SPB-BRiMS Challenge Paper Abstract

Title: Which fake fits? How the diffusion of disinformation varies by campaign type and community in the *Black Panther* Twitter discussion.

Authors: Matthew Babcock, Ramon Alfonso Villa Cox, Sumeet Kumar
Institution: Carnegie Mellon University

Introduction

There have recently been important public policy discussions over how to confront the growing problem of disinformation being shared through social media such as Twitter [1]. Recent research has focused on and been helpful in distinguishing between "true" and "false" information and comparing the characteristics of both on Twitter [2]. Unfortunately, not all "false" information is the same and not all false information affects 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. To inform future work in this area, we conducted a case study to explore how different communities reacted to the different types of disinformation that were present in the Twitter conversation related to the opening of the *Black Panther* film.

Data Description

Black Panther was a financial and critical success that was promoted in part by its status as the first Marvel Cinematic Universe movie to have a predominately African and African-American cast and focus. The opening of the Black Panther movie makes for a good case study because of the high level of Twitter activity surrounding the movie (it is currently the most tweeted about movie [3]) and due to the presence of multiple types of disinformation campaigns within the Twitter conversation. Four types of disinformation stories were identified: 1) Fake Attack posts claiming racially-motivated physical violence at movie theaters which were debunked, 2) Satirical Fake Attack posts making similar but more exaggerated claims in an attempt to mock or shame the original Fake Attack posts, 3) Fake Scene posts claiming the film contained scenes it did not, and 4) Alt-Right posts claiming the movie was supportive of Alt-Right ideology when in the film such policies are questioned and repudiated. A total of 5,229,739 tweets related to #BlackPanther and the various fake posts were collected from February 8 to March 16, 2018 using Twitter's API.

Analytical Methods

Community detection in networks is analogous to clustering of data. Clustering aims to group data-points with similar attributes together. In a similar way, community detection in networks tries to group nodes with similar connectivity patterns together. One popular model used to obtain these communities is the Stochastic Block Model (SBM) [4]. In the simplest form of the model, it is assumed that K different latent classes exists and the membership of each node (or vertex) in the network is determined independently. Moreover, edges are placed between each of the nodes with probabilities that are a function of their

respective classes. However, it is important to recognize that not all types of communications between agents are equal. In the context of Twitter, retweets imply endorsement of an idea while direct replies or quotes can signal either disagreement or endorsement. To tackle this limitation, we also estimate a Mixed Membership SBM [5], where each node has a soft assignment probability vector to the different clusters modeled.

Regardless of flavor of SBM used, its estimation not only provides estimates of the network's block structure (either by soft or hard assignments), but also information of the likelihood of communication within and between each of the blocks (measured by the interaction probabilities). Hence, allowing us to assess how conductive each of the communities is to the various types of disinformation identified. This is achieved by measuring the probability of observing the diffusion pattern of a particular story given the simulated distribution of diffusion patterns (based on the estimated edge probabilities). Contrasting how each of these types of stories is placed on the diffusion distribution for each community allows us to determine which is more structurally conductive to the class of message.

Expected Results

Following the above analytical methods, we intend to discover the following:

1. Identification of the various communities that existed within the *Black Panther* Twitter conversation and which communities are home to the originators of the disinformation posts.
2. A description of how the four disinformation campaigns spread (were retweeted or replied to) within each community.
3. An improved understanding of which communities are more conducive to each type of disinformation campaign.

References

1. Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D.: The science of fake news. Science. 359, 1094–1096 (2018)
2. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science. 359, 1146–1151 (2018)
3. The Associated Press: Twitter: "Black Panther" is most tweeted about movie ever, https://wtop.com/social-media/2018/03/twitter-black-panther-is-most-tweeted-about-movie-ever/
4. Karrer, B., Newman, M.E.J.: Stochastic blockmodels and community structure in networks. Phys. Rev. E. 83, (2011). doi:10.1103/PhysRevE.83.016107
5. Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P.: Mixed Membership Stochastic Blockmodels. 34