![JASSS logo]

Levent Yilmaz (2011)

Toward Multi-Level, Multi-Theoretical Model Portfolios for Scientific Enterprise Workforce Dynamics

Abstract

Development of theoretically sound methods and strategies for informed science and innovation policy analysis is critically important to each nation's ability to benefit from R&D investments. Gaining deeper insight into complex social processes that influence the growth and formation of scientific fields and development over time of a diverse workforce requires a systemic and holistic view. A research agenda for the development of rigorous complex adaptive systems models is examined to facilitate the study of incentives, strategies, mobility, and stability of the science-based innovation ecosystem, while examining implications for the sustainability of a diverse science enterprise.

Keywords:
Agent-Based Model, Complexity, Innovation, Science Studies, Diversity

Introduction

1.1 In a recent publication by the National Academy of Sciences (NAP 2010a), new knowledge, human capital (Latour 1992), and an environment that promotes innovation and entrepreneurship (Brown and Duguid 1991, 2000, 2001) are promoted as the key pillars for science-based innovation. It has increasingly become recognized that being competitive in the global technology-dominated marketplace requires a sustained highly qualified and diverse workforce (NAE 2006;NSF 2006). On the other hand, there is substantial empirical evidence, demonstrating the degree to which certain demographic groups are underrepresented in the scientific enterprise (NIGMS 2007).

1.2 Nobel Laureate Gary Becker developed a theory, which empirically established that people are as important to an economy as physical capital (Becker 1964/1993;Porter & Lawler 1968). Becker's now obvious observation is central to conscious attempts to influence human capital development for sustained and resilient innovation. Understanding the dynamics of scientific education, attainment and sustainment of diversity, leveled playing field for coopetition (i.e., cooperation and competition), social communication/ integration mechanisms (Monge & Contractor 2003), and their impact on the demography, as well as integrated differentiation of emergent collaboration networks, in which human capital is embedded (Powel & Snellman 2004) can help decision-makers improve innovation capacity and potential. Such understanding can also take advantage of network recruitment and management principles (Dhanaraj & Parkhe 2006).

1.3 An imminent issue of particular interest is the attainment of broad participation in science and engineering workforce (NAP 2010b). For instance, a demographic challenge in regard to scientific workforce development is that minorities are seriously underrepresented in science and engineering, yet they are also the most rapidly growing segment of the population. One of the conclusions of the workshop on Modeling Scientific Workforce Diversity (NIGMS 2007) was that model-building process could stimulate thinking about relationships, critical questions, and information needs in studying human capital of the scientific enterprise. Although coarse-grain system dynamics models were used in the past to predict the growth and development of scientific research, among the limitations of their use include (1) lack of heterogeneity in terms of individuals' decisions, actions, career choices, as well as learning and adaptation, (2) difficulty in analyzing the problem at different scales and levels, and (3) inability to capture interaction and social influence between members of the workforce.

1.4 This essay advocates development of hybrid, multi-level, and multi-theoretical portfolio of models that can be customized to study scientific workforce system as it is (e.g., empirical insight and validation), discover normative mechanisms to improve diversity (e.g., variety, disparity, and balance), innovation capacity, and stability, and perform what-if analysis to explore

alternative incentives to observe emergent population and network patterns. Proposed strategy involves staged development of increasingly sophisticated models that simulate workforce dynamics within epistemic cultures, followed by aggregate models that integrate multiple disciplines to examine their interaction with each other and hub organizations such as NSF. The strategy culminates in the integration of the workforce model with K-12 education systems dynamics model (i.e., stemNetwork) to close the loop. The types of models promoted in this essay aim to improve development of policies effective in promoting a sustained diverse science enterprise workforce. This effort would direct attention to the underlying dynamics that produce successful scientists, help identify questions in need of additional research, and stimulate the collection of the data necessary to answer them.

