Breithaupt, Patrick; Hottenrott, Hanna; Rammer, Christian; Römer, Konstantin

Working Paper

Mapping employee mobility and employer networks using professional network data

ZEW Discussion Papers, No. 23-041

Provided in Cooperation with:
ZEW - Leibniz Centre for European Economic Research

Suggested Citation: Breithaupt, Patrick; Hottenrott, Hanna; Rammer, Christian; Römer, Konstantin (2023) : Mapping employee mobility and employer networks using professional network data, ZEW Discussion Papers, No. 23-041, ZEW - Leibniz-Zentrum für Europäische Wirtschaftsforschung, Mannheim

This Version is available at:
https://hdl.handle.net/10419/279575

// NO.23-041 | 10/2023

DISCUSSION PAPER

// PATRICK BREITHAUPT, HANNA HOTTENROTT,
CHRISTIAN RAMMER, AND KONSTANTIN RÖMER

Mapping Employee Mobility and Employer Networks Using Professional Network Data

Leibniz Association

ZEW

Mapping Employee Mobility and Employer Networks using Professional Network Data

Patrick Breithaupt[1,2,*], Hanna Hottenrott[1,3], Christian Rammer[1] and Konstantin Römer[3]

[1] ZEW – Leibniz Centre for European Economic Research, L7 1, 68161 Mannheim, Germany
[2] Justus-Liebig-University Giessen, Faculty of Economics, Licher Straße 64, 35394 Gießen, Germany
[3] Technical University of Munich, TUM School of Management, Dept. of Economics & Policy, 80333 Munich, Germany

This version: September 15, 2023

Abstract

The availability of social media data is growing and represents a new data source for economic research. This paper presents a detailed study on the use of data from a career-oriented social networking platform for measuring employee flows and employer networks. The employment data are exported from user profiles and linked to the Mannheim Enterprise Panel (MUP). The linked employer-employee (LEE) data consists of 14 million employments for 1.5 million employers. The platform-based LEE data is used to create annual employer networks comprised of data from 9 million employee flows. Plausibility checks confirm that career-oriented social networking data contain valuable data about employment, employee flows, and employer networks. Using such data provides opportunities for research on employee mobility, networks, and local ecosystems' role in economic performance at the employer and the regional level.

Keywords: social networks, platform data, lee data, labour mobility, network analysis
JEL-Classification: C81, J60, L14

* *Corresponding author.* Full address: Patrick Breithaupt, Digital Economy Department, ZEW Mannheim, P.O. Box 103443, 68034 Mannheim, Germany. E-mail: patrick.breithaupt@zew.de.

Acknowledgments: The authors would like to thank the German Federal Ministry of Education and Research (BMBF) for providing funding for the research project (Networks of Innovative Firms (NETINU); funding identifiers: 16|FI105 and 16|FI006). We thank Sandra Gottschalk, Irene Bertschek, Thorsten Doherr, Thomas Niebel, Jan Kinne, Julian Castaldi, Janna Axenbeck, and Julian Dörr for their valuable input. Special thanks are owed to New Work SE (XING) for providing the data and cooperation. The research was conducted independently of New Work SE. All remaining errors are ours.

1 Introduction

The growing use of social media provides new sources of data for research purposes and the development of new economic indicators. Prominent examples of such sources are career-oriented social media platforms like LinkedIn or XING[1]. Career-oriented platforms are a natural candidate for the generation of large-scale employment indicators as such platforms - through network effects - attract many actors. There are incentives on both the users' and providers' sides to grow the platform through collecting and assembling data. Successful platforms are, therefore, a rich source of data on agents active on the platform. In the case of the career-oriented platform XING, these agents are, on the one hand, employees or job seekers and, on the other hand, employers such as companies, research institutes or public administration.

The data available on such platforms allows for identifying employer-employee relationships over time and space, hence allowing for tracking individuals' mobility from one employment to another. It is important to note that employment types listed on such platforms are not limited to those with social security contributions but also capture unpaid, freelance, and entrepreneurial activities, typically unobserved in administrative employer-employee data. Further, the employment and employee mobility data are available immediately after adding it to the user profile.[2] Thus, there is no time lag in data provision, as is typically the case with administrative data. However, outdated information may also be contained in the data. Furthermore, administrative data is often subject to stringent legal requirements and may not, for example, be linked to all types of third-party data. So far, it remains unclear whether data extracted from social networks is sufficiently representative for research purposes.

This study aims to assess the usefulness of career-oriented social media data for mapping and tracking employee mobility (here, also including non-conventional types of employment). Moreover, we explore the usefulness and plausibility of the employee flow data between employers by analysing the resulting networks. Measuring networks between employers through labour mobility is vital in innovation research (Balsvik 2011; Görg and Strobl 2005; Hottenrott and Lopes-Bento 2016). However, such networks are typically constructed from linked administrative employer-employee data (Collet and Hedström 2013; Kaiser et al. 2015; Maliranta et al. 2009), patent data, i.e. measuring inventor mobility (Rahko 2017; Somaya et al. 2008; van der Wouden and Rigby 2021), or data on

[1] LinkedIn and XING are both employment-focused social media platforms. The former was launched in 2002 and is owned by Microsoft. XING is operated by New Work SE and was founded in August 2003. These platforms entail user and employer profile data managed by the users or employer representatives. XING is particularly popular in German-speaking countries.

[2] User profiles represent the online profile of employees on the platform XING.

scientific publications, i.e. capturing author mobility (Edler et al. 2011; Franzoni et al. 2014). In this study, we show that social network data presents a valuable data source for exploring networks between employers resulting from employee flows. Exploring those network data is valuable for many research applications, particularly for research on the performance of employers or regions (Giuliani 2011; Ozman 2009; Schilling and Phelps 2007). The approach has the potential to augment data collected via surveys. While survey data are generally not well suited for mapping networks due to incomplete coverage and non-response, combining networks generated from a big-data source with survey data enriches the data portfolio and hence the scope of addressable research questions.

The data preparation consists of multiple consecutive steps: First, we disambiguate employers listed in publicly accessible employments.[3] The data originates from the user profiles stored in the data warehouse of the platform XING. We link and classify employers using the names and addresses (including employers, research institutes, public administration, (non-) governmental institutions, etc.) to identify these employers in the Mannheim Enterprise Panel (MUP) and other data sources such as the Mannheim Innovation Panel (MIP). The total number of available employments is about 46M.[4] We create a LEE data set by matching around 1.5M employers to 14M publicly accessible employments.[5] Second, we calculate employee flows based on the matched employments. We create an employee flow for each employee moving from one employer to another. As a result, we extract 9M employee flows between employers or into/out of employment, e.g., students entering the labour market or retirements out of the labour market. Third, the flow data are used to create annual flow networks that cover the period from 2010 to 2020. For the series of networks, we calculate a wide range of network measures, i.e., cliques, transitivity, reciprocity, and density. The resulting database contains high-quality employment, employer, employer flow, and annual network data.[6]

Next, we check the plausibility and representativeness of the data. For this purpose, we use MUP data, which covers almost all businesses in Germany[7] and a share of other organisations such as universities, research institutes, hospitals, and non-profit organisations. We compare the MUP employer data with the XING data. The results suggest that the coverage of the employers contained in the data is sufficiently representative of all employers in Germany, regarding age, size, sector, legal form and region. Moreover,

[3] In simplified terms, employment is a tuple that consists of one user/employee and one employer.

[4] We use the abbreviations 'K' for thousand and 'M' for million, e.g., 2M instead of 2 million.

[5] About two-thirds of the employments are not considered because of a restriction to high-quality matches to the MUP and a limitation to publicly accessible user profiles.

[6] For legal reasons, we decided against creating and analysing an employee data set.

[7] However, we mark some sectors with missing/others according to their NACE classification. This concerns, for example, agriculture (NACE A), private households (NACE T) and offshore organisations and bodies (NACE U).

it captures a significant share of employers in Germany. It is important to note that young employers are to some extent underrepresented. This is not surprising since newly founded employers have few employees and are less represented on employee platforms.

We analyse the employments regarding experience, discipline, career stage, and type, as well as the average length of stay. The matched employers are investigated concerning employer size, age, sector and region. In both cases, the distribution patterns considering the expected distribution are plausible. Furthermore, we test the flow data for validity by analysing the career level, employment discipline, employment type, employer size, employer sector, and employer region before and after the employment change. E.g., most people switch employments within a discipline and typically move upwards on the career ladder and towards metropolitan areas. Thus, the mobility patterns across employers and regions are plausible. Lastly, we analyse the network data through graph metrics such as the number of nodes, edges, and cliques. As an additional check, we visualise local networks.

Finally, since research shows a positive link between knowledge exchange through employee mobility and employer performance (Abbasiharofteh et al. 2021; Almeida and Kogut 1999; Godart et al. 2014; Wu et al. 2018), we also test the plausibility of selected network measures against employer data. The analyses yield promising results, e.g., changes in the degree centrality of employers are positively linked with changes in employment counts. In summary, with minor limitations, the data represents a valuable novel research data source for studying the role of employee mobility and employer networks. However, the coverage goes beyond paid employment, including internships, freelance work, and entrepreneurial activities. For network analyses, the coverage is sufficiently high and network measures can be derived for employers that are neither active in patenting nor engaged in larger, visible alliances.

The remainder of this paper is structured as follows. In Section 2, we relate our work to the literature and Section 3 presents the data processing. Section 4 describes the resulting data sets and Section 5 discusses the findings and concludes.

2 Related literature

With the broader adoption of the internet, the conceptualisation and development of real-time economic indicators became increasingly popular. Choi and Varian (2012), for instance, use Google trends for forecasting near real-time indicators for measuring queries about unemployment claims as an indicator of economic activity. They demonstrate the viability of online search queries as indicators, leading the way to novel economic indica-

tors from data other than Google trends. Allcott et al. (2019) adopted a similar approach and analysed the magnitude of the fake-news problem on Facebook. In difference to Twitter, they find that the magnitude of the problem on Facebook had declined over time, but the spread of fake-news was still predictable. Real-time indicators also provide new opportunities for innovation research, augmenting the portfolio of traditional innovation indicators. In traditional innovation studies, the data collection often relies on publicly available data such as surveys or statistical data provided by the government (e.g., Rammer, Doherr, et al. 2021). For example, innovation studies often rely on accounting or company survey data on expenditures for innovation or patent (application) counts collected from patent office databases. While innovation surveys like the Community Innovation Survey by the European Union have substantially deepened and improved our understanding of innovation activities (Hong et al. 2012), a key disadvantage of surveys are the cost of data collection, time lag, and problem of data availability for a sample of individuals or companies. The first two factors often limit reaching a sufficiently large sample, as asking thousands of companies directly costs time and money (Rammer and Es-Sadki 2022). In addition, the national statistical offices often limit access to raw data and only publish summary statistics or reports.

One approach to tackle these issues is to use available web data provided by employers or employees. While scraping (employer) web data and processing retrieved data takes time, and the quality of the available data differs widely from employer to employer, web data has been shown to offer valuable data. For instance, Gök et al. (2015) construct a relatively accurate web-based R&D indicator. Kinne and Lenz (2021) extended the approach and illustrated that company websites can be used to predict an innovation probability. In particular, the study uses survey data as a training sample to predict innovation activities for all German employers with a website. The authors illustrate the value of web mining and extend the previous approaches through deep learning, resulting in reliable innovation indicators available for employers that do not participate in innovation surveys or do not patent. In line with these ideas, Axenbeck and Breithaupt (2021) uses the web-mining approach to identify website characteristics predicting company-level product and process innovation activity, and Schwierzy et al. (2022) show that website data can be used for the mapping of specific technologies such as additive manufacturing. Axenbeck and Breithaupt (2022) show that employers' website data can be used to measure innovations and other activities, such as those related to digitisation. Other publications use, for example, employer-related data from the social media platforms Twitter, Kununu[8], and Facebook (e.g., Breithaupt et al. (2020) and Veltri (2013)).

[8]https://www.kununu.com/de

The objective of this paper is to contribute to the recent developments in web-based indicators by analysing a little-explored source of large-scale data on employee and employer activities. In particular, we test the use of data from a career-oriented social media platform for mapping the flows of individuals (knowledge transfer) between employers.

To do so, we build on graph theory. Graph theory is a branch of discrete mathematics and theoretical computer science (Diestel 2005). It builds upon Leonhard Euler and the famous 'Seven Bridges of Königsberg' problem. Graph theory is related to social network analysis (SNA) and is frequently used in the innovation literature (Abbasiharofteh et al. 2021; Axenbeck and Breithaupt 2021). While Wasserman and Faust (1994) and Scott (2017) provide a more theoretical view on SNA, this paper focuses on the application of SNA methods by analysing employee flows between employers and resulting employer networks. Furthermore, we aim to account for characteristics of the employee flows, such as the duration of the employments and the career level of the employees.

In addition, we contribute to the 'Linked Employer Employee' (LEE) literature in which employees' individual data is linked with data on their employers. Our contribution is the creation of a LEE database from non-official data, i.e., the combination of web and proprietary data. Our LEE data has the advantage of using up-to-date data and being updated quickly. Furthermore, it includes employees who immigrate from abroad, migrate abroad or enter the labour market from education into account. However, in contrast to the official LEE data, we have only limited data on employees, such as wages, and employers, such as financial indicators. Moreover, we do not have complete coverage of all employers and employees in Germany, and historical data might be sparse. Multiple LEE data sets are already created from official data for Germany. For example, the SOEPP-LEE extends the SOEP data by linking the employees' individual data with data on their employers (Weinhardt (2016), Weinhardt, Meyermann, et al. (2016), and Weinhardt, Meyermann, et al. (2017)). The Socio-Economic Panel (SOEP) is a large, long-running multidisciplinary longitudinal study in Germany. Other German IAB data, such as WeLL-ADIAB and LIAB, provide, for example, historically linked employer and employee data (Heining et al. (2016) and Schmucker et al. (2014)). Furthermore, the LEEP-B3 and linked ALLBUS data sets are also available in Germany (Abendroth et al. (2014) and Gerhards et al. (2010)). Lastly, there are LEE datasets for many more countries, e.g., 'US Worker Establishment Characteristics Database', the 'New Zealand's Linked Employer–Employee Database', and the 'Norwegian Linked Employer–Employee Database', as well as for the European region[9] (Jensen (2010)).

[9]'European Structure of Earnings Survey': https://ec.europa.eu/eurostat/web/microdata/structure-of-earnings-survey.

3 Data Processing

For our data analysis, we connect two data sources: First, we use data from the social and professional network XING. It provides detailed information on users – mostly professionals – who create profiles on the platform primarily for professional networking. The profiles comprise personal, employment-related data as well as data about the employer. The data access was granted in close cooperation with the platform provider New Work SE. Second, we use data from the Mannheim Enterprise Panel (MUP), which provides data about the population of registered businesses in Germany[10]. The data are maintained in collaboration with Creditreform, Germany's largest credit rating agency (Bersch et al. 2014). The MUP provides employer-level data that, besides others, contains addresses, employee counts, founding dates, and website URLs. We can combine both data sources, i.e., we link data about employees and their mobility between employers from XING to employer data from the MUP. In this project, we are interested in the aggregate flows of employees between employers and not individual users. Fortunately, we are allowed to use aggregate data on the career level, employment type, employment volume, length of stay in one employment (position & employer), and professional experience. The remainder of this section describes the data processing of employments and employers (Section 3.1), the derived employee flows (Section 3.2) and annual employer networks (Section 3.3).

3.1 Employers and Employments

In the first step, employer and user employment data are exported from the data warehouse of XING. This involves about 1.9M employers, whereby not every observation has to be a valid or active employer. These observations are not directly excluded, as we are also interested in historical data. XING users have deposited about 46M employment data points. About one-third of the employments are linked to the XING employer database.[11] The employment data is partly maintained by the users and includes the employer name, employer URL, industry, employment type, career level, and field of activity (discipline). Some of the fields are optional, e.g., the employer URL. Thus, we also export employment data points with unclear quality. XING's employer data show specific features: First, the employer database is not standardised or linked to a uniform database like the MUP. Second, outdated, duplicate or invalid/fake employers are listed. Examples include insolvent employers whose XING profiles are later deleted. Third, operating sites, subsidiaries, and employer groups can have their own XING profiles. Fourth, some

[10] We use the term employers instead of companies because the MUP includes not only companies but also a subset of public institutions, universities, research institutes, and nonprofit organisations.

