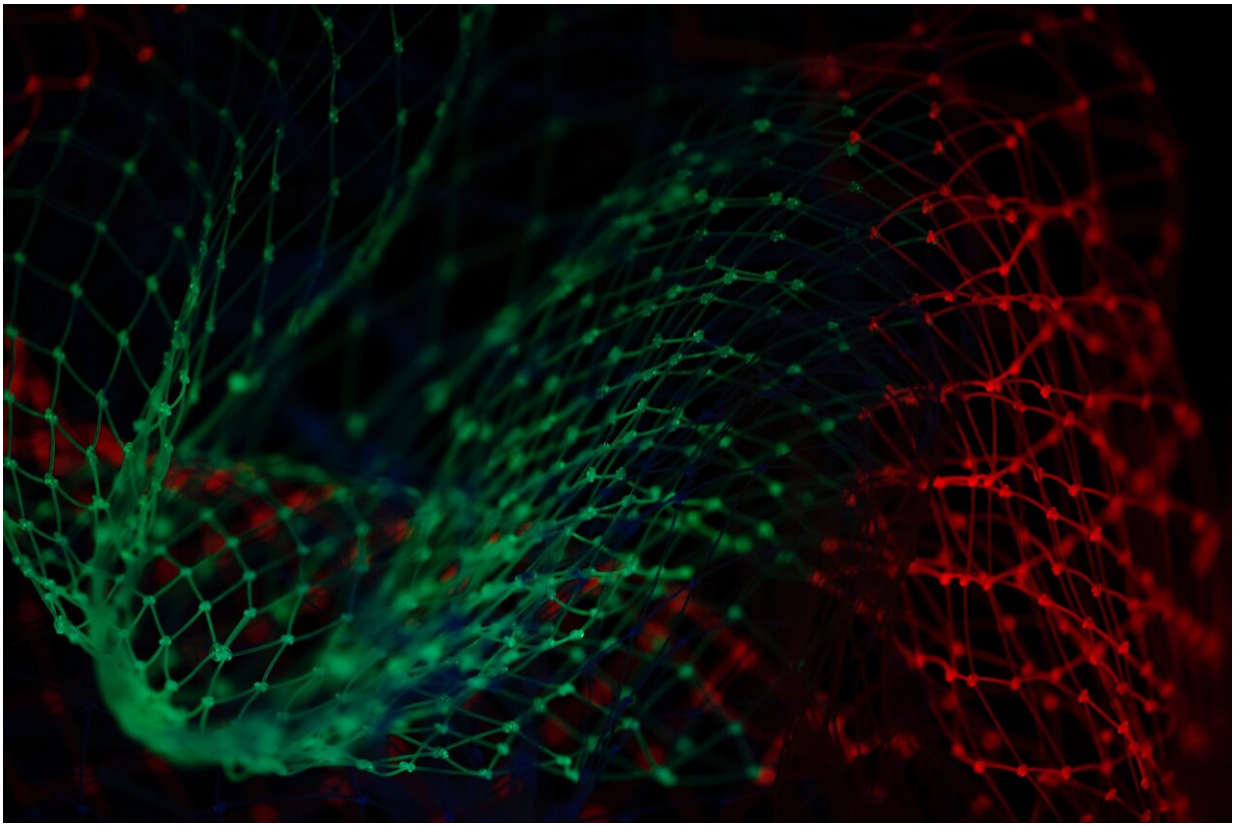
Researcher develops better tools for understanding, protecting big data

April 5 2021, by Kim Horner



Credit: Unsplash/CC0 Public Domain

Patterns and anomalies in big data can help businesses target likely customers, reveal fraud or even predict drug interactions. Unfortunately, these patterns are often not easily observable. To extract the needles of

useful information out of haystacks of data, data scientists need increasingly powerful methods of machine learning.

Dr. Aria Nosratinia, the Erik Jonsson Distinguished Professor of electrical and computer engineering at The University of Texas at Dallas, has received two grants from the National Science Foundation totaling $749,492 to uncover relationships hiding in big data via machine learning and to develop methods to keep data communications safe.

"The contribution of my lab is to expand the universe of tools and techniques so we can discover new connections in the data," said Nosratinia, who is associate department head of electrical and computer engineering in the Erik Jonsson School of Engineering and Computer Science.

Many machine learning and data mining algorithms use graphs, which are simply lists of connections between people, groups or objects. Examples include "friend," "like" or "follow" relationships in social networks, or the list of videos streamed or marked as favorites in a streaming subscription service.

These mountains of data hide useful information whose extraction belongs to an area known as graph inference. Graph inference has many interesting and useful applications—for example, suggesting movies in a streaming service based on viewing history or purchasing suggestions in online shopping. It also can reveal patterns in the spread of epidemics, or provide insights into the folding of proteins, which is important in understanding how proteins function.

Nosratinia's work for the first time proposes and analyzes techniques to improve graph inference by absorbing nongraph information, whose efficient blending with graph information was previously not well understood. Examples of non-graph information include a person's age

and residence ZIP code, which are individual attributes.

"In almost every practical application involving graphs, there exist nongraph data of great relevance," Nosratinia said. "The kind of work we do is further upstream, developing the mathematical models, theory and techniques, but it has widespread applications."

In several published works, Nosratinia describes the mathematical models he and members of his lab have developed that can improve the estimation of the information hidden in the graph with the aid of side information. Nosratinia and co-author Hussein Saad Ph.D."19, now a senior engineer with Qualcomm Inc., recently analyzed how to identify a small cluster or community hidden in a large graph. Their latest work appeared in the December 2020 issue of the journal IEEE Transactions on Information Theory.

The second component of Nosratinia's research addresses data security. His work harnesses the natural variations of wireless channels to provide layers of security for data transmission. This area of work, known as physical layer security, aims to leverage the imperfections of the communication channel as a tool for security. Part of this research is aimed at developing techniques for making the presence of electronic communication undetectable to cybercriminals.

"To give a simple example, a password works by leveraging the difference between what is known by a legitimate user versus cybercriminals who want to steal information," Nosratinia said. "Our work creates, amplifies and analyzes statistical asymmetry of information against adversaries in ways that do not involve passwords or keys, and uses them for securing communications."

More information: Hussein Saad et al. Recovering a Single Community With Side Information, *IEEE Transactions on Information*

Theory (2020). DOI: 10.1109/TIT.2020.3030764