A Complex Adaptive Model of Information Foraging and Preferential Attachment Dynamics in Global Participatory Science

Ozgur Ozmen
Industrial and Systems Engineering Department
Auburn University
Auburn, Alabama 36849
Email: ozgur@auburn.edu

Jeffrey Smith
Industrial and Systems Engineering Department
Auburn University
Auburn, Alabama 36849
Email: jsmith@auburn.edu

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

Alice E. Smith
Industrial and Systems Engineering Department
Auburn University
Auburn, Alabama 36849
Email: smithae@auburn.edu

***Abstract*—Recent developments in cyber-infrastructure and emerging virtual science collaboratories are enabling scientists to transparently co-develop, share, and communicate diverse forms of knowledge artifacts in real-time. Using collective action theory as a basis, we introduce an agent-based model of such collaborative environments as complex adaptive social communication systems. By examining empirical data from the Open Biomedical Ontologies (OBO) Foundry, we present a conceptually grounded agent-based model of what we call Global Participatory Science (GPS). The model represents the dynamics of GPS in terms of the information foraging, social exposure, and preferential attachment mechanisms. We monitor social network metrics and activity patterns as proxy metrics to infer innovation potential of collaboration networks. In this paper, we introduce our CollectiveInnoSim model and demonstrate the impact of foraging and preferential attachment mechanisms on emergent social network structures. The objective is to further our understanding of the dynamics of GPS and facilitate developing informed policies fostering innovation potential.**

I. Introduction

Science is becoming increasingly global and participatory due to online collaboration opportunities such as e-mailing, web-based social networking, and open-access collaboration platforms. Hence, scientists interact not only locally, but also globally by constructing self-organizing collaboration networks. We call scientific knowledge creation in such communities Global Participatory Science (GPS) [1].

[2] state that "one of the most significant problems in organizational scholarship is to discern how social collectives govern, organize, and coordinate the actions of individuals to achieve collective outcomes." This work explores micro-level (inter-scientist) socio-technical processes and mechanisms that explain emergent behaviors observed in scientific communities that collaborate over the cyber-infrastructure. Based on the views advocated by [3] and [4], we interpret the structure and behavior of GPS as a complex adaptive system (CAS). We leverage recent ethnographic studies, which suggest that GPS is a collective action undertaken by autonomous self-organizing scientists [5], [6].

It is demonstrated that science is complex because researchers interact in both competitive and cooperative ways, with no imposed blueprint. Furthermore, it is adaptive because scientists respond to environmental changes such as funding preferences or new discoveries [3]. We conceptualize information foraging, preferential attachment, and population dynamics as the underlying self-organization mechanisms of knowledge creation in GPS.

The understanding of CAS is more likely to arise with the help of computer-based models [7] and Agent Based Modeling (ABM) provides us with the opportunity to directly identify individual entities along with their relationships and capabilities. Hence, we simulate these mechanisms adopting the ABM worldview as a bottom-up approach that has a top-down guidance of the objectives we measure.

Our objective is to explain operational behavior of GPS and its socio-technical processes in the form of a computational model to gain empirical insight and perform exploratory analysis measuring innovation potential. The emergence of new knowledge structures, new channels of communication, and new network topology can be described as innovation in virtual scientific communities. We know most of the outputs of an innovation system, like the number of publications or patents etc. and the inputs, like resources allocated, but we do not really know much about the process that transforms inputs into outputs [8]. The next generation innovation metrics are more focused on emergence. We perceive emerging social-network structures as innovation indicators in our analysis.

In this paper, we present a complex adaptive model of GPS

that is conceptually grounded on self-organization mechanisms we built. In section 2, we introduce conceptual background of our model. In section 3, we present the conceptual model structure in detail. Section 4 discusses validation processes and the impact of different preferential attachment and the foraging mechanisms on emergent social network structures. In section 6, we conclude with a summary of our findings and the future work in progress.

II. Background and Related Work

This section provides a brief overview of the concepts of complex adaptive systems and self-organization mechanisms such as preferential attachment, information foraging, and collective action.

A. Complex Adaptive System models

Complex Adaptive Systems (CAS) can be described as a framework to understand the world around us. CAS are formed of elements that have wide range in both form and capability [7]. [9] describe CAS as composed of interacting thoughtful (but perhaps not brilliant) agents. "Not brilliant" raises concerns about "bounded rationality" principle that states that individuals should not necessarily be rational and can give their decisions with the perfect information [10]. In addition to "bounded rationality," [4] describe the main elements of complex systems in terms of the network of agents, their attributes or traits, the rules of interaction, and the structures that emerge from these micro-level interactions.

ABM captures emergent phenomena because it has a holistic approach that perceives a system as more than the sum of its constituent parts. The macro-level emerging behavior cannot be explained by the properties of the units in the system. Since ABM is used more with the behavioral entities, it provides an opportunity to model more realistically.

