Which Fake Fits?

How the Diffusion of Disinformation Varies by Campaign and Community in the *Black Panther* Twitter Discussion

Carnegie Mellon

Matthew Babcock, PhD
mbabcock@andrew.cmu.edu

Sumeet Kumar (student)
sumeetku@andrew.cmu.edu

Ramon Villa-Cox (student)
rvillaco@andrew.cmu.edu

Summary

We conducted a case study to explore how different communities reacted to four types of false stories present in the *Black Panther* conversation. We found that 1) the Alt-Right community and campaign were isolated, whereas each of the three other communities reacted to each of the three other campaigns, 2) the total number of responses and response half-life to each campaign varied by community, and 3) the satire community was the most active in responding to the non-satire campaign.

Motivation

Recent research has focused on and been helpful in distinguishing between "true" and "false" information and comparing the characteristics of both on Twitter. Unfortunately, not all "false" information is the same and not all false information affects and is reacted to by different communities in the same way. Further investigation of the different types of disinformation and its diffusion through various types of communities is needed to help stakeholders make more informed decisions about the appropriateness, effectiveness, and efficiency of possible responses.

Black Panther and Four Types of False Stories

Black Panther opened to much social media fanfare and financial success in Feb 2018. It is the 3rd highest grossing film in the US and the most tweeted about movie ever. The film was hyped on social media due in part to its representation of African and African-American actors and creators. We identified a total of four types of false stories that were shared and reacted to on Twitter:

Type of False Story	Description
Fake Attacks (Non-Satire)	claimed race-based assaults at movie theaters used images that were debunked by community
Fake Attacks (Satire)	mocked the original fake attack posts used more unbelievable images from pop-culture
Fake Scene	claimed movie contained false sexual/racial scenes
Alt-Right	claimed the movie promoted alt-right philosophy

