Proceedings of the 2010 Winter Simulation Conference
B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, eds.

EXPLORATORY SIMULATION OF COLLECTIVE INNOVATIVE BEHAVIOR IN GLOBAL PARTICIPATORY SCIENCE COMMUNITIES

Guangyu Zou

Auburn University
3101 Shelby Center
Auburn, AL 36849, USA

Levent Yilmaz

Auburn University
3116 Shelby Center
Auburn, AL 36849, USA

ABSTRACT

Better understanding of how and why networks of open innovation and global participatory science communities form and evolve, and how they can be governed or influenced toward sustainable innovation and productive states are critical questions. To this end, a simulation-based exploratory study is conducted to better understand the conditions that confer increased rates of innovation in such socio-technical systems. Three types of open science communities are identified and simulated using agent simulation as a method of inquiry. Simulation results show that centrality, as a measure of degree of connectedness, exhibits positive influence for innovation output in exploratory and service communities up to a point. Also, utility-oriented communities have social network structures with low density and high centrality, suggesting high potential for innovation.

1 INTRODUCTION

A diverse set of factors are driving the globalization of science. Factors that catalyze these trends include the emergence of new information architectures and collaboratories that enable the emergence of a new mode of collaboration among scientists distributed over the globe to share and co-develop knowledge over the cyber-infrastructure. The practice of science is more open and global, as the access to knowledge, as well as its production is becoming increasingly transparent. Service oriented science (Foster 2005) and e-Science initiatives lead to scientific communities, where shared domain knowledge is no longer solely documented in the journal articles or patents, but is also documented in software, simulations, and databases that represent an evolving collective knowledge-base that is governed and maintained by community members. Just like open source software communities, "SourceForge for science" style in scientific production and collaboration provide the requisite infrastructure that encompass community membership services, catalogs, storage services, and workflow orchestration service. We denote this style that revolves around open and distributed communities of scientific practice as *Global Participatory Science* (GPS).

GPS is a socio-technical practice. Scientists and engineers progress through an orientation phase, during which they become familiar with the norms and governance mechanisms of the community of practice. Recently a number of virtual scientific collaboratories emerged and continue to successfully bring together scientists over the globe to collaborate to not only share and aggregate data, but also create new knowledge (Smith and Ashburner 2007, NanoHUB 2009). For instance, OBO Foundry (Smith and Ashburner 2007) is collaboration among a group of communities that are active in developing ontologies to standardize data acquisition and use in the health sciences community. As collaboratories over the cyberinfrastructure become sophisticated in terms of capabilities that support remote access, collaboration, and cooperative activity management, virtual organizations as open science socio-technical systems are becoming prominent and increasingly central to science and engineering projects. To harness their full potential and the promise they offer in learning, discovery, and innovation, it is critical to understand what structural and behavioral conditions confer sustained innovation and facilitate enhancing scientific and engineering production in such communities.

Using agent simulation, we examine three types of communities (i.e., exploratory, service, utility) and observe emergent socio-technical networks to determine relations between emergent network structures and innovation output. Specifically, we are interested in exploring the following questions:

> what is the impact of alternative scientific community cultures and leadership styles on the sustainability and innovation potential of GPS?

Network metrics such as centrality, density, and clustering coefficient are used to determine if they are related to innovation output under alternative community governance mechanisms. Simulation results indicate that centrality, as a measure of integration, positively influences innovation up to a point. That is communities that generate high degrees of innovation output tends to have more focal nodes in their network structure than communities that have low or moderate degrees of innovation output. Social networks in utility-oriented community exhibit low density and high centrality, which suggest high potential for innovation. Also, utility-oriented communities exhibit large degree of variation in terms of average knowledge fitness defined in terms of knowledge utility and embeddedness.

2 BACKGROUND

2.1 Open Science Communities

The loss of richness in communication, diversity in membership, and increased mobility in open science communities pose unique challenges. Such challenges involved in sustaining scientific collaboratories are examined in (Bos, Zimmerman, Olson, Yew, Yerkie, Dahl, and Olson 2003). Traditions of scientific independence, difficulties in sharing knowledge, and formal organizational barriers are presented among the challenges that influence the sustainability of such communities. On the other hand, recently a number of virtual scientific collaboratories emerged and continue to successfully bring together scientists over the globe to collaborate to not only share and aggregate data, but also create new knowledge. The following are among such growing and active open science collaboratories and innovation networks.

- OBO Foundry- Open Biomedical Ontologies
- NanoHUB - Simulation, Education, Technology for Nano Technology
- NEES Grid - Network for Earthquake Engineering Cyberinfrastructure
- CABIG - Cancer Biomedical Informatics Grid

OBO Foundry (Smith and Ashburner 2007) comprises over 60 ontologies, and its role as an ontology information resource is supported by the NIH Roadmap National Center for Biomedical Ontology (NCBO). OBO Foundry serves as a hub for a network of communities to foster interoperability and to align and integrate their efforts. Some of the participating communities such as Open Biomedical Inveisitigations (OBI), are international, collaborative efforts to build artifacts for annotation of Biomedical Investigations. NanoHub (NanoHUB 2009) is a network comprised of scientists, engineers, and educators that share and exchange resources related to Nano techology. NanoHub is based on an NSF-funded initiative that aims to establish a virtual organization and network for computational nanotechnology. Shared resources include simulations, learning modules such as course materials, publications, presentations, and tools. Similarly NEESGrid, which is an active consortium of earthquake engineering centers, institutes, researchers, and practitioners in United States and around the world and is the result of a collaboration effort led by the National Center for Supercomputing Applications (NCSA).

