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

Nikoleta E. Glynatsi, Vincent A. Knight

1 Timeline

2 Analysing a large corpus of articles

The focus of the paper has been the academic publications on the topic the iterated prisoner's dilemma. Whilst in Section 1 we covered several publications of specific interest and manually partitioned the literature in different sections, in the second part of this paper we analyse the publications using a large dataset of articles. The data collection process is covered in Section 2.1 and a preliminary analysis of the data is conducted in Section 2.2. In Section 2.3, we present the methodology which will be used to analyse the authors relationships. In summary, we will be using graph theoretical methods to ascertain the level of collaborative nature of the field and identify influence, relative to:

- Two other sub fields of game theory: auction games [11] and the price of anarchy [17].
- A temporal analysis.

Finally in Section 2.4, the results of the analysis are presented.

2.1 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 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 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 [14]. 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 [12]. 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 [12] allow us to collect articles from a list of APIs by specifying just a single keyword. Four prominent journals 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, 23% of article considered here were included because any of the above terms existed within the abstract, 50% within the main text and 27% within the title. As will be described in Section 2.2, two other game theoretic sub fields were also considered in this work, auction games and the price of anarchy. For collecting data on these sub fields the search terms used were “auction game theory” and “price of anarchy”. The three data sets are archived and available at. Note that the latest data collection was perform on November 2018.

2.2 Preliminary Analysis

A summary of each of the three data sets used 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]. It consists of 3089 articles with unique titles. In case of duplicates the preprint version of an article (collected from arXiv) was dropped. Of these 3089 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 1. A more detailed summary of the articles’ provenance is given by Table 1. Only 3% of the data set consists by articles that were manually added and 33% of the articles were collected from arXiv. The rest four journals have contributed 9%-21% percent of the articles.

provenance	# of Articles	Percentage
Manual	89	2.88
IEEE	295	9.55
PLOS	482	15.60
Springer	572	18.52
Nature	673	21.79
arXiv	1056	34.19

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 49 articles are published per year on the topic. The most significant contribution to this appears to be from arXiv with 16 articles per year, followed by Nature with 10 and Springer with 9.

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 between 1950, when the game was introduced, and 2018 (Figure 1). Two observations can be made from Figure 1.

	Av. publication
IEEE	5.0
PLOS	8.0
Springer	9.0
Nature	11.0
arXiv	16.0
Overall	49.0

Table 2: Average publication for main data set.

1. A steady increase to the number of publications since the 1980s and the introduction of computer tournaments.
2. A decrease in 2017-2018. This is due our data set being incomplete. Articles that have been written in 2017-2018 have either not being published or have are not retrievable by the APIs yet.

These observations can be confirmed by studying the time series. Using [6], an exponential distribution is fitted to the data from 1980-2016. The perfect fitting proves that since 1980 there has been an exponential increase to the number of publications till 2016 (Figure 2). The fitted model can also be used to project the behaviour of the field for the next 5 years. The forecasted periods are plotted in Figure 3 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 that the number of publications will keep increasing, thus indicating that the field of the iterated prisoner's dilemma still attracts academic attention.

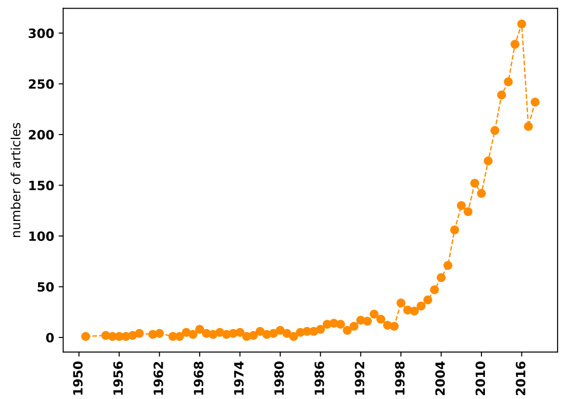

Figure 1: Line plot; # of articles published on the PD 1950-2019.

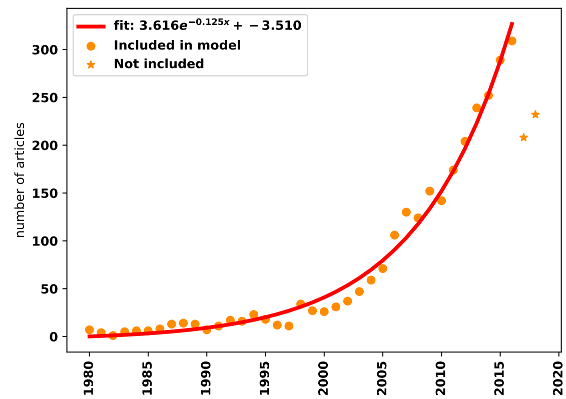

Figure 2: Scatter plot; # of articles published on the PD 1980-2019.

	Forecast
2017	371.0
2018	421.0
2019	478.0
2020	542.0
2021	615.0

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

Moreover, two sub fields of game theory have been chosen for this work; auction game and the price of anarchy.

- Auction theory is a branch of economics which deals with how people act in auction markets and researches the properties of auction markets. Game theory is being used for years to study auctions and the behaviour of the bidders [18]. The earlies entry in our data set [ref] goes back to 1974 (Figure 4).

