# A bibliometric study of research topics, collaboration and influence in the field of the Iterated Prisoner's Dilemma

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

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

**Abstract**

This manuscript explores the research topics and collaborative behaviour of authors in the field of the Prisoner's Dilemma using topic modeling and a graph theoretic analysis of the co-authorship network. The analysis identified five research topics in the Prisoner's Dilemma which have been relevant of the course of time. These are human subject research, biological studies, strategies, evolutionary dynamics on networks and modeling problems as a Prisoner's Dilemma game. Moreover, the results demonstrated the Prisoner's Dilemma is a field of continued interest, and although it is a collaborative field, it is not necessarily more collaborative than other scientific fields. The co-authorship network suggests that authors are focused on their communities and not many connections across the communities are made. The Prisoner Dilemma authors also do not influence or gain much information by their connections, unless they are connected to a "main" group of authors.

## 1 Introduction

The Prisoner's Dilemma (PD) is a well known game used since its introduction in the 1950's [25] as a framework for studying the emergence of cooperation; a topic of continued interest for mathematical, social [60], biological [74] and ecological [80] sciences. This manuscript presents a bibliometric analysis of 2,420 published articles on the Prisoner's Dilemma between 1951 and 2018. It presents the dominant topics in the PD publications, which have been identified using Latent Dirichlet Allocation [14], and it explores the changes in the dominant topics over time. The collaborative behaviour of the field is explored using the co-authorship network, and furthermore, the Latent Dirichlet Allocation topic analysis is combined with the co-authorship network analysis to assess the relative influence of authors in these topics. Assessing the collaborative behaviour of the field of collaboration itself is the main aim of this work.

As discussed in [82], bibliometrics (the statistical analysis of published works originally described by [61]) has been used to support historical assumptions about the development of fields [62], identify connections between scientific growth and policy changes [20], develop a quantitative understanding of author order [68], and investigate the collaborative structure of an interdisciplinary field [47]. Most academic research is undertaken in the form of collaborative effort and as [43] points out, it is rational that two or more people have the potential to do better as a group than individually. Indeed this is the very premise of the Prisoner's Dilemma itself. Collaboration in groups has a long tradition in experimental sciences and it has be proven to be productive according to [23]. The number of collaborations can be different between research fields and understanding how collaborative a field is not always an easy task. Several studies tend to consider academic citations as a measure for these things. A blog post published by Nature [56] 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 measuring collaborative behaviour, and to studying the development of a field is to use the co-authorship network, as described in [47]. The co-authorship network has many advantages as several graph theoretic measures can be used as proxies to explain author relationships. For example the average degree of a node corresponds to the average number of an authors' collaborators, and clustering coefficient corresponds to the extent that two collaborators of an author also collaborate with each other. In [47], the approach was applied to analyse the development of the field "evolution of cooperation", and in [82] to identify the subdisciplines of the interdisciplinary field of "cultural evolution" and investigate trends in collaboration and productivity between these subdisciplines. Moreover, [46] examined the long-term impact of co-authorship with established, highly-cited scientists on the careers of junior researchers. This paper builds upon the work done by [47] and [82], and extends their methodology. In [47, 82], a data set from a single source, Web of Science, is considered whereas the data set described here, archived at [29], has been collected from five sources.

Latent Dirichlet Allocation (LDA) is a topic modeling technique proposed in [14] as a generative probabilistic model for discovering underlying topics in collections of data. Applications of the technique include detection in image data [3, 19] and detection in video [55, 79]. Nevertheless, LDA has been applied by several works on publication data for identifying the topic structure of a subject area. In [39], it was applied to the publications on mathematical education of the journals "Educational Studies in Mathematics" and "Journal for Research in Mathematics Education" to identify the dominant topics that each journal was publishing on. The topics of the North American library and Information Science dissertations were studied chronologically in [72], and the main topic of the scientific content presented at EvoLang conferences was identified in [13]. In [13] the LDA approach is combined with clustering and a co-authorship network analysis. A clustering analysis is applied to the LDA topics, and the co-authorship network is analysed as a whole where the clusters are only used to differentiate between the authors' topics. In comparison, this works applies LDA to identify dominant topics in the Prisoner's Dilemma fields and analyses the networks corresponding to these topics individually.

The methodology used in this manuscript (including the data collection) is covered in Section 2 and a preliminary analysis of the data set is presented in Section 3. This manuscript makes usage of the methodology and data to address the following questions:

1. What are the research topics of the Prisoner's Dilemma?
2. Is one topic currently more in fashion?
3. How have the research topics changed over the years?
4. Is the Prisoner's Dilemma a collaborative field?
5. Are some topics more collaborative than others?
6. Are there authors which benefit more from their position in the network?

Results regarding questions 1-3 are presented in Section 4 and questions 4-6 are addressed in Section 5. The results are summarised in Section 6.

## 2 Methodology

Academic articles are accessible through scholarly databases. Several databases and collections today offer access through an open application protocol interface (API). An API allows users to query directly a journal's database and bypass the graphical user interface. Interacting with an API has two phases: requesting and receiving. The request phase includes composing a url with the details of the request. For example, `http://export.arxiv.org/api/query?search_query=abs:prisoner'sdilemma&max_results=1` represents a request message. The first part of the request is the address of the API. In this example the address corresponds to the API of arXiv. The second part of the request contains the search arguments. In this example it is requested that the word 'prisoners dilemma' exists within the article's title. The format of the request message is different from API to API. The receive phase includes receiving a number of raw metadata of articles that satisfies the request message. The raw metadata are commonly received in extensive markup language (xml) or Javascript object notation (json) formats [57]. 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 [30]. 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 [30] allow users to collect articles from a list of APIs by specifying just a single keyword. Articles for which any of the terms "prisoner's dilemma", "prisoners dilemma", "prisoner dilemma", "prisoners evolution", "prisoner game theory" existed within the title, the abstract or the text are included in the analysis. Four prominent journals in the field and a preprint server were used as sources to collect data for this analysis:

- arXiv [52]; a repository of electronic preprints. It consists of scientific papers in the fields of mathematics, physics, astronomy, electrical engineering, computer science, quantitative biology, statistics, and quantitative finance, which all can be accessed online.
- PLOS [1]; a library of open access journals and other scientific literature under an open content license. It launched its first journal, PLOS Biology, in October 2003 and publishes seven journals, as of October 2015.
- IEEE Xplore Digital Library (IEEE) [38]; a research database for discovery and access to journal articles, conference proceedings, technical standards, and related materials on computer science, electrical engineering and electronics, and allied fields. It contains material published mainly by the Institute of Electrical and Electronics Engineers and other partner publishers.
- Nature [32]; a multidisciplinary scientific journal, first published on 4 November 1869. It was ranked the world's most cited scientific journal by the Science Edition of the 2010 Journal Citation Reports and is ascribed an impact factor of 40.137, making it one of the world's top academic journals.
- Springer [53]; a leading global scientific publisher of books and journals. It publishes close to 500 academic and professional society journals.

The data set has been archived and is available at [29]. Note that the latest data collection was performed on the 30$^{\text{th}}$ November 2018.

The relationship between the authors within a field will be modeled as a graph $G = (V_G, E_G)$ where $V_G$ is the set of nodes and $E_G$ is the set of edges. The set $V_G$ represents the authors and an edge connects two authors if and only if those authors have written together. This co-authorship network is constructed using the main data set [29] and the open source package [34]. The PD network is denoted as $G$ where the number of unique authors $|V(G)|$ is 4226 and $|E(G)|$ is 7642 . All authors' names were formatted as their first name and last name (i.e. Martin A. Nowak to Martin Nowak). This was done to avoid errors such as Martin A. Nowak and Martin Nowak being treated as a different person. There are some authors for which only their first initial was found. These entries are left as such.

The collaborativeness of the authors will be analysed using measures such as, isolated nodes, connected components, clustering coefficient, communities, modularity and average degree. These measures show the number of connections authors can have and how strongly connected these people are. The number of isolated nodes is the number of nodes that are not connected to another node, thus the number of authors that have published alone. The average degree denotes the average number of neighbours for each nodes, i.e. the average number of collaborations between the authors. A connected component is a maximal set of nodes such that each pair of nodes is connected by a path [22]. The number of connected components as well as the size of the largest connected component in the network are reported. The size of the largest connected component represents the scale of the central cluster of the entire network, as will be discussed in the analysis section. Clustering coefficient and modularity are also calculated. The clustering coefficient, defined as 3 times the number of triangles on the graph divided by the number of connected triples of nodes, is a local measure of the degree to which nodes in a graph tend to cluster together in a clique [22]. It shows to which extent the collaborators of an author also write together. In comparison, modularity is a global measure designed to measure the strength of division of a network into communities. The number of communities will be reported using the Clauset-Newman-Moore method [18]. Also the modularity index is calculated using the Louvain method described in [15]. The value of the modularity index can vary between $[-1, 1]$, a high value of modularity corresponds to a structure where there are dense connections between the nodes within communities but sparse connections between nodes in different communities. That means that there are many sub communities of authors that write together but not across communities. Two further points are aimed to be explored in this work, (1) which people control the flow of information; as in which people influence the field the most and (2) which are the authors that gain the most from the influence of the field. To measure these concepts centrality measures are going to be used. Centrality measures are often used to understand different aspects of social networks [44]. Two centrality measures have been chosen for this paper and these are closeness and betweenness centrality.

