Does the Internet Change the Competition of Scholarly Communities? -- Evidence from a Citation Analysis

Meng Ma
University of Maryland

Follow this and additional works at: http://aisel.aisnet.org/amcis2004

Recommended Citation
Ma, Meng, "Does the Internet Change the Competition of Scholarly Communities? -- Evidence from a Citation Analysis" (2004). AMCIS 2004 Proceedings. Paper 264.
http://aisel.aisnet.org/amcis2004/264

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Does the Internet Change the Competition of Scholarly Communities? -- Evidence from a Citation Analysis

Meng “Jessie” Ma
Robert H. Smith School of Business
University of Maryland, College Park
mma@rhsmith.umd.edu

ABSTRACT
Recent research on search costs in electronic markets documents the wide use of Internet has reduced consumer search costs and has increased market efficiency. This study, considering scholarly publications as a special type of product, investigates the Internet’s impact on knowledge usage and transfer in academic communities. Based on the work on search theory and knowledge transfer, this paper proposes that lower search costs of scientific publications can increase scholars’ knowledge usage and change their capability or means to compete in the scientific marketplace for audiences. Scholarly articles not freely available online may be quoted less because of the easy and free access to a growing number of substitutable publications. Analysis of randomly selected publications from the ACM digital library provides preliminary support for my propositions. In addition, results also suggest that the Internet has a major impact on the speed of knowledge transfer among scholars.

Keywords
Search cost, Accessibility, Scholarly Community, Knowledge Use.

INTRODUCTION
In spite of the increasing research on the impact of the Internet on various retail markets, such as books (Bailey, 1998, Brynjolfsson and Smith, 2000), cars (Morton et al., 2000), airline tickets (Clemons et al., 2002), electronic products (Baylis and Perloff, 2002), prescription drugs (Sorensen, 2000), and insurance markets (Brown and Goolsbee, 2000), very little attention has been paid to a particular type of good — academic publications. The Internet-based electronic resources are growing dramatically. More and more publications are available online now, with or without a fee, leading scholarly communication to a fascinating stage of the transformation to an electronic format (Odlyzko, 2000). The motivation of this study is to further investigate the impact of this transformation of scholarly communication.

Extant work has not yet examined the impact of the Internet on scholarly communities in terms of a researcher’s benefits and costs. In a scholarly community, scientists can be considered as both the seller and the buyer of information goods (publications). Similar to other products, academic publications cost scholars (“buyers”) in terms of time and efforts to search. The Internet has reduced scholars’ search costs to locate and access publications. On the other hand, scholars are also the “seller” of academic publications. The reduced search costs of publications may also change the competition in a scholarly community (“scientific marketplace”). My research questions are: (1) Does reduced search cost increase knowledge use and transfer? (2) How do the differences in accessibility influence the impact of an article in terms of the number of citations? (3) Does reduced search cost change the distribution of article impacts? This paper presents an overview of the propositions I have developed, and some preliminary quantitative results.

THEORETICAL BACKGROUND
Search Theory
Based on Stigler’s (1961) original work on search theory, there have been numerous models and empirical research analyzing the impact of search costs on market efficiency (e.g., Brynjolfsson and Smith 2000). Stigler’s model suggests that costs exist for searching price and quality information and the customer is not fully informed. A consumer visits a particular number of
stores, depending on her search cost, and then buys from the store with a better quality/price ratio, which may not be the best in the market. Due to search costs, retailers selling inferior products with a high price may still make a profit from the uninformed consumers. Most current research suggests that the wide use of the Internet can reduce buyers’ search costs and increase the accessibility of product information (Brynjolfsson and Smith, 2000, Karen et al., 2002, Clemons et al., 2002, Bailey, 1998). As a result, products with lower quality/price ratios are less likely to be purchased in the Internet Age than in the pre-Internet Age.

The knowledge development process in scholarly communities can be viewed as a marketing process, whereby ideas are marketed in the form of alternative theories (Ozlyzko, 2000). From this perspective, we can use citation analysis to study the relative success of various theories or authors in the scientific marketplace (Zinkhan and Leigh, 1999). There is a cost (purchasing a journal or article) associated with the use of academic publications and a search cost (finding the right paper) for each knowledge user. The characteristics of the Internet have dramatically changed the search cost of scholarly publications. I believe, like the Internet’s impacts on other retail markets, the increasing use of the Internet in so-called “scientific marketplace” may change the way of competition and market equilibrium.

