UC SANTA BARBARA

THE Current

October 29, 2013
Sonia Fernandez

Of Big Data and Network Science

A message appears on a profile of a popular social network in a country suffering from political unrest. Seconds later it is repeated on another profile, and then another, and another. Within minutes the message spreads to hundreds, even thousands of profiles, from a core group of closely related users, to loosely affiliated contacts. Before the day is out, people are in the streets chanting for the end of the regime and those in power are seriously considering their next steps.

Somewhere else in the world, a child begins to sniffle and cough. Thinking it a minor bug, her mother sends her to school anyway, and then to day care after that. Within days the child's classmates are getting sick and by week's end a serious epidemic is underway, shutting down schools, slowing down business and causing flights to be canceled.

How did these events unfold? Where is the tipping point? Is it possible to predict the next event, and can they be controlled, prevented or in some cases engineered? These questions, and many like these, are what researchers at UC Santa Barbara are asking within the burgeoning field of network science.

"Network science is the study of how networks evolve, how they affect, how they are modeled," said Ambuj Singh, UCSB professor of computer science and also of biomolecular science and engineering.

The relatively new field of research also encompasses questions of how networks behave, what common elements they share, and how to visualize and interpret the

data in these networks –– much of which move and change over time.

With funding from a recently granted $3 million Integrative Graduate Education and Research Traineeship (IGERT) award from the National Science Foundation, Singh and several co-principal investigators from a wide range of disciplines have established a new graduate program focused on network science, and they are working to recruit and educate students into the growing field.

The goal? To prepare, train and mentor these graduate students to be effective and innovative researchers in the new era of Big Data, amid the growing realization that processes –– be they biological, computer or societal –– have effects that go beyond the individual units in which they occur.

"The systems view, we believe, matters much more than how a single unit behaves," Singh said.

Putting it to the test

Under the leadership of Singh; computer science professors Subhash Suri and Divy Agrawal; sociology professor John Mohr; Stephen Proulx, professor in the Department of Ecology, Evolution and Marine Biology; and 15 more participants in a wide array of disciplines, 25 graduate fellows will undertake courses of study over the two years of their fellowship.

Over the course of the program, starting in the fall of 2014, students will be immersed in classes, seminars, summer internships and workshops relevant to their areas of emphasis but using the wide-ranging perspective of network science. These students will rotate through three different modules involving Big Data problems from different domains, and with different faculty members. At the same time, they will be learning the necessary professional skills to put their knowledge to work, particularly in the areas of Big Data, interdisciplinary science, problem solving and innovation.

"Based on the anecdotal evidence, my sense is that the job prospects in the field of Big Data are excellent and this is only going to get better," Agrawal said. "Our NSF IGERT proposal to train students in the area of network science and Big Data is exactly what the industry would like us to do. Availability of trained data scientists will ensure that the United States maintains its competitive edge."

Additionally, the program aims to recruit a cohort of 70 percent women and underrepresented minorities from UCSB and through active engagement with four Hispanic Serving Institutions: California State University, San Bernardino, California State University, Los Angeles, UC Merced and the University of New Mexico.

"Today we live in a cloud of complex connectedness that is critical to understand as we attempt to solve the critical issues facing society today," said Cynthia Stohl, professor in the Department of Communication. "Developing competencies in network conceptualizations and analysis will open the doors to a multitude of professional opportunities that were never before available, especially for women and underrepresented minorities."

Now is the time

For Mohr, a sociologist, network science isn't an entirely new idea but perhaps one whose time has come.

"The phenomenon of people being connected to each other is really foundational to how social groups work," he said.

The difference for him and the other investigators is the unprecedented availability of data at their fingertips. Thanks to the Internet, documents from archives all over the world are accessible, as is real-time data from social networks, which enable one to follow the evolution of events or to uncover relationships that have widespread impacts, like terrorist attacks or large-scale protests.

However, the speed-of-light velocity of information comes with speed-of-light spread of misinformation as well, and this is one of several types of problems Mohr and graduate student researchers in his field might explore.

"While network science is not a new mathematical field, its scope, relevance and importance have grown significantly in the last two decades," Suri said.

Engineered networks –– the Internet, telecommunications and social networks, for instance –– are large-scale complex systems that have become in themselves important objects for network science. Their design, organization, behavior and control pose new types of questions and require new algorithmic paradigms, he said.

A goal of network science is to develop general insights and tools that can be applied across many, seemingly disparate networks.

According to Suri, a particularly daunting challenge in studying these systems is their scale (networks spanning billions of "nodes") and dynamics (changes in connectivity or interaction over time), both of which cause a "combinatorial explosion" in search of optimal solutions. Taming this complexity and designing computationally feasible algorithms is a major focus of research in Suri's lab.

Looking within

Perhaps nowhere else is the idea of network science more readily observable than it is in our own bodies and in our natural environment, where the existence and health of one unit –– be it a species, organism, body part, tissue or cell –– rely on the functioning of the units around it. For biologist Proulx, one of the major questions facing today's life scientists has been around for a long time.

"One of the challenges for people looking at biological networks is understanding the changes in the systems over short or long time scales, and how that depends on the different components of the network," he said. Think effects of ocean acidification on the various ecologies of the Earth, for instance, or the prolonged influence of the environment on genetics. Studying individual, isolated units has been less than adequate in finding patterns over time, but network science could light the way.

"The challenge has been to try to understand how the structure of a network relates to how that network functions," Proulx said. "So taking a more holistic network sciences approach –– combining tools from computer science and data sets from biology –– shows some promise in terms of figuring out how to answer these kinds of questions."

While funding for the program spans five years, the graduate emphasis is expected to continue and grow from the momentum generated by this initial IGERT award.

"National Science Foundation IGERT awards open doors for an innovative graduate research experience," said Rod Alferness, dean of the UCSB College of Engineering.

"This award is focused on preparing a generation of Ph.D. students to address the area of Big Data –– one of our major research thrusts in engineering and in the Department of Computer Science," Alferness said. "This is a cross-campus initiative that goes beyond just engineering and science to include sociology, biology, geography and communications, to provide unique opportunities for UCSB graduate students."

About UC Santa Barbara

The University of California, Santa Barbara is a leading research institution that also provides a comprehensive liberal arts learning experience. Our academic community of faculty, students, and staff is characterized by a culture of interdisciplinary collaboration that is responsive to the needs of our multicultural and global society. All of this takes place within a living and learning environment like no other, as we draw inspiration from the beauty and resources of our extraordinary location at the edge of the Pacific Ocean.