There are many inspiring implementations of agent based simulation models that are explaining different systems and are creating understanding for different contexts [11], [12], [13]. Additionally, different scholars use simulation to study scientific domains. [14] introduces a model to determine whether it is possible to reproduce observed regularities in science using a small number of simple assumptions, [15] continue on top of Gilbert' s model [14] and explore how different cognitive settings may affect the aggregate number of scientific articles produced. In the context of collective knowledge creation and diffusion, [16] simulate the knowledge exchange process to examine the relationship between network performance and the network architecture. [17] perceive science as problem solving including machine learning techniques. But in these studies, the social interactions (mechanisms) were not taken into account.

B. Social Mechanisms and Social Networks

Collective action is focused mainly on mutual interests and the possibility of benefits from coordinated action [4]. There are also social dilemmas introduced by [18], in which he asserts that the mutual interest and individual-interest conflict resulting in dissolution of the collective action. The dilemma between mutual and self interest is essential. Using the collective action theory, which includes models of self-interest, exposure, cognitive burden, and tension in scientific knowledge generation, we develop theoretically-grounded formalization of individual behavior of scientists and engineers.

Metaphorically, scientists are informavores like food foragers in the nature. Predators are expected to abandon their current patch (e.g., domain) when local capture rate (e.g., problem solving success) is lower than estimated capture rate in the overall environment [19]. Information foraging theory, which is derived from this evolutionary phenomenon developed by [20] assumes that people, if they have an opportunity, would adjust their strategies or the topology of their environment to maximize their rate of information gain.

All intelligible ideas, information, and data that can be delivered or gathered in a format can be referred to as knowledge [6]. The introduction of new ideas through weak ties can foster innovation and development of the system [3]. Artifacts are products of the collaboration, which can be in forms of document, code, bug-report, data etc. In addition to the artifacts, GPS has interactive communication outputs [4]. In other words, connectivity of the members (the network itself) and communality can be identified as the products of the collective action.

[21] states that "most real networks, however, exhibit preferential attachment, such that the likelihood of connecting to a node depends on the degree of the node." However, [22] suggest four different types of models of network: regular lattice, small-world, scale- free, and random. Then they argue that social networks are not random since people link with others who are similar. They also argue that people do not only use preferential attachment, in which new people link to the ones who already have many links because people do not necessarily know who has many links, so scale-free networks are not completely realistic. The real networks are formed by a mixture of different mechanisms and still preferential attachment is an essential process in networks. What information is available to the agents and how capable they are in processing it are the questions of interest.

III. Conceptual Model

Scientists join or leave a problem domain on the basis of problems to be solved and tasks to be accomplished, and their position in the scientific landscape depends upon their knowledge, levels of interest, personal learning objectives, resources, and commitments [23]. We leverage Collective Action Theory [24] as the socio-cognitive interaction mechanism in GPS. It basically asserts, when the sum of benefits an individual gains is more than the costs he/she is burdened with, that individual will join the collective action.

We perceive GPS as a collective action because artifacts as a knowledge-product of the collaboration are "public good" which are owned by the community and have the features "jointness of supply" and "impossibility of exclusion." This means that the knowledge produced is open, shared, everybody

can benefit from it, and benefit of someone does not diminish the benefit that can be gained by the others.

Scientists follow their self-interest on the theme of an artifact. But self-interested people are likely to get things with bearing the lowest cost. Additionally, the participation to GPS is not compulsory but if someone is in an open science community, there is social pressure that means scientists are exposed to the collective behavior. We can also call this phenomenon as "exposure" to the mutual-interest. The conflict between mutual-interest and self-interest is essential. While mutual-interest on an artifact is driving an individual to participate, self-interest might cause avoidance from participation or vise-versa.

A. Simulation Environment

We use Repast (Recursive Porous Agent Simulation Toolkit) as a tool and our simulation context is a grid. Scientists perceive their limited environment and search for artifacts to contribute. Analogically, movement of the scientists can be perceived as browsing on a web tool or a forum website. The snapshot below represents the grid and network visualizations in Repast.

