

*Searching for  
a Heavy Tail*

## MARIANA OLVERA- CRAVIOTO

Assistant Professor of Industrial  
Engineering and Operations Research

Search Google and within a fraction of a second it will return a list of the most popular websites on the subject. Or will it? Mariana Olvera-Cravioto has been trying to answer that question by understanding what makes a website popular on Google.

The principles behind Google's page-ranking system are well known. It weighs links to and from a page, as well as links of other pages on the same website. Yet the details remain unclear. For example, what counts more, a few links from such important websites as Wikipedia or Technorati, or many links from less significant pages? And what happens to the rankings if you change Google's search algorithm ever so slightly?

To probe those questions, Olvera-Cravioto relies on a form of probabilistic theory called heavy tail theory. To understand it, consider a normal bell curve. Most samples are grouped close to the center, or mean, and decline rapidly towards the ends of the curve.

Heavy tails have far more outliers at the ends of their curves than normal distributions. They are surprisingly common. They show up in the distribution of wealth (few people own more assets), oil reserves (a few have the most value), insurance payouts, and the time supercomputers spend completing tasks.

"Internet video transmission is an example of a heavy tail distribution," she said. "Streaming video only transmits pixels that change. Most of the time, that's relatively few pixels. But then the camera changes angles and the whole screen is refreshed. It happens less often, but accounts for most of the transmitted data." Google search results also have a heavy tail distribution.

"The mathematical techniques needed to solve heavy tails are completely different from what we use with well-behaved distributions," Olvera-Cravioto said. Those techniques provided some deep insights into Google's page rankings.

"Before we started our analysis, it was not obvious what determined the relative rankings of websites," Olvera-Cravioto said. Heavy tail analysis, for example, shows how large numbers of links outweigh important links in popularity rankings.

"Once you understand how it works, you can engineer search algorithms for specific purposes," she said. "Maybe you want to rank stores by number of sales rather than links, or measure the importance of a paper by how many times it is cited by reliable websites. Adding these things to a search algorithm could make it easier to find the page you're after."

*B.S., Instituto Tecnológico Autónomo de México, 2000; M.S., Stanford, 2004; Ph.D., Stanford, 2006*

