Your very private job agency: Job referrals based on residential location networks

FIRST DRAFT

Franziska Hawranek[a,*], Norbert Schanne[b]

[a] *University of Regensburg, Universitätsstr. 31, 93053 Regensburg. Phone +49 941 943 5688*
[b] *Institute for Employment Research (IAB), Weddigenstr. 20-22, 90478 Nuremberg, Germany. Phone +49 911 179 5904, Fax +49 911 179 3297*

Abstract

This paper analyzes job referral effects that are based on residential location. We use geo-referenced record data for the entire working population (liable to social security) and the corresponding establishments in the Rhine-Ruhr metropolitan area, which is Germany's largest (and EU's second largest) metropolitan area. We estimate the propensity of two persons to work at the same place when residing in the same neighborhood (reported with an accuracy of 500m×500m grid cells), and compare the effect to people living in adjacent neighborhoods. We find a significant increase in the probability of working together when living in the same neighborhood, which is stable across various specifications. We differentiate these referral effects for socioeconomic groups and find especially strong effects for migrant groups from former guestworker countries and new EU countries. Further, we are able to investigate a number of issues in order to deepen the insight on actual job referrals: distinguishing between the effects on working in the same neighborhood and working in the same establishment – probably the more accurate measure for job referrals – shows that the latter yield overall smaller effects. Further, we find that clusters in employment although having a significant positive effect play only a minor role for the magnitude of the referral effect. When we exclude short distance commuters, we find the same probabilities of working together, which reinforces our interpretation of this probability as a network effect.

Keywords: Job referrals, Labor market, Neighborhood effects, Network effects, Social interactions

*Corresponding author
Email addresses: franziska.hawranek@wiwi.uni-regensburg.de (Franziska Hawranek), norbert.schanne@iab.de (Norbert Schanne)

1. Introduction

In social sciences the interest of interactions between individuals has increased: how do people influence one another and how can we measure this interaction? In labor economics, the importance of social interactions for the determination of labor market outcomes has drawn attention in the last years. One aspect of social interactions is interaction on a very local level: how does sharing a residential neighborhood (and therefore facing the same institutions and infrastructure) affect labor market outcomes? The channels hereby can be diverse including e.g. spatial mismatch (Kain, 1968), discrimination, differences in access to resources (such as education) or differences in attitudes and role models across neighborhoods (Jencks and Mayer, 1990). In this paper, we look at how residential neighborhoods can serve as a pool of information for an informal labor market and investigate the effect of job referrals through one's residential location.

In particular, we analyze the relationship between living and working together in the context of job referrals in the Rhine-Ruhr metropolitan area. The Rhine-Ruhr is Germany's largest and the EU's second largest agglomeration, which is located in North Rhine-Westphalia, spread across 7,110 km^2 including big cities like Cologne, Düsseldorf and Dortmund. The metropolitan area is home to over 11 million inhabitants and is especially interesting for urban analysis due to its densely populated nature and the economic diversity.[1]

Our empirical framework is possible due to a novel data set covering geo-coded record data for the entire working population (liable to social security) and the corresponding establishments. As social interaction is not measurable directly with any kind of administrative data, we use a convincing and well-established approach to approximate a local network effect: We estimate the propensity of two individuals to work at the same place when residing in the same neighborhood (reported with an accuracy of 500m×500m grid cells) with a linear probability model (LPM), and compare propensity effects of living in the same grid cell (an unconditional effect) with propensity effects conditional on a super-neighborhood fixed effect (where super-neighborhoods are all adjacent neighborhood grid cells). The empirical design follows Bayer et al. (2008), who found strong positive effects for job referrals using US American data.[2] We find very

[1] Traditionally, the Rhine-Ruhr was specialized in heavy industry and mining. The structural change lead to a specialization in the service sector and in education and development. Until today, the area is economically contrasting with high unemployment rates in Dortmund and Gelsenkirchen on the one hand and the prospering Rhine area on the other hand. See figure .3 in the Appendix.

[2] Bayer et al. (2008) use Census data for the Boston metropolitan area, which has 4.5 million inhabitants and is spread over 12,105 km^2, which means that the Rhine-Rhur area

similar effects: Bayer et al. (2008) estimate that sharing the same immediate neighborhood raises the propensity to work together by 0.12 percentage points, whereas the effect is 0.14 percentage points in our case. We rule out several other possible explanations for this propensity effect by conducting a number of robustness checks. The effects are robust throughout all specifications which makes us confident to interpret this effect as an indication for a job referral where information on an informal job market is circulated in one's residential neighborhood. To this point, we cannot say anything about who (within a pair) benefits from this local effect on one's information set but merely want to investigate the existence and credibility of a residential referral effect. Furthermore, we differentiate job referral effects by certain characteristics such as industry, nationality or age groups. The effects differ especially for pairs of different ethnicity: compared to Germans, the propensity to work together when sharing the same neighborhood is highly increased, in particular for immigrants from new EU countries but also from the former guest worker countries Spain and Italy. This is in line with previous empirical findings on the usage of informal channels for job search.

The goal of this paper is first to look at how referral effects based on residential location (or via weak ties) may differ for a European country as opposed to US American data, given that institutional backgrounds and cultural conventions are quite different with respect to the labor market and job search. In addition, we are able to investigate a number of issues in order to shed further light on actual job referral effects: First, our data allows us to distinguish between the effects on working in the same neighborhood and working in the same establishment - probably the more accurate measure for job referrals. The analysis shows that the effect is smaller throughout the specifications for referrals to firms. This indicates that referrals have been overstated when measuring only referrals to neighborhoods. Second, we analyze to what extent the findings are due to highly concentrated clusters of employment opportunities in central business districts. We investigate, whether we receive similar estimates regarding job referrals if we randomly reassign people to jobs while leaving the geography of workplaces unchanged. Finally, we address to what extent people tend to work in their residential neighborhood, and whether the evidence in the literature is affected by inadequately accounting for short-distance commuting behavior.

The remainder of the paper is structured as follows. Section 2 gives an overview on the related literature. Section 3 describes the data set we use for the German Rhine-Ruhr area. Section 4 presents the research design and the baseline

over all is more densely populated.

model. In section 5 we discuss our results as well as robustness checks and further specifications. Section 6 concludes.

2. Literature Review

Neighborhood effects describe interactions between people living in the surrounding area which influence the behavior or socioeconomic outcome of an individual (Dietz, 2002). As interactions themselves are rarely observable, the identification of such effects is difficult and there exists a broad variety in approaches and results to such neighborhood effects. One crucial problem of identifying causal neighborhood effects is the issue of self selection, as individuals usually choose residential location non-randomly according to their preferences, which are hard to measure with observational data. Apart from this, especially measuring peer effects bears another identification problem: Manski's reflection problem (Manski, 1993) formally states the general impossibility to distinguish in a linear model between peer effects generated as a result of belonging to a group (e.g. because of imitation, a so-called endogenous effect), and peer effects arising among people belonging to one group who take similar decisions because they face similar environmental conditions and institutions (contextual effect) and have similar characteristics, which leads them to take similar decisions (correlated effect). To overcome these two essential identification problems and approximate the unobservable interaction as good as possible, several strategies have been applied in the literature.

If the effect of peers on an individual's decision or action is in the focus of interest, Maximum Likelihood estimators with multivariate probability distributions or IV methods in which the endogenous peer/neighborhood effect is instrumented for may potentially provide one solution for identification (see e.g. Evans et al., 1992; Bramoullé et al., 2009). Motivated by Manski's critique, another strand of literature employs randomized control group experiments for investigating peer effects; the Gautreaux Program in Chicago in the 1970s and the Moving to Opportunity (MTO) in Baltimore, Boston, Chicago, Los Angeles and New York in the 1990s are prominent examples (see e.g. Katz et al., 2001; Ludwig et al., 2013). For the purpose of analyzing the economic outcomes though, the experimental design may not be applicable. First, it is doubtful whether such an experiment extends to neighborhood effects in general. Second, the external effect of a neighborhood may be undermined in the experiment, as also relocated individuals normally choose their own peers within a new neighborhood. As the difference between new and old neighborhood was intended to be big, artificially relocated individuals may have been isolated in their new environment. Another force influencing external effects in a neighborhood may thereby be ignored, namely the information flow within networks which can emerge within residential locations.[3]

Especially considering information on job opportunities, there is a broad literature covering referrals among potential employees[4]. Also within residential neighborhoods, these effects are studied for example by Zenou (2013). He argues that the disadvantage due to spatial separation between jobs and residential locations (spatial mismatch) can be amplified through the disproportionate usage of informal networks for job access. Weak ties[5] are important for job referrals as they bring new information to the network. Thus, people who live farther away from jobs also live farther away from potentially beneficial contacts, which prevents individuals from finding a job. Numerous other papers emphasize the importance of informal job markets like Ioannides and Datcher Loury (2004) and Corcoran et al. (1980) using US data. Glitz (2013) and Dustmann et al. (2011) investigate the effects of coworker networks on labor market outcomes using German record data. Glitz (2013) argues that weak ties are more important for finding a job, using former coworker networks to investigate the effect on own employment probability and a wage effect after a layoff. He finds strong positive effects, indicating significant effects of social networks in the German labor force on labor market outcomes.

