ISSN 1178-2293 (Online)

University of Otago
Economics Discussion Papers
No. 1302

January 2013

An Agent-based Model of Interdisciplinary Science and the Evolution of Scientific Research Networks

Dr. Dan Farhat
University of Otago
Department of Economics

Address for correspondence:

Dan Farhat
Department of Economics
University of Otago
PO Box 56
Dunedin
NEW ZEALAND
Email: dan.farhat@otago.ac.nz
Telephone: 64 3 479 8645

An Agent-based Model of Interdisciplinary Science and the Evolution of Scientific Research Networks

Dr. Dan Farhat
University of Otago
Department of Economics
PO Box 56
Dunedin, New Zealand 9054
dan.farhat@otago.ac.nz

[Preliminary – January, 2013]

Abstract

This study proposes an agent-based model of the impact of research success on the structure of scientific communities. In the model, heterogeneous scientists scattered about a 'social landscape' influence each other through networking. Peer networks are allowed to change based on the accumulated achievements (or prestige) of researchers. The dynamics of these networks are illustrated. The framework is then adjusted to allow for interdisciplinary practices (modelled as network links to more distant peers on the social landscape). Separate disciplines are shown to collapse into a single, large scientific network. Managing growing research networks, therefore, becomes a concern.

Introduction

The creation of science is very much a social activity. The notion that a group structure of some sort underlies the practice of research emerged in the 1970s with the early works by Kuhn (1970) on paradigms of science, by Lakatos (1969, 1970) on scientific research programs and by Laudan (1977) on research traditions. This idea continued to develop in the 1990s and 2000s with work by Hands (1994, 2004) on the sociology of scientific knowledge. The central lesson taken from these studies is that analyzing the way that researchers interact with each other (via publications, at conferences, etc.) within the institutions that they themselves create (such as professional organisations, journals, etc.) contributes to our understanding of what 'good science' is and how it came to be. The influence, however, may not be one-way. Can the practice of research affect the structure of research communities? If so, what sort of process might explain the how these communities evolve over time? Will current trends promoting interdisciplinary research practices drastically impact our science culture?

To address these questions theoretically, this study uses a novel computational method known as *agent-based modelling* [ABM].[1] In this approach, artificially intelligent individuals are created in a virtual environment. Each agent is provided with private information and a set of objectives to achieve. As agents interact with each other (as well as the virtual landscape around them), aggregate patterns begin to form (i.e. macroscopic phenomena are built from the 'bottom-up'). The output generated by the model is *explained* by the structure of the agents, the formulation of the environment, and the system of interaction within the

[1] There are several introductory guides on agent-based modelling. Readers interested in this approach can consult Gilbert & Conte (1995), Epstein (1999), Macy & Willer (2002), Tesfatsion (2002) and Macal & North (2010) for a user-friendly introduction.

framework. Since these are all designed by the programmer and subject to few structural limitations, agent-based models can accommodate more realism and complexity than other theoretical methods.

Other studies have used this method to study the 'sociology of science'. Gilbert (1997) models the production of academic science in which researchers write papers based on their own past work and the work done by randomly selected peers. This model produces realistic citation behaviour and generates 'specialties' within academic fields. Sun and Naveh (2009) build on Gilbert (1997) by adding more realistic agent cognition.[2] Edmonds (2007) also adds to this strand of research by allowing agents to use a set of logic rules to create new knowledge from publically available resources.[3] In these models, strong emphasis is placed on how social process *produce* science. Farhat (2011), on the other hand, illustrates how the production of science might affect the structure of scientific communities by allowing agents to organize themselves into evolving 'scientific research programs' [SRP] (à la Lakatos). Kuhn's (1970) notion of the rise and fall of scientific paradigms is illustrated as agents change allegiances to these SRPs over time. This study contributes to this latter line of research.

