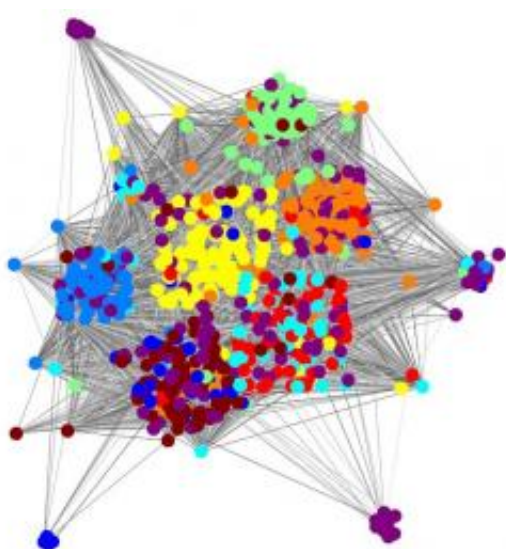# Untangling Facebook, decoding Congress: New mathematical method may help tame big data

May 13 2010

A visualization of a network of Facebook connections, from previous related research by Mucha and others. Credit: Amanda L. Traud, Christina Frost, UNC-Chapel Hill.

(PhysOrg.com) -- Networks permeate modern life, from Facebook to political allegiances. Now University of North Carolina at Chapel Hill mathematicians and colleagues have developed a new technique for examining networks to help identify patterns and see how connections evolve. A paper describing their research appears in the May 14, 2010 edition of *Science*.

One of the most prominent areas of network science is the study of what's called the "community structure" of a network. But until now, key methods could only detect "communities" (well-connected groups of nodes) in networks that don't change over time and only have one type of connection.

Of course, most networks in real life are more complicated, said Peter J. Mucha, Ph.D., associate professor of mathematics in the UNC College of Arts and Sciences and lead author of the paper. The new technique offers the ability to examine networks that vary over time and have multiple kinds of connections.

"It's 'connecting the dots' on steroids," Mucha said. "This method offers new potential for handling a fire hose of information, whether you're looking at an online social network or a real-world web of people or things."

Mucha and his colleagues derived their new method from mathematical principles and applied it to a few example datasets, including the complete historical roll call voting record in the U.S. Senate through 2008, and a set of Facebook profiles from almost 1,700 students at an anonymous American university including photo tags and housing information. Mucha said their community detection methodology identified some interesting details, including points of historical transition in the Senate and indications of different groups among Facebook users.

"Facebook is a good example of a tangled web of connections," he said. "Within it, there are groups of people who are more tightly connected to each other than they are to other groups. If you map out every individual 'friend' connection and trace one connection to another, you'll see some clumpiness to that network."

But a more complete analysis of the network would include information about the myriad of different types of connections. For example, by analyzing data such as individuals' profile details, photo tags, Facebook "likes" and recommendations and messages, it might be possible to identify other connections and groups that may be subtle or not explicitly obvious, Mucha said. (The paper in Science did not look at all such information)

The new method divides a network into multiple "slices," with each slice representing the network at one snapshot in time, or a different set of connections between the individuals within it. These slices are then combined and - by using a variety of computer algorithms - analyzed to identify communities.

Mucha's primary interest in network analysis is applying methodologies to real world data, including congressional relationships.

With the new community detection method, researchers should be able to dig deeper to examine the relationships among different groups in dynamic, multiplex data. Identifying community structures in a network might help to model processes and provides a signal about the underlying system, such as legislative polarization or the influence of various factors and forces, he said.

"Looking at the way legislators vote, it's usually easy to quickly group them into Republicans and Democrats, but that's really just a first pass at the data," he said. "Those legislators might be connected in many ways — the states they represent, who they've received political donations from, their caucuses or committee assignments, even where their offices are located in the building. Combining such information in a meaningful way helps us explore - and potentially make more sense of - legislative data."

Mucha believes another potential application for the new method is modeling the spread of diseases. He plans new research in that area.

Two UNC undergraduates - Thomas Richardson, class of 2008, and Kevin Macon, class of 2010 - are among the paper's co-authors, along with Mason A. Porter of Oxford University and Jukka-Pekka Onnela of Harvard University.

Provided by University of North Carolina at Chapel Hill