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

The articles contained in the data set ([29]) will be classified into research topics using LDA an unsupervised machine learning technique designed to summarize large collections of documents by a small number of conceptually connected topics or themes [14, 31]. The documents are the articles' abstracts and LDA was carried out using [63]. In LDA, each document/abstract is represented by a distribution over topics, and the topics themselves are represented by a distribution over words. More specifically, each topics is described by weights associated with words and each document by the probabilities of belonging to a specific topic. The probability of a document belonging to a topic is referred to as the percentage contribution denoted as $c$. For example the words and their associated weights for two topics A and B could be:

- Topic A: $0.039\times$"cooperation", $0.028\times$"study" and $0.026\times$"human".
- Topic B: $0.020\times$"cooperation", $0.028\times$"agents" and $0.026\times$"strategies".

The percentage contribution for a document with abstract "The study of cooperation in humans" has a $c_A = 0.039 + 0.028 + 0.026 = 0.093$ and $c_B = .020 + 0.0 + 0.0 = 0.020$. The topic to which a document is assigned to is based on the highest percentage contribution denoted as $c^*$. For the given example the dominant topic is Topic A $c^* = c_A$. LAD requires that the number of topics is specified in advance before running the algorithm. The number of topics can be chosen using the coherence value [65] or through subjective minimisation of the overlapping keywords between two topics. Both these approaches will be used in this work.

Several of the approaches described in this section have previously been carried out in [13, 47, 72, 82], the novelty here is combining the approaches as well as applying them to a new data set. A preliminary analysis of the data set is presented in the following section.

# 3 Preliminary Analysis

The data set [29] consists of 2422 articles with unique titles. In case of duplicates the preprint version of an article (collected from arXiv) was dropped. Similarly to [47], 76 articles have not been collected from the aforementioned APIs but have been manually added because they are of interest. Examples of such papers include [25] the first publication on the PD, [58, 70] two well cited articles in the field, and a series of works from Robert Axelrod [6, 7, 8, 9, 64] a leading author of the field. A more detailed summary of the articles' provenance is given by Table 1. Only 3% of the data set consists of articles that were manually added and 27% of the articles were collected from arXiv. The average number of publications is also included in Table 1. Overall an average of 43 articles are published per year on the topic. The most significant contribution to this appears to be from arXiv with 11 articles per year, followed by Springer with 9 and PLOS with 8.

| | Number of Articles | Percentage % | Year of first publication | Average number of publications per year |
|---|---|---|---|---|
| IEEE | 294 | 12.14% | 1973 | 5 |
| Manual | 76 | 3.14% | 1951 | 1 |
| Nature | 436 | 18.00% | 1959 | 8 |
| PLOS | 477 | 19.69% | 2005 | 8 |
| Springer | 533 | 22.01% | 1966 | 9 |
| arXiv | 654 | 27.00% | 1993 | 11 |
| Overall | 2470 | 100.00% | 1951 | 43 |

Table 1: Summary of [29] per provenance.

The data handled here is in fact a time series from the 1950s, the formulation of the game, until 2018 (Figure 1). Two observations can be made from Figure 1.

1. There is a steady increase of the number of publications since the 1980s and the introduction of computer tournaments [9] (work by Robert Axelrod).
2. There is a decrease in 2017-2018. This is due to our data set being incomplete. Articles that have been written in 2017-2018 have either not being published or were not retrievable by the APIs at the time of the last data collection.

These observations can be confirmed by studying the time series. Using [41], an exponential distribution is fitted to the data. The fitted model can be used to forecast the behaviour of the field for the next 5 years. Even though the time series has indicated a slight decrease, the model forecasts that the number of publications will keep increasing, thus demonstrating that the field of the PD continues to attract academic attention.

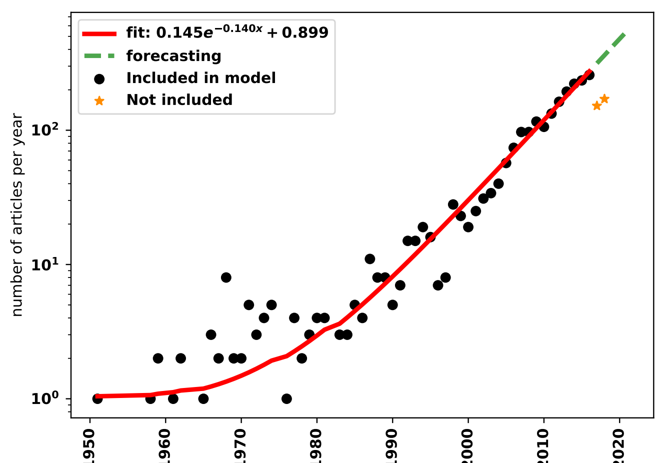



Figure 1: Number of articles published on the PD 1951-2018 (on a log scale), with a fitted exponential line, and a forecast for 2017-2022.

There are a total of 4226 authors in the data set ([29]) and several of these authors have had multiple publications collected from the data collection process. The highest number of articles collected for an author is 83 publications for Matjaz Perc. The distribution of the number of papers per author is given by Figure 2, and it can be seen that Matjaz Perc is an outlier. More specifically, most authors have 1 to 6 publications in the data set.

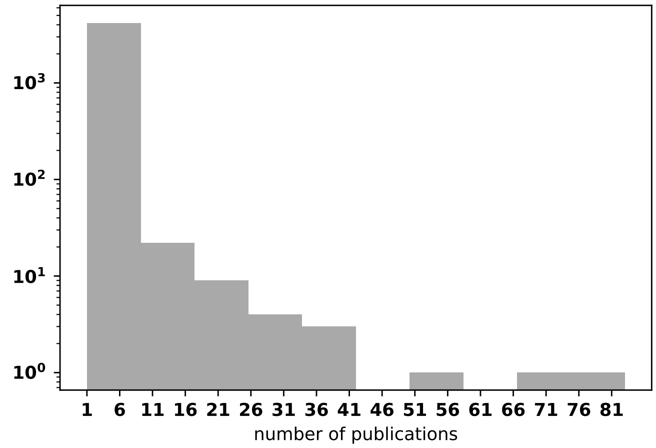



Figure 2: Distribution of number of papers per author (on a log scale).

The overall Collaboration Index (CI) or the average number of authors on multi-authored papers is 3.2, thus on average a non single author publication in the PD has 3 authors. This appears to be quite standard compared to other fields such as cultural evolution [82], Astronomy and Astrophysics, Genetics and Heredity, Nuclear and Particle Physics as reported by [50]. There are only a total of 545 publications with a single author, which corresponds to the 22% of the papers. It appears that academic publications tend to be undertaken in the form of collaborative effort, which is in line with the claim of [43]. From Figure 3 the trend of CI over the years is given. There are some peaks in the early years 1969 and 1980, however, a steady increase appears to happen after 2004. This could be an effect of better communication tools being introduced around that time which enabled more collaborations between researchers.

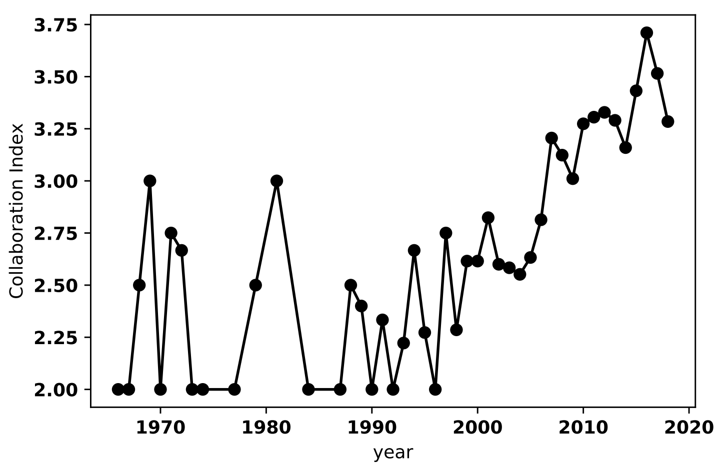



Figure 3: Collaboration index over time.

The collaborativeness of the authors is explored in more detail in Section 5 using the co-authorship network. The collaborative behaviour and relative influence of authors will also be explored in co-authorship networks which correspond to their publications research topics. These topics are presented in the next section.

