

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

4 J. Mario Siqueiros-García      Rodrigo García-Herrera  
5 Enrique Hernández-Lemus      Sergio A. Alcalá-Corona

6 November 9, 2016

7 **Abstract**

8        What would happen if researchers in a community were given the  
9 same amount of resources and then set free to interact among each  
10 other? How would these resources be distributed in the population  
11 after many interactions? The final result may be determined by the  
12 researchers' social and political power, an by their prestige. Neverthe-  
13 less, all these attributes depend, to a certain degree, on the research  
14 collaboration social network in which scientists are embedded. Here  
15 we develop a model based on the Prisoner's Dilemma implemented on  
16 a collaboration network. With this model we explore the distribution  
17 of resources and the dynamics of the strengthening or weakening of  
18 collaboration connections among partners, as scientists cooperate or  
19 defect in order to gain access to others' resources. The network is  
20 based on data about projects that were awarded with a grant. In this  
21 sense, we assume that resources are funds, equipment and time. We  
22 define Fitness as how much access a researcher has to the resources  
23 of his colleagues. We tested our simulation on the real biomedical  
24 research network and compared the results with an Erdős-Renyí, a  
25 Watts-Strogatz small-world and Barabási-Albert topologies. Different  
26 topologies display different fitness and connections strength distribu-  
27 tions. Moreover, the distribution of fitness and connections strength  
28 in the researchers network is similar to that of Barabási-Albert and  
29 Watts-Strogatz topologies, respectively. We believe that fitness distri-  
30 bution in the researchers network suggests that there are socio-cultural  
31 mechanisms governing the network that produce an asymmetric distri-  
32 bution of resources. The high distribution of strong connections might  
33 reflect some sort of subordination among researchers by which they are  
34 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.

## 1 Introduction

### Notes:

- We change utility for fitness. Utility seems a better concept in order to reflect players' preferences.
- Choosing a strategy depends on the balance  $\eta'$  equation –look at page 9. This equation states that the probability of an agent to cooperate increases as: a) its utility increases –as it moves away from 0, or; b) The strength of the connection with its peers increases –as it moves away from 0, or; c) both increases.

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

- En la pagina 9 decimos que “In order for the agent to choose to cooperate or defect, a degree of willingness to cooperate is assigned to each agent.” No estoy seguro que “degree of willingness to cooperate” sea la mejor manera de decirlo, mi duda est en la palabra “degree”, por lo que cambié la oración a: The probability for an agent to cooperate or defect depends on a number that referes to a historical balance between average utility and the average strength of the connections with its neighbors.

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[1]. 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[2, 3], whereas in natural history there has been data and samples exchange since the *XVII<sup>th</sup>* century[4, 5]. Despite the differences in culture and practices, the Human Genome Project made collaboration a

70 central feature of biology.

71

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

78

79 The analysis of cooperation in scientific research has been the subject of  
80 a number of studies [1, 7, 8, 9, 10, 3, 5]. This is not surprising since coopera-  
81 tion and competition are quite important in today’s academic success. How  
82 does collaboration happen within a competitive academic environment and  
83 what kind of payoff is present in these settings were questions considered  
84 recently by Wardil and Hauert [20] in the context of cooperation in multi-  
85 authored publications. Also, the role of game theory over complex scientific  
86 information and collaboration networks has attracted attention, mainly fo-  
87 cusing on how long-term strategies may shape different scenarios for Nash  
88 equilibria [21]. Prisoner’s Dilemma has been used in the study of impact  
89 factor and collaboration [13, 14].

90

91 Even with all these research efforts, cooperation in the context of sci-  
92 entific collaboration is still loosely defined and the long term dynamics of  
93 academic cooperation (and its consequences) are yet to be fully elucidated.  
94 Furthermore, to our current knowledge, there has been no use of game the-  
95 ory and complex network analysis for understanding how the topology of  
96 scientific collaboration networks affects access to resources among individ-  
97 uals present in the network <sup>1</sup>. Our work aims to contribute to our current  
98 understanding on the matter, specially when agents have to maximize their  
99 access to resources while taking care of their collaboration links.

100

101 In this article we explore the network effect on the distribution of play-  
102 ers having access to certain amount of resources from other players in the  
103 network and the distribution of the strength of connections among them.  
104 Particularly, we implemented two games played simultaneously: one for  
105 maximizing individual utility based on the iterated Prisoner’s Dilemma;  
106 the other, a coordination game for maximizing the connection strength be-  
107 tween players. We are interested in how they affect each other in the context

---

<sup>1</sup>For an account of scientific collaboration and definitions, please refer to [6].

108 of a network of scientific collaboration under the idea that while researchers  
109 are interested in maximizing their utilities, they also know that it is impor-  
110 tant to invest in building collaborative relationships. These two behaviors  
111 are explored in a biomedical research community of México.

112

113 In the context of our paper, utility represents the preference for having  
114 access to the resources of others. Maximizing utility means that the player  
115 is increasing its access to the resources of other players. The value of the  
116 Utility function for a player is the sum of the payoffs of playing with his  
117 neighbors. The opposing force comes from the other concurrent game: play-  
118 ers trying to maximize the strength of their connections to other players.  
119 In the coordinating game the best strategy is to adopt the same strategy  
120 as the other player, as it pays the most regardless of cooperation or de-  
121 fection in the utility game. When both cooperate the interaction gets a  
122 positive payoff, when both defect, the interaction doesn't get affected; but  
123 if they anti-coordinate, then the interaction loses. Finally, cooperation is a  
124 central feature of scientific work. For our biomedical network, cooperation  
125 can be thought of as sharing resources such as time, students, equipment,  
126 even money. Examples of defection to a cooperator are ghost authorship or  
127 prestige authorship.

128 The manuscript is structured in five sections. First we describe *FOS-*  
129 *ISS*, the main program for grants destined to biomedical applied research  
130 in México. This is the source of the database from which we created the  
131 researchers collaboration network. Next we describe our model and the dif-  
132 ferent network topologies on which we explored it. We then present our  
133 results and discuss them briefly. In the last section we draw some final  
134 remarks and conclusions.

## 135 2 Biomedical research: CONACyT and FOSISS

136 CONACyT (National Council of Science and Technology) is the Mexican  
137 government entity in charge of promoting the development of science and  
138 technology. Among CONACyT's functions are to develop science and tech-  
139 nology policies according to national needs and demands, to advise the dif-  
140 ferent instances of government on scientific and technological topics, to pro-  
141 mote the creation of research networks among the scientific community, to  
142 grant scholarships for masters and doctoral studies, and to manage different  
143 trusts intended to fund individuals and groups for scientific and technologi-  
144 cal research.

145

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

153

154 *FOSISS* or Sectoral Fund for Health and Social Security Research (*Fondo*  
 155 *Sectorial en Investigación en Salud y Seguridad Social*) is one of such funds.  
 156 FOSISS is constituted by CONACyT, SSA, IMSS and ISSSTE,<sup>2</sup> being all  
 157 of them the major public health providers and research institutions in the  
 158 country. Every year CONACyT opens a call for funds limited to a set of  
 159 health research areas previously defined by a group of experts. Such areas  
 160 range from public health issues to chronic and degenerative diseases.

