How does SFI encourage collaborations?

Wang Jing, Li Bingwei, Eliza Olivia Lungu, Zhang Nan, Yu Wei
s.s.f.i.2008@gmail.com

It is one part of a report which is under construction and recording what we planned to do and what we actually did or why and how we did what.

Abstract

During the CSSS, we conducted a series of interviews to know more about Santa Fe Institute (SFI). From the qualitative data (interviews, observations etc.) we gained the insights and formalized them into an agent base model (ABM). Our focus is the mechanism of collaboration and the function of SFI to help scientists know each other and then develop collaborations. In order to validate our ABM we used the real data from SFI working paper database.

Acknowledgement

We would like to show our gratitude to all the people who accepted our interviews.

This collaboration was initiated at the 2008 China Complex Systems Summer School, co-sponsored by the Santa Fe Institute and the Institute of Theoretical Physics of the Chinese Academy of Sciences.

The research and education programs of SFI are supported by core funding from the U.S. National Science Foundation and by gifts and grants from individuals, corporations, other foundations, and members of the Institute's Business Network for Complex Systems Research. SFI received direct support for the 2008 China CSSS from the US NSF under award OSIE-0623953.

Our Questions

Before we know everything about SFI we have already had some questions:

1) Will the SFI way, if there really is one, doing the science be the mainstream of academia in the future?
2) What is the relationship between SFI and the whole academia in which SFI is embedded?
3) What is complexity, interdisciplinary studies etc. actually? Or in the mind of scientists?
4) What is the structure of SFI and what culture or norm has evolved upon?
5) What is the interpretation of SFI individually or informally?
6) From all the above we would like to know that how people practice, namely what they do, within the cultural structure.
7) After all, how could we approach all the questions step by step and find something concrete and able to do the follow-ups within this three weeks' summer school?

… …

Since the first day or even before our project team had been formed, we began to interview scientists who was visiting the summer school and who are or are not doing research at SFI at the moment. Simultaneously, we were trying to collect the empirical data to testify our model.

We chose to observe everything going on, listen to their stories with SFI, to borrow their insights to construct our model. Then conceptualize what we found into two categories, network and knowledge. Networks, which might tie together researchers in a newly emerging synthesis, using both traditional and innovative forms of communication, raging from workshop and the exchange of graduate students and postdoctoral fellows to computer links, emerged as one of the initial foci of the Institute activity (David Pines, 1988). Knowledge is no doubt the outcome produced by SFI and flows in the conduits of networks. The first flow chart, network structure actually, in first proceeding of SFI is as followed,

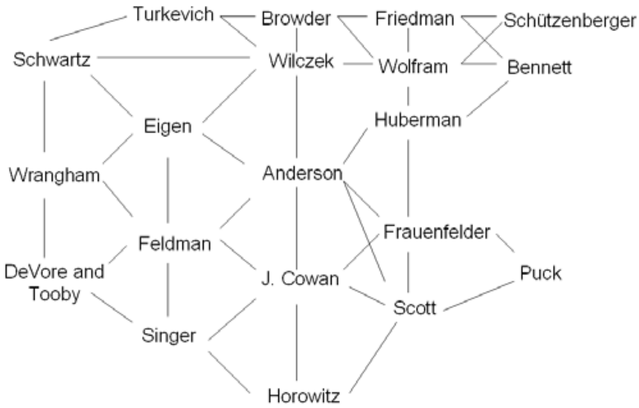

FIGURE 1 FLOW CHAT, illustrating some of the connections between talks and topics discussed at Santa Fe Institute Workshop on "Emerging Syntheses in Science."

The figure illustrate some of the connections between talks and topics discussed at SFI workshops on "Emerging Syntheses in Science", which does not have an elaborate illustration except for the sentence above. That will be quite helpful if we could know the criteria author employed to construct the network. Then we could know more about the early SFI and develop our studies upon it. We could not tell what each tie is about exactly but we can feel that the meaning of each tie is multiple and could be thought of as a combination of network and knowledge, the former of which is form while the latter is content.

Since Santa Fe Institute (SFI) focuses on the studies of complexity through interdisciplinary studies, collaboration should be one of the forms of most interdisciplinary studies and one of various kinds of cooperation. We decided to approach our topic by studying collaboration first. From the micro level we can study how people collaborate with others; from the macro level we can study the function of SFI in facilitating it.

Agent Based Model

Our purpose was to determine how collaborations are formed at SFI and especially how SFI as an institution encourage their formation. For this we used two approaches: we constructed the collaboration network between scientists that have connections with SFI and we tried to identify the mechanism that stands behind these collaborations. Then we formalized the mechanism observed into an agent based model and ran several simulations.

The discussions were held during the four weeks of the summer school and where focused on interviewer experience with interdisciplinary research, SFI and collaboration. The questions include: What is your criterion when choosing a collaborator? What is the difference between SFI and traditional research university? What do you consider to be the contribution made by SFI?

People are talking about "chemistry" when they meet someone they have a crush on. Actually, from our interviews, we found it is also important to choose a collaboration partner. "First, I should like the one then I will decide to collaborate… ". So we think the process of collaboration is more than one kind of rational, which is functioning. People want to find someone they like then they will see if there is anything they can do together. Another question pops up that why A has chemistry with B but not C? We are not very sure but it might be a cognitive problem. People would like to say the collaboration is considered to take place when two scientists share the same social values and have similar professional interests. We think that chemistry is more than just social values which could be not individual enough and should be more like a kind of feeling. As for the similarity of professional interests, it will be not enough if it is regarded as the knowledge people share only because it is more like the issue interesting them behind the knowledge. Finding the topic they can work together is described by the variable, profession values. More details of professional values will be discussed later and it is a dynamic process, which means people might change the values because of their experiences. Conducting a series of interviews with scientists presented at CSSS Beijing, 2008 we wanted to identify other conditions that activate the collaboration mechanism. All in all, there are two processes in collaborations, picking a partner and picking a topic. These two processes should be interacted with each other during the process. Since we don't have enough time to give a detailed analysis of the two each, we want to do it later.

Our model simulates the formation of the collaboration network in two cases: one when agents randomly connect to each other and another when there is a super agent that can create weak ties, which means people know each other, between them. The purpose is to see the impact of SFI over the interdisciplinary scientific community. In other words SFI is represented by the super agent and all the other agents are researchers from different fields.

Each agent represents a person except for one agent, SFI, which implies the SFI institution not person, and has the following parameters: Personal Value (PEV), Professional Coverage Radius (PRV.cr) and Professional Position Center (PRV.pc).

Personal Value represents the personal preferences, characteristics and reflects the social/human dimension of the agent. The Professional Position Center could be interpreted as an agent's major. Professional Coverage Radius describes agent's professional knowledge and interests. The first two parameters' values are randomly generated when the simulation starts, but the last one, Professional Coverage Radius modifies his values based on how collaborative or not, the agent is.

Between the agents can be established two types of connections: weak-ties and strong-ties. The establishment of a weak-tie means two agents "have met" each other. A strong-tie can be developed only when a weak-tie already exists and represents a stronger connection, like the ones established when working at a paper together.

The initial conditions are: n agents with given PEV, PRV.cr, PRV.pc and a random network of weak ties between them.

For our first model we consider that the initial random network, all weak ties, could not be added by any agents easily and we observe how the strong-tie network is formed based on the following rules:

1) $\left|PEV(x_i) - PEV(x_J)\right| \leq 1$

2) $\left|PRV.pc(x_J) - PRV.pc(x_i)\right| \leq \left|PRV.cr(x_i) + PRV.cr(x_J)\right|$

3) is connected with by a weak tie ;

Also, every time an agent establish a strong-time his Professional Coverage Radius grows, which means that he becomes more open to other collaborations, that he is more sensitive to the benefits of interdisciplinary work.

Weak-ties are developed continuously. If an agent has both links with its two neighbor agents, there will be another weak link between these two agents with a given probability Probability-of-creating-weak-ties.

Our second model, which is our main model, contains two kinds of agents, person agents and SFI agents. The latter agent could change the initial network structure, which means that SFI could let person agents meet each other to develop their social tie by SFI agents constructing ties with person agents (in real world we can say by inviting them to SFI for instance). In the second model, we could produce N SFI agents and produce ties with two kinds of agents, randomly, which facilitates the creations of strong-ties. For the time limit, our model is still under adjustment so we do not have comparatively fixed values for parameters. But we will make a comparison between the data produced by our second ABM (with SFI agents) and the data from the empirical part.

From the simulations ran on the first model (without SFI agents) we observed that the agents are clustering so they cannot develop further links between clusters even the collaborations. From the second model (with SFI) we found that the properties of agents (personal value and professional value) play important roles in developing collaborative ties.

Empirical Model

To analysis the scientific collaboration networks in the Sante Fe Institute (SFI), we first construct it by using the data from the SFI website. Here we use the working papers rather than published papers written by authors when they were in the SFI, partly because the most of working papers can be downloaded directly from the SFI website and partly because compare to the working papers, the published papers have a time delay, half year or more. If two scientists coauthored one paper, we make a connection between them.

Table 1. Summary of statistic analysis of the collaboration networks in the Sante Fe Institute

YEARS	NUMBER OF AUTHORS	NUMBER OF PAPERS	AVERAGE COLLABORATORS	AVERAGE DISTANCE	CLUSTERING COEFFICIENT
1989-1989	11	8	1. 2727	1. 2222	0. 8889
1989-1992	159	143	2. 4025	3. 0115	0. 8681
1989-1995	343	390	2. 4140	3. 2925	0. 8280
1989-1998	591	695	2. 7479	5. 5024	0. 8318
1989-2001	800	930	2. 8825	5. 2026	0. 8320
1989-2004	963	1112	3. 0550	5. 3848	0. 8350
1989-2008	1118	1292	3. 0394	5. 0077	0. 8350

A summary statistic for the collaboration networks in the SFI is illustrated in Table 1 for different time windows. In this table, we focus on the total number of the papers appearing, the total number of authors, average collaborators, average distance as well as clustering coefficient. In the following, we will highlight some of these results and discuss their possible implications.

The table clearly shows that the total number of papers and authors increase as the time passed. This fact is quite easy to understand because the SFI does not allow research scientists to stay there for very long period and more researchers from other institutes begin to join the SFI to do their research works. We also note that the change of the number of new scientists joined to the SFI each year is not very large, which may indicate that the available positions, such as post doctor in each year in SFI is to some intent fixed.

The average collaborators of one author also increase from 1.3 in the data of 1989 to 3.1 in the data of 1989~2008. To get more information about this thing, we also calculated the distribution of the degree for the data obtained during the period window of 1989~2008, as an example shown in Fig. 2. From this figure, one may find the value of $P(k)$ is larger as $k<10$. It means that most of authors have collaborators less than 10. Interestingly, we also find that few of them have a large number of collaboratorsFor instance, there is an author who has 90 collaborators. We also interestingly note that the authors who have large number of collaborators are generally the faculty of the SFI or the external professor. These people play an important role the collaboration networks and we call them bignames. They do exist in the previously studied collaboration networks (Newman, 2004).

Fig. 3 is drawn from the data produced by second ABM (with SFI agents), which shows a very different tendency. Firstly, it could be caused by the flaw of the working paper data for it only has the record of authors' collaboration when they were at SFI or at least one of the authors. Probably, two persons could collaborate even after their stay at SFI, which still thanks to SFI. Secondly, it could also be caused by the unreasonable values set for at least one of the parameters.

Another interesting point is average distance. From the table, one sees that the average distance is quite small at the beginning (during the year of 1989) and it is about 6 as the collaboration networks becomes larger. At the beginning, the scientific community is quite small and people are easier to get know each other and to do collaboration works. When the time is passed, some of them have to leave and go to other places, and at the same time, new scientists begin to join SFI. In this case, people will become a little difficult to know. So the distance between them becomes large. However, the average distance is about six implies that the community of people in SFI may have a small word effect.

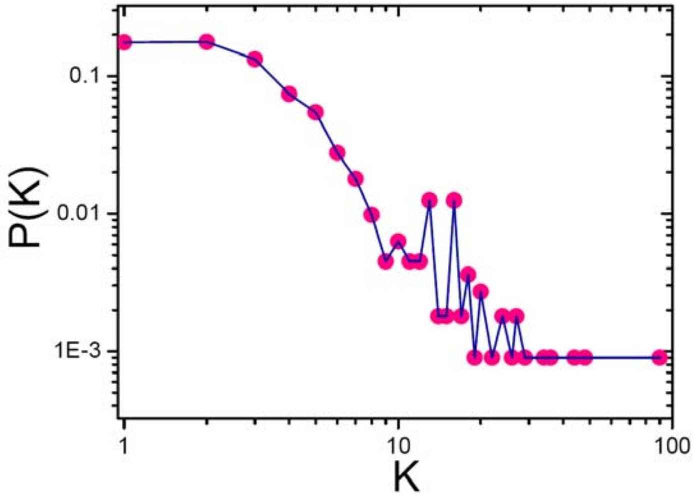

Fig. 2 The distribution of degree for the scientific collaboration networks of the Sante Fe Institute during the period 1989~2008.

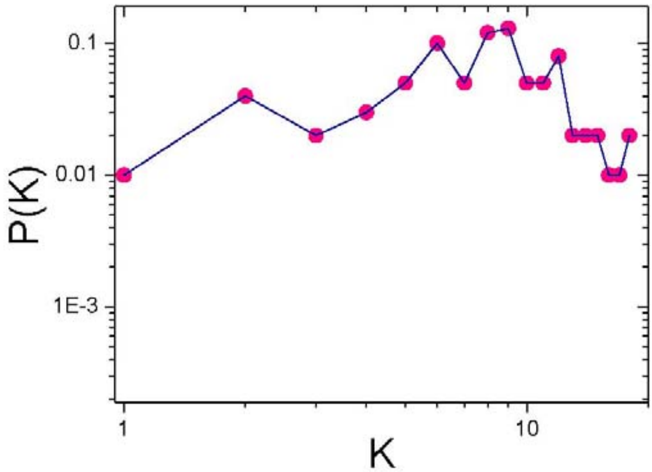

Fig. 3 The distribution of degree for the scientific collaboration networks of the second ABM.

We finally calculated the clustering coefficient of the studied network. This kind of clustering coefficient means that average fraction of pairs of a person's collaborators who have also collaborated with one another. Mathematically, we define the clustering coefficient as (Newman, 2001)

$$C=\frac{3\times number\ of\ triangles\ on\ the\ network}{number\ of\ connencted\ triples\ of\ vertices}$$

We find there is a very strong clustering coefficient in the scientific community in SFI. This means in SFI, two scholars have a great probability of collaborating if both have collaborated with a third scientist. One of possible explanations is that a lot of papers with three or more authors. This fact may be related to the free environment in SFI where scientists are encouraged to discuss with each other.

Further Thinking

The model we constructed takes two factors, power and resource, for granted: 1) no person agents enjoy more power than another; 2) SFI agents enjoy infinite resource. But it is quite impossible in the real world, which might be the task for future.

For further studies we are thinking of two sets of categories “SFI global network” and “SFI knowledge”. The former could be represented as various kinds of networks, i.e. alumni network, working paper network and invitation network (from workshops and conferences) while the latter could be the content through and outcome out of the conduits constructed by SFI global network. The links fostered by invitation network or alumni etc. could be seen as weak ties and the ones developed from working paper should be the strong ties.

References

David Pines, 1988, Emerging Syntheses in Science, Addison-Wesley Publishing Company Inc.

Newman, M. E. J., 2004, Proc. Natl. Acad. Sci. USA **101**, 5200.

Newman, M. E. J., 2001, Proc. Natl. Acad. Sci. USA **98**, 404.