Master's thesis - master Innovation Sciences

Social and cognitive patterns in science

An evolutionary approach to explore mechanisms of cumulative advantages and knowledge accumulation

Student name:	Wouter Westerdijk
Student number:	6312020
Email:	w.s.k.westerdijk@students.uu.nl
Address:	Oudwijkerveldstraat 33, 3581 JH Utrecht, The Netherlands
Thesis & *R* code:	https://thesis.wouterwesterdijk.nl
Supervisor:	dr. Gaston Heimeriks

Utrecht University, 2021

Abstract

Introduction: The Matthew effect is a cumulative advantage mechanism in the scientific system. This implies that successful scientists have an increased chance to become more successful, as they generally have gained more recognition, easier access to resources and funding, increased productivity, and are more likely to have their articles published. High status scientists have accrued power and influence on the scientific search regimes. Science is found to comprise strong social hierarchies. As the spread of knowledge is affected by social status and cognitive similarity, the interest of this exploratory research is to examine the relation between social inequalities and cognitive inequalities within scientific fields over time. Cognitive inequality implies that the focus of research is unevenly distributed over the subtopics of a scientific field.

Literature: Cumulative advantage mechanisms are linked to evolutionary economics. It is argued that scientists are subjected to selection pressures and seek to provide contributions to the scientific field as success in order to 'survive'. Contributions are considered novel and valuable knowledge combinations, which are retained also by other scientists as such combinations have proven to be successful. Additionally, the scientists who successfully contributed will receive recognition, which increases their chances for survival. Selection pressures urge scientists to avoid taking risks, which implies that they stick close to previous successes, such as contributions of highly recognised scientists or contributions with which they have been recognised. It is therefore hypothesized that scientific fields show an increase in social inequality and cognitive inequality over time, and that the increase of inequality differs for the amount of selection pressures present in a specific field.

Method: Four scientific fields with different selection pressures are compared based on publication data from articles published between 1980 - 2012 with an increment of two years. The inequality of citations among scientists, articles and keywords is calculated over the years using the Gini coefficient. This also applies to the inequality of productivity. Additionally, the statistics of the collaboration networks among scientists and keyword co-occurrences per year are calculated and compared.

Results: All fields show increasing inequalities in the social and cognitive domain. Fields have sublinear correlations of inequality with the number of annual publications.

Conclusion/Discussion: The annual productivity of the field showed the strongest correlations with both social and cognitive inequality. Each scientific field showed distinct

inequality patterns. It is discussed how inequality and cumulative advantage mechanisms in evolutionary theory relate.

Keywords: Matthew effect; cumulative advantage; inequality; organisational structure; sociology of science; evolutionary economics; scientometrics; knowledge production

Content

Abstract 2
Content 4
Introduction 7
Literature review 10

Evolutionary economics in the scientific system 10

The survival of the scientist 11

The scientist's legacy: retention patterns in knowledge production 13

Selection pressures from the scientific system 15

Method 17

Research design 17

Data collection and sampling strategy 17

Context scientific fields 18

Nanoscience & Nanotechnology 18

Biotechnology & Applied Microbiology 18

Astronomy & Astrophysics 19

Organic Chemistry 19

Operationalisation 19

Field characteristics 19

Social structure 19

Knowledge heterogeneity 21

Operationalisation table 22

Data analysis 23

Research quality 24

Results 25
Field characteristics 25
Annual productivity 25
Scientist involvement 27
Social structure 28
Recognition 28
Productivity 30
Collaboration 32
Knowledge heterogeneity 39
Retention 39
Relatedness 41
Research direction 43
Pattern comparison 52
Inequality and annual publications 52
Inequality and citations 57
Inequality and scientist productivity 60
Conclusion 66
Discussion 71
Practical insights 71
Theoretical considerations 72
Limitations / issues 74
Future research 76
References 78

Introduction

Science is considered by many as the backbone of innovation and economic development (Kuo et al., 2019). The knowledge created through academic research often formed the building blocks for societal and technological advancements as part of an innovation system (Etzkowitz, 2008; Etzkowitz & Leydesdorff, 2000). This brings a sense of responsibility for science to adhere to certain norms, e.g. those drafted by Robert K. Merton (1942) in his essay *A Note on Science and Democracy*. He believed knowledge produced through scientific research should be a common property, and results should remain detached from personal and subjective views. Additionally, he argued that scientists should remain indifferent to emotional, political or financial attachments to their work. Lastly, a scientist should always remain skeptical about their own work and others', including long-established research. These four norms suggest the scientific system to incorporate communism, universalism, disinterestedness and organised skepticism (Merton, 1942). This proposed scientific ideology seemed largely supported by scientists, although barriers with commercial and political involvement are fading (Macfarlane & Cheng, 2008; Perkmann et al., 2013). Additionally, while Merton (1942) acknowledged that scientists are entitled to recognition for sharing their academic works, he realised two and a half decades later that the distribution of such recognition is highly uneven (Merton, 1968).

The Matthew effect[1], as Merton labelled it, regards "the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark" (Merton, 1968, p. 58). It is a term nowadays widely used for the great status inequalities in the social domain of science. Aside from receiving disproportionally more recognition, high status scientists additionally require less effort in collecting resources for further research (Allison et al., 1982). For example, Bol et al. (2018) found the Matthew effect responsible for two mutually reinforcing processes; highly recognised scientists had better luck in receiving academic funding, while also maintaining a higher rate of applying for subsequent funding. Prior successes have a positive influence on the evaluation of future works, as well as the rate at which these future works can be produced as resources are more easily available (Allison et al., 1982; Bol et al., 2018; Larivière & Gingras, 2010; Merton, 1968; Petersen & Penner,

[1] After St. Matthew's quote of the Gospel: "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath." (Matthew 25:29)

2014; Sauder et al., 2012). As a result, the Matthew effect displays skewed distributions in citation patterns (Larivière & Gingras, 2010; Newman, 2005; Perc, 2014; Petersen & Penner, 2014; Price, 1976) or author appearance frequency in journals (Siler et al., 2018). The Matthew effect can be considered a fundamental mechanism behind bibliometric laws (Mingers & Leydesdorff, 2015). Social dynamics in science are considered important steering instruments for the direction of research (Whitley, 2000), such as the emergence of new fields (Sun et al., 2013) or change in scientific paradigms (T. S. Kuhn, 1970). The Matthew effect facilitates highly recognised scientists with power and influence (Siler et al., 2018), possibly significantly affecting these directions.

We are interested in whether the evolution of social status differences in scientific fields presents similar dynamics in the distribution of academic focus among research topics. Prior research emphasised the importance of social dynamics and scientific communities within the evolution of scientific fields, but never focused on direct correlations between social and cognitive inequalities (Jansen et al., 2010; Kauffman, 1993; T. S. Kuhn, 1970; Sun et al., 2013; Whitley, 2000; Zeng et al., 2017). The spread of ideas and knowledge is linked to social status and similarity (Azoulay et al., 2013; Heylighen, 1993; T. Kuhn et al., 2014; Schlaile, 2020; Watts & Gilbert, 2011). Additionally, the accumulation of knowledge is considered to follow evolutionary patterns, such as the selection and retention of favourable 'knowledge blocks' (Arthur, 2007; Dawkins, 2016; Heylighen, 1993; Kauffman, 1993; Nelson & Winter, 1982; Rigby, 2015). The perceived quality of a product, e.g. a research article, is partly influenced by the status of its producer, the scientist (Azoulay et al., 2013). The recital of our previous arguments seem to suggest a causal relation between status and knowledge accumulation. However, the intensity of research in certain topics attracts further research also as a result of exploitation, sunk costs and (cognitive) path dependence (Arthur, 1994, 2007; Heimeriks & Boschma, 2014; Kuo et al., 2019; March, 1991; Tijssen & Van Raan, 1994). Our aim is to explore correlations between the social and cognitive domains in scientific fields, based on inequality in concentration of resources and recognition among scientists and research topics. This leads to our research question:

- ***RQ****: How do social and cognitive patterns of inequality in science relate?*

Our research is of exploratory nature and aims to form the first bridge by addressing different gaps in the literature. To start, we could set an example of research conducted on the correlation between social cumulative advantages such as the Matthew effect and the path dependency of knowledge production. Most importantly, we attempt to light up the discussion on the emergence of Matthew effects as reiterated accumulation of social

advantages, since such cumulative advantage mechanisms are not yet fully understood (DiPrete & Eirich, 2006). We model this through an agent-based system inspired by literature on evolutionary economics (Fagerberg, 2003; Nelson & Winter, 1982; Shiozawa et al., 2019). Literature from both studies may offer clarifications for questions that have not yet been solved. In the theory section, we elaborate how an agent-based model from evolutionary economics could cause the emergence of cumulative advantage mechanisms. Additionally, while evolutionary economics literature is focused on the dynamics of system growth, there remains discussion about the functions of system optimisation and origins of novelty (Andersen, 2001; Fagerberg, 2003), which we address. Our insights may provide fuel for further research in the understanding of cumulative advantage mechanisms in emerging and growing systems. We compare four scientific fields based on differences in scientists' reliance upon prior knowledge and peers (Whitley, 2000), which is expected to represent selection differences. We use quantitative methods to analyse bibliographic publication data.

The societal relevance of this exploratory study is to raise awareness that the self-organisation of science might lead to gatekeeping and intolerance of novelty, decreasing chances of scientific breakthroughs and fruitful knowledge combinations (Siler et al., 2018). We also put social inequality into perspective in relation to growth and competition for resources. In our literature review, we describe the behavioural processes in evolutionary economics combined with empirical evidence in the scientific system and hypothesize the patterns we should find. In our method section we describe the four chosen scientific fields and elaborate on our variables and indicators. In the result section we describe the patterns of inequality we found over time and productivity, and test for correlations. After that, we provide our conclusion and further discuss our contributions to the literature and practical insights. We describe our research limitations and how literature on evolutionary economics and sociology of science could benefit from future research based on ours.

Literature review

This section adopts an evolutionary approach to examine how and why the scientific system generates cumulative advantage mechanisms as a byproduct in its search for new knowledge. A key factor in our consideration is the interaction between the system and its individuals, i.e., *micro-macro loops*. The Matthew effect as cumulative advantage mechanism is strengthened not only because of the scientific system rewarding successful scientists, but also because these scientists "accrue power and influence, enabling leaders in scientific fields to judge scientific work according to their preferred principles" (Siler et al., 2018, p. 230). High-status scientists tend to receive more favourable evaluations as status affects scientific assessments (Bravo et al., 2018; Ceci & Peters, 1982; Leahey, 2004; Simcoe & Waguespack, 2011; Tomkins et al., 2017). Especially in situations with high uncertainty, status plays a crucial role in decision-making (Sauder et al., 2012). Scientists, through status, thus seem to affect their system's rewarding mechanism.

Next to the social aspects of science dynamics, we discuss patterns of knowledge accumulation and path-dependency. Drawing upon evolutionary economics literature allows us to examine the interaction of social and cognitive patterns and system-agent dynamics (Nelson & Winter, 1982; Shiozawa et al., 2019). Knowledge is considered the backbone of economic development, and its accumulation patterns regulate the course of progress (Arthur, 1994; Breschi et al., 2000; Fagerberg, 2003; Schumpeter, 1943). Novelty builds upon existing knowledge through recombinations, which not only creates path-dependency (Arthur, 1994, 2007), it also foreshadows possible future recombinations through what Kauffman (1993) has coined the "*adjacent possible*". The costs associated with searching for new knowledge combinations are related to prior knowledge accumulation (Heimeriks & Balland, 2016; Heimeriks & Boschma, 2014; Rigby, 2015). The third subsection expands upon this topic in relation to science dynamics. Lastly, a perspective of the scientific system as selection environment for academic research is adopted.

Evolutionary economics in the scientific system

Evolutionary theory in the socio-economic context is to model the dynamics of change over time in which the system contains a source of variation and selection mechanisms (Dosi et al., 1988; Dosi & Nelson, 1994; Nelson & Winter, 1982). The system comprises phenotypes, the agents, that undergo selection pressures in which their chance of survival is largely dependent on the fitness of their genes (Dosi & Nelson, 1994). What constitutes these genes

should be modifiable, recombinable and replicable, i.e. knowledge blocks or memes[2] (Dawkins, 2016; Heylighen, 1993; T. Kuhn et al., 2014). The genes form the roots of the agent's behaviour which in combination with environment interactions create *decision rules* (Nelson & Winter, 1974). Through search and selection, agents modify, copy and adapt these rules to increase their chances of survival, where seemingly successful rules shape behavioural *routines* (Dosi & Nelson, 1994; Nelson & Winter, 1974, 1982). The differences in agents' routines constitute the variation of a system, where variation is considered the fuel of development (Andersen, 2001; Fagerberg, 2003).

The scientific system can be understood as a global body of knowledge continuously expanding (Fujigaki, 1998; Heimeriks & Balland, 2016). The universalistic and self-organising approach of science is to grow by transforming its finite resources into new knowledge (Sun et al., 2013). A parallel with economic growth can be drawn here as it concerns the input, output and prices over time (Nelson & Winter, 1974, 1982). The production of new knowledge is associated with costs of search as a result of finite possible knowledge recombinations and prior combinations (Heimeriks & Balland, 2016; Heimeriks & Boschma, 2014; Kauffman, 1993; Rigby, 2015). To reduce these costs, the scientific system allocates its resources to those who have already proven their research abilities (Allison et al., 1982; Cole & Cole, 1974; DiPrete & Eirich, 2006; Merton, 1968; Zuckerman, 1977). As such, scientists receive recognition as a currency with which they can obtain resources for future research (Azoulay et al., 2013; Bol et al., 2018; Merton, 1968; Siler et al., 2018). Therefore, we presume scientific development to largely rely on competition among scientists for shares of recognition.

The survival of the scientist

We consider science as an agent-based socio-economic system for which we turn to the evolutionary economics literature. In this section we apply theory of agents in an evolutionary system to empirical findings that concern cumulative advantages in science. The premise of evolutionary development is the continuous effort of agents to comply with the system's selection pressures and competition (Nelson & Winter, 1982; Shiozawa et al., 2019). Agents are considered unable to compute an optimal solution that would guarantee their survival. Instead, a nonoptimal but *satisficing* solution can be reached which, at least for the time

[2] Another branch of literature on the evolution of knowledge, evolutionary epistemology, considers knowledge pieces most valuable to the survival and reproduction of its carriers to be retained. An extension to this view, called *memetics*, considers knowledge to actively pursue its own survival. One of the most popular works of origin is *The Selfish Gene* (Dawkins, 2016). A piece of knowledge that can be transmitted and replicated to other carriers, thereby losing its reliance on the individual survival of its carrier, is defined as a *meme* (Heylighen, 1993).

being, would satisfy the selection pressures (Fagerberg, 2003; Shiozawa et al., 2019; Simon, 1972). While not theoretically the optimal solution, the agent could still apply one of the best solutions practically available.

The limitations to an agent's ability to compute the optimal solution, *bounded rationality*, is not the only aspect restraining their survival guarantees. Even if agents were not bound in their rationality, the myopic nature of their perception allows only a subset of all available knowledge to be taken into account (Nelson & Winter, 1982; Shiozawa et al., 2019). Therefore, agents prefer to rely on their existing routines for as long as the selection environment allows. The search for new routines is potentially costly and highly uncertain (Fagerberg, 2003; Nelson & Winter, 1974, 1982). However, agents can imitate each other's routines which would allow them to survive by copying routines of successful co-agents and thereby reducing costs and risks.

Mimetic and risk-aversive behaviour is expected to be present among scientists as well (T. Kuhn et al., 2014). Scientific paradigms build upon problem-solving rules that are generally accepted and unquestioned by scientists (T. S. Kuhn, 1970). A scientist can be considered successful when they contribute greatly to the scientific knowledge base (Merton, 1957, 1968). The evaluation of how 'greatly' a scientist contributes, e.g. the perceived quality of a journal article, is positively affected by the scientist's status (Azoulay et al., 2013; Sauder et al., 2012; Siler et al., 2018). A scientist's status can be derived from their social network position or the aggregation of perceived quality from previous research (Azoulay et al., 2013). Status and recognition seem to play a significant role in the survival successes of scientists, as this allows scientists to receive resources for the continuation of their research career (Allison et al., 1982; Merton, 1968; Sauder et al., 2012; Siler et al., 2018). Other scientists could imitate their high-status peers through social or cognitive replication, as status is obtainable through association (Gould, 2002; Sauder et al., 2012). Additionally, cognitive association requires some similarity to existing research to indicate relevance to peers (Watts & Gilbert, 2011) or receive resources in the first place (Kauffman, 1993). We are curious how this would be demonstrated in the social distribution of recognition within a scientific field. This brings us to the first subquestion:

- ***SQ1****: What are the dynamics of the social structure within a scientific field over time?*

We expect displays of the Matthew effect, which is supposed to result in some form of a power-law distribution (DiPrete & Eirich, 2006; Merton, 1968; Newman, 2005; Perc, 2014; Petersen & Penner, 2014; Price, 1976; Siler et al., 2018). Additionally, as high-status

scientists continue to receive greater increments of recognition over time, status inequality will increase, and the distribution will be increasingly skewed. The tail of the distribution can be less stable than the top, and therefore seem less fitting to the power-law (Newman, 2005; Perc, 2014; Petersen & Penner, 2014). From this we derive our first hypothesis:

- ***H1****: The social structure within a scientific field is strongly hierarchical and becomes increasingly unequal over time.*

The scientist's legacy: retention patterns in knowledge production

The evolution of knowledge is a process derived from the constant recombination of existing knowledge blocks (Arthur, 2007). Knowledge is accumulative and the search for new possibilities tends to create path-dependence (Arthur, 1994; Nelson & Winter, 1982). Existing knowledge has its limits and possibilities towards the production of new knowledge (Kauffman, 1993), where the costs of search for new combinations largely depend on the number of existing blocks and (past) interactivity between them (Heimeriks & Balland, 2016; Heimeriks & Boschma, 2014; Rigby, 2015). Risk-aversive behaviour tends to exploit related topics rather than explore new frontiers (Breschi et al., 2000; Dosi & Nelson, 1994; Fagerberg, 2003; Nelson & Winter, 1982), which has also been identified in science (Azoulay et al., 2013; Fuchs, 1993; March, 1991; Petersen & Penner, 2014; Siler et al., 2018). This behaviour resembles the creation and preservation of search regimes as a means to reduce costs and uncertainty (Bonaccorsi, 2008; Breschi et al., 2000; Gould, 2002; Kauffman, 1993; T. S. Kuhn, 1970). Therefore, we state our second subquestion:

- ***SQ2****: What are the dynamics of the cognitive structure within a scientific field over time?*

We expect the attempt to recombine existing knowledge to be limited by the myopic nature of the scientist's perception, implying that their efforts are only applied concerning a relatively small subset of the universal knowledge pool (Arthur, 1994; Heimeriks & Balland, 2016; Heimeriks & Boschma, 2014; Kauffman, 1993; T. S. Kuhn, 1970). We also expect the bounded rationality of scientists to limit their exertion for the optimal knowledge recombination to one that is satisficing (Simon, 1972; Watts & Gilbert, 2011). This shapes the scientific paradigms and search regimes to, at least in the short-term, reduce search costs and increase productivity and scientific output (Kauffman, 1993; T. S. Kuhn, 1970). However, the exploitative and accumulative nature of the paradigm may at some point lead to a 'cognitive lock-in' where novel combinations become (too) hard to find.

Scientists thus are the producers of new knowledge blocks, which are preserved and shared in the form of journal articles (Azoulay et al., 2013). Journals can be considered the heart of the rewarding system (Siler et al., 2018), which is why we consider it the place where (new) combinations are retained or go extinct. Cited journal articles display the retention patterns in knowledge accumulation (T. Kuhn et al., 2014). Knowledge combinations to be published as journal articles require both originality and similarity (Watts & Gilbert, 2011). This balance of originality (i.e. novel combinations) and similarity (i.e. prerequisite and overlapping knowledge) is what we consider to determine the level of *knowledge heterogeneity* within a certain unit of analysis. We define knowledge heterogeneity as the ratio of relatedness and uniqueness between a set of knowledge blocks. In a field focused on exploitation, research is more centered around similar topics and therefore should demonstrate more activity on overlapping knowledge. From this we derive our second hypothesis:

- ***H2****: Scientific fields focused more on the exploitation of existing topics show an increase in overlapping knowledge blocks over time.*

Lastly, we are curious whether the social and cognitive aspects within science have the tendency to affect one another, which leads to our third subquestion:

- ***SQ3****: How do scientist inequality and knowledge heterogeneity relate?*

The perception of a product quality, e.g. journal article, is affected by the status associated with its producer, e.g. the scientist (Azoulay et al., 2013). The Matthew effect strongly represents this, as papers from high-status scientists are disproportionately more cited (Azoulay et al., 2013; Merton, 1968; Perc, 2014; Petersen & Penner, 2014; Price, 1976). This effect also applies to journals' status (Larivière & Gingras, 2010; Petersen & Penner, 2014). Additionally, high-status scientists are more likely to repeatedly publish in journals (Siler et al., 2018). In terms of knowledge accumulation, this would imply that the knowledge combinations of high-status scientists are more likely to be retained. To scientists as risk-aversive and myopic agents, success is perceived when the rewards are generous. It could be an indication of fitness and as such, the agent is not inclined to change routines (Fagerberg, 2003; Nelson & Winter, 1982). Success tends to increase further specialisation through the combination of cognitively similar research topics, leading to the preference of exploitation over exploration (March, 1991). As Siler et al. (2018, p. 233) state: "The inverse relationship between success and exploration may cause successful scientists to be more conventional

and less innovative later in their careers." We expect this to show more explicitly in fields with greater status inequality among scientists:

- ***H3****: Scientific fields that show patterns of increasing scientist inequality produce less heterogenetic knowledge.*

Selection pressures from the scientific system

We have not been able to define what constitutes the selection environment. However, it is not the purpose of this thesis to do so. The selection environment in evolutionary theory is highly complex and can comprise both endo- and exogenous factors (Dosi & Nelson, 1994; Fagerberg, 2003; Nelson & Winter, 1974, 1982). We take inspiration from the ideas of search regimes and scientific paradigms as selection pressures concerning the fitness landscapes of scientific knowledge (Bonaccorsi, 2008; Kauffman, 1993; T. S. Kuhn, 1970). To address our research question, we explicitly focus on the internal social organisation and interaction rather than more exogenous factors such as government funding. Whitley (2000) proposes differences in the organisational structures of scientific fields to derive from the *reputational system* and *personal autonomy* within fields, which steers the direction of search. This brings us to our fourth and last subquestion:

- ***SQ4****: How does the organisational structure of a scientific field affect its social and cognitive patterns?*

We expect differences in the intensity of the Matthew effect among scientific fields (Perc, 2014), based on differences in organisational structures (Whitley, 2000). We follow the distinctions of *task uncertainty* and *mutual dependency* regarding the reputational system of science (Whitley, 2000). Task uncertainty is related to the cognitive predictability - or lack thereof - concerning the outcome and direction of new knowledge production, which is supposed to be present in any scientific endeavour, albeit expected lower for more accumulated and regulated fields (Heimeriks & Balland, 2016; Whitley, 2000). The mutual dependency concerns the extent to which scientists depend on each other for the production of new knowledge. This includes cognitive dependence, i.e. a strong reliance on existing literature, and the combined effort for the mobilisation of expensive infrastructure and equipment (Whitley, 2000). We expect differences in these dimensions to affect the selection pressures between scientific fields and therefore to result in different social and cognitive structures:

- ***H4**: Differences in task uncertainty and mutual dependency between fields show different social and cognitive structures.*

Method

Research design

This research aims to explore a plausible relation between the distribution of recognition and productivity among scientists (social structures) and the heterogeneity of knowledge (cognitive structures) in scientific fields. We used a longitudinal study, as the expected patterns have a reinforcing nature. The scientific field is our unit of analysis as we compared four different scientific fields to examine whether they produce different patterns of scientist recognition and knowledge heterogeneity. The observation unit is the journal article as it contains rich and codified information which is analysed through bibliometrics. The structure is exploratory and descriptive based on the longitudinal quantitative analysis of journal articles as we attempt to describe social and cognitive patterns in scientific knowledge production through an evolutionary lens. Since this is a fairly uncharted research approach, we intend to compare scientific fields based on descriptive quantitative results since we do not expect to find causal relationships.

Data collection and sampling strategy

The data was collected through the Web of Science (WoS), an online database containing citation data for millions of journal articles across various scientific disciplines. We used a purposive sampling strategy, selecting only the articles categorised in a relevant research field. The WoS have done this categorisation, which will be used for this research. The consideration of what comprises a scientific research field is elaborated in Construct validity. While the knowledge boundaries of research fields often are grey areas, more or less flowing over into neighbouring fields, collecting only the papers sorted in a relevant category will result in a clear distinction. For the sake of restricting this study to its research scope and time limitations, we will not go too much into detail about the ontology of scientific fields.

Two alternative methods of collecting and aggregating papers per field have been considered. The WoS offers the ability to search for 'topics', which returns results based on matching hits in the titles, abstracts and keywords. This method is disregarded as it searches on words, which has been criticised as a means for field distinction since words lack specification and can be used ambiguously among different (research) contexts (Leydesdorff, 1997; van den Besselaar & Heimeriks, 2006). The other method involves focusing purely on the field's core journals, which implies that boundary issues would no longer apply. However,

we expect that core journals generally amount to higher impact factors and citation density (Madhugiri et al., 2013). High impact journals are found to have a positive effect on the impact of individual articles as opposed to duplicate articles published in lesser-impactful journals, suggesting a Matthew effect between journals and their articles (Larivière & Gingras, 2010). This means that our data could be biased towards high impact articles and knowledge homogeneity, thereby downplaying any social or cognitive pattern we would expect to find. This possibility eliminates the option to collect core journal data.

The publication metadata was gathered for every article published in at least one of the four chosen categories within the years 1980 - 2012 with an increment of two years.

Context scientific fields

The four chosen fields were derived from Whitley's (2000) task uncertainty x mutual dependence two-dimensional matrix, meaning that for every combination, we analysed one field: Nanoscience & Nanotechnology (high, high); Biotechnology & Applied Microbiology (high, low), Astronomy & Astrophysics (low, high); and Organic Chemistry (low, low). The positioning of these fields has priorly been done by Heimeriks & Balland (2016). In their article, the central tendency method is used, where two core journals per field were analysed. As discussed, we could not use the same method. Therefore, comparable categories in the WoS have been collected.

Nanoscience & Nanotechnology

This is an emerging, highly diverse and interdisciplinary field (Leydesdorff & Schank, 2008). Research in this field requires a diverse set of expensive infrastructure and equipment and cooperation on the individual and organisational level (Bonaccorsi & Thoma, 2007). As an emerging and diverse field that requires expensive instruments, both task uncertainty and mutual dependence are argued to be high (Heimeriks & Balland, 2016).

Biotechnology & Applied Microbiology

This is also an emerging and interdisciplinary field, focused on applying knowledge produced and used by various actors (Heimeriks & Leydesdorff, 2012). The field is characterised by problem variety, instability and discrepancies concerning the priorities of technical standardisation, therefore task uncertainty is considered high (Heimeriks & Balland, 2016; Whitley, 2000). The field is relatively new and displays divergent dynamics (Bonaccorsi,

2008), which in combination with the wide variety of actors involved, is argued to contain low mutual dependence (Heimeriks & Balland, 2016).

Astronomy & Astrophysics

The field is highly cumulative and collaborative and as a traditional 'big science' it is considered to contain low levels of task uncertainty, but high mutual dependence (Heimeriks & Balland, 2016). The equipment required for research is large and expensive but the research objective is clear and the development of methods incremental (e.g. telescopes, but larger and better) (Price, 1963).

Organic Chemistry

This is a relatively older field with stable knowledge accumulation, similar to Astronomy & Astrophysics, but researchers are less mutually dependent for access to resources (Heimeriks & Balland, 2016; Whitley, 2000). Therefore the field is characterised by low levels of both task uncertainty and mutual dependence.

Operationalisation

Field characteristics

As we have described the broad field characteristics above, the first results should present insights for a general comparison. These insights are also to provide a perspective in which the social and cognitive structure patterns can be examined. We therefore looked at the yearly publication output per field, as well as the cumulative output to examine the growth and size. We also counted the number of individual scientists involved in publications per year and compared this with the number of individual scientists who only published multi-authored papers. Lastly, we tested whether the dynamics of publication output and individual scientist involvement correlate.

Social structure

Recognition. The social structure of a scientific field is considered to comprise status differences and the influence of status in decision-making. We considered *status* to be derived from the total amount of recognition a scientist received. As a result of the Matthew effect, where successful scientists have gained cumulative advantages over their peers, the distribution of recognition is highly skewed, with a small number of scientists having received the largest share of recognition (Allison et al., 1982; Larivière & Gingras, 2010; Merton, 1968;

Petersen & Penner, 2014; Sauder et al., 2012; Siler et al., 2018). A widely-used indicator for scientific recognition is the number of citations a scientist receives (Azoulay et al., 2013; Larivière & Gingras, 2010; Perc, 2014; Price, 1976; Siler et al., 2018). In principle, citing another person's work is considered payment to clear the intellectual debt that emerges as one utilises someone else's knowledge contribution to build upon further. It should be noted that although citing behaviour is much more complex than this rationale, the many reasons for citing did not constrain us from the ability to analyse citation patterns as indicators for recognition (Nicolaisen, 2007; Wouters, 1999).

The inequality in the distribution of recognition among scientists is measured by calculating the Gini coefficient for the total number of citations scientists received for papers they published in the year of observation (Carpenter, 1979). Similarly, the Gini coefficient is calculated for the distribution of appearances among first-author scientists of the cited articles in the year of observation. The Gini coefficient is 0 when resources (citations) are completely equally distributed over the population (scientists), and 1 when all resources are allocated to one actor (Pratt, 1977; Witlox, 2017). We used a scale of 0 to 100 for better comprehension of the values.

Additionally, we looked at the top twenty most-cited scientists within the cited papers as consequences of the Matthew effect are better depicted at the top rather than the tail of the distributions (Newman, 2005; Perc, 2014; Petersen & Penner, 2014). Examining the cited authors allows for better understanding of potential social inequality *within* the field. Finally, we calculated the stability of the top twenty ranking over the years.

The ranking stability was calculated based on the deviation from a theoretically perfect top twenty that does not change over the years. In a co-occurrence matrix of scientist (row) by rank (column) that is perfectly stable, each row would add up to 20. If we divide it by the number of ranking spots (20), we end up with 1 for each row. Therefore, the average of all columns summed per row divided by the number of ranks gives 1. Deviating from this means that there are less top twenty appearances per scientist, which always gives < 1. If we then multiply this value with the minimum number of unique scientist appearances (20) divided by the actual number of unique scientist appearances, a perfectly stable top twenty would also return 1, whereas a less stable top twenty always gives < 1 as well. If we multiply these two values we get a general impression of the matrix deviation from a perfectly stable matrix. Lastly, this value is multiplied with 100 to say that a ranking without change over the years has a 100% stability rate. The described reasoning led to the following equation:

$$ranking\ stability\ = \frac{mean(column\ sums\ per\ row)}{number\ of\ ranks} * \frac{number\ of\ ranks}{number\ of\ rows} * 100$$

Productivity. A consequence of (social) cumulative advantages is the increase in the productivity of recognised scientists (Allison et al., 1982; Bol et al., 2018; Larivière & Gingras, 2010; Merton, 1968). Therefore, we analysed the top twenty most productive scientists over the years per field and their combined share in total knowledge output. Additionally, we calculated the Gini coefficient to examine the inequality in number of publications per author over the years.

Collaboration. It has been found that the probability of a scientist finding new collaborations increases with the number of previous collaborations they have had (Newman, 2001). This follows the same reward accumulation phenomenon as identified in the Matthew effect (Perc, 2014). We built a collaboration network to examine the network statistics over the years. The average path length, network diameter, degree centrality, network density and transitivity of the network provided indications for the development of collaborations.

Knowledge heterogeneity

Our definition of knowledge heterogeneity implies an indication of relatedness among a subset of knowledge based on their 'genes'. More heterogeneous subsets would encompass less relatedness as opposed to less heterogeneous sets of knowledge. We acknowledge that there should be a certain minimum of relatedness among subsets when these comprise distinct scientific fields as cognitive similarities are a prerequisite for further knowledge production (Arthur, 2007; Leydesdorff & Schank, 2008; Watts & Gilbert, 2011). Constraints in resources meant that we could not thoroughly identify genes on a detailed level e.g. in Pfeiffer and Hoffmann's (2007) work, where they used gene names, synonyms and gene products to analyse abstracts and titles. The genes in our context are the cited journal articles or article keywords that are referred to in a collection of journal articles. The keywords used are *KeyWords Plus*® which have been automatically generated by the WoS based on index terms in titles of cited articles and are therefore better suited to objectively indicate research topics (Garfield, 1990; *Web of Science Core Collection Help*, n.d.; Zhang et al., 2016).

Retention. We used citation distributions to indicate the retention of knowledge as articles cite previous works when referring to claims and ideas that are being built upon (T. Kuhn et al., 2014; Nicolaisen, 2007; Pfeiffer & Hoffmann, 2007). To measure the inequality of successful retention among published papers, we calculated the Gini coefficient for the

number of citations papers have received from the year of observation up to the date of collection. We calculated the same for the total number of citations keywords have received after aggregating the total number of article citations by keyword.

Relatedness. We used the concept of bibliographic coupling, which implies that articles are bibliographically related based on the overlap of identical cited references (Kessler, 1963). We are interested in the distribution of the number of times these references have been cited. For each reference, we counted the number of published articles in the year of observation that has cited the reference. Using the Gini coefficient, we calculated the inequality among the appearances of references over the years to indicate whether the field is becoming more or less bibliographically coupled, thus related. We also examined the top twenty most cited references for each year based on growth and stability.

Research direction. We used a method for mapping the dynamics of scientific knowledge production through analysing the network statistics of keyword co-occurrences (Cheng et al., 2018; Li et al., 2016; Tijssen & Van Raan, 1994; Whittaker, 1989). This is a network with a similar philosophy as the scientist collaboration network (as it is a co-occurrence network but based on authors). Additionally, the co-occurrence of authors and keywords were drawn from the same set of articles, which allowed us to properly compare patterns between the two networks. As such, we calculated the same indicators: average path length, network diameter, degree centrality, network density and transitivity. Lastly, we calculated the Gini coefficient for the number of publications per keyword occurrence per year.

Operationalisation table

See table 1.

Table 1: Operationalisation table

Indicator	Description	Measurement
Recognition	Inequality in the distribution of total citations scientists received for each paper published in the year of observation until the date of sampling.	0 - 100 Gini coefficient
	Inequality in the distribution of total citations (appearances) first-author scientists received for each paper cited by articles in the year of observation.	0 - 100 Gini coefficient
	Stability of the annual top twenty most cited first-author scientists by articles published in the year of observation.	0 - 100 index of stability
Productivity	Share of the top twenty scientists with the most publications per year compared to the total number of publications per year.	0 - 100 percentage

	Inequality in the distribution of published articles per scientist per year.	0 - 100 Gini coefficient
Collaboration	The average length of all shortest paths between every scientist in the collaboration network.	1 - ∞ average path length
	The longest optimal distance between any two scientists.	1 - ∞ diameter
	Normalised distribution of the number of collaborations per scientist.	0 - 1 degree centrality
	Number of existing collaboration triplets among scientists divided by the number of theoretical triplets among scientists.	0 - 1 transitivity
	Total number of existing collaborations among scientists divided by the total number of potential collaborations.	0 - 1 network density
Retention	Inequality in the distribution of total citations articles received between the year of publication and date of sampling.	0 - 100 Gini coefficient
	Inequality in the distribution of total citations keywords received between the year of publication and date of sampling for each paper that included the keyword.	0 - 100 Gini coefficient
Relatedness	Inequality in the distribution of total citations (appearances) articles received by articles published in the year of observation	0 - 100 Gini coefficient
	Stability of the annual top twenty most cited articles by articles published in the year of observation.	0 - 100 index of stability
	Share of the top twenty articles with the most cited appearances per year compared to the total number of references per year.	0 - 100 percentage
Research direction	The average length of all shortest paths between every keyword in the co-occurrence network.	1 - ∞ average path length
	The longest optimal distance between any two keywords.	1 - ∞ diameter
	Normalised distribution of the number of co-occurrences per keyword.	0 - 1 degree centrality
	Number of existing co-occurrence triplets among keywords divided by the number of theoretical triplets among keywords.	0 - 1 transitivity
	Total number of existing co-occurrences among keywords divided by the total number of potential co-occurrences.	0 - 1 network density

Data analysis

The citation data was analysed using the statistical programming language *R*. We used the syntax library *bibliometrix* to make article metadata more accessible (Aria & Cuccurullo, 2017). Useful data implied for each journal article the author(s), keywords[3], cited references, publication date and WoS category. The data was aggregated per category and year.

[3] Using the *KeyWords Plus*® index terms which are automatically generated by the WoS based on titles of cited articles. Author keywords are available only from 1991 and onwards.

All indicators were calculated for all four research fields per year. We examined the patterns that evolve over time for each indicator. Additionally, we plotted the indicators over annual publication output. Lastly, we analysed the correlation of the social structure and knowledge heterogeneity per scientific field through a two-dimensional plot.

Research quality

The quality of this research is preserved by establishing validity (internal, external and construct) and mitigating any expected risks. The reliability of this thesis concerns its repeatability (Bryman, 2008). To ensure accurate repeatability, we adequately commented and included any code written in R and ensured its accessibility. Additionally, the data sources are available through the Web of Science and the sampling strategy is described. However, the interpretation of results relied on interpretation in combination with collected theory, and we too are bound in rationality and perception. Therefore, we cannot guarantee peers to reach similar conclusions. However, as this thesis is explorative, we only encourage any further discussion.

The internal validity concerns causality (Bryman, 2008), which does not apply to this thesis. For the external validity, which concerns the generalisability of our research, we expect applicability to other scientific fields, with the notion that these fields should be at least generally understood. We have also found evidence that status effects are not restrained within the scientific system (Sauder et al., 2012), and that cumulative advantage mechanisms similar to the Matthew effect are widely common throughout many social instances (DiPrete & Eirich, 2006), where skewed distributions of firms' sizes and market shares may be the result of these (DiPrete & Eirich, 2006; Santarelli et al., 2006). Lastly, the construct of a scientific field is based on the interactivity and networks between scientists and topics, making it dynamic and fluctuating over time (Sun et al., 2013). We elaborate upon the topics concerning the external and construct validation more thoroughly in the discussion.

Results

The sampling method returned metadata for a total of 701,356 articles. Download limitations from the Web of Science meant that the data could only be retrieved in sets of maximum 500 articles at a time. A total of 1.409 .txt-files were downloaded and converted into a single data frame using the *convert2df()*-function from the package '*bibliometrix*' in *R*. This resulted in a sample size of 701,279 unique articles. Converting the separate .txt-files to a single data frame thus lost 77 articles in the process, a loss of 0.01%. However, no error messages were returned during the data conversion and it could not be discovered which articles went missing. Given the relatively small loss, we expect no significant impact on our results. Finally, the Web of Science can assign multiple field categories to an article, meaning that a number of articles have been categorised in more than one of our chosen scientific fields. The cumulation of all publications per field therefore resulted in 704,453 articles, 3,174 more than the number of unique articles. All our analyses ran separately for each field, using a subset of articles all containing the corresponding field tag. We did not account for any overlap between these fields in our analyses.

Field characteristics

Annual productivity

The publication data shows superlinear cumulative growth patterns over the years for all four fields. As can be seen in Figure 1 from the continuous line, the total number of published articles between the fields Astronomy & Astrophysics (ASTRO) and Organic Chemistry (CHEM) remains fairly equal over the years, while the fields Biotechnology & Applied Microbiology (BIOT) and Nanoscience & Nanotechnology (NANO) lack behind but accumulate steeper in the last decade. In terms of yearly productivity, NANO advanced from the smallest output in 1980 (0.78%) to the largest in 2012 (29.44%). BIOT shows a similar pattern, albeit less extreme than NANO, contributing the second-smallest output in 1980 (14.46%) and second-largest in 2012 (26.36%). A more detailed overview with an increment of eight years is provided in Table 2. In Figure 1 the yearly productivity is plotted with the dotted lines.

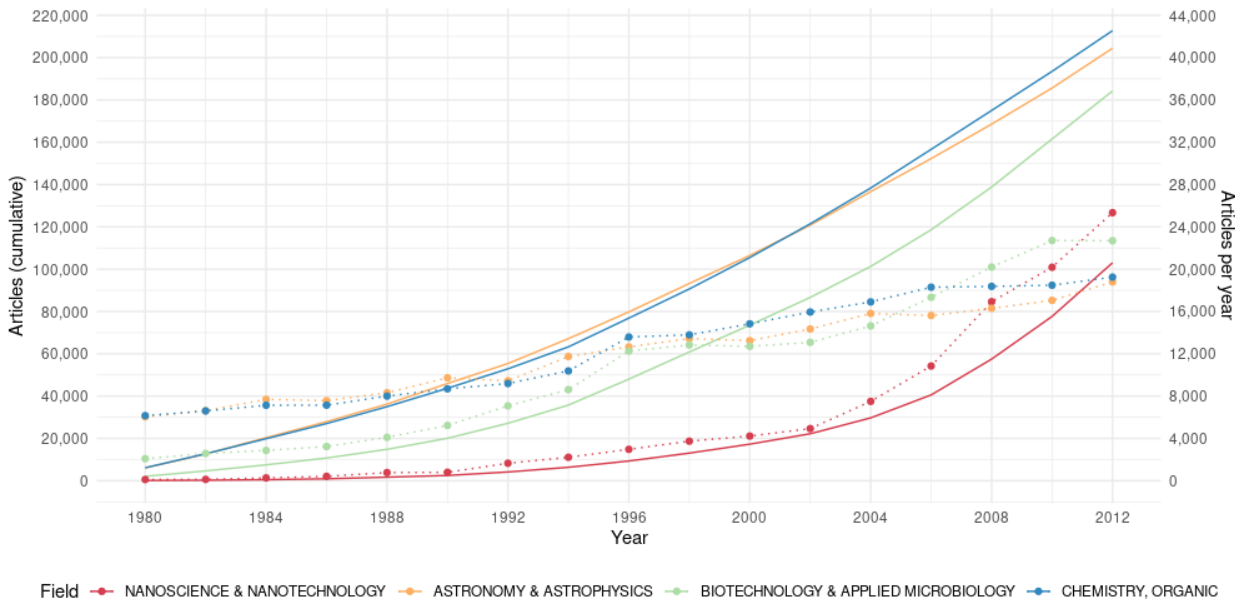

Figure 1: Continuous line: cumulative publications, dotted line: annual publications

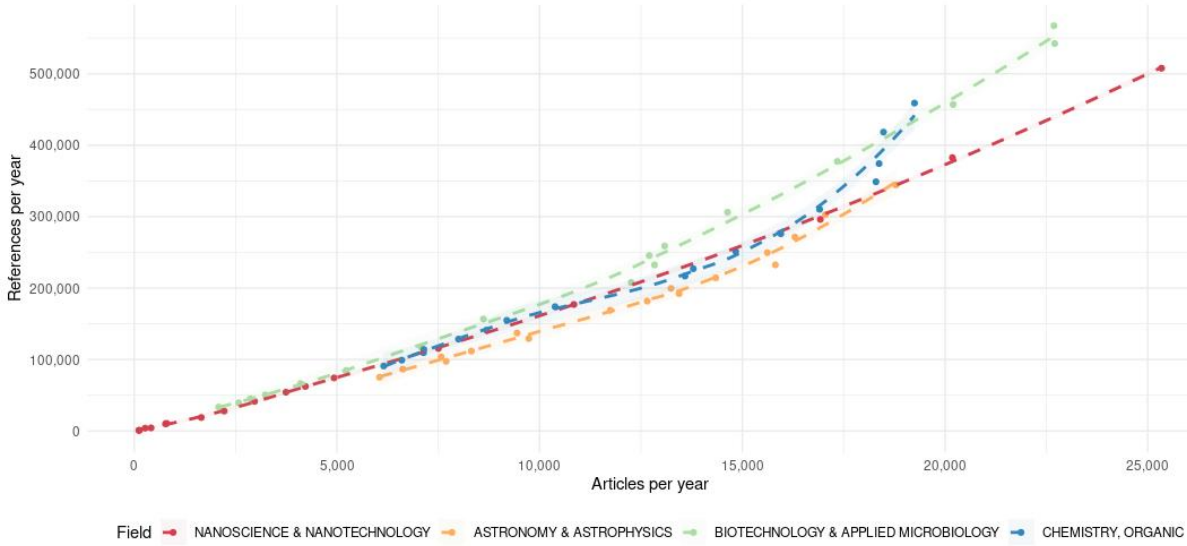

Figure 2: Yearly number of references over yearly publications

Table 2: Share in annual publications per field over the years

Field	**1980**		**1988**		**1996**		**2004**		**2012**	
NANO	112	*0.78%*	771	*3.64%*	2,969	*7.16%*	7,503	*13.68%*	25,343	*29.44%*
ASTRO	6,050	*42.02%*	8,316	*39.25%*	12,654	*30.51%*	15,816	*28.83%*	18,792	*21.83%*
BIOT	2,082	*14.46%*	4,103	*19.36%*	12,260	*29.56%*	14,635	*26.68%*	22,689	*26.36%*
CHEM	6,154	*42.74%*	7,998	*37.75%*	13,587	*32.76%*	16,907	*30.82%*	19,248	*22.36%*
TOTAL	14,398		21,188		41,470		54,861		86,072	

Scientist involvement

In Figure 3 the number of unique author scientists per field over the years is displayed by the continuous line. The dashed line shows the number of unique scientists excluding those who only published single-authored papers in that year. This indicates that the vast majority of scientists collaborate on their publications. Furthermore, the number of articles seems to correlate linearly with the number of scientists involved, as demonstrated by the black line in Figure 3. Pearson's correlation coefficients were calculated for the whole data set and per field, which are further disclosed in Table 3 including the p-values and confidence intervals.

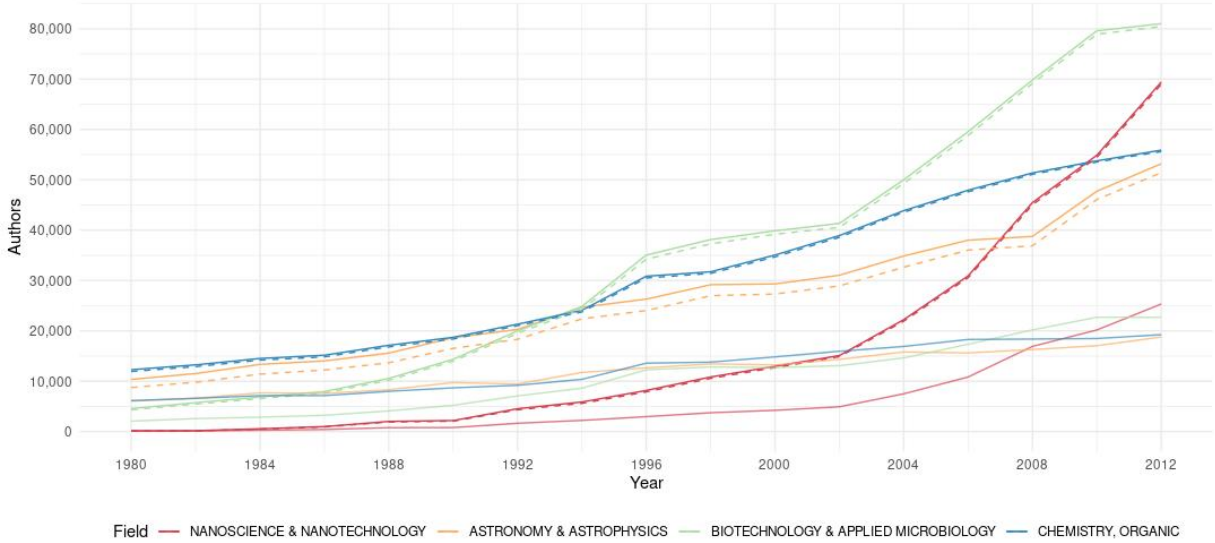

Figure 3: Continuous line: total individual authors. Dashed line: total individual authors of multi-authored papers

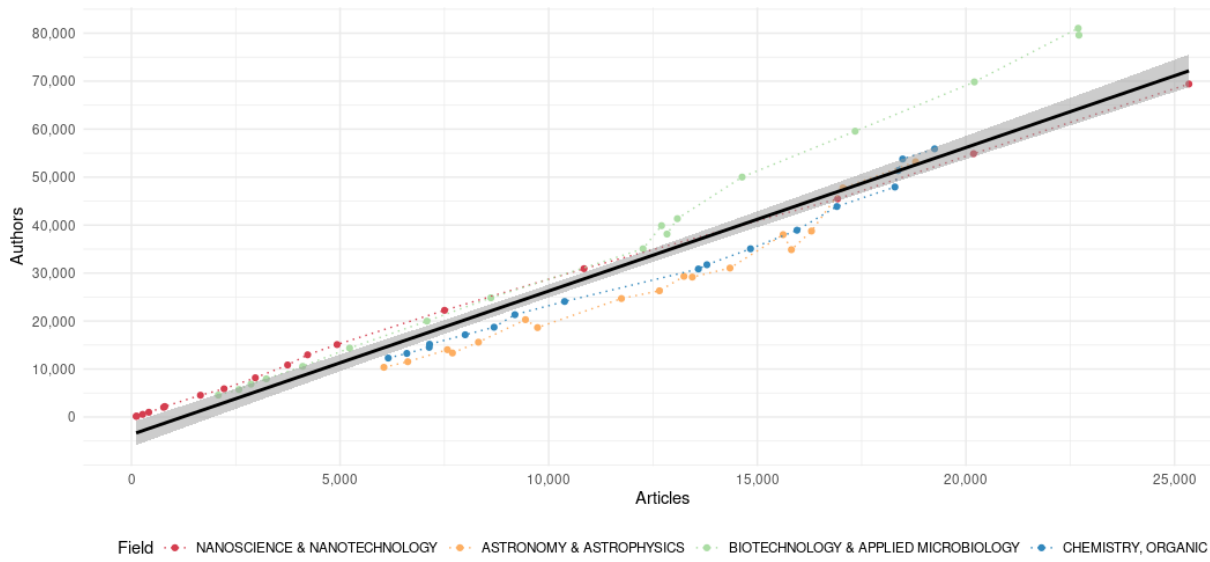

Figure 4: Number of yearly individual authors over yearly articles

Table 3: Pearson's correlation coefficient for annual articles and involved scientists

Field	**Pearson's *r***	**95% conf. int.**	**p-value**
NANO	0.9993624	0.9981833 - 0.9997763	< 2.2e-16
ASTRO	0.9793160	0.9421390 - 0.9926958	= 8.05e-12
BIOT	0.9967444	0.9907464 - 0.9988569	< 2.2e-16
CHEM	0.9729157	0.9272059 - 0.9900707	= 1.337e-11
TOTAL	0.9619656	0.9388786 - 0.9764380	< 2.2e-16

Social structure

Recognition

Based on the Gini coefficient for citations scientists received for their papers published per year, CHEM was the field with the least unequal distribution of citations among scientists with an average Gini coefficient of 58.17. BIOT was second with an average of 64.04, followed by NANO (68.28) and ASTRO (70.21). Moreover, CHEM was the only field with a downward trend in inequality since 2000, whereas all other fields have only become more unequal. NANO had a steep fluctuation between 1980 and 1984, with its highest Gini coefficient at 76.77 in 1982 and its lowest at 62.15 in 1984. From 1990 to 2012, the trend has been going upwards. The

Gini coefficient concerning citations scientists received for their papers published over the years is visualised in Figure 5.

The average Gini coefficient for the distribution of citations among first-author scientists of cited references indicates that NANO was least unequal with a value of 43.24, closely followed by BIOT (46.29). The fields CHEM and ASTRO were significantly more unequal with an average Gini coefficient of 61.12 and 70.42, respectively. However, as shown in FIG R5, NANO had the steepest upward trend, passing BIOT in 2004 and matching CHEM in 2012. All fields followed trends of increasing inequality among cited scientists with the fields ASTRO and CHEM seemingly paralleled, the former almost reaching a coefficient of 75 in 2012. The two fields also show the least fluctuation in inequality over the years. This is shown in Figure 6.