Methods

Community Selection
- Communities defined as the groups of users who exclusively retweeted a particular type of false story.
- A fifth community comprised of users that did not retweet any false story was also defined.
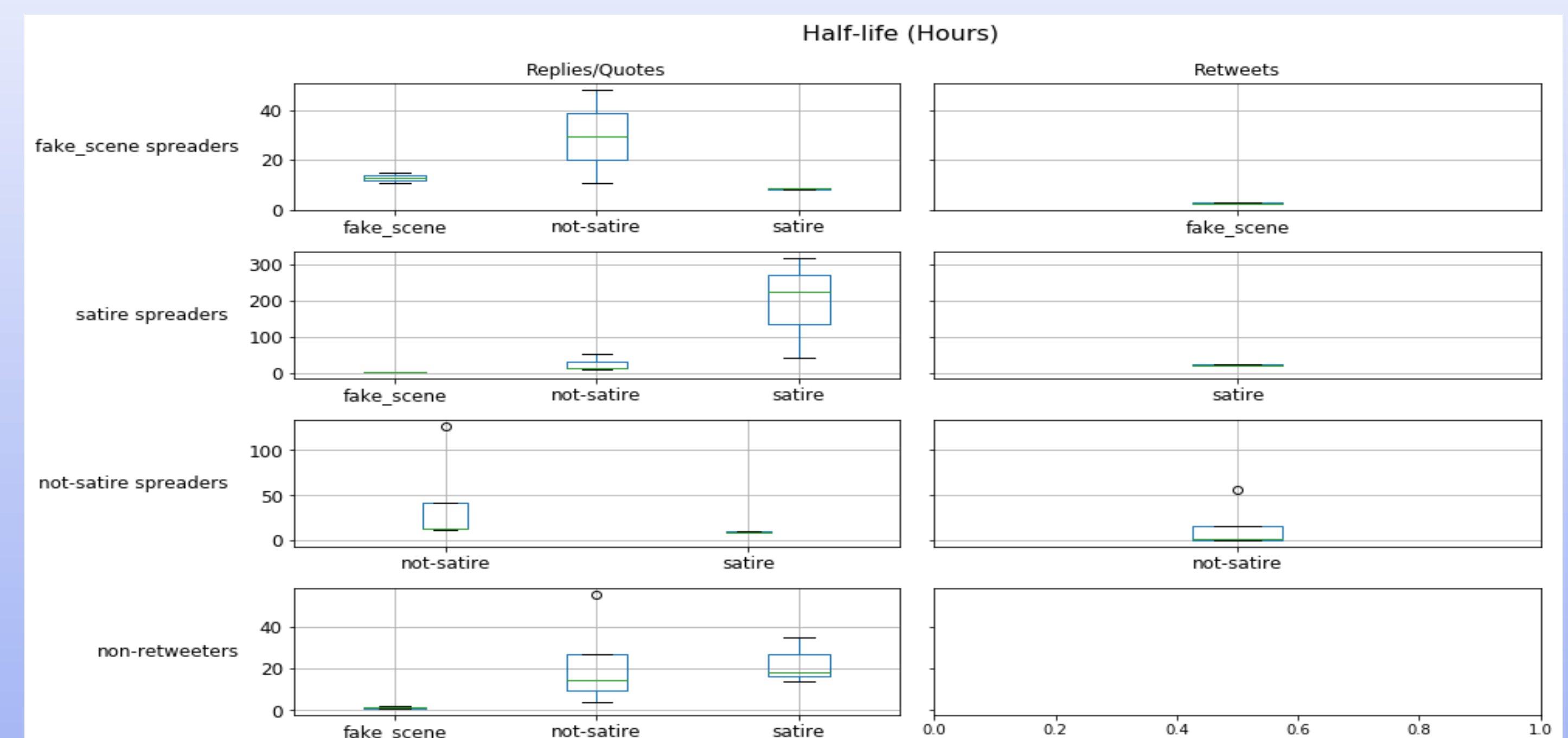
Types of Responses
- Retweets of original false tweet (assumed to be supportive)
- Replies of original false tweet (can be supportive or detracting)
- Quotes of original false tweet (can be supportive or detracting)
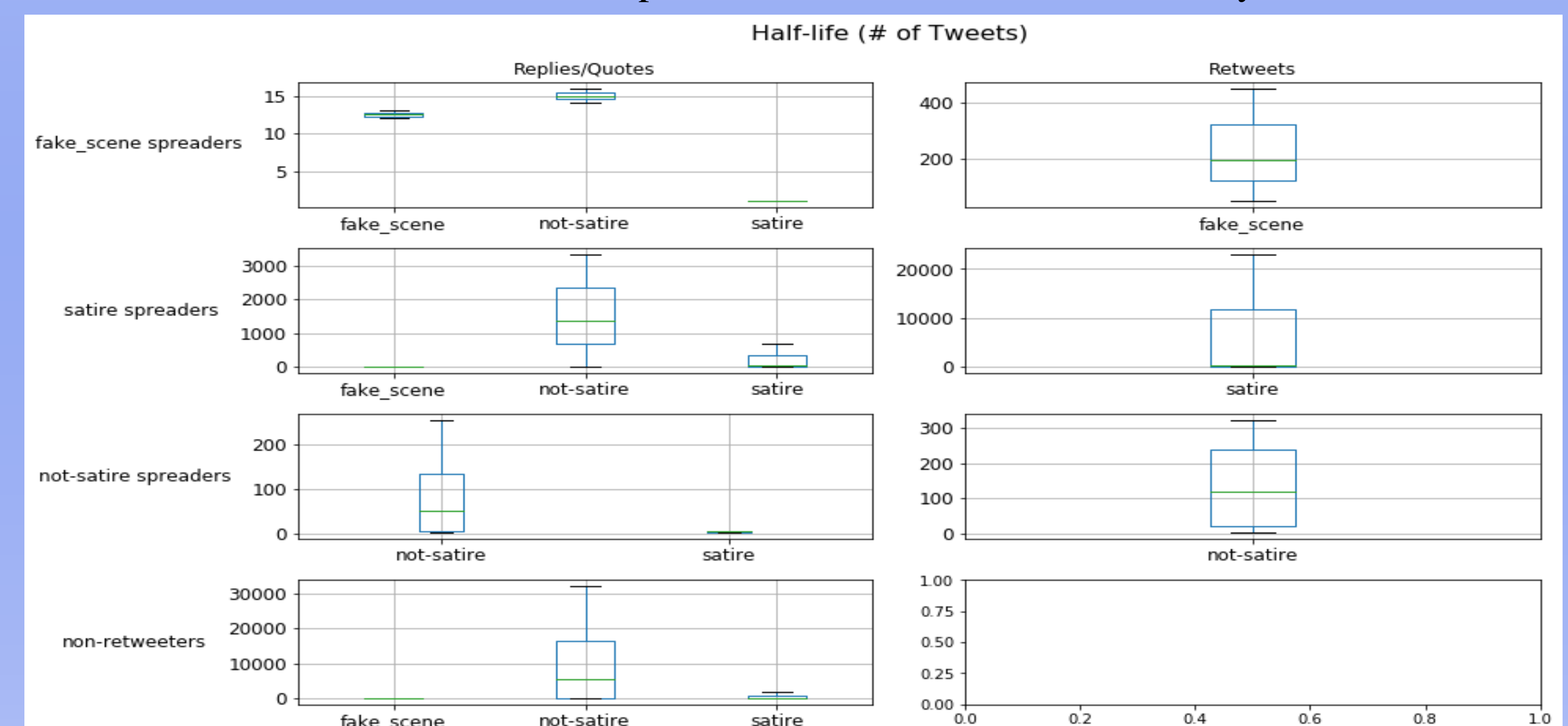- Retweets of reply or quote

