MAGYAR TUDOMÁNYOS AKADÉMIA
Közgazdaság- és Regionális Tudományi Kutatóközpont

MTA KRTK KTI

Centre for Economic and Regional Studies
HUNGARIAN ACADEMY OF SCIENCES

MŰHELYTANULMÁNYOK DISCUSSION PAPERS

MT-DP – 2015/50

Co-worker networks, labour mobility, and productivity growth in regions

BALÁZS LENGYEL - RIKARD ERIKSSON

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES,
HUNGARIAN ACADEMY OF SCIENCES BUDAPEST, 2015

Discussion papers
MT-DP – 2015/50

Institute of Economics, Centre for Economic and Regional Studies,
Hungarian Academy of Sciences

Co-worker networks, labour mobility, and productivity growth in regions

Authors:

Balázs Lengyel
research fellow
Institute of Economics - Centre for Economic and Regional Studies
Hungarian Academy of Sciences
e-mail: lengyel.balazs@krtk.mta.hu

Rikard Eriksson
associate professor
Department of Geography and Economic History
Umeå University
e-mail: rikard.eriksson@umu.se

September 2015

ISBN 978-615-5243-39-4
ISSN 1785 377X

Co-worker networks, labour mobility, and productivity growth in regions

Balázs Lengyel and Rikard Eriksson

Abstract

This paper provides a new empirical perspective for analysing the role of social networks for an economic geography approach on regional economic growth by constructing large-scale networks from employee-employee co-occurrences in plants in the entire Swedish economy 1990-2008. We calculate the probability of employee-employee ties at plant level based on homophily-biased random network assumptions and trace the most probable relations of every employee over the full period. Then, we look at the inter-plant ties for the 1995-2008 period because the network is already well developed after five years of edge construction. We argue that these personal acquaintances are important for local learning opportunities and consequently for regional growth. Indeed, the estimated panel Vector Autoregressive models provide the first systematic evidence for a central claim in economic geography: social network density has positive effect on regional productivity growth. The results are robust against removing the old and therefore weak ties from the network. Interestingly, the positive effect of density on growth was found in a segment of the co-worker network as well, in which plants have never been linked by labour mobility previously.

JEL: D85, J24, J61, R11, R23

Keywords: social network, homophily, probability of ties, labour mobility, regional productivity growth, panel vector autoregression

Acknowledgement

Help in geo-localization of the data was received from Magnus Strömgren. We thank Pierre-Alexandre Balland, Gábor Békés, Dániel Horn, Guilherme Kenjy Chihaya da Silva, Urban Lindgren and László Lőrincz for their suggestions. Useful comments have been received from participants of Institute of Economics seminar series at the Hungarian Academy of Sciences, ICELAB seminar at Umea University, and Economic Geography Session Series at the AAG Annual Meeting in Chicago, 2015. The work of Lengyel was supported by the Hungarian Scientific Research Fund (K112330).

Munkatársi kapcsolathálók, mobilitás és a regionális termelékenység növekedése

Lengyel Balázs és Rikard Eriksson

Összefoglaló

Ebben a tanulmányban egy új empirikus megközelítést kínálunk a társadalmi kapcsolathálózatok és a regionális gazdasági növekedés közötti összefüggések elemzésére. Svéd munkavállalói és telephelyi adatokból készítünk nagyméretű kapcsolathálózatot, ami lefedi a teljes svéd gazdaságot 1990-től 2008-ig. Homofílikus random hálózatokat feltételezve kiszámoljuk a munkavállalók azonos telephelyen dolgozó kollégáikkal való kapcsolatainak valószínűségét, és az összes munkavállaló minden évre vonatkozó legvalószínűbb kapcsolatait követjük végig a teljes időszakon. Ezután a telephelyek közötti kapcsolatokra koncentrálunk az 1995–2008 közötti időszakban, öt évet hagyva arra, hogy felépüljön a hálózat. Érvelésünk szerint ezek a személyes ismeretségek fontos katalizátorai a helyi cégek egymástól való tanulási folyamatainak, amelyek a regionális növekedésre is hatnak. Várakozásainknak megfelelően a becsült panel vektor autoregresszív modellek adják az első szisztematikus bizonyítékot a gazdaságföldrajz egyik központi állítására: a társadalmi kapcsolatháló sűrűsége pozitívan hat a régió termelékenységének növekedésére. Az eredményeink változatlanok maradnak akkor is, amikor a régi és ezért feltehetően gyenge kapcsolatokat eltávolítjuk a hálózatból. Érdekes módon akkor is megmarad a pozitív hatás, ha olyan telephelyek közötti kapcsolatokat nézünk, amelyek között sosem volt észlelhető munkavállalói mobilitás.

JEL: D85, J24, J61, R11, R23

Tárgyszavak: társadalmi kapcsolatháló, homofília, kapcsolatok valószínűsége, munkavállalói mobilitás, regionális termelékenységnövekedés, panel vektor autoregresszív modellek

INTRODUCTION

Following Marshall (1920) there is a general agreement in economic geography and related fields that the agglomeration of economic activities is essential for understanding regional innovation and growth. In this respect, face-to-face interaction is increasingly emphasized as essential for why proximity is still crucial for sustaining learning and innovation (Storper and Venables, 2004), and that more dense environments enhance the probability of "learning by seeing" (Glaeser, 2000). Human interaction and the social networks created thereof are thus expected to be key drivers behind regional economic growth. This is basically because the effectiveness of learning and co-operation of individuals are enhanced by personal relations and this is expected to have both direct and indirect effects on productivity growth since firms gain extra benefits when accessing external knowledge through social ties. However, despite the above theoretical claims on the role of face-to-face contacts and social networks for learning and growth, very little empirical work has actually been devoted to analysing the role of social networks on regional productivity growth. As argued by Huggins and Thompson (2014), the role of social networks for regional growth is still highly unresolved. Instead, scholars tend to proxy the socializing potential of regions by means of population density or industrial structure (Ciccone and Hall, 1996, Glaeser, 1999), and almost take the relation between density and social interaction for granted by assuming that the mere concentration of skilled workers automatically will increase the probability for social interaction and thus enhance learning and growth.

To address this potential shortcoming in the existing literature, the aim of this paper is to assess the influence of co-worker networks on productivity growth in 72 Swedish labour market regions 1995-2008. This is made possible by a unique longitudinal matched employer-employee database from which we construct a social network of employees based on their co-occurrence at workplaces 1990-2008 and analyse the effect of the network on productivity, proxied as regional income per capita These type of networks are frequently called co-worker networks in labour economics and scholars assume that two employees know each other when they have worked in the same workplace simultaneously in a certain period of their career (for an overview see Beaman and Jeremy, 2012). Evidence shows that information flow through these co-worker relations help people find better jobs and reduce unemployment time when dismissed (Calvo-Armengol and Jackson, 2004, Glitz, 2013, Granovetter, 1995, Hensvik and Nordström Skans, 2013). Given that the exchange of information and knowledge between workers and firms promotes the emergence and diffusion of innovation and subsequent productivity (Duranton and Puga, 2004), we claim that co-worker networks are important sources of regional economic dynamics. This is because valuable information flows more efficiently through co-worker relations and employees might learn more efficiently in dense co-worker networks (c.f. Breschi and Lissoni, 2009, Eriksson and Lindgren, 2009, Huber, 2012).

We claim to make two contributions to the existing literature. First, we develop a new probability measure of workplace-based acquaintance, building on the literature of homophily-biased random networks (Buhai and van der Lei, 2006, Currarini et al, 2009). We calculate tie probability using the concept of baseline homophily and rank employee co-occurrence according to this probability. Then, we trace a selected number of most probable individual ties of every employee. As result, we get a dynamically changing social network that represents the full economy and still captures social ties at the micro scale. Despite that co-worker networks and labour mobility networks presumably are interconnected because people establish new links in the co-worker network through mobility from one firm to another (Collet and Hedström, 2012), we illustrate in details that our approach differs from previous labour mobility studies in both conceptual and empirical concerns (e.g., Breschi and Lissoni, 2009, Eriksson and Lindgren, 2009).

The second contribution is that this paper provides the first empirical evidence that the density of the social network has a positive effect on productivity growth defined as regional income per capita. The findings are robust against removing the old and presumably weak ties from the network, as well as focusing only on a segment of the co-worker network, in which plants have never been linked by labour mobility previously.

LITERATURE AND HYPOTHESES

The spatial dimension of network-related learning and how that may influence regional growth is a core interest of economic geography (Bathelt and Glückler, 2003; Huggins and Thompson, 2014). It is well understood now that transaction costs are diminished by physical proximity as well as personal connections, which enhance the efficiency of mutual learning (Borgatti et al, 2009, Maskell and Malmberg, 1999, Sorensen, 2003). It is also claimed that most of the learning processes occur within certain spatial proximity despite distant, and presumably weak, ties might provide the region with new knowledge (Bathelt et al, 2004, Glückler, 2007). We also understand that not the social network per se but its' interplay with industry structure is crucial for learning because cognitive, institutional, and organizational proximities are very important for mutual understanding (Boschma, 2005, Sorensen et al, 2006). Despite the central interest, our knowledge about the network effect on regional productivity growth is still limited (Huggins and Thompson, 2014), which is partly due to data access difficulties. Our paper aims to contribute to the literature in this regard by constructing and analysing a large-scale co-worker network.

In generating co-worker networks, we assume that two employees know each other when they have worked in the same workplace simultaneously in a certain period of their career. Evidence shows that information flow through these co-worker relations help people find better jobs and reduce unemployment time when dismissed (Beaman and Jeremy, 2012, Calvo-Armengol and Jackson, 2004, Granovetter, 1995, Hensvik and Nordström Skans, 2013). However, the co-worker network approach in labour economics often focuses on small firms only because two randomly selected employees are less likely to know each other in a large firm than in a small firm. For example, Glitz (2013) only looked at firms with maximum 50 employees. The issue of firm size is still important in economic geography; however, we cannot eliminate co-worker networks generated at large firms when estimating the effect of the network on economic growth. Therefore, we develop a new probability measure of workplace-based acquaintance, building on the literature of homophily-biased random networks (Currarini et al, 2009). We calculate tie probability of every possible employee-employee pair accordingly and trace the most 50 probable individual ties of every employee.