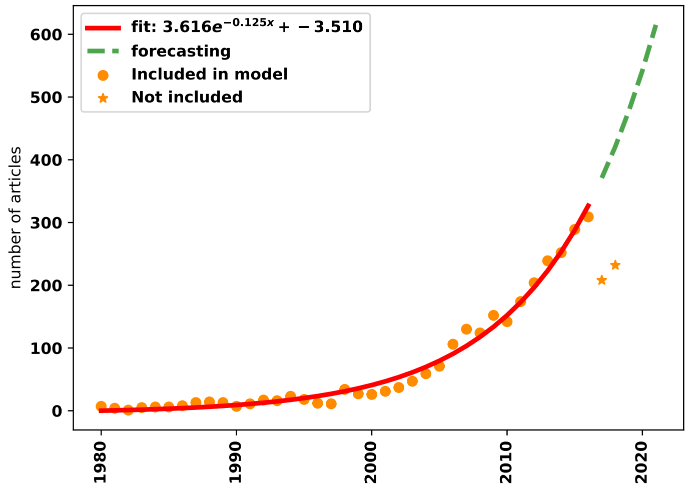

Figure 3: Forecast for 2017-2022

- Price of Anarchy is a concept in economics and game theory which measures how the efficiency of a system degrades due to selfish behaviour of it's agents. There is a variety of such measures however the price of anarchy has attracted a lot of attention since it's informal introduction in 1999 by [7]. The first entry in the data set [ref] is a year later in 2000 (Figure 5).

A summary of both data sets collected on both topics, in comparison to that of [ref], is given by Table 4.

The iterated prisoner's dilemma and auction theory are very well studied topics that have been publicising for decades. A large number of articles have been collected for both topics, 3089 and 34444 respectively. Though, auction games have a larger number of articles, the iterated prisoner's dilemma has almost 300 more authors. Auction games have an overall average publication of 93 articles compared to the prisoner's dilemma with 49. The 50% of articles for [ref] have been collected from the pre print server arXiv and no articles have been published in PLOS.

Compared to these two topics the price of anarchy is a fairly recent one. Only a total of 747 articles have been collected, however it has a large number of 1229 authors. Meaning that on average each paper had had at least two authors. It has an overall average publication of 39 articles and the biggest contribution has been made from Springer. Note that no article have been added manually for the data sets for the two extra sub fields.

	Num. Articles	Num. Authors	Manual (%)	PLOS (%)	Nature (%)	Springer (%)	IEEE (%)	arXiv (%)	Av. Publication
Prisoner's Dilemma	3089	5811	2.88	15.6	21.79	18.52	9.55	34.19	49.0
Auction Games	3444	5362	-	-	5.89	37.63	7.46	51.36	93.0
Price of Anarchy	746	1314	-	1.74	24.66	38.07	30.70	8.85	41.0

Table 4: Measures of all three data sets.

2.3 Methodology

As discussed in [19], bibliometrics or the statistical analysis of published works (originally described by [15]) have been used to support historical assumptions about the development of fields [16], identify connections between scientific growth and policy changes [3], and investigate the collaborative structure of an interdisciplinary field [10]. Most academic research is undertaken in the form of collaborative effort and as [8] 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 [4]. 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 [13] 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

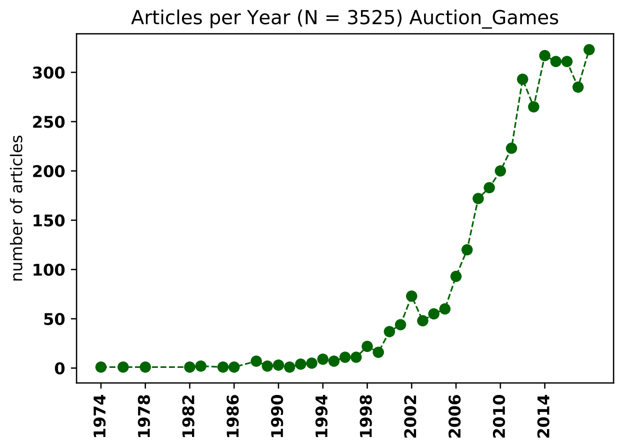

Figure 4: Line plot; # articles published on auction games 1974-2018.

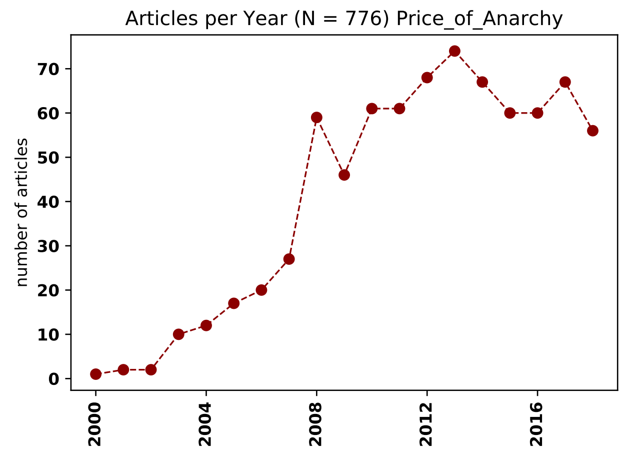

Figure 5: Line plot; # articles published on the price of anarchy.

