A systematic literature review of the Prisoner's Dilemma; collaboration and influence.

Nikoleta E. Glynatsi, Vincent A. Knight

2016

1 Analysing a large corpus of articles

The main focus of the second part of the paper still remains the academic publications on the iterated prisoner's dilemma. However, whilst in Section **??** we covered several publication of specific interest here we analyse the field using a large dataset of articles. We cover the data selection in Section 1.1 and we conduct a preliminary analysis of our data in Section 1.2. In Section 1.3, we describe how authors relationships are analysed graph theoretically to ascertain the level of collaborative nature of the field and identify influence, relative to:

- Two other sub fields of game theory: auction games [7] and the price of anarchy [13].
- A temporal analysis.

Finally in Section 1, we present the results of our study.

1.1 Data Collection

Academic articles are accessible through scholarly databases and collections of academic journals. 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 user interface side of the journal. Interacting with an API has two phases: requesting and receiving. The request phase includes composing a url with the details of what is wanted. 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 we are querying.

In this example the address corresponds to the API of arXiv. The second part of the request contains the search arguments. In our example we are requesting for a single article that the word 'prisoners dilemma' exists within it'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 satisfied the request message. The raw metadata are commonly received in extensive markup language (xml) or Javascript object notation (json) formats [10]. 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 [8]. 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 [8] allow us to collect articles from a list of APIs by specifying just a single keyword. Four prominent journal in the field and a pre print server were used as sources to collect data for this analysis. Those were PLOS, Nature, IEEE, Springer and arXiv.

A series of search terms were used to identify relevant articles. The terms used to collect the main data set were,

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

and articles for which any of these terms existed within the title, the abstract or the text are included in the analysis. More specifically, 30% of article considered here were included because any of the above terms existed within the abstract, 30% within the main text and 40% within the title.

As will be described in Section 1.2, two other game theoretic subfields were also considered in this work, auction games and the price of anarchy. For collecting data on these subfields the search terms used were "auction game theory" and "price of anarchy". The data that were collected and used in this work are archived and available at.

1.2 Preliminary Analysis

A total of three data sets are explored in this work. A summary of each data is presented in this section. The three data sets are:

- The main data set which contains articles on the prisoner's dilemma.
- A secondary data set which contains article on auction games.
- A secondary data set which contains articles on the price of anarchy.

The main data set and the main focus of this analysis is [ref]. This data set consists of 3052 articles with unique titles. In case of duplicates the preprint version of an article (collected from arXiv) was dropped. Of these 3052 article, 89 have not been collected from the aforementioned APIs. These articles were of specific interest and manually added to the dataset throughout the writing of Section **??**. A more detailed summary of the articles' provenance is given by Table 1.

Only 3% of the data set consists by articles we manually added and half the articles were collected from arXiv. The rest four journals have contributed 9%-17% percent of the articles. Note that the latest data collection was perform on November 2018.

provenance	# of Articles	Percentage
Manual	89	2.92
IEEE	295	9.67
Springer	458	15.01
PLOS	482	15.79
Nature	673	22.05
arXiv	1055	34.57

Table 1: Articles' provenance for main data set.

The average number of publications was calculated for the entire dataset and for each provenance. The average number of publications 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 2018 and the first published article, for each provenance, within the data. These averages are summarised in Table 2. Overall an average of 62 articles are published per year on the topic.

	Av. publication
IEEE	4.0
PLOS	7.0
Springer	7.0
Nature	10.0
arXiv	16.0
Overall	46.0

Table 2: Average publication for main data set.

The most significant contribution to this appears to be from arXiv with 32 articles per year, followed by Springer with 11 articles per year.

Though the average publication offers insights about the publications of the fields, it remains a constant number. The data we are handling here is a time series (plotted on Figure 1). The earliest entry is in 1950 when the game was introduced and the latest is in 2018. Though a small number of publication were made between 1950-1980 since the 1980s, and the introduction of computer tournaments, we can see a steady increase to the number of publications. Moreover there is a decrease in 2017-2018. We believe that this is because our data are incomplete for these time periods. Articles that have been written in 2017-2018 have either not being published yet or have are not retrievable yet by the APIs.

Our statements can be explored by studying the time series. Figure 2, illustrates a scatter plot showing the number of articles published from 198 onwards. Exponential fitting of the data, which was done using [], and a perfect fitting proofs that since 1980 to 2016 there has been an increase to the number of publications.

