Research as a Social Process: Considerations for Academic Libraries

Doug Horne

Doug Horne has been a librarian for twenty years, the first two at the Kitchener Public Library, and the last eighteen at the University of Guelph. His work and interests have varied over the years but have always revolved around new technologies, the nature of research, and public service. Most recently, Doug has put all of these interests together and taken on the role of User Experience Librarian.

Abstract

Academic Libraries for some time have attempted to come up with innovative ways to accomplish the mission of supporting research. Since research is a core activity of academia this would seem to be a natural goal for libraries, but it has been difficult at times to gain the attention and imagination of the academic community as libraries attempt to come up with new ways to serve the researcher. This paper suggests that our difficulty in gaining “traction” with our community members involved in research might be resolved if we take a closer look at what the practice of research involves, and attempt to more closely match the services we develop to the types of activities that our community undertakes. In particular, this paper suggests that libraries and those who provide us with information have lost sight of the fact that research is a social process in which researchers value interaction and confer prestige based on reputation and influence. Influence and reputation are vital concepts in the legitimation of ideas, the determination of the relative value of research, and the creation of knowledge. If academic libraries hope to be a central focus of research on campus, librarians need to build tools and offer services that take into account these social interactions and the importance of the community to the researcher.

While academic libraries create mission statements to reflect the reality of their local environment and the characteristics of their institutions, there are several themes that are common across these institutions. One theme that is often seen and easy to envision in everyday practice is that libraries strive to support teaching and learning through the collection of resources and the development of tools and services to assist users. Most librarians can easily identify with this concept as they have been involved in activities that can be seen to be working toward that goal. Another frequently occurring element of academic library mission statements that is more difficult to understand and put into action is the frequently mentioned idea that libraries should be actively working to support research. Fulfilling this mission is simple if the ‘research’ that is being supported is that work that users do in libraries as background work for their latest investigations into topics in their chosen field, or if ‘research’ is considered to be the activity of simply looking for information on any given topic. When thinking of research in these ways, libraries have fulfilled their mission in this area simply by making resources as accessible as possible to users by providing intensive help to researchers who elect to employ our expertise, and by thinking of that as a program of research support.
However, having managed a group of librarians responsible for liaising with faculty, it has become clear that providing support for research can become a significantly more complicated process. During recent years, research funding at the University of Guelph (my home institution) has grown in importance as a source of revenue and a measurable indicator of institutional status. The Library attempted to increase its focus on research support proportionately, but had only managed to come up with a few new ways to demonstrate our usefulness in this area. It seemed reasonable at the time to assume that if we could determine what academics actually do when performing this core activity, then we might better understand how to serve this aspect of their work. Thus I had asked the seemingly trivial question repeatedly: “What is it that academics do when they conduct research?” Having discussed this issue with experienced researchers and senior administrators responsible for campus-wide research, it became clear that even these individuals are uncertain as to how libraries might enhance their role in supporting researchers on campus beyond the traditional role of collecting and providing access to information resources. In many of the conversations with researchers it became clear that this question had simply not come up frequently in the past, and that the library is most commonly considered and valued as a storehouse of resources. Although this is a core and vital service to researchers, librarians also spend much time trying to determine if the library has more of a significant role in the process of research than just being a building where information is stored. This becomes an ever-pressing question for librarians with the advent of ubiquitous digital information. When even the simplest roles are being performed by online catalogues, as a profession, concerns are raised about whether or not some of these external resources may be more beneficial to users.

This paper takes a close look at the nature of what academics refer to as the activity of research, the nature and purpose of the research paper, as well as the types of services that libraries offer to support the activities of academic researchers. I illustrate how the history and practice of academic research comprises far more than just searching for information, and at its core continues to be a social activity practiced largely by communities of scholars, communicating and sharing ideas. These communities interact with each other and share information and ideas in ways that have been standard practice and, significantly, break the commonly held image of the scholar locked away in isolation working through a problem. In addition, I examine the nature of the scholarly research paper itself and what role publication plays in the process of research and the creation of knowledge. With an enhanced vision of the nature of research, I examine the possible roles of the Library relevant to this vital process in academia: how we have (or have not) provided products and services that serve this aspect of the research enterprise, and what are the major factors that might contribute to achieving that end?

