# 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 Section 1 has been the academic publications on the topic of 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 author relationships. In summary, we will be using graph theoretical methods to ascertain the level of collaborative nature of the field and identify influence. This type of analysis has been carried out in [**?**]. The novelty here is to consider approaches not considered in [**?**] and new origins of publications. A further comparison of the results are made, relative to:

- Two other sub fields of game theory: auction games [**?**] and the price of anarchy [**?**].
- 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. 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 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 [**?**]. 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 [**?**]. 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 [**?**] 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: PLOS, Nature, IEEE, Springer and arXiv.

A series of search terms were used to identify relevant articles:

- "prisoner's dilemma",
- "prisoners dilemma",
- "prisoner dilemma",
- "prisoners evolution",
- "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 data set which contains article on auction games.
- A data set which contains articles on the price of anarchy.

The main data set is archived at [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 of articles that were manually added and 33% of the articles were collected from arXiv.

| 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 the 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 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 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.

| | 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.

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.

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 to 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 [**?**], an exponential distribution is fitted to the data from 1980-2016. The exponential fitting proves that since 1980 there has been an exponential increase in 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. The time series has indicated a slight decrease however 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.

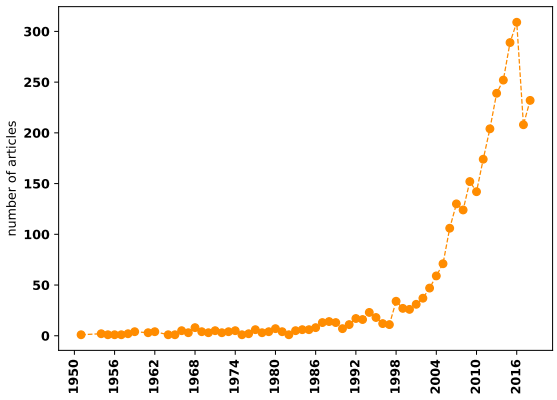



Figure 1: Line plot; # of articles published on the PD 1950-2019.

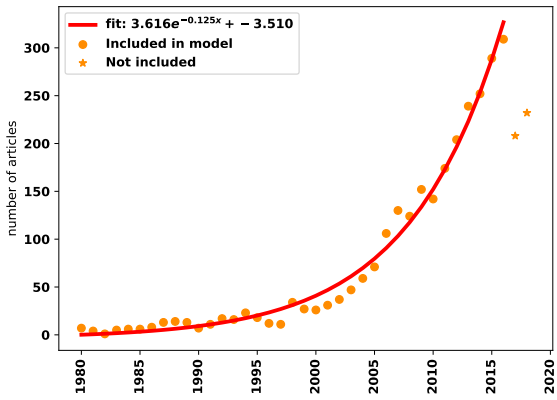



Figure 2: Scatter plot; # of articles published on the PD 1980-2019.

| | 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.

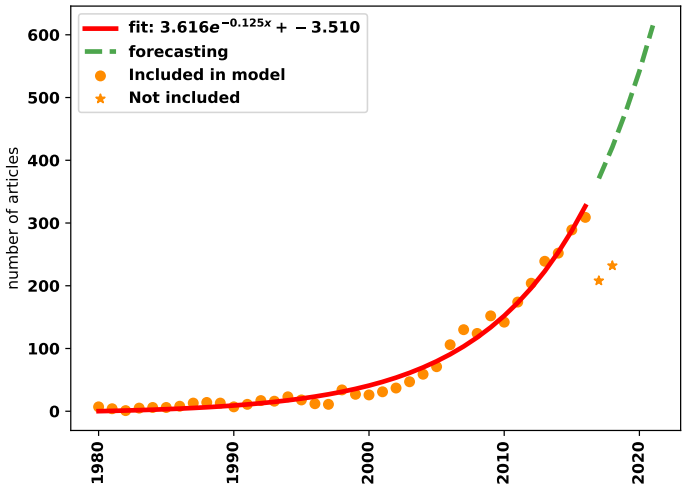



Figure 3: Forecast for 2017-2022

To allow for a comparative analysis two sub fields of game theory have been chosen for this work; auction games and the price of anarchy.

- Auction theory is a branch of game theory which researches the properties of auction markets. Game theory is used for years to study auctions and the behaviour of bidders [**?**]. The earliest entry in our data set [ref] goes back to 1974 (Figure 4).
- Price of Anarchy is a concept in 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 [**?**]. 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. 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. Yearly Publication |
|---|---|---|---|---|---|---|---|---|---|
| Prisoner's Dilemma | 3089 | 5811 | 2.88 | 15.6 | 21.79 | 18.52 | 9.55 | 34.19 | NaN |
| Auction Games | 3444 | 5362 | - | - | 5.89 | 37.63 | 7.46 | 51.36 | NaN |
| Price of Anarchy | 747 | 1315 | 0.13 | 1.74 | 24.63 | 38.02 | 30.66 | 8.84 | NaN |

Table 4: Measures of all three data sets.

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 yearly publication of 93 articles per year compared to the prisoner's dilemma with 49 per year. 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. 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 to Springer.

## 2.3 Methodology

As discussed in [**?**], bibliometrics or the statistical analysis of published works (originally described by [**?**]) have been used to support historical assumptions about the development of fields [**?**], identify connections between scientific growth and

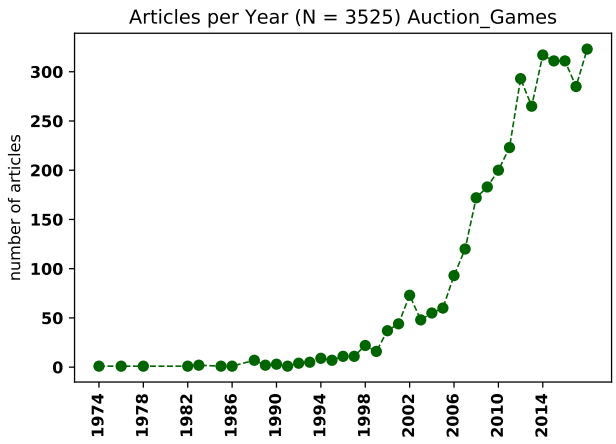



Figure 4: Line plot; # articles published on auction games 1974-2018.

