THE DYNAMICS OF THE DISTRIBUTION
OF CO-AUTHORS

In the present paper, the case of a database of scientific articles is described. There can be observed the quantitative effects of the increasing strength of cooperation between scholars. It is manifested in time-related features changes of the numbers distribution of a single co-authors' article. The distribution of the number of co-authors of an article recorded in the database evolves with time from a profile with no more than one author to a profile with several authors. A social dilemma model is proposed to explain the dynamics of changes in the distribution of the number of co-authors. The most successful strategy of the three considered alternative strategies of cooperation is selected.

Keywords: cooperation, PubMed, authorship, social dilemma

Streszczenie

W niniejszym artykule, na podstawie bazy artykułów naukowych, opisano ilościowy efekt wzrostu współpracy uczonych, która objawia się zmianami cech rozkładu liczby współautorów artykułu w czasie. Rozkład liczby współautorów artykułu dostępnego w bazie danych zmienia się w czasie z maksimum dla jednego autora do maksimum kilku autorów. Do wyjaśnienia dynamiki zmian rozkładu liczby współautorów zaproponowano model dylematu społecznego. Spóźnieniu, trzech zostało wybrana jedna strategia współpracy odnosząca największe sukcesy.

Słowa kluczowe: współpraca, PubMed, autorstwo, dylemat społeczny

DOI: 10.4467/2353737XCT.15.116.4153
1. Introduction

Cooperation is fundamental to numerous biological and social systems. Multicellular organisms, social insects, herds of animals and human societies are examples of systems that are dependent on cooperative interactions. Explaining the emergence of cooperation has been a challenge since Darwin. Apparently, cooperators help others at a cost to themselves, while defectors receive the benefits of altruism without providing any help in return. The emergence of cooperation has been studied based on tools developed within the framework or evolutionary game theory. Games such as prisoner’s dilemma games (PDGs) [1], snowdrift games (SGs) [2] and stag-hunt games (SHs) [3] form the standard set of social dilemmas. All aforementioned evolutionary games are played by a pair of players in which they receive a reward $R$ for mutual cooperation and a punishment $P$ for mutual defection. If one cooperates but the other defects, the defector (D) receives the payoff $T$, while the cooperator (C) gets the payoff $S$. In the case of PDG we have $TRPS$. In a SG, the payoff rank is $TRSP$. The stag-hunt game offers more support for cooperation because the rank of payoffs $(RTPS)$ favours cooperation over defection. Public goods games (PGGs) [4] engage many players and the number of players is an adjustable parameter. Cooperating players contribute to a common pool, which is then enhanced by a factor $\eta$ and equally distributed between cooperators and defectors. According to these models, a cooperator pays a price to receive a benefit; conversely, a defector pays no price but can receive extra benefits. It is commonly assumed that cost and benefit are measured in terms of fitness. It transpires that in any mixed population, defectors have a higher average fitness than cooperators and thus, after some time, cooperators vanish from the population. However, it is possible that under some circumstances, cooperators are not extinct in the course of system evolution.

In the present study, the case of a real system of cooperating agents, which adjust their cooperation preferences to the varying competitive pressure is described. As a result, the number of agents engaged in a single cooperation event changes over the course of the evolution of the system. The analysis of the system is based on real-world data – a database compiling bibliographic details of articles published in scientific periodicals over the last 60 years. Publishing an article is a cooperative act and the cooperation preferences of the scholars can be quantitatively characterised by the features of the distribution of the number of co-authors of a single article. From the analysis of the database, it follows that the average number of co-authors increases with time. The modal value of the number of co-authors of a single article increased from 1 to 4 in the period from 1950 to 2010 – this is a clear sign of the increasing strength of cooperation. We propose a novel social dilemma model to explain the observed dynamics of the distribution of the number of co-authors; possible further developments of the model are also discussed.