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 [10]. Using this approach has many advantages as several graph theoretic measures can be used as proxies to explain authors relationship. In [10], 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 prisoner's dilemma. In this paper we build upon the work done by [10] and extend their methodology. Though in [10], they considered a data set from a single source, Web of Science, our data have been collected from five different sources. Moreover, the collaborative results of our analysis are compared to those of two different sub fields. Co authorship networks have also been used in [19] for classifying topics of an interdisciplinary field. This was done using centrality measures, which will be covered below, here we use centrality measures in order to understand the influence an author can have and can receive by being part of an academic group.

The relationship between the authors within a field will be modelled 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. The co authorship network is constructed using the main data set described in Section 2.2 and the open source package Networkx [5]. The prisoner's dilemma network is denoted as G_1 where the number of unique authors $|V(G_1)|$ is 5811 and $|E(G_1)| = 10397$. Note that the names of all authors names were formatted as their last name and first initial (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. 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 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. We are interested in the number of connected components but also the size of the largest connected component in the network. The size 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 and modularity and are also calculated. Clustering coefficient defined by,

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

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 mainly write in groups and interact less with the rest of the network. We will be using the Louvain method described in [2] to calculate modularity.

Furthermore, the second part of the analysis focuses on the study of influence. Networks are commonly dominated by one person who controls information flow and people that receive a great amount of information due to their position. In this paper we aim to understand two things, (1) which people control the flow; 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 we will be using graph theoretic metrics, more specifically centrality measures. Centrality measures are often used to understand different aspects fo social networks [9]. In order to achieve that two centrality measures that have been chosen were closeness and betweenness centrality.

1. In networks some nodes have a short distance to other nodes and consequently are able to spread information on the network very effectively. A representative of this idea is **closeness centrality**, where a person is seen as centrally involved in the network if it requires only few intermediaries for contacting others and thus is structurally relatively independent. Here, we define this as influence. Authors with a high value of closeness centrality, are the authors that spread scientific knowledge easier on the network and we say that 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 regarded node 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 them, we define this as the gain from the influence, thus these authors gain the most from their networks.

In the next section we will be using all the metrics discussed here to provide insights on the field.

2.4 Analysis of co authorship network

As mentioned previously, G_1 denotes the co authorship network of the iterated prisoner's dilemma. The open source visualization software Gephi [1] has been used to plot the networks of this work, more specifically G_1 is given by Figure. It is evident that our network is disjoint, which is only natural as many authors write academic articles on their own. More specifically, a total of 176 authors, have had single author publications, which corresponds to the 0.033 (%) of authors in G_1. There are a total of 1356 connected components and the largest one has a size of 815 . The largest connected component is shown in Figure. The network as a clustering coefficient of 0.708 , thus authors are 70% likely to write with a collaborator's co author. The degree distribution, Figure 6, shows that the average degree is ≈ 4, however there are authors with far more connections, the largest one being 58.

How does these compare to other fields and more specifically to other fields of game theory? A summary of the two graphs,

- G_2 for auction games and
- G_3 for the price of anarchy,

are given by Figure. A summary of the collaborative metrics for all three co authorship networks is given by Table 5, and the following remarks can be made:

- Comparing to another well studied topic (G_2), the co authorship network G_1 appears to be more modular. This is due the high values of modularity, connected components and clustering coefficient. Authors in G_1 tend to write more in teams (modularity .977 > .958), separated from the main cluster and it's more likely for them to create smaller clusters of 3 (clustering .702 > .622). On the other hand, G_2 has main cluster of bigger size (1348 > 815), suggesting a more chained community than G_1.

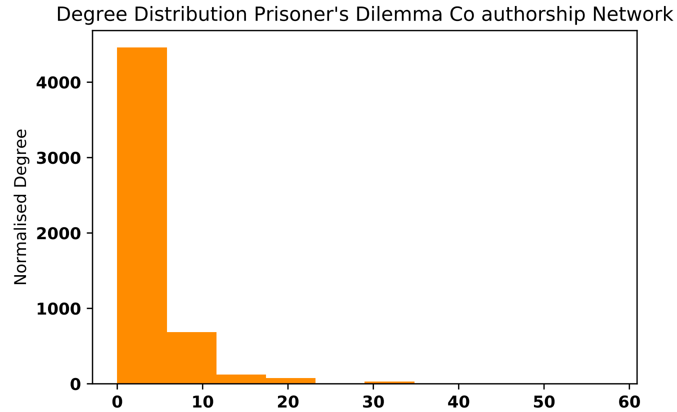

Figure 6: Degree distribution for network G_1.

	# Connected Components	# Edges	# Isolated	# Nodes	% Isolated	Av. Degree	Clustering	Largest cc	Modularity
Prisoner's Dilemma	1356	10397	176	5394	0.033	3.855	0.708	815	0.977
Auction Games	1272	7861	256	5165	0.050	3.044	0.622	1348	0.958
Price of Anarchy	245	1952	4	1154	0.003	3.383	0.713	221	0.964

Table 5: Network metrics for G_1, G_2, G_3.

