Gordon supercomputer used in 61-million-person Facebook experiment

September 20 2012

Credit: Alan Decker

(Phys.org)—A recently published study led by the University of California, San Diego in collaboration with Facebook and done in part using large-scale simulations on the San Diego Supercomputer Center's (SDSC) data-intensive *Gordon* supercomputer, confirms that peer pressure helps get out the vote while demonstrating that online social networks can affect important real-world behavior.

The study, published this month in the science journal *Nature*, found that about one-third of a million more people showed up at the ballot box in the United States on November 2, 2010 because of a single Facebook

message posted on that Election Day.

"Our study suggests that social influence may be the best way to increase voter turnout," said lead author James Fowler, UC San Diego professor of political science in the Division of Social Sciences and of medical genetics in the School of Medicine. "Just as importantly, we show that what happens online, matters a lot for the 'real world.'"

In the study, more than 60 million people on Facebook saw a social, non-partisan "get out the vote" message at the top of their news feeds on Nov. 2, 2010. The message featured a reminder that "Today is Election Day"; a clickable "I Voted" button; a link to local polling places; a counter displaying how many Facebook users had already reported voting; and up to six profile pictures of users' own Facebook friends who had reported voting.

About 600,000 people, or one percent, were randomly assigned to see a modified "informational message," identical in all respects to the social message except for pictures of friends. An additional 600,000 served as the control group and received no Election Day message from Facebook at all. Fowler and his colleagues then compared the behavior of recipients of the social message, recipients of the informational message, and those who saw nothing. The full article on the study can be read here .

Gordon Simulates Multi-Million-Person Social Networks

Though the main analysis was conducted on servers at Facebook, the research team turned to SDSC's Gordon supercomputer to optimize confirmatory Monte Carlo simulations – a process that generates thousands of probable outcomes or scenarios. They wanted to know if

they could really detect a treatment effect or correlation between two variables in the large-scale real world social network that they analyzed at Facebook. Those simulations produced randomly generated networks that were set up to look like the real-world networks observed by researchers in the study but did not actually contain any real Facebook data.

For each simulation a "true" value of the relationship between the treatment variable and a treated individual's behavior was randomly selected. Next, the researchers attempted to detect the "true" value of the relationship for each simulation using a statistical method that they describe in their study.

"If we could consistently find the treatment effect in the simulated networks then we'd have evidence that our statistical method was effective at estimating the treatment effect in the real-world network that we analyzed at Facebook," said Christopher J. Fariss, a Ph.D. candidate in the Political Science Department at UC San Diego and part of the research team. "The simulations provided us with pretty convincing evidence that we could indeed detect such a relationship."

Researchers used Gordon to simulate a 5,000,000-person network 1,000 times over to estimate the statistics. The process was then repeated nine times to account for different combinations of behavior types and network structures – meaning 9,000 simulations were run.

"Running the program that many times meant we needed a data-intensive resource as capable as Gordon," said Fariss. "At first, the program took a little more than one hour just to complete a single run of the simulation of the smaller network of only about 1,000,000 people. We then used vectorization or parallelization, in which software programs that perform only one operation at a time are modified to perform multiple operations simultaneously. That dropped the process to about one minute, which

dramatically cut the time needed to generate simulations of those larger social networks."

Fariss and the group also used additional multicore capabilities available in R, an open-source programming language widely used in statistical computing, to reduce compute times even further before generating the 5,000,000-person network simulations. That allowed researchers to complete all of the runs in about 8% of the time that the unmodified simulations would have taken.

"This project is a perfect example of how Gordon is assisting the research community on a wide range of data-intensive projects, and speeding their time to discovery," said SDSC Director Michael Norman, principal investigator for the Gordon project. "We are gratified to see fields other than science and engineering using supercomputing to do their research. This is a key objective of the Gordon project."

Additional co-authors of the study are Robert M. Bond, Jason J. Jones, and Jaime E. Settle of UC San Diego, and Adam D. I. Kramer and Cameron Marlow of [Facebook](). The study was supported in part by the James S. McDonnell Foundation and the University of Notre Dame and the John Templeton Foundation as part of the Science of Generosity Initiative.

More information: www.sdsc.edu/us/resources/gordon/

Provided by San Diego Supercomputer Center