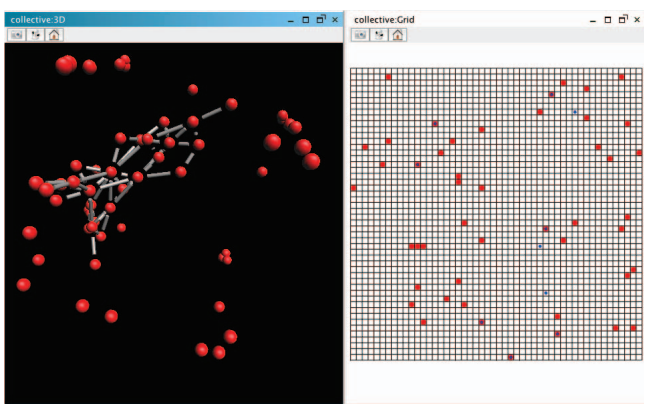


Fig. 1. Visualization of RePast contexts

B. Preferential Attachment Mechanism

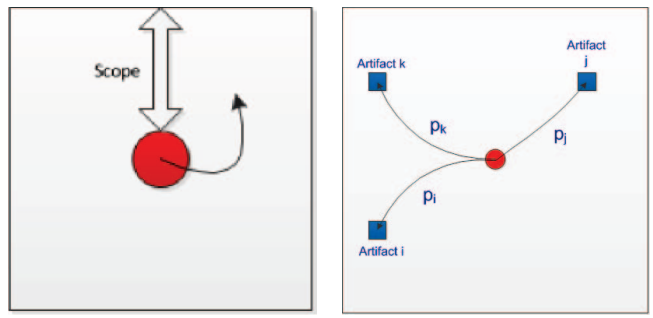
Let us imagine a web tool, in which scientists can browse the list of open artifacts, select one of them, contribute, and learn from it. Scientists are not homogeneous in terms of time spent in browsing. Some of them browse more titles and some browse less. Each scientist has a scope of environment that is limited (they do not have the perfect information about the whole environment) and they perceive that scope while searching for an artifact. The moving operator in figure 2 is basically browsing in the environment when there is no artifact to select within the scope of a scientist. The information they are exposed to is also limited. The preferential attachment mechanism is based on the calculation of three dimensions:

- *Popularity:* Scientists might choose an artifact according to the popularity of that particular artifact, which means the more elaborated the artifact is, the more likely it is to be selected ($0 < pa < 1$).
- *Self-interest:* According to familiarity of an artifact, a scientist is more likely to select it ($0 < si < 1$).
- *Imitation:* The artifacts with more active members are more likely to be selected ($0 < im < 1$).

Each dimension has a weight that signifies its importance in selection process of a scientist. Initially, each weight is equal and $w_{pa} + w_{si} + w_{im} = 1$. Each artifact j has an incentive $P_j = w_{pa} \times pa + w_{si} \times si + w_{im} \times im$. In case of being exposed to more than one artifact, roulette wheel selection algorithm is used to assign probability p_j to each artifact j and select one of them based on their probabilities.

$$p_j = \frac{P_j}{\sum_{i=0}^{N} P_i} \tag{1}$$

where N is the total number of artifacts that are within the scope of the scientist. In figure 2, we represent moving and artifact selection processes.



(a) Moving of a scientist (b) Artifact selection of a scientist

Fig. 2. Moving and artifact selection processes in the model

C. Collective Action Mechanism

As mentioned above, there are two driving forces for scientists: self-interest and mutual-interest. The scientists are more likely to benefit from familiar [4] or as a form of imitation they are more likely to follow the crowd (exposure mechanism). "Familiarity" is the parameter of self-interest and is average similarity of two lists: interest ($I_k[i]$) of scientist k and the theme ($S_j[i]$) of artifact j. Familiarity $F_{k,\,j}$ for scientist k to artifact j is calculated as following where N is the total number of interest/theme areas.

$$F_{k,\,j} = \frac{1}{N} \sum_{i=0}^{N} Min(I_k[i], S_j[i]) \tag{2}$$

"Exposure" as the parameter of mutual-interest is calculated as the proportion of active scientists in the network of a scientist. The formula is below, where $E_{k,\,t}$ is the "exposure" for scientist k at time t, $A_{k,\,t}$ is the number of active scientists in the social network of scientist k at time t and $TS_{k,\,t}$ is the total number of scientists in the social network of scientist k at time t:

$$E_{k,\,t} = \frac{A_{k,\,t}}{TS_{k,\,t}} \tag{3}$$

"Cognitive Burden" of a scientist is dependent on two lists: Expertise ($Ex_k[i]$) of scientist k and the Complexity ($C_j[i]$) of artifact j. Both of them are defined as a list of real numbers, which are between 0 and 1. For the sake of simplicity, we assume that each scientist k has a minimum cognitive burden $minCB_k$. Cognitive burden of a scientist k for artifact j is the following where N is the total number of areas, and $C_j[i]$ is the complexity of the artifact j on theme i.

$$CB_{k,\,j} = minCB_k + \frac{\sum_{i=0}^{N} Max(0, C_j[i] - Ex_k[i])}{N} \quad (4)$$

"Tension" is related with the artifact's saturation and is higher at the beginning of the artifact' s lifetime since at the early stages of a project it is difficult to have contributions. Then tension decreases with the increasing number of contributions and goes up again proportionally to the saturation (average complexity) of an artifact when the artifact gets more mature. Project life cycle approach of [25] is the underlying assumption while calculating the tension.

Some scientists believe in the necessity of the scientific collaboration in GPS more than the others. In order to capture this, we have an independent variable "Altruism." The decision to get active for a scientist is based on the statement of [24] in the case of shared costs, which says, "if the benefit is more than the costs of an action, people will participate." We build an analogy between multiplication of self-interest and exposure for a scientist and "Benefit" as well as between multiplication of tension in an artifact and cognitive burden of a scientist and "Cost." The condition to get active is below:

$$CB_{k,\,j} \times T_{j,\,t} - F_{k,\,j} \times E_{k,\,t} \leq Altruism \quad (5)$$

where Altruism is a value, which is fixed throughout the simulation and is different for each scientist. After finding an artifact, a scientist does the cost benefit analysis described above to decide on getting active or not.

