Leipzig University
Faculty for Social Science and Philosophy
Institute for Sociology

A thesis submitted for the degree of
Bachelor of Arts

Economic opportunities are distributed unequally: On the spatial concentration of complexity in German cities

Name	Clemens Baldzuhn
Matrikelnummer	3713306
E-Mail	clemens.baldzuhn@uni-leipzig.de
Adresse	Breitkopfstraße 18, 04317 Leipzig
Datum	8. August 2022
Betreuer	M.Sc. Felix Lennert, Prof. Dr. Marc Keuschnigg

Contents

Abstract

The massive growth in urbanisation carries with it a need to understand the economic attractiveness of cities and to structure the urban-rural transition socially just. It is well established that economically profitable opportunities and welfare concentrate in cities. Explanations of the phenomenon which relate the concentration of beneficial economic conditions in cities to their complexity have recently gained prominence. In this light, the present work investigates the relationship between the spatial concentration and the complexity of industry sectors and occupational groups on the German labour market using a novel dataset of job offerings. The results reveal that job offerings originating from more complex industries and occupations are significantly over-represented in large cities. Crucially, the results also show that with larger city size, jobs in the same occupational group get more complex and that this pattern gets more pronounced with increases in the required skill and specialisation level. Taken together, the findings stress the fact that cities are crucial for the emergence of highly profitable economic endeavours. Yet, it is illustrated that not all economic actors and social groups can benefit equally from the profitable economic conditions in cities. This poses challenges to the future dynamics of spatial inequality.

1 Introduction

Social life increasingly concentrates in large cities. In the last 70 years the fraction of the world population that lived in cities rose from 30% to about 50% and it is expected that the pace of urban expansion will accelerate in the future (United Nations, 2019). Accordingly, it is expected that the number of megacities – cities with more than 30 million inhabitants – likely exceeds 40 in the year 2030 (WBGU, 2019). This ongoing agglomeration of people and resources in space carries with it major challenges to nature and society. Cities produce 70% of the global CO_2 emissions (Seto et al., 2014) and thus contribute over-proportionately to climate change. Likewise, the growing urban-rural divide promotes increases in spatial inequality (Tickamyer, 2000). Already today, the access to healthcare, education, culture and the opportunities to participate in high-wage economic activities concentrates in cities (Keuschnigg and Wolbring, 2019, p. 34).

It shows that understanding urbanisation to tackle the drastic changes in the upcoming century touches various sociological subject matters. In effect, the groundwork for the present investigation of the societal effects of cities lies in classical sociological works. Max Weber (1958), for example, recognised early that cities are the engines which foster socioeconomic change. In his typology of cities, he recognised that changes in the societal function of cities stand in relation to social advances. Accordingly, ancient cities were consumer-cities that concentrated urban elites who exploited goods and rents from the countryside. In the course of the capitalist turn in the medieval ages, cities became producer-cities which sustained themselves through production and trade (Morris and Manning, 2005, p. 142). Hence, Weber established the idea that cities, above all, have an economic function for society (Weber, 1958, p. 67). Evidence that modern capitals contribute a majority to their countries' Gross-domestic product (GDP) underlines the sustained relevance of Weber's notions (IWD, 2022).

Besides, classic sociologist Georg Simmel laid the foundations to the research on the impacts and consequences of modern cities on attitudes and social behaviour. In his well-known treatise *The Metropolis and Mental Life* Simmel observes that the increased physical proximity in cities overstimulates the senses, which could lead to withdrawal from social life[1] (Simmel, 1903/1950, p. 410). Simmel's findings were later echoed in psychological studies (Milgram, 1970). Moreover, Simmel asserted that the intensification of social contact in cities promotes the multiplication and diversification of social roles.

In accordance with Simmel, American sociologist Louis Wirth pointed out similar effects of urban life on the individual in his work *Urbanism as a Way of Life* (Wirth, 1938). In his paper, he made the effort to separate cause and effect to gain an analytical understanding of the social dynamics in cities. His notions entail that the most important

[1] The concept of Anomia by Emile Durkheim is related to Simmel's notions but rather emphasises the erosion of social norms in the late 19th century as the principal cause for social isolation (Durkheim, 1897/2005).

factors that influence the attributes of cities are the total population and the population density. Besides their detrimental effects on the individual, for example, alienation, Wirth also stressed, that larger population sizes carry benefits on a macro-level. In particular, he identified that city size and density foster material welfare by permitting division of labour and specialisation, adopting the famous notions of economist Adam Smith (1776/2010).

Hence, early sociological thinking has contributed largely to the understanding of cities. In short, sociological accounts have stressed that (a) the growth of cities happens in unison with socioeconomic advances (Weber, 1958), (b) modernity made cities become the centres of social life which had far-reaching consequences for social norms, roles and attitudes (Simmel, 1950), and (c) city size promotes economic prosperity (Wirth, 1938). In the present work, particularly notions (a) and (c) are addressed to broaden the knowledge on the questions of

1. *Why do economic opportunities and welfare concentrate in cities?*
2. *What properties do cities have that foster their growth?*

Much has happened to the science of cities since early theorists established the field. Especially Louis Wirth's realisation that the city size is crucial for its functionality has been expanded. It is now widely accepted that economic productivity tends to increase over-proportionally with city size (Melo et al., 2009; Rosenthal and Strange, 2004). In the scientific discourse on the topic, these higher-than-expected outcomes that come with population size are referred to as agglomeration benefits. Interestingly, it was found that agglomeration benefits in large cities increase over-proportionately with city size in a large number of urban systems across the world (Bettencourt et al., 2007). This has led researchers in the field of urban scaling to propose that the association between population size and agglomeration benefits follows universal scaling laws (Bettencourt et al., 2010; Bettencourt and West, 2010).

In particular, the urban scaling paradigm has revealed that a multitude of beneficial socioeconomic outcomes Y are a function of city size N which can be modelled with a power-law of the form $Y(N) \sim Y_0 N^{\beta}$. In the formula, Y_0 denotes a constant, and β is the scaling coefficient, which describes the effects of a 1% change in population size on Y.

However, research has also demonstrated that not all economic activities profit from the agglomeration benefits in cities equally. Rather, there exist large variations among economic categories concerning their scaling with city size. While, for example, the number of industries in the informational and service sector increases over-proportionately with city size, industries in the primary sector are under-represented in large cities (Youn et al., 2016). Also, it shows that creative professions like scientist and artists exhibit larger scaling coefficients than other occupations (Balland et al., 2020; Bettencourt et al., 2007).

In recent years, explanations have gained prominence which relate scaling differences among economic activities to their complexity. Evidence accumulates that economic activities which are more complex – activities that require more complementary factors to be present simultaneously – tend to exhibit larger scaling coefficients and are over-represented in cities (Balland et al., 2020; Gomez-Lievano et al., 2016).

Thus, understanding the relationship between the complexity of economic activities and their spatial agglomeration may provide new insights into the question of how cities grow and how spatial inequality comes about. Furthermore, investigating the scaling of economic activities of varying complexity might permit a more fine-grained perspective on the question of which social domains profit most from the concentration of opportunities in cities.

The present work examines this particular notion. In specific, the hypothesis that the scaling (spatial concentration) of economic activities increases with their complexity is developed and tested empirically. For this purpose, selected findings from the fields of economic sociology, urban economics, economic complexity and urban scaling are highlighted and concepts are connected to substantiate theoretically why it is expected that complexity concentrates in cities. In a subsequent step, these theoretical ideas are tested against a novel data set of job advertisements from the German labour market.

The results of this investigation demonstrate that job advertisements originating from more complex industries and professions are over-represented in large cities. Hence, complex economic activities concentrate in large cities. Moreover, it is established that with an increase in city size, the complexity also increases within an economic activity. That is, the same job gets more complex with city size.

This study provides a novel contribution to the above-mentioned fields of research and the science of cities in general. Firstly, the sociological understanding of spatial inequality of economic opportunities is advanced. Secondly, research on economic complexity is extended by a new metric to asses complexity, that is the Shannon entropy of job advertisements. And thirdly, this work identifies the complexity of economic activities as an urban indicator which complies with scaling laws.

2 Concepts and Background

2.1 Complexity

In the section that follows, the concept of complexity complexity is specified and it is illustrated how complexity emerges in economic domains.

In her book *Complexity: A Guided Tour*, Melanie Mitchell illustrates the basic principles of complex systems. According to her, complex systems are characterised by 'large numbers of relatively simple entities [which] organise themselves, without the benefit of any central controller, into a collective whole that creates patterns, uses information, and, in some cases, evolves and learns' (Mitchell, 2009, p. 4). This definition of complexity is commonly boiled down to the idea that complex systems have macro-qualities that emerge from actions of independent entities[2]. Importantly, the macro properties of complex systems do not follow from the mere addition of the actions of micro-actions but exhibit non-linear, hard-to-predict dynamics (Thurner et al., 2018, p. 8).

Concepts from complex systems science were adopted in several research fields, including economics. The field of complexity economics is built on the notion that economies (macro-qualities) differ in terms of complexity due to their different capabilities to provide complementary inputs (entities) (Arthur, 2021). In a seminal paper Hidalgo and Hausmann (2009) propose a measure which determines economic complexity by the product portfolio of economies. Following studies pointed out the usefulness of these metrics to predict economic growth (Zhu and Li, 2017). However, production only pertains to one dimension of complexity. Another important dimension is the complexity of knowledge (Balland et al., 2022). This know-how dimension of complexity is particularly relevant to measuring economic complexity in developed countries which are characterised by tertiarisation and automation and therefore depend less on the production of material goods. Since the focus of this work lies on the German labour market, (economic) complexity means knowledge complexity hereafter.

Although knowledge complexity became subject to many studies in recent years, there is no common definition of it yet (Mewes and Broekel, 2020). In general, research on the concept revolves around the idea that 'the whole knows more because individuals know different' (Balland et al., 2022, p. 2) and that knowledge 'is more complex when it draws upon distinct and multiple kinds of components' (Zander and Kogut, 1995, p. 79).

Those definitions point to the testable assumption that knowledge complexity arises from specialised individuals who share their complementary knowledge. This suggests that the complementary division of knowledge should bring about productivity benefits analogously to the division of labour. A growing number of studies demonstrate this. Typically, these studies deploy network metrics to account for the relational, complementary dimension of knowledge.

Alabdulkareem et al. (2018) use data from the U.S. labour market to construct a bipartite network of high and low-wage occupations. Their analyses reveal that high-wage occupations exhibit the highest skill complementarities to other occupations and are most knowledge-intensive. A related study by Anderson (2017) underlines the fact that workers in knowledge-heavy industries earn more when they have complementary sets of skills. In the same vein, Neffke (2019) stresses that the combination of knowledge is more valuable than the sum of its parts. The author uses individual-level administrative data on the distribution of occupations and educational attainment in Swedish firms. Focusing

[2] These ideas resonate well with common explanatory models in sociology, in particular, the so-called 'Coleman diagram' (Coleman, 1986)

on interdependencies in the workplace, Neffke finds that co-workers with complementary knowledge earn higher wages than co-workers with substitutable knowledge. Interestingly, the results also show that the effect of complementary co-workers on wages becomes more pronounced with city size. These findings lead to the question of why knowledge complexity should concentrate in large cities.

Before presenting a selection of explanations on this matter, the next section highlights recent findings on the spatial concentration of complexity. In this regard, emphasis is put on works that highlight the knowledge dimension of economic complexity. For an extensive overview see Hidalgo (2021).

2.2 Background

Using the metrics developed by Hidalgo and Hausmann (2009), Lo Turco and Maggioni (2020) employ trade data to measure product-complexity in different industry and occupational fields in the US economy. Their results reveal that the most complex products require the most knowledge related to science, technology, engineering and mathematics. Furthermore, the authors show that knowledge-heavy industries and occupations tend to accumulate in specific regions in the United States. Similarly, Mewes and Broekel (2020) find a geographic concentration of the capabilities to produce complex technological products in European regions. To measure knowledge complexity in technological production, the authors apply network metrics to quantify the diversity of knowledge components involved in the submission of patents. Antonelli et al. (2020) and Pintar and Scherngell (2021) perform the same analyses on the spatial concentration of complex knowledge in European regions and also demonstrate that complexity predicts regional economic growth. Importantly, all those studies point to the fact that complexity does not concentrate randomly in space but sticks to densely populated areas. A study by Gomez-Lievano et al. (2016) demonstrates that complex phenomena occur in larger and more diverse urban centres. Using multiple data sources from the U.S., the researchers develop and test a statistical model built on the assumption that more complex phenomena require more complementary factors. Interestingly, their model explains why, for example, the number of innovations grows super-linearly with city size. Balland et al. (2020) use less sophisticated measures of complexity but also demonstrate that more complex activities concentrate in large U.S. cities. In their study, the authors measure the complexity in different economic domains like scientific fields and technologies. Concerning this work, their results on the spatial concentration of industry sectors and occupational fields are particularly relevant. Measuring knowledge complexity in terms of average years of education, it is shown that industries and occupations characterised by a high knowledge complexity are over-represented in large cities.

Having discussed recent evidence on the spatial agglomeration of complexity, the section that follows will feature selected explanations on why complexity should concentrate in large cities.

2.3 Why complexity concentrates in cities

Complexity should concentrate in cities because cities provide a beneficial environment for the production of economically complex products. Naturally, the properties of cities which stimulate economic activities are manifold and object to a large debate in many scientific fields. Duranton and Puga (2004) offer a conceptual division of the literature from an urban economics view. Keuschnigg and Wolbring (2019) provide a rather sociologically substantiated classification of mechanisms in cities which foster complex economic activities. In the following, a simple dichotomous classification of the literature will be put forward to summarise principal ideas and findings. Firstly, explanations which emphasises structural qualities of cities are summarised under the term *interconnectivity*. Secondly, explanations which point out the advantageous location factors and population characteristics in cities are listed under the term *composition*. Still, this analytical division of the literature can only be seen as a cognitive aid since it will become apparent that complexity in cities emerges through entangled and mutually reinforcing channels.

2.3.1 Interconnectivity

Explanations which pertain to the interconnectivity in cities emphasise structural qualities which stimulate the emergence of (complex) economic activities. Above all, prominence is given to the effect the population density has on the flow of information (Batty, 2008).

As mentioned above, the production of complex products relies on a deep division of knowledge between workers (Balland et al., 2022). Early accounts in economics have stressed that the division of tasks between workers is facilitated by spatial proximity which in turn fosters economic output. Becker and Murphy (1992) deliver a formalisation of the connection between economic output and city size. Importantly, Becker and Murphy also point to the underlying mechanism which drives this effect. According to them, the biggest advantage of cities is that they reduce the coordination costs which rise with the complexity of tasks: '[...] The division of labour may be greater in cities than in small towns not because markets are larger in cities, but because it is easier to coordinate specialists in more densely populated areas.' (Becker and Murphy, 1992, p. 1148). More recently, the mechanisms connecting population density and beneficial effects on economic activities were investigated in network science. Pan et al. (2013) build a network model to examine the effects of increases in population size on the number of social ties and the spread of information. They find that the number of social ties and the pace of information transmission increases exponentially with city size. Eagle et al. (2010) draw attention to the fact that the enhanced opportunities to form heterogeneous networks in cities are connected with beneficial outcomes for the inhabitants, particularly in terms of economic welfare. Coupling real-world communication network data with community-level socioeconomic output information, the authors find that social network diversity is a strong indicator of economic growth.

In the field of urban geography, the relation between close networks and beneficial economic outcomes is studied under the term spillover effects. These accounts stress that proximity between workers and firms promotes the transmission (spillover) of knowledge, the emergence of new ideas, innovation and, ultimately, economic development. In a renowned study, Audretsch and Feldman (1996) investigate the relationship between knowledge spillovers and spatial concentration from an industry perspective. The authors determine knowledge spillovers within industry sectors as a function of their economic activities in cooperation with non-industry establishments, for example, with universities. Their results show that industries which engage in activities that promote knowledge-spillover exhibit higher spatial clustering and are more innovative. From an employee perspective, it was also established that working in proximity has advantageous effects. Carbonell and Rodriguez (2006) find that proximity of teams is essential for face-to-face communication and determines the success of teams that manufacture complex technological products. Lengyel and Eriksson (2016) substantiate the importance of co-worker interactions with network measures. Using firm-level employment data from Sweden and network measures to estimate social-tie density, they show that dense working environments are associated with regional economic growth.

2.3.2 Composition

Explanations pertaining to composition emphasise the advantageous location factors and population characteristics in cities. As indicated previously, complexity requires the simultaneous presence of a large number of complementary factors. Certainly, these complementarities can not be found anywhere. Rather, companies and workers should move to areas where the location factors are advantageous for their business and promise high returns. This in turn shapes the economic landscape, which shows that the strategic decisions of economic actors have a large effect on spatial concentration of complex economic activities.

Alfred Marshall (1890/2013) famously pointed out why industries form localised clusters. According to him, physical conditions like the presence of charcoal or good soil initially propel the clustering of industries. Crucially, he called attention to the fact that initial beneficial conditions of this kind lead to the settlement of related industries which can generate profits from spatial proximity (ibid., p. 226). It is widely acknowledged today that especially complementary industries have an incentive to cluster in space. Recent investigations by Park et al. (2019) support this. The authors use data from LinkedIn and statistical metrics to cluster businesses which exhibit the strongest labour-flow between them. It shows that clusters of businesses relate to industry sectors and that individual businesses in these clusters also tend to be geographically adjacent. Moreover, it is found that the influx of highly-educated workers into industry clusters predicts economic growth.

Those latter findings draw attention to the composition of the workforce for innovative and profitable economic endeavours. In this regard, Jane Jacobs (1969/2016) prominently directed attention to the characteristics of people in cities and their relevance for innovation. Later, Richard Florida popularised these ideas and empirically demonstrated that especially the creative and highly-educated class agglomerates in cities and boosts innovative economic activities, firm formation and industrial development (Florida, 2004; Lee et al., 2004). Likewise, Boschma and Fritsch (2009) examined the association between the concentration of the creative class and beneficial economic outcomes on European employment data. Based on the classification of creative occupations developed by Florida 2004, they find that the creative class is unevenly distributed in European regions and that these agglomerations stand in relation to higher rates of start-up formation and total employment. The direction of causality, however, remains an open question. What comes first, places with favourable conditions or creative people which create these conditions? While for Florida, jobs and welfare follow people, findings from Keuschnigg et al. (2019) demonstrate reverse causality. Their results imply that the highly educated and smart class is attracted by beneficial socioeconomic conditions. Using geocoded microdata from Sweden, the authors show that the selective migration of the most productive and people partially explains the over-proportionate wages in cities.

2.4 Hypotheses and aims

Against this background, three hypotheses are put forward and tested to investigate the relationship between the complexity of different economic domains and their spatial concentration on the German labour market. Hypothesis 1a and 1a aim to verify the underlying assumptions established above. As stated before, differences in location factors, the composition of the workforce and structural variations should provoke that economic activities are distributed unequally in space. Hypothesis 1a explores whether economic activities indeed differ in terms of their spatial concentration.

H1a: Economic activities exhibit varying degrees of spatial concentration.

Furthermore, it is expected that economic activities differ in terms of their complexity – the extent to which they exhibit a deep division of knowledge between complementary co-workers. Hence, Hypothesis 1b examines whether economic activities display different levels of complexity.

H1b: Economic activities exhibit varying degrees of complexity.

Hypothesis 2 establishes the link between spatial concentration and complexity to investigate whether complex economic activities agglomerate in space. Thus, this hypothesis follows previously published research by Balland et al., 2020.

H2: Complex economic activities exhibit a greater degree of spatial concentration then less complex industries.

Theoretically, the relationship between complexity and spatial concentration should also hold when switching the focus of analysis from total quantities (the number of job advertisements) to average quantities (the average complexity) (Shalizi, 2011). Thus, it is expected that the average complexity should display higher spatial concentrations for more complex economic activities. In other words, the average knowledge complexity should concentrate more in cities only for complex economic activities.

H3: Within complex economic activities, the average complexity increases with city size to a greater degree than less complex economic activities.

3 Data and Methods

3.1 Data

Job advertisement data was collected from the German Federal Employment Agency (Bundesagentur für Arbeit, BAA). German administrative institutions have a mandate to make their data publicly available and machine-readable since the E-government law was passed in 2017. Since the BAA does not provide a way to access their data yet, an unaccredited Application Programming Interface (API) was used. The API was created by Lilith Wittmann and is maintained by an activist collective[3].

The *Jobsuche* API[4], which was employed here, provides information on job advertisements posted to the database of the BAA. According to the agency, their records of open job positions in the German labour market are the most comprehensive database on job offerings in Germany. On the 26th of June 2022, the database contained about 1.4 million job ads. Companies either submit job offers themselves or report open positions to the BAA which in turn adds them to the database. Although many alternative job boards exist, they usually do not offer APIs. The Google owned careers platform is an exception but only offers commercial plans to access their data.

Further, the BAA states that the distribution of job offers in the database concerning industry sectors is reflected in the current labour market situation. Hence, job ads were

[3] Numerous endpoints to German administrative data can be found at their website *bund.dev*.
[4] https://bit.ly/3OZ9cRV

retrieved from industry sectors to achieve a representative sample. In particular, the 10,000 most recent job ads for 26 industries were downloaded. A potential source of bias may arise from the fact that not all industries recorded 10,000 job ads on the platform. Some industries like the advertising sector (N = 3,980) are covered completely while for other industries like tourism, only the latest job ads are contained in the data (N = 10,000). Furthermore, biases resulting from over- or under-reporting of certain companies can not be fully excluded. Still, systematic errors from differential reporting are not expected to be significant for the scaling analyses given that the relationship between city size and the total number of job ads is almost proportional ($\beta = 1.04 \pm 0.024$). Moreover, while checking for such biases, plotting city size and total cases does not reveal significant outliers (Appendix Fig. 7). Table 3 in the appendix lists the industry sectors, exact retrieval dates and case numbers.

Each job ad consists of a multitude of variables. For the analysis, the most essential variables are the title of the occupation, the advertisement text and the exact location of the job, defined by latitude and longitude. Table 4 in the appendix lists all the variables available in the dataset. Advertisements for more than one open position were excluded to avoid bias.

In total, $N = 180,719$ job advertisements were considered after cleaning the data and excluding cases which distorted the analysis (see above). All data were acquired between the 5th and the 9th of May 2022. Access to the API was established using the R libraries *httr* (Wickham, 2022) and *jsonlite* (Ooms, 2022).

3.2 Classification of occupations

Job ads were classified according to the International Classification of Occupations 2008 (ISCO-08). The ISCO classification scheme is developed and maintained by the International Labour Organization (ILO, 2012) and ranks occupations hierarchically according to the skill, educational and the specialisation level required for the job. A four-digit code serves to classify occupations in major, sub-major and unit groups. For example, if a job identifier starts with a 9, the job belongs to the major group of elementary occupations. In turn, the second digit denotes the sub-major group. That is, whether the job belongs to, for example, sub-major group 94: cleaners and helpers. Therefore, lower grouping numbers reflect a higher job complexity.

Given that the data set only provides standardised job titles and the BAA uses its own coding scheme, several conversions had to be performed. In a first step, the standardised occupation title from the job was mapped to the German classification of Occupations (KldB-2010)[5]. Next, the KldB-2010 was mapped to the ISCO-08 classification using

[5] The tables for correspondence between BAA job titles and KldB-2010 can be downloaded here: https://bit.ly/3SCDi0D

ISCO-08 major groups	Skill level
1 Managers	3 + 4
2 Professionals	4
3 Technicians and Associate Professionals	3
4 Clerical Support Workers 5 Services and Sales Workers 6 Skilled Agricultural, Forestry and Fishery Workers 7 Craft and Related Trades Workers 8 Plant and Machine Operators, and Assemblers	2
9 Elementary Occupations	1
0 Armed Forces Occupations	1 + 2 + 4

Figure 1: Mapping of ISCO-08 major groups to skill levels according to International Labour Organization. Source: (ILO, 2012)

another correspondence table provided by the BAA[6].

In addition, a dichotomous classification in complex and non-complex jobs was carried out following the scheme of Boschma and Fritsch (2009) which is based on notions of Florida (2004) and based on the ISCO-08. Essentially, educational, scientific, managerial and artistic professions are classified as creative.

Since the classification by Boschma and Fritsch is based on outdated ISCO-88 codes, creative occupations from ISCO-08 were mapped manually to ISCO-08 codes using correspondence tables from the ILO[7]. Table 5 in the appendix lists the correspondence between creative occupations in ISCO-88 and ISCO-08. A major problem that comes with the conversions is that differences between the KldB-2010 and ISCO-08 do not permit a 1:1 mapping of jobs. The ISCO-08 classification provides more granularity which leads to multiple possible classifications. To account for this artefact, cases were only retained if they were distinct or their classifications did not vary to an extent that lead to a twofold grouping as creative and non-creative. In the latter case, only the first classification was kept.

Furthermore, ISCO-08 major group of armed forces occupations (group 0) was excluded since it only included four cases which distorted the later analyses.

3.3 Unit of analysis

German labour market areas (LMAs) were chosen as a spatial unit of analysis. LMAs provide functional delineations of urban centres based on the commuting flows of the local workforce. Administrative districts are pooled to a LMA if the centre of their regional labour market can be reached within commuting less than 45 minutes. Moreover, LMAs

[6] The tables for correspondence between KldB-2010 and ISCO-08 can be downloaded here: https://bit.ly/3zW5iol

[7] The tables for correspondence between ISCO-08 and ISCO 88 can be downloaded here: https://bit.ly/3oRqENh

have a minimum population of 50 thousand people. Changes in commuter behaviour or territorial reforms are taken into account in updated redefinitions of the LMAs (BBSR ,2017). For a detailed explanation of the formation of German LMAs see (Kosfeld and Werner, 2012).

Using functional delineations in the form of LMAs as spatial units accounts for the fact that people often don't live where they work, which would introduce significant bias to the analysis of scaling relationships. Using non-administrative boundaries to define agglomeration areas follows standard procedures in the literature on urban scaling (Bettencourt et al., 2013).

In total, 257 LMAs were considered in the analysis. The smallest LMA, Sonneberg, has a total population of 57044. Berlin is the largest LMA with a population of 3664088. Each job advertisement was mapped to its corresponding LMA (polygon) using the geocode provided in the data. Job ads without geocodes were excluded from the analysis.

For the mapping, the R libraries *rgdal* (Bivand et al., 2022) and *sp* (Pebesma & Bivand, 2022) were used. Shapefiles were taken from the Federal Office for Building and Regional Planning (Bundesamt für Bauwesen und Raumordnung)[8]. LMA-level population data was taken from the Federal Statistical Office (Statistisches Bundesamt)[9].

3.4 Knowledge complexity

It is assumed that the knowledge complexity required in a job is reflected in the complexity of its advertisement text. Textual complexity is computed by the Shannon entropy which in essence measures the information content or amount of surprise in an information source (Shannon, 1948)[10].

Claude Shannon developed the metric in the early 20th century to answer the question of how much information gets transmitted through communication channels like a telegraph or the telephone, which were revolutionising information transfer at that time (Mitchell, 2009, p. 52). In layman's terms, Shannon entropy can be understood as the average amount of surprise that is obtained from observing an information source (Thurner et al., 2018, p. 328). An information source is composed of messages, which can be any unit of communication, for example, a bit, a letter or a word. The Shannon entropy H of an information source X is defined as

$$H(X) = \sum_{x \in X} p(x) * \log_2(\frac{1}{p(x)}) \tag{1}$$

[8] The shapefiles can be downloaded here: https://bit.ly/3QfDktd

[9] LMA-level population data can be downloaded here: https://bit.ly/3zoTYj2

[10] The Shannon entropy was developed on basis of the Boltzmann entropy, a metric from statistical mechanics which estimates disorder in a system and is frequently applied in complexity science (Mitchell, 2009, p. 96).

where $p(x)$ describes the expected value of message x to occur. Importantly, the inverse of the expected value $\frac{1}{p(x)}$ is the surprise that is associated with receiving that specific message. It follows that messages which occur less frequently are more surprising.

Melanie Mitchell gives an illustrative example to understand the formula (Mitchell, 2009, p. 53). If a one-year old toddler speaks with her grandmother on the phone and says 'bla bla bla', the expected value $p(x)$ for each message x – 'bla' – is 1. Consequently, the surprise of each word is low: $\log_2(\frac{1}{1}) = 0$. Therefore, summing up the product of expected value and surprise of each message to compute the entropy gives zero. The sentence by her granddaughter leaves the grandmother completely unsurprised. This changes if the toddler learns the word 'icecream' and the sentence changes to 'bla bla icecream'. Now, the expected value for 'bla' is $\frac{2}{3}$ and that for 'icecream' is $\frac{1}{3}$. Both words gain surprise: Suprise(bla) $= \log_2(\frac{1}{\frac{2}{3}}) \approx 0.18$ and Suprise(icecream) $= \log_2(\frac{1}{\frac{1}{3}}) \approx 0.48$. Calculating the entropy gives H(bla bla icecream) ≈ 0.28 – an increase in information in comparison to the first sentence. This example depicts Shannon's major contribution to the understanding of information. That is, that information is linked to the frequency of events and that information must be interpreted in terms of alternative possibilities.

For this work, a variant of the Shannon entropy is employed which has been developed in the field of natural language processing (Dumais, 1991). To calculate the entropy of job ads within an industry sector, each job ad represents a document composed of terms. Employing the wording of (Khurana and Bhatnagar, 2022), the resulting entropy H of one job ad X can best be understood as a *document entropy in term space*. It has the following formula.

$$H(X_{ij}) = \sum_{x \in X} P_{xj} * \log_2(\frac{1}{P_{xj}}) \tag{2}$$

where

$$P_{x_j} = \frac{tf_{xj}}{gf_{xi}}$$

The probability P of term x to be present in a specific job ad j is calculated from the fraction of the term frequency tf of term x in j and the global frequency gf of term x in all the job advertisements of industry i. To put it simply, each term of a job ad is weighted by its overall occurrence in the whole collection of job ads from an industry before the entropy is calculated. Hence, the whole industry the job ad originates from serves as a reference for the calculation of information content. Thus, the entropy reflects complexity differences of the same occupation between industries. This follows theoretical ideas as it is expected that, for example, the job of cleaning in the paper and printing industry has lower complexity than cleaning in the chemical industry because of a lower co-worker complementarity.

Measuring the entropy at a document level entails two further properties. First, it has been empirically demonstrated that the entropy of a document is proportional to its length (Tweedie & Baayen, 1998). Accordingly, longer job ads will have higher entropy. Second, it was shown that the entropy of documents is closely related to measures of lexical richness and textual complexity (Thoiron, 1986). Measures of textual richness and complexity are applied in linguistics to estimate the sophistication of texts. For an overview see Lu et al. (2019) and Benoit et al. (2019). The best-known measure is the type-token ratio (TTR), which is calculated from the ratio of total words to unique words in a text. From this follows that entropy captures the complexity of texts in a multidimensional way.

Because knowledge complexity likely is mirrored in job ads in several ways, entropy is preferred over single-dimensional measures like the TTR in the context of this study. However, single-dimensional measures are also calculated for validation purposes in the appendix.

The R libraries *spacyr* (Benoit & Matsuo, 2020) and *quanteda.textstats* (Benoit et al., 2021) were employed for calculating text metrics.

3.5 Estimation of β

The spatial concentration of economic activities is operationalised in terms of the distribution of job advertisements over German LMAs.

The scaling coefficient beta is estimated following standard practice in urban scaling (Bettencourt et al., 2013). The power-law function $Y_a(N) \sim Y_0 N_a^\beta$ which models the relationship between LMA size and aggregate output is reformulated as a linearized model by taking its logarithm such that

$$\log(Y_a) = \log(Y_0) + \beta \log(N_a) + \epsilon_a. \tag{3}$$

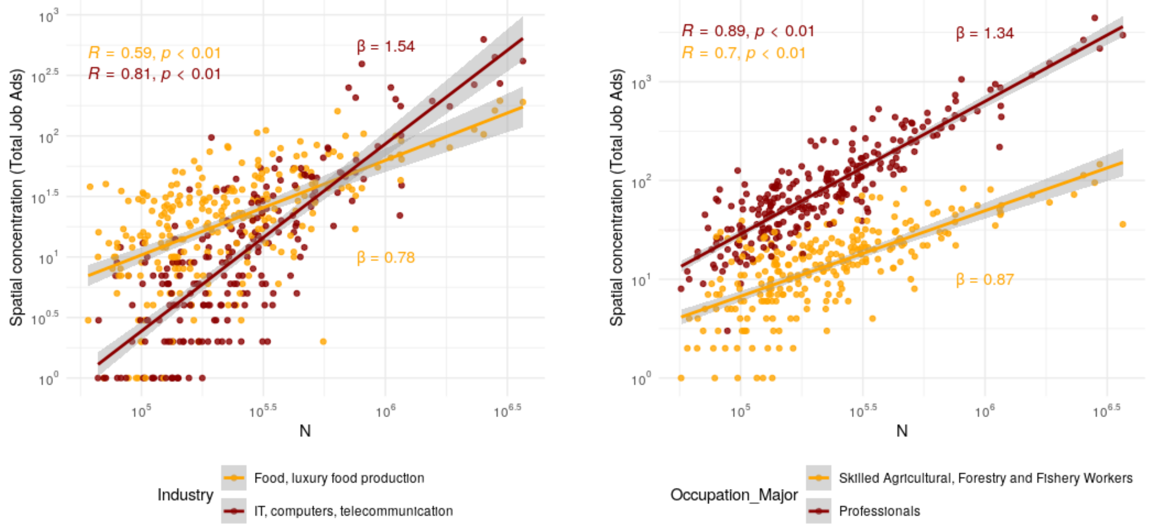
Here Y denotes the total number of job advertisements in LMA $a = 1, 2, ..., M$. N is the population size, Y_0 is the intercept and ϵ is a constant error term. Ordinary least squares regression is used to compute β, which represents the slope of the model. Superlinear scaling of Y implies $\beta > 1$. Log-log models render the coefficients unit free. This brings the advantage that the relationship between the dependent and independent variable can be interpreted in relative terms, such that a 1% change in N leads to a β% change in Y. This concept is known as elasticities in econometrics (Keuschnigg & Wolbring, 2019, p. 56).

Switching the focus of the analysis to the scaling of job complexity implies estimating β in a different model of the form

$$\log\left(\frac{Y_a}{N_a}\right) = \log(Y_0) + \beta \log(N_a) + \epsilon_a. \tag{4}$$

Now Y denotes the job complexity in LMA a. Note that here each LMA's mean job complexity $\frac{Y_a}{N_a}$ is estimated by the model instead of the total number of job advertisements as in equation 3. Using per-capita averages instead of LMA-wide totals implies superlinear scaling at $\beta > 0$ (Shalizi, 2011).

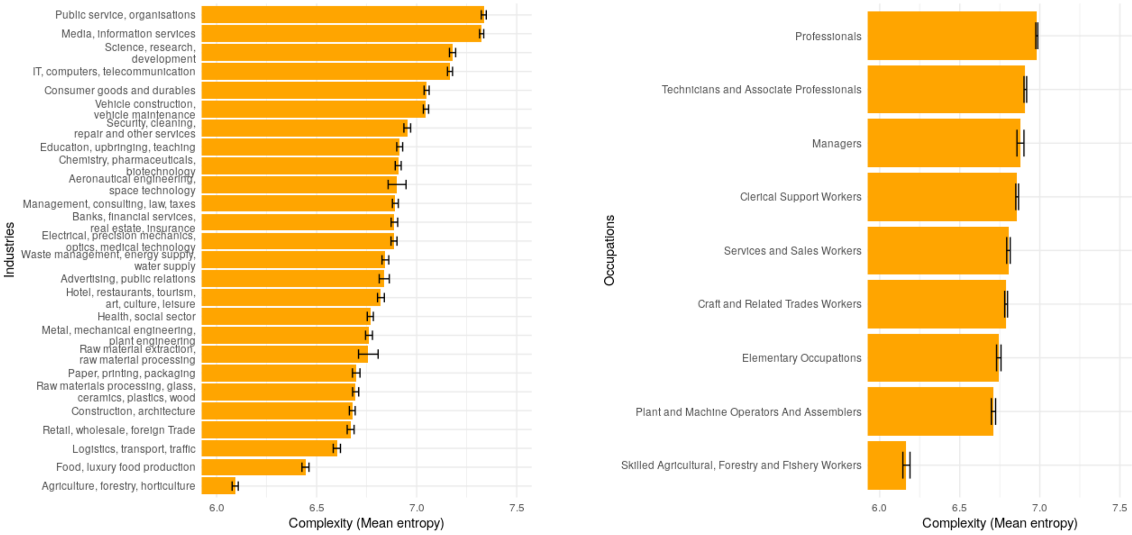
4 Results



(a) (b)

Figure 2: Spatial concentration of jobs on the German labour market. Scaling relations between the population of a LMA and the total number of job advertisements recorded in different **(a)** industries and **(b)** occupations.

Hypothesis 1a and 1b aimed to verify the assumptions that different industries and occupations exhibit varying degrees of (a) spatial concentration and (b) complexity. Figure 2 shows the relationship between city size and the total number of job advertisements for selected industries and occupations. Table 6 and 7 in the appendix list the beta coefficients for all industries and occupations respectively. In line with hypothesis 1a it becomes apparent from the data, that differences exists among industries and occupations concerning their spatial concentration. While, for example, the number of jobs offered in the IT industry scales super-linearly with city size ($\beta = 1.55 \pm 0.08$), the food production industry exhibit sub-linear scaling ($\beta = 0.78 \pm 0.07$). In general, it shows that the manufacturing industry (for example, raw material extraction ($\beta = 0.23 \pm 0.09$), agriculture ($\beta = 0.8 \pm 0.06$)) tends to be independent from city size growth. Industry sectors which centre around individuals (for example, retail ($\beta = 1.06 \pm 0.05$), waste management ($\beta = 1.0 \pm 0.05$)) display proportional scaling with city size. Tertiary, knowledge-intensive industry sectors (for example, information services ($\beta = 1.65 \pm 0.08$) and the banking sector ($\beta = 1.34 \pm 0.05$)) heavily agglomerate in cities.



(a) (b)

Figure 3: Differences in complexity among (a) industries and (b) occupations. Bootstrapped confidence intervals (10,000 re-sampling iterations) reveal significant differences in knowledge complexity (Shannon-entropy of job advertisements) among industries and occupations.

Likewise, it becomes apparent from 2b that job advertisements for professionals are over-represented in large cities ($\beta = 1.34 \pm 0.04$) while ads for skilled agricultural and fishery workers are under-represented ($\beta = 0.87 \pm 0.06$). In general, it shows that jobs with higher levels of specialisation, education and skill (as indicated by the ISCO-08 classification) demonstrate larger spatial concentrations.

Figure 3 presents the results with regard to hypothesis 1b. As can be seen in the chart, industries and occupations differ significantly in knowledge complexity. Confidence intervals are derived from estimating the distribution of the sampling means using bootstrapping with 10,000 re-sampling iterations. It shows that the tertiary and service sectors display higher levels of complexity than industries in the less knowledge-intensive production sector. Similarly, differences in complexity between occupational fields largely match the hierarchy in skill and specialisation levels in the ISCO-classification (ILO, 2012). Differences in complexity can also be observed with regard to creative and non-creative occupations (Fig. 4). Hence, the results confirm the assumptions that different industries and occupations vary in their extents of spatial concentration and display varying levels of complexity.

Figure 5 brings these findings together and shows the results of correlational analysis concerning the relationship between spatial concentration and complexity. A moderate positive correlation was found between complexity and spatial concentration of different industry sectors ($r = 0.52, p < 0.05$). A strong positive correlation was found between the complexity and the spatial concentration of occupations ($r = 0.95, p < 0.05$). Thus,

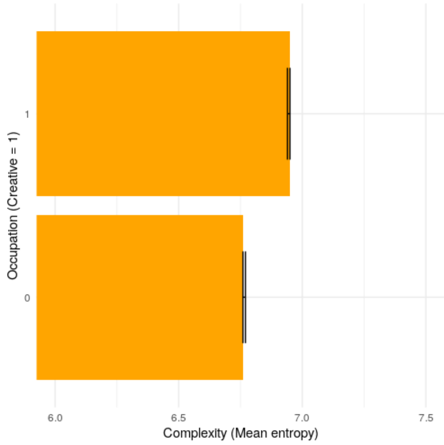


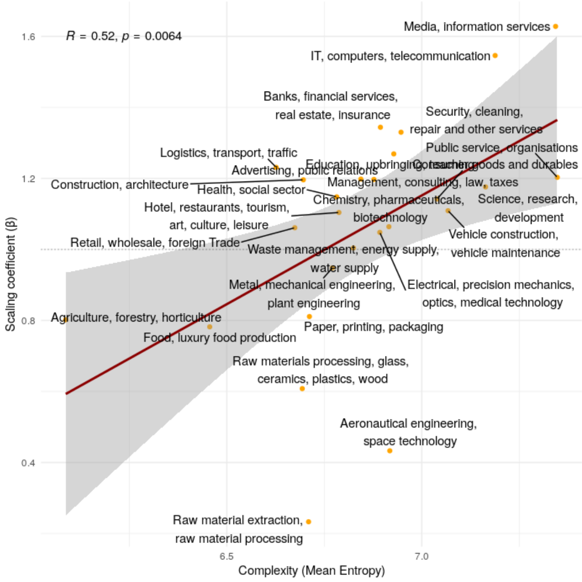
Figure 4: Differences in complexity between creative ($y = 1$) and non-creative ($y = 0$) occupations.

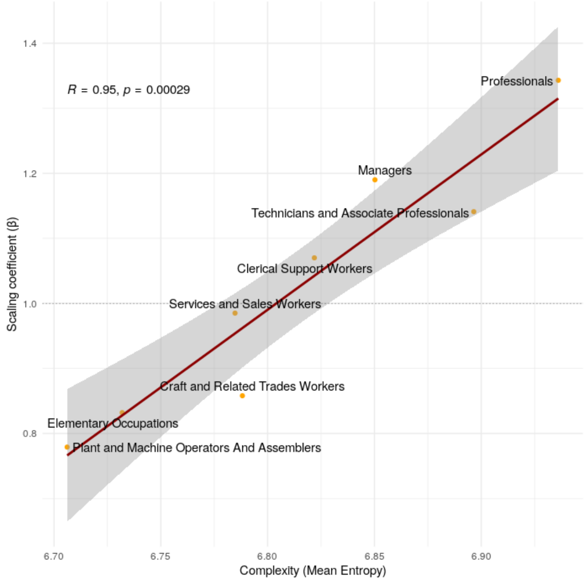
hypothesis 2 that more complex industries and occupations exhibit super-linear scaling with city size could be confirmed. In other words, more complex economic activities are over-represented in large cities.

The findings presented until now illustrate that more complex industries and occupations concentrate in large cities. Still, the question remains whether the complexity per se increases with city size. So far, it can not be excluded that economic activities gain in complexity with city size equally. As a matter of fact, it shows that the mean complexity within labour market areas exhibits super-linear scaling with LMA size ($\beta = 0.006 \pm 0.0008$, Appendix Fig. 9). Note that for analysing the spatial concentration of complexity with regard to H3, only occupational fields are examined since they are more homogeneous. Industry sectors are composed of a large number of divergent jobs which would distort the results considering complexity averages.

Table 1 lists the coefficients for the scaling relationship between complexity and city size for the different occupational fields. Figure 10 in the appendix presents the respective plots. Interestingly, it shows that complexity does not scale equally within the occupational fields. Rather, the findings demonstrate that knowledge complexity agglomerates to a greater extent for occupations which correspond to high rankings in the ISCO-classification. Also, only the estimates of more knowledge-intensive occupations are significant at $p < 0.05$. Whereas the knowledge complexity required in manual occupations like plant and machine operators and skilled agricultural workers does not change with city size, jobs of higher skill, educational and knowledge level like managers and professionals become more complex when they are located in large cities.

In summary, all hypotheses could be confirmed. Taken together it shows that complex industries and occupations tend to agglomerate in larger cities. Further, the results suggests that this pattern also translates to the complexity of jobs, although only the complexity of more knowledge-intensive jobs increases with city size.





(a) (b)

Figure 5: Spatial concentration increases with complexity. Higher spatial concentrations of jobs among **(a)** industries and **(b)** occupations relate to higher levels of knowledge complexity.

	Occupation major group	β	std.error	p.value
1	Skilled Agricultural, Forestry and Fishery Workers	-0.0006	0.0021	0.76
2	Plant and Machine Operators And Assemblers	0.0007	0.0012	0.59
3	Craft and Related Trades Workers	0.0009	0.0008	0.22
4	Technicians and Associate Professionals	0.0027	0.0006	0.00
5	Elementary Occupations	0.0040	0.0011	0.00
6	Managers	0.0045	0.0017	0.01
7	Services and Sales Workers	0.0052	0.0009	0.00
8	Clerical Support Workers	0.0054	0.0007	0.00
9	Professionals	0.0054	0.0005	0.00

Table 1: Spatial concentration of complexity within occupational fields.

5 Discussion

The main goal of the current study was to investigate the relationship between the complexity of economic activities and their spatial agglomeration. The presented results indicate that complex economic activities concentrate in large cities and thereby corroborate earlier findings of Balland et al. (2020) and Gomez-Lievano et al. (2016). In particular, it could be demonstrated that industry sectors and occupational fields of higher complexity are over-represented in large cities by analysing the distribution and complexity of job advertisements from the German labour market. The fact that prior findings were recreated using a different dataset and a new complexity metric underlines the universality of the relationship.

Crucially, this study extends earlier research by providing a micro perspective on the scaling of complexity. That is, that cities not only concentrate more complex industries and occupational fields, but also that the complexity within the same economic domains increases with city size. This suggests that identical jobs – jobs of the same occupational field – are less complex in small towns than in large cities.

The findings have several theoretical implications. Above all, the importance of cities for the emergence of complex economic activities is underlined. Bearing in mind that the most complex ventures are also the most lucrative (Hidalgo, 2021)[11], the results suggest that one reason why cities concentrate economic opportunities and welfare is that they give rise to the most profitable economic endeavours.

However, it shows that the mechanisms that advance these beneficial outcomes (complexity) in cities link to mechanism which foster inequality (spatial concentration) and not all market players can equally participate in complex economic activities. In terms of job complexity, it only makes a difference for employees in tertiary sectors whether they live in a large city or not. Manual workers gain nothing from a migration to cities while the highest ranking occupational fields in the ISCO classification can tremendously increase the complexity of their job by selective migration to agglomerated areas. Given that the occupational hierarchy of the ISCO classification correlates with occupational prestige (ISEI, Ganzeboom, 2010; Hoffmeyer-Zlotnik, 2003) also stresses that the class of individuals with the highest cultural capital can profit most from the agglomeration benefits in cities.

These implications point to important challenges regarding the future dynamics of spatial inequality. It can be expected that cities grow further largely due to their beneficial properties for the production of complex goods, which in turn sets of effects of cumulative advantage (Merton, 1968). On the one hand, the progressive concentration of welfare, capital and economic opportunities in cities means that for lower socioeconomic classes

[11] In fact, supplementary calculations show that – in the light of this work – the GDP of a LMA is well predicted by the complexity of job advertisements (Appendix Tab. 2).

it becomes compulsory to move to cities to secure income. On the other hand, upper classes will gain over-proportionately from the agglomeration benefits in cities and the expansion of complex professions. Hence, ensuring a socially just transition of from rural to urban areas must be a key political priority in the upcoming years.

In this regard, it will be interesting to observe whether the global pandemic and its impacts on the expansion of digital communication will break up the link between the production of complex products and spatial agglomeration by eliminating the need for spatial proximity of co-workers. Thus, future research should be dedicated to the question whether the results also hold when controlling for whether job advertisements offer the opportunity to work remotely.

Limitations

Because this study was limited to correlational analyses, it was not possible to determine the mechanisms that underlie the spatial concentration of complexity. Since conducting sociological research in a Weberian sense implies arriving at causal explanations, future research is needed to entangle the micro-foundations the phenomenon.

In addition, it must be questioned whether taking job advertisement texts for estimating knowledge-complexity is adequate. Yet, the complexity metric employed here demonstrates high face validity (Fig. 2, Fig. 3), proves to be highly valid in terms of criterion-relatedness (Appendix Tab. 2) and relates strongly to prior findings (Balland et al., 2020; Gomez-Lievano et al., 2016).

Data and code availability

All calculations were done using R, version 4.1.2. The code for retrieving data, reproducing figures and computing analyses is available on GitHub[12].

