Informatics

Data-driven network modeling of occupation mobility networks

A computational framework for modeling upward mobility and investigating social inequality in the Austrian labour market

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Data Science

by

Julia Putz, BSc.
Registration Number 51841360

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.-Prof. Dr. Allan Hanbury

Second advisor: Univ.-Prof. Fariba Karimi, M.Sc. Ph.D.

Vienna, September 3, 2025

Julia Putz

Allan Hanbury

Erklärung zur Verfassung der Arbeit

Julia Putz, BSc.

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.
Ich erkläre weiters, dass ich mich generativer KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Im Anhang „Übersicht verwendeter Hilfsmittel“ habe ich alle generativen KI-Tools gelistet, die verwendet wurden, und angegeben, wo und wie sie verwendet wurden. Für Textpassagen, die ohne substantielle Änderungen übernommen wurden, haben ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT- Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 3. September 2025

Julia Putz

Acknowledgements

I would like to thank my supervisors Fariba Karimi and Allan Hanbury for their guidance and constructive feedback. I am especially grateful for the opportunity to get to know the computational social science team lead by Fariba Karimi at TU Graz as well as colleagues from the complexity science hub Vienna. It was a pleasure to meet likeminded people and have inspiring chats which helped me view my thesis topic from different angles. I also want to thank Maria del Rio-Chanona who introduced me to occupational mobility networks and laid the foundation for my work by sharing her expertise and insights. Lastly, I am deeply grateful to my family, who have supported me not only during the writing of this thesis but throughout my entire studies. Their patience, encouragement, and ability to lift me up during difficult times have been invaluable.

Kurzfassung

In der heutigen Gesellschaft existieren nach wie vor eine Vielzahl struktureller Hindernisse für Personen aus sozial benachteiligten Schichten. In ökonomisch schwierigen Zeiten wie den heutigen, in denen es von wesentlicher Bedeutung ist, eine gut bezahlte und sichere Anstellung zu finden, ist es von besonderer Relevanz, die Ursachen für die bestehenden Ungleichheiten auf dem Arbeitsmarkt zu erfassen und zu verstehen. Die vorliegende Arbeit untersucht den Einfluss soziodemografischer Merkmale wie Alter, Geschlecht und Migrationshintergrund auf die berufliche Aufwärtsmobilität. Die vorliegende Arbeit stützt sich auf die österreichischen Mikrozensusdaten der Jahre 2011 bis 2022 und bedient sich dabei netzwerkbasierter Methoden, zu denen Simulationen mit kürzesten Wegen, Zufallsbewegungen und Markov-Bewegungen zählen, welche aus der Graphentheorie und Datenwissenschaft bekannt sind. Die Methoden werden adaptiert und evaluiert, um die berufliche Aufwärtsmobilität für acht soziale Gruppen zu erfassen, die Personen mit unterschiedlichen soziodemografischen Merkmalen repräsentieren. Die Validität der adaptierten Metriken wird durch den Vergleich mit etablierten Metriken, darunter der Gini-Koeffizient und eine adaptierte Version des durchschnittlichen Clustering-Koeffizienten, nachgewiesen. Die Ergebnisse dieser Arbeit zeigen erhebliche Unterschiede in Bezug auf Alter und Geschlecht. Diese Ergebnisse bestätigen die etablierten gesellschaftlichen Ungleichheiten auf dem Arbeitsmarkt und bieten gleichzeitig eine intuitivere Perspektive als die üblicherweise verwendeten abstrakten statistischen Werte, da sie die Analyse direkt mit realistischen Szenarien verknüpfen. Der in dieser Arbeit entwickelte Ansatz hat das Potenzial, zukünftige Analysen zu unterstützen und zur Entwicklung eines gerechteren Arbeitsmarktes beizutragen.

Abstract

Modern society still presents many structural barriers for individuals from disadvantaged backgrounds. When securing a well-paid, stable job is crucial, as it is in economically challenging times like today, it becomes particularly important to understand where inequalities in the labour market take root. The present thesis investigates the influence of sociodemographic features, such as age, gender, and migration background, on upward occupation mobility. It utilises Austrian microcensus data from 2011 to 2022, employing network-based methodologies, encompassing shortest paths, random walk and Markov walk simulations, as they are recognised in graph theory and data science. These methods are adapted and tested to capture upward occupational mobility for eight social groups, which represent individuals with different sociodemographic features. The validity of the adapted measures is demonstrated through their comparison with established metrics, including the Gini coefficient and an adapted version of the average clustering coefficient. The results of the study indicate significant disparities in age and gender. These findings serve to confirm established societal inequalities in the labour market, while also providing a more intuitive perspective, in contrast to commonly used abstract statistical values, by linking the analysis directly to realistic scenarios. The framework developed in this thesis has the potential to support future analyses and contribute to the development of more equitable labour market strategies.

Contents

CHAPTER 1

Introduction

This thesis employs a data-driven network approach to the analysis of labour market inequalities, proposing methods of measurement that are both reproducible and interpretable. Inequalities in terms of monthly wages are observed among individuals with diverse social demographics, with these disparities being attributed to features such as gender, age, migration background, and educational attainment. Different studies have focused on the impact of the mentioned sociodemographic factors on the Austrian labour market independently, but existing literature lacks a side by side statistical comparison of those factors. An examination of occupational mobility and income segregation from a more comprehensive standpoint rather than a narrow and detailed focus, is missing for recent years ([JLLW22], [GG15], [HTWA17]).

A major focus of this study is explainability. To this end, techniques from graph theory and network science, namely shortest path algorithms, Random walks, and Markov chain simulations, are adapted to the context of occupational mobility. These methods provide a structural lens for understanding the barriers and pathways to high-income occupations across different social groups. By utilizing real-world occupational data, the study integrates statistical inference with graph-based algorithms to construct an interpretable framework for analyzing group-based mobility variations. The emphasis on algorithmic transparency and reproducibility aligns with best practices in data science and contributes to a growing body of research at the intersection of social science and machine learning.

Research has shown that modern society still shows many structural barriers for people with under-privileged background ([RB11], [Die24], [HW23]). Especially the sociodemographic factors migration background, age and gender influence an individual's career path and job prospects. At least this is what societal prejudices tell us. But are there truly distinct differences observable when it comes to upward occupation mobility for individuals with different sociodemographic background or is this just something society assumes? Based on these societal assumptions it also seems evident that it is easier for e.g.

men to attain a high-income occupation than women. This research aims to investigate these assumptions and determine the impact of the identified factors on individuals' upward career trajectories by answering the following question.

What factors, including migration background, age, and gender determine upward occupation mobility?

This leads us to the question of how to actually measure and understand differences in upward occupation mobility. The utilisation of network structures for a multitude of purposes regarding labour market analysis has been demonstrated to yield valuable insights ([MdF18], [dMB$^+$21], [JL22]). However, the concept of comparing networks which coalesce career paths of individuals with certain sociodemographic characteristics has not yet been employed in the analysis of the Austrian labour market. This data-driven approach facilitates not only the application of network analysis concept but also enhances the clarity and comprehensibility of the subject matter to non-experts. Common state of the art methods summarize the complexity of the diverse labour market in a single number and fail to make its entire scope visible. This is why in this research we have explored multiple adaptions of network-centered approaches commonly known in the field of Data Science but not yet employed to describe upward occupational mobility the way it is presented in this study.

To ensure the validity of these approaches, we compare the results with a baseline, grounded in well-established measures. Overall, the applied approaches fall into two categories. In the first we observe approaches which generate a distinct score for each observed sociodemographic group. The underlying idea is that a bigger gap between the scores of two groups indicate more inequality between these groups. In the second, we want to focus more on the clear and comprehensible illustration of labour market inequalities. More precisely, we are interested in an individuals chances of attaining a well paid job. According to assumptions and prejudices mentioned above, we would assume that individuals with different sociodemographic characteristics have different chances in attaining a well paid job, especially when starting at the bottom in a low paid position. Therefore, we are simulating upward career paths and identify the first encounter with a high income occupation. In the field of stochastic modeling this is referred to as the first-passage time ([Red23]). Once more, the prevailing network analysis measures that are commonly employed for the description and identification of inequalities in social networks are utilised to assure the validity of the approach's results. With that in mind we ask ourselves:

How do group differences in distance measures and access to high income occupations capture labour market inequality?

Another major focus of this research lies on how to adapt the network centered approaches we already know to identify labour market inequalities and how the input data needs to be prepared. Generally, when analyzing social networks it is important to not only assume that the results are solid because they align with what society expects. In reality various aspects including the sample size on which such a network is built upon plays a crucial role. In other words, if a network represents more individuals, it also represents more career paths. This influences some of the methods applied in this research, since this does not necessarily infer that an individual has better chances in attaining a high-income occupation as these methods would then suggest. Another aspect refers to drawing those samples in the first place. This is a complex but crucial task. Moreover, when simulating one individual's upward career path, it is not representative of the total population. But how many simulations are necessary to obtain a representative result?

The results of this thesis should serve as a reference work on how to measure inequality in the labour market by using historic data. This work summarizes the most important factors and demonstrates the extent to which they are correlated with upward occupation mobility. Moreover, the implemented approaches are tested regarding their statistical significance and correctness based on existing measures of inequality, which reveals their potential in explaining inequalities.

CHAPTER 2

Related Research

This chapter provides an overview of the methods commonly employed in labour market research, particularly in the context of occupational mobility, and the insights gained. The sections on Random walks and Markov chains, as well as the section on shortest paths, provide an outline of where the central methodologies in this work are commonly applied, and how they have already been applied in other labour market-related research. The final section focuses on the Austrian labour market itself, covering important facts and its evolution over time.

2.1 Occupational Mobility

Occupational mobility is a field widely researched. It has been employed to model and explain various different labour market related aspects. The employment of a network approach, alongside network analysis tools and measures, is becoming increasingly prevalent.

Recent studies have examined the impact of automation on the labour market. Research utilising an occupation mobility network has demonstrated that network structure is a pivotal factor in determining the influence of automation on unemployment [dMB$^+$21]. Furthermore, working in an occupation which is highly connected to occupations less susceptible to automation has been demonstrated to yield better chances in finding a new job in case the current one is lost [Chr22]. Other studies demonstrated that occupation mobility networks can also serve as a valuable instrument in the modelling of human capital flows and the evaluation of career prospects, in addition to the assessment of career growth potential [LWY$^+$24]. Mobility boundaries are also shown to be able to be detected using occupation mobility networks. The application of a flow-based community detection algorithm has revealed that these boundaries facilitate the classification of individuals into occupation clusters based on their shared skills [CP20]. In recent years, transitions between occupational communities decreased, leading to a more fragmented labour market in the US. Based on this finding, research suggests that an individual's entrypoint into the labour market is more important nowadays since it can determine someone's career path [LH22]. The centrality of occupations within the French occupation mobility network has shown to be related to wages, 'the unemployment duration and the likelihood of finding a job when being unemployed' as demonstrated by Joyez and Laffineur [JL22]. It is evident that workers employed in more central occupations are positively influenced, leading to higher wages and reduced risk of long-term unemployment [JL22].

2.2 Wage inequality

From a sociological perspective, there are different principles on how to assess wage inequality. Baumgarten et al. [BH23] have shown in their study that Austrian citizens most commonly use the principle of performance to evaluate social inequality. This principle asserts that income should be based on performance, with higher performance resulting in a higher income. Participants in the mentioned study argued that income should be based on performance, as this principle states, and not on other factors, such as gender. The same study also reveals that low income is often associated with female dominated occupations. This association is indicative of the gender wage gap, a topic which is the subject of much discussion in Austria (see [FVL19], [BHMZ12], [BBF22]). It refers to the differences in income between male and female individuals. Whilst some of these disparities may be explainable by factors such as education, work experience, firm or industry characteristics, Böheim et al. [BHMZ13] has demonstrated that approximately 60% of the gender wage gap in Austria cannot be accounted for in this manner. Even though, as displayed in Figure 2.1, the gender wage gap is still large compared to other European countries, it shows a clear trend of shrinking [Bun].

From a statistical perspective, several different measures have been applied in previous research to determine wage inequality (see [BHMZ12], [KM04], [UZP+17]). A prominent measure is the gini coefficient, which is also applied in this research and explained in more detail in Section 3.3.1. Also the percentile ratios and the income share of the top 5 or 10 percent of all earners are often used to describe wage inequality. Another possibility is the interquartile range in relation to the median which describes the difference between the .75 and .25 percentile, where a higher value indicates more inequality in the underlying income distribution. This last measure has been applied by Unger et al. [UZP+17] to study inequality within industrial sectors. The findings of their study indicate that a greater proportion of women in certain industrial sectors is associated with higher levels of income inequality. Moreover, this study's results suggest that technological progress does not have a significant impact on income inequality, contradicting popular assumptions. Another popular approach, especially to explain where wage inequalities originate, are the application of logistic regression models. Where logarithmic wage is treated as the dependent variable, and other characteristics like gender or educational attainment account for independent variables (see [BHMZ12], [BHMZ13]). The most similar approach to the one applied in this research was applied by Hofer and Weber

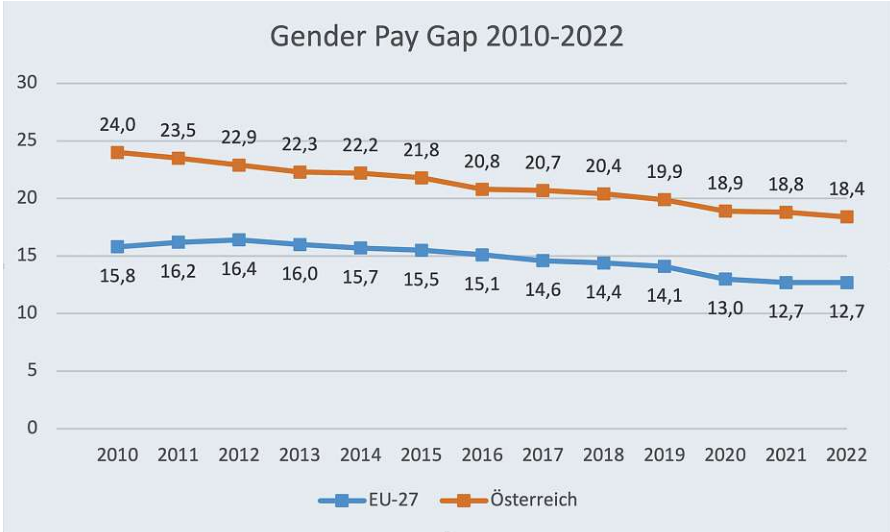


Figure 2.1: Evolution of the gender pay gap in percentage for 2010-2022 in Europe (EU-27) and Austria (Österreich) (source: [Bun])

[HW02], and utilizes transition matrices including transition probabilities to study wage mobility. This approach is based on the premise that an individual transitions from one quintile of the income distribution to another with a certain probability. The results obtained demonstrate higher mobility for younger individuals. However, it is evident that, in general, Austria exhibited low wage mobility in the early 90's, with the probability of transitioning more than one quintile within a six-year period being small [HW02]. In addition, an association was identified between occupational mobility and wage inequality. This finding was supported by a general equilibrium model, which was developed using data on the United States labour market. The model incorporated occupation-related human capital and work experience [KM04].

Besides focusing on overall and gender related wage inequality, such inequalities can also be observed regarding migration background. It is important to note that disparities may emerge when comparing immigrants who have accumulated human capital, such as education or professional experience, within their host or origin country. Holbrows' [Hol18] research indicates that the quality and wealth of an individual's country of origin are significant factors in this regard. Individuals immigrating from developing countries often have less quality education and encounter greater prejudicial attitudes, consequently resulting in lower income when compared to immigrants from higher quality origin countries. Additionally, unmeasurable factors, such as language barriers and a lack of familiarity with the host country's business practices, may contribute to this observed inequality [Hol18].

2.3 Random Walk and Markov Chains

The random walk theory has been employed to model the career progression of an individual by van Uem [vU23]. Consequently, this model is utilized for the purposes of forecasting and planning within organisational settings. Moreover, López et al. [LGA20] established that this method is also applicable in inter-organisational contexts, modelling the transition of workers between firms. Through the utilisation of existing data to establish a network that represents these movements, the constraints encountered by individuals during transitions can be effectively summarised. Random Walks are then applied on the resulting network to study occupation mobility which can have practical impact [LGA20]. Another significant application of random walks on social networks is the measurement and detection of segregation. Ballester and Vorsatz [BV14] utilized random walks for defining a novel segregation index, the probability of an individual meeting another individual from the same social group. This walk is performed on a graph representing geographic areas, where connected nodes represent neighbourhoods. The resulting probability is therefore a measure of residential segregation.

Other research has employed random walks to simulate career paths or social representations as mentioned below. In the labour market related research by Liu et al. [LWP+23] random walks are implemented as part of a deep occupational representation learning model with the aim of understanding the complexity of occupational connections. More precisely they applied biased random walks for simulating career paths. Meaning, the walk is not entirely random but is guided to some extent to represent the different attitudes of individuals. Individuals who are exploratory and transition to entirely different occupations are reflected by a depth-first search, which samples sequentially connected nodes that are distantly connected to the starting point. In contrast, a breadth-first search samples nodes that are directly adjacent to the starting point and therefore reflects more cautious individuals who seek to transition to similar occupations [LWP+23]. Similarly, Perozzi et al. [PAS14] model numerous short random walks to learn social representations, which are regarded as latent features of network nodes that encapsulate neighbourhood similarity and community membership. The simulation of these short walks enables the accommodation of minor alterations in network structure, circumventing the necessity for global recomputation. The underlying principle of Perozzi et al.'s algorithm known as DeepWalk is derived from language modelling, wherein walks are conceptualised as short sentences. DeepWalk has also been adapted for use in a variety of different contexts, including human mobility in geographic terms. In general, random walks are widely applied in this field not only because previous studies have shown that human mobility patterns are similar to scale-free random walks [BHG06]. Research on urban transportation networks has incorporated generalized travel costs into the original DeepWalk algorithm to ensure that the generated random walks are more realistic. This approach leads to improved results in this area, especially when the walk length is low [GLZ24].

Another study in the geographical context attempted to predict individuals future locations including residence time based on statistical Markov state space models [KRŠ+17].

Another algorithmic framework proposed by Grover and Leskovec [GL16] is called node2vec. It aims at learning feature representations of nodes within any social network. Again biased random walks are employed to explore neighborhoods. Its novelty lies in the introduction of two parameters that guide the walks. These parameters set the likelihood of immediately revisiting a node in a walk and can control the extend of depth-first or breadth-first search behavior. This makes node2vec flexible and controllable whilst results show promising performance [GL16].

2.4 Shortest Paths

The shortest path problem has not been employed in explaining inequalities in occupational mobility at the same level of granularity as presented in this study. However, it is used in labour market related studies and in the more general context of social network analysis (see [LZM13], [GVRZ12], [dSG21]). Lungu et al. [LZM13] observe occupational mobility of female and male higher education graduates by employing shortest paths. They created a separate occupation mobility network, similar to our approach, for male and female individuals. In addition to centrality and other network metrics, the average path length and the longest distance between any two occupations are used to highlight inequalities. When both metrics have a low value, it indicates a higher connection between the nodes [LZM13]. An alternative approach, however, applied to a different field, namely logistics, also employs the shortest path problem. The similarity to our study can be found in the fact that the overall network is divided into several sub-networks. In this approach, shortest paths are detected from a node starting in one sub-network and ending at a node in another. However, this is only part of a bigger algorithm presented [GVRZ12].

Determining shortest paths can also serve as a useful tool to facilitate the matching of individuals with occupations, a process that is of particular relevance to both job seekers and recruiters. When the shortest paths, based on a network representing skills and occupations, are computed, the most efficient transitions between all pairs of occupations can be identified. Edge weights, representing the similarity of the occupations in question based on their skill sets, are incorporated into the process. The implementation of this methodology can assist an individual seeking employment in their search process [dSG21]. Granovetters' [Gra73] introduction of the concept of strong ties, representing frequent contact, and weak ties, representing rare contact, between individuals, has allowed for the drawing of interesting conclusions connected to the employment rate and inequality. Özer and Perc [ÖP21] state that there are connections between shortest path length and such weak ties which often serve as bridges within networks. Where the removal of a weak tie from a network results in the removal of a path, which leads to an increase in shortest path lengths enhancing inequality. Adapted to the labour market, this means that increasing the average path length also increases the total unemployment in networks which already capture a high unemployment rate. Strong ties are also important, they play a pivotal in reducing network distance, as argued by Özer and Perc. Thereby their research highlights the necessity to identify the most significant ties within a network, given that the ratio of these ties can influence equality.