A particular usage of networks in the labor market is when searching a job: Ioannides and Datcher Loury (2004) summarize stylized facts on the usage of informal job search channels and find that about 15% of unemployed Americans use friends and acquaintances for job search.[6] They further state that there is a variance in the usage of such information channels among age and socioeconomic groups: e.g. woman and individuals with better education use friends and family less often whereas the findings for older people are opposing.[7]

[3]A third strand in the literature concentrates on estimating the intensity of social interaction and disentangling the network effect inherent in social interaction from the contextual and correlated component; the feedback from social interaction towards an individual's decision is postponed to subsequent analysis.

[4]See e.g. Topa, 2001 who explains clustering of unemployment in Chicago using a probabilistic approach, Calvo-Armegnol and Jackson, 2007 who investigate how an agent's information network influences one's own employment probability and expected wages or Montgomery, 1991 who develop a model in which social networks are used as a signal for otherwise noisy or unknown productivity.

[5]As introduced by Granovetter (1973), weak ties represent acquaintances whereas strong ties reflect family and closer friends.

[6]Using data from the PSID from 1993, Ioannides and Datcher Loury (2004) find that 15.5 percent of unemployed and 8.5 percent of employed ask friends and relatives about potential job openings.

[7]On the one hand Ports (1993) find increased usage of informal channels 45-55 year-olds and 55-65 year-olds in 1992 respectively analyzing CPS data. On the other hand, e.g. Corcoran et al. (1980) report that usage of informal job market declines with age and/or work experience. Holzer (1987) finds that especially young people aged 16 to 23 rely on friends and relatives in 60- 70% of all jobs they actually attained (using data on search methods from the 1981

Kramarz and Nordström Skans (2013) analyze how strong ties, namely family, and weak ties like classmates and neighbors affect the decisions of youths in Sweden who enter the labor market. They analyze this question using a population wide data set linking graduation records and family ties to longitudinal matched employer-employee data with information on the firms. They find that the effect of strong ties is important, but only significant if one parent is currently employed at the same plant. The effect is stronger for low educated youths, those with bad grades or bad training and for immigrants. The authors compare the effect of strong ties to those of weak ties and find a positive and significant effect[8] independent of level of education.

Pellizziari (2010) analyzes wage premiums and penalties for finding a job through personal networks comparing these effects for countries in the European Union. He uses the European Community Household Panel (ECHP) from 1994-2001 and identifies the efficiency of informal job search channels based on cross-country variation in institutions and formal labor market policy interventions.[9] Negative effect on wages estimated with OLS become insignificant with fixed effects. Pooling over all EU 14 countries, the effect is significant, negative and small. A provided comparison to the USA (using the NLSY between 1979 to 2000) shows that about 30% of Germans used personal contacts for finding a job whereas only about 15% of US Americans used such search channels. This suggests that job referrals might play an even more important role in European countries as compared to in the US[10]. Nevertheless, the evidence for Americans suggests that referral effects may differ between categories, which is why we will differentiate between industries, age groups, nationality and education categories.

The work most related to ours work is Bayer et al. (2008), who also estimate the propensity of working together, when living in the same as opposed to a nearby neighborhood, assuming that there is no correlation in unobservables affecting both work location and the choice of residential location within a super-neighborhood. They use the 1990 U.S. Census of Population for the Bo-

National Longitudinal Survey of the Youth).

[8]The estimation strategy is somewhat different, as they only compare employment probability in a plant where neighboring parents (as compared to own parents, the strong ties) work. Hence it is not surprising that the magnitude of the effect is substantially smaller than the strong tie's effect.

[9]The ECHP only incorporates information on whether or not individuals found jobs through formal or informal search channels but has no information on how this channel is characterize. Consequently, it is not possible to investigate the nature and referral effect itself using this data set.

[10]The difference in data sources limits the exact comparability of these numbers: in the NLSY only one cohort is interviewed and there are 12 possible answers to the question "How did you find your current job?" from which multiple can be selected, whereas the ECHP is representative for the whole population and offers only 6 exclusive choices for the same question. See (Pellizziari, 2010).

ston metropolitan area and define census blocks as neighborhoods and census block groups as super-neighborhoods. We follow their empirical design which is described in detail in section 4. Bayer et al. (2008) find robust evidence for social interaction on a very local level: living together on the same census block increases the probability of working together by 33 percent[11]. We choose this paper as a point of departure, as the authors make a strong case for identifying social interaction in a very specific way, given the assumption of no correlation in unobservables within super-neighborhoods. This identifying assumption is crucial but testable. As Bayer et al. (2008) are able to identify a neighborhood effect given their empirical design, they create a measure for neighborhood quality: some neighborhoods have a better quality of "referral opportunities" as opposed to others, which can be seen as a proxy for neighborhood quality in general. These differences in quality translate to advantages or disadvantages in the labor market and can possibly also be used to measure differences in the context of intergenerational mobility.

The goal of this paper is first to look at how referral effects based on residential location (or via weak ties) may differ for a European country as opposed to US American data, given that institutional backgrounds and cultural conventions are quite different with respect to the labor market and job search. We extend the analysis by Bayer et al. (2008) in several ways. Most importantly, our data set allows us to not only observe the location of the workplace, but also the exact establishment. This gives a much more precise indication of an actual job referral, as we assume that individuals mainly have information about job opportunities at their own employer. Taking only location of workplaces may lead to bias in the referral effect including also the effect of concentrated employment as in CBDs.

Further, our data set is more detailed in several other ways: we observe the entire German working population (subject to social security) for a densely populated metropolitan area, the Rhine-Ruhr area. This is an advantage, as we can compute pairs of individuals with all of their neighbors and therefore calculate more accurate propensities of working together. Additionally we are interested in the effect of sorting into jobs, which we analyze using a simulation of random assignment to jobs, while leaving the geographic distribution of employment unchanged. From this exercise we can infer whether the propensity to work together really is driven by referrals or whether it is an effect of geographic concentration and maybe accessibility.

[11]They estimate various specifications and estimates of different size, but the 33% correspond to their most conservative specification.

3. Data

In this study we employ register data which are collected in the administrative processes of the German Federal Employment Agency (FEA, Bundesagentur für Arbeit) and maintained in the Integrated Employment Biographies (IEB) of the Institute of Employment Research (Institut für Arbeitsmarkt- und Berufsforschung, IAB). The IEB cover all employed persons who pay statutory social security contributions, all recipients of benefits from unemployment security (according to Social Code III) or from basic life support (according to Social Code II), all participants in active labor market policy, as well as all persons who approach FEA for job-search support.[12] Due to the parallel nature of the various data bases stemming from different processes, multiple spells may coexist for each person at the same time (e.g. because a person searches for a new job while being employed). If existing, the employment spell with the highest salary is defined as the *main spell*.

To ease computation, we use data only for one part of Germany. We select the Rhine-Ruhr region as it is Germany's largest metropolitan area. It is a very densely populated area reflecting several aspects that also represent the whole of Germany. The area is diverse in its wealth and socioeconomic structure. It includes on the one hand prospering university cities like Bonn and on the other hand former heavy industry and mining centers, which have a high population of immigrants and also a high proportion of unemployment like Gelsenkirchen. The IAB Research Data Centre geo-coded both the work-place and the residential address corresponding to each person's *main spell* at June 30th 2008 (see Scholz et al. (2012)). Each person is assigned to a quadratic grid cell of 500m length to warrant anonymity compulsory in social security data provision. The area covered by the squares corresponds more to census block groups rather than to census blocks. We use these grid cells as our basic definition of a neighborhood, supposedly adequate for an agglomeration like the Rhine-Ruhr metropolitan area, our region of interest.

Figure 1 shows the structure of the neighborhood definition: According to the exact address, every individual is assigned to a grid cell (the small squares correspond to 500m × 500m grid cells). Individuals A and B are immediate neighbors here, whereas C shares what we will further on call "super-neighborhood" with A and B. D lives within a super-neighborhood of C but not with A and B. In contrast to Bayer et al. (2008) who use predefined census

[12] The IEB enclose information on basic life support since 2005, on programme participation since 2000 and on job search since 2001.

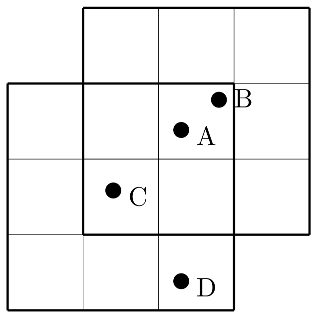

Figure 1: Defining neighborhood by a regular grid

blocks (neighborhoods) which belong to a fixed census block group (super-neighborhood), every grid cell (neighborhood) in our design is the centroid of a super-neighborhood and thus every grid cell belongs to several super-neighborhoods.
Within the geocoded IEB for Rhine-Ruhr, we observe roughly 4 million persons, dispersed across 21,509 grid cells, who are aged 15-65 and participate in the labor force (without self-employed, civil servants and members of the armed forces). Of these persons, roughly 3.5 million persons are employees. To get a file with individual data that is feasible for computation, we draw a 2 percent random sample from all employed persons and will further denote these individuals as i. They are combined with all persons residing within their own neighborhood or in one of the eight contiguous neighborhoods; we denote all possible neighbors as j and will further analyze pairs ij, who reside in the same super-neighborhood. Compared to working with (possibly larger) samples for both individuals and neighbors, the one-sided sampling has the advantage to enable conclusions on job referrals in the population more easily (with one-dimensional sampling probabilities, respectively univariate rather than bivariate cumulated densities). All in all, we observe approximately 3.4 million persons living in one of the super-neighborhoods. Figure 2 shows the distribution of neighborhood and super-neighborhood sizes. The mass of the neighborhood-size distribution lies in the range between 150 and 700 persons per grid cell; the average neighborhood size is around 320. However, the average pair is observed in a neighborhood with more than 900 inhabitants because larger neighborhoods have a higher probability to be represented in the sample, and a person in a large neighborhood has more neighbors.