2. Related works

Mechanisms that facilitate cooperation, include kin selection [5–9], direct and indirect reciprocity [1, 10–13], network reciprocity [14–17], trust [18, 19], group selection [20, 21], cooperation in game theory [22, 23], and evolution of cooperation [24, 25].
An important quantitative aspect of cooperation is the number of players engaged in a single play. The number of players playing the aforementioned social dilemma games is always fixed, while in the case of real systems, the number of players engaged in a single cooperation event is a quantity that should be set by the dynamics of the game. For example, in studies of the behaviour of predators [26‒29], the group size is partially determined by food intake, which is maximized for an optimal group size [30]. The problem of an optimal group size has been considered recently in the context of another social dilemma – the volunteer’s dilemma [31], but in that case, the group size is not set dynamically in the course of the system evolution. A mechanism of setting group size, based on ecology-inspired principles, has also been proposed within the context of the public goods game [32, 33]. An emerging field of co-evolutionary games [34, 22] can become a convenient framework for studying the problem of how the output of the games played by the players may affect the interaction network.

The evolution of the research of cooperation could be extended into crawling systems [35], where such cooperation would improve crawling performance. The cooperation processes are able to be modelled using Petri nets [36] to visualise various cooperation mechanisms. Computational intelligence methods [37] appear to be usable for evolution games, they are both based on nature laws and should be combined into one research strategy.
the available quantitative data is the distribution of the number of co-authors and one can only speculate as to how the fitness of the scholars is related to that distribution.

Because it appears that the standard evolutionary models of cooperation (PDGs, SGs, SHs or PGGs) do not have enough features to determine the outcome of cooperation in the form, for example, joint publications, there is a need of the new tool to model the kind of data described above. It is important that possible models account for the competition between scholars for limited resources as the total volume of accessible journals is limited. Additionally, to build their prestige, scholars attempt to prepare publications that are of as high a standard as possible, but an author cannot generally compare the quality of his or her publications to the quality of publications by other authors – this is the task of the reviewers who score the submitted publications.

3.1. Model

To qualitatively explain the empirical data discussed in the previous section, we consider a population consisting of \( N \) agents. Each agent \( A \) is located on one site of a network \( G \) and can cooperate with other agents, provided that they are connected to agent \( A \) by an edge in \( G \). Throughout this paper, it is assumed that \( G \) is a random graph, although other types of complex network underlying the model are clearly possible.

An agent is an abstract object characterised by the properties and actions which they can perform. The properties of an agent define their state, which is modified in the course of simulation. According to the proposed model, an agent’s properties are:

1. credit \( C \)
2. activity \( E \)
3. quality $Q$
4. standard deviation of quality $sdQ$
5. probability of cooperation $P$
6. list of neighbours $L$ in the underlying network $G$

An agent can perform two actions. First, they can update their probability of cooperation $P$. The formula for probability of cooperation update can possibly depend upon an agent’s properties (e.g. their actual credit) and defines an agent’s strategy of cooperation. Secondly, an agent can compete for resources, submitting proposals for resource allocation. A proposal can be submitted individually or in cooperation with other agents, depending on probability of cooperation $P$.

The state of the system (defined as the state of all agents) is modified iteratively. At the beginning of the simulations, credit $C_A$, activity $E_A$, quality $Q_A$ and standard deviation of quality $sdQ_A$ of an agent $A$ are drawn from normal distributions with the mean and standard deviation being the parameters of the model. Activity $E_A$, quality $Q_A$ and the standard deviation of quality $sdQ_A$ are kept fixed during the simulations. The probability of cooperation $P_A$ of agent $A$ is initially equal to 0 and is updated in every simulation step in accordance with the cooperation strategy of agent $A$. The list of neighbours $L_A$ is also fixed – it is assumed that the topology of $G$ does not change in the course of simulation.

At every simulation step, the credit of each agent is decreased by a constant amount $dC$ (which, for example, represents the costs of participating in the game). Then, with probability $E_A$, every agent generates a proposal for resource allocation. A proposal for resource allocation $R$ is an abstract object characterised by its quality $QR$ and the list of agents $L_R$, which have authorised it. When an agent $A$ generates a proposal $R$, the quality of proposal $Q_R$ is drawn from a normal distribution with mean and standard deviation equal to $Q_A$ and $sdQ_A$ respectively. The list of agents, which have authorised see above note $R$ contains initially only one element – the agent $A$. Then for every agent $B$ in $L_A$, the agent $A$ invites the agent $B$ to participate in the proposal $R$. The invitation is accepted with probability $P_A \times P_B$. After accepting an invitation, the quality $Q_R$ of $R$ is increased by amount $q$ which is drawn from a normal distribution with mean and standard deviation equal to $Q_B$ and $sdQ_B$ respectively. Agent $B$ is then added to $L_R$.