161

162 Eligibility is open to public and private health research sectors, how-  
 163 ever most applicants are public universities and research institutions. From  
 164 2002 through 2013, there were 91 institutions funded that comprised 4988  
 165 researchers.

166

167 From these data some important considerations should be made clear.  
 168 The population represented in the data include principal investigators (PIs),  
 169 associate researchers, postdoctoral associates, postgraduate and undergrad-  
 170 uate students. Unfortunately, this information is not specified in the database.  
 171 Still, this is something we acknowledge and it's important because researchers  
 172 in our network are under different circumstances and we know that this di-  
 173 versity has a real impact on the structure and eventually on the dynamics  
 174 of the network, as well as on the results of our model.

175

176 Our database includes the name of the project, the year it was approved  
 177 for funding and the research area to which it was assigned. It specifies  
 178 the names of PIs or the people responsible for the project and the names  
 179 of collaborators. Researchers can be PIs in one project and collaborators  
 180 on a different project. The institutional affiliation of all participants is in-

---

<sup>2</sup>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*)

181 cluded. Through this affiliation we know the principal institution behind  
182 every project.

183  
184 Even though curation and analysis work of the database is still ongoing,  
185 some relevant facts about the biomedical research can be said. Over the  
186 period of 12 years, 32 general research areas have been defined, the three  
187 most funded research areas are *chronic and degenerative diseases*, *malignant*  
188 *neoplasms*, and *infectious and parasitic diseases*. The least funded area is  
189 *Ethics and medicine*. The area with the most researchers is *malignant neo-*  
190 *plasms*. Other areas of relevance for México are *diseases related to poverty*  
191 and *Health and vulnerable groups*.

192  
193 From the institutions that have participated in a protocol funded by  
194 FOSISS, less than one fifth have been responsible for a project and more  
195 than 95% of them are Mexican, public institutions. There is also an impor-  
196 tant presence of foreign institutions as collaborators, most of them from the  
197 United States, though institutions from the UK, France, Spain, Netherlands,  
198 Colombia and Cuba are also in the database.

199  
200 Besides the characteristics of the population there are some other bound-  
201 ary conditions that play an important role on the network topology and dy-  
202 namics, that motivated the development of our model. Biomedical research  
203 in México constitutes a vibrant community and collaboration is part of ev-  
204 eryday work. However, México does not have public biobanks for research  
205 purposes (which are specially relevant for research in genomics, for exam-  
206 ple), there is no regulation on the access to biological samples such as tissue,  
207 cells, DNA, RNA, etc.<sup>3</sup> Something similar happens with data. There have  
208 been some attempts to create open data repositories for biomedical research,  
209 but they have not been established yet. Regulation on these subjects is still  
210 missing. Finally, technology (e.g., PCR, sequencing and expression profiling  
211 technology) is in the domain of the institutions with the highest research  
212 profiles and sometimes PIs see technology as a personal good.

213  
214 From our ethnographic work to date, we have been able to see that bio-  
215 logical samples, data and technology can become instruments for negotiat-  
216 ing collaboration. For example, among people involved in research projects,  
217 there are researchers that do not have direct access to samples, simply be-

---

<sup>3</sup>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.

218 cause their institution does not offer clinical services. Many of them are non  
219 medical doctors but chemists, biologists, physicists, and mathematicians.  
220 There is another group of researchers that are placed on hospitals who are  
221 able to do research and have access to biological samples from their own  
222 patients. It seems that this group is the most privileged one, and the least  
223 pressured to establish collaboration at whatever cost. Finally, there is one  
224 more group formed by those who work as clinicians at small hospitals with  
225 no research infrastructure whatsoever. This group may have an interest in  
226 research and the way for them to become part of a project and be listed as  
227 authors in scientific papers is by giving researchers who do not have access  
228 to biological samples access to patients.

229  
230 Due to these differences in the access to resources, researchers in general  
231 are compelled to build strategic alliances through which samples, data, tech-  
232 nology and authorship, among other assets, become part of a constant flow  
233 through the network. Social and political capital, as well as concentrations of  
234 resources become fundamental tools for establishing fruitful collaborations.

### 235 3 Methodology

236 Our model is based on the iterated version of the Prisoner’s Dilemma (PD)  
237 and a coordination game instantiated on networks. Implementing games  
238 on networks is not new and it’s an active area of research aimed to under-  
239 stand the evolution of cooperation in networks populated by selfish agents  
240 [22, 23, 24, 28, 27]. In many network models on which some of game theory  
241 games are simulated, agents’ decision to cooperate or defect depend on a  
242 specific strategy, such as the well known *tit-for-tat* [25, 26]. In some other  
243 cases, agents can modify the weight of the interactions with their neighbors  
244 [27]. From a different perspective others have explored the effect of different  
245 topologies on the emergence of cooperation [28, 29]. **In our case, an agent’s  
246 decision to cooperate or defect is an outcome that depends on a balance  
247 between utilities and the current strength of its collaboration relationships.  
248 Such balance reflects the overall success or failure of its strategies.** We study  
249 the behavior of the system under different topologies, including a real-world  
250 network.

251  
252 In our model, agents are embedded in a network with varying number  
253 of neighbors. Following the traditional PD game, the strategy chosen by an  
254 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 connection, 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 because it wins or it doesn't lose.

#### $w$ 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 connection ( $w_{ji}$ ) 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  $w$ . Metaphorically, the strength of connection 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 connection among people. The second one is that we think that selfishly maximizing resources and strenghting 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.

### 3.1 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 assigned 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 ( $\eta$ ) that referes to a historical balance between average utility and the average strength of the connections with its neighbors. 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.

$\eta$  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,  $\eta$  is compared to a global threshold  $\nu$ . If the agents'  $\eta > \nu$ , then the agent will cooperate, otherwise he will be suspicious and will defect.  $\nu$  is a global parameter that establishes a threshold that an agents'  $\eta$  must cross in order to decide to cooperate. In this way,  $\eta$  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,  $\eta$  represents the carrying capacity of the system for the population of cooperators.

Our simulation was tested on an Erdős-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  $\nu$ . 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. For this experiment 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  $\nu$  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.

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

### 4.1 Erdős-Renyí

Erdős-Renyí networks [30] (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  $\langle l \rangle = \frac{\ln N}{\ln \langle k \rangle}$ .

## 361 4.2 Small-World

362 Watts-Strogatz networks [31] (small-world networks) are in a regime between  
363 a fully regular grid (lattice) and a random network (Erdős-Rényi). In order  
364 to build them, a node is chosen from a lattice (a ring) and the edge that  
365 connects it to nearest neighbor in a clockwise sense. With probability  $p$ , this  
366 edge is reconnected to a node chosen uniformly at random over the entire  
367 ring, with duplicate edges forbidden; otherwise the edge is left in place. This  
368 process is repeated by moving clockwise around the ring, considering each  
369 node in turn until one lap is completed. Next, the edges connect nodes to  
370 their second-nearest neighbors clockwise. And as before, each of these edges  
371 is randomly rewired with probability  $p$ , and continue this process, circulating  
372 around the ring and proceeding outward to more distant neighbors after each  
373 lap, until, each edge in the original lattice has been considered once. The  
374 main characteristic of these networks is that the average shortest path length  
375 is small and grows as  $\log(N)$  ( $\langle l \rangle \sim \log(N)$ ). Also, the average clustering  
376 coefficient  $\langle C \rangle$  remains large in terms of  $p$ . For  $p < 0.1$ ,  $\langle C \rangle \sim 1$ .

