A bibliometric study of collaboration and influence in the field of the Iterated Prisoner's Dilemma

Nikoleta E. Glynatsi[1] and Vincent A. Knight[1]

[1] *Cardiff University, School of Mathematics, Cardiff, United Kingdom*

Abstract

This manuscript explores the collaborative behaviour of authors in the field of the Prisoner's Dilemma using bibliometric analysis, topic modeling, and the co-authorship network. The results demonstrate that the Prisoner's Dilemma is a field of continued interest and a collaborative field. Authors are very like to cooperate with a collaborator's co-author and single author publications are rare. The co-authorship network however suggests that authors do not influence or gain much information by their connections, unless they are connected to a "main" group of authors. The articles collected for this work are classified based on the language of their abstracts in five different topics, and the results are additionally verified by the topics' networks.

1 Introduction

The Prisoner's Dilemma (PD) is a well known game used since its introduction in the 1950's [13] as a framework for studying the emergence of cooperation; a topic of continuing interest for mathematical, social [33], biological [41] and ecological [43] sciences. This manuscript presents a bibliometric analysis of the Prisoner's Dilemma field, in the interest of understanding the number of publications over the years, the research trends within the field, and mainly to assess the collaborative behaviour of the authors in the field of cooperation itself.

As discussed in [44], bibliometrics (the statistical analysis of published works originally described by [35]) has been used to support historical assumptions about the development of fields [36], identify connections between scientific growth and policy changes [9], develop a quantitative understanding of author order [39], and investigate the collaborative structure of an interdisciplinary field [27]. Most academic research is undertaken in the form of collaborative effort and as [24] points out, it is rational that two or more people have the potential to do better as a group than individually. Collaboration in groups has a long tradition in experimental sciences and it has be proven to be productive according to [12]. The number of collaborations can be very different between research fields and understanding how collaborative a field is not always an easy task. Several studies tend to consider academic citations as a measure for these things. A blog post published by Nature [31] argues that depending on citations can often be misleading because the true number of citations can not be known. Citations can be missed due to data entry errors, academics are influenced by many more papers than they actually cite and several of the citations are superficial.

A more recent approach to measure collaborative behaviour is to use the co-authorship network, as described in [27]. Using this approach has many advantages as several graph theoretic measures can be used as proxies to explain authors relationship. In [27], they analyse the development of the field "evolution of cooperation" using this approach. The topic "evolution of cooperation" is a multidisciplinary field which also includes a large number of publications on the PD. This paper builds upon the work done by [27] and extends their methodology. Though in [27], they considered a data set from a single source, Web of Science, the data set described here, archived at [15], has been collected from five different sources. Co-authorship networks have also been used in [44, 26]. In [44] it is used for identifying the subdisciplines the interdisciplinary field of cultural evolution so that they can describe each subdiscipline and investigate trends in collaboration and productivity, and in [26] for examining the long-term impact of co-authorship with established, highly-cited scientists on the careers of junior researchers.

The Prisoner's Dilemma, much like cultural evolution and evolution of cooperation, has attracted the attention of researchers across fields, and more specifically its iterated version, the Iterated Prisoner's Dilemma (IPD). This is due to the game not only being mathematically intriguing [34] but also being used a model of social [45] and biological [10] interactions and having applications in numerous fields such as healthcare [41]. The game was chosen for this work as a result of it being a specific topic with publications across the academic fields.

The type of analysis which is carried out in this work has been carried out before in [27] and [44]. The novelty here is to consider a new data set on a specific game theoretic topic and approaches not considered in either works, such as topic modeling [5] and centrality measures. The data set of this work contains 2422 articles' metadata on the IPD that have been collected from five different prominent sources. The methodology used in this manuscript, which includes the data collection, is covered in Section 2. The results regarding, the amount of publications and the topics of the field over the years are described in Section 3. In Section 4, graph theoretical methods are used to ascertain the level of collaborative nature of the field and identify influence.

2 Methodology

Academic articles are accessible through scholarly databases. Several databases and collections today offer access through an open application protocol interface (API). An API allows users to query directly a journal's database and bypass the graphical user interface. Interacting with an API has two phases: requesting and receiving. The request phase includes composing a url with the details of the request. For example, `http://export.arxiv.org/api/query?search_query=abs:prisoner'sdilemma&max_results=1` represents a request message. The first part of the request is the address of the API. In this example the address corresponds to the API of arXiv. The second part of the request contains the search arguments. In this example it is requested that the word 'prisoners dilemma' exists within the article's title. The format of the request message is different from API to API. The receive phase includes receiving a number of raw metadata of articles that satisfies the request message. The raw metadata are commonly received in extensive markup language (xml) or Javascript object notation (json) formats [32]. Similarly to the request message, the structure of the received data differs from journal to journal.

The data collection is crucial to this study. To ensure that this study can be reproduced all code used to query the different APIs has been packaged as a Python library and is available online [16]. The software could be used for any type of projects similar to the one described here, documentation for it is available at: `http://arcas.readthedocs.io/en/latest/`. Project [16] allow users to collect articles from a list of APIs by specifying just a single keyword. Articles for which any of the terms

- {"prisoner's dilemma", "prisoners dilemma", "prisoner dilemma", "prisoners evolution", "prisoner game theory"}

existed within the title, the abstract or the text are included in the analysis. Four prominent journals in the field and a preprint server were used as sources to collect data for this analysis:

- arXiv [29]; a repository of electronic preprints. It consists of scientific papers in the fields of mathematics, physics, astronomy, electrical engineering, computer science, quantitative biology, statistics, and quantitative finance, which all can be accessed online.
- PLOS [1]; a library of open access journals and other scientific literature under an open content license. It launched its first journal, PLOS Biology, in October 2003 and publishes seven journals, as of October 2015.
- IEEE Xplore Digital Library (IEEE) [22]; a research database for discovery and access to journal articles, conference proceedings, technical stan-

dards, and related materials on computer science, electrical engineering and electronics, and allied fields. It contains material published mainly by the Institute of Electrical and Electronics Engineers and other partner publishers.

- Nature [18]; a British multidisciplinary scientific journal, first published on 4 November 1869. It was ranked the world's most cited scientific journal by the Science Edition of the 2010 Journal Citation Reports and is ascribed an impact factor of 40.137, making it one of the world's top academic journals.
- Springer [30]; a leading global scientific publisher of books and journals. It publishes close to 500 academic and professional society journals.

The data set is archived and available at [15]. Note that the latest data collection was performed on the 30$^{\text{th}}$ November 2018.

The data set contains a total of 2422 articles. These will be classified into research topics using **topic modeling**, a computational method designed to summarize large collections of texts by a small number of conceptually connected topics or themes [5, 17]. The text used for each article is its abstract. Topic modeling analyses the patterns with which words co-occur in the abstracts, and then a topic is defined by a probability distribution over words. So in one topic the word "cooperation" would have a high weighting, and in another topic it would have a low weighting. Similarly for all the words that have been captured by the abstracts. Note that the words have been simplified by ignoring order and removing stop words. Once the topics are defined, as a distribution over words, articles are assigned to them based on which topic has the highest percentage contribution for a given article. For example let the words that contribute to Topic A be: $0.039\times$"cooperation", $0.028\times$"study" and $0.026\times$"human" and for Topic B: $0.020\times$"cooperation", $0.028\times$"agents" and $0.026\times$"strategies". A document with abstract "The study of cooperation in humans" is assigned to Topic A because $(0.039 + 0.028 + 0.026) > 0.020$.

A issue regarding the technique is that the number of topics need to be specified in advance before running the algorithm. The suitable number of topics presented here are based on the topic coherence measure [38] and on the minimisation of overlapping words within two topics. Note that the minimisation of overlapping words is done manually by the authors.

The relationship between the authors within a field will be modeled as a graph $G = (V_G, E_G)$ where V_G is the set of nodes and E_G is the set of edges. The set V_G represents the authors and an edge connects two authors if and only if those authors have written together. The co-authorship network is constructed using the main data set [15] and the open source package [20]. The PD network is denoted as G where the number of unique authors $|V(G)|$ is 4226 and $|E(G)|$ is 7642 . All authors' names were formatted as their last name and first name (i.e. Martin A. Nowak to Martin Nowak). This was done to avoid errors such as Martin A. Nowak and Martin Nowak being treated as a different person. There are some authors for which only their first initial was found and are left this way.