The fitted model can be used to project the behaviour of the field of the next 5 years. The forecasted periods are plotted in Figure and their exact values are given by Table 3. Thought the time series has indicated a slight decrease we can see that the model forecasts an increase over the next years.

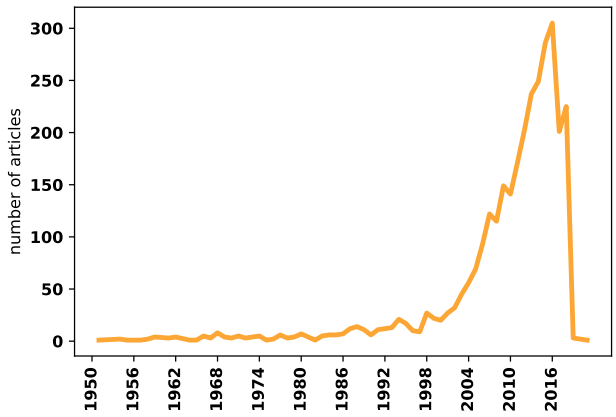


Figure 1: Line plot showing the number of articles published on the PD 1950-2018.

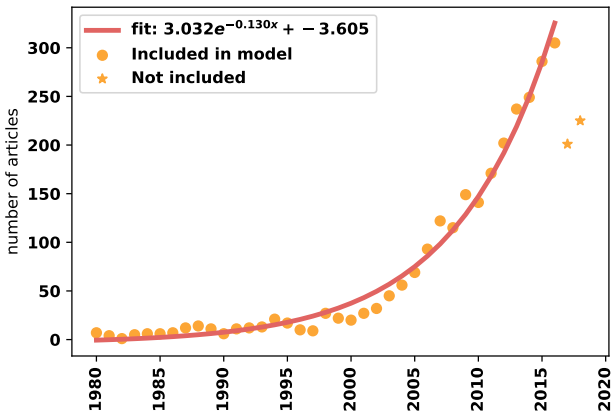
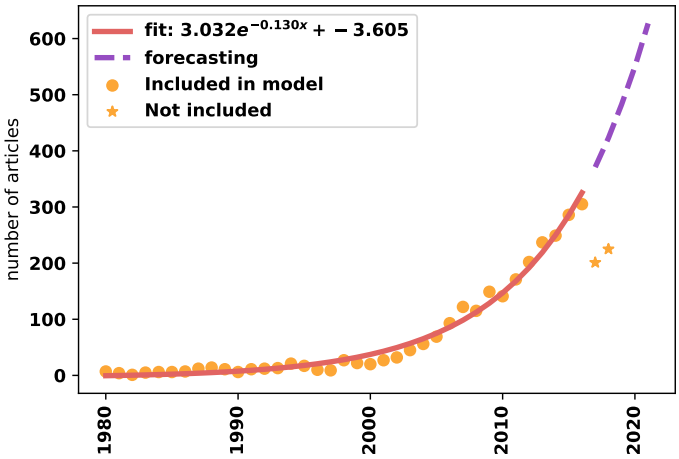


Figure 2: Scatter plot showing the number of articles published on the PD 1980-2018.

Moreover, two subfields of game theory have been chosen for this work; auction game and the price of anarchy. A summary of both data sets collected on this two topics is given by Table 4 in comparison to that of the PD.

In this section we have described the three data sets that we are going to use in the following sections in order to identify collaborative behaviour and influence. Two data sets of different topics are used for comparison reasons. The frequency of articles and authors differs within the three data sets which is ideal.



	Forecast
2017	371.0
2018	423.0
2019	483.0
2020	550.0
2021	627.0

Table 3: Forecasting the number of publications over the next 10 years.

	Prisoner's Dilemma	Price of Anarchy	Auction Games
Num. Articles	2975.00	774	3449
Num. Authors	4918.00	1262	4779
Nature (\%)	22.62	27.26	5.89
IEEE (\%)	9.92	29.59	7.45
Springer (\%)	15.39	36.82	37.72
arXiv (\%)	35.46	8.53	51.29
PLOS (\%)	16.20	1.68	-
Manual (\%)	2.99	-	-
Av. Publication	2975.00	774	3449

Table 4: Measures of all three data sets.