[12] https://github.com/baldzuhnc/BAA

References

Alabdulkareem, A., Frank, M. R., Sun, L., AlShebli, B., Hidalgo, C., & Rahwan, I. (2018). Unpacking the polarization of workplace skills. *Science Advances*, *4*(7), eaao6030. https://doi.org/10.1126/sciadv.aao6030

Anderson, K. A. (2017). Skill networks and measures of complex human capital. *Proceedings of the National Academy of Sciences*, *114*(48), 12720–12724. https://doi.org/10.1073/pnas.1706597114

Antonelli, C., Crespi, F., & Quatraro, F. (2020). Knowledge complexity and the mechanisms of knowledge generation and exploitation: The European evidence. *Research Policy*, 104081. https://doi.org/10.1016/j.respol.2020.104081

Arthur, W. B. (2021). Foundations of complexity economics. *Nature Reviews Physics*, *3*(2), 136–145. https://doi.org/10.1038/s42254-020-00273-3

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review*, *86*(3), 630–640. https://www.jstor.org/stable/2118216

Balland, P.-A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O'Clery, N., & Rigby, D. (2022). The new paradigm of economic complexity. *Research Policy*, *51*(3), 104450. https://doi.org/10.1016/j.respol.2021.104450

Balland, P.-A., Jara-Figueroa, C., Petralia, S. G., Steijn, M., Rigby, D. L., & Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour*, *4*(3), 248–254. https://doi.org/10.1038/s41562-019-0803-3

Batty, M. (2008). The size, scale, and shape of cities. *Science*, *319*(5864), 769–771. https://doi.org/10.1126/science.1151419

Becker, G. S., & Murphy, K. M. (1992). The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics*, *107*(4), 1137–1160. https://doi.org/10.2307/2118383

Benoit, K., & Matsuo, A. (2020). *Spacyr: wrapper to the 'spacy' 'nlp' library* (Version 1.2.1). https://cran.r-project.org/web/packages/spacyr

Benoit, K., Munger, K., & Spirling, A. (2019). Measuring and explaining political sophistication through textual complexity. *American Journal of Political Science*, *63*(2), 491–508. https://doi.org/10.1111/ajps.12423

Benoit, K., Watanabe, K., Wang, H., Wei Lua, J., & Kuha, J. (2021). *Quanteda.textstats: textual statistics for the quantitative analysis of textual data* (Version 0.95). https://cran.r-project.org/web/packages/quanteda.textstats

Bettencourt, L. M., Lobo, J., & Youn, H. (2013). The hypothesis of urban scaling: formalization, implications and challenges. *ArXiv*. https://doi.org/10.48550/arXiv.1301.5919

Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, *104*(17), 7301–7306. https://doi.org/10.1073/pnas.061017210

Bettencourt, L. M., Lobo, J., Strumsky, D., & West, G. B. (2010). Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PloS One*, *5*(11), e13541. https://doi.org/10.1371/journal.pone.0013541

Bettencourt, L. M., & West, G. (2010). A unified theory of urban living. *Nature*, *467*(7318), 912–913. https://doi.org/10.1038/467912a

Bivand, R., Keitt, T., & Rowlingson, B. (2022). *Rgdal: bindings for the 'geospatial' data abstraction library* (Version 1.5). https://cran.r-project.org/web/packages/rgdal

Boschma, R. A., & Fritsch, M. (2009). Creative class and regional growth: Empirical evidence from seven European countries. *Economic Geography*, *85*(4), 391–423. https://doi.org/10.1111/j.1944-8287.2009.01048.x

Bundesinstitut für Bevölkerungsforschung (BBSR). (2017). *Wachsende und schrumpfende Arbeitsmarktregionen*. Retrieved July 9, 2022, from https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/regionen/wachsende-schrumpfende-arbeitsmarktregionen/amr_wachsen-schrumpfen.html

Carbonell, P., & Rodriguez, A. I. (2006). Designing teams for speedy product development: The moderating effect of technological complexity. *Journal of Business Research*, *59*(2), 225–232. https://doi.org/10.1016/j.jbusres.2005.08.002

Coleman, J. S. (1986). Social theory, social research, and a theory of action. *American Journal of Sociology*, *91*(6), 1309–1335.

Covington, M. A., & McFall, J. D. (2010). Cutting the gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics*, *17*(2), 94–100. https://doi.org/10.1080/09296171003643098

Dumais, S. T. (1991). Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments, & Computers*, *23*(2), 229–236. https://doi.org/10.3758/BF03203370

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (pp. 2063–2117). Elsevier. https://doi.org/10.1016/S1574-0080(04)80005-1

Durkheim, E. (2005). *Suicide: A study in sociology* [Original work published 1897]. Routledge.

Eagle, N., Macy, M., & Claxton, R. (2010). Network diversity and economic development. *Science*, *328*(5981), 1029–1031. https://doi.org/10.1126/science.1186605

Florida, R. (2004). *Cities and the creative class*. Routledge. https://doi.org/10.4324/9780203997673

Ganzeboom, H. B. (2010). A new international socio-economic index (ISEI) of occupational status for the international standard classification of occupation 2008 (ISCO-08) constructed with data from the issp 2002–2007. *Annual Conference of International Social Survey Programme*, *1*.

Gomez-Lievano, A., Patterson-Lomba, O., & Hausmann, R. (2016). Explaining the prevalence, scaling and variance of urban phenomena. *Nature Human Behaviour*, *1*(1), 1–6. https://doi.org/10.1038/s41562-016-0012

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, *3*(2), 92–113. https://doi.org/10.1038/s42254-020-00275-1

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, *106*(26), 10570–10575. https://doi.org/10.1073/pnas.0900943106

Hoffmeyer-Zlotnik, J. H. P. (2003). "Stellung im Beruf" als Ersatz für eine Berufsklassifikation zur Ermittlung von sozialem Prestige. *ZUMA Nachrichten*, *27*(53), 114–127.

Informationsdienst des Instituts der deutschen Wirtschaft (IWD). (2017). *Wohlstand: Ohne die Hauptstadt meist geringer*. Retrieved July 11, 2022, from https://www.iwd.de/artikel/wohlstand-ohne-die-hauptstadt-meist-geringer-362413/

International Labour Organisation (ILO). (2012). *International standard classification of occupations (ISCO-08)*.

Jacobs, J. (2016). *The economy of cities* [Original work published 1969]. Knopf Doubleday Publishing Group.

Keuschnigg, M., Mutgan, S., & Hedström, P. (2019). Urban scaling and the regional divide. *Science Advances*, *5*(1), eaav0042. https://doi.org/10.1126/sciadv.aav0042

Keuschnigg, M., & Wolbring, T. (2019). *Wirtschaftssoziologie: Städte-Märkte-Netzwerke*. UTB.

Khurana, A., & Bhatnagar, V. (2022). Investigating entropy for extractive document summarization. *Expert Systems with Applications*, *187*, 115820. https://doi.org/10.1016/j.eswa.2021.115820

Kosfeld, R., & Werner, A. (2012). Deutsche Arbeitsmarktregionen – Neuabgrenzung nach den Kreisgebietsreformen 2007–2011. *Raumforschung und Raumordnung*, *70*(1), 49–64. https://doi.org/10.1007/s13147-011-0137-8

Lee, S. Y., Florida, R., & Acs, Z. (2004). Creativity and entrepreneurship: a regional analysis of new firm formation. *Regional Studies*, *38*(8), 879–891. https://doi.org/10.1080/0034340042000280910

Lengyel, B., & Eriksson, R. H. (2016). Co-worker networks and productivity growth in regions. *Journal of Economic Geography*, *17*(3). https://ideas.repec.org/p/egu/wpaper/1513.html

Lo Turco, A., & Maggioni, D. (2020). The knowledge and skill content of production complexity. *Research Policy*, 104059. https://doi.org/10.1016/j.respol.2020.104059

Lu, C., Bu, Y., Dong, X., Wang, J., Ding, Y., Larivière, V., Sugimoto, C. R., Paul, L., & Zhang, C. (2019). Analyzing linguistic complexity and scientific impact. *Journal of Informetrics*, *13*(3), 817–829. https://doi.org/10.1016/j.joi.2019.07.004

Marshall, A. (2013). *Principles of economics* [Original work published 1890]. Palgrave.

Melo, P. C., Graham, D. J., & Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, *39*(3), 332–342. https://doi.org/10.1016/j.regsciurbeco.2008.12.002

Merton, R. K. (1968). The matthew effect in science. *Science*, *159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

Mewes, L., & Broekel, T. (2020). Technological complexity and economic growth of regions. *Research Policy*, 104156. https://doi.org/10.1016/j.respol.2020.104156

Milgram, S. (1970). The experience of living in cities: a psychological analysis. In F. Korten, S. Cook, & J. Lacy (Eds.), *Psychology and the Problems of Society* (pp. 152–173). American Psychological Association. https://doi.org/10.1037/10042-011

Mitchell, M. (2009). *Complexity: a guided tour*. Oxford University Press.

Morris, I., & Manning, J. G. (2005). The economic sociology of the ancient mediterranean world. In N. J. Smelser & R. Swedberg (Eds.), *The Handbook of Economic Sociology* (pp. 131–160). Princeton University Press. https://doi.org/10.1515/9781400835584.131

Neffke, F. M. H. (2019). The value of complementary co-workers. *Science Advances*, *5*(12), eaax3370. https://doi.org/10.1126/sciadv.aax3370

Ooms, J. (2022). *Jsonlite: a simple and robust json parser and generator for r* (Version 1.8). Retrieved August 3, 2022, from https://cran.r-project.org/web/packages/jsonlite/

Pan, W., Ghoshal, G., Krumme, C., Cebrian, M., & Pentland, A. (2013). Urban characteristics attributable to density-driven tie formation. *Nature Communications*, *4*(1), 1961. https://doi.org/10.1038/ncomms2961

Park, J., Wood, I. B., Jing, E., Nematzadeh, A., Ghosh, S., Conover, M. D., & Ahn, Y.-Y. (2019). Global labor flow network reveals the hierarchical organization and dynamics of geo-industrial clusters. *Nature Communications*, *10*(1), 3449. https://doi.org/10.1038/s41467-019-11380-w

Pebesma, E., & Bivand, R. (2022). *Sp: classes and methods for spatial data* (Version 1.5). https://cran.r-project.org/web/packages/sp/

Pintar, N., & Scherngell, T. (2021). The complex nature of regional knowledge production: Evidence on European regions. *Research Policy*, 104170. https://doi.org/10.1016/j.respol.2020.104170

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (pp. 2119–2171). Elsevier. https://doi.org/10.1016/S1574-0080(04)80006-3

Seto, K. C., Dhakal, S., Bigio, A., Blanco, H., Delgado, G. C., Dewar, D., Huang, L., Inaba, A., Kansal, A., & Lwasa, S. (2014). Human settlements, infrastructure and spatial planning. *Climate Change 2014: Mitigation of Climate Change. IPCC Working Group III Contribution to AR5*. http://pure.iiasa.ac.at/11114

Shalizi, C. R. (2011). Scaling and hierarchy in urban economies. *ArXiv*. https://doi.org/10.48550/arXiv.1102.4101

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Simmel, G. (1950). *The sociology of Georg Simmel* (K. H. Wolff, Ed.) [Original work published 1902]. Simon; Schuster.

Smith, A. (2010). *The Wealth of Nations: An Inquiry into the Nature and Causes of the Wealth of Nations* [Original work published 1776]. Harriman House Limited.

Thoiron, P. (1986). Diversity index and entropy as measures of lexical richness. *Computers and the Humanities*, *20*(3), 197–202. https://doi.org/10.1007/BF02404461

Thurner, S., Hanel, R., & Klimek, P. (2018). *Introduction to the theory of complex systems*. Oxford University Press.