- In the more recent topic of price of anarchy (G_3) there are hardly any people that have published a paper alone. There is already a small community that is chained in the main cluster of 221 authors. The network has a high value of modularity 0.964 and a high clustering probability (clustering coeff.= 0.713).
- Shown in Figure 7 are the degree distributions of all three networks. It has been statistically tested, and none of the distributions are normally distributed. More specifically, all three distributions are very skewed to the left side. Though the average degree is near 4 the medians for $G_1 - G_3$ are 3, 2 and 3. Based Kruskal Wallis test $p < 0.05$ there is significant difference in the medians. In G_1 and G_2 there are cases of high degree (> 20) but this could be an affect of the size of the data, networks and subsequently the size of the main clusters.

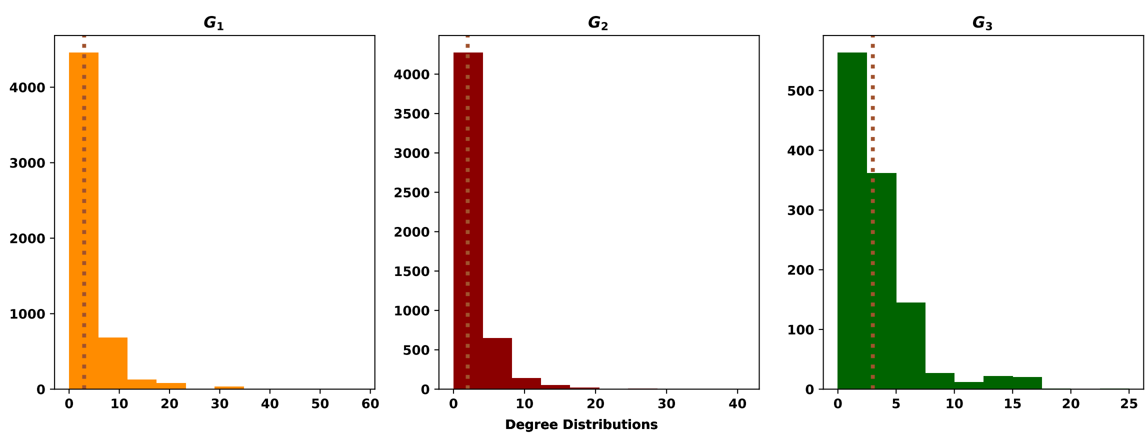

Figure 7: Degree distribution for networks G_1 - G_3.

The growth of the networks over time is studied by constructing the network cumulatively with a year (of publication) interval. A total of 64 subgraphs over 64 periods, starting in 1950, were created and all the collaborative metrics for each subgraph have been calculated. These are given by Table 6. Similar to the results of [10], it can been observed that the network G_1 grows over time. Metrics such as the number of nodes, the number of connected components and the degree all increase. The network seems to have always had a high value of modularity.

In Figure 8 the normalised number of nodes, which is calculated by dividing by the total number of nodes in G_1 (5811), is shown. A steep increase to the size of the network is spotted around the 2000s, this is was also briefly comment upon in [10]. However, that increase does not appear only in G_1. By comparing the normalised number of nodes of G_1 to the other networks it is shown that the growth rate for all three networks after the 2000s is a perfect match, Figure 9. This is evidence that something else shocked the academic community around that time.

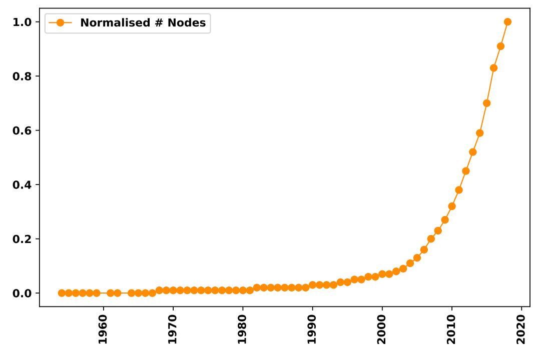

Figure 8: Normalized # Nodes over time for G_1.

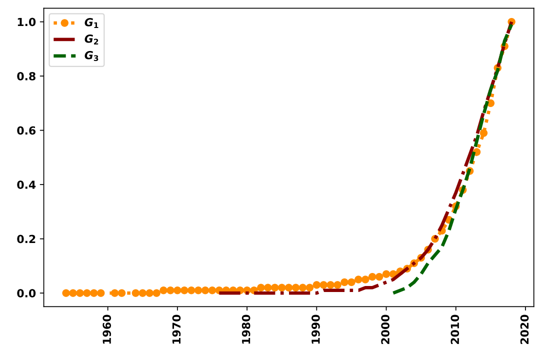

Figure 9: Normalized # Nodes over time for G_1 -G_3.

The change of the average degree over time has also been calculated for the subgraphs and is shown in Figure 10. The growth of the average degree appears to differ between the networks. In G_1 the first publications were single author ones, followed by a steep increase to a degree of 1 just before the 1970s and the average degree is steadily increasing since then. A similar trend appears in G_2, however in G_3 there has been a sharp increase since the beginning of the field, with a minor decrease.

