Appraise –
Beyond Threaded Conversations in Scholarly Publishing

Shay David
Cornell University / Information Science
301 College Ave. Ithaca, NY 14850
sd256 [at] cornell [dot] edu

With
Geri Gay, Carl Lagoze, Bing Pan,
and Simeon Warner

ABSTRACT
Scholarly communication is radically changing as new technologies like Blogs, Wikis, e-print archives and the like open up new multi-threaded conversation spaces. These alternatives challenge the ailing journal-based publication system which is imploding and the single-threaded conversations that use email to compensate for the latency of traditional publishing. Building models of hybrid-networks, which include documents and researchers alike, Appraise is a project that aims to articulate design principles and develop software prototypes that address the arising needs of this system in transition. Appraise will be deployed and validated at the arXiv electronic repository that includes over 300,000 articles and is by far the largest open scholarly database in all areas of science and technology. Appraise will explore the manner in which new technological interfaces have already changed scholarly communication, and the potential that they hold for even more fundamental change. Our investigations will be both social and technical and involve both the application of new models for analysis and the development of on-line user tools based on these network models.

Author Keywords
Scholarly Publishing, Hybrid Networks, Social Network Analysis, Design principles for collaborative spaces.

ACM Classification Keywords
H.5.3 Group and Organization Interfaces: Theory and models / K5.m Computers and Education: misc.

INTRODUCTION
The Web in general and new publishing technologies such as Wikis, Blogs and RSS in particular provide alternatives to the traditional publishing model. Importantly, these new technologies work particularly well with open-access journals (e.g. Public Library of Science -PLoS) and e-Print repositories (e.g. arXiv, PubMed Central) that are mechanisms for rapid publication and access to latest results. These new publication avenues give authors and institutions more control over intellectual property and new technologies for packaging and presenting scholarly results and for ensuring interoperability among distributed repositories. Appraise addresses key aspects of a world in transition. The system, which is at design stages, explores the relationship between the technological basis of scholarly communication and the structure of scholarly communities. It asks how documents, the artifacts of scholarly communication, provide the basis for *social worlds*, the context for ideation and scholarly activities. Understanding the nature of this relationship and building appropriate tools to address its needs are the primary goals of Appraise.

New technologies for scholarly communication facilitate social and scholarly interaction in a space that might be thought-of as the 'long tail'[1] of the scholarly publishing chain. Particularly, on the view that like many other phenomena in nature, scholarly ideas are developed according to a power-law distribution it is evident that the 'long tail' represents a knowledge realm whose depth and breadth might eclipse those of the head of the chain. Consequently, given the appropriate tools authors and readers can be freed from the narrow bandwidth of traditional journals like Nature and Science which are part of a collapsing system which is imploding due to financial, technical, legal and social difficulties. Instead, given the appropriate mechanisms to exploit this long tail, publication in open-access journals or even in blogs and wikis can become significant. This transition however, which, by and large, abandons the traditions of expensive peer review and a long editorial process raises significant questions regarding the management of authorship, reputation and expertise. It also illustrates the need for appropriate tools that do not exist today—in this context—to enable the discovery, evaluation and dissemination of such, so called, semi-formal scholarly artifacts.

INNOVATION IN APPRAISE
Appraise will explore and extend existing social networking models and link analysis techniques with a new models of hybrid networks that combines actors and digital artifacts. It will investigate, in the context of existing electronic publication systems, the utility of hybrid networks as the basis for tools that improve scholarly communication and ideation, including tools for recommendation, annotation, reputation evaluations, expert finding, and reputation management. It is our intention to learn from tools that have been developed in commercial contexts (eBay, Amazon, SlashDot, Wikipedia) and adopt them to the scholarly environment with its unique constraints. Significantly, the context in which the research will take place and where systems will be integrated and deployed is arXiv (http://www.arXiv.org) the largest e-repository in all areas

of science and technology, which includes over 300,000 articles published to date. Successful deployment in this context will provide an exemplary model for other types of scholar communities and thus hopefully cultivate a new revolution in scholarly publications.