Librarians have a good perspective from which to observe the activity of research. Sitting at a reference desk, we know that while the research activity of an undergraduate student may involve searching for a certain number of resources and then writing a paper once those are found, the advanced academic takes something of a different route. We often advise the senior undergraduate to pay more attention to
the citations that can be found in works already discovered, and to work one’s way into the web of related articles by following the paths of cited articles. Once a student is advanced enough to be doing more original research, the search is no longer for the article about their topic in particular, but rather the articles related to their topic that they synthesize to create new ideas. They can consult with the experts in an area by browsing the entire literature of their subjects, observing the patterns of influence among scholars, and ensure that their topic is indeed original. This is the leap that we often help students make at the reference desk as they struggle with topics in their more senior years that require familiarity with an entire literature, and no longer simply demand the search for an answer in a discreet article.

Experience from my own graduate research and assisting with that of others, suggests that a very important part of research is the identification of the relevant community of scholars, as well as understanding the relation of the concepts being considered and the connections and patterns of influence that exist among researchers doing the work. There is a long tradition in the history of science of suggesting that research is not accomplished in a vacuum, but only as one scholar understands another, adopts his ideas, and advances those ideas a step further than his predecessors. Most famously, perhaps, in the words of Isaac Newton in 1676 “If I have seen a little further it is by standing on the shoulders of giants” (Merton, 1986, p. 273). In fact, this acknowledgement of the importance of one’s predecessors and the centrality of taking an idea and advancing it as the essence of ‘original research’ has been traced as far back as the twelfth century or earlier, and repeated by many of the great thinkers in the philosophy of science.

In discussions about the nature of research, the social nature of this work has been worked out in detail in the idea of the invisible college, which has been recognized by scholars but only investigated at length by a few researchers. In fact, the identification of the community of scholars with similar interests, be they formal or informal groupings, could be accomplished by virtually any scholar for social and purely practical reasons. At an advanced level of research, the originality of work and receiving credit for it are of great importance; a researcher must monitor the work of peers and rivals to ensure that advances in research are not mere duplications of previous work, and be aware of the work of others in order to build on previous accomplishments. As described by one of the central researchers of this scholarly interaction, Derek de Solla Price (1963):

[T]hese groups devise mechanisms for day-to-day communication. There is an elaborate apparatus for sending out not merely reprints of publications but preprints and pre-preprints of work in progress and results about to be achieved. The existence of such a group might be diagnosed by checking the preprint list of every man and following this by a check of the list of each man mentioned. I think one would soon find a closed group, a small number of hundreds in membership strength, selected from a population of a large number of tens of thousands (p. 75).

Price continues:

Such groups constitute an invisible college, in the same sense as did those first unofficial pioneers who later banded together to form the Royal Society in 1660. In exactly the same way, they give each man status in the form of approbation from his peers, they confer
prestige, and above all, they effectively solve a communication crisis by reducing a large group to a small select one of the maximum size that can be handled by interpersonal relationships. Such groups are to be encouraged, for they give status payoff without increasing the papers that would otherwise be written to this end. I think one must admit that high-grade scientific commuting has become an important channel of communication, and that we must ease its progress (p. 76).

The groups that Price refers to are very familiar to librarians, and to all academics. They are both formal and informal, and vary in nature between disciplines and sub-disciplines. In fact, the social groups that exist for scholarly communication will be different for almost every individual researcher, but they seem to exist and play a role in the academic life of virtually everyone. These groups can be professional societies, groups of researchers that gather at annual conferences, scholars from similar disciplines that review each other’s work, research teams focused on one topic, or simply loose collections of researchers from the same or related disciplines who are aware of each other and communicate either formally or informally.

In fact, the term ‘invisible college’ is problematic, as no one up to this point has been able to definitively describe the nature of the relationships between researchers, and thus describe this ‘invisible college’ - a term first used by the Royal Society of London in the seventeenth-century to describe a group of scientists who did not all belong to a formal institution but whose relationships across institutions and over geographic distances constituted some type of communication network (Zuccala, 2005, p. 152). The term has since been used to describe the social institutions of science (both formal and informal) that exist for scientists to share work, collaborate, cite one another’s works, and receive affirmation and recognition from peers. These ‘social processes of research’ serve multiple purposes. Researchers not only communicate and share ideas in formal or informal forums, but also utilize various processes to legitimate ideas prior to publication in the acts of distributing pre-prints, delivering papers at conferences, and discussing ideas with fellow members of professional societies:

The basic phenomenon seems to be that in each of the more actively pursued and highly competitive specialties in the sciences there seems to exist an "in-group". The people in such a group claim to be reasonably in touch with everyone else who is contributing materially to research on this subject, not merely on a national scale, but usually including all other countries in which that specialty is strong. The body of people meet in select conferences, they commute between one center and another, they circulate preprints and reprints to each other, and they collaborate in research. Since they constitute a power group of everybody who is really somebody in a field, they might at the local and national level actually control the administration of research funds and laboratory space. They may also control personal prestige and the fate of new scientific ideas, and intentionally or unintentionally they may decide the general strategy of attack in an area (Price, 1963, 119-120).

