InnoScape: A Creative Artificial Ecosystem Model of Boundary Processes in Open Science

Guangyu Zou
Computer Science and Software Engineering
Auburn University
Auburn, Alabama 36849
Email: gzz0001@auburn.edu

Levent Yilmaz
Computer Science and Software Engineering
Auburn University
Auburn, Alabama 36849
Email: yilmaz@auburn.edu

***Abstract*—With the increasing use of cyberinfrastructure and popularity of e-Science initiatives, science is becoming truly globalized, reducing barriers to entry and enabling formation of open and global networked innovation communities. In this study, we characterize such networks as complex adaptive communication systems that exhibit traits of self-organized creative artificial ecosystems. A simulation-based exploratory study is conducted to better understand community traits that confer increased diversity and resilience in such global participatory systems. Five types of interaction topologies are identified and simulated using agent simulation as a method of inquiry. Simulation results show that scale-free network has the highest resilience as compared to random and random group network.**

I. Introduction

Scholarly communication and organizational forms of science is undergoing rapid and dramatic change. These changes are being driven by advances in computer and communication technologies and collective social, economic, cultural, and communicative processes that we call globalization. The practice of science is becoming more open and global, as the access to knowledge, as well as its production is becoming increasingly transparent. Service oriented science [1] and e-Science [2] initiatives lead to emergence of scientific communities, where domain knowledge is no longer solely documented in the journal articles or patents, but is also embodied in software, simulations, and databases. We call such open source science as Global Participatory Science (GPS). Among the examples of such communities include Open Biomedical Ontologies (OBO) Foundry [3] and NanoHub [4].

Sustainability of GPS communities require 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 (global) access to tools and instruments. What forms of collaboration structures and processes are conducive to sustained knowledge production and novelty in such open socio-ecological communities is critical to future development of virtual scientific collaboratories. It is recognized that the small world network structures with low absorption capacities result in higher innovation capacity and potential [5]. Hence, given the potential role of network recruitment and management on innovation potential, hub organizations (e.g., NSF, NIH) that orchestrate collective knowledge creation need to better understand the effects of different network topologies.

Since collective creativity and innovation are elusive and difficult to measure, we leverage two proxy metrics that are known to have positive influence on innovativeness: *diversity* and *resilience*. Diversity facilitates interdisciplinarity that fosters innovativeness and causes breakthrough in science [6]. Adaptability and transformability enables resilience, which is an attribute of socio-ecological systems with sustained innovative behavior [7]. Understanding the emergence and co-evolution of such global communities as a complex system requires exploring mechanisms that guide communication as well as the trajectories of the underlying fields guided by the typology of networks they are embedded and the rules of attachment that steer members' choice of collaborators.

To examine the impact of structural context and environment on diversity and resilience, we introduce a model in which scientific communities coevolve within selected topologies (i.e., 1D grid, 2D grid, random network, random group network, and scale-free network) and observe emergent communities in relation to their innovation performance. Future work will allow emergence of arbitrary network patterns based on alternative communication strategies. In this study, we are interested in exploring the following question:

> What is the impact of scientific community traits and environmental constraints (i.e. interaction topologies, carrying capacity, distribution of external funding) on the diversity and resilience of GPS?

Development of the model is based on understanding knowledge creation and community formation, growth, and dissolution as consequences of socio-ecological processes under given environmental constraints. Therefore, we first establish a metaphor between dynamics of GPS and artificial creative ecosystems. The artificial ecosystem concept [8] that is leveraged here is based on selected processes found in biology.

The rest of the paper is organized as follows. In section 2 we overview processes in artificial life and ecologies and relate them to open science to provide a basis for InnoScape model, which is introduced in section 3. Section 4 presents the implementation and validation of the model, followed by overview of preliminary simulation results and their discussion in section 5. We conclude in section 6 by summarizing findings and pointing out potential avenues of future research.

II. Background

In biological and ecological research, an ecosystem is defined in terms of interactions of species with each other and their physical environment. The idea of abstracting processes from biological phenomena and applying them to other domains is not new. Development and use of Genetic Algorithms to mimic evolutionary processes in search for optimal solutions and design exploration is one such example. In regard to application of the artificial ecosystem concept to open science, the central question in mapping ecological novelty to creative novelty involves identification of structural analogies and behavioral processes common to both systems.