A social network typically consists of nodes representing users and edges representing communication between said users. Calculating the shortest path between users has been demonstrated to facilitate the identification of indirect connections, a result which can be advantageous in a variety of contexts, including the recommendation of products or studying the behavior of information spreading [GLW+16]. A comparative analysis of the most commonly applied shortest path methods, including the Dijkstra and Bellman-Ford algorithms, has revealed that the former is particularly well-suited to social networks due to its faster execution time [ŠP23]. Ahmad et al. [AAV17] showed that in terms of social search algorithms in online social networks an influence based strongest path algorithm outperforms the traditional Dijkstra algorithm. They defined the influence of one node on another as the proportion of one node's investment on another node. The focus of Nawaz et al. [NUL14] has been on analysing the paths themselves to anticipate the behaviour of traversal algorithms on real-life networks, rather than on performance. Their empirical analysis shows that nodes with a high degree appear in the majority of shortest paths found, whilst those with only an average degree are not considered by the Dijkstra algorithm at all. However, it is important to note the poor performance, as presented by Zhao et al. [ZSZZ11], of the well-known and frequently mentioned algorithms of amongst others Dijkstra and Floyd-Warshall when scaled to network size. Despite this, they argue that the analysis of large networks is becoming increasingly important due to the numerous applications it has, such as recommending an item to a similar user. Their approach to address this issue is to use graph coordinates to approximate distances between nodes by embedding the network in question into a hyperbolic space. Zhao et al. have shown that this approach, known as Rigel, results in more accurate and faster results than traditional algorithms. Another approach presented by Maoguo et al. [GLW+16] aims at improving time efficiency by maintaining the accuracy scale of shortest paths. With the community feature being a very important characteristic of social networks, this approach is based on community detection. A community is defined as being a group of nodes with similar properties or roles within the network. They created a so called community graph based on the detected communities, where each node represents a community. Based on the community graph the community paths are calculated and a sub-graph, from which the shortest path is calculated, is defined. This narrows the search space which explains the time efficiency [GLW+16].

According to Rezavnian and Meybodi [RM15], sampling is a pivotal method in the analysis of social networks, particularly in the domains of preprocessing, characterisation and the study of real networks. It facilitates the observation of a network's components while preserving its essential characteristics. A recently developed sampling method has demonstrated enhanced performance in comparison to conventional approaches, such as random edge or node sampling, and random walk sampling. This method employs shortest paths to define a subgraph of the original network, encompassing a specified percentage of high-ranked edges. Edges are previously ranked based on their frequency of appearance in the shortest paths which is computed on the original network [RM15].

2.5 Austrian labour market

The Austrian labour market is subject to constant change influenced by digitalisation and technological change as well as demographic developments. Hence, much research is available about labour market related aspects like unemployment rates or income distribution (see previous Section about wage inequality) [EBFM18]. The following presents the most relevant facts and statistics about the current labour market situation.

The age at which an individual may first enter the labour force is dependent upon their educational trajectory. The earliest age at which a person may access the labour market is 15 years. Upon attaining a minimum insurance period, individuals of a certain age are entitled to a retirement pension, which is designed to ensure financial security. It should be noted that the legal retirement age differs for men and women and is subject to regular adjustments. While the retirement age for men is currently set at 65 years, women are required to continue working until the age of 61 [Bun24]. Consequently, Statistik Austria defines the prime working age as the interval between 15 and 64 years [KM24].

To ensure fairness to some extent in the labour market, collective agreements are negotiated on an annual basis between the chamber of commerce and the Austrian trade union federation for each economic sector. The specific collective agreement applicable is determined not by the individual's job but by the sector in which the company is located. Such agreements inter alia set the minimum wages and determine the number of standard working hours per week [Wir25].

2.5.1 Statistics

On a quarterly basis, Statistik Austria publishes statistics with regard to the labour market, which are based on the Austrian microcensus (further details can be found in Section 3.1). The results from each quarter are summarised in a yearly labour market statistic. The following facts and numbers, if not stated otherwise, are taken from the Arbeitsmarktstatistik 2023 [KM24], since this is the most recent yearly publication available.

In essence, the labour market has demonstrated notable robustness, despite a decline in economic activity in recent years. This is evidenced by comparable numbers of employed individuals. In 2024/3 (the third quarter of 2024), the employment rate was only 0.1 percentage points lower than in the previous year (2023/3) [Kla24]. In 2023, 74.1% of all individuals between 15 and 64 years were employed (incl. self-employment). The employment rate in 2023 was higher for men, 77.9%, compared to women, 70.3%. A comparison of the employment rate with that of 2022 reveals an increase, which can be partly explained by the rising number of employees in part-time jobs. However, distinct differences between men and women are observable, which can be explained by childcare and other care work typically carried out by women. Approximately 50% of employed women and only around 13.4% of employed men work part-time. Other factors, such as age and education, also influence part-time employment rates. For instance, 58% of women aged between 35 and 44 years work part-time, whereas men aged over 60 had a higher chance of working part-time than younger men. In addition, educational attainment appears to influence part-time employment rates. The data show that higher education is associated with a higher probability of working part-time in case of men, where the opposite holds for women.

Looking at occupational groups, such differences in gender become more distinct. Occupational groups are defined by the international standard classification of occupations (ISCO) on different granularity levels. The following refers to the level 1 groups, where a complete list can be found in the Appendix A.1. 90% of employees in *Craft and Related Trades* and 64.7% of individuals working as *Managers* were men. *Clerical support* (68.5%) and *Service and Sales* (66.1%) occupations were dominated by women. Out of all working individuals, the largest amount was employed, whilst only 11% were self-employed. Economic sectors are defined by the ÖNACE 2008 standard and again listed in the Appendix A.2. The most prominent sectors, which can be summarized as Service sectors (sectors G-U), saw the highest increase in the number of employees. 71.7% of employed individuals were working in this field. The largest decrease in number of workers was found in the sectors of industry and commerce (sectors B-F), in which still roughly a fourth of all employed individuals were working in.

Income tax data as well as social insurance data act as a data basis for the following income related statistics for 2022. The median net income in 2022 was € 2,330. Again differences between men and women are very distinct and hence highlighted in the annual labour market report. On average, men earn €2,651, while women's net income is much lower, averaging only €1,951. When dividing the net income into income categories, namely ten equally sized decile, this difference can be observed more granulary. Like also displayed in Figure 2.2, only about one third of all employed men, but nearly two thirds of all employed women are found in the lower half of income. Overall women are overrepresented in the lower and men in the upper half of the income decile.

Similar to what was mentioned above, educational attainment also influences an individual's median net income. The highest differences were found between individuals with at most compulsory education and those with a university degree. The latter received almost twice as much net income. Moreover, the occupational position is also correlated with income. Civil and contract workers had an average income one third above the average income among all employed individuals. However, also other aspects like the higher age and better education among employees in this position need to be considered to draw appropriate conclusions. Focusing on economic sectors, the highest income was found in industry and commerce, whilst agriculture has the lowest income.

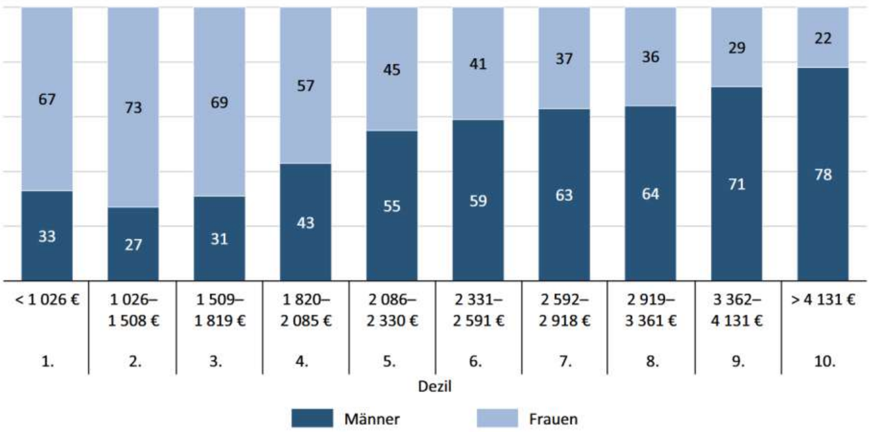


Figure 2.2: 2022 Annual average net income of employed individuals divided into 10 decile (Dezil) - Each bar shows the share of women (Frauen) and men (Männer) in the respective decile indicating an overrepresentation of men in high-income decile (source: Arbeitsmarktstatistik 2023 [KM24])

CHAPTER 3

Methods

The work can be split into 5 major steps which are summarized in the following framework. As displayed in Figure 3.1, it starts with pre-processing and getting familiar with the data. Part of it is dividing the Austrian population into 8 social groups based on sociodemographic features. A detailed explanation on the employed methods to define these groups is given in Section 3.1.3. Based on the clean data, occupation mobility networks are constructed. Such a network represents transitions between occupations, where more details will be explained in Section 3.2. This construction process is repeated for each social group separately. Lastly, a thorough analysis and comparison of these networks is conducted. Three different approaches, namely shortest paths, random walks,

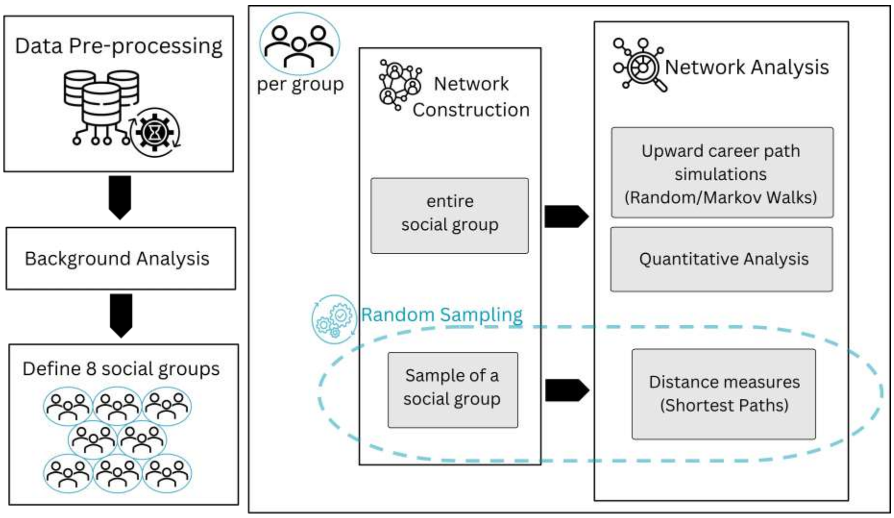


Figure 3.1: Framework - Main steps followed in this thesis to investigate inequality in upward occupation mobility

and Markov simulations, are introduced to describe discrepancies between the networks in an explainable manner. Each of the steps mentioned are explained in further detail in the following sections.

3.1 Data

All results presented in this thesis are based on the Austrian Microcensus Labour Force Survey / Housing Survey and income data provided by the Austrian Social Science Data Archive (AUSSDA). AUSSDA is a platform especially for the social science community offering services to support research. These include data archiving but also help with data re-utilization. Their main goal is to make data accessible and usable [The25]. The Austrian microcensus is a mandatory random sample survey that collects information on a range of subjects including work, living, and education. Being the largest regularly conducted survey in Austria, 22.500 households are interviewed each quarter based on a specified set of questions. The dataset used for this thesis is an anonymised and already imputed version of the microcensus data for scientific use [Aus22]. Imputed means that some variables were not explicitly asked for in the microcensus interview, but derived from an individual's answers to other questions asked in the interview. If selected from the central population register, one gets interviewed five quarters in a row. This means that each quarter one fifth of the respondents are new [MGK16]. Hence, it is possible to e.g. detect occupational transitions that happened within this time period. Additionally, income data provided by Statistik Austria are used to determine low and high-income occupations as well as wage discrepancies between groups.

3.1.1 Pre-processing

When starting this thesis, the microcensus data was available for the years 1970 to 2024 (1st quarter). The income data, which has to be requested at AUSSDA, however was available for the years 2011 to 2022. Moreover, the microcensus data before 2000 differ to more recent data in terms of inconsistent variable naming and different file formats. These are the reasons why the research focuses on the time frame 2011 to 2022.

Whilst the only cleaning necessary for the income data is to remove entries with missing information on the monthly net income, the microcensus data needs more preparation. Each year is provided in a separate file (see [Aus21a], [Aus21b], [Aus21c], [Aus21d], [Aus20a], [Aus20b], [Aus20c], [Aus20d], [Aus20e], [Aus20f], [Aus21e], [Aus22]). Some variables are only available for some years and others are named differently throughout the years. The first step therefore is identifying useful variables for this specific use case that are also available for the entire time frame. Due to the different naming schemes across the years, variable name mappings are implemented. Once each year is filtered and renamed accordingly, they are merged into one major dataframe.

After performing the necessary steps, income and microcensus date are merged. This is possible by a combination of the three variables, person/household id, year, and quarter, which also uniquely identifies a person. Lastly, the data is filtered to only contain people that are in the workforce, meaning everyone who is not unemployed, as well as entries that contain information on income, education, migration, age and gender.

3.1.2 Important Variables

The dataset offers ∼248 variables, representing an individual's answers as columns, and ∼43.000 observations, representing an individual's interview as rows, per year to analyze [Aus22]. Only a selection of variables that is personal, migration, education, and occupation related is relevant in this research. The most important variables are briefly discussed below.

In order to uniquely identify a person, a combination of three variables is necessary. Since the same person only gets interviewed once every quarter, the combination of **year** and **quarter** help identifying a person. The third variable is composed of the **household and person identifier**. Each household has a unique identifier. Within one household each person is assigned a consecutive number [Aus24]. If, e.g. two people live in household 1234, one person gets assigned the identifier 1 which leads to the household/person id of 123401, whilst the second person ends up with 123402. A look into the data shows that the household/person id is unique throughout the years and the year and quarter are solely used for to double check if an observation is truly unique.

Since only a small sample from the entire Austrian population is interviewed each quarter, the data obtained is not representative of the true population. This bias is reversed by introducing the variable **weight** as described by Meraner et al. [MGK16]. Each person in the dataset is assigned a weight based on how many people they represent. The basis of computing the weight is the so-called design weight, which is the inverse of the selection probabilities. In other words, the design weight states how likely a household was to be chosen. In addition to the design weights, sources such as the statistical population register are used to retrieve population totals, such as age, which are then used to calibrate the weights using the iterative proportional fitting procedure. This procedure computes calibration factors which adjust the initial weights to ensure that the population estimates align with the known population totals mentioned above. More precise details about how the weighting process works can be found in ([MGK16], [MGK15]).

Employment related variables are based on a standard from the International Labour Organisation (ILO). This offers the possibility to compare the data with that of other countries [MGK16]. The occupation is defined by the international standard classification of occupations 2008 (ISCO-08). ISCO-08 classifies all jobs into 436 categories, where these categories are aggregated into several levels of less detailed categories [Off12]. In this research the four-digit level is used to define the **occupations**, and the two-digit level to define **occupational groups** [Aus24].

The monthly gross and **net income** are the two available income variables. It was decided to use the net income for the analysis. This is because it represents the actual amount of money available for an individual and is therefore better suitable to analyze real financial discrepancies between people.

3.1.3 Social Groups

Since the thesis aims at explaining labour market discrepancies for different social groups, these groups need to be defined first. This grouping process is based on individuals' sociodemographic features. The observed features are as follows: migration background, level of education, age, and gender. The dataset already provides a categorization for those features as shown in Table 3.1. Using all possible combinations would lead to comparing 432 groups. Moreover, the group sizes would become too small to obtain statistically appropriate results. It is therefore necessary to combine certain categories into social groups.

features	migration background	education level	age	gender
available categories	no migration background	primary	20-24	male
	first generation	lower secondary	25-29	female
	second generation	upper secondary	30-34	
		post-secondary	35-39	
		short-cycle tertiary	40-44	
		bachelor's or equivalent	45-49	
		master's or equivalent	50-54	
		doctoral or equivalent	55-59	
			60-64	

Table 3.1: Observed sociodemographic features and their categories provided by the dataset

The challenge which arises here is to ensure internal coherence of the new groups in socioeconomic terms and to not falsify the results by e.g. combining two categories which have a different impact on a person's income. An analysis is carried out to check which categories are commonly grouped in previous studies and which categories have a similar income distribution and can therefore be grouped together. The use of significance tests for comparing underlying income distributions such as the Kolmogorov-Smirnov test or ANOVA is not an option due to the large and varied sample sizes. Applying statistical tests on large samples, even small and irrelevant differences can appear statistically significant. This might lead to misinterpretation of the two samples as being significantly different [SF12]. To avoid such Type II errors, wrongly concluding there is a difference, the following decisions rely on the effect size Cohen's d, the Wasserstein distance and visual aids. The resulting grouping decision is presented in Section 4.1.2.

Cohen's d

The effect size measures the 'magnitude of the difference between groups' as stated by Sullivan and Feinn [SF12]. Cohen's d, which is a common effect size measure, is used in this research. Sullivan and Feinn recommend that this measure is utilised during the planning stage in order to ascertain a suitable group size for the research being conducted. Essentially, it quantifies the discrepancy between the means of two groups, where a small value (∼0.2) signifies higher and a larger value (∼0.8) indicates less overlap of the underlying distributions [SF12]. As introduced by Cohen [Coh87], Cohen's d is defined as:

$$d = \frac{m_A - m_B}{\sigma} \tag{3.1}$$

where m is the sample mean of group A and B. The σ represents either standard deviation. Cohen assumed the standard deviation of both groups equal. According to Aoki [Aok20] the assumption of equal standard deviations between two groups is often inaccurate in practical applications and is therefore a fragile assumption. Hence, Aoki refers to using the pooled standard deviation. The adapted Cohen's d is defined as:

$$d = \frac{m_A - m_B}{\sqrt{\frac{(n_A-1)*\sigma_1^2+(n_B-1)*\sigma_2^2}{n_A+n_B-2}}} \tag{3.2}$$

where m is the sample mean of group A and B. The denominator incorporates the sample size n of group A and B as well as the standard deviation σ of each group. This version incorporates the variance of both observed groups and therefore leads to a more reliable measure [Lak13].

Wasserstein distance

The Wasserstein distance represents the required effort of transforming one distribution into another no matter the kinds of underlying distributions. Contrary to other popular statistics, it considers differences in spread, skewness and location of the underlying distribution. Therefore, it is utilized to compare income distributions which often tend to be skewed or for goodness-of-fit testing [PZ19]. Comparing the distribution u and v Ramdas et al. [RGC15] define the Wasserstein distance as:

$$l_1(u,v) = \underset{\pi\in\Gamma(u,v)}{inf} \int |x-y| d\pi(x,y) \tag{3.3}$$

where $\Gamma(u,v)$ denotes the set of all joint distributions π whose marginals are u and v. The x stands for the probability of u at position x and y denotes the probability of v and position y. A higher distance indicates larger dissimilarity of the compared distributions or more effort of transforming one distribution into the other [PZ19].

Violin Plots

As visual aids, violin plots are employed. Introduced by Hintze and Nelson [HN98], violin plots combine a boxplot and the density trace into one single plot making it intuitive to compare distributional factors of multiple variables.

3.1.4 Random Sampling

The social groups that have been defined above vary in size. Within the observed time frame, the smallest group that are women with migration background below 30 contains ∼332.000 unique people. Whilst the largest one consists of ∼3.220.000 unique people and represents men without migration background above 30. In order to avoid any possible group size bias, simple random sampling (SRS) as described by Noor et al. [NTG22] is employed. It ensures that every individual is selected to be part of the sample with equal probability, which is not necessarily true in random sampling. SRS is especially useful for scenarios where the research participants are randomly selected from a homogeneous population as stated by Noor et al. This is especially crucial for comparing the shortest path measures between groups calculated later in this work. Larger samples capture more connections resulting in more dense networks. These additional connections and nodes are not caused by structural differences but by scale. The sample size of the random samples is the sum of weights of the smallest social group. The entire sampling process described below and shown in Figure 3.2 is performed 10 times for each social group. The repetition is employed to again avoid a possible bias that might occur due to the random choice of individuals.

It starts with filtering the data by social groups. Then, unique individuals are selected on the basis of their identifier. All individuals are shuffled and the first individual's weight is taken as a starting point. Every subsequent individuals' weight is added to the cumulative weight of the preceding individuals. Until the sample size is reached, this process is continued iteratively. This step ensures samples of comparable size, each sample consisting of a selection of individuals with the same social group. Based on each sample a network is constructed as described in Section 3.2 and the analysis as described in Section 3.3 is carried out for each network. This leads to 10 random samples per social group as well as one network and analysis per random sample. To enable comparison across social groups, the measures calculated in the analysis process are aggregated within each social group. Specifically, the average of each measure obtained from the analysis is calculated across the ten random samples to provide a representative summary of each social group.