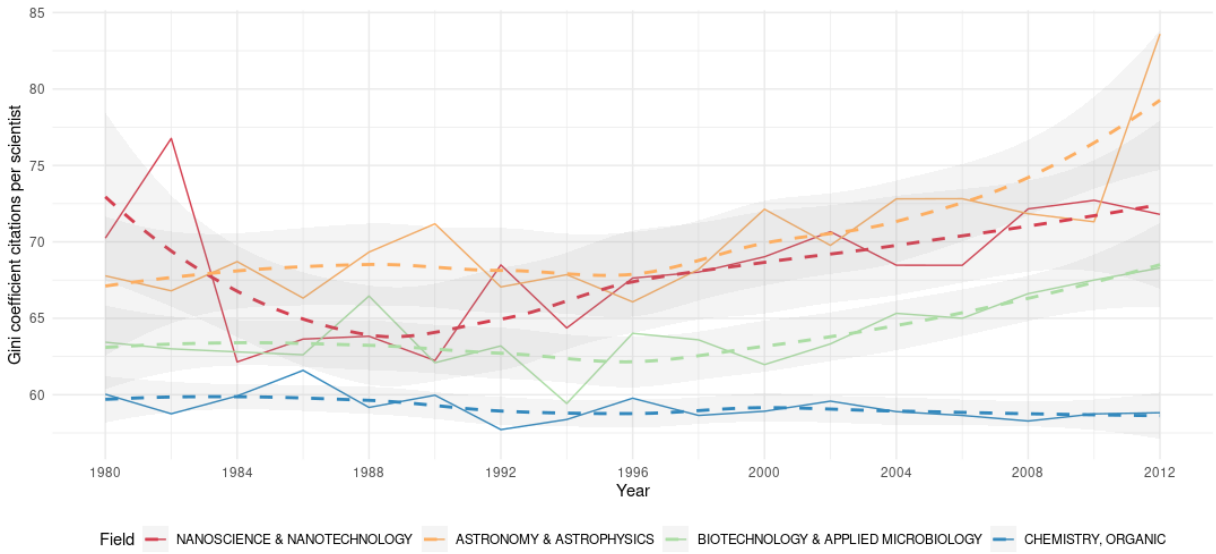

Figure 5: Citation inequality among scientists over the years

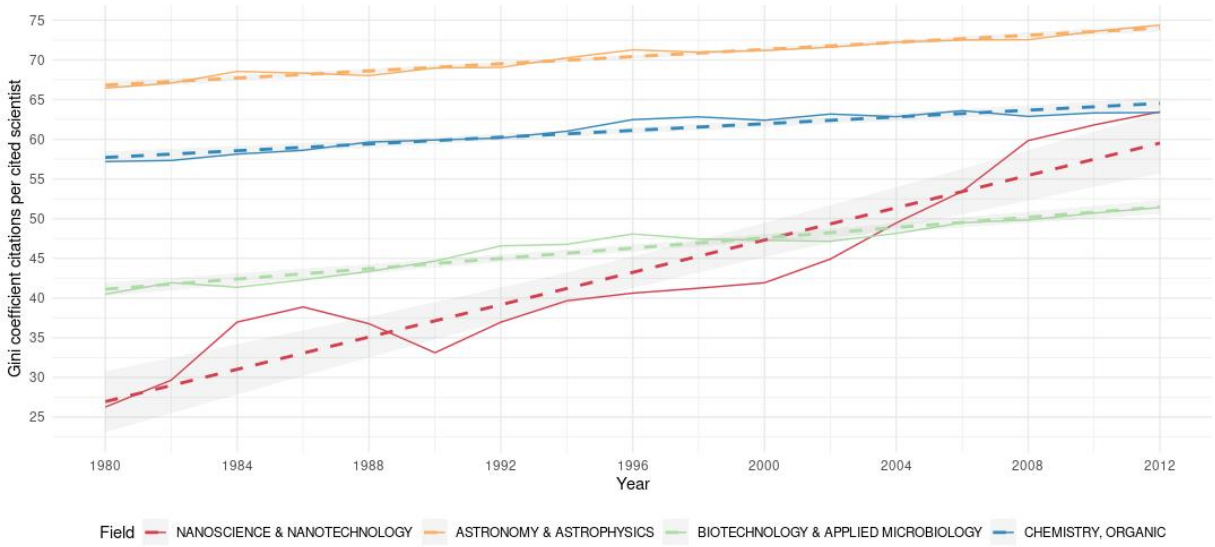

Figure 6: Citation inequality among cited scientists over the years

The field with the least stable top twenty was NANO, where our stability metric is indexed at 0.90. ASTRO and BIOT both have a stability index of 3.14 and CHEM can be considered most stable with an index of 6.04. The index increases for all fields when calculating the stability for the top five most cited authors, with NANO (3.85) still indexed the least stable, followed by ASTRO (8.84), BIOT (21.25) and CHEM (26.23).

Productivity

The twenty most productive scientists in NANO published 7,706 articles combined over the years, making up 7.48% of all papers published in the field. In ASTRO, the top twenty published the most articles compared to the other fields, with a total of 12,752. Compared to the total number of articles published by all scientists in ASTRO, the top twenty contributed for 6.24%. The top twenty scientists in the fields BIOT and CHEM were comparably productive with a total of 8,743 (4.75%) and 8,524 (4.01%), respectively. Throughout the years, the relative share of publications by the top twenty scientists did not change much over time. However, as the number of yearly publications and scientist involvements did increase for each field, the absolute number of publications by the top twenty had to increase as well. The twenty most productive scientists in NANO had a relatively large share in the number of publications but with a steep decline in the first two and a half decades, which is visualised in Figure 7. This can partly be explained by the fact that NANO was significantly smaller when compared to the other fields; e.g. in 1990, both ASTRO and CHEM had an output more than ten times as much as NANO. See Table 2 on page 27 for better comprehension.

The Gini coefficients regarding the inequality among the number of publications per scientist are plotted in Figure 8. Compared to the inequality among citations, the distribution of publications per scientist seems slightly more equal. ASTRO embedded the highest inequality each year and it seems to have increased the most over the years as well. NANO and BIOT embedded nearly similar inequalities until around 2002 after which the coefficient for NANO increased more. BIOT and CHEM show the least fluctuation over the years and a less steep increasing trend than the other two fields.

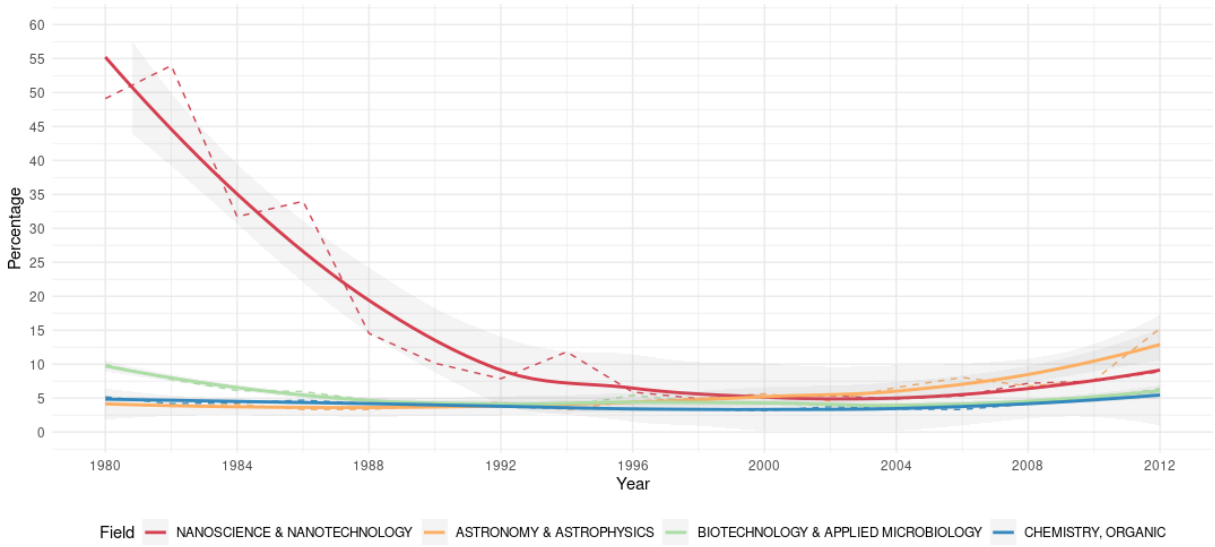

Figure 7: Share in publications of the top twenty most productive scientists per year

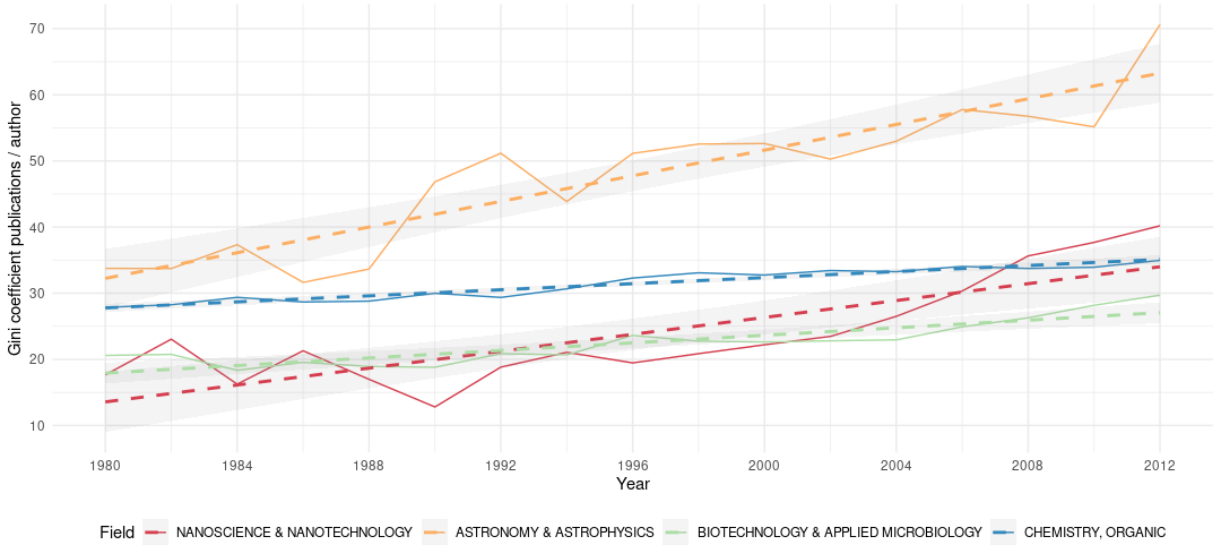

Figure 8: Productivity inequality among scientists per year

Collaboration

All fields showed superlinear growth in their network size, which represents the total number of individual scientists involved each year. The number of scientists in these networks is slightly lower than the total number of scientists involved (see Figure 3). This is the result of our method where we first calculated the mean number of scientists per paper for each year and field and deleted all co-author scientists listed higher than the mean. See our limitations for further elaboration. Figure 9 shows the number of scientists per year in the collaboration network.

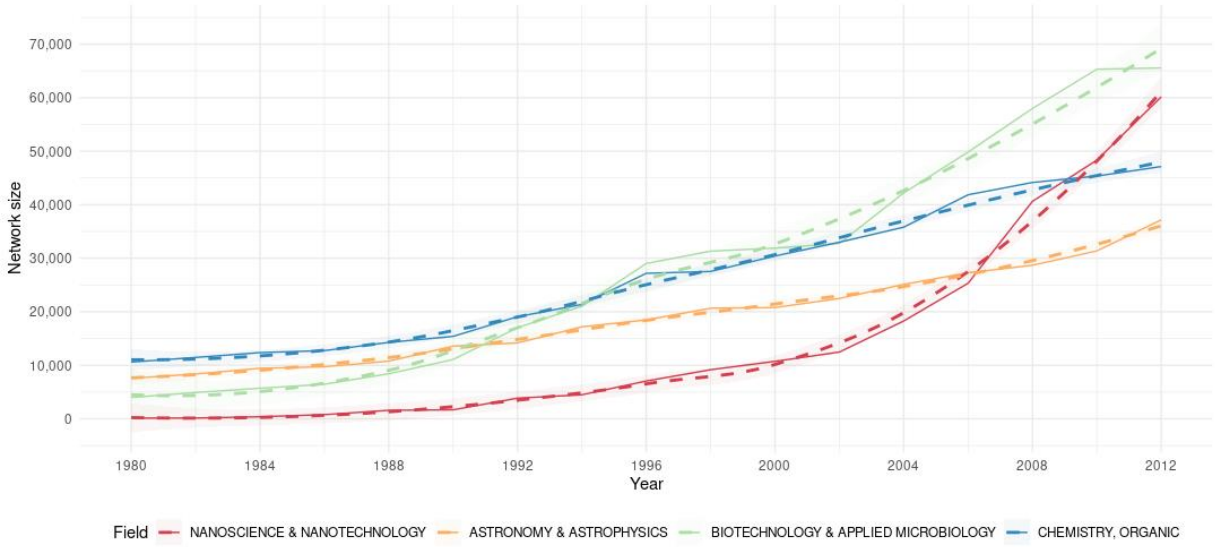

Figure 9: Number of individual scientists in the collaboration network over the years

Average path length

Over the years, the average path length increased for every field suddenly at some point in time after which it decreased again, more subtly, as shown in Figure 10. When we compared the average path length over the size of the collaboration network, we found that for all fields the longest average length was reached between 10,000 and 20,000 individual scientists. As the network increased after that, the average path length decreased again. BIOT showed the most subtle decrease, whereas the average path length in ASTRO and CHEM decreased more rapidly as the network size increased. NANO showed an initially steep decline, but after 25,000 individual scientists the decrease slowed down. This is visualised in Figure 11.

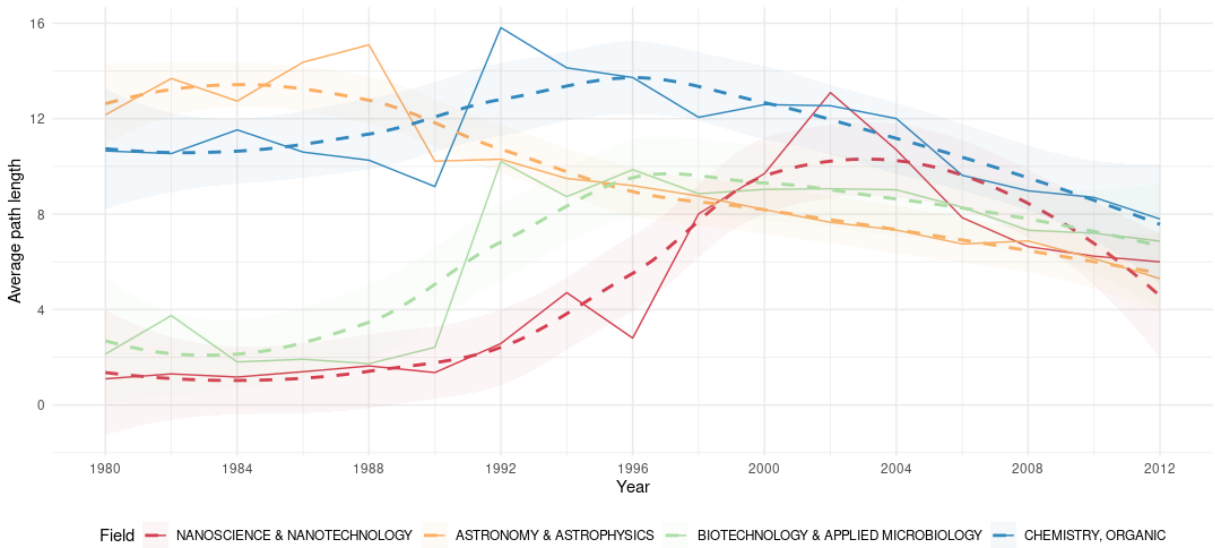

Figure 10: Average path length over the years

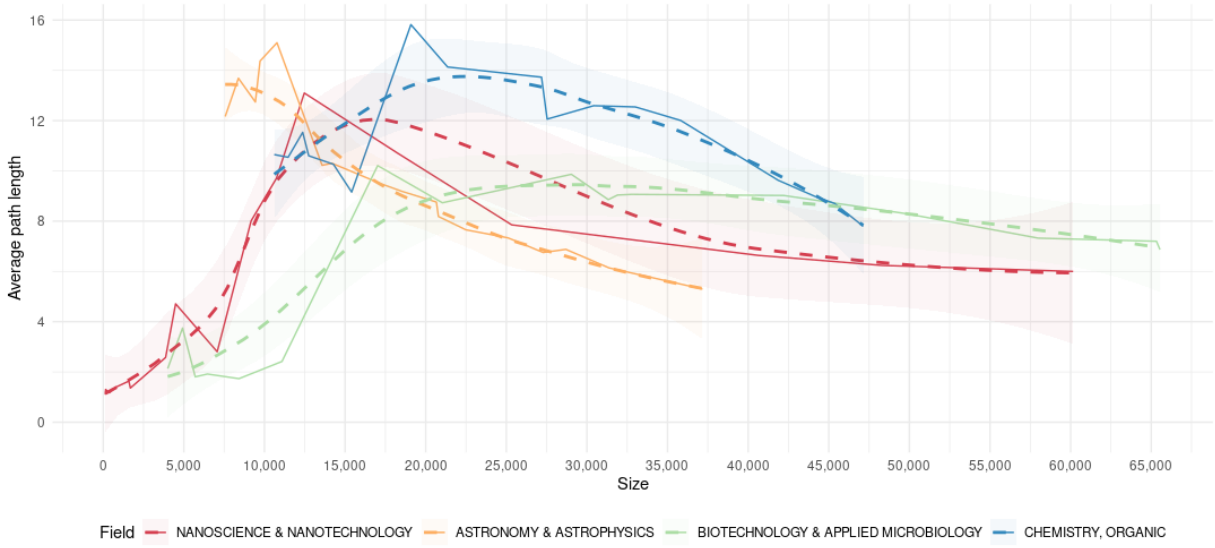

Figure 11: Average path length over collaboration network size

Degree centrality

The degree centrality seemed to make a U-shaped pattern over time for all fields except ASTRO. This was most clearly present in NANO, whereas the degree centrality in ASTRO seemed to increase superlinearly over time. Concerning the degree centrality over the size of the collaboration network, the U-shaped pattern becomes less visual. The lowest degree

centrality for all fields except ASTRO is reached with a network size between 10,000 and 20,000 scientists. As the network grew, the degree centrality increased seemingly superlinearly. The degree centrality in ASTRO seemed to only increase as the network grew, comprising the steepest increase as well. BIOT and CHEM showed similar degree centralities over the collaboration network size. Figure 12 and Figure 13 show the degree centrality over the years and size.

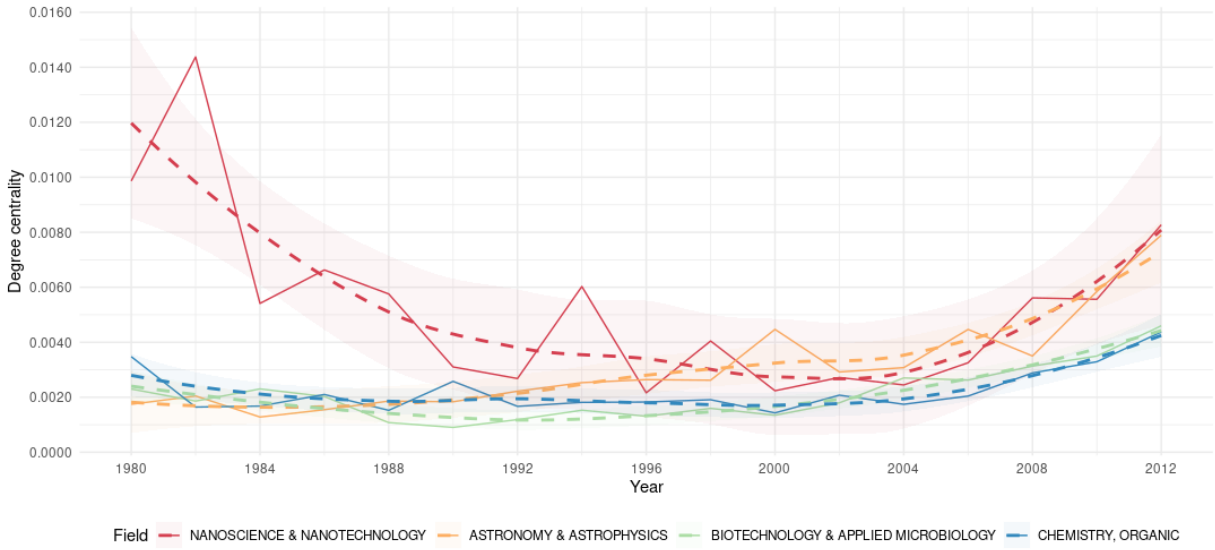

Figure 12: Degree centrality over the years

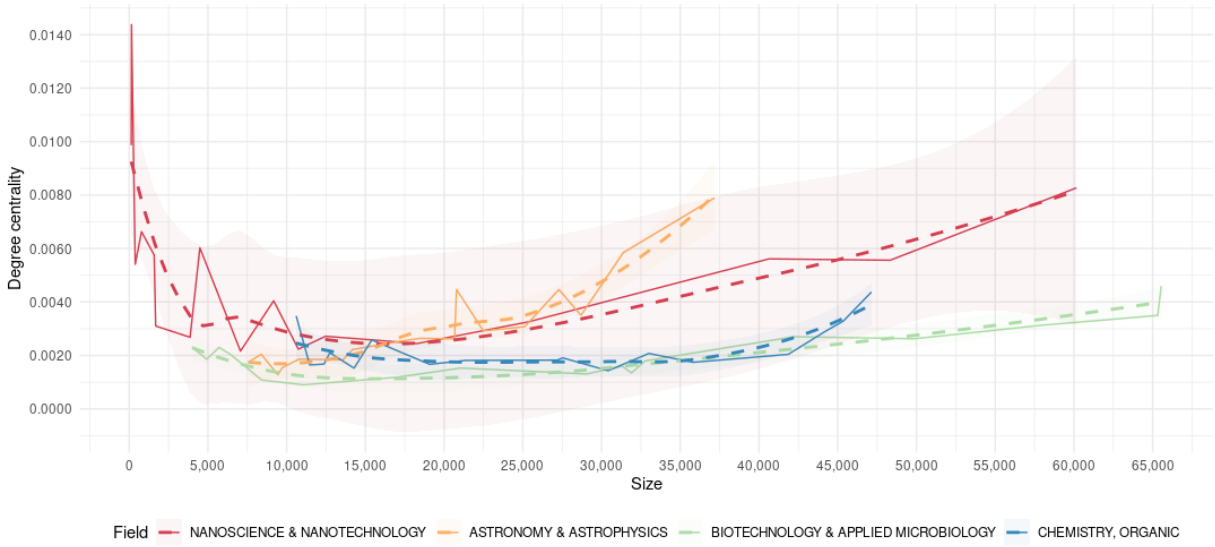

Figure 13: Degree centrality over collaboration network size

Transitivity

The transitivity, or clustering coefficient, decreased over time for each field with ASTRO and CHEM showing only a slight decline. BIOT and NANO had a higher transitivity but that started to decrease more rapidly after 1996. Additionally, NANO did not have a transitivity value until 1986. This is shown in Figure 14. When we considered the transitivity over the collaboration network size, in Figure 15, we found that the decrease of the clustering coefficient showed a superlinear pattern for BIOT, whereas we found a sublinear pattern for NANO with similar network sizes. For ASTRO and CHEM, the transitivity increased slightly up to a network size of 15,000 and 20,000, respectively, after which it decreased.

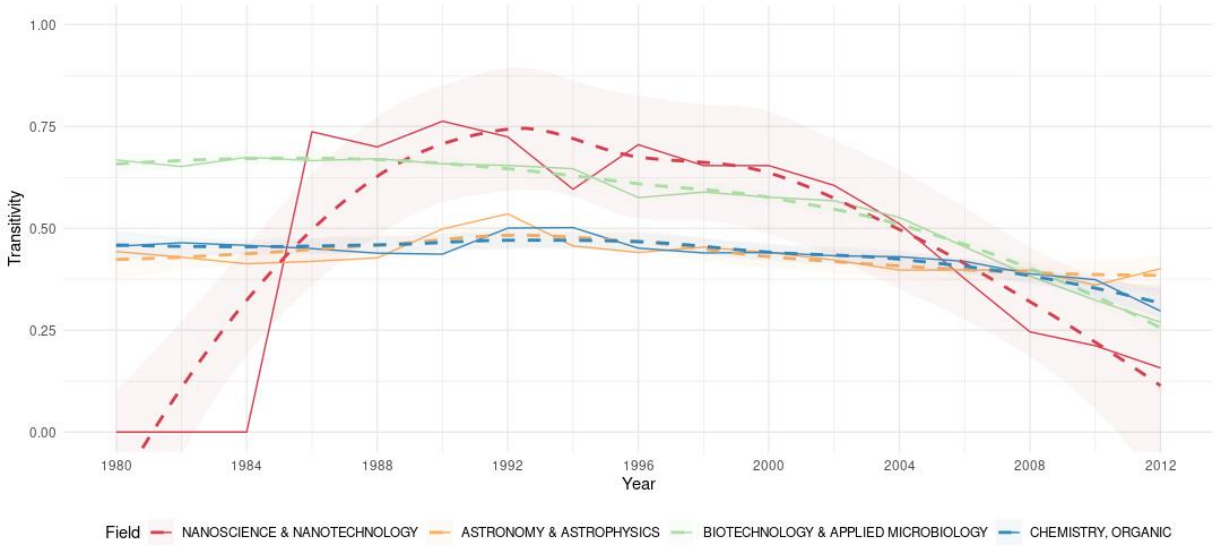

Figure 14: Clustering coefficient (transitivity) over the years

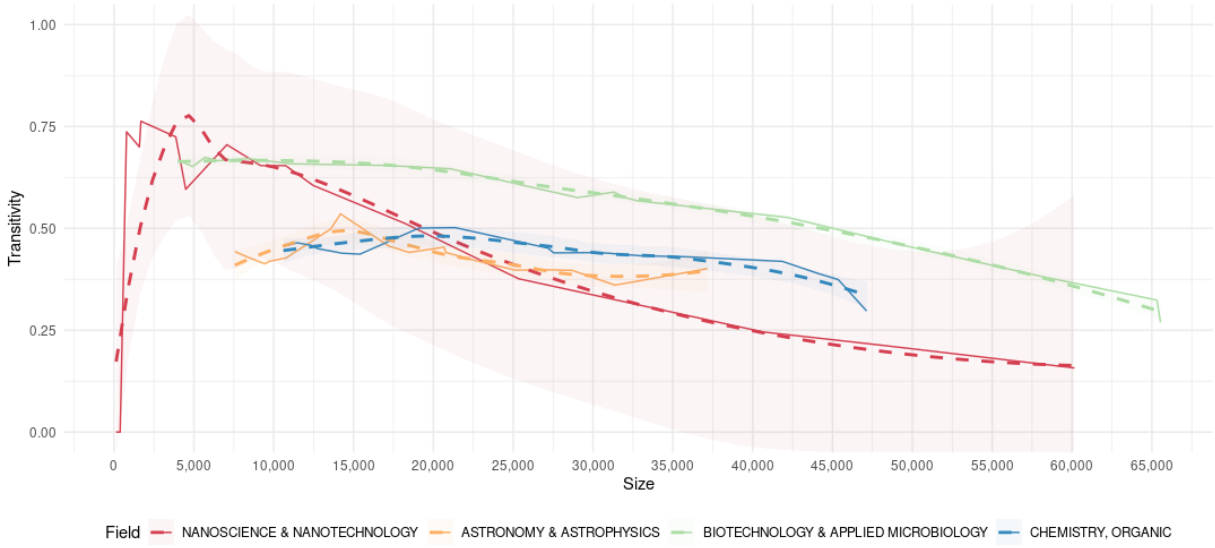

Figure 15: Clustering coefficient (transitivity) over collaboration network size

Network density

The network density seemed to remain the same over the years for all fields except NANO, with ASTRO slightly increasing between the years 2004 and 2012, see Figure 16. The network density in BIOT slightly decreased between the years 1980 and 1988. Furthermore, the

network density of NANO started considerably higher but decreased rapidly to similar values in the year 2012.

