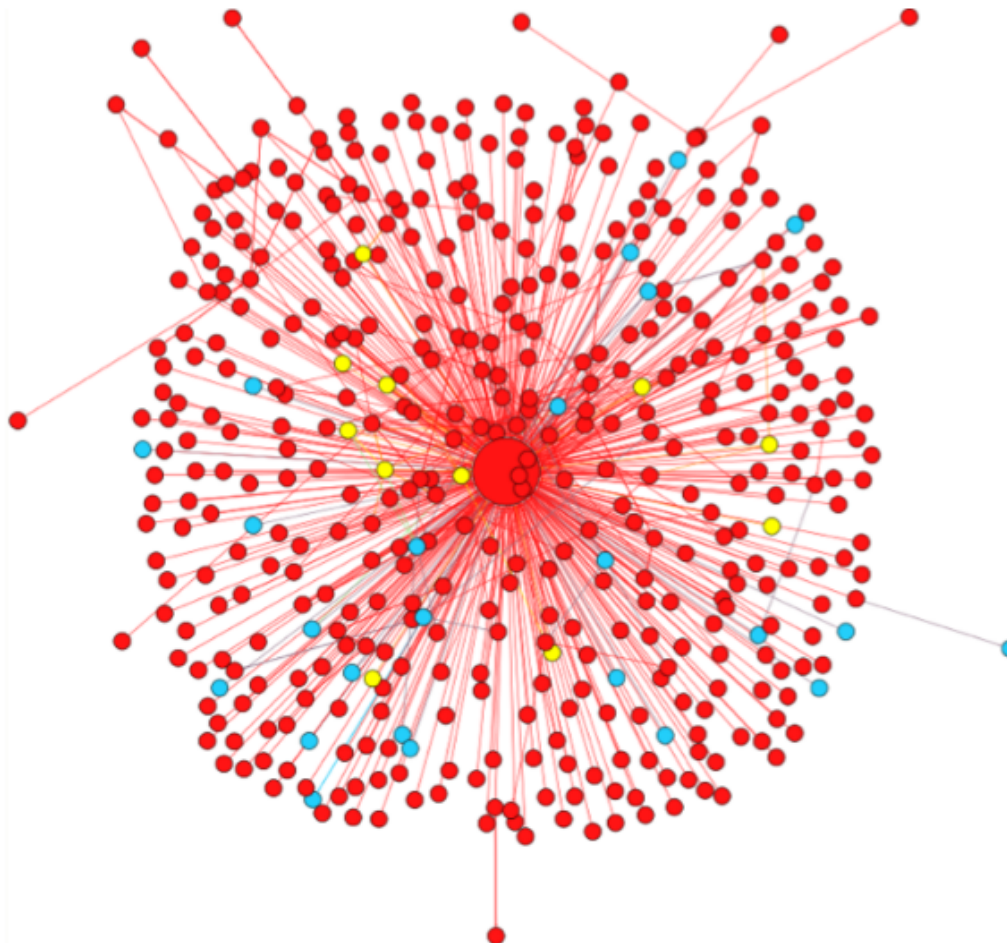# Rumor-detection software IDs disputed claims on Twitter

May 29 2015, by Glenda Bullock

One minute after the hacked twitter account sent out a rumor of explosions at the White House, users were already inquiring about its accuracy. Blue nodes show inquiry tweets. Yellow represent correcting tweets. Red represent spreading tweets. Credit: Zhe Zhao

A week after the Boston marathon bombing, hackers sent a bogus tweet from the official Twitter handle of the Associated Press. It read: "Breaking: Two Explosions in the White House and Barack Obama is injured."

Before the AP and White House could correct the record, the stock market responded, dropping more than 140 points in a matter of minutes. Losses mounted into the billions.

The market recovered just as quickly, but analysts said the timeframe could well have been long enough for in-the-know perpetrators to profit through trading.

Rumors and their negative effects can spread rapidly in these hyperconnected times, says Qiaozhu Mei, an associate professor in the University of Michigan School of Information and Department of Electrical Engineering and Computer Science.