[11] An internal database by XING that lists employers. Not all employers are linked to the database. These employers are only mentioned in the employment data.

employers are not from Germany or the DACH region[12]. To provide some examples of the difference between valid German and available employers: The employer-level country data on XING is available for around 1.1M employers, and, of these, 450K have entered a country other than Germany. Furthermore, about 650K of the 1.9M employers have an invalid employer name, e.g., the names consist exclusively of dots and hyphens.

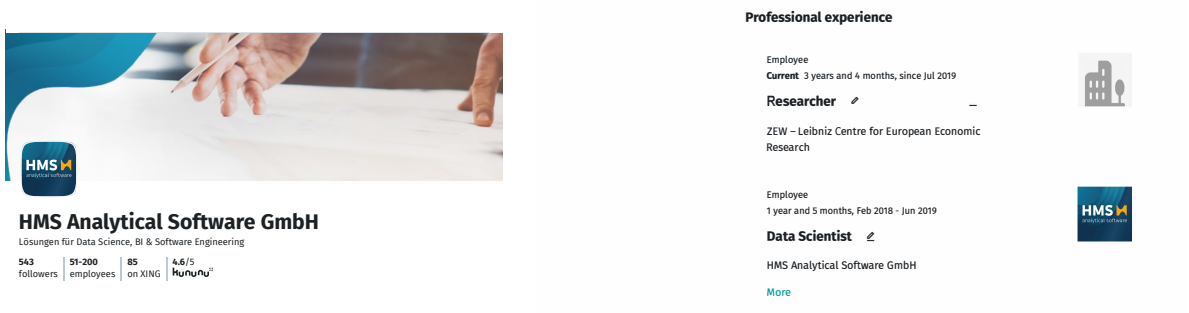


Figure 1: Example for a XING employer profile (left) and a 'professional experience' timeline for a platform user (right). Source: The images were taken from the platform XING (www.xing.com).

Figure 1 (left) shows the XING profile of an employer. An employer size group indicates the number of employees within the respective group (51-200 employees). Figure 1 (right) shows the work experience of a XING user. The user has had two employments. He has worked first as a Data Scientist and then as a researcher. The change of employment took place seamlessly in 2019 and is illustrated by the end date of the first employment and the start date of the second. The employer 'HMS Analytical Software GmbH' maintains a XING profile. However, the employer 'ZEW - Leibniz Centre for European Economic Research' does not. For example, the employer logo is missing (not a sufficient condition).

The data preparation consists of two consecutive steps. First, the employers (1.9M) and employments (46M) need to be linked to the Mannheim Enterprise Panel (MUP). This is done with the SearchEngine tool[13] developed by Thorsten Doherr. The linking relies on the text fields 'employer name' or 'employer URL', if available. Employer profiles and employment data are separately linked with the tool, as their quality is different. We assume that the data from the employer profiles has a higher quality than the employment data. As a result, we receive candidates from the Mannheim Enterprise Panel for the XING employer profiles and employers mentioned in employments on XING. Each candidate has a unique identifier called 'crefo' that represents the employer identifier of the MUP. There are 86M candidates for the employment data points based on employer name and 12M based on the URL.[14] Furthermore, about 187K employer profiles are linked

[12]DACH region: Germany, Austria, Switzerland.
[13]The GitHub project is available at https://github.com/ThorstenDoherr/searchengine (Doherr 2023).
[14]A XING company may have multiple candidates (potential matches) from the MUP database.

to MUP employers. Second, the MUP candidates for the employments are enriched with additional employer-level data like exit dates, if available. Some outliers are removed, e.g., implausible data like employments before 1900 and after 2020 (after the data export), leading to a subset of 44M employments. Then, we apply the 'group-crefo', which combines affiliated employers within the MUP, e.g., subsidiaries. Lastly, binary indicators are created, which, for example, indicate if a 'crefo' exists in the Markus database[15].

Table 1: Processing Steps for Employment Data

Step	Description	Data
1	Concatenate match candidates (MUP) for employments based on URLs and employer names.	Input: 86M + 12M = 98M candidates.
2	Drop duplicates based on unique XING 'employ ment id' and 'crefo'; select unambiguous matches.	11M matches are selected and 83M are left.
3	Use candidates with 'exist' or 'missing' exit status and select unambiguous matches.	6M matches are selected and 58M are left.
4	Use candidates existent in 'markus' data base and select unambiguous matches.	2M matches are selected and 38M are left.
5	Use candidates with highest fuzzy-matching score (employer name) and select unambiguous matches.	1M matches are selected and 28M are left.
6	Concatenate with disambiguated employer profile matches. Prefer employer profile matches.	21M employment matches.
7	Select matches for a subset of public employments.	Result: 14M employment matches (i.e., the employer listed in employment is linked to the MUP).

Multi-stage selection of the best employment matches to the Mannheim Enterprise Panel (MUP). Includes the observation counts of input and after processing in million employments (M).

Next, a five-step heuristic is used to select the best candidate from the MUP for each employment data point (see Table 1). As a result, about 21M employments are matched to the MUP. Thus, the matching rate is about 47 percent. However, we deliberately extracted fewer matches than possible to ensure higher quality, as there is a trade-off between the observation count and data quality of the matches. For legal reasons, all employments from users without a public profile must be removed. This reduces the number of mapped employments to 14M.[16] We do not delete the 10M unmatched employments of users with public profiles but assign an artificial employer identifier based on the employer name.

If users list the same employer name in their employment history and the respective employments, have not been linked to the MUP. The employers will receive the same artificial identifier. In our hypothetical example, we assume that the employers 'HMS' and 'IW' could not be matched to the MUP (see Table 2). Each row corresponds to

[15]https://www.bvdinfo.com/en-gb/our-products/data/national/markus.

[16]We export a list of all publicly available user profiles (privacy setting) to ensure that we use only publicly accessible employments as this is a legal requirement by New Work SE. Public employments can be viewed and saved by every visitor of XING. Each employment data point has a user reference, which is used to remove non-public data. Around 31M of 46M employment data points are publicly available.

Table 2: Example List of Matched Employments

User	Employer-Name	Start-Year	End-Year	Matched	Employer identifier
0	zew mannheim	2010	2015	Yes	crefo: 2344
0	DIW	2015	Missing	Yes	crefo: 9988
1	ZEW Gmbh	2010	2012	Yes	crefo: 2344
1	IW	2012	Missing	No	artificial: 1
2	ZEW - Mannheim	2009	2015	Yes	crefo: 2344
2	IFO	2015	2020	Yes	crefo: 2885
2	DICE	2020	2021	Yes	crefo: 2367
3	IW Koeln	1980	2010	Yes	crefo: 7781
3	IAB	2010	2015	Yes	crefo: 4887
3	DIW	2015	Missing	Yes	crefo: 9988
4	zew mannheim	2015	2018	Yes	crefo: 2344
4	HMS	2018	Missing	No	artificial: 2

Example list of employments matched to the MUP or received an artificial identifier. Employer characteristics are not shown. The presented employer identifiers (crefo) are made up dummy data and do not match MUP data.

employment and consists of a user identifier, a start and end year and an 'employer identifier', a reference to the MUP (crefo) or was generated artificially. For example, the employers named 'ZEW Gmbh' and 'ZEW - Mannheim' were linked to the same employer identifier. For employers successfully linked to the MUP, we have additional data like the employee count, the location of the employer, and the year of foundation from this data preparation step on. The data is usually available as a panel. For employers with an artificial id, the characteristics are not available.[17]

The employer data consists of three types: Employers only identified in XING, employers only identified in the MUP and employers identified in both databases. In the following step, we are only interested in employers listed in public employments linked to the MUP or non-matched employers (XING). For the subsequent statistical analyses, we only use the matched employers. About 1.5 million unique employers have been successfully linked to the MUP. The number refers to the matched employers listed in XING employments because we do not count employers with a profile but are not mentioned in at least one employment. Furthermore, employers with an artificial identifier are not counted as well. In the MUP, some variables have missing values. The missing rates for the variables of the matched employers are: Founding date (14%), district id (4%), legal form (<1%), two-digit NACE code (12%), exit data (<1%). The employment counts are rather sparse. For example, 23% of matched employers have not even one employment count from 2000 to 2021. However, many of these employers were not yet or no longer active. A possible

[17]The matched employment data has the following characteristics (missing rate in parentheses): Employment type (0%), employment title (<1%), career level (47%), discipline (60%), start year (14%), and end year (35%). Some of the employment characteristics were re-coded for this project (see Table A.1).

solution is imputing missing numbers or carrying forward the last existing employment count. About 9 million employers are identified in the XING database without finding a link to the MUP, e.g., organisations outside the business enterprise sector or foreign employers. These employers have been issued an artificial identifier. The employer count is overestimated because unmatched employers like 'ZEW' and 'ZEW Leibniz Centre' do not receive the same identifier. For these observations, there are no MUP characteristics, such as the founding year of the employer. However, the non-matched employers are not directly used in the following analyses.[18]

3.2 Flows

Employee flows between employers are extracted from the employment data on XING. Figure 2 shows a schematic representation of the previous and subsequent data preparation steps. The steps are: Data export, computation of match candidates, enrichment of candidates with external data, selection of the candidate with the best match, deletion of non-public data, and computation of flows. For this, we select the employment data of users with a public profile and define a flow as the switch between two successive employments. Temporal breaks between employments, such as unemployment, are ignored.

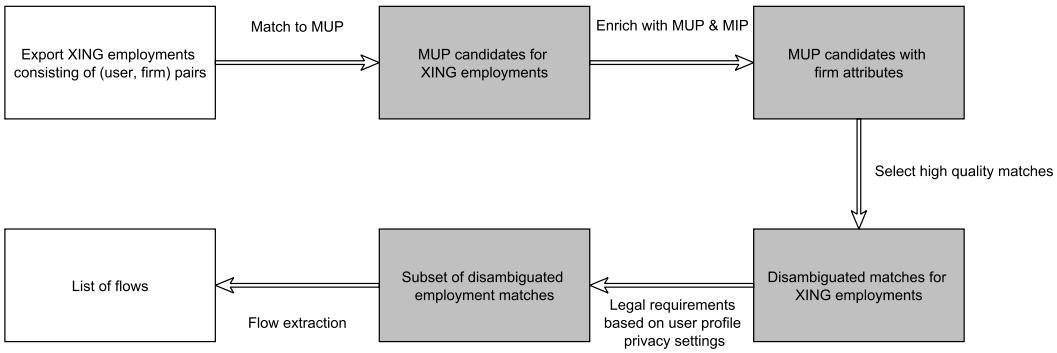


Figure 2: Flow extraction procedure. Steps: (1) Export data, (2) Get match candidates, (3) Enrich data, (4) Disambiguate match candidates, (5) Select subset following legal requirements, (6) Extract flow data. Own illustration.

For example, if an individual was unemployed for three years between two employments, then a flow between the employment before and after unemployment exists in the data set. We create an employer identifier ('missing') for the initial entry and final exit, e.g., for employment starters or retired individuals. Furthermore, only user profiles that list at least two employments are used. The described process leads to about 21M flows for the matched and unmatched employments. Flows extracted from matched employments

[18] Lastly, we also re-code employer characteristics of the MUP database (see Table A.2 in the Appendix).

Table 3: Flow Extraction Steps

Step	Description	Data
1.	Start with all employments.	46M raw employments are exported from the Data Warehouse.
2.	Keep employments of users with at least two entries.	34M employments are selected from the user profiles.
3.	Keep employments of users with public profiles.	23M employments: 14M are matched; 10M employments have artificial identifiers. Numbers do not sum up, due to rounding errors.
4.	Extract flows from employments.	21M flows are retrieved from employments. The flow data includes first employment entrances and last employment exits (modelled as sink nodes).
5.	Extract matched flows.	9M flows between matched employers (incl. sink nodes). 7M flows model employment changes between two matched employers (excl. sink nodes).

Includes observation counts for each step in million employments (M). Input data: Employment and employer data.

comprise about 9M observations, where 7M observations are flows between employers (see Table 3). Each flow entails the year of the employment change and the old and new employment characteristics. In the following, the flow extraction processing is presented in a simplified way. Table 4 shows a simplified list of employee flows retrieved from the employments in Table 2. Each row corresponds to a flow between employers and contains a user identifier, the year of the employment switch, and a reference to the old and new employer. The employees who work for the first time (e.g., previously in school) or do not have follow-up employment (e.g., retired or pensioned) link to the 'missing' node. As a time stamp for a flow, the start year of the new employment is used instead of the end year of the old employment.[19] After this transformation, the user reference is deleted to meet the privacy requirements.

3.3 Networks

Lastly, we create a temporally ordered set of graphs using the flow data.[20] We do not include the unmatched flow data as employer-level characteristics are missing. In addition, a higher number of employers is significantly lengthening the subsequent calculations, and the data quality of those observations is lower. We create a series of graphs on an annual level.[21] For this purpose, we include a flow into an annual graph if the employment change has occurred within the respective year. Internships and student employments are not considered. We model the data as a weighted (each edge has a weight), directed (edges have a direction), and simple (no loops; at most, one edge per node pair) graph (Diestel 2005). The graph nodes are the employers, and the edges denote the employee flows

[19]Exception: For the switch to pension/retirement, we use the end year of the last employment.

[20]The terms 'graph' and 'network' are often used interchangeably. We use the term graph to refer to the mathematical model. For the analytical applications, we use the term network.

[21]There are also dynamic time series methods, e.g., the sliding window model (Datar et al. 2002).

Table 4: Example List of Extracted Flows

User	Old Employer	New Employer	Year of switch
0	Missing	crefo: 2344	2010
0	crefo: 2344	crefo: 9988	2015
1	Missing	crefo: 2344	2010
1	crefo: 2344	artificial: 1	2012
2	Missing	crefo: 2344	2009
2	crefo: 2344	crefo: 2885	2015
2	crefo: 2885	crefo: 2367	2020
2	crefo: 2367	Missing	2021
3	Missing	crefo: 7781	1980
3	crefo: 7781	crefo: 4887	2010
3	crefo: 4887	crefo: 9988	2015
4	Missing	crefo: 2344	2015
4	crefo: 2344	artificial: 2	2018

The list of extracted flows retrieved from the (un)matched employments that are presented in Table 2. Employer characteristics are not shown. The presented employer identifiers (crefo) are made up dummy data and do not match MUP data.

between employers. Loops are deleted, e.g. a switch of employments or promotion within an employer, and multi-edges are aggregated to weighted edges. If multiple employees move between two employers, we model this with an edge characteristic named 'weight'.[22] Furthermore, we model the direction of the employee flow in the graph (directed edge).

Table 5: Filtered Flow List

User	Old Employer	New Employer	Intern/Student	Year of switch	Keep
0	Missing	crefo: 2344	No	2010	No
0	crefo: 2344	crefo: 9988	No	2015	Yes
1	Missing	crefo: 2344	No	2010	No
1	crefo: 2344	artificial: 1	No	2012	No
2	Missing	crefo: 2344	Yes	2009	No
2	crefo: 2344	crefo: 2885	No	2015	Yes
2	crefo: 2885	crefo: 2367	No	2020	No
2	crefo: 2367	Missing	No	2021	No
3	Missing	crefo: 7781	No	1980	No
3	crefo: 7781	crefo: 4887	No	2010	No
3	crefo: 4887	crefo: 9988	No	2015	Yes
4	Missing	crefo: 2344	No	2015	Yes
4	crefo: 2344	artificial: 2	No	2018	No

The filtered list of flows is based on Table 4. The graph edges kept for 2015 are marked in the last column (fulfilled all requirements). The presented employer identifiers (crefo) are made up dummy data and do not match MUP data.

[22] An alternative are so-called multigraphs, where multiple edges are allowed between pairs of nodes.