# 4 Research topics in the Prisoner's Dilemma research

In order to identify the topics which are being discussed in the field of the PD, the LDA algorithm implemented in [63] is applied to the abstracts of the data set. As mentioned before, the number of topics, which will be denoted as $n$, needs to be specified before running the algorithm. The appropriate number of topics is chosen based on the coherence value [65]. Figure 4 gives the coherence values of 18 models where $n \in \{2, 3, \ldots, 19\}$, and it can be seen than the most appropriate number of topics is 6 with a coherence value of 0.418.

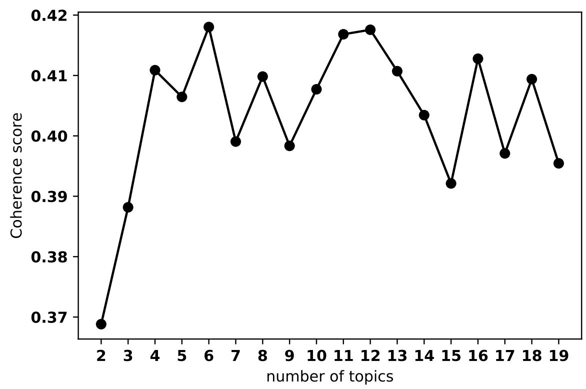



Figure 4: Coherence for LDA models over the number of topics.

An LDA model outputs an $N \times n$ matrix - $N$ rows for $N$ abstracts and $n$ columns for $n$ topics. The cells contain the percentage contributions for each topic for each abstract, $c_i^j$ for $i \in \{1, 2, \ldots, n\}$ for $j \in \{1, 2, \ldots, N\}$. In essence, LDA maps every paper to a vector space of dimension the number of topics. In the case of 6 topics it is difficult to visualise the clustering of topics. To overcome this a dimensionality reduction approach called t-Distributed Stochastic Neighbor Embedding (t-SNE) [48] is applied to the LDA model outputs. More specifically, t-SNE is used to reduce the dimensions of each $c^j$ from $n$ to 2. Figure 5, gives the visualisation of LDA for $n = 6$. Each point represents a single document and its color corresponds to the topic with the highest percentage contribution. The documents which are clustered together have a similar percentage contribution distribution over the topics.

Even though the LDA model with $n = 6$ has the highest coherence value, Figure 5 shows that documents of the same topic are closer to documents from other topics than each other. For example the documents of topic 2 are divided into two clusters. The one cluster is closer to documents from topic 4 and the other has a few documents closer to topic 1. In the case of $n = 6$ topic 4 appears to be on "evolution of cooperation on networks", and the papers from topic 2 surrounded from topic 4 include the articles "Evolutionary prisoner's dilemma game on hierarchical lattices" [76] and "Social evolution in structured populations" [21]. Publications that clearly also fit topic 4.

In comparison, 6 gives the visualisation of LDA $n = 5$ where the separation of the documents is more clear. Though several models, Figure 4, have a higher coherence value than the LDA model with $n = 5$, the separation of topics is not as clear for any model as it is for $n = 5$. Thus, $n = 5$ is chosen to carry out the analysis of this work, and moreover the LDA model for $n = 5$ has a coherence value 0.406 which is close to 0.418.

**What are the research topics of the Prisoner's Dilemma?**

For $n = 5$ the articles are clustered and assigned to their dominant topic, based on the highest percentage contribution. The keywords associated with a topic, the most representative article of the topic (based on the percentage contribution) and its academic reference are given by Table 2. The topics are labelled as A, B, C, D and E, and more specifically:

- Based on the keywords associated with Topic A, and the most representative article, Topic A appears to be about **human subject research**. Several publications assigned to the topic study the PD by setting

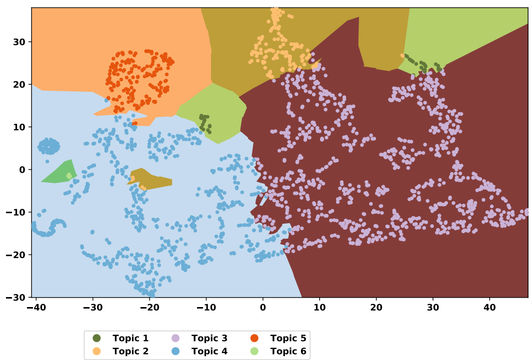



Figure 5: Visualisation of LDA with $n = 6$ on 2 dimensions.

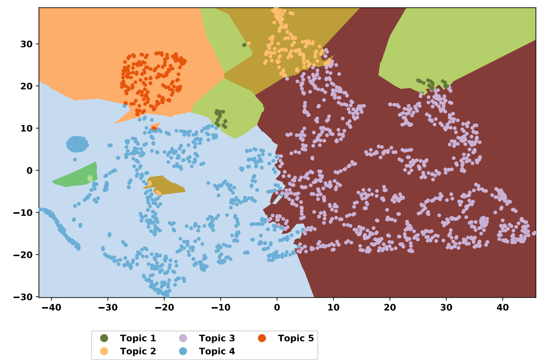



Figure 6: Visualisation of LDA with $n = 5$ on 2 dimensions.

experiments and having human participants simulate the game instead of computer simulations. These articles include [51] which showed that prosocial behavior increased with the age of the participants, [45] which studied the difference in cooperation between high-functioning autistic and typically developing children, [54] explored the gender effect in highschool students and [12] explored the effect of facial expressions of individuals.

- Though it is not immediate from the keywords associated with Topic B, investigating the papers assigned to the topic indicate that it is focused on **biological studies**. Papers assigned to the topic include papers which apply the PD to genetics [66, 69], to the study of tumours [4, 67] and viruses [75]. Other works include how phenotype affinity can affect the emergence of cooperation [81] and modeling bacterial communities as a spatial structured social dilemma.
- Based on the keywords and the most representative article Topic C appears to include publications on PD **strategies**. Publications in the topic include the introduction of new strategies [71], the search of optimality in strategies [10] and the training of strategies [40] with different representation methods. Moreover, publications that study the evolutionary stability of strategies [2] and introduced methods of differentiating between them [5] are also assigned to C.
- The keywords associated with Topic D clearly show that the topic is focused on **evolutionary dynamics on networks**. Publications include [37] which explored the robustness of cooperation on networks, [78] which studied the effect of a strategy's neighbourhood on the emergence of cooperation and [17] which explored the fixation probabilities of any two strategies is spatial structures.
- The publication assigned to Topic E are on **modeling problems as a PD game**. Though Topic B is also concerned with problems being formulated as a PD, it includes only biological problems. In comparison, the problems in Topic E include decision making in operational research [59], information sharing among members in a virtual team [24], the measurement of influence in articles based on citations [36] and the price spikes in electric power markets [33], and not on biological studies.

Note that the whilst for the choice of 5 topics the actual clustering is not subjective (the algorithm is determining the output) the interpretation above is.

**Five topics in the PD publications identified by the data set of this work are human subject research, biological studies, strategies, evolutionary dynamics on networks and modeling problems as a PD.**

These 5 topics nicely summarise the PD research. They highlight the interdisciplinarity of the field; how it brings together applied modeling of real world situations (Topic B and E) and more theoretical notions such as evolutionary dynamics and optimality of strategies.

**Is one topic currently more in fashion?**

| Dominant Topic | Topic Keywords | Most Representative Article Title | Reference | # Documents | % Documents |
|---|---|---|---|---|---|
| A | social, behavior, human, study, experiment, cooperative, cooperation, suggest, find, behaviour | Facing Aggression: Cues Differ for Female versus Male Faces | [26] | 496.0 | 0.2008 |
| B | individual, group, good, show, high, increase, punishment, cost, result, benefit | Genomic and Gene-Expression Comparisons among Phage-Resistant Type-IV Pilus Mutants of Pseudomonas syringae pathovar phaseolicola | [69] | 309.0 | 0.1251 |
| C | game, strategy, player, agent, dilemma, play, payoff, state, prisoner, equilibrium | Fingerprinting: Visualization and Automatic Analysis of Prisoner's Dilemma Strategies | [69] | 561.0 | 0.2271 |
| D | cooperation, network, population, evolutionary, evolution, interaction, dynamic, structure, cooperator, study | Influence of initial distributions on robust cooperation in evolutionary Prisoner's Dilemma | [16] | 556.0 | 0.2251 |
| E | model, theory, base, system, problem, paper, propose, information, provide, approach | Gaming and price spikes in electric power markets and possible remedies | [33] | 548.0 | 0.2219 |

Table 2: Keywords for each topic and the document with the most representative article for each topic.

Figure 7 gives the number of articles per topic over time. The topics appear to have had a similar trend over the years, with topics B and D having a later start. Following the introduction of a topic the publications in that topic have been increasing. There is no decreasing trend in any of the topics. All the topics have been publishing for years and they still attract the interest of academics. Thus, **there does not seem to be any given topic more or less in fashion**.