The patterns were similar when considering the network densities over the collaboration network size, see Figure 17, where the most considerable change was the increased steepness of decline in the network density for NANO. All fields except ASTRO showed similar correlations between the densities and sizes of the collaboration networks with the density declining as the network grew. However, ASTRO showed the opposite pattern, as the density increased superlinearly over its network growth.

Figure 16: Network density over the years. Top: total, bottom: zoomed in for densities < 0.0005

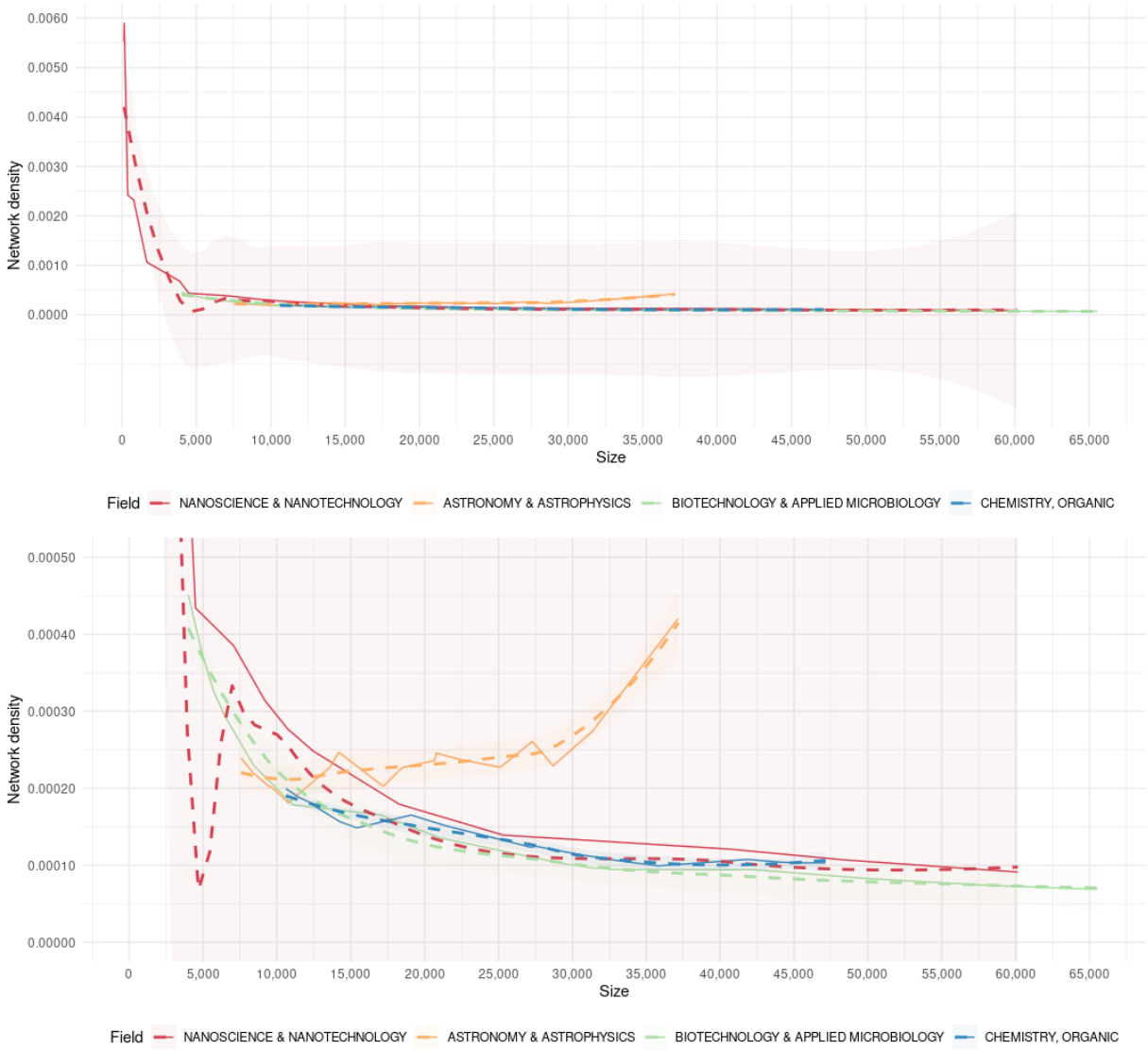

Figure 17: Network density over collaboration network size. Top: total, bottom: zoomed in for densities < 0.0005

Knowledge heterogeneity

Retention

In all four fields, the Gini coefficient for the distribution of citations per article had a decreasing trend, as shown in Figure 18. This concerns the inequality in citations articles received between the publication year and date of data collection. The fields BIOT and CHEM have shown a less steep decline in inequality with nearly parallel linear trendlines. ASTRO had the steepest decline, followed by NANO. On average, CHEM was least unequal with a Gini coefficient of 53.00. The other three fields were relatively close to each other in terms of

inequality, with BIOT (62.79) as second most equal followed by ASTRO (64.15) and NANO (67.39).

Keyword data was sparse if not unavailable for any field prior to 1992, as visualised in Figure 19. The averages of the Gini coefficients for citation inequality among keywords are significantly higher than the averages among articles described above. CHEM (75.95) and BIOT (77.99) were least unequal within the years 1992 - 2012, followed by NANO (78.37) and ASTRO (79.74). These averages are relatively close to each other. All four fields showed increasing trends: NANO had the steepest increase of inequality, encompassing the least inequality among citations per keyword in the early 1990's and the most inequality from 2006 onwards. ASTRO seems to have increased little in inequality over the twenty years period. CHEM and BIOT have increased almost evenly, with the former showing a slightly steeper trendline, but remaining the least unequal field after the year 2000 when NANO showed a sudden increase.

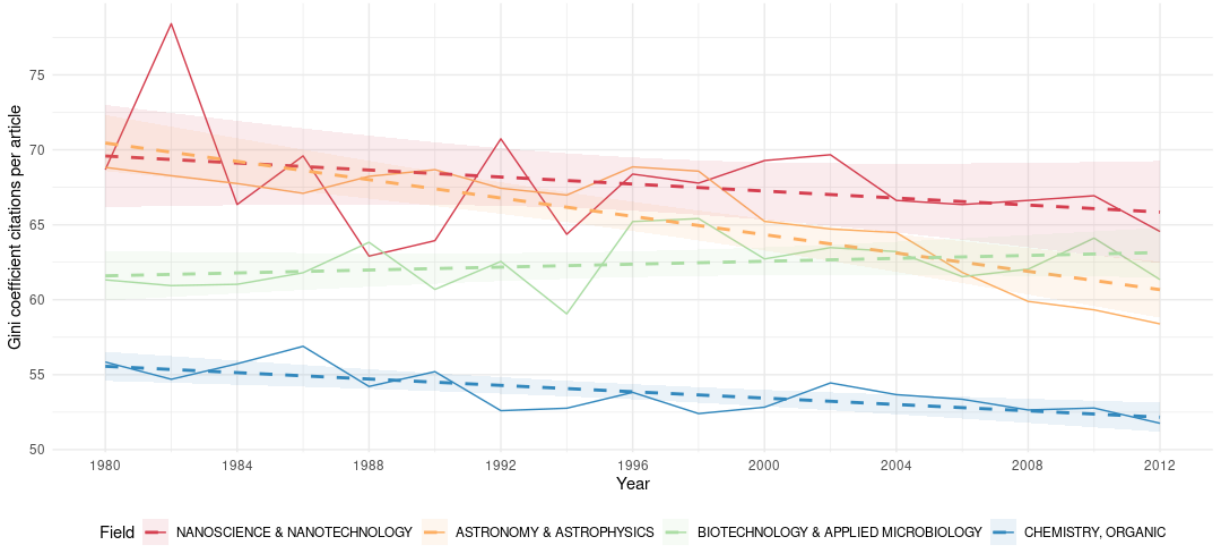

Figure 18: Citation inequality among published articles per year

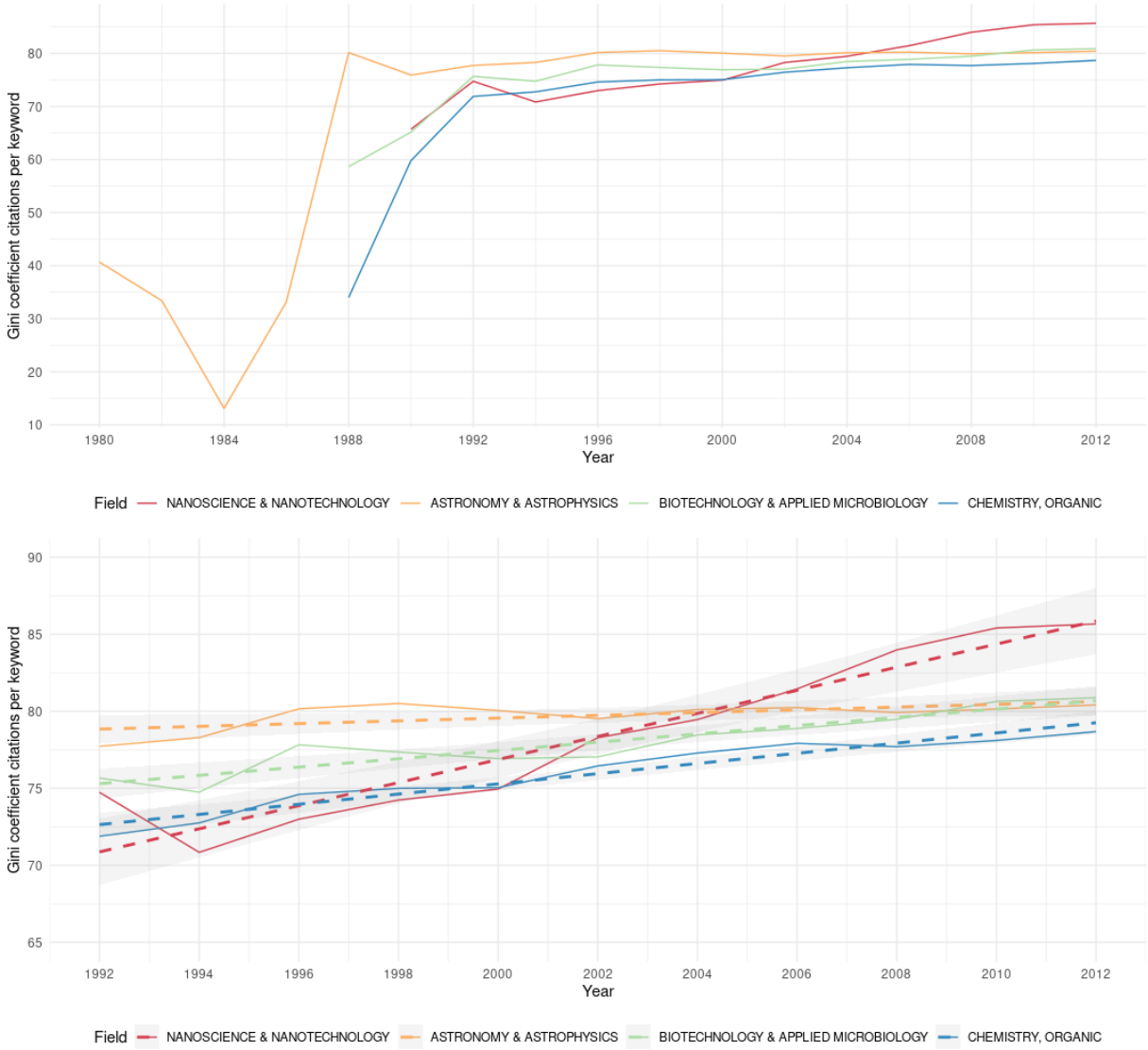

Figure 19: Citation inequality among keywords in published articles per year. Top: total, bottom: for period 1992 - 2012 including linear trend line

Relatedness

Compared to the other three fields, NANO increased most in terms of inequality among cited references over the years. As can be seen in Figure 20, the field also shows the most fluctuation around its trendline. Until 2006, it encompassed the least inequality, whereas from 2008 onwards it jumped to become the second most unequal field. The second-steepest trend appeared in ASTRO, which was substantially more unequal than any other field. The lowest Gini coefficient calculated for ASTRO was 39.26 in 1982, which is marginally higher than the highest Gini coefficient of the other three fields, which was 37.92 for NANO in 2012 when

ASTRO had reached its peak at 51.51. The averages of the Gini coefficients are 22.89 (NANO), 25.48 (BIOT), 31.61 (CHEM) and 45.46 (ASTRO).

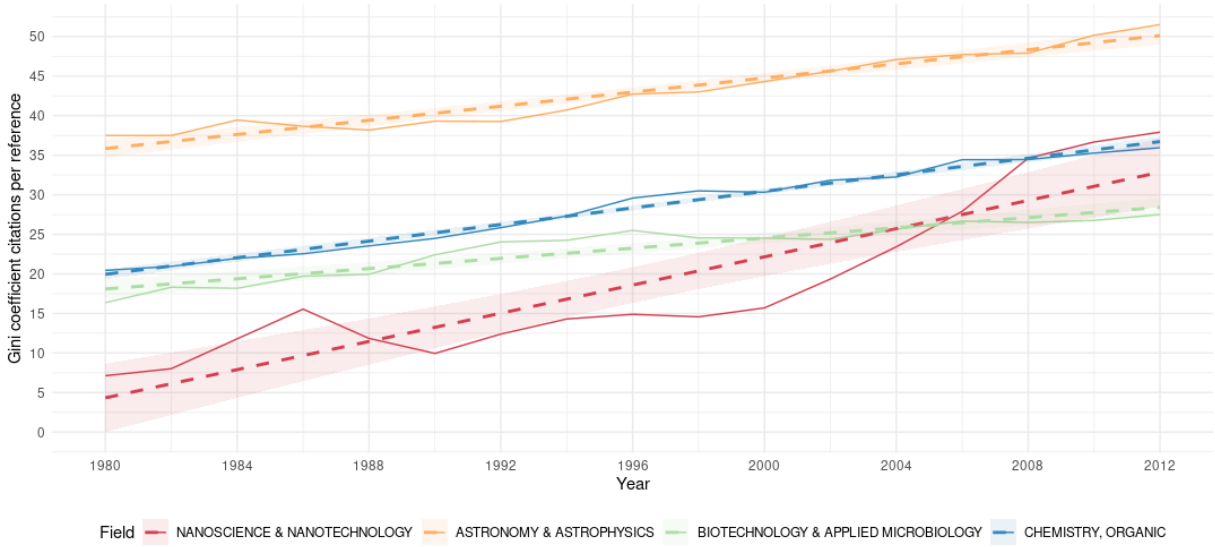

Figure 20: Citation inequality among cited articles per year

The share of the top twenty most cited references compared to the total number of references in percentages did not indicate linear trends. This is shown in Figure 21. The average share of the top cited references per field is relatively low, where the top cited articles in ASTRO contributing the most with 1.88%, followed by BIOT (1.59%), NANO (1.10%) and CHEM (0.92%). The contribution of the most cited references seemed most stable in CHEM, whereas the biggest difference over the years was present in NANO between 1980 (6.57% and 2000 (0.65%). NANO was also considered less stable with our index calculated at 0.63 (top twenty) and 2.71 (top 5). This was followed by ASTRO (1.73 and 4.20), CHEM (2.08 and 13.60) and BIOT (3.77 and 21.25). The order in ranking stability did not change when only taking into account the top five cited references, but the differences did become larger.

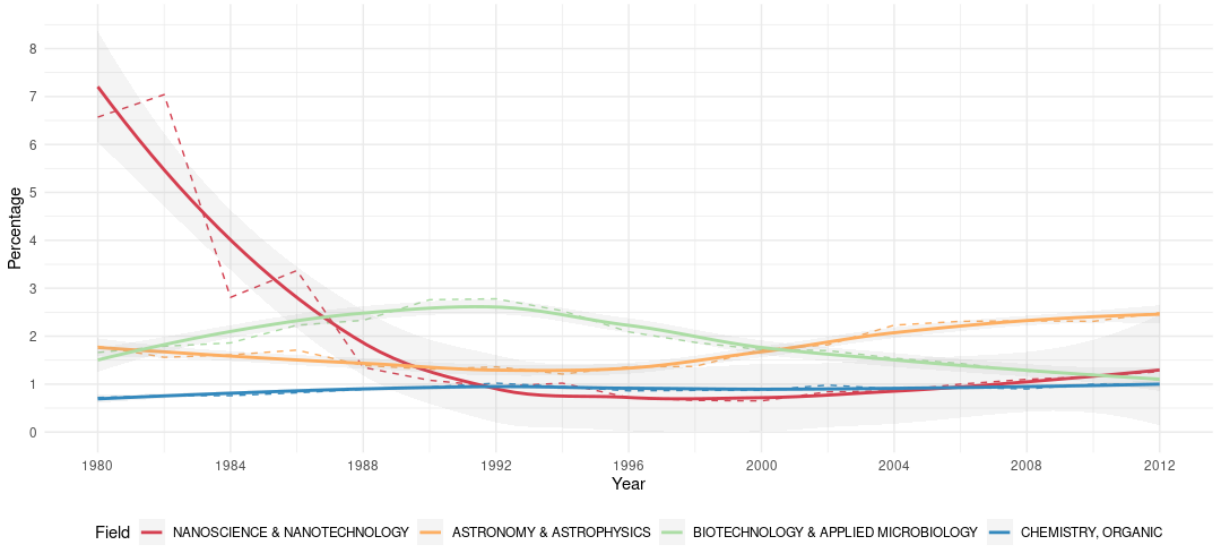

Figure 21: Share of citations of the top twenty most cited articles per year

Research direction

The network size, i.e. the number of unique keywords per year, is marginal for any field up to 1990 and seems to be representative from 1992 onwards. This matches previous results, indicating that our collection from the WoS did not have proper keyword data up to 1992. The number of unique keywords increased similarly for BIOT and CHEM between 1992 and 2004, after which the increase slowed down for CHEM. After 2002, the keyword network growth of NANO seemed to match BIOT. Both ASTRO and CHEM showed less growth in the last decade. This is shown in Figure 22.

Figure 22: Number of unique keywords per year

The Gini coefficient for the publication inequality among keywords, visualised in Figure 23, remained similar over the years for BIOT and CHEM. All fields saw an increase in publication inequality, with NANO encompassing the steepest increase. The other three fields increased comparably, with ASTRO showing a slightly more withheld increase. Publication inequality was largest for keywords in ASTRO until 2008 when NANO surpassed the field, the latter being the least unequal field until 2004.

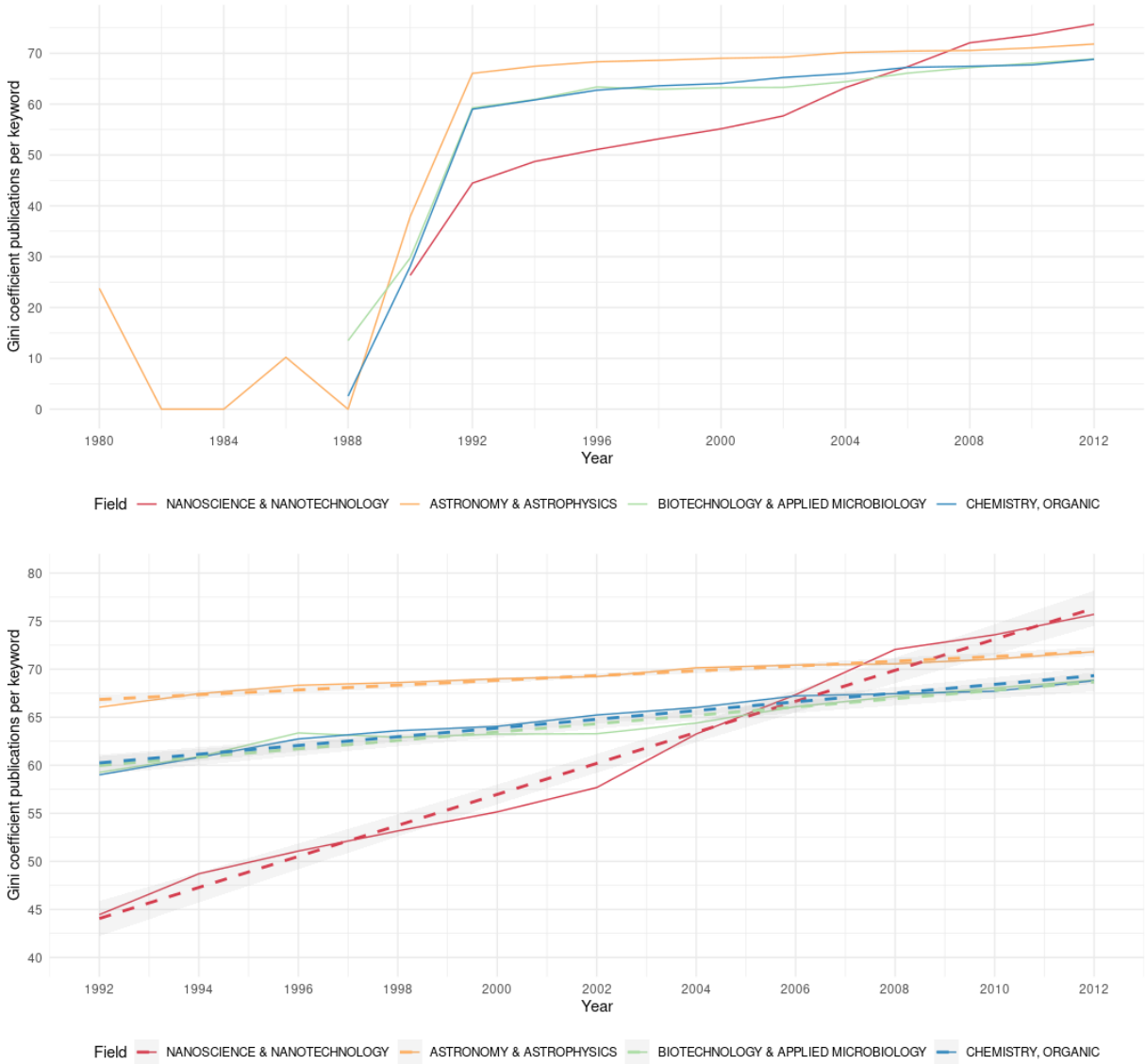

Figure 23: Publication inequality among keywords per year. Top: total, bottom: from 1992

Average path length

The average path length showed similar pattern behaviour for all four fields, especially when considering changes in average path length over network size. All fields had an average path length around 3.5 with 1,000 - 2,500 unique keywords. As the networks grew, the decrease in average path lengths slowed down and seemed to stabilize between 2.85 and 2.95. NANO and ASTRO showed similar decreases after the network grew from around 10,000 unique keywords. This also applied to BIOT and CHEM after around 15,000 keywords. See Figure 24 and Figure 25.

