Understanding the Patterns of Health Information Dissemination on Social Media during the Zika Outbreak

Improving Population Health Using Novel Data Sources

S47

Xinning Gui¹ and Yue Wang²

1. University of California, Irvine
2. University of Michigan

Twitter: #AMIA2017
Disclosure

We and our spouses/partners have no relevant relationships with commercial interests to disclose.
Learning Objectives

After participating in this session the learner should be better able to:

• Utilize a mixed methods approach, in addition to machine learning, to monitor and assess the dissemination of health information (e.g. that related to Zika) on social media
• Formulate effective strategies for communicating public health information on social media
• Identify opportunities and challenges that social media present to risk communication
Research Goals

• To analyze the risk communication on Twitter during 2016 Zika outbreak

• To understand the information created by **general public** and **public health authorities**, and different information dissemination patterns

• To provide implications for effective risk communication strategies on social media

Collecting Zika-related tweets

10% of Twitter in 2016

"zika"

1,495,480 tweets

Top languages
- English 54%
- Spanish 27%
- Portuguese 12%

10% of Tweets
Zika tweets & Google trend

Cases reported in 20 countries in the Americas

Brazil 2016 Summer Olympics
Worldwide Zika discussion on Twitter

mid January, 2016

mid February, 2016

WHO declared emergency on Feb 1st, 2016
Content analysis roadmap

1,495,480 tweets containing “zika”

54% in English

popular tweets (3,581)

#retweets >= 100

authoritative tweets (1,227)

17 authoritative accounts (CDC, WHO, NIH, federal/state/local officials)

Machine learning-assisted content analysis

4,808 tweets
Machine Learning-Assisted Content Analysis

Open coding; Axial coding  →  Codebook  →  Coded tweets  →  Train machine learning model (multiclass text classification)

285 sampled tweets (200 popular + 85 authoritative)

4,808 tweets (3,581 popular + 1,227 authoritative)

Use learned model to assign codes to large data set
Machine Learning-Assisted Content Analysis

Open coding; Axial coding → Codebook → Coded tweets → Train machine learning model (multiclass text classification)

- 285 sampled tweets (200 popular + 85 authoritative)
- 4,808 tweets (3,581 popular + 1,227 authoritative)

Use learned model to assign codes to large data set

Cross-validation accuracy ~80%

Scale up analysis to big data with modest amount of manual effort
285 coded tweets: 8 categories

- Joke
- Policy
- Research Progress
- Infection Update
- Sports Events
- Consequence
- Scientific Knowledge
- Pharmaceutical Progress

Watch GeoVax $GOVX will hit #WSJ front page: as a #Zika #vaccine swoops the world before #REO2016 #Olympics to create a #trading powerhouse.
Typology of Zika-related Tweet Content

Descriptive Statistics of Top Retweeted Tweets (# of retweets >=100)

<table>
<thead>
<tr>
<th>Category</th>
<th>Retweet Count Average (std. dev.)</th>
<th>Like Count Average (std. dev.)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joke</td>
<td>1362 (3706)</td>
<td>1425 (3040)</td>
<td>1.6</td>
</tr>
<tr>
<td>Sports Events</td>
<td>336 (452)</td>
<td>317.0 (311)</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Descriptive Statistics of Authoritative Tweets

<table>
<thead>
<tr>
<th>Category</th>
<th>Retweet Count Average (std. dev.)</th>
<th>Like Count Average (std. dev.)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joke</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sports Events</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Selecção Brasileira @BrazilStat

USA goalkeeper joked about Brazil being dangerous due to Zika. Every time she hit the ball, Brazil fans scream ZIKA.
twitter.com/idiotforddl/st ...

12:00 AM - 7 Aug 2016

20,809 Retweets 14,580 Likes

Read @CDCgov's latest #Zika reports on the @CDCMMWR website:
cdc.gov/mmwr/zika_repo...