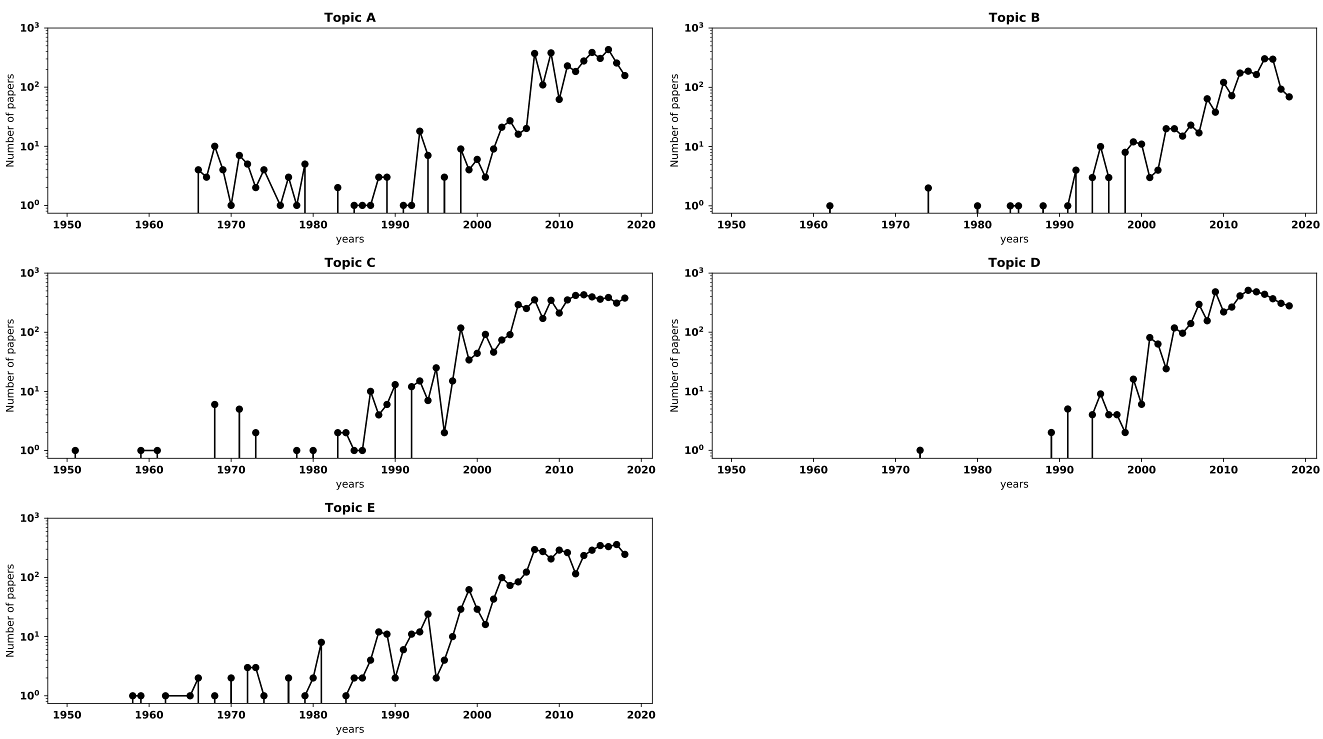



Figure 7: Number of articles per topic over the years (on a logged scale).

**How do the research topics change over the years?**

To gain a better understanding regarding the change in the topics over the years, LDA is applied to the cumulative data set over 8 time periods. These periods are 1951-1965, 1951-1973, 1951-1980, 1951-1988, 1951-1995, 1951-2003, 1951-2010, 1951-2018. The number of topics for each cumulative subset is chosen based on the coherence value and no objective approach is used. As a result, the period 1951-2018 has been assigned $n = 6$ which had the highest coherence value instead of 5. The chosen models for each period including the number of topics, their keywords and number of articles assigned to them are given by Table 3.

| Period | Topic | Topic Keywords | Num of Documents | Percentage of Documents |
|---|---|---|---|---|
| 1951-1965 | 1 | problem, technology, divert, euler, subsystem, requirement, trace, technique, system, untried | 3 | 0.375 |
| 1951-1965 | 2 | interpret, requirement, programme, evolution, article, increase, policy, system, trace, technology | 2 | 0.25 |
| 1951-1965 | 3 | equipment, agency, conjecture, development, untried, programme, trend, technology, weapon, technique | 1 | 0.125 |
| 1951-1965 | 4 | variation, celebrated, trend, untried, change, involve, month, technique, subsystem, research | 1 | 0.125 |
| 1951-1965 | 5 | give, good, modern, trace, technique, ambiguity, problem, trend, technology, system | 1 | 0.125 |
| 1951-1973 | 1 | study, shock, cooperative, money, part, vary, investigate, good, receive, equipment | 12 | 0.3243 |
| 1951-1973 | 2 | cooperation, level, significantly, sequence, reward, provoke, descriptive, principal, display, argue | 4 | 0.1081 |
| 1951-1973 | 3 | player, make, effect, triad, experimental, motivation, dominate, hypothesis, instruction, trend | 3 | 0.0811 |
| 1951-1973 | 4 | ss, sex, male, female, dyad, design, suggest, college, factor, tend | 3 | 0.0811 |
| 1951-1973 | 5 | result, research, format, change, operational, analysis, relate, understanding, decision, money | 2 | 0.0541 |
| 1951-1973 | 6 | condition, give, high, treatment, conflict, cc, real, original, replication, promote | 2 | 0.0541 |
| 1951-1973 | 7 | group, competitive, show, interpret, scale, compete, escalation, free, variable, individualistic | 2 | 0.0541 |
| 1951-1973 | 8 | outcome, strategy, choice, type, pdg, difference, dummy, conclude, compare, consistent | 2 | 0.0541 |
| 1951-1973 | 9 | game, difference, pair, approach, behavior, person, weapon, occur, advantaged, differential | 2 | 0.0541 |
| 1951-1973 | 10 | response, present, dilemma, influence, cooperate, bias, point, amount, participate, factor | 2 | 0.0541 |
| 1951-1973 | 11 | trial, problem, previous, involve, prisoner, experiment, follow, tit, increase, initial | 1 | 0.027 |
| 1951-1973 | 12 | matrix, behavior, rational, black, model, research, broad, distance, complex, trace | 1 | 0.027 |
| 1951-1973 | 13 | play, finding, individual, noncooperative, white, nature, race, ratio, represent, prisoner | 1 | 0.027 |
| 1951-1980 | 1 | play, trial, group, follow, white, interpret, scale, black, trend, small | 14 | 0.25 |
| 1951-1980 | 2 | outcome, level, effect, type, dyad, vary, pdg, participate, understanding, arise | 9 | 0.1607 |
| 1951-1980 | 3 | game, strategy, cooperation, significant, difference, sentence, text, occur, differential, hypothesis | 4 | 0.0714 |
| 1951-1980 | 4 | male, female, find, result, sex, subject, experimental, situation, treatment, computer | 4 | 0.0714 |
| 1951-1980 | 5 | research, problem, influence, matrix, format, model, analysis, year, crime, equipment | 4 | 0.0714 |
| 1951-1980 | 6 | condition, dilemma, bias, free, attempt, book, year, dummy, prison, design | 4 | 0.0714 |
| 1951-1980 | 7 | variable, result, factor, individual, ability, triad, half, migration, change, investigate | 3 | 0.0536 |
| 1951-1980 | 8 | show, present, suggest, rational, compete, approach, characteristic, examine, person, conduct | 3 | 0.0536 |
| 1951-1980 | 9 | behavior, high, finding, relate, obtain, assistance, ratio, good, weapon, competition | 3 | 0.0536 |
| 1951-1980 | 10 | ss, shock, money, competitive, part, difference, pair, amount, man, information | 3 | 0.0536 |
| 1951-1980 | 11 | player, conflict, theory, decision, determine, produce, maker, cooperate, specialist, programming | 2 | 0.0357 |
| 1951-1980 | 12 | study, prisoner, make, response, experiment, noncooperative, standard, separate, conclude, initial | 2 | 0.0357 |
| 1951-1980 | 13 | give, cooperative, choice, cognitive, real, operational, set, subject, ascribe, concern | 1 | 0.0179 |
| 1951-1988 | 1 | trial, difference, find, choice, significant, competitive, effect, triad, interact, occur | 24 | 0.2553 |
| 1951-1988 | 2 | ss, shock, money, pair, response, part, high, tit, receive, amount | 13 | 0.1383 |
| 1951-1988 | 3 | suggest, paper, case, debate, view, achieve, framework, natural, assumption, finitely | 10 | 0.1064 |
| 1951-1988 | 4 | prisoner, dilemma, behavior, model, present, involve, person, increase, trust, experiment | 8 | 0.0851 |
| 1951-1988 | 5 | game, player, show, approach, repeat, previous, move, tat, related, include | 8 | 0.0851 |
| 1951-1988 | 6 | cooperation, level, mutual, equilibrium, standard, provide, information, human, real, question | 6 | 0.0638 |
| 1951-1988 | 7 | play, result, male, subject, female, cooperative, sex, experimental, treatment, computer | 5 | 0.0532 |
| 1951-1988 | 8 | research, study, variable, ability, factor, conflict, matrix, year, student, interpret | 4 | 0.0426 |
| 1951-1988 | 9 | problem, group, small, scale, social, issue, large, base, bias, party | 4 | 0.0426 |
| 1951-1988 | 10 | game, strategy, outcome, type, cooperate, ethical, pdg, explain, dependent, separate | 4 | 0.0426 |
| 1951-1988 | 11 | give, condition, individual, major, dyad, behaviour, produce, conflict, assistance, collectively | 3 | 0.0319 |
| 1951-1988 | 12 | situation, iterate, statement, rational, card, side, paradox, true, consequence, front | 2 | 0.0213 |
| 1951-1988 | 13 | inflation, hypothesis, rate, run, change, demand, nominal, cost, output, growth | 2 | 0.0213 |
| 1951-1988 | 14 | theory, make, analysis, decision, system, examine, work, soft, lead, hard | 1 | 0.0106 |
| 1951-1995 | 1 | strategy, population, evolution, iterate, tit, opponent, evolve, dynamic, set, tat | 31 | 0.1732 |
| 1951-1995 | 2 | game, repeat, assumption, rule, person, equilibrium, general, finitely, indefinitely, analyze | 24 | 0.1341 |
| 1951-1995 | 3 | inflation, long, rate, hypothesis, run, policy, cost, nominal, demand, programming | 20 | 0.1117 |
| 1951-1995 | 4 | condition, outcome, trial, find, difference, cooperation, experiment, level, significant, response | 15 | 0.0838 |
| 1951-1995 | 5 | rational, result, receive, statement, money, paradox, shock, iterate, consequence, common | 14 | 0.0782 |
| 1951-1995 | 6 | cooperation, show, competitive, high, probability, conflict, simulation, altruism, yield, natural | 14 | 0.0782 |
| 1951-1995 | 7 | prisoner, dilemma, give, point, defect, form, cooperator, increase, relate, ethical | 10 | 0.0559 |
| 1951-1995 | 8 | player, give, decision, provide, cooperative, game, previous, pair, determine, interact | 9 | 0.0503 |
| 1951-1995 | 9 | play, cooperate, result, male, subject, female, time, relationship, suggest, student | 8 | 0.0447 |
| 1951-1995 | 10 | problem, group, theory, good, approach, society, large, scale, issue, level | 8 | 0.0447 |
| 1951-1995 | 11 | study, situation, behaviour, computer, argue, change, implication, characteristic, real, associate | 8 | 0.0447 |
| 1951-1995 | 12 | model, paper, behavior, examine, present, mutual, expectation, develop, type, variable | 7 | 0.0391 |
| 1951-1995 | 13 | make, research, system, analysis, choice, work, base, relation, world, wide | 6 | 0.0335 |
| 1951-1995 | 14 | individual, social, behavior, standard, choose, evolutionary, partner, payoff, defection, small | 5 | 0.0279 |
| 1951-2003 | 1 | game, player, dilemma, prisoner, theory, give, paper, make, group, problem | 151 | 0.4266 |
| 1951-2003 | 2 | cooperation, result, play, show, cooperate, condition, cooperative, high, level, time | 106 | 0.2994 |
| 1951-2003 | 3 | strategy, model, agent, study, behavior, individual, population, evolutionary, state, player | 97 | 0.274 |
| 1951-2010 | 1 | model, theory, paper, base, make, present, problem, provide, human, decision | 325 | 0.3454 |
| 1951-2010 | 2 | game, strategy, player, agent, play, dilemma, system, behavior, show, state | 322 | 0.3422 |
| 1951-2010 | 3 | cooperation, network, study, population, individual, evolutionary, social, evolution, interaction, structure | 294 | 0.3124 |
| 1951-2018 | 1 | model, theory, system, base, paper, problem, propose, present, approach, provide | 556 | 0.2251 |
| 1951-2018 | 2 | behavior, social, human, decision, study, experiment, make, suggest, result, behaviour | 482 | 0.1951 |
| 1951-2018 | 3 | individual, group, good, social, punishment, level, cost, mechanism, dilemma, cooperative | 428 | 0.1733 |
| 1951-2018 | 4 | game, strategy, player, agent, play, dilemma, state, prisoner, payoff, equilibrium | 380 | 0.1538 |
| 1951-2018 | 5 | population, evolutionary, dynamic, model, selection, result, evolution, evolve, show, process | 351 | 0.1421 |
| 1951-2018 | 6 | cooperation, network, interaction, structure, study, evolution, find, behavior, cooperative, simulation | 273 | 0.1105 |