A. *Science as an Artificial Creative Ecosystem*

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 [9], ecology [10], and social science [11]. Basic concepts and processes of artificial ecosystems as they relate to science are as follows:

- Each scientific community 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.
- 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.
- Individuals are distributed and (optionally) migrate across communities and disciplines that serve as cognitive niches to individuals seeking environments conducive to creative problem solving.
- 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.
- Individuals and scientific communities are associated with a scalar health or *fitness* measure indicating success in their environment.
- Scientific communities undergo stages of coalescing, growth, stability, and renewal that may affect its behavior.
- An explicit model of *environment* (e.g., funding agencies) that influences decisions of individu-

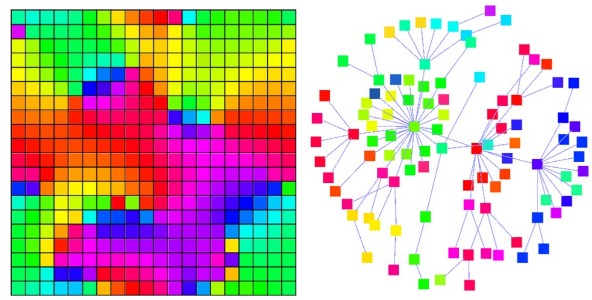
als and communities by altering the availability and distribution of resources.
- An explicit model of knowledge production that converts human, financial, and knowledge capital into resources (e.g., open problems skills), which are then transformed into solutions and products. This is similar to the *energy-metabolism dynamics* in ecological systems where energy is converted into resources utilized by the species in the environment to perform actions.

B. Science as a Communication System

Communication is the essence of scholarship. With the dynamics of globalization and increasing role that it plays to shape the flow of knowledge, expertise, and resources, there is further impetus to study social communication within the context of scientific knowledge ecosystems. 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. These communities have been described as *thought collectives* [12], *communities of practice* [13], and *knowledge value collectives*.

C. Open Science Communities

Recently a number of virtual scientific collaboratories emerged and continue to successfully bring together scientists over the globe to not only share and aggregate data, but also to create new knowledge. Such virtual collaboratories include Open Source Science (OSS) communities such as OBO Foundry (Open Biomedical Ontologies) [3], which is a form of GPS. Furthermore, compared with traditional scientific teams, OSS is driven by a distributed network of scientists with an open and transparent decentralized decision-making style. Besides OBO, the following are among successful open science communities: NanoHUB (Simulation Education Technology for Nano Technology) [4], and NEES Grid (Network for Earthquake Engineering Cyberinfrastructure).



(a) 2D (b) Scale Free

Fig. 1. Snapshots of InnoScape Model

III. InnoScape: Modeling Boundary Processes

Globalization is theorized in [14] as a confluence of multiple flows, called *scapes*: ethnoscape, technoscape, financescape, mediascape, and ideoscape. Inspired by this characterization, InnoScape focuses on modeling flow of knowledge, expertise, and skills among communities of practice through boundary processes such as communication, learning, and innovation. InnoScape is conceptualized as a multi-community ecosystem of knowledge producing and diffusing communities.

A. Genotype

Three major components are used to specify growth and development of scientific communities: *domain*, *maturity*, and *resources*. Domain refers to discipline, whereas maturity indicates the degree of development in that specific domain. Resources hold by a community are vital to undertake scientific activities. In order to visually depict the evolving states of communities, the *HSB* color model is used. Hue indicates the domain of a community. Saturation within the color spectrum represents maturity. The degree of brightness corresponds to the level resource.

Figure 1 depicts the snapshots of our model with grid and network topology respectively, where each cell represents a community whose color corresponds to its internal state. As shown in the figure, the state space of communities exhibits a color landscape.

B. Process

As shown in Figure 2 the behavior of scientific communities in InnoScape is comprised of six sub-processes: resource allocation, interaction within community, learning, innovation, growth, and fade. Resource allocation refers to strategies to distribute resources to communities. Interaction within community refers to scientific activities at the macro level i.e. community is driven by funding to improve its maturity. Learning and innovation between communities mimic the boundary processes among communities i.e. communities affect and are influenced by peer communities. Growth is defined as the process through which communities improve their sizes so as to increase their influences. Fade refers to disappearance of the community due to loss of resources. These six sub-processes are discussed in detail in the following sections.