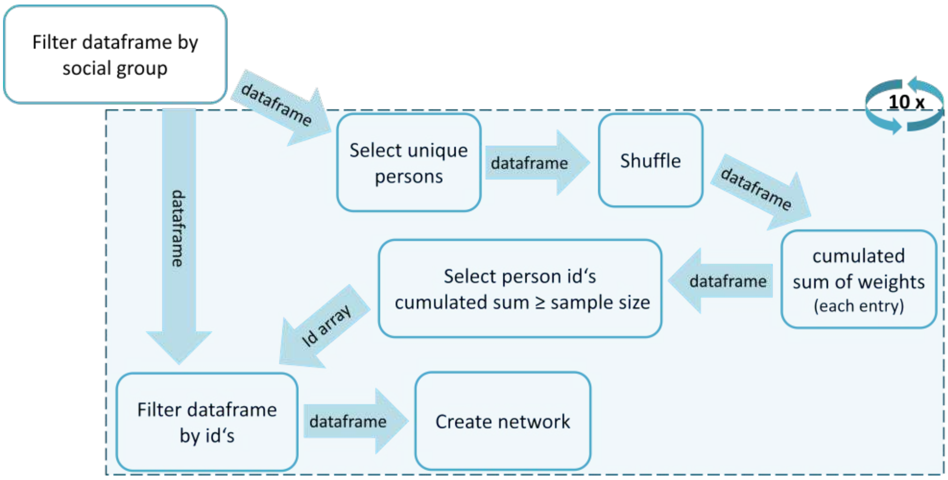


Figure 3.2: Simple Random Sampling Process for a social group

3.2 Network Construction

The construction of the occupation mobility networks is performed for each sample of every social group separately. This process is wrapped into a pipeline, which comprises the random sampling process described in the previous section and the construction process outlined in Section 3.2.3. This facilitates the reproducibility and efficiency of the construction process. The following explains the basics of occupation mobility networks, how their transition probabilities are being calculated as well as the construction process itself and which tools are used.

3.2.1 Occupation Mobility Network

An occupation mobility network as presented by Mealey et al. [LWY$^+$24] is a data-driven approach to analyze the transition behavior of a group of people in the labour market. It is a directed Graph G that consists of a set of nodes V that represent occupations and a set of edges E which represent transitions between occupations as displayed in Figure 3.3. For instance edge $e(i,j) \in E$ represents a transition between occupation i and occupation j, where $i,j \in V$ [LWY$^+$24]. Moreover, each node and edge can be addressed attributes. These are useful properties to store additional information like the transition probability $P_{(i,j)}$ between the occupations i and j, namely the probability that the corresponding edge $e(i,j)$ is followed [MdF18].

3.2.2 Transition probabilities

The probability of any person transitioning from one occupation to a connected other, is given by the transition probability. This probability P of transitioning from occupation i to occupation j, as stated by Mealy et al. [MdF18] is denoted by:

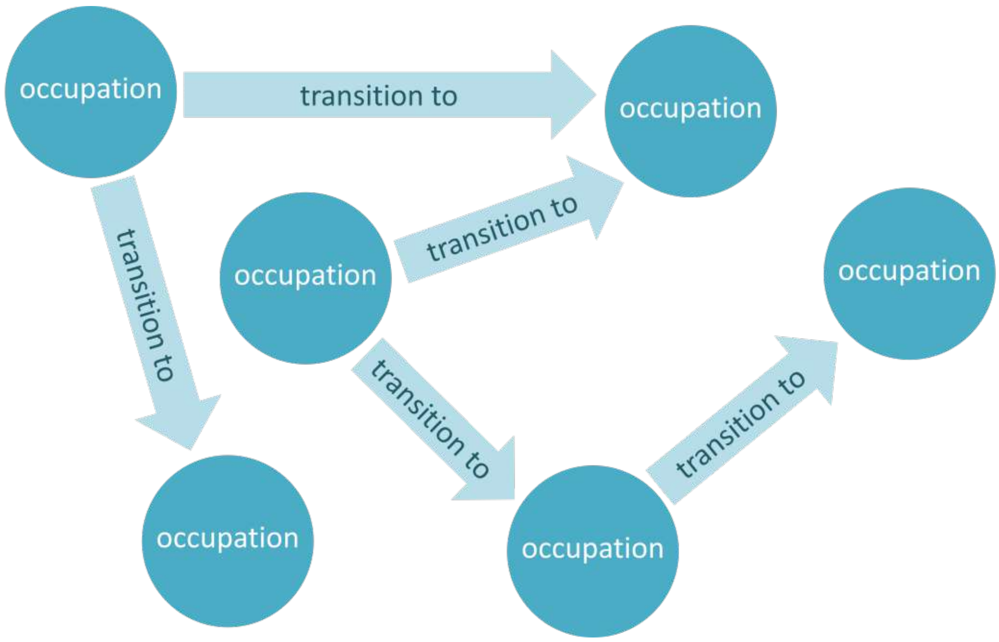


Figure 3.3: Basic structure of an occupation mobility network where nodes represent occupations and directed edges represent transitions from one occupation to another (based on del Rio-Chanona et al. [dMB+21]

$$P_{ij} = \frac{T_{ij}}{T_i} \tag{3.4}$$

Namely the number of individuals that transitioned from occupation i to occupation j (T_{ij}) divided by the number of people that transitioned away from occupation i (T_i). In this work, each individual is additionally weighted according to their assigned weight. For example, if an individual has a weight of ten, this indicates that they represent ten Austrian individuals with similar characteristics. Consequently, such a case is interpreted that not one but ten individuals transitioned from occupation i to occupation j. Incorporating the weights of individuals leads to the weighted transition probabilities denoted by the following formula:

$$P_{ij} = \frac{\sum\limits_{x \in \tau_{ij}} w_x}{\sum\limits_{y \in \tau_{i*}} w_y} \tag{3.5}$$

where τ_{ij} represents a set of individuals that transitioned from occupation i to occupation j and τ_{i*} denotes a set of individuals that transitioned away from occupation i to any other occupation. The weight of an individual is denoted by w. The calculated weighted transition probabilities are stored as edge attributes. Following the execution of this calculation, the occupation mobility network can also be regarded as a Markov Chain.

3.2.3 Construction process

After regrouping the variables holding the key features migration background, age, and education according to the final groups from Section 3.1.3, random samples are drawn for each social group and separate networks are constructed as described in the following. Since the personal information like the unique person identifier is lost and only aggregated values per occupation are kept in this process, the construction is based on the samples created previously.

It starts with determining occupation transitions. As stated previously, each individual is represented in the dataset up to five times. Therefore, it is possible to filter those people who have stated different occupations in consecutive interviews. Next, the variables **new occupation** and **old occupation**, which hold information about an individuals current and previous occupation, are introduced. Based on this information, the transition probabilities are calculated according to Equation 3.5. This results in a transition probability dataframe, where each entry represents an old and a new occupation as well as the weighted probability of transitioning from the old to the new one.

Each unique occupation code, combining old and new ones, is added to an empty directed graph as a node. All transitions extracted from the previously mentioned transition probability dataframe are added to the graph as edges. On average the created networks contain ∼100 nodes and ∼570 edges. Lastly, several node attributes are added. Those include the average income per occupation. In this study, the weighted mean monthly net income for each occupation is calculated based on the social group as a whole. Even for the same occupation, the mean monthly net income can vary across different social groups, indicating inequality. Furthermore, the occupation name and more general occupation category are added as node attributes. This facilitates a clearer interpretation and allows analysis results to be more directly linked to real-world scenarios and assumptions. In addition, the percentage of people working in each occupation, as well as the share of people with low, medium and high education working in each occupation, are added as node attributes. The share of people S with a certain feature, e.g. low education level, is calculated as follows:

$$S_i = \frac{F_i}{T_i} * 100 \tag{3.6}$$

where F_i stands for the amount of people with the feature in question working in occupation i, and T_i being the total number of people working in occupation i. In order to express the result as a percentage, it further is multiplied with 100.

Finally the network is saved and can be reloaded for further use either using networkx for calculating measures or using gephi for a visual analysis.

3.2.4 Tools

For constructing and later analyzing the network, the open source python package networkx is used. Networkx can be used for creating, manipulating and analyzing complex network structures. The software under discussion also provides implementations of numerous graph algorithms, including shortest paths and structure measures. It is well-documented and straightforward to integrate into Python code. Furthermore, it offers the option to save the network in several popular file formats, thus facilitating further use in a range of tools, including Gephi [HSS08].

3.3 Network Analysis

This section focuses on the difference between occupation mobility networks for the social groups defined previously. Starting off with an exploratory analysis, centrality measures are computed for each group. The analysis offers insights and shows inequalities between social groups, which justifies further investigation.

The following presents three distinct methodologies for demonstrating an individual's chances of securing a position in a high-income occupation after starting out in a low-income occupation. These methodologies are shortest paths, Random walks and Markov simulations. The other approaches listed are utilised as a baseline for the purpose of verifying the results. However, before applying the necessary computation, the step of categorizing occupations into low and high-income occupations needs to be conducted. Based on the entire Austrian population within the observed time frame of 2011-2022, the thresholds for an income being considered as low or high are defined. Whilst an income is considered low if it is below the 25 percentile (< € 1386,-), it is considered high if it is above the 75 percentile (> € 2692,-). With those values set, for each network individually the occupations are categorized into low, medium, and high-income occupations considering the average income per occupation per social group, which was calculated and stored as node attribute when constructing the network.

3.3.1 Gini Index

The Gini index is a popular measure to highlight inequalities of any distribution, like the income distribution of a population [Has23]. Throughout the years, many different formulations of the Gini index were published [CV12]. Originally it was developed in 1912 by Gini and is highly associated with the Lorenz curve. In other words, the Gini index measures the ratio of the area between the Lorenz curve and the equidistribution line to the area of maximum concentration [BL06]. The Lorenz curve is a line representing the cumulative shares. Namely, the share of the total income earned by which percent of the population. In a perfect world where income is shared equally, this would be a straight line, as e.g. 30% of the population would earn 30% of the total income. Hence, the Lorenz curve describing this perfect scenario can also be referred to as the line of equality or equidistribution line as stated by Donovan [Don25]. The Gini index measures the difference between this perfect scenario and the actual curve. In case the calculated result is close to 0, the area between the curves is small, the income can be interpreted as being equally distributed. A 1 indicates that the income is not equally distributed amongst the population [Has23].

Being such a popular measure of income inequality, the Gini index is used as a baseline for the further computed measures.

3.3.2 Shortest Paths

A common problem in Graph theory is the one of shortest paths. It aims at finding the shortest route between two vertices or nodes within a graph. This problem can also be found in many real-world scenarios, like maps or in case of this research occupational connections. There exist several well-known algorithms to conquer this task. They include Dijkstra, Bellman-Ford and other algorithms [MMJ13]. For this research, the Dijkstra algorithm is used to detect shortest paths between low and high-income occupations. This algorithm is based on the assumptions that all edge weights are positive, there is a finite number of nodes and the starting point is always a single node but the target can be any other node [Che03]. The basic process looks as follows [MMJ13]:

```
1. start at any low-income node S
        mark distance = 0
        mark all other nodes distance = ∞

2. select non-visited node with smallest distance
3. for each neighbor
        calculate distance
        ∑ distances from previous to current node

        if new distance < currently known distance
             update distance
4. if there are unvisited nodes
        continue with 2.
   if all nodes are visited
        end
```

This research focuses on paths between low an high-income occupations. Hence, all possible shortest paths, starting at an occupation categorized as low income and targeting high-income occupations, are calculated. That way it is possible to determine the *number of shortest paths* available. The rationale behind this value is to compare if each social group has similar number of connections between low and high-income occupations.

Secondly, the *average shortest path length between low and high-income occupations* is computed. Again, discrepancies might occur since one social group might averagely need to visit more nodes to reach a node considered high income when starting at a low income node.

Possible significant disparities observed between the two calculated measures for the designated social groups suggest the presence of inequalities in occupational mobility. These disparities serve to illustrate the varying opportunities available to individuals from diverse backgrounds in attaining high-income occupations.

3.3.3 Random Walk

The objective of the random walk is to simulate the career path of an individual. Where the walker represents the individual person. At each step a random successor of the current node, representing an occupation, is selected as this individual's next occupation. Although in practice individuals do not typically select their subsequent occupation at random, such approaches have been demonstrated to yield favourable and surprising outcomes in similar contexts. Such models are popular not only because of their simplicity but also generality [RM21]. On the other hand, one can argue that there is a certain randomness involved when selecting a new occupation. A person might be introduced to some colleague or friend who influences them choosing a different career, or they simply want to completely reorient themselves professionally. Hence, the new occupation choice may seem random from the outside.

Contrary to related work, our walks always start on a randomly selected low-income occupation. The stopping criteria is defined based on the expected outcome. The two outcomes capture different aspects of upward occupation mobility. Outcome 1 captures the required effort measured in the number of steps of an individual to secure a high-income occupation. In this scenario a random walk stops once a high-income occupation or a maximum number of steps is reached. A high number indicates bad prospects in terms of reaching a high-income occupation. In other words, a large number of steps represents more complex career paths, suggesting limited upward occupation mobility. Whilst a lower number showcases high connectivity between low and high-income occupations and indicate better chances in working in a high-income occupation. Large discrepancies between the resulting values for the different social group can hence indicate inequalities in the labour market.

Outcome 2 captures the likelihood of success or in other words the probability of reaching a high-income occupation after a certain number of steps. In that case a walk stops after the defined number of steps or if a high-income occupation is reached. A high probability on reaching a high-income occupation, indicates a high connectivity between low and high-income occupations. The larger this value, the higher the chance of reaching a high-income occupation after the given amount of steps.

Combining the two outcomes provides a comprehensive overview of the labour market. Together, they demonstrate the length of career paths and the likelihood of securing a high-income occupation. In order to obtain statistically relevant results, 500 random walks are performed.

```
randomly select low income occupation

while True:
    if occupation has successor
        select random successor from current occupation
    else
        break

    if selected occupation is considered high income
       break
    if maximum number of steps (30) is reached
       break
    else
       continue
```

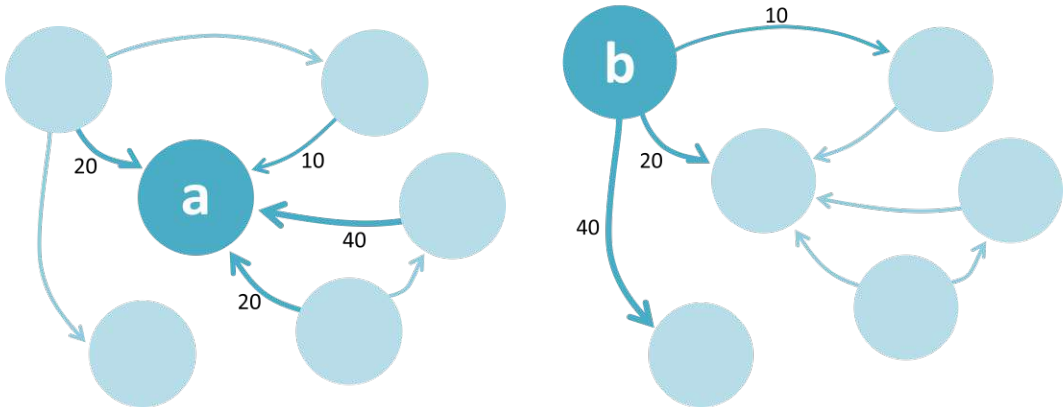
3.3.4 Markov Simulation

A Markov process or Markov chain is a discrete-time stochastic process that fulfills the Markov property. This property defines a process for which each state is determined exclusively by the preceding state and does not depend on any further past values [Pri18]. Mathematically this property requires that for all subsets S of states s the following holds:

$$Prob\{s_{t+1} \in S | s_t = P_t\} \tag{3.7}$$

where $t \in \mathbb{N}$ being a time sequence, where in each time step the process is in a specific state s. The state s_{t+1} solely depends on the previous state s_t and no further historic information is necessary [Her02]. Transition probabilities are used to describe the likelihoods of moving between all states from one time step t to the following $t + 1$. By moving a walker, whose decision on where to move next is defined by the specified transition probabilities, is then a simulation of the Markov chain [AH19].

In the context of this research, the potential states are defined as occupations, with the transition probabilities denoting the probability of switching from one occupation to another. In the absence of a connection between occupations, the transition probability is considered to be equal to zero. The simulation setup is similar to the one in Section 3.3.3. Namely, each simulation starts in a state or occupation considered low income. It ends as soon as a state labeled as high income or a specified number of steps is reached. Again, two different outcomes are sought. One being the average number of steps needed until a high-income occupation is reached, and the other one being the probability of reaching a high-income occupation after a certain amount of steps.



(a) in-degree centrality - Dark blue node a has an in-degree centrality of 4, since 4 arrows point towards it. Its weighted in-degree is 20 + 10 + 40 + 20 = 90.

(b) out-degree centrality - Dark blue node b has an out-degree centrality of 3, since 3 arrows lead away from it. Its weighted out-degree is 10 + 20 + 40 = 70.

Figure 3.4: Visualization of how to calculate in- and out-degree centrality in a weighted occupation mobility network

3.3.5 Centrality and Clustering

Centrality measures are widely used in network analysis to describe phenomena like social influence [PZC+18], detect opinion leaders [PLDP17] or detect the source of rumor [SA19]. The bare concept of centrality is very important in understanding social networks [PZC+18]. There exists a variety of metrics by which the centrality of a node can be ascertained. These metrics serve as indicators of the node's relevance within the network. The selection of an appropriate metric is determined by the specific context [PLDP17]. In this research the utilization of centrality measures is twofold. Firstly, they are employed for exploratory analysis of networks, and secondly, they facilitate quantitative analysis, thereby contributing to the interpretation of results. The measures employed are weighted in-degree centrality, weighted out-degree centrality, and betweenness centrality.

The **degree** of a node represents the number of edges connected with that node [SA19]. In the context of a directed network, the degree can be subdivided into in-degree and out-degree. The in-degree centrality metric quantifies the number of edges directed towards the observed node (see Figure 3.4a), whereas the out-degree centrality metric calculates the number of edges directed away from the node (see Figure 3.4b) [PZC+18]. If a directed network is additionally weighted, the weights of incoming and outgoing edges are taken into account. Hence, in this work the number of transitions between the occupations are incorporated into the calculation as such weights [OAS10].

Betweenness describes how often a node appears in the shortest paths between any other nodes. Namely it measures the number of times a node acts as a bridge between others [SA19]. Considering the shortest paths between a start node $i \in V$ and a target node $j \in V$. In case of $i = j$ the shortest path between those nodes would be equal to 0. The betweenness centrality B for the node v is then defined as follows:

$$B_v = \sum_{i,j \in V} \frac{\sigma(i,j|v)}{\sigma(i,j)} \tag{3.8}$$

where $\sigma(i,j|v)$ denotes the number of shortest paths between node i and node j passing through some node $v \neq i,j$. $\sigma(i,j)$ represents the overall number of shortest paths between i and j [Bra08].

These centrality measures help us to understand why groups have a low or high number of shortest paths or a high or low chance for an individual to reach a high-income occupation. Depending on whether low-, medium- or high-income occupations are among the most central for any social group, this can indicate the chances of an individual belonging to this group of reaching a high-income occupation. For example, a low chance could be explained by a high in-degree indicating that many people transition to low or medium-income occupations. Social groups with many central high-income occupations are expected to have a higher chance of reaching a high-income occupation.

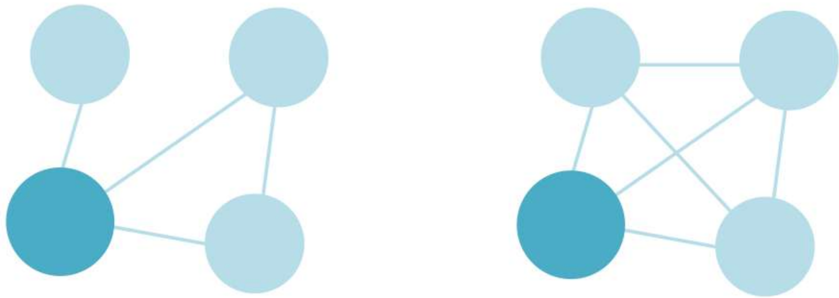
The **Clustering coefficient** is another measure applied to assist in interpreting and explaining results obtained in terms of network connectivity. There are two different versions, namely the local and global clustering coefficient. The local version focuses on a single node and measures its closeness to its neighbors, whereas the global version represents the overall connectivity of nodes within the network [BSD+23]. In essence, a high global clustering coefficient is indicative of a high density of connections within the network. To illustrate this concept, consider a network that represents individuals as nodes and their friendships as edges. In such a network, a low clustering coefficient would imply that the friends of one individual are not necessarily friends with each other [HSS11]. In mathematical terms the global clustering coefficient can be described as:

$$C = \frac{1}{N} \sum C_i \tag{3.9}$$

with N being the number of nodes in the network and C_i representing the local clustering coefficent of node i [Kai08]. The local clustering coefficient is defined by

$$C_i = \frac{\frac{T_i}{deg_i(deg_i - 1)}}{2} \tag{3.10}$$

where T_i is the number of connections between neighbors of node i and deg_i represents the degree of node i [AM22]. Figure 3.5 shows an example of how the local clustering coefficient is calculated.