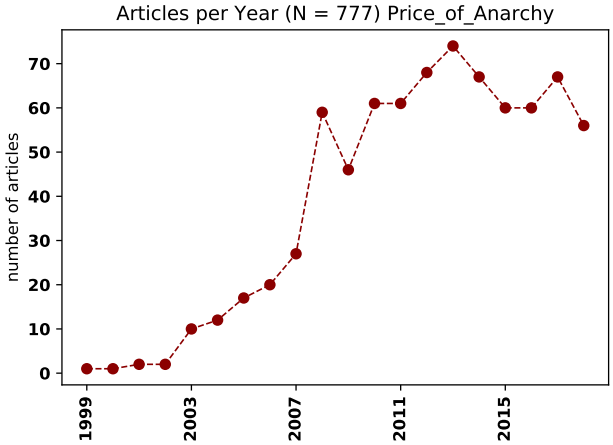



Figure 5: Line plot; # articles published on the price of anarchy.

policy changes [**?**], and investigate the collaborative structure of an interdisciplinary field [**?**]. Most academic research is undertaken in the form of collaborative effort and as [**?**] 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 [**?**]. The number of collaborations can be very different between research fields and understanding how collaborative a field is, 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 [**?**] 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 [**?**]. Using this approach has many advantages as several graph theoretic measures can be used as proxies to explain authors relationship. In [**?**], 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 [**?**] and extend their methodology. Though in [**?**], 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 [**?**] 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 [**?**]. The prisoner's dilemma network is denoted as $G_1$ where the number of unique authors $|V(G_1)|$ is 5394 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.

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. The clustering coefficient, defined as 3

times the number of triangle 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. 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 [**?**] to calculate modularity.

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 further 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 [**?**]. In order to achieve that two centrality measures that have been chosen are 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 to contact 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 into the field.

## 2.4 Analysis of co authorship network

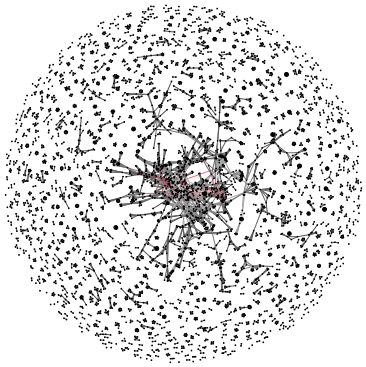
As mentioned previously, $G_1$ denotes the co authorship network of the iterated prisoner's dilemma. The open source visualization software Gephi [**?**] has been used to plot the networks of this work, more specifically $G_1$ is given by Figure 6a. It is evident that our network is disjoint, which is 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 3.3000000000000003 (%) 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 6b. 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 7, shows that the average degree is $\approx 4$, 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,

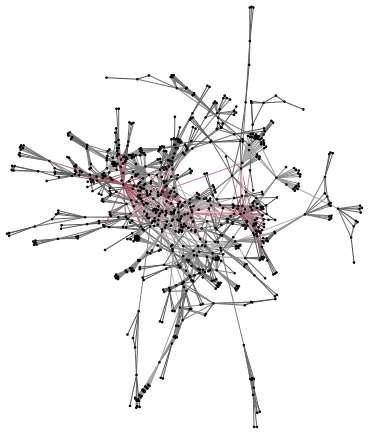
- $G_2$ for auction games and
- $G_3$ for the price of anarchy,

are given by Figures 6c and 6e. A summary of the collaborative metrics for all three co authorship networks is given by Table 5, and the following remarks can be made:

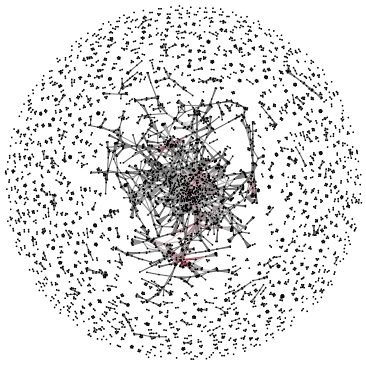
- Comparing to another well studied topic ($G_2$), the co authorship network $G_1$ appears to be more modular. This is due the high values of modularity, connected components and clustering coefficient. Authors in $G_1$ tend to write more in teams (modularity $.977 > .958$), separated from the main cluster 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 $G_1$.



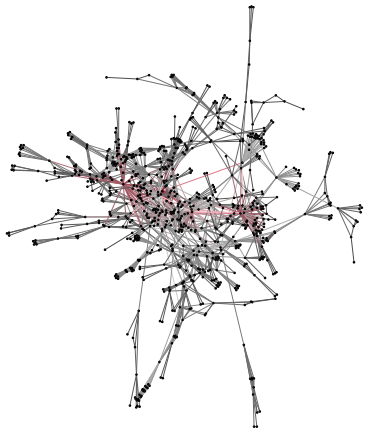
(a) $G_1$ network.



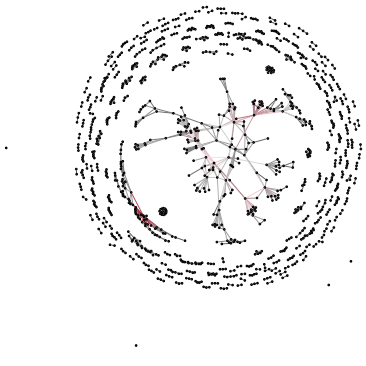
(b) $G_1$ larger connected component.