Those seeking information have long been aware of the importance of the citation in the process of research, and how tracing patterns of citation can allow one access to the ‘approved’ research in any given field. In fact, citation patterns in the sciences are so important that impact factors have been developed as a measure of the importance of a scholarly work, and the study of the frequency of citation may be used in the process of justifying tenure for a faculty member. The same could be said
of invited speaking engagements, conference presentations to groups of peers, and other ways in which the work of a scholar is recognized by a community, and reputation is built and confirmed.

In what really shouldn’t be surprising to anyone who has been involved in the process of research, Diana Crane (1972) states:

Studies of citations show that about half the references in each group of new scientific papers link them to a small group of earlier publications, most of which are close to them in time. The other half of the references links [sic] the papers apparently randomly to a very sizable part of the scientific literature. These findings indicate that the literature of basic science consists of tightly knit clusters of papers, each of which is loosely linked to a large number of other clusters. The clusters represent research areas, sets of closely related problems that, as will be shown in subsequent chapters, are viewed by the scientists who study them as discrete entities (p. 12).

She continues:

If scientific growth represents the accretion of many small innovations, and if, in producing these innovations, authors are indeed building upon each other’s work (as analysis of their citations to each other’s publications suggests), then it would appear that such authors are adopting some of each other’s innovations. In this sense, the growth of scientific knowledge is a kind of diffusion process in which ideas are transmitted from person to person (pg. 22).

Crane concludes this thought with a statement that is central to that idea that research is far from a solitary practice as it is often perceived, but rather, that:

scientific growth is both a social and a cognitive process. Social interaction facilitates the diffusion of ideas that in turn makes possible cumulative growth of knowledge in a research area (p. 26).

A common view from outside the scholarly community is that the academic paper is primarily a communication device in which scholars display their work so others in the field might discover what they have been researching. The concept of the invisible college suggests, however, that this is an overly simplistic idea of publication, as a great deal of communication has gone on prior to final publication and many scholars in the field will already be aware of the work by this point in the process. It would be a relatively uninformed or out of touch scholar who would be surprised to see an article in a journal in his field published by a member of his own community. Admittedly, a scholar researching an unfamiliar field would be navigating through works by authors that he had never heard of on novel topics, but a scholar who had been admitted to the community in his field as a recognized scholar should not be seeing completely unfamiliar things as he browses scholarly journals. Derek J. de Solla Price (1963), has considered the following question:

If then the prototype of the modern scientific paper is a social device rather than a technique for cumulating quanta of information, what strong force called it into being and kept it alive? Beyond a doubt, the motive was the establishment and maintenance of intellectual property. It was the need which scientists felt to lay claim to newly won knowledge as their own, the never gentle art of establishing priority claims (p. 58-59).

He continues that “scientists have a strong urge to write papers but only a relatively mild one to read them” (Price, 1963, p. 62). Being members of a group of researchers already and being involved in the ‘conversation’ of research throughout the process, the paper takes on a unique role not primarily as a communication device,
but more importantly as the officially recognized place to stake one’s intellectual claim and to mark the next step in the process of knowledge creation. While the undergraduate student at the reference desk will see the article found by searching as an isolated document and the result of a research project, the researcher may very well see it as a snapshot of the state of research at the moment of publication.

It is important to add that the scholarly paper, if it is successful, is not the end of the conversation. The document will be one moment of an ongoing conversation of the community of researchers and, if successful, will be referred to in the process of citation as that conversation continues. In fact, if the work does not contribute to the continuing conversation in the field, it is unlikely to be seen as terribly successful as it will not be cited and it will have a very low impact factor. In the words of Price (1963): A scholarly publication is not a piece of information but an expression of the state of a scholar or group of scholars at a particular time. We do not, contrary to superstition, publish a fact, a theory, or a finding, but some complex of these. A scientific paper is at the same time more and less than a concept or a datum or a hypothesis. If the paper is an expression of a person or several persons working at the research front, we can tell something about the relations among the people from the papers themselves (p. 160).

By the time a paper reaches a journal and may be retrieved by a search, the author is likely further down the path of research and does not treat this article as ‘news.’ The discussion has likely moved on, but the intellectual claim has been staked. While treating the articles in journals as the finished product of a research project is suitable for the undergraduate or the casual information-seeker (who is likely to be searching for the ‘answer’ or a supporting fact for a point in an assignment), does this type of searching equally serve the more serious researcher?