🌍 Population Modeling of Scientific Workforce

2.1 Population models should explore factors and computational mechanisms that can help explain the changes over time of various types of members in selected domains of the complex scientific enterprise. Two types of models can provide a basis for such study. First, scientific workforce preparation strategies that support postsecondary matriculation of underrepresented minority students who are qualified can be studied by leveraging models of multi-group ecological dynamics.

2.2 By using the ecological dynamics metaphor, one can examine relations between connectivity, stability, and diversity with possible interpretations in the context of STEM Education. Second, a social simulation based on cognitive niche construction can demonstrate how diversity can be improved and sustained within the scientific workforce dynamics. The hypotheses space (i.e., mechanisms) and the experiment/parameter space of models need to be calibrated and evolve to discover the most fit models that explain relevant empirical data collected from databases such as SESTAT, The Scientists and Engineers Statistical Data System, and the WebCASPAR database. These public domain systems provide access to a large body of statistical data resources for science and engineering (S & E) at U.S. academic institutions. Specifically, using the agent-based modeling paradigm, one can focus on the following questions:

- What is a useful and abstract, yet sufficiently powerful population dynamics model that could explain workforce dynamics, as well as how constituent components of the scientific enterprise emerge, grow, and dissolve?
- How can we evolve ensembles of computational models to discover most plausible exogenous (experiment/context space) and endogenous (hypotheses space) configurations that explain the empirical data so as to reproduce historical dynamics regarding workforce population growth, demographics, and variation over time.
- Are there high-level macro patterns such as power-laws, temporal and relational order, punctuated equilibrium, logistic growth, scale-free distribution in the empirical data that may suggest specific self-organization mechanisms such as preferential attachment?

Figure 1. Scientific Discourse as a Network of SEIR Models
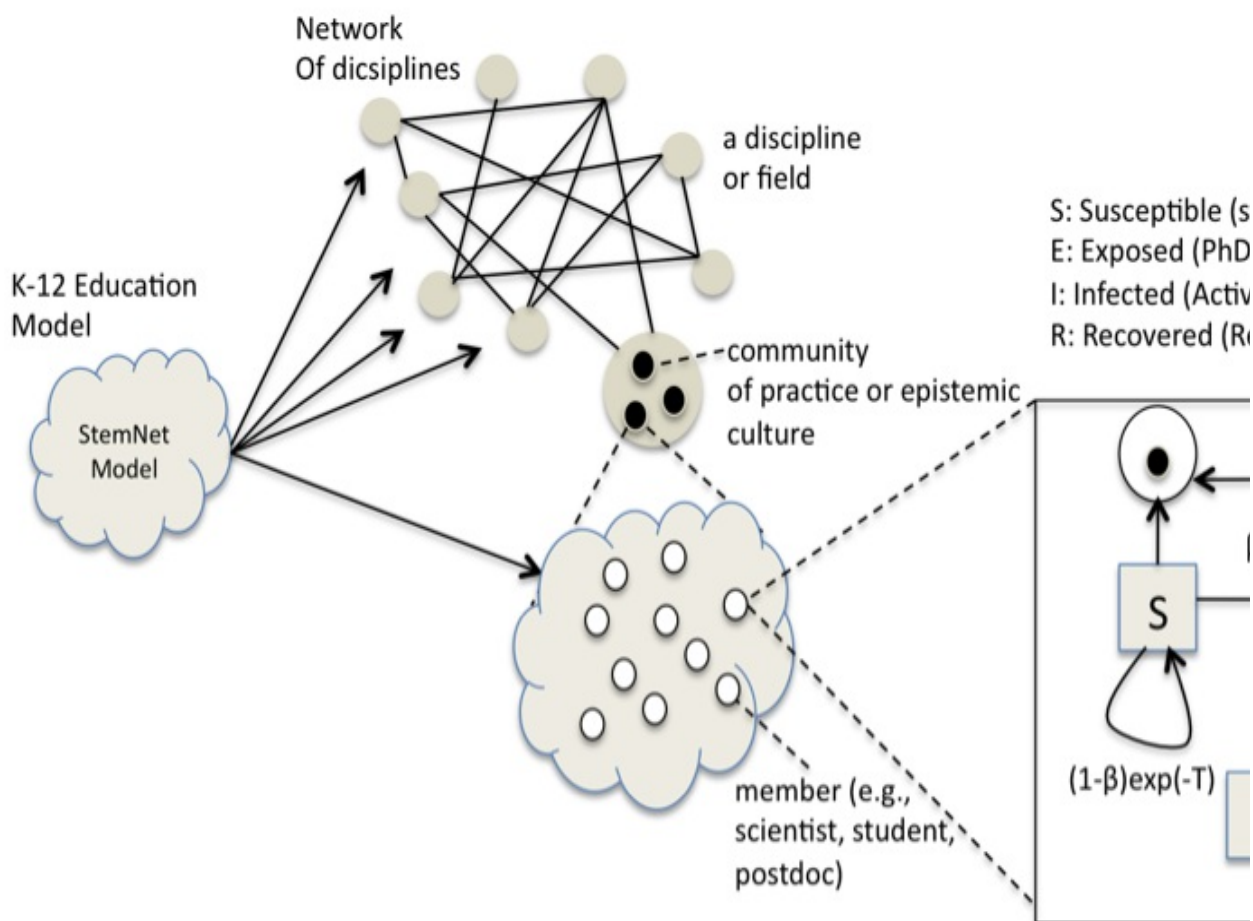
An Abstract Model