D. Foraging Mechanisms

Every scientist has different levels of expectations for the amount of time they should spend on a patch until they have a successful contribution. Each scientist has a different initial expectation, which is called "timeToReward" and shown as $TC_{k,\,t}$ for scientist k at time t. If the time passed without success on a patch is more than the expectation, then the scientist forages. Foraging is basically increasing the scope (e.g., 3 times) and moving to a different cell within the extended scope.

[26] states that a forager should leave a patch if the rate of gain (in terms of energy etc.) within the patch forager resides in drops below the rate of gain that can be achieved by traveling to a different patch. In Charnov' s Marginal Value theorem, the gain starts after a certain time t that is the amount of time forager spends to travel to a new patch. Analogically, in GPS, the amount of time spent for traveling to another patch is almost instantaneous. So in our case, the tradeoff between time spent in traveling and the expected rate of gain is different.

We have two kinds of foraging strategy in our model. Optimal foraging behavior, which is inspired by [27], checks the rate of return in terms of expertise a scientist gains from the environment. If the rate of return drops consecutively below the maximum rate of return achieved so far for a certain amount of time, then the scientist forages. In the more basic foraging strategy, if a scientist can not make a contribution for a certain amount of time then he/she forages. Every scientist has a different expectation regarding the amount of time until the success criteria is achieved.

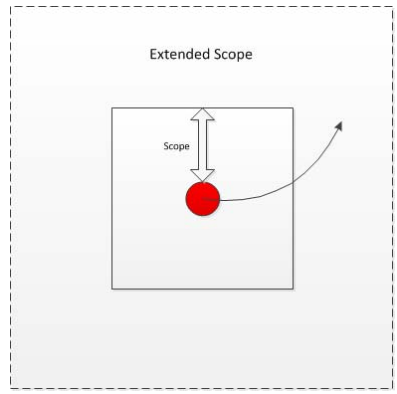


Fig. 3. Foraging Behavior

E. Other Mechanisms

SEIR model is a widely known epidemiology model [28]. It stands for 4 states an individual has transition to: "*Susceptible(S)*, *Exposed(E)*, *Infected(I)*, and *Recovered(I)*." We build a metaphor between SEIR models and our environment. All scientists start the simulation in a "Susceptible" state. When they find an artifact, they switch to "Exposed." After they get active with a contribution on an artifact, they change to "Infected." If scientists forage for a certain amount of time (e.g., 4), with certain probability (e.g., 0.01) they transfer in "Recovered" state, which means they leave the environment.

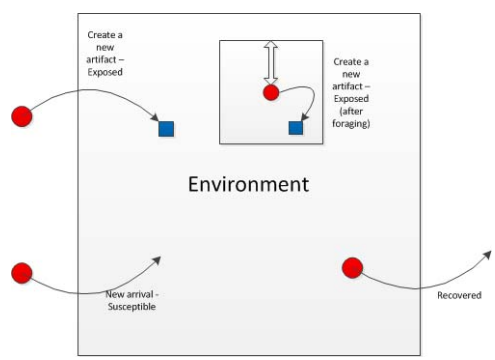


Fig. 4. Population dynamics in the system

Our simulation environment is not a closed system. Like the web platforms in real life, our model has new user arrivals. For simplicity, we do not use any recruiting by scientists. New scientists, who start to browse the system, are created with a

certain arrival rate in the context. With a certain probability (e.g., 0.2), new arrival enters the system, creating a new artifact (with probability of 0.05) or just browsing the environment. The contributions also influence the theme and complexity of the artifacts while scientists are gaining expertise from their contributions. In figure 4, we represent the new actor arrivals to the system.

IV. Preliminary Experiments and Analysis

A. Validation Processes

Along with the conceptual validation, we define operating mechanisms based on open source software community governance and observations we had in Open Biomedical Ontologies (OBO). In OBO, scientists are forming communities and domains related with different areas of health sciences while collaborating on the ontology data to standardize the shared terminology. It is a ”Sourceforge” style science development activity. In OBO data, we assume that if two scientists collaborated on the same artifact in the same month, then they are connected. OBO log-data (between 2000 - 2009) is parsed from “Sourceforge” and the network data is constructed in Auburn Modeling and Simulation Lab.

