

Professional Networks in the Life Sciences: *Linking* the Linked

Thomas S. Deisboeck and Jonathan Sagotsky

Complex Biosystems Modeling Laboratory, Harvard-MIT (HST) Athinoula A. Martinos Center
for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, 02129, USA.
Corresponding author email: deisboec@helix.mgh.harvard.edu

Abstract: The world wide web has furthered the emergence of a multitude of online expert communities. Continued progress on many of the remaining complex scientific questions requires a wide ranging expertise spectrum with access to a variety of distinct data types. Moving beyond peer-to-peer to *community-to-community* interaction is therefore one of the biggest challenges for global interdisciplinary Life Sciences research, including that of cancer. Cross-domain data query, access, and retrieval will be important innovation areas to enable and facilitate this interaction in the coming years.

Keywords: Life Sciences, web community, data-sharing

Cancer Informatics 2010:9 189–195

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

In recent years, community websites such as Facebook [<http://facebook.com>]^{1,2} and MySpace [<http://myspace.com>] have moved social networking into the spotlight.^{3,4} Although these two communities are broad, others are focused on specific interests, for instance photography (Flickr [<http://flickr.com>]), microblogging (Twitter [<http://twitter.com>]),^{5,6} video (YouTube [<http://youtube.com>]),^{7,8} and shared book-marking ([<http://del.icio.us>]).⁹ And while industry is still figuring how to capitalize on this phenomenon, professional communities such as LinkedIn [<http://linkedin.com>]¹⁰ have already begun to take a more business oriented approach to social networking.

LinkedIn goes a step farther than other social networking sites in how it manages connections between users. Most social networking sites have a single distinction for connections—a friend or contact list. Any user in the system can indicate attachment to any other user in the system. Depending on the social network (and user security settings), attachment is either granted immediately or upon confirmation from the other user. LinkedIn however requires a confirmation of a professional connection to the other person. A professional connection can be verified by an up to date email address, a common employer, or a mutual friend who brokers the connection.⁶ This maintains LinkedIn as a web of coworkers, past and present, as opposed to Facebook's web of people who may have once met socially.

Many of these communities are open to the public while others are intentionally kept private. Sites, such as ExpertMapper [<http://expertmapper.com>],¹¹ maintain large lists of experts organized according to region and interest. These lists are useful for finding a local specialist, but provide no means for communication. While it is possible to use this kind of site to make connections and form a network, the site itself provides no infrastructure for communication. So while ExpertMapper and similar directories may include hundreds or thousands of experts, those experts are not actively participating community members, but entries in a virtual phone book.

In the Life Sciences, communities such as Biomed Experts [<http://www.biomedexperts.com/>]¹² bring professional social-networking to the medical research world. Others include Spidera [<http://spidera.eu>], a European Commission-supported social network that

operates as an interface of academia and small-to-medium enterprises in the health care and life sciences domain, Nature Network [<http://network.nature.com>], which is a professional networking site for scientists worldwide, and CViT.org [<https://www.cvit.org>]^{13–15} (see Fig. 1) an NIH/NCI-supported online community dedicated to supporting cancer modeling and simulation.

While these web-based expert networks claim to facilitate interaction and promote data sharing in the interest of optimizing collective output, such web-based communities also experience significant challenges. The obstacles faced by online communities are broad in range: a provenance system that maintains ownership of shared data and addresses copyright issues; securing sustained funding, necessary to provide fresh and relevant content for a growing community; and to continue the buildup of infrastructure to distribute this content.

