

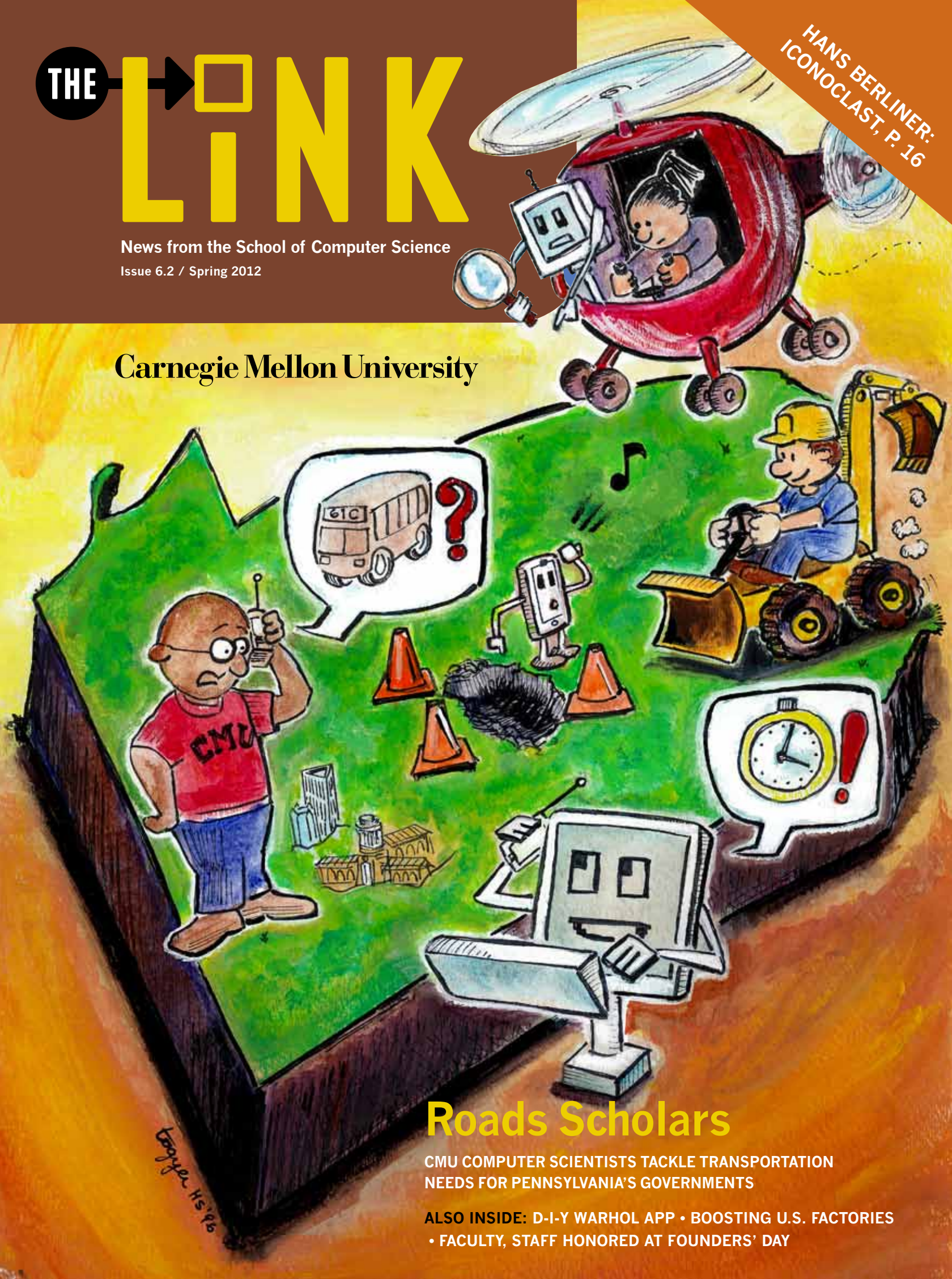
THE

LINK

News from the School of Computer Science
Issue 6.2 / Spring 2012

HANS BERLINER:
ICONOCLAST, P. 16

Carnegie Mellon University



Roads Scholars

CMU COMPUTER SCIENTISTS TACKLE TRANSPORTATION
NEEDS FOR PENNSYLVANIA'S GOVERNMENTS

ALSO INSIDE: D-I-Y WARHOL APP • BOOSTING U.S. FACTORIES
• FACULTY, STAFF HONORED AT FOUNDERS' DAY

Carlos

Carlos Guestrin is Finmeccanica associate professor of machine learning and computer science and co-directs the Sense, Learn, and Act (Select) Lab with Geoff Gordon. A graduate of the Polytechnic School of the University of São Paulo, Brazil, Guestrin earned his Ph.D. in computer science from Stanford University in 2003.