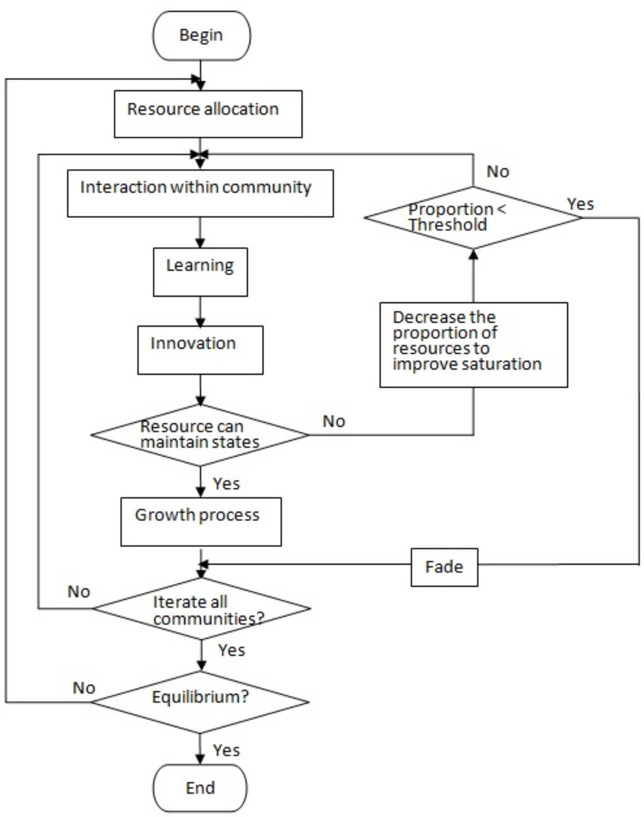


Fig. 2. Process Model

1) Resource Allocation: Although it can be varied, the strategy for resource allocation adopted in the current model is uniform allocation, that is the total resources are distributed among all communities equally. The total amount of resources available for allocation is equal to sum of the contributions of communities and external funding. Contributions by communities are based on the assumption that produced knowledge can be transferred to technology which in turn results in economic growth. Contributions provided by a community is moderated by the product of its maturity and resource, based on the hypothesis that communities with higher maturity and resources are expected to be more productive This is expressed as follows:

$$R_t = \sum_{i=1}^{\#communities} (F_{i,t} + S_{i,t} \times B_{i,t}) \quad (1)$$

where, R_t indicates the total resource available at time t. $F_{i,t}$ denotes the external funding allocated to community i at the time t. $S_{i,t}$ and $B_{i,t}$ indicate maturity and resources of community i respectively.

2) Sustaining the Community: Every time step during the simulation, each community receives resources via funding. However, not all available resources can be used to improve maturity of community i.e. only part of the resource helps advance maturity, because learning and innovation processes also require resource. How much saturation the community can gain by these resources is determined as follows:

$$S_{t+1} = S_t + \alpha_t \times (1 - S_t) \times R_{s,t} \quad (2)$$

where, S_{t+1} is the maturity of the community at the time $t + 1$. $R_{s,t}$ is the resources that could be used to increase maturity, which is a proportion of available remaining resources after maintenance (e.g., equipment, infrastructure). α_t adjusts increase in saturation, which is an exponential decay function over time, to reflect inertia and the increasing cost of further maturation after dominant norms are settled. As maturity increases, it is necessary to consume resources because technology develops based on research and development costs.

$$B_{t+1} = B_t + R_t - R_{m,t} - R_{s,t} \quad (3)$$

where, B_{t+1} is the new resource level. $R_{m,t}$ is the resource needed to maintain the current state.

3) Learning: According to homophily theory [15], influences that communities exert or receive are based on their interaction frequency. Interaction frequency between communities is depicted by the weights associated with links between them in the evolving communication graph. The intensity of community j's influence on community i is defined as follows:

$$\begin{cases} W_{ji,t} &= W_{ji,t-1} + C_W I_{ji,t}(1 - W_{ji,t-1}) \;\; I_{ji,t} \geq 0 \\ W_{ji,t} &= W_{ji,t-1} + C_W I_{ji,t} W_{ji,t-1} \;\;\; otherwise \end{cases} \tag{4}$$

where, $W_{ji,t}$ is at the current time. C_W is between 0 and 1 and is inversely proportional to inertia (resistance to change in a community). $I_{ji,t}$ is the intensity of change in the influence.

$$I_{ji,t} = (1 - D_{ji,t})^4 - (1 - \overline{D_{i,t}})^4 \tag{5}$$

where, $D_{ji,t}$ is the dissimilarity defined as the distance between community i and community j in terms of their current hue (e.g., discipline tendency) at the time t. $\overline{D_{i,t}}$ is the average distance between community i and all its neighbors at the time t. This function grows much faster when dissimilarity between i and j becomes smaller in comparison to average dissimilarity, resulting in more intense influence $I_{ji,t}$.

