

Contact  
Lynn A. Nystrom  
[tansy@vt.edu](mailto:tansy@vt.edu)  
540 231 4371

## **Researchers study new ways to forecast critical societal events**

Blacksburg, Va., DATE – University and industry scientists are determining how to forecast significant societal events, ranging from violent protests to nationwide credit-rate crashes, by analyzing the billions of pieces of information in the ocean of public communications, such as tweets, web queries, oil prices, and daily stock market activity.

The Institute for Critical Technology and Applied Science at Virginia Tech has assembled a team that is creating mathematical models, algorithms, and software to harvest public data, integrate relevant information from multiple sources, and generate alerts.

"We are automating the generation of alerts, so that intelligence analysts can focus on interpreting the discoveries rather than on the mechanics of integrating information," said Naren Ramakrishnan, the Thomas L. Phillips Professor of Engineering in the computer science department at Virginia Tech. He is leading the team of computer scientists and subject-matter experts from Virginia Tech, the University of Maryland, Cornell University, Children's Hospital of Boston, San Diego State University, University of California at San Diego, and Indiana University, and from the companies, CACI International Inc., and Basis Technology.

Within Virginia Tech, the team spans the departments of computer science, mechanical engineering, statistics, and agricultural and applied economics, the Virginia Bioinformatics Institute, and the Institute for Critical Technology and Applied Science.

The project is supported by a potential \$13.36 million three-year contract from the Open Source Indicators (OSI) Program of the Intelligence Advanced Research Projects Activity (IARPA), a research arm of the Office of the Director of National Intelligence. Three teams were awarded contracts, with continuation after the first year contingent upon satisfactory progress.

“Research shows that many significant societal events are preceded by population-level changes in communication, consumption, and movement. Some of these changes may be indirectly observable from diverse, publicly available data, but few methods have been developed for anticipating or detecting unexpected events by fusing such data,” said Jason Matheny, OSI Program Manager at IARPA. “OSI’s methods, if proven successful, could provide early warnings of emerging events around the world.”

Each OSI research team will be required to make a number of warnings/alerts that will be judged on their lead time, or how early the alert was made; the accuracy of the warning, such as the where/when/what of the alert; and the probability associated with the alert, that is, high vs. very high.

The Virginia Tech-led team calls its project EMBERS, for early model-based event recognition using surrogates. Surrogates are accessible pieces of information that mirror or precede events of interest. For example, one might assess the sentiments of tweets to assess economic indicators. Johan Bollen, associate professor of informatics with the Center for Complex Networks and Systems Research at Indiana University and an EMBERS co-investigator, has devised a way to evaluate the tone of tweets – calm, alert, vital, etc. -- to predict stock market trends.

The team intends to organize a huge database of surrogates predictive of real events and to apply these surrogates to public data sources.

The focus of the IARPA program is on Latin American countries. "Latin America is one of the most dynamic regions of the world. Changes are taking place without being adequately noticed by the rest of the world," said Dipak Gupta, a Distinguished Professor of Political Science at San Diego State University and EMBERS team member.

“Latin America is a great place to do this type of research because the region is one that is not overwhelmed by significant events, yet strikes, ethnic empowerment, and infectious diseases are occurring. The challenge of distinguishing the normal level of conflict in a democratic society from events that could signal threats to governance makes this project exciting,” said EMBERS

subject matter expert David Mares, Institute of the Americas Chair for Inter-American Affairs at the University of California, San Diego.

A key theme in the EMBERS project is the use of models to capture population-level behavioral changes. Tracking or identifying individuals is strictly excluded from the research.

"The models must be expressive enough to capture many important behaviors. For instance, how many people and what other factors result in a protest becoming violent? When do a few reported cases of dengue fever become an epidemic? But we do not want a model that is so complex that it becomes intractable. So finding the right balance is important," said Madhav Marathe, professor of computer science and deputy director of the Network Dynamics and Simulation Science Laboratory at the Virginia Bioinformatics Institute, and EMBERS co-investigator.

"When the model provides an alert, we want to be able to trace back to the tipping points, to ask 'what if ...' questions and change variables, to understand the elements of the prediction," said Achla Marathe, associate professor of agricultural and applied economics, also with the Virginia Bioinformatics Institute and an EMBERS co-investigator.