Table 3: Topic modeling result for the cumulative data set over the periods

But how well do the five topics which were presented earlier fit the publications over time? This is answered by comparing the performance of three LDA models over the cumulative periods' publications. The three models are LDA models for the entire data set for $n$ equal to 5, 6 and the optimal number of topics over time. For each model the $c^*$ is estimated for each document in the cumulative data sets. The performance of the models are then compared based on:

$$\bar{c^*} \times n \tag{1}$$

where $\bar{c^*}$ is the median highest percentage contribution and $n$ is the number of topics of a given period. A model with more topics will have more difficulty to assign papers. Thus, equation (refeq:ratio) is a measure of confidence in assigning a given paper to its topic weighted by the number of topics. The performances are given by Figure 8.

The five topics of the PD presented in this manuscript appear to always be less good at fitting the publications compared to the six topics of LDA $n = 6$. Moreover, there are less good than the topics of the optimal number of topics from 1951 to 1995. The difference in the performance values, equation (1), however are small. **The relevances of the five topics has been increasing over time, and though, the topics did not always fit the majority of published work over time, there were still papers being published on those topics**.

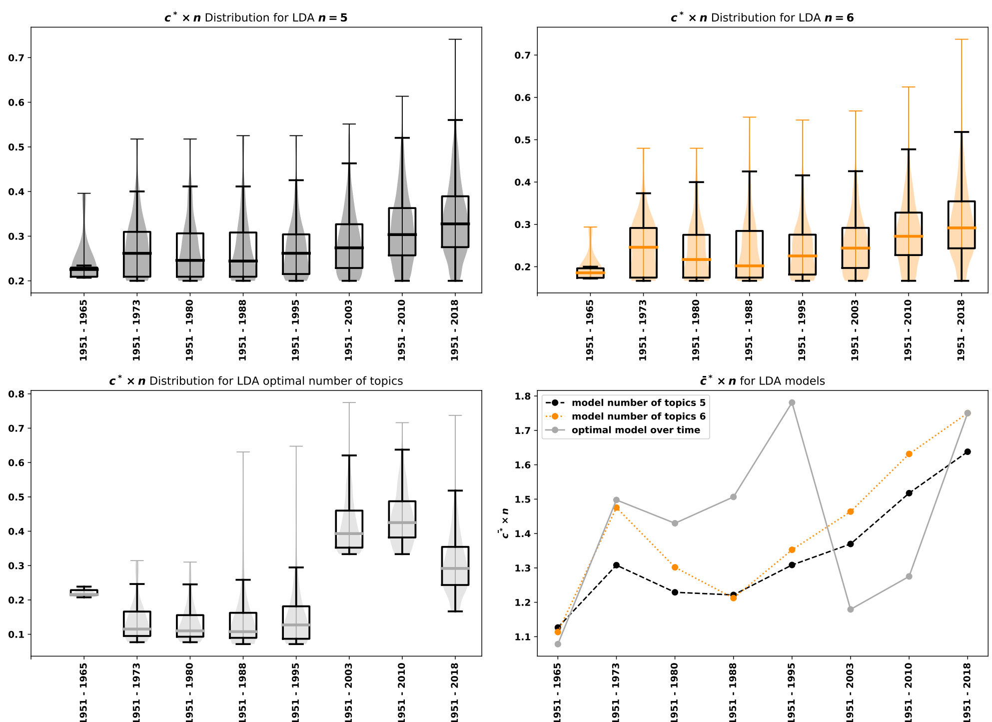



Figure 8: Maximum percentage contributions ($c^*$) over the time periods, for the LDA models for the entire data set for $n$ equal to 5, 6 and the optimal number of topics over time.

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

# 5 Analysis of co-authorship network

The collaborative behaviour of authors in the field of the PD is assessed using the co-authorship network, which as mentioned in Section 2 is denoted as $G$. There are a total of 947 connected components in $G$ and the largest component has a size of 796 nodes. The largest connected component is going to be refereed to as the main cluster of the network and is denoted as $\bar{G}$. A graphical representation of both networks is shown in Figure 9 and a metrics summary is given by Table 4.