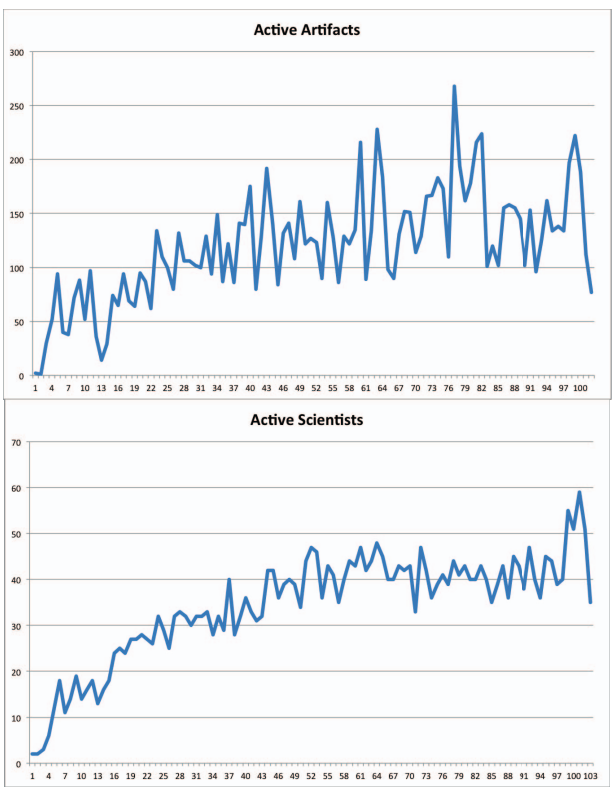


Fig. 5. Number of active artifacts and active scientists over time - OBO

In Figure 5, we plot the number of active scientists and artifacts through time for a single OBO group (Gene Ontology). We run our simulation for 500 time ticks, which can be perceived as 10 years of collaboration period. Then we measure the number of active users and the number of active artifacts over time. We represent simulation results for

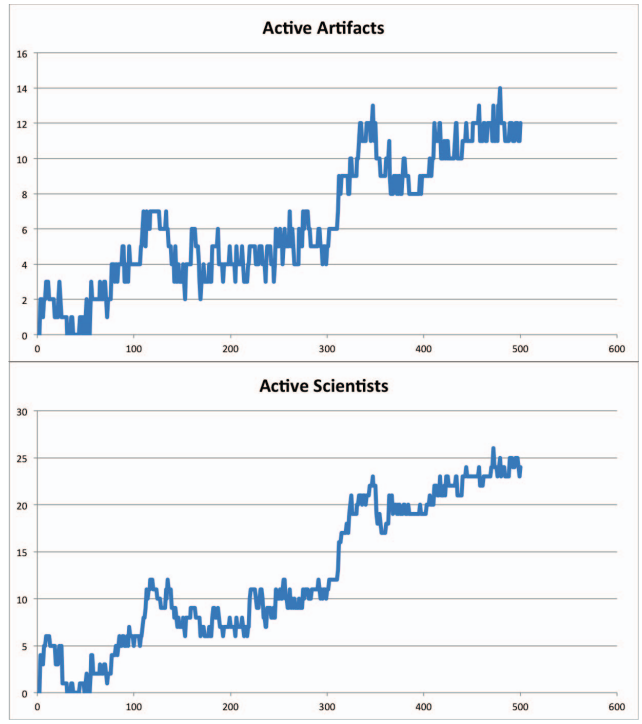


Fig. 6. Number of active artifacts and active scientists over time - Simulated

a single run in Figure 6 for illustration of similar fluctuating time series we observed in OBO. This pattern is observed at each run in our simulations. The number of active artifacts and scientists are increasing because of new arrivals in the community over time but later we observe limit cycle that means the numbers oscillate around the same values. We can also talk about adaptive renewal cycles in our simulation data that is one of the hallmarks of CAS, which we do not observe but we expect to see in OBO data if we could observe OBO communities for a longer period of time.

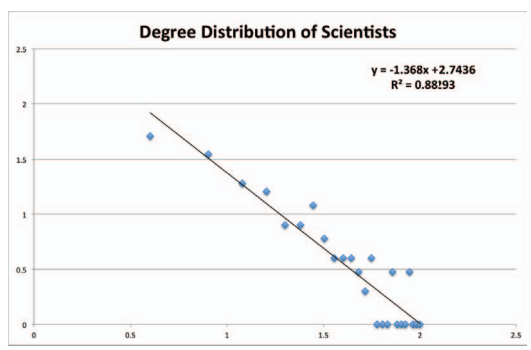


Fig. 7. Degree distribution Log-Log plot - OBO

Another phenomenon we look for is Scale-Free network structures, which creates power law distribution. We expect to observe small number of highly central users with substantial number of ties to others while most of the network members have small number of ties. We also suspect that contribution

data has the same behavior that means small number of scientists/artifacts have high number of contributions while most of the scientists/artifacts have small number of contributions. Observation of power law distributions are also typical characteristic of CAS.