## 377 4.3 Barabási-Albert

378 Barabási-Albert networks [33] (scale-free networks) are generated by adding  
379 new nodes to a network. Each new node is added connecting it to an existing  
380 node with a probability proportional to the degree  $k$  (connectivity) of each  
381 node (*preferential attachment*). The result is a a power law distribution for  
382 connectivity of nodes  $P(k)$  where few nodes have many connections and the  
383 most have very few connections. Furthermore these networks are also small  
384 world networks, showing a quite small  $\langle l \rangle$ .

## 385 4.4 FOSISS: Biomedical research community network

386 The biomedical research network on which we are running our model was  
387 generated with data from collaborative projects. Our data was obtained  
388 from CONACyT and includes information for twelve years of *FOSISS* grants.  
389 Data included names of Principal Investigators, collaborators, research top-  
390 ics, etc. The network we are using here has researchers as nodes and edges  
391 represent the connection of two scientists when they colaborate in the same  
392 project. Edges are also weighted according to the number of projects shared  
393 by any pair of scientists.

394

## 395 5 Results

396 In this section we present the main results of the study, namely the topolog-  
397 ical structure of the underlying network models, the dynamics of the games  
398 under different parameters and network topologies and the distribution of  
399 utility and of the strength of interaction resulting of playing the games in  
400 all the different scenarios considered, including the real FOSISS network.

401

402 FOSISS network summed-up a total of 145 components or subnetworks,  
403 but we ran the model on the giant component made-up of 4122 researchers,  
404 and 23391 edges. The giant component was analyzed using *Cytoscape* **Fig-**  
405 **ure 1.** Results show that it is a well integrated network, with a clustering  
406 coefficient  $\langle C \rangle = 0.870$ , an average shortest path length of  $\langle l \rangle = 5.493$  and  
407 a density of  $p = 0.003$ . Such properties recall a small-world topology [31],  
408 and a great deal of self-organization when compared to a random network  
409 with the same density and number of nodes. Network centralization is 0.023,  
410 since there are no visible researchers that play as hubs in the network. Nev-  
411 ertheless the network heterogeneity is 0.873, which means that the network  
412 is highly hierarchical. When the degree distribution is analyzed, degree  
413 decreases as a power-law with an exponent of 1.7, similar to other social  
414 networks described as scale-free topology networks [33]. Finally, the aver-  
415 age number of neighbors of each node is 11.39 [34].

416

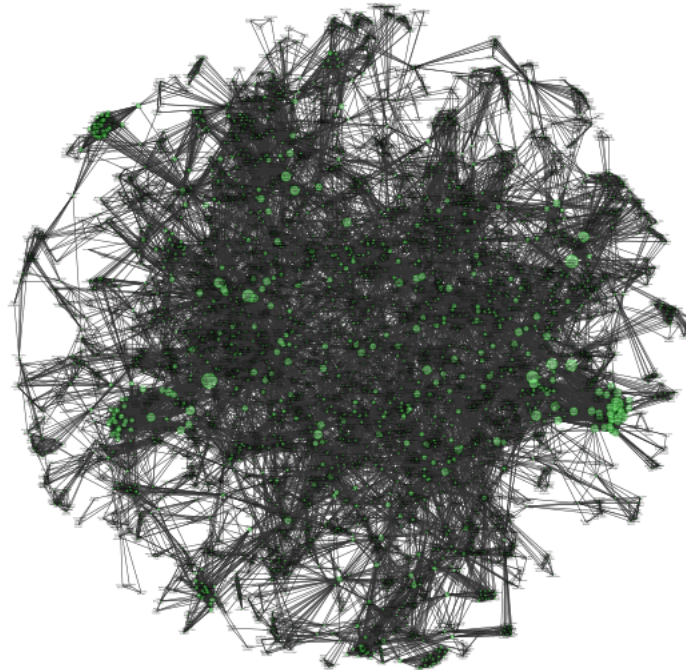


Figure 1: Bioemdicall research collaboration network (FOSISS) giant component.

Other results to report are those of the dynamics of different variables as the carrying capacity  $\nu$  changes. The most salient result is that for  $\nu$  between 0.19 and 0.24, there is an apparent phase transition in all different topologies and for all the different variables. Nevertheless, it is worth noticing that the shape of the phase transition is different according to the topology of the network at stake. When  $\nu$  is between 0.0 and 0.2, that is, when there is no space or a very short one for suspiciousness, all agents cooperate, when *carrying capacity* is above 0.25, all agents defect. *Utility*, *strength of interactions*, and *changing state population* replicate that same behavior for the same limits.

In **Figure 2**, we present how the number of cooperators in the population change as  $\nu$  changes. In the Erdős-Renyí, network, between 0 and 0.18 approximately, all agents converge to a cooperative behavior, from 18 to 20, convergence to cooperative state takes longer but eventually all agents are cooperating. Close to  $\nu \approx 0.21$  there is a sharp fall to a point in which

433 around half the population is cooperating and the rest is defecting. Reach-  
434 ing  $\nu \approx 0.25$  there is another sharp fall of cooperators and all agents turn  
435 into a defecting state.

436

437 In the case of the Watts-Strogatz, small-world network, the whole popu-  
438 lation remains cooperating for ranges between 0 and 0.2 but as it gets closer  
439 to 0.2, more time is needed for the population to become full of coopera-  
440 tors. In  $\nu \approx 0.2$  the cooperators will represent only half of the population  
441 and such number of them will be constant up to  $\nu \approx 0.25$  forming a short  
442 plateau. From  $\nu \approx 0.25$  to  $\nu \approx 0.6$  cooperators will be present at the begin-  
443 ning of the simulation but will go diminishing as time goes on. In the case  
444 of the Barabási-Albert network, crossing the threshold of  $\nu \approx 0.2$ , there is a  
445 sharp decrease in the number of cooperators, but stays constant over time.  
446 Such behavior is present for a very short range of  $\nu$ , and before  $\nu \approx 0.24$ ,  
447 cooperators disappear for the rest of values of  $\nu$ . Finally, FOSISS network  
448 behaves similar to the other networks in that there is a fall in the number of  
449 cooperators close  $\nu \approx 0.2$ . Different to the other networks, FOSISS network  
450 lacks the sharp reduction of cooperators, instead this population declines  
451 smoothly and progressively; specially, when it reaches a  $\nu \approx 0.25$  coopera-  
452 tors go decreasing in a less dramatic manner all the way to  $\nu \approx 0.5$ . It is also  
453 noteworthy that from  $\nu = 0$  to  $\nu \approx 0.5$  the number of cooperators converge  
454 to a certain amount and stays constant for the rest of the simulation.