**Is the Prisoner's Dilemma a collaborative field?**

Based on Table 4 an author in $G$ has on average 4 collaborators and a 70% probability of collaborating with a collaborator's co-author. An author of $\bar{G}$ on average is 7% more likely to write with a collaborator's co-author and on average has 2 more collaborators. Moreover, there are only 3.2 % of authors in the PD that has no connection to any other author.

How does this compare to other fields? Two more data sets for the topics "Price of Anarchy" and "Auction Games" have been collected in order to compare the collaborative behaviour of the PD to other game theoretic fields. A total of 3444 publications have been collected for Auction games and 748 for Price of Anarchy. Price of Anarchy is relatively a new field, with the first publication on the topic being [42] in 1999. This explains the small number of articles that have been retrieved. Both data sets have been archived and are available in [27, 28]. The networks for both data sets have been generated in the same way as $G$. A summary of the networks' metrics are given by Table 5.

The average degrees for the Price of Anarchy and for Auction games are lower than the PD's. In Auction games an author is more likely to have no collaborators, and in the Price of Anarchy there are almost no authors that are not connected to someone. This could be an effect of the field being introduced in more modern days. Overall, an author in the PD has on average more collaborators and there are less isolated authors compared to another well established game theoretic field. These results seem to indicate that the PD is a *relatively* collaborative field.

However, both $G$ and $\bar{G}$ have a high modularity (larger than 0.84) and a large number of communities (967 and 25 respectively). A high modularity implies that authors create their own publishing communities but not many publications from authors from different communities occur. Thus, author tends to collaborate with authors in their communities but not many efforts are made to create new connections to other communities and spread the knowledge of the field across academic teams. The fields of both Price of Anarchy and Auction games also have high modularity, and that could indicate that is in fact how academic publications are.

Thus, **the PD is indeed a collaborative field but perhaps it is not more collaborative than other fields**, as there is no effort from the authors to write with people outside their community.

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

Table 4: Network metrics for $G$ and $\bar{G}$ respectively.

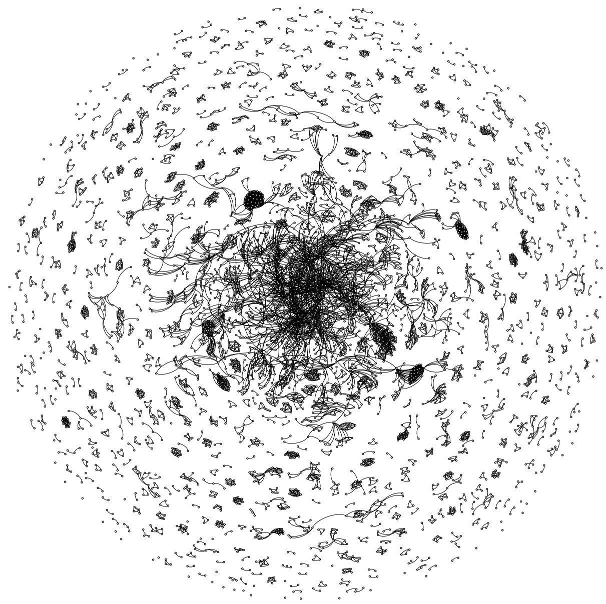
| | # Nodes | # Edges | # Isolated nodes | % Isolated nodes | # Connected components | Size of largest component | Av. degree | # Communities | Modularity | Clustering coeff |
|---|---|---|---|---|---|---|---|---|---|---|
| Auction Games | 5165 | 7861 | 256 | 5.0 | 1272 | 1348 | 3.044 | 1294 | 0.957 | 0.622 |
| Price of Anarchy | 1155 | 1953 | 4 | 0.3 | 245 | 222 | 3.382 | 253 | 0.965 | 0.712 |

Table 5: Network metrics for auction games and price of anarchy networks respectively.

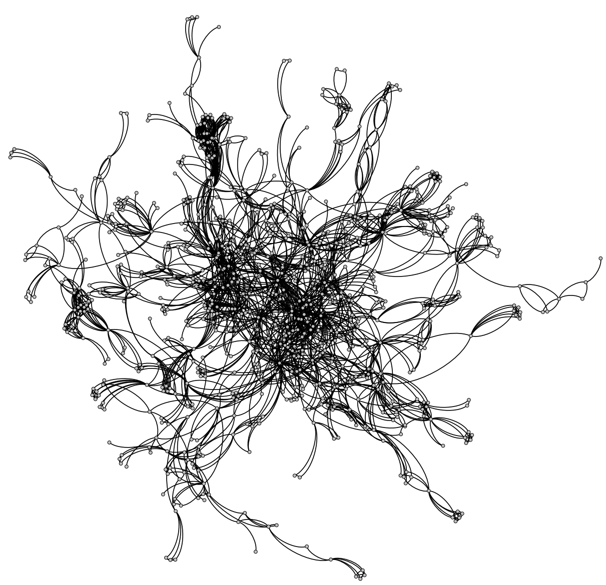
The evolution of the networks was also explored over time by constructing the network cumulatively over 51 periods. Except from the first period 1951-1966 the rest of the periods have a yearly interval (data for the years 1975 and 1982 were not retrieved by the collection data process). The metrics of each sub network are given in the Appendix A.

The results, similarly to the results of [47], confirm that the networks grow over time and that the networks always had a high modularity. Since the first publications authors tend to write with people from their communities, and that is not an effect of a specific time period.

**Are some topics more collaborative than other?**



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



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

Figure 9: A graphical representation of $G$ and $\bar{G}$

The networks corresponding to the topics of Section 3 have also been generated similarly to $G$. Note that authors with publications in more than one topic exist, and these authors are included in all the corresponding networks. A metrics' summary for all five topic networks is given by Table 6.

Topic B is the network with the highest average degree followed by Topic A. The topic with the smallest average degree, 2.5, is Topic C. In topics A and B the number of isolated nodes is very small $lessthan(0.2)$ compared to Topic E where the percentage of isolated nodes is approximately 6%. Moreover, in topics C and E an author is 10% more likely to collaborate with a collaborator's co-author. Thus, **topics "human subject research" and "biological studies" tend to be more collaborative than the topic of "strategies", and an authors in these are less likely to have at least one collaborator compared to the topic of "modeling problems as a PD"**.

**"Evolutionary dynamics on networks" also appear to be a collaborative topic**. In fact the network of the topic is a sub graph of $\bar{G}$, the main cluster of $G$ and it will be demonstrated in the following section that authors in this network are more like to gain from the influence of the network compared to any other topic network.

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

Table 6: Network metrics for topic networks.

**Are there authors which benefit more from their position in the network?**

There are two centrality measures reported in this work, closeness and betweenness centrality. Closeness centrality is a measure of how easy it is for an author to contact others, and consequently affect them; influence them. Thus closeness centrality here is a measure of influence. Betweenness centrality is a measure of how many paths pass through a specific node, thus the amount of information this person has access to. Betweenness centrality is used here as a measure of how much an author gains from the field. All centrality measure can have values ranging from 0 to 1. The influence and the amount of information an author has access to are used to explore which authors benefit more from their position.

For $G$ and $\bar{G}$ the most central authors based on closeness and betweenness centralities are given by Table 7. The most central authors in $G$ and $\bar{G}$ are the same. This implies that the results on centrality heavily rely on the main cluster (as expected). Matjaz Perc is an author with 83 publications in the data set and the most central authors based on both centrality measures. The most central authors are fairly similar between the two measures. The author that appear to be central based on one measure and not the other are Martin Nowak, Franz Weissing, Jianye Hao, Angel Sanchez and Valerio Capraro which have access to information due to their positioning but do not influence the network as much, and the opposite is true for Attila Szolnoki, Luo-Luo Jiang Sandro Meloni, Cheng-Yi Xia and Xiaojie Chen.

It is obvious that in $G$ the centralities values are low which suggests that in the PD authors do not benefit from their positions. This could be an effect of information not flowing from one community to another as authors tend to write with people from their communities. Nevertheless, **there are authors that do benefit from their position, but these are only the authors connected to the main cluster**.

The centrality measures for the topic networks have also been estimated and are given in Tables 8-9. If information was flowing between the communities of the research topics then there would be an increase to the values of centralities for the sub networks. However, the only topic where authors gain from their positions are the authors of Topic D (topic on evolutionary dynamics on network). From the list of names it is obvious that these authors are part of $\bar{G}$, and that the network of Topic D is a sub network of $\bar{G}$. This confirms the results. The people benefiting from their position in the co-authorship networks corresponding to research topics of the PD are only the people from the main cluster of $G$.

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

Table 7: 10 most central authors based on betweenness and closeness centralities for $G$ and $\bar{G}$.

