

Thesis Proposal

Institute for Software Research
Societal Computing



Leveraging Stances in Conversations for the Assessment of Contentious events in Twitter

Ramon Villa-Cox

Monday, May 27th, 2021

3:00-6pm EDT

Via Zoom:

<https://cmu.zoom.us/j/7014965451?pwd=SXFTOFBsRWwreGtWc1N0VFpkZkVEQT09>

Meeting ID: 701 496 5451

Passcode: 688533

There is currently an ongoing policy discussion regarding the impact of the observed polarization online, how it affects the spread of false information, and what if anything should be done to curtail it. Options have been presented in academia and the public discourse ranging from those that focus on government and platform-based interventions to those seeking to educate and empower individuals in their interactions with social media. To design and implement effective and efficient interventions requires a more detailed understanding of how false information and the response to it spreads on digital social media. This understanding is complicated by the reality that there are different kinds of false information being shared from and through different online communities. It may also be affected by the diffusion of negative responses to the false information that occur prior to the news about the false information going more public. In this thesis I focus on the study of polarized Twitter communities, and the spread of disinformation through them, during contentious events.

Due to its large number of users, Twitter has been the primary social media platform for acquiring, sharing, and spreading information. However, it has also inevitably become a source for misinformation spread and polarization. The effect might not be crucial when the subject in hand is a trivial one, however, during globally concerning events, the spread of information through Twitter gains an undeniable importance. Most of the research on information diffusion in social media has focused on retweeting, despite it being only one potential reaction to information found on Twitter. Moreover, people often use other mechanisms such as quoting to resend messages. In doing so, they can change the context of the original tweet. Examining replies to tweets and the support of those replies also assists in creating a more complete picture of the discourse path. As shown in this work, it can be misleading, particularly when characterizing the spread of disinformation. We find that negative responses towards fake or contentious tweets tend to manifest through replies and quotes and that diffusion through this medium can dwarf what is observed through retweets.

For this reason, we first build a comprehensive labeled dataset of these contentious messages and their responses, collection that focused around the South American Protests that paralyzed the region in 2019, and which consequences are being felt still today. We seek to develop classifiers that leverage the interactions between users, and their stances in the event, to: improve early detection of rumorous tweets, and improve identification of polarized communities. These will be used to build validation datasets for a Twitter Agent-Based simulation in which we will carefully implement the different ways of interaction allowed in the platform and the way that content is broadcasted between users. By doing so, we can evaluate the effect of a social platform's infrastructure on the diffusion process of different types of stories and the relevant user interaction.

**Committee: Dr. Kathleen M. Carley (Chair), Dr. Hirokazu Shirado,
Dr. Alex Davis, Dr. Ignacio Arana, Dr. Enrique Pelaez (ESPOL)**