User Clickstreams Lead To A 'Map Of Science'

By News Staff | March 10th 2009 10:00 PM | 🖨 Print | ✉ E-mail

News Staff

[Search This Blog]

While science is of tremendous societal importance, it is difficult to probe the often hidden world of scientific creativity. Most studies of scientific activity rely on citation data, which takes a while to become available because both the cited publication and the publication of a particular citation can take years to appear. In other words, citation data observes science as it existed years in the past, not the present.

What we need is a Map Of Science.

Enter a group of Los Alamos researchers who created a high-resolution graphic depiction of the virtual trails scientists leave behind when they retrieve information from online services.
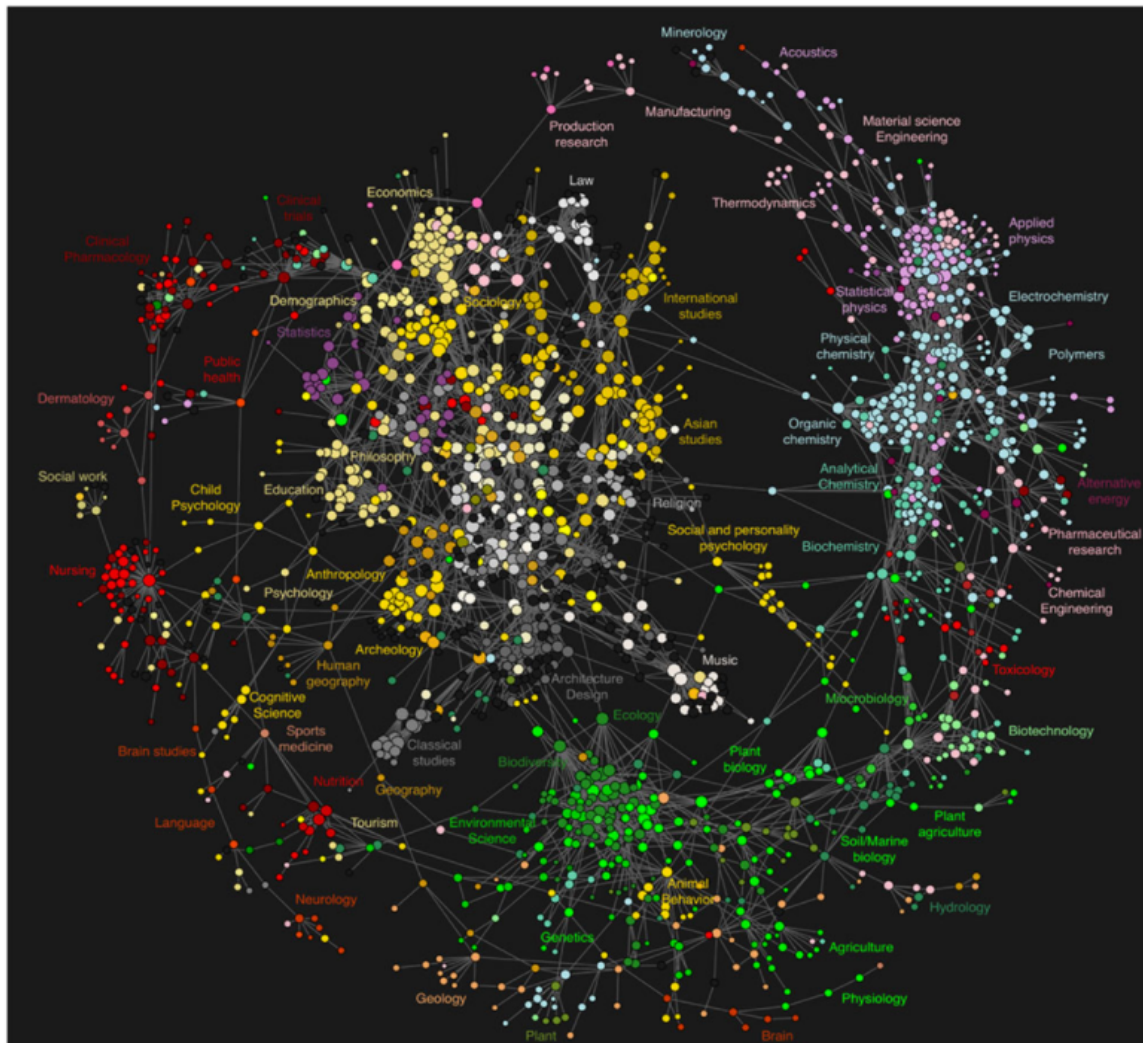
Johan Bollen and colleagues from LANL and the Santa Fe Institute collected usage-log data gathered from a variety of publishers, aggregators, and universities spanning a period from 2006 to 2008. Their collection totaled nearly 1 billion online information requests. Because scientists typically read articles online well before they can be cited in subsequent publications, usage data reveal scientific activity nearly in real-time. Moreover, because log data reflect the interactions of all users—such as authors, science practitioners, and the informed public—they do not merely reflect the activities of scholarly authors.

Whenever a scientist accesses a paper online from a publisher, aggregator, university, or similar publishing service, the action is recorded by the servers of these Web portals. The resulting usage data contains a detailed record of the sequences of articles that scientists download as they explore their present interests. After counting the number of times that scientists, across hundreds of millions of requests, download one article after another, the research team calculated the probability that an article or journal accessed by a scientist would be followed by a subsequent article or journal as part of the scientists' online behavior. Based on such behavior, the researchers created a map that graphically portrays a network of connected articles and journals.

They were surprised by the map's scope and detail. Whereas maps based on citations favor the natural sciences, the team's maps of science showed a prominent and central position for the humanities and social sciences, which, in many places, acted like interdisciplinary bridges connecting various other scientific domains. Sections of the maps were shaped by the activities of practitioners who read the scientific literature but do not frequently publish in its journals.

The maps furthermore revealed unexpected relations between scientific domains that point to emerging relationships that are capturing the collective interest of the scientific community—for instance a connection between ecology and architecture.

"We were surprised by the fine-grained structure of scientific activity that emerges from our maps," said Bollen.

According to Bollen, future work will focus on issues involved in the sustainable management of large-scale usage data, as well the production of models that explain the online behavior of scientists and how it relates to the emergence of scientific innovation. This information will help funding agencies, policy makers, and the public to better understand how best to tap the ebb and flow of scientific inquiry and discovery.

The research team includes Bollen, Herbert Van de Sompel, Ryan Chute, and Lyudmila Balakireva of LANL's Digital Library Research and Prototyping Team and Aric Hagberg, Luis Bettencourt and Marko A Rodriguez of LANL's Mathematical Modeling and Analysis Group, and LANL's Center for Nonlinear Studies. Bettencourt also is part of the Santa Fe Institute.

Bollen and colleagues received funding from the Andrew W. Mellon foundation to examine the potential of large-scale usage data. The study is part of the MESUR (Metrics from Scholarly Usage of Resources) project of which Bollen is the principal investigator. The MESUR usage database is now considered the largest of its kind.

Source: Science 2.0
https://www.science20.com/news_releases/user_clickstreams_lead_map_science