Formally, dissimilarity between communities i and j is defined as follows.

$$D_{ji,t} = Dissimilarity(H_{i,t}, H_{j,t}) \tag{6}$$

where, $H_{i,t}$ is the hue of community i at the time tick t. $H_{j,t}$ is the hue of community j at the time j.

$$Dissimilarity(x, y) = \begin{cases} \frac{|x-y|}{180} & |x - y| \leq 180 \\ \frac{360-|x-y|}{180} & otherwise \end{cases} \tag{7}$$

Normalization for the weights of neighbors is required, where β is the receptivity of the community.

$$W'_{ji,t} = \beta \times \frac{W_{ji,t}}{\sum_{k=1}^{\#neighbors} W_{ki,t}} \tag{8}$$

Learning among communities affects both the saturation and the discipline. Maturity of a domain in a discipline is affected as scientists use cross lineage to adopt and transfer knowledge from other domains, resulting in knowledge spillover. The closer two interacting domains in the discipline (i.e., hue) spectrum, the larger the positive impact of spillover on maturity. Alternatively, as the distance increases, new knowledge may result in reconsideration of earlier assumptions and result in reconstruction and redirection. This is formalized as

$$S_{i,t+1} = S_{i,t} + S \times \sum_{j=0}^{\#neighbors} W_{ji,t} \times S_{j,t} \times cos(\alpha_{ji,t}) \tag{9}$$

where, $S_{i,t+1}$ refers to the saturation of community i at the time $t+1$. $\alpha_{ji,t}$ refers to the angle between hues of community i and community j. S is the susceptibility of the community i to influence, which defined as an exponential decay function of resources held by the community. Furthermore, learning leads the current community to change its hue i.e. discipline (specific norms, practices, and relevant skills) due to the influences of neighbor communities. At the same time, the community itself is inclined to realize its own target goals within the objective discipline:

$$H_{i,t+1}^{current} = H_{i,t}^{current} + S \times H_c \tag{10}$$

$$H_c = \sum_{j=1}^{\#neighbors} (W_{ji,t}(H_{j,t}^{current} - H_{i,t}^{current})) + C \tag{11}$$

$$C = W_{i,t}(H_{i,t}^{target} - H_{i,t}^{current}) \tag{12}$$

where, $H_{i,t+1}^{current}$ refers to the new hue after the learning process and H_c is the change in hue moderated by susceptibility to change, S. The parameter C represents degree of conservativeness of the discipline, which pulls the community toward original target hue set for the discipline.

4) Innovation: Innovation changes the norms of the community i.e. target hue in the InnoScape model because changing of target hue is a strategy for a community to adapt to its environment. Moving target hue of a community toward its current hue can decrease resource consumption during the learning process, which in turn improve its sustainability. When the distance between current and target hue of a discipline exceeds the tolerance threshold, $T_{innovation}$, conditions for innovation is established.

$$D_{ii,t} \geq T_{Innovation} \tag{13}$$

In the InnoScape model, a community innovates in two ways: reorganization or specialization. By reorganization a community transforms itself by moving its accepted target toward the current state. On the other hand, specialization involves branching out new subcommunities. In the InnoScape model, whether reorganization or specialization happens is determined by a parameter, called reorganization tendency.

5) Fade: After the innovation process, if the resource of a community cannot maintain its current state, then $R_{s,t}$ is decreased, and the processes of interaction, learning and innovation are started over. The iteration process continues until the resources left can maintain the current state or $R_{s,t}$ is equal to 0. When $R_{s,t}$ is equal to 0, the community fades and is removed from the current context.

6) Grow: If the community has enough resources to maintain and the neighbor cell is empty, then the community is likely to extend to occupy the neighbor cells with a small probability. This captures evolutionary dynamics by retaining those communities that are fit to survive in the current environment.

IV. Metrics And Indicators For Measuring Innovation Potential

Since we are interested in observing potential relations between the structure of the social network and innovation capacity of a community, two types of metrics are considered: innovation metrics and network structure metrics that pertain to integrated differentiation that is related to innovation potential [16].

The process of knowledge creation is based on combination and elaboration of existing knowledge. Diverse sources of knowledge challenge existing solutions, ignite new ideas and lead to novel solutions [17]. So, diversity is a proxy indicator for innovation potential and capacity. There are three dimensions related to diversity: *variety*, *balance*, and *disparity* [18]. Variety can be computed as the number of clusters of communities within the environment. Each cluster is comprised of similar communities and derived using the *QT* clustering algorithm [19]. Balance is defined as the equality in terms of resources each community holds, which is calculated using the Gini coefficient [20]. Disparity refers to the degree of difference of each community, which is formalized using the dissimilarity metric discussed earlier.

Innovation is the process of finding alternative, more effective ways to address challenges and seize opportunities. On the other hand, resilience is the capacity to adapt, restore in constructive ways while undergoing changes so as to still retain essentially the same function. Innovation is change, but resilience is survival. Due to presence of uncertainty in the evolution of the innovation landscape, resilience is an essential property for a scientific community to sustain its innovation capacity. Resilience is the capacity of a system to absorb disturbance and reorganize while undergoing change so as to still retain its identity and feedbacks [7]. Based on this definition, we define resilience as the extent of disturbance of the system that reduces the ratio of active communities to initial set of communities below a specific threshold.