**Ease of Access and Knowledge Use**

Research on knowledge transfer has shown that knowledge accessibility significantly influences knowledge usage (Allen, 1966, Allen and Reilly, 1973, Gerstber and Allen, 1968, Morrison and Vancouver, 2000). Knowledge users may settle for knowledge with easy access even though it may not be the best knowledge out there. In the context of scholarly communities, Odlyzko (1999, 2000) has suggested that substitutes exist for most publications. For instance, if the researchers are interested in a certain theory, some of them may find the original literature, while the others may be satisfied with a related reference or review. Further, Lawrence (2001) has empirically tested that the computer science literature with free online full-text was cited more because of its ease of access.

**RESEARCH PROPOSITIONS**

Citations are normatively expected in scholarly communities when a concept, result, or idea is adopted, and thus provide a behavioral record of knowledge use and objective measures of impact (Zinkhan and Leigh, 1999, Garfield, 1955). Although citations have been examined for several decades, a complete theory of citation is still lacking (Leydesdorff, 1998, Cozzens, 1981). Drawing upon the extant literature, I summarize the factors that influence an article’s number of citations in Figure 1. First of all, most scientists agree that paper quality is correlated with citations (e.g., Garfield, 1955, Garfield, 1996). Second, paper accessibility, defined as how easily readers can retrieve and obtain full-text of an article, will influence citations as an indicator of knowledge use. As argued in the second section, ease of access will lead to more knowledge use (Allen, 1966). Besides quality and accessibility, two control variables are also witnessed in extant research. First, the total number of publications in the same subject reflects the competition for audiences in the same academic field, which may influence the citations of each article (Garfield, 1955). Second, paper published in more highly ranked journals may be cited more because those journals are believed to publish high quality work and thus attract more knowledge seekers (Morrison and Vancouver, 2000, Leckie et al., 1996).

The volume of scientific literature typically far exceeds the ability of scholars to identify and utilize all relevant information in their research area. As discussed above, the Internet has the potential to significantly reduce search costs by allowing the users to search articles online and obtain full-text for free. It can be more difficult for an individual paper not freely available online to be cited by other researchers, who now have easier access to wider choices. Hence, I propose,

**P1: Ceteris paribus**, when the use of the Internet increases, articles not freely available online will be cited less than those freely available online.
Relevance of an article is the relation between two articles in such dimensions as information, context, and time (Mizzaro, 1997, Saracevic, 1970). Ideally, scholars should cite an article solely based on its quality and relevance. However, when search costs of publications are high and accessibility is low, researchers may also cite articles based on their accessibility (Lawrence, 2001). As the Internet dramatically increases the accessibility of publications, access becomes a much less important factor in a scholar’s decision of which work to cite. As a result, citation behavior becomes more efficient, based on just quality and relevance. In other words, a researcher may cite some low quality or less relevant articles in the pre-Internet age due to their easy access, but with accessibility to an increasing amount of literature, she is less likely to cite those low quality or less relevant articles. Therefore, along with the penetration of the Internet, the distribution of citation frequency may cluster around those limited number of high quality classic articles in the field (i.e., citation dispersion become higher), or scattered evenly in different sub-topic that are more relevant to each knowledge user (i.e., citation dispersion become lower). Hence, two possible changes may apply to the distribution of citations:

P2: Ceteris paribus, the more people (scholars) use the Internet, the higher the degree of citation dispersion is.

P2’: Ceteris paribus, the more people (scholars) use the Internet, the lower the degree of citation dispersion is.

METHODOLOGY

Citation analysis will be used in this study. Citation-based measures have been used to evaluate the impact of journals (Zinkhan and Leigh, 1999) and research institutions (Erkut, 2002). It also plays a role in personnel decisions, such as hiring and tenure. Recently, a couple of studies adopted citation analysis to explore the visibility or impact of publications available online (Lawrence, 2001, Goodrum et al., 2001).