The geographic scale in the IEB data set differs from that in the role model paper. While Bayer et al. (2008) use census blocks (which on average measure 160m of length) as a definition for neighborhoods, our neighborhoods are considerably larger measuring 500m×500m. Nevertheless, we believe that this

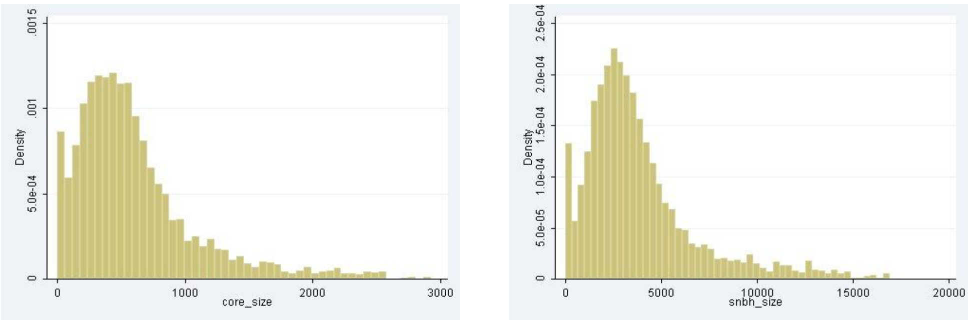

Figure 2: Size distribution of neighborhoods and super-neighborhoods

extent is small enough to guarantee the possibility of individuals actually interacting with each other. For example the length of a grid cell corresponds to the standard distance between bus stops for medium and highly populated urban areas (see Köhler and Bertocchi, 2010), which is what the whole Rhine-Ruhr area can be classified as. The grid cell size approximates a walking distance of five minutes, which in general should be small enough for people to actually meet. We define social interaction as the possible enhancement of individuals' information set on job opportunities through their residential environment. Such information transfer is said to happen between individuals' "weak ties", which we believe to coincide with one's residential neighborhood (see literature review for more details). The formation of weak ties within one's residential neighborhood can occur through meeting points such as sport clubs, churches or elementary schools[13], which are the places where one could potentially meet ones neighbors and interact with them.

Although the classification of neighborhoods and super-neighborhoods does not depend on geographic factors such as big roads or rivers, the flexible design guarantees an assignment for each grid cell to be the centroid of a super-neighborhood as well as part of the surrounding for all neighboring grid cells. We believe that this sampling scheme is an advantage as measured interaction is still very local but the conditioning surrounding is flexible. Furthermore, using a neighborhood definition that is based on real distances rather than the number of people sharing a neighborhood (as it is the case for census blocks and

[13]The whole Rhine-Ruhr area compasses 1,774 elementary schools, which differ in their dispersion: on the basis of municipalities (German "Kreise" and "kreisfreie Städte"), there is one elementary school per 2,522 inhabitants and a maximum 7,840 inhabitants per elementary school (data from the ministry of education in North Rhine-Westphalia at www.schulministerium.nrw.de). 2,522 inhabitants correspond to less than 1,100 employees when using the ratio of 35.9 Mio employees over 82 Mio inhabitants in Germany 2008 as an approximation. If we believe that e.g. parents meet when picking up their children and possibly form social contacts there, the extent of the draw area is larger than that of a residential neighborhood in our definition but smaller than a super-neighborhood.

census block groups) makes accounting for distances to workplaces and reflecting commuting behavior more realistic.
In table 1 we compare groups in the population, in the sample, and in the neighborhoods and super-neighborhoods of the sampled persons. As table 1 shows, the 2% sample is almost identical to the population. The groups considered here correspond to the covariates in our estimations. We differentiate between population groups, to see how network usage may change with groups. In line with previous literature, we expect especially ethnic and education groups to differ with respect to their usage of networks for informal job search. The countries and groups of countries are the largest immigrant groups and those who traditionally came to Germany as guest workers (southern European countries). Therefore we would expect those groups to have formed particularly strong networks within Germany.

Table 1: Group sizes in population and sample

Group	Population	Sample	Neighbors	Super-neighbors
Male	0.5181	.5168	.5182	.5180
Age 15-24	0.0985	.0991	.0993	.0993
Age 25-34	0.2000	.2001	.2053	.2023
Age 35-54	0.5392	.5405	.5417	.5439
Age 55-65	0.1531	.1515	.1536	.1545
Unskilled	0.1485	.1477	.2809	.2788
Med. skilled	0.4750	.4780	.1509	.1493
Highskilled	0.0963	.0932	.4722	.4752
German	0.9032	.9018	.8992	.9023
Greek	0.0051	.0052	.0053	.0051
Italian	0.0086	.0083	.0089	.0086
Spanish	0.0019	.0021	.0020	.0020
Turkish	0.0352	.0355	.0372	.0357
Yugoslavian[a]	0.0104	.0105	.0108	.0104
From new EU[a]	0.0068	.0070	.0067	.0068
Other nationality	0.0288	.0296	.0298	.0290
Primary sector	0.0395	.0405	.0388	.0394
Manufacturing	0.1763	.1779	.1754	.1764
Construction	0.0458	.0455	.0455	.0457
TTC[b]	0.2622	.2643	.2630	.2620
Business Services	0.1757	.1746	.1765	.1758
Other Services	0.3005	.2972	.3007	.3008
# employees	3,459,941	68,947	3,169,180	3,397,929

[a]: Yugoslavian covers immigrants from the territory of former Yugoslavia (including Slovenia and Croatia); these are not included in the group of immigrants from new EU members (which come from Estonia, Latvia, Lithuania, Poland, Czech Republic, Slovakia, Hungary, Bulgaria, Romania, Malta and Cyprus).
[b]: Trade, Transportation and Communication (TTC).

4. Empirical Design

4.1. Baseline specification

Our goal is to compare the propensity of individuals to work together for those living in the same neighborhood with individuals living close by. Our empirical design allows to identify a social interaction effect based on within super-neighborhood variation. The baseline model can be summarized as follows:

$$W^a_{ij} = \rho_s + \alpha_0 R^n_{ij} + \varepsilon_{ij} \text{ with } a = \{n, f\} \tag{1}$$

i and j denote individuals living in the same super-neighborhood (block of 9 grid cells) and W^a_{ij} is an indicator for both individuals sharing the same work place. W^a_{ij} takes on the values 0 or 100 so that parameters in the LPM directly represent changes in percentage points. We differentiate W^a_{ij} over $a = \{n, f\}$: first, we follow Bayer et al. (2008) and define the same work place as the neighborhood n where an individual works. Second, we use exact information on the establishments, where $W^f_{ij} = 100$ if a pair of individuals works at the exact same firm. All specifications are estimated with heteroscedasticity and cluster robust standard errors[14]. Therefore we can interpret α_0, the social interaction effect, as the increase in probability of working together when sharing a neighborhood. R^n_{ij} is equal to 1 if both i and j live in the same grid cell and zero otherwise. ρ_s denotes a fixed effect for the super-neighborhood. Including this fixed effect deals with sorting into residential location which leads to selection bias due to correlation in unobservable factors in neighborhoods (such as amenities or the access to public transportation), which is an important issue in the neighborhood effects literature (see section 2). If we assume that individuals freely choose their reference group in form of a super-neighborhood but only have a restricted choice within this super-neighborhood s, α_0 can be identified as the social interaction effect given that the two key assumptions are fulfilled: first, social interaction within a neighborhood is a local phenomenon. Second, individuals are able to choose their residential location freely but there is no correlation in unobservable characteristics affecting both work place and residential location between individuals living in the same neighborhood within a super-neighborhood.

To meet the requirement of the second key assumption, no correlation in un-

[14]Following Angrist and Pischke (2009), including robust standard errors deals with most of the problems when applying an LPM. Additional to the more straight forward interpretation of LPM estimating e.g. a Probit model would make computation more difficult given the extent of the data set.

observables within super-neighborhoods, Bayer et al. (2008) argue that on a very local level, the housing market is comparably thin. When individuals are choosing their residential location, it may be hard to observe variation within super-neighborhoods, whereas it is easier to see this variation between the larger super-neighborhoods. Furthermore, as with 500m length a grid cell is considerably small, such that it is not necessarily the case that one can find a suitable dwelling given an appropriate search period in an exact small neighborhood, but rather has to look for something in a more spacious area (such as the super-neighborhood). Germans in general are less mobile compared to US Americans: 16% of Germans have changed their residence within the last two years and only 9% moved within a city (Böltken et al., 2013)[15], which gives rise to the assumption that the thinness of the housing market is plausible even within cities.