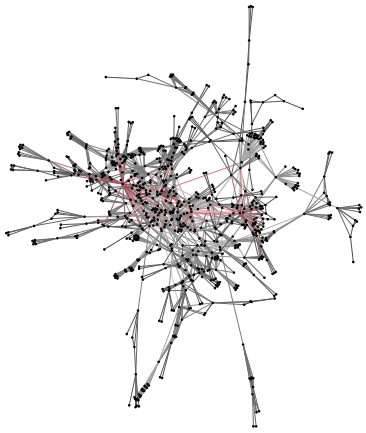
(c) $G_2$ network.



(d) $G_2$ larger connected component.



(e) $G_3$ network.



(f) $G_3$ larger connected component.

| | # Nodes | # Edges | # Isolated nodes | % Isolated nodes | # Connected components | Size of largest component | Av. degree | Modularity | Clustering coeff |
|---|---|---|---|---|---|---|---|---|---|
| Prisoner's Dilemma | 5394 | 10397 | 176 | 3.3 | 1356 | 815 | 3.855 | 0.977 | 0.708 |
| Auction Games | 5165 | 7861 | 256 | 5.0 | 1272 | 1348 | 3.044 | 0.957 | 0.622 |
| Price of Anarchy | 1155 | 1953 | 4 | 0.3 | 245 | 222 | 3.382 | 0.965 | 0.712 |

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

| | # Nodes | # Edges | # Isolated nodes | % Isolated nodes | # Connected components | Size of largest component | Av. degree | Modularity | Clustering coeff |
|---|---|---|---|---|---|---|---|---|---|
| Prisoner's Dilemma | 815 | 2300 | 0 | 0.0 | 1 | 815 | 5.644 | 0.856 | 0.775 |
| Auction Games | 1348 | 3158 | 0 | 0.0 | 1 | 1348 | 4.685 | 0.858 | 0.699 |
| Price of Anarchy | 222 | 521 | 0 | 0.0 | 1 | 222 | 4.694 | 0.817 | 0.711 |

Table 6: 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.

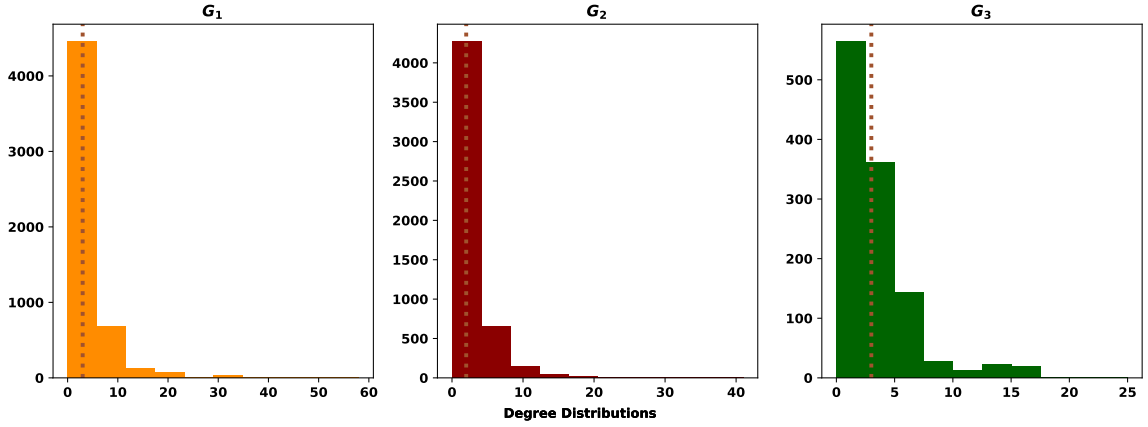
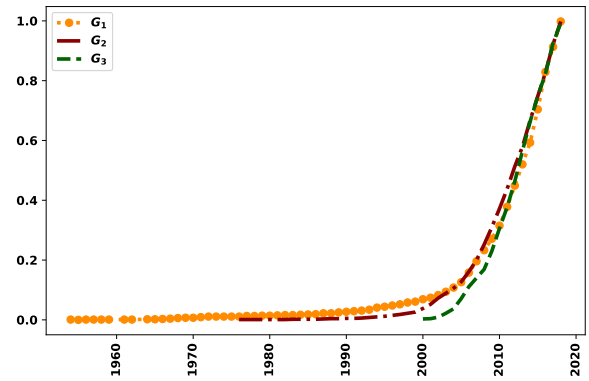


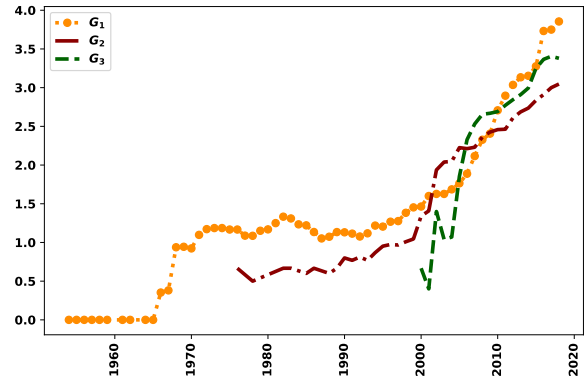

Figure 7: Degree distribution for networks $G_1$ - $G_3$.

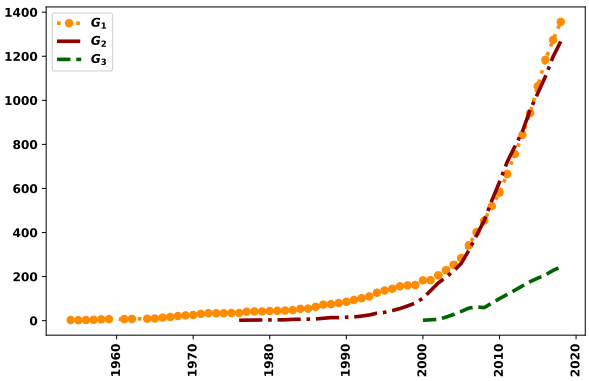
The growth of the networks over time is studied by constructing the network cumulatively with a year (of publication) interval. A total of 64 subgraphs over 64 periods, starting in 1950, were created and all the collaborative metrics for each subgraph have been calculated. These are given by Table 7. Similar to the results of [**?**], 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.

