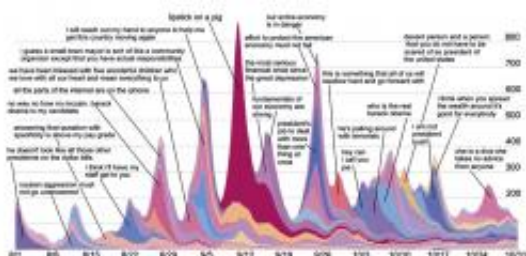


'Lipstick on a pig' -- tracking the life and death of news

July 13 2009, by Bill Steele



Top 50 threads in the news cycle with highest volume for the period Aug. 1 to Oct. 31, 2008. The total number of posts was constant through the period; peaks represent stories that took up the largest proportion of all posts at a given time. An interactive version of the graph is available at <http://memetracker.org>.

(PhysOrg.com) -- By observing the global flow of news online, Cornell computer scientists have managed to track and analyze the "news cycle" - the way stories rise and fall in popularity.

Jon Kleinberg, Cornell professor of [computer science](#), Jure Leskovec, postdoctoral researcher, and graduate student Lars Backstrom tracked 1.6 million online news sites, including 20,000 mainstream media sites and a vast array of blogs, over the three-month period leading up to the 2008 presidential election - a total of 90 million articles, one of the largest analyses anywhere of online news.

They found a consistent rhythm as stories rose into prominence and then

fell off over just a few days, with a "heartbeat" pattern of handoffs between blogs and mainstream media. In mainstream media, they found, a story rises to prominence slowly then dies quickly; in the blogosphere, stories rise in popularity very quickly but then stay around longer, as discussion goes back and forth. Eventually though, almost every story is pushed aside by something newer.

"The movement of news to the Internet makes it possible to quantify something that was otherwise very hard to measure - the temporal dynamics of the news," said Kleinberg. "We want to understand the full news ecosystem, and online news is now an accurate enough reflection of the full ecosystem to make this possible. This is one [very early] step toward creating tools that would help people understand the news, where it's coming from and how it's arising from the confluence of many sources."

The researchers also say their work suggests an answer to a longstanding question: Is the "news cycle" just a way to describe our perception of what's going on in the media, or is it a real phenomenon that can be measured? They opt for the latter, and offer a mathematical explanation of how it works.

The research was presented at the Association for Computing Machinery Special Interest Group on Conference on Knowledge Discovery and Data Mining Conference June 28-July 1 in Paris.

The ideal, Kleinberg said, would be to track "memes," or ideas, through cyberspace, but deciding what an article is about is still a major challenge for computing. The researchers sidestepped that obstacle by tracking quotations that appear in news stories, since quotes remain fairly consistent even though the overall story may be presented in very different ways by different writers.

Even quotes may change slightly or "mutate" as they pass from one article to another, so the researchers developed an algorithm that could identify and group similar but slightly different phrases. In simple terms, the computer identified short phrases that were part of longer phrases, using those connections to create "phrase clusters." Then they tracked the volume of posts in each phrase cluster over time. In the August and September data they found threads rising and falling on a more or less weekly basis, with major peaks corresponding to the Democratic and Republican conventions, the "lipstick on a pig" discussion, rising concern over the financial crisis and discussions of a bailout plan.

The slow rise of a new story in the mainstream, the researchers suggest, results from imitation - as more sites carried a story, other sites were more likely to pick it up. But the life of a story is limited, as new stories quickly push out the old. A mathematical model based on the interaction of imitation and recency predicted the pattern fairly well, the researchers said, while predictions based on either imitation or recency alone couldn't come close.

Watching how stories moved between mainstream media and blogs revealed a sharp dip and rise the researchers described as a "heartbeat." When a story first appears, there is a small rise in activity in both spheres; as mainstream activity increases, the proportion blogs contribute becomes small; but soon the blog activity shoots up, peaking an average of 2.5 hours after the mainstream peak. Almost all stories started in the mainstream. Only 3.5 percent of the stories tracked appeared first dominantly in the blogosphere and then moved to the mainstream.

The mathematical model needs to be refined, the researchers said, and they suggested further study of how stories move between sites with opposing political orientation. "It will be useful to further understand the roles different participants play in the process," the researchers concluded, "as their collective behavior leads directly to the ways in

which all of us experience [news](#) and its consequences."

Source: Cornell University ([news](#) : [web](#))

Citation: 'Lipstick on a pig' -- tracking the life and death of news (2009, July 13) retrieved 9 April 2024 from <https://phys.org/news/2009-07-lipstick-pig-tracking-life.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.