2.3   A population dynamics model can focus on the nature of postgraduate scientific discourse using simplified assumptions regarding multiple behavioral life-cycle phases experienced by members of scientific communities during their tenure from graduate school to retirement. These assumptions can be refined by developing conceptually valid as well as hypothetical decision rules to mimic heterogeneous cognitive preferences of individual agents to incrementally and iteratively increase the level of fidelity of the model.

- **Actors:** The abstraction can leverage the metaphor between spread of ideas/influence in academia and epidemics in terms of the SEIR epidemic model, where members of the scientific workforce have individual life-cycle stages that are classified as susceptible, exposed, infected, and recovered. By changing their modes between these alternative states, future scientists (i.e., agents) will transition between student (i.e.,susceptible), apprenticeship (i.e., exposed—PhD student, postdoc), active membership (i.e., infected—publishing), and retirement (e.g., recovered) states. The premise of this metaphor is based on the observation that much as a virus spreads via contact from person-to-person, throughout a susceptible population, so too do ideas infect individuals within epistemic cultures.
- **Social Organization:** The realization of the SEIR phase transitions in terms of individuals' own decision rules and available perceived opportunities require developing behavioral axioms that relate to interactive behavior. That is, instead of purely random and stochastic behavior, agents should increasingly rely on realistic behavior that is grounded on mechanisms that explain scholarly practice. Among the possible extensions include operational specification of contagion mechanism (Monge & Contractor 2003) and exposure via interaction (Yilmaz 2009) that can formalize transitions into exposure and infected states, respectively. Furthermore, integration of multiple SEIR community models into a network of domains within an enterprise, each with multiple communities, requires explicit specification of the organizational structure. Alternative network topologies (e.g., random, random-group, scale-free, grid network, regular graph) can be generated. Individual agents within SEIR models can be augmented with new rules to enable mobility across communities and disciplines. Such extension can facilitate examination of the affects on population distribution of alternative incentives that manage entry thresholds into communities of practice. Findings would be relevant for decision-makers that aim to influence network recruitment and management activities.