The annual graphs do not represent employers with less than one flow in the respective year.[23] Each node contains employer characteristics like the employee count and location. The edges are affiliated with user characteristics, e.g., the year of the employment change. Table 5 presents an example of flows extracted from matched employments. The flows fulfilling all requirements are marked in the last column. In our example, we are interested in employee flows within 2015 that fulfil some additional requirements. For example, employments of students and internal employment changes (loops) are removed. The remaining employee flows are modelled as a graph consisting of five nodes and four edges (see Figure 3). The annual networks contain employer- and flow-level characteristics, as those are not shown for reasons of simplicity. User data are no longer needed as the employment data has been mapped to the nodes and edges. Table 6 shows the number of nodes and edges in the annual XING networks from 2010 until 2020. The data for 2020 was not fully available at the time of the data export.[24] The sequence of networks has between 75K and, at most 164K nodes. The edge count is at a minimum of 100K and at most 289K (see $Edges_1$). We find a slight decline in 2019, although the complete XING data was exported. The employer and edge counts may be lower in the most recent years, as there is a time lag in the MUP data until newly established employers are available. The number of edges (including duplicates) is similar to the number of unique edges (see $Edges_2$ vs. $Edges_1$) as only a few employees move between the same employers in one year. The ratio between nodes and edges changes over time, i.e., the edge count per node increases. The edge count does not add up to 7M, as we ignore, for example, unpaid workers and do not consider all available years.

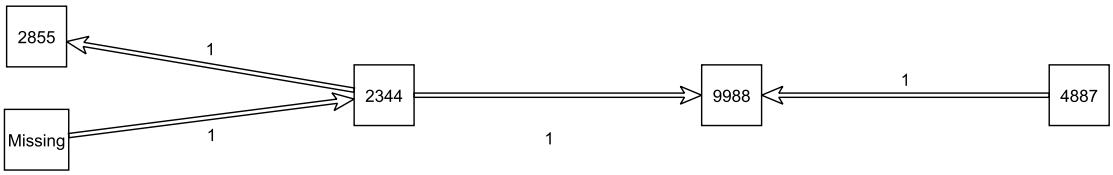


Figure 3: Weighted, directed, and simple graph for 2015. Nodes represent employers, and edges model employee flows. The edge weight is the employee count moving between two employers. The figure is based on flows from Table 5. The presented employer identifiers (crefo) are made up dummy data. Own illustration (created with yEd - graph editor; https://www.yworks.com).

Lastly, the degree centrality measure is introduced (Freeman 1978; Nieminen 1973). The measure is calculated at the node level, i.e., at the level of employers. The definition depends on the network type and indicates how strongly or often a node is connected to

[23]The isolated nodes can easily be added. However, we decided against it because the runtime and required memory for some graph algorithms scale quadratically with the number of nodes in the graph.

[24]Date of data export: 10/26/2020.

Table 6: Annual Count for Nodes and Edges

Year	$\#Nodes$	$\#Edges_1$	$\#Edges_2$
2010	127,988	162,264	176,324
2011	137,723	186,636	203,013
2012	142,236	197,865	215,600
2013	147,168	207,711	227,067
2014	154,134	229,854	252,824
2015	159,902	253,068	281,432
2016	163,706	277,176	311,493
2017	163,364	288,805	324,304
2018	157,730	287,170	322,554
2019	135,470	241,574	271,399
2020*	74,367	99,602	110,319

Number of nodes, unique edges ($Edges_1$) and edges counting duplicates ($Edges_2$) for the XING networks. Networks are retrieved from the presented flow data. Source: TUM and ZEW based on XING data.
* Only partially covered.

other nodes. For directed networks, an in-degree and out-degree centrality is available in addition to the degree centrality. For each node, the weights of the adjacent edges are summed, i.e., total, incoming or outgoing edges.[25] If there is no weight, each edge is assumed to have a score of one. Table 7 shows the degree centrality for the example network in Table 5. The centrality scores are calculated for a directed network and represent the number of people moving between two employers within the year 2015. In addition to the degree centrality, further measures determine the relevance of nodes in a network, e.g., eigenvector and PageRank centrality (Page et al. 1999). However, this work does not use these due to more complex definitions and interpretations. The number of

Table 7: Degree Centrality for Example Nodes

Node Identifier	Degree	In-Degree	Out-Degree
Missing	1	0	1
crefo: 2855	1	1	0
crefo: 2344	3	1	2
crefo: 9988	2	1	1
crefo: 4887	1	0	1

Degree centrality measure for the weighted, directed, and simple network illustrated in Figure 3. The metrics are calculated at the node level. The presented employer identifiers (crefo) are made up dummy data and do not match MUP data.

centrality scores corresponds to the node (employer) count of the network. As a result, we construct an unbalanced panel based on the centrality scores available for a series of years. However, not every XING employer linked to the MUP also exists in all annual networks, e.g., if no employee left the employer or the employer closed within a year.

[25]Some definitions standardise the centrality score by dividing it with the node count of the network.

4 Data Description and Results

In this section, we describe the data set. The focus lies on the employments (Section 4.1), employers (Section 4.2), flows (Section 4.3), and networks (Section 4.4).

4.1 Employments

Figure 4 shows selected characteristics of the platform users. The analyses improve our understanding of the users active on the platform XING.[26] The most recent employment per user is utilised for the discipline, career stage, and employment type analyses. For the analysis of the user experience, the oldest employment per user is considered. Our findings suggest: First, most employees have between zero and nineteen years of professional experience. Few users have more than 39 years of professional experience.[27] Second, the

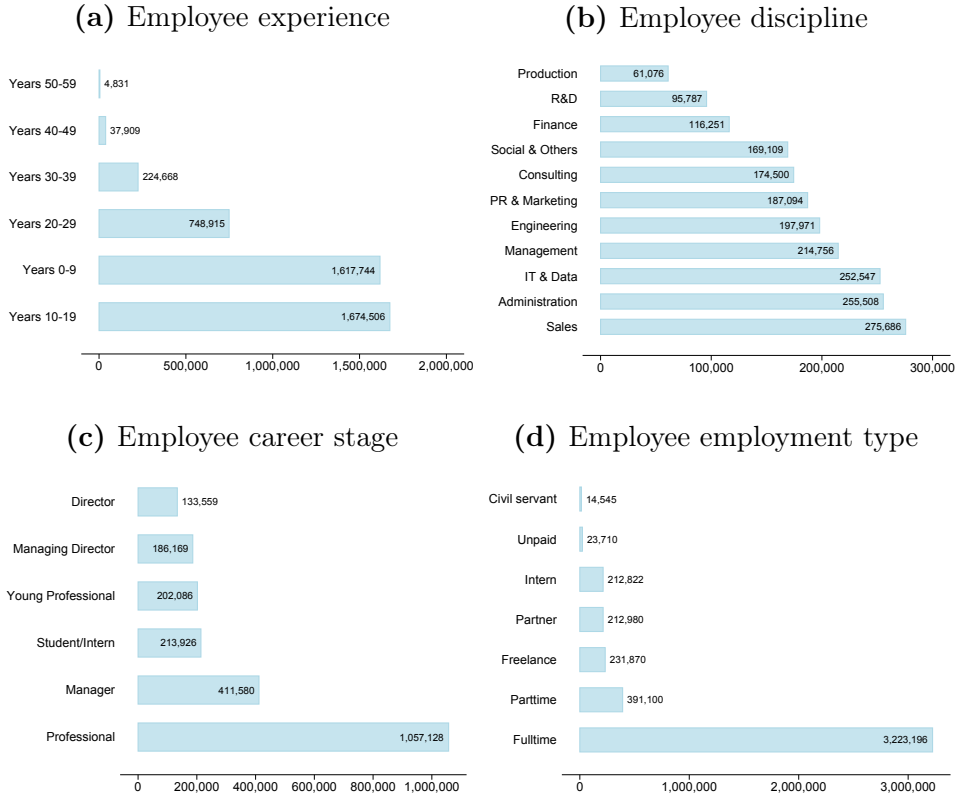


Figure 4: XING users by different characteristics. Note: (a): Per experience class; (b): Per discipline; (c): Per career stage; (d): Per employment type. For the analyses, only matched employments from publicly accessible user profiles are used. Source: TUM and ZEW based on XING data. Own illustration.

[26] We restrict the analysis to subsets of the data: Employments have to be publicly available, matched to the MUP and the respective characteristic, e.g., the career stage, needs to be known.

disciplines *sales*, *administration*, and *IT & Data* have the most employees. Surprisingly, few employees are working in *Production*, *R&D* and *Finance*. This may result from the respective employees rarely using the platform (production) or comparatively few employments in the field (R&D). Third, most employees have professional experience but no managerial position. Managing directors and directors are found least often. Fourth, the majority of the employees have full-time employment.[28] Civil servants and unpaid workers are by far the least frequent. It is important to note that the XING data do not represent German employees. Figure 5 shows the employment length for employer and employment characteristics. First, employees stay longest at employers founded before 1990. For example, these employers are often larger and more established. However, we did not account for employments that were changed within an employer.[29] Second, the employment length is highest for partners and civil servants. By far, the shortest employment length is found for interns.[30] Figure 6 shows the number of employees by employer size, region,

(a) Employment length by employer founding year class

(b) Employment length by employment type

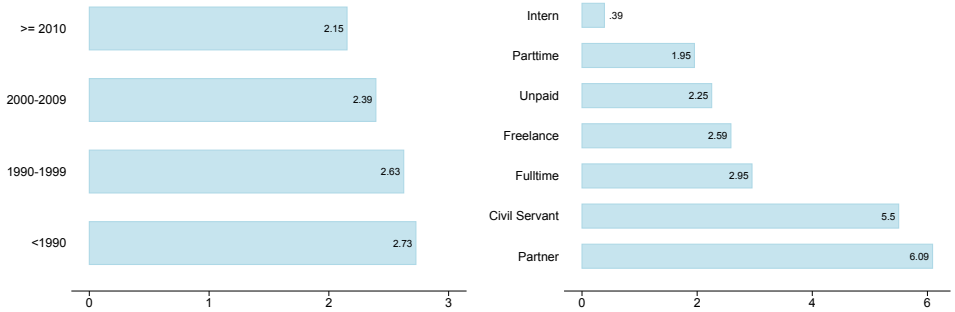


Figure 5: Average length of employment (in years) for users by employer age (left) and employment type (right). The start and end year of the employment are used as a heuristic; based on public and matched employments. Source: TUM and ZEW based on data by XING and MUP. Own illustration.

legal form, sector, and founding period. First, small employers (<10 employees) are most often linked to employments. We expected the pattern as a large proportion of employees in Germany work for small and medium-sized enterprises (SME). The smallest number of employments is found for employers between 250 and 999 employees. Second, most

[27] German employees must work for 35 or 45 years to be considered long-term or very long-term insured to receive their full pensions. Primarily, better-educated employees are present on XING, who often start their careers later and thus work for fewer years overall.

[28] Students/Interns most often have the employment types 'intern' and 'part time'.

[29] Larger employers often provide more opportunities to climb the career ladder in-house.

[30] The median length of employment relationships subject to social security contributions (excluding apprenticeships) is slightly larger than four years for 2017-2021. Source: https://statistik.arbeitsagentur.de/Statistikdaten/Detail/202112/iiia6/beschaeftigung-sozbe-dauern/dauern-d-0-202112-xlsx.xlsx?__blob=publicationFile&v=1.

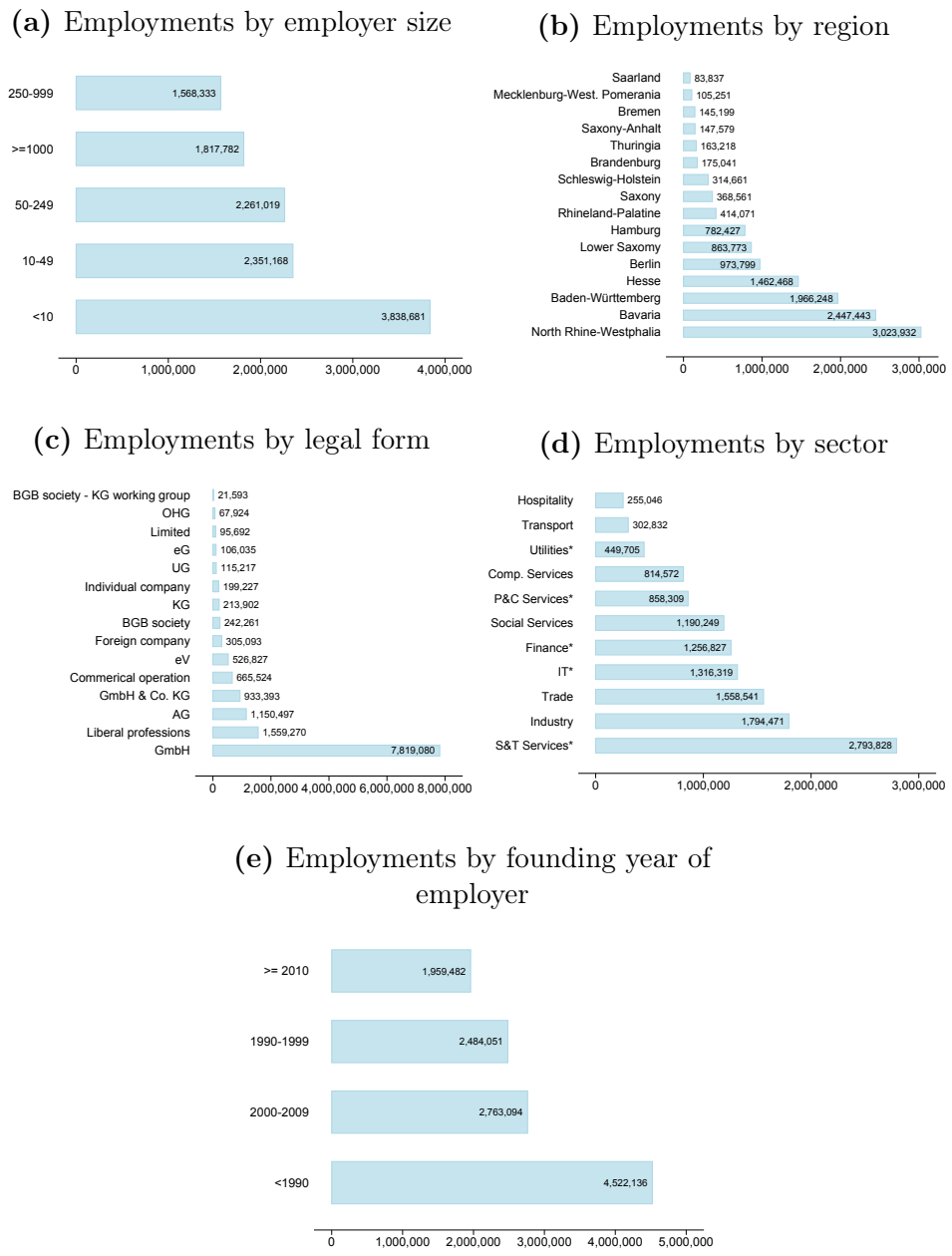


Figure 6: Employment counts (XING) w.r.t. median employer size (top left), region (top right), legal form (middle left), employer sector (middle left), employer founding year (bottom). Only matched and public employments are used. Labels marked with a star (*) are listed in abbreviated form (see Table A.2 for a list of all labels). Source: TUM and ZEW based on data by XING and MUP. Own illustration.

employments are linked to employers in North Rhine-Westphalia, Bavaria and Baden-Württemberg. The fewest employments are found in Saarland, Mecklenburg-Western Pomerania and Bremen. This is plausible because these states have the highest/lowest population counts. Third, most employments are linked to limited liability companies (GmbH) and liberal professions. Furthermore, employments within the 'Freelance, scien-

tific and technical services' (S&T Services) sector and for employers founded before 1990 are the most frequent. Only a few employments exist in the hospitality and transport sector and for employers founded since 2010.[31]

4.2 Employers

Figure 7 shows selected MUP characteristics for the matched XING employers.[32] The XING employer data set cannot be compared directly with the total stock of German employers in the MUP. For example, the XING data contains employers that are no longer economically active. Our findings suggest: First, most employers on XING have less than ten employees and only few employers have at least 250 employees. Second, most employers are found in the sector *scientific and technical Services (69-75)*; the fewest employers are found in the sector *transport (49-53)*[33]. Third, most employers are founded in 2000 - 2009 or 2010 & later; there is a small decrease in the number of founded employers before 2000. Fourth, most employers are located in the German regions North Rhine-Westphalia, Bavaria and Baden-Württemberg (in the former territory of West Germany). The fewest are located in Mecklenburg-Western Pomerania, Bremen and Saarland. Many regions of the former German Democratic Republic (GDR) are in the lower half of the ranking illustrating an east-west divide. The city of Berlin and Hamburg are in the midfield of the ranking. In summary, the XING employer data seems plausible and representative in Germany.[34]

4.3 Flows

Figure 8 shows the employee flows by XING users by the employment characteristics, career level, discipline, employment type, employer size, sector, and region. First, we analyse the flows between employments for professional experience levels. Most users transfer out of employments with professional experience. Users usually switch employments within an experience class or into a higher experience class, e.g., from *young professional*

[31]The presented figures pool historical XING data. Therefore, they are no representation of the current employments in Germany. For comparison, the employer characteristics are described in Figure A.4 (Appendix) for the stock of MUP employments (2002-2019) in Germany. Larger deviations in the MUP and XING data are found for the founding period, employer size, and industry sector.