Discussion

However, the central challenge for many of these highly focused communities is that the very essence of their appeal, a high level of expertise paired with the prospect of facilitated peer-to-peer interaction, determines the content that they have to continue to offer to keep their core user base engaged. The result is an initial period of accelerated growth at the price of self-selecting isolation; that is, eventually most of these communities will have to slow their expansion as the available highly trained talent in that expertise segment is limited: growth in an expert community stagnates once the supply of experts is exhausted. Students participating in the community will eventually graduate, advance and can become experts in their field thus raising the total level of available expertise. While this may seem like a very reasonable long term investment, it is risky short-term. That is, arguably, a well known expert has little to gain instantly from joining a community of students or lay personnel; by diluting expertise a community risks lessening its appeal to established experts who can contribute now. Life Sciences research, however, is not a members-only club. Work in isolation cannot be a road to success as there is growing consensus that most remaining grand challenges in the Life Sciences require large-scale interdisciplinary approaches that per definition will exceed the expertise, data, and tool sets available in any one of these communities.

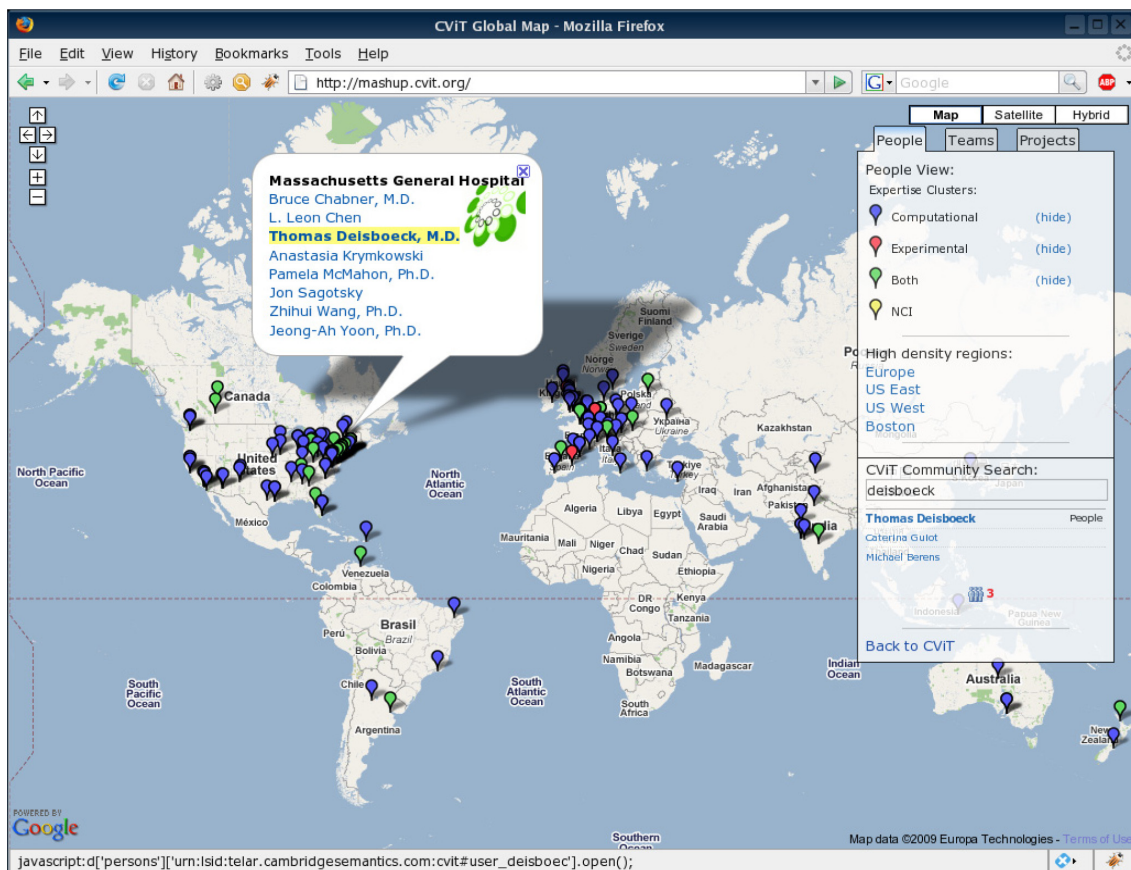


Figure 1. CViT global map. CViT global map lists all CViT users. Also available are displays of projects and files stored in the semantic layered Digital Model Repository.