(a) low clustering coefficient: $\frac{\frac{1}{3(3-1)}}{2} = \frac{1}{3}$ (b) high clustering coefficent $\frac{\frac{3}{3(3-1)}}{2} = 1$

Figure 3.5: Calculation of local clustering coefficient - example given for the dark blue node

In this research, the global clustering coefficient is utilised as a means to gain insights into the degree of interconnectedness amongst disparate social groups within their respective occupation mobility networks. A high connectedness among low-income occupations compared to other occupations can indicate lower chances of transitioning to a higher income occupation, due to the disproportionate large number of connections to other low-income occupations. Whilst a similar average clustering coefficient among income-related occupational subgraphs is expected to yield higher chances of reaching a high-income occupation when starting at a low-income occupation.

CHAPTER 4

Results

This chapter presents the results obtained by this study. It begins with descriptive statistics, outlining the evolution of key features over the observed period and examining the distribution of monthly net income across individuals with different socioeconomic features. Based on these findings eight social groups are defined. Results derived from occupation mobility network analysis are then reported, focusing on shortest paths, Random walks and Markov walks. These results are accompanied by a robustness check that supports the methodological choices such as how many walks should be performed to obtain reliable and robust results. Finally, the chapter concludes with a quantitative analysis utilizing the clustering coefficient which serves as an additional validation check of the insights obtained by the walk and shortest path measures. All results are based on the Austrian microcensus and income data provided by AUSSDA [The25].

4.1 Descriptive Statistics

Over the entire observed time period, the original dataset contains 2.127.197 entries - per year there are on average ~177.266 entries available. The Austrian population grew over the years. Whilst in 2011, the weighted sum of people given in the dataset was ~8.269.175, this number increased up to ~8.900.823 in 2022. The weighted sum of people should account for the correct number of inhabitants in Austria.

The dataset records 125 unique occupations and 10 occupational categories into which the occupations are grouped. Overall, the most common occupation is *Shop Salespersons*. On the other hand least common occupations include *Ships' Deck Crews and Related Workers* and *Traditional and Complementary Medicine Professionals*. More generally, the most common occupation category in 2022 are *Professionals*, which is displayed in Figure 4.1. Least amount of people are found to work in an occupation categorized as *Armed Forces Occupations*.

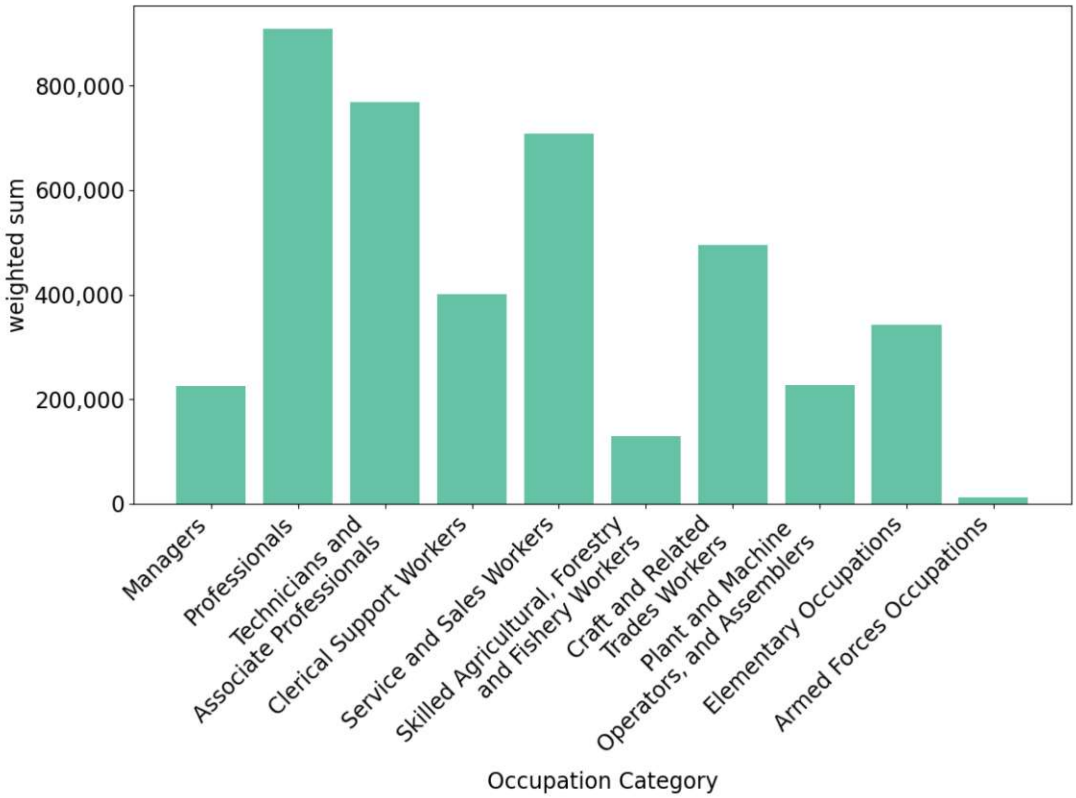


Figure 4.1: Amount of Austrian individuals being occupied in each occupation category in 2022 based on the Austrian microcensus data [Aus22]

Before performing the cleaning and pre-processing process described in Section 3.1, some possible combinations of the key feature categories (e.g. male non-migrants with low education between 20 and 24 years) do not even contain individuals. Moreover, the weighted frequencies also vary greatly, e.g. in 2022 between ∼17 and ∼471.197. This is why it is necessary to take these preparatory steps. They eliminate the existence of empty groups and reduce the difference between the weighted frequencies of each possible combination. Afterwards, the smallest group found are male migrants with high education below 30 years. In 2011 this group only contains ∼5.107 and in 2012 ∼5.487 weighted individuals. According to Statistik Austria ([Aus24]), assumptions made about this group in these two years have to be handled with care. Groups with a weighted sum of individuals of less than 6.000 are subject to randomness. All other combinations for the remaining years do exceed this threshold, where as a reference the weighted values for 2022 are displayed in Table 4.1. Moreover, since the main grouping does not involve the educational aspect, resulting sizes of the groups used for the main analysis largely exceed the mentioned threshold. As stated in Section 3.1, this aspect is only used for explaining results in depth. Hence, even though only ∼40% of the original dataset is left after cleaning, the group sizes are still sufficient to draw statistically correct conclusions.

Overall, an average Austrian has held 1.1 occupations over the observed time period. Among those who have held at least two occupations, the average number is 2.2. The group with the highest average number of occupations are male individuals without a migration background over 30, with 2.3 occupations, whilst the group with the lowest average number of occupations are migrant women under 30, with 2.1 occupations.

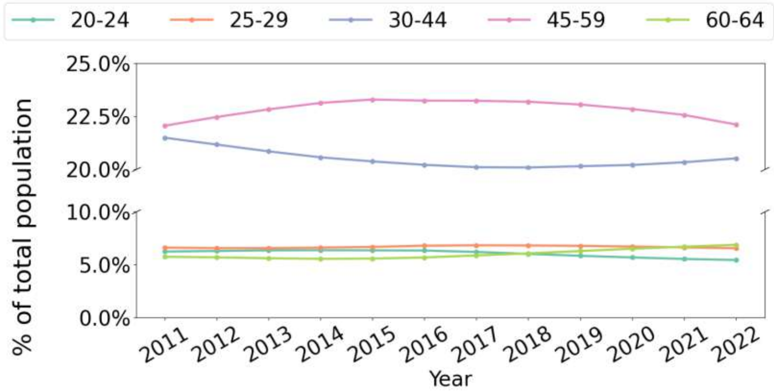
			low education	medium education	high education	total
no migration background	male	< 30	∼18.111	∼165.778	∼98.485	∼282.374
		> 30	∼62.105	∼601.580	∼411.039	∼1.074.723
	female	< 30	∼10.677	∼126.903	∼117.988	∼255.568
		> 30	∼91.391	∼532.094	∼371.476	∼994.961
migration background	male	< 30	∼26.206	∼58.3834	∼30.425	∼115.014
		> 30	∼88.929	∼192.276	∼146.863	∼428.068
	female	< 30	∼9.865	∼54.454	∼34.572	∼98.892
		> 30	∼78.174	∼164.508	∼147.554	∼390.235

Table 4.1: Weighted number of people per key feature category combination (weighted contingency table) after pre-processing including cleaning and re-grouping of the features migration background, age and gender (based on Austrian microcensus 2022 [Aus22])

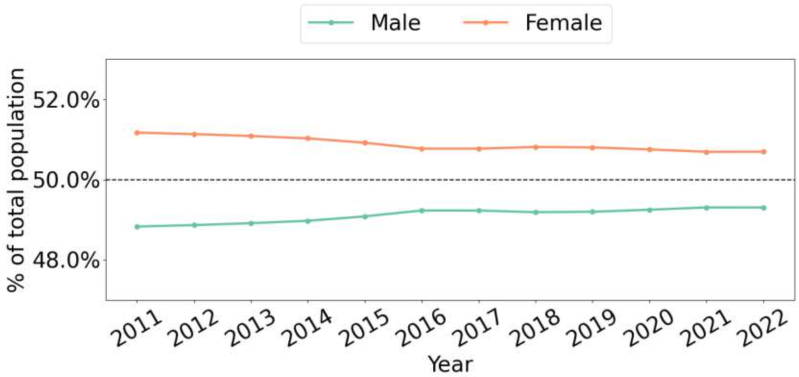
4.1.1 Evolution

In total the focus lies on four key features, which are an individual's age, gender, migration background, and educational attainment. Looking at those features individually, it can be seen that they have evolved similarly. Figure 4.2 represents the share of people per feature category of the entire Austrian population. Overall, the shares remained very stable over the years. As shown in Figure 4.2b, there are slightly more women than men in Austria, with the difference of the share becoming smaller in the last years.

In terms of migration background a slight shift towards an increasing percentage of individuals with migration background can be observed in Figure 4.2c. Taking a closer look at educational attainment in Figure 4.2d, it shows that the highest education level is upper secondary education for over 40% of Austrians, remaining constant. It needs to be considered that from 2014 onwards variables related to education are mapped according to ISCED 11, whilst before they were mapped according to ISCED 97. There are only 7 levels in ISCED 97, whilst in ISCED 11 the level 'tertiary' is split into 'short-cycle tertiary', 'bachelors equivalent' and 'masters equivalent' to account for the Bologna-system which introduced Bachelor and Master degrees in Austria in 1999. This is why in 2014 a distinct increase in individuals with a masters degree is observable, since this attainment level did not exist before. Moreover, the transition from ISCED 97 to



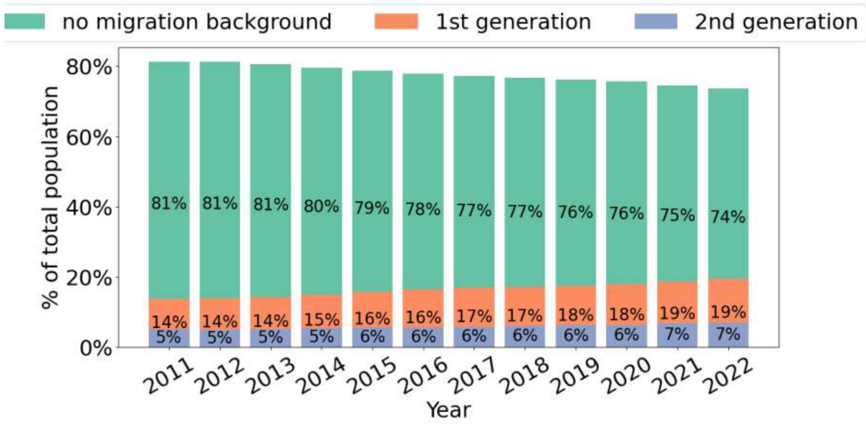
(a) Age groups



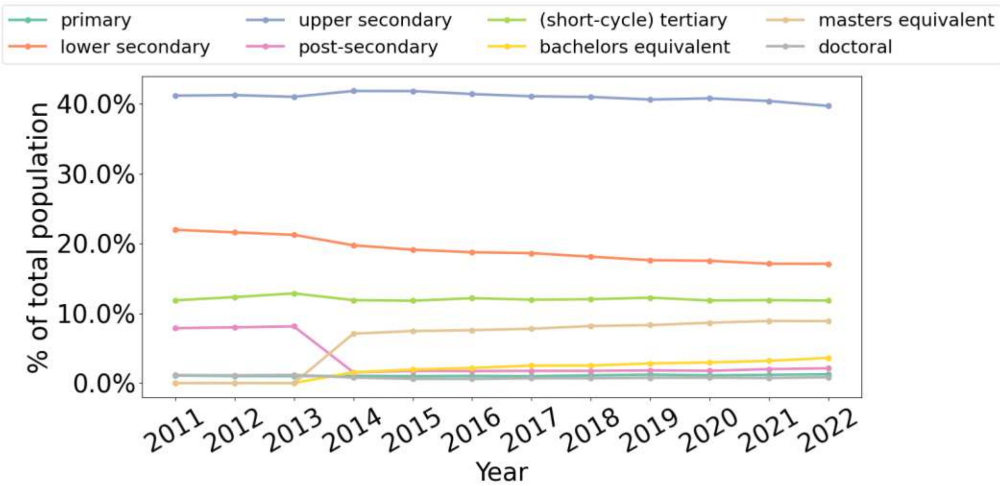
(b) Gender

Figure 4.2: Evolution of the share of individuals of the Austrian population belonging to a certain category per key feature within the years 2011 to 2022 based on the Austrian microcensus data (Part 1)

ISCED 11 led to changes in the classification of education programmes, particularly at post-secondary and higher levels [Hip12]. Therefore, educational attainment before and after 2014 must be compared cautiously.



(c) Migration Background



(d) Educational attainment

Figure 4.2: Evolution of the share of individuals of the Austrian population belonging to a certain category per key feature within the years 2011 to 2022 based on the Austrian microcensus data (Part 2)

4.1.2 Income Distribution and Social Group Definition

The disparities between different social groups are measured on the basis of their monthly net income. Therefore, it is relevant to investigate the underlying income distributions of individuals with the same feature categories, e.g. male and female individuals. The observed features migration background, age, gender and educational attainment are divided into several categories in the original source dataset. However, using these original categories to divide the Austrian population into social groups would lead to defining too many groups. Hence, different categories are introduced for each feature based on the following results which are derived from the 2022 income statistics provided by AUSSDA [The25], leading to 8 social groups.

Migration background

The first feature observed is an individuals migration background. The variable holding information about migration background is divided into three different categories. People who are born in a country other than Austria are considered 'first generation' migrants. If a person was born in Austria but both of their parents were born in another country, they are categorized as 'second generation' migrants. People who themselves and at least one of their parents were born in Austria are considered to have no migration background [Aus24]. Figure 4.3 visualizes the monthly net income distributions for the three original categories of migration background. First and second generation migrants show a similar distribution and their difference in their median is only €53. The Wasserstein distance which is closely related to the median, shows a similar result. Moreover, the standard deviation between these two categories is more similar and shows a much smaller effect size compared to the category no migration background as seen in Table 4.2. This small effect size together with the income distribution shows that the income distributions of these two categories have a high degree of overlap. They are therefore grouped together, as seen in Table 4.3, and the final categories are: migration background, no migration background.

Categories		effect size	Wasserstein distance	Median difference	Standard deviation difference
no	1st generation	0.31	353.91	313.0	186.41
1st generation	2nd generation	0.06	74.16	53.0	39.05

Table 4.2: Income distribution similarity of original migration background categories (2022)

original	customized
no migration background	no migration background
1st generation 2nd generation	migration background

Table 4.3: Original and customized categories of migration background

Age

Age is already categorized into 5-year groups as presented in a previous Section. Focusing on adults in the usual work age leaves us with people between 20 and 64 years. Table 4.4 shows that comparing the underlying income distribution of older age groups tends

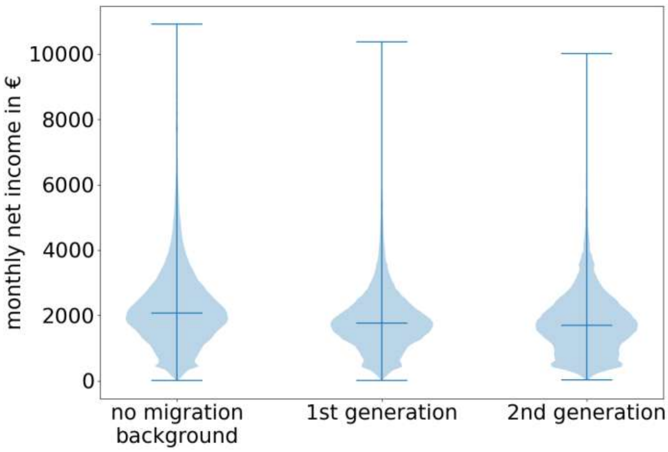


Figure 4.3: Monthly net income distribution by original migration background categories (based on Austrian microcensus and income data 2022 [Aus22])

to have a very small effect size indicating their similarity which is also visible in Figure 4.4. Younger age groups, especially the youngest one between 20 and 24 years, tend to have a lower median monthly net income than older age groups. These results show, that in 2022 income discrepancies are more evident for younger people. Looking at Table 4.4 the largest income dissimilarities based on their respective median and Wasserstein distance are found between the age groups '20-24' and '25-29', '25-29' and '30-34', as well as '55-59' and '60-64'. It is decided to only use 2 categories, since the low effect size underlines that people above 30 show a similar income distribution. Hence, age is split in people below and above 30 as shown in Table 4.5.

Categories		effect size	Wasserstein distance	Median difference	Standard deviation difference
20-24	25-29	-0.43	361.20	307.0	107.66
25-29	30-34	-0.21	213.03	143.0	184.74
30-34	35-39	-0.10	125.86	32.0	172.89
35-39	40-44	-0.05	67.86	17.0	82.46
40-44	45-49	-0.08	114.28	81.0	80.73
45-49	50-54	-0.07	100.62	80.0	63.19
50-54	55-59	-0.06	111.62	75.0	97.31
55-59	60-64	-0.09	425.24	152.0	439.54

Table 4.4: Income distribution similarity of original age groups (2022)

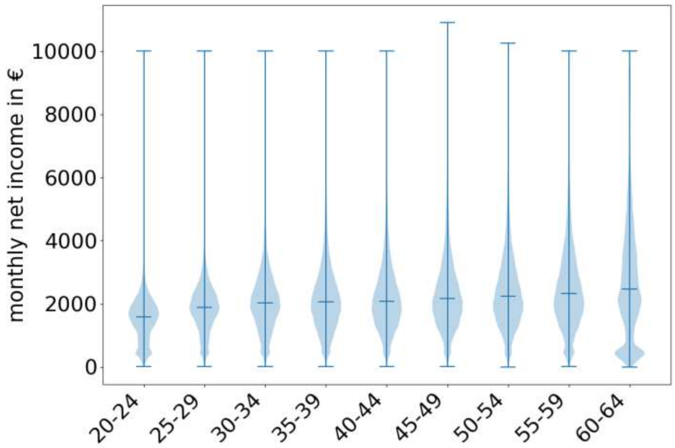


Figure 4.4: Monthly net income distribution by original age groups ((based on Austrian microcensus and income data 2022 [Aus22]))

original	customized
20-24 25-29	below 30
30-34 35-39 40-44 45-49 50-54 55-59 60-64	above 30

Table 4.5: Original and customized categories of age

Gender

The microcensus questionnaires exclusively provided the options of 'male' or 'female' as gender categories. Starting in 2020, a third option,'other', was introduced. According to Statistik Austria [Sch25], the number of non-binary individuals surveyed is insufficient for separate evaluation. Consequently, imputation rules are employed to assign a binary gender category to each interviewed individual [Sch25]. Hence, the dataset used provides two different categories. The medium, nearly large, effect size displayed in Table 4.6 indicates a large dissimilarity between the underlying monthly net income distributions of men and women. Figure 4.5 as well as the other calculated statistics underline this disparity by resulting in large values. In 2022, the median income of men is € 655,- higher than the median income of women. These great income differences indicate that no further grouping is necessary (4.7).