[32]The employers are extracted from the XING employments previously matched to the MUP.

[33]The numbers in parentheses are the two-digit NACE codes. The NACE codes are the 'Statistical Classification of Economic Activities in the European Community'. For a definition, see https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF.

[34]Figure A.3 (Appendix) shows the share of MUP employers (2002-2019) and XING employers in Germany for regions, founding year periods, legal forms, employer size classes, and sectors. The distributions of the XING and MUP data appear to be roughly similar. However, some deviations exist, such as the legal form and sector. For us, it seems plausible that, for example, employers in the utility sector are less covered as these employers and their workforce are less reliant on online platforms.

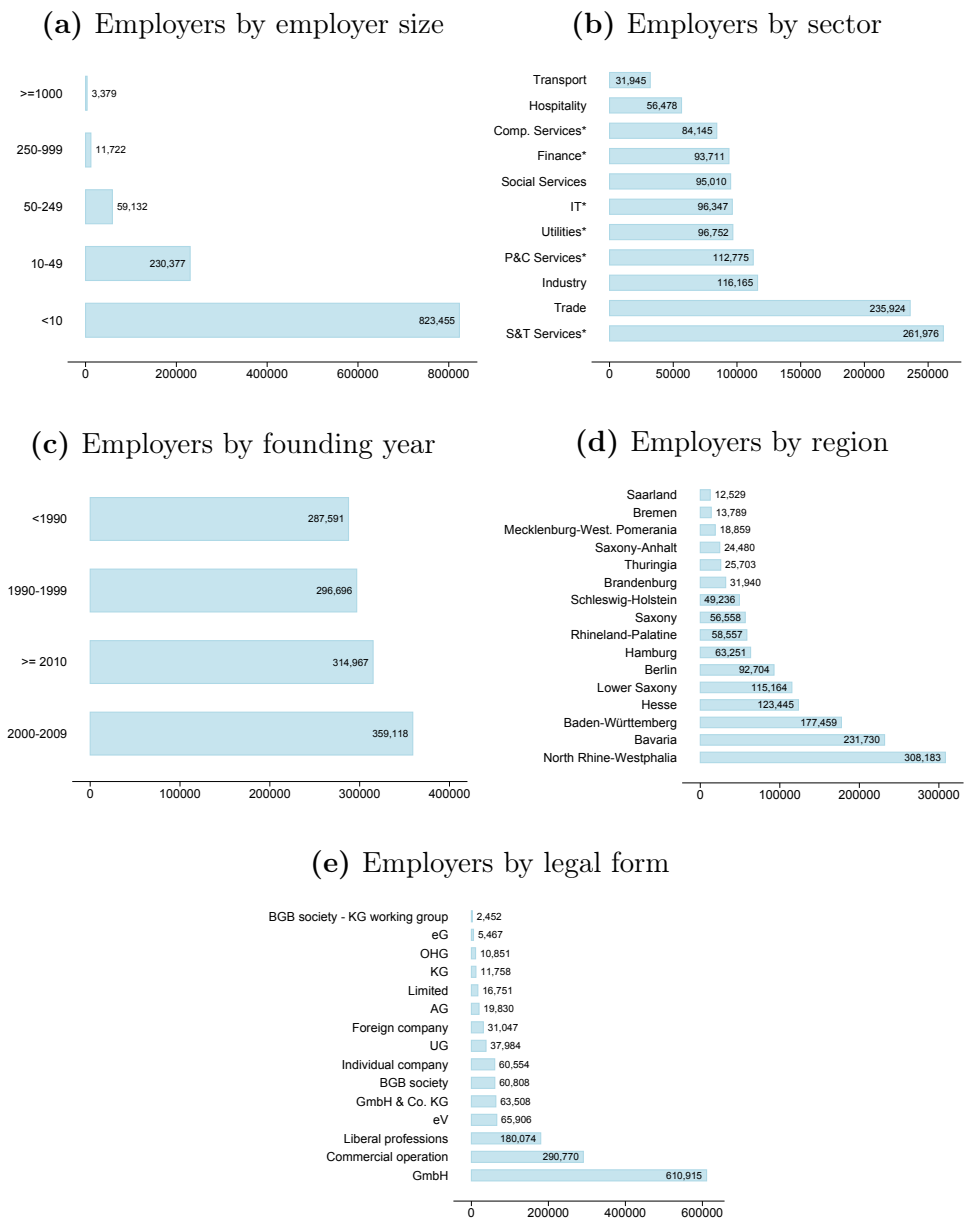


Figure 7: Number of XING employers w.r.t. MUP characteristics; based on matched employers. Note: (a): Per median employer size group; (b): Per sector; (c): Per founding year group; (d): Per German region; (e) Per legal form. Labels marked with star (*) are listed in abbreviated form (see Table A.2 for full label). Source: TUM and ZEW based on XING data. Own illustration.

to *professional*. Employees rarely reduce their experience level, for example, if an employee is moving into a new sector. Second, we analyse the flows between employments for employment disciplines. Most users switch employments within their discipline. However, we find flows for all discipline combinations in the data set. Third, we analyse the flows between employments by employment types. Here, we find a deviating pattern: Many of the employment types show a flow into the 'Full-time' class. Further validity checks show that

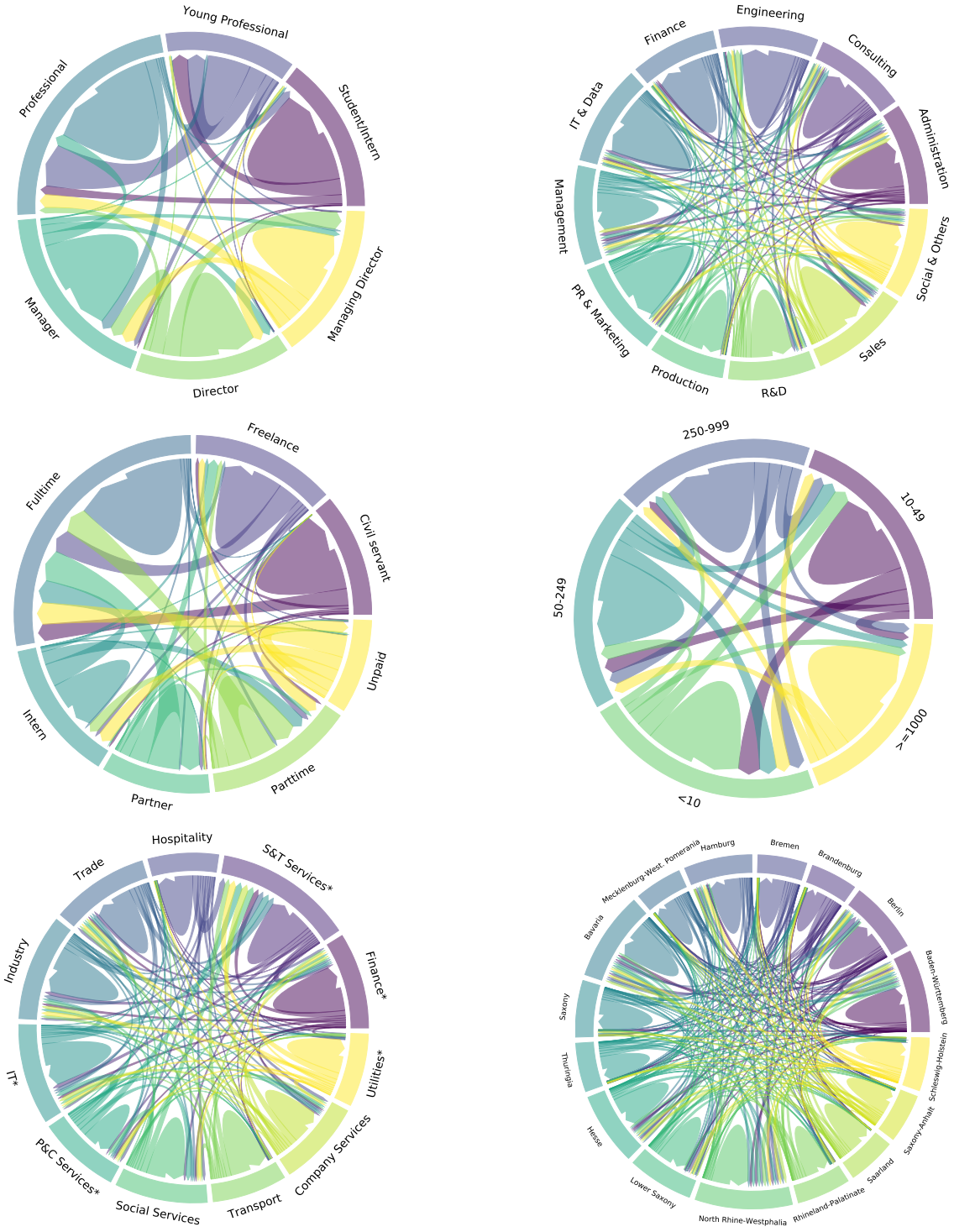


Figure 8: Employee flows illustrated as chord diagrams. Top left: Experience levels. Top right: Employment discipline, Middle left: Employment type, Middle right: Employer size, Bottom left: Employer sector, Bottom right: Employer region. Flows are based on matched employments from public user profiles. Labels marked with a star (*) are listed in abbreviated form (see Table A.2 for a list of all labels). Table A.6 and A.7 (Appendix) present the respective flow matrices. Source: TUM and ZEW based on XING data. Own illustration.

civil servants move only occasionally into unpaid employments and young professionals rarely move into (managing) director positions. Fourth, many flows occur within the same employer size class and relatively few flows are found for employment changes to large-sized employers. Fifth, many employees are moving within sectors. However, employers in the *scientific and technical Services* (69-75) receive substantial inflows from all other sectors. Sixth, the majority of employees change employers within regions. However, the economically strong regions Bavaria, North Rhine-Westphalia, and Baden-Württemberg have substantial inflows from all other regions. Figure A.1 (Appendix) shows the flows for the regions of West and East Germany.[35] Most employees change employers within West Germany. Inflows from the West do not offset the outflows from the East. Lastly, we present the flows by district types (see Table A.5 (Appendix) for definitions). Most employees switch employers within big cities. The data does not show a rural exodus, although some employees move from less populated regions into urban districts and big cities.

4.4 Networks

Table 9 presents selected metrics for the employee flow networks. The annual networks are available for the time period 2010 to 2020. The nodes represent the employers while the edges are based on the employee flows. The definitions of the network metrics and references to their implementation are presented in Table 8.[36]

The clique counts of the graphs vary between 181K and 649K. These cliques may be, for example, highly inter-connected employers within the same field. Large cliques consist of smaller cliques as subgraphs (therefore, the number is quite high). The maximum clique size varies between 8 and 16 and is on average 12. The cluster counts lie between 64K and 133K, and the graphs' densities vary between $9.6\text{x}10^{-6}$ and $1.8\text{x}10^{-5}$. Thus, many pairs of employers have no employee flows and may also not be indirectly linked by employee flows. The transitivity lies between 0.10 and 0.12, and the reciprocity is between 0.05 and 0.07. Mutual, e.g., between direct competitors, and transitive employee flows between employers are thus not very common. Exemplary, Figure 9 presents the intra-city flows between Munich employers in 2019.[37] The network comprises 2,982 employers and 4,235 unique employee flows and is a subgraph of the data described in Table 6. Large parts of the network are connected and some central employers have many incident edges. Some employers have only one incident edge and are not or only sparsely connected to

[35] Berlin is not considered as the assignment to West or East Germany is not clear.

[36] Figure A.7 provides a directed example graph and the metrics from Table 9. Further details about the implementation of the metrics are available at https://igraph.org/python/doc/api/igraph.Graph.html.

[37] Table A.4 (Appendix) describes the fuzzy matching of XING employers to geographical coordinates.

Table 8: Applied Network Metrics

Metric	Definition	Function
#Nodes	Number of nodes in the network.	graph.vcount()
#Edges	Number of unique edges in the network.	graph.ecount()
#Cliques	Number of complete subgraphs, where an edge is present between any two nodes (excl. loops).	len(list(graph.cliques()))
Max. clique size	The number of nodes in the largest clique.	graph.clique_number()
Transitivity	Measures the probability that two neighbors of a node are connected. Calculated for each node and then averaged. Vertices with less than two neighbors are ignored.	graph.transitivity_avg local_undirected()
Reciprocity	It is defined as the probability that a directed edge's opposite counterpart (other direction) is also included in the network.	graph.reciprocity()
#Clusters	Number of strongly connected components in the network. A strongly connected component is a subgraph, where every node is reachable from every other node.	len(list(graph.clusters()))
Density	Ratio of the edge count by the maximum possible edge count.	graph.density()
Girth	Length of the shortest circle in the network. Circles consist of at least three nodes.	graph.girth()

Description of network metrics. The table provides the definitions of the metrics and the igraph function (Python; https://igraph.org/python/).

