correction-twitter.

CCS CONCEPTS

• **Information systems** → Social networks.

KEYWORDS

Misinformation, Counter-misinformation, Social Correction, Twitter, COVID-19 vaccines

ACM Reference Format:
Yingchen Ma, Bing He, Nathan Subrahmanian, and Srijan Kumar. 2023. Characterizing and Predicting Social Correction on Twitter. In *15th ACM Web Science Conference 2023 (WebSci '23), April 30–May 01, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3578503.3583610

WebSci '23, April 30–May 01, 2023, Austin, TX, USA

ACM ISBN 979-8-4007-0089-7/23/04.
https://doi.org/10.1145/3578503.3583610

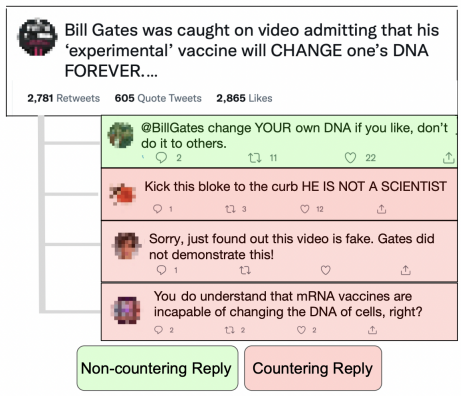

Figure 1: Examples of misinformation tweets and counter-misinformation replies.

1 INTRODUCTION

Online misinformation leads to societal harm including diminishing trust in vaccines and health policies [6, 49], damaging the well-being of users consuming misinformation [35, 63], encouraging violence and harassment [5, 60], and posing a danger to democratic processes and elections [57–59]. The problem has been exacerbated during the COVID-19 pandemic [40, 56]; particularly, COVID-19 vaccine misinformation including false claims that the vaccine causes infertility, contains microchips, and even changes one's DNA and genes has fueled vaccine hesitancy and reduced vaccine uptake [56]. Therefore, it is crucial to restrain the spread of online misinformation [36, 40]. In this work, we use a broad definition of misinformation which contains rumors, falsehoods, inaccuracies, decontextualized truths, or misleading leaps of logic [35, 68].

To combat misinformation, various countermeasures have been developed [40, 42, 65]. Recent work has shown that ordinary users of online platforms play a crucial role in countering misinformation. According to the research study by Micallef et al. [40], the vast majority (96%) of online counter-misinformation responses are made by ordinary users, with the remainder being made by professionals such as fact-checkers and journalists. While fact-checking from these professionals has been widely used due to its prominent and measurable impact [40, 65], this process typically does not involve engaging with the actors spreading misinformation. Instead, the ordinary users' counter-misinformation efforts complement those from professional fact-checkers by directly engaging in countering