The fact that most authors of the main cluster are primarily publishing in evolutionary dynamics on networks indicates that publishing in this specific topic differs from the other topics covered in this manuscript. There appears to be more collaboration and influence in the publications on evolutionary dynamics and authors are more likely to gain from their position, though it is not clear as to why.

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

Table 8: 10 most central authors based on betweenness centrality for topics' networks.

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

Table 9: 10 most central authors based on closeness centrality for topics' networks.

The distributions of both centrality measures for all the networks of this work are given in the Appendix B.2.

# 6 Conclusion

This manuscript has explored the research topics in the publications of the Iterated Prisoner's Dilemma, and moreover, the authors' collaborative behaviour and their influence in the research field. This was achieved by applying network theoretic approaches and a LDA algorithm to a total of 2422 publications. Both the software [30] and the data [30] have been archived and are available to be used by other researchers. In fact [30] has been used by [49] and [73].

The data collection and an introduction to the methodology used in this work were covered in Section 2. Section 3 covered an initial analysis of the data set which demonstrated that the PD is a field that continues to attract academic attention and publications. In Section 4 LDA was applied to the data set to identify topics on which researchers have been publishing. The LDA analysis showed that the data could be classified into 5 topics associated with human subject research, biological studies, strategies, evolutionary dynamics on networks and modeling problems as a PD. These topics summarize the field of the PD well, as they demonstrate its interdisciplinarity and applications to a variety of problems. A temporal analysis explored how relevant these topics have been over the course of time, and it revealed that even though there were not the necessarily always the most discussed topics they were still being explored by researchers.

The collaborative behaviour of the field was explored in Section 5 by constructing the co authorship network. It was concluded that the field is a collaborative field, where authors are likely to write with a collaborator's co-authors and on average an author has 4 co-authors, however it not necessarily more collaborative than other fields. The authors tend to collaborate with authors from one community, but not many authors are involved in multiple communities. This however might be an effect of academic research, and it might not be true just for the field of the PD. Exploring the influence of authors and their gain from being in the network of the field demonstrated that authors do not gain much, and the authors with influence are only the ones connected to the main cluster, to a "main" group of authors. This 'main" group of authors consists of authors publishing in evolutionary dynamics on networks. Thus, an author would be aiming to publish on this topic if they were interested in gaining from their position in the publications of the PD.

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

## 7 Acknowledgements

A variety of software have been used in this work:

- The Matplotlib library for visualisation [35].
- The Numpy library for data manipulation [77].
- The Networkx [34] package for analysing networks.
- Gephi [11] open source package for visualising networks.
- The Gensim library for the topic modeling [63].
- The louvain library for calculating the networks modularity `https://github.com/taynaud/python-louvain`.

## References

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

[2] Christoph Adami and Arend Hintze. Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything. *Nature communications*, 4:2193, 2013.

[3] Ankur Agarwal and Bill Triggs. Multilevel image coding with hyperfeatures. *International Journal of Computer Vision*, 78(1):15–27, 2008.

[4] Marco Archetti. Evolutionary game theory of growth factor production: implications for tumour heterogeneity and resistance to therapies. *British journal of cancer*, 109(4):1056, 2013.

[5] Daniel Ashlock and Eun-Youn Kim. Fingerprinting: Visualization and automatic analysis of prisoner's dilemma strategies. *IEEE Transactions on Evolutionary Computation*, 12(5):647–659, 2008.

[6] Robert Axelrod. Effective choice in the prisoner's dilemma. *Journal of conflict resolution*, 24(1):3–25, 1980.

[7] Robert Axelrod. More effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, 24(3):379–403, 1980.

[8] Robert Axelrod et al. The evolution of strategies in the iterated prisoner's dilemma. *The dynamics of norms*, pages 1–16, 1987.

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

[10] Dipyaman Banerjee and Sandip Sen. Reaching pareto-optimality in prisoner's dilemma using conditional joint action learning. *Autonomous Agents and Multi-Agent Systems*, 15(1):91–108, 2007.

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

[12] Raoul Bell, Laura Mieth, and Axel Buchner. Separating conditional and unconditional cooperation in a sequential prisoner's dilemma game. *PloS one*, 12(11):e0187952, 2017.

[13] TILL Bergmann and RICK Dale. A scientometric analysis of evolang: Intersections and authorships. In *The evolution of language: Proceedings of the 11th international conference (EVOLANGX11). http://evolang.org/neworleans/papers/182. html. Retrieved*, volume 22, 2018.

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

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

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

[17] Yu-Ting Chen, Alex McAvoy, and Martin A Nowak. Fixation probabilities for any configuration of two strategies on regular graphs. *Scientific reports*, 6:39181, 2016.

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

[19] Luis Pedro Coelho, Tao Peng, and Robert F Murphy. Quantifying the distribution of probes between subcellular locations using unsupervised pattern unmixing. *Bioinformatics*, 26(12):i7–i12, 2010.

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

[21] Florence Débarre, C Hauert, and M Doebeli. Social evolution in structured populations. *Nature Communications*, 5:3409, 2014.

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

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

[24] Xiuzhen Feng and Yijian Liu. Trilateral game analysis on information sharing among members in a virtual team. In *2008 IEEE Symposium on Advanced Management of Information for Globalized Enterprises (AMIGE)*, pages 1–5. IEEE, 2008.

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

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

[27] N. E. Glynatsi. Articles' meta data on auction games. https://doi.org/10.5281/zenodo.3406544, September 2019.

[28] N. E. Glynatsi. Articles' meta data on the price of anarchy. https://doi.org/10.5281/zenodo.3406542, September 2019.

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

[30] N. E. Glynatsi and V. Knight. Nikoleta-v3/arcas: Arcas v 0.0.4. https://doi.org/10.5281/zenodo.1127684, December 2017.

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

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

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

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

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

[36] B Ian Hutchins, Xin Yuan, James M Anderson, and George M Santangelo. Relative citation ratio (rcr): A new metric that uses citation rates to measure influence at the article level. *PLoS biology*, 14(9):e1002541, 2016.

[37] Genki Ichinose, Yuto Tenguishi, and Toshihiro Tanizawa. Robustness of cooperation on scale-free networks under continuous topological change. *Physical Review E*, 88(5):052808, 2013.

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

[39] Matthew Inglis and Colin Foster. Five decades of mathematics education research. *Journal for Research in Mathematics Education*, 49(4):462–500, 2018.

[40] Hisao Ishibuchi, Hiroyuki Ohyanagi, and Yusuke Nojima. Evolution of strategies with different representation schemes in a spatial iterated prisoner's dilemma game. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(1):67–82, 2011.

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

[42] E. Koutsoupias and C. 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.

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

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

[45] Jing Li, Liqi Zhu, and Michaela Gummerum. The relationship between moral judgment and cooperation in children with high-functioning autism. *Scientific Reports*, 4:4314, 2014.

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

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

[48] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[49] Smriti Mallapaty. The brane. `https://thebrane.com`, 2016.

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

[51] Yoshie Matsumoto, Toshio Yamagishi, Yang Li, and Toko Kiyonari. Prosocial behavior increases with age across five economic games. *PloS one*, 11(7):e0158671, 2016.

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

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

[54] J Alberto Molina, J Ignacio Giménez-Nadal, José A Cuesta, Carlos Gracia-Lazaro, Yamir Moreno, and Angel Sanchez. Gender differences in cooperation: experimental evidence on high school students. *PloS one*, 8(12):e83700, 2013.

[55] Juan Carlos Niebles, Hongcheng Wang, and Li Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. *International journal of computer vision*, 79(3):299–318, 2008.

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

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

[58] Hisashi Ohtsuki, Christoph Hauert, Erez Lieberman, and Martin A Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092):502, 2006.

[59] Richard J Ormerod. Or as rational choice: A decision and game theory perspective. *Journal of the Operational Research Society*, 61(12):1761–1776, 2010.

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

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

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

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

[64] Rick L Riolo, Michael D Cohen, and Robert Axelrod. Evolution of cooperation without reciprocity. *Nature*, 414(6862):441, 2001.

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

[66] Lorenzo A Santorelli, Christopher RL Thompson, Elizabeth Villegas, Jessica Svetz, Christopher Dinh, Anup Parikh, Richard Sucgang, Adam Kuspa, Joan E Strassmann, David C Queller, et al. Facultative cheater mutants reveal the genetic complexity of cooperation in social amoebae. *Nature*, 451(7182):1107, 2008.

[67] Javad Salimi Sartakhti, Mohammad Hossein Manshaei, David Basanta, and Mehdi Sadeghi. Evolutionary emergence of angiogenesis in avascular tumors using a spatial public goods game. *PloS one*, 12(4):e0175063, 2017.

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

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

[70] Alexander J Stewart and Joshua B Plotkin. Extortion and cooperation in the prisoner's dilemma. *Proceedings of the National Academy of Sciences*, 109(26):10134–10135, 2012.