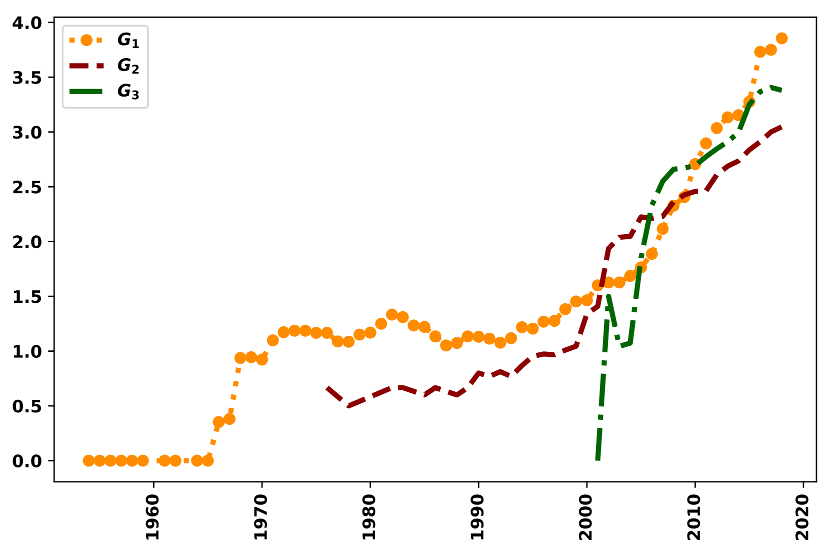

Figure 10: Degree distribution of cumulative subgraphs for G_1 - G_3.

The next results discussed here are on centrality measures. For G_1 the most central author based on closeness and betweenness are given by Tables 11 and 12 respectively. Centrality measures range between $[0, 1]$. The betweenness centrality of the most central authors in G_1 are rather low with the highest ranked author being Matjaz Perc with a between centrality of 0.008, Table 12. Matjaz Perc is also the first ranked author based on closeness centrality, with a centrality of 0.04. Perc's work has been briefly discussed in Section, and the centrality measure suggest that the network

	# Connected Components	# Isolated	# Nodes	% Isolated	Av. Degree	Clustering	Largest cc	Modularity
1954 - 1950	3	3	3	1.00	0.00	0.00	1	-
1954 - 1955	2	2	2	1.00	0.00	0.00	1	-
1955 - 1956	3	3	3	1.00	0.00	0.00	1	-
1956 - 1957	4	4	4	1.00	0.00	0.00	1	-
1957 - 1958	6	6	6	1.00	0.00	0.00	1	-
1958 - 1959	7	7	7	1.00	0.00	0.00	1	-
1959 - 1961	7	7	7	1.00	0.00	0.00	1	-
1961 - 1962	8	8	8	1.00	0.00	0.00	1	-
1962 - 1964	9	9	9	1.00	0.00	0.00	1	-
1964 - 1965	10	10	10	1.00	0.00	0.00	1	-
1965 - 1966	14	11	17	0.65	0.35	0.00	2	0.666667
1966 - 1967	17	13	21	0.62	0.38	0.00	2	0.75
1967 - 1968	21	13	32	0.41	0.94	0.14	5	0.684444
1968 - 1969	24	16	36	0.44	0.94	0.14	6	0.629758
1969 - 1970	26	17	39	0.44	0.92	0.13	6	0.666667
1970 - 1971	31	18	51	0.35	1.10	0.28	6	0.826531
1971 - 1972	34	19	58	0.33	1.17	0.34	6	0.866782
1972 - 1973	34	18	59	0.31	1.19	0.34	6	0.873469
1973 - 1974	34	18	59	0.31	1.19	0.34	6	0.873469
1974 - 1975	35	19	60	0.32	1.17	0.33	6	0.873469
1975 - 1976	35	19	60	0.32	1.17	0.33	6	0.873469
1976 - 1977	41	23	68	0.34	1.09	0.29	6	0.885318
1977 - 1978	42	23	70	0.33	1.09	0.29	6	0.890582
1978 - 1979	42	23	73	0.32	1.15	0.29	6	0.893424
1979 - 1980	44	25	77	0.32	1.17	0.31	6	0.899753
1980 - 1981	45	26	80	0.32	1.25	0.32	6	0.8928
1981 - 1982	46	26	84	0.31	1.33	0.35	6	0.903061
1982 - 1983	48	27	87	0.31	1.31	0.34	6	0.906125
1983 - 1984	54	32	94	0.34	1.23	0.31	6	0.909037
1984 - 1985	55	33	95	0.35	1.22	0.31	6	0.909037
1985 - 1986	63	40	104	0.38	1.13	0.28	6	0.911807
1986 - 1987	73	48	116	0.41	1.05	0.25	6	0.916958
1987 - 1988	75	48	121	0.40	1.07	0.27	6	0.924497
1988 - 1989	80	47	134	0.35	1.13	0.27	6	0.937673
1989 - 1990	86	49	145	0.34	1.13	0.27	6	0.944676
1990 - 1991	94	53	158	0.34	1.11	0.27	6	0.950413
1991 - 1992	102	59	169	0.35	1.08	0.25	6	0.953025
1992 - 1993	110	62	186	0.33	1.12	0.27	6	0.95932
1993 - 1994	127	72	220	0.33	1.22	0.32	6	0.965471
1994 - 1995	137	74	239	0.31	1.21	0.30	6	0.969329
1995 - 1996	145	77	257	0.30	1.27	0.32	6	0.970831
1996 - 1997	156	81	279	0.29	1.28	0.34	6	0.974309
1997 - 1998	160	65	311	0.21	1.38	0.35	6	0.979773
1998 - 1999	162	58	329	0.18	1.45	0.38	6	0.981741
1999 - 2000	183	67	373	0.18	1.46	0.39	6	0.983778
2000 - 2001	184	54	400	0.14	1.60	0.41	7	0.983066
2001 - 2002	206	61	450	0.14	1.63	0.42	7	0.984547
2002 - 2003	229	58	509	0.11	1.63	0.42	7	0.987083
2003 - 2004	253	58	580	0.10	1.69	0.43	10	0.988052
2004 - 2005	284	57	679	0.08	1.76	0.46	19	0.98891
2005 - 2006	342	66	854	0.08	1.89	0.50	21	0.990724
2006 - 2007	402	76	1056	0.07	2.12	0.53	24	0.989663
2007 - 2008	454	85	1255	0.07	2.33	0.55	32	0.989753
2008 - 2009	520	104	1462	0.07	2.41	0.55	56	0.987517
2009 - 2010	581	114	1700	0.07	2.71	0.57	99	0.979084
2010 - 2011	665	121	2040	0.06	2.90	0.60	121	0.980477
2011 - 2012	756	126	2422	0.05	3.04	0.63	210	0.979199
2012 - 2013	843	138	2807	0.05	3.13	0.64	330	0.977857
2013 - 2014	942	148	3199	0.05	3.15	0.65	406	0.974512
2014 - 2015	1064	159	3798	0.04	3.28	0.67	514	0.975909
2015 - 2016	1184	169	4472	0.04	3.73	0.69	614	0.974724
2016 - 2017	1274	173	4925	0.04	3.75	0.70	703	0.976258
2017 - 2018	1356	176	5385	0.03	3.85	0.71	815	0.97731