The argument stresses three points. First, regional income per capita growth is positively related to co-worker network density as it is claimed in the first hypothesis. Second, the positive effect of network density remains significant when old and presumably weak ties have been eliminated from the network, as stated by the second hypothesis. Third, although co-worker networks are generated by means of inter-firm labour mobility, the effect of co-worker network density on regional growth is independent from labour mobility networks, which is claimed by the third hypothesis.

Density indicators – population density, in particular – have been repeatedly found to have a positive effect on regional economic growth. This is because spatial agglomeration unburdens the sharing of common facilities, increase the chances of a productive job-worker matching, and enhances interactive learning through the concentration of firms and workers (Duranton and Puga, 2004), which has a direct effect on productivity growth differences (Ciccone and Hall 1996, Glaeser 1999). As argued by Glaeser (2000) workers in dense environments are more likely to acquire human capital through learning by seeing which make dense regions more productive as well as more attractive for skilled workers with large potential returns for learning which will further increase productivity. We argue that looking at not only the co-location of individuals but investigating also the density of social networks will improve our understanding because face-to-face relations and personal acquaintance are important for knowledge sharing (Storper and Venables, 2004). Workplaces and consequently the co-worker networks that bind workplaces together are major fields of such knowledge sharing even after the termination of the co-worker relation because people maintain their professional contacts over time and might even follow the career of former colleagues in order to map out the knowledge-base they have potential access to (Dahl and Pedersen, 2003). Thus, co-worker networks are important for local learning and consequently on regional economic growth.

H1: Density of the local co-worker network enhances regional income per capita growth.

The first hypothesis refers to a central debate in the social networks literature. Network density has been considered as a major indicator of social capital for decades in sociology (Burt, 1992, Coleman, 1990, Walker et al, 1997, Wasserman and Faust, 1994) because the closure of social relations enhances trust, authority and sanctions among local actors, all of which supports learning from contacts. Certainly, density alone does not sufficiently describe the full horizon of information-flow tendencies in a network. The strength of social ties is a crucial factor and results in two fundamental processes (Granovetter, 1973). On the one hand, weak ties offer access to new information and combination of non-redundant knowledge, which can lead to radical innovations (Ahuja, 2000). On the other hand, people trace strong ties frequently, which offers the possibility of incremental innovation and increase in individual productivity because they learn effectively from each other (Balkundi and Harrison, 2006, Borgatti and Cross, 2003). The above issue of tie strength in the co-worker network and local learning is addressed by removing the old and presumably weak ties from the network and focusing only on the recent and strong ties, which is a process suggested in the sociology and network science literatures (Burt, 2000, Murase et al, 2015).

H2: The positive effect of network density on regional economic growth remains significant even if we eliminate the old and presumable weak ties from the network.

Similar ideas to the network-related learning have been present in the economic geography literature (for an overview see Ter Wal and Boschma, 2009). For example, strong social ties within certain sectors in specialized industrial districts enhance incremental innovation and productivity growth (Amin, 2000, Asheim, 1996, Malmberg, 1997), whereas diverse networks across industries in urban areas are associated with potential new combinations of information, creation of new knowledge and radical innovation (Feldman, 1999). More recently, the emerging literature of evolutionary economic geography suggests that spatial learning depends on a complex combination of various proximity dimensions between individual firms and that regional productivity growth is the result of technological proximities among co-located firms (Boschma, 2005, Frenken et al, 2007). Labour flows have been used extensively to proxy technological proximities or relatedness across industries

(Neffke and Henning, 2013); and a growing number of papers consider spatial labour mobility between firms as a major source of learning due to the transfer of embodied knowledge (Almeida and Kogut, 1999, Eriksson and Lindgren, 2009) and assess the effect of related labour flows on regional and firm dynamics (Boschma et al, 2009, Timmermans and Boschma, 2014). Apart from improving the potential regional matching of skills, Boschma et al (2014) also show that high concentrations of skill-related flows in a region strongly influence productivity growth in Sweden due to the production complementarities produced by such labour market externalities.

The co-worker approach is closely connected to the labour mobility approach because we assume that former colleagues maintain their relations even after moving from one workplace to another (Collet and Hedström, 2012), which is a proposition made in evolutionary economic geography as well (Boschma and Frenken, 2011). Related empirical evidence shows lasting co-inventor relations are important for later patenting collaborations (Agrawal et al, 2006, Breschi and Lissoni, 2009). The recent paper is the first attempt to analyze co-worker networks in economic geography. We aim to show that not only the transfer of embodied knowledge and labour flows, but also social networks that are independent from labour flows, have an effect on regional growth. Therefore, we decompose the co-worker network into two segments: (1) links have been preceded by labour mobility and (2) links that have not been preceded by labour mobility.

> *H3: Co-worker network density enhances regional income per capita growth even if the ties across plants have not been preceded by labour flows among the concerned plants.*

METHODOLOGY

We propose that employee i and employee j working for in the same workplace at the same period of time know each other with probability P_{ij} [0,1] and maintain a tie Lij even after the termination of the co-workership. For practical reasons, we select the most probable 50 co-workers of highest P_{ij} for each employee in each year and trace these co-occurrences over the full period and look at those L_{ij} when employee i and employee j work for two different firms.

Probability calculation starts from the assumption of random tie formation at workplaces, which means that a tie between every pair of employees is established with equal probability. Intuition suggests that the larger workplace the less likely that employees know each other. Thus, we first set tie probability proportional to the size of workplace. However, this tie probability creates a large fraction of isolated ties in random network simulations, which is not our intention. Therefore, we use the probability threshold where isolated nodes tend to disappear in a random network setting (Erdős and Rényi, 1959, Jackson, 2008) and formulate random probability (P^r_{ij}) by

$$P^r_{ij} = \frac{\ln N}{N}; \qquad (1)$$

where N is the number of employees in the workplace.

In a second step, we consider that individual similarity increases the probability of tie formation, which is called homophily in the large range of social sciences (for an overview see McPherson et al, 2001). It has been shown repeatedly that much more friendship ties are formed across those individuals who are similar in terms of age, gender, race, education, occupation etc. than expected by random tie establishment (Blau, 1977, Blau et al, 1982, Blum, 1985, Feld, 1982, Granovetter, 1995, Kossinets and Watts, 2006, Lincoln and Miller, 1979, McPherson and Smith-Lovin, 1987, Sias and Cahill, 1998). Two types of homophily are distinguished in the literature: baseline homophily and inbreeding homophily. Baseline homophily means that individual choice of selecting friends is generated by the structure of the group because the larger subgroup of similar individuals the larger possibility of choosing similar friends. Thus, baseline homophily (H_b) can be measured by the share of subgroup in the firm by

$$H_b = \frac{N_m}{N}; \qquad (2)$$

where Nm denotes the size of the subgroup characterized by feature m.

We will assume that H_b influences P_{ij} because relations are more likely between those employees who are of similar age and sex and have the similar educational background. However, Currarini et al. (2009) showed that friendship ties usually exhibit larger homophily than H_b due to additional inbreeding homophily and individuals' choice is even more biased towards akin. Thus, using H_b we will most likely underscore the real probability of the tie between co-workers. We define employee characteristics like age, gender, and education as those subgroup features that are expected to increase tie probability then we can calculate H_b in a repetitive manner as explained above.

In the third step, we have to realize that the size of the subgroups – defined by employee characteristics – has a similar effect on tie probability than the firm size itself. Thus, we have to diminish the probability by $ln\ (N_m\ /\ N)$ in each case when employee i and j are similar.

Finally, we simply sum the probabilities calculated from firm size, baseline homophilies and group size effects in order to get probability of co-worker ties (Buhai and van der Lei, 2006). Probability is formulated as

$$P_{ij} = \frac{\ln N}{N} + \sum_{G=1}^{M} \left(\frac{\ln N_m}{N_m} \Big/ \frac{N_m}{N} \right) \times \delta_{ij}; \qquad (3)$$

where G $\in \{1, 2, \dots M\}$ denotes those characteristics we use for similarity measurement, N denotes plant size, N_m denotes subgroup size according to feature m and δ_{ij} equals 1 if employee i and j are similar according to feature m and 0 otherwise.

We maximize co-worker tie probability at 1, rank co-workers for every employee and follow the 50 most probable co-workers of every employee over time. Density of the network can be calculated for every region and every year. Then, for investigating *H2*, we eliminate those ties that are older than 5 years. In order to investigate *H3*, we also distinguished employee-employee links by concerning the presence or absence of previous labour mobility between the plants, as we further explain in Section 5.

NETWORK CREATION AND DESCRIPTION

DATA AND NETWORK CREATION

We use matched employer-employee data obtained from official registers from Statistics Sweden that –among a wide variety of data– contains age, gender, and detailed education code of individual employees and enables us to identify employee-employee co-occurrence at plants for the 1990-2008 period. Data is generated on a yearly basis and if employees change workplace over the year, they are listed repeatedly with different plant codes in the same year. Geo-location of plants is defined by transforming the data from a 100m x 100m grid setting into latitudes and longitudes.

For practical reasons, and in order to keep the size of the sample at the limit the analysis can handle, we exclude those without tertiary education from the data. Including all employees would exponentially increase computation demand without contributing much to the analysis. This is motivated by the fact that skilled workers (bachelors) are assumed to benefit more from learning by seeing and interacting (Glaeser, 2000). We therefore propose that workers without bachelor degrees rely to a greater extent on tacit or embodied knowledge and therefore might learn less from an individual level social network with colleagues at other plants. If an employee who has already been in the data obtains graduation at a later point in time, she will be included in our sample afterwards. As a result, the data contains 366.336 individuals in 1990 and 785.578 individuals in 2008 and those plants are excluded where none of the employees had BA degree or above (Table 1).

Table 1

Number of employees, plants, and co-occurrence in 1990 and 2008

		1990	2008
All employees	Employees	2,628,306	3,824,182
	Plants	254,445	402,610
Employees with BA degree or above	Employees	366,336	785,578
	Plants	52,872	113,441

We first generated the list of employee pairs as co-occurrence at plants for every year, then calculated the probability of the co-worker relation for each employee pair using Equation 3. Three characteristics of employees were used to generate subgroups: Direction of education (6 groups), gender (2 groups) and age (3 groups). For further information of group definitions and descriptive statistics, see Appendix 1.

