

# Algorithm tests your Twitter skills and predicts virality of tweets

February 12 2015

---



As countless political orators have demonstrated, it's not just what you say, it's how you say it. Using automated text analysis, Cornell University researchers have identified an array of features that can make a tweet more likely to get attention, and have created a website that will predict which version of a tweet will be more popular:

[chenhaot.com/retweetedmore](http://chenhaot.com/retweetedmore)

The researchers tested their ideas on Twitter, where their computer algorithm predicted more accurately than human observers which

version of a tweet would be retweeted more. The results might be applied to longer forms of discourse, from essays to getting your idea accepted in a committee meeting. "We're looking at persuasion everywhere," said Lillian Lee, professor of computer science and information science.

Twitter enabled the researchers to conduct a controlled experiment to eliminate the effects of the popularity of the poster or the topic: Many posters will [tweet](#) on the same topic more than once, with different wording. The researchers collected and compared thousands of these pairs, and after taking into account the effect of repetition, the experiment showed that wording still matters.

Lee, graduate student Chenhao Tan and Google researcher Bo Pang reported their results in the Proceedings of the Association for Computational Linguistics, and Lee will present the research during a symposium, "The Linguistics of Status, Influence, and Innovation: A Computational Perspective," Feb. 15 at the annual meeting of the American Association for the Advancement of Science in San Jose, Calif.

On Twitter, many posts are links to websites the poster thinks are important. "You want to say something about it to make people look at it," Lee explained. Previous [social science research](#) has identified several features of an argument that might make it more effective.

The Cornell researchers tested each one separately, then combined them into their algorithm.

The computer looks for the occurrence of certain keywords and compares "bigrams" – combinations of two words. Such combinations may reflect a linguistic style. "Cornell," for example, might often be followed by "Chronicle" within the campus community, but on Twitter

you might find "Cornell research," "Cornell students," "Cornell ornithologists."

The researchers found these features most likely to generate retweets:

- Ask people to share. Words like "please," "pls," "plz" and, of course, "retweet" were common in successful messages.
- Be informative (often measured by length).
- Use the language of the community, and be consistent with the language you usually use yourself, with which your followers are familiar. The researchers are also testing on Reddit, where users form distinct communities.
- Imitate the style of newspaper headlines. (In their tests, the researchers used the *New York Times* as a model.)
- Use words that appear often in other retweeted messages.
- Use words that express positive or negative sentiment.
- Refer to other people, not just yourself. Use third person pronouns.
- Use generalizations. Statements that can be applied to a variety of situations are the most often repeated.
- Make it easy to read. The researchers applied a formula used to measure the grade level of a text.

"We would love to capture amusingness or cleverness, but we haven't found a way to do that yet," Lee added.

The researchers conclude with a challenge: Social scientists should try to find out why these tactics work. The research was supported by the National Science Foundation and Google.

Provided by Cornell University

Citation: Algorithm tests your Twitter skills and predicts virality of tweets (2015, February 12)  
retrieved 7 May 2024 from

<https://phys.org/news/2015-02-algorithm-twitter-skills-virality-tweets.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.