To account for differences in the usage of informal networks between socioeconomic groups, we include individual characteristics in 1.

$$W_{ij}^a = \rho_s + \beta'(X_i - \bar{X}) + (\alpha_0 + \alpha_1'(X_i - \bar{X}))R_{ij}^n + \varepsilon_{ij} \text{ with } a = \{n, f\} \qquad (2)$$

Here we can investigate how belonging to a certain group adds to the propensity of working together. α_1 depicts the effect of being part of a particular group and working together - a "one-sided" social interaction effect. To interpret the effect of sharing a neighborhood at the mean of the categorical variables X, we center all covariates around zero.[16] We use categorical variables for personal characteristics such as sex, age groups[17], skill groups[18], categories of nationality, different industries, and a control for the size of the neighborhood. β can be interpreted as the baseline propensity of residing in the same super-neighborhood (belonging to the same reference group) but not sharing an immediate neighborhood

[15] Bayer et al. (2008) argue, that only 11 percent of the owner occupants in their census sample had changed owners. As the data we use is registry data, we cannot observe how people live and have to rely on additional data for motivational reasons. In Germany, the owner occupancy rate is considerably smaller - about 50% (Böltken et al., 2013) - as compared to the US where the rate is about 70% (Ihrke and Faber, 2012). Both in Germany and the US, owner-occupants are less mobile: in the German data, only 6.3% moved in the last two years and only 3.6% moved within a city. As moving rates for Germans are comparably smaller anyway, we believe that the argumentation of Bayer et al. (2008) holds for our data set, too.

[16] Wooldridge (2002) argues that subtracting the sample mean from each component allows identification of α_0 as the average treatment effect of R_{ij} on the dependent variable.

[17] Young adults from 15-24, career entrants aged 25-34, those established in the work force from 35-54 and senior workers between 55 and 65.

[18] Low skilled refers to lower secondary education with and without apprenticeship. Medium skilled individuals have higher secondary education (German "Abitur"), with and without apprenticeship. The high skilled group refers to individuals with a university degree.

on working together for different characteristic groups (X_i).

$$W_{ij}^a = \rho_s + \beta'(X_{ij} - \bar{X}) + (\alpha_0 + \alpha_1'(X_{ij} - \bar{X}))R_{ij}^n + \varepsilon_{ij} \text{ with } a = \{n, f\} \quad (3)$$

In equation 3, we examine whether the propensity to work together varies with the characteristics of the matched pair (as opposed to the individual characteristic measured by equation 2). Including this specification aims to investigate whether e.g. more similar pairs are more likely to profit from social interaction and whether certain groups have higher probabilities to work together, because of a stronger attachment to the labor market. Both equation 2 and 3 can be used to validate our estimates with evidence from the informal job market and network literature presented in section 2.

4.2. Robustness

The baseline model presented above has two major issues for identifying a causal social interaction effect: self selection and a potential simultaneity bias. Self selection arises when individuals sort themselves into residential location, such that sharing a neighborhood (R_{ij}^n) is not randomly assigned. A simultaneity bias could arise if we cannot rule out a referral effect on the housing market or in other words that people might actually live together because they work together, not the other way around. In the following, we discuss strategies to reduce these problems.

4.2.1. Sorting within super-neighborhoods

To deal with self selection into residential location, we first include the super-neighborhood fixed effect ρ_s in all estimation equations. Fixed effects deal with selection at least to some extent: on the basis of super-neighborhoods, all observable and unobservable factors influencing both work place and residential location are held constant. What remains a concern then is the sorting within super-neighborhoods. Therefore, we want to make sure the key assumption for identification, that there is no correlation in unobservables affecting work location within a super-neighborhood, can be regarded as reasonable.

First we analyze the sorting behavior with respect to observable characteristics. Following Altonji et al. (2005), the selectivity in observables is proportional to selectivity in unobservables and can therefore be seen as an indication of sorting on the basis of unobservable characteristics. We compute correlations of observable characteristics (age groups, gender, nationality groups, skill groups and industry groups) for both pairs that reside in a neighborhood together and for pairs who share a super-neighborhood but are not immediate neighbors. We test whether these correlations differ significantly between the two groups. If

correlations for pairs sharing an immediate neighborhood are significantly higher than those for pairs living in the same super-neighborhood, we interpret this as a sign for sorting with respect to observables within super-neighborhoods.
Second we test whether there is sorting within super-neighborhoods with respect to unobservables. To test this, we analyze the residuals from estimating equation 2, which represent everything which is unobservable with respect to the choice of residential and working location and therefore proxy sorting on the basis of unobservables. By construction, the residuals should have an average value of zero on the basis of super-neighborhoods. Comparing the mean residuals for those pairs sharing a neighborhood (i.e. $R_{ij} = 1$) with those sharing a super-neighborhood gives a direct test for sorting on the basis of unobservables: if the mean of residuals for pairs sharing neighborhood is significantly different from zero, we can expect there to be sorting within super-neighborhoods on the basis of unobservables.

4.2.2. Reverse Causality

Another important issue is to eliminate the possibility of reverse causality, meaning that the estimated effects are actually no job referrals but individuals receive referrals at the workplace for a place of residence. To check which direction of the effect is the most plausible, we select four different subsamples and estimate the baseline specification of equation 1. As in Bayer et al. (2008), we first select individuals who have a stable residence: as the IEB data set is geo-referenced only for the cross section of 2008, we have to rely on residential location in form of zip codes two years prior to our main sample, in 2006. Zip codes refer to districts within cities or municipalities; hence the residential areas are larger than that of our main specification but still represent movements within cities.
First we check how the propensity to work together is changed when regarding only pairs where both individuals have lived in the same zip code area in the last two years and refer to them as "residential stayers". If the propensity is significantly smaller than that of the baseline estimation with the whole sample, this would be a sign that referrals actually take place on the housing market. Second we use a subset of "job movers": we select only pairs of which one individual has changed the workplace (workplace here is defined as the zip code where an individual works). This specification includes individuals who move to find a new job, but it should give us a more precise feeling for the magnitude of the third effect: here, we select a subsample of individuals, who have all lived in the same zip code in the last two years and use only pairs where one individual has changed the working location, i.e. "residential stayers with a job move". Whenever one individual has changed working location and both individuals

have stayed at their residence, it is more likely that the effect we observe is induced by an actual job referral. Fourth, we select a subsample where it is most likely to observe a referral on the housing market: we use pairs where one individual has lived in the zip code area for the last two years whereas the other has changed zip code area but both individuals have worked in the same zip code area during that period, i.e. there is one change in residential location but no change in employment for both. This is a circumstance where it is most likely that the estimated social interaction effect is actually induced by co-workers exchanging information on the housing market.

4.2.3. Random Reassignment to Jobs

Is it possible that the correlation we observe is induced by something other than referrals by neighbors? Workplaces are neither evenly nor randomly allocated over space. They follow a certain structure because firms settle up more frequently in the central business district, subcentral business districts, or particular business zones (see e.g. Fujita et al. (1999) for an overview). As a consequence, a certain correlation with regard to workplaces may arise because people optimize their commuting distance. In order to disentangle this spurious correlation from the correlation due to job referrals, we randomly reassign a workplace neighborhood to the persons i according to the workplace probabilities in their super-neighborhood. To do so, we determine for each super-neighborhood s the specific relative frequencies (i.e. the probabilities) for each workplace neighborhood, $p_{n|s}$, with cumulated frequencies $F_{n|s} = \int \bigcup_{m \in [1,\ldots,n]} p_{m|s}$; the frequencies add up to the unit interval as $\int \bigcup_{n \in [1,\ldots,N]} p_{n|s} = 1$). Then we draw for each person i from a uniform distribution. The realization of this draw corresponds to a unique workplace n-specific partition on the unit interval (as $\{u_i \in (F_{n-1|s}, F_{n|s}]\} \mapsto n$) which determines for each person i a counterfactual workplace. Then we can construct a new variable for the hypothetical workplace coincidence, $\tilde{W}_{ij}^n$, and reestimate equation 1:

$$\tilde{W}_{ij}^n = \rho_s + \alpha_0 R_{ij} + \varepsilon_{ij} \tag{4}$$

This allows us to test whether α_0 from equation 1 differs from that in equation 4. Our approach then could show that the effect we estimate as a referral effect is actually driven by clusters in employment.

4.2.4. Short Distance Commuting

To get further insight on the nature of the measured referral effects, we want to explicitly address the effect of commuting behavior. We suspect, that a reasonable number of people works close to where they live and therefore commutes only very short distances. We analyze whether the increases in propensity to work together are driven by a disproportionately high number of short distance commuters and first analyze the commuting behavior descriptively. Then we exclude all individuals who work in the neighborhood of their residence and reestimate equation 1 with this restricted sample and test whether the coefficient of social interaction α_0 differs from that in the full sample.

5. Results

5.1. Baseline model on Job Referrals

Table 2 summarizes the results from our baseline model as presented in section 4.1. Estimating an unconditional (without super-neighborhood fixed effects) gives some first impressions on the baseline probability of working together[19]: when residing in the same super-neighborhood the probability of working in the same neighborhood is 1.8% and 0.22% for working in the same firm. Estimating equations 1-3 can then be interpreted as an increase in this baseline probability by residing in the same neighborhood.

We differentiate all specifications between two types of referrals: one where the referral goes to a neighborhood ($a = n$) and one where we interpret the increase of probability of working at the same firm ($a = f$) when sharing a neighborhood. Across all specifications, the magnitude of the effect is about 0.06 percentage points smaller for referrals to firms.