Table 6: Collaborativeness metrics for cumulative graphs.

is very influenced by him. He is connected to a total of 58 nodes and he has published to all five of the different sources we are considering in the study. Though he also gains from his position in the network, the gain is minor. An author is not in the top influencers but gain much from his position is Martin Nowak.

	Name	Closeness
1	Matjaz Perc	0.048447
2	Yamir Moreno	0.044840
3	Zhen Wang	0.044005
4	Long Wang	0.043770
5	Attila Szolnoki	0.043338
6	Luo-Luo Jiang	0.042148
7	Arne Traulsen	0.041790
8	Valerio Capraro	0.041257
9	Cheng-Yi Xia	0.040791
10	Angel Sanchez	0.040562

Figure 11: Authors that gain the most influence in G_1.

	Name	Betweeness
1	Matjaz Perc	0.008331
2	Zhen Wang	0.006356
3	Yamir Moreno	0.004806
4	Long Wang	0.003538
5	Martin Nowak	0.003230
6	Valerio Capraro	0.002739
7	Arne Traulsen	0.002479
8	Angel Sanchez	0.002319
9	Jianye Hao	0.002188
10	Franz Weissing	0.002186

Figure 12: Ten most influenced authors in G_1.

Overall, the values of closeness centrality appear to be higher than those of betweenness. These can be better explored by considering the centralities' distributions for all three networks. The distributions for both centralities are plotted in Figures 13 and 14, and more over for closeness centrality a violin plot is also given by Figure 15. Several remarks are made from the centralities' distributions.

Regarding between centrality,

- None of the three distributions, Figure 13, is normally distributed. The medians of the distributions are compared and found to be statistically different based on a Kruskal Wallis test. Authors in G_3 gain more from being in the network than others in G_1 and G_2.
- All three distributions are skewed to the left (at 0). That implies that in all three networks, authors do not gain much from the influence of their fields.

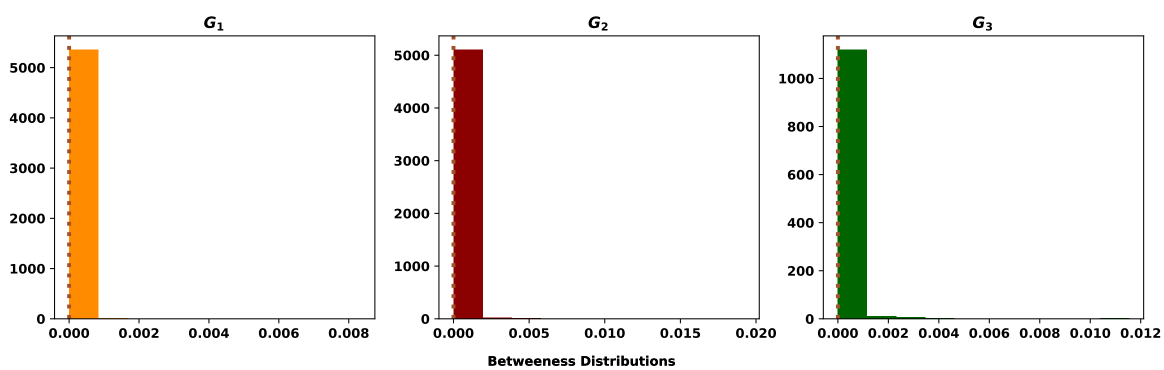

Figure 13: Betweenness centrality distributions $G_1 - G_3$.