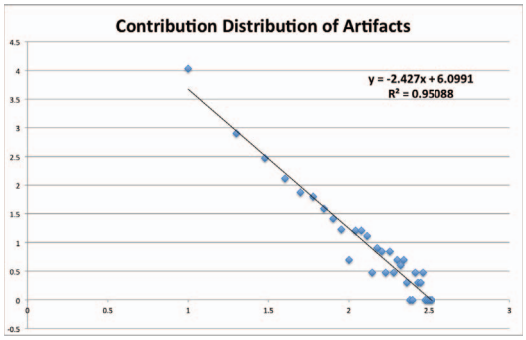


Fig. 8. Contribution distribution Log-Log plot of artifacts - OBO

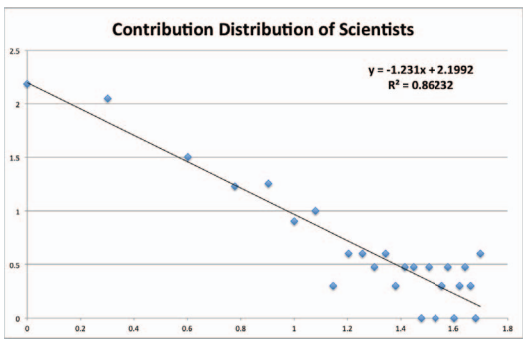


Fig. 9. Contribution distribution Log-Log plot of scientists - OBO

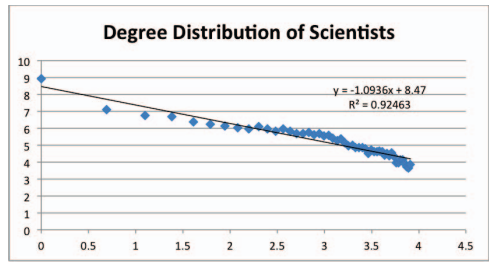


Fig. 10. Degree distribution Log-Log plot - Simulated

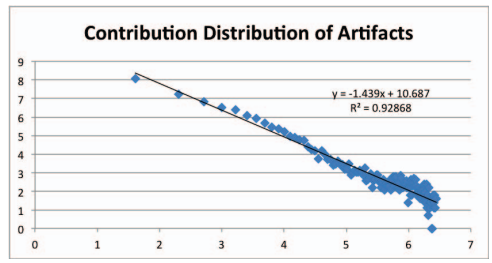


Fig. 11. Contribution distribution Log-Log plot of artifacts - Simulated

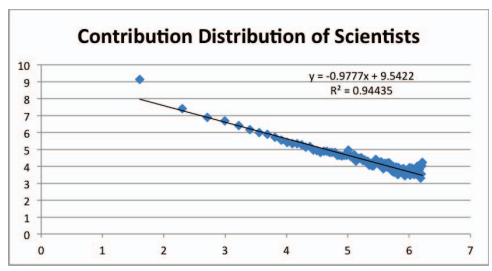


Fig. 12. Contribution distribution Log-Log plot of scientists - Simulated

We present the log-log diagrams of the contribution distribution of a scientist, degree of a scientist, and contribution distribution of an artifact in figures 7, 8, and 9 considering whole OBO community for better illustration. Because of multiple observations of the same value (or zeros), there is a noise in the tail. We excluded outliers in the tail for better illustration. There are two ways to create bins of data while looking for power law distribution. First way is to have equal width for each bin and second way is to normalize the widths of bins (logarithmic bins etc.). In each figure, we used bins of equal width. In figures 10, 11, and 12, in order to generate more data and better illustration, we ran our simulations for 200 times. OBO data and simulation data are indicative of the existence of power law distribution as we expect.

B. Preliminary Analysis

In this section, we simulate 8 different scenarios. The impact of different scenarios on innovation potential is discussed by measuring 3 social network metrics:

- *Degree Centrality of Network* is $DC_{Network} = \frac{\sum_{i=0}^{N} DC_{max} - DC_i}{N-2}$, where N is the total number of nodes and DC_{max} is the maximum degree centrality a scientist has in the network.
- *Density* is $\frac{2|E|}{N(N-1)}$, where $|E|$ the total number of edges in a network and N is the total number of nodes.
- *Clustering Coefficient* is the number of edges in a neighborhood divided by the maximum number of edges that could exist in that neighborhood. Basically, for each scientist i we define his/her neighborhood and assuming that this neighborhood is a network itself, we measure the proportion of possible ties existing between neighbor nodes. Clustering Coefficient of whole network is defined as the average of the clustering coefficients of individual scientists.

As mentioned before, we have 2 different information foraging strategies; Optimal Foraging and Basic Foraging strategies. Additionally, 4 preferential attachment scenarios are created by setting the weights of dimensions to different numbers in the artifact selection process. We conducted 30 runs for each scenario.

Higher density suggests us higher connectivity/group cohesion [29]. It promotes information sharing with high con-

nectivity that enhances innovation potential. But we are also interested in quantifying the variability of the individual indices so we calculate degree centrality of the network. Because in terms of promoting innovation, there are two competing hypothesis that are mentioned in [30]: (1) High Centrality-Low Density networks are desired because they have unique sources/actors that connect different clusters with more structural holes and (2) High Density-Fewer structural holes (with moderate level degree centrality) networks reflect better trust and connectedness so more innovation mobility and fewer leaders can facilitate the network [31].