Categories		effect size	Wasserstein distance	Median difference	Standard deviation difference
male	female	0.75	768	655.0	361.09

Table 4.6: Income similarity of gender (2022)

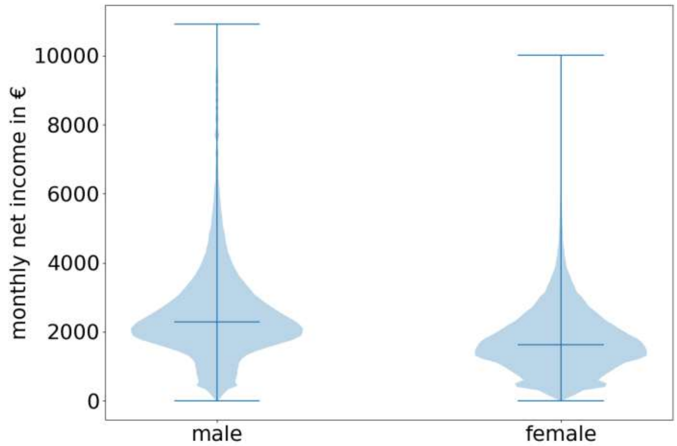


Figure 4.5: Monthly net income distribution by gender (based on Austrian microcensus and income data 2022 [Aus22])

original	customized
male	male
female	female

Table 4.7: Original and customized categories of gender

Educational attainment

The educational attainment level is categorized based on the 'International Standard Classification of Education' (ISCED) [eur]. Up until including 2013 ISCED 97 and starting from 2014 ISCED 11 was applied [Aus21d]. There are several levels of granularity, where this work uses the 1-digit coding scheme which encodes the educational attainment. Other coding schemes use more digits and encode more detailed information, e.g. 3-digit coding schemes encodes educational attainment and education programmes ([UNE12]). Here, educational attainment is divided into 9 levels. Except for the lowest level 'early childhood education', all other categories (listed in Table 3.1) are found in the dataset. Eurostat [eur] suggests the three aggregation levels, low, medium and high education. Primary and lower secondary education are considered low. Upper secondary and post-secondary would be grouped as medium education, where the rest is considered high education [eur]. Figure 4.6 indicates that the higher someones' educational attainment, the higher their income. However, having primary or lower secondary education results in similar income distributions, which is also proven by the very low effect size displayed in Table 4.8. The largest differences, indicated by a medium effect size and larger Wasserstein distances than for the other comparisons, are found between 'lower secondary' and 'upper secondary' as well as 'bachelor's' and 'masters'. The final regrouping is displayed in Table 4.9.

Categories		effect size	Wasserstein distance	Median difference	Standard deviation difference
primary	lower secondary	-0.04	77.53	38.0	91.60
lower secondary	upper secondary	-0.58	593.39	575.0	238.49
upper secondary	post-secondary	-0.19	212.99	187.0	115.79
post-secondary	short-cycle tertiary	-0.25	346.94	273.0	237.0
short-cycle tertiary	bachelor's	0.21	281.85	214.0	43.30
bachelor's	master's	-0.59	1009.25	784.0	468.88
master's	doctoral	-0.32	647.36	544.0	190.08

Table 4.8: Income distribution similarity of educational attainment categories (2022)

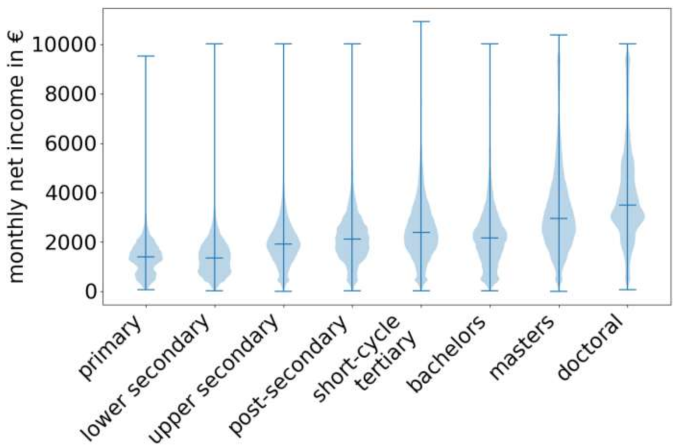


Figure 4.6: Monthly net income distribution by educational attainment category (based on Austrian microcensus and income data 2022 [Aus22])

original	customized
primary lower secondary	low education
upper secondary post-secondary	medium education
short-cycle tertiary bachelor's or equivalent master's or equivalent doctoral or equivalent	high education

Table 4.9: Original and customized categories of educational attainment

Group Definition

Finally the following social groups are defined as shown in Table 4.10.

migration background	female	below 30
migration background	female	above 30
migration background	male	below 30
migration background	male	above 30
no migration background	female	below 30
no migration background	female	above 30
no migration background	male	below 30
no migration background	male	above 30

Table 4.10: Social groups defined by migration background, gender and age as used in further analysis

Additionally, the education level is used for parts of the analysis that are more in depth. Especially when it comes to explaining why certain results occur.

4.1.3 Feature Correlation

Another aspect being analyzed is how the individual features correlate with each other. The following results are based on the year 2022. According to the chi square statistic calculated based on the weighted contingency table, no correlation is found indicated by a p-value below 0.05. The contingency table contains the weighted amount of individuals for every possible category combination of all key features, as already discussed and displayed above. However, taking a closer look at each feature combination still yields valuable insights. The values presented are rounded values, resulting in small shifts to the exact values, but are nevertheless meaningful in understanding the broad picture. Table 4.11 shows the overlap of features in percentage. The rows account for the total population with a given feature, while the columns represent the share of individuals that additionally have the corresponding feature. In other words the table indicates the share of people with the row feature that also have the column feature. For example out of all people with migration background (row 1), 21% are below 30 and 79% are above 30 years old.

		Migration		Age		Gender		Education		
		yes	no	< 30	> 30	female	male	low	medium	high
Migration	yes	100%	0%	21.0%	79.0%	51.3%	48.7%	24.2%	43.0%	32.9%
	no	0%	100%	18.9%	81.0%	49.2%	50.8%	9.9%	54.3%	35.8%
Age	< 30	31.5%	68.5%	100%	0%	48.6%	51.4%	11.6%	52.6%	35.9%
	> 30	28.8%	71.2%	0%	100%	50.1%	49.9%	14.7%	50.6%	34.7%
Gender	female	30.2%	69.8%	19.0%	81.0%	100%	0%	15.8%	49.1%	35.0%
	male	28.4%	71.6%	20.0%	80.0%	0%	100%	12.4%	52.8%	34.8%

Table 4.11: Share of individuals with one feature (row) and another (column) (based on the Austrian microcensus 2022 [Aus22]). Rows represent the total group with a given feature, and columns represent the percentage of that group that also has the corresponding feature - e.g. 21% of individuals with migration background are below 30 years old.
The percentages are based on the observed population only (individuals between 20-64 years containing valid information in the observed features Migration, Age and Gender are considered being the total population)

As seen in Table 4.11 (row Gender), a slightly higher share of female individuals have a migration background. In total, as already seen in Section 4.1.1, there are slightly more women than men. Overall ∼70% of all Male and Female individuals do not and about ∼30% do have migration background.

Comparing migration background with educational attainment, it can be seen in Table 4.11 that for both people with and without migration background, having at most a medium education is most common. A closer look shows that a similar share of people among migrants and non migrants have a high education. However, differences become clear when focusing on low education. Whilst only ∼10% of individuals without migration background do have low education, even ∼24% of people with migration background have this educational attainment level. Taking a look from a different perspective reveals that about half of the individuals with low education do and the other half does not have migration background. On the other hand more than 70% of people with medium and high education do not have any migration background.

When looking at age in combination with migration background, as shown in Table 4.11, it can be seen that around 80% of individuals with and without migration background are above 30 years old. For people below 30 having a migration background is slightly more common than for those above 30 which is indicated by the fact that ∼31% of those below 30 and only ∼29% of those above 30 do have migration background.

Gender and Educational attainment are related in the way that female individuals are overrepresented in low education. Where ∼56% of people with low education are women. Diving deeper shows, that whilst ∼12% of all men, ∼16% of all women have at most a low education. On the other hand men are slightly overrepresented in medium and high education. This is shown by the fact that ∼52% of individuals with medium education and slightly above 50% of those with high education are men. However, the discrepancy in high education is very minimal.

Among both male and female individuals, the age groups are distributed similarly, as shown in Table 4.11. The main difference found is that younger age groups below 30 are slightly dominated by men, whilst the older age groups above 30 are slightly dominated by women. Slightly in this context means that the share of male and female individuals within those groups differs by ∼1-3%.

The last possible combination investigated is age and educational attainment. In general having a medium education is most common among all age groups. Whilst having only low education is least common for both age groups. However, out of all individuals with low education, over 80% are above 30. In case of people with medium and high education slightly below 80% are above 30.

Overall, this analysis shows that the largest discrepancies lie between gender and education. It is observed that women are overrepresented when it comes to having low education. Important to note is to, especially when considering migration background, also take into account that the categories of each feature are not always evenly distributed. Taking migration background as an example, it is known from the previous section that in total there are more individuals without than with migration background, leading to higher values when e.g. looking at the share of (non) migrant individuals per gender.

4.2 Network Analysis

This main part of the analysis is based on the occupation mobility network capturing the time period between 2011-2022 explained in Section 3.2. The separate networks for each social group are analysed separately and the combined results are presented below.

4.2.1 Central occupations

The weighted in-degree, weighted out-degree and betweenness centrality are calculated for each occupation per social group. The 5 most central occupations for the social groups above 30 including their calculated centrality value are listed (see Table 4.12, 4.13), whilst the results for the remaining groups can be found in the Appendix B.1.

Major similarities are found within male groups and female groups. The most central occupation among male individuals without migration background, regardless of their age group, is *Physical and Engineering Science Technicians*, as shown in Table 4.12. Differences for this group arise in age when looking at the second and third most popular occupations. Whilst *Shop Salespersons* record a high influx, many individuals below 30 leave the occupation *Machinery Mechanics and Repairers*. The older generation records a high influx and outflow in *Building Frame and Related Trades Workers*.

Female individuals without migration background of any age most prominently switch to and away from *Shop Salespersons*. Whilst *General Office Clerks* record a high influx in each age group. *Waiters and Bartenders* have a high beetweenness centrality and weighted out degree among those below 30 (see Appendix B.1). For individuals above 30, whose top occupations are displayed in Table 4.12, *Domestic, Hotel and Office Cleaners and Helpers* reaches a higher beetweenness centrality.

Among people with migration background, *Shop Salespersons* are central. Men, regardless their age, record a high influx in assistance jobs like *Transport and Storage Labourers* and *Mining and Construction Labourers*.

Female individuals with migration background record a high influx in *Administrative and Specialized Secretaries* and *Waiters and Bartenders* among the younger generation. For those above 30, *Domestic, Hotel and Office Cleaners and Helpers* is the most central occupation among all centrality measures, as displayed in Table 4.13. But also *Shop Salespersons* and different assistance jobs record a high influx in both age groups.

4.2.2 Upward mobility

Based on the occupation mobility network representing the Austrian labour market, various measures are calculated to model paths from low-income occupations to high-income occupations. The underlying concepts for those measures are shortest path, Random walk and Markov walk, where their results are compared to the Gini coefficient used as a baseline.

weighted in degree	weighted out degree	betweenness
Male		
Physical and Engineering Science Technicians (1.0)	Physical and Engineering Science Technicians (1.0)	Physical and Engineering Science Technicians (0.077)
Building Frame and Related Trades Workers (0.552)	Building Frame and Related Trades Workers (0.605)	Heavy Truck and Bus Drivers (0.036)
Sales and Purchasing Agents and Brokers (0.531)	Shop Salespersons (0.602)	Machinery Mechanics and Repairers (0.033)
Shop Salespersons (0.499)	Machinery Mechanics and Repairers (0.563)	Shop Salespersons (0.032)
Administrative and Specialized Secretaries (0.468)	Sales and Purchasing Agents and Brokers (0.501)	Administrative and Specialized Secretaries (0.031)
Female		
Administrative and Specialized Secretaries (1.000)	Shop Salespersons (1.000)	Shop Salespersons (0.142)
General Office Clerks (0.999)	General Office Clerks (0.870)	Domestic, Hotel and Office Cleaners and Helpers (0.088)
Shop Salespersons (0.978)	Administrative and Specialized Secretaries (0.848)	General Office Clerks (0.087)
Domestic, Hotel and Office Cleaners and Helpers (0.638)	Nursing and Midwifery Associate Professionals (0.602)	Administrative and Specialized Secretaries (0.065)
Nursing and Midwifery Professionals (0.631)	Domestic, Hotel and Office Cleaners and Helpers (0.562)	Waiters and Bartenders (0.0378)

Table 4.12: Top 5 occupations by centrality measure for individuals without migration background above 30, the top representing male and the bottom female individuals (based on Austrian microcensus 2011 - 2022 provided by AUSSDA [The25])

The results are presented in Table 4.14. It shows that groups above 30 do have a higher number of shortest paths between low and high-income occupations while the average shortest path length is similar for all social groups. The number of shortest paths takes alternate routes into consideration, reflecting real-world scenarios which are not always straightforward due to environmental influences. The average path length however only describes the structural distance between low and high-income occupations. A similar average shortest path length therefore indicates that it is possible for everyone to reach a high-income occupation, while the number of shortest paths reflects actual real-

weighted in degree	weighted out degree	betweenness
Male		
Mining and Construction Labourers (1.000)	Building Frame and Related Trades Workers (1.000)	Shop Salespersons (0.097)
Building Frame and Related Trades Workers (0.952)	Mining and Construction Labourers (0.720)	Physical and Engineering Science Technicians (0.087)
Building Finishers and Related Trades Workers (0.788)	Building Finishers and Related Trades Workers (0.646)	Sales and Purchasing Agents and Brokers (0.065)
Transport and Storage Labourers (0.743)	Manufacturing Labourers (0.639)	Car, Van and Motorcycle Drivers (0.053)
Car, Van and Motorcycle Drivers (0.717)	Car, Van and Motorcycle Drivers (0.602)	Mining and Construction Labourers (0.053)
Female		
Domestic, Hotel and Office Cleaners and Helpers (1.000)	Domestic, Hotel and Office Cleaners and Helpers (1.000)	Domestic, Hotel and Office Cleaners and Helpers (0.200)
Shop Salespersons (0.680)	Shop Salespersons (0.632)	Shop Salespersons (0.193)
Food Preparation Assistants (0.448)	Waiters and Bartenders (0.416)	Manufacturing Labourers (0.084)
Personal Care Workers in Health Services (0.418)	General Office Clerks (0.365)	General Office Clerks (0.063)
Manufacturing Labourers (0.388)	Food Preparation Assistants (0.360)	Administrative and Specialized Secretaries (0.063)

Table 4.13: Top 5 occupations by centrality measure for individuals with migration background above 30, the top representing male and the bottom female individuals (based on Austrian microcensus 2011 - 2022 provided by AUSSDA [The25])

world scenarios. Besides few outliers, having no migration background results in higher shortest path measures. A higher number of shortest paths between low- and high-income occupations indicates that individuals over 30 have a higher chance of transitioning to a high-income occupation. Conversely, having only a few shortest paths, as is the case for individuals with migration background under 30, showcases boundaries in upward career mobility. These boundaries suggest that individuals with such sociodemographic characteristics may find it more challenging to reach a high-income occupation, as only a few paths lead from a low-income occupation to a high-income one.

Results obtained from Random and Markov walks indicate that people above 30 on average need fewer steps to reach a high-income occupation when starting at a low-income occupation. Again, similar to what the number of shortest paths tells us, these few steps indicate less boundaries in upward career mobility for individuals above 30. Except for individuals with migration background below 30, Male groups averagely need to take less steps than the respective female groups. This suggests that, on average, women need to secure more jobs than men before reaching a high-income occupation. Assuming that people only transition between occupations a limited number of times, the greater number of steps required to reach a high-income occupation reduces the likelihood of ever achieving a high income.

			nr of shortest paths	avg shortest path length	avg Random walk steps	avg Markov walk steps
no Migration	< 30	Male	67	3.0	46	55
		Female	59	3.0	58	60
	> 30	Male	107	3.3	2	2
		Female	284	3.0	6	9
Migration	< 30	Male	14	2.6	49	45
		Female	54	3.1	22	29
	> 30	Male	98	3.2	4	5
		Female	200	3.2	9	15

Table 4.14: Mean measures per social group - Shortest path measures are the mean based on 19 samples per social group drawn from the original social group; Walk measures are based on the original occupation mobility networks containing the entire social group not only a sample

Group Differentials

The following is performed to assess how well shortest path as well as the number of steps needed in a Random or Markov walk capture income inequality similar to the Gini coefficient. The Gini coefficient measures inequality based on the distribution of income within a group. By merging two social groups and calculating their respective Gini coefficients, the result highlights the extent of inequality related to income within these two groups. Then for each pair of social group the following differences between the two groups are being calculated to determine income and career path inequalities and further referred to as difference measures:

- number of shortest paths between low and high-income occupations
- average shortest path length between low and high-income occupation
- average number of steps needed in a Random walk starting at any low-income occupation to reach any high-income occupation
- average number of steps needed in a Markov walk starting at any low-income occupation to reach any high-income occupation

For example, calculating the difference in the number of shortest paths between groups 1 and 2 showcases the disparity in chances of reaching a high-income occupation. If the difference is large, it indicates that one group has a better chance of reaching or at least has more opportunities to reach a high-income occupation than the other, which is seen as an indicator of inequality. The correlation between all measures is now computed. If the correlation between the Gini coefficient and any other difference measure is high, this suggests that the difference measure captures information that is similar to that captured by the Gini index. This indicates disparities in access or opportunities related to income. Conversely, if the correlation is low, the two measures reflect different dimensions of inequality. This indicates that the difference measure in question provides complementary insights beyond those revealed by common income-based measures and that the two measures are not appropriate for measuring the same thing, but rather shed different perspectives. As shown in Figure 4.7, the Pearson correlation coefficient indicates a moderate correlation between the Gini coefficient and the number of shortest paths, indicating these two measures encode similar insights and conclusions about income inequality and accessibility of high-income occupations. The average number of steps needed to reach a high-income occupation when starting from a low one shows a lower correlation with the baseline. This indicates that they not necessarily carry the same information as the baseline, but can be used as a complementary perspective to draw a fuller picture of the labour market.

Probability

A novel approach in identifying and visualizing the differences in upward mobility between different social groups is the probability of reaching any high-income occupation when

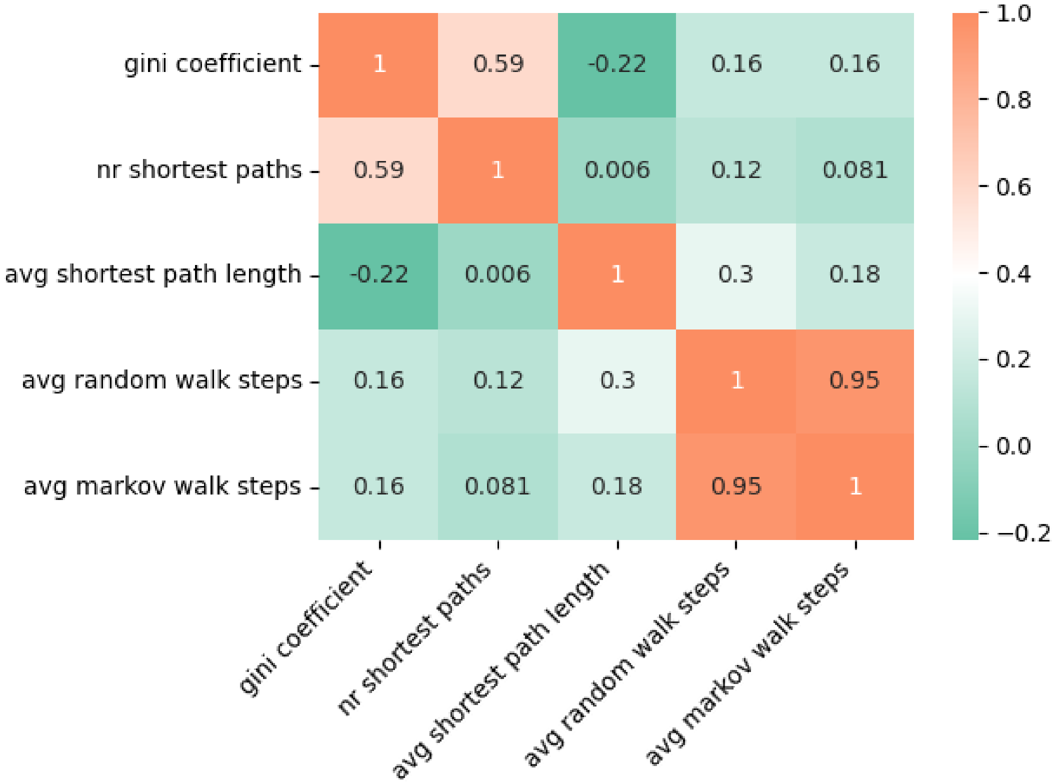


Figure 4.7: Pearson Correlation of inter-group differences of each measure. For example the difference in the number of shortest paths between group 1 and 2 in correlation with the Gini coefficient of the intersection of group 1 and 2.

starting at any low-income occupation after a certain number of steps. Since Markov walks incorporate transition probabilities calculated based on the Austrian population, they are closer to the real-world scenario and are therefore used for showcasing the probability results. Therefore, a walker starts at any low-income occupation and walks until either the specified amount of steps are reached or it reached a high-income occupation. The decision on which neighboring occupation to transition to is based on transition probabilities. In total a simulation consists of 500 walks, where the results are averaged across all walks. The decision about the right number of walks to assure robust results is argued in Section 4.2.3.

Figure 4.8 shows the results of this model applied to the Austrian labour market. It shows that in general, individuals over 30 have a higher chance of reaching a high-income occupation. Even after only taking 5 steps 80% of male individuals without migration background above 30 have reached a high-income occupation. Whilst individuals below 30 at most reach a probability of 19% when taking 30 steps.

