

Networks and Links

■ LISE GETOOR

Sometimes the best way to understand something is to see how it relates to others. Lise Getoor, an assistant professor of computer science and member of UMIACS, specializes in studying networks or how entities—whether people, events, research projects, or points on a map—are connected to each other. She examines how these connections can be used for machine learning and probabilistic reasoning.

“Traditionally, people have looked more at individuals instead of the links among them,” says Getoor. “We can label individuals based not only on their attributes but on what they’re connected to.” In other words, Getoor studies relationships. The work of “link mining” to extract information can be looked at in an abstract, mathematical way as well as in very practical terms. Getoor has developed algorithms for classifying all manner of data.

Exploring links in an academic context, she has analyzed how research publications are connected to each other through their authors and citations. From such bibliographic data, one can identify collaborative groups—one type of social group. Going a step further,



Networks and Links

■ LISE GETOOR

To contact any researcher in UMIACS, go to www.umiacs.umd.edu.

Getoor has looked at how social relationships among academics, deduced from evidence such as how professors are grouped in committees organizing a conference, affect individuals' social capital—and, in turn, publication rates. The same kind of analysis can be used to explore relationships between lawmakers and lobbyists, she says.

In another sort of link mining, collaborating with researchers at Carnegie Mellon University, Getoor has looked at the hyperlinks among the Web sites of professors and students. It turns out that one strong sign that a Web site belongs to a professor is that it does not include links to the Web sites of other professors! This is just one example of how one can deduce the roles of individuals by looking at the types of interactions they have. In another example, Getoor has analyzed how e-mail history reveals people's roles in a company as well as the professional and social groups to which they belong.

In addition to link mining, one of Getoor's main interests is entity resolution, or developing tools to deduce

whether two separate mentions of a person, place, object, or event refer in fact to the same thing. For example, are references to a Joe Smith and another to a Joseph Smith talking about the same person? Entity resolution is also important in geography, and Getoor is collaborating with the U.S. government's National Geospatial-Intelligence Agency to go through databases of locations and peg which mentions are redundant, referring to the same underlying place. For example, the agency receives updates from local governments and needs to distinguish whether new information refers to a place meriting a new entry or refers to an entity that already exists in the database, if under a slightly different name or mismatched coordinates. "It's common for names or coordinates to be off," notes Getoor.

Whether dealing with people or geographical locations, entity resolution requires algorithms that assess the commonalities and distinctions between two records. Clearly, if two entities' attributes and relationships do not overlap extensively, they are more likely to be distinct rather than duplicates.

Getoor speaks about concrete problems but also about the abstract essence of those problems. "I do like the theory and algorithms as well as the compelling applications," she says. At the level of theory, she talks about statistical relational learning and link mining driving machine learning, artificial intelligence, and reasoning under uncertainty. At the

level of applications, she talks about entity resolution, group discovery, and role discovery.

"Lise has a unique way of breaking down problems and teasing out the essence of a question," says Chris Diehl, a senior research scientist at the Johns Hopkins Applied Physics Laboratory who has collaborated with Getoor.

For her research, Getoor receives funding from the National Science Foundation as well as a consortium of intelligence agencies. "I'm also interested in how all this relates to privacy and what you can say about the privacy you can guarantee," she says. In some cases, such as with health or financial records, the goal is to guarantee that no one can know that two records refer to the same person. In other cases, distinguishing entities can be helpful—for example, to thwart financial identity theft. In early 2007, Getoor received a Google Research Award for her work on entity resolution.

Ultimately, Getoor's work revolves around graph identification—or taking noisy, redundant data and identifying the true underlying network of nodes and links. Getoor works to distinguish, label, and rank nodes by better assessing, categorizing, and inferring the links among them, at the same time always asking exactly how much and what kind of data is necessary to get accurate and reliable analysis.

— Profile written by Karin Jegalian