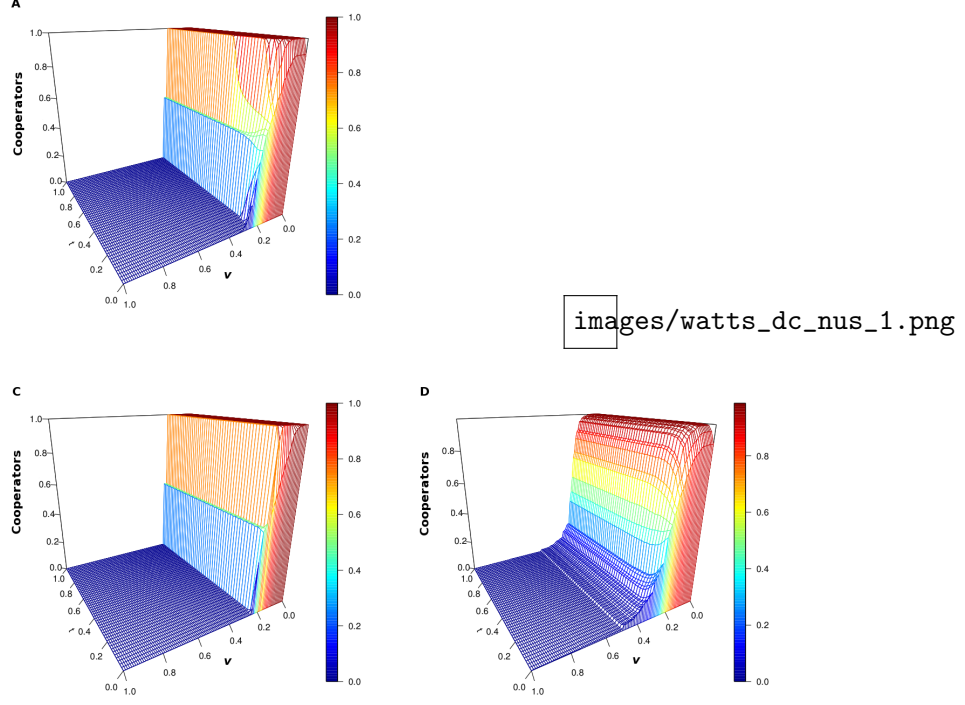
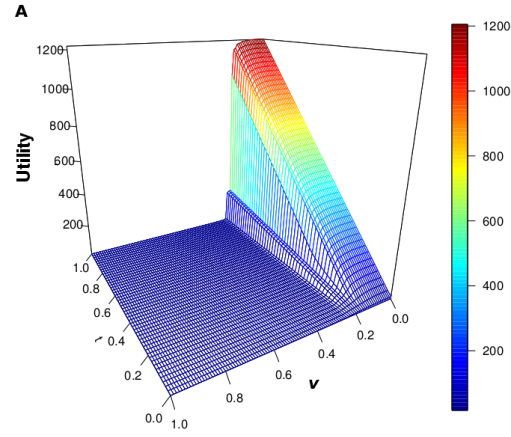


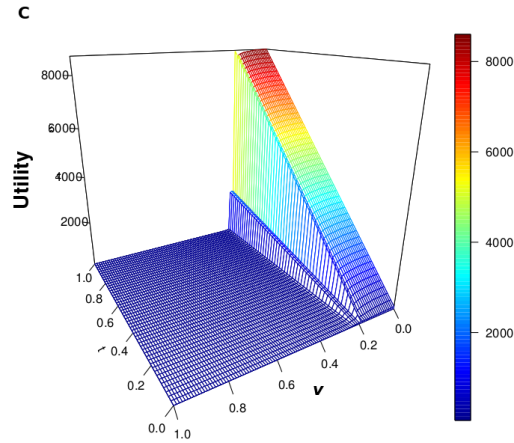
Figure 2: Proportion of cooperators as a function of carrying capacity and time.

455 Utility and strength of interactions dynamics under different  $\nu$  emulate  
 456 the phase transition found before. In **Figures 3 and 4**, it can be observed  
 457 that there is a drop in utility and strength of interactions according to  
 458 the drop in the number of cooperators. Erdős-Renyí and Barabási-Albert  
 459 networks are quite similar in the way these variables fall in two steps, the  
 460 first one at  $\nu \approx 0.2$  and the next one at  $\nu \approx 0.23$ . The fall is sharper  
 461 still in the Barabási-Albert topology. Utility and strength of interactions  
 462 phase transition in Watts-Strogatz network is significantly more staggered  
 463 compared to the former networks. In the case of utility, there is a region  
 464 in the limits of  $\nu \approx 0.25$  and  $\nu \approx 0.3$ , before utility goes to 0, in which it  
 465 remains low but stable over time. In general, strength of interactions follows  
 466 the same pattern as utility but in the same  $\nu \approx 0.25$  and  $\nu \approx 0.3$ , strength of  
 467 interactions grows to a value that is higher than the one given by default but  
 468 soon starts to decrease as the simulation runs. For FOSISS network, utility

469 and strength of interactions fall is quite steep but smooth, without sharp  
 470 cuts. Between  $\nu \approx 0.23$  and  $\nu \approx 0.28$ , utility and strength of interactions  
 471 start at their lowest, but there is a slight increase in both of them.



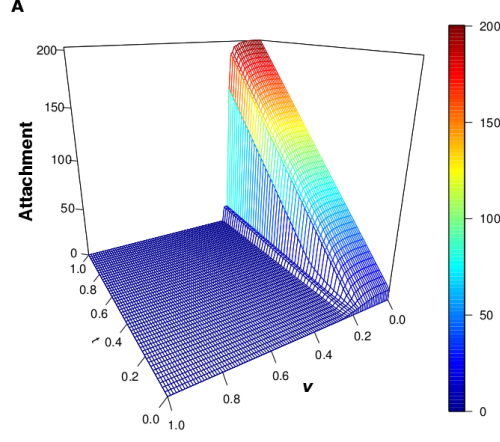
watts\_fitness\_nus\_1.png



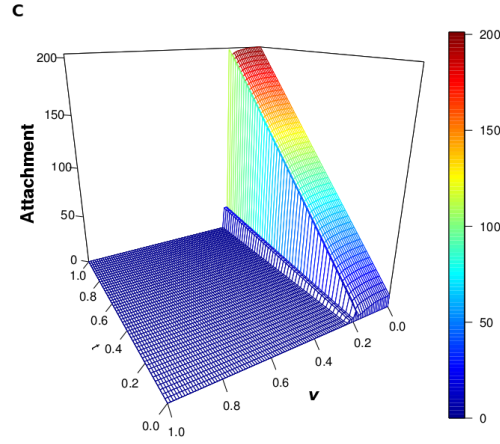
fosiss\_fitness\_nus\_1.png

Figure 3: Utility  $U$  dynamics as a function of carrying capacity  $\nu$  and time.





watts\_trust\_nus\_1.png



fossiss\_trust\_nus\_1.png

Figure 4: Strength of interactions  $w$  dynamics as a function of carrying capacity  $\nu$  and time.