Before collecting large amounts of data, I have conducted a pretest in order to provide some initial support for my propositions. The level of analysis is publication. Eighty publications in IT and Computer Science, from 1992-1997, were randomly selected from the ACM digital library. Based on Figure 1, two linear regressions were run:

Model 1: \( \text{CITE5} = \beta_1 + \beta_2 \text{IntPen} + \beta_3 \text{J_Ranks} + \beta_4 \text{RefRatio} + \beta_5 \text{TtlPubs} \)

Model 2: \( \text{CITE2} = \beta_1 + \beta_2 \text{IntPen} + \beta_3 \text{J_Ranks} + \beta_4 \text{RefRatio} + \beta_5 \text{TtlPubs} \)

Citations were regressed on Internet penetration (IntPen), journal ranks (J_Ranks), reference ratio (RefRatio), and the total number of publications in the same subject (TtlPubs).

1 ACM digital library is an online article search engine. Full-text content is not free and only available for subscribed users.
PRELIMINARY RESULTS

Internet Impact

No heteroskedasticity was detected using a Park test. OLS results show that Internet Penetration has a significant negative effect on the number of citations an article had in both 5 and 2 years after publication ($p = 0.024$–$0.03$). P1 is preliminarily supported. Consistent with our intuition, TtlPubs has a significant negative effect on a publication’s citation number ($p = 0.017$–$0.046$), implying that the competition for audiences in IT and computer science is escalating with a growing number of publications. RefRatio has a positive link to citations ($p = 0.001$–$0.015$). This finding is consistent with previous study (Garfield, 1996) and indicates that paper (such as a review) with more references may serve as a substitute of those articles it refers. Overall, two estimation models explain a fair amount of the variance ($R^2 = 0.235$–$0.305$) considering the small sample size of the preliminary study.

Citation Dispersion

I compare the ranges, interquartile ranges, and standard deviations of CITE2 and CITE5 from year 1992 to 1997. Overall, the citation dispersions tend to decrease overtime. This result provides some evidence for P2’--knowledge users are quoting more relevant articles instead of well-known classic pieces that may not be very applicable.

Citation Delay

Since the ACM digital library was first launched online in 1997, I am particularly interested in whether this event reduced the delay or increased the speed with which an article was cited. RATIO is defined as the ratio of citations in 2 years and citations in 5 years. The bigger the RATIO, the faster an article is cited. I believe that easy access provided by online digital library will reduce the delay of knowledge use and transfer. A T-test was conducted ($p=0.007$) and results show that on average, 37.3% citations happened in the first 2 years for paper published in 1996, while 75.2% in 1997. Pretest supports my argument that the Internet reduced the delay of knowledge usage and made knowledge communication more efficient.

CONCLUSION AND DISCUSSION

There is no doubt that the Internet provides a fast and efficient means of distributing and accessing scientific publications. The search cost literature suggests that the Internet increases market efficiency and competition. In this study, I apply an analogy between commercial product markets and the “scientific marketplace”, and empirically investigate the Internet’s impact on it. The preliminary results show that easy access did not change the amount of knowledge used by scientists when they wrote a paper (RefRatio did not change overtime) but may have improved the relevance of the references and the speed of knowledge transfer. I also find that scientific articles not freely available online were less cited with the increasing penetration of the Internet.

There are some limitations in this research in progress. First of all, the sample is small and limited to the publications from the ACM digital library. Second, whereas I argue that the citation decline of ACM articles is due to the unavailability of free access, alternative explanations exist and need to be considered carefully. For instance, it is possible that most classic and fundamental studies, which are cited repeatedly, were published in early 90s. However, Erkut (2002) argued that classic works in a field may become common wisdom and authors may stop citing them. Also, authors may cite reviews as a substitute of the original research paper (Garfield, 1996). I am continuing to work on eliminating alternative explanations. A citation comparison between publications freely available online and those not is also in progress.

Finally, this study sheds light on some policy issues regarding copyright. Free online availability facilitates knowledge transfer and offers substantial benefits to both science and scientists. Making a publication freely available online may boost its impact and offer scientists an advantage in the “competition” of knowledge market. Scholars may prefer posting their works online for a higher impact, though copyright problems with the publisher restrict such conduct. Thus, current copyright policy that means to protect knowledge owners may diminish the social welfare.
REFERENCES