



# CS 8803 Social Computing: Credibility

*Munmun De Choudhury*

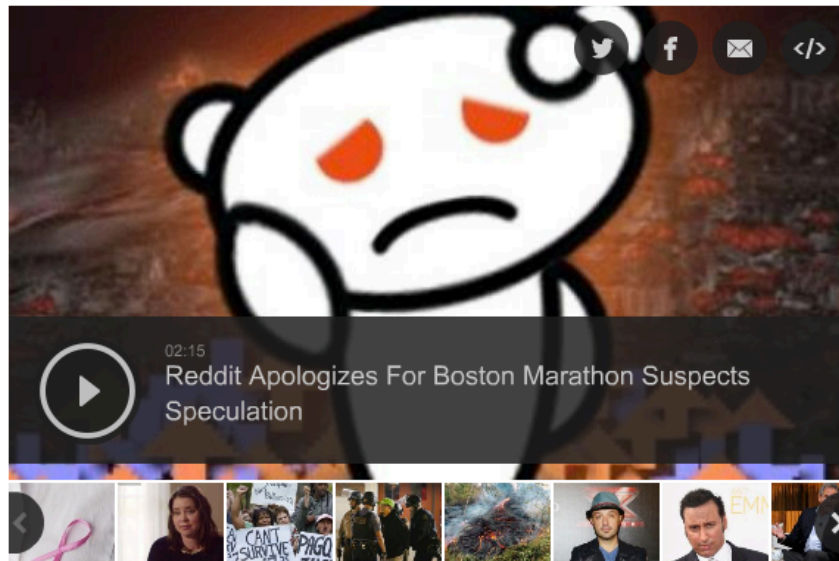
[munmund@gatech.edu](mailto:munmund@gatech.edu)

Week 12 | November 3, 2014

# Reddit Apologizes For Speculating About Boston Marathon Suspects

The Huffington Post | By Katherine Bindley

Posted: 04/22/2013 5:18 pm EDT | Updated: 04/23/2013 7:31 pm EDT



292 80 45 0 3 133

Like Share Tweet LinkedIn Email Comment

Speculation on social media last week over who was responsible for the [bombing at the Boston Marathon](#) produced its own set of innocent victims: the falsely accused.

Reddit -- which was fiercely criticized for its ["Findthebostonbombers" thread](#) that called out specific people standing near the scene as suspects -- has now issued a [public apology](#) to those whose names were dragged through the mud.

"We all need to look at what happened and make sure that in the future we do everything we can to help and not hinder crisis situations," the statement reads.

**Sponsored Links**

**Odd Trick Fights Diabetes**  
 "Unique" Proven Method To Control Blood Sugar In 3 Weeks. Watch Video.  
 Smart-Consumer-Lifestyles.com

▶ The message, which was published in a blog post on Monday, continues by mentioning the serious ramifications of the rampant speculation that occurred on the site.

**Know Where You Stand**

# Information Credibility on Twitter

# Summary

- The paper investigates use of machine learning methods to assess credibility of tweets – distinguish news from rumor or fake information.
- Features used: message, user, topic, propagation.
- Findings:
  - User based features perform better than message features
  - Presence of URL in a post was the feature with highest information gain followed by estimate of negative sentiment and question-centric nature of posts.
- Human assessment to create ground truth for the algorithm.  
70-80% accuracy.
- Limitations:
  - Only focused on tweets belonging to trending topics.

Tweeting is Believing?  
Understanding Microblog  
Credibility Perceptions

# Summary

- The article presents results of a survey study that gauged users' perceptions of tweet credibility.
- 26 features identified to impact credibility assessment based on a pilot sample of five individuals.
- Two controlled experiments to measure the impact of several tweet features (message topic, user name, and user image) on perceptions of message and author credibility
  - Participants recruited from inside Microsoft and through a message board of CMU
- Main findings:
  - Users are poor judges of tweet credibility based on content alone; reliance on cues such as username and presence/absence of profile pictures.

# Finding and Assessing Social Media Information Sources in the Context of Journalism

# Summary

- The article develops a tool SRSR( Seriously Rapid Source Review) which enables journalists to search for and assess sources in social media around news events
- User-centered design approach; pilot interviews followed by design, deployment, and evaluation of SRSR
- Different cues about a source and content of tweet are taken into account: *aggregated, derived, computed*
  - Source classifier
  - Eyewitness classifier
  - Content attributes – named entities, type of URL shared, sharing from mobile devices
  - Spatial attributes – location field match with location of event, concentration of friends in the location of event



Three aspects not examined in much depth:

- 1) **who** is consuming the information whose credibility is being judged;
- 2) ease of **verifiability** of the information (e.g., would a simple search help?)
- 3) **motivation** behind spreading the credible (or not) information.

How can these attributes we incorporated in credibility assessment?

Most of the papers focus on assessing credibility of news. Would same observations apply to judging credibility of non-real time information? E.g., health myths

More question-centric nature of posts around a topic may also be because the topic is controversial (ref. Castillo et al.). It doesn't necessarily mean it is less credible. How can we distinguish between these subtle cases?

Credibility is, after all, a domain-dependent attribute. What additional new feature would you consider, in addition to the ones raised in the papers, that could be useful for this purpose?

That presence of an URL indicates credibility (ref. Castillo et al.) may be exploited by ill-intentioned users to increase “clicks”. What are the dangers of relying on URL presence as a feature of credibility alone?

Morris et al. found that in general, a Twitter user's displayed image did not make much difference to their credibility perception. Could the same observation be extended to other social media platforms? [Hint: internet user name styles were found less credible]

The papers examined and studied credibility on Twitter. How would the different cues change if we look at the host of new mobile-centric social apps?

Except Diakopoulos et al., to some extent, none of the other papers exploit the wealth of information embedded in the network structure of a user. This can be very useful for credibility purposes. Discuss some ways you'd utilize this information.



Could Reddit's link or comment karma be useful to infer credibility? How would you use it? What are its limitations?

On Facebook and Twitter, users can't assign trust scores to others, except Facebook's "like" feature. Could the "like" feature be used for credibility inference in any way? What other cues on Facebook would be useful?

... and we haven't even talked about "gaming" the system to make a certain information look credible!

# Next class

- Wednesday 11/05
- Topic: “Polarization and Selective Exposure”
- Assigned readings due by 11:59 pm Tuesday
- Assignment III