It is clear that employee co-occurrence is exponentially higher in large plants than in small plants and that co-workers know each other at a lower probability in large plants than is small plants. However, there is no clear suggestion in the literature regarding a reasonable number of ties per person, which can be handled by the analysis. Management papers report on task-oriented ego-networks based on survey data and the number of personal ties in these networks are below ten on average (Brass, 1985, McPherson et al, 1992, Lincoln and Miller, 1979, Morrison, 2002). Recent papers in labour economics tend to construct much larger co-worker networks assuming that everyone knows everyone in a firm not larger than 500 (Hensvik and Nordström Skans, 2013) or 3000 employees (Saygin et al, 2014), while Glitz (2013) only looked at firms with between 5 and 50 employees.

Our approach is based on the labour economics literature; however, we handle the problem of co-worker ties in large plants in a novel way and rank employee pairs based on their P_{ij} values[1] in every plant regardless of size. P_{ij} values are calculated and relations are ranked on a yearly basis, which most likely make co-worker ties appear and disappear from the employees' portfolio in large plants from year to year. To handle this problem, we trace all those co-worker ties that were ranked among the top 50 at least in one year over the full period. We exclude the tie if at least one employee is already above 65 years of age, if either one or both individuals are not present in the labour market and if the employees work in the same plant.

[1] In case employee pairs have the same probability, we rank those with same educational background and smaller age difference higher, respectively.

The unbalanced panel of 155,671,574 employee pairs constitute a dynamic co-worker network over the 1991-2008 period we look at in the analysis. This network can be analysed on the individual level, and ties can be aggregated on the plant and industry levels. However, we must keep in mind, that this is a constantly growing network, because the number of employees in the sample increases monotonically, which is not balanced by labour market exits[2].

PROPERTIES OF THE CO-WORKER NETWORK

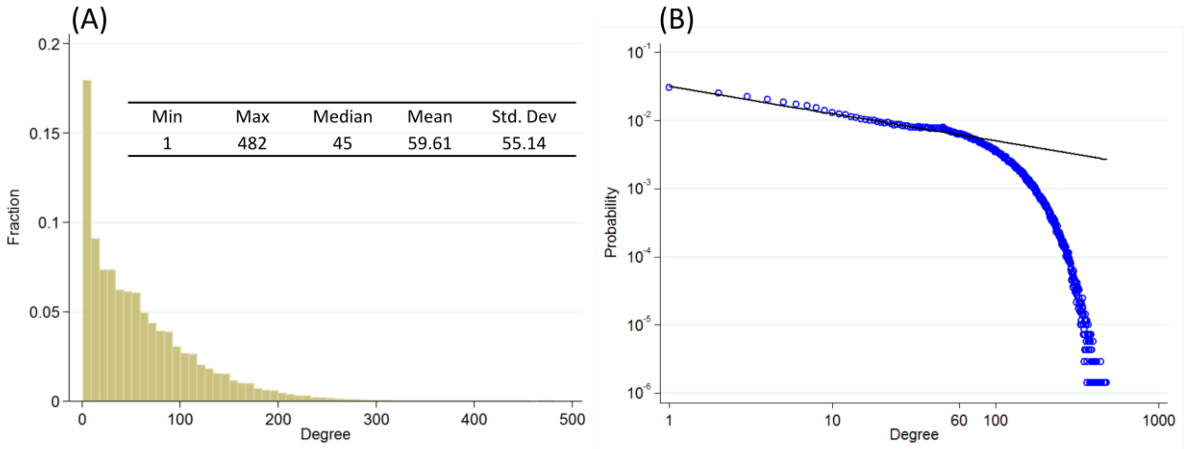
The analysis is based on the assumption that the co-worker network resembles social networks embedded in spatial environments. In this section we show that both the degree distribution and the spatial dimension of the network fulfil these criteria.

We find a negative exponential degree distribution of the co-worker network in year 2008, which has some very nice properties. For example, the expected degree can be approximated by the average degree in the network. Furthermore, we find that the probability of finding employees who has more degrees than the average decreases sharply. Thus, the mean is not only the expected value but also a turning point in the distribution.

Figure 1.

Degree distribution and summary statistics

of the individual level network, 2008



Note: The slope of the solid line is -0.4 in (B).

[2] See Lengyel and Eriksson (2015) for a more detailed discussion about the 50 best friend approach and more information regarding the co-worker network generation.

The histogram of degrees on a natural scale resembles a negative exponential distribution, where the fraction of nodes decreases monotonically as degree grows (Figure 1A). The degree varies on a large scale from a minimum value of 1 to a maximum value of 482. The mean is larger than the median and standard deviation almost equals to the mean, which are well-known properties of exponential distributions. Furthermore, the approximated rate parameter proxies the median quite well[3].

The degree distribution in 2008 illustrated on a log-log scale (Figure 1B) resembles degree distributions in other large-scale social networks (Adamic and Adar, 2005). The majority of employees have small number of connections and the probability that the employee has degree d decreases exponentially with an exponent -0.4 until d is around 60. This exponent is very similar to the exponent (-0.35) found previously in a large-scale online social network (Lengyel et al, 2015). The break in the distribution suggests that the probability of larger degrees than the turning point falls sharply as degree grows, which implies that there are very few employees with many connections and the number of these employees is proportional to their degree. Interestingly, the turning point of the distribution coincides with the mean. Cumulative degree distribution can be found in Appendix 2.

The spatial level of the regional growth model will be selected on the basis of the network geography and here we provide information on how co-worker ties scatter across space. Not surprisingly, the network is spatially concentrated. More than 30% of all individual links were within municipality borders (the smallest administrative division in Sweden) in 2008 and this share is 60% when we look at functional regions (Table 2). The latter regions represent labour market areas defined by The Swedish Agency for Economic and Regional Growth. This regional definition covers the whole territory of Sweden without overlapping each other and stem from observed commuting distances between the 289 Swedish municipalities. When we aggregate the network on the plant level we find a very similar pattern.

Table 2.

Number of ties within regional borders, 2008

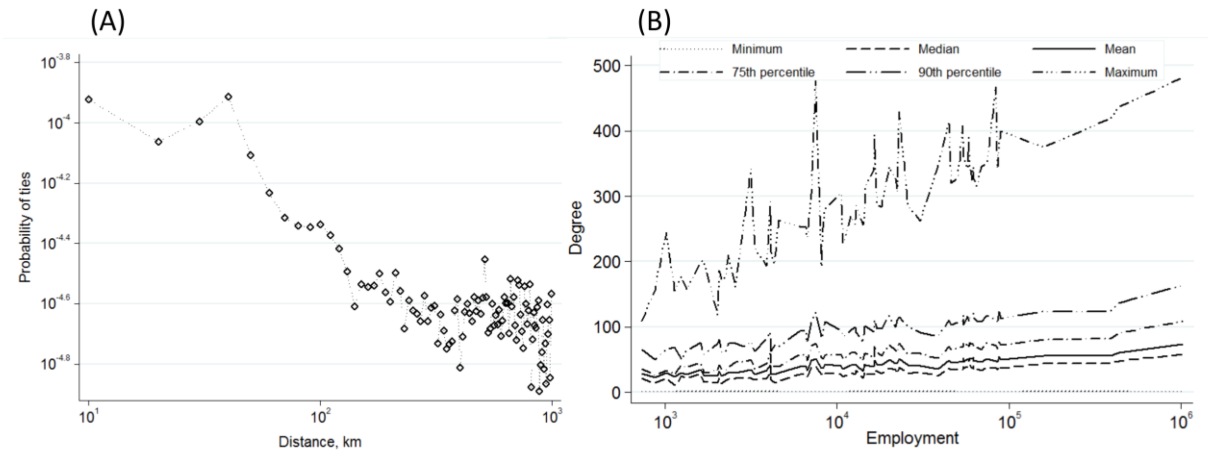
	Number of links	
	Individual level	Plant level
Full network	20,855,160	5,574,879
Within functional region (N=72)	14,066,872	3,170,695
out of which within municipalities (N=289)	7,826,977	1,470,603
Across functional regions	6,788,288	2,404,184

[3] The mean in exponential distributions is $E[X]=1/\lambda$. Approximating the rate parameter by reciprocating the mean gives us $\lambda = 0.02$. Then, substituting the rate parameter into $m[X]=ln(2)/\lambda$ gives us 40 as median, which is a fair approximation.

The previous observation gets further support when we look at the probability of having a tie between two arbitrary employees as a function of distance. We define L_d as the number of observed ties between employees separated from each other by distance d; and N_d the number of possible ties at distance d. Then, we can calculate the probability that individuals have links to others given distance d by the formula $P_d = L_d / N_d$. A 10 km resolution was used for binning distance distribution. The probability of a co-worker tie is close to be constant until 40-50 kilometres, after which it falls sharply (Figure 2A). Since the average distance of commuting to another town in Sweden is 45 km, we find that labour market areas and thus functional regions are the proper ground for testing our hypothesis.

Figure 2.

Distance effect and degree distribution by region size, 2008



Note: (A) The effect of distance on probability of ties. (B) The degree distribution in the region is depicted by minimum, median, mean, 75th and 90th percentile and maximum values.

The degree distribution does however not only depend on region size. We have plotted the minimum, median, mean, 75th percentile, 90th percentile and maximum values of degrees against the number of employees in the region. Figure 2B demonstrates that these values grow as the size of the region increases. However, we find that except the line connecting the maximum values, degree distribution in larger regions is only a little bit pushed to the right compared to smaller regions. The sharp increase of maximum degree in regions implies that the distribution has a longer and longer tail as the size of the region grows.

LABOUR MOBILITY AND DEGREE IN THE CO-WORKER NETWORK

Labour mobility is considered one of the major factors behind co-worker networks (Collet and Hedström, 2012). Indeed, labour mobility is the most influential of the free factors that might drive degree of individuals in our method.