Collaborativeness will be analysed using measures such as, isolated nodes, connected components, clustering coefficient, communities, modularity and average degree. These measures show the number of connections authors can have and how strongly connected these people are. The number of isolated nodes is the number of nodes that are not connected to another node, thus the number of authors that have published alone. The average degree denotes the average number of neighbours for each nodes, i.e. the average number of collaborations between the authors. A connected component is a maximal set of nodes such that each pair of nodes is connected by a path [11]. The number of connected components as well as the size of the largest connected component in the network are reported. The size of the largest connected component represents the scale of the central cluster of the entire network, as will discussed in the analysis section. Clustering coefficient and modularity are also calculated. The clustering coefficient, defined as 3 times the number of triangles on the graph divided by the number of connected triples of nodes, is a local measure of the degree to which nodes in a graph tend to cluster together in a clique [11]. It is precisely the probability that the collaborators of an author also write together. In comparison, modularity is a global measure designed to measure the strength of division of a network into communities. The number of communities will be reported using the Clauset-Newman-Moore method [8]. Also the modularity index is calculated using the Louvain method described in [6]. The value of the modularity index can vary between $[-1, 1]$, a high value of modularity corresponds to a structure where there are dense connections between the nodes within communities but sparse connections between nodes in different communities. That means that there are many sub communities of authors that write together but not across communities. Two further points are aimed to be explored in this work, (1) which people control the flow of information; as in which people influence the field the most and (2) which are the authors that gain the most from the influence of the field. To measure these concepts centrality measures are going to be used. Centrality measures are often used to understand different aspects of social networks [25]. Two centrality measures have been chosen for this paper and these are closeness and betweenness centrality.

1. In networks some nodes have a short distance to a lot of nodes and consequently are able to spread information on the network very effectively. A representative of this idea is **closeness centrality**, where a node is seen as centrally involved in the network if it requires only few intermediaries to contact others and thus is structurally relatively independent. Here, this is interpreted as a influence. Authors with a high value of closeness centrality, are the authors that spread scientific knowledge easier on the network and they have high influence.
2. Another centrality measure is the **betweenness centrality**, where the determination of an author's centrality is based on the quotient of the number of all shortest paths between nodes in the network that include the node in question and the number of all shortest paths in the network. In betweenness centrality the position of the node matters. Nodes with a higher value of betweenness centrality are located in positions that a lot of information pass through, this is interpreted as the gain from the influence, thus these authors gain the most from their networks.

In the next sections the methodology discussed here is applied to the data set and results regarding the publications and collaborativeness of the field are presented.

3 Initial Analysis

The data set [15] consists of 2422 articles with unique titles. In case of duplicates the preprint version of an article (collected from arXiv) was dropped. Of these 2422 articles, 76 have not been collected from the aforementioned APIs but have been manually added because they are of interest to the authors. A similar approach was used in [27] where a number of articles were manually added to the data set. A more detailed summary of the articles' provenance is given by Table 1. Only 3% of the data set consists of articles that were manually added and 27% of the articles were collected from arXiv.

provenance	Number of Articles	Percentage%
IEEE	294	12.14%
Manual	76	3.14%
Nature	436	18.00%
PLOS	477	19.69%
Springer	533	22.01%
arXiv	654	27.00%

Table 1: Articles' provenance for the main data set [15].

The average number of publications per year is denoted as $\mu_P = \frac{N_A}{N_Y}$ where N_A is the total number of articles and N_Y is the years of publication. The years of publication is calculated as the range between the collection date and the first published article, for each provenance, within the data. These averages are summarised in Table 2. Overall an average of 43 articles are published per year on the topic. The most significant contribution to this appears to be from arXiv with 11 articles per year, followed by Springer with 9 and PLOS with 8. Though the average publication offers insights about the publications of the fields, it remains a constant number. The data handled here is a time series starting in 1950s, when the game was introduced, until 2018 (Figure 1). Two observations can be made from Figure 1.

	μ_p
Overall	43.0
IEEE	5.0
Nature	8.0
PLOS	8.0
Springer	9.0
arXiv	11.0

Table 2: Average yearly publication (μ_P) for main data set [15].

1. There is a steady increase to the number of publications since the 1980s and the introduction of computer tournaments [3].
2. There is a decrease in 2017-2018. This is due to our data set being incomplete. Articles that have been written in 2017-2018 have either not being published or were not retrievable by the APIs at the time of writing this manuscript.

These observations can be confirmed by studying the time series. Using [23], an exponential distribution is fitted to the data from 1980-2016 (Figure 2). The fitted model can be used to forecast the behaviour of the field for the next 5 years. The forecasted periods are plotted in Figure 3. The time series has indicated a slight decrease however the model forecasts that the number of publications will keep increasing, thus demonstrating that the field of the IPD continues to attract academic attention.

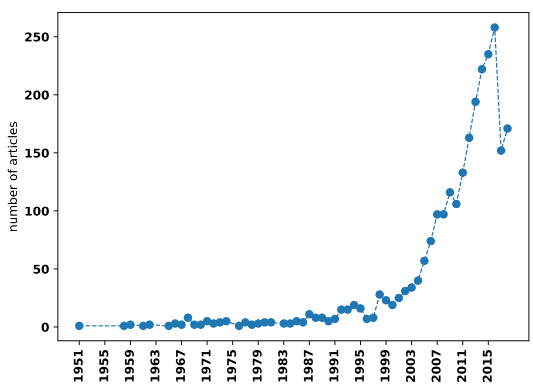


Figure 1: Line plot; number of articles published on the PD 1950-2019.

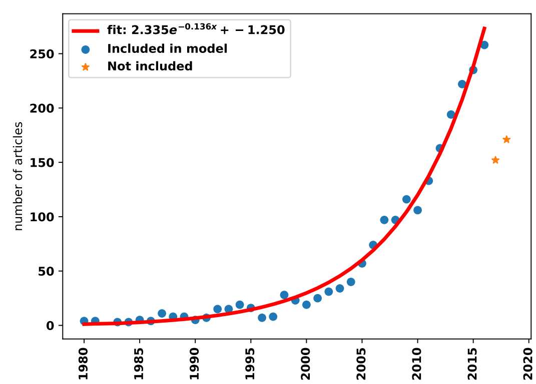


Figure 2: Scatter plot; number of articles published on the PD 1980-2019.

The overall collaboration index (CI) or the average number of authors on multi-authored papers is 3.2. Thus on average a non single author publication in the PD has 3 authors. The interdisciplinary field of cultural evolution was also reported to have a CI of 3 [44] and from a blog post published in Nature [28] it can be seen that other fields such as Astronomy and Astrophysics, Genetics and Heredity, Nuclear and Particle Physics have most likely 3 authors. The CI index over time is also given in Figure 4. There are some peaks in the early years 1969 and 1980, however, a steady increase appears to happen after 2004. This could be an effect of better communication tools being introduced around that time which enabled more collaborations between researchers. The collaborative behaviour will be explored in more details in the following section using the co-authorship network.

There are a total of 4226 unique authors in [15] and several of them have had multiple publications collected. The highest number of articles collected for an author is 83 publications for Matjaz Perc. The logged distribution of number of papers per author is given by Figure 5 and it can be seen that Matjaz Perc is an outlier. Mainly the authors have had 1 to 6 publications gathered by the data collection.

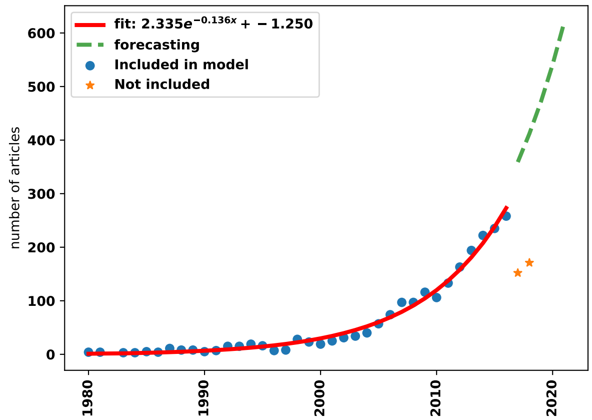


Figure 3: Forecast for 2017-2022.

