UCSB engineering researchers turn to geometry in their quest to map social networking data in real time

Living Story of Social Graphs

by Sonia Fernandez
From flash mobs at the local mall to trending activist hashtags, social networks have quickly integrated themselves into modern human life and become a tool for instantaneous global communication. Every day, an estimated 700 million people (out of billions of registered users) worldwide are weighing in on the top social networking sites, swaying others, making decisions and forming relationships in a constant torrent of information.

While it’s true that we can analyze the complexity of networks, given time, the deluge of data is too massive, too complex and too time-consuming for current technology to sort in real time.

Which is why UC Santa Barbara professors of computer science Ben Zhao [2], Subhash Suri [3] and Heather Zheng [4], along with electrical and computer engineering professor Upamanyu Madhow [5], have teamed up for an ambitious project that not only aims to further understand social networks but also creates a means for analyzing them as they happen. It will provide a deeper comprehension of an increasingly "real" virtual world, as well as ways to monitor or prevent viral outbreaks, both in the real world and online, or track systems like transportation or biological protein networks.

Their project, titled Social Network Analysis: Geometry, Dynamics and Inference for Very Large Data Sets (SNAG-IT), was awarded a $6.2 million grant from the U.S. Department of Defense’s Defense Advanced Research Projects Agency (DARPA). SNAG-IT’s obvious challenge ? considering the
sheer size of the network and the enormous amount of information is unraveling data in real time.

To help Zhao and colleagues in this task, they have partnered with information technology giant Hewlett-Packard, the project’s primary contractor, which is researching and building scalable graph processing systems.

“When you look at Facebook, or LinkedIn or Twitter, you’re talking about networks of more than a billion people,” said Zhao, who leads the four-year project. Traditional algorithms developed and proved near-optimal decades ago no longer apply, he said. Developed for smaller sets of data, current algorithms scale poorly when the amount of data skyrockets.

To compound the problem, social networks are based on constantly changing relationships, which affects what kind and how much profile information can be seen by others. Meanwhile, some people gain popularity, others lose clout, and events have immediate impact on topics of cyberspace discussions and real-life decisions.

For example, a LinkedIn profile page could tip a company toward hiring a certain individual if his or her list of connections was populated with influential people in the industry. Conversely, an offhand comment, a change of profile picture, even a “like” could lose a user friends, connections or followers and therefore influence in the social network, and opportunities in real life.

The power of geometry

To understand this modern kind of dataset, the researchers are using an ancient system: geometry.

“Geometry is a powerful way of visualizing complex relationships,” said Suri, who specializes in computational geometry. User profiles can be plotted as points—nodes—on a coordinate space, with distance and dimension representing relationships, for instance. Other information deemed relevant can also dictate the node’s positioning or interaction with other nodes around it.

The group is also interested in teasing out data that is not explicitly mentioned from the flow of information: inferred associations from professional affiliations or shared skill sets, for instance, or implicit relationships from timing of events—not just the presence of a connection, but also its quality.

“We seek to develop a systematic framework for teasing out information from spatiotemporal patterns of activity on social networks,” Madhow said. “As one example, by correlating the timing and volume of activity with the timing of a class of external events, for example baseball games, it may be feasible to make inferences about a user’s interests, such as, is he or she a baseball fan? Furthermore, such inferences can be strengthened and extended by examining the patterns of activities for groups of linked users. As another example, by looking at the spatiotemporal spread of a rumor, can one make systematic statistical inferences about its source?”

But dealing with massive and rapidly growing amounts of sometimes seemingly disparate information is no small feat, and current technology does not have the power necessary to analyze such vast amounts of data at a meaningful speed. Even now, queries for profiles on current social media websites like LinkedIn, for instance, return precomputed, sometimes days-old information, which may or may not reflect up-to-the-moment developments, the scientists say.

As an example, Suri said, take GPS navigation systems, with main roads and side roads all plotted out in relationship to the driver’s coordinates, and measurements taken and relayed continuously via satellite as directions are sent to the driver while he moves from one location to the next.
Road networks are large graphs that people are just now getting comfortable with in terms of real-time response, he said. But road information relayed via GPS is minuscule and simple compared to the quantities that flow through networks like Facebook and LinkedIn, with profile views, public and private interactions, status updates, evolving relationships, responses to external events, timing of communication and constant changes over time.

?Current systems simply limit the power of the queries you can execute. LinkedIn for example does not let you query more than three hops away from yourself,? Zhao said. ?Others simply limit functionality. For example, you cannot yet search for all users on Facebook while sorting by social distance away from you. Enabling that would significantly improve your chances of finding friends you already know, especially those with common names.?

Breaking down complex data structures

Enter modeling and algorithms, meant to efficiently and elegantly describe and approximate behaviors; reveal elements like influential thought leaders and communities; and potentially even predict events, whether it?s the next Internet meme or the next Arab Spring.

At the same time, these complex data structures have to be condensed into as small a dimensional space as possible to allow for rapid computations while sacrificing the least amount of accuracy.

?We are going to have errors,? Zhao said, explaining that capturing a data structure with up to 100 dimensions or more ? depending on how comprehensive the social network is trying to be with its users ? into a small number of dimensions that can be visualized in a graph ?fundamentally just cannot be done perfectly.? Some nodes in the data may just be out of place, he said.

The intensity of information will also make it more difficult for people to lie about their cyberselves,
Zhao said, because even if a person changes his or her information in one sense, the other dimensions, relationships and inferences drawn from those associations still exist.

?In a practical sense, it?s very difficult to mislead the data in a meaningful way,? he said. ?Unless you move your location, change your job and change your circle of friends, that closeness with certain people or things will still remain.?

To evaluate and validate these algorithms and models, the group will be using preexisting datasets from previous projects, the largest of which is a 40 million-node graph of anonymized user profiles from a Chinese online network.

?It becomes a mathematical problem,? said Zhao, who specializes in modeling and mining massive graphs as well as analysis of social networks and Internet communities.

It also becomes a laborious process in which they take a ?brute force? approach to get to the ground truth: Run lengthy computations with the preexisting data and see how close they get with their algorithms. Computations for even a small, 20,000-node network can run for weeks.

Ultimately, however, the result will be powerful programs, applications and systems that can run fast, compute enormous amounts of data and do it with today?s machines, with all the physical constraints they face.

The research could also lead to uses in other fields. For instance, the high-speed computing and real-time capacity could be used to observe transportation systems and biological protein interaction networks. The algorithms would prove useful in the monitoring and possible prevention of viral outbreaks, both biological and online.

Such research into the complex dance of social media networks can provide a foundation from which social scientists might study a variety of behavior patterns and interactions in an increasingly ?real? virtual world.

**Latest Issues**

**Tag Cloud**

global warming conservation catch share system electromagnetics spectrum CAMP silicon semiconductors TOEC oil spill preservation Matthew Tirrell III-V LED NEMS parasites autism photonics neurogenesis detection bioengineering hydrocarbon brain Computer Science cells biomedical microRNA woman in mechanical engineering
Source URL: http://convergence.ucsb.edu/article/living-story-social-graphs

Links: