RESEARCH

A game-theory modeling approach to utility and strength of interactions dynamics in biomedical research social networks

J. Mario Siqueiros-García[??,??*†], Rodrigo García-Herrera[??], Enrique Hernández-Lemus[??] and Sergio Alcalá-Corona[??]

[*]Correspondence:
jmario.siqueiros@iimas.unam.mx
[??]IIMAS, Universidad Nacional Autónoma de México, Circuito Escolar 3000, Cd. Universitaria, 04510 Mexico City, Mexico
Full list of author information is available at the end of the article
[†]Equal contributor

Abstract

Collaboration has become a cornerstone in biomedical research today. In contrast to physics which has a long history and experience in collaborative projects, biology is only recently becoming an evermore collaborative discipline. In this article we explore the effect of a collaboraion network on the distribution of players having access to certain amount of resources from other players in the network and the distribution of the strength of connections among them. Particularly, we implemented two games played simultaneously: one for maximazing individual utility based on the iterated Prisoner's Dilemma; the other, a coordination game for maximazing the connection strength between players. We are interested in how they affect each other in the context of a network of scientific collaboration under the idea that while researchers are interested in maximazing their utilities, they also know that it is important to invest in building collaborative relationships. We tested our simulation on a biomedical research community network of México. and compared the results with an Erdös-Renyí, a Watts-Strogatz small-world and Barabási-Albert topologies. Different topologies display different utility and attachment distributions. Moreover, the distribution of utility and attachment in the researchers network is similar to that of Barabási-Albert and Watts-Strogatz topologies, respectively. We believe that utility distribution in the researchers network suggests that there are socio-cultural mechanisms governing the network that produce an asymmetric distribution of resources. The high distribution of strong connections might reflect some sort of subordination among researchers by which they are morally obliged to cooperate by the same socio-cultural mechanisms. The range around the threshold that regulates the decision to cooperate or defect according to the agent's historical balance between utility and strength of collaborative relationships and carrying capacity of the system is small, suggesting that there is a region in which a phase transition takes place from a population of cooperators to a population of defectors. Simulations like this may help to develop science policies to promote fair distribution of resources.
Text for this section.

Keywords: Collaboration networks; Social Networks Analysis; Biomedicine; Game theory

Introduction

Collaboration has become a cornerstone in biomedical research today. In contrast to physics which has a long history and experience in collaborative projects, biology

is only recently becoming an evermore collaborative discipline[**?**]. Biology has an interesting record in such matters because scientific collaboration means something different to different branches of biology: molecular biology has traditionally been a research activity of small laboratories[**?**, **?**], whereas in natural history there has been data and samples exchange since the $XVII^{th}$ century[**?**, **?**]. Despite the differences in culture and practices, the Human Genome Project made collaboration a central feature of biology.

Nowadays it is widely acknowledged that collaboration takes many forms, from sharing of biological samples and biobanking to international groups in charge of helping research communities to harmonize and share their data. Sharing resources such as equipment, funds, and time is critical; building trust among scientists is fundamental. Also, resources are mobilized in order to create strategic alliances.

The analysis of cooperation in scientific research has been the subject of a number of studies [**?**, **?**, **?**, **?**, **?**, **?**, **?**]. This is not surprising since cooperation and competition are quite important in today's academic success. How does collaboration happen within a competitive academic environment and what kind of payoff is present in these settings were questions considered recently by Wardil and Hauert [**?**] in the context of cooperation in multi-authored publications. Also, the role of game theory over complex scientific information and collaboration networks has attracted attention, mainly focusing on how long-term strategies may shape different scenarios for Nash equilibria [**?**]. Prisoner's Dilemma has been used in the study of impact factor and collaboration [**?**, **?**].

Even with all these research efforts, cooperation in the context of scientific collaboration is still loosely defined and the long term dynamics of academic cooperation (and its consequences) are yet to be fully elucidated. Furthermore, to our current knowledge, there has been no use of game theory and complex network analysis for understanding how the topology of scientific collaboration networks affects access to resources among individuals present in the network [1]. Our work aims to contribute to our current understanding on the matter, specially when agents have to maximize their access to resources while taking care of their collaboration links.

In this article we explore the network effect on the distribution of players having access to certain amount of resources shared by other players in the network and the distribution of the strength of interactions among them. Particularly, we implemented two games played simultaneously: one for maximazing individual utility based on the iterated Prisoner's Dilemma; the other, a coordination game for maximazing the connection strength between players. We are interested in how they affect each other in the context of a network of scientific collaboration under the idea that while researchers are interested in maximazing their utilities, they also know that it is important to invest in building collaborative relationships. These two behaviors are explored in a biomedical research community of México.

[1] For an account of scientific collaboration and definitions, please refer to [?].

In the context of our paper, utility represents access to resources shared by others. The value of the Utility function for a player is the sum of the payoffs of playing with its neighbors. The opossing force comes from the other concurrent game: players trying to maximize the strength of their interactions with other players. In the coordinating game the best strategy is to adopt the same strategy as the other player, as it pays the most regardless of cooperation or defection in the utility game. When both cooperate the interaction gets a positive payoff, when both defect, the interaction doesn't get affected; but if they anti-coordinate, then the interaction looses. Finally, cooperation is a central feature of scientific work. For our biomedical network, cooperation can be thought of as sharing resources such as time, students, equipment, even money. Examples of defection to a cooperator are ghost authorship or prestige authorship.

The manuscript is structured in five sections. First we describe *FOSISS*, the main program for grants destined to biomedical applied research in México. This is the source of the database from which we created the researchers collaboration network. Next we describe our model and the different network topologies on which we explored it. We then present our results and discuss them. In the last section we draw some final remarks and conclusions.

Biomedical research: CONACyT and FOSISS

CONACyT (National Council of Science and Technology) is the Mexican government entity in charge of promoting the development of science and technology. Among CONACyT's functions are to develop science and technology policies according to national needs and demands, to advise the different instances of government on scientific and technological topics, to promote the creation of research networks among the scientific community, to grant scholarships for masters and doctoral studies, and to manage different trusts intended to fund individuals and groups for scientific and technological research.

In the year 2002 CONACyT, along with other government agencies and entities, created sectoral funds to cover and equally promote research capacities of different areas such as energy, agriculture and health. Technological innovation is fostered by the generation of human resources and by helping research groups to consolidate. It is expected that the knowledge generated under the sponsorship of these funds will be the product of applied research that attends national public needs, and promotes economic growth.

FOSISS or Sectoral Fund for Health and Social Security Research (*Fondo Sectorial en Investigación en Salud y Seguridad Social*) is one of such funds. FOSISS is constituted by CONACyT, SSA, IMSS and ISSSTE,[2] all of them being the major public health providers and research institutions in the country. Every year

[2]SSA is the acronym for Secretariat of Health *Secretaría de Salud*; IMSS is the acronym for Social Security Mexican Institute (*Instituto Mexicano del Seguro Social*); ISSSTE stands for Institute for Social Security and Services for State Workers (*Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado*)

CONACyT opens a call for funds limited to a set of health research areas previously defined by a group of experts. Such areas range from public health issues to chronic and degenerative diseases.

Most applicants are public universities and research institutions, but eligibility is open to public and private health research sectors. From 2002 through 2013, there were 91 institutions funded that comprised 4988 researchers.

From these data some important considerations should be made clear. Scientists in the database take on the roles of principal investigators (PIs), associate researchers, postdoctoral associates, postgraduate and undergraduate students. Unfortunately, information on these roles is not specified in the database. We acknowledge the importance of this deficiency because researchers in our network act under different circumstances and we know that this diversity has a real impact on the structure and eventually on the dynamics of the network, as well as on the results of our model.

Our database includes the name of the project, the year it was approved for funding and the research area to which it was assigned. It specifies the names of PIs or the people responsible for the project and the names of collaborators. Researchers can be PIs in one project and collaborators on a different project. The institutional affiliation of all participants is included. Through this affiliation we determine the principal institution behind every project.

Even though curation and analysis work of this database is still going on, some relevant facts about the biomedical research can be said. Over the period of 12 years, 32 general research areas have been defined, the three most funded research areas are *chronic and degenerative diseases*, *malignant neoplasms*, and *infectious and parasitic diseases*. The least funded area is *Ethics and medicine*. The area with the most researchers is *malignant neoplasms*. Other areas of relevance for México are *diseases related to poverty* and *Health and vulnerable groups*.

From the institutions that have participated in a protocol funded by FOSISS, less than one fifth have been responsible for a project and more than 95% of them are Mexican, public institutions. There is also an important presence of foreign institutions as collaborators, most of them from the United States, though institutions from the UK, France, Spain, Netherlands, Colombia and Cuba are also in the database.

Besides the characteristics of the population there are some other boundary conditions that play an important role on the network topology and dynamics, that motivated the development of our model. Biomedical research in México constitutes a vibrant community and collaboration is part of everyday work. However, México does not have public biobanks for research purposes (which are specially relevant for research in genomics, for example), there is no regulation on the access to biological samples such as tissue, cells, DNA, RNA, etc.[3] Something similar happens

[3]Regulation exists regarding researcher-subject relations based on legal and ethical grounds. Also, all projects need to be approved by the Ethics Committee and IRB.

with data. There have been some attempts to create open data repositories for biomedical research, but they have not been established yet. Regulation on these subjects is still missing. Finally, technologically advanced equipment such as high throughput sequencers are kept by institutions with the highest research profiles and sometimes PIs manage them in a self serving way.

From our ethnographic work to date, we have been able to see that biological samples, data and technology can become instruments for negotiating collaboration. For example, among people involved in research projects, there are researchers that do not have direct access to samples, simply because their parent institution does not offer clinical services. Many of them are non medical doctors but chemists, biologists, physicists, and mathematicians. There is another group of researchers that are placed on hospitals which is able to do research and have access to biological samples from their own patients. It seems that this group is the most privileged one, and the one with the least pressure to establish collaboration at whatever cost. Finally, there is one more group formed by those who work as clinicians at small hospitals with no research infrastructure whatsoever. This group may have an interest in research and the way for them to become part of a project and be listed as authors in scientific papers is by giving researchers who do not have access to biological samples access to patients.

Due to these differences in the access to resources, researchers in general are compelled to build strategic alliances through which samples, data, technology and authorship, among other assets, become part of a constant flow through the network. Social and political capital, as well as concentrations of resources become fundamental tools for establishing fruitful collaborations.

Methodology

Our model is based on the iterated version of the Prisoner's Dilemma (PD) and a coordination game instantiated on networks. Implementing games on networks is not new and it's an active area of research aimed to understand the evolution of cooperation in networks populated by selfish agents [**?**, **?**, **?**, **?**, **?**]. In many network models on which some of game theory games are simulated, agents' decision to cooperate or defect depend on a specific strategy, such as the well known *tit-for-tat* [**?**, **?**]. In some other cases, agents can modify the weight of the interactions with their neighbors [**?**]. From a different perspective others have explored the effect of different topologies on the emergence of cooperation [**?**, **?**]. In our model, an agent's decision to cooperate or defect depends on a balance between utilities and the current strength of its collaboration relationships. Such balance reflects the overall success or failure of its strategies. We study the behavior of the system under different topologies, including a real-world network.

In our model, agents are embedded in a network with varying number of neighbors. Following the traditional PD game, the strategy chosen by an agent and the strategy chosen by its neighbors will produce a pay-off. Pay-off follows the traditional PD rule: $T > R > P > S$. T is for temptation to defect. It is the highest

pay-off and it takes place when the player defects and the other cooperates. R is for reward for when both players cooperate. P is the punishment for when both players defect. And S is for suckers pay-off, the worst outcome that takes place when the player cooperates but its neighbor defects. Utility is a property of agents in which pay-off is accumulated.

PD utility pay-off matrix

	Cooperate	**Defect**
Cooperate	R, R	S, T
Defect	T, S	P, P

The strength of the interaction, represented by w, is a property of the link between two agents and gets updated according to an A_{ij} matrix of a coordination game. In the w matrix, the highest value goes to an edge when both agents cooperate, getting an R for reward, if one of them defects, the connection gets weaker getting P for the collaborative connection being punished. If both agents defect, the value w doesn't change, which means that agents didn't interact or that the interaction gets nullified N. In this game, the best action for any agent is to coordinate with its neighbor, either beacuse it wins or because it doesn't loose.

w **pay-off matrix**

	Cooperate	**Defect**
Cooperate	R	P
Defect	P	N

After each game, the agent adds-up utility (u), which is the sum of the pay-offs following the PD matrix. A pair of neighbors will add-up to the strength of their interaction (w_{ji}) as they coordinate or anti-coordinate, being w also cumulative. We measure global utility and connection strength for the whole network. Global utility U is the sum of all individual utilities and global strength of connections or W is the sum of every pair of agents' links w. The strength of interaction can be thought as some sort of "trust".

It should be noted that the same actions or behaviors work for both u and w. There are two reasons for this decision in the design of the model. The most general one is that we believe that in the real world, actions such as cooperating and defecting affect the strength of the interaction among people. The second one is that we think that selfishly maximizing access to resources and strengthening relationships are *opposing forces* acting on the same set of behaviors. The actions of an agent imply a trade-off in which defecting may increase its utility at the expense of its collaborative relationships. If collaborators have nourished their relationships, they might be strong enough to endure occasional defection. Cooperating may build up relationships but it can be expensive for the player.

Network initialization and agent state update

All networks are initialized equally. The number of nodes for every network is 4122, the same as in the FOSISS network. The same utility is given to every agent and all edges are asigned the same weight. In the case of the FOSISS network, edge weight is given by the number of collaborations among researchers, utility remains the same for all nodes as in the other networks.

The probability for an agent to cooperate or defect depends on a number (η) that referes to a historical balance between average utility and the average strength of the connections with its neigbors. This is so because we assume that whatever the result in utility or strength of connection, as long as one of them increases, the player will be confident in the strategy followed so far.

η is calculated as:

$$\eta_i = \frac{\langle f_i \rangle + \langle w_{ij} \rangle_j}{2}$$

For the agent to decide whether to cooperate or not, η is compared to a global threshold ν. If the agents' $\eta > \nu$, then the agent will cooperate, otherwise he will be suspicious and will defect. ν is a global parameter that establishes a threshold that an agents' η must cross in order to decide to cooperate. In this way, η can limit the size of the population of cooperators. Due to what the system and the game can offer to agents in terms of utilities and the strength of collaboration relationships, η represents the carrying capacity of the system for the population of cooperators.

Our simulation was tested on an Erdös-Renyí, a Watts-Strogatz small-world and Barabási-Albert topologies, as well as on the real biomedical research collaboration network. The simulation was run in a synchronous manner, in which all agents update their behavior simultaneously.

We ran two different experiments. In the first we simulated different values of carrying capacity ν. With this experiment we were able to see how the number of cooperators, utility, strength of connections among agents and the ratio of shifting state population would change in the range of the carrying capacity. The states of the agents were the same at initialization, for all values of the carrying capacity. Since the model is deterministic, it will return the same result if run under the same conditions.

The second experiment consisted in running the simulation under the same degree of carrying capacity ν but randomizing the initial states of the agents. This would show that the system converges to a global state. For every network, the simulation was run 100 times and results were averaged.

Implementation of the model in different topologies

We built three classical topologies for networks besides the FOSISS network, their parameters are shown in the following table.

Topology	m	$\langle k \rangle$	$\langle C \rangle$	$\langle l \rangle$
Erdös-Rényi	25591	12.4	0.003	3.6
Watts-Strogatz	206100	100	0.7	3.4
Barbási-Albert	183465	89	0.06	2.13
FOSISS	23391	11.39	0.87	5.49

Erdös-Renyí

Erdös-Renyí networks [**?**] (random networks) are constructed by randomly selecting a pair of N possible nodes and attaching them with an edge, given a probability p, as long as there is no edge between them. The result is a Poisson distribution for connectivity of nodes $P(k)$, where each node has a degree quite close to the average $\langle k \rangle$. Also for this type of network, average clustering coefficient $\langle C \rangle$ is small, actually it is equal to p (the probability of connecting two nodes) and the average shortest path length $\langle l \rangle = \frac{lnN}{ln\langle k \rangle}$.

