Kelly Servick | Stanford Engineering

Assistant Professor Jure Leskovec uses information collected from sites like Twitter, Wikipedia and Facebook to tackle big questions about how society works.



For the more than two billion people who are now online, the Internet is an answer factory. The right keywords and navigation skills can unlock more data than any mind could hold.

But for <u>Stanford</u> [1]Computer Science Professor Jure Leskovec, the most interesting answers are under the hood, in the data of our online habits. His research uses information collected from sites like Twitter, Wikipedia and Facebook to tackle big questions about how society works.

Leskovec is exhilarated by the possibilities hidden in the data. When he talks about his work, he speaks a mile a minute, as if he can't describe all these possibilities fast enough. As one of the youngest members of computer science faculty, Leskovec says he feels fortunate to be at Stanford, where collaborations with Silicon Valley create what he calls a "let's-do-something-cool-together" mentality.

In 2011, he was named one of "Al's 10 to Watch" by Intelligent Systems magazine for his work designing computer systems that navigate the web and mine its information. This year, he was selected for an Alfred P. Sloan Foundation Fellowship, which provides a two-year research grant of \$50,000 to promising early-career scientists.

### **Signing On to the Network**

None of these possibilities could have crossed Leskovec's mind when, as a 12-yearold in Slovenia, he saved up about \$150 to buy his first computer. While still in high school, he built a text-to-speech system dubbed "best innovation for the disabled" by the government of the Republic of Slovenia.

As an undergraduate, Leskovec made connections with institutions in the United States. He spent two summers as an intern in Silicon Valley, and recalls being astounded by the abundance of technology headquarters while driving down Highway 101. For a student in Central Europe, these big-name companies had

Published on Wireless Design & Development (http://www.wirelessdesignmag.com)

seemed mythical and out of reach, "like things from the sky," he said.

During his doctoral research at Carnegie Mellon University, Leskovec turned his eye toward social networks. With the rise of the Internet, he saw that computer science was not just about engineering new systems, but also about using online data to test all kinds of hypotheses.

"The Internet is really where computer science becomes a science," he said. In Leskovec's view, the links we click, the news we share, and the votes we cast online represent raw material waiting to be organized and explored to identify fundamental patterns of human behavior.

The scale of that raw material is mind-boggling: hundreds of millions of interconnected Facebook accounts, three billion tweets posted to Twitter in the course of month. One of Leskovec's recent studies mines a collection of six billion news articles and blog posts collected daily for the last four years.

Some of this data is "crawled" from the Internet, meaning that a program systematically browses the web and collects relevant information. Other times, companies like Facebook, Twitter and Microsoft provide data to Leskovec and his colleagues, hoping to gain useful information about how people make sense of the Internet and social media. For example, Facebook collaborated with Leskovec's research group to find a method for predicting which friends Facebook users would add next and implemented it as the "people you may know" feature.

#### A Amall World after All

Leskovec often uses Internet data to test sociological models and theories of the past. In the 1960s, Stanley Milgram – the same psychologist who studied authority structures – devised an experiment to test the interconnectedness of strangers. He mailed letters to a group of people in Nebraska, hoping these letters would eventually all end up in the mailbox of one person in Massachusetts.

The letter instructed recipients to forward it to an acquaintance whom they thought could get it one step closer to the unknown Bostonian stockbroker. One average, the chain from Nebraska to Massachusetts took six steps — the source of the famous "six degrees of separation" idea.

In 2008 Leskovec used the largest social network of that time, the communication network of 240 million users of Microsoft Instant Messenger and verified that it is a small world after all --- in the Messenger network the average degree of separation is 6.6. However, in original Milgram's experiment people couldn't see beyond their immediate acquaintances, so they had to make inferences about how best to direct the letter. Milgram called this the "small world problem."

In the digital age, this small world plays out in web browsers, Leskovec says. When we navigate in search of information, we are flying blind, just like the people in Milgram's experiment. We can only see the links available from our current page, and we don't always know how to get to what we need.

Published on Wireless Design & Development (http://www.wirelessdesignmag.com)

Leskovec saw this scenario as a chance to learn about how people look for things. One of his students created a game called Wikispeedia, in which a player starts at one Wikipedia article and tries to navigate to a random target article just by clicking links. For example, if you started at "Mozart" and had to get to "The Terminator," one possible path would be to click the link from "Mozart" to "Austria," then "Schwarzenegger," which would lead you to your target.