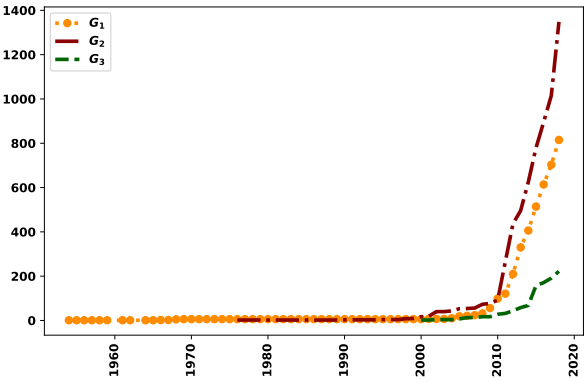
In Figure **??** the normalised number of nodes, which is calculated by dividing by the total number of nodes in $G_1$ (5394 ), is shown. A steep increase to the size of the network is spotted around the 2000s, this is was also briefly comment upon in [**?**]. However, that increase does not appear only 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 **??**. This is evidence that something else shocked the academic community around that time.

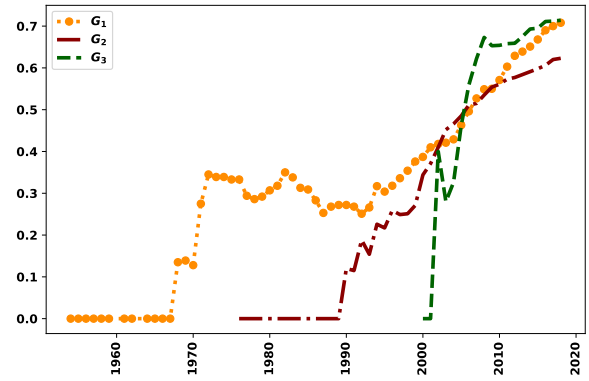
The change of the average degree over time has also been calculated for the subgraphs and is shown in Figure **??**. The growth of the average degree appears to differ between the networks. In $G_1$ the first publications were single author ones, followed by a steep increase to a degree of 1 just before 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.

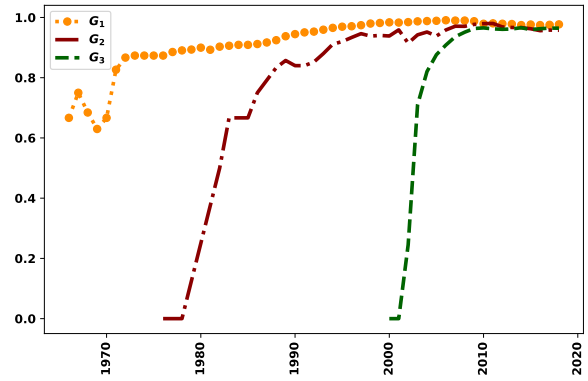

$G_1$
$G_2$
$G_3$
1.0
0.8
0.6
0.4
0.2
0.0
1960
1970
1980
1990
2000
2010
2020

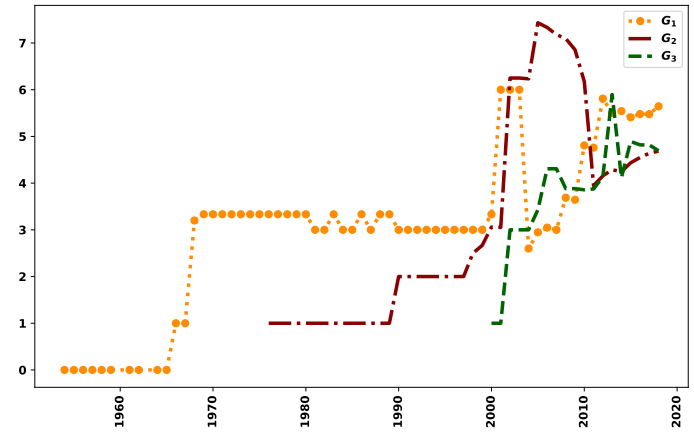

$G_1$
$G_2$
$G_3$
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1960
1970
1980
1990
2000
2010
2020

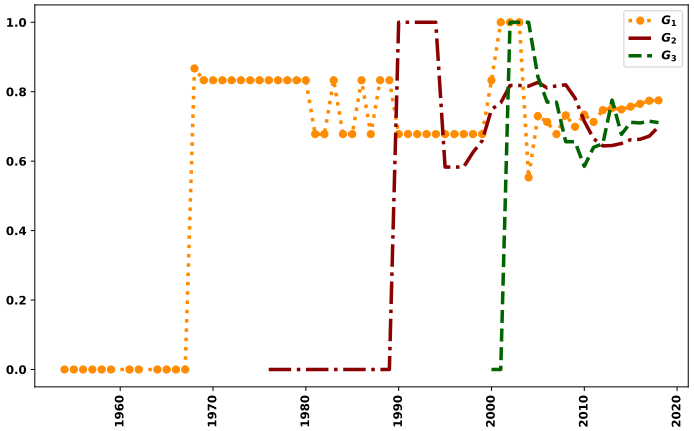

$G_1$
$G_2$
$G_3$
1400
1200
1000
800
600
400
200
0
1960
1970
1980
1990
2000
2010
2020

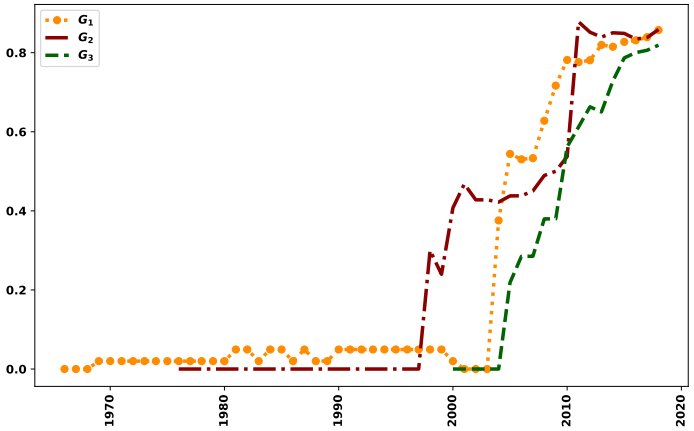

$G_1$
$G_2$
$G_3$
1400
1200
1000
800
600
400
200
0
1960
1970
1980
1990
2000
2010
2020


$G_1$
$G_2$
$G_3$
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1960
1970
1980
1990
2000
2010
2020