472 We also measured the number of agents shifting states –between cooper-  
 473 ating and defecting- under different  $\nu$  values. We found that for all networks  
 474 there is a critical point around  $\nu \approx 0.2$  in which all agents are shifting states.  
 475 For Erdős-Renyí and Barabási-Albert networks, for this region, agents never  
 476 settle to a single state. Contrary to the former cases, the number of shifting  
 477 agents decreases considerably for the Watts-Strogatz and FOSISS networks,

478 and find an equilibrium state. Once the the limits of this region are crossed  
 479 as  $\nu$  increases, the number of agents shifting states falls to 0 and all nodes  
 480 become defectors.

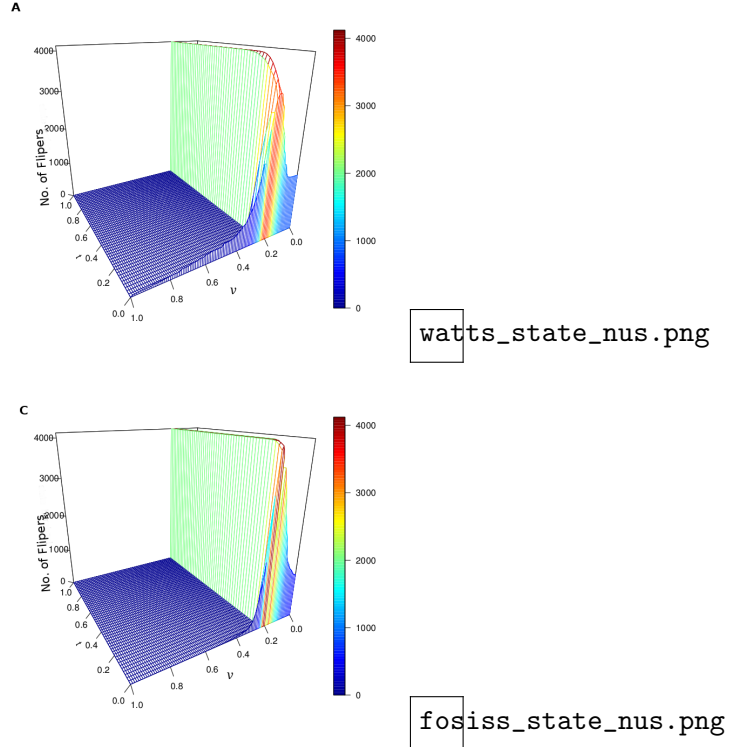


Figure 5: Shifting population between cooperators and defectors as a function of carrying capacity  $\nu$  and time.

481 Central to our argument is the differences in utility and strength of in-  
 482 teractions distribution at the end of the simulation, for every topology. We  
 483 found that utility distribution for the *FOSISS* network, resembles quite ac-  
 484 curately to the distribution of utility in the Barabási-Albert network.

485  
 486 The distribution of utility on each topology is induced by the degree  
 487 distribution. This is so, since a given agent (node) will interact with its  
 488 neighbors to either cooperate or defect, in such a way that connectivity in-  
 489 fluences the number of events played and thus the likelihood of increasing  
 490 its corresponding utility. For instance, utility distribution in the random,  
 491 Erdős-Rényi network displays a normal-like curve. The algorithm that gen-

492 erates this kind of topology, assigns to every node the same probability of  
 493 connecting with any other node, which produces a *poissonian* degree dis-  
 494 tribution [30]. Since the Watts-Strogatz degree distribution is described by  
 495 a function that is midway between a random distribution and a scale-free  
 496 network [32] one may expect also an intermediate behavior of the utility  
 497 distribution. This assumption seems to be fulfilled by the distribution in  
 498 Figure 6B.

499 The resemblance of the degree-distribution and utility distribution also  
 500 holds for the Barabási-Albert network. As mentioned in the methods sec-  
 501 tion, the degree distribution of a Barabási-Albert topology follows a power-  
 502 law that describes the fact that there are a small number of nodes with  
 503 large  $k$  and most nodes have a small  $k$  [33]. As it is shown in the following  
 504 figure, most utility is concentrated in a few number of agents, while most  
 505 agents have a small amount of it. This is consistent with other research in  
 506 which concentration of resources, fame or citations in science decreases as a  
 507 power-law [37, 38, 39]. FOSISS network utility distribution is also skewed  
 508 to the left, similar to that of the Barabási-Albert network.

509

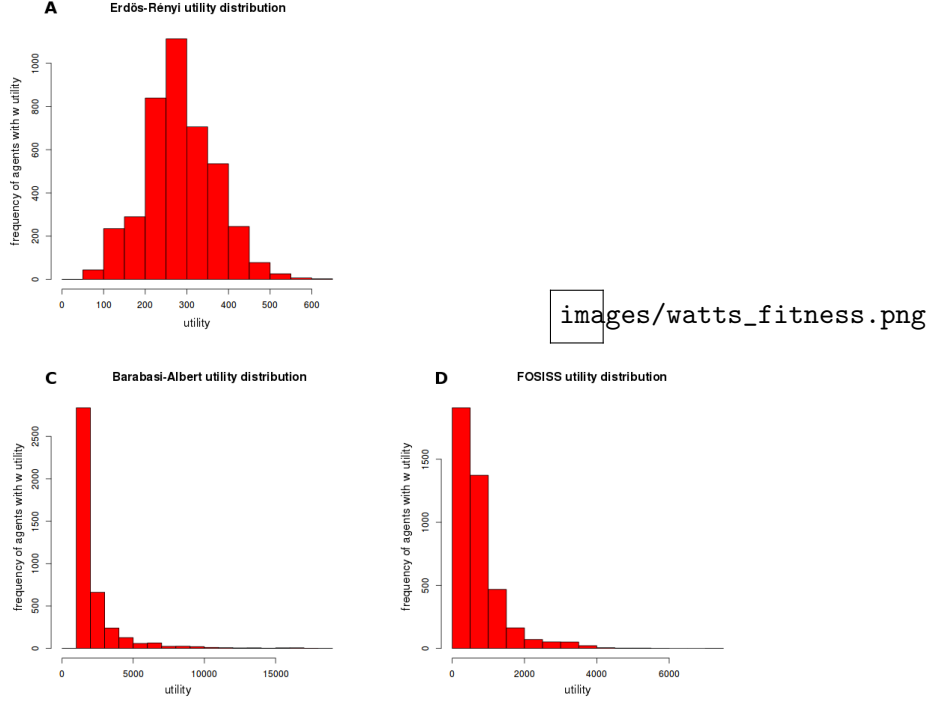


Figure 6: Utility distribution on different topologies. A. Random, Erdős-Rényi network displays a normal-like distribution. B. Watts-Strogatz network utility distribution is highly skewed to the right. C. Barabási-Albert network. Utility is distributed highly skewed to the left. D. FOSISS, biomedical researchers collaboration network distribution of utility resembles to Barabási-Albert network.