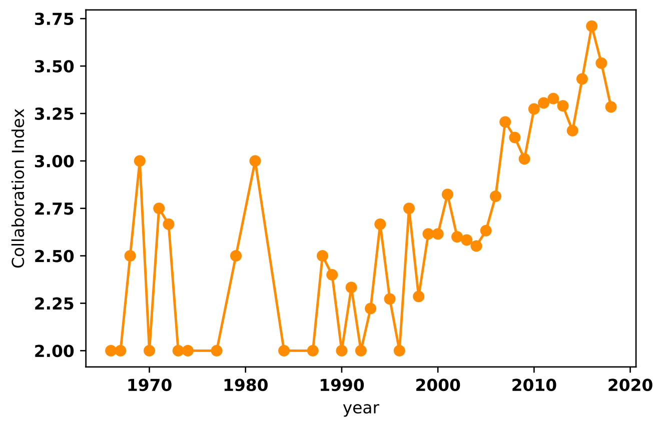


Figure 4: Collaboration index over time.

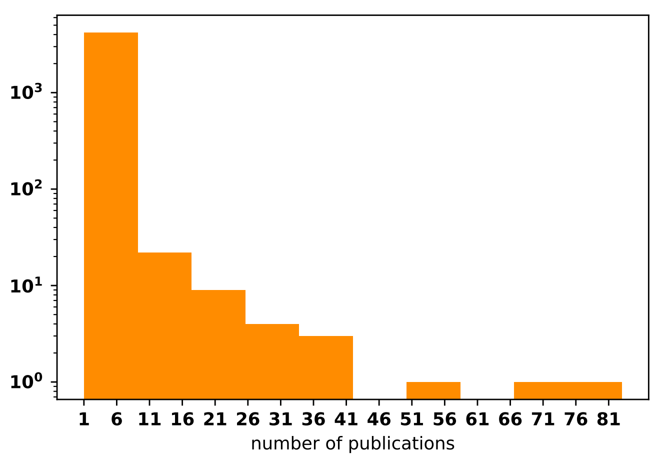


Figure 5: Logged distribution of number of papers per author.

Topic modeling has been applied to the data set using [37] and the 2422 articles have been classified into 5 different topics. The number of total topics was chosen based on the coherence value and the minimisation of overlapping words within two topics. The keywords associated with a topic, the most representative article of the given topic based on the percentage contribution and its academic reference are given by Table 3. The topics are labelled as A, B, C, D and E, and based on the keywords and the most represented articles appear to be about:

- A: human subject research
- B: biological studies
- C: strategies and agent based simulations
- D: evolutionary dynamics on networks
- E: modeling problems as a PD

The number of articles, and their equivalent percentage, assigned to each topic are also given in Table 3. Most of the topics have been assigned 500-560 articles or 20% of the articles, except topic B which has 309 entries assigned to it and is the topic with the smallest number of articles. The number of articles per topic over the years are given by Figure 6, however, no much can be concluded regarding the temporal change of the topics.

Dominant Topic	Topic Keywords	Article Title	References	Number of Documents	Percentage of Documents
A	social, behavior, human, study, experiment, cooperative, cooperation, suggest, find, behaviour	Facing Aggression: Cues Differ for Female versus Male Faces	[14]	496.0	0.2008
B	individual, group, good, show, high, increase, punishment, cost, result, benefit	Genomic and Gene-Expression Comparisons among Phage-Resistant Type-IV Pilus Mutants of Pseudomonas syringae pathovar phaseolicola	[40]	309.0	0.1251
C	game, strategy, player, agent, dilemma, play, payoff, state, prisoner, equilibrium	Fingerprinting: Visualization and Automatic Analysis of Prisoner's Dilemma Strategies	[40]	561.0	0.2271
D	cooperation, network, population, evolutionary, evolution, interaction, dynamic, structure, cooperator, study	Influence of initial distributions on robust cooperation in evolutionary Prisoner's Dilemma	[7]	556.0	0.2251
E	model, theory, base, system, problem, paper, propose, information, provide, approach	Gaming and price spikes in electric power markets and possible remedies	[19]	548.0	0.2219

Table 3: Keywords for each topic and the document with the most representative article for each topic.

To gain a better understanding regarding the number of topics over the years, a topic modeling algorithm is applied to the cumulative data set over 8 periods. These periods are 1951-1965, 1951-1973, 1951-1980, 1951-1988, 1951-1995, 1951-2003, 1951-2010, 1951-2018. The number of topics for each cumulative subset is chosen based on the coherence value and only. The overlapping of keywords has not been manually minimised. As a result, the period 1951-2018 has been assigned 6 topics instead of 5. Table 6 gives the topics, their keywords, the number of articles and their respective percentage for each period. For period 1951 to 1965 the topics are labelled as 1 to 5 indicating that a total number of 5 topics was selected for that period.

There are several changes in the keywords over time, which is only natural as the scientific language within a research field itself changes, and equivalently the number of topics change as well. For example the word 'sex', referring to the male and female sex, does not appear in any keywords after 1988. This implies that studying the effect of a participant's sex when playing an PD became less frequently and possibly less important. The word 'evolution' is a keyword of topic 1 in the period 1951-1995. Topic 1 is the largest topic of that period, implying that researchers were starting to publish more on evolution. The word network is a keyword for the first time following 2003. Suggesting that applications of the PD in social and computer networks became more relevant after 2003.

In the following section the collaborative behaviour of authors in the field, and within the field's topics as were presented in this section, are explored using a network theoretic approach.