$G_1$
$G_2$
$G_3$
1.0
0.8
0.6
0.4
0.2
0.0
1970
1980
1990
2000
2010
2020


$G_1$
$G_2$
$G_3$
0
1
2
3
4
5
6
7
1960
1970
1980
1990
2000
2010
2020


$G_1$
$G_2$
$G_3$
0.0
0.2
0.4
0.6
0.8
1.0
1960
1970
1980
1990
2000
2010
2020


$G_1$
$G_2$
$G_3$
0.0
0.2
0.4
0.6
0.8
1970
1980
1990
2000
2010
2020

| | # Nodes | # Edges | # Isolated nodes | % Isolated nodes | # Connected components | Size of largest component | Av. degree | Modularity | Clustering coeff |
|---|---|---|---|---|---|---|---|---|---|
| 1954 - 1950 | 3 | 0 | 3 | 100.0 | 3 | 1 | 0.000 | - | 0.000 |
| 1954 - 1955 | 2 | 0 | 2 | 100.0 | 2 | 1 | 0.000 | - | 0.000 |
| 1955 - 1956 | 3 | 0 | 3 | 100.0 | 3 | 1 | 0.000 | - | 0.000 |
| 1956 - 1957 | 4 | 0 | 4 | 100.0 | 4 | 1 | 0.000 | - | 0.000 |
| 1957 - 1958 | 6 | 0 | 6 | 100.0 | 6 | 1 | 0.000 | - | 0.000 |
| 1958 - 1959 | 7 | 0 | 7 | 100.0 | 7 | 1 | 0.000 | - | 0.000 |
| 1959 - 1961 | 7 | 0 | 7 | 100.0 | 7 | 1 | 0.000 | - | 0.000 |
| 1961 - 1962 | 8 | 0 | 8 | 100.0 | 8 | 1 | 0.000 | - | 0.000 |
| 1962 - 1964 | 9 | 0 | 9 | 100.0 | 9 | 1 | 0.000 | - | 0.000 |
| 1964 - 1965 | 10 | 0 | 10 | 100.0 | 10 | 1 | 0.000 | - | 0.000 |
| 1965 - 1966 | 17 | 3 | 11 | 64.7 | 14 | 2 | 0.353 | 0.666667 | 0.000 |
| 1966 - 1967 | 21 | 4 | 13 | 61.9 | 17 | 2 | 0.381 | 0.75 | 0.000 |
| 1967 - 1968 | 32 | 15 | 13 | 40.6 | 21 | 5 | 0.938 | 0.684444 | 0.135 |
| 1968 - 1969 | 36 | 17 | 16 | 44.4 | 24 | 6 | 0.944 | 0.629758 | 0.139 |
| 1969 - 1970 | 39 | 18 | 17 | 43.6 | 26 | 6 | 0.923 | 0.666667 | 0.128 |
| 1970 - 1971 | 51 | 28 | 18 | 35.3 | 31 | 6 | 1.098 | 0.826531 | 0.275 |
| 1971 - 1972 | 58 | 34 | 19 | 32.8 | 34 | 6 | 1.172 | 0.866782 | 0.345 |
| 1972 - 1973 | 59 | 35 | 18 | 30.5 | 34 | 6 | 1.186 | 0.873469 | 0.339 |
| 1973 - 1974 | 59 | 35 | 18 | 30.5 | 34 | 6 | 1.186 | 0.873469 | 0.339 |
| 1974 - 1975 | 60 | 35 | 19 | 31.7 | 35 | 6 | 1.167 | 0.873469 | 0.333 |
| 1975 - 1976 | 60 | 35 | 19 | 31.7 | 35 | 6 | 1.167 | 0.873469 | 0.333 |
| 1976 - 1977 | 68 | 37 | 23 | 33.8 | 41 | 6 | 1.088 | 0.885318 | 0.294 |
| 1977 - 1978 | 70 | 38 | 23 | 32.9 | 42 | 6 | 1.086 | 0.890582 | 0.286 |
| 1978 - 1979 | 73 | 42 | 23 | 31.5 | 42 | 6 | 1.151 | 0.893424 | 0.292 |
| 1979 - 1980 | 77 | 45 | 25 | 32.5 | 44 | 6 | 1.169 | 0.899753 | 0.307 |
| 1980 - 1981 | 80 | 50 | 26 | 32.5 | 45 | 6 | 1.250 | 0.8928 | 0.318 |
| 1981 - 1982 | 84 | 56 | 26 | 31.0 | 46 | 6 | 1.333 | 0.903061 | 0.350 |
| 1982 - 1983 | 87 | 57 | 27 | 31.0 | 48 | 6 | 1.310 | 0.906125 | 0.338 |
| 1983 - 1984 | 94 | 58 | 32 | 34.0 | 54 | 6 | 1.234 | 0.909037 | 0.313 |
| 1984 - 1985 | 95 | 58 | 33 | 34.7 | 55 | 6 | 1.221 | 0.909037 | 0.309 |