Tickamyer, A. R. (2000). Space matters! Spatial inequality in future sociology. *Contemporary Sociology*, *29*(6), 805–813. https://doi.org/10.2307/2654088

Tweedie, F. J., & Baayen, R. H. (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities*, *32*(5), 323–352. https://doi.org/10.1023/A:1001749303137

United Nations. (2019). World population prospects 2019. *Department of Economic and Social Affairs, Population Division*, *141*.

WBGU Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen. (2019). *Der Umzug der Menschheit: Die transformative Kraft der Städte*. WBGU.

Weber, M. (1958). *The city* (D. Martindale & G. Neuwirth, Eds.). New York: Free Press.

Wickham, H. (2022). *Httr: tools for working with urls and http* (Version 1.4.3). https://cran.r-project.org/web/packages/httr

Wirth, L. (1938). Urbanism as a way of life. *American Journal of Sociology*, *44*(1), 1–24.

Youn, H., Bettencourt, L. M., Lobo, J., Strumsky, D., Samaniego, H., & West, G. B. (2016). Scaling and universality in urban economic diversification. *Journal of The Royal Society Interface*, *13*(114), 20150937. https://doi.org/10.1098/rsif.2015.0937

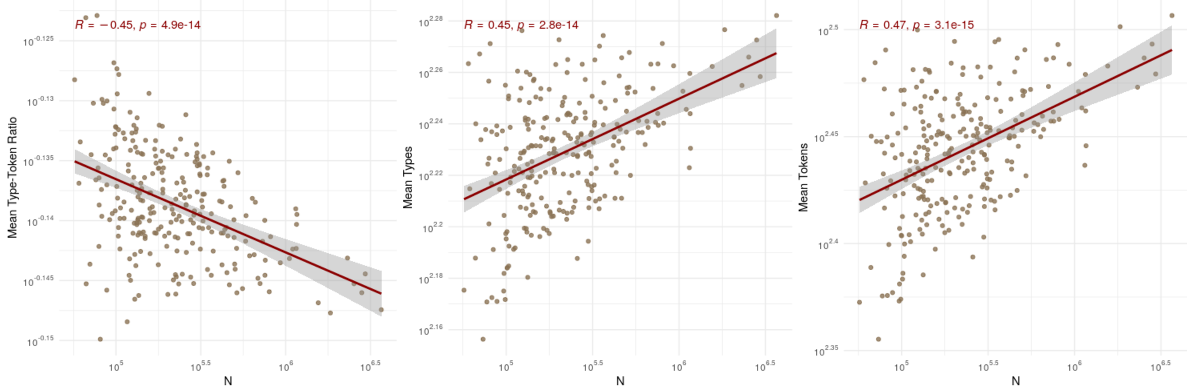
Zander, U., & Kogut, B. (1995). Knowledge and the speed of the transfer and imitation of organizational capabilities: an empirical test. *Organization Science*, *6*(1), 76–92. https://doi.org/10.1287/orsc.6.1.76

Zhu, S., & Li, R. (2017). Economic complexity, human capital and economic growth: empirical research based on cross-country panel data. *Applied Economics*, *49*(38), 3815–3828. https://doi.org//10.1080/00036846.2016.1270413

6 Appendix

6.1 Robustness

The type-token ratio is calculated for validating the complexity measure (Fig. 6). Values closer to one signal higher textual complexity. Contrary to expectations and the results reported in Figure 9, it shows that the relationship between spatial concentration and complexity changes direction when using the TTR as a metric. Yet, these findings may be explained by the fact that the TTR is very sensitive to the text length (Covington and McFall, 2010). Indeed Fig. 6b and 6c demonstrate that the text length (number of tokens) scales slightly more with population size than the types.



(a) TTR **(b)** Types **(c)** Tokens

Figure 6: Spatial concentration of complexity using the TTR.

Still, confidence in the Shannon entropy complexity measure arises from the fact that the metric is highly predictive for work productivity ($p < 0.05$). Table 2 reports the results from regressing the GDP per wage earner for 125 German LMAs on the the average complexity[13].

[13] LMA-level GDP data can be found here: https://bit.ly/3Qh4gIR

	Dependent variable:
	log(GDP per wage earner)
log(mean_entropy)	5.238^{***}
	(1.018)
Constant	-5.813^{***}
	(1.955)
Observations	125
R^2	0.177
Adjusted R^2	0.170
Residual Std. Error	0.118 (df = 123)
F Statistic	26.457^{***} (df = 1; 123)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: Regressing the average work productivity in 125 LMAs on the average job complexity.

6.2 Figures

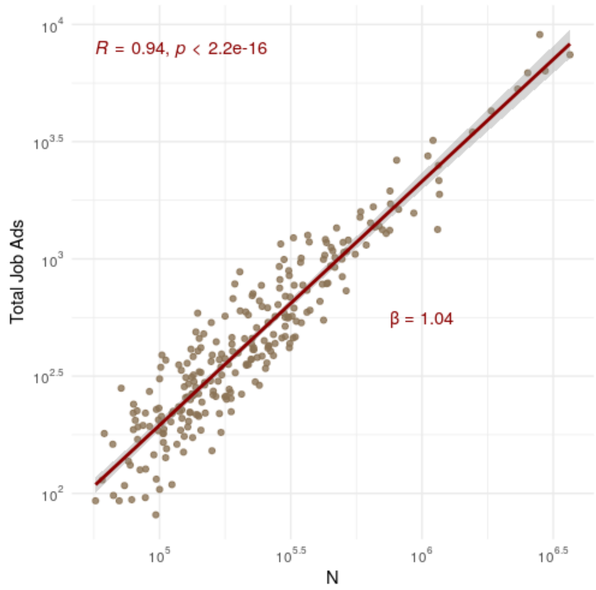


Figure 7: Spatial concentration of job advertisements. The total number of job ads scales almost proportionately with LMA size.

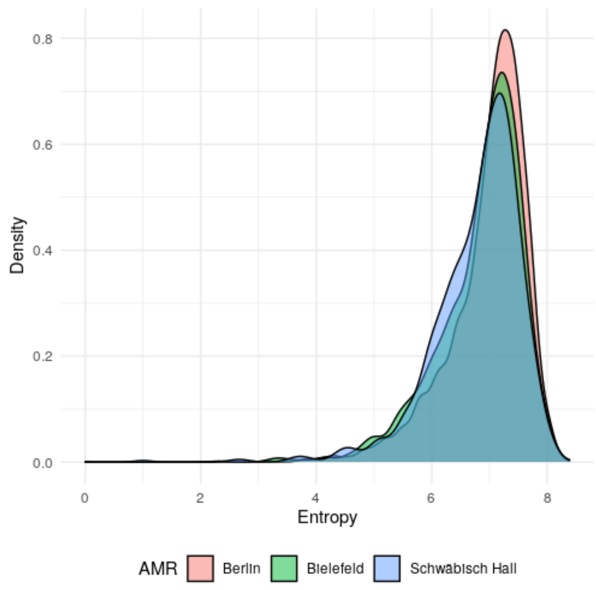
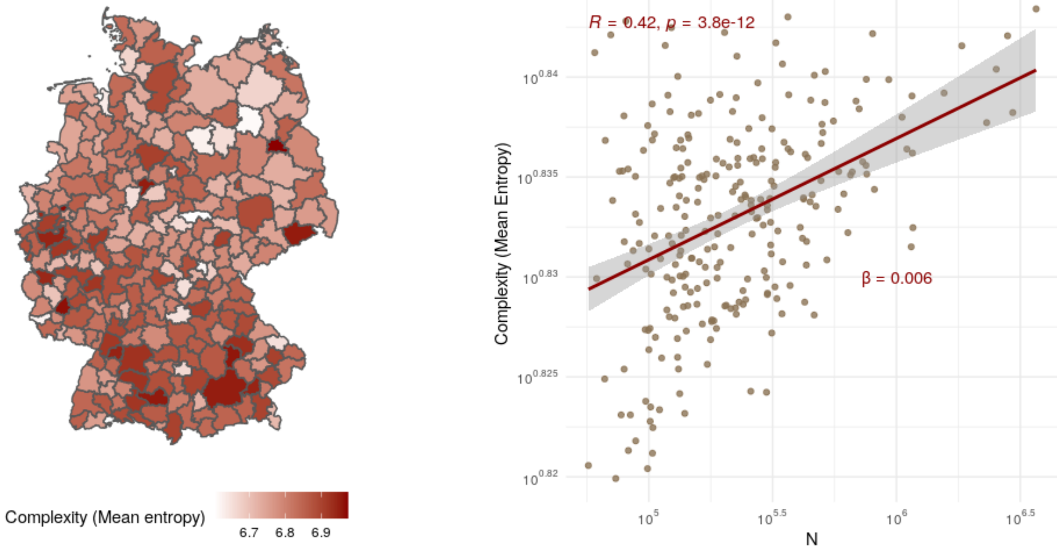


Figure 8: Distribution of the complexity measure.



(a) (b)

Figure 9: Spatial concentration of complexity on the German labour market. **(a)** Knowledge complexity (Shannon-entropy of job advertisements) agglomerates in large metropolitan areas and **(b)** scales super-linearly with LMA size.