Regarding strength of interactions distribution, the Erdős-Rényi random network displays a normal distribution of strength of interactions, as expected. Again strength of interactions values are highly influenced by the corresponding degree distribution (Figure 7A). However, in the Watts-Strogatz network topology (Figure 7B), the strength of interactions distribution is a highly asymmetric bimodal, with a really-low frequency mode of low strength of interactions and a highly probability mode for high strength of interactions. A possible explanation for this phenomenon is that under network topologies maximizing inter-node communication (by minimizing the average distance between nodes) such as the Watts-Strogatz, strength of interactions is favored both among the cooperators (constituting the ma-

521 jority of players) and the defectors.

522

523 The Barabási-Albert network (7C) presents also a symmetric unimodal  
 524 distribution with values higher (on average) than those of the Erdős-Rényi  
 525 random network, this may be the effect of increased communication due to  
 526 more efficient network navigability. Interestingly, the network corresponding  
 527 to the real FOSISS collaborations (7D) is an asymmetric unimodal distribu-  
 528 tion in which moderate to high values of strength of interactions are more  
 529 likely. We hypothesize that this effect is also due to the communication  
 530 properties of the network.

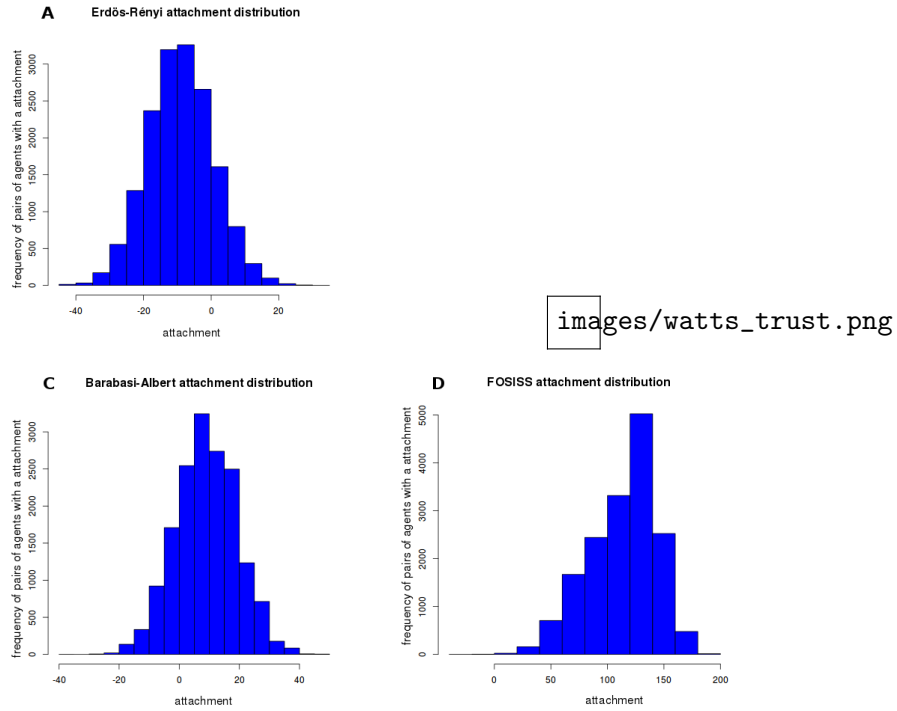


Figure 7: Strength of interactions distribution on different topologies. A. Random, Erdős-Rényi network displays a normal-like distribution. B. Watts-Strogatz network strength of interactions distribution is bimodal and highly skewed to the right. C. Barabási-Albert network strength of interactions distribution is also a normal-like curve. D. FOSISS, biomedical researchers collaboration network distribution of strength of interactions is skewed to the right.

531 An interesting feature of highly communicated networks (characterized  
532 by high values of clustering coefficient) is the fact that certainty among play-  
533 ers seems to be enhanced, that is, in such networks the rate of change of  
534 strategy is significantly lower (and smoother) than in poorly connected net-  
535 works. This is another instance in which easier communication (i.e. lower  
536 average minimum path lengths) lead to better performance of the whole  
537 collaborative research system.

538

## 539 6 Discussion and Conclusions

540 In this work we have analyzed the influence of parameters given by the  
541 underlying social structure of a science collaborative network on collective  
542 strength of interactions and utility dynamical behavior, based on a class of  
543 iterated PD and coordination games. Such parameters include mainly local  
544 and global connectivity like the degree centrality and average clustering co-  
545 efficient, as well as communication patterns.

546

547 Under the assumptions given by the model, we were able to notice that,  
548 in general, communication within the social collaborative networks has a  
549 positive correlation with average strength of interactions between the indi-  
550 viduals partaking in the games and also with the global collective utility  
551 (given by the sum of the individual payoffs). The better the communication  
552 among players, the higher the strength of interactions and the utility lead-  
553 ing to an optimized functioning of the whole *scientific collaboration system*.  
554 This is an important result that may be useful for scientific policy planning  
555 and may set a foundation for the optimal use of social networks in scientific  
556 collaboration as a means to improve the relationships among collaborating  
557 peers and thus ultimately the performance of research systems. We obvi-  
558 ously need more qualitative work in order to validate these results from a  
559 sociological and anthropological perspective.

560

561 To close this article, we would like to comment on three issues we con-  
562 sider important. The first one is about confronting our results with data  
563 available to us from the biomedical research community. Second, some re-  
564 marks we believe are important about the role of computational simulations  
565 in the social sciences. Finally we would like to comment on our future work.

566

567 The two main results of our model are the smoothness in the phase

568 transition-like behavior for different parameters and the distribution of util-  
569 ity and strength of interactions in the FOSISS network compared to other  
570 networks and their topologies. We are certain that the particular structure  
571 of FOSISS network is playing a central role in the results and because of  
572 that we would like to discuss it a little bit further.

573

574 FOSISS network is highly hierarchical according to its heterogeneity of  
575 0.873. Because of it, one would expect to find the presence of important  
576 hubs [35], that is, a few researchers control the whole network, as has been  
577 reported in some other places [36]. Surprisingly, FOSISS network central-  
578 ization is very low 0.023, i.e. there are no researchers that centralize the  
579 majority of connections. Our guess is that the network is composed of  
580 many small communities or groups with a central researcher or Principal  
581 Investigator (PI). If this is the case in the FOSISS network, it means that  
582 those groups have a very hierarchic structure as well.

583

584 Under such structure, when the clustering coefficient is considered, it can  
585 be said that groups are also well connected but that inter-group connections  
586 are sparse. In other words, individual groups are strongly connected but the  
587 network as a whole is supported by a small number of links. We came by  
588 this idea partially from another study about scientific collaborations based  
589 on co-authorships in one of the research centers that is part of the FOSISS  
590 network [10]. In the cited reference an apparently well integrated community  
591 was found (high clustering coefficient and a very short characteristic path  
592 length). Such integration was mostly superficial, since it depended on the  
593 presence of external collaborators from other research institutions (most of  
594 them from overseas). Removing these external collaborators broke down the  
595 network into small subgraphs that worked independently. Remarkably those  
596 subgraphs corresponded to the real groups of that research center. What is  
597 more, several groups had a hierarchical structure as the one we suspect is  
598 common in FOSISS subgraphs. The results showed that collaboration was  
599 poor between groups but strong among the members of each group, and that  
600 collaboration among groups doesn't emerge bottom-up, instead it seems to  
601 be promoted from the top, from the administrative authorities.