Clustering coefficient is also an important metric to understand the network topology. It is indicative of presence of different communities/groups in the network [32]. Higher values might indicate sparsely clustered groups or a high connectivity in the whole network as a structure. High level of cohesion and clustered structures with structural holes between clusters who fosters the dissemination of ideas are desired.

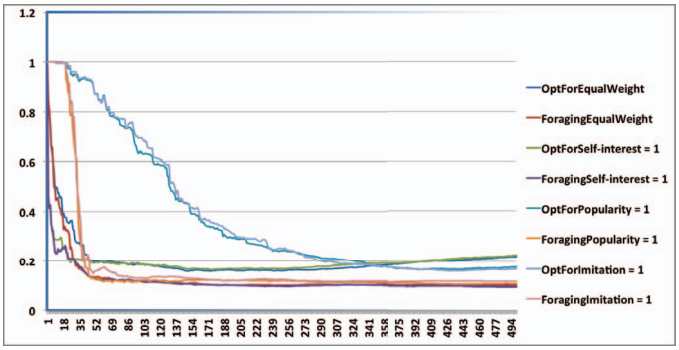


Fig. 13. Density of Social Network

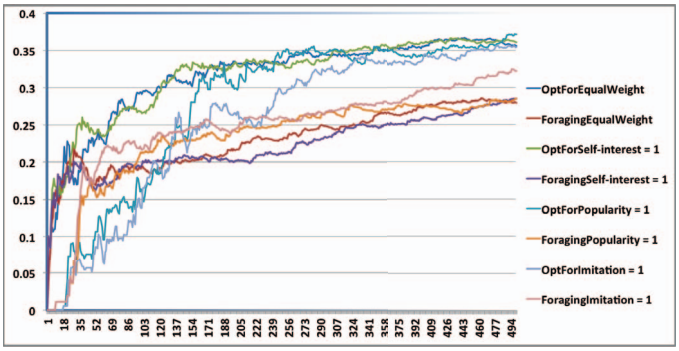


Fig. 14. Degree Centrality of Social Network

After long run, all the network metrics are fluctuating around the same values, which we observed in OBO and core/periphery networks. Regarding figure 13, optimal foraging strategy with only popularity and optimal foraging strategy with only imitation result in connected networks as a whole, which have high density. They have high connectivity during the early stages of the network that decreases gradually. But as a consequence, the clustering coefficient is lower in the early stages, which is caused by the lack of clusters and having

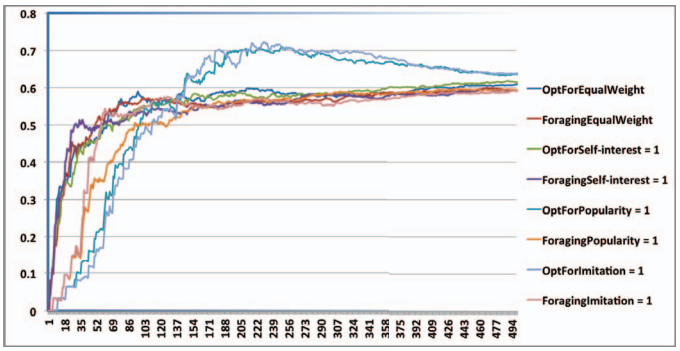


Fig. 15. Clustering Coefficient of Social Network

a one huge cluster. The best connectivity is created by the optimal foraging strategy with self-interest.

If we look at figure 14, we can see that optimal foraging strategy outperforms the basic foraging strategy, creating highly central actors in the network. In figure 15, we can observe that in long run, optimal foraging strategy with only popularity and optimal foraging strategy with only imitation are resulting in highly clustered networks more successfully than the others. Optimal foraging strategy again outperforms the other foraging strategy in terms of clustering coefficient.

V. Conclusion

In this study, we introduced CollectiveInnoSim model and briefly described the self-organizing mechanisms we demonstrated in the model. We adopt CAS approach and ABM mindset in modeling process and we discussed about some of the validation opportunities and efforts we implemented. We analyzed implications of different foraging and preferential attachment mechanisms on selected social network metrics. It can be indicated that optimal foraging with only popularity and optimal foraging with only imitation support "High Centrality-Low Density" hypothesis better than the others. "High Density-Fewer structural holes" hypothesis can be achieved by lower clustering coefficient in the network because structural holes are connections between different clusters, which result in lower clustering coefficient. Hence, optimal foraging with only self-interest supports the second hypothesis better since it has high density and relatively lower clustering coefficient. But the interpretation of the metrics in this analysis are relative to the decision-maker and could be interpreted along with different dimensions such as diversity in the network and robustness. The future work would be to analyze the impact of different scope values along with different foraging scenarios. We can also explore if there is a diminishing return or outperforming combination of weights regarding our preferential attachment mechanisms. One other potential extension would be to add an adaptation mechanism that updates the expected amount of time that scientists wait before foraging continuously.

ACKNOWLEDGMENT

The research was partially supported by NSF-SBE-0830261. We thank to members of Auburn Modeling and Simulation Lab for the stimulating environment and their support.