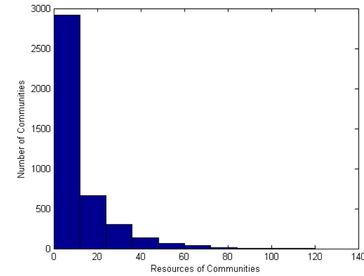
Structural properties of networks as they relate to creative output pertain to integrated differentiation [16]. As a general measure of the degree of social interaction, we use density, centrality, clustering coefficient so as to determine their potential role in and relation to innovativeness.

V. IMPLEMENTATION AND FACE VALIDITY

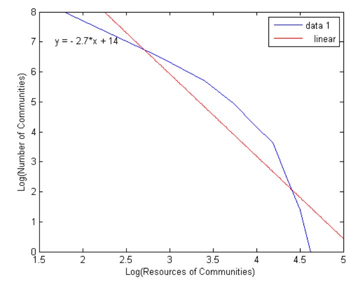
Figure 3 presents evolving states of communities over time during a single run of the InnoScape model. Initially, the colors of communities are grey due to their low maturity. As the simulation unfolds, states of communities become increasingly colorful due to increasing maturity through community sustainment, interaction, learning, and innovation processes. After a long run, clusters with similar color patterns emerge, which suggest formation of related disciplines as a result of communication and boundary processes.

Table I is the comparison of network metrics generated by InnoScape model and the corresponding metrics from empirical OBO data (expected values in the table). Since the confidence intervals

Fig. 3. Growth and Formation of Community Clusters



(a)



(b)

Fig. 4. Resource Distribution

of metrics derived from the simulation data contain the corresponding values of OBO network. we can conclude that InnoScape model can generate similar network structures compared to OBO. In addition, the best configuration parameters against the network of OBO has a medium level tolerance (0.6), high receptivity (0.9), and high degree of communication frequency (1.0). These are peculiar characteristics of open source science communities.

TABLE I
SIMULATION VS. OBO DATA

Metrics	Mean Value	Standard Deviation	Confidence Interval at 90%	OBO
Number of Communities	55.633	25.594	[46.076, 65.190]	49
Density	0.605	0.225	[0.521, 0.689]	0.549
Clustering Coefficient	0.846	0.092	[0.812, 0.881]	0.880
Centrality	0.355	0.140	[0.302, 0.407]	0.405

Figure 4(a) depicts the inequality of communities in terms of resources. Most communities hold the relatively few resources, while a small part of communities hold the relatively many resources. This observation is indicative of the presence of power law. The power law exists in many social systems, for instance, the number of papers published by authors, the citation index of papers etc [21]. Figure 4(b) shows the relationship between the log value of number of communities and their resources, as well as the corresponding linear regression curve. Since the R^2 for this fitting is 0.86, the InnoScape model suggests the presence of power law in resource distribution.

VI. Preliminary Experiments and Evaluation

A. Configuration Parameters

Table II denotes the configuration parameters and their initial values. We conducted a series of sensitivity analysis experiments to gain insight about diversity and resilience.

B. Topologies of Interaction Context

The experiments in this section test five types of interaction topologies and their effects on diversity and resilience:

1) One-dimensional grid: each community has two neighbors, one on the left and one on the right.
2) Two-dimensional grid: each community has eight neighbors around it. Figure 1(a) presents a snapshot of the 2D grid environment.
3) Random network: the edges between any pair of nodes are created with equal probability.
4) Random group network: the nodes within a group have higher probability to build links than those between different groups.
5) Scale-free network: the nodes with more links are more likely to be selected to build links.

TABLE II
CONFIGURATION PARAMETERS AND THEIR INITIAL VALUES

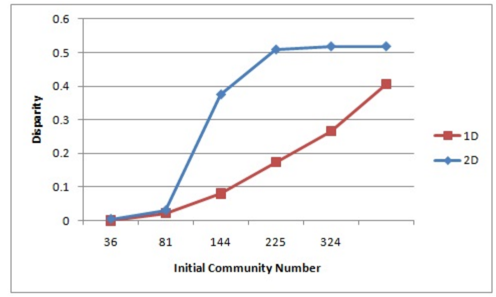
Parameter	Description	Range	Initial Value
Carrying Capacity	Initial number of communities	[10, 200]	100
Stop Time	Time to run	$[1, \infty)$	1000
$F_{i,t}$ in Eq. 1	External funding	[0.1, 1]	0.5
$T_{Innovation}$ in Eq. 13	Tolerance	[0, 1]	0.2
Reorganization Tendency	Reorganization happening frequence	[0, 1]	0.5
Neighbor Size	Influential radius of community	$[1, \infty]$	1
$W_{ji,t}$ in Eq. 4	Initial weight of neighbor	[0, 1]	Random
C_W in Eq. 4	Resistance to change	[0, 1]	0.5
β in Eq. 8	Receptivity	[0, 1]	0.5
Current Color	State of community	HSB range	Random
Target Color	Target of community	HSB range	Random

Figure 1(b) illustrates a snapshot of scale-free network.