The scholarly paper serves a different purpose for the student than it does for the community of scholars. The question naturally arises of how libraries might better support the actual process of academic research, much of which will be accomplished pre-publication and prior to the inclusion of a finished document in a library collection. The world of publishing that we deal with in libraries (and in information-seeking activities in general) is a series of walled gardens, defined by the ownership of a vendor, with search functions defined by that vendor to search their material and that of nobody else. This searching is mostly a matter of the most basic of string-matching in the form of keyword searching, and relevance ranking of results is often somewhat less than optimized. Citation lists in these articles are generally not composed of live links between related papers as that might be, in fact, a connection to another vendor’s collection in another walled garden. The ‘conversation’ that is academic research is thus hobbled in this environment as the conceptual and social links between articles and disciplines are clumsy or absent, and search relevance is calculated by keyword ‘hits’ rather than conceptual linking, the approval of the research community, or the sharing of information between like-minded researchers. The social process that is key to the legitimization of academic work is thus entirely absent from what we in libraries most often offer as ‘search.’

Currently, it seems safe to state that many of the systems provided by the publishers themselves and offered to libraries are losing ground to ‘third-party’ search
systems that can be accessed by anyone, with Google being the obvious destination of most. It is being praised most commonly for the simplicity of its interface and the relevance of the results of searches. The sometimes startling relevance of Google Scholar results is the result of a carefully designed algorithm intended to reflect the social process of research and to tap the very social systems that researchers have created to communicate with their community. This connection to the community of researchers is not an accident as the Google founders clearly spelled out their intentions early on in the development of their search algorithm. In addition to developing methods for allowing search to be fast and the interface to be simple, the creators of Google, Sergey Brin and Lawrence Page, put much thought into relevance and the importance or ‘reputation’ of an article. In a 1998 article they state:
The citation (link) graph of the Web is an important resource that has largely gone unused in existing Web search engines. We have created maps containing as many as 518 million of these hyperlinks, a significant sample of the total. These maps allow rapid calculation of a Web page’s “PageRank,” an objective measure of its citation importance that corresponds well with people’s subjective idea of importance […] Academic citation literature has been applied to the Web, largely by counting citations or backlinks to a given page. This gives some approximation of a page’s importance or quality (Brin and Page, 1998, p. 109).

Similarly, Chen, et al. suggest that “the PageRank algorithm implements, in an extremely simple way, the reasonable notion that citations from more important publications should contribute more to the rank of the cited paper than those from less important ones” (Chen et al, 2007, p.15). This makes it clear that a search engine like Google is using a measure of influence or importance when retrieving and ranking documents, and this parallels how researchers in practice determine the importance of works.

As Chen, et al. (2007) state:
The situation in citation networks is not that dissimilar from that in the World Wide Web, where hyperlinks contained in popular websites and pointing to your webpage would bring more Internet traffic to you and thus would contribute substantially to the popularity of your own webpage. Scientists commonly discover relevant publications by simply following chains of citation links from other papers. Thus, it is reasonable to assume that the popularity or “citability” of papers may be well approximated by the random surfer model that underlies the PageRank algorithm (p. 14-15),

At the reference desk lately, and in my own personal searching, the effectiveness of the relevancy ranking of Google seems to have become clearly superior to virtually any vendor-supplied front-end search. Although anecdotal, and possibly varying by discipline, over the course of the past year, I have repeatedly turned to Google Scholar simply to benefit from the relevancy algorithm, and Google’s ability to calculate influence or reputation. The most obvious reason for this perceived effectiveness is clearly laid out in Brin and Page’s article of so many years ago. Their system was designed with the idea of the citation map of the literature first and foremost. ‘Relevancy’ to them is a measure of what they called PageRank, ‘citation importance,’ or the importance or influence of that document measured by its citation relation to other documents. The common concept of ‘relevance’ in library or vendor systems, on the other hand, is typically based on something called ‘TF-IDF’ or
‘term frequency - inverse document frequency.’ Although search algorithms are routinely kept secret, it is common knowledge that the principles employed to determine relevance in most of these systems involves the frequency of term use (TF), combined with the infrequency of its use through the database in question (IDF). So if a term shows up many times in a document, it has a high TF score, and if it is an infrequently utilized term generally, it is considered to be ‘unique’ or ‘important’ (Schneider, 2010). While Google will, of course, also look for matching terms in its search, the crucial measure of relevance, and thus importance or influence, is a very human measure of importance based on links and citations between documents. These are, in effect, expressions of the importance of a document suggesting that it is worthy of being linked to or cited by the community of scholars. One of the complaints against Google’s tactics is that they routinely ignore the ownership that splits our literatures into segments defined by the vendors who happen to own each piece. It should be clear by now that for Google to be effective at mapping the relationship of influence between documents the division of the literature by corporate ownership cannot split information into collections defined by access rights. As much as possible, the entire universe of intellectual output needs to be mapped, in order to maximize the effectiveness of the measurement of influence between scholars who are not separated by vendor agreements, but grouped by common research interests.