Differences are also observable between men and women. Groups including female individuals on average have a 15% lower probability than male groups.

Groups with migration background always show slightly lower probabilities than the respective groups without migration background. This gap is shrinking with probability. Meaning, the chances for reaching a high-income occupation are more distinct between migrants and non-migrants for male individuals above 30, who have large probabilities, than for female individuals below 30 who only show very low probabilities.

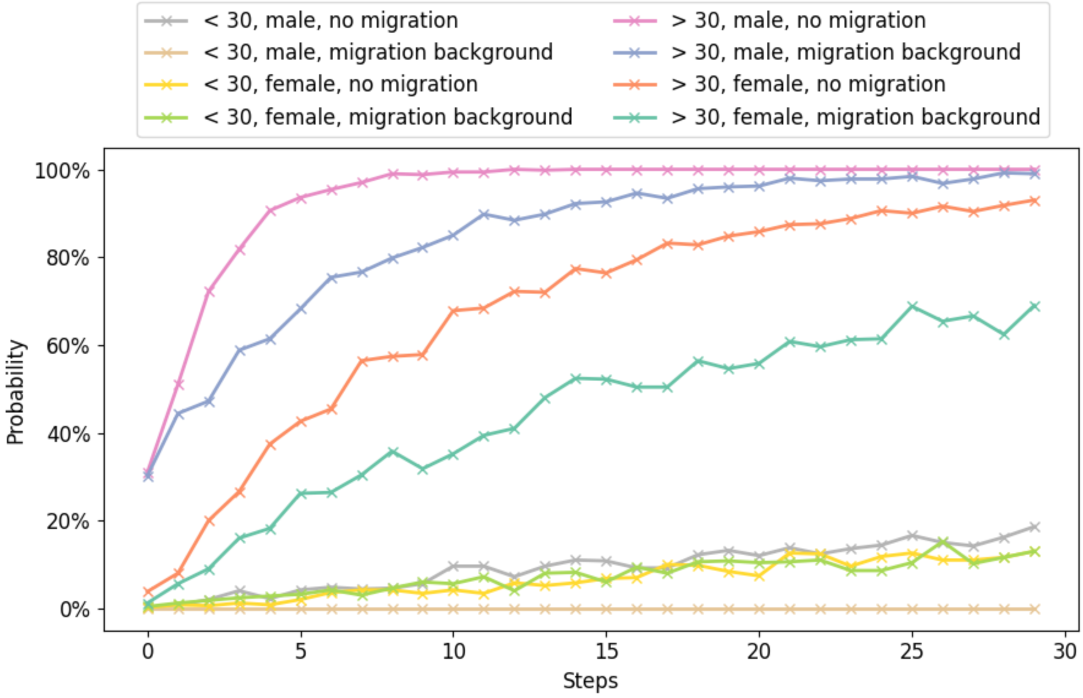


Figure 4.8: Probability of reaching a high-income occupation when starting in a low-income occupation in a Markov walk after at most the specified number of steps

4.2.3 Model Robustness

Walks

When simulating behaviour in the labour market using Random and Markov walks, the simulation consists of multiple walks. One walk involves starting from a random low-income node and selecting a neighbouring node until a high-income node is reached. This process is repeated multiple times, and the final simulation result is the average number of steps taken per walk.

However, the number of walks needs to be predefined. To figure out which is the correct number of walks necessary to reach a stable result, relative standard error of the mean is used. Therefore, for each social group, the entire simulation process is performed multiple

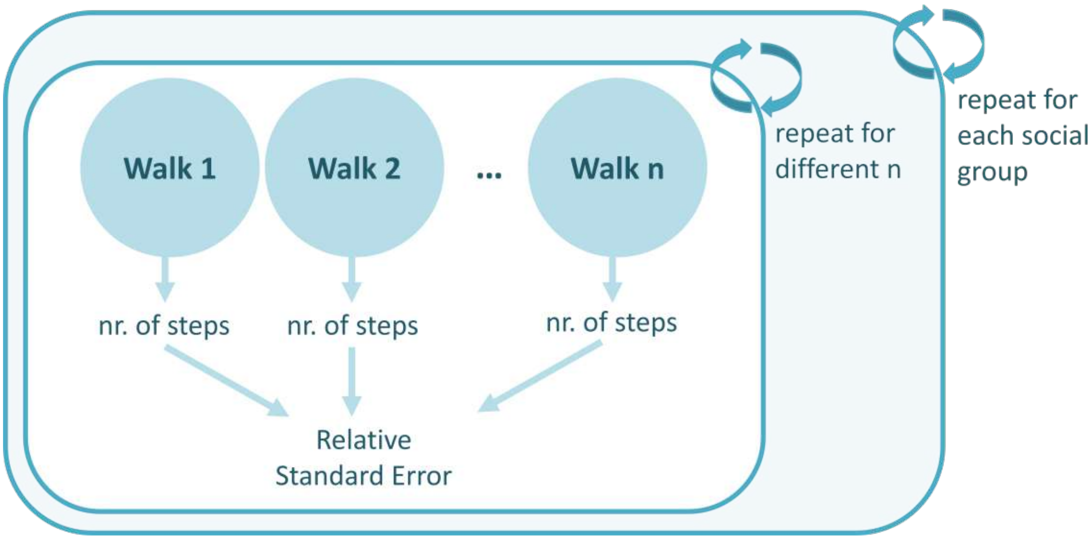


Figure 4.9: Process of robustness check of Random/Markov walks

times with n walks, as shown in Figure 4.9. Where n increases with each iteration and takes values between 100 and 3000. Comparing the resulting relative standard error of the mean (RSEM) shows that it quickly stabilises. This behaviour is displayed in Figure 4.10. Even though the results differ for the different social groups, the trend of how the RSEM evolves is similar. The threshold used for considering obtained estimates as accurate or relevant differs across research. In labour market research, a RSEM of 8% is often found to indicate accurate results [Sta23]. It is evident from the findings that 500 walks per simulation are an adequate trade-off for deriving valid conclusions and computation time. This is evidenced by the fact that for each sociodemographic group network, the RSEM is below 8% when 500 walks are run.

Sampling

In order to reach stable results for the shortest path related measures, a certain number of samples to perform the analysis process on are necessary. Therefore, new networks are sampled the respective amount of times for each social group. The analysis process is performed on each sample and its results are combined by calculating the mean per social group. The entire procedure of sampling, analyzing and combining is carried out 10 times. Finally, the standard deviation and the RSEM of the 10 obtained results is calculated and compared for the different number of samples. The lowest standard deviation indicates low variability in the results and hence more reliability. Similarly a lower RSEM indicates more reliable results. As above, the threshold for the RSEM to consider the obtained estimate accurate is 8%. Due to computational constraints, the number of samples was limited to a maximum of 20, since the computation time increases significantly with each additional sample.

As displayed in Figure 4.11, in case of the average shortest path length between low and

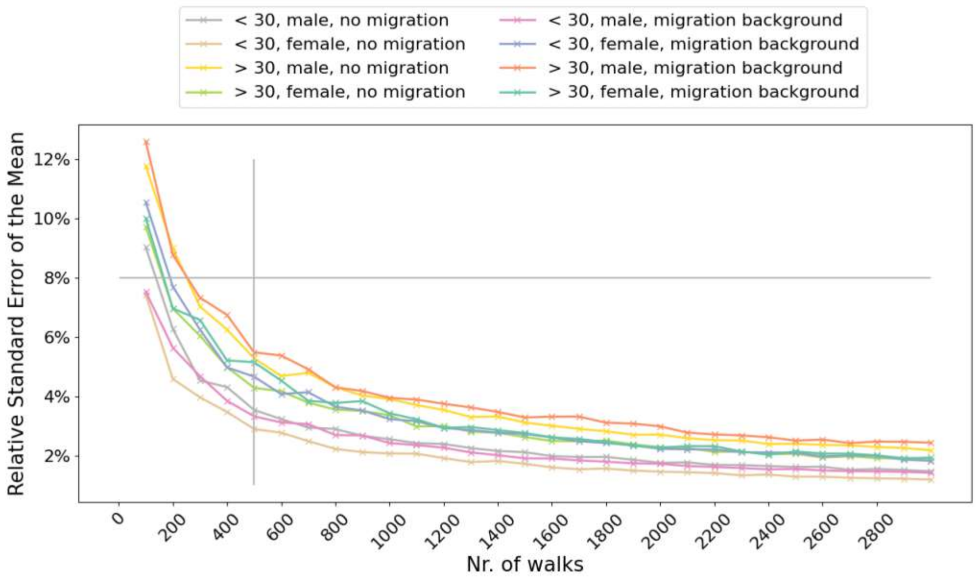


Figure 4.10: Relative Standard Error of the Mean of number of steps needed to reach a high-income occupation starting at a low-income occupation across walks for different number of walks - black horizontal line indicates threshold and black vertical line marks the number of walks selected

high-income occupations, the number of samples is not important. The obtained results for this measure deviate constantly across all sampling amounts tested. Whereas, the lowest standard deviation of ~0.045 can be found when using 18 samples and the second lowest is 0.051 when using 20 samples.

Resulting in higher standard deviations, the number of samples influence the results obtained by the number of shortest paths. The standard deviation and the RSEM calculated for this measure show an overall decreasing trend when the number of samples increases. Hence, a larger number of samples is recommended to obtain stable results for this measure. The most optimal number based on this robustness check is 19 with a standard deviation of ~7 and a RSEM of ~4.7% which is below the 8% threshold. Using 19 samples in this study strikes a balance, ensuring computation time remains manageable while the RSEM still indicates valid and reliable results.

4.2.4 Quantitative Analysis

This part of the analysis delves deeper into the network structures of each social group. This provides insight into group behaviour and helps to draw the right conclusions and understand the results presented earlier. It also helps to determine whether the methods used are appropriate for this use case and can serve as a measure of labour market inequality.

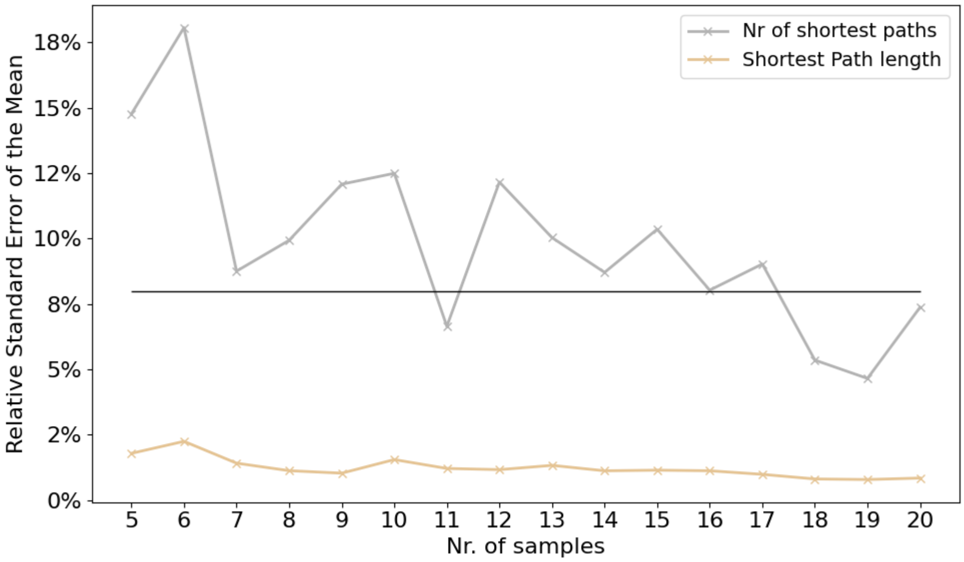


Figure 4.11: Maximum Relative Standard Error of the Mean of calculated measures per number of samples - black line indicates the threshold

The average clustering coefficent shows that male individuals without migration background above 30 are tightly connected in occupational terms. It also shows that in case of all other groups, the female ones are slightly more connected than their respective male groups. This is displayed in Figure 4.12. In other words, tightly connected means that if there are three occupations, if transitions occur from occupation A to B and from occupation A to C, it is highly probable that transitions will also occur between C and B. In a Random or Markov walk, more tightly knit neighbourhoods encourage movements between these neighbourhoods. The interpretation of the coefficient however is influenced by the location of these densely connected neighbourhoods within the network. Therefore, it is zoomed in on low-, medium- and high-income occupations separately.

As seen in Figure 4.13, female groups have a higher average clustering coefficent among low-income occupations than any male group. This indicates that in a Random Walk scenario, a walker working in a low-income occupation is presented many opportunities to transition to another low-income occupation. This reduces the chances of actually transitioning to an occupation belonging to another income category creating boundaries and limiting upward occupation mobility. Especially male groups above 30 have a high coefficient among high-income occupations. In terms of individuals with migration background, results show that they are less connected within all subgraphs than their respective groups without migration background.

Figure 4.14 illustrates that most occupations are considered being a medium-income occupation in nearly all social groups. Whilst the highest amount of high income occupations are found in the group of male individuals above 30 without migration

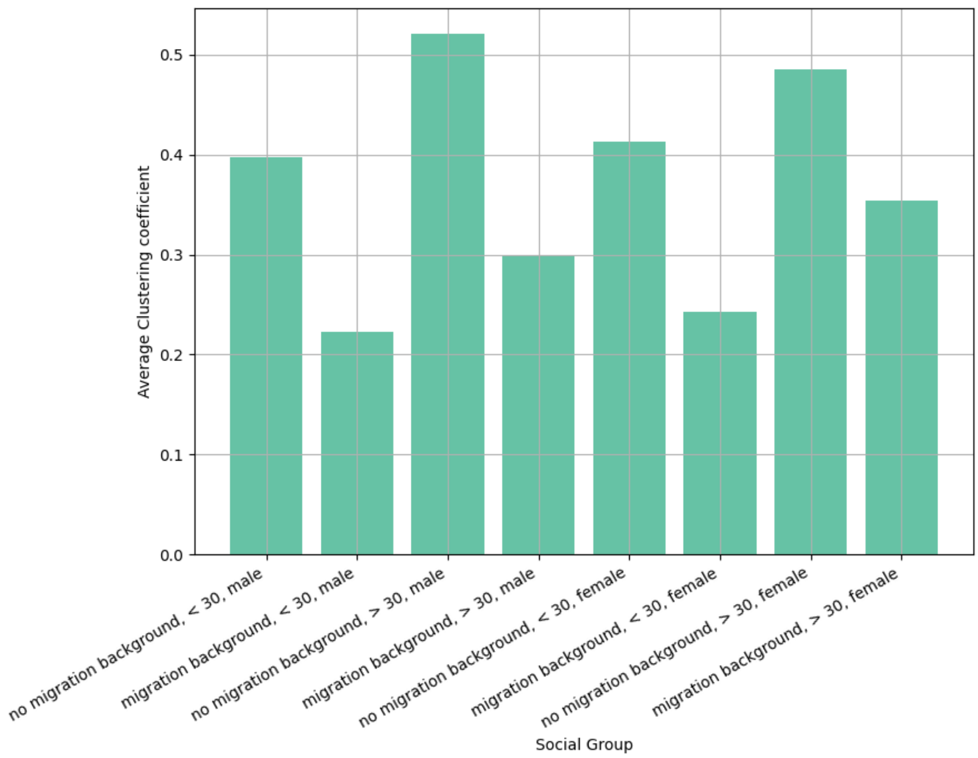


Figure 4.12: Average Clustering coefficient among the entire occupation mobility network, including low-, medium- and high-income occupations, per social group

background. 50% of the occupations in which individuals of this group work are considered high-income ones. The highest percentage of low-income occupations on the other hand are found in the group of female individuals below 30 with migration background. Focusing on age shows that younger groups have a higher percentage in low and medium-income occupations than their respective older groups.

4.3 Summary

This chapter has presented a range of methods aimed at detecting inequalities between social groups in upward occupation mobility at the example of the Austrian labour market. **Descriptive statistics** showing the sociodemographic composition of the population over time, alongside a comparison of income distributions across each sociodemographic category, revealed significant differences between men and women, with men earning on average €655,- more than women. A slightly smaller, yet still significant, median difference in monthly net income of around €313,- was observed between individuals with and without a migration background. Based on these observations, eight social groups were defined to structure the subsequent analyses.

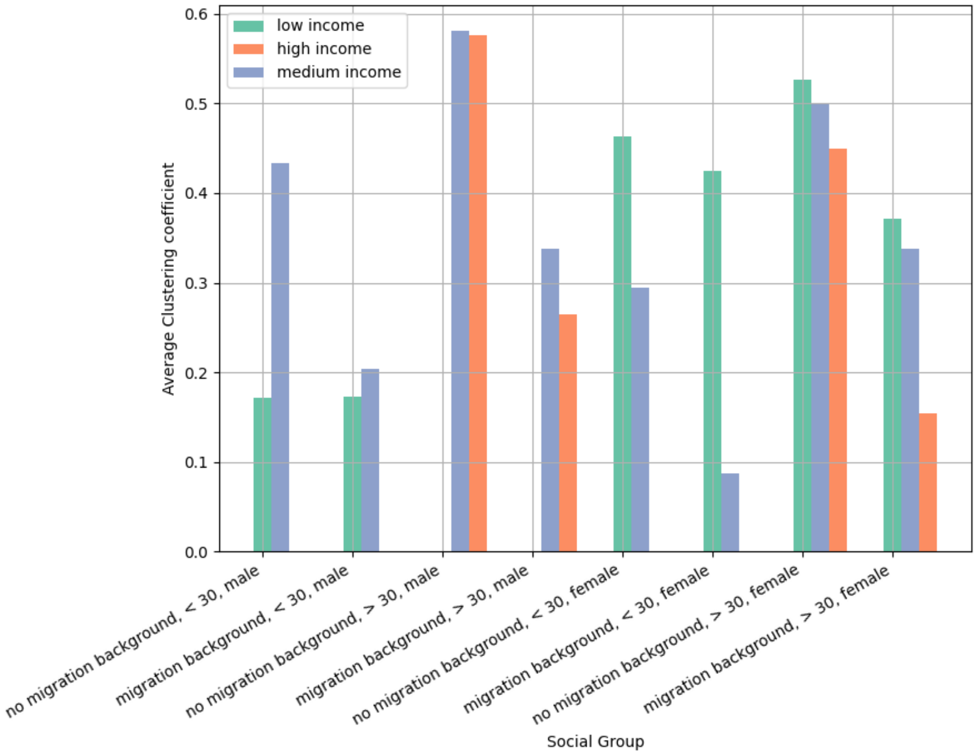


Figure 4.13: Average clustering coefficient for subgraphs of the occupation mobility network containing only low-, medium- or high-income occupations per social group

The network-based analyses provided insights into the structural properties of occupational transitions. Observing **central occupations** for each social group based on the weighted in and out-degree and betweenness centrality reveals that individuals with migration background tend to work more in trade, assistant and service occupations. Central occupations for men with and without migration background of any age include *Physical and Engineering Science Technicians*, *Sales and Purchasing Agents and Brokers* and *Blacksmiths, Toolmarkers and Related Workers*, which have an average monthly net income of more than €2.000,-. Most occupations that are found to be central for women such as *Shop Salespersons* and *General Office Clerks* only have a mean monthly net income below €2.000,-. This already reveals inequalities as women tend to work in lower paid occupations than men.

Focusing on **upward occupation mobility**, paths from low-income occupation to high-income occupations are analysed in detail. Age groups below 30, no matter gender and migration background, only have few shortest paths between 14 and 59. Individuals above 30 on the other hand do have at least 98 shortest paths from low to high-income occupations, indicating more chances of reaching a high-income occupation. The group with the lowest number of shortest paths, 14, are male individuals with

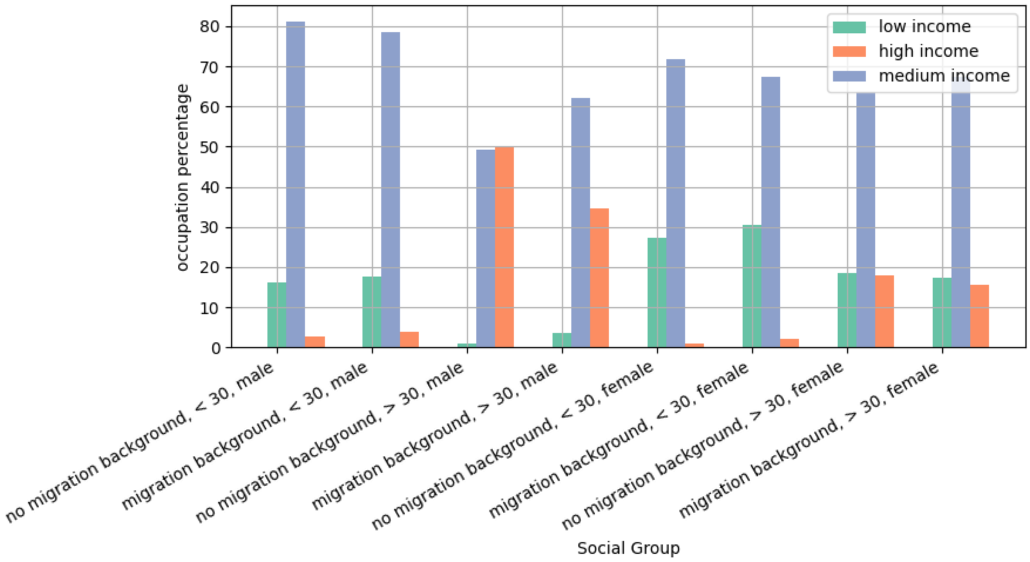


Figure 4.14: Percentage of low-, medium- and high-income occupations per social group

migration background below 30. Women without migration background above 30 have 284 shortest paths. The simulations utilizing Markov walks also indicate large differences in the probability of reaching a high-income occupation in terms of age. From the best performing social group, Men above 30 without migration background, more than 80% manage to reach a high-income occupation within only 5 occupation transitions. The least chances of reaching a high-income occupation has the group of male individuals with migration background below 30 since only 3.7% of occupations represented in their mobility network are high-income occupations. Also, young men with migration background are the smallest of the eight social groups leading to less occupations of any income being represented in their network. These probabilities also reveal discrepancies between men and women. From women of any age or migration background at most 93% reach a high-income occupation, even after taking 30 steps. Male individuals above 30 without migration background even reach a probability of 100% after only 12 steps. Inequalities are also clearly visible for individuals above and below 30. Individuals below 30 at reach a probability of at most 19%. This can again be connected to the fact that individuals below 30 on average earn less that individuals above 30 and therefore have fewer high-income occupations present in their occupation mobility network.