Column (1) corresponds to equation 1, where sharing a neighborhood is the single explanatory variable. The social interaction effect is positive and highly significant for both specification cases of $a = (n, f)$, which means that we find evidence for a significant positive impact of sharing a residential neighborhood on the propensity to work together. For a referral to a neighborhood ($a = n$), the propensity of working together is increased by 0.14 percentage points, which corresponds to an increase by 8 percent. Despite the different definition of neighborhoods the magnitude of the social interaction effect is similar compared to the effect estimated by Bayer et al. (2008), who find it to be 0.12.

Interpreting the referral to a firm can be seen as an even higher indication for an actual job referral: in general, we assume that individuals rather have information on available jobs at their own firm, not of establishments in the same neighborhood as their firm[20]. The estimated social interaction effect is somewhat smaller as compared to the referral to neighborhood effect, albeit

[19] Here, we estimate $W_{ij}^a = \alpha_0 + \alpha_1 R_{ij}^n + \varepsilon_{ij}$ and interpret α_0 as the baseline probability of working together when sharing the same super-neighborhood.

[20] Although there could be scenarios, where e.g. people commute to work together with other people working in the same neighborhood, but not at the exact firm and where people hear about potential job openings in public transportation.

Table 2: Estimation of Referral Effects

Variable	Referral to neighborhood ($a = n$)				Referral to firm ($a = f$)			
	No FE	(1)	(2)	(3)	No FE	(1)	(2)	(3)
Constant	1.7982***	1.7941***	2.2207***	2.0837***	.2169***	.2189***	.2347***	-.1196
	(.0011)	(.0037)	(.4000)	(.3811)	(.0004)	(.0038)	(.1063)	(.1054)
R_{ij}	.1111***	.1368***	.1432***	.1238***	.0868***	.0746***	.0784***	.0605***
	(.0027)	(.0238)	(.0238)	(.0185)	(.0010)	(.0240)	(.0241)	(.0182)
Sex	–	–	${X_i}^{***}$	${X_{ij}}^{***}$	–	–	${X_i}^{***}$	${X_{ij}}^{***}$
			$R_{ij}X_i$	$R_{ij}{X_{ij}}^{***}$			$R_{ij}X_i$	$R_{ij}{X_{ij}}^{***}$
Qualification	–	–	${X_i}^{***}$	${X_{ij}}^{***}$	–	–	${X_i}^{***}$	${X_{ij}}^{***}$
			$R_{ij}X_i$	$R_{ij}{X_{ij}}^{***}$			$R_{ij}X_i$	$R_{ij}{X_{ij}}^{***}$
Age group	–	–	${X_i}^{***}$	${X_{ij}}^{***}$	–	–	${X_i}^{***}$	${X_{ij}}^{***}$
			$R_{ij}X_i$	$R_{ij}X_{ij}$			$R_{ij}X_i$	$R_{ij}{X_{ij}}^{***}$
Ethnicity	–	–	${X_i}^{***}$	${X_{ij}}^{***}$	–	–	${X_i}^{***}$	${X_{ij}}^{***}$
			$R_{ij}{X_i}^{**}$	$R_{ij}{X_{ij}}^{***}$			$R_{ij}{X_i}^{**}$	$R_{ij}{X_{ij}}^{***}$
Industry	–	–	${X_i}^{***}$	${X_{ij}}^{***}$	–	–	${X_i}^{***}$	${X_{ij}}^{***}$
			$R_{ij}{X_i}^{**}$	$R_{ij}{X_{ij}}^{***}$			$R_{ij}{X_i}^{***}$	$R_{ij}{X_{ij}}^{***}$
lncoresize	–	–	X_i	X_{ij}	–	–	X_i	X_{ij}
			$R_{ij}{X_i}^{***}$	$R_{ij}{X_{ij}}^{***}$			$R_{ij}{X_i}^{***}$	$R_{ij}{X_{ij}}^{***}$
σ_u	–	5.2934	36.6795	31.1610	–	2.5382	5.1042	4.6731
σ_ε	13.351	13.2881	13.2821	13.2703	4.7956	4.7654	4.7623	4.7485
# pairs	179.7 Mio	179.7 Mio	179.7 Mio	179.7 Mio	179.7 Mio	179.7 Mio	179.7 Mio	179.7 Mio
# groups	–	11757	10159	10159	–	11757	10159	10159
Corr(u,Xb)	–	-.0072	-.0042	-.9996	–	.0083	.0249	-.9984

Heteroscedasticity-consistent standard errors in parentheses.

*/**/***mark significance at the 90%/95%/99% confidence level.

Asterisks at the control variables and interaction terms mark significant differences between the groups. Reference is a female German in the age between 35 and 54 with unknown qualification working in Manufacturing, or a pair of two women with these characteristics, respectively. Coefficient estimates are reported in the appendix.

still positive and highly significant. We estimate, that the propensity to work together at the same firm increases by 0.07 percentage points if a pair of individuals lives in the same neighborhood. This is equivalent to a 30% increase in probability compared to the unconditional baseline probability.[21]

Columns (2) and (3) refer to equation 2 and equation 3, where we are interested in how the social interaction effect reacts first for different socioeconomic groups and second for pairs of socioeconomic groups. For expositional purpose, we only report joint significance in this table; full outputs are presented in the appendix. A noticeable result is that the social interaction effect of living together in a neighborhood (R_{ij}) is relatively stable across specifications. Column (2) shows the one-sided interaction effect. Here, only some of the interactions are jointly significant: there is no statistically significant effect of sharing a neighborhood and qualification, age group or gender both for referrals to neighborhoods and to firms. An individual's own ethnicity[22] and in which industry[23] one works has overall a significant effect on the propensity to work together. The larger a neighborhood in which i lives, the smaller the propensity to work together; this probably corresponds to the likelihood of interaction the more individuals reside in a neighborhood.

Column (3) describes how pairs of certain groups interact in residential neighborhoods. The effect of X_{ij} describes the "baseline" propensity to work together when sharing a super-neighborhood: as expected, we see higher propensities for young and old pairs of workers, as well as unskilled pairs and matches for several industry sectors, but almost no effect of ethnic groups. Again the interaction term determines the local referral effect. Apart from age groups, the impact of all categories are jointly significant which indicates that matching pairs with respect to socioeconomic categories at least play some role for job referrals. The interaction effects (α_1 in equation 3) can be interpreted as the additional effect of being both in the same socioeconomic group and sharing the same neighborhood. There are no big differences across gender and age groups (meaning that the interaction effects are either small or insignificant). Consistent with the literature on informal job markets, pairs of unskilled workers have a comparatively higher propensity to work together both at the same neighborhood

[21] A 30% increase in probability is what Bayer et al. (2008) find, too. For working in the same neighborhood, the effect rises only by 8%, which is probably a consequence of the larger reference group definition.

[22] The effect differs between referrals to neighborhoods, where Greeks have the most significant increase in probability of working together, and referrals to neighborhoods, where Turks seem to profit the most from referral effects. For all other groups, the effects are positive but rather noisy.

[23] Compared to women working in manufacturing, working in all other industry sectors has a negative effect on working together when sharing a neighborhood, with business related services having the largest and most significant effect.

and at the same firm. For different ethnic groups the effect varies, too: especially for people from the new EU countries, the propensity to work together increases by over 20% as compared to Germans (the reference group) both for referrals to neighborhoods and referrals to firms. Also Italians and people from former Yugoslavia have a higher probability to work together when sharing a neighborhood. In contrast, albeit being the biggest migrant group in Germany, Turkish do not seem to behave differently than Germans, with the interaction effect being insignificant. For the different types of industries[24], the propensity to work together is increased in a similar way across groups. The size of the residential neighborhood of pairs seems to have no effect on working together; it has a significant negative effect on the interaction (the referral), however. This is in line with the decreasing probability to meet when living in a higher populated neighborhood or the more extensive usage of residential networks in more sparsely populated areas.

5.2. Robustness

To check whether the estimated effects are stable, we apply several robustness checks as described in 4.2.

5.2.1. Sorting within Super-neighborhoods

Table 3 presents correlations on the basis of observables. We compute correlations as $E(D_i \frac{1}{n_i} \sum_j D_j) = E(X_{ij})$, which is the expected value of observing two individuals i and j belonging to the same group D[25]. For the purpose of sorting, we look at the difference in conditional probability between neighborhood and super-neighborhood: we see that there are no big differences with the super-neighborhood having slightly less correlations. This indicates, that there is sorting on the basis of observables but that there is no difference in the patterns of sorting between neighborhoods and super-neighborhoods. Apart from that, especially Turkish, people from former Yugoslavia and the new EU countries sort themselves together into neighborhoods. In contrast, immigrants from other southern European countries tend to sort away from each other. This is remarkable when thinking about the interpretation of the interaction effects presented above: Turkish, who seem to sort themselves together do not tend to be more likely to work together. In contrast, Italians and Spanish who have an increased probability to work together tend to sort away from each

[24]An exception is the Primary Sector. Here the increase in propensity to work together can probably be accounted for – at least to some extent – by disproportionally many people living very close to their workplace.

[25]Therefore some of the correlations are very high just because the group is comparatively big, which is why the probability to be matched into a pair with your own group is high.