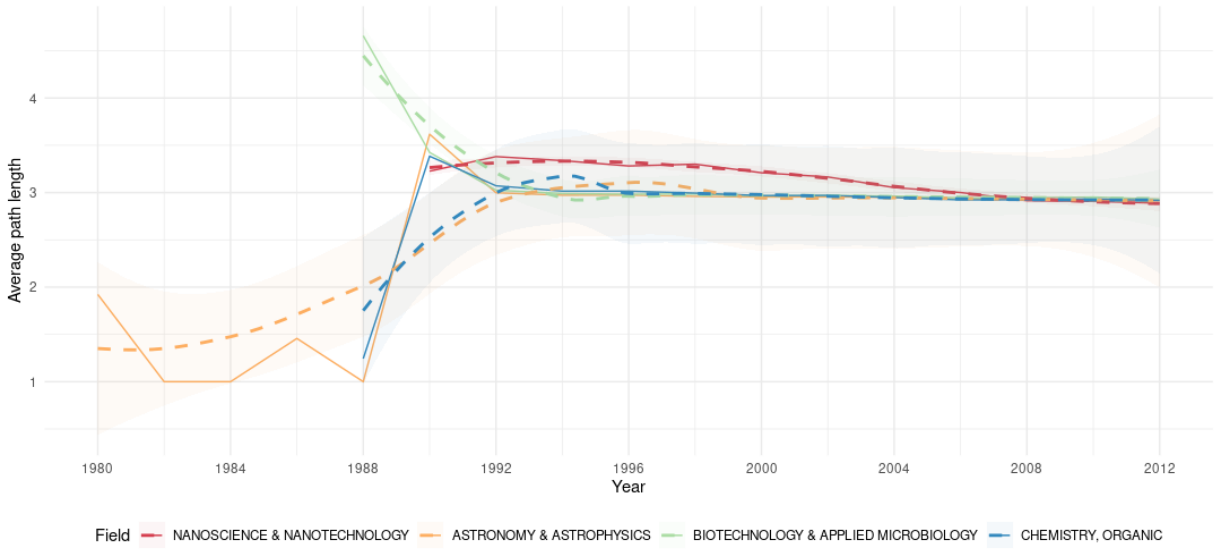

Figure 24: Average path length over the years

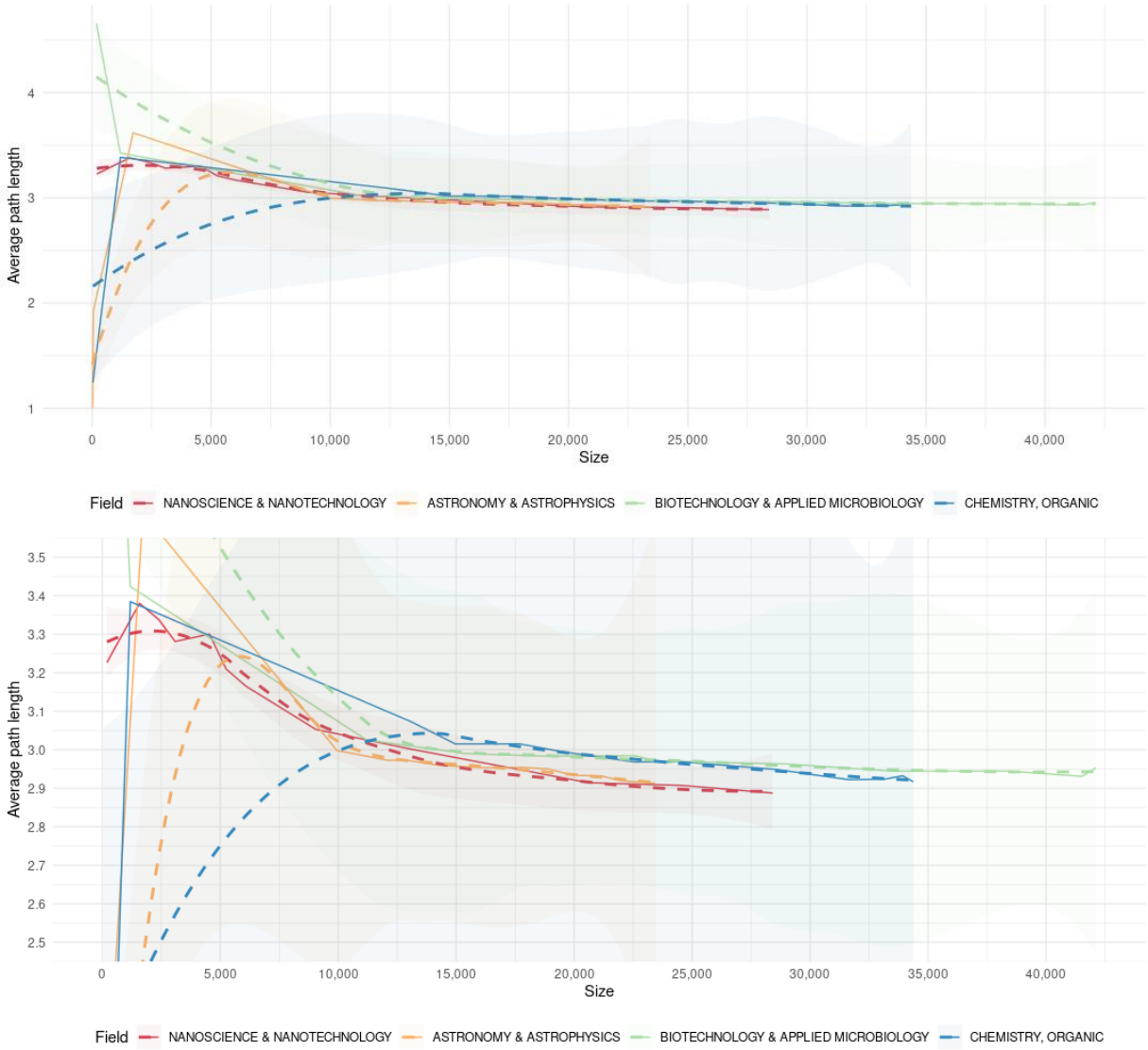

Figure 25: Average path length over number of keywords. Top: total, bottom: zoomed in for average path lengths between 2.5 - 3.5

Degree centrality

Over the years, the change in degree centralities in NANO and CHEM seemed to be opposite of each other: the pattern in NANO was more U-shaped, whereas CHEM seemed more inverse U-shaped. However, over network growth the patterns of the two fields seemed similar, with the degree centrality in CHEM being the highest for all network sizes. ASTRO and BIOT showed similar degree centralities over time and network size. See Figure 26 and Figure 27.

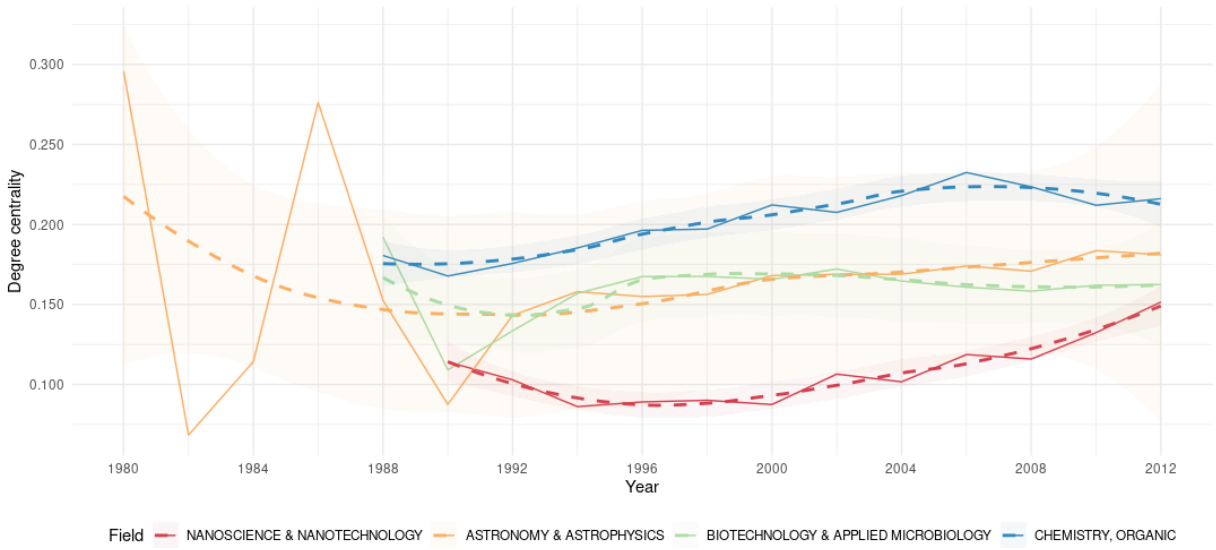

Figure 26: Degree centrality over the years

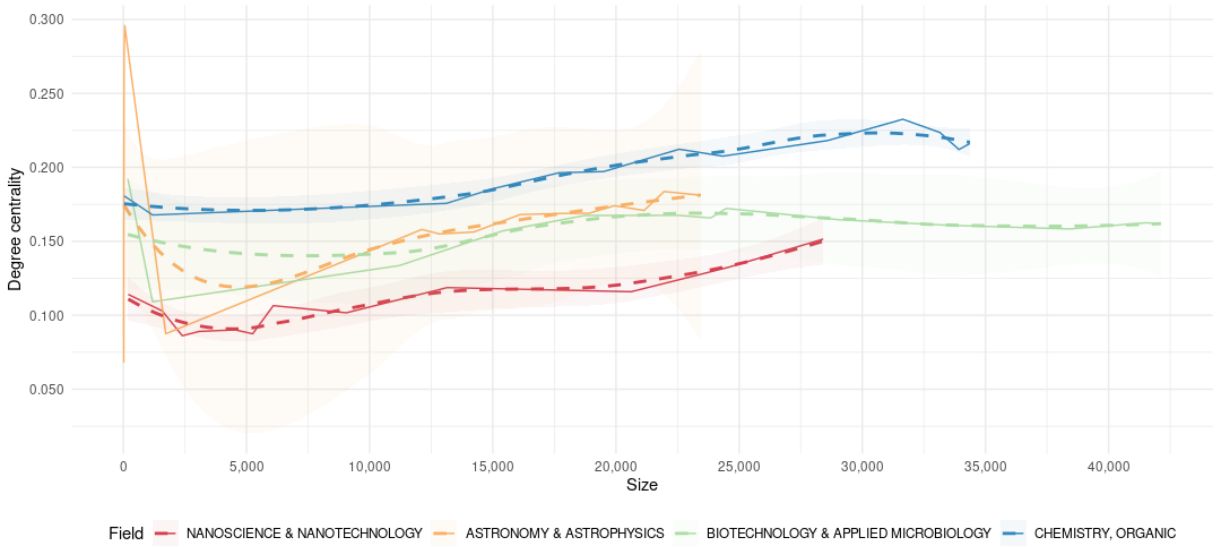

Figure 27: Degree centrality over number of keywords

Transitivity

The transitivity, or clustering coefficient, decreased for every field over the years. Considering only the period after 1992, the decrease was slightly sublinear. Over the years, but most considerably over the increase in number of keywords, we saw the pairings ASTRO and NANO, and CHEM and BIOT containing almost identical clustering coefficients when the number of unique keywords in the network became larger than 15,000. The transitivity was

slightly higher in the former two fields. All fields showed stabilising transitivity over time and network size, which can be seen in Figure 28 and Figure 29.

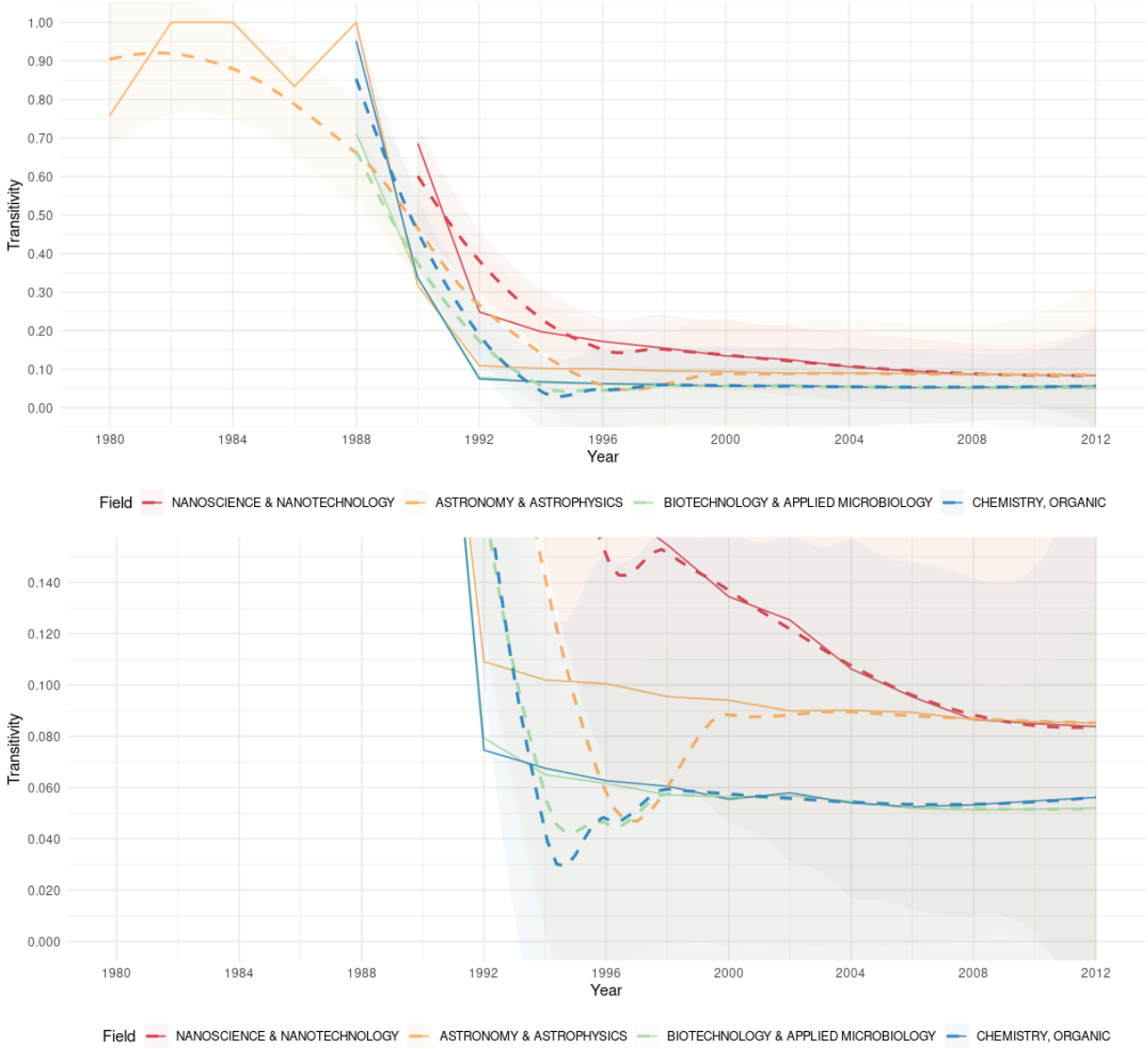

Figure 28: Transitivity over the years. Top: total, bottom: zoomed in for transitivity < 0.15

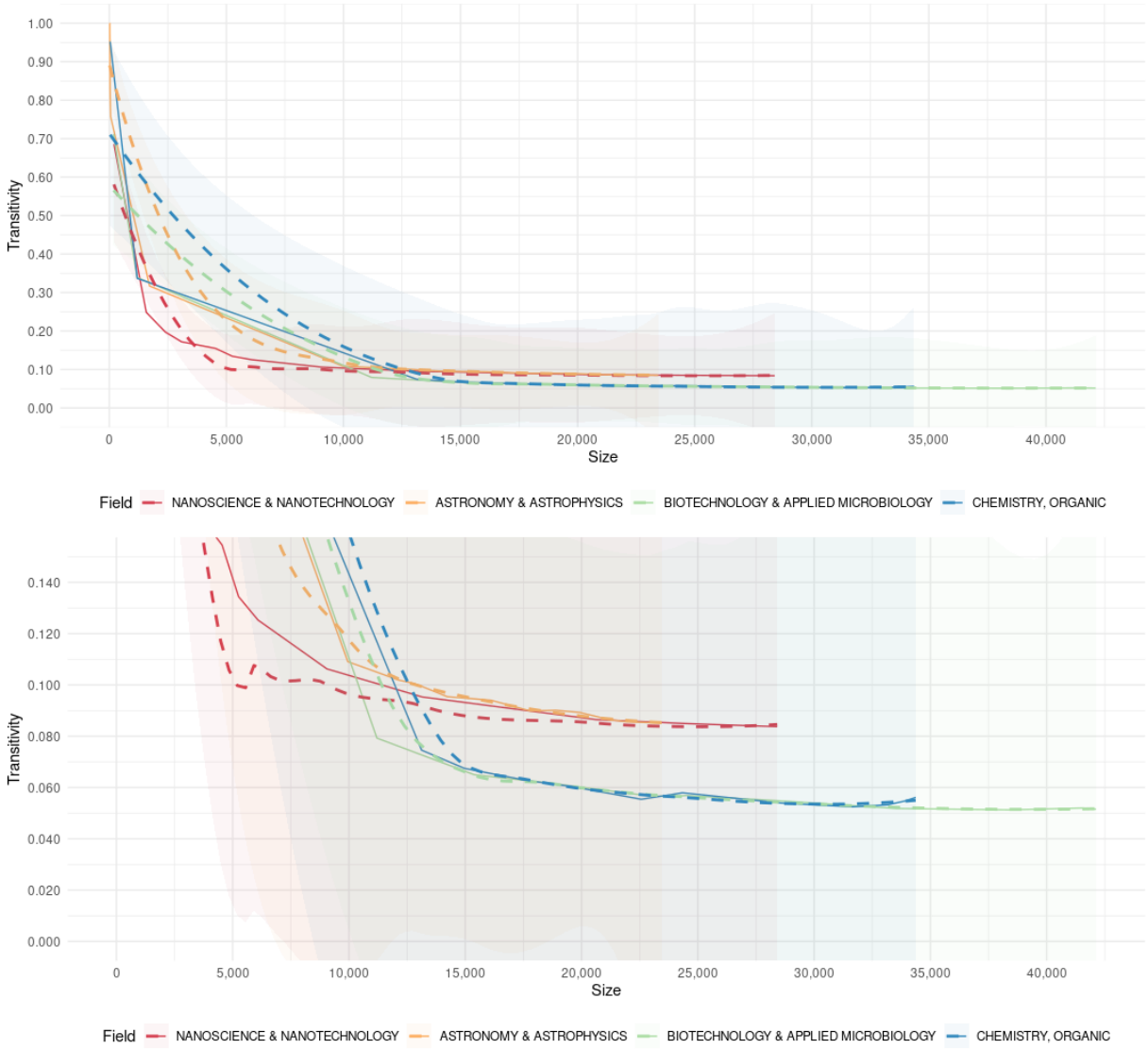

Figure 29: Transitivity over number of keywords. Top: total, bottom: zoomed in for transitivity < 0.15

Network density

The patterns were similar to those found for transitivity, both over time and network growth. BIOT and CHEM both had very similar densities after the network grew beyond 15,000 keywords. The density for NANO was slightly lower than ASTRO over network size. This is shown in Figure 30 and Figure 31.

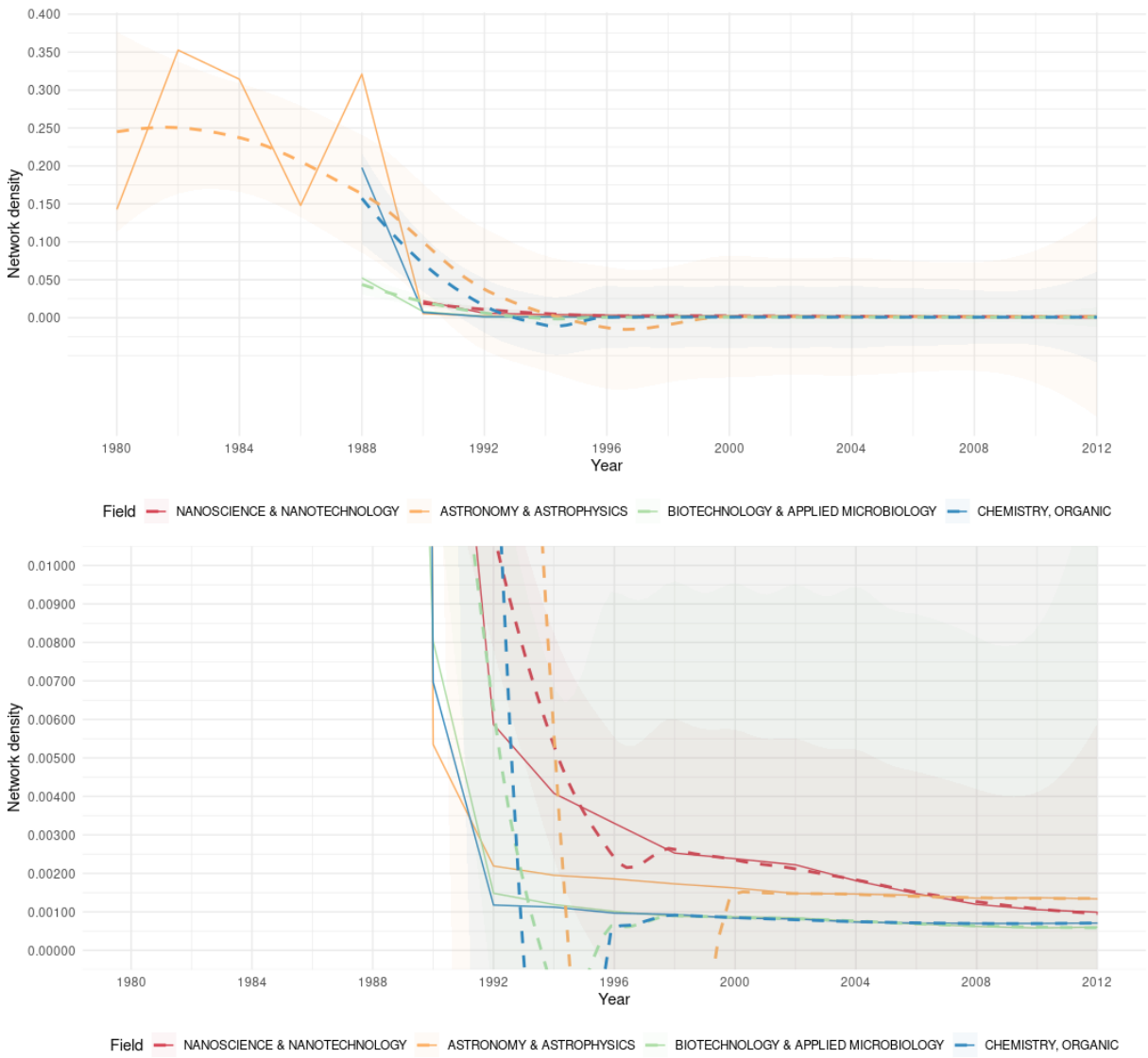

Figure 30: Network density over the years. Top: total, bottom: zoomed in for network density < 0.01

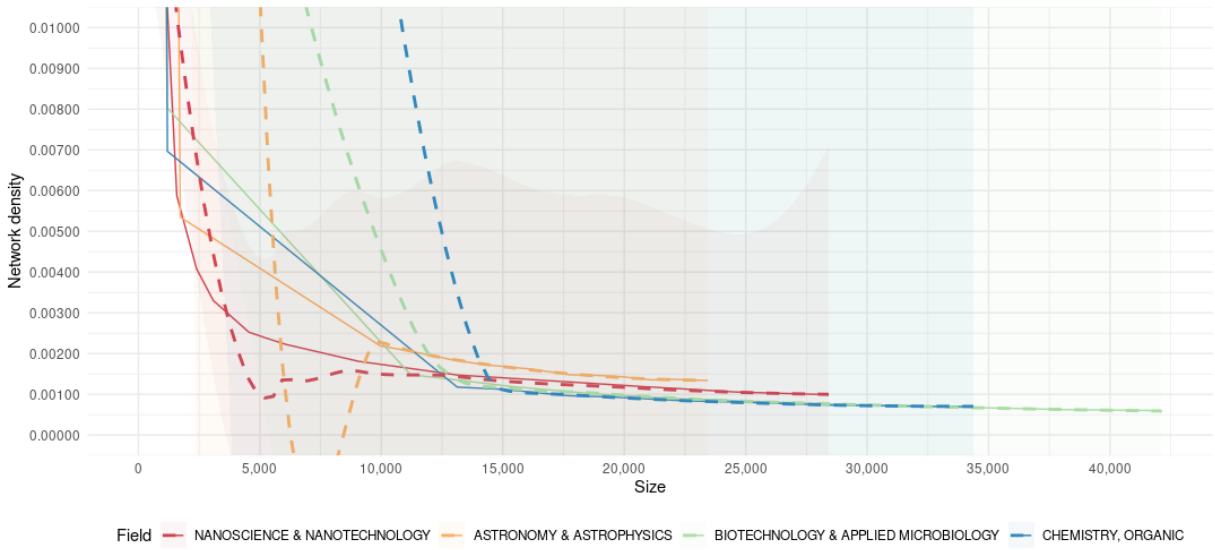

Figure 31: Network density over number of unique keywords, zoomed in for network density < 0.01

Pattern comparison

Inequality and annual publications

The Gini coefficients for cited first-author scientists and cited articles in the year of observation both seemed to positively correlate with the number of publications for each field. The coefficients regarding cited scientists over annual publication output are shown in Figure 32. All fields except ASTRO clearly showed a sublinear increase in inequality over productivity. In the years with an output of over 15,000 articles, the increase in inequality within ASTRO seemed higher than expected when following a sublinear trend similar to CHEM. This extra increase after 15,000 articles is also visible for BIOT. The emergence of inequality is most visible in the fields NANO and BIOT, especially before the annual productivity surpassed 7,500 articles. Additionally, it can be seen that when the annual productivity was similar to other years, e.g. in fields such as ASTRO and CHEM, the corresponding Gini coefficient remained similar as well.

The trends described above regarding the citation inequality among first-author scientists seem to apply to cited articles in similar fashion. This is visualised in Figure 34. However, the additional increase in inequality within ASTRO was steeper for cited articles and seems to have initiated at around 12,500 yearly articles. Inequality in cited articles also showed a steeper increase for CHEM after the annual productivity surpassed 15,000 articles. Most notably, the citation inequality among cited articles over annual productivity was similar

between NANO and CHEM, the latter was slightly less unequal. Considering the citation inequality among cited scientists, CHEM was significantly more unequal than NANO - especially when the annual productivity was lower.