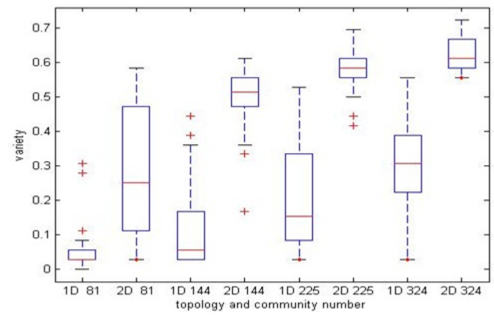
C. Diversity vs. Carrying Capacity

In this experiment, we explore variation of diversity in relation to initial community numbers within a specific topology. Figure 5 evaluates variety and disparity across combination of two factors, number of communities and 1D/2D topology, and their varying levels.

In Figure 5, we observe that variety and disparity increase with the initial community size. In the 2D topology, disparity increases as the size of the community increases up to a critical threshold, after which further increase in community size does not result in further dissimilarity. Computation of variety is based on clustering algorithm which in turn is based on the pre-selected diameter that is defined as the maximum difference of members within a cluster. In this experiment, the diameter is set to 10, which means the differences of hue of communities within a cluster can be up to 10. Therefore, the maximum variety is 360/10= 36 i.e. diversity cannot increase infinitely along with initial community number. Furthermore, based on Figure 5(b), the comparison between 1D and 2D suggests that 2D topology is not only more effective in fostering variety with a lower degree of uncertainty in comparison to 1D topology, which has a more restricted sphere of communication. This limitation inhibits propagation of influence and hence takes more time to reach equilibrium.



(a)



(b)

Fig. 5. Diversity vs. Initial Community Numbers

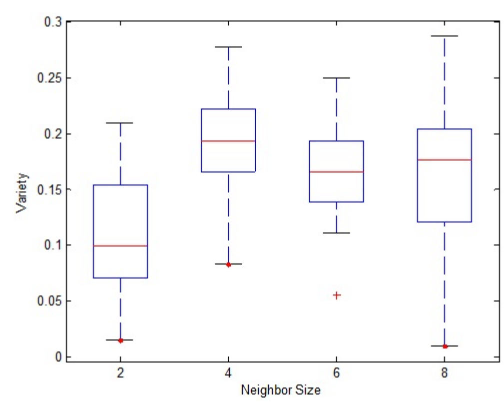


Fig. 6. Variety vs. Neighbor Size in 1D

To test the impact of neighbor size in the 1D topology, we gradually increased the interaction window from 2 to 8 neighbors. Observations depicted in Figure 6 suggest that interaction window positively affects variety and underlying uncertainty (i.e., dispersion) up to a level, beyond which variety stops improving and uncertainty starts increasing.

D. Diversity vs. External Resource

Figure 7 depicts the trend of diversity with respect to size of external resource injected into the environment.

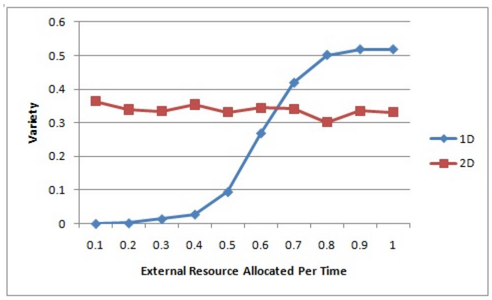


Fig. 7. Diversity vs. Resource Allocated Per Time

The abscissa indicates the amount of resources allocated to each community per time tick. In the 1D topology, the rate of increase in variety slows and stabilizes over time. On the other hand, 2D topology seems to be less sensitive to external resource, which suggests higher degree of potential for resilience than 1D.

E. Diversity vs. Receptivity

In this experiment, we considered multiple interaction contexts based on the selected network topologies. Figure 8 shows the change in diversity with respect to varying levels of receptivity and connectedness. Receptivity of a community is defined as the ratio of neighbor influence to inertia. Connectedness is defined as the probability of building links between nodes. Figure 8 indicates that there is a critical receptivity threshold, after which the behavior of low and high density communities diverge. Under environments with high receptivity, variety favors low connectivity. However, communities with various levels of connectivity examined in this experiment converge to the same stable level of variety. Similar patterns are observed in both random and random group networks.

