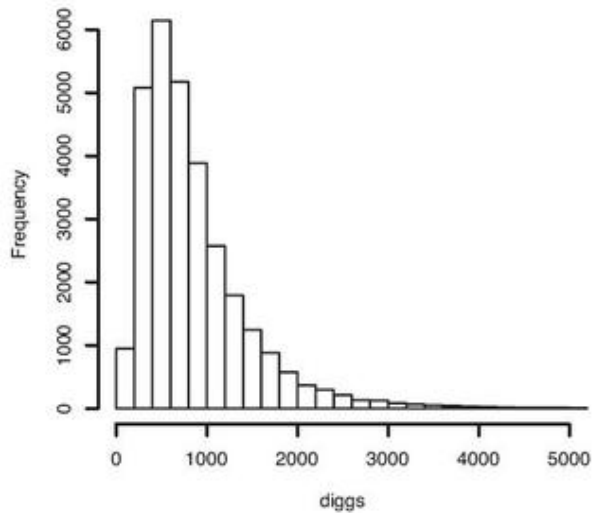


Researchers Study Digg.com to Investigate Collective Attention



One result of the study: The distribution of 29,864 popular stories on digg.com by the number of diggs each story received. The log-normal distribution shows that more than half of the stories received 200-800 diggs, with the number of stories with more than 800 diggs gradually decreasing up to about 4000 diggs. Credit: Fang Wu and Bernardo Huberman. ©2007 PNAS.

In a world where millions of people are bombarded with thousands of messages daily, understanding how some messages become popular among large populations is vital for successful advertising, marketing and journalism strategies.

Researchers Fang Wu and Bernardo Huberman from Hewlett-Packard Laboratories in Palo Alto, California, thought that the popularity of an idea or item in a large population should depend on its novelty, which tends to decrease over time due to a variety of factors. In order to test this idea, the researchers analyzed how a news story propagates on the interactive Web site digg.com.

Their results, which are published in *PNAS*, analyze 29,864 popular digg stories during the year 2006, with more than 1 million users determining the popularity of the submissions.

“The greatest significance of our results is that we were able to establish a connection between novelty and collective attention in large groups of people,” Huberman told *PhysOrg.com*. “Psychologists have studied the interplay between novelty and attention with human subjects for many years, but never at the group level on such a scale.”

Among Wu and Huberman’s results is the existence of a natural time scale over which attention grows and fades. Since the researchers only considered popular stories (those which made it to the front page, based on a complex algorithm), the initial growth must be rapid. However, once a story makes it to the front page, the data indicate a straight correlation between the average log of diggs and the time variance. In other words, the more popular a story becomes, the faster it spreads.

“Basically, at any given time, a group of n people will read the story, digg it and tell it to others, but only some random fraction of those told will actually read and digg the story,” Huberman explained. “This process, which naively would lead to multiplicative growth in the number of diggs, is counteracted by the fact that the novelty of the story is decaying in time. When put together, this process leads to a log-normal distribution of the number of diggs and also to a fast increase at the beginning that tapers off with a rate predicted by our equations, and verified to high precision with the data.”

As for the decline of interest in a story, a story’s digg rate decays, on average, following a stretched exponential law. This is because many factors contribute to the decreasing novelty of a successful story

(such as topic category, time of day when it was made popular, etc.). Since individual stories tend to fade exponentially, the combined result looks like a stretched exponential. Interestingly, stretched exponentials have been used to explain other decays, such as the discharge of a capacitor and luminescence decays.

The researchers also analyzed the probability distribution of stories based on the total number of diggs received, which is approximately log-normal. Specifically, most popular stories quickly climb to 200 diggs, but the largest section (20%) tops off at between 400 and 600 diggs. After that, the distribution of diggs drops off more gradually, with a handful reaching up past the 3000-digg mark.

This study is one of the first to analyze data on collective attention from very large groups of people in a non-laboratory, yet isolated, setting. As such, the researchers hope that the results will help experts in choosing which items to display in different types of news media, advertising, or any other medium that relies on social networking to spread its content.

“Our theory points to a methodology for displaying content on a portal,” Huberman said. “If one is interested in maximizing the number of views of the portal, then given our theory and a few measurements, they can decide on posting the newest stories first (like digg does now) or the more favorite ones.”

More information:

HP's Information Dynamics Lab: <http://www.hpl.hp.com/research/idl/>

Wu, Fang, and Huberman, Bernardo A. “Novelty and collective attention.” *Proceedings of the National Academy of Sciences*. November 6, 2007, vol. 104, no. 45, 17599-17601.

Copyright 2007 PhysOrg.com.

All rights reserved. This material may not be published, broadcast, rewritten or redistributed in whole or part without the express written permission of PhysOrg.com.

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