Table 3: Correlation between individual and average characteristics across neighbors

	Within neighborhood		Super-neighborhood, without neighborhood	
	unconditional	cond. on neighborhood	unconditional	cond. on neighborhood
female	0.4919	0.4920	0.4954	0.4955
male	0.5062	0.5060	0.5050	0.5048
Age 15-24	0.1153	0.1235	0.1101	0.1145
Age 25-34	0.2038	0.2071	0.1949	0.1953
Age 35-54	0.5288	0.5254	0.5266	0.5255
Age 55-65	0.1739	0.1764	0.1720	0.1730
Unskilled	0.1267	0.1340	0.1226	0.1244
Med. skilled	0.3958	0.3913	0.4024	0.4012
Highskilled	0.1182	0.1074	0.1104	0.1014
Unknown skill	0.2611	0.2684	0.2576	0.2618
German	0.8766	0.8731	0.8830	0.8817
Greek	0.0042	-0.0210	0.0062	-0.0229
Italian	0.0082	0.0188	0.0053	-0.0068
Spanish	0.0041	-0.0577	0.0024	-0.0373
Turkish	0.0623	0.1157	0.0481	0.0817
Yugoslavian[a]	0.0254	0.0626	0.0161	0.0257
From new EU[a]	0.0159	0.0787	0.0089	0.0336
Other nationality	0.0687	0.0933	0.0538	0.0706
Primary sector	0.0208	0.0177	0.0241	0.0200
Manufacturing	0.1372	0.1349	0.1352	0.1337
Construction	0.0288	0.0283	0.0288	0.0252
TTC[b]	0.2595	0.2612	0.2604	0.2619
Business Services	0.1672	0.1715	0.1687	0.1706
Other Services	0.2326	0.2287	0.2365	0.2346

[a]: Yugoslavian covers immigrants from the territory of former Yugoslavia (including Slovenia and Croatia); these are not included in the group of immigrants from new EU members (which come from Estonia, Latvia, Lithuania, Poland, Czech Republic, Slovakia, Hungary, Bulgaria, Romania, Malta and Cyprus).
[b]: Trade, Transportation and Communication (TTC).

other[26]. This indicates that although we clearly see some sorting in observables, it does not seem to bias our interaction estimates systematically. In addition, we check whether there is sorting on the basis of unobservables within super-neighborhoods. We use the residuals $\hat{\varepsilon}_{ij}$ from equation 1 and test whether these have a mean of zero on the basis of neighborhoods. For the estimation of a referral to a neighborhood, the mean is .13 and the median -.08, for referrals to a firm the mean residual is .07 and the median -.06. In both cases, the null

[26]The very high positive effect for new EU migrants, however, seems to be inflated by positive sorting bias. Nevertheless, as it is big and statistically highly significant, we believe that there should still be some effect generate by referrals.

hypothesis of the mean being zero can not be rejected[27]. Therefore, we conclude that there is no sorting on the basis of unobservables affecting both workplace and residential location within super-neighborhoods. This means that our empirical design deals successfully with self selection of residential location, the most important issue in identification of neighborhood effects.

5.2.2. Reverse Causation

Table 4 summarizes the results for the strategy presented in 4.2.2. The social interaction effect R_{ij} for "residential stayers", those individuals who have been living in the same zip code area in the last two years. The effect for a referral to workplace ($a = n$) rises slightly to 0.1552 percentage points and .0828 for referrals to a firm. Here also the constant rises which is associated with an overall increase in probability, because restricting the sample to only residential stayers mainly excludes pairs not working together. This is an indication that the estimated effect is not driven by referrals for housing at the workplace, as we would expect the magnitude to be smaller for residential stayers.

Second we look at pairs of which one individual has changed job location (on the basis of zip codes). Both for referrals to a neighborhood and referrals to a firm the effect is very close to the one estimated with the whole sample.

For the subsample of pairs of whom both have lived in the same neighborhood in the last two years and one has changed the job in the last two years, the effect decreases slightly for both kinds of referrals but remains statistically significant. This is the group where job referrals are most likely, as one of the pair is supposed to have been seeking a job in the previous two years. Nevertheless, the sample differs from the whole sample, which is why we should not suspect the effect to be as big as that for the whole sample: this is in line with Bayer et al. (2008), who find a social interaction effect of 0.09 percentage points for job movers. The difference between the estimated referral effects in the baseline specification and in this restricted sample are not statistically different from each other (for both cases of $a = \{n, f\}$), which makes us confident that even when restricting the sample to this specific case we still find the same social interaction effect.

Finally, we look at those pairs, where it is most likely to observe a referral effect on the housing market: we select a sample of pairs of which both have worked in the same zip code area two years ago and of whom one has changed residential location (as based on the zip codes). In this case, the referral effect is increased and highly significant both for $a = n$ and $a = f$. This means, that we cannot rule

[27]The z-statistic yields .0596 for $a = n$ and .0506 for $a = f$, which is substantially smaller than any critical value to reject the null.

Table 4: Referral Effects amongst job movers and residential stayers

Variable	Res. stayers		Job movers		Res. stayers with job move		Housing referral	
	$(a=n)$	$(a=f)$	$(a=n)$	$(a=f)$	$(a=n)$	$(a=f)$	$(a=n)$	$(a=f)$
Constant	2.095***	.2756***	1.8667***	.1593***	1.8066***	.1513***	2.3407***	.3278***
	(.0036)	(.0035)	(.0054)	(.0055)	(.0051)	(.0051)	(.0057)	(.0047)
R_{ij}	.1552***	.0828***	.1351***	.0796***	.1183***	.0739**	.1861***	.0980***
	(.0228)	(.0223)	(.0347)	(.0355)	(.0326)	(.0329)	(.0368)	(.0305)
σ_u	5.3345	2.5287	3.5007	.6615	3.2159	.6691	3.2471	1.4941
σ_ε	14.3187	5.3321	13.5379	4.0968	13.3088	3.9935	15.0251	5.7943
# pairs	102.9 Mio	102.9 Mio	105.3 Mio	105.3 Mio	62.1 Mio	62.1 Mio	9.6 Mio	9.6 Mio
# groups	10662	10662	10792	10792	10165	10165	9262	9262
Corr(u,Xb)	-.0079	.0073	-.0059	.0085	-.0070	.0063	-.0045	.0057

Heteroscedasticity-consistent standard errors in parentheses.

*/**/***mark significance at the 90%/95%/99% confidence level.

Residential stayers: Pairs who both live in the same zip code area and have lived there for at least two years. Job movers: Pairs of which at least one has moved her workplace across zip code districts within the previous two years. Housing referral: Pairs both working in same zip code area for two years, one of them has changed residential location.

out a reverse causality. Nevertheless the sample size is considerably smaller than in all other cases and the sample seems to be inherently different from those before: regarding the magnitude of the constant suggests, that by selecting this specific subsample, we exclude primarily individuals not working together (i.e. zeros for W_{ij}), which could be a reason why the estimated interaction effect is bigger than in the estimation with the whole sample. For referrals to a neighborhood, the constant (which can be interpreted as the baseline probability of working together when sharing a super-neighborhood) increases by 15% compared to the baseline estimation with the whole sample, for referrals to a firm it is even increased by 50%. Apart from this, the people in this subsample should differ from those in the whole sample, as we explicitly select individuals with a stable employment. This gives rise to believe that interpreting these numbers can actually not tell us a lot about reverse causation. In other words, up to this point we cannot reject the hypothesis that what we measure in equation 1 as a job referral effect is biased by referrals on the housing market. Still when modelling an environment, where a job referral is most likely (residential stayers with a job move), the magnitude and significance of the social interaction effect are very stable, which means that we have also evidence for a referral effect where the job referral is most likely.

5.2.3. Random Reassignment to Jobs

As described in 4.2.3, we are concerned that a substantial part of the measured referral effect is actually driven by clusters in employment. When individuals minimize their commuting time, living and working together could only be a side effect and the interpretation of the network effect misleading due to this spurious correlation. We compute $\tilde{W}_{ij}^{n}$ (in equation 4) as an artificial workplace (in the sense of neighborhood of employment) and estimate the correlation of living in the same neighborhood as a result of clusters in employment. Table 5 shows, that the spurious correlation is positive and statistically highly significant. Nevertheless, the magnitude of the effect is small compared to what we estimate to be the overall referral effect. When subtracting this spurious correlation of 0.03 percentage points, we observe a referral effect (to a neighborhood) that is comparable in extent to what we find for referrals to a firm. All in all, this indicates that what we measure as a referral effect using the design of Bayer et al. (2008), is probably a little bit too high but the effects seem to be robust to additional specifications and checks.

5.2.4. Short Distance Commuting

Apart from the impact of jobs being clustered in central business districts, another driver of the measured referral effect could be individuals working at the

Table 5: Baseline estimation for artificial workplaces

Variable	$\tilde{W}_{ij}^{n}$ for $a = n$
Constant	1.8195***
	(.0016)
R_{ij}	.0278***
	(.0104)
σ_u	2.1095
σ_ε	13.2895
# pairs	155.7 Mio
# groups	11376

Heteroscedasticity-consistent standard errors in parentheses. */**/***mark significance at the 90%/95%/99% confidence level.