1. Intra-plant changes across employee categories might increase the degree, because we have three age categories and people gain or loose similarity to other colleagues at the same plant over years in their career. This might be especially true in big plants, and therefore we use YEARS IN CAREER (total number of years spent in work) and AVERAGE PLANT SIZE (the average size of plants the employee worked for weighted by the years spent at the plant) variables to address this problem.
2. Labour mobility of the employee herself has an effect on her degree because the more one moves the more friends we count over time. Thus, we measure the effect of JOB CHANGES (the number of entries to new plants) on degree.
3. Labour mobility at the plant-level might influence the degree as well, because the employee can get co-workers if a new colleague arrives to the plant and she gets new connections in the network if someone leaves. We expect that the more people come and go over time the more friends we count; thus, we use the MOVEMENTS variable (the aggregate number of mobility to and from the plant at the time when the employee was working for the plant) to address this issue.

In fact, if projecting degree distribution on any of the above variables, the degree grows as years in career, average plant size, job changes and movements increase (see Lengyel and Eriksson 2015). To show the dominance of labour mobility in generating the co-worker network, we carry out a multivariate analysis, in which the degree of employees is the dependent variable and the indicators introduced above are used as explanatory variables. We include the size of the region into the analysis (Employment in the region) in order to double check its' effect on individual degree. We transform all the above variables to the logarithm of base 10. Since Average plant size and Movements are highly correlated (0.94) they have been inserted separately.

Results of the cross-sectional OLS regression, in which Degree was set as dependent variable, imply the higher values of factors the higher degree (Table 3). Nevertheless, Job changes and Movements are found to have the strongest effects on degree. These two variables together explain 61% of the variation of individual degree in the co-worker network (Model 3). Therefore, labour mobility needs to be considered explicitly when estimating the effect of the co-worker network on regional dynamics.

Table 3.

Drivers of degree (log) in the co-worker network,

cross-sectional OLS regression, 2008

	Model 1	Model 2	Model 3
Years in career (log)	0.566***	0.379***	
	(0.001)	(0.001)	
Average plant size (log)	0.380***		
	(0.001)		
Job changes (log)	0.850***	0.769***	1.010***
	(0.002)	(0.002)	(0.001)
Movements (log)		0.382***	0.417***
		(0.001)	(0.001)
Employment in the region (log)	0.026***	0.014***	
	(0.001)	(0.001)	
Constant	-0.131***	-0.125***	0.125***
	(0.004)	(0.004)	(0.001)
N	696,354	696,354	696,354
R^2	0.669	0.655	0.609
F	373,634.071	355,138.141	543,257.6

Note: Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001

MODELLING THE INFLUENCE OF NETWORK DENSITY

DENSITY EFFECT

The widely known formula that gives us the density of a network is the following

$$D = \frac{2 \times L}{N \times (N-1)} \; ; \qquad (4)$$

where L is the number of observed links and N is the number of nodes. However, the above formula handles intra-plant ties as observable, which is not the case in the co-worker network because we only observe inter-plant ties. Therefore, we have to reduce the nominator with the number of potential employee-employee pairs at same plants. Thus, density of the co-worker network in the region (D_c) is

$$D_c = \frac{2 \times L}{N \times (N-1) - \sum_k N_k \times (N_k - 1)} \; ; \qquad (5)$$

where N_k is the number of employees at plant k and $\sum_k N_k$ equals N.

To estimate whether social network density is related to productivity growth we resort to a panel vector autoregression (pVAR) in a generalized method of moments (GMM) framework. Since all variables typically are treated as endogenous (e.g., Holtz-Eakin, et al. 1988, Canova and Ciccarelli, 2013), it is regarded particularly suitable in our case given that the network density itself may be driven by productivity, population size and density, and labour flows. Thus, to understand the role of network density for regional productivity we need to assess how a number of different covariates co-evolve. In short, a pVAR model fits a multivariate panel regression of each dependent variable on lags of itself and of lags of all other dependent variables by means of GMM estimation through either first-differencing or forward orthogonal deviation. Instead of using deviations from past realizations of each variable, the latter deviation subtracts the average of all future observations, which also make past realizations valid instruments (see Love and Zicchino, 2006, Abrigo and Love, 2015 for further information).

The variables used in the model are the log of network density as defined above to capture the network effect. This variable is then estimated together with regional productivity, which is defined as regional per capita wages. This is motivated by the fact that wages tend to be held as the best available proxy for worker productivity (Feldstein, 2008), and because worker productivity tend to be expressed in higher regional wage levels (Combes et al., 2005, Kemeny and Storper, 2015). Since the potential for network formation may be driven by the turnover rates in regions, which in turn can be driven by the size of the region, we include two further variables reflecting these issues. PopDens is defined as the total number of employees per square kilometre in each region, while MobAcc is defined as the total number of job switches per region from the beginning of the investigated period until the observed year. Apart from potentially influencing the role of network density on regional productivity, both variables are also often held responsible for regional growth. Population density is as mentioned in previous sections often used as a proxy for regional socializing potential that drives knowledge spillovers (e.g., Storper and Venables, 2004, Ciccone and Hall, 1996, Glaeser, 1999), while job related mobility is regarded as a direct mean to transfer embodied knowledge between firms, which in turn tend to be productivity enhancing (Eriksson and Lindgren, 2009, Boschma et al, 2014). All variables are logged to reduce the impact of skewed distribution and we only model the years 1995-2008 since the network is not developed fully until after a couple of years as illustrated in Appendix 4.

The pVAR modelling requires the optimal lag order to be chosen for both the VAR specification and the moment conditions. This was calculated by using the first to third order lags for all variables together with lags 3-5 of each variable as instruments. Based on the model selection criteria we could conclude that a second-order pVAR is the preferred model since all tests (MBIC, MAIC and MQIC) were smallest for the second lag. Further, a key criterion of the pVAR is that the model needs to fulfil the stability condition. This was not the case when running the models on levels since at least one eigenvalue exceeded 1, thus indicating that a unit-root is present. To remedy that, we first-differenced all variables, which then produced stable estimates. By first-differencing we also remedy the influence of unobserved heterogeneity in the form of time-invariant regional-specific effects (see e.g., Coad and Broekel, 2012).

Table 4 presents the results of the pVAR models with two lags included and GMM estimation through forward orthogonal deviation using lags 3-5 as instruments. All models are estimated with cluster robust standard errors at regional level. Compared to Holtz-Eakin et al. (1988) we only use instruments with observations with valid instruments, thus omitting observations with missing values instead of substituting missing values with the value zero. The latter approach produced identical results but with slightly higher Hansen J statistics, which is an indication of overidentified restrictions.

Model 1 in Table 4 estimate the relation between productivity, population density, regional turnover and the density of the full network. Based on the first column, we can conclude that previous realisations of productivity are highly influential for explaining future realisations. This is expected and in line with previous studies in Sweden (e.g. Boschma, et al, 2014). As stated in previous studies (e.g., Storper and Venables, 2004; Ciccone and Hall, 1996) we also find that population density is positively influencing productivity, which points to the fact that density in itself may contribute to spillover effects. We can however not find a statistically significant relation indicating that high regional mobility per se would influence productivity, which confirms earlier studies stating that it is not mobility per se but the type of labour flows that positively influence productivity (e.g., Boschma, et al. 2014). Finally, and most importantly, both lags of network density are positively significant which points to the fact that given past realizations of both productivity and population density, network density has a strong and positive influence on productivity.

However, from the three following columns, it is evident that some of these variables are co-evolving. Both mobility and in particular network density is significantly negatively correlated to population density. This finding suggests that the social network is sparser in population-wise denser regions, which is evident because the number of potential links grows exponentially with population size, while the number of observed links does not. This finding and implications will be discussed more in detail in the following subsection. Further, mobility does explain the density of the co-worker network meaning that mobility indeed is a driver of link formation in the network. Thus, while mobility per se had no influence on productivity, it has an indirect effect mediated by the co-worker network. Productivity is on the other hand not causing any of the variables, except mobility, which indicates that very high levels of mobility is more related to low productivity that could be assumed to involve lay-offs rather than volunteered moves. The model suffers from over-identification, because too many instrument are used as it is indicated by the Hansen J statistic.

In the second model of Table 4, we have removed all the ties older than five years since strength of relationships may weaken over time (Burt, 2000, Jin et al, 2015). Edge removal is a reasonable method to solve the problem of link ageing because characteristics of social networks are better reproduced when old links are deleted as compared to keeping these ties (Murase et al, 2015)[4]. By removing all old, and thereby weak, ties the network effect becomes stronger and even the second lag of network density is positively significant at the 1% level as shown in column 1. Moreover, based on the output in the fourth column, we find no signs that any of the other variables influence the density of the network since the effects of both mobility and population density are no longer significant. Moreover, the Hansen J statistic is not significant, meaning that the model is not over-identified.

However, it may still be so that the network effects are partly driven by mobility because young and recent ties across plants are more likely to be preceded by labour flow between the same plants then old and weak ties that can connect employees who have moved across several plants after establishing the connection. . In other words, we need to explicitly handle the effect of mobility across plants on the co-worker network.

[4] The five years threshold for link deletion was chosen by measuring tie weights and using exponential time decay curves as explained in Eriksson and Lengyel (2015).

Table 4.

Co-worker network density and productivity growth, panel vector autoregressive models, 1995-2008.

