"Six Degrees of Separation" Explained In New Computer Algorithm

September 12 2005

University of Massachusetts Amherst researchers have invented a new algorithm, solving a network-searching conundrum that has puzzled computer scientists and sociologists for years.

The scientists created an algorithm that helps explain the sociological findings that led to the theory of "six degrees of separation," and could have broad implications for how networks are navigated, from improving emergency response systems to preventing the spread of computer viruses.

Dubbed expected-value navigation, the algorithm describes an efficient way of searching a particular class of networks and was presented by doctoral student Özgür Simsek, and David Jensen, professor of computer science, at the 19th International Joint Conference on Artificial Intelligence in Edinburgh, Scotland.

The algorithm is applicable to a number of networks say the researchers. Ad-hoc wireless networks, peer-to-peer file sharing networks and the World Wide Web are all systems that could benefit from more efficient message-passing. The algorithm could work especially well with dynamic systems such as ad-hoc wireless networks where the structure may change so quickly that a centralized hub becomes obsolete.

The work was inspired by research pioneered in the late 1960s that focused on navigating social networks, explains Simsek. In a now famous study by psychologists Milgram and Travers, individuals in

Boston and Omaha, Neb., were asked to deliver a letter to a target person in Boston, but via an unconventional route: the message had to be passed through a chain of acquaintances.

The people starting the chain had some basic information about the target individual—including name, age and occupation—and were asked to forward the letter to someone they knew on a first-name basis in an effort to deliver it through as few intermediaries as possible. Of the letters that reached the target, the median number of people in the message-passing chain was a mere six.

"What came out of that study was that we are all connected," says Simsek. But the findings also raised a number of questions about how we are connected, she says. What are the properties of these networks and how do people efficiently navigate them?

The social network exploited by Travers and Milgram isn't a straightforward, evenly patterned web. For one thing, network topology is only known locally—individuals starting with the letter did not know the target individual—and the network is decentralized—it didn't use a formal hub such as the post office.

If navigating such a network is to succeed—and tasks such as searching peer-to-peer file sharing systems or the navigating the Web by jumping from link to link do just that—there must be parts of the underlying structure that successfully guide the search, argue Jensen and Simsek.

Participants in the Travers and Milgram study who efficiently sent the message probably acted intuitively by combining two human traits that apply to computerized network-searching as well, say the researchers.

People tend to associate with people who are like themselves, and some individuals are more gregarious than others. "Searching" using both of

these factors, one can efficiently get to a target even when little is known about the network's structure.

The tendency of like to associate with like, or homophily, means that attributes of a node—an individual in the Travers and Milgram study—tend to be correlated. Bostonians often know other Bostonians, and the same holds true for qualities such as age or occupation.

The second important characteristic of these networks is that some people have many more acquaintances than others. This "degree disparity" leads to some individuals acting as hubs.

Taking these factors into account simultaneously results in a searching algorithm that gets messages to the target by passing it to gregarious individuals who are most like the target. Or in the language of network-searching, it favors nodes that maximize the probability of linking directly to the target, which is a function of both degree and homophily, say the scientists.

Previous research had explored these aspects separately, but Simsek and Jensen are the first to step back and incorporate both these qualities into one broadly applicable algorithm with a strong basis in probability theory. And the combination yields a powerful punch. It is remarkably efficient at finding the short paths between nodes without knowing the central network's structure, say the researchers

"In this case, one plus one is more than two," says Simsek.