Analysis
- Only false tweets that had 100+ responses were considered (total of 10).
- No Alt-Right false stories reached this threshold
- We calculated and analyzed the half-life (time in hours taken for the number of responses to reach half of the total responses) and the number of responses within each community to each type of false story.
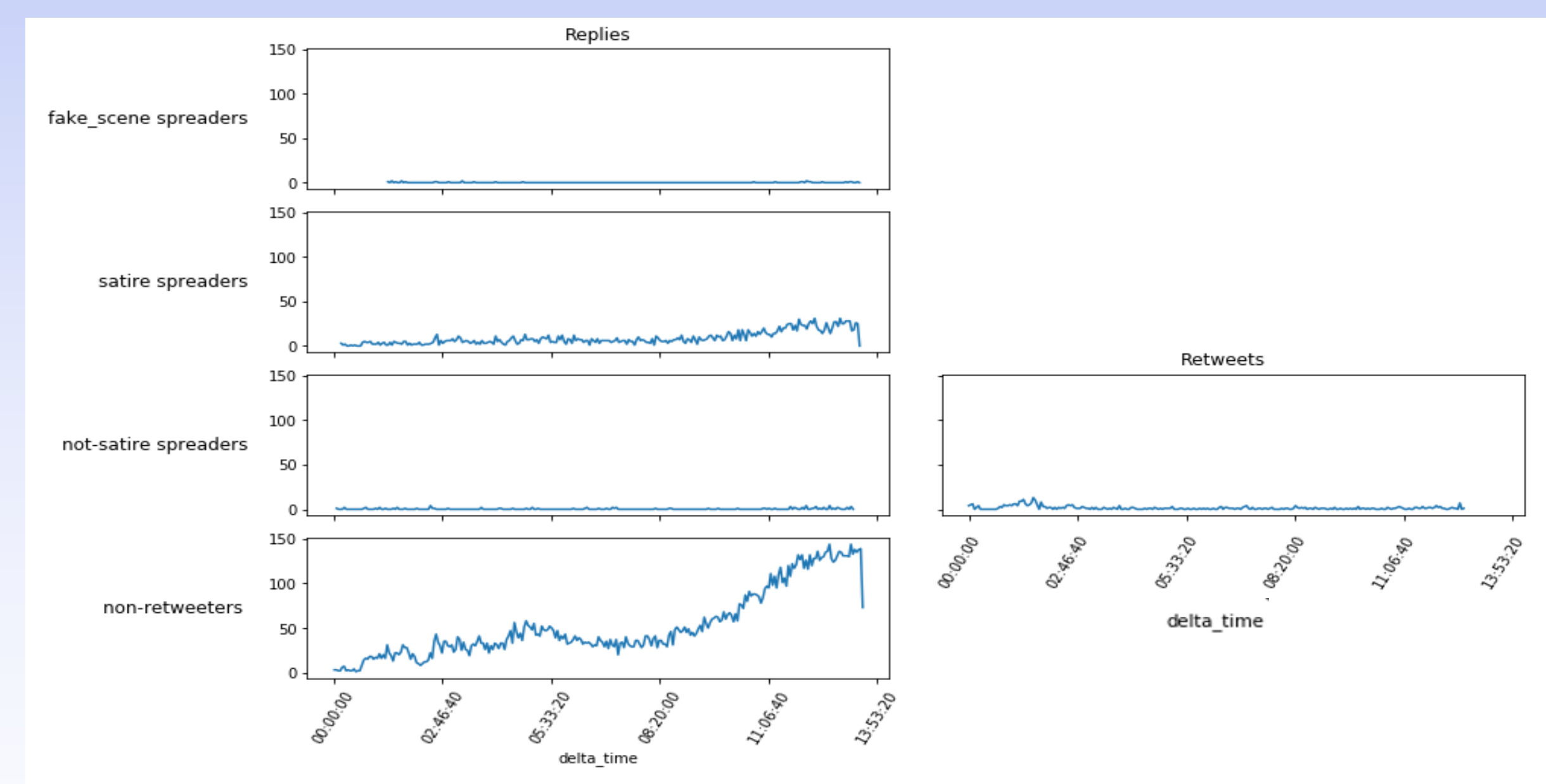
Results

In general, we observe that stories that have longer half-lives fail to reach "viral" status, as observed by the number of tweets. However, there appears to also be a difference in the half-life between retweets and replies/quotes, the later having longer half-lives for comparable numbers of tweets.

We also observed different response behaviors from the identified communities. First, while not-satire spreaders replied to each other almost exclusively, not-satire tweets evoked direct replies/quotes from all other groups. Second, while satire-type tweets tend to be primarily retweeted, most of the other observed responses come from its own community.

These results suggest that careful consideration is needed when aggregating the different types of interactions between users, especially when the purpose is community detection. Moreover, even within the reply/quote interactions, a distinction has to be made on the stance the response has to the original message. This is made apparent below for one of the most retweeted not-satire stories. While replies to this message where overall positive within its own community, the responses originating from other communities were negative and dwarfed even its retweets. This could have played a role in the eventual shutdown of the account.

www.casos.cs.cmu.edu

isr institute for SOFTWARE RESEARCH