Period	Topics	Topic Keywords	Num of Documents	Percentage of Documents
1951-1965	1	problem, technology, divert, euler, subsystem, requirement, trace, technique, system, untried	3	0.375
1951-1965	2	interpret, requirement, programme, evolution, article, increase, policy, system, trace, technology	2	0.25
1951-1965	3	equipment, agency, conjecture, development, untried, programme, trend, technology, weapon, technique	1	0.125
1951-1965	4	variation, celebrated, trend, untried, change, involve, month, technique, subsystem, research	1	0.125
1951-1965	5	give, good, modern, trace, technique, ambiguity, problem, trend, technology, system	1	0.125
1951-1973	1	study, shock, cooperative, money, part, vary, investigate, good, receive, equipment	12	0.3243
1951-1973	2	cooperation, level, significantly, sequence, reward, provoke, descriptive, principal, display, argue	4	0.1081
1951-1973	3	player, make, effect, triad, experimental, motivation, dominate, hypothesis, instruction, trend	3	0.0811
1951-1973	4	ss, sex, male, female, dyad, design, suggest, college, factor, tend	3	0.0811
1951-1973	5	result, research, format, change, operational, analysis, relate, understanding, decision, money	2	0.0541
1951-1973	6	condition, give, high, treatment, conflict, cc, real, original, replication, promote	2	0.0541
1951-1973	7	group, competitive, show, interpret, scale, compete, escalation, free, variable, individualistic	2	0.0541
1951-1973	8	outcome, strategy, choice, type, pdg, difference, dummy, conclude, compare, consistent	2	0.0541
1951-1973	9	game, difference, pair, approach, behavior, person, weapon, occur, advantaged, differential	2	0.0541
1951-1973	10	response, present, dilemma, influence, cooperate, bias, point, amount, participate, factor	2	0.0541
1951-1973	11	trial, problem, previous, involve, prisoner, experiment, follow, tit, increase, initial	1	0.027
1951-1973	12	matrix, behavior, rational, black, model, research, broad, distance, complex, trace	1	0.027
1951-1973	13	play, finding, individual, noncooperative, white, nature, race, ratio, represent, prisoner	1	0.027
1951-1980	1	play, trial, group, follow, white, interpret, scale, black, trend, small	14	0.25
1951-1980	2	outcome, level, effect, type, dyad, vary, pdg, participate, understanding, arise	9	0.1607
1951-1980	3	game, strategy, cooperation, significant, difference, sentence, text, occur, differential, hypothesis	4	0.0714
1951-1980	4	male, female, find, result, sex, subject, experimental, situation, treatment, computer	4	0.0714
1951-1980	5	research, problem, influence, matrix, format, model, analysis, year, crime, equipment	4	0.0714
1951-1980	6	condition, dilemma, bias, free, attempt, book, year, dummy, prison, design	4	0.0714
1951-1980	7	variable, result, factor, individual, ability, triad, half, migration, change, investigate	3	0.0536
1951-1980	8	show, present, suggest, rational, compete, approach, characteristic, examine, person, conduct	3	0.0536
1951-1980	9	behavior, high, finding, relate, obtain, assistance, ratio, good, weapon, competition	3	0.0536
1951-1980	10	ss, shock, money, competitive, part, difference, pair, amount, man, information	3	0.0536
1951-1980	11	player, conflict, theory, decision, determine, produce, maker, cooperate, specialist, programming	2	0.0357
1951-1980	12	study, prisoner, make, response, experiment, noncooperative, standard, separate, conclude, initial	2	0.0357
1951-1980	13	give, cooperative, choice, cognitive, real, operational, set, subject, ascribe, concern	1	0.0179
1951-1988	1	trial, difference, find, choice, significant, competitive, effect, triad, interact, occur	24	0.2553
1951-1988	2	ss, shock, money, pair, response, part, high, tit, receive, amount	13	0.1383
1951-1988	3	suggest, paper, case, debate, view, achieve, framework, natural, assumption, finitely	10	0.1064
1951-1988	4	prisoner, dilemma, behavior, model, present, involve, person, increase, trust, experiment	8	0.0851
1951-1988	5	game, player, show, approach, repeat, previous, move, tat, related, include	8	0.0851
1951-1988	6	cooperation, level, mutual, equilibrium, standard, provide, information, human, real, question	6	0.0638
1951-1988	7	play, result, male, subject, female, cooperative, sex, experimental, treatment, computer	5	0.0532
1951-1988	8	research, study, variable, ability, factor, conflict, matrix, year, student, interpret	4	0.0426
1951-1988	9	problem, group, small, scale, social, issue, large, base, bias, party	4	0.0426
1951-1988	10	game, strategy, outcome, type, cooperate, ethical, pdg, explain, dependent, separate	4	0.0426
1951-1988	11	give, condition, individual, major, dyad, behaviour, produce, conflict, assistance, collectively	3	0.0319
1951-1988	12	situation, iterate, statement, rational, card, side, paradox, true, consequence, front	2	0.0213
1951-1988	13	inflation, hypothesis, rate, run, change, demand, nominal, cost, output, growth	2	0.0213
1951-1988	14	theory, make, analysis, decision, system, examine, work, soft, lead, hard	1	0.0106
1951-1995	1	strategy, population, evolution, iterate, tit, opponent, evolve, dynamic, set, tat	31	0.1732
1951-1995	2	game, repeat, assumption, rule, person, equilibrium, general, finitely, indefinitely, analyze	24	0.1341
1951-1995	3	inflation, long, rate, hypothesis, run, policy, cost, nominal, demand, programming	20	0.1117
1951-1995	4	condition, outcome, trial, find, difference, cooperation, experiment, level, significant, response	15	0.0838
1951-1995	5	rational, result, receive, statement, money, paradox, shock, iterate, consequence, common	14	0.0782
1951-1995	6	cooperation, show, competitive, high, probability, conflict, simulation, altruism, yield, natural	14	0.0782
1951-1995	7	prisoner, dilemma, give, point, defect, form, cooperator, increase, relate, ethical	10	0.0559
1951-1995	8	player, give, decision, provide, cooperative, game, previous, pair, determine, interact	9	0.0503
1951-1995	9	play, cooperate, result, male, subject, female, time, relationship, suggest, student	8	0.0447
1951-1995	10	problem, group, theory, good, approach, society, large, scale, issue, level	8	0.0447
1951-1995	11	study, situation, behaviour, computer, argue, change, implication, characteristic, real, associate	8	0.0447
1951-1995	12	model, paper, behavior, examine, present, mutual, expectation, develop, type, variable	7	0.0391
1951-1995	13	make, research, system, analysis, choice, work, base, relation, world, wide	6	0.0335
1951-1995	14	individual, social, behavior, standard, choose, evolutionary, partner, payoff, defection, small	5	0.0279
1951-2003	1	game, player, dilemma, prisoner, theory, give, paper, make, group, problem	151	0.4266
1951-2003	2	cooperation, result, play, show, cooperate, condition, cooperative, high, level, time	106	0.2994
1951-2003	3	strategy, model, agent, study, behavior, individual, population, evolutionary, state, player	97	0.274
1951-2010	1	model, theory, paper, base, make, present, problem, provide, human, decision	325	0.3454
1951-2010	2	game, strategy, player, agent, play, dilemma, system, behavior, show, state	322	0.3422
1951-2010	3	cooperation, network, study, population, individual, evolutionary, social, evolution, interaction, structure	294	0.3124
1951-2018	1	game, theory, model, provide, make, problem, present, give, framework, equilibrium	462	0.1871
1951-2018	2	social, behavior, human, experiment, study, suggest, result, decision, participant, make	430	0.1742
1951-2018	3	base, model, information, time, paper, propose, learn, method, system, work	418	0.1693
1951-2018	4	strategy, player, game, agent, play, payoff, state, result, iterate, show	323	0.1308
1951-2018	5	cooperation, individual, group, cooperative, level, mechanism, good, punishment, cost, social	308	0.1247
1951-2018	6	population, model, evolutionary, evolution, result, selection, show, environment, interaction, dynamic	279	0.113
1951-2018	7	network, cooperation, structure, interaction, cooperator, dynamic, evolutionary, evolution, rule, simulation	249	0.1009

Table 4: Topic modeling result for the cumulative data set over the periods

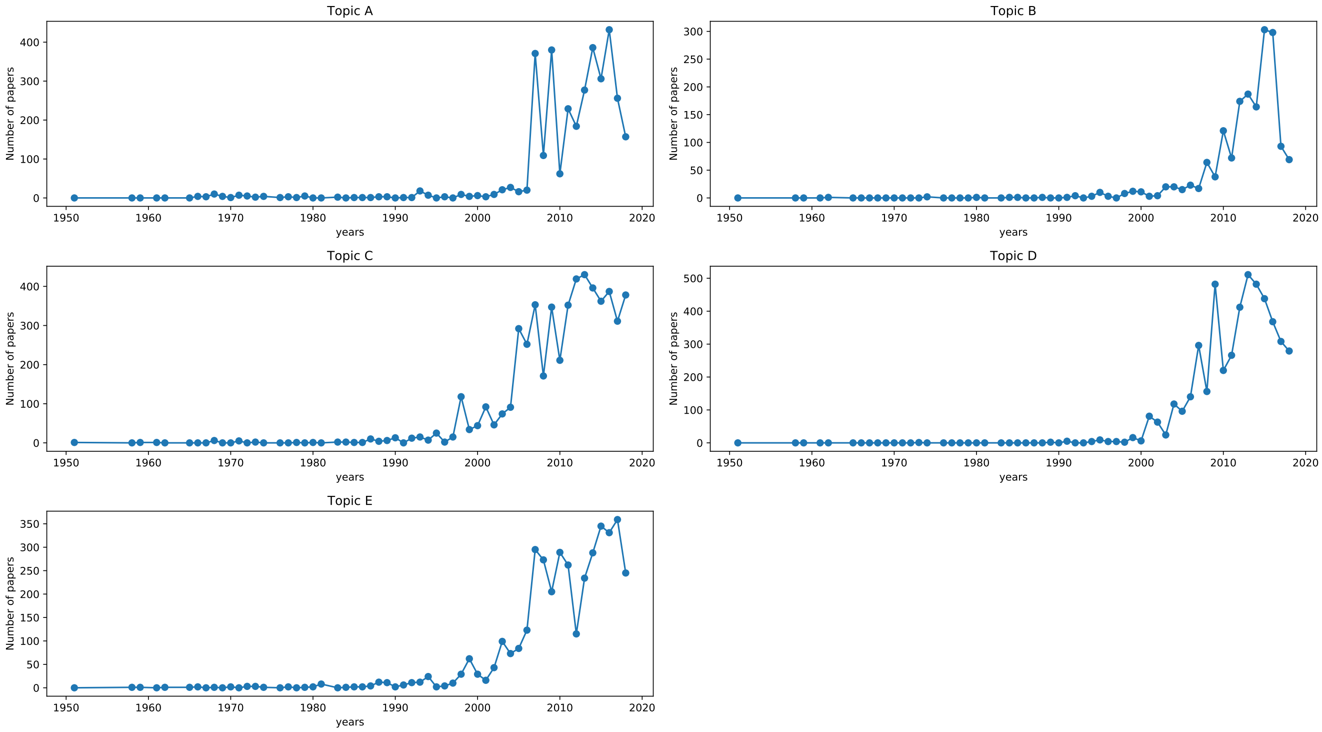


Figure 6: Number of articles per topic over the years.