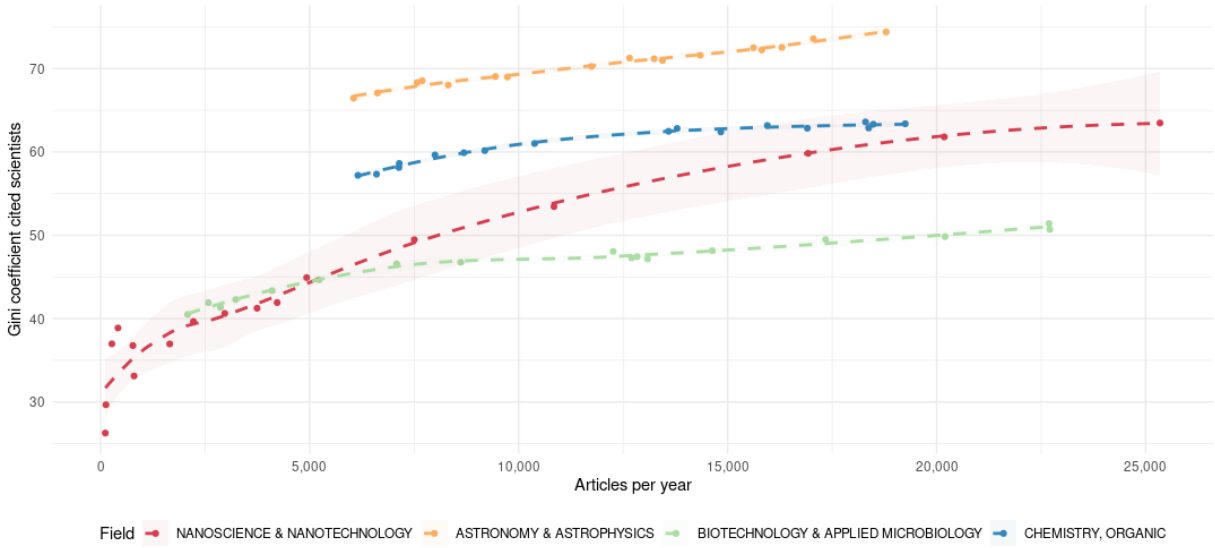

Figure 32: Citation inequality among cited scientists over annual publications

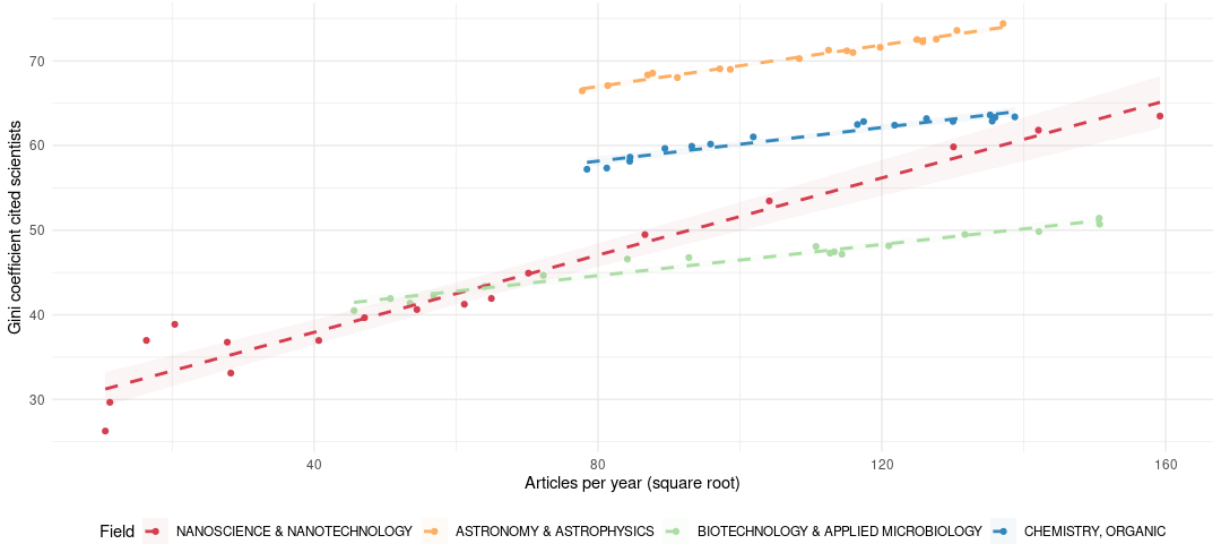

Figure 33: Citation inequality among cited scientists over the square root of annual publications

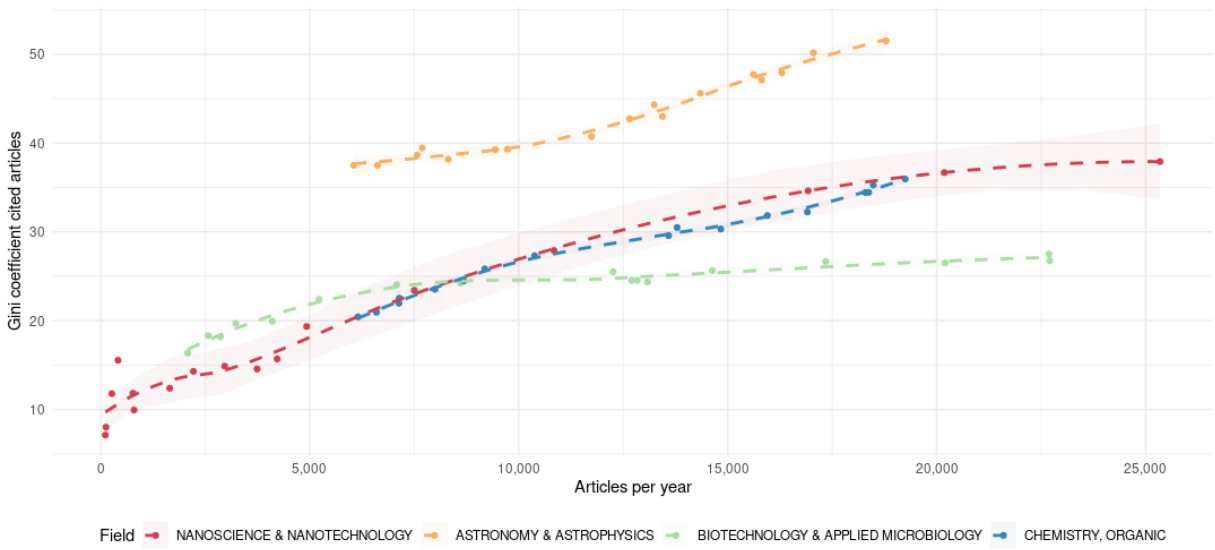

Figure 34: Citation inequality among cited articles over annual publications

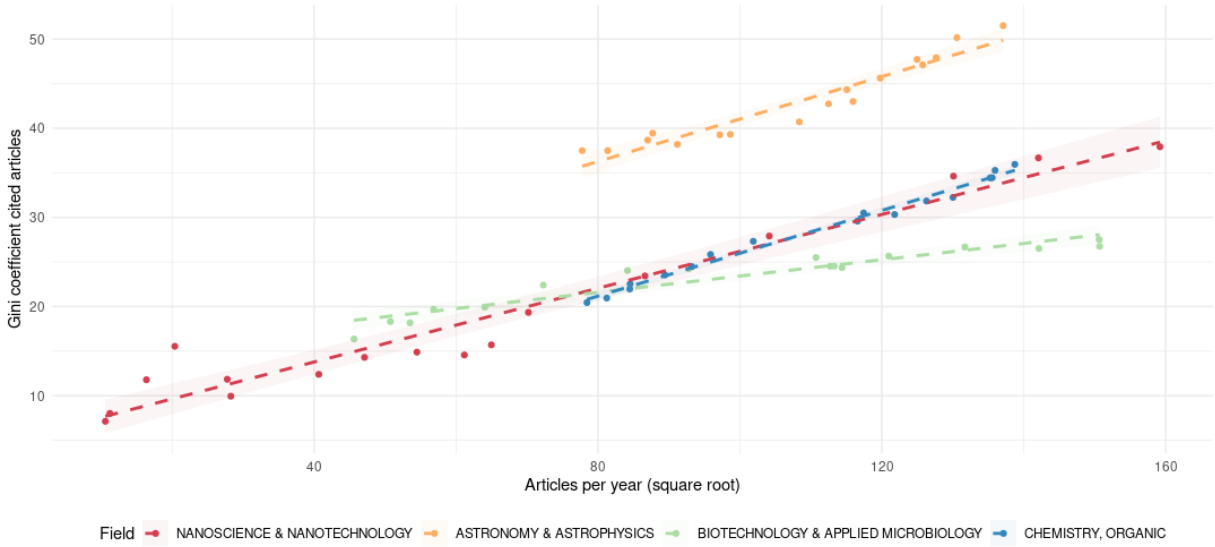

Figure 35: Citation inequality among cited articles over the square root of annual publications

We saw a sublinear increase in productivity inequality among scientists over growth in annual publications for the fields NANO and CHEM, as visualised in Figure 37. ASTRO showed most fluctuations in the Gini coefficients over growth, but the coefficients also seemed to increase most rapidly. ASTRO and BIOT showed an additional increase in inequality as the annual number of publications grew, whereas CHEM and NANO seemed to stabilise.

All fields showed a steep increase in keyword publication inequality up to approximately 12,500 annual publications. This is shown in Figure 39. ASTRO and NANO seemed to stabilise around the same Gini coefficients per yearly publications, as did BIOT and CHEM with slightly less keyword publication inequality.

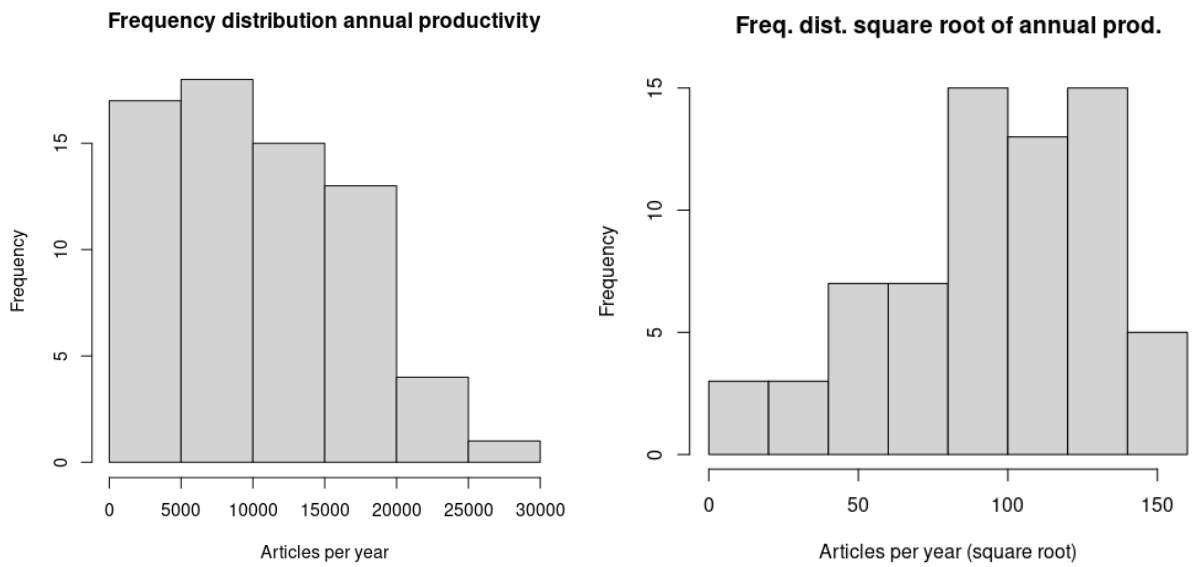

Figure 36: Frequency distribution annual productivity (left: normal, right: square root)

To test for correlations using Pearson's *r*, we first had to assume linearity. We found that the frequency distribution of the annual output was skewed, and that calculating the square root here would improve the normality of the frequency distribution. This is shown in Figure 36. Additionally, the plots in Figure 33, Figure 35 and Figure 38 show enhanced linearity. The correlation statistics from Pearson's *r* test are embedded in Table 4. As all plots showed positive correlations, we used the alternative hypothesis of the estimated measure of association being $0 < r \leq 1$. All correlations for all fields were significant and showed great correlation with $r > 0.9$ for every instance.

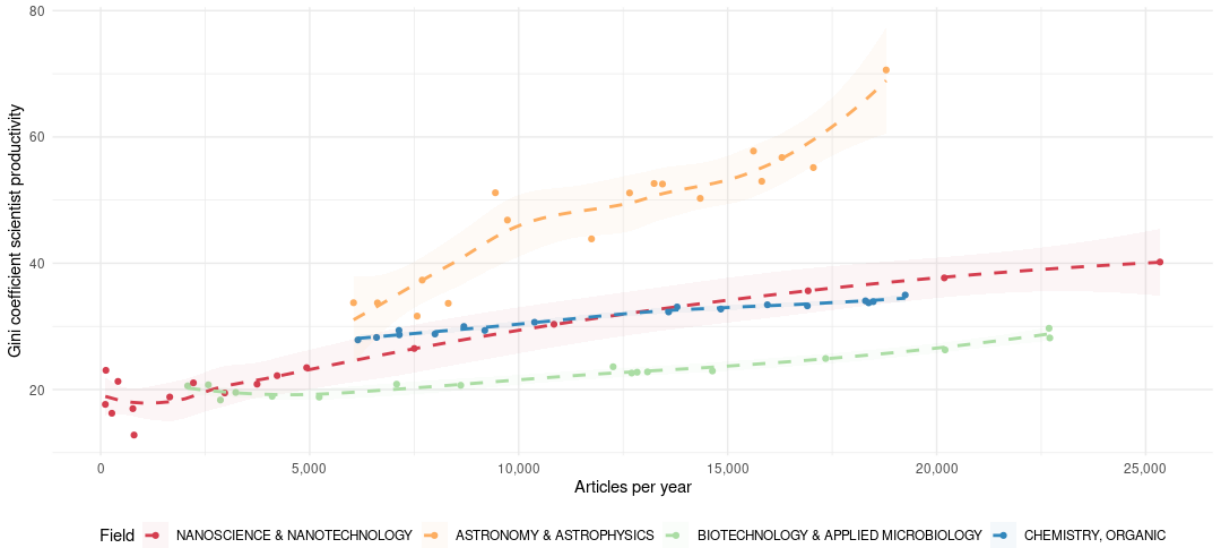

Figure 37: Productivity inequality among scientists over annual publications

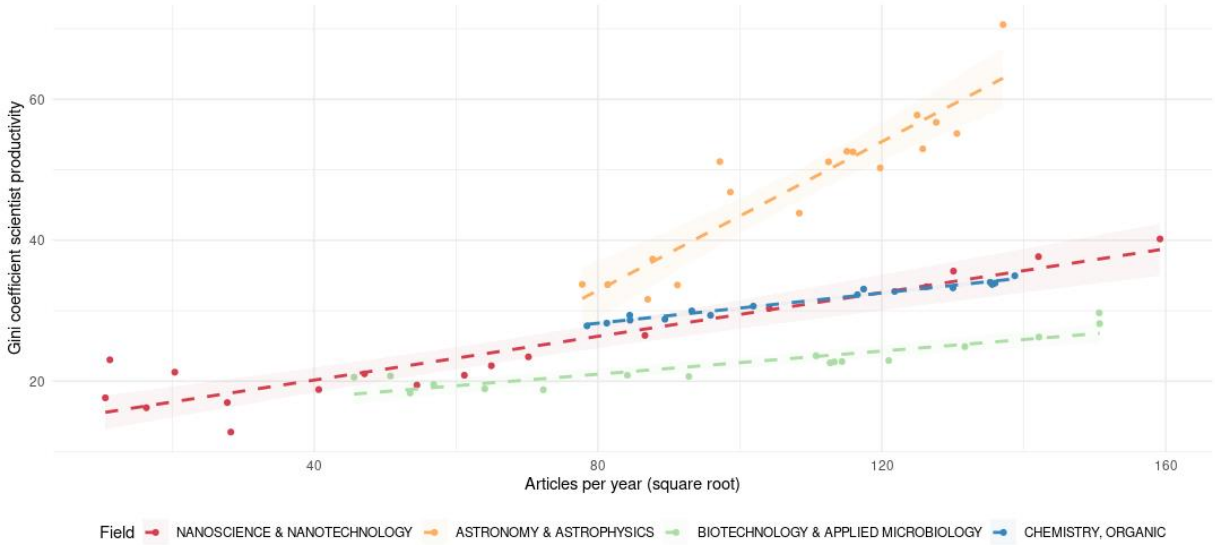

Figure 38: Productivity inequality among scientists over the square root of annual publications

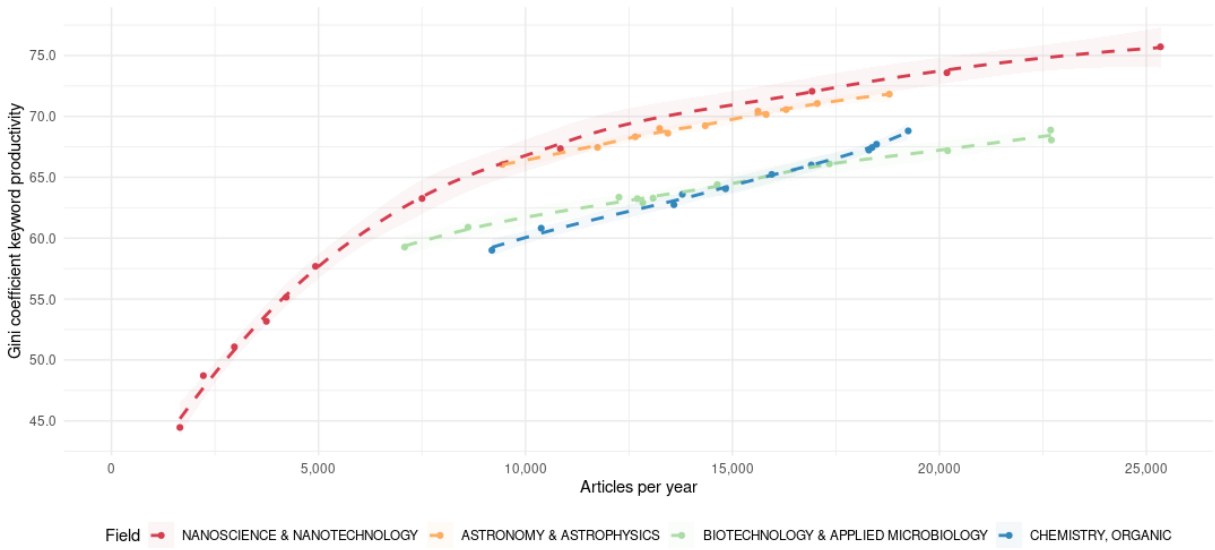

Figure 39: Publication inequality among keywords over annual publications

Table 4: Pearson's correlation coefficients for square root of annual publications

Field	**Pearson's *r***	**95% conf. int.**	**p-value**
		Correlation cited scientists:	
NANO	0.9743396	0.9392817 - 1.0000000	= 1.999e-11***
ASTRO	0.9893737	0.9745914 - 1.0000000	= 2.805e-14***
BIOT	0.9801537	0.9528495 - 1.0000000	= 2.959e-12***
CHEM	0.9645722	0.9167326 - 1.0000000	= 2.183e-10***
		Correlation cited articles:	
NANO	0.9728699	0.9358693 - 1.0000000	= 3.022e-11***
ASTRO	0.9597886	0.9057993 - 1.0000000	= 5.565e-10***
BIOT	0.949883	0.8833854 - 1.0000000	= 2.82e-09***
CHEM	0.9943311	0.986398 - 1.0000000	= 2.556e-16***
		Correlation productivity scientists:	
NANO	0.9265142	0.8316864 - 1.0000000	= 4.645e-08***
ASTRO	0.9189279	0.8152496 - 1.0000000	= 9.489e-08***
BIOT	0.9026189	0.7804699 - 1.0000000	= 3.574e-07***
CHEM	0.9861411	0.9669368 - 1.0000000	= 2.037e-13***
		Correlation publication keywords:	
NANO	0.9790186	0.9343782 - 1.0000000	= 7.972e-08***
ASTRO	0.9940245	0.9810045 - 1.0000000	= 2.86e-10***
BIOT	0.9948487	0.9836097 - 1.0000000	= 1.468e-10***
CHEM	0.9916036	0.9733791 - 1.0000000	= 1.317e-09***

Inequality and citations

We found no clear correlation patterns between citation inequality among scientists and articles, with all fields showing much fluctuation except CHEM. Most data points seemed to cluster by scientific field, with CHEM considerably least unequal in both citation distributions,

followed by BIOT. NANO and ASTRO seemed evenly unequal for most years, with ASTRO seeming slightly more unequal in scientist citations and NANO more unequal in article citations. This is shown in Figure 40.

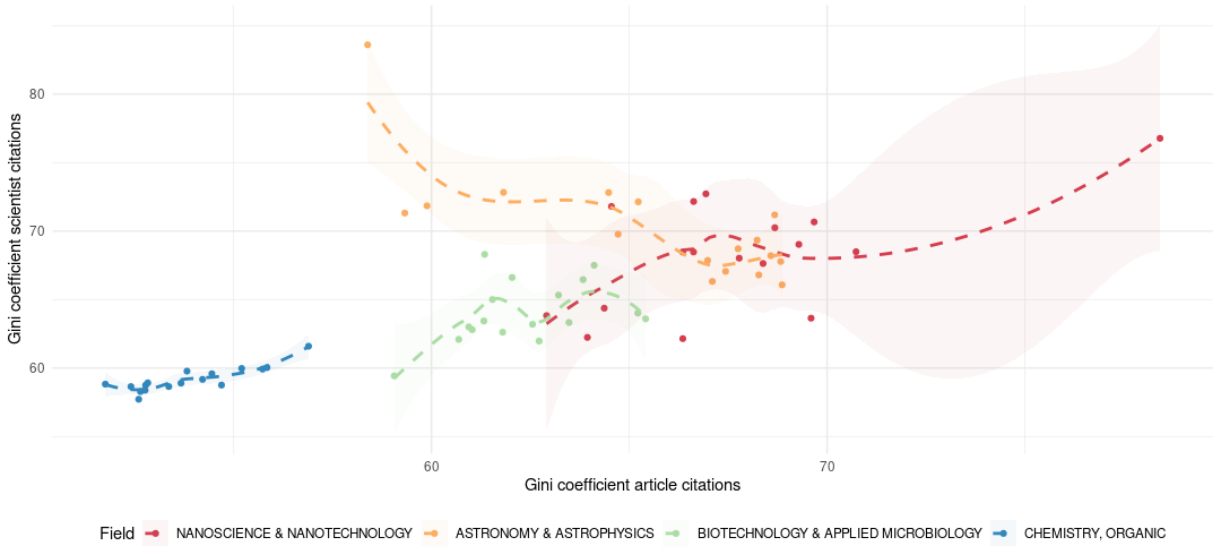

Figure 40: Citation inequality among scientists over citation inequality among articles

Both NANO and ASTRO showed sublinear correlation for the citation inequality among cited scientists over the inequality among cited articles, where both inequalities were at their relative lowest, which is visualised in Figure 41. NANO then showed a linear correlation pattern, whereas ASTRO seemed slightly superlinear. CHEM showed sublinear correlation pattern and BIOT superlinear.

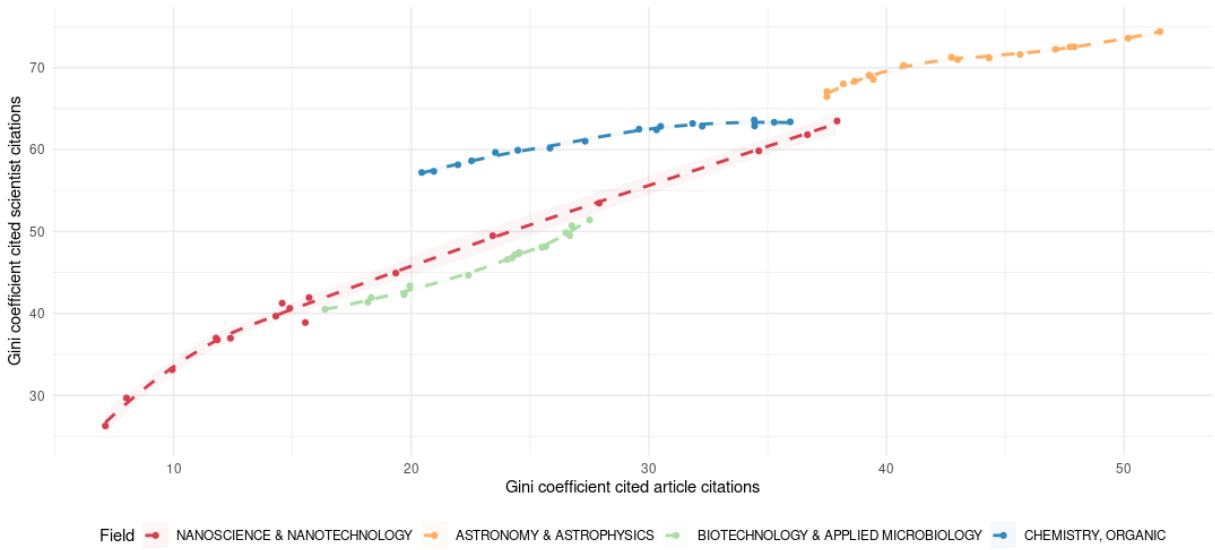

Figure 41: Citation inequality among cited scientists over citation inequality among cited articles

As shown in Figure 42 (and Figure 32), we found similar patterns in citation inequality among cited scientists over annual publications and annual references. Inequality growth seemed to stabilise in CHEM after approximately 200,000 references per year, whereas for NANO this is expected after 500,000 references. The increase in inequality over yearly references showed slightly sublinear correlation for ASTRO. BIOT showed a sublinear increase up to 200,000 yearly references, after which the inequality per yearly references seemed to increase more linearly.

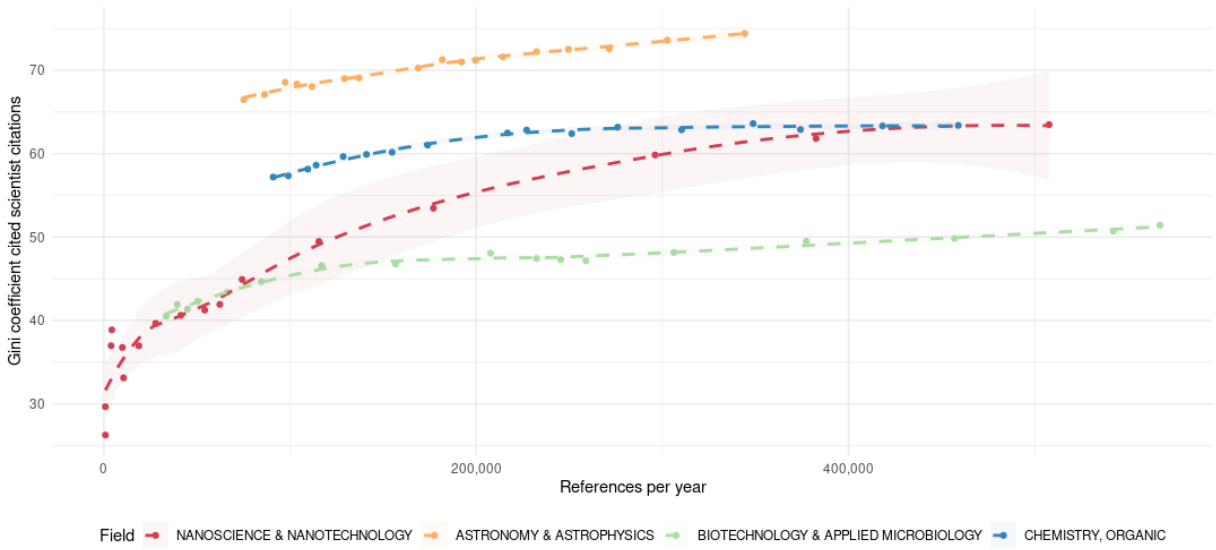

Figure 42: Citation inequality among cited scientists over yearly number of references

Inequality and scientist productivity

We fitted a linear correlation trend line in our results, as shown in Figure 43. NANO showed the steepest increase in citation inequality among cited scientists as the inequality in productivity among scientists increased. The data fluctuated more around the trend line for NANO than the other three fields. ASTRO saw the biggest increase of productivity inequality while the increase of inequality among cited scientists was one of the smallest.