| 1985 - 1986 | 104 | 59 | 40 | 38.5 | 63 | 6 | 1.135 | 0.911807 | 0.283 |
| 1986 - 1987 | 116 | 61 | 48 | 41.4 | 73 | 6 | 1.052 | 0.916958 | 0.253 |
| 1987 - 1988 | 121 | 65 | 48 | 39.7 | 75 | 6 | 1.074 | 0.924497 | 0.268 |
| 1988 - 1989 | 134 | 76 | 47 | 35.1 | 80 | 6 | 1.134 | 0.937673 | 0.272 |
| 1989 - 1990 | 145 | 82 | 49 | 33.8 | 86 | 6 | 1.131 | 0.944676 | 0.272 |
| 1990 - 1991 | 158 | 88 | 53 | 33.5 | 94 | 6 | 1.114 | 0.950413 | 0.268 |
| 1991 - 1992 | 169 | 91 | 59 | 34.9 | 102 | 6 | 1.077 | 0.953025 | 0.251 |
| 1992 - 1993 | 186 | 104 | 62 | 33.3 | 110 | 6 | 1.118 | 0.95932 | 0.266 |
| 1993 - 1994 | 220 | 134 | 72 | 32.7 | 127 | 6 | 1.218 | 0.965471 | 0.317 |
| 1994 - 1995 | 239 | 144 | 74 | 31.0 | 137 | 6 | 1.205 | 0.969329 | 0.304 |
| 1995 - 1996 | 257 | 163 | 77 | 30.0 | 145 | 6 | 1.268 | 0.970831 | 0.318 |
| 1996 - 1997 | 279 | 178 | 81 | 29.0 | 156 | 6 | 1.276 | 0.974309 | 0.336 |
| 1997 - 1998 | 311 | 215 | 65 | 20.9 | 160 | 6 | 1.383 | 0.979773 | 0.354 |
| 1998 - 1999 | 329 | 239 | 58 | 17.6 | 162 | 6 | 1.453 | 0.981741 | 0.376 |
| 1999 - 2000 | 373 | 273 | 67 | 18.0 | 183 | 6 | 1.464 | 0.983778 | 0.387 |
| 2000 - 2001 | 400 | 320 | 54 | 13.5 | 184 | 7 | 1.600 | 0.983066 | 0.410 |
| 2001 - 2002 | 450 | 366 | 61 | 13.6 | 206 | 7 | 1.627 | 0.984547 | 0.418 |
| 2002 - 2003 | 509 | 414 | 58 | 11.4 | 229 | 7 | 1.627 | 0.987083 | 0.421 |
| 2003 - 2004 | 580 | 489 | 58 | 10.0 | 253 | 10 | 1.686 | 0.988052 | 0.429 |
| 2004 - 2005 | 679 | 599 | 57 | 8.4 | 284 | 19 | 1.764 | 0.98891 | 0.463 |
| 2005 - 2006 | 854 | 806 | 66 | 7.7 | 342 | 21 | 1.888 | 0.990724 | 0.496 |
| 2006 - 2007 | 1056 | 1117 | 76 | 7.2 | 402 | 24 | 2.116 | 0.989663 | 0.527 |
| 2007 - 2008 | 1255 | 1460 | 85 | 6.8 | 454 | 32 | 2.327 | 0.989753 | 0.549 |
| 2008 - 2009 | 1462 | 1759 | 104 | 7.1 | 520 | 56 | 2.406 | 0.987517 | 0.550 |
| 2009 - 2010 | 1700 | 2301 | 114 | 6.7 | 581 | 99 | 2.707 | 0.979084 | 0.571 |
| 2010 - 2011 | 2040 | 2954 | 121 | 5.9 | 665 | 121 | 2.896 | 0.980477 | 0.603 |
| 2011 - 2012 | 2422 | 3676 | 126 | 5.2 | 756 | 210 | 3.036 | 0.979127 | 0.629 |
| 2012 - 2013 | 2807 | 4398 | 138 | 4.9 | 843 | 330 | 3.134 | 0.978211 | 0.639 |
| 2013 - 2014 | 3199 | 5044 | 148 | 4.6 | 942 | 406 | 3.153 | 0.97468 | 0.651 |
| 2014 - 2015 | 3798 | 6221 | 159 | 4.2 | 1064 | 514 | 3.276 | 0.976572 | 0.668 |
| 2015 - 2016 | 4472 | 8344 | 169 | 3.8 | 1184 | 614 | 3.732 | 0.974694 | 0.690 |
| 2016 - 2017 | 4925 | 9235 | 173 | 3.5 | 1274 | 703 | 3.750 | 0.975539 | 0.700 |
| 2017 - 2018 | 5385 | 10379 | 176 | 3.3 | 1356 | 815 | 3.855 | 0.977076 | 0.708 |