On the other hand, closeness distributions have more variation. The following observations are made from the distributions:

- Neither are normally distributed and there is a significant difference between the medians of all three distributions, with G_3 having a larger median.
- There are clusters from all three networks for which a number of authors have a closeness centrality greater than 0.02. The authors in these clusters were explored but not pattern was found behind their publications. The provenance and the year of publication were checked.

- The authors in these clusters, are the authors which are in the main clusters of their relative networks. Thus, the people that influence the field the most are the most central authors in the main cluster of the co authorship network of a field.
- In Figure 15, the distributions are plotted in a violin plot. The network with the highest value of closeness centrality is the network with the largest cluster, G_2.
- Both G_2 and G_3 have more people influencing the field compared to G_1.

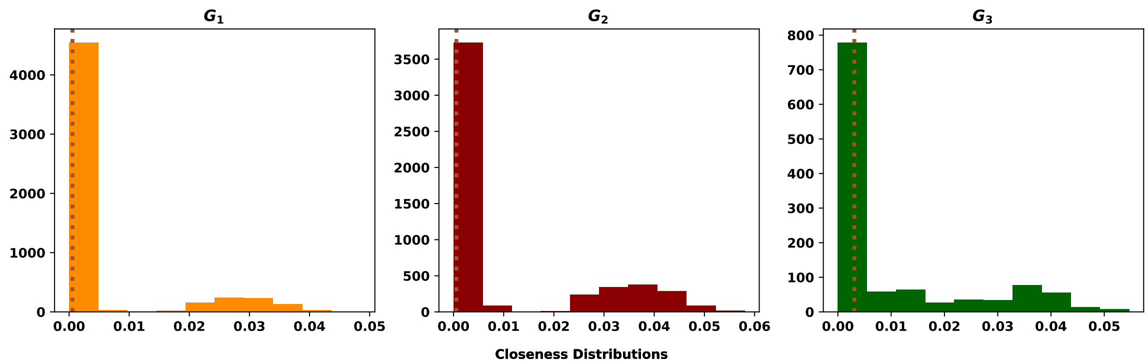

Figure 14: Closeness centrality distributions $G_1 - G_3$.

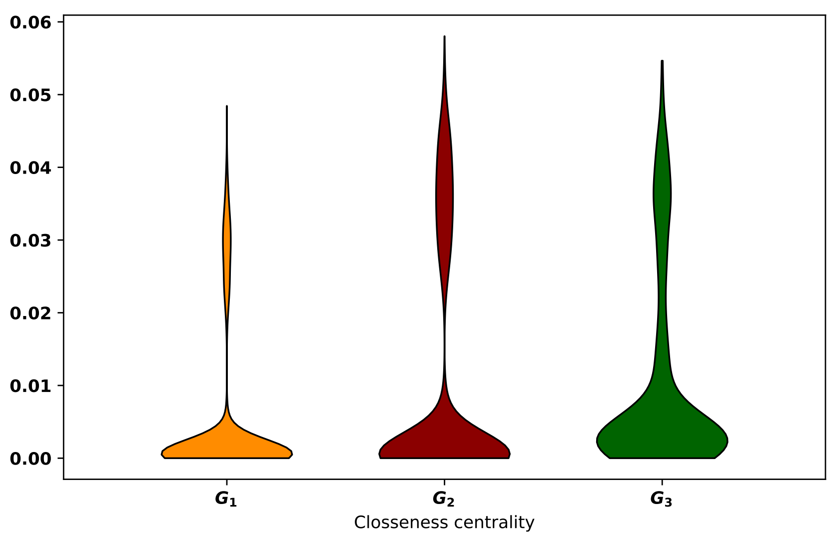

Figure 15: Violin plots of closeness centralities.

These results can be extend to the cumulative sub graphs. The distributions for both centralities for each sub graph of G_1 have been plotted. In Figure 16 we observe that though there are a few outliers, betweenness centrality has always been low for G_1. In comparison, the closeness centrality of the network changes over times. There are several periods when authors influenced the field more than today, for example in 1972-1973. The centrality does not appear to follow a pattern, Figure 17.