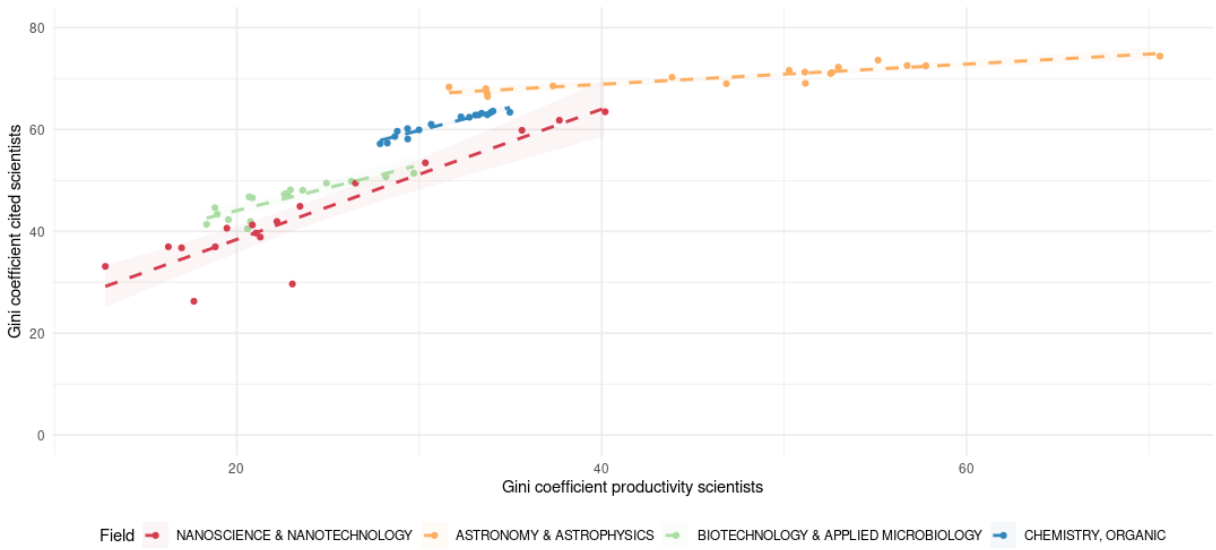

Figure 43: Citation inequality among cited scientists over productivity inequality among scientists

We could not find clear correlation trends between the share in annual publications of the top twenty most productive authors and the productivity inequality among scientists. NANO and BIOT showed great fluctuation in the share of the top twenty productive scientists, especially on the lower end of productivity inequality, this is shown in Figure 44. CHEM and ASTRO showed less fluctuation. The two fields showed a decline in the publication share of the top twenty scientists before it seemingly suddenly increased as the Gini coefficient for productivity increased. Both fields showed a U-shaped pattern, albeit considerably skewed.

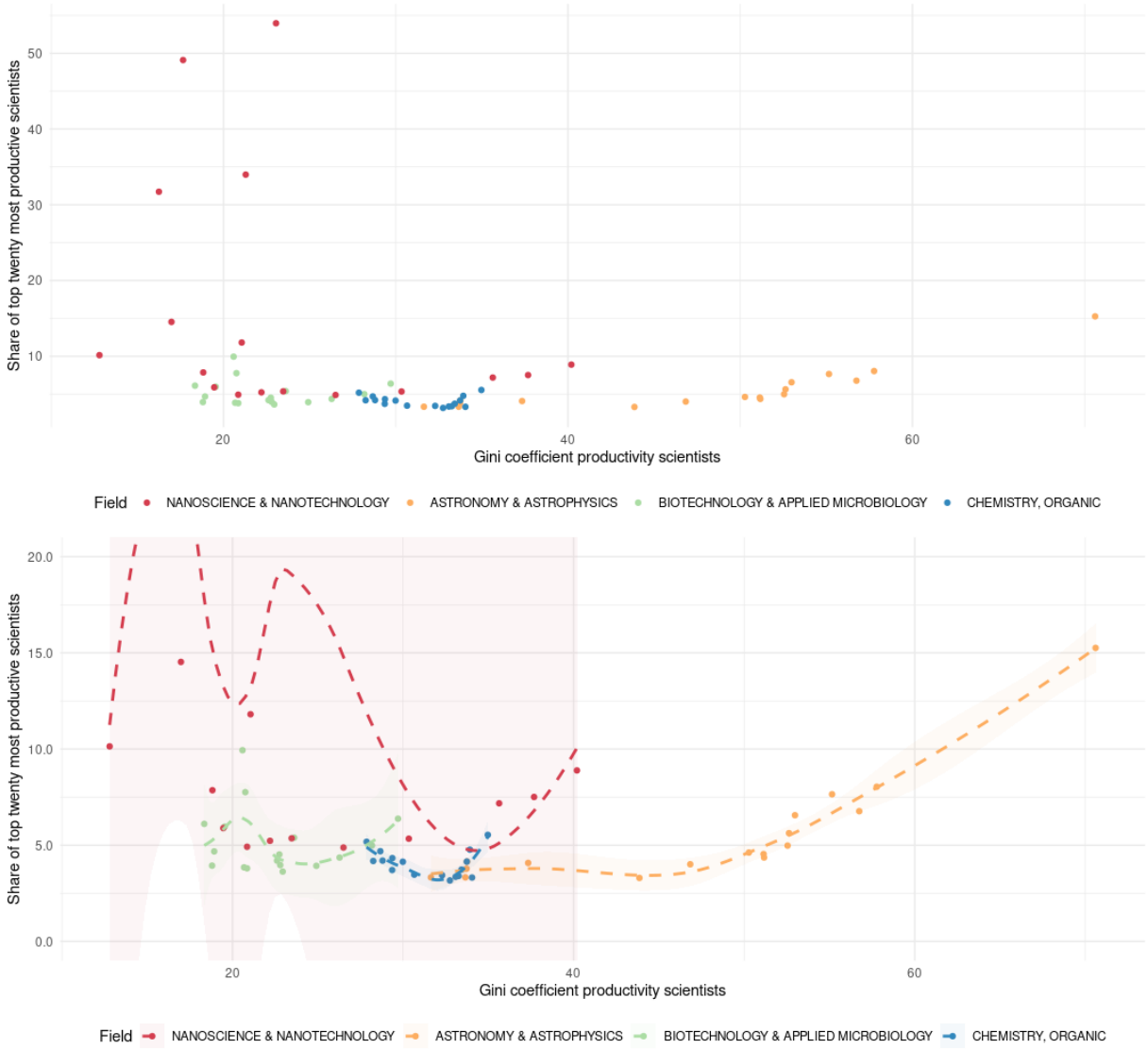

Figure 44: Share in publications of the top twenty most productive scientists over the productivity inequality among scientists. Top: total, bottom: zoomed in for percentage share < 20.0, including trend line

We could not find clear correlation trends between the share in citations of the top twenty most referenced articles and the productivity inequality among scientists. There are no clear indications of a correlation and the data points fluctuate significantly around the attempted trend line, as can be seen in Figure 45.

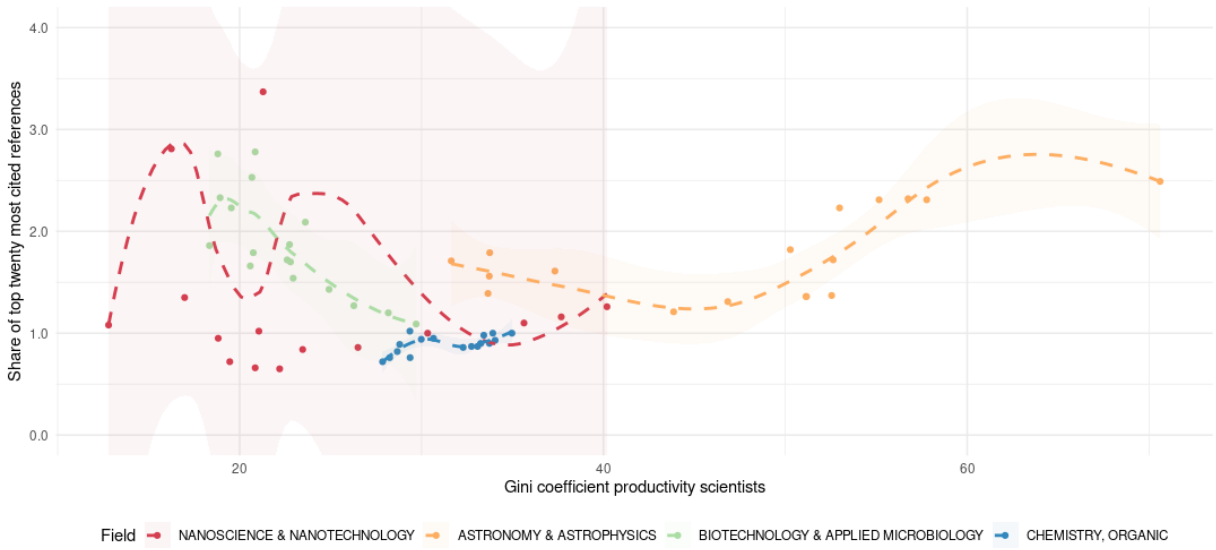

Figure 45: Share in references of the top twenty most cited articles over the productivity inequality among scientists

For all fields, especially BIOT and CHEM, the first increase of the Gini coefficients for inequality among scientists' productivity saw a relatively large increase in inequality among the number of publications per keyword. Additionally, almost each field seemed to have a different pattern. NANO and BIOT both showed a sublinear increase in keyword publication inequality, CHEM showed an S-shaped correlation and ASTRO seemed more linearly correlated albeit with considerable fluctuation. This is visualised in Figure 46.

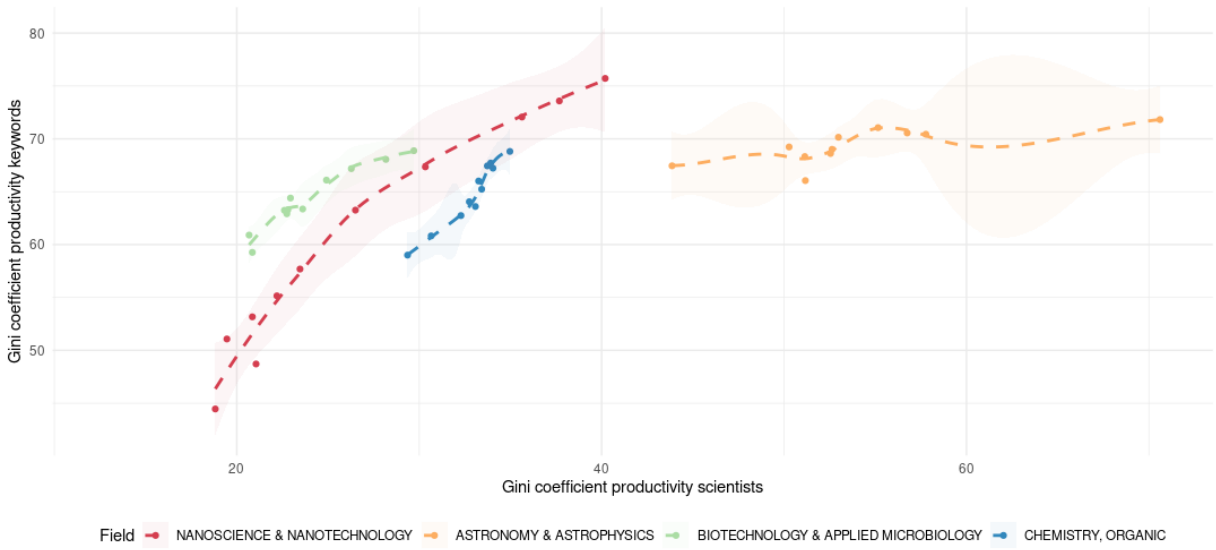

Figure 46: Publication inequality among keywords over productivity inequality among scientists.

For all data points with a percentage share of the top twenty most productive scientists being less than 10%, approximately, we could not find clear correlation patterns. The data fluctuates too much to be able to distinguish correlational trends. NANO did show a superlinear correlation as the share of the top twenty most productive scientists increased. This is visualised in Figure 47.

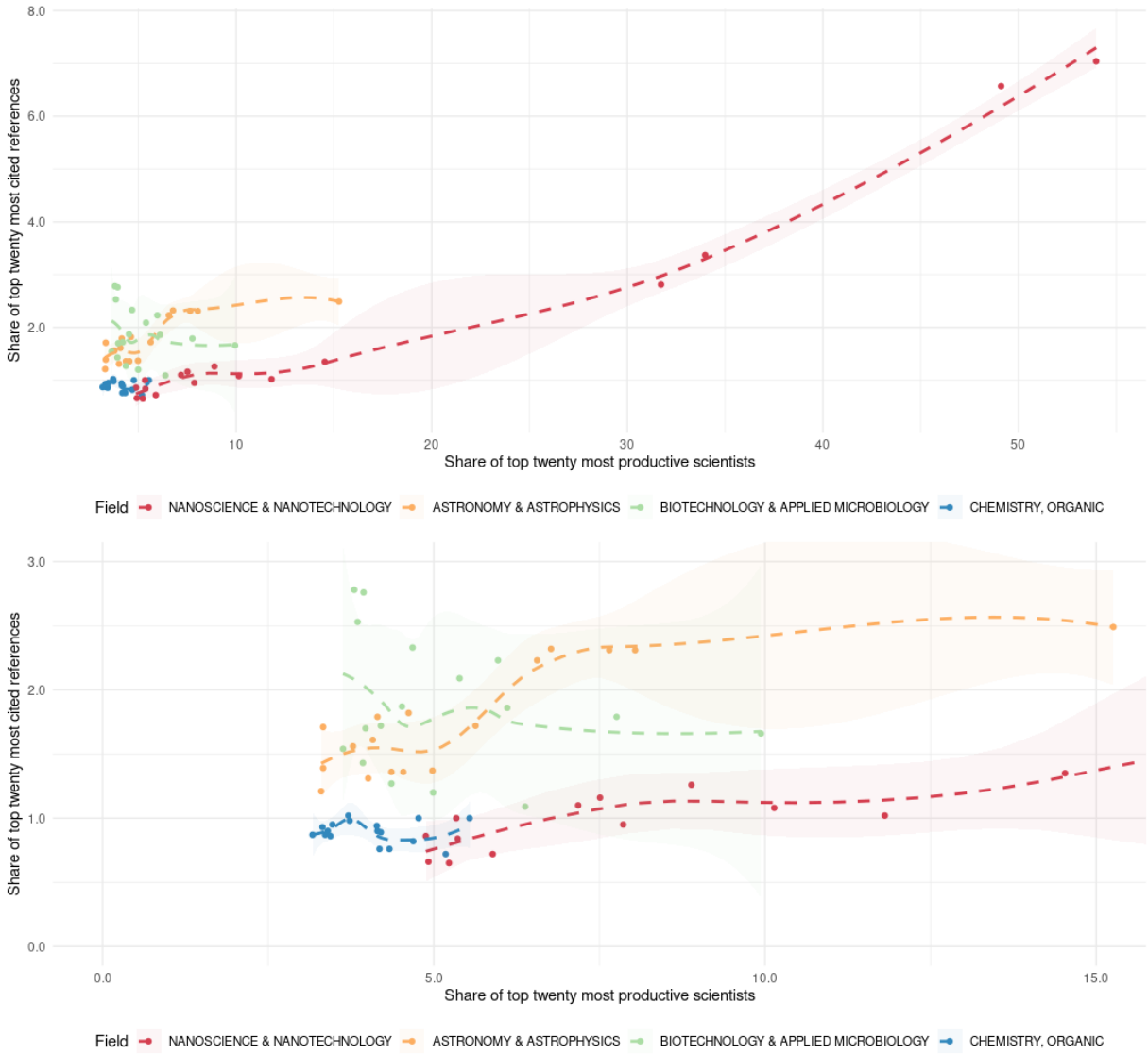

Figure 47: Share in references of top twenty most cited articles over share in publications of top twenty most productive scientists. Top: total, bottom: zoomed in for share in references < 3 and share in publications < 15

Conclusion

We examined the distribution inequalities of recognition and productivity among scientists, articles and keywords over the years. Additionally, we compared these inequalities with annual productivity as well as with each other. We explored differences between these variables for the fields Nanoscience & Nanotechnology (NANO), Astronomy & Astrophysics (ASTRO), Biotechnology & Microbiology (BIOT), and Organic Chemistry (CHEM). In this section, we compare our results with existing literature in order to answer our research question:

How do social and cognitive patterns in science relate?

We first answer our sub questions based on the results of our data analyses. Thereafter, we combine our theoretical understanding of the research context with the partial conclusions derived from answering the sub questions.

***SQ1**: What are the dynamics of the social structure within a scientific field over time?* We expected scientific fields to become more unequal over time. This was with the underlying assumption that as time passes, these fields grow in terms of knowledge production and scientist involvement. We found that for all four examined fields this was the case. However, the more established fields ASTRO and CHEM grew less expeditiously than the emerging fields BIOT and NANO. The annual production of ASTRO and CHEM has been stable for the last 4 - 6 years, especially considering the growth in yearly publications for NANO.

We measured social inequality through citation and productivity distributions among scientists based on the publications and references each year, as well as the statistics of the collaboration network. We found that the appearance of cited first-author scientists became more unequal over time. Citation inequality among cited scientists had a strong correlation with annual publication output. This implies that as a scientific field grows over time, it becomes more reliant upon the same scientists. A strong correlation also was found for publication inequality. Additionally, the inequality among cited first-author scientists and scientists' productivity correlated linearly. This indicates that when more articles are published by less scientists (productivity inequality), the number of references to a smaller group of scientists (citation inequality) increases, and vice versa.

In terms of scientist collaboration, we found that the overall network density and clustering declined as the network grew. This either indicates a decline in existing

collaborations and dissolving communities, or that the addition of new scientists and collaborations forms more rapidly than the addition of collaborations among established scientists. The average path length increased up to a certain number of involved scientists and then more slowly declined. The degree centrality increased for each field after a rapid dip around 10,000 scientists. These two variables indicate more connectivity within the network, which either implies either more scattered collaborations throughout the network or the formation of highly interconnected individuals. We saw the collaboration network density and clustering decline, which contradicts the former possibility. Therefore, we can conclude that we found indications of collaboration inequality through the formation of scientists acting as hubs. Additionally, we found that for all fields except ASTRO the involvement of new scientists through collaborations (network growth) increased more than (additional) collaborations among established scientists (community and hub-forming).

Overall, we can conclude that the social structure of a scientific field becomes more unequal over time, but we found evidence that social inequality has stronger correlation patterns with growth.

SQ2*: What are the dynamics of the cognitive structure within a scientific field over time?* The citation inequality among keywords increased for all fields, which indicates scientists' favouritism to cite articles that contain certain keywords or topics. However, we did not normalise the keyword citation data for the number of times it appeared each year. Publication inequality among keywords increased over the years as well, but in general the inequality was lower and increased less rapidly compared to keyword citation inequality. Scientists thus favour certain keywords and topics in both occasions for publishing their own research as well as building upon existing findings.

Citation inequality among cited articles increased over the years, but less so as the number of annually published articles enlarged. Moreover, the inequality among cited articles was considerably lower compared to the inequality among keywords or cited scientists. The article itself appeared to have less of a retention (citation) value than its topic(s) or author(s). We found that citation inequality among published articles declined over the years. As these citations accumulate from the year of publication until the date of our data collection, we could say that citation inequality among articles increased as their publication date was further back in the past. This only holds if the number of citations an article has received increases its probability for more citations, especially compared to other articles published in that year. When we consider that citation inequality increases the older a set of articles become, and each year's publications encompass an increasing inequality among cited articles, this could

be indications for path dependency. A field leans more towards the same (older) articles as its output increases.

Additionally, the keywords as indicators for research topics become more unequal in the distributions of both total citations and publications. We found similar network characteristics over time and growth for the co-occurrence of keywords in publications as we did for the collaboration network. However, the keyword co-occurrence network contained less clustering and was in general more dense and interconnected. The average path length quickly stabilised and was considerably lower for keyword co-occurrences. With a stable average path length, network density and clustering coefficient over growth, we conclude that we did not find indications of extensive hub-forming. Rather, the dynamics of the variables seem to indicate that the network consists of many 'sub-hubs' with minor communities. As the network grows, co-occurrences among existing keywords seem to increase with similar pace as co-occurrences with new keywords.

The cognitive structure of a scientific field seems to be built around the origins of its founding articles and reinforces the direction of its past research. As the annual knowledge production increases, the inequality of retention and relatedness among knowledge does so as well.

***SQ3**: How do scientist inequality and knowledge inequality relate?*
We found curvilinear correlations between productivity inequality from the social structure and research direction inequality from the cognitive structure. Furthermore, the collaboration social network structure showed more signs of inequality than the keyword co-occurrence network, which implies that research directions tend to exceed scientists' collaboration communities. Both social and cognitive structures inherited more inequality over time and growth. We found indications of the Matthew effect and exploitative research dynamics, but we cannot definitively conclude how these patterns directly relate. However, both social and cognitive inequality positively correlate with annual productivity.

***SQ4**: How does the organisational structure of a scientific field affect its social and cognitive patterns?*
Comparing inequalities between the scientific fields over the years, we found that the more mature and stable fields with low task uncertainty, ASTRO and CHEM, with most comparisons were or at least started as the two most unequal fields. These two fields also showed similarity in growth of annual publications and scientists' involvement. In both aspects, the fields had the largest annual productivity and involvement in the first decade. ASTRO and CHEM

remained the largest fields in terms of accumulated publications, but the emerging fields BIOT and NANO both overtook them in 2010 concerning annual publications and involvement. While ASTRO and CHEM showed considerable similarity in growth and accumulated size, the emergence of BIOT and NANO did not show many overlapping dynamics. The initial size and growth of BIOT in the first sixteen years was significantly larger than NANO, the latter saw its biggest increase in the last decade.

ASTRO, the field with low task uncertainty and high mutual dependence, showed high and stable inequalities in both scientist recognition and productivity as well as knowledge retention and relatedness. Together with CHEM, low in both task uncertainty and mutual dependence, the two fields had the highest inequality among their references to scientists as well as the publications per scientist over the years. However, considering productivity inequality over annual publications, NANO and CHEM were nearly identical - this was also found for inequality among cited articles. Other than that, we most often found similar patterns or inequalities between the pairs based on mutual dependence (ASTRO and NANO, high; CHEM and BIOT, low). Task uncertainty seemed the least affecting organisational characteristic, where we cautiously conclude that a low task uncertainty goes paired with highly accumulated fields, but stagnating annual productivity and great social inequality. Mutual dependency appeared to have led to the pattern similarities between two fields most often where social and cognitive inequalities saw the largest increases over growth. All variables considered, however, we found that the four fields individually showed their own characteristics. The scientific fields all presented distinctive trends for each variable. Ultimately, we must conclude that there is no definitive answer to this sub question based on our findings.

We now proceed to answer our research question based on the answers above in light of forgoing literature used in the theory section. We found relations between social and cognitive patterns as both structures became more unequal over the increasing size of annual knowledge output. While we discovered correlations between social and cognitive inequalities, the most fruitful results were the distinctive nature of the different organisational structures found between the fields over their annual productivity. Considering a scientific field as a system incorporating both competition among its actors and a collective desire for output optimisation, we found that inequality has a sublinear relation with its system's productivity and size. The intensity of inequality increase is segregated by the organisational structures, which act as leveragers for the amount of selection pressures. Generally, the fields classified with lower mutual dependence showed less cognitive inequalities whereas a lower task uncertainty seemed to impede less social inequality. However, the differences were not always

substantially apparent. Nevertheless, we can conclude that the social and cognitive patterns in science relate based on the organisational structure of the field as both social and cognitive inequality have a positive sublinear correlation with the field's productivity.

Discussion

Practical insights

As our research was of exploratory nature and more theoretically focused, direct practical implications seem limited. We also explicitly examined correlations rather than causations, which restricts us in providing guidance with changing dynamics. However, we did offer a view of the inequality within four scientific fields. The Gini coefficients indicating citation and productivity inequality are high, especially compared to Gini coefficients indicating countries' income inequality (Luebker, 2010). Perhaps the most useful practical contribution is the realisation that inequality is a reinforcing mechanism that seems to increase when the focus lies on productivity and rewards. If the focal point of scientific research, to optimise limited resources and produce impactful knowledge, comes with increasing social and cognitive inequality then we should critically ask ourselves whether this is indeed the route we desire to maintain. It is a converging route that sees a selective group of scientists become more cited and productive, as the cognitive output becomes more centered around the same topics. We understand that these dynamics are jointly responsible for the rapid emergence of completely new fields and possibilities such as nanotechnology - but we have not seen mechanisms that prevent science from being carried away.

If variation is the fuel of a healthy system (Fagerberg, 2003), inequality might be the side effect of an engine that runs too fast. If the same scientists and the same research topics receive more citations and publications we may increase similarities and reduce odds for scientific breakthroughs (Siler et al., 2018). Ultimately, it is a highly complex matter, but if we can interpret our theoretical understanding for the increase in inequality we will point to two aspects: rewards and selection. Scientists desire citations and publications to remain relevant and will receive such rewards by the selection mechanism. To maintain or increase equality, the selection mechanism should be more lenient towards dissimilarity and not only reward what has been proven successful. This would imply, e.g., that scientific fields should engage more in funding explorative research and lesser-known scientists. It may not be the quickest route to success, but we believe it is the most sustainable. And to the scientists, if it can be afforded, reserve some resources to explore a broader spectrum of research possibilities every once in a while. After all, we should not forget the importance of taking time to refuel.

Theoretical considerations

In our introduction, we explained our concern for converging and more unequally becoming patterns in the sciences. We formed a novel bridge between models of evolutionary economics and the Matthew effect in our theory section. While the latter has been paralleled with other social and economic phenomena aggregated as cumulative advantage mechanisms (DiPrete & Eirich, 2006), our thesis attempts to find a fundamental explanation for the occurrence of such cumulation based on systemic processes. We modelled the behaviour of individual scientists, their relation to the system and how the micro and macro levels affect one another through co-evolution. We introduced the mechanisms of variation, selection and retention in an economic system (Dosi & Nelson, 1994; Nelson & Winter, 1982; Schumpeter, 1943), and adapted it to fit the academic setting. In doing so, we discovered a potential explanation for a missing piece in the evolutionary economics literature which, according to Fagerberg (2003, p. 147) is "an issue on which many of the evolutionary theorists have relatively little to say".

The issue concerns the source of variation that refuels the system as system growth leads to optimisation through natural selection processes that improve the average fitness of the actors, which ultimately results in extreme similarities as each actor has adopted the best practices available (Andersen, 2001; Dosi & Nelson, 1994; Nelson & Winter, 1982). Some potential solutions are presented (Fagerberg, 2003), such as exogenous inventions that leak variety back into a system, and a change in search strategies by actors to become more explorative. However, our results suggest that variety does not leave nor deplete - quite the opposite. We have seen the scientific system grow in annual productivity, which also led to an increase in the number of individual actors as well as unique keywords. We propose the solution that variation grows together with the system. Additionally, we argue that the optimisation of a system's growth in productivity is not through each actor adopting identical routines and thereby performing equally well. Our proposition to explain the productivity growth of a system is increasing hierarchy and efficiency through system inequality. Successful actors on top of the pyramid have accrued nearly system-level power and can therefore easily steer the direction of search. While actors may eventually perform equally well, their functions and role within the system may differ, similar to food- or supply-chains. Selection pressures may come from the level of hierarchical inequality, the intensity of competition for each level and the amount of available resources to share.