Table 7: Collaborativeness metrics for cumulative graphs.

| | # Nodes | # Edges | # Isolated nodes | % Isolated nodes | # Connected components | Size of largest component | Av. degree | Modularity | Clustering coeff |
|---|---|---|---|---|---|---|---|---|---|
| 1954 - 1950 | 1 | 0 | 1 | 100.0 | 1 | 1 | 0.000 | - | 0.000 |
| 1954 - 1955 | 1 | 0 | 1 | 100.0 | 1 | 1 | 0.000 | - | 0.000 |
| 1955 - 1956 | 1 | 0 | 1 | 100.0 | 1 | 1 | 0.000 | - | 0.000 |
| 1956 - 1957 | 1 | 0 | 1 | 100.0 | 1 | 1 | 0.000 | - | 0.000 |
| 1957 - 1958 | 1 | 0 | 1 | 100.0 | 1 | 1 | 0.000 | - | 0.000 |
| 1958 - 1959 | 1 | 0 | 1 | 100.0 | 1 | 1 | 0.000 | - | 0.000 |
| 1959 - 1961 | 1 | 0 | 1 | 100.0 | 1 | 1 | 0.000 | - | 0.000 |
| 1961 - 1962 | 1 | 0 | 1 | 100.0 | 1 | 1 | 0.000 | - | 0.000 |
| 1962 - 1964 | 1 | 0 | 1 | 100.0 | 1 | 1 | 0.000 | - | 0.000 |
| 1964 - 1965 | 1 | 0 | 1 | 100.0 | 1 | 1 | 0.000 | - | 0.000 |
| 1965 - 1966 | 2 | 1 | 0 | 0.0 | 1 | 2 | 1.000 | 0 | 0.000 |
| 1966 - 1967 | 2 | 1 | 0 | 0.0 | 1 | 2 | 1.000 | 0 | 0.000 |
| 1967 - 1968 | 5 | 8 | 0 | 0.0 | 1 | 5 | 3.200 | 0 | 0.867 |
| 1968 - 1969 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1969 - 1970 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1970 - 1971 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1971 - 1972 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1972 - 1973 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1973 - 1974 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1974 - 1975 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1975 - 1976 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1976 - 1977 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1977 - 1978 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1978 - 1979 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1979 - 1980 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1980 - 1981 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1981 - 1982 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1982 - 1983 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1983 - 1984 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1984 - 1985 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1985 - 1986 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1986 - 1987 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1987 - 1988 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1988 - 1989 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 1989 - 1990 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1990 - 1991 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1991 - 1992 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1992 - 1993 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1993 - 1994 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1994 - 1995 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1995 - 1996 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1996 - 1997 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1997 - 1998 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1998 - 1999 | 6 | 9 | 0 | 0.0 | 1 | 6 | 3.000 | 0.0493827 | 0.678 |
| 1999 - 2000 | 6 | 10 | 0 | 0.0 | 1 | 6 | 3.333 | 0.02 | 0.833 |
| 2000 - 2001 | 7 | 21 | 0 | 0.0 | 1 | 7 | 6.000 | 0 | 1.000 |
| 2001 - 2002 | 7 | 21 | 0 | 0.0 | 1 | 7 | 6.000 | 0 | 1.000 |
| 2002 - 2003 | 7 | 21 | 0 | 0.0 | 1 | 7 | 6.000 | 0 | 1.000 |
| 2003 - 2004 | 10 | 13 | 0 | 0.0 | 1 | 10 | 2.600 | 0.37574 | 0.553 |
| 2004 - 2005 | 19 | 28 | 0 | 0.0 | 1 | 19 | 2.947 | 0.544005 | 0.730 |
| 2005 - 2006 | 21 | 32 | 0 | 0.0 | 1 | 21 | 3.048 | 0.530273 | 0.713 |
| 2006 - 2007 | 24 | 36 | 0 | 0.0 | 1 | 24 | 3.000 | 0.533179 | 0.678 |
| 2007 - 2008 | 32 | 59 | 0 | 0.0 | 1 | 32 | 3.688 | 0.627837 | 0.732 |
| 2008 - 2009 | 56 | 102 | 0 | 0.0 | 1 | 56 | 3.643 | 0.716792 | 0.699 |
| 2009 - 2010 | 99 | 238 | 0 | 0.0 | 1 | 99 | 4.808 | 0.781539 | 0.734 |
| 2010 - 2011 | 121 | 288 | 0 | 0.0 | 1 | 121 | 4.760 | 0.776054 | 0.713 |
| 2011 - 2012 | 210 | 610 | 0 | 0.0 | 1 | 210 | 5.810 | 0.7809 | 0.747 |
| 2012 - 2013 | 330 | 908 | 0 | 0.0 | 1 | 330 | 5.503 | 0.81959 | 0.753 |
| 2013 - 2014 | 406 | 1125 | 0 | 0.0 | 1 | 406 | 5.542 | 0.81494 | 0.749 |
| 2014 - 2015 | 514 | 1390 | 0 | 0.0 | 1 | 514 | 5.409 | 0.827358 | 0.757 |
| 2015 - 2016 | 614 | 1682 | 0 | 0.0 | 1 | 614 | 5.479 | 0.831149 | 0.765 |
| 2016 - 2017 | 703 | 1925 | 0 | 0.0 | 1 | 703 | 5.477 | 0.839165 | 0.774 |
| 2017 - 2018 | 815 | 2300 | 0 | 0.0 | 1 | 815 | 5.644 | 0.857237 | 0.775 |

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 **??** and **??** 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 **??**. 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 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 10: 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 11: 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 **??** and **??**, and more over for closeness centrality a violin plot is also given by Figure **??**. Several remarks are made from the centralities' distributions.

Regarding between centrality,

- None of the three distributions, Figure **??**, 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.

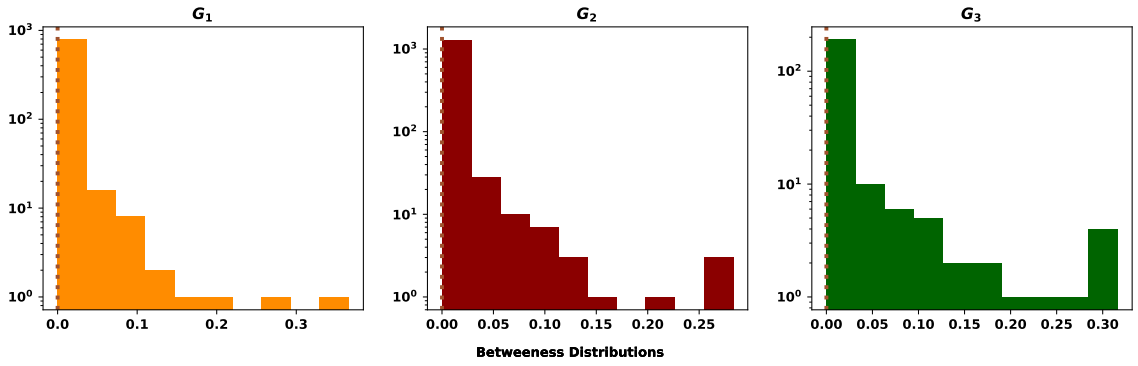
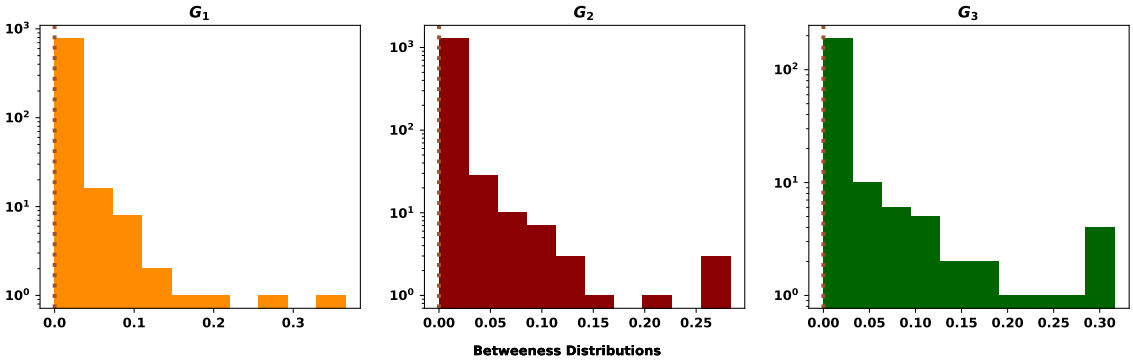



