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

Following the literature review we conducted in Section **??**, in this section we will focus on the analysis of the study of the prisoner's dilemma using a large dataset of articles. In Section 1.1 the data collection is covered, followed by a preliminary analysis of the 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. This will be done relative to:

- Two other sub fields of game theory: auction games [5] and the price of anarchy [9].
- A temporal analysis.

Finally in Section, we present the results of our analysis.

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 [8]. 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 [6]. 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 [6] 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 3637 articles with unique titles. In case of duplicates the preprint version of an articles (collected from arXiv) were dropped. Of these 3637 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	Percentage
Manual	2.45
IEEE	8.74
Springer	8.96
PLOS	11.44
Nature	17.10
arXiv	51.31

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

	Av. publication
IEEE	5.0
Springer	6.0
PLOS	7.0
Nature	11.0
arXiv	32.0
Overall	62.0

Table 2: Average publication for main data set.

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 byTable 3. Thought the time series has indicated a slight decrease we can see that the model forecasts an increase over the next years.

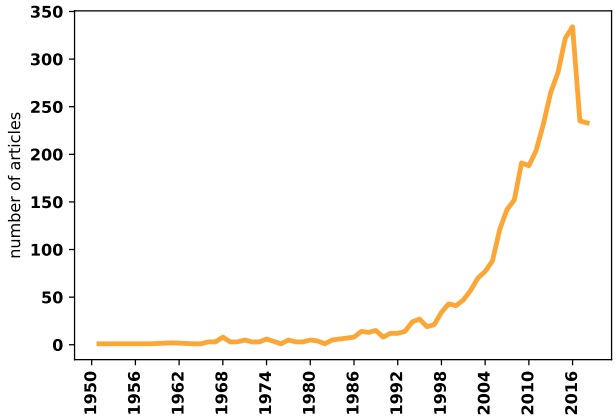


Figure 1: Line plot showing the number of articles published on the PD 1950-2018.

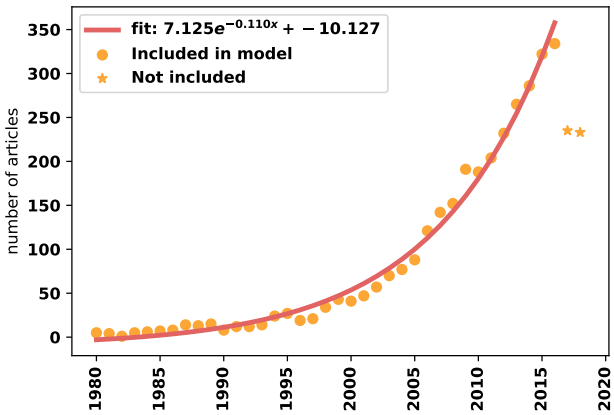
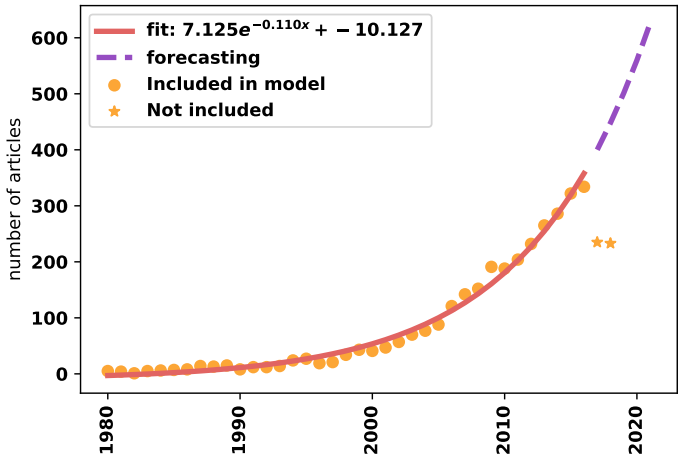


Figure 2: Scatter plot showing the number of articles published on the PD 1980-2018.