ANALYTIC FRAMEWORK - HYBRID NETWORKS IN SCHOLARLY PUBLISHING

Roosendaal and Geurts [7] used the metaphor of scholarly activity as *agora*, or communities for the generation of ideas. As we know, the nature of these communities changes as the technology for publication and result documentation changes. In Appraise we propose studying the relationship between the technology for scholarly communication and the intellectual activity that it enables. Our work will advance the technology for documentation and develop new tools based on social network analysis and link analysis on a hybrid network of authors and documents as a means of promoting more dynamic interaction of documents and researchers. We aim to further our understanding of the interaction between ideation and documentation in the context of ever-growing amounts of publication.

To clarify the interplay between information artifacts, the forming of scholarly communities, and their institutionalization, we propose a conceptual framework based on information flows. We argue that scholarly publishing and communication are all information exchanges through information artifacts: either formal and materialized artifacts such as journal articles and conference proceedings, or informal and non-conclusive artifacts such as email discussions, Wikis, Blogs, RSS feeds and informal meetings. This model, we argue, is essential to the understanding of the true potential of these new modes of scholarly communication as multi-threaded spaces of conversation.

Three current classes of documents can be identified: formal documents, in the forms of officially published journal papers and conference proceedings; semi-formal documents, such as the e-prints in arXiv; and informal documents, such as comments, reviews, ratings, discussions, in emails, wikis, blogs, personal web pages, etc. The official and semi-formal publication systems form document networks that can be investigated using citation analysis. As we know, the development of electronic publishing exemplified by arXiv has collapsed the time dimension of citations: one hundred years ago you might only be able to cite papers published in the past year; now you can cite contemporaneous papers stored in e-print. Evaluation of a citation graph reveals popularity of documents and even suggests quality [3-5]. The reasons for citing others include quality considerations, similarity of arguments, availability, self-promotion, and social bias, indicating that the system is far from perfect [2]. However in existing analysis techniques, a large amount of informal communication and the related scholarly artifacts are neither “materialized” nor analyzed, and are consequently excluded from quality and popularity measurements. These channels of informal communication compose a rich array of information flows, which foster the creation of original ideas and help mature scientific theories. Adding both researchers and documents of all classes to the same graph creates a Hybrid Network model. By constructing such a hybrid network model we build a more complete view of the communication process than we cold have done by looking at communication via any single medium.

By formalizing the hybrid network as a graph we see that the connections are bi-directional, the in- and out- degrees follow a power-law distribution, and that the network has a high cluster and low diameter, possessing the “small-world” attributes [6, 8].This transformation to a hybrid network is illustrated in Figure 1.

SOFTWARE PROTOTYPES TO BE BUILT

Software development will focus around four areas: (a) annotations; (b) recommendations; (c) measures of reputation; and (d) social network visualization. Each node in the network (traditional document, actor, new document) has an identity and metadata associated with it, which can be used for discovery, selection, citation, etc. We aim to build search and visualization tools based on social network and link analysis algorithms that use this metadata in order to facilitate the discovery of high quality and high impact authors and documents; help to visualize different sub-communities of authors and documents; plot the visualization of the evolution of this scholar community in terms of both social network and document network. It is important to stress that these tools will be available for the use of the scholars in real time.

CONCLUSION

In summary, we hope to encourage open communication and ideation through the creation of a cohesive framework that permits the integration of traditional articles with less formal and richer notions of documents. We hope to learn from some important steps that have been taken in this directions in a commercial context.. We hope that with the introduction of Appraise parts of the scholarly process that might otherwise be lost within the ‘long tail’ will become part of a hybrid network of actors and documents that will permit more fluid notions of discovery, selection and certification that include not only those associated with the conventional scholarly publishing system (journal ranking; publication type, citation amount) but also measures that depend on network structure as the basis of reputation and expertise. By shifting into a new analytical framework that views scholarly communication as a hybrid network we expect to gain a better understanding of scholarly communication and build tools that exploit what has hitherto been considered auxiliary activities. By building the system within the world’s leading open scientific repository we hope to demonstrate how such new tools are both applicable and useful.