Small-World

Watts-Strogatz networks [**?**] (small-world networks) are in a regime between a fully regular grid (lattice) and a random network (Erdös-Rényi). In order to build them, a node is chosen from a lattice (a ring) and the edge that connects it to nearest neighbor in a clockwise sense. With probability p, this edge is reconnected to a node chosen uniformly at random over the entire ring, with duplicate edges forbidden; otherwise the edge is left in place. This process is repeated by moving clockwise around the ring, considering each node in turn until one lap is completed. Next, the edges connect nodes to their second-nearest neighbors clockwise. As before, each of these edges is randomly rewired with probability p. This process continues, circulating around the ring and proceeding outward to more distant neighbors after each lap, until each edge in the original lattice has been considered once. The main characteristic of these networks is that the average shortest path length is small and grows as $log(N)$ ($\langle l \rangle \sim log(N)$). Also, the average clustering coefficient $\langle C \rangle$ remains large in terms of p. For $p < 0.1$, $\langle C \rangle \sim 1$.

Barbási-Albert

Barbási-Albert networks [**?**] (scale-free networks) are generated by adding new nodes to a network. Each new node is added connecting it to an existing node with a probability proportional to the degree k (connectivity) of each node (*preferential attachment*). The result is a a power law distribution for connectivity of nodes $P(k)$ where few nodes have many connections and the most have very few connections. Furthermore these networks are also small world networks, showing a quite small $\langle l \rangle$.

FOSISS: Biomedical research community network

The biomedical research network on which we are running our model was generated with data from collaborative projects. Our data was obtained from CONACyT and includes information for twelve years of *FOSISS* grants. Data included names of Principal Investigators, collaborators, research topics, etc. The network we are using here has researchers as nodes and edges represent the connection of two scientists

when they colaborate on the same project. Edges are also weighted according to the number of projects shared by any pair of scientists.

Content

Text and results for this section, as per the individual journal's instructions for authors.

Section title

Text for this section ...

Sub-heading for section

Text for this sub-heading ...

Sub-sub heading for section

Text for this sub-sub-heading ...

Sub-sub-sub heading for section Text for this sub-sub-sub-heading ... In this section we examine the growth rate of the mean of Z_0, Z_1 and Z_2. In addition, we examine a common modeling assumption and note the importance of considering the tails of the extinction time T_x in studies of escape dynamics. We will first consider the expected resistant population at vT_x for some $v > 0$, (and temporarily assume $\alpha = 0$)

$$E\big[Z_1(vT_x)\big] = E\bigg[\mu T_x \int_0^{v\wedge 1} Z_0(uT_x)\exp\big(\lambda_1 T_x(v-u)\big)\,du\bigg].$$

If we assume that sensitive cells follow a deterministic decay $Z_0(t) = xe^{\lambda_0 t}$ and approximate their extinction time as $T_x \approx -\frac{1}{\lambda_0}\log x$, then we can heuristically estimate the expected value as

$$\begin{aligned} E\big[Z_1(vT_x)\big] &= \frac{\mu}{r}\log x \int_0^{v\wedge 1} x^{1-u}x^{(\lambda_1/r)(v-u)}\,du \\ &= \frac{\mu}{r}x^{1-\lambda_1/\lambda_0 v}\log x \int_0^{v\wedge 1} x^{-u(1+\lambda_1/r)}\,du \\ &= \frac{\mu}{\lambda_1-\lambda_0}x^{1+\lambda_1/rv}\left(1-\exp\left[-(v\wedge 1)\left(1+\frac{\lambda_1}{r}\right)\log x\right]\right). \quad (1) \end{aligned}$$

Thus we observe that this expected value is finite for all $v > 0$ (also see [**?**, **?**, **?**, **?**, **?**]).

Competing interests
The authors declare that they have no competing interests.

Author's contributions
Text for this section ...

Acknowledgements
Text for this section ...

Figure 1 Sample figure title. A short description of the figure content should go here.

Figure 2 Sample figure title. Figure legend text.

Table 1 Sample table title. This is where the description of the table should go.

	B1	B2	B3
A1	0.1	0.2	0.3
A2	...	..	.
A3	..	.	.

Figures
Tables
Additional Files
Additional file 1 — Sample additional file title
Additional file descriptions text (including details of how to view the file, if it is in a non-standard format or the file extension). This might refer to a multi-page table or a figure.

Additional file 2 — Sample additional file title
Additional file descriptions text.