Dynamics of skills demand and job transition opportunities: A machine learning approach[1]

Patricia Prüfer, Pradeep Kumar, and Marcia den Uijl[2]

12 April 2019

FOR CONFERENCE PURPOSES ONLY – PLEASE DO NOT CITE OR DISTRIBUTE

What are consequences of the ongoing digitalization and automatization for the labor market? We analyze to which extent types of occupations and skill requirements change over time and how these insights can be used to substantiate demand for reskilling of several groups of employees. To answer these questions, we make use of a novel approach in which we combine unstructured data from the Internet with structured data from labor market forecasts. Based on a dataset of 95% of all job vacancies in the Netherlands over a 6-year period with 7.7 million data points, we show which skills are particularly important for which type of profession. Besides, we provide job transition opportunities for employees from shrinking sector or occupations to sectors and professions not affected negatively by technological change. Our results suggest that the labor market is undergoing a transitions from degree-based to skill-based demand. This has consequences for both the participants and the institutions connected to the labor market.

Introduction

Today, our world is full of devices and applications that generate, store, and transmit huge amounts of information, producing enormous quantities of data available for all kinds of analyses. Together with technological innovations, such as the Internet of Things or sensors in numerous types of devices, for example cars, homes, and offices, these fast growing volumes and varieties of available data, paired with cheaper and more powerful computational processing and affordable data storage, have constituted the "rise of big data" or "datafication" (Mayer-Schönberger and Cukier, 2013).

These developments have not gone unnoticed by many organizations that are consequently inclined to work more data driven and fact based. In parallel, data and technology ubiquity is likely to have direct effects on individual as well as group behavior and on the nature of social, organizational, and economic structures (Bessen, 2015). The common denominators in these developments are digitalization, automation, and Information and Communications Technology (ICT). On the other hand, there are risks of undesirable applications of new technology posing

[1] This research is partly funded by the Dutch Ministry of Social Affairs and Employment under the DWSRA grant and partly by a large consortium consisting of CA-ICT, ICT Netherlands, CIO Platform Netherlands and six of the ten so-called Dutch 'top sectors': Chemical Industry, Energy, High-tech Industries, and Life Sciences and Health.

[2] CentERdata, Tilburg University.

security challenges, such as cybercrime or privacy breaches. Digitalization, automation and the development of new (adaptable) technologies also have an increasing impact on the labor market where the boundaries between 'ICT people' and other professions in which ICT-related skills are required are becoming increasingly blurred and the specific skills demands in all occupations have considerably changed in recent years. Moreover, the specific skills demanded and the tasks that have to be fulfilled in all occupations have changed considerably in recent years (Spitz-Oener, 2006).

This is already reflected in a large quantitative and qualitative demand for ICT experts and employees with sufficient digital skills in many countries, including the Netherlands (ROA, 2017). The shortage in personnel creates not only pressure within the ICT sector, but also between sectors and job categories. For employees who, in the longer term, cannot acquire the necessary digital skills through training and retraining, suitable measures must be found that offer a solution from a situation of insufficient qualifications and, eventually, unemployment. On the other hand, research has shown that employees can adapt sufficiently to the changes on the labor market and that the effects of digitalization and automation might be exaggerated, as many jobs may change but also new jobs will be created (Autor, 2015; Arntz et al., 2016). Given that innovative capacity is directly related to economic growth, a lack of people with sufficient digital, technical, and ICT skills, in combination with a broader set of so-called 21st century skills, limits innovative capacity (Obschonka et al., 2017). Alternatively, abundance of these types of employees helps to mitigate not negative effects on innovative capacity and the labor market (Elliott, 2017; Mc Afee and Brynjolfsson, 2017).

What do these developments mean for the (Dutch) labor market and for the various sectors? What are the dynamics in skill demand on the labor market? How do they affect certain types of professions such as managers, ICT professionals, and employees in non-IT/technical jobs? Will these developments eventually lead to new labor market structures, for example by offsetting clearly separated sectors? To answer these questions, we make use of a novel approach of 'labor market analytics' in which information from online vacancies, thus from unstructured (big) Internet data, is combined with information from labor market forecasts, that is, with structured data from administrative sources.

In a recent report, the World Economic Forum (WEF, 2018) used a similar approach for the US labor market. This research on the 'reskilling revolution' in the US investigated the possibilities for switching for employees in shrinking occupations and to what extent retraining and reskilling is necessary, on the basis of their big data job bank and analysis tools. Based on a data-driven approach, WEF (2018) identifies unexpected switching possibilities from shrinking sectors and occupations to sectors and/or occupations that are not shrinking. Apart from that, technological progress also has negative effects on some professions and sectors, and disproportionately on women, who often work in shrinking occupations and sectors. Our paper extends the WEF approach by also analyzing dynamics in the demand for skills whereby we allow for a distinction between digital and technical skills on the one hand and general cognitive and non-cognitive skills on the other hand.

What we do in this paper

In this paper, we derive insights on the consequences of the ongoing digitalization and automatization for the Dutch labor market and on the extent to which occupations have change over time. More specifically, we analyze the effects of technological change on demand for and supply of specific occupations, competencies, skills, and tasks, and the implications for employees with varying experience and education levels. We distinguish three types of professions: managers, ICT jobs, and non-ICT jobs. This helps to understand how the requirements have changed over time and among types of professions and, thus, not only provides insights into ongoing skills dynamics, but also on the need for additional qualifications and retraining of specific groups.

The following research questions are answered.

RQ 1: How are digital skills affected by digitalization and automation?

The objective of this research is to study the impact of digitalization in the key sectors such as ICT, Life Sciences and Health, Chemical Industry, Creative Industry, Energy, and High-tech Industries. This part of the study aims to measure the variation in skill requirements within these sectors over time. We consider technical and digital skills, more general cognitive and non-cognitive skills, and different types of abilities. This helps to understand how requirements have changed over time. It also gives insight in the need for retraining due to digitalization and automation and helps to estimate the need for modifications in educational programs.

RQ 2: What is the transition viability, i.e. similarity, between different job pairs?

The second objective of this study is to theoretically assess the feasibility of various job transitions on the Dutch labor market. Our data-driven approach is based on measuring the similarity of two jobs based on their published job descriptions on different job portals. We analyze the texts of job descriptions, containing information about candidates' required knowledge, skills, abilities, and experiences. This information is used to construct a similarity index between any pair of jobs.

RQ 3: How are career perspectives affected by trends on the labor market?

In a final step, we intend to match (the insights from) big data with forecasts on labor market dynamics for sectors and occupations. This will allow us to derive insights on how general trends on the labor market influence the career perspectives of employees in various sectors and occupations. Moreover, we will show to which extent employees in occupations that may disappear, can transfer to alternative jobs and, thus, what the impact of digitalization and automation will be on unemployment and need for training. Will we face an overwhelming increase in unemployment, or will the problem solve itself in a natural way, such that employees can utilize their skill sets in alternative jobs (potentially after some extra training)?

Data and methods

Data

Online vacancy data

We use data from the vacancy database *Jobfeed*, which is administered by TextKernel, a tech company.[3] This online job portal contains more than 95% of all vacancies published on the Dutch labor market in the last ten years. Therefore, it offers a nearly complete - and hence nearly representative - data set of online job ads in the Netherlands. Jobfeed searches the Internet for new vacancies on a daily basis and applies ML algorithms to crawl for vacancies and filter out redundancies. The data mainly contain (unstructured) text, but Jobfeed also extracts structured data such as profession, education, location and company name. Table 1 gives an overview of the information contained in *Jobfeed*.

Table 1: Information used from Jobfeed database

Variable	**Description**
Date	Date on which the vacancy was found, for analyzes per year
ISCO-08 code	ISCO-08 code, indicate type of job
Organization activity	The main activity of the organization in the Standard Industrial Classification (SBI) indicates the sector
Job description	Description of the job
Candidate description	Description of the candidate, also contains requested skills, training, experience, and education level

We use data for a period of 6 years, from January 2012 until December 2017, in total about 7.7 million vacancies. Most of the vacancies are written in Dutch; about 8% are in English. As long as a candidate or job description is available, we use all vacancies in our analyses; this holds for 7.32 million vacancies relating to 371 different occupations. The candidate and job descriptions contain relevant information about required skills, experience, and education.

In addition, we use information gathered from multiple sources, including the Occupational Information Network (O*NET), an online database with information about the knowledge, skills, tasks, training and experience required for a large number of occupations. Another data source is ISCO (International Standard Classification of Occupations; version ISCO-2008), a classification of 436 professions supplied by the International Labor Organization (ILO). In the ISCO-08 classification, a profession has a skill level (1 to 4) and is a combination of the nature of the work,

[3] See https://www.textkernel.com/hr-software/jobfeed/.

the required training, and the required experience. Other sources for skills data we used include the EU skills framework, Stackoverflow, and Dbpedia, Wikipedia's skills database.[4] Table 2 provides an overview of the numbers and types of skills included in this analysis and gives some examples.

Table 2: Number of types of skills and examples

Type	Category	# skills	Examples
Digital	One symbol/letter, one word or abbreviation	1837	C, C++, R, Java, CAD/CAM, SAP, ERP, Excel, SQL
	More than one word	3860	Object-oriented programming, test automation, internet of things, windows XP, Microsoft exchange server, IT governance, database management system
Other	One symbol/letter, one word or abbreviation	195	Communication, flexibility, leadership
	More than one word	121	Critical thinking, planning and organisation, problem solving

This approach is not without caveats. Vacancy data are not necessarily representative and we do not know who applies for a certain vacancy and who is employed in the end. On the other hand, (online) vacancies give a much more fine-grained and real-time picture of labor market demand and, therefore offers a "great opportunity for real-time monitoring of the labor market" (Boselli et al., 2017). This type of data source can provide information over a longer period of time, for a larger sample, and across various locations. Moreover, vacancy data are less prone to response and recall bias, which are eminent in survey data --- even more so as it is fairly expensive to place a (clearly visible and widely distributed) vacancy. Finally, vacancy data are much cheaper than other sources of information such as questionnaires within a representative sample or register data that have to be linked from multiple sources. Vacancy data are, thus a very rich and suitable source for what is called Labor Market Intelligence (LMI) nowadays (Kurekova et al., 2015).

Labor market forecast data

We received data on labor market projections from the Research Center for Education and the Labor Market (ROA). For its medium-term, six-year period, labor market forecasts ROA uses a flow-based approach of labor force flows to and from the labor market. The most recent forecasts until 2022 are based on labor market developments by economic sector, occupational group, and type of education between 1996 and 2016.The forecasting model itself is based on (explanatory)

[4] More information can be found on the various websites: https://www.onetonline.org/help/onet/database; http://dbpedia.org/page/Category:Skills; https://stackoverflow.com/tags?page=1&tab=popular.

econometric models developed by ROA. The most important data inputs of the model include: 1) the Labor Force Survey (LFS) provided by Statistics Netherlands (CBS), 2) employment forecasts by industry sectors (by Panteia), 3) baseline forecasts from the Ministry of Education, Culture and Science (OCW) for the inflow of students in the labor market, and 4) data from ROA's School Leaver Information System (SIS).

The ROA forecasts allow for the dynamic interplay between submarkets as the method includes substitution between different labor market segments. Forecasts are generated for a total of 113 occupational groups and 90 types of education, and span the entire labor market. In line with the ILO-definition of employment, the forecasting model counts all persons between the age of 15 and 74 at work for at least 1 hour per week.

Table 3: Information used from labor market forecasts

Variable	**Description**
Expected expansion demand until 2022	Demand for new manpower that arises from growth in employment. If there is a fall in employment, the expansion demand is negative.
Expected replacement demand until 2022	Replacement demand is the demand for new workers that arises from retirement, (temporary) retirement due to care duties, disability, professional mobility, or transfer to other training.
Expected job openings until 2022	Job openings are the total demand for newcomers to the labor market, as determined by employment growth (positive expansion demand) and replacement demand.
Indicator Future Staffing Bottlenecks (ITKB)	This reflects the expected tension by occupation. The ITKB indicates the chance that the desired personnel composition by training can be realized within professional groups, taking into account the expected supply per training. The lower the value of the indicator, the larger the expected bottlenecks.
Average gross hourly wage	Average gross hourly wage of employees in 2016 in euros; based on LFS 2016.