We therefore argue that there is significant and as of yet untapped value in interacting with communities in neighboring expertise areas and thus accessing additional expertise, data, metadata, models and tools that have the potential to accelerate domain-specific research.¹⁶ The question is then twofold: (i) how to maintain community growth from sources outside the core expertise area which will require opening up the content spectrum, while (ii) minimizing alienation to its core constituency? How can separately developed professional networks communicate with each other? In other words, how can we move from the current peer-to-peer to community-to-community interaction?

To answer this, we must first consider what exactly community-to-community interaction is. Perhaps the closest example is an academic journal. Members of a research community compile their knowledge into a series of papers; if peer-review asserts that the papers are of sufficient quality and interest they are published in a journal and disseminated to neighboring

communities with access to this journal. This is a community-to-community broadcast, rather than interaction. Any interaction or feedback that results from the paper takes place outside of the journal and becomes peer-to-peer interaction instead. We would like to enable this level of back and forth communication, without taking a step back from intersubjective community space to do so.

Solutions for this problem will have to be both technological and sociological. The core user group will have to understand and ‘buy’ into the added value of growth beyond the immediate domain expertise which in turn necessitates not only a conceptual shift in approaching the problem but also a multitude of technological advancements. Those will include sophisticated search tools that can query, access, and retrieve data from domain-specific databases across a variety of overlapping communities, which at the minimum requires shared ontologies, standardized metadata, and cross-authentications of access credentials.



A number of blogs and other informal communities have achieved *cross-authentication* by using OpenID [<http://openid.net>] for authentication. OpenID provides a single identity which can be used to log in to any OpenID enabled site. This kind of service is useful for traversing the blogosphere, but is it trustworthy enough to handle provenance of data with potential intellectual property value? An authentication service such as OpenID can only be as secure as its weakest link. Furthermore, even if OpenID can serve as authentication across communities, authentication is the only problem it will be able to solve. OpenID allows for authenticated logins of an individual user. While this is sufficient for a person to visit a community, the ensuing interaction would be at the community-to-individual level, not at community-to-community—that sort of interaction is beyond the scope of OpenID.

Another innovation of the blog world which brings us closer to community-to-community interaction is Linkback (also known as Trackback or Pingback depending on implementation). Each of these methods is used to inform blog authors of references to their articles. When a blog is published on a Linkback enabled server, the server sends a message to all links mentioned in the blog post. If the recipient of one of these messages is also a Linkback-enabled blog server, that server becomes aware that its posts are being mentioned elsewhere on the web. What happens at that point depends on the blog software. Usually the original blog post is updated with links to the referring posts. This effectively opens a portal between the two communities (and if OpenID is involved visiting the other side of that portal becomes several steps easier) but requires an active effort on the part of members of each side to communicate with the other side. Perhaps a more active way to encourage intercommunity participation would involve blog software which, instead of representing the Linkback as a link, displayed its entirety as a follow-up blog post. Blog software could even aggregate user comments together into one discussion between two or more distinct groups.

Searching another community presents a greater challenge. While it is trivial to index the text of a blog post, many online scientific communities will be engaged in more than just blogging. They will be uploading and sharing their research data which

may be useful to neighboring communities. Without violating data privacy,¹⁷ these community sites will need to be able to index each other's shared content which will be stored in a variety of different formats and storage schemas. Consider that this should encompass more than just a text or media search. For instance, a researcher who stores the results of his simulations in a database should be able to provide access to that database to neighboring communities. Not only should the contents of the database be discoverable, its meaning should be accessible as well. The use of semantic web ontologies would be an excellent way of accomplishing this. One such ontology is the Semantically-Interlinked Online Communities project (SIOC) [<http://sioc-project.org/>] which provides machine readable descriptions of web communities such as blogs and forums. SIOC is limited however in that it only publishes indexes of content—it does not address where that content goes or who gets to read it.