Innovation Ecosystem Models of Scientific Workforce Dynamics

3.1   As a socio-ecological system, sustainability of an enterprise depends on the motivation and satisfaction of individuals, as well as coherence of emergent groups. Artificial ecosystem models involve interacting agents (e.g., species) that compete to gain resources from their environment to survive and grow, while also cooperating to develop symbiosis and improve their chance for survival. By including resource/fund allocation and distribution patterns, one can mimic the energy metabolism resource dynamics to extend the basic population model. This extension allows studying interaction of the workforce dynamics with the intentional changes (e.g., incentives) in the environment. Knowledge and human capital in innovation now follows a bottom-up rather than a top-down model and the only way to ensure it can flourish is to create the best possible environment and then get out of the way. Building on lessons from earlier models such as *InnoScape* (Zou & Yilmaz forthcoming), it is possible to conceptualize human capital as a multi-community ecosystem of knowledge producing and diffusing organizations. The primary question of interest in this context is as follows:

- How do we simulate the motivational and competitive behavior over resources, cognitive niche construction, as well as reward system of scientific production and its influence on the scientific workforce dynamics?

A Strategy: A Creative Artificial Ecosystem Model of Workforce Dynamics

3.2   Sustainability of scientific workforce requires availability of open challenges, problems, and resources so that participants can get motivated and stimulated. Participants seek resources by inhabiting and constructing cognitive niches or parts of the knowledge ecosystem where such resources exist. In science, resources may include knowledge, people with skills and abilities, financial support and/or access to tools and instruments. Artificial ecosystem models involve interacting agents (e.g., species) that compete to gain resources from their environment to survive and grow, while also cooperating to develop symbiosis and improve their chance for survival. Among the application of such methods to modeling problems include economics, ecology, and social science.
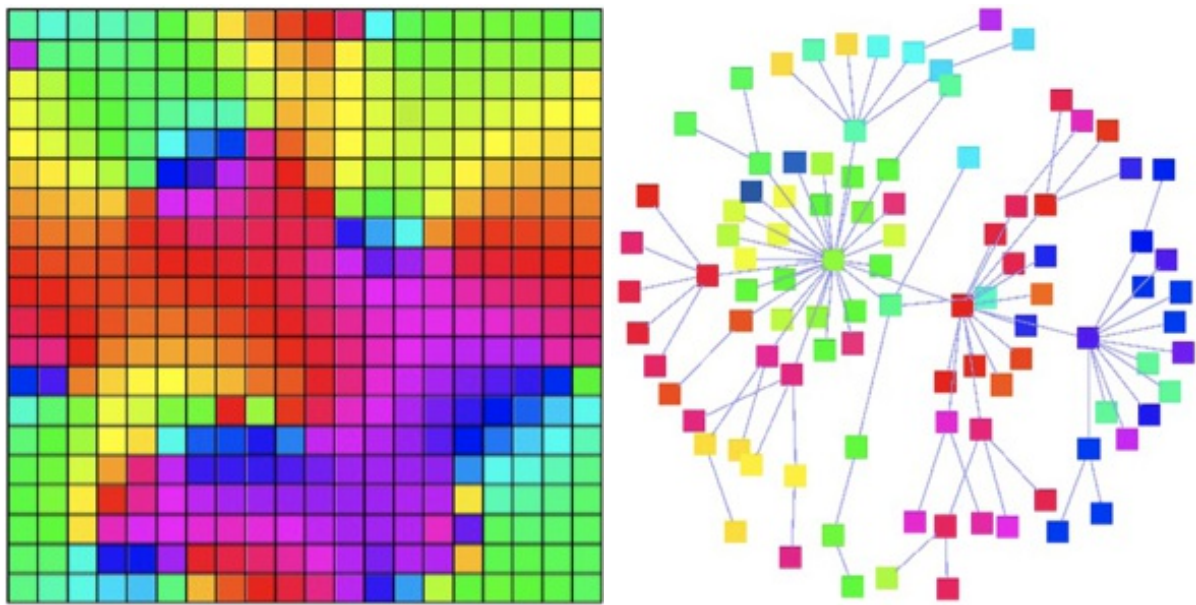
Figure 2. Formation of Scientific Disciplines in InnoScape

3.3 Basic concepts and processes of artificial ecosystems, which are also the premises of our evolving *InnoScape* model (Zou & Yilmaz, forthcoming), are as follows:

1. Each scientific community within the enterprise has a phenotype that defines its domain or discipline that is comprised of norms, practices, and skills that are deemed to be critical to collective creativity within the field of study.
2. Each community is comprised of individuals that relate to members of species in ecological models. Science accommodates multiple communities that interact and co-evolve with each other through processes of learning, transformation, and mobility.
3. Individuals are distributed and (optionally) migrate across communities and disciplines that serve as cognitive niches to individuals seeking environments conducive to creative problem solving.
4. Individuals within scientific communities have the ability to change and modify their environment as a result of their development within, and interaction with, the environment.
5. Individuals and scientific communities are associated with a scalar health or fitness measure indicating success in their environment.
6. Scientific communities undergo stages of coalescing, growth, stability, and renewal that may affect its behavior.
7. An explicit model of environment (e.g., funding agencies) that influences decisions of individuals and communities by altering the availability and distribution of resources.
8. An explicit model of knowledge production that converts human, financial, and knowledge capital into resources (e.g., open problems, skills), which are then transformed into solutions and products. This is similar to the energy metabolism dynamics in ecological systems where energy is converted into resources utilized by the species in the environment to perform actions.

Socio-cultural Dynamics of Scientific Commons

3.4 Science is also a socio-cultural practice. Communication is the essence of scholarship. With the dynamics of improved scholarly communication technologies and increasing role that they play in shaping the flow of knowledge, expertise, and resources, there is further impetus to study social communication within the context of scientific workforce dynamics. Science is inherently a social activity, because generic processes of creation and leveraging of knowledge such as knowledge sharing and combination are contextual and relational. Communities construct knowledge as they interact in a social context, which in turn influences future preferences and behavior of scientists. Influenced by this observation, socio-cultural models could help explore the following.

- What is the impact of alternative scholarly communication strategies (e.g., legitimization, dissemination, access) promoted by hub organizations such as NSF on the diversity, sustainability, stability, and innovation capacity of scientific enterprise?

🌐 Robust Long Term Policy Analysis with Adaptive Model Portfolios

4.1 Decision-makers need to factor in environmental uncertainty to identify strategies and incentives that are resilient across multiple futures. That is, rather than seeking strategies that are optimal for a specific set of expectations about the long-term future, a trustworthy approach would seek near and long-term strategies that are robust i.e., that perform reasonably well compared to alternatives across a wide range of plausible scenarios. Any one or small number of scenarios about the environmental context is bound to be wrong. Hence, the main objective of this phase is the following.

- What are the critical pillars of a robust long-term scientific enterprise workforce policy analysis framework, methodology, and associated computational framework that can leverage and integrate models developed during the three phases above so as to effectively search for incentives and course of actions that promote diversity and resilience?

4.2   Any policy carefully optimized to address a "best guess" forecast or well-understood risks may fail in the face of inevitable surprise. We promote four key elements for successful long-term policy analysis:

1. Examine large number of simulation ensembles, not one simulation model.
2. Instead of targeting optimal solutions within predefined environments, aim for robust strategies that are resilient under uncertainty.
3. To achieve robustness, incorporate adaptive schemes within the decision policy discovery process
4. Facilitate interactive exploration of the multiplicity of plausible environmental conditions.