F. Resilience of Different Network Topologies

Resilience is defined as the extent of disturbance of the system that significantly reduces the ratio of active communities to number of communities observed when external resource is set to maximum. To calculate resilience, the number of communities under maximum resource availability is set as the base reference level for each topology. As resources are incrementally reduced, the ratio (ρ) of number of communities to the reference is computed. The loss ratio is defined as $1 - \rho$ and ranked to identify resilient topologies. According to Table III, scale-free network has the highest resilience, and random group network has higher resilience than random network, because the loss ratio of scale free network is smallest and the loss ratio of random network is largest when external resources decrease to 0.7. Based on Figure 9, it is clear that random group network has higher resilience than random network .

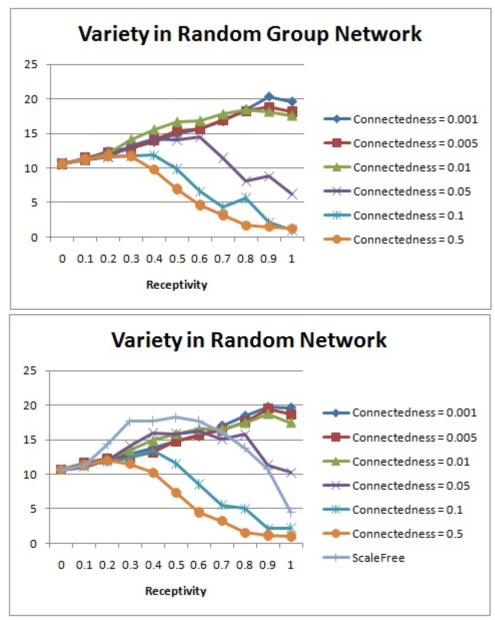


Fig. 8. Variety in Random and Random Group Network

G. Relationship between Diversity and Network Metrics

To examine the relationship between variety and density, we plotted in Figure 10 average variety values for each level of density in both rando and random group networks. The data suggests that variety improves with increasing density up to a point, which can be considered as a low connectivity. As the density increase beyond 0.2, we observe a general trend toward non-monotonic reduction in degree of variety.

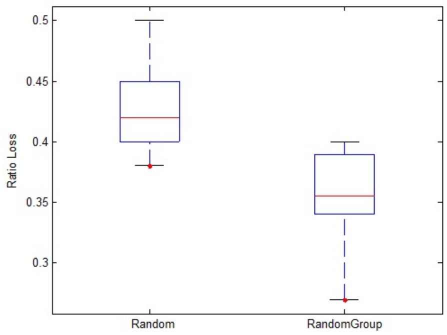


Fig. 9. Comparison of Random and Random Group Network on Resilience

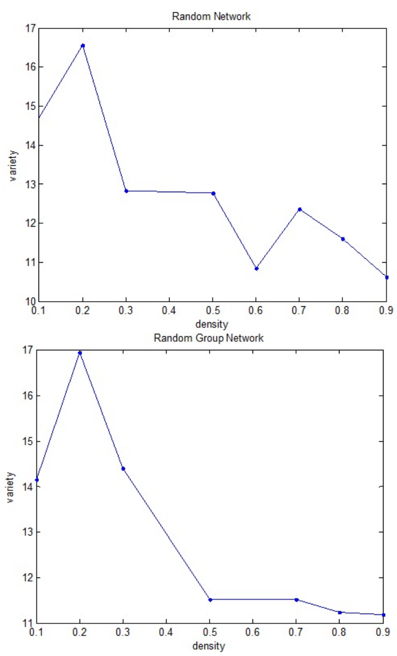


Fig. 10. Variety vs. Density in Random and Random Group Network

As shown in Figure 11, a similar pattern is observed in regard to average degree centrality and variety. Both observation support our expectation that increasing degree of connectivity beyond a critical threshold results in more homogeneity due to loss of differentiation.

VII. Conclusion

In this study, we conceptualized growth and development of scientific communities in terms of a complex adaptive communication system that follow principles of creative artificial ecosystems. Based on the preliminary experiments, some of which are not presented in this paper, we draw the following conclusion. The size of carrying capacity

