Accelerating Science Using the Science Network

John Harney
National Center for
Computational Sciences
Oak Ridge National Laboratory
Oak Ridge, Tennessee, USA
Email: harneyjf@ornl.gov

Roselyne Tchoua
University of Chicago
Chicago, Illinois, USA
Email:

Gavin Bell
Lawrence Livermore National Laboratory
Livermore, CA
Email:

***Abstract*—Social networking is a phenomenon that has revolutionized the way in which the world communicates. Applications like Facebook have become platforms for individuals to connect and share content with each other. Intuitively, many scientific disciplines may find utility in social networks, as they provide ideal environments for supporting collaboration. A number of efforts were initialized to bring such applications into production. Many, however, have received a lukewarm reception from the scientific community. This is due in part to the fact that these scientific social network applications, borrowing from established social networks like Facebook, remain human-centric, where content is meant to be processed by humans. As a result, scientists still must conduct science offline. We address this problem by introducing the Science Network, which enhances the traditional social network model with scientific discovery-centric associations. This provides scientists a platform for conducting their scientific activities online while seamlessly sharing their findings with others within the scientific community. Furthermore, the Science Network presents an opportunity to leverage algorithms that allow machines to process content within the network. This may supply interesting trends and patterns in the nature of scientific study in general. In this proposal, we will implement and research tools and technologies to realize the Science Network vision.**

I. Introduction

The potential of social networks have not been leveraged effectively in the context of science, the scientific community and scientific collaboration. This proposal seeks to bring the true utility of social networks to science, particularly with respect to scientific discovery, sharing and collaboration. The Science Network will bring together these aspects, along with a practical platform for performing experiments, peer review and vetting of scientific discovery. Social networking is a phenomenon that has revolutionized the way in which the world communicates. Applications like Facebook [1], Twitter [2], and del.icio.us [3] have become so popular that they have profoundly influenced the way people conduct their daily lives. People are able to represent themselves publicly through the web and connect to others by sharing content ranging from simple comments to other more complex media such as video streams. Often, an important and serendipitous benefit of social networks is the ability to create concrete models that enable knowledge discovery possibilities about the nature of how people interact with each other. Intuitively, scientific communities should find utility in social networks [4] . Canady [5 he value of social networks for scientists lies in faster access to information relevant to their research and the communities that are made more available by new tools. These networks provide a forum for scientists to converse and engage each other on current scientific activities. Connections can be conveniently made for people in a more expedient manner online, rather than rely on traditional mechanisms that are hampered by geospatial (e.g. face-to-face meetings) and temporal (e.g. telecommunications) constraints. For example, climate modelers that participate and contribute to the Earth System Grid Federation (ESGF) data portal system may want to share observations about particular simulations and experiments. Rather than enforce teleconferences or all-hands meetups, users of ESGF can contribute these scientific activities at their leisure. This type of openness leads to faster, more collaborative efforts that can be used to optimize the scientific discovery process.] states that t The concept of a social network for scientists is hardly novel, as a number of implementation efforts have been initiated. One of the first applications to take advantage of the growing popularity of social networks is Mendeley [6], which allows scientists to manage and share publications related to their research. Users of Mendeley – typically scientists who have published in their respective fields – have access to a number of key social networking technologies (e.g. news feeds, comments, profile pages, collaborative tagging mechanisms, etc) that they can utilize to make connections with other Mendeley users. Mendeley then has the computational capacity to make intelligent associations between users of the system based on their publications and their research interests. Another interesting scientific social network is ourSpaces [7] . OurSpaces builds on the concept of the Web-based Virtual Research Environment (VRE) [8] to allow scientists to log their activities online and socially connect those activities to other scientists. They build upon the Open Provenance Model [9] to log these activities so that users can view historical content of scientific study and utilize principles of the semantic web to resolve heterogeneities over different activity definitions. Perhaps the most ambitious effort currently in production is ResearchGate [10], whose user base exceeds 1.4 million scientists from many different disciplines. Researchers are able to connect to a global scientific community and make their work visible. This is especially appealing to younger, inexperienced scientists that would like to make relevant connections to top people in their fields. It harnesses the power of the social graph [11] and presents a user interface similar to popular Facebook, providing new users a relatively small barrier of entry. Users receive feeds of scientific activities of their friends and are able to post messages and comments about their activities. Despite the early successes of the aforementioned social networks, there are plenty of skeptics who suggest that the concept of the scientific social network is doomed to fail [12] [13]. While bringing several pertinent points to their arguments, these and many other critics fail to realize that their criticism 1 is directed at what we call human-centric social networks. Human-centric social networks, despite leveraging the available social mechanisms and Web 2.0 technologies that foster human-to-human collaboration and communication, share a vital, fundamental flaw – they assume that all information available on the network is readable and processable only by humans. Therefore, scientists who utilize these networks are still forced to work in isolated environments. Individually, they discern the content that they extract from the network and use it in their individualized work. In other words, the majority of their science is still conducted offline. Conversely, a stronger scientific insight can be derived from the ability to read and understand the interactions and scientific discoveries online. This will allow machines to process information about the knowledge that is being generated by and presented to scientists. We thus present the Science Network (SN). The idea behind the Science Network is similar to the human-centric social network with one notable exception - scientific discoveries are now the focal point. We say that a scientific discovery is any advance made in the scientific process. The concept of a discovery-centric social network is derived from the idea that no discovery is an island – discoveries are often a product of one or more discoveries that preceded them. For example, The US Constitution was created by combining ideas from the works, Two Treatises for Government by John Locke, and the The Spirit of the Laws by Charles de Montesque1 . The scientific discovery process works in a similar manner. As a concrete example, a climate scientist may discover a new anomaly in a dataset stored as a netCDF [14] file by using the subsetting tool OpenDAP [15]. If we enrich the human-centric social network model with the discovery-centric methodology, we add an extra dimension full of knowledge discovery possibilities. In addition to people being connected, discoveries can be connected to each other. The relationships between discoveries, in addition to the relationships between people and their discoveries can provide us with invaluable information about scientific activities online. We can then use machine processable methods over this information to infer important aspects about the nature of scientific effort and collaboration . In this proposal, we will implement and research methodologies in order to realize the Science Network vision. The proposal is organized in the following manner. In Section 2, we introduce our main use case, the Earth System Grid Federation (ESGF) Peer-to-Peer (P2P), in which we will integrate the Science Network. Section 3 describes the overall Science Network framework. Section 4 explores avenues of research paths that may be taken to build the Science Network. Section 5 presents a rudimentary milestone and deliverables list that we intend to follow throughout the course of the projects lifespan. Section 6 gives outlines a management plan for the project. Section 7 demonstrates a success matrix. We list our external references in Section 8.

II. Motivating Scenario

Our initial release of the Science Network will be applied to the next generation Earth Science Grid system [Williams et. al.]. The Earth System Grid Federation (ESGF) Peer-to-Peer (P2P) enterprise system is a collaboration that develops, deploys and maintains software infrastructure for the management, dissemination, and analysis of model output and observational data. ESGF's primary goal is to facilitate advancements in Earth System Science. It is an interagency and international effort led by the Department of Energy (DOE). The ESGF mission is to

III. Science Network Overview

A. Data Model

1) The Science Network Model:

2) The Science Network Model Example:

B. Execution Platform

C. User Interface

1) The Human-Centric User Interface:

2) The Discovery-Centric User Interface:

IV. Science Network Research

A. Science Network Structural Learning

B. Science Network Data Learning

C. User Interface

V. Preliminary Results

VI. Related Work

VII. Conclusion

Acknowledgment

The authors would like to thank...

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.