In the model below, the structure of scientific communities is represented by networks between heterogeneous peers. A networks-based approach is of growing popularity amongst researchers who use agent-based models to study the process of innovation. For example, Gilbert, Pyka and Ahrweiler (2001) and Pyka, Gilbert and Ahrweiler (2007) have illustrated how realistic innovation networks can form using agent-based models of knowledge production. This study uses a simpler algorithm for producing knowledge and illustrates the properties of formed networks using geographical space (or 'social landscape'). Researchers are first scattered across this space, then allowed to form links with others to construct a personal research network. The size and composition of this network affects how successful the researcher is at producing science. Once research is performed and the results are disseminated, agents are allowed to adjust their network by moving across the social landscape. Movements depend on the achievements occurring in the scientific community. The formulation of each agent's network and the accuracy of each agent's research activity are monitored during simulation.

In the remainder of this paper, details of the agent-based framework are presented and select simulation results are described. These results show that clustered networks appear in the social landscape under multiple types of networking behaviours. A positive relationship emerges between the number of peers a researcher has and their accumulated success by the end of the simulation. Incorporating a notion of 'interdisciplinary research' by allowing agents to link with far-away researchers leads to a single, large cluster forming (i.e. distinct disciplines collapse and personal connections are formed with all other agents in the virtual environment). The paper concludes with directions for future work.

[2] Agents actively identify lucrative strands of research, establish private values for different research outputs, and either become successful or die out. Similar results to Gilbert (1997) are generated.

[3] Agents pull theorems from a public repository, apply a rule of logic for creating new theorems, then publish a set of their results.

An Agent-based Model of Scientific Networks

Consider a virtual environment where I agents perform academic research in an attempt to earn prestige. As in Farhat (2011), 'performing science' involves successfully matching one or more of D stylised facts. Denoting the probability that agent i successfully reproduces stylised fact d in period t as $p_{i,d,t}$, the practice of science is modelled as pulling a random number from a uniform distribution with minimum 0 and maximum 1. If the value of the draw is less than $p_{i,d,t}$, the researcher is awarded a success (valued at 1). Otherwise, they are awarded a fail (valued at 0). Each agent maintains a record of the number of attempts and the corresponding outcomes for each stylised fact. From this data, they calculate their total prestige at period t as the total number of successes they have achieved since the start of the simulation. They compute their success rate for each fact, $sr_{i,d,t}$, as the total number of successes agent i obtained for stylised fact d since the start of the simulation relative to the total number of attempts for that fact.

In the model, an agent's private ability is partially influenced by the research successes of their peers. Denoting $sp_{i,d,t}$ as the total number of successes accumulated by time t by agent i's peer group for stylised fact d and $ap_{i,d,t}$ as the total number of attempts that that group has made to match fact d, the overall success rate of agent i's peer group for fact d, as $srp_{i,d,t}$, can be computed ($= sp_{i,d,t} / ap_{i,d,t}$). Note that this value changes with the composition of the agent's peer group. The agent's private ability is assumed to evolve according to the expression: $p_{i,d,t+1} = g \times srp_{i,d,t} + (1-g) \times p_{i,d,t}$ where $g \in [0,1]$ affects how magnitude of influence the peer group exerts on the agent's ability.[4]

Gilbert (1997) uses a geographic space to describe the relationship between newly published papers and those they cite. Hamill and Gilbert (2009) model general social network formation in geographic space. This study uses a similar spatial method[5] to model the 'scientific community'. At the start of the simulation (t = 0), each agent is randomly assigned a pair of coordinates ($x_{i,0}$ and $y_{i,0}$) on a plane of width X and length Y. Agents are allowed to form a personal network by establishing one-way links to all other agents in their immediate vicinity. Denoting R as an agent's social reach, agent j will become a peer of agent i if the Euclidean distance between the two is no greater than R.[6] The network formed from this process can be thought of as a 'regular lattice' where each agent is only connected with its immediate neighbours. If allowed, agents can also make a connection with a randomly-selected distant peer in each period, t, with probability $F > 0$. As a result, a 'small-world' network structure emerges (see figure 1 for an illustration). In networks of this sort, an agent may be influenced by a large number of other agents via a relatively low number of interpersonal links. These two types of network structures are commonly used with varying success.[7]

[4] Note that if $g > 0$ and the agent's peer group consistently generates $srp_{i,d,t} = 0$ (either by always failing to reproduce fact d or by never attempting it), the agent's ability to replicate fact d will erode to zero. As a result, the agent will come to 'specialise' in a subset of stylised facts.
[5] Hamill & Gilbert (2009) use an unbounded grid (a sphere). To make the model simple to program, this study uses a bound grid (a plane) similar to that in Gilbert (1997).
[6] Note that establishing one-way links in this fashion generates reciprocation between local agents. In other words, if agent j is in the vicinity of agent i, then agent i is also in the vicinity of agent j. Hence, agent i will also be a member of agent j's peer group and the links between them will appear to be two-way.
[7] See Hamill and Gilbert (2009) for a concise overview of commonly-modelled network structures and the features they exhibit.

Agents actively move around the social space in hopes of improving the quality of their network. Three particular forms of movement will be considered:

a. *Move-to.* In this regime, agents move towards the most successful peer in their network. First, the agent identifies who in their network has the most accumulated prestige (denoted as j^+) and then closes the Euclidean distance between them[8]:

$$x_{i,t+1} = x_{i,t} + m(x_{j+,t} - x_{i,t})$$
$$y_{i,t+1} = y_{i,t} + m(y_{j+,t} - y_{i,t})$$

b. *Move-from.* In this regime, agents move away from the least successful peer in their network. After identifying the colleague with the least accumulated prestige (denoted as j^-), the expand the Euclidean distance between them:

$$x_{i,t+1} = x_{i,t} - m(x_{j-,t} - x_{i,t})$$
$$y_{i,t+1} = y_{i,t} - m(y_{j-,t} - y_{i,t})$$

c. *Mixed.* This approach is a mixture of the two above movements (moving towards the most successful and away from the least successful):

$$x_{i,t+1} = x_{i,t} + m(x_{j+,t} - x_{j-,t})$$
$$y_{i,t+1} = y_{i,t} + m(y_{j+,t} - y_{j-,t})$$

In the strategies above, $m > 0$ is a scale parameter that guides the magnitude of movement. As agents move around the social space, the local peer network is redefined. Distant peers (from the establishment of small-world networks) remain in the network unless they come to be the least prestigious peer under b and c, in which case ties with them are cut and the agent keeps their present position.

The model is simulated for T periods, where the following occurs in each period:

1. Agents identify their peer group.
2. Agents update their private abilities for each stylised fact.
3. A set of stylised facts is randomly chosen for each agent to attempt. To limit the scale of research activity for any agent, each fact is chosen for an attempt with a small probability $\delta > 0$.
4. Agents perform science. Successes and failures are awarded for attempted stylised facts and accumulated prestige is updated.
5. Agents adjust their position in the social space using updated prestige values.
6. Repeat at 1.

At each round, data regarding prestige, research success and social location for each agent is recorded.

[8] Gilbert (1997) uses a similar rule to establish the position of a new paper relative to the papers it cites.

Experiments

In the simulations that follow, parameters of the model are calibrated as follows: I = 1000, D = 1000, T = 100, the social space has dimensions X = 400 and Y = 300, R = 25, g = 0.1, m = 0.5 and δ = 0.5%. These values are arbitrary, but are sufficient for the thought experiments below.[9]

Experiment 1 – No Social Movement in a Regular Lattice

The model is first simulated for agents that have neither the ability to move from their initially-allocated location on the social landscape nor the ability to make connections to distant peers. As shown by figure 2(a), which illustrates the accumulated prestige achieved by each agent by the end of the simulation, agents remain evenly distributed across the social landscape (hence personal networks are fixed). Figure 2(b), indicates that there is a weak positive relationship between network size and accumulated prestige in the final period of simulation (agents located either near the edge of the social space or in sparse neighbourhoods have fewer successes). This is likely due to fixed networks and the algorithm by which scientific abilities update. Since an agent will lose the capability to match a stylised fact if their network consistently fails at matching it or if attempts by their network to match it are low, the skills of agents with small networks erode quite quickly.

Experiment 2 – Towards Success or Away from Failure in a Regular Lattice

In this experiment, the three movement regimes described above are compared under the assumption that agents cannot make distant friends. In all experiments, distinct scientific communities form.

Figure 3(a) shows that the ‘move-to’ strategy leads to the formation of ‘cliques’ where members are connected with each other, but not with any outsiders.[10] These cliques, which form quickly and then remain fixed during the simulation, can be thought of as scientific communities with isolationist doctrines (explicitly defined ‘disciplines’ or ‘specialties within disciplines’). Figure 3(d) shows that a positive correlation between network size and accumulated success emerges by the end of the simulation (i.e. the most successful tend to be the most popular), but the size of an agent’s peer groups is limited (to 60 or less).

Figure 3(b) shows that the ‘move-from’ strategy produces loose clusters, with the worst performers becoming excluded from joining any community. These sorts of scientific communities contain members who may not know each other, but who can still influence one another via shared peers. Further, these communities also change in composition as the simulation progresses as movements occur perpetually during the simulation.[11] As before, figure 3(e) shows a positive relationship between network size and prestige. Network size for the majority of researchers tends to be quite small (50 or less), but some large networks form (with agents having up to 150 peers).

---

[9] Evaluating the impact of changing these parameters on the virtual world is left for future work.
[10] A point in figure 3(a) represents several agents stacked on top of one another.
[11] Figure 3(b) shows the clusters as they appeared in the final period of the simulation.

When agents adopt the 'mixed' movement regime, tight clusters emerge (where agents band together with prestigious peers) with many least successful agents becoming isolated (as groups of agents collectively shun least successful peers).[12] See figure 3(c). Note that the clusters which form for this movement regime tend to contain more members than the cliques formed by the 'move-to' strategy. Figure 3(e) again reveals a weak positive relationship between network size and prestige. Most researchers have moderately sized peer groups (50 – 100), but some have built extensive networks (up to 250 peers).

Figure 4 compares the average number of science successes across researchers in each simulation period, t, for each movement strategy. The mixed strategy provides the largest number of successes on average, followed next by the move-to strategy, then by the move-from strategy. No movement produces the lowest number of successes on average. The differences between the movement regimes on this dimension, however, are quite small.

The agent-based model in this experiment attempts to explain how the current structure of scientific networks may have formed. Tight-knit groups of researchers organised into disciplines or specialties (either within fields of study or across multiple fields) emerge when agents gravitate towards successful practitioners. The propensity to shun unsuccessful researchers loosens these networks with some agents becoming completely isolated. With this behaviour, the structure of personal networks perpetually changes over time. In all simulations, however, those who connect with a large number of others are generally the most accurate.

Experiment 3 – Interdisciplinary Science and the Creation of Small-World Networks

In this experiment, agents are selected to make a connection with a randomly-chosen distant peer in each period with probability $F = 5\%$. The virtual environments produced when there is no social movement and when movements are guided by the 'move-to' strategy are compared. Doing this allows us to identify any perpetual movement produced exclusively by the small-world network structure.[13] Knowing that the 'move-to' regime generates distinct 'disciplines', distant connections are allowed to form only after period $t = 25$ to give these research cliques a chance to form. This makes the model relatable to the emergence of interdisciplinary research.

Figure 5(a) and 5(c) show that the small-world network structure produces a scientific environment quite similar to that of the regular lattice when there is no social movement at all. Connecting to distant agents simply gives each researcher the chance to increase the size of their personal network. As before, a positive relationship between network size and accumulated prestige emerges by the end of the simulation.

[12] Unlike figure 3(a), the majority of distinct points in figure 3(c) represent a single agent isolated from the tight clusters.
[13] Recall the movement regimes where agents run away from weaker peers produced perpetual moves across the social landscape in regular lattices whereas the regime where agents move towards their most successful peer resulted in a 'stable' social environment.

Figure 5(b) and 5(d) show that the small-world network structure results in a single loose cluster forming over time.[14] Isolated research cliques (which emerge prior to period 25 in the simulation) completely collapsed into sparser interconnected private networks. The production of this loose cluster, driven in part by the opportunity to make distant friends, results in a dramatic increase in personal network size. Notably, any positive correlation between network size and accumulated success disappears (i.e. being popular is no longer associated with being successful). Were the simulation allowed to continue beyond $T = 100$, or if the propensity to make distant connections (F) were to increase, this cluster would tighten and individual network size would approach the population size (i.e. a researcher would have a connection to every other person in the virtual world).

Figure 6 shows that the average number of successes per researcher is higher when social movement is permitted in the presence of interdisciplinary practice. Although the difference is distinct, it is not large. Also, careful comparison with figure 4 shows that the presence of interdisciplinary practise has little impact on achieving successful outcomes. The behaviour guiding network evolution drives the differences in achievement.

The agent-based model in this experiment attempts to predict how interdisciplinary practices may affect the research community. When the ability to like to peers outside their ‘neighbourhood’, moving towards successful peers causes separate scientific communities to collapse into a single mega-community over time. In other words, interdisciplinary science drives isolated disciplines to extinction.

Conclusion

Agent-based models, which explain the emergence of complex phenomena from the bottom-up, have produced insightful contributions in a variety of research areas. The agent-based model in this study uses spatial network formation to identify the potential impact of research success on the structure of scientific communities. The model produces clusters of networked scientists whose ‘closeness’ depend on how links between researchers responds to research success. A positive correlation between personal network size and accumulated prestige emerges. Adding the potential for interdisciplinary study produces a single, grand scientific community (provided researchers close the social distance between themselves and successful peers). This poses an important policy question; namely, how can scientists efficiently manage extensive research networks. With modern research technology (including search engines for scholarly journals and social networking websites for scientists), maintaining large networks seems feasible.

Much is left for future work. As noted above, the model is designed to explore the impact of research outputs on the scientific community. When looking at the impact of scientific network evolution on the number of successes achieved by researchers, results are quite mild. Additional elaboration on how science is produced in the presence of research networks (along the lines of Gilbert, Pyka and Ahrweiler (2001) and Pyka, Gilbert and Ahrweiler (2007)) may improve the model on this dimension. Also, network formation is

[14] Increasing the propensity to form networks with distant peers (F) can result in this cluster tightening.

costless in the model above whereas in reality it is not. Adding 'social costs' into the framework and asking agents to form links in a cost-effective way[15] is also left as a future project. Connecting simulation output to the empirical regularities associated with scientific collaboration (as seen in such studies as Newman (2001) and Barabási et al. (2002)) is a necessary next step.

References

Barabási, A. L. , Jeong, H., Néda, Ravasz, E., Schubert, A. & Vicsek, T. (2002). Evolution of the Social Network of Scientific Collaborations, *Physica A*, 311, 590-614.

Edmonds, B. (2007). Artificial Science: A Simulation to Study the Social Process of Science, in *Social Simulation: Technologies, Advances and New Discoveries*, (eds.) B. Edmonds, K. G. Troitzsch & C. H. Iglesias, Hershey: IGI Publishing, 61-67.

Epstein, J. M. (1999). Agent-based Computational Models and Generative Social Science, *Complexity*, 4(5), 41-60.

Farhat, D. (2011). Virtually Science: An Agent-based Model of the Rise and Fall of Scientific Research Programs, *Journal of Economic Methodology*, 18(4), 363-385.

Gilbert, N. & Conte, R. (eds.)(1995). *Artificial Societies: The Computer Simulation of Social Life*, London: University College London Press.

Gilbert, N. (1997). A Simulation of the Structure of Academic Science, *Sociological Research Online*, 2, available at http://socresonline.org.uk/2/2/3.html.

Gilbert, N., Pyka, A. & Ahrweiler, P. (2001). Innovation Networks – A Simulation Approach, *Journal of Artificial Societies and Social Simulation,* 4(3)8, available at http://jasss.soc.surrey.ac.uk/4/3/8.html.

Hamill, L. & Gilbert, N. (2009). Social Circles: A Simple Structure for Agent-based Social Network Models, *Journal of Artificial Societies and Social Simulation*, 12(2)3, available at http://jasss.soc.surrey.ac.uk/12/2/3.

Hands, D. W. (1994). The Sociology of Scientific Knowledge, in R.E. Blackhouse (ed.), *New Directions in Economic Methodology*, New York: Routledge, 75-106.

Hands, D. W. (2004). Constructivism: The Social Construction of Scientific Knowledge, in J. B. Davies, A. Marciano & J. Runde (eds.), *The Elger Companion to Economics and Philosophy,* Northampton: Edward Elgar, 197-212.

Kuhn, T. (1970). *The Structure of Scientific Revolutions (2nd ed.)*. Chicago: The University of Chicago Press.

Lakatos, I. (1969). Criticism and the Methodology of Scientific Research Programmes, *Proceedings of the Aristotelian Society*, 69, 149-186.

Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programmes, in I. Lakatos & A. Musgrave (eds.), *Criticism and the growth of knowledge*, Cambridge: The Cambridge University Press, 91-196.

Laudan, L. (1977), *Progress and its Problems: Toward a Theory of Scientific Growth*, Berkeley: University of California Press.

Macal, C. M. & North, M. J. (2010). Tutorial on Agent-based Modelling and Simulation, *Journal of Simulation*, 4, 151-162.

Macy, M. & Willer, R. (2002). From Factors to Actors: Computational Sociology and Agent-based Modelling, *The Annual Review of Sociology*, 28, 143-166.

Newman, M. E. J. (2001). The Structure of Scientific Collaboration Networks, *Proceedings of the National Academy of Sciences*, 98(2), 404-409.

Pyka, A., Gilbert, N. & Ahrweiler, P. (2007). Simulating Knowledge – Generation and distribution Processes in Innovation Collaborations and Networks, *Cybernetics and Systems: An International Journal*, 38(7), 667-693.

Sun, R. & Naveh, I. (2009). Cognitive Simulation of Academic Science, in *Proceedings of the 2009 International Joint conference on Neural Networks*, 418-424.

Tesfatsion, L. (2002). Agent-based Computational Economics: Growing Economies from the Bottom Up, *Artificial Life*, 8(1), 55-82.

[15] As in both Gilbert, Pyka and Ahrweiler (2001) and Pyka, Gilbert and Ahrweiler (2007).

Figures

Figure 1: Geographical Network Structure

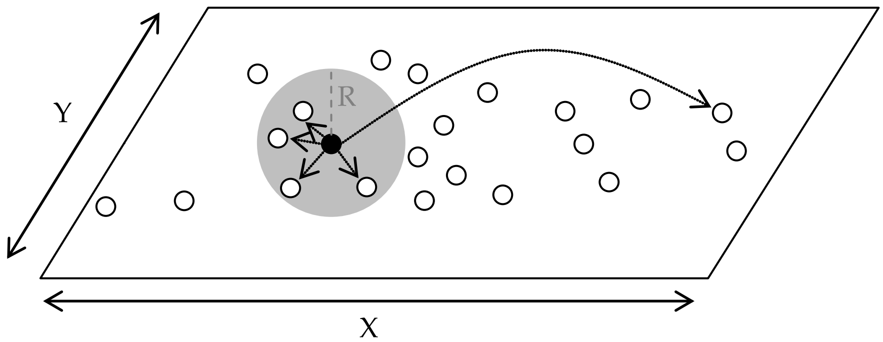
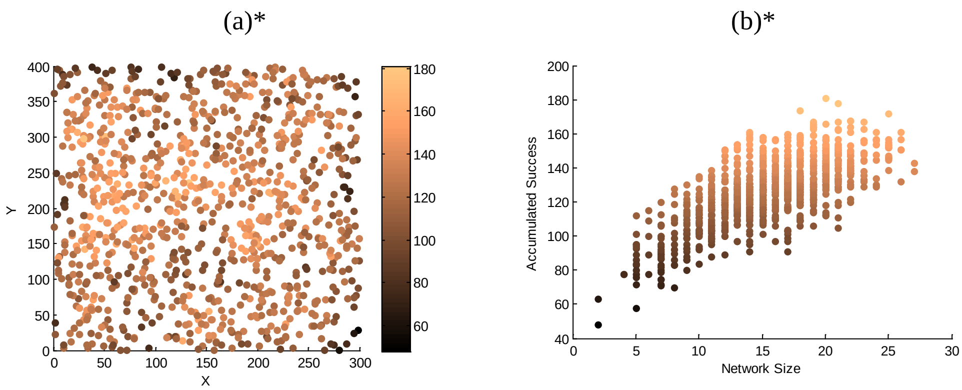
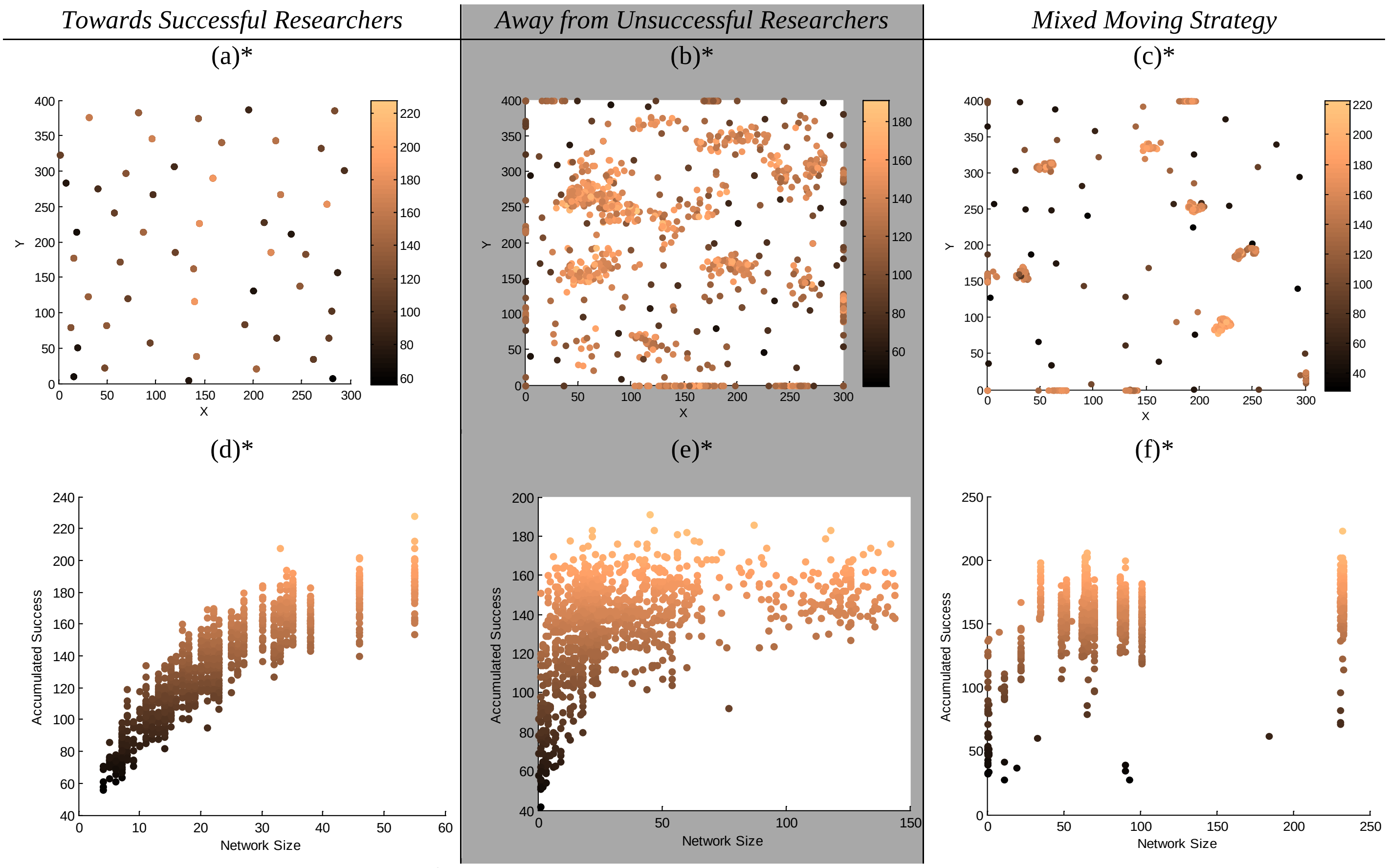


Figure 2: No Social Movement



* Illustrates the state in period T, the last period of the simulation.

Figure 3: Social Movement



* Illustrates the state in period T, the last period of the simulation.

Figure 4: Research Success and Social Movement

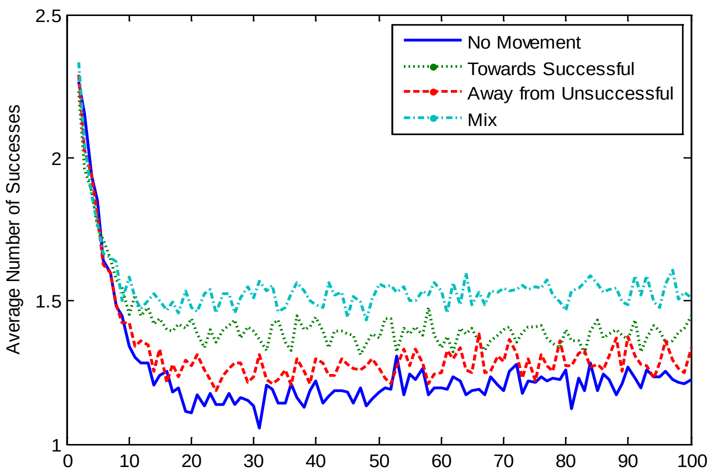
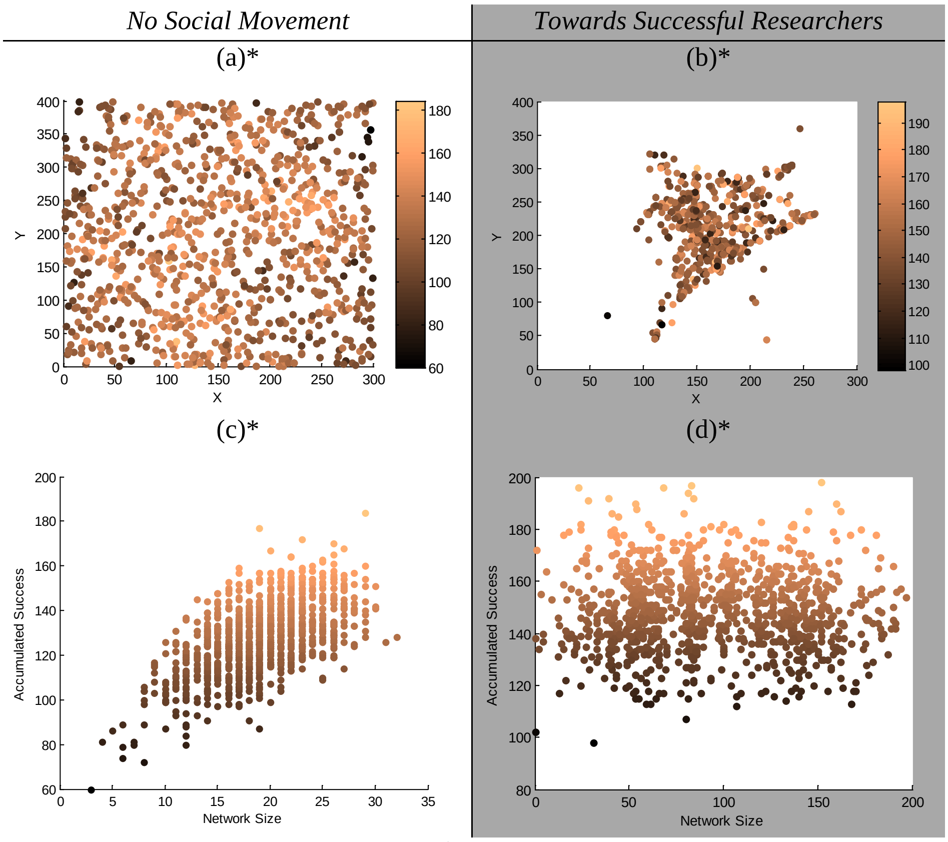


Figure 5: Interdisciplinary Science



* Illustrates the state in period T, the last period of the simulation.

Figure 6: Research Success, Social Movement and Interdisciplinary Practice