4 Analysis of co-authorship network

In this section the collaborative behaviour of authors in the field of the IPD is assessed using the co-authorship network. As mentioned in Section 2 the co-authorship network is denoted as G. There are a total of 947 connected components in G and the largest component has a size of 796 nodes. The largest connected component is going to be refereed to as the main cluster of the network. The main cluster is denoted as $\bar{G}$ and a graphical representation is shown in Figure 7b. The metrics for both G and $\bar{G}$ are given by Table 5.

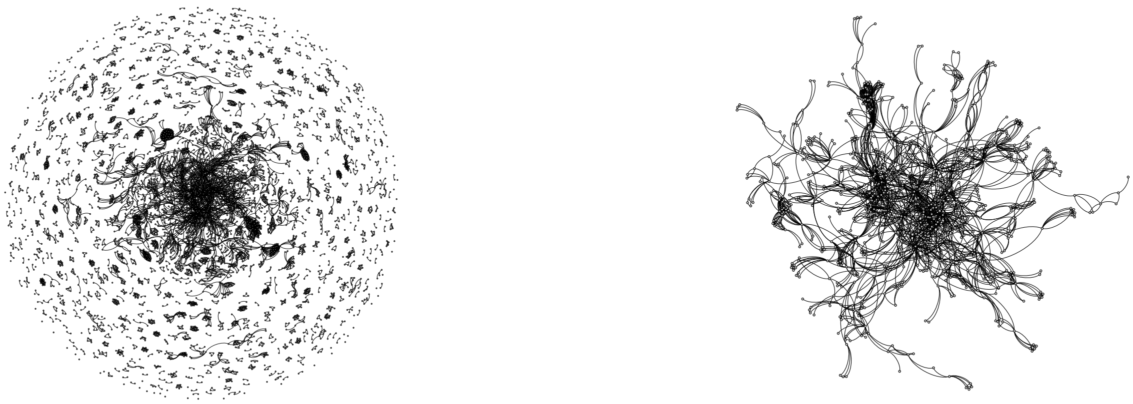
The networks have a high modularity and a large number of communities. A high modularity implies that authors create their own publishing communities but not many publications from authors from different communities occur. However, the communities of the network are collaborative. An author on average has 4 co-authors and 0.708 likelihood of collaborating with a collaborator's co-author. There are only 128 authors with single author publications, which corresponds to only the 3.2 % of authors in G. Moreover, in the PD an author from the main cluster is 7% more likely to write with a collaborator's co-author and on average has 2 more co-authors.

In [27] the collaborative metrics for the "evolution of cooperation" co-authorship network were also reported. Though their network is of smaller size (number of nodes $3670 < 4226$), the collaborative metrics are fairly similar between the two graphs (clustering coeff. 0.632 and modularity 0.95 close to 0.96), indicating that for the same field the same remarks can be made from a different co-authorship network.

	# Nodes	# Edges	% Isolated nodes	# Connected components	Size of largest component	Av. degree	# Communities	Modularity	Clustering coeff
G	4011	7642	3.2	947	796	3.811	967	0.96491	0.701
$\bar{G}$	796	2214	0.0	1	796	5.563	25	0.84406	0.773

Table 5: Network metrics for G and $\bar{G}$ respectively.

The change of the network over time is also studied by constructing the network cumulatively over 51 periods. Except from the first period 1951-1966 the rest of the periods have a yearly interval (data for the years 1975 and



(a) G the co-authorship network for the IPD.

(b) $\bar{G}$ the largest connected component of G.

1982 have not been collected). The metrics of each sub network are given by Table 6. This is also done for $\bar{G}$ over the same periods and the results are given by Table 7. Similar to the results of [27], it can been observed that the network G grows over time and that the network always has modularity. This is also confirmed by $\bar{G}$.

In Section 3 the articles of this work were classified into 5 different topics. The co-authorship networks for each topic have been constructed in the same way as G. Note that authors with publications in more than one topic exist. These authors have not been removed. The metrics of the five co-authorship networks are given by Table 8.

Though Topic B has the smallest number of articles and is the smallest network as well, with 695 authors, it has the highest average degree. The network of Topic A is the largest network, moreover, it has the smallest percentage of isolated nodes (1.3%) and an average degree of 3.8. The networks of topics E and C have a large number of isolated nodes, 5.9% and 4.6% respectively, a small clustering coefficient and Topic C has the smallest average degree. The authors with publications in Topics B and A appear be more collaborative, with authors having 4 co-authors and hardly writing at all, compared to Topics E and C. Topic D has the largest connected component though it is one of the smallest networks. The connected component of topic D is made up by the same authors as the main cluster $\bar{G}$, this will be clear in the next result.

	# Nodes	# Edges	# Isolated nodes	% Isolated nodes	# Connected components	Size of largest component	Av. degree	# Communities	Modularity	Clustering coeff
Topic A	1124	2137	15	1.3	264	56	3.802	265	0.983	0.759
Topic B	695	1382	13	1.9	157	80	3.977	158	0.950	0.773
Topic C	900	1141	41	4.6	281	29	2.536	281	0.981	0.636
Topic D	880	1509	17	1.9	174	312	3.430	183	0.918	0.701
Topic E	1045	1964	59	5.6	354	31	3.759	354	0.926	0.664

Table 8: Network metrics for topic networks.

Though the authors with publications in topics C, D and E have exhibited to be less collaborative it must be taken into account that authors had their publications assigned to multiple topics, and thus the edges of a node have been divided between the different networks. More specifically, a total of 1975 authors have articles assigned to multiple topics. Out of these 1975 authors, 55% of them had their publications assigned to two different topics, 30% in three, 10% in four and 5% in five, Figure 8. The probabilities of having publications being assigned to a given combination of topics based on the different number of topics are given by Table 9. Table 9 shows that authors are most likely to have their publications assigned to topics C and D, C and E (for $n = 2$), C, D and E (for $n = 3$) and A, C, D and E (for $n = 4$). Indeed the authors with publications in topics C, D and E could have