Similarly, *data privacy*¹⁸ could be ensured with software inspired by OpenID. Each community would need to be its own authentication provider. In addition to authentication, this software would have to manage access control. And it would have to do this at both the user and group level. Taking the simulation researcher with a database example, initially the database would only be shared with his principal investigator (PI) and coauthors (see Fig. 2). The PI would likely authenticate against the same community as the researcher. Coauthors on the other hand may be members elsewhere. Since they lack accounts in the researcher's community, the server should be able to message the coauthor's community to authorize a login. Should the author decide to publish the research database when the paper sees print, the server would accept any login from the coauthor's community, not just the coauthor himself. Sharing these resources is not just common courtesy—it is good science as more distinct data pieces can be accessed in an effort to addressing the remaining, more complex scientific questions.

This sharing paradigm is not dissimilar to the Digital Model Repository (DMR) run by CViT.org. The semantic-layered DMR infrastructure allows users to upload modeling files and to share those with other DMR users, either as individuals or as members of specific institutions. Because data files often have

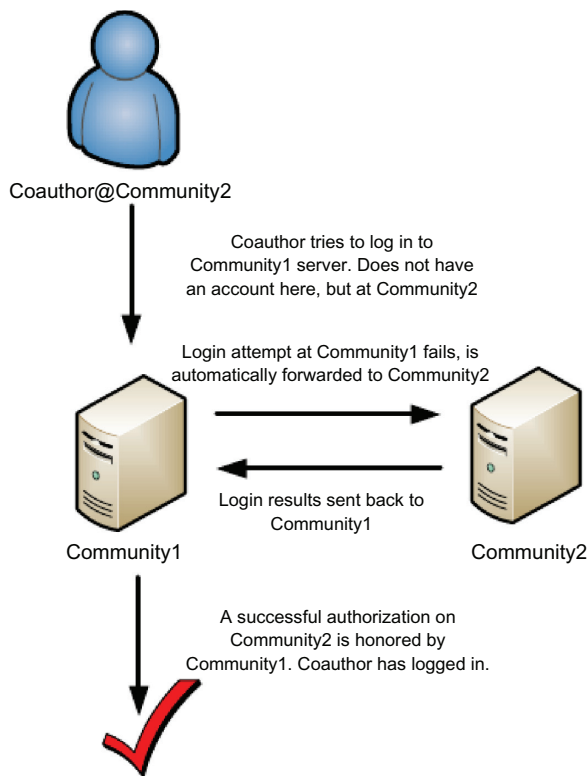


Figure 2. Login authentication forwarding. The coauthor of a paper is able to log in to the author's community site if it is configured to honor authorizations from the coauthor's own community.

intellectual property value, all outgoing sharing has to be approved by an institution's licensing officer. This approval is tracked to ensure provenance of data ownership. This allows data to be shared between neighboring groups, while respecting the intellectual property rights of each group. That said the DMR is a single site. Even though the groups within CViT

can be seen as distinct communities (see Fig. 3), they all exist in the same system on the same server. A true community-to-community interaction would have to take place between different community sites on different servers. Nonetheless, CViT allows and encourages discussion between researchers who would otherwise have no contact with each other. Furthermore this discussion takes place in near real time, allowing for research to be shared immediately, instead of at a rate determined by publishers.

As always, *data security* is a concern, particularly as it relates to protecting clinical data and/or data with intellectual property value. Any service designed to share such data between disparate communities would be under scrutiny. Such a service would have to be absolutely compliant to a series of security measures including but not limited to ensuring data readability privileges, security in data transmission authenticating both the source and recipient of the transmission, and secure user accounts. This is tedious but it is also a significant step forward over the current standard—local storage with file transfer taking place over email. As long as files are passed back and forth between collaborators by email there will exist the potential for a digital break-in. Those files can stay in an inbox indefinitely. Even users who tidy their inboxes regularly may be vulnerable if their mail provider's backup archives are compromised. Instead of local storage with email based transport and authentication, researchers should be using a system designed to store and share data securely. While securing such a system is difficult, it is less difficult

Table 1.

