Nevanlinna Prize

Jon Kleinberg

Jon Kleinberg's work has brought theoretical insights to bear on important practical questions that have become central to understanding and managing our increasingly networked world. He has worked in a wide range of areas, from network analysis and routing, to data mining, to comparative genomics and protein structure analysis. In addition to making fundamental contributions to research, Kleinberg has thought deeply about the impact of technology, in social, economic, and political spheres.

One of Kleinberg's most important research achievements focuses on the network structure of the World Wide Web. His insights have greatly influenced how all of today's major search engines operate. While individual web pages have a degree of structure imposed by the creators of those pages, the structure of the World Wide Web as a whole is completely unplanned; the structure is continually emerging as new content and links are added. As a result, it is a challenging problem to design effective ways of selecting relevant web pages to offer to search-engine users. Prior to Kleinberg's work, search engines focused only on the content of web pages, not on the link structure. In 1996, Kleinberg introduced the idea of "authorities" and "hubs": An authority is a web page that contains information on a particular topic, and a hub is a page that contains links to many authorities. For example, consider the search query "digital camera". A web site offering guides and reviews of digital cameras would be a hub, and the web site of a manufacturer of digital cameras would be an authority. Although the idea of hubs and authorities is clear intuitively, defining these notions mathematically is difficult because of the circularity that arises: A hub is a site that points to many authorities; an authority is a site with links from many hubs. Kleinberg's key contribution was to figure out a way to break this circularity, thereby opening the way for a mathematical analysis of the link structure of networks. He developed algorithmic tools for this problem and demonstrated their effectiveness by testing them on the World Wide Web.

Another area in which Kleinberg has made fundamental advances is in that of "small-world" networks. These networks were first noticed in experiments carried out in the 1960s by the psychologist Stanley Milgram and became the focus of a series of mathematical models based on work of Duncan Watts and Steve Strogatz in the 1990s. In Milgram's experiments, a person A would be asked to forward a letter to person B, whom A does not know personally, in as few steps as possible. Each person in the chain could forward the letter only to someone he or she knew personally. Milgram observed that on average the letter would reach B after being forwarded by 6 people (this is the origin of the phrase "six degrees of separation"). The number 6 is surprisingly small, given the size of the population overall, and given that

each person operated on the basis only of local information about the social network. Up through the 1990s, there was a good deal of research in the social science literature about the existence of short paths between individuals in social networks. But there was little consideration of how to find short paths given only local information. How do people who know only their immediate friends and do not have a a global view of the whole network find shortest paths? It is this question that Kleinberg addressed, and he found some surprising things.

The nodes of a social network are the people in it; two nodes are connected if the two people know each other. Kleinberg observed that in small-world networks the probability two nodes will be connected decreases as the geographical distance between them increases. And he showed that, if the probability of a connection between two nodes in a network decreases with the square of the distance between them, then there are efficient algorithms for finding the shortest path between the two nodes. Most surprisingly, he showed that if the decrease is faster or slower than the square of the distance, no efficient algorithm exists. In addition to elucidating theoretical aspects of small-world networks, Kleinberg's work has proven useful in designing peer-to-peer file sharing systems, where information has to be located without a central index.

One of Kleinberg's early results also concerned efficient algorithms, this time for the problem of finding "nearest neighbors" in high-dimensional data sets. This problem arises in the following kind of setting. Suppose you have a set of documents and a dictionary of n words. For each document, you count the number of times the first word in the dictionary appears, the number of times the second word appears, the number of times the third word appears, and so on. In this way you obtain n counts, and they can be thought of as an n-dimensional vector that represents the document. For any particular document A, how can you identify its nearest neighbor---that is, the document B that is the most similar to A of all the documents in the set? Thinking of A and B as vectors, you want to find B such that the distance from A to B is as small as possible. One solution is of course to simply compute the distance between A and every other vector and then see which distance is smallest. But if n---the number of words in the dictionary---is very large, then one is dealing with a very high-dimensional space. In high-dimensional spaces the brute-force technique of simply checking all the possibilities runs quickly into the "curse of dimensionality" and becomes time-consuming and expensive. Kleinberg broke this impasse by developing an
ingenious new approach to the nearest neighbor problem, by randomly combining one-dimensional projections of the vectors.

Another highlight of Kleinberg's work is his development of a mathematical model to recognize "bursts" in data streams. One way of analyzing the structure of information contained in a text data stream is to look for bursts of activity that appear suddenly and are sustained for a period of time. For example, even if you did not speak English, if you observed the sudden and frequent appearance of the word "Katrina" in news wire data from August

2005, you would know that something significant was happening. Kleinberg began studying bursts for a very practical reason: He wanted a better way of organizing the huge archive of personal email that he was accumulating. While the idea of a burst of information activity is intuitively clear, defining the notion rigorously is not easy, because of the difficulty of distinguishing bursts from statistical fluctuations in a sea of random information. Kleinberg's key contribution was to supply such a rigorous definition of bursts, using the mathematical concept of Markov models. He tested his ideas on a range of different data sets, such as the titles of research papers spanning several decades of research in a given area. In this case, the bursts correspond to the appearance of and attention paid to new topics in that research area.

The interplay between understanding networks theoretically and observing actual networks in action has informed Kleinberg's research and brought new perspectives about the role of technology in society. "[T]he Internet and the Web force us to think about the social consequences of a world in which information is more plentiful and travels more widely than ever before, and in which anyone has the potential, through new kinds of media and at very little cost, to become an author with a global audience," Kleinberg wrote in an email interview conducted with Technology Research News online magazine. "But there are a number of fundamental challenges here. We know that on-line discourse can be highly polarized; is it the case that the on-line tools we've created are contributing to a rising level of polarization in civic dialogue more generally? How might we accurately assess this phenomenon, and how might we think about designing new tools that make on-line discourse more productive?"

Kleinberg's work is distinguished by its richness and diversity, and also by his ability to use theoretical insights and deep understanding to address practical problems. This powerful combination ensures Kleinberg's status as one of the leading thinkers in theoretical computer science in the years to come.

1380 words

BIOGRAPHICAL SKETCH

Jon Kleinberg was born in 1971 in Boston, Massachusetts, USA. He received his Ph.D. in 1996 from the Massachusetts Institute of Technology. He is a professor of computer science at Cornell University. Among his distinctions are a Sloan Foundation Fellowship (1997), a Packard Foundation Fellowship (1999), and the Initiatives in Research Award of the U.S. National Academy of Sciences (2001). In 2005, Kleinberg received a MacArthur "genius" Fellowship from the John D. and Catherine T. MacArthur Foundation.

PICTURE CAPTION

hubs-authorities.pdf: The red dots represent "hubs" and the black dots represent "authorities". Image courtesy of Jon Kleinberg.

networks.gif: Link structures in networks. [Permission is needed from Science magazine. See

http://www.cs.cornell.edu/home/kleinber/sci01.pdf for the full article.]