Robustness checks confirmed the stability of the upward occupation mobility results. RSEM computations for up to 3.000 walks per social group have been performed, with valid results being achieved when 500 walks are performed per simulation. The shortest path measures are computed on samples of each social group to ensure similar group sizes and avoid any bias that may arise from it. To ensure stable results, the measures are computed for multiple samples per social group, from which the mean is obtained to compare the social groups with each other. The RSEM indicates that 19 samples per social group is sufficient to derive valid results in a feasible computation time.

The **quantitative analysis** utilizing the average clustering coefficient added evidence that male and female individuals are similarly well connected in their occupation mobility networks. However, female individuals show a high average clustering coefficient in low and medium-income occupations. Men on the other hand have a low average clustering coefficient below 0.2 and higher coefficients in high-income occupations than respective female groups.

CHAPTER 5

Conclusion

Especially in economically challenging times like the ones we are currently experiencing including living costs that tend to increase, attaining a well paid occupation assumes greater importance. Despite the Gini index for Austria falling below the EU average and thus indicating a comparatively low level of income inequality in comparison to other countries, issues such as the gender pay gap remain a topic of significant debate in Austria. Consequently, it is timely and pertinent to make a contribution to these debates by observing the underlying structure of these labour market inequalities that still persist. However, existing measures and approaches, like the Gini index mentioned above, do not do justice to the complexity of the labour market and at the same time showcasing upward career paths. This study sheds new light on how labour market inequalities are represented and measured. This is achieved by applying methods from data science and related studies to a sociological topic. It has been demonstrated that networks which capture the dynamics of the labour market, known as occupation mobility networks, yield valuable insights with regard to labour market research. It is for this reason that they form the foundation of the present work. Established methods known from graph theory and data science are employed. These include shortest paths as well as career path simulations based upon Random Walks and Markov Simulations. This work demonstrates that these methods are valid approaches for this topic. It underlines results obtained by using existing and popular measures like the Gini index and average clustering coefficient. Additionally, the presented methods yield intuitive results. By simulating occupational transitions using real-world data, the outcomes can be expressed in more comprehensible terms suitable for non-experts. For instance, we can determine the average number of steps required for an individual to reach a high-income occupation. These types of measure are more intuitive than abstract statistical values because they link the analysis directly to realistic scenarios that most people are familiar with. The following summarizes its outcome.

To answer the first research question asking *what factors determine upward occupation* *mobility*, results are obtained from the number of shortest path, average shortest path length and the Random and Markov Walk simulations. It is essential to acknowledge the role of upward occupational mobility as a network effect when interpreting the results and answering this question. It is evident that an individual's position within a network significantly influences their probability and capacity to transition to a high-income occupation. For instance, an individual employed in an occupation directly connected with high-income occupations within the occupation mobility network is more likely to advance in their career than an individual employed in an occupation not directly connected to a high-income occupation. Consequently, the interpretations derived from this study describe an average person belonging to a certain social group, while certain individuals experience greater or lesser ease in attaining a high-income occupation due to their current position.

The number of shortest paths as well as the average shortest path length between low and high-income occupations, indicate distinct discrepancies between the observed social groups. Where the discrepancies are more distinct when focusing on the number of shortest paths available.

Based on these measures and the probabilities of reaching a high-income occupation, it can be concluded that age has the highest influence on upward occupation mobility. This can be said because the probability of reaching a high-income occupation has a steeper incline with increasing number of steps for individuals above 30 than those below. Moreover, the number of shortest paths reveals that age is the main source of variation, suggesting that it plays the most important role in reaching a high-income occupation. Of the three factors under observation, age is the only one for which income disparities can be adequately explained. As is apparent, with advancing age comes greater experience. Higher income being the result of greater experience and higher performance is a reasonable justification as already mentioned in Chapter 2 known as the principle of performance.

The second most determent factor is gender. Whilst for men the probability of reaching a high-income occupation increases fast with the number of steps, probabilities for female groups have a more shallow increase. Moreover, even after 30 steps the probability never reaches such a high value for older female groups than for older male ones. The second most noticeable differences in the number of shortest paths were also observed between groups that differed in gender. However, no rational explanation can be proposed for why men should have higher chances in reaching a high-income occupation than women, despite the fact that such disparities are observed. In contrast to age, where increased experience provides a plausible rationale for higher income, gender-related income disparities cannot be attributed to individual capabilities such as experience. Instead, these outcomes are predominantly attributable to societal factors and structural inequalities that persist within the labour market.

Lastly, having a migration background or not is still a crucial factor, as differences in the number of shortest paths and the probability of achieving a high-income occupation are observable. However, these differences are not as significant as those observed for the other factors. Similar to gender-related disparities, migration-related disparities cannot be attributed to inherent experience, ability or performance, but rather, they are shaped by societal factors, including unequal access to opportunities, discrimination, and structural barriers within the labour market.

In summary, while age exerts a more pronounced influence on the probability of attaining a high-income occupation following an initial low-income occupation, these disparities can be at least justified. Gender and migration background have been demonstrated to be significant factors in the attainment of high-income occupations, with no apparent underlying rationale such as experience. Consequently, these two factors are a robust indication of the inequality present in the Austrian labour market.

To prove the validity of these results, we asked *how group differences in distance measures and access dynamics to high-income occupations capture labour market inequality* in our second research question. Comparing the differences in the two shortest path measures between any two social groups with the Gini coefficient of the combined groups suggests that the number of shortest paths is more closely correlated with the Gini coefficient than the average shortest path length. Moreover, the results from the first research question are backed up by a quantitative analysis as well as an analysis of the central occupations of each group.

It demonstrates that groups above 30 have more connections and a higher proportion of high-income occupations than the respective groups below 30. The most central occupations of groups above 30 tend to have a higher income than those of groups below 30. Therefore, it is evident that the probability of achieving a high-income occupation and the number of shortest paths are higher for older age groups.

Even though female groups have similar or higher average clustering coefficients than respective male groups, these groups are highly connected within low-income occupations which hinders upward mobility. In highly clustered low-income sub-networks, individuals are more likely to encounter other low-income occupations when making transitions, due to the dense interconnections between them. This structural feature reduces the chances of reaching a high-income occupation, as the abundance of low-income alternatives disproportionately draws transitions away from higher-income occupations. This explains why the probability of reaching a high-income occupation is lower for female than for male groups. Moreover, female groups tend to have a high in-degree towards occupations either considered low-income or close to being a low-income occupation, with only few exceptions. In case of those exceptions the respective male groups include central occupations which have an even higher average income. However, most male groups have fewer shortest paths between low- and high-income occupations than female groups. Nevertheless, male groups have such a large share of high-income but a very low share of low-income occupations explaining this result. Those shares in combination with lower average clustering coefficients draws transitions towards higher-income occupations. Hence, their chances of reaching a very well-paid occupation are still greater.

The larger probabilities of reaching high-income occupations for groups without migration background can be explained in a similar manner. Even though the differences are less distinct than for other features, groups with a migration background have slightly larger share of low-income occupations. Older age groups with a migration background also have a lower share of high-income occupations. Additionally, all groups with a migration background have a lower average clustering coefficient in each occupation income category than their respective non-migrant groups. Taking a look at the central occupations of groups with migration background reveals an overall higher centrality for lower-income occupations than for respective non-migrant groups. For the same reasons as above, their probability of reaching a high-income occupation is always slightly lower than that of their respective non-migrant groups.

It is evident that there are certain limitations to the present study that could be addressed in future work. One such aspect pertains to the data. The final, cleaned version of the dataset, which is then usable for the purpose of constructing the occupation mobility networks, represents a mere 40% of the original dataset. This is due to the presence of missing values in the observed variables, which necessitated the removal of corresponding entries. Furthermore, the construction of the network exclusively incorporated individuals for whom job transitions have been documented. This is a significant restriction, given that individuals at most occur 5 times in the data. Future work may also investigate and incorporate the probability of staying at the same occupation instead of transitioning to another one. This particular aspect has not been addressed in the present research study, partly due to the aforementioned limitation concerning the duration of observation periods of a person in the dataset, which did not extend beyond five quarters. With regard to the applied approaches, basic versions were adapted for use in this particular case. It is recommended that this aspect be explored further by extending the present versions or testing more complex versions. It would be a valuable contribution to the field to conduct further research into the various types of random walks. At the same time improving the implementation in terms of efficiency and computational time are suggested. Even though attempts were taken to parallelize certain parts of the construction and analysis pipeline, its computational time still has room for improvement. Especially when extending the presented approaches or applying it to larger datasets, this will be a necessity.

Overall the labour market is a complex system in which individuals' decisions about which occupation to transition to depend on multiple factors such as educational attainment, household context, career trajectory, friends and colleagues, including a random aspect. To fully capture the dynamics of occupational mobility, it is therefore essential to observe the labour market from multiple perspectives by applying diverse methods, such as shortest path analysis, upward mobility simulations, as well as clustering coefficients and interpreting their results in combination. Compared to measures already used in this regard, the presented results prove the validity of the proposed adapted approaches. The findings derived from these approaches are congruent with those derived from existing measures utilised as a baseline. They also align with societal expectations such as income inequality between men and women (see [FRB23], [DMW21]) as well as migrants facing difficulties in labour market integration (see [All16]) and hence upward occupation mobility. They can help society understand why certain sociodemographic groups have a harder time in following an upward career path. At the same time this understanding can help to push boundaries in terms of occupation mobility. Further developing the approaches presented here would mean gaining an even deeper and better understanding of inequalities in career advancement opportunities. This in turn helps to define boundaries even more precisely in order to break them and eliminate inequalities in the future.

APPENDIX A

Standards

A.1 International Standard Classification of Occupations ISCO-08

Code	Title
1	Managers
111	Legislators and Senior Officials
112	Managing Directors and Chief Executives
121	Business Services and Administration Managers
122	Sales, Marketing and Development Managers
131	Production Managers in Agriculture, Forestry and Fisheries
132	Manufacturing, Mining, Construction and Distribution Managers
133	Information and Communications Technology Service managers
134	Professional Services Managers
141	Hotel and Restaurant Managers
142	Retail and Wholesale Trade Managers
143	Other Services Managers
2	Professionals
211	Physical and Earth Science Professionals
212	Mathematicians, Actuaries and Statisticians
213	Life Science Professionals
214	Engineering Professionals (excluding Electrotechnology)
215	Electrotechnology Engineers
216	Architects, Planners, Surveyors and Designers
221	Medical Doctors
222	Nursing and Midwifery Professionals
223	Traditional and Complementary Medicine Professionals

Code	Title
224	Paramedical Practitioners
225	Veterinarians
226	Other Health Professionals
231	University and Higher Education Teachers
232	Vocational education teachers
233	Secondary Education Teachers
234	Primary School and Early Childhood Teachers
235	Other Teaching Professionals
241	Finance Professionals
242	Administration Professionals
243	Sales, Marketing and Public Relations Professionals
251	Software and Applications Developers and Analysts
252	Database and Network Professionals
261	Legal Professionals
262	Librarians, Archivists and Curators
263	Social and Religious Professionals
264	Authors, Journalists and Linguists
265	Creative and Performing Artists
3	Technicians and Associate Professionals
311	Physical and Engineering Science Technicians
312	Mining, Manufacturing and Construction Supervisors
313	Process Control Technicians
314	Life Science Technicians and Related Associate Professionals
315	Ship and Aircraft Controllers and Technicians
321	Medical and Pharmaceutical Technicians
322	Nursing and Midwifery Associate Professionals
323	Traditional and Complementary Medicine Associate Professionals
324	Veterinary Technicians and Assistants
325	Other Health Associate Professionals
331	Financial and Mathematical Associate Professionals
332	Sales and Purchasing Agents and Brokers
333	Business Services Agents
334	Administrative and Specialized Secretaries
335	Government Regulatory Associate Professionals
341	Legal, Social and Religious Associate Professionals
342	Sports and Fitness Workers
343	Artistic, Cultural and Culinary Associate Professionals
351	Information and Communications Technology Operations and User Support Technicians
352	Telecommunications and Broadcasting Technicians
4	Clerical Support Workers
411	General Office Clerks

Code	Title
412	Secretaries (general)
413	Keyboard Operators
421	Tellers, Money Collectors and Related Clerks
422	Client Information Workers
431	Numerical Clerks
432	Material Recording and Transport Clerks
441	Other Clerical Support Workers
5	Service and Sales Workers
511	Travel Attendants, Conductors and Guides
512	Cooks
513	Waiters and Bartenders
514	Hairdressers, Beauticians and Related Workers
515	Building and Housekeeping Supervisors
516	Other Personal Services Workers
521	Street and Market Salespersons
522	Shop Aalespersons
523	Cashiers and Ticket Clerks
524	Other Sales Workers
531	Child Care Workers and Teachers' Aides
532	Personal Care Workers in Health Services
541	Protective Services Workers
6	Skilled Agricultural, Forestry and Fishery Workers
611	Market Gardeners and Crop Growers
612	Animal Producers
613	Mixed Crop and Animal Producers
621	Forestry and Related Workers
622	Fishery Workers, Hunters and Trappers
631	Subsistence Crop Farmers
632	Subsistence Livestock Farmers
633	Subsistence Mixed Crop and Livestock Farmers
634	Subsistence Fishers, Hunters, Trappers and Gatherers
7	Craft and Related Trades Workers
711	Building Frame and Related Trades Workers
712	Building Finishers and Related Trades Workers
713	Painters, Building Structure Cleaners and Related Trades Workers
721	Sheet and Structural Metal Workers, Moulders and Welders, and Related Workers
722	Blacksmiths, Toolmakers and Related Trades Workers
723	Machinery Mechanics and Repairers
731	Handicraft Workers
732	Printing Trades Workers
741	Electrical Equipment Installers and Repairers

Code	Title
742	Electronics and Telecommunications Installers and Repairers
751	Food Processing and Related Trades Workers
752	Wood Treaters, Cabinet-makers and Related Trades Workers
753	Garment and Related Trades Workers
754	Other Craft and Related Workers
8	Plant and Machine Operators, and Assemblers
811	Mining and Mineral Processing Plant Operators
812	Metal Processing and Finishing Plant Operators
813	Chemical and Photographic Products Plant and Machine Operators
814	Rubber, Plastic and Paper Products Machine Operators
815	Textile, Fur and Leather Products Machine Operators
816	Food and Related Products Machine Operators
817	Wood Processing and Papermaking Plant Operators
818	Other Stationary Plant and Machine Operators
821	Assemblers
831	Locomotive Engine Drivers and Related Workers
832	Car, Van and Motorcycle Drivers
833	Heavy Truck and Bus Drivers
834	Mobile Plant Operators
835	Ships' Deck Crews and Related Workers
9	Elementary Occupations
911	Domestic, Hotel and Office Cleaners and Helpers
912	Vehicle, Window, Laundry and Other Hand Cleaning Workers
921	Agricultural, Forestry and Fishery Labourers
931	Mining and Construction Labourers
932	Manufacturing Labourers
933	Transport and Storage Labourers
941	Food Preparation Assistants
951	Street and Related Service Workers
952	Street Vendors (excluding Food)
961	Refuse Workers
962	Other Elementary Workers
0	Armed Forces Occupations
011	Commissioned Armed Forces Officers
021	Non-commissioned Armed Forces Officers
031	Armed Forces Occupations, Other Ranks

Table A.1: ISCO-08 Codes and Titles of 1-digit and 3-digit occupation group levels [Int12]

A.2 Austrian Classification of Economic Activities ÖNACE 2008

Code	Title
A	Agriculture, Forestry and Fishing
B	Mining and Quarrying
C	Manufacturing
D	Electricity, Gas, Steam and Air Conditioning Supply
E	Water supply; Sewerage, Waste management and Remediation Activities
F	Construction
G	Wholesale and Retail Trade; repair of Motor Vehicles and Motorcycles
H	Transportation and Storage
I	Accommodation and Food Service Activities
J	Information and Communication
K	Financial and Insurance Activities
L	Real Estate Activities
M	Professional, SCientific and Technical Activities
N	Administrative and Support Service Activities
O	Public Administration and Defence; Compulsory Social Security
P	Education
Q	Human Health and Social Work Activities
R	Arts, Entertainment and Recreation
S	Other Service Activities
T	Activities of Households as Employers; Undifferentiated Goods- and Services-producing Activities of Households for own use
U	Activities of ExtraterritoriaL Organisations and Bodies

Table A.2: ÖNACE 2008 (1-digit) Codes and Titles of first level economic sectors [Aus25]

APPENDIX B

Central Occupations

B.1 Top 5 Occupations - Social Groups below 30

weighted in degree	weighted out degree	betweenness
Physical and Engineering Science Technicians (1.000)	Physical and Engineering Science Technicians (1.000)	Physical and Engineering Science Technicians (0.162)
Shop Salespersons (0.433)	Machinery Mechanics and Repairers (0.594)	Shop Salespersons (0.066)
Machinery Mechanics and Repairers (0.426)	Shop Salespersons (0.548)	Administrative and Specialized Secretaries (0.050)
Building Frame and Related Trades Workers (0.377)	Building Frame and Related Trades Workers (0.514)	Machinery Mechanics and Repairers (0.049)
Building Finishers and Related Trades Workers (0.297)	Blacksmiths, Toolmakers and Related Trades Workers (0.431)	Building Frame and Related Trades Workers (0.038)

Table B.1: Top 5 occupations by centrality measure for male individuals without migration background below 30 (based on Austrian microcensus 2011 - 2022 provided by AUSSDA [The25])

weighted in degree	weighted out degree	betweenness
Shop Salespersons (1.000)	Shop Salespersons (1.000)	Shop Salespersons (0.204)
General Office Clerks (0.986)	Waiters and Bartenders (0.765)	Waiters and Bartenders (0.116)
Administrative and Specialized Secretaries (0.796)	General Office Clerks (0.672)	General Office Clerks (0.104)
Waiters and Bartenders (0.723)	Administrative and Specialized Secretaries (0.522)	Administrative and Specialized Secretaries (0.065)
Nursing and Midwifery Professionals (0.425)	Client Information Workers (0.344)	Physical and Engineering Science Technicians (0.053)

Table B.2: Top 5 occupations by centrality measure for female individuals without migration background below 30 (based on Austrian microcensus 2011 - 2022 provided by AUSSDA [The25])

weighted in degree	weighted out degree	betweenness
Transport and Storage Labourers (1.000)	Shop Salespersons (1.000)	Shop Salespersons (0.152)
Shop Salespersons (0.979)	Transport and Storage Labourers (0.729)	Waiters and Bartenders (0.102)
Physical and Engineering Science Technicians (0.933)	Waiters and Bartenders (0.679)	Physical and Engineering Science Technicians (0.089)
Mining and Construction Labourers (0.677)	Physical and Engineering Science Technicians (0.556)	Manufacturing Labourers (0.063)
Waiters and Bartenders (0.632)	Machinery Mechanics and Repairers (0.548)	Transport and Storage Labourers (0.062)

Table B.3: Top 5 occupations by centrality measure for male individuals with migration background below 30 (based on Austrian microcensus 2011 - 2022 provided by AUSSDA [The25])

weighted in degree	weighted out degree	betweenness
Shop Salespersons (1.000)	Shop Salespersons (1.000)	Shop Salespersons (0.278)
Administrative and Specialized Secretaries (0.570)	Waiters and Bartenders (0.412)	Waiters and Bartenders (0.123)
Waiters and Bartenders (0.551)	Domestic, Hotel and Office Cleaners and Helpers (0.400)	Administrative and Specialized Secretaries (0.103)
General Office Clerks (0.450)	General Office Clerks (0.323)	General Office Clerks (0.100)
Domestic, Hotel and Office Cleaners and Helpers (0.388)	Client Information Workers (0.309)	Domestic, Hotel and Office Cleaners and Helpers (0.099)

Table B.4: Top 5 occupations by centrality measure for female individuals with migration background below 30 (based on Austrian microcensus 2011 - 2022 provided by AUSSDA [The25])

B.2 Average Income

Occupation	Average Income
Domestic, Hotel and Office Cleaners and Helpers	1043.83
Waiters and Bartenders	1176.04
Food Preparation Assistants	1254.54
Car, Van and Motorcycle Drivers	1360.80
Shop Salespersons	1492.52
Client Information Workers	1530.21
Personal Care Workers in Health Services	1616.09
General Office Clerks	1665.95
Transport and Storage Labourers	1677.34
Manufacturing Labourers	1820.31
Mining and Construction Labourers	1897.63
Building Frame and Related Trades Workers	1985.92
Building Finishers and Related Trades Workers	2009.91
Nursing and Midwifery Associate Professionals	2050.92
Heavy Truck and Bus Drivers	2107.60
Administrative and Specialized Secretaries	2120.61
Machinery Mechanics and Repairers	2195.01
Blacksmiths, Toolmakers and Related Trades Workers	2353.27
Nursing and Midwifery Professionals	2511.79
Physical and Engineering Science Technicians	2642.82
Sales and Purchasing Agents and Brokers	2868.25

Table B.5: Average monthly net income per top occupation based on the entire Austrian population with available information about income and occupation (based on Austrian microcensus 2011 - 2022 provided by AUSSDA [The25])

Overview of Generative AI Tools Used

ChatGpt - This tool was used for researching, to find interesting sources and gain a better understanding of the topic. Later, it was also used to rephrase sentences or short paragraphs to make them more concise and understandable to the reader. It was also used to support the coding aspect of this work. This includes general coding support as well as debugging and improving code readability.