[71] Alexander J Stewart and Joshua B Plotkin. From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proceedings of the National Academy of Sciences*, 110(38):15348–15353, 2013.

[72] Cassidy R Sugimoto, Daifeng Li, Terrell G Russell, S Craig Finlay, and Ying Ding. The shifting sands of disciplinary development: Analyzing north american library and information science dissertations using latent dirichlet allocation. *Journal of the American Society for Information Science and Technology*, 62(1):185–204, 2011.

[73] Ilyass Tabiai. Exploring the evolution of open access publications with arcas. `http://iltabiai.github.io`, 2019.

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

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

[76] Jeromos Vukov and György Szabó. Evolutionary prisoner's dilemma game on hierarchical lattices. *Physical Review E*, 71(3):036133, 2005.

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

[78] Juan Wang, ChengYi Xia, YiLing Wang, Shuai Ding, and JunQing Sun. Spatial prisoner's dilemma games with increasing size of the interaction neighborhood on regular lattices. *Chinese science bulletin*, 57(7):724–728, 2012.

[79] Xiaogang Wang, Xiaoxu Ma, and W Eric L Grimson. Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. *IEEE Transactions on pattern analysis and machine intelligence*, 31(3):539–555, 2008.

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

[81] Te Wu, Feng Fu, and Long Wang. Phenotype affinity mediated interactions can facilitate the evolution of cooperation. *Journal of theoretical biology*, 462:361–369, 2019.

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

# A Cumulative Networks Metrics

## A.1 Collaborativeness metrics for cumulative graphs, $\tilde{G} \subseteq G$

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

## A.2 Collaborativeness metrics for cumulative graphs' main clusters, $\tilde{G} \subseteq \bar{G}$

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

# B Centrality Measures Distributions

## B.1 Distrubutions for $G$ and $\bar{G}$

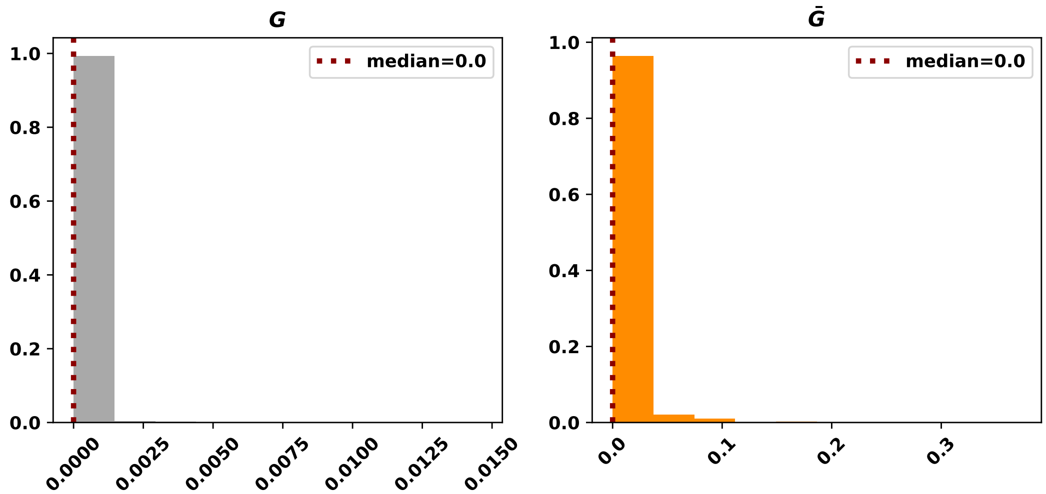



Figure 10: Distributions of betweenness centrality in $G$ and $\bar{G}$

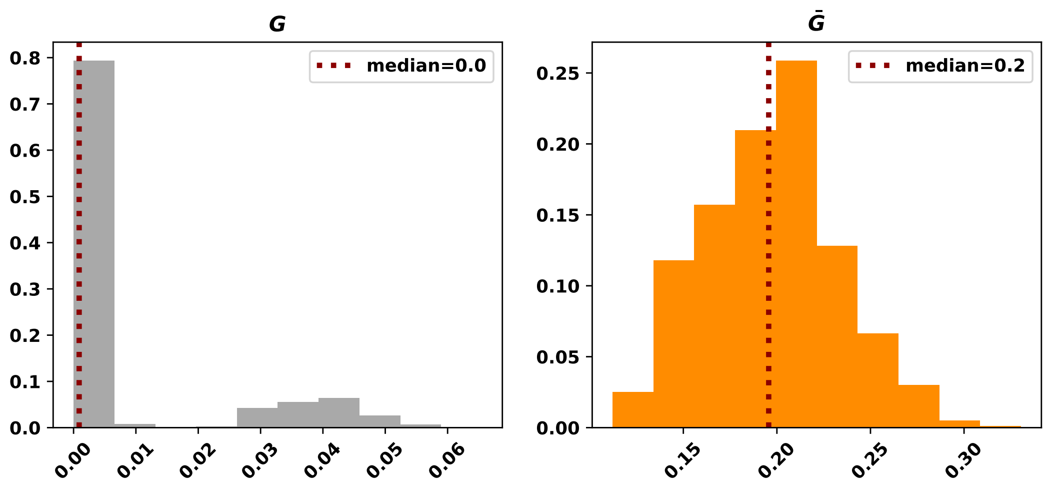



Figure 11: Distributions of closeness centrality in $G$ and $\bar{G}$

## B.2 Distrubutions for Topic Networks

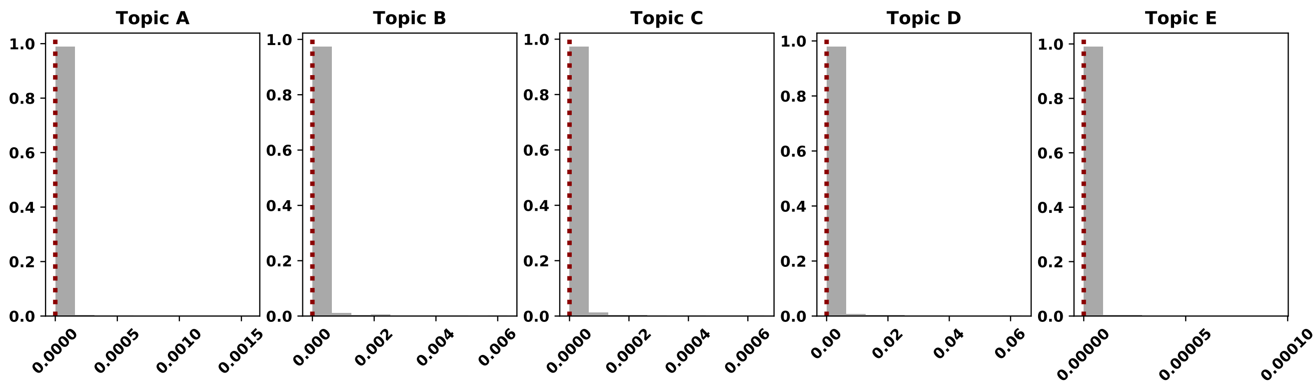



Figure 12: Distributions of betweenness centrality in topics' networks.

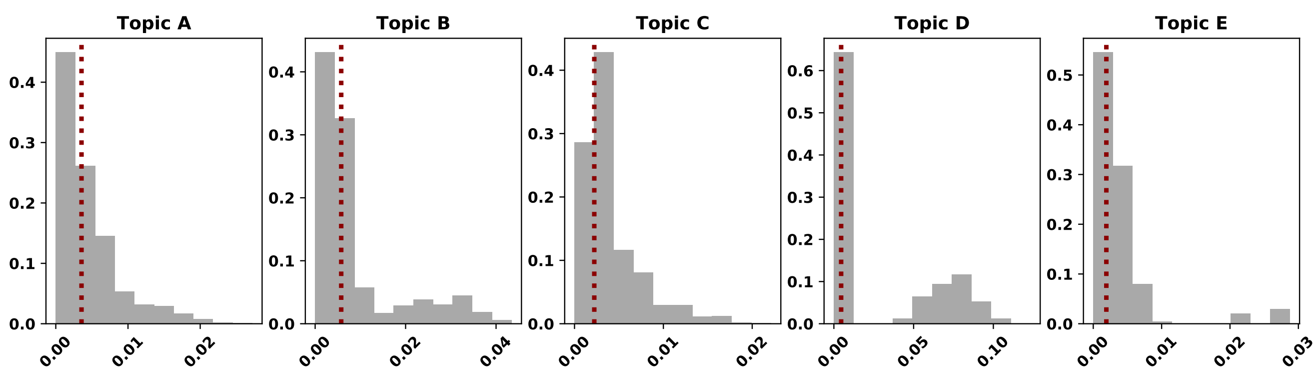



Figure 13: Distributions of closeness centrality in topics' networks.