Table 9: Annual Network Metrics

Metric	Min.	Max.	Mean
#Nodes	74,367	163,706	142,162
#Edges	99,602	288,805	221,065
#Cliques	181,363	649,221	456,504
Max. clique size	8	16	11.72
Transitivity	0.10	0.12	0.11
Reciprocity	0.05	0.07	0.06
#Clusters	64,501	133,471	118,768
Density	$9.6\text{x}10^{-06}$	$1.8\text{x}10^{-05}$	$1.1\text{x}10^{-05}$
Girth	3	3	3

Network metrics for the series of annual graphs (2010-2020).

the rest of the network. Employers without flows are not represented in the network. Comparable analyses can be performed for arbitrary time periods and cities or regions, see Figure A.2 (Appendix) for Mannheim. For 2019, the network for Mannheim includes 236 employers and 221 unique employee flows. Therefore, it is considerably smaller than the network for Munich. Again, a large share of the employers and flows are found in the city centre. In contrast, Figure A.5 and A.6 (Appendix) show the relative flow

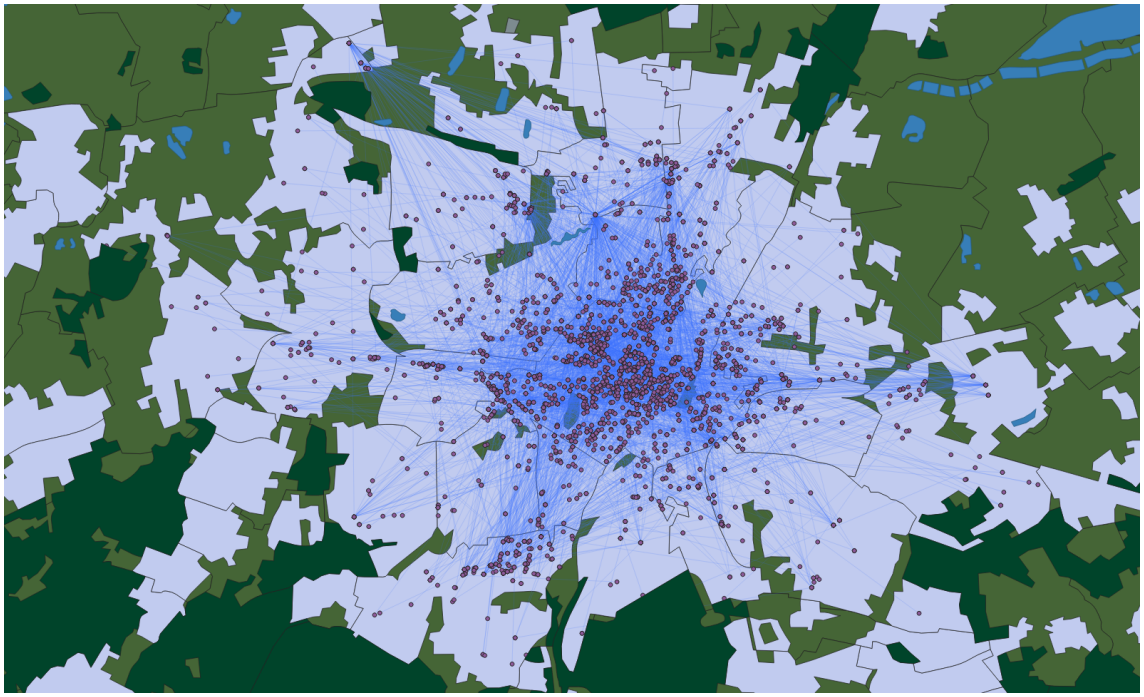


Figure 9: Employee flow network for the city of Munich. The flow data is restricted to flows within the city in 2019. Source: TUM and ZEW based on XING data. Own illustration (created with QGIS; https://www.qgis.org).

counts for the cities Berlin, Cologne, Hamburg, and Munich. We present the ten districts with the most inflows and outflows for each city. Berlin, Hamburg and Munich play an important role in employee mobility through a high employee exchange between the cities. Furthermore, regional differences exist, thus districts in the closer neighbourhood have a special role. For example, the city of Cologne has a high level of employee exchange with the nearby districts Bonn and 'Rhein-Sieg-Kreis'. Figure 10 shows network-based measures for German regions. The figure on the left shows the log-scaled number of clusters per district. Larger cities and districts in western Germany have the most regional clusters, i.e., groups of employers without flows from/to other groups. However, the number of clusters does not say anything about their size. The figure on the right shows the network density scores per district. Large parts of eastern Germany have the highest scores. The maximal edge count of a network scales quadratically with the node count. As a result, the density score is directly impacted in sparse networks. A possible explanation is that there are fewer employers in eastern Germany and, for example, in the region Saarland, but those are better connected by employee flows. Figure 11 presents the degree centrality by employer characteristics. First, we analyse the degree centrality by employer size. The employer size is the median number of employees per employer in the panel. Employers with larger employee counts have higher degree centrality scores. Second, employer age is positively linked with high degree centrality scores. We expected this link

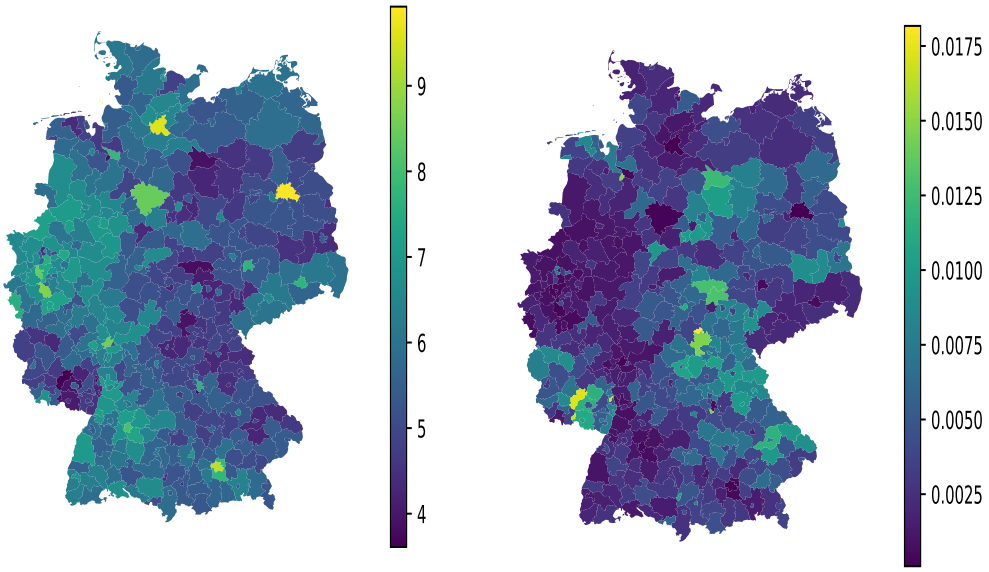


Figure 10: Left: Number of clusters per district (log-scaled). Right: Network density per district. Both measures are calculated for district-level networks without temporal restrictions. Data: Public and matched XING employments and flows. Source: TUM and ZEW based on XING and MUP data. Own illustration (created with geopandas; https://geopandas.org/en/stable/).

as employer age correlates with employer size. Third, employers listed as publicly traded ('AG') have, by far, the highest degree centrality. We found the lowest centrality scores for entrepreneurial companies with limited liability ('UG') and commercial operations.[38] Fourth, the sectors 'Social Services', 'Finance, Insurance & Real Estate' and 'Information and Communication' have the highest centrality score. The sectors 'Hospitality' and 'Utilities/Construction' have the lowest scores. Fifth, big cities and urban districts have higher scores than rural districts. Sixth, the regions Berlin and Hamburg have the highest scores. Brandenburg and Saarland have the lowest scores. The centrality scores show an east-west divide.

In the last step, we perform two plausibility checks at the employer level to verify that the XING and MUP data are statistically related. This indicates that the flow data is externally valid and models the actual in- and outflows of German employers, although being by no means complete. Unfortunately, the absolute number of employees and flows cannot be observed over time on XING, as many employees and employers are not active on XING. Table 10 shows the relationship between the changes in the employee counts based on the MUP data and XING data.[39] The variables are positively and significantly

[38]The German and partly translated labels (English) are listed in Table A.3 (Appendix).

[39]We start with the 1.6M observations from Table 6, i.e., the sum of network nodes over the period from 2010 to 2020. Next, we delete observations if sector, region, founding date or legal form are missing, leaving us with 1.4M observations; other observations are omitted due to missing employee data.

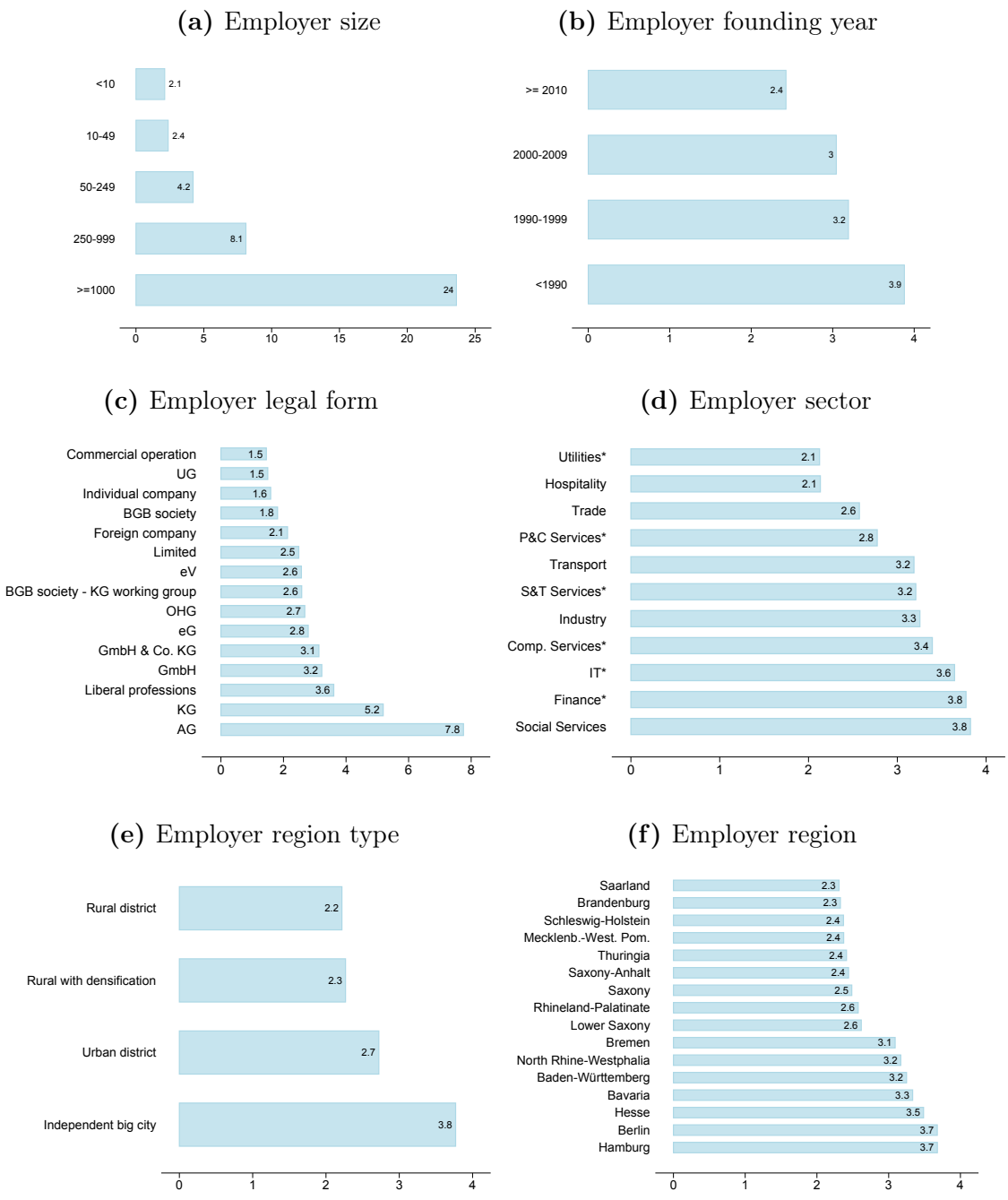


Figure 11: Degree centrality by employer characteristics. The employer data is structured as a panel. Missing values are not shown. Employers without flows in the individual annual networks are not considered. Labels marked with a star (*) are listed in abbreviated form (see Table A.2 for a list of all labels). Source: TUM and ZEW based on XING and MUP data. Own illustrations.

related, as we measure the changes in the employment counts (= in-degree minus out-degree of an employer within one year) with the flow data from the platform. However, the coefficients for ΔEmployees are smaller than one indicating that we do not capture the entire inflow and outflow of employers. The relationship persists for the original data

(Columns 1 and 3) and log-transformed variables (Columns 2 and 4). Unfortunately, the R^2 scores are relatively low.[40]

Table 10: Regression by Degree Centrality

	(1) ΔDegree	(2) Δlog(Degree+1)	(3) ΔDegree	(4) Δlog(Degree+1)
ΔEmployees	0.00266***		0.00267***	
	(<0.001)		(<0.001)	
Δlog(Employees+1)		0.176***		0.174***
		(0.00237)		(0.00237)
Year			0.00154	0.00105***
			(0.00118)	(<0.001)
Constant	-0.0341***	-0.0292***	-2.869	-2.104***
	(0.00335)	(<0.001)	(2.373)	(0.623)
Sector Dummies	No	No	Yes	Yes
Legal form Dummies	No	No	Yes	Yes
Region	No	No	Yes	Yes
Founding Dummies	No	No	Yes	Yes
N	767,652	767,652	767,652	767,652
R^2	0.004	0.009	0.005	0.011
adj. R^2	0.004	0.009	0.005	0.011

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard errors in parentheses. Regressing degree centrality as dependent variable (XING) on employer characteristics (MUP). Source: TUM and ZEW based on XING and MUP data.
Definitions:
ΔDegree(t) = In-Degree(t) - Out-Degree(t)
Δlog(Degree+1) = log(In-Degree(t)+1) - log(Out-Degree(t)+1)
ΔEmployees(t) = Employees(t+1) - Employees(t-1).

5 Conclusions

The availability of social media data creates new opportunities for empirical economic research. This paper presents a detailed exploration of the usage data from the career-oriented social networking platform XING for measuring employee flows and employer networks. We obtain employment data from public user profiles and link them based on employer data to the Mannheim Enterprise Panel (MUP). This novel link creates a unique platform-based LEE data set that allows tracking employee flows between employers. The data set comprises about 14M employments for 1.5M disambiguated employers. Furthermore, the matched employment data is used to extract 9 million employee flows

[40] Robustness check: Consider in every year all 964K employers with at least one employee flow between 2010 & 2020. Set missing centrality scores to zero if no flows for the employers are found in the respective year. Results show that the coefficients for the employee variables are positive and highly significant.

and create eleven annual flow networks for 2010 to 2020.

We check the plausibility of the data set and show that career-oriented social networking data contains meaningful and valuable data about employments, employers, employee flows, and employer networks. In doing so, we test the plausibility of selected network measures against employer data and find that they show plausible patterns. For example, the employer-level degree centrality (in-degree minus out-degree) positively correlates with changes in MUP employment counts. Thus, the professional network data appears to be externally valid and can model the in- and outflows of German employers. Notably, the coverage goes beyond paid employment by including, for example, freelance work and some entrepreneurial activities. For the analysis of networks, the coverage is sufficiently high, and network measures can be derived for employers that are neither active in patenting nor engaged in larger, visible alliances.

However, using such data for research purposes should be done with a lot of care. Platforms like XING and LinkedIn may have some biases concerning their user base. For example, older employees, employees without training, and employees from specific sectors, such as household services, are less frequently represented. A better overview of all such employments can probably be achieved, for example, with linked employer-employee data from official sources such as the IAB data from the German Labor Office[41] even though this also comes with other restrictions as discussed at the beginning of the paper. Furthermore, self-employed workers may not be correctly matched in our data set because they could not be linked to an MUP employer. In principle, however, it is possible to recognise these employments and treat them separately automatically. Furthermore, we only use publicly available data from the XING platform to comply with the user's desire to keep their data non-public. This may potentially result in an additional bias as, for example, specific user groups are more concerned about their privacy, e.g., employees in the fields of law and IT. Unfortunately, investigating this hypothesis in more detail is not straightforward. Also, some users do not update their profiles frequently. This creates the image that an employer employs a user for longer than actual. This can also happen if a user changes platforms and maintains his LinkedIn profile. However, our platform-based approach has a multitude of advantages. The resulting data is subject to reasonable legal constraints, is updated regularly, and is publicly accessible.

We want to highlight that employers on XING and MUP can be linked using fuzzy string matching. However, our method is not error-free. The matching approach can be further improved, for example, by adjusting the parameters or using different ways. There is a discrepancy in the unit of observation. On XING, employer sites are often listed and

[41] Institute for Employment Research: https://iab.de/.

linked to user employments. The MUP, however, consists of employers that are legally independent entities. Therefore, a link of multiple employer sites to one employer is necessary. Furthermore, employers with few employees may have a biased representation in the network, as they may not have publicly listed employment on XING. In addition, the networks are constructed on annual basis so that there are hard boundaries (between years) for the networks of a series. For example, employment switches on New Year's Eve or New Year will thus end up in two different networks. Smooth transitions between the annual networks would improve the quality of the data, e.g., by including the previous year's flows with a smaller weight. The extraction of flows does not consider many exceptional cases, for example, concurrent employments or gaps between employments. Gaps between employments are, so far, ignored. In the future, we might model them as separate nodes representing unemployment.

In conclusion, despite these challenges, we can link a large share of employers due to careful disambiguation of names, URLs and profiles. The plausibility checks suggest that the resulting data has no major shortcomings. Hence, the new database provides opportunities for being used in subsequent research on the role of employee mobility, networks, and local ecosystems for economic performance both at the employer and the regional level. The micro-nature of the data also allows, for example, the calculation of indicators on the level of the network nodes. These include centrality measures for individual employers as well as aggregate measures for network characteristics at the regional level. Data availability over time further facilitates analyses of network development and drivers of these changes.