602

603 We believe that the the situation just refered is also true for the whole  
604 biomedical research community in México. The amount of PIs who have also  
605 been collaborators in other projects is about one fifth of the total number  
606 of participants. This is a number big enough to connect the whole network  
607 in one giant component. Yet, due to the topological characteristics of the

608 FOSISS network, it appears to be the case that researchers can be the lead-  
609 ers in one project and collaborators of different project of *its own research*  
610 *group*. CONACyT's funding policies makes it impossible for a researcher  
611 to get a grant from a fund if that researcher has an ongoing project with a  
612 grant from that same fund. That is, a researcher can ask for a grant from  
613 FOSISS if and only if at the moment he doesn't have a grant from FOSISS  
614 already. This policy has lead researchers from the same group to ask for  
615 grants from the same fund in order to rise their budgets. A consequence of  
616 this behavior is that group interactions get reinforced, but intergroup con-  
617 nections not necessarily so.

618  
619 As is common among scientific communities in biomedical research, PIs  
620 play a central role in the network. Strong PIs and well connected groups  
621 seem to be somehow responsible for the high levels of strength of interactions  
622 and the centralization of utility in our simulation. As for what seems to be  
623 phase transitions, in the case of FOSISS networks, these are smoother than  
624 those in the other networks with different topologies, even for those with a  
625 small-world topology. We think that this behavior is also the result of the  
626 hierarchical structure already mentioned. If this is true, strength of interac-  
627 tions is first lost in the edges that link different groups and then in the edges  
628 that connect members of the groups. Connections between groups would  
629 not be as dense as those inside the groups, which means that there would  
630 not be enough information of the behavior of one group regarding its neigh-  
631 bors to constrain them as it seems to happen with individual researchers  
632 inside their communities. Nevertheless, if values of the carrying capacity  
633  $\nu$  keep increasing and it becomes more difficult to strengthen interactions,  
634 then strength of interactions begins to diminish inside groups.

635  
636 Another issue that we would like to mention about FOSISS network  
637 topology is that it is not a robust collaboration network. At the level of  
638 groups, these might be well connected and consolidated but at the level of  
639 the network, this could be no more than an aggregate of individual groups.  
640 A robust network would be resistant to changes in the connections between  
641 groups but in the case of FOSISS, it seems that the network would brake  
642 down into small research groups by cutting some edges, as it happened in our  
643 co-authorship collaboration network [10]. The lack of robustness might be  
644 indicative of the fact that resources stay inside the groups, that is, they do  
645 not circulate through or articulate different communities. For example, one  
646 may think of certain expensive technologies for genomic research that could  
647 be bought once and shared among research groups, however, this doesn't



648 seem to happen very often. There are some other consequences, such as  
649 low communication among groups, atomization of practices and know-hows,  
650 redundancy in equipment tenancy, difficulties for implementing community-  
651 wide infrastructures such as biobanks, etc.

652

653 On these grounds, part of our future work is based on some of the results  
654 presented here. We would like to identify researchers in our simulations and  
655 corroborate their situation in the model and in the real world. We are also  
656 interested in going back to the field and interviewing those groups with an  
657 interesting behavior found in our simulations, probably we would follow a  
658 similar strategy as the one developed in [42]. Finding communities beyond  
659 the level of the groups is an important task. We think that there are many  
660 possibilities that emerge from the integration of different methodologies.  
661 Moreover, studying social processes in science is particularly attractive due  
662 to the amounts of data already available that can be easily collected. This  
663 is a privilege because simulations can be designed on real world data, some-  
664 thing that only very recently has become possible [43].

665

666 No doubt, social sciences are getting more “computational”. This can  
667 be seen everywhere, but curiously enough, disciplines like physics and com-  
668 puter science are the ones who are moving towards the social sciences and  
669 not so much the other way around. It might be the case that the pioneer-  
670 ing disciplines in the computational social sciences will set the agenda, an  
671 agenda that will apparently be mostly based on taking advantage of big data  
672 and on hypothesis-free approaches. We believe that the social sciences have  
673 important questions that should be added to that agenda and those ques-  
674 tions may not be answered only by big data techniques but they may require  
675 creating models and simulations in the style of the best hypothesis-driven  
676 research.

## 677 **7 acknowledgements**

678 The authors gratefully acknowledge support by grants: CB-222220-R/2013  
679 and CB-179431/2012 (Consejo Nacional de Ciencia y Tecnología). We would  
680 also like to acknowledge CONACyT for letting us access their database.  
681 Finally, we acknowledge the help of Francisco Allende for his work curating  
682 databases of researchers.

## 683 References

- 684 [1] Vermeulen, Nikki; Parker, John N & Penders, Bart, Understanding  
685 life together: A brief history of collaboration in biology, *Endeavour*,  
686 37(3): 162-171 (2013)
- 687 [2] Knorr-Cetina, Karol; *Epistemic cultures: how the sciences make*  
688 *knowledge*, MIT Press, Cambridge, MA (1999)
- 689 [3] Strasser, B.J. Collecting and Experimenting: The Moral Economies  
690 of Biological Research, 1960s-1980s. *Preprint no. 310*. Berlin: Max  
691 Planck Institute for the History of Science, (2006)
- 692 [4] Müller-Wille, S. & Charmantier. I. Natural history and information  
693 overload: The case of Linnaeus. *Studies in History and Philosophy of*  
694 *Biological and Biomedical Sciences*. 43:4-15, (2012)
- 695 [5] Strasser, BJ. Data-driven sciences: From wonder cabinets to elec-  
696 tronic databases. *Studies in History and Philosophy of Biological and*  
697 *Biomedical Sciences* 43:85-87, (2012)
- 698 [6] Sonnenwald, DH. Scientific Collaborations. *Annual Review of Infor-*  
699 *mation Science and Technology*, 41.
- 700 [7] Newman, MEJ. Clustering and preferential attachment in growing net-  
701 works. *Physical Review E*, 64:025102-1/025102-4, (2001)
- 702 [8] Newman, MEJ. Coauthorship networks and patterns of scientific col-  
703 laboration. *PNAS*, 101, no. suppl 1, 5200-5205 (2004)
- 704 [9] Elango, B. & Rajendran, J. Authorship Trends and Collaboration Pat-  
705 tern in the Marine Sciences Literature: A Scientometric Study. *Inter-*  
706 *national Journal of Information Dissemination and Technology*, 2(3),  
707 166-169, (2012)
- 708 [10] Hernández-Lemus, Enrique & Siqueiros-García JM, Information the-  
709 oretical methods for complex network structure reconstruction. *Com-*  
710 *plex Adaptive Systems Modeling*. 1: 8, doi:10.1186/2194-3206-1-8  
711 (2013).
- 712 [11] Gilbert, N. A Simulation of the Structure of Aca-  
713 demic Science. *Sociological Research Online*, 2(2) 3,  
714 <http://www.socresonline.org.uk/2/2/3.html>