His main research interest is in developing efficient algorithms and methods for designing, analyzing and controlling complex real-world systems.

A member of the DARPA Information Sciences and Technology advisory group, Guestrin was the recipient of an NSF CAREER Award and the Presidential Early Career Award for Scientists and Engineers and was twice awarded an IBM Faculty Fellowship. In 2008, Popular Science magazine named Guestrin to its “Brilliant 10” list of the “brightest researchers” in the United States.

He spoke to Link Managing Editor Jason Togyer.

Your undergraduate degree is in mechanical engineering. How did you wind up in computer science?

I wanted to do something that seemed “futuristic”—even though the word “futuristic” now seems retro—so one day I went into my professor’s office and said, “I’m really interested in robotics.” He said, “Well, I work in computer vision, maybe you want to work with me?” I did wind up working on a robot, and it was a very interesting experience, but at the end of my undergraduate years, I decided I wanted to do something more theoretical for my Ph.D., and artificial intelligence was very interesting to me.

Why not stay in robotics?

I switched to AI because I was building things that demonstrated “intelligence,” but I didn’t understand where it was coming from. AI also allows me to examine the frameworks of why we perceive intelligence.

What is “perceived intelligence”?

In even the simplest things, such as a spam filter, we might say “Oh, that was smart how it figured out that message was spam.” But the

Guest

JOHN BARNA PHOTO



Carlos Guestrin

way an algorithm does spam filtering is not the way that a human being would do spam filtering. Although we may work with mathematical models, not rule-based systems, our challenge is to make the right mathematical models—the right assumptions.

What led to your work on the Cascades algorithm?

When I finished my Ph.D., I had built a theoretical foundation for working on AI, but I was missing an application domain. I went to Intel for a year with a group that was deploying sensors in a forest to understand a microclimate. In such a situation, you want to capture the data—say, water contamination—as close to the source as possible, but sensors are expensive. I decided there must be some way to balance the cost of collecting the information with the need for putting out enough sensors. We developed what we thought was a really nice theory that evolved into the Cascades algorithm, which you can use for a variety of applications where you might need to develop a sensor network.

How did you apply the Cascades algorithm to analyzing the spread of news?

A blog is kind of like a sensor. It's trying to capture a story as early as possible. Two students working with me, Andreas Krause and Jure Leskovec, said, "Where else would the Cascades algorithm be

useful?" One of them said, "Well, I have this blog data to analyze, maybe it would be useful there." We applied the same algorithm to the spread of information on the web, and it turns out that the way stories spread in the blogosphere is very much like contaminates spreading through water. We were able to identify the top 100 blogs that report news stories as early as possible.

What's the practical application of knowing that?

Information overload is perhaps one of the most important sensing problems of the coming decade. Ten years ago, we were already talking about the explosion of the Internet. Who knows where we're going to be 10 years from now? And this is not just an issue with the Web—it's an issue with the scientific process, with the political process. It's a very pressing challenge, and computer science is able to deal with this kind of problem.

Can't we just use search engines to filter information?

Right now, when you look for information on the Web, you do a keyword search. You might get 10 results and if you're not happy with them, you change your keywords. It's an iterative process. What are better ways to look for information besides changing your queries? Then there's another problem—I may not know which

sources to trust. I'm just overwhelmed from every direction.

How can computer science address those problems?

Think about the economic crisis triggered by the collapse of AIG, and about the health care debate. There must be a way those two stories are connected, right? One way to find those connections is to get an article on each topic and find the shortest path connecting the two. If you do that now, you'll find strong but superficial connections—it's like a stream of consciousness, or a conspiracy theory generator. What we're trying to do is give you more comprehensive information about how a connection comes about.

We're also working on a way to use networks to determine the trustworthiness of sources. There are many ways to give feedback on what to trust, and what not to trust. You can look at who else cites a particular source, for instance.

Another area of our research is ways of suggesting new sources of information. We often have a very biased perspective—we may go to the same websites all the time, or maybe we always read the same influential researchers' papers. We don't really have a good way to learn which things we don't know about. Maybe our model can suggest other people or websites we might be interested in.

By analyzing those things you're interested in, I can get a very good sense of your biases and build a model and push you to discover things that you don't know. There's a real opportunity to help all of us be exposed to multiple points of view.

You create paintings, sculptures and collages. Why are you so passionate about art?

Artists typically use their art for emotional expression, and there's definitely a joy to having a creative outlet—although my work here is also a creative outlet. CMU has a lot of very interesting art-related activities, both in our department and across departments, and I think it's really nice to be in an environment such as this. Recently I taught a class with Osman Khan, a visiting art professor, and that was very exciting for me. When I was deciding what to do career-wise, I considered pursuing a career in art—and now I say, if CS doesn't work out, I can always fall back on art!