Figure 12: Betweenness centrality distributions $G_1 - G_3$.

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 **??**, 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 $G_1$.

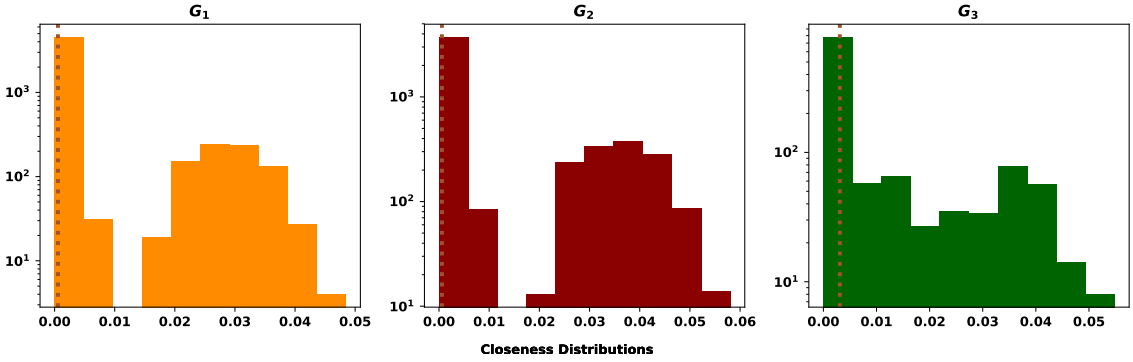
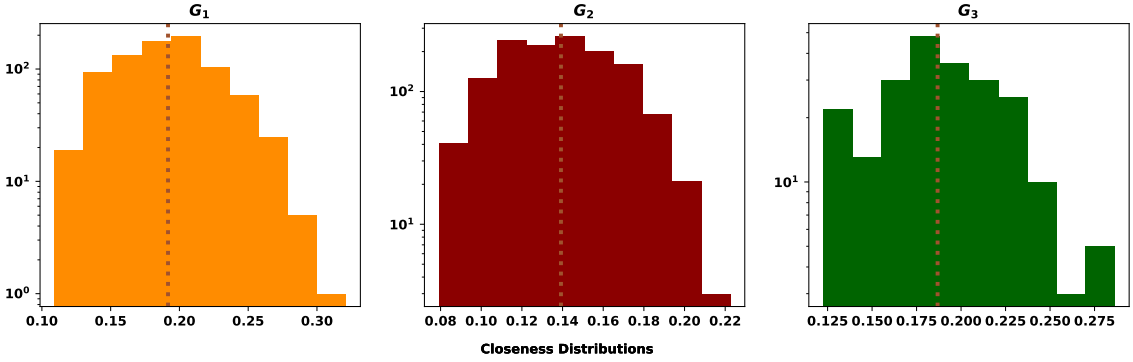



Figure 13: Closeness centrality distributions $G_1 - G_3$.




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 **??** 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

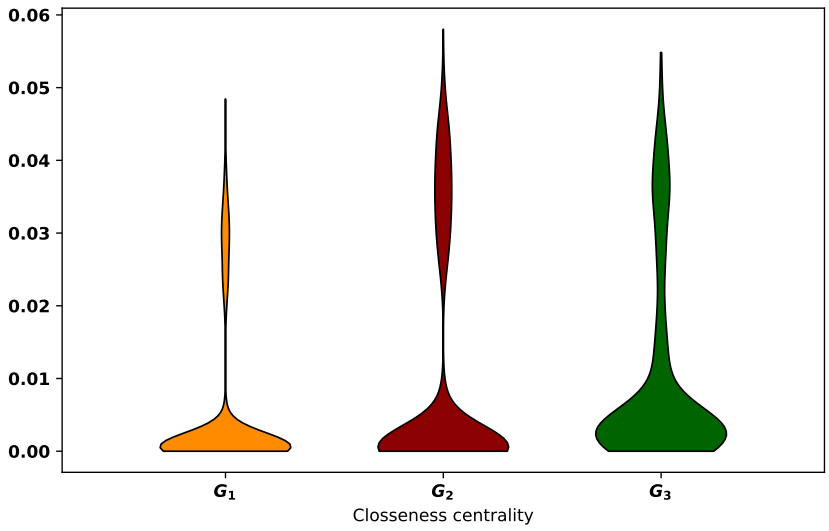



Figure 14: Violin plots of closeness centralities.

when authors influenced the field more than today, for example in 1972-1973. The centrality does not appear to follow a pattern, Figure **??**.

### 2.5 Conclusion

## 3 Acknowledges

- Networkx will also be used the following section to conduct our analysis.

# 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 historical 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.