	Forecast
2017	400.0
2018	448.0
2019	501.0
2020	560.0
2021	626.0

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

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 **??** in comparison to that of the PD.



Data Set	Num. Articles	Num. Authors	Nature (%)	IEEE (%)	PLOS (%)	Springer (%)	arXiv (%)	Overall Av. publication
Prisoner's Dilemma	3637	2668	17.10	8.74	11.44	8.96	51.31	62
Auction Games	1386	1041	13.92	0.72	-	65.51	19.84	29
Price of Anarchy	432	367	48.38	-	-	32.18	19.44	9

Table 4: Measures of all three data sets.

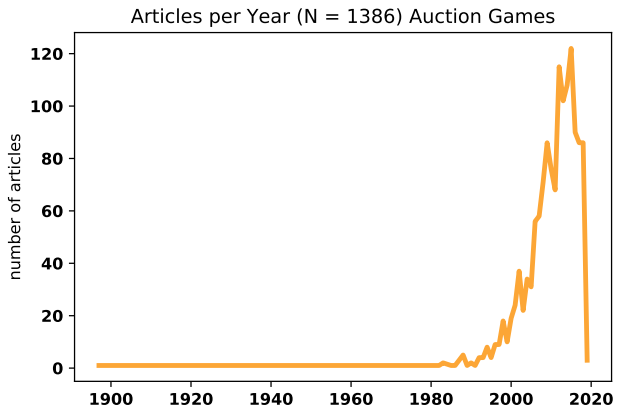


Figure 3: Line plot showing the number of articles published on Auction Games.

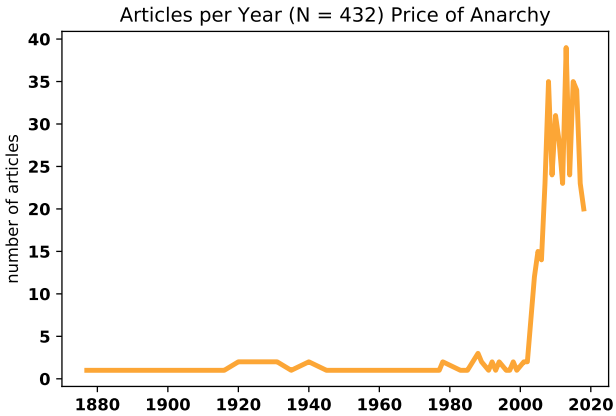


Figure 4: Line plot showing the number of articles published on the Price of Anarchy.

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.

1.3 Methodology

Most academic research is undertaken in the form of collaborative effort. As discussed in [4], it is rationale that two or more people have the potential to do better as a group than individually. Academic collaborations have many different forms. Researchers might have immediately collaborated and written together. Others might have collaborated through a common co author.

Collaboration in groups has a long tradition in experimental sciences and it has be proven to be productive according to [2]. Even so, the number of collaborations can be very different between research fields and measuring collaboration is not always an easy task. Another aspect of collaborative behaviour is influence. For example academics can influence through workshops, talks or by collaborating with people in our environment.

Several studies tend to consider academic citations as a measure for these things. As discussed in [7], 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.
- Academics are influenced by many more papers than they actually cite.
- Several citations are superficial.

We suggest an alternative measure of collaboration and influence by looking at the co authorship network. A co authorship network, is a network where academics that have written and published together are connected.

Using graph theoretic concepts this network will be analysed to undestand:

- Collaborativeness; for example the number of connections an author has as well as more sophisticated measures of closeness.
- Influence; how many connections are made possible because of an author.

We introduced several network measures that we will be using such as:

- Number of connected components.
- Clustering coefficient.
- Degree distributions.
- Centrality.

References

[1] Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer Science & Business Media, 2013.

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

[3] R I D Harris. Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test. *Economics Letters*, 38(4):381–386, 1992.

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

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

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

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

[8] 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.

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