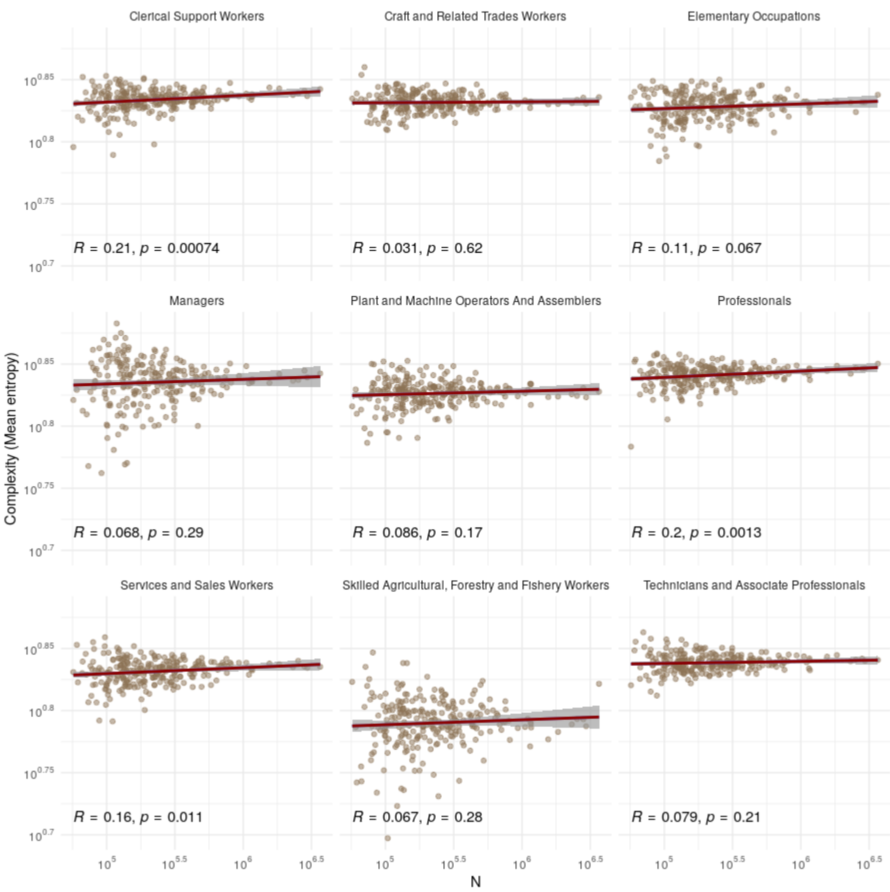


Figure 10: Spatial concentration of complexity within occupational fields.

6.3 Tables

	Industry	Date retrieved	N
1	Construction, architecture	2022-05-05 00:25	8601
2	Chemistry, pharmaceuticals, biotechnology	2022-05-05 00:48	7353
3	Electrical, precision mechanics, optics, medical technology	2022-05-05 13:00	7957
4	Education, upbringing, teaching	2022-05-05 14:35	8162
5	Vehicle construction, vehicle maintenance	2022-05-05 16:20	8495
6	Banks, financial services, real estate, insurance	2022-05-05 16:45	7267
7	Health, social sector	2022-05-05 00:10	8472
8	Raw materials processing, glass, ceramics, plastics, wood	2022-05-05 18:00	8491
9	Retail, wholesale, foreign Trade	2022-05-05 18:35	7668
10	Hotel, restaurants, tourism, art, culture, leisure	2022-05-05 19:40	8006
11	IT, computers, telecommunication	2022-05-05 11:20	7161
12	Agriculture, forestry, horticulture	2022-05-05 23:10	8684
13	Management, consulting, law, taxes	2022-05-05 23:20	8292
14	Media, information services	2022-05-06 09:15	8109
15	Metal, mechanical engineering, plant engineering	2022-05-06 09:38	7863
16	Consumer goods and durables	2022-05-06 11:40	8655
17	Food, luxury food production	2022-05-06 00:17	7687
18	Public service, organisations	2022-05-06 00:47	8721
19	Paper, printing, packaging	2022-05-09 10:17	4508
20	Raw material extraction, raw material processing	2022-05-09 10:19	879
21	Logistics, transport, traffic	2022-05-09 10:30	6560
22	Waste management, energy supply, water supply	2022-05-09 10:39	7281
23	Security, cleaning, repair and other services	2022-05-05 11:45	7794
24	Advertising, public relations	2022-05-09 10:42	3189
25	Science, research, development	2022-05-09 10:47	4333
26	Aeronautical engineering, space technology	2022-05-09 10:48	531

Table 3: Number of cases and retrieval dates. Total $N = 180,719$.

	Fields	Values
1	ID	6027
2	Titel	Berufskraftfahrer/innen CE (m,w,d)
3	Beruf	Berufskraftfahrer/in
4	Veroeffentlicht	2022-05-03
5	Branche	Einzelhandel, Großhandel, Außenhandel
6	Angebotsart	1
7	AnzahloffeneStellen	1
8	Arbeitgeber	Baustoff Dietrich GmbH und Co. KG
9	Region	Hessen
10	Ort	Kassel, Hessen
11	PLZ	34127
12	Lat	51.32329
13	Lon	9.493716
14	Befristung	2
15	Uebernahme	FALSE
16	Verguetung	16,00 €
17	Tarifvertrag	
18	Arbeitszeit	40 Wochenstunden Mo.-Fr. von 7:00-17:00 Uhr in der Regel bis 16:00 Uhr
19	Stellenbeschreibung	Kraftfahrer/innen (m, w, d) Anforderungen: - Führerscheinklasse CE mit gültigen Modulen - Deutschkenntnisse in Wort und Schrift - Anhängererfahrung/Drehschemel Anhänger - Kran wird von uns geschult bzw. externer Führerschein übernommen. Wir bieten: - geregelte Arbeitszeiten Mo.-Fr. (7:00-17 Uhr) 40 Std. Woche - keine Schicht-oder Wochenendarbeit - 30 Tage Urlaub - Gewinnbeteiligung - pünktliche und leistungsgerechte/übertarifliche Bezahlung - Spesen für Kraftfahrer - Nahverkehr - Berufsbekleidung - unbefristete Arbeitsverträge
20	FuerGefluechtete	FALSE
21	NurFuerSchwerbehinderte	FALSE
22	Staerken	Flexibilität, Belastbarkeit, Zuverlässigkeit, Selbständiges Arbeiten, Motivation/ Leistungsbereitschaft
23	Sprachkentnisse	NA, NA
24	Fuehrerschein	

Table 4: Example of a job advertisement.

Occupation	ISCO-88	ISCO-08
Creative Core		
Physicists, chemists, and related professionals	211	211; 226
Mathematicians, statisticians, and related professionals	212	212
Computing professionals	21	251; 252
Architects, engineers, and related professionals	214	214; 215; 216
Life science professionals	221	221; 213; 225
Health professionals (except nursing)	222	221; 225; 226
College, university, and higher education teaching professionals	231	231; 232
Secondary education teaching professionals	232	232
Primary and preprimary education teaching professionals	233	234
Special-education teaching professionals	234	235
Other teaching professionals	235	235
Archivists, librarians, and related information professionals	243	262
Social sciences and related professionals	244	263
Public service administrative professionals	247	335
Creative Professionals		
Legislators, senior officials, and managers	1	1
Nursing and midwifery professionals	223	223
Business professionals	241	241; 226; 242; 243; 333
Legal professionals	242	261
Physical and engineering science associate professionals	31	31; 35
Life science and health associate professionals	32	32; 226; 223
Finance and sales associate professionals	341	331; 332; 333
Business services agents and trade brokers	342	332; 333
Administrative associate professionals	343	331; 333; 334; 335
Police inspectors and detectives	345	335; 341
Social work associate professionals	346	341
Bohemians		
Writers and creative or performing artists	245	243; 264; 216; 265; 263
Photographers and image and sound recording equipment operators	3131	3431
Artistic, entertainment, and sports associate professionals	347	216; 343; 264; 265; 343; 342
Fashion and other models	521	5241

Table 5: Correspondence between creative occupations in the ISCO-88 and ISCO-08 classification scheme. The categorisation of creative occupations follows Boschma and Fritsch (2009).

	Industry	β	std.error	p.value
1	Raw material extraction, raw material processing	0.23	0.09	0.01
2	Aeronautical engineering, space technology	0.43	0.17	0.02
3	Raw materials processing, glass, ceramics, plastics, wood	0.61	0.07	0.00
4	Food, luxury food production	0.78	0.07	0.00
5	Agriculture, forestry, horticulture	0.80	0.05	0.00
6	Paper, printing, packaging	0.81	0.07	0.00
7	Metal, mechanical engineering, plant engineering	0.95	0.07	0.00
8	Waste management, energy supply, water supply	1.00	0.05	0.00
9	Electrical, precision mechanics, optics, medical technology	1.05	0.07	0.00
10	Retail, wholesale, foreign Trade	1.06	0.05	0.00
11	Chemistry, pharmaceuticals, biotechnology	1.06	0.07	0.00
12	Hotel, restaurants, tourism, art, culture, leisure	1.10	0.07	0.00
13	Vehicle construction, vehicle maintenance	1.11	0.06	0.00
14	Consumer goods and durables	1.14	0.07	0.00
15	Health, social sector	1.15	0.05	0.00
16	Science, research, development	1.18	0.09	0.00
17	Construction, architecture	1.20	0.07	0.00
18	Management, consulting, law, taxes	1.20	0.05	0.00
19	Advertising, public relations	1.20	0.06	0.00
20	Public service, organisations	1.20	0.05	0.00
21	Logistics, transport, traffic	1.23	0.07	0.00
22	Education, upbringing, teaching	1.27	0.06	0.00
23	Security, cleaning, repair and other services	1.33	0.05	0.00
24	Banks, financial services, real estate, insurance	1.34	0.05	0.00
25	IT, computers, telecommunication	1.55	0.08	0.00
26	Media, information services	1.63	0.08	0.00

Table 6: Spatial concentration of industry sectors.

	Occupation	β	std.error	p.value
1	Plant and Machine Operators And Assemblers	0.78	0.04	0.00
2	Elementary Occupations	0.83	0.03	0.00
3	Craft and Related Trades Workers	0.86	0.04	0.00
4	Skilled Agricultural, Forestry and Fishery Workers	0.87	0.06	0.00
5	Services and Sales Workers	0.98	0.03	0.00
6	Clerical Support Workers	1.07	0.03	0.00
7	Technicians and Associate Professionals	1.14	0.03	0.00
8	Managers	1.19	0.05	0.00
9	Professionals	1.34	0.04	0.00

Table 7: Spatial concentration of occupations.

Eidesstaatliche Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, alle Ausführungen, die anderen Schriften wörtlich oder sinngemäß entnommen wurden, kenntlich gemacht sind und die Arbeit in gleicher oder ähnlicher Fassung noch nicht Bestandteil einer Studien- oder Prüfungsleistung war. Ebenfalls versichere ich, dass die elektronische und die gedruckte Fassung dieser Arbeit übereinstimmen.

Leipzig, den 8. August 2022 Clemens Baldzuhn

Ort, Datum Clemens Baldzuhn