ROA uses the ROA CBS 2014 Occupational Classification (abbreviated BRC-2014) which is a classification derived from ISCO-08. An occupational group contains 1 or more ISCO-08 unit groups (436 in total) and can be linked to the Jobfeed vacancy data also classified in ISCO-08 categories. The average gross hourly wage and ITKB of an ISCO-08 unit group is taken from the BRC-2014 professional group in which the unit group falls. The expected replacement demand, expansion demand, and job openings are calculated in proportion to the share of the ISCO-08 unit group within the BRC-2014 profession group.

Methods

Text analytics

As the collected vacancies from the Internet consist of unstructured text, we apply *Natural Language Processing* (NLP) techniques. In order to be able to use text for quantitative analyses, initially *data pre-processing* has to be done, such as the *stop words (i.e.* articles and prepositions) or high-frequency words that do not provide meaningful information, for instance 'experience' or 'knowledge'. Moreover, we removed structured fields in the Jobfeed database, such as e-mail addresses, telephone numbers and links to websites, by using so-called regular expressions.[5] In a final pre-processing step we normalize the text by applying a stemming algorithm because words can appear in various forms, such as 'required,' 'require,' and 'requiring' or as derived words with a similar meaning, such as 'entrepreneurial,' 'entrepreneur,' and 'entrepreneurship.' To do this in Dutch language we apply an existing algorithm.[6]

The specific NLP tool we use for our analysis is the *bag-of-words* algorithm. This model helps to retrieve information from an unstructured data source by representing a text as the bag (multiset) of its words, disregarding grammar and word order, but keeping information on the frequency of each word and using it as a feature for training a classifier. To make text suitable for analysis, we transformed it into a vector of numbers that relate to the meaning of each word and how it relates to other words. By applying a mathematical distance measure, the difference (distance) between all the words in our text fragments can be calculated.

After the pre-processing steps, we can finally extract all necessary information from our text data. Therefore, we categorize the skills into two unique lists: *digital and technical skills* and *other skills*. Because the vacancies are partly in English, we use both Dutch and English skill labels. We also include as many different forms and expressions of skills as possible based on the frequency of words in the vacancy texts. In addition, to make the extraction process of skills more reliable and robust, the entire list of skills is normalized and divided into two parts - skills that contain one character, one word or an abbreviation, and a second list with skills with more than one word. For both categories, the skills are searched within one vacancy. If the exact skill is found in the text, it is counted and if a skill occurs several times within one vacancy, this counts as one. A *unigram model* was used for the first category. In this model, the text (candidate and job description) is fragmented word by word after all the noise, for instance special characters or white spaces, has been removed by regular expressions. The splitting into words then only needs to happen on a single space while the words can be looked up in the list of skills.[7] For the skills from the second

[5] *Regular expressions are* a sequence of characters that define a search pattern and is, thus, useful for dictionary-based skills extraction. It is also useful for text cleaning operations by matching the specified sequence of characters, for example website links or email addresses, which we remove from our data because this type of information is not relevant for our analysis and could even have negative effects (for instance, web links could be incorrectly recognized as HTML-skills).

[6] This is called the *Dutch Snowball stemmer*, available in the Python NLTK package.

[7] This approach helps to prevent that skills are incorrectly recognized on the basis of only part of a word or sentence, thus to avoid any spurious matching of skills. A special exception is the 'Microsoft Word' skill. This skill is sometimes referred to only as 'Word'. But the Dutch word 'word' and 'Word' is also common. Only the word 'Word', case sensitive, is recognized as a skill, with the exception of cases where it is followed by 'you' or 'then'. The same problem occurs with the programming language 'C', which cannot be erroneously recognized when asked for a driving license C or the Dutch nursing diploma C.

category, skills with more than one word (so-called bigrams or trigrams), we use regular expressions to match the skills after having done the necessary cleaning.

Although lots of frameworks and definitions exist about different types of skills, there is no generally accepted definition of either digital/technical skills. Moreover, there are semantic problems, for example, one job ad could mention 'solution-oriented,' whereas another one requires applicants to be 'capable of solving problems;' often multiple skills fall into the same category. Therefore, we created broader 13 categories for the digital and technical skills and 14 categories for general cognitive and non-cognitive skills (named other skills) reflecting frequently mentioned skills in the standard frameworks, for instance of digital skills and of 21st century skills, and in the literature (see Table A1 and A2 in the Appendix). The digital category 'Digital transformation skills' includes skills and techniques related to recent developments in digitization, such as 3D printing, artificial intelligence (AI), blockchain, cloud computing, cybersecurity, internet of things and robotics.

Calculating the transition viability, i.e. similarity score, between jobs

The vacancy data contains vacancies of 377 ISCO unit groups. As the O*NET data has no information for 6 ISCO unit groups, this results in a 371 x 371 (symmetric) matrix with similarity scores. Following the method established by WEF (2018), our study aggregates these components of job-fit into an index of similarity, or 'similarity scores'. Individual similarity scores are first calculated for both the vacancy data and the O*NET data to combine the advantages of having access to standardized descriptions (in O*NET) and to current up-to-date professional requirements (in Jobfeed). These individual scores are then weighted for an overall similarity score.

The combined candidate and job descriptions of Jobfeed job data and the structured information from O*NET contain information about the required skills, experience, training, knowledge and education, the profile, for a specific profession. The profile of each profession can be expressed in a vector. Two professions can be compared with each other by comparing their vectors with each other by using the cosine similarity as distance metrics. Thus, the similarity score ranges between 0 and 1, where a score of 1 means identical profiles. They can be seen as a proxy measure for the feasibility of transitioning between the two jobs. Job pairs that have a similarity score of 1 can be said to have a perfect fit, while job pairs with a similarity score of 0 have the most remote and imperfect fit. For example, a software developer and an application programmer have a high job-fit with a similarity score of 0.95, while an office clerk and an aerospace engineering technician have a low job-fit with a similarity score of 0.81. We describe high similarity scores as scores of at least 0.9 or higher, medium similarity scores as those between 0.85 and 0.9, and low similarity scores as those below 0.85. We use these similarity scores as a tool to objectively measure the similarity between each pair of our 371 unique job types and create a schema (in essence, a matrix) to identify the job-fit between all jobs in our dataset. Figure A2 in the Appendix depicts the overall job-fit matrix between all 371 types of occupations (categorized by job family) in the Netherlands our dataset. Where a zone is highlighted in dark blue, the corresponding row and column define two occupations with a combined profile that suggests a high degree of job-fit.