Period	# Nodes	# Edges	# Isolated nodes	% Isolated nodes	# Connected components	Size of largest component	Av. degree	# Communities	Modularity	Clustering coeff
1951 - 1966	6	3	0	0.0	3	2	1.000	3	0.667	0.000
1951 - 1967	8	4	0	0.0	4	2	1.000	4	0.750	0.000
1951 - 1968	19	15	0	0.0	8	5	1.579	8	0.684	0.228
1951 - 1969	20	17	0	0.0	8	6	1.700	8	0.630	0.250
1951 - 1970	22	18	0	0.0	9	6	1.636	9	0.667	0.227
1951 - 1971	33	28	0	0.0	13	6	1.697	13	0.827	0.424
1951 - 1972	39	34	0	0.0	15	6	1.744	15	0.867	0.513
1951 - 1973	42	35	1	2.4	17	6	1.667	17	0.873	0.476
1951 - 1974	42	35	1	2.4	17	6	1.667	17	0.873	0.476
1951 - 1976	42	35	1	2.4	17	6	1.667	17	0.873	0.476
1951 - 1977	44	36	1	2.3	18	6	1.636	18	0.880	0.455
1951 - 1978	44	36	1	2.3	18	6	1.636	18	0.880	0.455
1951 - 1979	47	40	1	2.1	18	6	1.702	18	0.884	0.454
1951 - 1980	47	40	1	2.1	18	6	1.702	18	0.884	0.454
1951 - 1981	50	46	1	2.0	18	6	1.840	18	0.889	0.497
1951 - 1983	51	46	2	3.9	19	6	1.804	19	0.889	0.487
1951 - 1984	53	47	2	3.8	20	6	1.774	20	0.894	0.469
1951 - 1985	53	47	2	3.8	20	6	1.774	20	0.894	0.469
1951 - 1986	53	47	2	3.8	20	6	1.774	20	0.894	0.469
1951 - 1987	56	48	3	5.4	22	6	1.714	22	0.898	0.443
1951 - 1988	62	52	4	6.5	25	6	1.677	25	0.909	0.449
1951 - 1989	75	62	5	6.7	31	6	1.653	31	0.926	0.424
1951 - 1990	79	64	5	6.3	33	6	1.620	33	0.930	0.403
1951 - 1991	87	69	6	6.9	37	6	1.586	37	0.937	0.400
1951 - 1992	95	72	10	10.5	42	6	1.516	42	0.941	0.367
1951 - 1993	106	81	12	11.3	47	6	1.528	47	0.947	0.366
1951 - 1994	124	95	16	12.9	56	6	1.532	56	0.955	0.394
1951 - 1995	135	102	17	12.6	61	6	1.511	61	0.960	0.384
1951 - 1996	142	105	18	12.7	65	6	1.479	65	0.962	0.365
1951 - 1997	155	115	20	12.9	71	6	1.484	71	0.966	0.392
1951 - 1998	191	140	21	11.0	87	6	1.466	87	0.973	0.367
1951 - 1999	221	169	25	11.3	99	6	1.529	99	0.977	0.397
1951 - 2000	250	195	27	10.8	110	6	1.560	110	0.979	0.418
1951 - 2001	287	235	30	10.5	125	7	1.638	125	0.977	0.419
1951 - 2002	335	278	36	10.7	146	7	1.660	146	0.979	0.428
1951 - 2003	381	310	40	10.5	168	7	1.627	168	0.982	0.413
1951 - 2004	437	370	40	9.2	185	10	1.693	185	0.983	0.424
1951 - 2005	532	476	41	7.7	214	19	1.789	214	0.985	0.458
1951 - 2006	640	603	43	6.7	246	22	1.884	246	0.987	0.486
1951 - 2007	793	877	46	5.8	283	25	2.212	283	0.985	0.532
1951 - 2008	948	1170	50	5.3	318	33	2.468	319	0.985	0.558
1951 - 2009	1108	1442	54	4.9	356	71	2.603	358	0.982	0.573
1951 - 2010	1300	1936	66	5.1	402	133	2.978	405	0.965	0.592
1951 - 2011	1560	2375	79	5.1	472	157	3.045	475	0.970	0.613
1951 - 2012	1837	2865	80	4.4	534	209	3.119	537	0.969	0.634
1951 - 2013	2149	3420	93	4.3	603	322	3.183	609	0.965	0.644
1951 - 2014	2481	3971	103	4.2	683	399	3.201	694	0.962	0.658
1951 - 2015	2938	4877	110	3.7	765	504	3.320	779	0.965	0.675
1951 - 2016	3469	6532	114	3.3	850	613	3.766	863	0.964	0.696
1951 - 2017	3735	7072	119	3.2	895	706	3.787	912	0.964	0.700
1951 - 2018	4011	7642	128	3.2	947	796	3.811	967	0.966	0.701

Table 6: Collaborativeness metrics for cumulative graphs, $\tilde{G} \subseteq G$.

Periods	# Nodes	# Edges	# Isolated nodes	% Isolated nodes	# Connected components	Size of largest component	Av. degree	# Communities	Modularity	Clustering coeff
1951 - 1966	2	1	0	0.0	1	2	1.000	1	0.000	0.000
1951 - 1967	2	1	0	0.0	1	2	1.000	1	0.000	0.000
1951 - 1968	5	8	0	0.0	1	5	3.200	1	0.000	0.867
1951 - 1969	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1970	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1971	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1972	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1973	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1974	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1976	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1977	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1978	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1979	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1980	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1981	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1983	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1984	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1985	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1986	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1987	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1988	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1989	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1990	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1991	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1992	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1993	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1994	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1995	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1996	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1997	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1998	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 1999	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 2000	6	10	0	0.0	1	6	3.333	2	0.020	0.833
1951 - 2001	7	21	0	0.0	1	7	6.000	1	0.000	1.000
1951 - 2002	7	21	0	0.0	1	7	6.000	1	0.000	1.000
1951 - 2003	7	21	0	0.0	1	7	6.000	1	0.000	1.000
1951 - 2004	10	13	0	0.0	1	10	2.600	2	0.376	0.553
1951 - 2005	19	28	0	0.0	1	19	2.947	3	0.544	0.730
1951 - 2006	22	35	0	0.0	1	22	3.182	4	0.527	0.720
1951 - 2007	25	39	0	0.0	1	25	3.120	5	0.558	0.686
1951 - 2008	33	62	0	0.0	1	33	3.758	4	0.623	0.736
1951 - 2009	71	148	0	0.0	1	71	4.169	6	0.697	0.698
1951 - 2010	133	387	0	0.0	1	133	5.820	7	0.726	0.749
1951 - 2011	157	465	0	0.0	1	157	5.924	8	0.727	0.725
1951 - 2012	209	611	0	0.0	1	209	5.847	11	0.733	0.737
1951 - 2013	322	892	0	0.0	1	322	5.540	12	0.780	0.743
1951 - 2014	399	1109	0	0.0	1	399	5.559	15	0.794	0.742
1951 - 2015	504	1368	0	0.0	1	504	5.429	24	0.811	0.751
1951 - 2016	613	1677	0	0.0	1	613	5.471	21	0.819	0.761
1951 - 2017	706	1935	0	0.0	1	706	5.482	29	0.830	0.772
1951 - 2018	796	2214	0	0.0	1	796	5.563	25	0.845	0.773

Table 7: Collaborativeness metrics for cumulative graphs' main clusters, $\tilde{G} \subseteq \bar{G}$.

suffered from their connections being divided into different networks, and thus the respective networks could be more collaborative than the analysis has suggested.

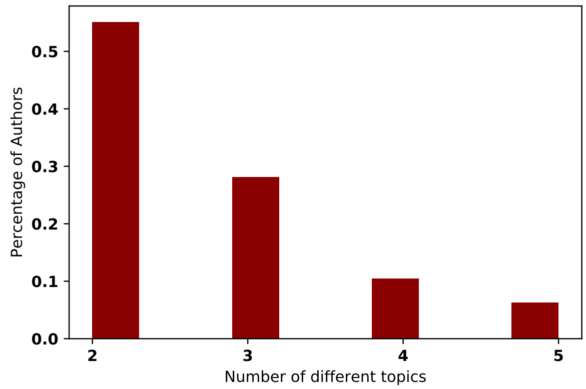


Figure 8: Distribution of number of different topics authors have written in excluding single topic.

Number of different topics authors have publications n	Topics	Probabilities
$n = 2$	B - E	0.04
	C - D	0.24
	C - E	0.15
	A - B	0.08
	D - E	0.13
	A - E	0.07
	B - C	0.03
	B - D	0.07
	A - D	0.12
	A - C	0.07
$n = 3$	A - D - E	0.10
	C - D - E	0.29
	A - B - C	0.04
	A - C - D	0.13
	B - C - D	0.11
	A - B - E	0.03
	A - B - D	0.10
	B - C - E	0.04
	A - C - E	0.09
	B - D - E	0.06
$n = 4$	A - C - D - E	0.35
	A - B - C - D	0.14
	B - C - D - E	0.33
	A - B - D - E	0.13
	A - B - C - E	0.06
$n = 5$	A - B - C - D - E	1.00

Table 9: Probabilities of an author writing in any combinations of topics based on the number of topics the author has publications in.

The next result discussed here are on centrality measures. As a reminder, two centrality measures are reported, these are the closeness centrality and the betweenness centrality. Closeness centrality is a measure of how easy it is for an author to contact others, and consequently affect them; influence them. Thus closeness centrality here is a measure of influence. Betweenness centrality is a measure of how many paths pass through a specific node, thus the amount of information this person has access to. Betweenness centrality is used here as a measure of how much an author gains from the field. All centrality measure can have values ranging from 0 to 1.

For G and $\bar{G}$ the most central authors based on closeness and betweenness centralities are given by Table 10. For G the most central author, based on both measures, is Matjaz Perc. A few authors appear to be central based on one measure and not the other. Martin Nowak, Franz Weissing, Jianye Hao, Angel Sanchez and Valerio Capraro are authors that have access to information due to their positioning but do not influence the network as much. The opposite is true for the authors Attila Szolnoki, Luo-Luo Jiang Sandro Meloni, Cheng-Yi Xia and Xiaojie Chen.

Overall the values of both measures are low. The medians of the distributions, given in Figures 9 and 10, are both zero. From Figure 10 it is evident that some authors have higher values of closeness centrality and thus do influence the network. These can be verified to be the authors from the main cluster. The most central authors in $\bar{G}$ are the same people as in G. This implies that the results on centrality heavily rely of the main cluster.

In conclusion the authors of the PD do not gain much from the influence of their fields. This can even be verified by the topics networks. Table 11 the authors that gain the most from their respective topics and the values are once again low. In some cases even zero (Topic A, C and E). In relation to authors influencing their field. There are authors in G that do influence the network. The group of people that do are the authors of the main cluster. These authors even gain from being in the main cluster and part of this "main" group of authors. From Table 12 it can be seen that Topic D has the highest values of closeness centrality and this is because the authors are a subset of $\bar{G}$. In comparison the other networks have very low values of centrality and the results are verified further by the topics' networks.

	G				$\bar{G}$			
	Name	Betweenness	Name	Closeness	Name	Betweenness	Name	Closeness
1	Matjaz Perc	0.015	Matjaz Perc	0.066	Matjaz Perc	0.373	Matjaz Perc	0.330
2	Zhen Wang	0.011	Long Wang	0.060	Zhen Wang	0.279	Long Wang	0.301
3	Long Wang	0.007	Yamir Moreno	0.059	Long Wang	0.170	Yamir Moreno	0.299
4	Martin Nowak	0.006	Attila Szolnoki	0.059	Martin Nowak	0.159	Attila Szolnoki	0.297
5	Angel Sanchez	0.004	Zhen Wang	0.059	Angel Sanchez	0.114	Zhen Wang	0.296
6	Yamir Moreno	0.004	Arne Traulsen	0.056	Yamir Moreno	0.110	Arne Traulsen	0.281
7	Arne Traulsen	0.004	Luo-Luo Jiang	0.055	Arne Traulsen	0.107	Luo-Luo Jiang	0.280
8	Franz Weissing	0.004	Sandro Meloni	0.055	Franz Weissing	0.101	Sandro Meloni	0.278
9	Jianye Hao	0.004	Cheng-Yi Xia	0.055	Jianye Hao	0.094	Cheng-Yi Xia	0.276
10	Valerio Capraro	0.004	Xiaojie Chen	0.055	Valerio Capraro	0.093	Xiaojie Chen	0.276

Table 10: 10 most central authors based on betweenness and closeness centralities for G and $\bar{G}$.

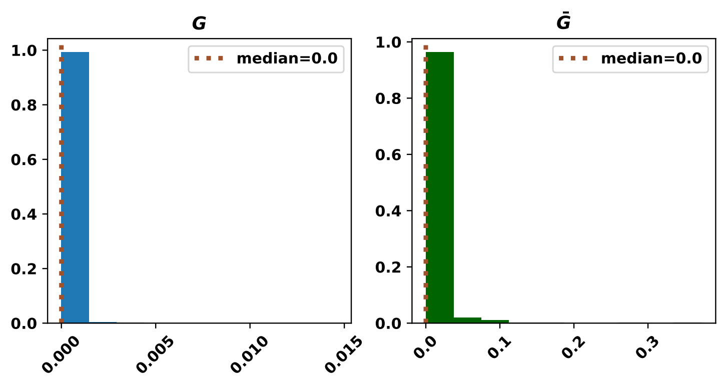


Figure 9: Distributions of betweenness centrality in G and $\bar{G}$

	Topic A		Topic B		Topic C		Topic D		Topic E	
	Name	Betweeness	Name	Betweeness	Name	Betweeness	Name	Betweeness	Name	Betweeness
1	David Rand	0.002	Long Wang	0.006	Daniel Ashlock	0.001	Matjaz Perc	0.064	Zengru Di	0.0
2	Valerio Capraro	0.001	Luo-Luo Jiang	0.005	Matjaz Perc	0.000	Luo-Luo Jiang	0.037	Jian Yang	0.0
3	Angel Sanchez	0.001	Martin Nowak	0.004	Karl Tuyls	0.000	Yamir Moreno	0.031	Yevgeniy Vorobeychik	0.0
4	Feng Fu	0.001	Matjaz Perc	0.003	Philip Hingston	0.000	Christoph Hauert	0.027	Otavio Teixeira	0.0
5	Martin Nowak	0.000	Attila Szolnoki	0.003	Eun-Youn Kim	0.000	Long Wang	0.024	Roberto Oliveira	0.0
6	Nicholas Christakis	0.000	Christian Hilbe	0.002	Wendy Ashlock	0.000	Zhen Wang	0.024	M. Nowak	0.0
7	Pablo Branas-Garza	0.000	Yamir Moreno	0.002	Attila Szolnoki	0.000	Han-Xin Yang	0.023	M. Harper	0.0
8	Toshio Yamagishi	0.000	Xiaojie Chen	0.002	Seung Baek	0.000	Martin Nowak	0.020	Xiao Han	0.0
9	James Fowler	0.000	Arne Traulsen	0.002	Martin Nowak	0.000	Angel Sanchez	0.017	Zhesi Shen	0.0
10	Long Wang	0.000	Zhen Wang	0.002	Thore Graepel	0.000	Zhihai Rong	0.016	Wen-Xu Wang	0.0

Table 11: 10 most central authors based on betweenness centrality for topics' networks.

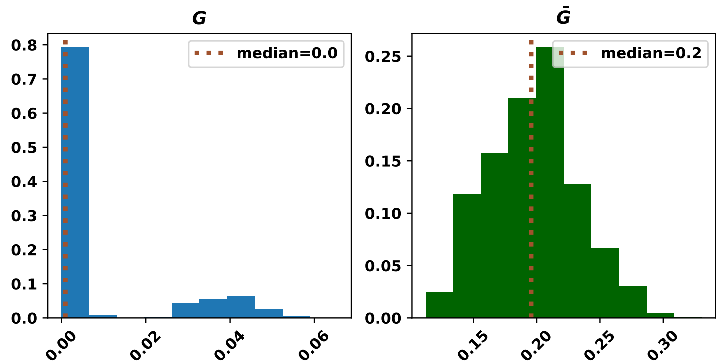


Figure 10: Distributions of closeness centrality in G and $\bar{G}$

	Topic A		Topic B		Topic C		Topic D		Topic E	
	Name	Closeness	Name	Closeness	Name	Closeness	Name	Closeness	Name	Closeness
1	David Rand	0.027	Long Wang	0.043	Karl Tuyls	0.022	Matjaz Perc	0.123	Stefanie Widder	0.029
2	Valerio Capraro	0.023	Matjaz Perc	0.041	Thore Graepel	0.019	Zhen Wang	0.109	Rosalind Allen	0.029
3	Jillian Jordan	0.022	Attila Szolnoki	0.040	Joel Leibo	0.018	Long Wang	0.107	Thomas Pfeiffer	0.029
4	Nicholas Christakis	0.021	Martin Nowak	0.040	Edward Hughes	0.017	Yamir Moreno	0.105	Thomas Curtis	0.029
5	James Fowler	0.020	Olivier Tenaillon	0.038	Matthew Phillips	0.017	Luo-Luo Jiang	0.104	Carsten Wiuf	0.029
6	Martin Nowak	0.020	Xiaojie Chen	0.038	Edgar Duenez-Guzman	0.017	Attila Szolnoki	0.103	William Sloan	0.029
7	Angel Sanchez	0.019	Bin Wu	0.038	Antonio Castaneda	0.017	Gyorgy Szabo	0.102	Otto Cordero	0.029
8	Gordon Kraft-Todd	0.019	Yanling Zhang	0.037	Iain Dunning	0.017	Xiaojie Chen	0.102	Sam Brown	0.029
9	Akihiro Nishi	0.019	Feng Fu	0.037	Tina Zhu	0.017	Guangming Xie	0.101	Babak Momeni	0.029
10	Anthony Evans	0.019	David Rand	0.037	Kevin Mckee	0.017	Lucas Wardil	0.101	Wenying Shou	0.029

Table 12: 10 most central authors based on closeness centrality for topics' networks.