Table 6: Baseline estimation excluding short distance commuters

Variable	$a = n$	$a = f$
Constant	1.9528***	.2390***
	(.0040)	(.0041)
R_{ij}	.1298***	.0787***
	(.0269)	(.0262)
σ_u	4.9131	2.2322
σ_ε	13.8424	4.9754
# pairs	154.2 Mio	154.2 Mio
# groups	11325	11325

Heteroscedasticity-consistent standard errors in parentheses. */**/***mark significance at the 90%/95%/99% confidence level.

same location where they live. One way to investigate the impact of those short distance commuters is to exclude all individuals working at the same zip code as they live in. Table 6 summarizes the results. The results correspond to estimating the baseline estimation of equation 1 both for referrals to a neighborhood and referrals to a firm. The sample size is restricted to 154.2 million pairs ij, which means that excluding all individuals who live where they work does not restrict the data set fundamentally, but there seems to be only a minority of individuals working at their residential location. Furthermore, both the constant and the social interaction effect remain at a comparable level. The baseline probability of working together when sharing the same super-neighborhood (constant) is slightly higher both for referrals to a neighborhood and referrals to a firm. The social interaction effect in contrast is slightly lower for a referral to a neighborhood, but still in a very similar range with .13 versus .14 with the whole sample. For referrals to a firm, the effect is even a bit higher as compared to the estimation with the whole sample (.0787 versus .0746). Over all, the results stay very much the same when excluding short distance commuters, which suggests that disproportionately many people working where they live are not the main drivers in the referral effect and we do not observe a spurious correlation here.

6. Conclusion

Most of the empirical work on the economic effects of neighborhoods so far has been on US American data; in contrast, we look at labor market effects for the Rhine-Ruhr area, one of the biggest agglomerations in Europe. We use the research design proposed by Bayer et al. (2008) to compare propensities to work together when sharing an immediate neighborhood while holding the surrounding neighboring area constant. This design allows us to identify a social interaction effect using the within variation of the so-called super-neighborhoods.
The results of our baseline specification are very similar to those for the Boston metropolitan area: we estimate a significant increase of 0.14 percentage points in the probability to work together when sharing a neighborhood while Bayer et al. (2008) find the increase to be 0.12 percentage points. So the first question whether the extent of referral effects based on residential location differs for a European country as compared to the US can be denied: although we use a different definition of neighborhood and super-neighborhood, we find very similar results. As our neighborhoods encompass a greater geographic entity, we would expect the magnitude of the referral effect to be smaller, as with more people in a neighborhood the probability to meet decreases. As our estimates are slightly higher, we can reject the hypothesis that Germans use weak ties for job information less intensively.
The novel geo-coded data set we use allows us further to differentiate two kind of referral effects: as in Bayer et al. (2008), we also estimate a "referral to neighborhood" effect; the increased propensity to work in the same neighborhood when living in the same residential neighborhood. Additionally, we estimate a "referral to firm" effect; this effect is about 0.06 percentage points below that for referrals to a neighborhood and stable across specifications. We interpret this second effect as the more precise measure for job referrals, as information on available jobs should be restricted mostly to one's own firm. Hence, we argue that the previously estimated effect overstates actual network effects.
Our estimates for referral effects are stable across several specifications: we first analyze whether different types of socioeconomic groups have different probabilities to work together when being neighbors. We find that only one's nationality and the sector of employment have significant impact on our residential referral effect. Second, we look at how pairs of different groups interact: especially for several ethnic groups, the residential referral effect is big and significant. Also for industry groups and pairs of low qualified, the probability to work together

is increased when sharing a residential neighborhood.

We address possible shortcomings of the design in several ways: we check for potential sorting within super-neighborhoods and find some sorting on the basis of observables. Nevertheless, the extent of sorting within super-neighborhoods is not systematically different from that between super-neighborhoods, which is why we think we can address this problem by using the fixed effects. Apart from that, we would expect an upward bias for positive sorting, but especially those groups which tend to sort themselves together have lower or insignificant probabilities to work together, which makes us confident about the robustness of our findings. Furthermore, we find that there is no sorting on the basis of unobservables within super-neighborhoods, which means that including the fixed effects should deal with the issue of self selection into residential location.

Although we cannot rule out completely the possibility of a bias in our estimated referral effect due to simultaneity, we argue that it is very plausible that what we observe accounts for an actual referral effect on the job market, as we can show to find very similar results for a subset of individuals, for whom job referrals are most likely.

To comment on the extent to which the estimated effects are a result of clusters in employment and differences in accessibility, we reassign jobs randomly to people, while leaving their location unchanged. We find positive and significant spurious correlation due to the geographical distribution of workplaces. However, the greater portion can be attributed to an actual referral effect. The effect of spurious correlation amounts to only 0.03 percentage points for a spurious referral to a neighborhood, which means that even when subtracting this from our estimated interaction effect, we still find a positive and significant referral effect which is comparable in magnitude to what previous literature found.

Finally, we plan to analyze whether the correlation we measure is a result of disproportional many people working at their residential neighborhood. To address this issue, we exclude all short distance commuters and reestimate our baseline specification. The results do not change substantially, which indicates that our estimates are not driven by short distance commuters.

The paper investigates the effect of living together on the probability of working together. We find strong evidence for a positive and highly significant relationship, which is robust across several specifications and robustness tests, addressing common issues on the identification of neighborhood effects.

References

Altonji, Joseph, Todd Elder, and Christopher Taber (2005) "Selection of Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, Vol. 113, No. 1, pp. 151 – 184.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*: Princton University Press.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa (2008) "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, Vol. 116, No. 6, pp. 1150–1196.

Böltken, Ferdinand, Gabriele Sturm, and Antje Walther (2013) "LebensRäume - Bevölkerungsumfrage des BBSR (Bundesamt für Bauwesen und Raumforschung) 2000-2011," GESIS Datenarchiv, ZA5611 Datenfile Version 2.0.0.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009) "Identification of peer effects through social networks," *Journal of Econometrics*, Vol. 150, pp. 41–55.

Calvo-Armegnol, Antoni and Matthew O. Jackson (2007) "Networks in Labor Markets: Wage employment dynamics and Inequality," *Journal of Economic Theory*, Vol. 132, No. 1, pp. 27 – 46.

Corcoran, Mary, Linda Datcher, and Greg Duncan (1980) *Information and Influence Networks in Labor Markets*, five thousand american families: patterns of economic progress edition, Vol. 7.

Dietz, Robert D. (2002) "The estimation of neighborhood effects in the social sciences: an interdisciplinary approach," *Social Science Research*, Vol. 31, pp. 539–575.

Dustmann, Christian, Albrecht Glitz, and Uta Schönberg (2011) "Referral-based Job Search Networks," *IZA discussion papers*, Vol. 5777.

Evans, William N., Wallace E. Oates, and Robert M. Schwab (1992) "Measuring Peer Group Effects: A Study of Teenage Behavior," *Journal of Political Econonmy*, Vol. 100, No. 5, pp. 966–991.

Fujita, Masahisa, Paul Krugman, and Anthony J. Venables (1999) *The Spatial Economy: Cities, Regions, and International Trade*, Vol. 1: MIT Press Books.

Glitz, Albrecht (2013) "Coworker Networks in the Labour Market," *Working Paper*.

Granovetter, Mark S. (1973) "The Strength of Weak Ties," *American Journal of Sociology*, Vol. 78, No. 6, pp. 1360–1380.

Holzer, Harry J. (1987) "Informal Job Search and Black Youth Unemployment," *American Economic Review*, Vol. 77, No. 3, pp. 446–452.

Ihrke, David K. and Carol S. Faber (2012) "Geographical Mobility: 2005 to 2010,"Technical report, United States Census Bureau.

Ioannides, Yannis M. and Linda Datcher Loury (2004) "Job Information Networks, Neighborhood Effects, and Inequality," *Journal of Economic Literature*, Vol. 42, No. 3, pp. 1056 – 1093.

Jencks, Christopher and Susan E. Mayer (1990) *Inner-City Poverty in the United States*, Chap. The Social Consequences of Growing up in a Poor Neighborhood, Washington D.C.: National Academy Press.

Kain, John F. (1968) "Housing Segregation, Negro Employment, and Metropolitan Decentralization," *Quarterly Journal of Economics*, Vol. 82, No. 2, pp. 175–197.

Katz, Lawrance F., Jefferey Kling, and Jefferey Liebman (2001) "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics*, Vol. 116, No. 2, pp. 607–654.

Köhler, Uwe and Timo Bertocchi (2010) "Empfehlungen für Planung und Betrieb des öffentlichen Personennahverkehrs," Forschungsprojekt des Forschungsprogramms Stadtverkehr 70.837, Forschungsgesellschaft für Strassen- und Verkehrswesen.

Kramarz, Francis and Oskar Nordström Skans (2013) "When Strong Ties are Strong: Networks and Youth Labor Market Entry," *Review of Economic Studies*, forthcoming.

Ludwig, Jens, Greg J Duncan, Lisa A. Gennetian, Lawrance F Katz, Ronald C. Kessler, Jeffrey R Kling, and Lisa Sanbonmatsu (2013) "Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity," *American Economic Review: Papers and Proceedings*, Vol. 103, No. 3, pp. 226 – 231.

Manski, Charles F. (1993) "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, Vol. 60, pp. 531–542.

Montgomery, James D. (1991) "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis," *American Economic Review*, Vol. 81, No. 5, pp. 1408–1418.

Pellizziari, Michele (2010) “Do Friends and Relatives Really Help in Getting a Good Job?” *Industrial and Labor Relations Review*, Vol. 63, pp. 494–510.