Linear optimization for viable job transitions

By themselves, similarity scores provide a useful tool for a systematic and comprehensive comparison of job-fit and for identifying viable job transition options. However, as with any composite index, the scores provide a highly aggregated summary view of the theoretical viability of any given job transition. Additional filter criteria are needed to ensure that the job-fit indicated by the aggregate similarity score stays realistic.

For example, prospective job movers are unlikely to be hired when their work experience and educational background are significantly divergent from the requirements of a job. Therefore, we restrict any job transition to move only 1 skill level up or down us to control for unrealistic or unrewarding moves. The restriction also ensures consistency in the actual level of skills and knowledge use within any given occupation.

Within the full range of possible job transitions, there are a number of transitions that may be viable options—in the sense detailed in the previous section—but which are nevertheless unlikely to represent sustainable or attractive options for the individuals seeking to move jobs concerned. Two parameters capture these concerns: the long-term stability of the target job and its capacity to financially uphold (or improve) the standard of living to which the prospective job mover is currently accustomed.

Some theoretically viable job transitions are unsustainable and undesirable simply because the number of people projected to be employed in this job category is set to decline. In the medium term, a number of current occupations in the Netherlands are forecast to shrink or fully disappear due to technological change (ROA 2017). To identify job transitions that are undesirable due to declining target job numbers, we have used Dutch employment figures for 2016 as well as projections of expected employment change by 2022 from ROA.

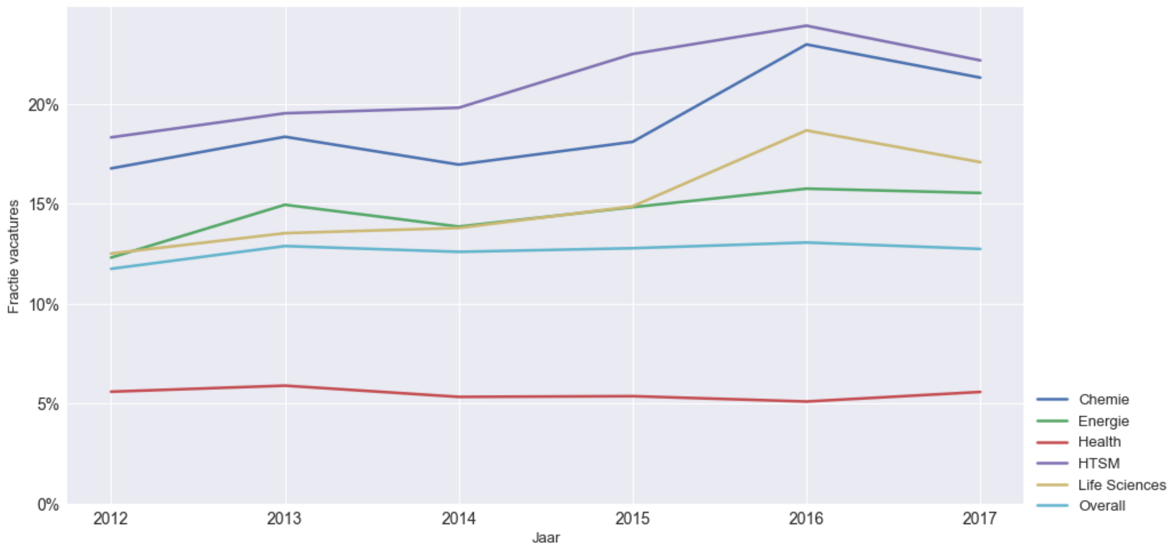
To summarize, in order to be able to say that a viable job transition opportunity represents a desirable job transition option, we require a pairing of a starting job and target job that involves: (1) stable long-term prospects, i.e. a job transition into an occupation with job numbers that are forecast not to decline; (2) relative wage continuity, i.e. the (absolute) difference in wage is not larger than 0.3; (3) there is a high similarity between the starting and target job (at least 0.85); and there is only a move +1 or -1 in skill level (where a job transition from the highest skills level (e.g. CEO's) is only possible at the same level.

What we find in this paper

Results on dynamics in skills demand

Our analyses show that digital and technical skills are becoming increasingly important.[8] This can be seen both in a slight increase from around 12% in 2012 to 13% in 2017 in vacancies in which these types of skills are requested (Figure 1), and in an increase in the total demand for these skills. Thus, more and more digital and technical skills are required per profession meaning that professions become more and more technical. These findings apply to all the sectors we focus at, except for the Health sector which has a stable demand. For the other the increase over the years varies from 15-23% in 2017.[9]