ACKNOWLEDGMENTS

Some of the early stages of this research were funded by a grant from the Social Science Research Council as part of the Culture and Creativity Program / Digital Cultural Institutions Project.

REFERENCES

1. Anderson, C. The Long Tail. *Wired Magzine, 12* (10).
2. Borgman, C.L. and Furner, J. Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology, 36*. 3-72.
3. Garfield, E. *Citation indexing: Its theory and application in science, technology, and humanities*. John Wiley, New York, NY, 1979.
4. Garfield, E. and Welljams-Dorof, A. Citation data: Their use as quantitative indicators for science and technology evaluation and policy-making. *Science and Public Policy, 19* (5). 321-327.
5. Garfield, E. and Welljams-Dorof, A. Of Nobel class: A citation perspective on high impact research authors. *Theoretical Medicine, 13* (2). 117-135.
6. Kleinberg, J. Navigation in a small world. *Nature, 406*. 845.
7. Roosendaal, H.E. and Guerts, P.A.T.M., Forces and functions in scientific communities: an analysis of their interplay. in *CRISP 97: Cooperative Research Information Systems in Physics*, (Oldenburg, Germany, 1997).
8. Watts, D.J. *Six Degrees: The science of a connected age*. W.W. Norton & Company, New York, 2003.

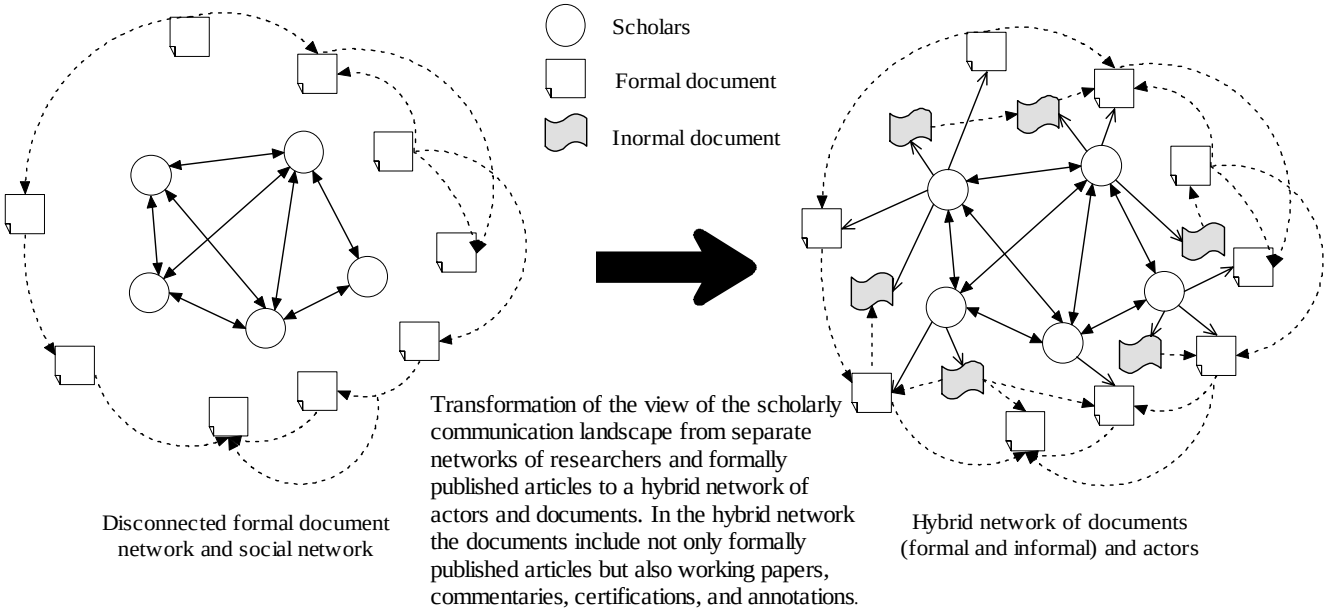

Figure 1: Transformation of a Scholarly Network