Ports, Michelle H. (1993) “Trends in Job Search Methods, 1970-92,” *Monthly Labor Review*, Vol. 116, No. 10, pp. 63–67.

Scholz, Theresa, Cerstin Rauscher, Jörg Reiher, and Tobias Bachteler (2012) “Geocoding for German Administrative Data - The Case of the Institute for Employment Research,” FDZ Methodenreport 09, Research Data Centre of the German Federal Employment Agency at the Insitute for Employment Research.

Topa, Giorgio (2001) “Social Interactions, Local Spillovers, and Unemployment,” *Review of Economic Studies*, Vol. 62, No. 2, pp. 261 – 295.

Wooldridge, Jeffrey M. (2002) *Econometric Analysis of Cross Sections and Panel Data*, Cambridge, Massachusetts: The MIT Press.

Zenou, Yves (2013) “Spatial versus Social Mismatch,” *Journal of Urban Economics*, Vol. 74, pp. 113 – 132.

Appendix

Table .7: Estimation of Heterogenous Referral Effects, Full Output

	Refferal to neighborhood ($a = n$)		Refferal to firm ($a = f$)	
Variable	(2)	(3)	(2)	(3)
Constant	2.2207***	2.0837***	.2347***	-.1196
	(.2998)	(.3811)	(.1063)	(.1054)
R_{ij}	.1432***	.1237***	.0784***	.0605***
	(.0238)	(.0185)	(.0241)	(.0182)
male	-.4281***	-.3378***	.0266***	.2031***
	(.0322)	(.0204)	(.0068)	(.0094)
male x R_{ij}	-.0047	.0457***	-.0065	.0426***
	(.0205)	(.0138)	(.0152)	(.0076)
Age 15-24	.1423***	.4552***	-.0300**	.0173
	(.0539)	(.0550)	(.0119)	(.0125)
15-24 x R_{ij}	.0018	.0482	-.0030	.0738***
	(.0298)	(.0493)	(.0176)	(.0263)
Age 25-34	-.1018***	-.1012***	-.0204**	.0513***
	(.0381)	(.0371)	(.0093)	(.0161)
25-34 x R_{ij}	.0256	.0245	-.0030	.0072
	(.0182)	(.0264)	(.0176)	(.0183)
Age 55-65	.2559***	.5401***	.0471***	.1673***
	(.0414)	(.0379)	(.0109)	(.0119)
55-65 x R_{ij}	.0253	.0382	-.0268	.0203
	(.0310)	(.0344)	(.0109)	(.0213)
Unskilled	.2130***	.7077***	.1592***	.4913***
	(.0485)	(.0436)	(.0117)	(.0232)
Uskill x R_{ij}	.0106	.1874***	.0136	.1640***
	(.0224)	(.0389)	(.0108)	(.0283)
Medium Skilled	.0335	.1684***	.1085***	.1859***
	(.0370)	(:0195)	(.0069)	(.0064)
Mskill x R_{ij}	-.0206	.0199	.0122	.0353***
	(.0193)	(.0158)	(.0127)	(.0118)
Highskilled	-.3452***	-.1757**	.0887***	.3355***
	(.0656)	(.0806)	(.0310)	(.0601)
Hskill x R_{ij}	.1225	.7734	.1639	.8573
	(.1632)	(.7215)	(.1705)	(.7615)

Greek	.0714	.6593**	.0544	.6890***
	(.1938)	(.3124)	(.0442)	(.1534)
Greek x R_{ij}	.2102**	1.1252***	.0544	.9341***
	(.0924)	(.3955)	(.0442)	(.3205)
Italian	.3245**	.8777***	.0388	.4942***
	(.1563)	(.2147)	(.0508)	(.1168)
Italian x R_{ij}	.2061	1.3011**	.2183	.9314***
	(.2092)	(.5590)	(.2092)	(.2962)
Spanish	.4197	.2624	.0565	.5697*
	(.4024)	(.5732)	(.1050)	(.3239)
Spanish x R_{ij}	-.1935	1.0928	.0036	.4120
	(.1337)	(1.0306)	(.0508)	.7877)
Turkish	.1791**	1.0417***	.1543***	.9615***
	(.0793)	(.1300)	(.0247)	(.0911)
Turkish x R_{ij}	.0355	.1888	.0392**	.1672**
	(.0404)	(.1221)	(.0160)	(.0677)
Yugoslavian[a]	.1747	.5328**	.0206	.3085***
	(.1409)	(.2055)	(.0309)	(.0728)
Yugo. x R_{ij}	.1214*	1.0888***	.0012	.6416***
	(.0665)	(.2647)	(.0189)	(.1313)
From new EU[a]	-.0339	1.3035*	-.0571	.3241
	(.1657)	(.6390)	(.0384)	(.3214)
New EU x R_{ij}	.5642*	23.8789***	.4976	23.4559***
	(.3133)	(5.5630)	(.3144)	(6.9550)
Primary Sector	.1152*	6.1700***	-.2440*	5.0837***
	(.0697)	(.3083)	(.0191)	(.2351)
PSector x R_{ij}	-.0936**	1.777***	-.0268	1.8385***
	(.0373)	(.3862)	(.0373)	(.3411)
Construction	-.1783**	.6681***	-.3861**	.4942***
	(.0593)	(.0993)	(.0181)	(.1169)
Constr. x R_{ij}	-.0293	.7193***	-.0703**	.1385***
	(.0364)	(.1031)	(.0215)	(.0259)
TTC[b]	.3323***	.7939***	-.3072***	.4300***
	(.0434)	(.0336)	(.0173)	(.0163)
TTC x R_{ij}	-.0493*	.2150***	-.0661*	.1827***
	(.0279)	(.0356)	(.0205)	(.0321)
Buisness Services	.3367***	1.1018***	.3367***	1.1018***
	(.0530)	(.0563)	(.0530)	(.0563)

Buisness x R_{ij}	-.0996***	.1754***	-.0996***	.1754***
	(.0304)	(.0404)	(.0304)	(.0404)
Other Service	.6980***	1.7521***	-.1078***	.8796***
	(.0120)	(.1761)	(.0205)	(.0307)
Services x R_{ij}	-.0171	.03045*	.0112	.3091*
	(.0622)	(.1761)	(.0613)	(.1853)
coresize	32.5784	27.6795	-4.3344	-3.9490
	(20.6601)	(19.6010)	(5.4710)	(5.4285)
csize x R_{ij}	-.0822***	-.0882***	-.0500***	-.0571***
	(.0120)	(.0124)	(.0072)	(.0083)
σ_u	36.6795	31.1610	5.1042	4.6731
σ_ε	13.2821	13.2703	4.7623	4.7485
# pairs	179.7 Mio	179.7 Mio	179.7 Mio	179.7 Mio
# groups	10,159	10,159	10,159	10,159
Corr(u,Xb)	-.0042	-.9996	.0249	-.9984

Heteroscedasticity-consistent standard errors in parentheses.

*/**/***mark significance at the 90%/95%/99% confidence level.

[a]: Yugoslavian covers immigrants from the territory of former Yugoslavia (including Slovenia and Croatia); these are not included in the group of immigrants from new EU members (which come from Estonia, Latvia, Lithuania, Poland, Czech Republic, Slovakia, Hungary, Bulgaria, Romania, Malta and Cyprus).

[b]: Trade, Transportation and Communication (TTC).

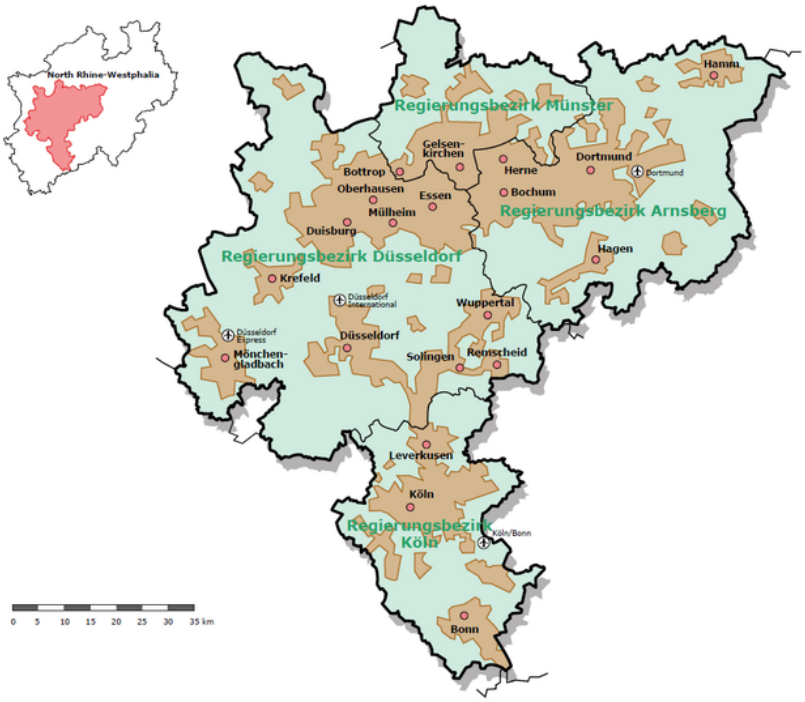

Figure .3: Rhein-Ruhr Metropolitan Area