Figure 1: Fraction of vacancies with at least one digital and technical skills

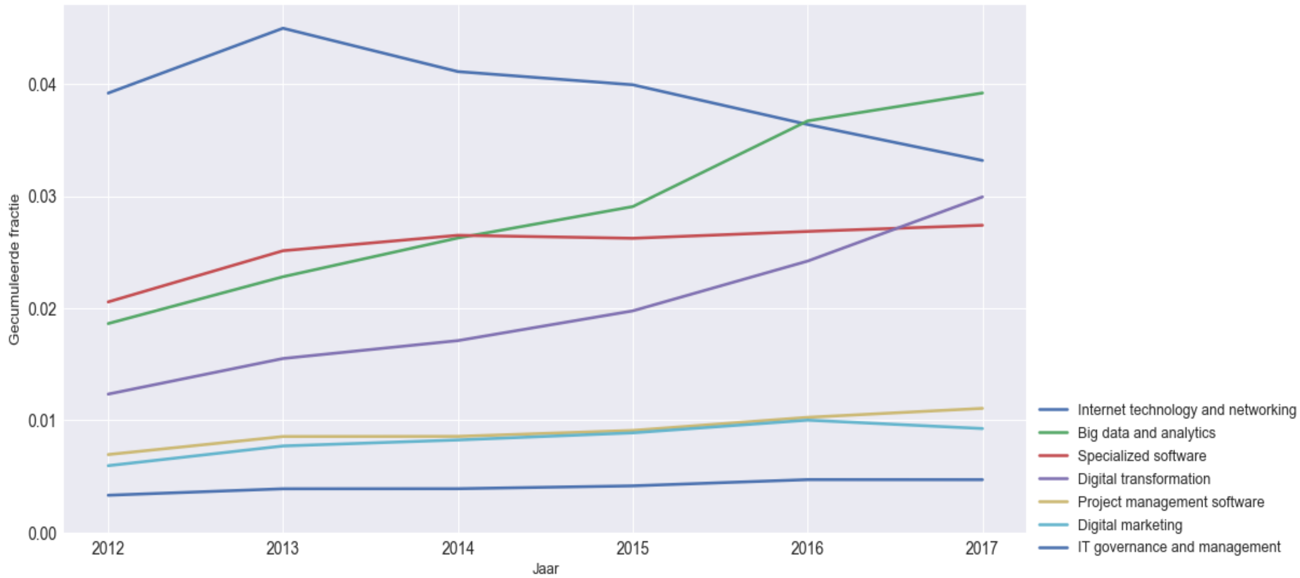


As Figure 2 shows the demand some types of digital and technical skills has increased steeply between 2012 and 2017. This is especially true for the demand for the newest types of digital and technical skills, summarized in 'Big data and analytics' and 'Digital transformation'. Moreover, demand for specialized software skills and project management software is decreasing.

[8] Among the many results we derive in this paper, we only highlight a few, which are also interesting from an Institutional and Organizational Economics (IOE) point of view. The rest of our results can be found in Prüfer et al. (2019), the Dutch policy report we wrote for the Dutch Ministry of Social Affairs and Employment and their partners (see footnote 1).

[9] In some of the sectors, we see a somewhat surprising decline from 2016 to 2017 (in High-tech Industry (HTSM), Chemistry, and Life Sciences). According to TextKernel, this could be due to increased automation which affected demand for certain technical and ICT occupations negatively, for example software testers. Moreover, some of the nowadays very standard skills such as Microsoft Word or Excel could be mentioned less in vacancies as experience in these skills is considered 'natural' these days.

Figure 2: Dynamics in digital skills demand between 2012 and 2017, per category



As can be seen from Figure 3, these developments apply not only to all sectors, but also to all types of professions.[10] In the ranking of demand for at least one digital skill required for a specific type of profession, we see that 'Basic computer skills' and 'Programming skills' are overall very important, except for ICT occupations where 'Basic computer skills' are obviously not important. The categories of 'Digital transformation' and 'Big data and analytics' skills are relatively more demanded by a manager vacancy, while 'Enterprise resource planning (ERP) and management software' and 'Computer-aided design' are most demand for the other occupations.

[10] 10% of all vacancies are ICT professionals, around 7% are managerial positions, and the rest are other positions (including 'unknown'). Managerial positions are defined by the ISCO codes 11, 12, 13, and 14; the codes for executive functions. The number of ICT professions varies greatly per sector. The largest share is in the HTSM, with 19% ICT professions. This is followed by the Energy sector with 9%, Life Sciences with 5%, and Chemistry with 4.3%. In the Health sector 1 in 50 vacancies is an ICT vacancy. The number of vacancies for ICT professions and the total number of vacancies in the Jobfeed database have almost doubled from 2012 to 2017. This could be the result of more frequent posting of vacancies on the internet or due to technical improvements allowing for a more precise crawling of vacancies from the Internet.

Figure 2: Digital and technical skills per profession (managers, ICT positions, others)

	Overall	Manager	ICT	Other
Basic computer skills	1	1	5	1
Programming skills	2	2	1	6
Resource management software	3	3	4	2
Database management, design and query	4	4	3	4
Web platform development software	5	8	2	9
Computer-aided design	6	6	11	3
Big data and analytics	7	5	7	7
Internet technology and networking	8	9	6	8
Specialized software	9	10	9	5
Digital transformation	10	7	8	11
Project management software	11	12	10	12
Digital marketing	12	11	13	10
IT governance and management	13	13	12	13

Technological and economic developments also have changed the demand for general cognitive and non-cognitive skills ('other skills'), especially the demand for non-cognitive skills. A well-known framework for this type of skills are the so-called *21st century skills*, which in addition to media literacy, basic ICT and information skills also include skills such as creative and critical thinking, communication and collaboration or problem-solving ability. If we look at the trend in general skills between 2012 and 2017 (Figures 4 and 5), we observe an increase in the demand for cooperation (related to *communication* and *collaboration* skills) and in skills for *planning and organization*, *self-starter*, *computational thinking*, *problem solving*, and *active learning*. *Flexibility* and *leadership skills* are also in increasing in demand, while the remaining skills remain more or less stable. Overall, the demand for *active learning skills* is rising most in this period, indicating an increasing need for employees that are intrinsically interested in achieving higher skill levels and in lifelong learning. This also highlights the repercussions from the ongoing digitization and automation, which lead to faster technological change and, therefore, impose higher demand for a highly skilled, self-managing, and continuously learning labor force.

Figure 4: Dynamics in general skills demand between 2012 and 2017, per category, 6 largest categories