While such a system may seem more cooperative than competitive, the hierarchical stability does not come from lenient actors and a utopian ideology. A system is stable so long

as the productivity still grows. The top-level actors are responsible for the growth of the system. As such, the pressure for optimisation and short-term results leads to risk-aversive and satisficing behaviour. Those who made it to the top exploit the successes that have rewarded them their place. This would explain the highly unequal cognitive structure of the scientific fields. Consider the total citation inequalities for keywords and scientists (see FIG and FIG). The latter fluctuated more over the years and showed more decreasing or stagnating patterns than the former. This is also apparent when comparing the productivity inequality between the two, albeit a more stable increase is present for both (see FIG and FIG). Moreover, we found an interesting pattern in the field Biotechnology & Microbiology. For both the annual number of publications as well as the number of involved scientists, between 1980 and 1996 a superlinear - almost exponential - increase is visible. However, after 1996 the growth suddenly stagnated until around 2002. If we look at the social inequality of the field, we see a consistent drop in the Gini coefficients for citations per cited scientists and their productivity. This is also the case for the cognitive inequality, where the Gini coefficients for citations and productivity per keyword slightly, but consistently, decreased. This may have indicated a destabilisation in the hierarchical top as response to the sudden stagnation in productivity.

Science can be considered a self-organising system (Kauffman, 1993; T. S. Kuhn, 1970; Whitley, 2000), although societal, economical and political topics may steer the general course of the universal scientific system (Heimeriks & Leydesdorff, 2012), with regional differences of focus and specialization (Heimeriks & Balland, 2016). Similar to capitalistic markets, successful efforts are rewarded with currencies that allow access to further resources and productivity. This may be the fundamental process that leads to inequality in influence and market share as, e.g., larger firms grow disproportionally quicker than smaller firms (Samuels, 1965; Santarelli et al., 2006). Large and powerful incumbents tend to hold small and potentially disruptive firms in check through political and economic power (Aldrich & Fiol, 1994; Penna & Geels, 2015; Smink et al., 2015). Literature suggests that such counterplay requires the targeted new firms to establish *legitimacy* for their product (Kukk et al., 2016; Weber & Rohracher, 2012). In science, similar procedures are discovered where high-status scientists effectively act as gatekeepers since their power enables them to "judge scientific work according to their preferred principles, in a sort of 'victor's history'" (Siler et al., 2018, p. 230). We acknowledge that such distribution of power contributes to the stability and productivity of the system, however this also implies that in systems thriving on self-organisation, inequality is imminent and deposits the steering wheel for research directions into the hands of the powerful few. This could develop strong search regimes, which are found

to remain incremental and converging on the long-term (Bonaccorsi, 2008; Breschi et al., 2000; Dosi et al., 1988; Heimeriks & Leydesdorff, 2012; T. S. Kuhn, 1970). Additionally, over time the system becomes more and more path dependent and therefore more resistant to change (Arthur, 1994).

Our biggest theoretical consideration is the effect of accruing power of successful agents and favouritism for cognitive similarities as the productivity and stability of a self-organising system increases. We have learned that both social and cognitive inequalities increase as a scientific field grows. Given the similarities to other self-organising and reward-based systems, there could at some point be an exogenous cry for change that will not be heard; e.g. climate change and the (too) slow responses from the industry to become more environmentally-friendly or develop sustainable alternatives (Penna & Geels, 2015; Smith et al., 2010). We should thus focus on methods to counteract inequality while maintaining stability and productivity. Perhaps the system is required to change at its organisational structure, by including a regulating body. However, regulations can still bottleneck innovation (Blind, 2012; Boon et al., 2010; Faulkner & Poort, 2017). We would suggest that mission-oriented regulations could be a better solution as this rewards agents for their attempts, rather than successes. This could decrease inequality, since it reduces the strength of the cumulative advantages of agents being able to build upon their successes and simultaneously receiving rewards for them. Instead, rewards are given to agents who pursue novel alternatives to achieve the same objectives, laid-out by the regulating body (picking the willing instead of the winners) (Mazzucato, 2018).

Limitations / issues

The Gini coefficient as measurement for inequality over time provided sufficient insights combined with our theoretical overview for a general exploration of cumulative inequalities within scientific fields. However, we relied upon the assumption that as a field became more unequal, the distribution remained path dependent. This assumption implies that the unit to distribute skews towards those who already received the larger share as part of cumulative advantage mechanisms. We did not control whether the distribution rankings remained similar over time. We acknowledge that this forms a weak spot in the fundamentals of our conclusion, but it can partly be parried by the consistency of our results that showed increased inequality for nearly all variables. The possibility of a scientific field becoming more unequal over time with similar increments while ultimately reorganising the annual distribution to different recipients is existent, but minuscule. Especially considering the vast amount of empirical

evidence for cumulative advantage mechanisms in science, we can be fairly certain that such inequality indicates skewed accumulation. However, the lack of such insights in our results limits the firmness of our conclusions. We could have included more descriptive statistics for the annual distributions that we analysed to provide better understanding.

We used raw data downloaded from the Web of Science (WoS) which we then converted into a single large database using the *bibliometrix*-package (Aria & Cuccurullo, 2017). Our data consisted mostly of textual data, where some variables such as authors, keywords and cited references were listed per row in a single variable column, separated by a semicolon. Due to the sheer size of our database and lack of resources from our side, we did not clean our data and trusted the WoS and the *bibliometrix*-package to have provided us with proper data and analysis functions. Generally, this seemed to be correct but we did notice that the cited references missed information such as first-authorship in some cases. This became apparent in our 'top twenty cited scientists' results. Some names were missing and substituted by publication year, or appeared twice; with and without dots separating the initials. This may have slightly lowered the Gini coefficients for first-author cited scientists, but unlikely changed its dynamics over time. Our top twenty metric is more likely to have been affected.

Concerning our analyses for the top twenty rankings, we should disclose three flaws that likely affected our results. The first one is unclean data as described above, which potentially led to a less stable ranking over the years. Second, our stability metric calculation does not take into account the individual changes in ranks over the years. Rather, we focused on the total number of individuals who appeared at least once in the top twenty between 1980 and 2012, and the average times of top twenty appearances. A more thorough stability ranking should for each individual calculate their changes in rank over time and provide an average for the whole field. Third, we calculated the percentage share of the top twenty compared to the entire field to compare the share of the top twenty over the years. However, for nearly every instance twenty was an insignificant sample compared to the total appearances. For example, the top twenty most productive scientists were compared to tens of thousand of their colleagues. While the top still held a significant percentage share, the chance of significant deviation in the total number of publications by the top over the years was considerable - especially taking into account the slight discrepancies in our data. While these computational errors might seem insignificant for the impression compared to the whole field, it could have affected the patterns for the percentage shares compared to each other. Comparing the percentages with each other could contain relatively large differences. This could be an explanation for the lack of noticeable (correlation) patterns.

Our data contained 11,068 articles without cited references and 169,815 articles without indexed keywords (*Keywords Plus*). Most articles without references or keywords were published prior to 1992. All but three articles without cited references also did not contain keywords. Our analyses on keywords (mainly for research direction) therefore were based on a smaller sample size. This mainly impacted results for the years prior to 1992, which were therefore excluded from further analyses.

Towards the later years, especially 2010 and 2012, we noticed some articles with an extraordinary number of authors. It also seemed that some fields were more keen on including a large number of authors. This could highly impact the collaboration network analyses per field over the years, especially considering the vast majority of articles (605,482) included five or less authors. There were 98 articles with more than one thousand authors and 1,957 with more than one hundred. As a countermeasure, we decided to set a maximum number of authors per paper and exclude authors from the collaboration network analyses after this cut off point. However, the bias in our analysis would only be present if the ratio of authors per article were too different over a too small sample. We also did not want to unnecessarily exclude co-authors. Therefore, we calculated the average number of authors per article for each year per field and used this as the cut off point.

Lastly, we have to point out a more fundamental issue with our research. This concerns the data we used to analyse the knowledge structure in particular. The keyword data that we used was indexed for each article by the WoS based on words in the titles and abstracts of cited references. However, these words largely remain concepts with definitions interpreted by the scientists of that time. However, that would likely be the case with any form of labelling as this will always be affected by the social realm, which makes it difficult to provide the full context and dynamics of the actual cognitive patterns.

Future research

We suggest three different directions for future research based on our thesis. The first direction is to improve the certainty and reliability of our research. We suggest deeper comparisons and analyses based on our data and overcoming the limitations described above. This includes distribution statistics of citations and productivity, such as testing for power-laws and the preferential attachment coefficient (Perc, 2014). Additionally, calculating the overlap of (prior) knowledge among keywords or scientists could improve the indication of cognitive similarities. This could be calculated using the Jaccard Index, which calculates the proportion of overlap to uniqueness of two sets (Heimeriks & Balland, 2016; Li et al., 2016;

Zeng et al., 2017). Adding insights in how much fields overlap cognitively over time or size would improve certainty in the interpretations of our Gini coefficients.

The second direction for future research is conducting a case study on a specific scientific field, preferably one that has also shown a declining productivity. The aim of this research is to map the selection environment and analyse the field from its emergence up to its decline. Besides bibliographic analyses that examine the social and cognitive dynamics over time as part of the organisational structure (preferably including the suggestions above, not only inequality indexes), such future research should include a timeline of relevant circumstances that could affect the evolution of the field, e.g. sudden changes in government funding, significant public exposure, industry interest, or prestigious prices. The research should provide proper context of exogenous influence as well as internal steering. Given the complexity of such scientific endeavour, we suggest focusing on a specific region.

Our third and final suggested direction concerns further exploring the link between inequality, competition and reward mechanisms. We have discovered that self-organising systems, such as a scientific field, that reward success with an increased chance of achieving more success, inherit inequality amongst its agents and knowledge genes over growth. As agents compete for rewards, they are driven by achieving as much success as possible. Here lies the question of what happens when rewards are distributed differently. This could be researched through modelling a self-organising system, since this allows to change the reward distributions and keeping all other variables constant. The model should include a form of currency as reward and requirement to access resources for which the agents compete. Additionally, agents should be able to collaborate and negotiate, such that their knowledge blocks can be shared and retained, similar to research articles or patent licensing.

References

Aldrich, H. E., & Fiol, C. M. (1994). Fools rush in: The institutional context of creation. *Academy of Management Review*, *19*(4), 645–670. https://doi.org/10.2307/258740

Allison, P. D., Long, J. S., & Krauze, T. K. (1982). Cumulative Advantage and Inequality in Science. *American Sociological Review*, *47*(5), 615–625. JSTOR. https://doi.org/10.2307/2095162

Andersen, E. S. (2001). Toward a Multiactivity Generalisation of the Nelson–Winter Model. *Unpublished*, 42.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press. https://doi.org/10.3998/mpub.10029

Arthur, W. B. (2007). The structure of invention. *Research Policy*, *36*(2), 274–287. https://doi.org/10.1016/j.respol.2006.11.005

Azoulay, P., Stuart, T., & Wang, Y. (2013). Matthew: Effect or Fable? *Management Science*, *60*(1), 92–109. https://doi.org/10.1287/mnsc.2013.1755

Blind, K. (2012). The influence of regulations on innovation: A quantitative assessment for OECD countries. *Research Policy*, *41*(2), 391–400. https://doi.org/10.1016/j.respol.2011.08.008

Bol, T., de Vaan, M., & van de Rijt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences*, *115*(19), 4887–4890. https://doi.org/10.1073/pnas.1719557115

Bonaccorsi, A. (2008). Search Regimes and the Industrial Dynamics of Science. *Minerva*, *46*(3), 285. https://doi.org/10.1007/s11024-008-9101-3

Bonaccorsi, A., & Thoma, G. (2007). Institutional complementarity and inventive

performance in nano science and technology. *Research Policy*, *36*(6), 813–831. https://doi.org/10.1016/j.respol.2007.02.009

Boon, W. P. C., Moors, E. H. M., Meijer, A., & Schellekens, H. (2010). Conditional approval and approval under exceptional circumstances as regulatory instruments for stimulating responsible drug innovation in Europe. *Clinical Pharmacology and Therapeutics*, *88*(6), 848–853. https://doi.org/10.1038/clpt.2010.207

Bravo, G., Farjam, M., Moreno, F. G., Birukou, A., & Squazzoni, F. (2018). Hidden connections: Network effects on editorial decisions in four computer science journals. *Journal of Informetrics*, *12*(1), 101–112.

Breschi, S., Malerba, F., & Orsenigo, L. (2000). Technological Regimes and Schumpeterian Patterns of Innovation. *The Economic Journal*, *110*(463), 388–410. https://doi.org/10.1111/1468-0297.00530

Bryman, A. (2008). *Social research methods* (3rd ed). Oxford University Press.

Carpenter, M. P. (1979). Similarity of Pratt's measure of class concentration to the Gini index. *Journal of the American Society for Information Science*, *30*(2), 108–110. https://doi.org/10.1002/asi.4630300210

Ceci, S. J., & Peters, D. P. (1982). Peer review: A study of reliability. *Change: The Magazine of Higher Learning*, *14*(6), 44–48.

Cheng, F.-F., Huang, Y.-W., Yu, H.-C., & Wu, C.-S. (2018). Mapping knowledge structure by keyword co-occurrence and social network analysis: Evidence from *Library Hi Tech* between 2006 and 2017. *Library Hi Tech*, *36*(4), 636–650. https://doi.org/10.1108/LHT-01-2018-0004

Cole, J. R., & Cole, S. (1974). Social Stratification in Science. *Science and Society*, *38*(3), 374–378.

Dawkins, R. (2016). *The Selfish Gene: 40th Anniversary edition* (Fourth Edition). Oxford University Press.

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative Advantage as a Mechanism for Inequality:

A Review of Theoretical and Empirical Developments. *Annual Review of Sociology*, *32*(1), 271–297. https://doi.org/10.1146/annurev.soc.32.061604.123127

Dosi, G., Freeman, C., Nelson, R., Silverberg, G., & Soete, L. (1988). *Technical Change and Economic Theory* [LEM Book Series]. Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies. https://econpapers.repec.org/bookchap/ssalembks/dosietal-1988.htm

Dosi, G., & Nelson, R. R. (1994). An introduction to evolutionary theories in economics. *Journal of Evolutionary Economics*, *4*(3), 153–172. https://doi.org/10.1007/BF01236366

Etzkowitz, H. (2008). The triple helix: University-industry-government innovation in action. In *The Triple Helix: University-Industry-Government Innovation in Action* (illustrate ed.). Routledge. https://doi.org/10.4324/9780203929605

Etzkowitz, H., & Leydesdorff, L. (2000). *The dynamics of innovation: From National Systems and '" Mode 2 "' to a Triple Helix of university – industry – government relations*. 109–123.

Fagerberg, J. (2003). Schumpeter and the revival of evolutionary economics: An appraisal of the literature. *Journal of Evolutionary Economics*, *13*(2), 125–159. https://doi.org/10.1007/s00191-003-0144-1

Faulkner, A., & Poort, L. (2017). Stretching and Challenging the Boundaries of Law: Varieties of Knowledge in Biotechnologies Regulation. *Minerva*, *55*(2), 209–228. https://doi.org/10.1007/s11024-017-9326-0

Fuchs, S. (1993). A Sociological Theory of Scientific Change. *Social Forces*, *71*(4), 933–953. https://doi.org/10.2307/2580125

Fujigaki, Y. (1998). Filling the gap between discussions on science and scientists' everyday activities: Applying the autopoiesis system theory to scientific knowledge. *Social Science Information*, *37*(1), 5–22. https://doi.org/10.1177/053901898037001001

Garfield, E. (1990). Keywords Plus®: ISI's breakthrough retrieval method. Part 1.

Expanding your searching power on Current Contents on Diskette. *Current Contents®*, *1*(32), 5–9.

Gould, S. J. (2002). *The structure of evolutionary theory*. Belknap Press of Harvard University Press.

Heimeriks, G., & Balland, P.-A. (2016). How smart is specialisation? An analysis of specialisation patterns in knowledge production. *Science and Public Policy*, *43*(4), 562–574. https://doi.org/10.1093/scipol/scv061

Heimeriks, G., & Boschma, R. (2014). The path- and place-dependent nature of scientific knowledge production in biotech 1986–2008. *Journal of Economic Geography*, *14*(2), 339–364. https://doi.org/10.1093/jeg/lbs052

Heimeriks, G., & Leydesdorff, L. (2012). Emerging search regimes: Measuring co-evolutions among research, science, and society. *Technology Analysis & Strategic Management*, *24*(1), 51–67. https://doi.org/10.1080/09537325.2012.643562

Heylighen, F. (1993). Selection Criteria for the Evolution of Knowledge. *Proceedings of the 13th International Congress on Cybernetics*, 524–528.

Jansen, D., von Görtz, R., & Heidler, R. (2010). Knowledge production and the structure of collaboration networks in two scientific fields. *Scientometrics*, *83*(1), 219–241. https://doi.org/10.1007/s11192-009-0022-1

Kauffman, S. A. (1993). *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10–25. https://doi.org/10.1002/asi.5090140103

Kuhn, T., Perc, M., & Helbing, D. (2014). Inheritance Patterns in Citation Networks Reveal Scientific Memes. *Physical Review X*, *4*(4), 041036. https://doi.org/10.1103/PhysRevX.4.041036

Kuhn, T. S. (1970). *The structure of scientific revolutions* ([2d ed., enl, Vol. 2). University of Chicago Press.

Kukk, P., Moors, E. H. M., & Hekkert, M. P. (2016). Institutional power play in innovation systems: The case of Herceptin®. *Research Policy*, *45*(8), 1558–1569. https://doi.org/10.1016/j.respol.2016.01.016

Kuo, C.-I., Wu, C.-H., & Lin, B.-W. (2019). Gaining from scientific knowledge: The role of knowledge accumulation and knowledge combination. *R&D Management*, *49*(2), 252–263. https://doi.org/10.1111/radm.12322

Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology*, *61*(2), 424–427. https://doi.org/10.1002/asi.21232

Leahey, E. (2004). The role of status in evaluating research: The case of data editing. *Social Science Research*, *33*(3), 521–537.

Leydesdorff, L. (1997). Why words and co-words cannot measure the development of the sciences. *Journal of the American Society for Information Science*, *48*, 418–427.

Leydesdorff, L., & Schank, T. (2008). Dynamic animations of journal maps: Indicators of structural changes and interdisciplinary developments. *Journal of the American Society for Information Science and Technology*, *59*(11), 1810–1818. https://doi.org/10.1002/asi.20891

Li, H., An, H., Wang, Y., Huang, J., & Gao, X. (2016). Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network. *Physica A: Statistical Mechanics and Its Applications*, *450*, 657–669. https://doi.org/10.1016/j.physa.2016.01.017

Luebker, M. (2010). *TRAVAIL Policy Brief No. 3*. 8.

Macfarlane, B., & Cheng, M. (2008). Communism, Universalism and Disinterestedness: Re-examining Contemporary Support among Academics for Merton's Scientific Norms. *Journal of Academic Ethics*, *6*(1), 67–78. https://doi.org/10.1007/s10805-008-9055-y

Madhugiri, V. S., Ambekar, S., Strom, S. F., & Nanda, A. (2013). A technique to identify

core journals for neurosurgery using citation scatter analysis and the Bradford distribution across neurosurgery journals: A review. *Journal of Neurosurgery*, *119*(5), 1274–1287. https://doi.org/10.3171/2013.8.JNS122379

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.

Mazzucato, M. (2018). Mission-oriented innovation policies: Challenges and opportunities. *Industrial and Corporate Change*, *27*(5), 803–815. https://doi.org/10.1093/icc/dty034

Merton, R. K. (1942). A Note on Science and Democracy. *Journal of Legal and Political Sociology*, *1*(1 and 2), 115–126.

Merton, R. K. (1957). Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *American Sociological Review*, *22*(6), 635–659. https://doi.org/10.2307/2089193

Merton, R. K. (1968). The Matthew Effect in Science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, *246*(1), 1–19. https://doi.org/10.1016/j.ejor.2015.04.002

Nelson, R. R., & Winter, S. G. (1974). Neoclassical vs. Evolutionary Theories of Economic Growth: Critique and Prospectus. *The Economic Journal*, *84*(336), 886. https://doi.org/10.2307/2230572

Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change* (SSRN Scholarly Paper ID 1496211). Social Science Research Network. https://papers.ssrn.com/abstract=1496211

Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E*, *64*(2), 025102. https://doi.org/10.1103/PhysRevE.64.025102

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351. https://doi.org/10.1080/00107510500052444

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology*, *41*(1), 609–641. https://doi.org/10.1002/aris.2007.1440410120

Penna, C. C. R., & Geels, F. W. (2015). Climate change and the slow reorientation of the American car industry (1979-2012): An application and extension of the Dialectic Issue LifeCycle (DILC) model. *Research Policy*, *44*(5), 1029–1048. https://doi.org/10.1016/j.respol.2014.11.010

Perc, M. (2014). The Matthew effect in empirical data. *Journal of The Royal Society Interface*, *11*(98), 20140378. https://doi.org/10.1098/rsif.2014.0378

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Krabel, S., Kitson, M., Llerena, P., Lissoni, F., Salter, A., & Sobrero, M. (2013). Academic engagement and commercialisation: A review of the literature on university–industry relations. *Research Policy*, *42*(2), 423–442. https://doi.org/10.1016/j.respol.2012.09.007

Petersen, A. M., & Penner, O. (2014). Inequality and cumulative advantage in science careers: A case study of high-impact journals. *EPJ Data Science*, *3*(1), 1–25. https://doi.org/10.1140/epjds/s13688-014-0024-y

Pfeiffer, T., & Hoffmann, R. (2007). Temporal patterns of genes in scientific publications. *Proceedings of the National Academy of Sciences*, *104*(29), 12052–12056. https://doi.org/10.1073/pnas.0701315104

Pratt, A. D. (1977). A measure of class concentration in bibliometrics. *Journal of the American Society for Information Science*, *28*(5), 285–292. https://doi.org/10.1002/asi.4630280508

Price, D. J. D. S. (1963). Little Science, Big Science. In *Little Science, Big Science*. Columbia University Press. http://www.degruyter.com/document/doi/10.7312/pric91844/html

Price, D. J. D. S. (1976). A general theory of bibliometric and other cumulative advantage

processes. *Journal of the American Society for Information Science*, *27*(5), 292–306. https://doi.org/10.1002/asi.4630270505

Rigby, D. L. (2015). Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes. *Regional Studies*, *49*(11), 1922–1937. https://doi.org/10.1080/00343404.2013.854878

Samuels, J. M. (1965). Size and the Growth of Firms. *The Review of Economic Studies*, *32*(2), 105–112. https://doi.org/10.2307/2296055

Santarelli, E., Klomp, L., & Thurik, A. R. (2006). Gibrat's Law: An Overview of the Empirical Literature. In E. Santarelli (Ed.), *Entrepreneurship, Growth, and Innovation: The Dynamics of Firms and Industries* (pp. 41–73). Springer US. https://doi.org/10.1007/0-387-32314-7_3

Sauder, M., Lynn, F., & Podolny, J. M. (2012). Status: Insights from Organizational Sociology. *Annual Review of Sociology*, *38*(1), 267–283. https://doi.org/10.1146/annurev-soc-071811-145503

Schlaile, M. P. (2020). *Memetics and Evolutionary Economics: To Boldly Go Where no Meme has Gone Before*. Springer Nature.

Schumpeter, J. A. (1943). *Capitalism, socialism, and democracy,*. G. Allen & Unwin ltd.

Shiozawa, Y., Morioka, M., & Taniguchi, K. (2019). Microfoundations of Evolutionary Economics. In Y. Shiozawa, M. Morioka, & K. Taniguchi (Eds.), *Microfoundations of Evolutionary Economics* (pp. 1–52). Springer Japan. https://doi.org/10.1007/978-4-431-55267-3_1

Siler, K., Sugimoto, C. R., & Larivière, V. (2018). *Cumulative Advantage in Scientific Visibility: Citation Performance of Repeat Authors in Economics Journals*. 228–254.

Simcoe, T. S., & Waguespack, D. M. (2011). Status, quality, and attention: What's in a (missing) name? *Management Science*, *57*(2), 274–290.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization*, *1*(1), 161–176.

Smink, M. M., Hekkert, M. P., & Negro, S. O. (2015). Keeping sustainable innovation on a leash? Exploring incumbents' institutional strategies. *Business Strategy and the Environment*, *24*(2), 86–101. https://doi.org/10.1002/bse.1808

Smith, A., Voß, J. P., & Grin, J. (2010). Innovation studies and sustainability transitions: The allure of the multi-level perspective and its challenges. *Research Policy*, *39*(4), 435–448. https://doi.org/10.1016/j.respol.2010.01.023

Sun, X., Kaur, J., Milojević, S., Flammini, A., & Menczer, F. (2013). Social Dynamics of Science. *Scientific Reports*, *3*(1), 1069. https://doi.org/10.1038/srep01069

Tijssen, R. J. W., & Van Raan, A. F. J. (1994). Mapping Changes in Science and Technology: Bibliometric Co-Occurrence Analysis of the R&D Literature. *Evaluation Review*, *18*(1), 98–115. https://doi.org/10.1177/0193841X9401800110

Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Reviewer bias in single-versus double-blind peer review. *Proceedings of the National Academy of Sciences*, *114*(48), 12708–12713.

van den Besselaar, P., & Heimeriks, G. (2006). Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. *Scientometrics*, *68*(3), 377–393. https://doi.org/10.1007/s11192-006-0118-9

Watts, C., & Gilbert, N. (2011). Does cumulative advantage affect collective learning in science? An agent-based simulation. *Scientometrics*, *89*(1), 437. https://doi.org/10.1007/s11192-011-0432-8

Web of Science Core Collection Help. (n.d.). Retrieved October 1, 2021, from https://images.webofknowledge.com/WOKRS535R83/help/WOS/hp_full_record.html

Weber, K. M., & Rohracher, H. (2012). Legitimizing research, technology and innovation policies for transformative change: Combining insights from innovation systems and multi-level perspective in a comprehensive "failures" framework. *Research Policy*, *41*(6), 1037–1047. https://doi.org/10.1016/j.respol.2011.10.015

Whitley, R. (2000). *The Intellectual and Social Organization of the Sciences* (Second

Edition). Oxford University Press.

Whittaker, J. (1989). Creativity and Conformity in Science: Titles, Keywords and Co-word Analysis. *Social Studies of Science*, *19*(3), 473–496. https://doi.org/10.1177/030631289019003004

Witlox, F. (2017). Gini Coefficient. In *International Encyclopedia of Geography* (pp. 1–4). American Cancer Society. https://doi.org/10.1002/9781118786352.wbieg0855

Wouters, P. F. (1999). *The citation culture* [PhD thesis, Universiteit van Amsterdam]. https://dare.uva.nl/search?identifier=b101b769-100f-43e5-b8d2-cac6c11e5bbf

Zeng, A., Shen, Z., Zhou, J., Wu, J., Fan, Y., Wang, Y., & Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Physics Reports*, *714–715*, 1–73. https://doi.org/10.1016/j.physrep.2017.10.001

Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., & Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology*, *67*(4), 967–972. https://doi.org/10.1002/asi.23437

Zuckerman, H. (1977). *Scientific elite: Nobel laureates in the United States*. Transaction Publishers.