TwitterSim: A policy-oriented test-bed for the spread of Contentious Messages in Twitter

- Black Panther opened to much social media fanfare and financial success in Feb 2018.
- 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.
- User communities were defined based in the retweeters of each type of story (as it signals tacit approval of the message).

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

Identifying the Stance of Responses in Twitter

- We developed a neural network classifier that uses the conversational structure of Twitter threads to classify the stance of responses.
  - Responses are classified as commenting, supporting, denying or querying.
- We seek to improve on the state of the art in two ways, by including an attention mechanism over conversation threads.
  - This can improve accuracy over longer threads.
- We also designed a collection methodology oriented towards denials (which is considerably under-sampled in available datasets).

TwitterSim

- A simulation that incorporates the structural properties of the platform as well as mechanisms for individual user behavior.
- These properties impose a series of rules for interaction and content promotion that influence the diffusion of information.
- We model the main types of interactions available:
  - Quotes, Replies and Retweets
- An economy of attention is introduced by modelling the timeline of each users and their limited capacity to read them.
- The behavioral model of agents allow for stance in the type of responses in order to explore the effect of these interactions on the diffusion of contentious information.
- Validation: we will evaluate the simulated diffusion process of the different types of stories against what is observed on the different case studies.
- The case studies, based on known rumors, are determined based on our stance classifier.