	Model 1: Full network				**Model 2: Old ties excluded**			
	dRegProd	dPopDens	dMobacc	dNetDens	dRegProd	dPopDens	dMobAcc	dNetDens
L.dRegProd	0.354***	-0.433	-2.000	2.683	0.277**	-0.365	-3.266**	-0.918
	(0.123)	(0.865)	(1.633)	(1.699)	(0.135)	(0.880)	(1.584)	(1.725)
L2.dRegProd	-0.047	0.132	-3.385***	-0.026	-0.039	0.192	-2.852**	0.048
	(0.057)	(0.357)	(0.654)	(0.600)	(0.066)	(0.383)	(0.710)	(0.910)
L.dPopDens	-0.040**	1.040***	-0.128	0.265	-0.040*	1.040**	-0.398	0.193
	(0.020)	(0.232)	(0.222)	(0.196)	(0.023)	(0.283)	(0.293)	(0.236)
L2.dPopDens	0.059***	-0.236	-0.118	-0.280*	0.055**	-0.213	-0.051	0.028
	(0.018)	(0.207)	(0.245)	(0.167)	(0.021)	(0.282)	(0.252)	(0.099)
L.dMobAcc	-0.009	-0.029	0.030	0.278	-0.021	0.041	0.085	-0.081
	(0.011)	(0.069)	(0.158)	(0.210)	(0.013)	(0.065)	(0.156)	(0.272)
L2.dMobAcc	-0.000	-0.045*	-0.126**	0.117**	-0.007	-0.015	-0.178**	-0.189
	(0.004)	(0.023)	(0.063)	(0.055)	(0.005)	(0.024)	(0.067)	(0.210)
L.dNetDens	0.058***	-0.186***	0.181	0.083	0.077***	-0.224*	0.173	0.239
	(0.017)	(0.066)	(0.228)	(0.182)	(0.028)	(0.124)	(0.254)	(0.206)
L2.dNetDens	0.008*	-0.012	-0.050	0.014	0.015***	-0.032	0.038	0.060
	(0.005)	(0.024)	(0.065)	(0.053)	(0.004)	(0.022)	(0.043)	(0.069)
Hansen J	54.735**				36.807			
Stable	Yes				Yes			
N	792				735			

Note: Cluster robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

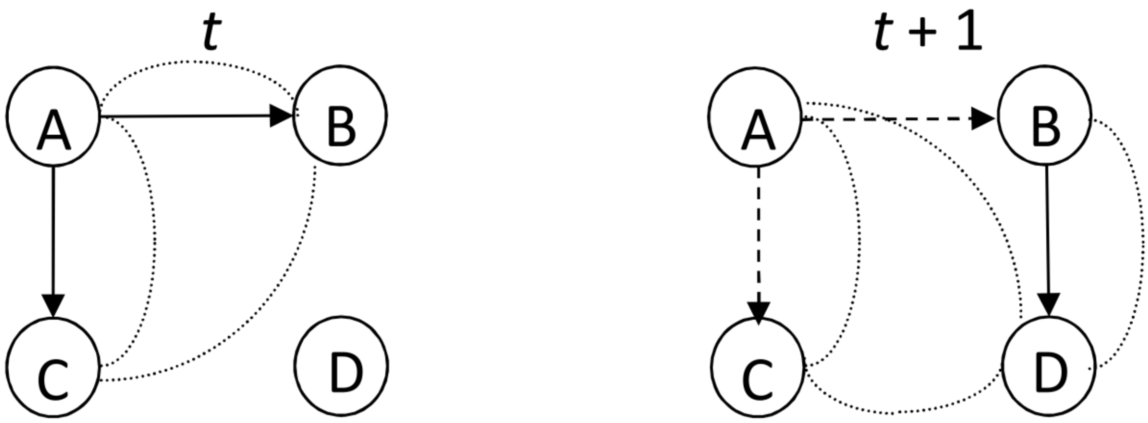
DENSITY AND LABOUR MOBILITY

Labour mobility has an influential effect on the co-worker network, because an employee establishes co-worker ties to distinct plants if she moves or if one of her colleagues moves across plants. Due to the above fact, labour mobility ties across plants in the region might have a strong influence on co-worker ties across plants in the region.

However, co-worker ties can be independent from labour mobility ties for two reasons: (1) co-worker ties can be established between plants with no previous labour flow; (2) previous labour flow does not necessarily mean subsisting co-worker ties across plants. For example, consider plant *A* that has at least three employees out of which employee *i* moves to plant *B* and employee *j* moves to plant *C* in time *t* (*B* and *C* have at least one employees before the arrival of *i* and *j*). Then, there will be co-worker ties between plants *A* and *B*, *A* and *C*. Additionally, there will be a co-worker tie between *B* and *C* without any employee moving from *B* to *C* or vice versa (Figure 3). Furthermore, if employee *i* moves from plant *B* to plant *D* in time *t*+1, then the link between *A* and *B* will disappear despite the previous labour flow.

Figure 3.

Labour mobility and co-worker ties across plants



Note: the solid arrow denotes actual mobility of 1 employee, the dashed arrow denotes previous mobility and dotted line denotes co-worker ties across plants.

To address how labour mobility across plants may influence the co-worker network, we first calculated the share of those plant-level co-worker links that were not preceded by any labour mobility between the certain plants, and then, for every year in our data, we repeated the calculation for the individual level co-worker network. Appendix 4 illustrates that the ratio of links without being preceded by mobility increases almost monotonically over time (e.g., 33% by year 1996 and almost reaches 50% by year 2008). This large and growing share of co-worker ties suggests that the co-worker network becomes increasingly independent of previous labour mobility. Zooming into regions we find that the bigger the region the larger the share of those individual co-worker links that were not preceded by mobility (Figure 4A). The effect of region size on the rate of co-worker links without being preceded by labour mobility becomes stronger and clearer over time: both the co-efficient and R2 of the linear fit increases.

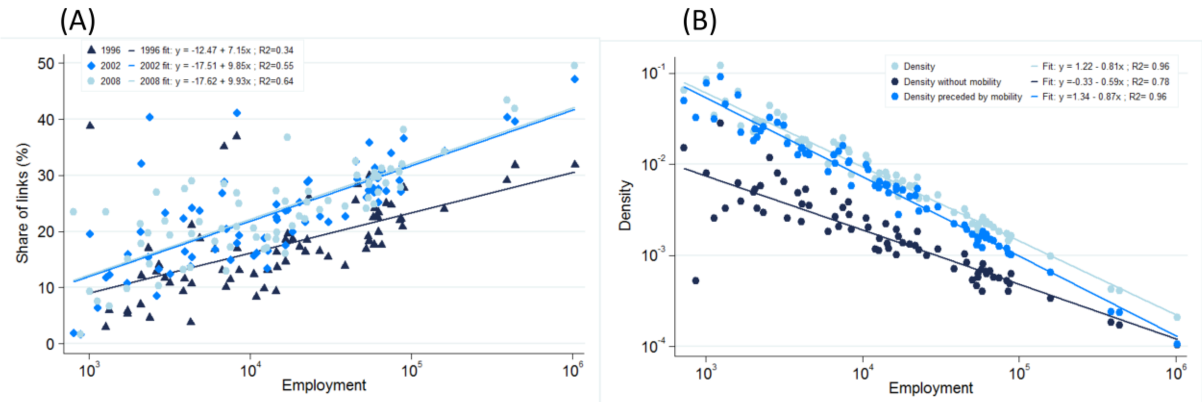
Then, we can decompose D_c into two segments: (1) in which inter-plant links have been preceded by labour mobility, and (2) in which links are present between plants without previous labour mobility. The formula for that is

$$D_c = \sum_{ab}^{l} \frac{2 \times L_{ab}}{N_a \times N_b} \times \frac{N_a \times N_b}{N \times (N-1) - \sum_a N_a \times (N_a - 1)} \times \delta_{ab}^{l} ; \qquad (6)$$

where L_{ab} is the number of observed links between plants a and b and $\sum_{ab} L_{ab}$ equals L; N_a and N_b are number of employees at plants a and b; l denotes the different network segments described above and δ_{ab}^{l} equals 1 if the ab link belongs to the respective segment and 0 otherwise. Consult Appendix 5 for a visual explanation of network density decomposition.

Figure4.

Mobility-independent co-worker links and density by size of the region



Note: (A) Region size and share of co-worker links not preceded by labour mobility, 1996-2002-2008. Size of the region was captured by the maximum number of employees in the region over the full period. (B) Density and density decomposition by size of the region in 2008 as described in Section 6.3.

We find that the log of network density is proportional to the log of the size of the region: the larger the region, the smaller the density (Figure 4B). This is an important finding because it suggests that the vast majority of possible regional links are actually not observed and that this share increases as the size of the region grows. Thus, the frequently accepted intuition that social networks are denser in densely populated areas than in sparsely populated areas is not true. Density is higher in small regions because there are less people and less possible links. Although there are much more observed links in big regions than in small regions, the number of possible links is higher with magnitudes, which produces low network density. The network segment in which co-worker ties have been preceded by labour mobility prevails in terms of contribution to overall density. However, the co-worker network segment without previous mobility is more and more apparent as the size of the region grows.

In Table 5 we present the results of the decomposed network indicator. Models 3 and 4, respectively, estimate the full network and the network were old ties are excluded. Since the decomposed network indicators are highly correlated, it is not possible to estimate their effect in the same model. Thus, in Models A we only estimate networks that are not preceded by mobility, while Models B show the results of ties preceded by mobility. Similar to the results in Model 1, we find in Model 3A that the lagged values of productivity, population density, as well as network density independent of mobility are positively correlated with productivity. Thus, the results indicate that inter-plant ties that are not directly preceded by mobility triggers productivity growth. Further, based on the findings in column 3 we can confirm previous evidence (e.g., Calvo-Arengol and Jackson, 2004, Granovetter, 1995) regarding social networks stimulation on mobility on the regional level.. As noted from the descriptives above, we also find that these types of networks are more prevalent in less densely populated regions.

Model 3B assesses the impact of network density preceded by mobility, we still find a positive effect of the network on productivity but also that mobility per se is hampering productivity. This latter finding points to the fact that, as shown in previous studies, it is not mobility per se that triggers productivity but the social ties created by mobility); lagged observations of mobility have positive effect on network density (column 8). However, the line of causality in this argument is not straightforward since there is a positive relation between productivity and network density (column 8).

Turning to the lower part of Table 5, Models 4A and 4B show the results when all old ties are excluded. We still find a positive, but much weaker, relationship between network density not preceded by mobility and productivity. It also shows that productivity is negatively associated with this type of network, pointing towards the fact that in less prosperous regions, co-worker ties not preceded by mobility tend to be more dense. Similar to the findings on the full network, strong ties preceded by mobility has a positive influence on productivity as well as the lags of productivity and population density while the effect of mobility is negative (column 5 in Model 4B). However, in contrast to the full network, network density is not significantly related to any other variable, neither when being on the right hand side in the model nor on the left hand side (column 8). Thus, based on our findings it appears that it is particularly strong ties that are preceded by labour flows that have the strongest influence on productivity.

A final observation regarding the results concerns the overall model fit, and the Hansen J statistic on the issue of over-identification. In general, all models on the full network seem to suffer from over-identification, meaning that too many instruments are used to being able to remove the endogenous components of the variables (Roodman, 2007). This is however less prevalent for Model 4A, where Hansen J has a lower significance, and it is not the case for Model 4B that estimates the strong ties preceded by mobility only, which means that these models can be considered to be the most robust. To remedy this particular problem we also ran the models on the full network when only using lags 3-4 as instrument (rather than lags 3-5 which were chosen to allow for a long time span between the observation and the instrument). This procedure did not influence the overall interpretation of the models, while the Hansen statistic turned insignificant.

Table 5.

Decomposed network density and productivity growth, panel vector autoregression models, 1995-2008

	Model 3A: Full network, without mobility				**Model 3B: Full network, with mobility**			
	dRegProd	dPopDens	dMobAcc	dNetDens Indep	dRegProd	dPopDens	dMobAcc	dNetDens Mob
L.dRegProd	0.518***	-0.445	-1.821	4.070	0.376***	-0.354	-1.638	3.063*
	(0.163)	(0.657)	(1.589)	(3.686)	(0.142)	(0.881)	(1.831)	(1.843)
L2.dRegProd	-0.001	0.122	-3.547***	-0.476	-0.020	0.138	-3.278***	0.115
	(0.062)	(0.346)	(0.873)	(1.091)	(0.060)	(0.356)	(0.671)	(0.695)
L.dPopDens	-0.025	1.159***	-0.247	0.649	-0.040*	1.036***	-0.197	0.215
	(0.019)	(0.215)	(0.197)	(0.439)	(0.021)	(0.231)	(0.241)	(0.204)
L2.dPopDens	0.032*	-0.241	0.038	-1.000**	0.060***	-0.219	-0.043	-0.247
	(0.018)	(0.223)	(0.204)	(0.390)	(0.019)	(0.211)	(0.253)	(0.170)
L.dMobAcc	-0.024	-0.021	0.085	0.421	-0.020*	-0.004	-0.051	0.215
	(0.021)	(0.061)	(0.215)	(0.479)	(0.011)	(0.064)	(0.164)	(0.207)
L2.dMobAcc	-0.006	-0.031*	-0.152**	-0.001	-0.001	-0.044*	-0.127**	0.101*
	(0.004)	(0.017)	(0.077)	(0.080)	(0.005)	(0.023)	(0.061)	(0.054)
L.dNetDensIndep	0.021**	-0.035	0.174*	0.000				
	(0.009)	(0.025)	(0.102)	(0.184)				
L2.dNetDensIndep	0.005	-0.011*	0.027	-0.090				
	(0.003)	(0.006)	(0.039)	(0.109)				
L.dNetDensMob					0.065***	-0.195***	0.207	-0.068
					(0.020)	(0.076)	(0.247)	(0.212)
L2.dNetDensMob					0.011**	-0.029	-0.020	-0.043
					(0.006)	(0.024)	(0.073)	(0.063)
Hansen J	51.984**				51.590**			
Stable	Yes				Yes			
N	792				792			

Note: Cluster robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 5.
continued from previous page

	Model 4A: Old ties excluded, without mobility				Model 4B: Old ties excluded, with mobility			
	dRegProd	dPopDens	dMobAcc	dNetDens IndepM5	dRegProd	dPopDens	dMobAcc	dNetDens MobM5
L.dRegProd	0.468***	-1.213**	-2.337**	0.819	0.369**	-0.929	-3.264*	-0.627
	(0.113)	(0.557)	(1.011)	(2.536)	(0.145)	(0.784)	(1.975)	(1.744)
L2.dRegProd	-0.180***	0.298*	-2.829***	-3.131***	-0.063	0.311	-2.922***	-0.557
	(0.048)	(0.398)	(0.875)	(1.141)	(0.069)	(0.383)	(0.857)	(1.047)
L.dPopDens	-0.067***	1.190***	-0.152	-0.239	-0.034	0.984***	-0.328	0.036
	(0.027)	(0.171)	(0.223)	(0.432)	(0.022)	(0.271)	(0.343)	(0.282)
L2.dPopDens	0.052**	-0.250	-0.165	-0.383	0.049***	-0.155	-0.089	-0.362
	(0.025)	(0.181)	(0.261)	(0.404)	(0.019)	(0.259)	(0.305)	(0.230)
L.dMobAcc	-0.021*	0.102*	-0.180	-0.436	-0.034**	0.056	0.162	0.159
	(0.012)	(0.060)	(0.127)	(0.285)	(0.016)	(0.067)	(0.183)	(0.233)
L2.dMobAcc	-0.006	-0.013*	-0.149***	-0.134	-0.008	-0.019	-0.165**	0.082
	(0.004)	(0.023)	(0.054)	(0.097)	(0.005)	(0.022)	(0.079)	(0.087)
L.dNetDensIndep	0.022*	0.007	0.016	-0.197				
	(0.012)	(0.051)	(0.174)	(0.231)				
L2.dNetDensIndep	0.002	-0.003	0.042	-0.142***				
	(0.003)	(0.010)	(0.040)	(0.055)				
L.dNetDensMob					0.073***	-0.147	0.207	0.267
					(0.024)	(0.092)	(0.250)	(0.229)
L2.dNetDensMob					0.014***	-0.021	0.005	-0.031
					(0.003)	(0.017)	(0.053)	(0.141)
Hansen J		50.44*				32.992		
Stable		Yes				Yes		
N		735				735		

Note: Cluster robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

CONCLUSION AND DISCUSSION

The paper provides the first systematic evidence that social networks are important for regional productivity growth. In order to establish that argument, a new way of constructing social networks (e.g. co-worker networks) from employee-employer co-occurrence databases was introduced. Then, we described the steps of the co-worker network construction for the entire economy of Sweden for the period 1990-2008 and demonstrated that this network can be considered as a spatially embedded social network, indeed. As a next step, we showed how labour mobility influences the co-worker network on the individual level. Then, panel vector autoregressive models were estimated in which the co-worker network density is used as main explanatory variable. The next model was built on a network from which all the old ties were excluded because these ties might be weak and therefore ineffective for productivity. In the final models network density was decomposed into those links that have been preceded by labour mobility and those that are independent from labour mobility.

A crucial finding implies that the constructed co-worker network is similar to other large-scale social networks. This makes us believe that the approach introduced in this paper can offer a wide variety of new answers for questions addressing the role of social networks in regional economic development. The co-worker network methodology opens up the possibility of employing a micro perspective, one can analyse networks aggregated on various levels including individuals, plants, firms or industries. The current paper however focused on three hypotheses: (1) there is a positive effect of co-worker network density on productivity growth; (2) the positive effect holds when the network contains strong ties only; (3) the positive effect holds even if we look at the segment of the co-worker network that cannot be observed by labour mobility.

People might learn more efficiently from those they have been in a co-worker relation with previously rather than from co-location *per se*. Thus, learning through the co-worker network is expected to enhance the productivity of the region. Indeed, our empirical analysis indicates that – along with population density – the density of the co-worker network is also important for regional productivity growth. This finding verifies our first hypothesis claiming that network density triggers productivity growth, and underlines the importance of related policy implications. For example, productivity gains shall motivate public authorities to develop such environments that encourage employees to establish more professional connections at workplaces and also trace them over their career.

In relation to our second hypothesis we do find that network density is triggering productivity if only those ties are considered that are younger than five years. In fact, the model becomes more stable when all the old ties have been excluded. These results correspond with the previous literature and intuition as well. The strong co-worker ties are more efficient when it comes to learning and productivity growth because co-located previous colleagues might communicate more if only short period of time passed since the end of their shared job.

Concerning the third hypothesis, we find that network density still has a positive effect, despite that links are not preceded by mobility. However, the most robust model is built on the co-worker network segment where links were preceded by mobility between the plants, whereas mobility itself does not trigger productivity growth. This finding confirms previous studies showing that regional job flows per se is not an economic blessing for regions since that may produce sunk-costs for both the involved firms and individuals unless the flows are between skill-related industries characterised by cognitive proximity (e.g., Boschma et al, 2014). These findings do however indicate the indirect influence of mobility since co-worker ties are indirectly driven by mobility. In this respect future studies could pay more attention to the different ties that are established between technologically related industries and whether the degree of social proximity may influence to what extent learning across related industries are present. It shall be noted in policy implications as well that recent attempts to make the labour market more flexible to facilitate mobility are not hitting the target since mobility only has an indirect effect.

Further research might devote attention to the effects of the co-worker network's structure on other aspects of regional dynamics like firm entry, investment flows, entrepreneurship or employment growth introducing sector-specific characteristics into the analysis. For example, employees might learn more in those co-worker networks where the industry-specific knowledge is easier to transfer. Another potential in the co-worker approach is calculating the tie strength instead of removing edges and one might be interested how the strength of weak ties – as Granovetter put it – applies to the effect of co-worker networks on innovation performance. Another aspect related to this study is whether these processes are shaped by the Swedish context or are more generalizable. For example, population density at the regional scale may not be a perfect indicator in the Swedish case due to the relatively sparse population distribution. Analysing the performance of industries or plants instead would not only open up for greater heteregoneity but also allows controlling for further aspects influencing performance, which are industry- or plant-specific. One might be interested to introduce co-worker network into regional economic growth frameworks (Huggins and Thompson, 2014) because we have to understand how such networks influence growth on the long run, which is missing from our approach. It might be also straightforward to use the new co-worker network for a closer look at the spatial dimension of information spreading and learning across firms (Jackson, 2008). Last but not least, we shall further develop our homophily-biased random network approach by introducing the effect of group diversity, time and triadic closure and fit the model to real social networks in firms. These future steps might increase the precision of link probability calculation, which is important for creating social networks from a wide variety of co-occurrence data.

REFERENCES

Abrigo, M.R.M., Love, I., 'Estimation of Panel Vector Autoregression in Stata: a Package of Programs', 2015, available at http://paneldataconference2015.ceu.hu/Program/Michael-Abrigo.pdf

Adamic, L., Adar, E., 'How to search a social network', Social Networks, 27, 2005, pp. 187-203.

