



- [Home](#)
- [Videos](#)
- [Blogs](#)
- [Briefings](#)
- [Community](#)
- [Magazine](#)
- [Newsletters](#)
- [Events](#)
- [Resources](#)

-
-
-
-
-
-
- [Subscribe](#)

- [Computing](#)
- [Web](#)
- [Communications](#)
- [Energy](#)
- [Materials](#)
- [Biomedicine](#)
- [Business](#)

-

Mims's Bits

Christopher Mims is a journalist who covers technology and science for just about everybody.

[View Complete Bio](#)

Recent Posts

- [» Whatever Happened to ... Virtual Reality?](#)
- [» Is Microsoft Going to Go After the Chinese Government?](#)
- [» How to Handle the World's Largest Digital Images](#)
- [» Supercomputer Digests Twitter in Real-Time](#)
- [» Why There Can Never Be a Competitor to Google Books](#)

Top Stories Of The Day

- [» Cheap Electronics on Paper Diagnostic Chips](#)

Mims's Bits

Wednesday, October 20, 2010

Supercomputer Digests Twitter in Real-Time

Specialized hardware and software could provide new insights into the social Web.



Cray XMT supercomputer courtesy [Cray](#)

Determining the most influential users of Twitter is probably not what the creators of the [Cray XMT](#) supercomputer had in mind when they designed their machine. But when you're packing this much computational heat, you go where the hard problems are. Twitter, Facebook and the rest of the social Web have become the modern-day equivalent of the water cooler, albeit with an automatic transcriptionist present. And processing all the data that conversation generates turns out to be a very hard problem.

For example, as of February 2010, Facebook included 400 million active users with an average of 120 "friend" connections each, all of whom collectively shared 5 billion pieces of information in a single month.

Log In

[f](#) [t](#) [Forgot your password?](#) [Register](#)

RSS

[Subscribe to the Mims's Bits RSS Feed](#)

[» A Cell-Phone Network without a License](#)

[» Cheap Diesel-Powered Fuel Cells](#)

Advertisement

SPAIN
New Technologies in Spain Series
 An eight-part series focusing on Spanish innovation
[CLICK HERE](#)

Find out about the latest trends in clean energy.
[CLICK HERE](#)

Figuring out who the "influencers" are in such a massive social networks requires creating a gigantic social graph, where each user is a vertex and the connections between them are lines. Ranking users within such a graph requires a determination of their "centrality". That is, how many other people are connected to them, and how many people are connected to them, and so on, until you get to the trunk of the tree structure underlying connectedness on a service like Twitter.

It turns out this is not the sort of problem that is readily handled even by the usual go-to workstations of the scientific supercomputing world -- the GPGPU-powered supercomputers that leverage the graphics chips usually used to render lush 3D environments in videogames. These GPGPU workstations simply don't allow enough control over how many processes are running in parallel to efficiently churn through social graphs as big as the one represented by Twitter or Facebook.

That's why David Ediger of Georgia Tech, [with the help of a long list of collaborators](#), turned to the 128-CPU Cray XMT housed at the Pacific Northwest National Laboratory. The XMT is a favorite of supercomputing hot-rodders and uber-geeks who appreciate its fine-grained massively multithreaded tunability. This machine is usually pressed into service for solving problems like "Hierarchical Bayesian Modeling for Text Analysis" or [analyzing the stability of America's power grid](#), but Ediger had it cogitating on every stray thought from a single day's worth of the [Twitter firehose](#).

The Cray made short work of Twitter, disposing of an entire day's worth of connections in under an hour. The results will surprise no one -- on Twitter, a tiny fraction of sources are retweeted widely, mostly government and media, while the rest of the service is either people talking in small groups or literally talking to themselves.

The point, though, is that throwing a finely-tuned Cray running Ediger's custom software -- GraphCT -- at Twitter allowed the researchers to digest the service in something like real time. Which is exactly the sort of capability that intelligence agencies, marketers and perhaps even Twitter itself might want to have.

[Follow Mims on Twitter](#) or [contact him via email](#).

Tags: [social networks](#), [Facebook](#), [twitter](#), [supercomputer](#), [real-time search](#)

[Close Comments](#)

Comments

[Reply](#)

Advertisement

The Future of Energy Webcast at EmTech 2010
Watch On Demand Now.
 Sponsored by
Technology Review
PUBLISHED BY MIT

[« Previous Post](#)

[Next Post »](#)

- [Computing](#)
- [Web](#)
- [Communications](#)
- [Energy](#)
- [Materials](#)
- [Biomedicine](#)
- [Business](#)
- [Today's Stories](#)

Advertisement

SUBSCRIBE TODAY

Daily NEWSLETTERS

Monday - Friday

Daily e-mail and mobile newsletter summaries of our top stories and blog.

Stay informed with our **FREE** daily newsletters.

Technology Review
PUBLISHED BY MIT

Current Issue



The TR35

Our annual selection of the world's top innovators under the age of 35.

- [Subscribe](#) Save 36%
- [Table of Contents](#)
- [MIT News](#)

- [» Gift Subscription](#)
- [» Digital Subscription](#)
- [» Reprints, Back Issues](#)
- [» Subscribe](#)
- [» Table of Contents](#)
- [» MIT News](#)

More Technology News from [Forbes](#)

- [10 Tech Fixes You Can Make Right Now](#)
- [10 Dream HDTVs On A Budget](#)
- [20 Gadgets For Your 'Man Cave'](#)
- [10 Ways We're Really Spending Our Time](#)
- [20 Social Media Blunders](#)

Advertisement

Click here to learn how biometrics is being used to help secure our future.

eflorida.com

- [About Us](#) |
- [Privacy](#) |
- [Terms of Use](#) |
- [Subscribe](#) |
- [Advertise](#) |
- [Customer Service](#) |
- [Sitemap](#) |
- [Contact Us](#) |
- [Feedback](#)



© 2010 Technology Review. All Rights Reserved.