5 Conclusion

This manuscript has explored the number of publications, the authors' collaborative behaviour and their influence in the research topic of the Iterated Prisoner's Dilemma. This was achieved by applying network theoretic approaches and a bibliometric analysis in a data containing more than 2000 publications. The data set was automatically collected from five different sources using a bespoke piece of software written for this purpose [16].

The data collection and an introduction to the methodology used in this work were covered in Section 2. The data set contains a total of 2422 articles, it has been archive and it available in [15] for further analysis. Section 3 covered an initial analysis of the data set, and applied a topic modeling algorithm which classified the articles into topics. The initial analysis demonstrated that the field of the Prisoner's Dilemma remains a prominent field with several papers being published in journals. The topic modeling analysis identified five different topics which appeared to be human subject research, biological studies, strategies and agent based simulations, evolutionary dynamics on networks and modeling problems as a PD. A temporal topic modeling analysis showed that over time not only the number of topic changed, but also the scientific language and the most important the topics authors were publishing on.

Following Section 3, the collaborative behaviour of the field was explored in Section 4. It was concluded that the field of the Iterated Prisoner's Dilemma is a collaborative field where authors are likely to write with a collaborator's co-authors and on average an author has 4 co-authors. The results on collaborativeness were verified when studying the co-authorship network for each of the five topics defined in Section 3 as well. Exploring the influence of authors and their gain from being in the network of the field demonstrated that authors do not gain much, and the authors with influence are the ones connected to the main cluster, to a "main" group of authors.

The study of the Prisoner's Dilemma is the study of cooperation and investigating the cooperative behaviours of authors is what this work has aimed to achieve. Interesting areas of future work would include extending this analysis to more game theoretic sub fields, to evaluate whether the results remain the same.

6 Acknowledgements

A variety of software have been used in this work:

- The Axelrod library for IPD simulations [2].
- The Matplotlib library for visualisation [21].
- The Numpy library for data manipulation [42].
- The Networkx [20] package for analysing networks.
- Gephi [4] open source package for visualising networks.
- The Gensim library for the topic modeling [37].
- The louvain library for calculating the networks modularity `https://github.com/taynaud/python-louvain/issues`.

References

[1] PLOS public library of science. https://www.plos.org/.

[2] The Axelrod project developers . Axelrod: v.4.5.0, April 2016.

[3] Robert Axelrod and William D Hamilton. The evolution of cooperation. *science*, 211(4489):1390–1396, 1981.

[4] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks, 2009.

[5] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[6] V. D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[7] Xiaojie Chen, Feng Fu, and Long Wang. Influence of initial distributions on robust cooperation in evolutionary prisoner's dilemma. *arXiv preprint physics/0701318*, 2007.

[8] A. Clauset, M. E.J. Newman, and C. Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.

[9] R. das Neves Machado, B. Vargas-Quesada, and J. Leta. Intellectual structure in stem cell research: exploring brazilian scientific articles from 2001 to 2010. *Scientometrics*, 106(2):525–537, 2016.

[10] Michael Doebeli and Christoph Hauert. Models of cooperation based on the prisoner's dilemma and the snowdrift game. *Ecology letters*, 8(7):748–766, 2005.

[11] D. Easley, J. Kleinberg, et al. *Networks, crowds, and markets*, volume 8. Cambridge university press Cambridge, 2010.

[12] H. Etzkowitz. Individual investigators and their research groups. *Minerva*, 30(1):28–50, 1992.

[13] M. M. Flood. Some experimental games. *Management Science*, 5(1):5–26, 1958.

[14] Shawn N Geniole, Amanda E Keyes, Catherine J Mondloch, Justin M Carré, and Cheryl M McCormick. Facing aggression: Cues differ for female versus male faces. *PLOS one*, 7(1):e30366, 2012.

[15] N. E. Glynatsi. Articles' meta data on the prisoner's dilemma. https://doi.org/10.5281/zenodo.3406536, September 2019.

[16] N. E. Glynatsi and V. Knight. Nikoleta-v3/arcas: Arcas v 0.0.4, December 2017.

[17] Justin Grimmer and Brandon M Stewart. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3):267–297, 2013.

[18] Nature Publishing Group. Nature. https://www.nature.com/, 1869.

[19] Xiaohong Guan. Gaming and price spikes in electric power markets and possible remedies. In *Proceedings. International Conference on Power System Technology*, volume 1, pages 188–vol. IEEE, 2002.

[20] A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA, August 2008.

[21] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

[22] IEEE. Ieee xplore digital library. http://ieeexplore.ieee.org/Xplore/home.jsp.

[23] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed ¡today¿].

[24] S. Kyvik and I. Reymert. Research collaboration in groups and networks: differences across academic fields. *Scientometrics*, 113(2):951–967, 2017.

[25] A. Landherr, B. Friedl, and J. Heidemann. A critical review of centrality measures in social networks. *Business & Information Systems Engineering*, 2(6):371–385, Dec 2010.

[26] Weihua Li, Tomaso Aste, Fabio Caccioli, and Giacomo Livan. Early coauthorship with top scientists predicts success in academic careers. *Nature Communications*, 10(1):2041–1723, 2019.

[27] P. Liu and H. Xia. Structure and evolution of co-authorship network in an interdisciplinary research field. *Scientometrics*, 103(1):101–134, Apr 2015.

[28] Smriti Mallapaty. Paper authorship goes hyper. `https://www.natureindex.com/news-blog/paper-authorship-goes-hyper`, 2018.

[29] G. McKiernan. arxiv. org: the los alamos national laboratory e-print server. *International Journal on Grey Literature*, 1(3):127–138, 2000.

[30] Mannheim Media. Springer publishing. http://www.springer.com/, 1950.

[31] R. Van Noorden. The science that's never been cited. `https://www.nature.com/articles/d41586-017-08404-0`, 2017.

[32] N. Nurseitov, M. Paulson, R. Reynolds, and C. Izurieta. Comparison of json and xml data interchange formats: a case study. *Caine*, 2009:157–162, 2009.

[33] Matjaž Perc and Attila Szolnoki. Social diversity and promotion of cooperation in the spatial prisoner's dilemma game. *Physical Review E*, 77(1):011904, 2008.

[34] William H Press and Freeman J Dyson. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413, 2012.

[35] A. Pritchard et al. Statistical bibliography or bibliometrics. *Journal of documentation*, 25(4):348–349, 1969.

[36] D. Raina and B. M. Gupta. Four aspects of the institutionalization of physics research in india (1990–1950): Substantiating the claims of histortical sociology through bibliometrics. *Scientometrics*, 42(1):17–40, 1998.

[37] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. `http://is.muni.cz/publication/884893/en`.

[38] Michael Röder, Andreas Both, and Alexander Hinneburg. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408. ACM, 2015.

[39] V. Sekara, P. Deville, S. E. Ahnert, A. Barabási, R. Sinatra, and S. Lehmann. The chaperone effect in scientific publishing. *Proceedings of the National Academy of Sciences*, 115(50):12603–12607, 2018.

[40] Mark Sistrom, Derek Park, Heath E O'Brien, Zheng Wang, David S Guttman, Jeffrey P Townsend, and Paul E Turner. Genomic and gene-expression comparisons among phage-resistant type-iv pilus mutants of pseudomonas syringae pathovar phaseolicola. *PloS one*, 10(12):e0144514, 2015.

[41] Paul E Turner and Lin Chao. Prisoner's dilemma in an rna virus. *Nature*, 398(6726):441, 1999.

[42] S. Walt, S. C. Colbert, and G. Varoquaux. The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.

[43] Te Wu, Feng Fu, and Long Wang. Moving away from nasty encounters enhances cooperation in ecological prisoner's dilemma game. *PLoS One*, 6(11):e27669, 2011.

[44] M. Youngblood and D. Lahti. A bibliometric analysis of the interdisciplinary field of cultural evolution. *Palgrave Communications*, 4(1):120, 2018.

[45] Martín G Zimmermann and Víctor M Eguíluz. Cooperation, social networks, and the emergence of leadership in a prisoner's dilemma with adaptive local interactions. *Physical Review E*, 72(5):056118, 2005.