11:25 AM - 29 Jul 2016

3 Retweets
CDC has updated its interim guidance for U.S. health care providers caring for pregnant women with possible Zika virus exposure, to include the emerging data indicating that Zika virus RNA can be detected for prolonged periods in some pregnant women. To increase the proportion of pregnant women with Zika virus infection who receive a definitive diagnosis, CDC recommends expanding real-time reverse transcription-polymerase chain reaction (rRT-PCR) testing. Possible exposures to Zika virus include travel to or residence in an area with active Zika virus transmission, or sex with a partner who has traveled to or resides in an area with active Zika virus transmission without using condoms or other barrier methods to prevent infection. Testing recommendations for pregnant women with possible Zika virus exposure who report clinical illness consistent with Zika virus disease (symptomatic pregnant women) are the same, regardless of their level of exposure (i.e., women with ongoing risk for possible exposure, including residence in or frequent travel to an area with active Zika virus transmission, as well as women living in areas without Zika virus transmission who travel to an area with active Zika virus transmission, or have unprotected sex with a partner who traveled to or resides in an area with active Zika virus transmission). Symptomatic pregnant women who are evaluated <2 weeks after symptom onset should be tested. Symptomatic pregnant women who are evaluated 2–12 weeks after symptom onset should first receive a Zika virus immunoglobulin (IgM) positive or equivocal serum and urine rRT-PCR testing should be performed. Testing recommendations for pregnant women with possible Zika virus exposure who report clinical illness consistent with Zika virus disease (asymptomatic pregnant women) differ based on the circumstances of possible exposure. For asymptomatic preconception Zika virus transmission and who are evaluated <2 weeks after last possible exposure, rRT-PCR testing should be performed. If the rRT-PCR result is positive or equivocal, serum and urine rRT-PCR testing should be performed 2–12 weeks after the exposure. Asymptomatic pregnant women who do not live in an area with active Zika virus transmission, who have no possible exposure, and who report a Zika virus IgM antibody test result is positive or equivocal, serum and urine rRT-PCR testing should be performed within 2–12 weeks after the exposure. These recommendations will be updated when additional data become available.

Introduction

Zika virus continues to spread worldwide, and as of July 21, 2016, 50 countries and territories reported active Zika virus transmission (locations with active transmission). Although most persons with Zika virus infection are asymptomatic or have mild clinical disease, infection during pregnancy can cause congenital Zika virus infection, and it has also been linked to other adverse pregnancy outcomes, including miscarriage and stillbirth. The U.S. Zika Pregnancy Registry (ZAPPS) was established in collaboration with state, tribal, local, and territorial health departments to monitor the spectrum of conditions associated with congenital Zika virus infection and determine the risk for Zika virus infection during pregnancy and the spectrum of conditions associated with congenital Zika virus infection in the 50 U.S. states and the District of Columbia, and 378 women in all U.S. territories (aggregated territories' data from the USZPR and ZAPPS).
Implications

• Conduct more interactive and engaging communication strategies

  • E.g., CDC’s viral Zombie Apocalypse campaign

• Consider including more information in tweets and restating scientific messages in plain language
## Typology of Zika-related Tweet Content

### Descriptive Statistics of Top Retweeted Tweets (≥100 retweets)

<table>
<thead>
<tr>
<th>Category</th>
<th>Retweet Count Average (std. dev.)</th>
<th>Like Count Average (std. dev.)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharmaceutical Progress</td>
<td>1192 (1509)</td>
<td>607 (787)</td>
<td>0.4</td>
</tr>
<tr>
<td>Consequence</td>
<td>704 (922)</td>
<td>450 (680)</td>
<td>1.5</td>
</tr>
<tr>
<td>Research Progress</td>
<td>419 (672)</td>
<td>301.4 (529)</td>
<td>9.5</td>
</tr>
</tbody>
</table>

### Descriptive Statistics of Authoritative Tweets

<table>
<thead>
<tr>
<th>Category</th>
<th>Retweet Count Average (std. dev.)</th>
<th>Like Count Average (std. dev.)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scientific Knowledge</td>
<td>43 (177)</td>
<td>22 (65)</td>
<td>47.6</td>
</tr>
<tr>
<td>Infection Update</td>
<td>25 (45)</td>
<td>87 (56)</td>
<td>24.8</td>
</tr>
<tr>
<td>Policy</td>
<td>26 (38)</td>
<td>19 (30)</td>
<td>17.0</td>
</tr>
</tbody>
</table>
**Implications**

- Consider monitoring information dissemination trends on social media to
  - gain familiarity with major conversations and debates that take place among the general public
  - evaluate the effectiveness of social media efforts
Temporal Development between Popular and Authoritative Tweets

Research Progress

Infection Update

popular authoritative

popular authoritative
Implication

- Consider publishing more timely content related to influential news and major events
Conclusion

• providing more engaging and straightforward health message contents that attend to people’s information needs

• adopting more interactive communication strategies

• delivering messages timely after related news and major events

• monitoring information dissemination trends on social media and evaluating the effectiveness of social media efforts
Acknowledgement

The Co-authors

Yubo Kou
PhD, Purdue University

Tera Leigh Reynolds
MPH, MA, University of California, Irvine

Yunan Chen
PhD, University of California, Irvine

Qiaozhu Mei
PhD, University of Michigan

Kai Zheng
PhD, University of California, Irvine

And Anonymous Reviewers
THANKS

raywang@umich.edu

guix@uci.edu