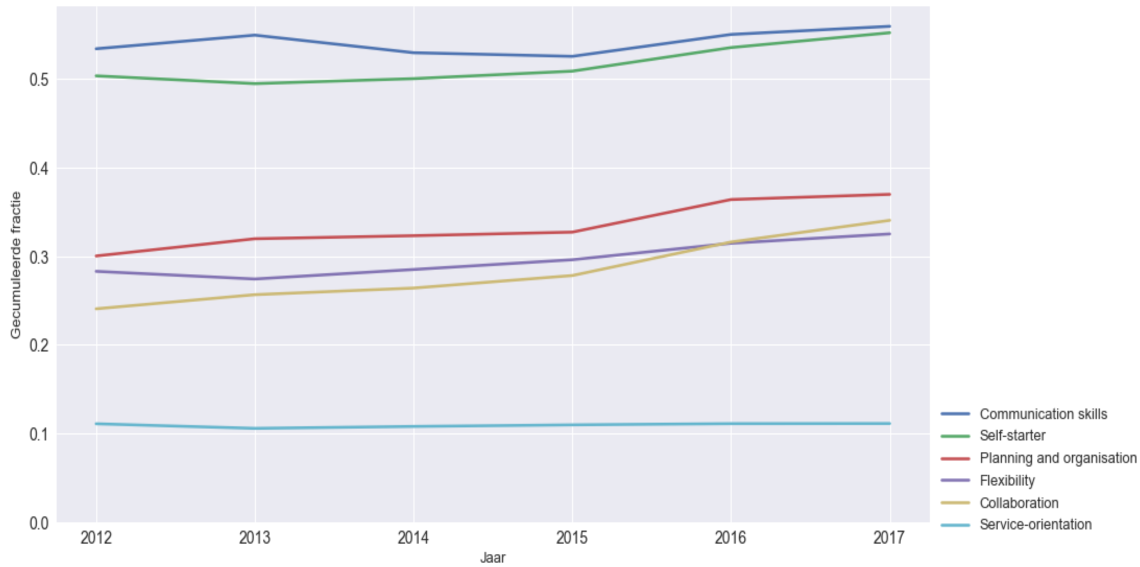
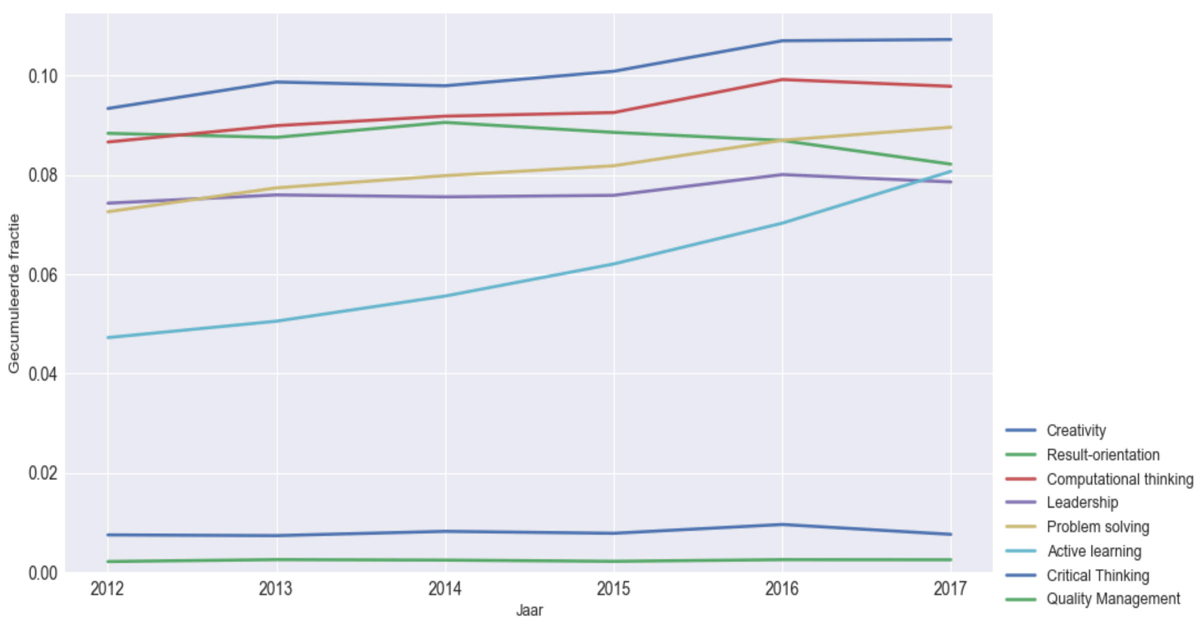


Figure 5: Dynamics in general skills demand between 2012 and 2017, per category, rest of categories



Results regarding viable and preferable job transitions

The second part of our study relates to the development in occupations over time and to the question how certain types of professions are affected by technological change. From there, we investigate potentially viable and preferable job transitions employees in shrinking sectors or occupational classes have in order to prevent unemployment. Luckily, we find suitable transition possibilities for a large majority of shrinking occupations. In the scenario where we fix the salary range and allow only for transitions to occupations in the same salary range, we find for 79% of all the employees a viable transition possibility. If we loosen this assumption and allow for transitions also to occupations with lower salaries, even 91% of all employees can make a preferable transition. As an illustration, Figure 5 shows part of a Sankey diagram with 'good-fit' job transition options.[11]

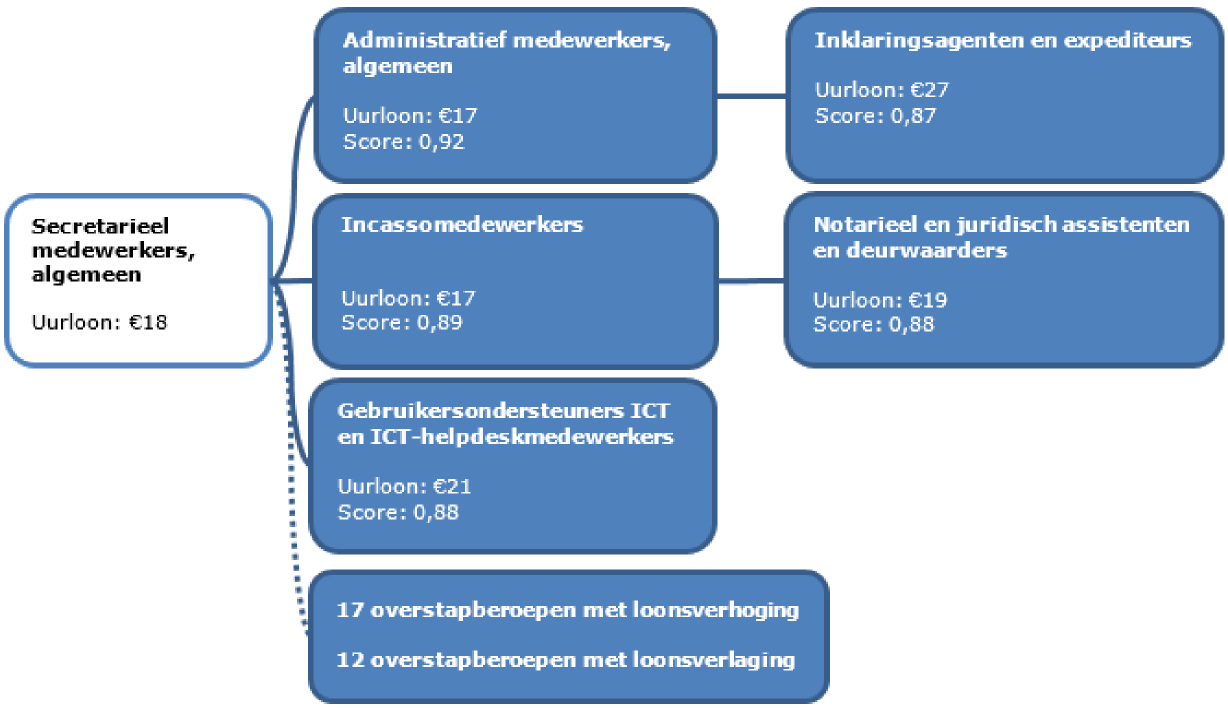
Figure 6: (Part of) Sankey diagram for 'good fit' job transitions