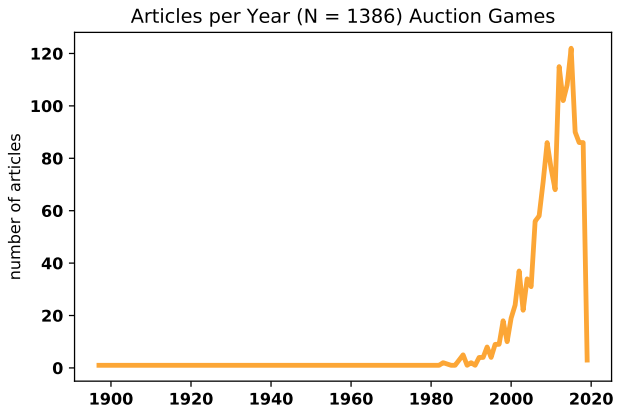


Figure 3: Line plot showing the number of articles published on Auction Games.

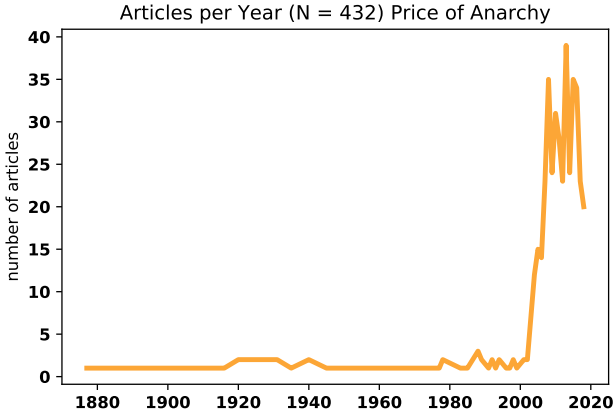


Figure 4: Line plot showing the number of articles published on the Price of Anarchy.

1.3 Methodology

As discussed in [14], bibliometrics or the statistical analysis of published works (originally described by [11]) have been used to support historical assumptions about the development of fields [12], identify connections between scientific growth and policy changes [2], and investigate the collaborative structure of an interdisciplinary field [6]. Most academic research is undertaken in the form of collaborative effort and as [5] points out, it is rationale 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 [3]. The number of collaborations can be very different between research fields and understanding how collaborative a field is, it is not always an easy task. Several studies tend to consider academic citations as a measure for these things. A blog post published in Nature [9] argues that depending on citations can often be misleading. This is because, the true number of citations can not be known (citations can be missed due to data entry errors). Moreover, academics are influenced by many more papers than they actually cite and even so several citations are superficial.

A more recent approach to measure collaborative behaviour is to use the co authorship network, as described in [6]. Taking a the structural point of view, we can model relationship of authors within a field as a graph G with a set V_G of nodes and E_G of edges. The set V_G represents the authors and an edge connects two authors if and only if those authors have written together. Co authorship networks have had several applications, including classifying topic of an interdisciplinary field [14], but here we build upon the work done by [6] on collaborative behaviour. More specifically, here we explore the collaborativeness of the prisoner's dilemma field, and we also extend the approach in order to understand influence; how many connections are made possible because of an author. This possible only because several graph theoretic measures can be used as proxies.

The co authorship network is constructed using the main data set described in Section 1.2 and the open source package Networkx [4]. The prisoner's dilemma network is denoted as G_1 where the number of unique authors $|V(G_1)|$ is 4918 and $|E(G_1)| = 9680$. Note that the names of all authors were standardised to be their last name and first initial (i.e. Martin A. Nowak to M.Nowak). This was done to avoid errors such as Martin A. Nowak and Martin Nowak, being treated as a different person. Networkx will also be used the following section to conduct our analysis.

Collaborativeness, will be analysed using measures such as, isolated nodes, connected components, clustering coefficient, modularity and average degree. These measures allow us to understand the number of connections author can have and how strongly connected these people are. The number of isolated nodes allow us to understand the how many nodes are not connected to another node, thus the number of authors that had not had any known collaboration in the field. The average degree denotes the average number of neighbours for each nodes, i.e. the average number of collaborations between the authors and the number of largest connected component represents the scale of the central cluster of the entire network, as it will discussed in the analysis section. Clustering coefficient, modularity and the degree distribution are also measured. Clustering coefficient defined by,

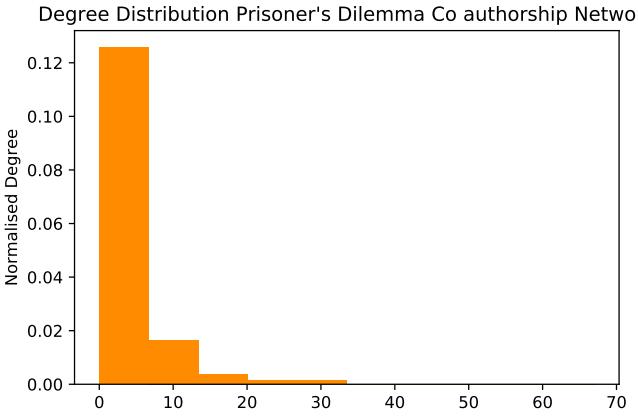
$$C = \frac{3 \times (\text{number of triangle on the graph})}{\text{number of connected triples of nodes}}$$

is a local measure of the degree to which nodes in a graph tend to cluster together in a clique. 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 modules. A high value of modularity corresponds to a structure where authors meanly write in groups and interact less with whole network. We will be using Clauset-Newman-Moore [1] greedy modularity maximization which is implemented within the Networkx library.

Furthermore, we study the infunce using centrality measures. Centrality Measures are .... and often sued to understnad different aspects fo social netwprks []. In this paper we wnat to gras an unrestanfin of two thins. (1) which people influce the field more, and (2) whcih people gain more form their licattion, how much asn other gains from the ... In order to achieve that two cetrality measures that have been choosen were the closeesn and the netween centrality. closeesn

Betwee.

In the next section we will be using all the metrics discussed here to provide insights on tou field. This will done also my compaoring to 2 other feidls as well as exploring te progress of the field over time.



	Prisoner's Dilemma	Price of Anarchy	Auction Games
Av. Degree	4.24	3.54	3.39
# Isolated	157.00	4.00	210.00
# Connected Components	989.00	198.00	949.00
Clustering	0.69	0.70	0.60
Largest cc	1425.00	421.00	2079.00
% Isolated	0.03	0.00	0.05

Table 5: Collaboration metrics comparison.

1.4 Analysis of co authorship network

As subsicss above G_1 is the network of the prisoner;s dilemma. This has been plotted using the open softwrae Gephi and an be seen in Fig.1. It is evidende that our network is disjoint. Which comes with no suprsai as there ate authors that have written on their own. More specificall a total of (%). The size if the network si. Moreover in Fig we can see that there is a by cluster, the largest connected component in the graps, with a aisze of . Overall the avegrage degree is, the clusteing coefficect and the distribution is by:

How these compare to other fields of game theory? A summary of the two graohs g2 and g3 is given by table. G2 and G3 is. We can see that.

Moreover, Table provides a clearer insight on the field recargind time. With a time 5 years we un contructed the cumulatice co authror ship network and Table summarizes all the metrics as discussed all ready.

- isolated 157
- isolated percentage 0.034362004815058
- connected 989
- largest 1425
- cc 0.689
- av. degree 4.2372510396147955

References

[1] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.

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

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

[4] Aric A. Hagberg, Daniel A. Schult, and Pieter 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.

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

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

[7] Flavio M Menezes and Paulo Klinger Monteiro. *An introduction to auction theory*. OUP Oxford, 2005.

[8] Nikoleta and Vince Knight. Nikoleta-v3/arcas: Arcas v 0.0.4, December 2017.

[9] Richard Van Noorden. The science thats never been cited. `https://www.nature.com/articles/d41586-017-08404-0`, 2017.

[10] Nurzhan Nurseitov, Michael Paulson, Randall Reynolds, and Clemente Izurieta. Comparison of json and xml data interchange formats: a case study. *Caine*, 2009:157–162, 2009.

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

[12] Dhruv Raina and BM Gupta. Four aspects of the institutionalization of physics research in india (1990–1950): Substantiating the claims of historical sociology through bibliometrics. *Scientometrics*, 42(1):17–40, 1998.

[13] Tim Roughgarden. *Selfish routing and the price of anarchy*, volume 174. MIT press Cambridge, 2005.

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