What does all of this mean for the library? As much as many of us hate to admit it, Google has understood the nature of the research community and made the relationship between documents the cornerstone of the PageRank algorithm that powers all of their search functionality. In Google Scholar, they have taken the same system into the controlled environment of scholarly journals and peer review, thus allowing the researcher to be comfortable with the idea that he or she is searching only scholarly material, and eliminating the concerns related to searching websites that do not employ peer review. This was truly the turning point for the legitimate use of search engines originating outside of the library to find materials that may very well be held by the library. To this point, libraries have merely reacted to these developments and have attempted to play catch-up while reassuring themselves that they are the centre of the academic research process. Search front-ends, with their own relevance ranking engines, have been promised for libraries, as has the ability to search across many formats, but these enhancements still show no sign of employing influence measures beyond traditional counts of matching terms in the process of search.

Now that the technology is available and familiar to users, libraries must take into account the full range of purposes of the academic paper when designing search. In addition to considering a document as a collection of words and phrases, value can be added if it is seen as a snapshot of a continuing conversation between researchers. This step would allow libraries to use the available data to map the academic conversation and describe the relationships between researchers. It is also possible that we might then be able to point out new possible relationships that had not been discovered but might be beneficial to researchers. In our increasingly interdisciplinary
academic environment, the discovery of patterns of influence across disciplines that might otherwise not be discovered in conventional searching could be a great service to researchers in new and fringe areas of study. Others have long understood the importance of reputation and influence and have been seeking out innovative ways to put those concepts to work. We are already at the technical point at which researchers can take advantage of the available data to mine the idea of reputation and influence to rank research projects and documents, and to seek out new communities that were previously hidden. The next generation of social software (and some already in production) are using the huge quantities of data being gathered to discover the relations between users, and make recommendations based on their networks, influences, and interactions. To date, these systems are not commonly being implemented in libraries.

While libraries have continued to refine their search with tagging, faceted searching and efforts to consolidate large quantities of disparate data, the use of reputation as a search factor is still limited to the manual input of ‘likes,’ when that data has been made available. The majority of the advances in searching in libraries have been efforts to optimize literal string matching, and the concept of what we are searching needs to be expanded if we hope to keep up with other readily available technologies. We need to move beyond mining collections of words, and begin to take advantage of data that describes the relationships between documents and the scholars who have produced them. This step could be a great contribution that libraries can make to the process of research. For this to be done effectively, however, libraries must begin to look at the literature of academia not as collections divided by ownership, but as clusters of work related by conceptual connections between researchers and patterns of influence that frequently exist across collections currently delineated by the laws of intellectual property. This should not be a great conceptual leap for librarians who have always campaigned for the freedom of information, for the ability to investigate without concern for legislated boundaries, and for academic freedom. Librarians also have the experience of lobbying for just these issues, should commercial, rather than academic, interests prevail.

Too often libraries become concerned with replacing an existing system that is already functioning in the interest of better serving their users. One of the great difficulties of promoting alternatives to the existing system of scholarly publishing (such as open access journals) is that the system of legitimization for each discipline is already in place, has been ingrained as part of an academic tradition within those disciplines, and tend to work for the primary purpose of staking an intellectual claim. It is possible for libraries to reestablish their central role in the academic process not by attempting to build a new structure to compete with one that is firmly in place, but rather to enhance that existing process, and reassert the principles that built them in the first place to facilitate research and the sharing of ideas. It should not be difficult to sell the idea that we need to identify and map the system of influence between researchers, and we have the data, knowledge, and infrastructure to do that well. We need to campaign for the breaking down of the walls that have divided academic literature in artificial ways so that we can clearly define relationships that do not
follow the lines of the corporate ownership of information. This effort would be very much in line with our traditional library principles and beliefs. In taking on this task we could truly support research, facilitate the relationships between people and their work that is vital to the process of knowledge creation, and help to forge new productive relationships between researchers.

References


Schneider, K.G. “How OPACs suck, part 1: relevance rank (or the lack of it)”, ALA Techsource, (March 13, 2006), http://www.alatechsource.org/blog/2006/03/how-opacs-suck-part-1-relevance-rank-or-the-lack-of-it.html