From an individual perspective of a single employee affected negatively by digitalization and automation, we observe on average 24 options to switch from one profession to another one, whereby 11 of these possibilities come with and 13 transition with equal or less salary. These positive findings of a large variety of good job transitions apply to women and men equally, and thus deviate from the results of the WEF (2018) on the US labor market, where men have more

[11] Due to time constraints, this Sankey diagram could not be translated into English yet. In the Dutch policy report accompanying this first draft of an academic paper, the complete diagram can be found.

options for viable and preferable job transitions than women. Figure 7 shows an example of a pathways for secretaries to demonstrate which type of results can be derived from our analysis.

Figure 7: Example of pathways for a secretary (due to time constraints not yet translated into English)



Conclusions

By using the big data vacancy database Jobfeed, we have been able to analyze a very rich source of information enabling us to calculate the similarity between 371 different professions based on the set of skills, the level of education, and the level of expertise that is required. This similarity is measured in a similarity score and the higher this score, the more similar the professions. Based on this score, we can thus compare the degree of overlap in skills, education level, and expertise required and therefore, to establish potential transitions pathways for employees from one profession to another one with a similar set of requirements. Moreover, we can also draw conclusions about viable and transferable transitions taking also into account whether there is a need for more employees in a given sector or job family, whether we can expect a surplus of employees given medium-term labor market forecasts, and whether salary conditions are satisfiable.

Apart from insights on dynamics in occupations, our fine-grained Labor Market Intelligence (LMI) approach allows to derive insights in the dynamics of skills demand. We find that digital and technical skills related to very recent trends in digitization and automation, particularly skills concerning 'Big data and analytics' and 'Digital transformation skills', are in high demand.

Moreover, there is an increasing demand in general cognitive and non-cognitive skills such as self-starter, flexibility, active learning, and leadership skills highlighting the ongoing repercussions related to faster technological change which also impose higher demand for a highly skilled, self-managing, and continuously learning labor force. Besides, there is an increased demand for the so-called 21st century skills, in particular the 4 Cs: creativity, collaboration, communication and computational thinking.

In summary, this research leads to positive outcomes with regard to the impact of digitization and automation that demonstrate the necessary opportunities for maintaining sustainable employability and economic growth options. By combining unstructured ‘big data’ from online vacancies with structure information from labor market forecasts, able to distinguish not only the effects of general trends on the labor market, but also show the impact on career prospects of employees in many different sectors and professions and provided clear-cut suggestions which job transitions an employees in a shrinking sector and/or professions can make. Thereby, our study enables both employees and employers to actively search for job market opportunities that are sustainable over the longer term and provides guidelines for additional training necessary to make a ‘future-proof’ switch. Thus, we can also demonstrate indirectly what the impact of digitization and automation is on unemployment and lifelong learning and how different sectors and stakeholders on the labor market, such as labor unions, guilds, or lobby groups from a specific sector could react. If the different stakeholders collaborate to create suitable "Lifelong Learning" policies that help all employees gather the necessary ‘skills of the future’ and facilities job transitions even across sectors, the potentially negative effects of technological progress can be mitigated.

Some of the insights found in this study and especially the LMI approach could also be used for follow-up research on suitable sources for new personnel (steeply) growing sectors. Think, for example, of vacancies for engineers or other technical occupations which will be hard to get filled in the near future. By reversing our approach one could also investigate where to get the most suitable candidates from when in need for if we need engineers. Further research could also help to outline an even sharper picture of required (clusters of) skills, apply our insights to future (or current) crap professions or substantiate that online vacancies can actually be a reliable source for labor market analyzes. In any case, we can conclude that, in combination with administrative data, for example from labor market surveys, they lead to reliable conclusions.

References

Arntz, M, Gregory, T. en U. Zierahn (2016), The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis, OECD Social, *Employment and Migration Working Papers*, No. 189, OECD Publishing, Paris.

Autor, D. (2015), Why Are There Still So Many Jobs? The History and Future of Workplace Automation, *Journal of Economic Perspectives*, 29(3), 3-30.

Bessen, J. (2015), Learning by Doing - The Real Connection between Innovation, Wages, and Wealth, New Haven: Yale University Press. EconTalk Episode, http://www.econtalk.org/archives/2016/05/james_bessen_on.html.

Boselli, R., Cesarini, M., Mercorio, F. and M. Mezzanzanica (2017), Using Machine Learning for Labour Market Intelligence, in Y. Altun et al. (Eds.): *Machine Learning and Knowledge Discovery in Databases*, New York: Springer, 330–342.

Cogburn, D. and Hine, M. (2017), Introduction to Text Mining in Big Data Analytics, *Proceedings of the 50th Hawaii International Conference on System Sciences*, HICSS 2017.

Elliott, S. (2017), *Computers and the Future of Skill Demand*, OECD Publishing, Paris.

Kurekova, L., Beblavy, M. en A. Thum-Thysen (2015), Using online vacancies and web surveys to analyse the labour market: a methodological inquiry, *IZA Journal of Labor Economics* 4, 2-20.

Mayer-Schönberger, V. and Cukier, K. (2013), The Rise of Big Data, *Foreign Affairs*, May/June Issue.

Mc Afee, A., Brynjolfsson, E. (2017), *Machine - Platform – Crowd: Harnessing our Digital Future*, New York: Norton.

Obschonka, M., Hakkarainen, K., Lonka, K. and K. Salmela-Aro (2017), Entrepreneurship as a twenty-first century skill: entrepreneurial alertness and intention in the transition to adulthood, *Small Business Economics*, 48: 487-501.

Prüfer, P., Kumar, P. and M. den Uijl (2019), Arbeidsmarktonderzoek Digitalisering in Topsectoren, mimeo, CentERdata, Tilburg.