- [12] Edmonds, B., Gilbert, N., Ahrweiler, P., & Scharnhorst, A. Simulating the Social Processes of Science. *Journal of Artificial Societies and Social Simulation*, 14(4) 14, <http://www.jasss.soc.surrey.ac.uk/14/4/14.html>
- [13] Hara, N. Solomon, P. Kim SL. Sonnenwald DH. An Emerging View of Scientific Collaboration: Scientists Perspectives on Collaboration and Factors that Impact Collaboration. *Journal of the American Society for Information Science and Technology*, 54(10):952965, (2003)
- [14] Lieberman, E. Hauert, C. & Nowak, MA. Evolutionary Dynamics on Graphs. *Nature*, 433, doi:0.1038/nature03204 (2005).
- [15] Lo, T.S., Chan, H.Y., Hui, P.M., Johnson, N.F., Theory of networked minority games based on strategy pattern dynamics, *Physical Review E* 70, 056012 (2004)
- [16] Buesser, P. and Tomassini, M., Evolution of cooperation on spatially embedded networks, *Physical Review E* 86, 066107 (2012)
- [17] Ichinose, G., Tenguishi, Y., and Tanizawa, T., Robustness of cooperation on scale-free networks under continuous topological change, *Physical Review E* 88, 052808 (2013)
- [18] Ahmed, A., Karlapalem, K., Inequity aversion and the evolution of cooperation, *Physical Review E* 89, 022802, (2014)
- [19] Vilone, D., Ramasco, J.J., Sánchez, A., San Miguel, M., Social imitation versus strategic choice, or consensus versus cooperation, in the networked Prisoners Dilemma, *Physical Review E* 90, 022810, (2014)
- [20] Wardil, L., Hauert, C., Cooperation and coauthorship in scientific publishing, *Physical Review E* 91, 012825, (2015)
- [21] Hanauske, M., Evolutionary Game Theory and Complex Networks of Scientific Information, in A. Scharnhorst et al. (eds.), *Models of Science Dynamics, Understanding Complex Systems Series*, Springer Berlin Heidelberg, (2012)
- [22] Szabó, György & Fáth, Gábor, Evolutionary games on graphs. *Physics Reports*, 446: 97-216 (2007)
- [23] Nowak, Martin A. & May, Robert M, Evolutionary Games and Spatial Chaos. *Nature*, 359: 826-829, (1992)

- 748 [24] Oshtuki, Hisashi, Hauert, Christoph, Lieberman, Erez & Nowak, Mar-  
749 tin A, A simple rule for the evolution of cooperation on social graphs.  
750 *Nature*, 441: 502-505 (2006)
- 751 [25] Axelrod, Robert, *Evolution of cooperation: Revised edition*, Basic  
752 Books, Cambridge, MA (2006)
- 753 [26] Nowak, Martin A. & Highfield, Roger, *Supercooperators. Altruism,*  
754 *Evolution, and Why We Need Each Other to Succeed*. Free Press, New  
755 York, NY (2011)
- 756 [27] Santos, Francisco C., Pacheco, Jorge M., Lenaerts, Tom, Cooperation  
757 prevails when individuals adjust their social ties. *PLoS Computational*  
758 *Biology*. 2(10): 1284-1291 (2006)
- 759 [28] Santos, F.C. & Pacheco, J.M., Scale-Free Networks Provide a Unifying  
760 Framework for the Emergence of Cooperation. *PRL*, 95: 098104-1 to  
761 098104-4, (2005)
- 762 [29] Hauert, Christoph & Doebeli, Michael, Spatial structure often inhibits  
763 the evolution of cooperation in the snowdrift game. *Nature*, 428: 643-  
764 646, (2004)
- 765 [30] Erdős, P. & Renyi, A. On Random Graphs. *Publ. Math*, 6:290 (1959).
- 766 [31] Watts, Duncan J. & Strogatz, Steven H, Collective dynamics of 'small-  
767 world' networks. *Nature*, 393: 440-442 (1998)
- 768 [32] Barrat, A.; Weigt, M., On the properties of small-world network mod-  
769 els, *European Physical Journal B* 13 (3): 547560, (2000)
- 770 [33] Barabási, Albert-László. & Albert, Réka, Emergence of Scaling in Ran-  
771 dom Networks. *Science*, 286: 509-512 (1999)
- 772 [34] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D,  
773 Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment  
774 for integrated models of biomolecular interaction networks. *Genome*  
775 *Research*, 13(11):2498-504, (2003)
- 776 [35] Jun Wu, Yue-Jin Tan, Hong-Zhong Deng, Da-Zhi Zhu. A new measure  
777 of heterogeneity of complex networks based on degree sequence. Eds.  
778 Ali Minai, Dan Braha, Yaneer Bar-Yam *Unifying Themes in Complex*  
779 *Systems. Proceedings of the Sixth International Conference on Com-*  
780 *plex Systems*, Springer Berlin Heidelberg, 66-73, (2008)

- 781 [36] Yousefi-Nooraie R, Akbari-Kamrani M, Hanneman RA, Etemadi A.  
782 Association between co-authorship network and scientific productiv-  
783 ity and impact indicators in academic medical research centers: A  
784 case study in Iran. *Health Research Policy and Systems* 6:9 (2008)  
785 doi:10.1186/1478-4505-6-9.
- 786 [37] Simon, HA. On a class of skew distribution functions. *Biometrika*, 42  
787 (3-4): 425-440, (1955)
- 788 [38] Price, Derek JS. Networks of Scientific Papers. *Science*, 149(3683):  
789 510-515, (1965)
- 790 [39] Merton, Robert K, The Matthew Effect in Science, *Science*, 159(3810):  
791 56-63 (1968)
- 792 [40] Epstein JM & Axtell R. *Growing artificial societies. Social science*  
793 *from the bottom up*. The Brookings Institution Press and The MIT  
794 Press, Washington, DC. (1996)
- 795 [41] Lansing JS. “Artificial Societies” and the Social Sciences. *Artificial*  
796 *Life*, 8:279-292, (2002)
- 797 [42] Hara N, Solomon P, Kim SL, Sonnenwald DH. An emerging view of  
798 scientific collaboration: Scientists’ perspectives on collaboration and  
799 factors that impact collaboration. *Journal of the American Society for*  
800 *Information Science and Technology*, 54(10):952965, 2003
- 801 [43] Barabási, AL. The network takeover. *Nature*, 8:1416, (2012)
- 802 [44] Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, Kertesz  
803 J, Loreto V, Moat S, Nadal J.-P, Sanchez A, Nowak A, Flache A, San  
804 Miguel M, Helbing D. Manifesto of computational social science. *The*  
805 *European Physical Journal, Special Topics* 214(1):325346, (2012)