2.5 Conclusion

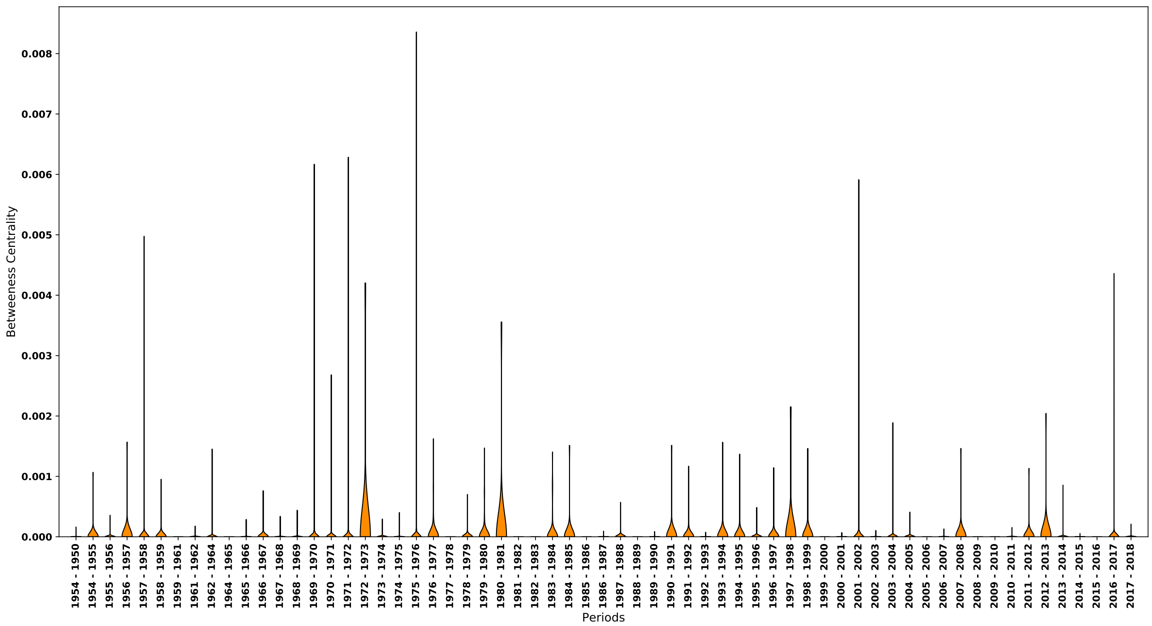

Figure 16: Betweenness centrality distributions cumulative sub graphs of G_1.

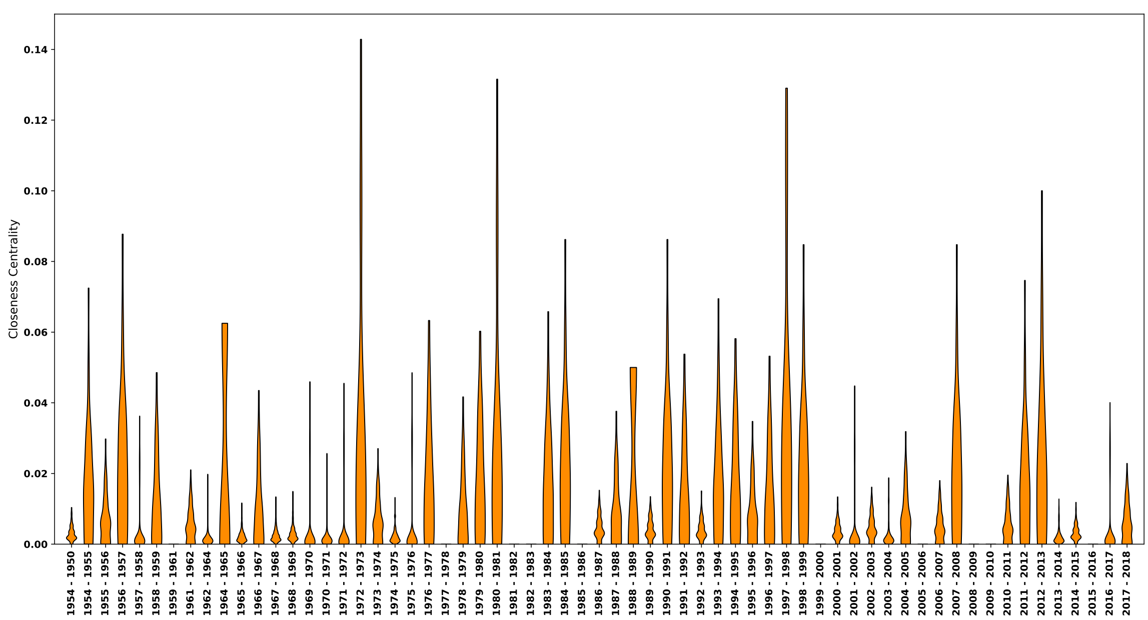

Figure 17: Closeness centrality distributions cumulative sub graphs of G_1.

References

[1] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks, 2009.

[2] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[3] 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.

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

[5] 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.

[6] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed ¡today¿].

[7] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Conference on Theoretical Aspects of Computer Science*, STACS'99, pages 404–413, Berlin, Heidelberg, 1999. Springer-Verlag.

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

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

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

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

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

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

[14] 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.

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

[16] Dhruv Raina and BM 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.

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

[18] Martin Shubik. The dollar auction game: a paradox in noncooperative behavior and escalation. *Journal of Conflict Resolution*, 15(1):109–111, 1971.

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