6 Literature

Abbasiharofteh, M., Kinne, J., & Krüger, M. (2021). The strength of weak and strong ties in bridging geographic and cognitive distances [Preprint (Online); accessed 16.02.2023]. *ZEW – Leibniz - Centre for European Economic Research Discussion Paper*, *21-049*. http://ftp.zew.de/pub/zew-docs/dp/dp21049.pdf

Abendroth, A.-K., Melzer, S. M., Jacobebbinghaus, P., & Schlechter, F. (2014). Methodenbericht Beschäftigten-und Partnerbefragung des Linked-Employer-Employee Panels (LEEP-B3) im Projekt B3 „Wechselwirkungen zwischen Verwirklichungschancen im Berufs- und Privatleben“ [Preprint (Online); accessed 18.04.2023]]. *SFB 882 Technical Report Series, No. 6*. https://core.ac.uk/download/pdf/211813627.pdf

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, *6*(2), https://doi.org/10.1177/2053168019848554.

Almeida, P., & Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, *45*(7), 905–917.

Axenbeck, J., & Breithaupt, P. (2021). Innovation indicators based on firm websites—which website characteristics predict firm-level innovation activity? *PLOS ONE*, *16*(4), e0249583.

Axenbeck, J., & Breithaupt, P. (2022). Measuring the Digitalisation of Firms – A Novel Text Mining Approach [Preprint (Online); accessed 10.01.2023]. *ZEW – Leibniz - Centre for European Economic Research Discussion Paper*, *22-065*. http://ftp.zew.de/pub/zew-docs/dp/dp22065.pdf

Balsvik, R. (2011). Is labor mobility a channel for spillovers from multinationals? Evidence from Norwegian manufacturing. *The Review of Economics and Statistics*, *93*(1), 285–297.

Bersch, J., Gottschalk, S., Müller, B., & Niefert, M. (2014). The Mannheim Enterprise Panel (MUP) and Firm Statistics for Germany [Preprint (Online); accessed 16.02.2023]. *ZEW – Leibniz - Centre for European Economic Research Discussion Paper*, *14-104*. https://ftp.zew.de/pub/zew-docs/dp/dp14104.pdf

Breithaupt, P., Kesler, R., Niebel, T., & Rammer, C. (2020). Intangible capital indicators based on web scraping of social media [Preprint (Online); accessed 16.02.2023]. *ZEW – Leibniz - Centre for European Economic Research Discussion Paper*, (20-046). http://ftp.zew.de/pub/zew-docs/dp/dp20046.pdf

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*, 2–9.

Collet, F., & Hedström, P. (2013). Old friends and new acquaintances: Tie formation mechanisms in an interorganizational network generated by employee mobility. *Social Networks*, *35*(3), 288–299.

Datar, M., Gionis, A., Indyk, P., & Motwani, R. (2002). Maintaining stream statistics over sliding windows. *SIAM Journal on Computing*, *31*(6), 1794–1813.

Diestel, R. (2005). Graph theory (3rd ed.) *, Electronic Edition, Berlin, New York: Springer-Verlag.*

Doherr, T. (2023). The SearchEngine: A Holistic Approach to Matching [Preprint (Online), accessed 15.02.2023]. *ZEW – Leibniz - Centre for European Economic Research Discussion Paper*, *23-001.* http://ftp.zew.de/pub/zew-docs/dp/dp23001.pdf

Edler, J., Fier, H., & Grimpe, C. (2011). International scientist mobility and the locus of knowledge and technology transfer. *Research Policy*, *40*(6), 791–805.

Franzoni, C., Scellato, G., & Stephan, P. (2014). The mover's advantage: The superior performance of migrant scientists. *Economics Letters*, *122*(1), 89–93.

Freeman, L. C. (1978). Centrality in Social Networks Conceptual Clarification. *Social Networks*, *1*(3), 215–239.

Gerhards, C., Liebig, S., & Elsner, J. (2010). Datenhandbuch: Projekt "Verknüpfte Personen-Betriebsdaten im Anschluss an den ALLBUS 2008": ALLBUS-Betriebsbefragung 2009 [Preprint (Online); accessed 18.04.2023]. *DSZ-BO Technical Report*, *1.* https://portal.fdz-bo.diw.de/sites/default/files/DSZ-BO-Technical-Report_Nr01_0.pdf

Gök, A., Waterworth, A., & Shapira, P. (2015). Use of Web Mining in Studying Innovation. *Scientometrics*, *102*(1), 653–671. https://doi.org/10.1007/s11192-014-1434-0

Giuliani, E. (2011). Networks of innovation. *Handbook of Regional Innovation and Growth.* Edward Elgar Publishing.

Godart, F. C., Shipilov, A. V., & Claes, K. (2014). Making the most of the revolving door: The impact of outward personnel mobility networks on organizational creativity. *Organization Science*, *25*(2), 377–400.

Görg, H., & Strobl, E. (2005). Spillovers from foreign firms through worker mobility: An empirical investigation. *The Scandinavian Journal of Economics*, *107*(4), 693–709.

Heining, J., Klosterhuber, W., Lehnert, P., & Seth, S. (2016). Linked-Employer-Employee-Daten des IAB: LIAB-Längsschnittmodell [Preprint (Online); accessed 18.04.2023]. *FDZ Datenreport 10/2016 DE (Dokumentation zu Arbeitsmarktdaten).* https://doku.iab.de/fdz/reporte/2016/DR_10-16.pdf

Hong, S., Oxley, L., & McCann, P. (2012). A survey of the innovation surveys. *Journal of Economic Surveys*, *26*(3), 420–444.

Hottenrott, H., & Lopes-Bento, C. (2016). R&D partnerships and innovation performance: Can there be too much of a good thing? *Journal of Product Innovation Management*, *33*(6), 773–794.

Jensen, P. H. (2010). Exploring the uses of matched employer–employee datasets. *Australian Economic Review*, *43*(2), 209–216.

Kaiser, U., Kongsted, H. C., & Rønde, T. (2015). Does the mobility of R&D labor increase innovation? *Journal of Economic Behavior & Organization*, *110*, 91–105.

Kinne, J., & Lenz, D. (2021). Predicting innovative firms using web mining and deep learning. *PLOS ONE*, *16*(4), e0249071.

Maliranta, M., Mohnen, P., & Rouvinen, P. (2009). Is inter-firm labor mobility a channel of knowledge spillovers? Evidence from a linked employer–employee panel. *Industrial and Corporate Change*, *18*(6), 1161–1191.

Nieminen, U. (1973). On the centrality in a directed graph. *Social Science Research*, *2*(4), 371–378.

Ozman, M. (2009). Inter-firm networks and innovation: a survey of literature. *Economic of Innovation and New Technology*, *18*(1), 39–67.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *PageRank: Bringing order to the web* (tech. rep.) [Preprint (Online); accessed 12.02.2023]. Stanford Digital Libraries Working Paper. http://ilpubs.stanford.edu/422/1/1999-66.pdf

Rahko, J. (2017). Knowledge spillovers through inventor mobility: The effect on firm-level patenting. *The Journal of Technology Transfer*, *42*(3), 585–614.

Rammer, C., Doherr, T., Krieger, B., Marks, H., Niggemann, H., Peters, B., Schubert, T., Trunschke, M., & von der Burg, J. (2021). Indikatorenbericht zur Innovationserhebung 2020 [Preprint (Online); accessed 30.01.2023]. https://ftp.zew.de/pub/zew-docs/mip/20/mip_2020.pdf

Rammer, C., & Es-Sadki, N. (2022). Using Big Data for Generating Firm-Level Innovation Indicators–A Literature Review [Preprint (Online); accessed 15.02.2023]. *ZEW – Leibniz - Centre for European Economic Research Discussion Paper*, (22-007). https://ftp.zew.de/pub/zew-docs/dp/dp22007.pdf

Schilling, M. A., & Phelps, C. C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management Science*, *53*(7), 1113–1126.

Schmucker, A., Seth, S., & Eberle, J. (2014). WeLL-Befragungsdaten verknüpft mit administrativen Daten des IAB [Preprint (Online); accessed 18.04.2023]. *FDZ Datenreport 01/2014 DE (Dokumentation zu Arbeitsmarktdaten)*. https://doku.iab.de/fdz/reporte/2014/DR_01-14.pdf

Schwierzy, J., Dehghan, R., Schmidt, S., Rodepeter, E., Stoemmer, A., Uctum, K., Kinne, J., Lenz, D., & Hottenrott, H. (2022). Technology mapping using WebAI: The case of 3D printing [Preprint (Online); accessed 15.02.2023]. https://arxiv.org/abs/2201.01125

Scott, J. (2017). *Social Network Analysis, 4th Edition* [ISBN: 9781473952126]. SAGE Publications Ltd.

Somaya, D., Williamson, I. O., & Lorinkova, N. (2008). Gone but not lost: The different performance impacts of employee mobility between cooperators versus competitors. *Academy of Management Journal*, *51*(5), 936–953.

van der Wouden, F., & Rigby, D. L. (2021). Inventor mobility and productivity: A long-run perspective. *Industry and Innovation*, *28*(6), 677–703.

Veltri, G. A. (2013). Microblogging and nanotweets: Nanotechnology on Twitter. *Public understanding of science*, *22*(7), 832–849.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and Applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815478

Weinhardt, M. (2016). *SOEP-LEE Betriebsbefragung-Datenhandbuch der Betriebsbefragung des Sozio-oekonomischen Panels* (tech. rep.) [Preprint (Online); accessed 18.04.2023]. SOEP Survey Papers 306: Series D. https://www.econstor.eu/handle/10419/130173

Weinhardt, M., Meyermann, A., Liebig, S., & Schupp, J. (2016). The Linked Employer–Employee Study of the Socio-Economic Panel (SOEP-LEE): Project Report [Preprint (Online); accessed 18.07.2023]. *SOEPpapers on Multidisciplinary Panel Data Research 829-2016*. https://www.diw.de/de/diw_01.c.530102.de/publikationen/soeppapers/2016_0829/the_linked_employer-employee_study_of_the_socio-economic_panel__soep-lee___project_report.html

Weinhardt, M., Meyermann, A., Liebig, S., & Schupp, J. (2017). The Linked Employer–Employee Study of the Socio-Economic Panel (SOEP-LEE): Content, Design and Research Potential. *Jahrbücher für Nationalökonomie und Statistik*, *237*(5), 457–467.

Wu, L., Jin, F., & Hitt, L. M. (2018). Are all spillovers created equal? A network perspective on information technology labor movements. *Management Science*, *64*(7), 3168–3186.

A Appendix: Additional Figures and Tables

Table A.1: Definitions of Employment Characteristics

	Original definition	New definition
Discipline	Administration, Clerking	Administration
	Purchasing, Materials Management, Logistics	Administration
	Human Resources	Administration
	Management and Corporate Development	Management
	Product Management	Management
	Project Management	Management
	Analysis and Statistics	IT & Data
	IT and Software Development	IT & Data
	Consulting	Consulting
	Law	Consulting
	Controlling and Planning	Finance
	Finance, Accounting and Controlling	Finance
	Customer Service and Support	Sales
	Distribution and Trade	Sales
	Research, Teaching and Development	R&D
	Health, Medicine and Social Affairs	Social & Others
	Other Fields of Activity	Social & Others
	Graphics, Design and Architecture	PR & Marketing
	PR, Public Relations and Journalism	PR & Marketing
	Marketing and Advertising	PR & Marketing
	Engineering and Technical Professions	Engineering
	Process Planning and Quality Assurance	Engineering
	Production and Craft	Production
Career level	Intern/Student	Intern/Student
	Job starter	Young Professional
	With Job Experience	Professional
	Managers (With/Without Pers. Responsibility)	Manager
	Director (Division Manager, VP, SVP etc)	Director
	Managing Director (CEO etc)	Managing director
Employment type	Civil Servant	Civil Servant
	Honorary	Unpaid
	Freelancer	Freelancer
	Freelancing	Freelancer
	Self Employed	Freelancer
	Recruiter	Freelancer
	Shareholder	Partner
	Shareholder/Partner	Partner
	Owner	Partner
	Partner	Partner
	Intern	Intern
	Parttime	Parttime
	Fulltime	Fulltime
	Member of the Board	Fulltime

Data processing of employment characteristics (XING). The original categories are translated from German into English. The discipline, career level and employ categories of ment type are mapped to more coarse granular groups. Further and non-existing values are re-coded as missing/others.

Table A.2: Data Processing of Employer Characteristics

	Original data	New data
Sector	05 $\leq$ NACE code $\leq$ 33	Industry
	35 $\leq$ NACE code $\leq$ 43	Utilities/Construction ('Utilities')
	45 $\leq$ NACE code $\leq$ 47	Trade
	49 $\leq$ NACE code $\leq$ 53	Transport
	55 $\leq$ NACE code $\leq$ 56	Hospitality
	58 $\leq$ NACE code $\leq$ 63	Information & Communication ('IT')
	64 $\leq$ NACE code $\leq$ 68	Finance/Insurance/Real Estate/Property and Housing ('Finance')
	69 $\leq$ NACE code $\leq$ 75	Freelance, Scientific and Technical Services ('S&T Services')
	77 $\leq$ NACE code $\leq$ 82	Company Services
	84 $\leq$ NACE code $\leq$ 88	Social Services
	90 $\leq$ NACE code $\leq$ 96	Personal/Cultural Services ('P&C Services')
Founding	founding year $<$ 1990	$<$ 1990
	1990 $\leq$ founding year $\leq$ 1999	1990 - 1999
	2000 $\leq$ founding year $\leq$ 2009	2000 - 2009
	founding year $\geq$ 2010	$\geq$ 2010
Employer size	employer size $<$ 10	$<$ 10
	10 $\leq$ employer size $\leq$ 49	10 - 49
	50 $\leq$ employer size $\leq$ 249	50 - 249
	250 $\leq$ employer size $\leq$ 999	250 - 999
	employer size $\geq$ 1000	$\geq$ 1000

Data processing of employer characteristics according to MUP. Two-digit NACE code, founding year, and employer size data are mapped to categorical variables. The abbreviations of long labels are shown in parentheses. Further and non-existing values are re-coded to missing/others.

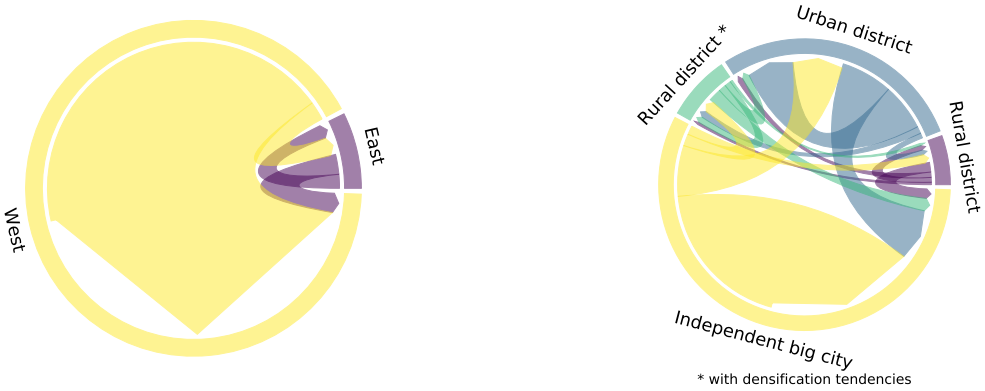


Figure A.1: Left: Flows between East and West German. Data from Berlin was excluded because of the unclear assignment. Right: Flows between the different district types. Table A.8 (Appendix) presents the respective flow matrices. Source: TUM and ZEW based on XING and MUP data. Own illustration.