Deepl Write - This tool was used to rephrase sentences or short paragraphs to make them more concise and academically appropriate.

List of Figures

List of Tables

Bibliography

[AAV17] Amreen Ahmad, Tanvir Ahmad, and Harsh Vijay. *Shortest Path Algorithms for Social Network Strengths*, pages 353–361. Springer Singapore, 03 2017.

[AH19] Jim Albert and Jingchen Hu. *Probability and Bayesian Modeling*, chapter 9 - Simulation by Markov Chain Monte Carlo. Chapman and Hall/CRC, 2019.

[All16] Bernadette Allinger. Austria: Labour market integration and competences of refugees. *Eurofound*, 2016.

[AM22] Sanchi Arora and Abhijit Majumdar. Machine learning and soft computing applications in textile and clothing supply chain: Bibliometric and network analyses to delineate future research agenda. *Expert Systems with Applications*, 200:117000, 2022.

[Aok20] Satoshi Aoki. Effect sizes of the differences between means without assuming variance equality and between a mean and a constant. *Heliyon*, 6(1):e03306, 2020.

[Aus20a] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2015 (SUF edition). `https://doi.org/10.11587/RRKEA9`, 2020.

[Aus20b] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2016 (SUF edition). `https://doi.org/10.11587/ROYRWO`, 2020.

[Aus20c] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2017 (SUF edition). `https://doi.org/10.11587/TT7HGH`, 2020.

[Aus20d] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2018 (SUF edition). `https://doi.org/10.11587/L33O5L`, 2020.

[Aus20e] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2019 (SUF edition). `https://doi.org/10.11587/XGE7TF`, 2020.

[Aus20f] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2020 (SUF edition). `https://doi.org/10.11587/GXAPI7`, 2020.

[Aus21a] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2011 (SUF edition). https://doi.org/10.11587/KALXOU, 2021.

[Aus21b] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2012 (SUF edition). https://doi.org/10.11587/5NFEWC, 2021.

[Aus21c] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2013 (SUF edition). https://doi.org/10.11587/IYZNKC, 2021.

[Aus21d] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2014 (SUF edition). https://doi.org/10.11587/AQKWM6, 2021.

[Aus21e] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2021 (SUF edition). https://doi.org/10.11587/AROIHY, 2021.

[Aus22] Statistics Austria. Microcensus Labour Force Survey / Housing Survey 2022 (SUF edition). https://doi.org/10.11587/XSWWSL, 2022.

[Aus24] Statistics Austria. Codebook - Mikrozensus ab 2021. https://doi.org/10.11587/EADSOF, 2024.

[Aus25] Statistics Austria. Önace 2008. http://data.europa.eu/88u/dataset/9aac6fe1-d1dd-3bac-8bea-8f1a0858fca7, 2025. Accessed: August 10, 2025.

[BBF22] Johann Bacher, Martina Beham-Rabanser, and Matthias Forstner. Can work value orientations explain the gender wage gap in Austria? *International Journal of Sociology*, 52(3):208–228, 2022.

[BH23] Florian Baumgarten and Julia Hofmann. Diffuse Gefühle: Wie Einkommensunterschiede in Österreich bewertet werden. *Wirtschaft und Gesellschaft*, 49(1):61—-90, May 2023.

[BHG06] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439:462–465, 02 2006.

[BHMZ12] René Böheim, Klemens Himpele, Helmut Mahringer, and Christine Zulehner. The gender wage gap in Austria: eppur si muove! *Empirica*, 40:585–606, 2012.

[BHMZ13] René Böheim, Klemens Himpele, Helmut Mahringer, and Christine Zulehner. The distribution of the gender wage gap in Austria: evidence from matched employer-employee data and tax records. *Journal for Labour Market Research*, 46:19 – 34, 2013.

[BL06] Lorenzo Giovanni Bellù and Paolo Liberati. Inequality Analysis: The Gini Index. *FAO*, 2006.

[Bra08] Ulrik Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145, 2008.

[BSD+23] Subhayan Bhattacharya, Sankhamita Sinha, Paramita Dey, Anindita Saha, Chandrayee Chowdhury, and Sarbani Roy. Chapter 5 - Online social-network sensing models. In Dipankar Das, Anup Kumar Kolya, Abhishek Basu, and Soham Sarkar, editors, *Computational Intelligence Applications for Text and Sentiment Data Analysis*, Hybrid Computational Intelligence for Pattern Analysis and Understanding, pages 113–140. Academic Press, 2023.

[Bun] Bundeskanzleramt. Einkommen und der Gender Pay Gap. https://www.bundeskanzleramt.gv.at/agenda/frauen-und-gleichstellung/gleichstellung-am-arbeitsmarkt/einkommen-und-der-gender-pay-gap.html. Accessed: February 2025.

[Bun24] Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz. `https://www.sozialministerium.at/Themen/Soziales/Sozialversicherung/Pensionsversicherung/Pensionsarten/Alterspension.html`, 2024. Accessed: February 2025.

[BV14] Coralio Ballester and Marc Vorsatz. Random Walk-Based Segregation Measures. *Review of Economics and Statistics*, 96:383–401, 07 2014.

[Che03] Jing-Chao Chen. Dijkstra's Shortest Path Algorithm. *Journal of formalized mathematics*, 2003.

[Chr22] Aleksandr Christenko. Automation and occupational mobility: A task and knowledge-based approach. *Technology in Society*, 70:101976, 2022.

[Coh87] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences - Second Edition*. Lawrence Erlbaum Associates, 1988 `https://doi.org/10.4324/9780203771587`.

[CP20] Siwei Cheng and Barum Park. Flows and boundaries: A network approach to studying occupational mobility in the labor market. *American Journal of Sociology*, 126(3):577–631, November 2020. Publisher Copyright: © 2020 by The University of Chicago. All rights reserved.

[CV12] Lidia Ceriani and Paolo Verme. The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *The Journal of Economic Inequality*, 10:421–443, 2012.

[Die24] Claudia Diehl. Ethnic and racial inequality in the UK: a comment from a German perspective. *Oxford Open Economics*, 3(Supplement_1):i474–i479, 07 2024.

[dMB+21] R. Maria del Rio-Chanona, Penny Mealy, Mariano Beguerisse-Díaz, François Lafond, and J. Doyne Farmer. Occupational mobility and automation: a data-driven network model. *Journal of The Royal Society Interface*, 18:20200898, 2021.

[DMW21] Francesco D'Acunto, Ulrike Malmendier, and Michael Weber. Gender roles produce divergent economic expectations. *Proceedings of the National Academy of Sciences*, 118(21):e2008534118, 2021.

[Don25] Sarah A. Donovan. A Guide to Describing the Income Distribution. *Library of Congress*, 2025.

[dSG21] Maurits de Groot, Jelle Schutte, and David Graus. Job Posting-Enriched Knowledge Graph for Skills-based Matching. `https://arxiv.org/abs/2109.02554`, 2021.

[EBFM18] Rainer Eppel, Julia Bock-Schappelwein, Ulrike Famira-Mühlberger, and Helmut Mahringer. Der österreichische Arbeitsmarkt seit der Wirtschaftskrise. *WIFO-Monatsberichte*, 91(3):191–204, 2018.

[eur] eurostat. International Standard Classification of Education (isced). `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_Standard_Classification_of_Education_(ISCED)`.

[FRB23] Daniela P Fernández, Michelle Ryan, and Christopher T Begeny. Gender expectations, socioeconomic inequalities and definitions of career success: A qualitative study with university students. *PloS one*, 18(2), 2023.

[FVL19] Nina-Sophie Fritsch, Roland Verwiebe, and Bernd Liedl. Declining Gender Differences in Low-Wage Employment in Germany, Austria and Switzerland. *Comparative Sociology*, 18(4):449 – 488, 2019.

[GG15] Thomas Grandner and Dieter Gstach. Decomposing wage discrimination in Germany and Austria with counterfactual densities. *Empirica*, 42:49–76, 2015.

[GL16] Aditya Grover and Jure Leskovec. node2vec: Scalable Feature Learning for Networks. `https://arxiv.org/abs/1607.00653`, 2016.

[GLW+16] Maoguo Gong, Guanjun Li, Zhao Wang, Lijia Ma, and Dayong Tian. An efficient shortest path approach for social networks based on community structure. *CAAI Transactions on Intelligence Technology*, 1(1):114–123, 2016.

[GLZ24] Jiaao Guo, Qinghuai Liang, and Jiaqi Zhao. F-deepwalk: A Community Detection Model for Transport Networks. *Entropy*, 26(8), 2024.

[Gra73] Mark S. Granovetter. The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[GVRZ12] Kai Gutenschwager, Sven Völker, Axel Radtke, and Georg Zeller. The shortest path: Comparison of different approaches and implementations for the automatic routing of vehicles. In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, pages 1–12, 2012.

[Has23] Joe Hasell. Measuring inequality: what is the Gini coefficient? *Our World in Data*, 2023. https://ourworldindata.org/what-is-the-gini-coefficient.

[Her02] Holger Hermanns. *Markov Chains*, pages 35–55. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.

[Hip12] Ute Hippach-Schneider. Indikatoren und Benchmarks im gemeinsamen Arbeitsprogramm „Allgemeine und berufliche Bildung 2020“ der EU. In *Datenreport zum Berufsbildungsbericht 2016 - Informationen und Analysen zur Entwicklung der beruflichen Bildung*. Bundesinstitut für Berufsbildung, 2012.

[HN98] Jerry L. Hintze and Ray D. Nelson. Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, 52(2):181–184, 1998.

[Hol18] Hilary J. Holbrow. Detangling capital from context: a critical investigation of human capital explanations for immigrant wage inequality. *Journal of Ethnic and Migration Studies*, 46(19):4043–4065, 2018.

[HSS08] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[HSS11] Derek L. Hansen, Ben Shneiderman, and Marc A. Smith. Chapter 3 - Social Network Analysis: Measuring, Mapping, and Modeling Collections of Connections. In *Analyzing Social Media Networks with NodeXL*, pages 31–50. Morgan Kaufmann, Boston, 2011.

[HTWA17] Helmut Hofer, Gerlinde Titelbach, Rudolf Winter-Ebmer, and Alexander Ahammer. Wage Discrimination Against Immigrants in Austria? *LABOUR*, 31(2):105–126, 2017.

[HW02] Helmut Hofer and Andrea Weber. Wage mobility in Austria 1986–1996. *Labour Economics*, 9(4):563–577, 2002. European Association of Labour Economists, 13th Annual Conference, University of Jyvaskyla, 13-16 September 2001.

[HW23] Martin Halla and Andrea Weber. Inequality in Austria in recent decades. *Country Studies: Inequalities in Europe and North America. A parallel study to the IFS Deaton Review*, 2023.

[Int12] International Labour Office. International Standard Classification of Occupations: ISCO-08: Structure, group definitions and correspondence tables. *International Labour Organization*, 2012.

[JL22] Charlie Joyez and Catherine Laffineur. The occupation space: network structure, centrality and the potential of labor mobility in the French labor market. *Applied Network Science*, 7(16), 2022.

[JLLW22] Stefan Jestl, Michael Landesmann, Sebastian Leitner, and Barbara Wanek-Zajic. Trajectories of Employment Gaps of Refugees and Other Migrants: Evidence from Austria. *Population Research and Policy Review*, 41:609–669, 2022.

[Kai08] Marcus Kaiser. Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks. *New Journal of Physics*, 10(8):083042, August 2008.

[Kla24] Karin Klapfer. Arbeitsmarktstatistik 3. Quartal 2024 Mikrozensus-Arbeitskräfteerhebung. *Statistik Austria*, 2024.

[KM04] Gueorgui Kambourov and Iourii Manovskii. Occupational Mobility and Wage Inequality. *IZA Discussion Papers, No. 1189*, 2004.

[KM24] Karin Klapfer and Cornelia Moser. Arbeitsmarktstatistiken 2023 - Ergebnisse der Mikrozensus-Arbeitskräfteerhebung und der Offene-Stellen-Erhebung. *Statistik Austria*, 2024.

[KRŠ+17] Blaž KažIč, Jan Rupnik, Primož ŠKraba, Luka BradšKo, and Dunja Mladenič. Predicting Users' Mobility Using Monte Carlo Simulations. *IEEE Access*, 5:27400–27420, 2017.

[Lak13] Daniel Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 2013.

[LGA20] Eduardo López, Omar A. Guerrero, and Robert L. Axtell. A network theory of inter-firm labor flows. *EPJ Data Science*, 9(33), November 2020.

[LH22] Ken-Hou Lin and Koit Hung. The Network Structure of Occupations: Fragmentation, Differentiation, and Contagion. *AJS; American journal of sociology*, 127(5), 2022.

[LWP+23] Jm Liu, Tao Wang, Yu Pan, Feng Yao, Renjie He, and Zhongshan Zhang. Deep Occupational Representations Based on Career Mobility. In *2023 IEEE 3rd International Conference on Digital Twins and Parallel Intelligence (DTPI)*, pages 1–6, 11 2023.

[LWY+24] Jiamin Liu, Tao Wang, Feng Yao, Witold Pedrycz, Yanjie Song, and Renjie He. Assessing growth potential of careers with occupational mobility network and ensemble framework. *Engineering Applications of Artificial Intelligence*, 127:107306, 2024.

[LZM13] Eliza-Olivia Lungu, Ana-Maria Zamfir, and Cristina Mocanu. Occupational mobility networks of female and male higher education graduates. *Theoretical and Applied Economics*, 0(8(585)):37–46, August 2013.

[MdF18] Penny Mealy, R Maria del Rio-Chanona, and J. Doyne Farmer. What You Do at Work Matters: New Lenses on Labour. *SSRN*, March 2018.

[MGK15] Angelika Meraner, Daniela Gumprecht, and Alexander Kowarik. Die neue Hochrechnung des Mikrozensus - Methodenbeschreibung. *Statistik Austria*, 2015.

[MGK16] Angelika Meraner, Daniela Gumprecht, and Alexander Kowarik. Weighting Procedure of the Austrian Microcensus using Administrative Data. *Austrian Journal of Statistics*, 45(3):3—-14, Jun. 2016.

[MMJ13] Kairanbay Magzhan and Hajar Mat Jani. A Review and Evaluations of Shortest Path Algorithms. *International journal of scientific technology research*, 2013.

[NTG22] Shagofah Noor, Omid Tajik, and Jawad Golzar. Simple Random Sampling. *International Journal of Education Language Studies*, 1:78–82, 12 2022.

[NUL14] Waqas Nawaz, Kifayat Ullah Khan, and Young-Koo Lee. Shortest Path Analysis in Social Graphs. `https://arxiv.org/abs/1412.7760`, 2014.

[OAS10] Tore Opsahl, Filip Agneessens, and John Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, 2010.

[Off12] International Labour Office. International Standard Classification of Occupations - Structure, group definitions and correspondence tables. *ISCO-08 Volume 1*, 2012.

[ÖP21] Mahmut Özer and Matjaž Perc. Impact of social networks on the labor market inequalities and school-to-work transitions. *Yükseköğretim dergisi*, 11(Št. 1):38—-50, 2021.

[PAS14] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '14, pages 701—710. Association for Computing Machinery, August 2014.

[PLDP17] P. Parau, C. Lemnaru, M. Dinsoreanu, and R. Potolea. Chapter 10 - Opinion Leader Detection. In Federico Alberto Pozzi, Elisabetta Fersini, Enza Messina, and Bing Liu, editors, *Sentiment Analysis in Social Networks*, pages 157–170. Morgan Kaufmann, Boston, 2017.

[Pri18] Nicolas Privault. *Understanding Markov Chains - Examples and Applications*. Springer Singapore, 2018.

[PZ19] Victor M. Panaretos and Yoav Zemel. Statistical Aspects of Wasserstein Distances. *Annual Review of Statistics and Its Application*, 6:405–431, 2019.

[PZC+18] Sancheng Peng, Yongmei Zhou, Lihong Cao, Shui Yu, Jianwei Niu, and Weijia Jia. Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, 106:17–32, 2018.

[RB11] Sean F. Reardon and Kendra Bischoff. Income Inequality and Income Segregation. *American Journal of Sociology*, 116(4):1092–1153, 2011.

[Red23] Sidney Redner. A first look at first-passage processes. *Physica A: Statistical Mechanics and its Applications*, 631:128545, 2023. Lecture Notes of the 15th International Summer School of Fundamental Problems in Statistical Physics.

[RGC15] Aaditya Ramdas, Nicolas Garcia, and Marco Cuturi. On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests. `https://arxiv.org/abs/1509.02237`, 2015.

[RM15] Alireza Rezvanian and Mohammad Reza Meybodi. Sampling social networks using shortest paths. *Physica A: Statistical Mechanics and its Applications*, 424:254–268, 2015.

[RM21] A P Riascos and José L Mateos. Random walks on weighted networks: a survey of local and non-local dynamics. *Journal of Complex Networks*, 9(5):cnab032, 10 2021.

[SA19] Sushila Shelke and Vahida Attar. Source detection of rumor in social network – A review. *Online Social Networks and Media*, 9:30–42, 2019.

[Sch25] Julia Schuster. Nicht-binäre Geschlechter in der Bevölkerungsstatistik - Zählen nur Männer und Frauen? *STATjournal Nr. 1*, 2025.

[SF12] Gail M. Sullivan and Richard Feinn. Using Effect Size - or Why the P Value Is Not Enough. *Journal of graduate medical education 4(3)*, pages 279—-282, 2012.

[ŠP23] Boštjan Šumak and Maja Pušnik. *Analysis of the Shortest Path Method Application in Social Networks*, pages 169–182. IOS Press, 01 2023.

[Sta23] Statistics Lithuania. State Data Agency. Labour market in Lithuania. *Statistics Dissemination and Communication Division of the State Data Agency*, 2023.

[The25] The Austrian Social Science Data Archive. `http://www.aussda.at/`, 2025.

[UNE12] UNESCO Institute for Statistics. International Standard Classification of Education ISCED 2011. ISBN: 978-92-9189-123-8 https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf, 2012.

[UZP+17] Maximilian Unger, Stella Sophie Zilian, Wolfgang Polt, Wilfried Altzinger, Timon Scheuer, and Karim Bekhtiar. Technologischer Fortschritt und Ungleichheit: eine empirische Analyse der Entwicklung in österreich 2008-2014. *Wirtschaft und Gesellschaft*, 43(3):405 – 437, 2017.

[vU23] Theo van Uem. Application of Random Walk in Manpower Planning. `https://arxiv.org/abs/2109.11335`, 2023.

[Wir25] Wirtschaftskammer österreich. Kollektivvertrag: Basis-Infos und Grundlagen. `https://www.wko.at/entlohnung/kollektivvertrag-info`, 2025. Accessed: February 2025.

[ZSZZ11] Xiaohan Zhao, Alessandra Sala, Haitao Zheng, and Ben Y. Zhao. Efficient Shortest Paths on Massive Social Graphs (Invited Paper). IEEE, 2011.