When people try Wikispeedia, they play the "six degrees of separation" game, strategizing about how the ideas might be connected. Leskovec says this works sometimes, but often leaves us floundering many miles (clicks) from our destination. He designed a computer agent that plays the game much more efficiently by deciding which links are likely to lead to more useful pages.

Using the data from 40,000 Wikispeedia games, he is working on programs that predict what people are trying to find based on the way they navigate and whether they will give up searching for the target. This could lead to technology that recognizes when you're not finding what you need, and nudges you in the right direction.

The enemy of my enemy is my (Facebook) friend?

Mining social networks led Leskovec to explore an even older theory about how people judge one another. An Austrian psychologist in 1950s proposed that relationships between friends and enemies tend to have a certain balance, as expressed in the proverb, "The enemy of my enemy is my friend." But Leskovec said the only real evidence behind the balance theory were diplomatic case studies, like the forming of alliances before World War I. He wanted to transform the theory into something computable.

He collected data about how people evaluate each other on Wikipedia, Epinions and the programming forum StackOverflow. On these sites, users express approval or disapproval either by explicitly trusting or distrusting someone's online reviews, voting to promote a user to an administrative position or by rating the usefulness of a user's advice.

Leskovec found that in the real world, balance theory was not very good at predicting positive and negative evaluations. People are just not consistent about how their "friends" and "enemies" are aligned, he said.

But the data suggested a more useful model, one based on the ideas of status and similarity.

Leskovec found that people tend to judge others by comparing themselves to them. It is as if an evaluator compares herself to the target and then based on the relative difference in status makes the evaluation. However, there is also a second effect. People who were highly similar – who wrote Wikipedia articles on the same subjects and who reviewed similar products or answered similar questions – tended to judge each other favorably regardless of status differences. But in the absence of such similarities, people resorted to notions of status – the number of articles or reviews

Published on Wireless Design & Development (http://www.wirelessdesignmag.com)

a user has written, for instance - to make their judgments.

Using this data, he can now judge with 90 percent accuracy how one person will evaluate another. And these predictions about our judgment process, which might often be subconscious, hold true across all kinds of networks – not just the ones he used in the experiment.

#### **Tracking News and Making News**

Leskovec and his research group are continuing research of social networks on the Internet. Their most recent projects aim to figure out how news and other information spreads through networks. He and his graduate students are building a system that automatically creates "a London underground map of news," in which the trajectory of a news story is like a subway line transecting other news stories as it moves across the network to reach new readers.

Their studies show that such maps complement today's search engines and are extremely useful when one want to understand a complex topic like the recent European debt crisis, U.S. health reform or the events of the Arab Spring.

He is also interested in modeling and predicting how different stories compete for attention, and what factors motivate people to share a story with others. His findings could help advertisers reach potential buyers, but could also lead to new ways of aggregating and summarizing news, helping people stay informed and consume news more efficiently.

Recently, Leskovec has begun extending his reach beyond Internet data alone.

"We're working with two million court cases to build models that help judges make decisions and with a set of 10 million medical records to better understand interactions between drugs, conditions and procedures," he said.

On the question of how his research will be applied in the business world, Leskovec said he is aware of concerns about how our online habits are recorded and monetized. He admits that the way some advertisers obtain information is "a cause of serious concern," and is especially concerned about sites that store data called "cookies" in web browsers to track a single user across multiple sites.

But Leskovec sees his research as fundamentally different. He looks at the habits of millions of anonymous users within a single web community. In his view, the users who choose to be a part of this community know that their activities are being documented online and mined for insights.

Along similar lines, Leskovec and his group have established a number of active industrial collaborations and work with companies ranging from Twitter and LinkedIn to Samsung and Volkswagen, applying models to their data to gather insights. Likewise, he views starting a company as a distinct possibility. Leskovec sees it as "betting on a horse" in order to make a bigger impact in the industry.

Published on Wireless Design & Development (http://www.wirelessdesignmag.com)

For now, however, he wants to focus on all the interesting questions his terabytes of raw material have to offer.

"I can think of the web as a telescope, or a sensor into human life," he says. "What wasn't visible before is visible now."

Kelly Servick is a science-writing intern for the Stanford University School of Engineering.

For more information visit <a href="http://engineering.stanford.edu">http://engineering.stanford.edu</a> [1]

Assistant Professor Jure Leskovec (Photo: Andraž Kavčič)

#### Source URL (retrieved on 12/08/2013 - 2:21pm):

http://www.wirelessdesignmag.com/news/2013/04/big-questions-big-data

#### Links:

[1] http://engineering.stanford.edu