

School of Computer Science and Informatics
Cardiff University
2021



**Predicting Information Flow and Survival of Malicious Posts
Around Popular Events**

Author: Paige Lowe
Supervisor: Amir Javed

Project description

In recent years, social media platforms such as Twitter, Facebook and Instagram have largely increased their presence in our daily lives. As of the 2nd January 2020, Twitter alone reported 145 million daily users and 500 million tweets daily. [1] That amasses to 6000 tweets per second through one platform alone. Though social media has a great impact on the connectivity we may feel in our lives it unfortunately leaves us open to many forms of cyber-attacks.

One such form of attack is a drive by download, whereby a malicious URL is posted to a social platform and shared across the network unknowingly by its users. A malicious URL is a clickable link embedded in some document, in this case a social media post, that once clicked has the purpose of compromising the user in some way. By clicking a malicious URL, the user may download malicious software that may cause harm to the device or its data. [2].

In order to prevent these forms of attacks, we first need to successfully identify and recognize the factors that aid their survival. As part of this research project my proposed solution will focus on building a program to identify the factors that aid in the propagation of these URL's across a social network. In particular, I will use data gathered from the online platform Twitter and focus my research on how attackers make use of significant events to spread malicious content. I have chosen to focus the research on popular events as they tend to be a prevalent opportunity for attackers to spread their content.

The first part of the project will be collecting malicious tweets around the topic of Covid-19 and incorporating these with existing datasets captured from various sporting events. By combining data from various types of events I hope to draw conclusions from a more complete model. I will then be building a program to analyse and identify factors these tweets have in common and that correlate to survival. Such factors may include...

- The URL itself - domain name - content/keywords, length.
- The posting account - friends, age of account, profile image, how active the account is on twitter.
- The tweets content - sentiment, nouns/verbs/keywords, images, hashtags, likes, time and day of the week it was tweeted - posting at times of high user activity may correlate to survival.

I then aim to build a model of the above findings and in doing so hope to identify the features that correlate with information flow and survival of malicious posts. In the future, this research could then be applied and coupled with machine-learning methods to train classification models to help identify malicious posts and prevent their spread across social platforms.

Ethical considerations

A key consideration of ethicality in this project is the dataset that will be used. Given that the data will be gathered from a public domain for research purposes, ethical approval is not required. Nonetheless, throughout the project an ethical approach shall be maintained. In particular, the final statistical model developed will aim to be transparent, fair and accountable in the conclusions it draws and where these conclusions are commented on in the final report this will be done in an ethical way.

Goals and Objectives

The main goal in this project is to build a program that identifies the key factors that aid in propagation and survival of malicious information across social platforms. The research objectives and primary goals that need to be achieved in order to meet this are covered below.

Research Objectives

- Research methods for collecting data from twitter.
- Research to identify content-based factors that could aid in survival of tweets. This could include the number of likes, sentiment analysis, emotion extraction, type of content (image etc.)
- Gain knowledge of existing resources for analysing the features mentioned above.
- Research methods for parallelising URL checks against virustotal.
- Research potential modelling software to build the final statistical model (STATA/SPSS).

Primary Goals

- Collect a dataset of Covid based tweets that contain URLs.
- Filter the dataset for those with malicious URLs using virustotal.
- Build a program that analyzes the resulting dataset to identify each malicious tweets sentiment and expressive emotion and also extract content and account based features that could be used to build statistical models.
- Build statistical model to identify factors contributing to size (virality) and survival of tweet.

Desirable Goals

- Visualise correlations between the factors that affect survival and flow of malicious content.

Work Plan

The following is a simplified timeline plan of the project, the key milestones have been identified and mapped between their appropriate tasks. By measuring my progress against this plan, I aim to stay on track to meet my goal of submitting the project and report by the end of week 12.

| Week | Tasks | Milestone |
|-----------------------|--|---|
| Week 1 (27/01/20) | Initial meeting with supervisor Apply for twitter developer account Write initial plan Submit initial plan to supervisor for approval | |
| Week 2 (03/02/20) | Submit initial plan Begin research into methods of collecting data from twitter Decide on the key attributes to capture | Submit initial project plan |
| Week 3 (10/02/20) | Begin collecting data from twitter Fortnightly supervisor meeting | Begin project implementation (collecting data) |
| Week 4 (17/02/20) | Finish collection of data Filter and parse/format data ready for processing | Finish collecting data |
| Week 5 (24/02/20) | Begin building a program to extract sentiment, emotion, content and various account-based features from the dataset. Fortnightly supervisor meeting | Begin identifying important factors |
| Week 6 (03/03/20) | Finalise implementation of the above analysis program | Finish building main program |
| Week 7 (10/03/20) | Begin building statistical model from the data Fortnightly supervisor meeting | |
| Week 8 (17/03/20) | Completion of coding side of project Analyse the statistical model - begin thinking of conclusions that could be drawn | Finish modelling findings Finish coding of project |
| Week 9 (24/03/20) | Begin writing final report Fortnightly supervisor meeting throughout duration | Begin final report |
| Week 10 (31/03/20) | | |

References

[1] - Brandwatch. 2021. *60 Incredible and Interesting Twitter Stats and Statistics*. [online] Available at: <<https://www.brandwatch.com/blog/twitter-stats-and-statistics/#:~:text=Twitter%20usage%20statistics,That's%206%2C000%20tweets%20every%20second.>> [Accessed 29 January 2021].

[2] - Spambrella. 2021. *What is a malicious URL? - Spambrella*. [online] Available at: <<https://www.spambrella.com/what-is-a-malicious-url/>> [Accessed 29 January 2021].