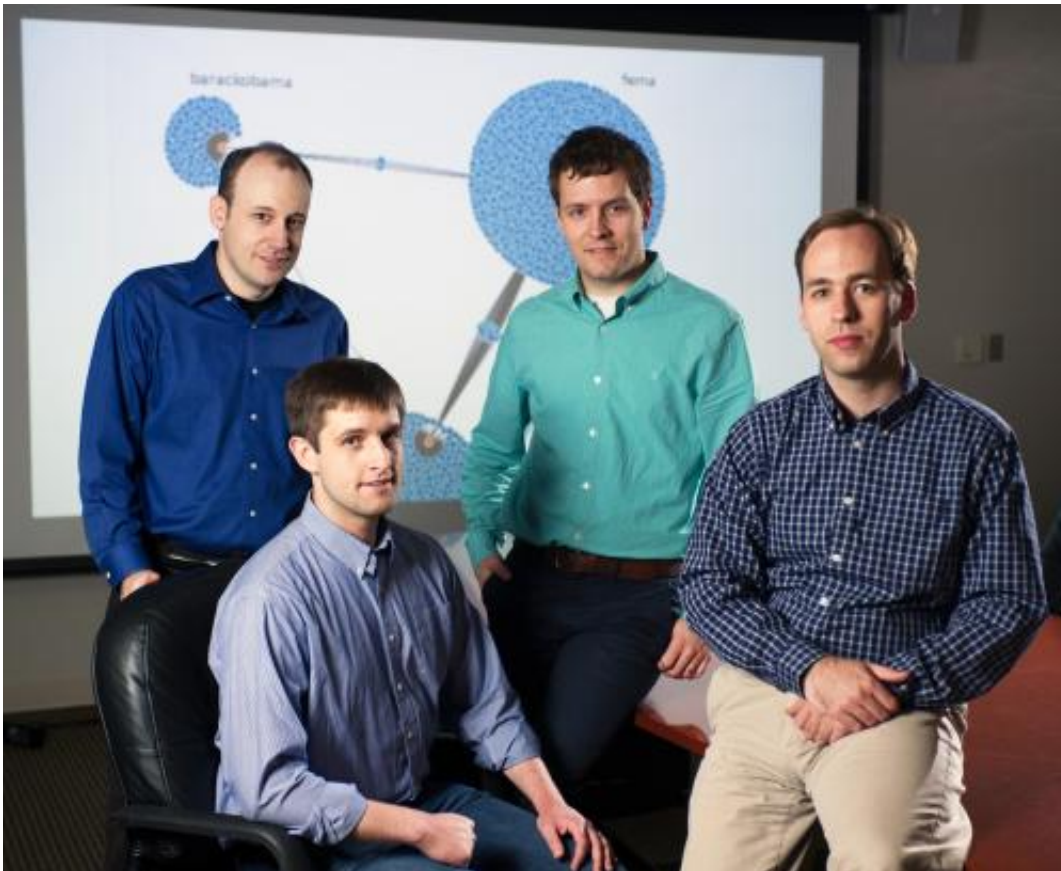# Dynamic graph analytics tackle social media and other big data

November 7 2014, by Rick Robinson

A GTRI team consisting of (left to right) Dan Campbell, Rob McColl, Jason Poovey, and David Ediger is bringing graph analytics to bear on a range of data-related challenges including social networks, surveillance intelligence, computer-network functionality, and industrial control systems.

Today, petabytes of digital information are generated daily by such

sources as social media, Internet activity, surveillance sensors, and advanced research instruments. The results are often referred to as "big data" – accumulations so huge that highly sophisticated computer techniques are required to identify useful information hidden within.

Graph analysis is a prime tool for finding the needle in the data haystack. This potent technology – not to be confused with simple illustrations like bar graphs and pie charts – utilizes mathematical techniques that represent relationships in the data more efficiently than traditional statistical analyses.

Researchers at the Georgia Tech Research Institute (GTRI) are bringing graph analytics to bear on a range of data-related challenges. They're developing advanced technology that can help investigate social networks, surveillance intelligence, computer-network functionality, industrial control systems, and more.

"Our first task is to look at the interesting properties of a graph – to find the important questions we can ask of that graph," said Dan Campbell, a GTRI principal research engineer who heads the High Performance Computing Branch. "The second task is to find the answers as quickly as possible, and then put them to practical use."

A graph is a type of data structure comprised of entities – meaning anything that can be represented digitally – and their relationships. In graph terminology, an entity is a vertex or a node; the connections between it and other vertices are edges or arcs. Graphs are constructed using software algorithms that represent both the data points and the relationships between them, and also enable computers to manipulate and analyze that information.

GTRI researchers make extensive use of a graph-analysis framework called STINGER, built specifically to tackle dynamic, ever-changing

applications such as social networks and Internet traffic. STINGER was created by a team led by David A. Bader, a professor in the School of Computational Science and Engineering; key members of that team included David Ediger and Robert McColl, who are now part of Campbell's GTRI group. STINGER, which is open-source software (STINGERgraph.com), continues to be developed at Georgia Tech and in the broader graph analytics community.

"We've done a great deal of work on analyzing openly available social media in real time," said Ediger."Social media analysis clearly has an important role to play in emergency response to both natural disasters like Hurricane Sandy and to potential terrorist attacks, and we're actively researching applications in those areas, among others."

STINGER helps support GTRI's focus on streaming or dynamic-graph technology, which can store very large databases and then update them in real time as new data come in. This novel approach allows users to monitor social media on a massive scale, and can also be utilized to simulate very large networks.

Georgia Tech researchers have presented this technology at several recent conferences including the 1st Workshop on Parallel Programming for Analytics Applications, which was held in February in Orlando, Fla., in conjunction with the 19th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming.

"Unlike traditional graph databases, STINGER's streaming-graph technology lets us store very big graphs and analyze them at high speed using fairly modest computing capability," said Jason Poovey, a GTRI research scientist in Campbell's group. "In half a terabyte of main memory – a pretty reasonable size today – we can handle billions of nodes and edges. Our benchmark tests show we can represent, update and analyze a graph in real time that's essentially the size of all the data

in Twitter."

GTRI is focusing on multiple efforts in which graph analysis plays a key role. These projects include:

- Behavioral Modeling and Computational Social Systems (BMCSS) Strategic Initiative – A GTRI team led by senior research scientist Erica Briscoe has used STINGER to study real-time social media analytics, as part of research aimed at predicting human behavior on a large scale.
- BlackForest – Members of Campbell's group are using graph analytics to support the BlackForest project led by GTRI researcher Chris Smoak. The aim of this externally funded project involves forming coherent intelligence pictures from disparate types of data obtained from multiple sources.
- Nextcache – This externally funded project focuses on developing new CPU, cache and memory designs tailored for graph-based applications.
- Real-time Business Intelligence – Using streaming graph technology, members of Campbell's group are working with GTRI researcher Erica Briscoe to develop a business-intelligence dashboard that monitors social media in real time and helps businesses gauge consumer sentiment.
- XDATA – Working with researchers from the School of Computational Science and Engineering, GTRI senior research scientists Barry Drake and Richard Boyd are helping to address big-data challenges by studying the computational demands of processing machine-learning algorithms.

Provided by Georgia Institute of Technology

Citation: Dynamic graph analytics tackle social media and other big data (2014, November 7)