All agents submit their proposals synchronously and a pool of all submitted proposals is created. The proposals are then scored by their quality and the authors of each proposal receive the payoffs based on the scores of their proposals. Each payoff is shared in equal parts between all authors of a proposal and the credit of each author is increased accordingly. It is assumed in the model that the total sum of payoffs is adjusted so as to keep the total sum of credits constant. In the simulations, it was also assumed that the payoff depends linearly on the rank of a proposal in the pool – the proposal with the lowest quality receives zero payoff, while the proposal with the highest quality receives the highest payoff.

Three strategies of cooperation were studied. Agents adopting the strategy ‘Never cooperate’ keep their probability of cooperation always equal to zero. Agents adopting the strategy ‘Cooperate when in trouble’ increase their probability of cooperation whenever their credit decreases, otherwise they decrease their probability of cooperation. Agents adopting the strategy ‘Cooperate’ update their probability of cooperation whenever their credit decreases but never decrease it below the actual value. Every agent updates their
probability of cooperation in every simulation step. It was assumed that the credit-probability of cooperation dependence is given by a sigmoid-like function approximating between 1 for low credits and 0 for large credits.

4. Social dilemma in the network of co-authors

To see that the model described above belongs to a class of social dilemmas, consider the following simplified version of the model. Let each agent be characterised by the same constant probability of cooperation $P$ and a constant number of cooperators $N$. Assume for now that $P$ determines the probability that a given agent acquires another agent to authorise his proposal. Let every agent accept the invitation to authorise other agent’s proposal with probability equal to 1. It should also be assumed that all agents submit their proposals at every time step and the payoff of the proposal authorised by $k \geq 1$ co-authors is equal to $f(k)$. In the following equation, it is assumed that $f(k) = k^\alpha$, where $\alpha \geq 0$ is an exponent. Then the average payoff $S$ of a single agent can be split into two parts: the payoff $S_d$ from the proposal submitted by the agents and the payoff $S_c$ from the proposals of the agents, which have invited the agent to write their proposals. One has:

$$S_d = \sum_{k=0}^{N} \binom{N}{k} \cdot P^k \cdot (1 - P)^{N-k} - k \cdot \frac{f(k+1)}{k+1}$$

(1)

$$S_c = N \cdot P \cdot \sum_{k=0}^{N-1} \binom{N-1}{k} \cdot P^k \cdot (1 - P)^{N-1-k} - k \cdot \frac{f(k+2)}{k+2}$$

(2)

After basic manipulation, it can be shown that:

$$S = S_d + S_c = \sum_{k=0}^{N} \binom{N}{k} \cdot P^k \cdot (1 - P)^{N-k} - k \cdot f(k+1)$$

(3)

It follows immediately from Eq. 1 and 2 that if $f(k) = k^\alpha$, then for any $\alpha \geq 0$ smaller than 1, the payoff $S_d$ of an agent which cooperates ($P > 0$) is smaller than 1, which is the payoff $S_d$ of a defector ($P = 0$). Consequently, a defector facing cooperators would have higher fitness than his cooperating co-players; thus, a rational player will choose the strategy to defect. On the other hand, it follows from Eq. 3 that the average payoff (equal to 1) in a network of defectors is lower than the average payoff in a network of cooperators. Thus, in the described system, the agents are facing a social dilemma. Note that the cooperation strategy is always advantageous if $\alpha > 0$.

To solve the social dilemma described above, an additional constraints must be introduced to promote cooperation. In the present study, the reciprocity mechanism is adopted. In the case described above, cooperation cannot emerge because the probability $P$ of acquiring a co-author is independent on the probability (equal to 1) of accepting other agents’ invitations to authorise their proposals. If each agent is instead characterised by
a single probability of cooperation which determines both the probability of proposing authorship to other agents and the probability of accepting other agents invitations, then cooperation emerges, as is described in the next section.

4.1. Results

The results presented in the current section were obtained for the following values of model parameters. A random network of 1000 agents was created with the average number of neighbours being equal to 20 (the total number of network links was 10 times larger than the number of agents). The initial credit of the agents was drawn from normal distribution $N(30, 2)$ with the mean and standard deviation equal to 30 and 2 respectively. At every simulation step, the credit was decreased by 0.1. The quality of agents was drawn from normal distribution $N(2, 0.5)$. The standard deviation of quality was set to 0.2 and the activity was set to 0.5 for all agents. The credit-probability of cooperation dependence was given by the function:

$$P(C) = \frac{1}{1 + e^{C/10}}$$  \hspace{1cm} (4)

where $P$ stands for the probability of cooperation and $C$ stands for an agent’s credit.