Agrawal, A., Cockburn, I., McHalle, J., 'Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships', Journal of Economic Geography, 6, 2006, pp. 571-591.

Ahuja, G., 'Collaboration networks, structural holes, and innovation: A longitudinal study', Administrative Science Quarterly, 45, 2000, pp. 425-455.

Almeida, P., Kogut, B., 'Localization of knowledge and the mobility of engineers in regional networks', Management Science, 45, 1999, pp. 905-917.

Amin, A., 'Industrial districts'. In Sheppard, E., Barnes T.J. (eds.) A Companion to Economic Geography, Oxford: Blackwell Publishing, 2000, pp. 149-168..

Arellano, M., Bond, S., 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations', Review of Economic Studies, 58, 1991, pp. 277–297.

Asheim, B., 'Industrial districts as learning regions: a condition for prosperity', European Planning Studies, 4, 1996, pp. 379-400.

Balkundi, P., Harrison, D.A., 'Ties, leaders, and time in teams: strong inference about network structure's effects on team viability and performance', Academy of Management Journal, 49, 2006, pp. 49-68.

Bathelt, H., Glückler, J., 'Toward a relational economic geography', Journal of Economic Geography, 3, 2003, pp. 117-144.

Bathelt, H., Malmberg, A., Maskell, P., 'Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation', Progress in Human Geography, 28, 2004, pp. 31-56.

Beaman, L., Magruder, J., 'Who gets the job referral? Evidence from a social networks experiment', American Economic Review, 102, 2012, pp. 3757-3593.

Blau, P., 'Heterogeneity and Inequality: Towards a Primitive Theory of Social Structure'. New York: Free Press, 1977.

Blau, P., Blum, T., Schwartz, H., 'Heterogeneity and intermarriage', American Sociological Review, 47, 1982, pp. 45–62.

Blum, T.C., 'Structural constraints on interpersonal relations: a test of Blau's macrosociological theory', American Journal of Sociology, 91, 1985, pp. 511–521.

Borgatti, S.P., Cross, R., 'A Relational view of information seeking and learning in social networks', Management Science, 49, 2003, pp. 432-445.

Borgatti, S.P., Mehra, A., Brass, D.J., Labiance, G., 'Network Analysis in the Social Sciences', Science, 323, 2009, pp. 892-895.

Boschma, R.A., ‘Proximity and innovation: a critical assessment’, Regional Studies, 39, 2005, pp. 61-74.

Boschma, R.A., Eriksson, R., Lindgren, U., ‘How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity’, Journal of Economic Geography, 9, 2009, pp. 169-190.

Boschma, R., Eriksson, R.H., Lindgren, U., ‘Labour market externalities and regional growth in Sweden: the importance of labour mobility between skill-related industries’, Regional Studies, 48, 2014, pp. 1669-1690

Boschma, R.A., Frenken, K., ‘The emerging empirics of evolutionary economic geography’, Journal of Economic Geography, 11, 2011, pp. 295-307.

Brass, D.J., ‘Men’s and women’s networks: a study of interaction patterns and influence in an organization’, Academy of Management Journal, 28, 1985, pp. 327-343.

Breschi, S., Lissoni, F., ‘Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows’, Journal of Economic Geography, 9, 2009, pp. 439-468.

Buhai, S., van der Lei, M., ‘A social network analysis of occupational segregation’. Tinbergen Institute Discussion Paper, No. 06-016/1, 2006.

Burt, R. L., ‘Structural Holes’. Cambridge, MA: Harvard University Press, 1992.

Burt, R.L., ‘Decay functions’, Social Networks, 22, 2000, pp. 1-28.

Calvo-Armengol, A., Jackson, M.O., ‘The Effects of Social Networks on Employment and Inequality’, American Economic Review, 94, 2004, pp. 426–54.

Canova, F., Ciccarelli, M., ‘Panel Vector Autoregressive Models: A Survey’. European Central Bank Working Paper Series No. 1507, 2013.

Ciccone, A., Hall, R.E., ‘Productivity and the Density of Economic Activity’, American Economic Review, 86, 1996, pp. 54-70.

Coad, A., Broekel, T., ‘Firm growth and productivity growth: Evidence from a panel VAR’, Applied Economics, 44, 10, 2012, pp. 1251-1269

Coleman, J. S., ‘Foundations of Social Theory’. Cambridge, MA: Harvard University Press, 1990.

Combes, P., Duranton, G., Overman, H., ‘Agglomeration and the adjustment of the spatial economy’, Papers in Regional Science, 84, 2005, pp. 311–349.

Currarini, S., Jackson, M.O., Pin, P., ‘An Economic Model of Friendship: Homophily, Minorities, and Segregation’, Econometrica, 77, 2009, pp. 1003-1045

Collet, F., Hedström, P., ‘Endogenous tie formation mechanisms in a directed network generated by employee mobility’. IFAU Working Paper, 25, 2012.

Dahl, M.S., Pedersen, C.O.R., ‘Knowledge flows through informal contacts in industrial clusters. Myth or reality?’, Research Policy, 33, 2003, pp. 1673–1686.

Duranton, G. and Puga, D., ‘Micro-foundations of urban agglomeration economies’. In: Henderson J.V., Thisse J.V. (eds.) Handbook of Regional and Urban Economics, vol. 4, Amsterdam: North-Holland, 2004, pp. 2063–2117..

Erdős, P., Rényi, A., ‘On random graphs’, Publicationes Mathematicae, 6, 1959, pp. 290-297.

Eriksson, R., Lengyel, B., ‘Co-worker networks and local knowledge externalities. Paper presented at NetWorkShop, Pécs University, Pécs, 2015. Available at http://ktk.pte.hu/sites/default/files/hir_mellekletek/2015/07/eriksson-lengyel_2015june22.pdf

Eriksson, R., Lindgren, U., ‘Localized mobility clusters: impacts of labour market externalities on firm performance’, Journal of Economic Geography, 9, 2009, pp. 33-53.

Eriksson, R., Lindgren, U., Malmberg, G., 'Agglomeration mobility: effects of localisation, urbanisation, and scale on job changes', Environment and Planning A, 40, 2008, pp. 2419–2434.

Feld, S.L., 'Social structural determinants of similarity among associates', American Sociological Review, 47, 1982, pp. 797–801.

Feldman, M.P., 'The new economics of innovation, spillovers and agglomeration: a review of empirical studies', Economics of Innovation and New Technology, 8, 1999, pp. 5-25.

Feldstein, M., 'Did wages reflect growth in productivity?', Journal of Policy Modeling 30, 2008, pp. 591–594.

Frenken, K., Van Oort, F.G., Verburg, T., 'Related variety, unrelated variety and regional economic growth', Regional Studies, 41, 5, 2007, pp. 685–697.

Glaeser, E. L., 'Learning in cities', Journal of Urban Economics, 46, 1999, pp. 254-277.

Glaeser, E.L., 'The New Economics of Urban and Regional Growth'. In: Clark, G.L., Gertler, M.S., Feldman, M. (eds.) The Oxford Handbook of Economic Geography. Oxford: Oxford University Press, 2000 , pp: 83-98

Glitz, A., 'Co-worker networks in the labour market. CESifo Working Paper Series No. 4250, 2013. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2274613

Glückler, J., 'Economic geography and the evolution of networks', Journal of Economic Geography, 7, 2007, pp. 619-634.

Granovetter, M., 'The strength of weak ties', American Journal of Sociology, 78, 1973, pp. 1360-1380.

Granovetter, M., 'Getting a Job: a Study of Contacts and Careers'. Chicago: The University of Chicago Press, 1995.

Hensvik, L., Nordström Skans, O., 'Social networks, employee selection and labor market outcomes'. IFAU Working Paper, 2013: 15.

Holtz-Eakin, H.H., Newey, W., Rosen, H.S. 'Estimating vector autoregressions with panel data', Econometrica, 56, 6, 1988, pp. 1371-1395

Huber, F., 'Do clusters really matter for innovation practices in Information Technology? Questioning the significance of technological knowledge spillovers', Journal of Economic Geography, 12, 2012, pp. 107-126

Huggins, R., Thompson, P., 'A Network-based view on regional growth'. Journal of Economic Geography, 14, 3, 2014, pp. 511-545

Jackson, M.O., 'Social and Economic Networks'. Princeton, NJ: Princeton University Press, 2008.

Jin, E.M., Girvan, M., Newman, M.E.J., 'Structure of growing social networks'. Physical Review E, 64, 2004, 046132. doi: 10.1103/PhysRevE.64.046132

Kossinetz, G., Watts, DJ., 'Empirical analysis of an evolving social network', Science, 311, 2006, pp. 88-90.

Lengyel, B., Eriksson, R., 'Co-worker networks and productivity growth in regions'. Papers in Evolutionary Economic Geography #15.13., 2015. Available at https://peeg.wordpress.com/2015/05/13/15-13-co-worker-networks-and-productivity-growth-in-regions/

Lengyel, B., Varga, A., Ságvári, B., Jakobi, A., Kertész, J., 'Geographies of an online social network'. arXiv preprint, 2015, available at http://arxiv.org/abs/1503.07757?context=physics.soc-ph.

Lincoln, J.R., Miller, J., 'Work and friendship ties in organizations: a comparative analysis of relation networks', Administrative Science Quarterly, 24, 1979, pp. 181-199.

Love, I., Zicchino, L., 'Financial Development and Dynamic Investment Behaviour: Evidence from Panel VAR'. The Quarterly Review of Economics and Finance, 46, 2, 2006, pp. 190-210

Malmberg, A., 'Industrial geography: location and learning', Progress in Human Geography, 21, 1997, pp. 573-582.

Marshall, A., 'Principles of Economics – An Introductory Volume'. London: MacMillan, 1920.

Maskell, P., Malmberg, A., 'Localised Learning and Industrial Competitiveness', Cambridge Journal of Economics, 23, 1999, pp. 167-185.

McPherson, M., Popielarz, P.A., Drobnic, S., 'Social networks and organizational dynamics', American Sociological Review, 57, 1992, pp. 153-170.