2.2 Computational Models of Science and Innovation

Earlier studies pertaining to the application of computational models to scientific discovery processes focus on simulating cognitive processes and re-enact discoveries (Klahr and Simon 1999). Specifically, computational modeling of concept formation is viewed as central to discovery and has a long history (Hovland and Hunt 1960).

Although significant research has been conducted on social aspects of scientific communities, simulation modeling of such communities is rare. One notable example is (Gilbert 1997), where the citation patterns and growth of knowledge are simulated to exhibit empirical regularities observed in scientific communities. Yet, this study does not aim to consider social processes pertaining to enculturation and innovation. On the other hand, the simulation study presented in (Edmonds 2007) views scientific discovery as a social process. However, unlike the model presented in this article, its underlying generative process does not take interactions between agents (i.e., scientists) into account. In the context of innovation, the use of simulation of collective invention and innovation diffusion (Cowan and Jonard 2004) revealed the significance of social network structure in knowledge creation and diffusion. Building on these earlier studies, the model introduced herein (1) explicitly specifies the underlying generative mechanisms related social dynamics of knowledge creation and (2) examines the implications of these mechanisms within the context of emergent virtual forms of scientific communities.

3 A COMPUTATIONAL MODEL OF GLOBAL PARTICIPATORY INNOVATION COMMUNITIES

Following the basic tenets of systems model of creativity (Csikszenthmihalyi 1999), Figure 1 presents components of the conceptual model. *Individuals* are scientists that operate on knowledge structures that are generated by *creation*, *combination*, and *elaboration* operations (Ward, Smith, and Vaid 1997).

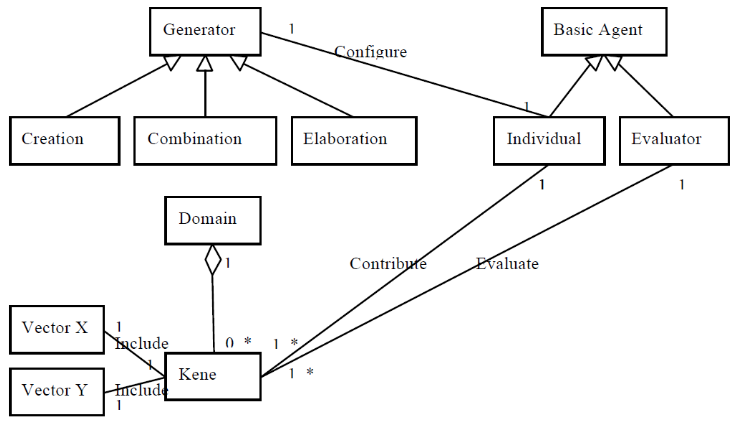

Figure 1: Components of the Model

The *domain* is comprised of knowledge units, called *kenes* (Gilbert 1997) that are contributed by scientists and evaluated for inclusion in the domain based on quality (i.e., novelty) and fitness (i.e., usefulness) of the generated knowledge. As shown in Figure 2, each scientist, modeled as an agent, involves in enculturation (i.e., orientation) and innovation processes.

3.1 Entry and Enculturation

Scientists go through an orientation process during which they adopt certain knowledge and skills, so as to develop new ideas and knowledge built on the existing knowledge base. The influence that scientists that are active in the discipline exert on others and the collaboration that ensues during knowledge creation process yields a continuously self-organizing relation between scientists that in turn effects knowledge generation process.

At every time interval, a random number of new agents enter the community and begin the enculturation process. During the enculturation process, agents move randomly within the knowledge space. As they move, at each time interval they select a random number of community members in their neighborhood to interact with. Using the components of the socialization model presented in (Harrison and Carroll 2006), we model the change in fitness level of an agent as a function of its susceptibility to influence and the intensity of influence it receives from the agents that it interacts with. The first parameter is the *susceptibility* of agent i, which is defined as follows:

$$S(i) = \beta_0 + e^{-\beta_1 - \beta_2 T}, \tag{1}$$

where T is the tenure of the agent in the community, with $0 < \beta_1 < 1$, and $0 < \beta_2 < 1$ so that initial susceptibility is high and it decreases exponentially to an asymptote defined by $0 < \beta_0 < 1$. We also take into account the enculturation intensity (EI) of scientists in terms of their interaction with other scientists. In the following formulation, enculturation intensity $EI(i)$ of agent i is computed in terms of how much influence each agent j exerts based on its reputation α_j. Specifically, agent j pulls agent i based on its current enculturation level $E(j)$ and its reputation.

$$EI(i) = \sum_{j=1}^{K} I_{ji} \tag{2}$$

$$I_{ji} = (E(j) - E(i))\alpha_j \tag{3}$$

Finally, new enculturation level of an agent i at time t is computed by using enculturation intensity $EI(i)$ and susceptibility $S(i)$:

$$E(i,t) = E(i,t-1) + S(i)EI(i) \tag{4}$$