### **How will the models and alerts be used?**

The team builds upon many successful projects in modeling and prediction. John Brownstein, associate professor in the Department of Emergency Medicine and Informatics at Children's Hospital of Boston and an EMBERS co-investigator, created and manages HealthMap [www.healthmap.org](http://www.healthmap.org), an internet-based global infectious disease alert system. Chang-Tien Lu, associate professor of computer science at Virginia Tech and EMBERS co-investigator has created spatio-temporal knowledge discovery systems to model crime and traffic. Luis Rocha, associate professor of informatics and cognitive science at Indiana University, member of the Center for Complex Networks and Systems, and EMBERS co-investigator, has developed bio-inspired methods to predict associations in biochemical, social, and knowledge networks, including web and e-mail systems. EMBERS co-investigators Aravind Srinivasan and Jennifer

Golbeck, of the University of Maryland, College Park, have developed models of social sentiment evolution and of trust and distrust in social networks.

EMBERS co-investigators Achla Marathe, Anil Vullikanti, Madhav Marathe, Chris Barrett, Stephen Eubank, Bryan Lewis, and Jiangzhuo Chen, members of the Network Dynamics and Simulation Science Laboratory at Virginia Bioinformatics Institute, have created flu and pandemic models to capture disease spread in different settings. These models can be used to recommend containment strategies, such as who to vaccinate, whether to close schools, and what information the public needs. “Contagion processes can be broadly applied from infectious diseases to populist uprisings” said Vullikanti.

Although designed with specific contexts in mind, the underlying principles of these previous models can easily be adapted to accept other data streams and elements. “Given micro pieces of information on particular events, our research aims to develop models that can capture the macro level trends,” said Tanzeem Choudhury, associate professor of information science at Cornell University, and EMBERS co-investigator.

The university partners will create the models, with research and development into the software frameworks led by the industrial partners.

CACI International Inc provides professional services and IT solutions in the areas of defense, intelligence, homeland security, and IT modernization and government transformation.

According to Kristen Summers, CTO of the CACI Advanced Knowledge Solutions Division Group, “As the system integrator on this project, we will leverage our expertise in cloud computing technologies and data analysis tools to combine the academic models into an automated warning system for significant real-life events. CACI is proud to support Virginia Tech in this important intelligence effort.”

Basis Technology, headquartered in Cambridge, Mass., has a long history of providing customized, multilingual text analysis solutions to the Department of Defense and intelligence community. Jeff Godbold, Federal Solutions director at Basis Technology and EMBERS co-investigator, said, “Basis Technology is excited to support the (Open Source Indicators) program

with the use of our Rosette linguistics platform which will help identify the languages of incoming feeds, analyze the text, extract entities and manage the variations and ambiguities that naturally occur within human language.”

The new and huge benefit of the EMBERS project is that it will combine socio-economic-political-natural disaster-health information for more realistic alerts, said Ramakrishnan who leads the Discovery Analytics Center at Virginia Tech (<http://dac.cs.vt.edu>) and has had considerable success in creating systems for knowledge discovery.

"A significant portion of our effort thus goes into information integration, such as how to fuse predictions from diverse sources," said Lise Getoor, associate professor of computer science at the University of Maryland, College Park, and EMBERS co-investigator.

A broad range of techniques will be used for combining initial alerts, including Bayesian multi-modal sensor integration, the specialty of Scotland Leman, assistant professor of statistics and Michael Roan, associate professor of mechanical engineering at Virginia Tech, and EMBERS co-investigator.

“Our system can also be used to understand how policy interventions are likely to affect future possibilities. A purely data-driven or phenomenological approach is unlikely to generate such insights,” Ramakrishnan said.

“Extracting valuable information from massive data sets is the new frontier of computing. This project demonstrates the power of well-led interdisciplinary teams in developing new knowledge discovery and data analytics algorithms and systems to address important problems,” said Barbara Ryder, J. Byron Maupin Professor of Engineering and head of the Department of Computer Science at Virginia Tech.

“Large-scale analytics is considered to be one of the emerging technologies that will have transformative impact on lives. The capacity of this exceptional team to leverage analytics to generate automated alerts of events is symbolic of the vast potential for new markets that will

drive society's future," said Roop Mahajan, director, Institute for Critical Technology and Applied Science at Virginia Tech.

"Naren Ramakrishnan has established a powerhouse team of leading experts from academia and industry. This team will use its expertise to deliver rapid ways to arrive at solid analytical decisions and quantitative predictions to our nation's intelligence analysts," said Richard C. Benson, the Paul and Dorothea Torgersen Chair and Dean of Virginia Tech's College of Engineering. "Virginia Tech is honored to be leading such an accomplished group of investigators."

The team response to the IARPA solicitation was led by Jon Greene, director of National Security Research and Program Management at the Institute for Critical Technology and Applied Science. Christine Tysor at the institute will lead project management for EMBERS. "These individuals are invested in ensuring superior performance in all aspects of the formulation and execution of this project" said Mahajan.