Building successful online research networks with the Last.fm model

Jason J. Hoyt[1], Jan Reichelt[1,2], and Victor Henning[1,3]
1. Mendeley Ltd., London, UK 2.University of Cologne, Germany 3. Bauhaus-University of Weimar, Germany

The emergence of online social networks such as Friendster, MySpace, Facebook, and LinkedIn changed the way people socially engaged and interacted with each other, both in private and in professional life. Could the principles driving these websites be carried over to scientists and the world of research though? Between 2006 and 2008, more than a dozen online social networks for scientists launched. And for the most part, these could all be considered a failure. So, what happened? What was so different from the mainstream social networks that didn't work with academic social networks? Could the lessons learned be used to create "valuable" researcher networks? Are there any other types of online networks that could in fact be mimicked successfully?

As of June 2009, Facebook had more than 200 million members with 50% of those logging on at least once a day. LinkedIn had nearly 12 million unique visitors in May 2009 and more than 45 million visits overall in the same month. Meanwhile, it is estimated that there are up to 145 million academics, students, information professionals, and industry researchers who could benefit from online collaboration. One of the common themes in online research is to make use of social bookmarking for research papers. Usually, a commenting system is attached to the bookmarked citations. These commenting forums invariably go unused and are abandoned soon after inception. We argue that there are three elements, which have contributed to the failure of academic social networks. First, there is little to no value-add in a pure social network for researchers. Second, there was the chicken and egg problem. And third, to establish an online social network for researchers, there was more a psychological/social problem to overcome, rather than a technical one.

A site such as Facebook was able to create value by, initially, offering a private online place for college friends. LinkedIn offered value in the way of business networking. A large percentage of that estimated 145 million academic/researcher sized market is already a member of Facebook and/or LinkedIn. That membership is in fact one reason they do not join

an academic network, because there is no additional benefit in joining yet another network for pure social networking reasons. The online network problem has already been solved, and so there is no additional value-add in joining a niche science network, even if this offered a special set of research-specific features. So, the first contributing element was that the social network was created first, rather than focusing on tools a researcher could immediately use without the critical mass requirement.

So if academics wouldn't join for the social aspect, as all their friends and colleagues were already on other social networks, another reason would be because specialized social networks for academics would offer additional functionality. As already mentioned, many academic social networks often featured commenting systems around citations and abstracts. The intention was that academics would engage in online discussions and make post-publication review a dynamic social event, rather than just a static individual pursuit. However, there was an unforeseen chicken and egg problem with this idea: In order to get a commenting system up and running, there must already be a crowd present. And without that crowd, there is very little incentive to participate when there are no other comments.

The third element was that the psychological and behavioral norms of academics were not taken into account. What the architects of these academic networks didn't account for was that scientists are semi-open when it comes to sharing. This difference in persona directly led to a lack of participation in online discussion forums. Additionally, there were no incentives created, such as a reputation awards system to encourage active participation. Even efforts to make commenting systems anonymous proved to be the wrong turn. The main driver here is that there is still a large, unfounded fear in the academic community about Open Science.

Make the data social, not the scientist

In our opinion, the answer to overcome these problems is to confer some type of utility to

researchers beyond the idea of participating in a "social network." More specifically, the research network should only create a social network as a secondary purpose, and must have some other primary usage value to a researcher as an individual. The focus should be on the data and resources researchers work with regularly. The other requirement is that the data and materials need to be in common usage with the community of researchers. And finally, it needs to be an application that would work in a world of one user.

We propose that by making research papers (items that researchers work with on a daily basis) 'social objects,' networks will grow organically around the literature and its associated data. To start with, we can take a look at other objects that have achieved tremendous growth as just described, one such object is music. Last.fm, which bills itself as a "social music service," tracks music listening behavior on the user's computer. The software enables the user to receive a personalized radio stream and music recommendations from Last.fm's collaborative filtering system, as well as discover other users with a similar taste in music. Last.fm, in turn, is able to generate usage-based metrics (pervasiveness of songs, number of repeat plays, and tags) for songs, bands, and genres of music. With this model, over a period of five years, Last.fm has achieved to build one of the world's biggest music databases.

We have extended the Last.fm model to academic literature. By making the data or paper the object, rather than the person, there is an immediate value-add, even when it is a network of one person. We have developed a system called "Mendeley" that, upon initial inspection, appears to be just a reference manager for organizing PDF files on the computer. This system is seen as beneficial to the lone researcher who may not see any added value in joining another online social network. And in fact, it isn't a social network, but rather an incubator for one to naturally occur should the community move in that direction. With the reference manager in place, usage statistics and trends are then discovered, just as with the Last.fm music service. Again, this adds additional value in contrast to a stand-alone network.

The end goal is to build collaborative tools for researchers and encourage sharing, be it through commenting or even primary research. To overcome the chicken and egg problem discussed above, we have designed the system to support the natural habits of researchers. For instance, in the offline world a research paper will undergo individual annotation and a form of categorization. Only then will that paper be shared with colleagues in the lab or during journal clubs. Community discussions then follow. To replicate that in the online world, users can directly annotate and tag the PDF itself within the reference manager. Those annotations can then be collaboratively shared with colleagues in "collections," which are in effect, documents grouped by topic.

We are also following the semi-open nature of researchers. There is a fine balance between their desire to be open with thoughts and data and fear of being scooped or ridiculed. Rather than bending the social norms to the technology, we are shaping the technology to the social norms. For instance, these document collections are not publicly accessible for anyone to annotate. They are kept private amongst research groups where sharing is already occurring offline. We have also created a secure environment in which documents and annotations need not be uploaded online to a central server. The Mendeley software consists of an online component, but the security and real power is in the desktop client. Thus, a researcher can choose to keep documents offline and never sync them to an online server (although syncing does offer additional value-add). Currently, the desktop client-online server is a hosted solution, but we are working towards an enterprise version, such that institutions can host the client-server themselves for additional security.

Mendeley has been in open beta testing since January 2009. Through June 2009, it has seen more than 30,000 client downloads spread across multiple countries. It is growing at approximately 35% each month and users have uploaded more than 2.25 million unique documents, doubling every seven weeks. Perhaps most importantly, it is being adopted not just by graduate students who are digital natives, but by tenured faculty as well. While still early, the success here suggests that the Last.fm model can be applied to academic literature, and perhaps other resources, to build large online research networks.