REFERENCES

[1] G. Zou, "SpringerLink - Computational & Mathematical Organization Theory, Volume 17, Number 1," *Computational and Mathematical Organization Theory*, no. 17, pp. 35–58, 2011.

[2] S. O'Mahony and F. Ferraro, "The emergence of governance in an open source community," *Academy of Management Journal*, vol. 50, no. 5, pp. 1079–1106, 2007.

[3] C. Wagner, *The new invisible college: Science for development*. Brookings Inst Pr, 2008.

[4] P. Monge and N. Contractor, *Theories of communication networks*. Oxford University Press, USA, 2003.

[5] M. Nielsen, "The Logic of Collective Action," 2010.

[6] E. Ostrom and C. Hess, *Understanding knowledge as a commons: from theory to practice*. MIT Press, 2007.

[7] J. R. Koza, "Hidden Order: How Adaptation Builds Complexity.," *Artificial Life*, vol. 2, pp. 333–335, jan 1995.

[8] E. Milbergs and N. Vonortas, "Innovation Metrics: Measurement to Insight," tech. rep., IBM Corporation, IBM Corporation, jan 2006.

[9] M. F. Shlesinger, "Complex Adaptive Systems: An Introduction to Computational Models of Social Life," *Journal of Statistical Physics*, vol. 129, pp. 409–410, sep 2007.

[10] R. Axtell and J. Epstein, "Coordination in transient social networks: an agent-based computational model of the timing of retirement," in *Generative social science: Studies in agent-based computational modeling*, Princeton Univ Press, 2006.

[11] X. Cui, J. Beaver, J. Treadwell, T. Potok, and L. Pullum, "A Stigmergy Approach for Open Source Software Developer Community Simulation," *Computational Science and Engineering, 2009. CSE'09. International Conference on*, vol. 4, pp. 602–606, 2009.

[12] L. Yilmaz, "On the synergy of conflict and collective creativity in open innovation socio-technical ecologies," in *Proceedings of the 2008 Winter Simulation Conference*, 2009.

[13] J. Dron and T. Anderson, "On the Design of Collective Applications," in *Computational Science and Engineering, 2009. CSE'09. International Conference on*, (IEEE), pp. 368–374, IEEE, 2009.

[14] N. Gilbert, "A simulation of the structure of academic science," *Sociological Research Online*, vol. 2, 1997.

[15] I. Naveh and R. Sun, "A cognitively based simulation of academic science," *Computational and Mathematical Organization Theory*, vol. 12, no. 4, pp. 313–337, 2006.

[16] R. Cowan and N. Jonard, "Network structure and the diffusion of knowledge," *Journal of economic Dynamics and Control*, vol. 28, no. 8, pp. 1557–1575, 2004.

[17] J. Shrager and L. Pat, "Computational approaches to scientific discovery," *Computational models of scientific discovery and theory formation*, pp. 5–25, 1990.

[18] R. Hardin, *Collective action*. Resources for the Future, 1982.

[19] C. Bernstein, A. Kacelnik, and J. Krebs, "Individual decisions and the distribution of predators in a patchy environment," *The Journal of Animal Ecology*, vol. 57, no. 3, pp. 1007–1026, 1988.

[20] P. Pirolli and S. Card, "Information foraging.," *Psychological review*, vol. 106, no. 4, p. 643, 1999.

[21] A. Barabasi, *Linked: The New science of networks*. Perseus Books, 2002.

[22] L. Hamill and N. Gilbert, "Social Circles: A Simple Structure for Agent-Based Social Network Models," *Journal of Artificial Societies and Social Simulation*, vol. 12, 2009.

[23] A. Hollingshead, J. Fulk, and P. Monge, "Fostering intranet knowledge sharing: An integration of transactive memory and public goods approaches," *Distributed work*, pp. 335–355, 2002.

[24] M. Olson, *The logic of collective action: Public goods and the theory of groups*. Harvard University Press, 1974.

[25] D. Wynn, "Organizational structure of open source projects: A life cycle approach," in *7th Annual Conference of the Southern Association for Information Systems, Georgia*, 2003.

[26] E. Charnov, "Optimal foraging, the marginal value theorem," *Theoretical population biology*, vol. 9, pp. 129–136, 1976.

[27] P. Pirolli, *Information foraging theory*. adaptive interaction with information, Oxford University Press, USA, apr 2007.

[28] M. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[29] P. Blau, *Inequality and heterogeneity: A primitive theory of social structure*. Free Press New York, 1977.

[30] L. Yilmaz, "Innovation systems are self-organizing complex adaptive systems," in *Association for the Advancement of Artificial Intelligence*, 2008.

[31] R. Burt, *Structural holes: The social structure of competition*. Harvard University Press, 1995.

[32] T. Schank and D. Wagner, "Approximating clustering coefficient and transitivity," *Journal of Graph Algorithms and Applications*, vol. 9, pp. 265–275, 2005.