Table A.3: Legal Form of Employers (MUP)

German label	English label
Aktiengesellschaft (AG)	Public Limited Company
Kommanditgesellschaft (KG)	Limited Partnership
Gesellschaft mit beschränkter Haftung (GmbH)	Limited Liability Company
freie Berufe	Liberal Professions
Gmbh & Co. KG	Gmbh & Co. KG
Eingetragene Genossenschaft (eG)	Registered Cooperative
Limited	Limited
BGB-Gesellschaft - Arbeitsgemeinschaft KG	BGB Society - KG Working Group
Offene Handelsgesellschaft (OHG)	Open Trading Company
Eingetragener Verein (eV)	Registered Association
Firma (Ausland)	Foreign Company
BGB-Gesellschaft	BGB Society
Einzelfirma	Individual Company
Unternehmergesellschaft (UG)	Entrepreneurial Company with Limited Liability
Gewerbebetrieb	Commercial Operation
Einzelperson	Single Person
Privatperson (Ausland)	Private Person (Foreign)

Legal form of employers according to MUP. The legal forms are not necessarily always all available in the data analyses, for example, private person (foreign). Left column: Original labels (German). Right column: Translated labels (English).

Table A.4: Geographical Matched Subset

Steps	Observations	Description
1	870K	Load geo-referenced address data (longitude and latitude) for Germany from external data sources. Jan Kinne kindly provided the data.
2	58K	Load XING employers matched to the MUP. We restrict the data to employers located in Mannheim or Munich. The employers are selected based on district numbers 8222 and 9162.
3	57K	Load location data for matched XING employers. We remove all observations if no postal code or address data is available in the MUP.
4	57K	The MUP addresses consist of a postal code and address. We split the addresses into street names and numbers (heuristic).
5	57K / 870K	Standardise the address text data. For example, transform the text into lowercase and treatment of special characters (ä, ü., ö, ß).
6	48K	Link 57K employers to 870K geo-referenced address candidates. High-quality matches: Check if postal code, street name & number is substring of candidate addresses. Medium quality matches: Check if the postal code and street name are substrings of candidate addresses. Merge both match data sets, but prefer high quality over medium quality matches. If there are several matches within one quality class, select the first one (heuristic). Result: Mapping of geo-coordinates to XING employers.

Matching a subset of XING employers to geographical coordinates. The match is based on a heuristic and does not claim to be error-free. Number of employers in thousand (K).

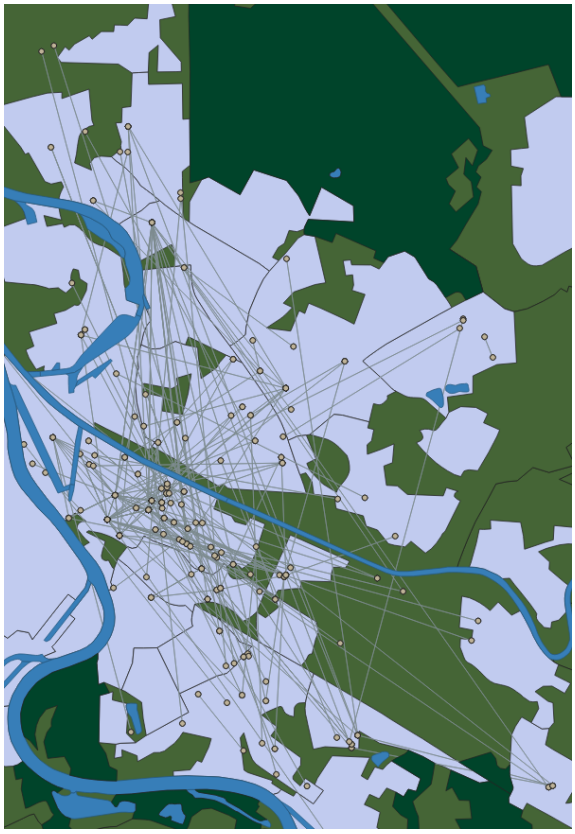


Figure A.2: Employee flow network for the city of Mannheim. The flow data is restricted to flows within the city in 2019. Source: TUM and ZEW based on XING data. Own illustration (created with QGIS; https://www.qgis.org).

Table A.5: Definition of District Types

Type	Description
Rural district	Population share in large and medium-sized cities below 50% and population density excluding large and medium-sized cities below 100 inhabitants/km^2.
Rural district with densification tendencies	Population share in large and medium-sized cities of at least 50%, but a population density of fewer than 150 inhabitants/km^2, a population share in large and medium-sized cities of less than 50% with a population density without large and medium-sized cities of at least 100 inhabitants/km^2.
Urban district	Population share in large and medium-sized cities of at least 50% and a population density of at least 150 inhabitants/km^2; a population density without large and medium-sized cities of at least 150 inhabitants/km^2.
Independent big city	At least 100K inhabitants.

Source: https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/kreise/siedlungsstrukturelle-kreistypen/kreistypen.html.

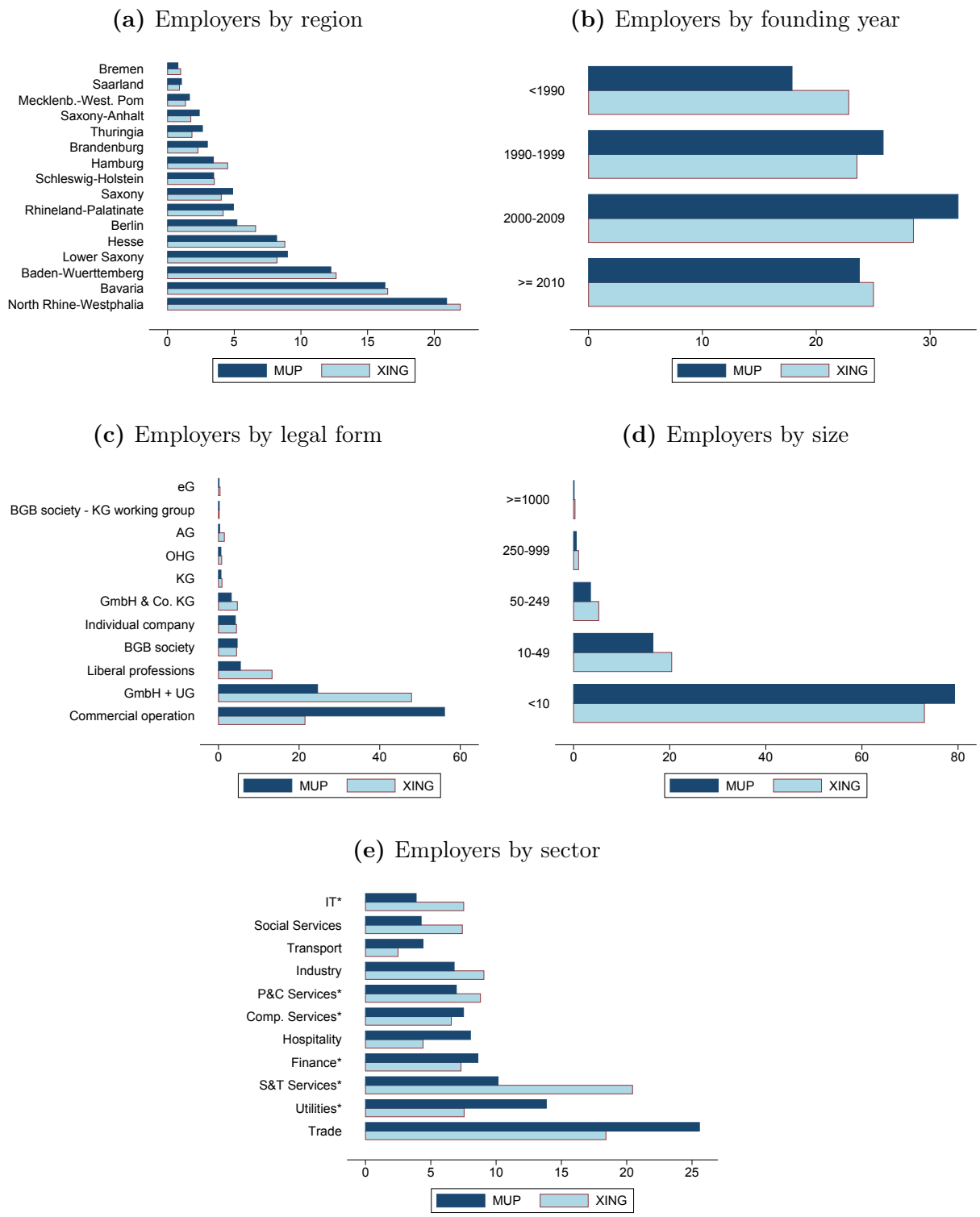


Figure A.3: Share of MUP and XING employers with respect to the employer characteristics region, founding year, legal form, size, and sector. The numbers sum up to one hundred percent for each data source. The MUP data is restricted to unique German employers that are listed in 2002 to 2019 (we use the latest entry). Missing values are not shown. Labels marked with star (*) are listed in abbreviated form (see Table A.2 for full labels). Source: TUM and ZEW based on XING and MUP data. Own illustrations.

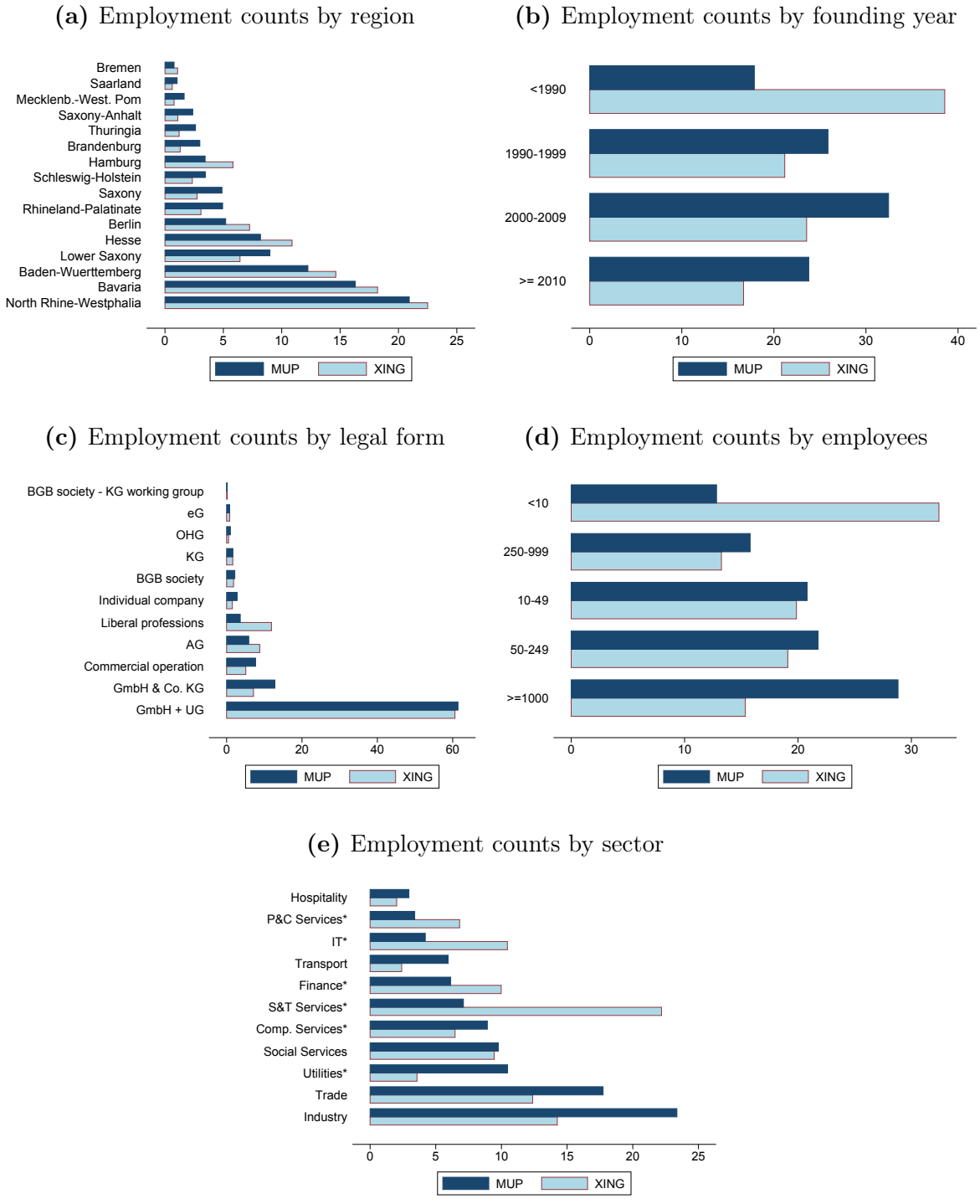
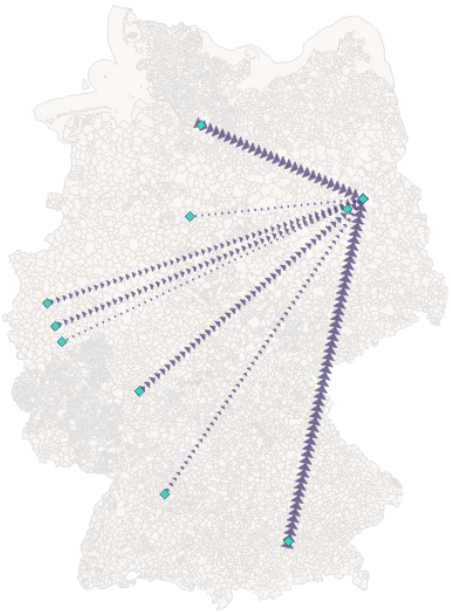
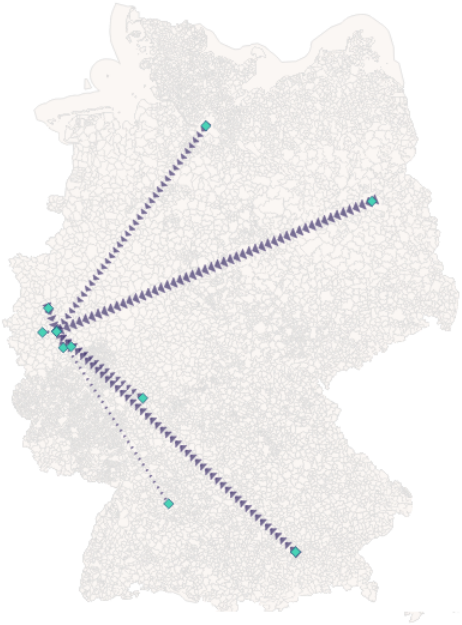


Figure A.4: Share of MUP and XING employees by employer characteristics region, founding year, legal form, size, and sector. The numbers sum up to one hundred percent for each data source. The MUP data is restricted to unique German employers listed from 2002 to 2019 (we use the latest entry for the employment counts). Not extrapolated and, therefore, includes only a subset of all employments. Missing values are not shown. Labels marked with a star (*) are listed in abbreviated form (see Table A.2 for list of all labels). Source: TUM and ZEW based on MUP and XING data. Own illustrations.