Community	Blogs	Messaging	Contacts	Profiles	User-base	Focus
Biomed Experts http://www.biomedexperts.com/		✓	✓	✓	*	Life Sciences
BioSpace http://www.biospace.com				✓	?	Biotechnology and pharmaceutical industries
CViT https://www.cvit.org	✓		✓	✓	309	Cancer modeling
Expert Mapper http://www.expertmapper.com/				✓	*	Finding experts
Nature Network http://network.nature.com	✓	✓	✓	✓	25000	Professional networking and collaboration
Spidera http://www.spidera.eu/	✓	✓		✓	365	Life Sciences

Notes: *Biomed Experts and Expert Mapper both contain a great many user profiles. These profiles are not user submitted. They are scraped from published papers. In the case of Expert Mapper, researchers who publish under different names will have duplicate listings.

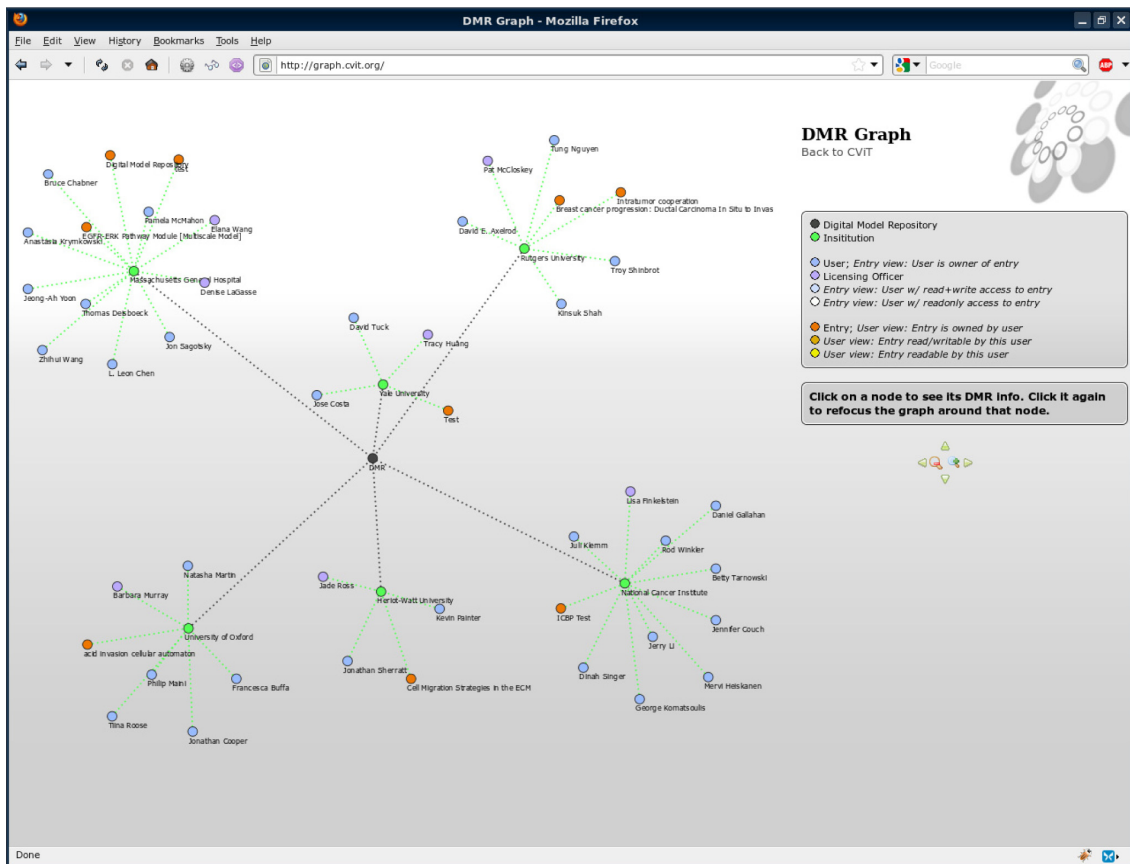


Figure 3. DMR graph. CVIT graph shows institutions in the semantic layered Digital Model Repository and their respective users/projects.

than securing millions of lab computers and research inboxes. A service like this is necessary for communities of researchers to be able to interact online in a meaningful way.

