Mapping intellectual milestones

21 January 2004: Not sure which direction to take to get the most from your literature search? Chaomei Chen says he can help. His new program CITESPACE can map a path to all the landmark papers and pivotal points in any scientific topic area. Chen's work, published in the *Proceedings of the National Academy of Sciences*, is part of the developing field of 'knowledge domain visualization' aimed at creating a picture of how science grows and evolves over time. "It could be useful for people who are not quite familiar with the domain, to help identify the most important papers, so they have a start," said Chen, associate professor at the College of Information Science and Technology at Drexel University in Philadelphia, USA.

"Visualization of scientific frontiers is a relatively new field," says Kevin Boyack of the Computation, Computers and Mathematics Center at Sandia National Laboratories in Albuquerque, New Mexico, who produces and analyses science domain visualizations. "The application of science to science itself has been undertaken for decades," notes Boyack. "What is new is the field of information visualization and application of its techniques to help us understand the process of science in the making." Chen's aim is to identify key changes in a 'knowledge domain', and to simplify the search for significant papers in the domain's literature. "One of our main objectives is...to improve visualization techniques so that groundbreaking articles can be characterized by distinguishable visual features," he says. He prefers to leave the definition of a 'knowledge domain' fairly vague, but essentially it can be thought of as a scientific field, he says - although large fields might have several domains.

"If you have lots of closely related publications they form the basis of a knowledge domain," he explained. "The point is not to make it too clear cut at this stage, but to present an image provided by the connectivity [between papers]." He analyses the networks formed by co-citation of papers; that is, when two papers (A and B, say) appear on the same reference list, they are linked. "The point is to figure out the connection between A and B, rather than just how many times they happen," he told *BioMedNet News*. The frequency of citation doesn't tell you anything about how A is related to B or anything else, he says, but co-citation is the way to get at how they are related.

The number of co-citations, even in a small field, is vast. According to Chen, the trick to seeing relevant links is being able to prune away all the excess noise to find patterns. "The patterns become apparent because scientists form a consensus - they independently make the connections between papers A and B. This kind of pattern will be shown in the picture, and then we prune to get rid of extraneous noise."

CITESPACE analyzes how the co-citation relationships change over time. "Think of it as a movie," Chen suggested. "You're aiming your camera at the domain and you take a snapshot every few years, or if you want more detail, every week or month." As the scientific domain develops over time, each picture will be different from the next. "The focus is what's the difference," said Chen. "How can you go from one time picture to the next - and make links."

The shape of the networks can alter dramatically over time, with nodes and links being added or removed, so Chen developed a method of 'merging' the different snapshots, matching sequential pictures together to form a complete map of how the field changes over time. "We can finally find the common points and join them so we can follow, we can walk from yesterday to today to tomorrow, like a path," he said. "You can trace the tracks of the most significant papers." The major links become like highways, he says, so you know they are important and can follow that

route. And important articles in the network become obvious as 'landmark', 'hub' or 'pivot' nodes. The large landmark nodes tend to represent highly cited articles, and hub nodes show where a particular article has connections to many different papers. "A widely co-cited article is a good candidate for significant intellectual contributions," pointed out Chen.

The pivot nodes form joins between two networks, or gateways from one to another. "The presence of pivot nodes enables us to narrow down the visual search quickly to a small number of good candidate nodes for intellectual turning points," explains Chen. "Ultimately, it's just like watching a movie - you don't have to read individual papers - and then you can get an idea of the position of each of the papers in the whole of that domain," he concluded. Jon Kleinberg, associate professor of computer science at Cornell University in Ithaca, New York, is impressed. "Given the complexity inherent in a scientific community, it is difficult to construct 'global' views of its structure and its evolution over time. [Chen's] technique provides a nice way to obtain such global views, and it identifies several critical roles that individual papers can play in transforming a field of study," he said.

Kleinberg's approach to the problem is to study the usage of technical terms to study how a field develops. "[My method] is based on the idea that many technical terms exhibit a characteristic 'bursty' pattern in their usage over time," he said. "They increase sharply in frequency as the term becomes prevalent and then decline again as the discourse of the field moves on." By scanning the literature over several decades, he identifies the "burstiest" terms and the time interval when they were active, to generate a timeline and follow the hotspots in a field. Chen acknowledges that one of the shortcomings of his program is the detection of very abrupt changes in citations within a short period - essentially a paradigm shift in the domain. He suggests that Kleinberg's method, among others, could be used to overcome this deficiency.

Kleinberg agrees. "I think there is definitely an opportunity to see whether techniques to analyze 'burstiness' could be usefully applied in conjunction with Dr Chen's methods," he said.