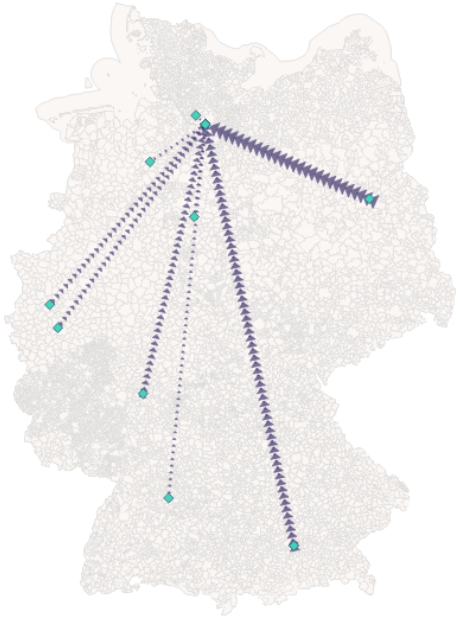
(a) Inflows for Berlin



(b) Inflows for Cologne



(c) Inflows for Hamburg



(d) Inflows for Munich

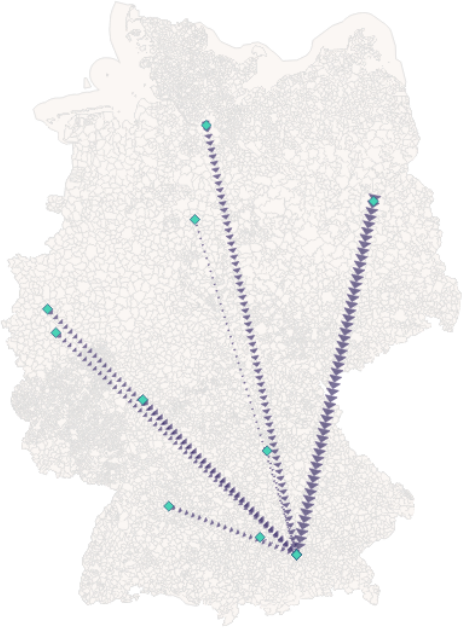
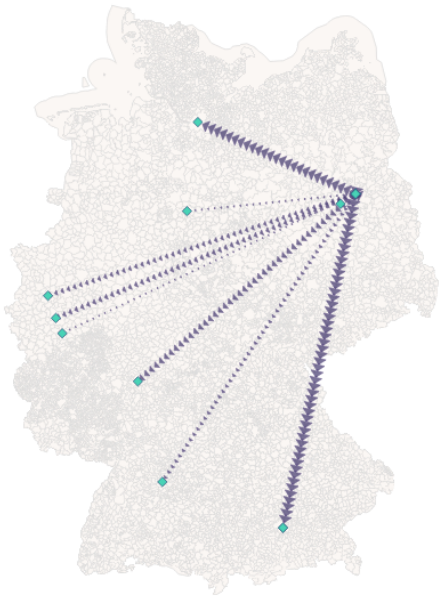
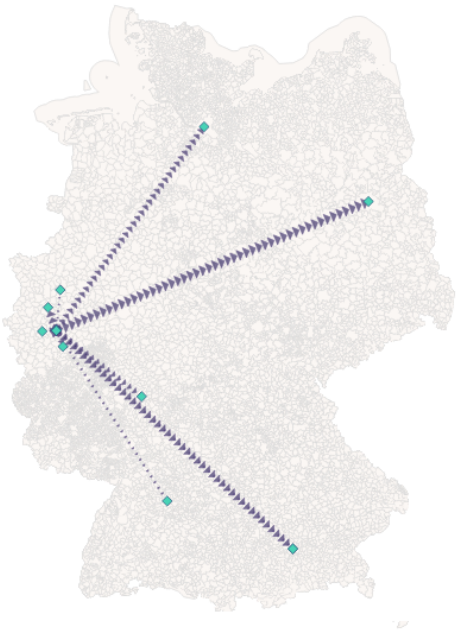


Figure A.5: Top 10 districts with the most flows to the German cities Berlin, Cologne, Hamburg, and Munich. Some districts may overlap in their location, e.g., the city and region of Munich. The edge thickness illustrates the relative flow count in the flow set. Source: TUM and ZEW based on XING and MUP data. Own illustration (created with QGIS; https://www.qgis.org).

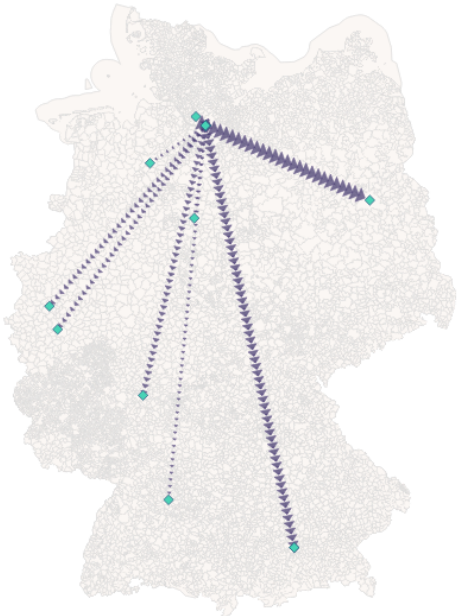
(a) Outflows for Berlin



(b) Outflows for Cologne



(c) Outflows for Hamburg



(d) Outflows for Munich

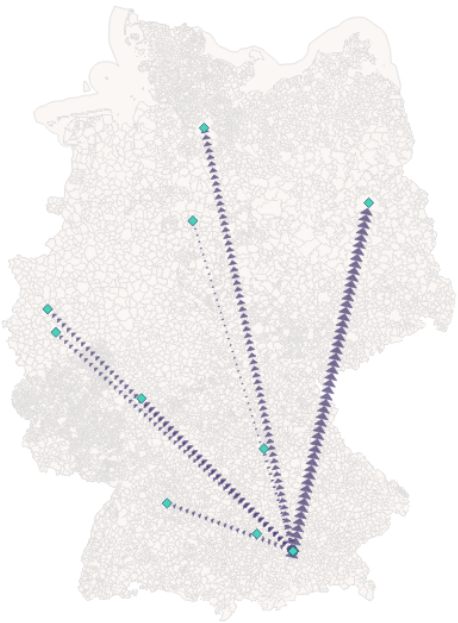


Figure A.6: Top 10 districts with the most flows from the German cities Berlin, Cologne, Hamburg, and Munich. Some districts may overlap in their location, e.g., the city and region of Munich. The edge thickness illustrates the relative flow count in the flow set. Source: TUM and ZEW based on XING and MUP data. Own illustration (created with QGIS; https://www.qgis.org).

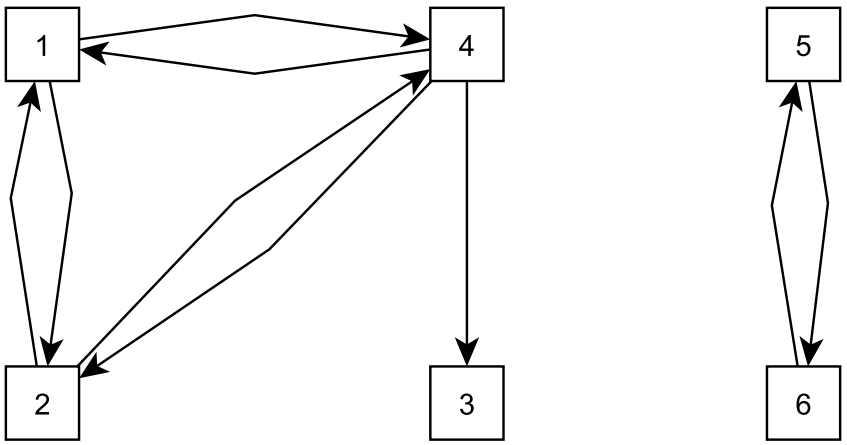


Figure A.7: Example graph and corresponding network metrics: #Nodes = 6, #Edges = 9, #Cliques = 12, Maximum clique size = 3, Transitivity = 0.777, Reciprocity = 0.888, #Clusters = 3, Density = 0.3, Girth = 3. Own illustration (created with yEd - graph editor; https://www.yworks.com).

Table A.6: Employee Flow Matrices

Discipline		New employment										
		Administration	Consulting	Engineering	Finance	IT & Data	Management	PR & Marketing	Production	R&D	Sales	Social & Others
Old employment	Administration	**57.79**	4.71	2.12	4.26	3.40	6.01	4.20	1.41	2.09	8.69	5.32
	Consulting	6.59	**54.01**	2.30	4.52	7.16	9.00	3.91	0.57	2.68	5.40	3.85
	Engineering	2.64	2.20	**65.71**	1.05	4.10	7.44	1.11	2.71	7.19	3.00	2.85
	Finance	8.33	6.83	1.89	**59.65**	3.30	6.20	2.40	0.80	1.62	5.45	3.54
	IT & Data	2.57	4.87	3.18	1.45	**73.13**	4.55	1.98	0.53	3.23	2.47	2.05
	Management	6.34	8.10	6.18	2.98	5.53	**45.89**	7.38	1.70	3.21	8.23	4.46
	PR & Marketing	4.37	3.30	0.96	1.13	2.65	6.93	**68.04**	0.69	2.22	5.51	4.20
	Production	6.29	1.76	12.46	1.65	2.58	6.86	2.59	**47.16**	5.84	7.04	5.77
	R&D	3.66	4.27	12.09	1.25	6.49	6.48	3.45	1.85	**52.23**	2.63	5.60
	Sales	9.56	4.19	2.06	2.80	3.21	7.14	5.31	1.40	1.35	**57.73**	5.25
	Social & Others	8.88	3.89	3.86	2.56	3.29	5.26	5.84	2.48	4.84	8.42	**50.68**

Type		New employment						
		Civil servant	Freelance	Fulltime	Intern	Partner	Parttime	Unpaid
Old employment	Civil servant	**61.27**	4.66	25.46	2.54	1.04	4.03	0.97
	Freelance	0.10	**44.15**	35.68	4.91	5.57	8.34	1.21
	Fulltime	0.10	4.47	**84.64**	2.53	2.57	5.17	0.50
	Intern	0.03	5.14	34.46	**38.28**	0.68	19.49	1.88
	Partner	0.06	15.29	47.47	1.87	**28.79**	5.22	1.27
	Parttime	0.11	6.31	40.64	15.59	1.56	**34.35**	1.41
	Unpaid	0.20	9.42	33.30	20.14	2.38	18.20	**16.31**

Career level		New employment					
		Student/Intern	Young Professional	Professional	Manager	Director	Managing Director
Old employment	Student/Intern	**64.22**	21.27	11.83	1.73	0.22	0.70
	Young Professional	10.60	**23.65**	56.94	6.76	0.69	1.34
	Professional	2.71	3.42	**74.46**	15.16	1.57	2.65
	Manager	0.85	0.91	16.26	**63.40**	11.48	7.08
	Director	0.56	0.44	7.57	17.76	**50.72**	22.91
	Managing Director	1.91	1.70	17.59	18.11	14.66	**46.00**

Top: Discipline. Middle: Type. Bottom: Career level. Flows are based on matched and public employments. Missing values are not shown. Source: TUM and ZEW based on XING data.

Table A.7: Employee Flow Matrices

Sector		New employment										
		Finance*	S&T Services*	Hospitality	Trade	Industry	IT*	P&C Services*	Social Services	Transport	Company Services	Utilities*
Old employment	Finance*	**56.62**	12.79	0.74	5.10	5.84	5.48	3.14	3.79	1.17	3.72	1.63
	S&T Services*	5.65	**52.86**	0.76	6.47	8.78	8.57	4.19	5.38	1.14	4.27	1.93
	Hospitality	5.53	13.28	**44.12**	6.48	4.87	4.27	5.73	6.41	1.69	6.30	1.32
	Trade	4.83	13.64	0.76	**48.50**	10.24	7.26	3.45	3.94	1.36	4.20	1.82
	Industry	3.98	13.82	0.43	7.57	**57.20**	4.38	2.71	4.13	0.93	3.09	1.76
	IT*	4.62	15.65	0.41	6.08	4.84	**53.60**	4.74	3.99	0.84	4.16	1.07
	P&C Services*	5.15	16.61	1.27	6.13	6.85	9.94	**34.34**	11.59	1.29	5.14	1.68
	Social Services	4.31	15.64	0.92	4.65	7.28	6.31	7.84	**46.52**	1.01	4.02	1.50
	Transport	5.43	12.01	0.95	6.70	6.66	4.81	3.39	4.05	**48.82**	5.35	1.83
	Company Services	5.84	15.83	1.26	6.85	7.27	8.44	4.84	5.59	1.91	**40.19**	1.97
	Utilities*	5.72	16.15	0.79	6.88	9.75	4.51	3.65	5.16	1.37	4.51	**41.52**

Employees		New employment				
		<10	10-49	50-249	250-999	>=1000
Old empl.	<10	**50.97**	17.47	14.35	8.44	8.74
	10-49	23.11	**45.66**	15.25	8.15	7.81
	50-249	18.28	14.71	**49.62**	8.98	8.38
	250-999	15.83	11.53	13.14	**49.27**	10.20
	>=1000	14.81	10.11	11.13	9.24	**54.68**

Region		New employment															
		Baden -Württemberg	Berlin	Brandenburg	Bremen	Hamburg	Bavaria	Saxony	Thuringia	Hesse	Meckl.-West. Pomerania	Lower Saxony	North Rhine -Westphalia	Rhineland -Palatinate	Saarland	Saxony -Anhalt	Schleswig -Holstein
Old employment	Baden-Württemberg	**65.40**	2.89	0.42	0.37	1.98	9.35	0.97	0.81	4.81	0.21	2.05	7.58	1.81	0.33	0.37	0.65
	Berlin	5.49	**55.05**	2.57	0.44	4.26	8.87	1.51	0.54	5.32	0.53	2.54	9.90	1.22	0.19	0.62	0.96
	Brandenburg	5.76	18.35	**36.93**	0.52	3.27	7.82	3.03	0.81	4.57	1.00	3.17	10.67	1.33	0.25	1.14	1.37
	Bremen	5.47	3.43	0.48	**47.39**	6.84	6.78	1.04	0.42	4.24	0.54	10.77	9.11	1.03	0.16	0.47	1.83
	Hamburg	4.61	5.13	0.53	1.09	**56.13**	7.13	0.82	0.36	4.55	0.77	4.13	9.18	0.89	0.15	0.38	4.13
	Bavaria	7.38	3.72	0.46	0.36	2.52	**65.93**	1.22	0.57	4.83	0.29	2.15	8.20	1.07	0.19	0.43	0.68
	Saxony	6.07	4.84	1.16	0.44	2.17	8.92	**55.00**	1.56	4.06	0.46	2.72	8.38	1.13	0.26	2.04	0.81
	Thuringia	11.07	4.26	0.80	0.35	2.17	9.85	3.78	**45.60**	5.37	0.40	3.37	8.90	1.54	0.25	1.38	0.92
	Hesse	6.95	4.05	0.47	0.45	3.07	8.86	0.99	0.57	**56.92**	0.28	2.49	10.93	2.37	0.32	0.42	0.86
	Meck.-West.Pomerania	5.00	6.53	1.49	0.92	7.47	7.90	1.81	0.67	4.33	**43.51**	4.44	9.61	1.12	0.22	0.88	4.09
	Lower Saxony	5.49	3.62	0.60	1.98	4.90	6.87	1.17	0.64	4.47	0.51	**54.13**	11.69	1.15	0.19	0.89	1.69
	North Rhine-Westphalia	5.35	3.65	0.53	0.45	2.83	7.11	1.03	0.47	5.24	0.30	3.23	**66.75**	1.50	0.24	0.44	0.91
	Rhineland-Palatine	10.03	3.41	0.53	0.40	2.21	7.24	1.01	0.57	9.60	0.27	2.44	11.98	**47.93**	1.12	0.44	0.84
	Saarland	9.46	3.04	0.46	0.34	2.01	6.92	1.27	0.56	6.16	0.26	2.12	9.74	5.74	**50.65**	0.41	0.86
	Saxony-Anhalt	5.95	5.42	1.32	0.49	2.75	8.19	5.65	1.76	4.41	0.62	5.71	9.84	1.29	0.23	**45.16**	1.20
	Schleswig-Holstein	4.94	3.62	0.71	0.91	13.18	6.25	1.10	0.43	3.96	1.26	4.68	9.06	1.06	0.21	0.49	**48.14**

Top: Sector. Middle: Employer size. Bottom: Region. Flows are based on matched and public employments. Each row has been normalised, to sum up to 100 percentage points. Missing values are not shown. Labels marked with a star (*) are listed in abbreviated form (see Table A.2 for the list of all labels). Source: TUM and ZEW based on XING data.

Table A.8: Employee Flow Matrices

East-West		New employment	
		East	West
Old empl.	East	**41.10**	58.89
	West	4.48	**95.51**

	District type	New employment			
		Rural district	Rural district*	Urban district	Independent big city
Old empl.	Rural district	**21.71**	10.76	23.93	43.58
	Rural district*	7.99	**21.73**	25.74	44.52
	Urban district	4.43	6.38	**42.62**	46.55
	Independent big city	3.94	5.33	22.73	**67.99**

Top: East-West. Bottom: District type. Flows are based on matched and public employments. Each row has been normalised, to sum up to 100 percentage points. The class 'Rural district*' is the abbreviation for 'Rural district with densification tendencies'. Missing values are not shown. Source: TUM and ZEW based on XING data.

Download ZEW Discussion Papers:

https://www.zew.de/en/publications/zew-discussion-papers

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim
ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de