Conclusion

In summary, community-to-community interaction is one of the biggest challenges faced by developing scientific online communities. Global interdisciplinary Life Sciences research must figure out smart ways to have separately evolved expert networks and their infrastructures communicate. Cross-domain search, data retrieval, and ontology issues will remain important innovation areas to enable and facilitate this interaction.

Acknowledgements

This work has been supported in part by NIH grant CA 113004 and by the Harvard-MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging and the Department of Radiology at Massachusetts General Hospital.

Disclosures

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers of this paper report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

References

- Gorrindo T, Gorrindo PC, Groves JE. Intersection of online social networking with medical professionalism: can medicine police the facebook boom? *J Gen Intern Med.* 2008 Dec;23(12):2155; author reply 2156.
- Thompson LA, Dawson K, Ferdig R, et al. The intersection of online social networking with medical professionalism. *J Gen Intern Med.* 2008 Jul;23(7):954-7.
- Dwyer C, Hiltz SR, Passerini K. Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. *AMCIS.* Vol Paper 339. Keystone, Colorado; 2007.
- Cain J. Online social networking issues within academia and pharmacy education. *Am J Pharm Educ.* 2008 Feb 15;72(1):10.
- Gamble KH. Just a tweet away. *Healthc Inform.* 2009 May;26(5):30, 32, 34 passim.
- Graham LK. What is social networking? And how do I get clued in to LinkedIn? *J Am Diet Assoc.* 2009 Jan;109(1):184.



7. Farnan JM, Paro JA, Higa J, Edelson J, Arora VM. The YouTube generation: implications for medical professionalism. *Perspect Biol Med*. Autumn 2008;51(4):517–24.
8. Steinberg PL, Wason S, Stern JM, Deters L, Kowal B, Seigne J. YouTube as source of prostate cancer information. *Urology*. Mar;75(3):619–22.
9. Clink K. Del.icio.us. *Refernce Reviews*. 2008;22(5):1.
10. Skeels MM, Grudin J. When social networks cross boundaries: a case study of workplace use of facebook and linkedin. *Conference on Supporting Group Work*. Sannibel Island, Florida, USA: ACM; 2009.
11. Bailey DS, Zanders ED. Drug discovery in the era of Facebook—new tools for scientific networking. *Drug Discov Today*. 2008 Oct;13(19–20):863–8.
12. Whitaker IS, Shokrollahi K. BiomedExperts: Unlocking the Potential of the Internet to Advance Collaborative Research in Plastic and Reconstructive Surgery. *Annals of Plastic Surgery*. 2009 Aug;63(2):2.
13. Deisboeck TS, Zhang L, Martin S. Advancing Cancer Systems Biology: Introducing the Center for the Development of a Virtual Tumor, CViT. *Cancer Inform*. 2007;5:1–8.
14. Sagotsky JA, Zhang L, Wang Z, Martin S, Deisboeck TS. Life Sciences and the web: a new era for collaboration. *Mol Syst Biol*. 2008;4:201.
15. Deisboeck TS. Personalizing medicine: a systems biology perspective. *Mol Syst Biol*. 2009;5:249.
16. Schleyer T, Spallek H, Butler BS, et al. Facebook for scientists: requirements and services for optimizing how scientific collaborations are established. *J Med Internet Res*. 2008;10(3):e24.
17. Elger BS, Iavindrasana J, Lo Iacono L, et al. Strategies for health data exchange for secondary, cross-institutional clinical research. *Comput Methods Programs Biomed*. Jan 18.
18. Malin B, Karp D, Scheuermann RH. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. *J Investig Med*. 2010 Jan;58(1):11–18.

Publish with Libertas Academica and every scientist working in your field can read your article

“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”

“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I’ve never had such complete communication with a journal.”

“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>