4.3   As shown in Figure 3, having specified (context) conditions (e.g., resource allocation strategy, availability of resources, incentives, socialization intensity, contact and recruitment rates) and building block mechanisms of the model (e.g., communication theories, decision rules of individual agents), a robust strategy discovery system can employ a scenario generator to create a portfolio of plausible scenarios. Each such scenario consists of one particular choice of mechanism, often adaptive, and one particular manifestation of an environmental condition sampled from the context space. Once initial results about robust strategies are formed, they can be revised through computer search within the experiment/context space to invalidate claims about robust strategies (path b). Such invalidation triggers exploring alternative promising mechanisms (path c). Searches for invalidating mechanisms may occur entirely within the space of contexts and strategies contained in the initial portfolio; that is, within the capabilities of the current scenario generators. As the users hypothesize about alternative mechanisms that might prove more robust than the current alternatives (path d), hypothesized scenario space can expand. Alternatively, decision-makers may suggest surprises that might occur in the context space to cause an apparently robust mechanism to fail (path e).
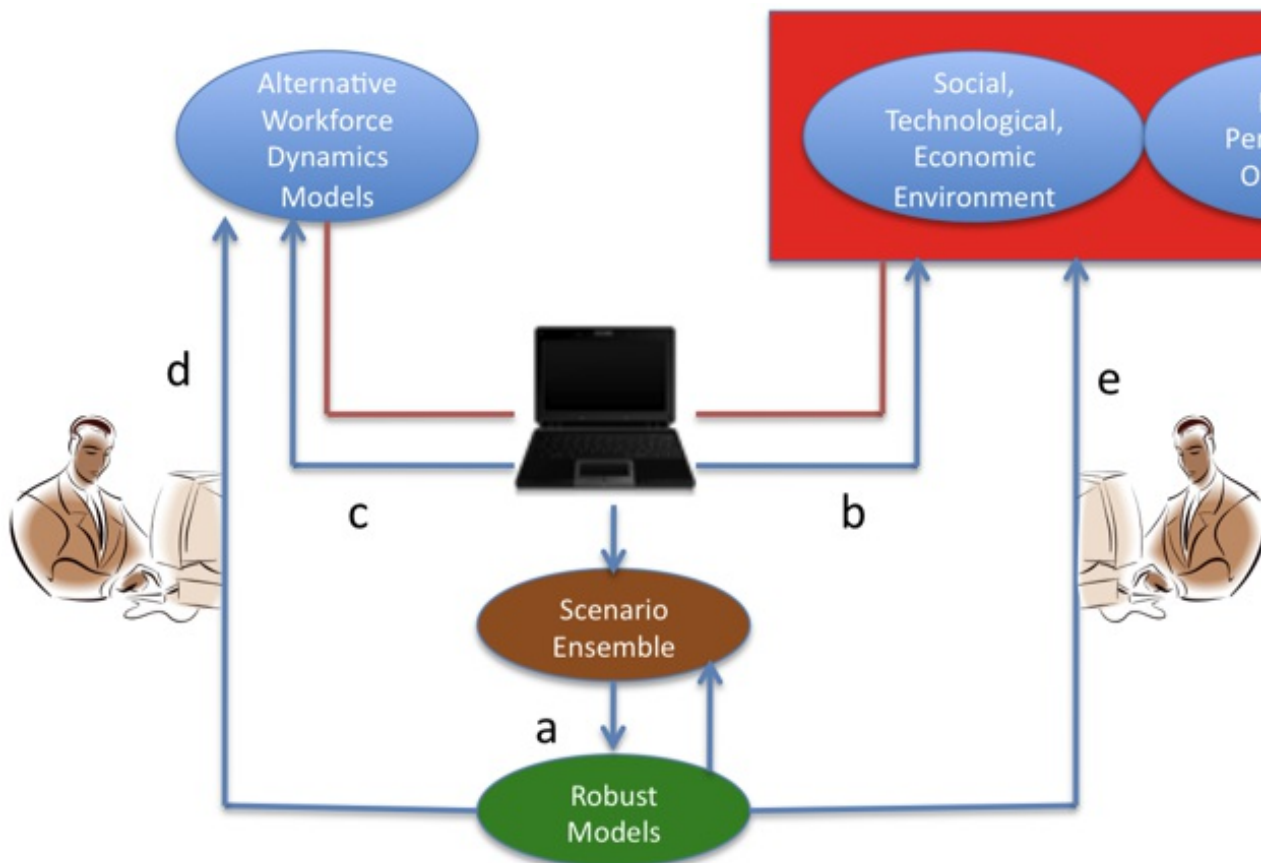

Figure 3. Robust Science and Innovation Policy Analysis Framework

4.4   These new candidates can be added to the scenario generator and their implications explored by the computational discovery system. Exploratory modeling software can enable users to navigate through the large numbers of scenarios required to make up a scenario portfolio and to formulate rigorous arguments about intervention choices based on these explorations. A scenario generator will use the relationship among the variables to create members of scenario ensembles. In contrast to traditional model development methodologies that are typically framed to produce a comparatively small number of predictive conclusions, a scenario generator should yield a full range of plausible alternatives through the above incremental and iterative refinement protocol.