The credit of agents adopting the strategy ‘Never cooperate’ is plotted against their quality in Fig. 2 for three points in time. Note that the destiny of an agent is completely determined by their quality; thus, in a sense, an agent has no way to escape their destiny. Inequality of credit share or social inequality increases over time. The distribution of credit values is more or less uniform after an initial transient period (Fig. 3).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{credit_vs_quality.png}
\caption{Credit plotted against the quality of agents in a network of agents adopting the strategy ‘Never cooperate’. The results for time (measured in simulation steps) equal to 1 unit (triangles), 5,000 units (circles) and 10,000 units (squares).}
\end{figure}
The credit of agents adopting the strategy 'Cooperate when in troubles' is plotted against their quality in Fig. 4 for four points in time. Also in this case a strong dependence between an agent’s quality and their credit is observed in some range of quality values, but that range shrinks with time. Comparison of Fig. 4 and Fig. 5, in which probability of cooperation is plotted vs. an agent’s quality indicates that that strong dependence is observed for agents which actually do not cooperate, keeping their probability of cooperation equal to zero. The group of non-cooperating agents disappears with time and a more or less egalitarian
Fig. 4. Credit plotted against the quality of agents in a network of agents adopting the strategy ‘Cooperate when in trouble’. The results for time (measured in simulation steps) equal to 1 unit (top-left panel), 5,000 units (top-right panel), 10,000 units (bottom-left panel) and 30,000 units (bottom-right panel).

Fig. 5. Probability of cooperation plotted against the quality of agents in a network of agents adopting the strategy ‘Cooperate when in troubles’. The results for time (measured in simulation steps) equal to 1 (top-left panel), 5,000 units (top-right panel), 10,000 units (bottom-left panel) and 30,000 units (bottom-right panel).
society arises (Fig. 6). Although cooperation becomes yet more important in a network of agents adopting the strategy ‘Cooperate when in troubles’, the distribution of the number of co-authors of a single proposal converges to a profile which has a maximum for one author (Fig. 7). The average number of co-authors of a single proposal converges to $1.78 \pm 0.04$ (Fig. 8).

Fig. 6. The distribution of credits in a network of agents adopting the strategy ‘Cooperate when in trouble’. The results for time (measured in simulation steps) equal to 1 unit (top-left panel), 5,000 units (top-right panel), 10,000 units (bottom-left panel) and 30,000 units (bottom-right panel)
Fig. 7. Model distributions of the number of co-authors of a single article for the case of agents adopting the strategy ‘Cooperate when in troubles’

Fig. 8. Average number of the co-authors of a single article plotted against the simulation step ‘time’ for the case of agents adopting the strategy ‘Cooperate when in trouble’. The error bars represent the mean square error of the average number of authors.
Fig. 9. Credit plotted against the quality of agents in a network of agents adopting the strategy ‘Cooperate’. The results for time (measured in simulation steps) equal to 1 unit (top-left panel), 5,000 units (top-right panel), 10,000 units (bottom-left panel) and 80,000 units (bottom-right panel).

Fig. 10. Probability of cooperation plotted against the quality of agents in a network of agents adopting the strategy ‘Cooperate’. The results for time (measured in simulation steps) equal to 1 unit (top-left panel), 5,000 units (top-right panel), 10,000 units (bottom-left panel) and 80,000 units (bottom-right panel).
The credit of agents adopting the strategy ‘Cooperate’ is plotted against their quality in Fig. 9 for four points in time. Relatively early, two branches develop on the quality-credit plot. The branch observed for high values of quality is the trace of the existence of the high-quality agents which have not yet decided to actively cooperate, thus keeping their probability of cooperation low (Fig. 10). On the other hand, can be observed a branch

Fig. 11. The distribution of credits in a network of agents adopting strategy ‘Cooperate’. The results for time (measured in simulation steps) equal to 1 unit (top-left panel), 5,000 units (top-right panel), 10,000 units (bottom-left panel) and 80,000 units (bottom-right panel)
Fig. 12. Model distributions of the number of co-authors of a single article for a case of agents adopting the strategy ‘Cooperate’