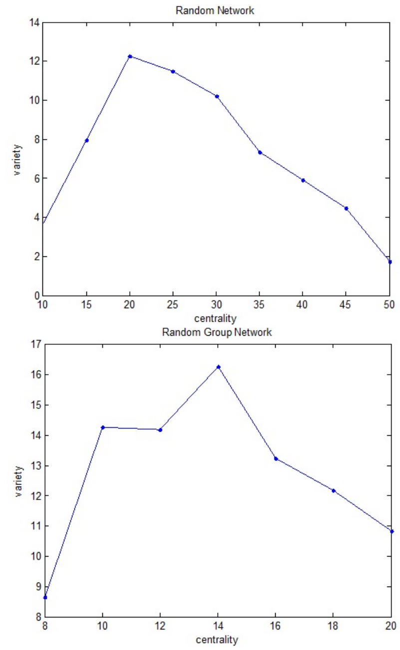


Fig. 11. Variety vs. Centrality in Random and Random Group Network

of the knowledge ecosystem has positive effects on diversity. Yet, there is a point of diminishing returns. Also, diversity does not monotonically increase with increasing levels of external resource. Additionally, our observations suggest that the 2D topology is more resilient than 1D topology, and scale-free networks have higher resilience than random and random group networks. Furthermore, In low density networks, increasing levels of receptivity improves diversity up to a level. Similarly, variety increases with density and centrality up to a point, beyond which diversity is inhibited.

Acknowledgment

This work was supported in part by the National Science Foundation Grant No. NSF-SBE-0830261. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. government.

References

[1] I. Foster, "Service-oriented science," *Science*, vol. 308, pp. 814–817, May 2005.

[2] P. A. David and M. Spence, "Toward institutional infrastructures for e-science: The scope of challenges," *Technical Report: Oxford Technical Institute*, May 2003.

[3] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, T. O. Consortium, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, and S. Lewis, "The obo foundry: coordinated evolution of ontologies to support biomedical data integration," *Nature Biotechnology*, vol. 25, no. 11, pp. 1251–1255, 2007.

[4] NanoHUB, "Network for computational nanotechnology," *http://nanohub.org/*, May 2009.

[5] R. Cowan and N. Jonard, "Network structure and the diffusion of knowledge," *Journal of Economic Dynamics & the Diffusion of Knowledge*, vol. 28, pp. 1557–1575, SAGE 2004.

[6] S. E. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools and Societies*. Princeton University Press, 2007.

[7] B. Walker, "Resilience, adaptability and transformability in social–ecological systems," *Ecology and Society*, 2004.

[8] J. McCormack, "Artificial ecosystems for creative discovery," *In Proceedings of the GECCO'07*, pp. 301–307, 2007.

[9] W. B. Arthur, S. Durlauf, and D. A. Lane, *The economy as an evolving complex system*. Reading, MA: Addison-Wesley, 1997, vol. 2.

[10] M. Mitchell and C. E. Taylor, "Evolutionary evolutionary computation: An overview," *Annual Review of Ecology and Systematics*, vol. 30, pp. 593–616, 1999.

[11] J. M. Epstein, *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton, NJ: Princeton University Press, 2006.

[12] L. Fleck, *Genesis and Development of Scientific Fact*. The University of Chicago Press, 1979.

[13] S. J. Brown and P. Duguid, "Knowledge and organization: A social practice perspective," *Organization Science*, vol. 122, pp. 198–213, 2001.

[14] A. Appadurai, "Disjuncture and difference is the global cultural economy," *Global Culture: Nationalism, Globalization, and Modernity*, pp. 295–310, 1990.

[15] P. R. Monge and N. S. Contractor, *Theories of Communication Networks*. Oxford University Press, 2003.

[16] R. Stankiewicz, "Technology as an autonomous socio-cognitive system," *Dynamics of Science-Based Innovation*, Springer-Verlag 1992.

[17] S. A. Mohrman and C. S. Wagner, *The Dynamics of Knowledge Creation: Phase One Assessment of the Role and Contribution of the Department Energy's Nanoscale Science Research Centers*. Department of Energy: Technical Report, 2008.

[18] C. S. Wagner, J. D. Roessner, K. Bobb, J. T. Klein, K. W. Boyack, J. Keyton, I. Rafols, and K. Börner, "Approaches tounderstandingandmeasuringinterdisciplinary scientific research (idr):a review of the literature," *Journal of Informetrics*, vol. doi:10.1016/j.joi.2010.06.004, 2010.

[19] L. J. Heyer, S. Kruglyak, and S. Yooseph, "Exploring expression data: identification and analysis of coexpressed genes," *Genome Research*, 1999.

[20] I. Buchan, "Calculating the gini coefficient of inequality," *Northwest Institute for BioHealth Informatics*, 2002.

[21] C. R. S. A. Clauset and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, pp. 661–703, 2009.

TABLE III
RESILIENCE OF DIFFERENT NETWORK TOPOLOGIES

	Random		Random Group		Scale Free	
Resources	Number of Communities	Loss Ratio	Number of Communities	Loss Ratio	Number of Communities	Loss Ratio
1	43.77	0	35.57	0	71.43	0
0.9	42.3	0.03	34.27	0.04	66.47	0.07
0.8	37.43	0.14	30.63	0.14	60.7	0.15
0.7	25.33	0.42	22.83	0.36	52.2	0.27