Concluding Remarks

5.1 Modeling strategies discussed in this essay aim to improve development of policies effective in promoting a sustained diverse scientific workforce. This effort could direct attention to the underlying dynamics that produce a creative and diverse science-based innovation ecosystem, help identify questions that suggest further research, and stimulate the collection of the data necessary to answer them. Gaining deeper insight into complex socio-technical processes that influence the growth and formation of scientific enterprise and understanding their impact on the workforce dynamics requires a systemic and holistic view. Proposed strategies build and extend on our continuing agent-based modeling and simulation research that involves boundary processes such as learning, innovation, and mobility in global scientific communities.

References

BECKER, S. G. (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. (3rd ed.) Chicago, University of Chicago Press. (First edition 1964).

BROWN, J. S. & Duguid, P. (1991). Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation Organization. *Science,* 2, 40-57.

BROWN, J. S. & Duguid, P. (2000). *The Social Life of Information*. Harvard Business School Press. Boston.

BROWN, J. S. & Duguid, P. (2001). Knowledge and Organization: A Social Practice Perspective Organization. *Science,* 12, 198-213.

DHANARAJ, C. & Parkhae, A. (2006). Orchestrating Innovation Networks. *Academy of Management Review*, 31(3), 659-669. [doi:10.5465/AMR.2006.21318923]

HUGHES, J. (1986). *The Vital Few: The Entrepreneur and American Economic Progress*. Oxford University Press.

LATOUR, B. (1992). *Science in Action. How to Follow Scientists and Engineers through Society*. Cambridge, Mass. and London: Harvard University Press.

MILLER, H. J. & Page, E. S. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press. Princeton.

MONGE, P. R. & Contractor, S. N. (2003). *Theories of Communication Networks*. Oxford University Press.

NAP (2010a). *Rising Above the Gathering Storm, Revisited.* By Members of the 2005 "Rising Above the Gathering Storm" Committee; Prepared for the Presidents of the National Academy of Sciences, National Academy of Engineering, and Institute of Medicine.

NAP (2010b). *Expanding Underrepresented Minority Participation: America's Science and Technology Talent at the Crossroads*. Committee on Underrepresented Groups and the Expansion of the Science and Engineering Workforce. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine.

NAE (2006). *Assessing the Capacity of the U.S. Engineering Research Enterprise*. National Academy Press.

NIGMS (2007). Modeling Scientific Workforce Diversity. National Institute of General Medical Sciences. October, 2007. http://www.nigms.nih.gov/News/Reports/workforcediversity_100307.htm.

NSF (2006). *Final Report from the NSF Innovation and Discovery Workshop: The Scientific Basis of Individual and Team Innovation and Discovery*. Technical Report. National Science Foundation.

PORTER, W. L. & Lawler E. E. (1968). *Managerial Attitudes and Performance*. Doursey Press. Homewood, IL.

POWELL, W. W. & Snellman K. (2004). The Knowledge Economy. *Annual Review of Sociology*, 30. 199-220. [doi:10.1146/annurev.soc.29.010202.100037]

YILMAZ, L. (2009). An Agent Simulation Study on Conflict, Community Climate, and Innovation in Open Source Communities. *International Journal of Open Source Software and Processes* 1(3), 1-24.

ZOU G & Yilmaz L (2011). InnoScape: A Creative Artificial Ecosystem Model of Boundary Processes in Open Science. Accepted, to appear in Proceedings of the *IEEE Atlanta Conference on Science and Innovation Policy*, September 15-17, 2011 Atlanta, USA.