McPherson, M., Smith-Lovin, L., 'Homophily in voluntary organizations: status distance and the composition of face-to-face groups', American Sociological Review, 52, 1987, pp. 370-379.

McPherson, M., Smith-Lovin, L., Cook, J.M., 'Birds of a feather: homophily in social networks', Annual Review of Sociology, 27, 2001, pp. 415-444.

Morrison, E.W., 'Newcomers relationship: the role of social network ties during socialization', The Academy of Management Journal, 45, 2002, pp. 1149-1160.

Murase, Y., Jo, H-H., Török, J., Kertész, J., Kaski, K., 'Modeling the role of relationship and breakup in social network formation', PLOS ONE, 10, 7, 2015: e0133005. doi:10.1371/journal.pone.0133005

Neffke, F., Henning, M., 'Skill relatedness and firm diversification', Strategic Management Journal, 34, 2013, pp. 297-316.

Saygin, P.O., Weber, A., Weynandt, M.A., 'Coworkers, networks, and job search outcomes'. IZA Discussion Paper No. 8174, 2014.

Roodman, D., 'A Short Note on the Theme of Too Many Instruments'. Center for Global Development Working paper No 125, August 2007.

Sias, P.M., Cahill, D.J., 'From co-workers to friends: the development of peer friendship in the workplace', Western Journal of Communication, 62, 1998, pp. 273-299.

Sorensen, O., 'Social networks and industrial geography', Journal of Evolutionary Economics, 13, 2003, pp. 513-527.

Sorensen, O., Rivkin, J. W., Fleming, L., 'Complexity, networks and knowledge flow', Research Policy, 35, 2006, pp. 994-1017.

Storper, M., Venables, A.J., 'Buzz: face-to-face contact and the urban economy', Journal of Economic Geography, 4, 2004, pp. 351-370.

Ter Wal, A., Boschma, R.A., 'Applying social network analysis in economic geography: framing some key analytic issues', Annals of Regional Science, 43, 2009, pp. 739-756.

Timmermans, B., Boschma, R.A., 'The effect of intra- and inter-regional labour mobility on plant performance in Denmark: the significance of related labour inflows', Journal of Economic Geography, 14, 2013, pp. 289-311.

Walker, G., Kogut, B., Shan, W., 'Social Capital, Structural Holes and the Formation of an Industry Network', Organization Science, 8, 1997, pp. 109-125.

Wasserman, S., Faust, K., 'Social Network Analysis: Methods and Applications'. Cambridge: Cambridge University Press, 1994.

Appendix 1A.

Categories of employee education by direction of studies

			1990		2008		1990	2008
		code	N	%	N	%	%	%
1	Pedagogy and teaching	14	107,853	29,441	168,497	21,44879	29,44	21,45
2	Arts and media	21	5.100	1.392165	12.018	1.529829	6.91	5.84
	Journalism and media	32	3.491	0.95295	11.053	1.40699		
	Humanities	22	16.725	4.565481	22.825	2.905504		
3	Social sciences	31	27.273	7.444805	47.950	6.103786	22.43	21.40
	Business. trade and administration	34	40.262	10.99046	92.489	11.77337		
	Law	38	14.640	3.996331	27.662	3.521229		
4	Biology and environment	42	1.821	0.497085	9.571	1.218339	4.54	6.08
	Physics and chemistry	44	3.191	0.871058	10.265	1.306681		
	Mathematics	46	9.381	2.560764	10.637	1.354035		
	Data	48	2.256	0.615828	17.288	2.200673		
5	Engineering	52	36.910	10.07545	105.734	13.45939	14.68	18.09
	Manufacturing	54	1.476	0.402909	4.072	0.518344		
	Construction	58	10.915	2.979505	23.481	2.989009		
	Agriculture and forestry	62	2.835	0.77388	5.767	0.734109		
	Environmental protection	85	467	0.127479	1.828	0.232695		
	Transport services	84	1.175	0.320744	1.265	0.161028		
6	Animal care	64	807	0.22029	1.865	0.237405	21.00	24.37
	Health care	72	58.451	15.95557	151.420	19.27498		
	Social work	76	17.647	4.817162	36.679	4.669046		
	Personal services	81	42	0.011465	1.472	0.187378		
0	Security and military	86	52	0.014195	3.634	0.462589	0.99	2.77
	Unknown	99	3.566	0.973423	18.106	2.3048		
	SUM		366.336	100	785.578	100	100.00	100.00

Note: Employees with educational background code 0 are excluded from the analysis.

Appendix 1B.

Number of employees by gender categories

Gender	1990	2008
0	182874	451303
1	183462	334275
SUM	366336	785578

Appendix 1C.

Number of employees by age categories

Age	1990	2008
-34	79437	217813
35-49	201334	317635
50-	85565	250130
SUM	366336	785578

Appendix 2.

Cumulative degree distribution in 2008, individual level network

Degree	Number of employees	Rate (%)
< 10	133,967	19.2
< 20	208,255	29.9
< 40	323,033	46.4
< 60	423,128	60.8
< 80	500,711	71.2
< 100	558,777	80.2
< 200	678,637	97.5
SUM	696,354	100

The distribution implies that almost two-third of the employees have less connections than the average degree; 80% of employees have less than 100 connections and only 2.5% of employees have more than 200 connections.

Appendix 3.

Descriptive statistics and correlation of variables, 1991-2008

		N	Mean	St.Dev	Min	Max	Pairwise Pearson correlation, pooled							
RegProdCap	Productivty defined as gross income per capita (log)	1296	7.855	0.104	7.638	8.233								
PopDens	Population density	1296	22.169	27.911	0.241	147.701	0.392							
MobAcc	The number of job changes accumulated over time(log)	1296	5.904	2.226	0	12.367	0.655	0.709						
NetDens	Network density (log)	1296	-4.839	1.415	-9.497	0.154	-0.391	-0.756	-0.858					
NetDensMob	Network density, preceded by mobility (log)	1296	-5.056	1.477	-9.579	0.154	-0.431	-0.767	-0.881	0.997				
NetDensIndep	Network density, not preceded by mobility (log)	1296	-6.523	1.789	-12.037	0	-0.209	-0.526	-0.621	0.783	0.763			
NetDensM5	Network density, ties older than 5 years excluded (log)	1293	-1.007	0.611	-3.013	1.286	-0.620	-0.602	-0.707	0.785	0.797	0.522		
NetDensMobM5	Network density, preceded by mobility, ties older than 5 years excluded (log)	1293	-1.043	0.615	-3.034	1.286	-0.620	-0.588	-0.688	0.763	0.779	0.499	0.993	
NetDensIndepM5	Network density, not preceded by mobility, ties older than 5 years excluded (log)	1246	-2.464	0.723	-4.901	0.637	-0.257	-0.445	-0.409	0.623	0.592	0.790	0.623	0.565

Note: all correlation coefficients are significant at the 0.01 level.

Appendix 4.

The share of individual ties with or without labour mobility links across plants, 1991-2008

Year	Plant level			Individual level		
	Only co-worker link (%)	Co-worker link preceded by mobility (%)	Number of links	Only co-worker link (%)	Co-worker link preceded by mobility (%)	Number of links
1991	57.6	42.4	63,016	8.8	91.2	1,119,684
1992	70.2	29.8	160,299	16.9	83.1	2,084,934
1993	75.1	24.9	241,860	21.7	78.3	2,691,104
1994	79.8	20.2	379,556	26.1	73.9	3,554,774
1995	83.3	16.7	560,507	29.8	70.2	4,514,635
1996	85.4	14.6	761,416	32.9	67.1	5,682,175
1997	87.1	12.9	986,641	35.4	64.6	6,640,262
1998	88.7	11.3	1,111,434	39.1	60.9	6,633,685
1999	89.8	10.2	1,378,353	41.6	58.4	7,711,355
2000	90.8	9.2	1,838,224	43.2	56.8	9,624,640
2001	91.6	8.4	2,237,292	45.5	54.5	11,103,743
2002	92.0	8.0	2,498,506	46.6	53.4	12,172,480
2003	92.3	7.7	2,744,254	47.0	53.0	13,266,549
2004	92.5	7.5	2,935,742	47.9	52.1	13,895,050
2005	99.2	0.8	2,995,758	47.7	52.3	15,187,785
2006	99.2	0.8	3,332,845	47.4	52.6	16,586,603
2007	93.0	7.0	4,232,703	47.7	52.3	19,411,643
2008	93.6	6.4	4,623,753	49.0	51.0	20,855,161

Appendix 5.

Density decomposition

Consider an adjacency matrix of 12 employees working for plants a, b, and c, in which X denotes if there is a connection between employees. Because co-worker ties are non-directed, we see exactly the same pattern on both size of the matrix diagonal. Then, the density of the network is twice the observed number of connections over the number of possible connections. In this case it equals: 2*10/12*11=0.152.

However, because only inter-plant ties can be observed in the co-worker network and one has to eliminate those employee-employee pairs that are within plant borders. Thus, the number of possible ties decreases and density grows: 2*10/(12*11)-(3*2+4*3+5*4)=20/94=0.213.

		a			b				c				
		1	2	3	4	5	6	7	8	9	10	11	12
a	1				X	X							
	2								X	X			
	3							X					
b	4	X											
	5	X											
	6										X		X
	7			X					X	X	X		
c	8		X					X					
	9		X					X					
	10						X	X					
	11												
	12						X						

Density of the matrix can be decomposed to the sum of the densities in its submatrices weighted by the proportion of the submatrix size to the full matrix size. We can write the decomposition of density in the sequence of $a \times b$, $a \times c$, $b \times c$ submatrices:

0.213 = {(2*3)/(4*3)*(4*3)/ 94} + {(2*2)/(5*3)*(5*3)/ 94}+ {(2*5)/(4*5)* (4*5)/ 94} = 6/94+4/94+10/94 = 0.064+0.043+0.106

Let us assume (in accordance with Figure 5) that labour mobility occurred previously between plants a and b, between a and c, but there was no mobility between b and c. Consequently, the density of the mobility-dependent segment is 0.107 (aggregate of $a \times b$ and $a \times c$ submatrix densities) and the density of the mobility-independent segment is 0.106 (density of $b \times c$ submatrix).