Fig. 13. Average number of the co-authors of a single article plotted against simulation step ‘time’ for a case of agents adopting the strategy ‘Cooperate’. The error bars represent the mean square error of the average number of authors.
in the range of low qualities is the trace of the action of low quality agents, that cooperate and join their attempts to gain a higher share in the resource allocation. After a long time the credit of agents is decoupled from their quality and the cooperation is the main mechanism, which determines an agent’s credit. The distribution of the agents’ credits is broader than in the case of agents adopting the strategy ‘Cooperate when in trouble’ and has an exponential tail (Fig. 11). Because of high competitive pressure (agents never decrease their probability of cooperation) in a network of agents adopting the strategy ‘Cooperate’, the modal value of the distribution of the number of co-authors of a single proposal moves to the right in time (Fig. 12). Obviously, there exists a natural limit for the evolution of this distribution, namely, the distribution of the neighbours in the underlying network of contacts $G$. The average number of co-authors of a single proposal increases with time (Fig. 13).

![Fig. 14. Average credit of agents adopting the strategy ‘Never cooperate’ (squares) and ‘Cooperate’ (circles) in a network consisting of agents of both types plotted against the fraction of agents adopting the strategy ‘Cooperate’. The error bars represent the mean square error of the average credit values](image)

Finally, we studied competition between agents adopting different strategies within the same network of contacts. Even for the fairly small fraction of agents adopting the strategy ‘Cooperate’, these agents gain a much higher share of the total credit compared to agents adopting the strategy ‘Never cooperate’ (Fig. 14). Similar results (not shown) were obtained for networks consisting of agents adopting the strategies ‘Cooperate’ and ‘Cooperate when in trouble’.
5. Discussion

In the present paper, we have described the case of a database of scientific articles in which the emergence of cooperation between the scholars and the increasing strength of cooperation between them is manifested by an increasing average number of co-authors of a single article, recorded in the database. We have also proposed a social dilemma model to explain the dynamics of the changes in the distribution of the number of co-authors. Among three considered play strategies, the most successful strategy was selected.

Within the framework of the proposed model, the agents compete for common, limited resources, which are allocated to them based on their quality (or fitness). Based on the acquired fitness, the agents adjust their behaviour. Commonly, a generation exchange is implemented within the framework of evolutionary games – this feature is not embedded in the presented model because the time scale of the dynamics of the average number of co-authors observed in the PubMed database is of the same order as the typical duration of an academic career.

The proposed model demonstrates the mechanisms which can induce changes in the distribution of the number of co-authors. It is worth noticing that the dynamics exhibited by the system of agents competing for resources and developing cooperation in the course of simulation is an example of Red Queen dynamics [38]. The agents with low quality must cooperate in order to be allocated the resources they request, otherwise, the resources would be allocated to high quality agents. After a time, groups of cooperating low-quality agents become more effective in gaining resources; thus, high-quality agents are forced to cooperate under increasing competition pressure. However, this has an effect on the low-quality agents, which must strengthen their cooperation.

Many further developments are possible. First, in the present model there is only one probability of cooperation, which determines both the probability of inviting others to cooperate and the probability of accepting cooperation. In reality, the probability of invitation can be different from the probability of accepting invitations. Next, the probability of cooperation can be associated not with an agent but with an arc of the graph of contacts – there is no need to assume an undirected graph in this case because an agent can prefer cooperation with one of his neighbours over cooperation with another individual. In that case, the probability of cooperation can depend on the history of joint publication (successful publication increases probability of cooperation, while a fault decreases it) and on the quality of both sides the link. Next, the network structure was static in the present model, but in reality, it can change at the time scale during the simulations – the number of agents, the number of links and the topology of the graph can evolve over time. Additionally, the resources and the strategy for allocating the resources are fixed and this assumption can be easily relaxed. Finally, the proposed model can be reformulated within the evolutionary framework with credit interpreted as fitness. Another dynamic variable is the function which specifies how the payoff depends on the number of players. In the present paper, a simple form of a function was assumed but there are other choices possible, leading to the formation of stable groups. This problem is currently under study.
The dilemma one is facing when investigating models like PDG is how cooperation arises in the population of selfish individuals. In the model described here, the individuals cooperate because they are selfish – to gain the best resources for themselves, they must create alliances.

References