ROA (2017), De Arbeidsmarkt naar Opleiding en Beroep tot 2022, ROA-R-2017/10, Maastricht.

Spitz-Oener, A. (2006), Technical Change, Job Tasks, and Rising Educational Demands: Looking Outside the Wage Structure, *Journal of Labor Economics*, 24, 235-270.

World Economic Forum (2018), *Towards a Reskilling Revolution - A Future of Jobs for All*, Davos, Zwitserland.

Appendix

Table A1: Categories of general cognitive and non-cognitive skills with examples

Category	Skill Examples
Critical thinking	Reasoning, Research, judgment and decision making, critical thinking, systems analysis, systems evaluation, Business analysis, Business modelling, Detail oriented, Business process improvement
Creativity	Creative, Innovation, Originality
Collaboration	active listening, team-oriented, Participation in discussions
Communication	Speaking, Writing, Reporting, Reading comprehension
Computational thinking	Mathematics, Analytical, Science, Econometrics,
Flexibility	Adapting, Flexibility, Resistant to stress
Leadership	Coordination, Negotiation, Leadership, Delegating, Coaching, persuasiveness,
Self-starter	Self motivated, Initiative, Inquisitive, Enthusiast, Independence
Result-orientation	Goal setting, go-getter, Perseverance
Problem solving	Root Cause Analysis, Problem management, Problem Sensitivity, Problem solving
Active learning	Active learning, Learning strategies, Learning assessment and evaluation, Development management
Planning and organization	Time management, Risk management, Organization design and implementation, Project management, Facility Management, Strategic thinking, Systemic thinking, Change management, Program management, Sustainability strategy, Requirements definition and management, requirement gathering
Service-orientation	Customer focus, Service orientation, Service level management, Relationship management, Customer service, Service acceptance
Quality Management	Quality control, Quality assurance, Quality management

Table A2: Categories of digital and technical skills with examples

Category	Skill Examples
Specialized software	File versioning software, industrial control software, medical software, map creation software and compliance software like: Blackboard, Git, SVN, arcgis, SCADA
Computer-aided design	Computer-aided design and manufacturing (CAD/CAM) software, computer based training software and pattern design software, like: Catia, CAD, E-plan, Cadence, Autocad, civil 3d
Resource management software	Inventory management software, customer relationship management (CRM) software, materials requirements planning logistics and supply chain software like: SAP,ERP,Primavera
Basic computer skills	Word processing, presentation and spreadsheet software, internet browser software, electronic mail software, operating system software and backup or archival software like: Microsoft Office, Windows operating system, Solaris, Unix, TextPad, Ubuntu
Database management, design and query	Query and processing language, database user interface and query software, object oriented data base management software, metadata management software and database reporting software like: SQL, MYSQL, datawarehouse, netezza, database management, RDBMS, NoSQL
Big data and analytics	Data analytics, natural language processing, parallel computing, machine learning, artificial intelligence, business intelligence and data analysis software like: Hadoop, Spark, Hive, Pig, Tableau, Rapidminer, Logistic regression, Support vector machine, K means, Text analytics
Programming skills	Object or component oriented development software, development environment, program testing software, compiling software like: C++,C#, Perl, Java, Lisp, prolog++, Julia, Python
Internet technology and networking	Network monitoring software, network security and virtual private network, application server software VPN, internet protocol IP multimedia subsystem software and equipment software like: LAN, WAN, DNS,webserver,traceroute,weblogic
Web platform development software	Bv: HTML, javascript, django, angularjs, php, css, drupal, joomla, Typescript, requirejs, dhtml, Ruby on Rails, ngrx
Project management software	Bv.: Devops, Content workflow software, Microsoft Project

Category	Skill Examples
IT governance and management	Bv.: Information management, IT governance, IT infrastructure, data governance, IT frameworks, Information systems coordination
Digital marketing	Bv.: Adsense tracker, Digital marketing
Digital transformation skills	3D printing, Artificial intelligence, Blockchain, Cloud computing, Cybersecurity, Docker, internet of things, Robotics

Similarity score and transition matrix

The final result is a 371x371 similarity score matrix. The average similarity score between professions is 0.72 with a standard deviation of 0.08. See Figure A1 for the distribution of similarity scores. Because a similar score has been determined for 371 professions, this is a total of 68,635 pairs of professions. Figure A2 shows the overall job transition matrix for the Netherlands.

Figure A1: Distribution similarity score

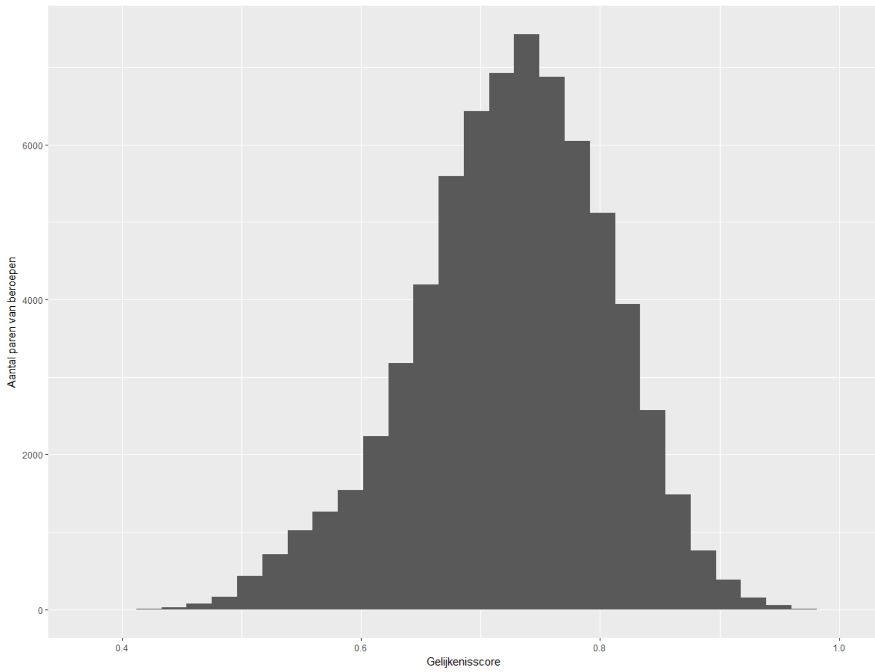


Figure A2: Job transition matrix between 371 jobs in the Netherlands