Discovery of Knowledge Flow in Science

By Hai Zhuge

Recognizing and understanding knowledge flow between scientists is valuable for science. Discovering, managing, and utilizing such knowledge are advanced services of the e-science knowledge grid environment.

Whether we are aware of it or not, knowledge flows within human society and in the Internet-mediated interconnection environment scientists increasingly rely on for research. Knowledge flows influence the evolution of culture and language, promote international collaboration, and hasten the development of science.

Scientists have developed many approaches to the static representation of knowledge, and to extracting, discovering, learning, and reasoning about it. However, knowledge is dynamic—it goes through human brains for knowing, invention, propagation, fusion, generalization, and problem solving. Scientific articles are the major medium that carries knowledge between scientists.

The Citation Network. The citations in scientific articles are objective data used, for example, by the Institute for Scientific Information's *Journal Citation Report*, for assessing scientists, scientific articles, and journals [5]. Statistics from ISI shows that top scientists are always most-cited.

Scientific articles' citations follow a power-law distribution [9], which is coincidentally similar to the distribution of the hyperlinks of the Web [1, 2]. Neither the citation nor the hyperlink relationship is transitive; that is, "A cites/links to B" and "B cites/links to C" does not imply "A cites/links to C."

Hyperlinking is arbitrary—"anything can link to anything" [3]—but citation is not. Citations among scientific documents form a *time-constrained non-redundant content net* with the following characteristics:

- The content of published papers is fixed.
- Citations between published papers are fixed.
- Published articles cannot cite those not yet completed.
- No two scientific papers may be completely identical.
- Authors who cited each other share some