That's why he and a team of researchers have developed software to help society identify and correct erroneous claims on Twitter. They introduced the software recently at the International World Wide Web Conference in Florence, Italy. Later this summer, they hope to put it in practice at a website they're developing called Rumor Lens.

"One post of a rumor in social media can sometimes spread beyond anyone's control," said Mei, an expert on text mining and natural language processing. "Our goal is to detect emerging rumors as quickly as possible."

The team demonstrated what its software is capable of by analyzing two sets of tweets: 30 million sent relating to the Boston Marathon bombing in April 2013 and a random sample of 1.2 billion tweets sent during November of the same year.

Two seconds after the first denial from an AP employee and two minutes before the of- ficial denial from AP, the rumor had already gone viral. Red represents the rumor spreading. Blue shows questioning tweets and yellow nodes are correcting tweets. Credit: Zhe Zhao

They gathered the second set from Twitter's Gardenhose—10 percent of its real-time stream. The datasets represent both an unpredictable, high-profile event that would likely spawn rumors and a relatively uneventful span of time.

The software successfully detected 110 rumors from the stream of tweets about the Boston Marathon bombing, with an average accuracy of more than 50 percent. Its average accuracy was 33 percent for Twitter Gardenhose data.

Both percentages are significantly higher than the less-than-10-percent accuracy of rumor-detecting through hashtag tracking and trending topics, the researchers point out. Furthermore, their software finds fishy statements a lot faster.

"Our method can detect rumors 3.6 hours earlier than methods that use trending topic detection, and 2.8 hours earlier than methods using hashtags as signals," said Zhe Zhao, a doctoral student in the Department of Electrical Engineering and Computer Science.

The researchers' key insight is that before social media users decide whether to believe a piece of information is true, many will ask for more information or express skepticism.

So they designed their software to listen in on Twitter traffic for signs that users are "questioning the truth value of information." Words and phrases the program has an ear for include "unconfirmed," "Is this true?" and "Really?"
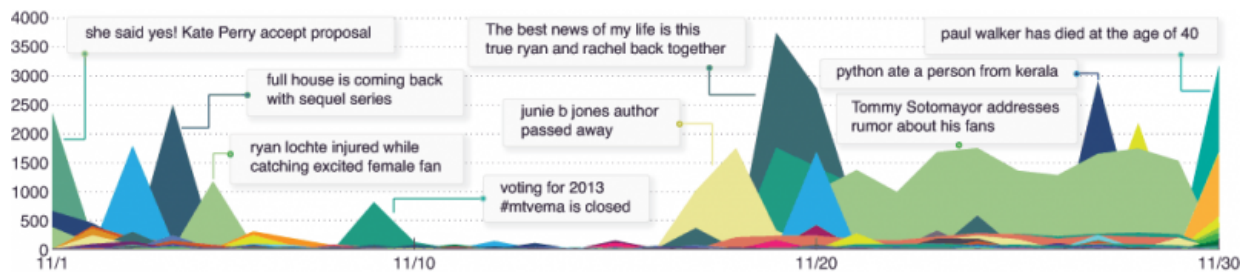
Tracking rumors on Twitter during November, 2013, a relatively uneventful

Once it zeroes in on a potential rumor, it looks for more tweets about the topic to gauge how widespread the conversation is. The researchers then rely on humans to fact-check.

The point of the effort isn't for a computer to determine whether a claim is true or false, but rather to highlight disputed information before it ends up on popular debunking sites like Snopes.com.

"By the time a rumor gets to Snopes, it's often too late," Mei said.

Tracking rumors on Twitter about the Boston marathon bombing. Credit: Zhe Zhao

Rumor Lens—the researchers' own website—is expected to be available in the next couple of months. The team envisions it serving as a Snopes-like online community of social media observers, academics and reporters who have an interest in following and debunking rumors.

The algorithms would highlight potential rumors and the people in the

community would do the fact-checking. The researchers define a rumor as a controversial statement that can be fact-checked.

The team presented a paper about the research at the World Wide Web Conference. Paul Resnick, a professor in the School of Information, is also a co-author. The work is supported in part by the National Science Foundation and the Defense Advanced Research Projects Agency.

More information: "Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts." www-personal.umich.edu/~qmei/pub/www2015-zhao.pdf

Provided by University of Michigan