Papers in Evolutionary Economic Geography

15.13

Co-worker networks and productivity growth in regions

Balázs Lengyel & Rikard H. Eriksson

Utrecht University
Urban & Regional research centre Utrecht

Balázs Lengyel [a,b], Rikard H. Eriksson [c]

[a] Centre for Economic and Regional Studies, Hungarian Academy of Sciences, Budaörsi út 45, 1112 Budapest, Hungary. Corresponding author: lengyel.balazs@krtk.mta.hu

[b] International Business School Budapest, Záhony utca 7, 1031 Budapest, Hungary

[c] Department of Geography and Economic History, Umeå University, SE-901 87 Umeå, Sweden

Co-worker networks and productivity growth in regions

This version: May 5, 2015

Abstract

This paper provides a new empirical perspective for analysing the role of social networks for 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. We argue that these personal acquaintances are important for local learning opportunities and consequently for regional growth. Indeed, the paper provides the first systematic evidence for a central claim in economic geography: social network density has positive effect on regional growth defined as productivity growth. Interestingly, the most robust effect of density on growth was found in a segment of the co-worker network in which plants have never been linked by labour mobility previously.

JEL codes: D85, J24, J61, R11, R23

Keywords: social network, random network with homophily bias, probability of tie, labour mobility, regional productivity growth, panel regression

Acknowledgment: 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).

1. 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. 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 to what extent co-worker networks influence productivity growth in 72 Swedish labour market regions 1990-2008. 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 and analyse the effect of the network on regional dynamics. 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 as compared to the technological externalities assumed to be residing "in the air" of agglomerations (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 but that 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 regional productivity growth. Interestingly, we find the most robust effect of density on productivity growth in a segment of the co-worker network, in which plants have never been linked by labour mobility previously. This finding implies that the effect of co-worker network density is independent of labour flow networks.

2. Literature and hypotheses

The spatial dimension of network-related learning is a core interest of economic geography (Bathelt and Glückler, 2003). 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, 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. The argument stresses two points: first, the network density is very important for regional productivity growth as it is claimed in the first hypothesis; and second, the co-worker network becomes more and more independent from labour mobility networks over time, which is increasingly true for large regions, and provides the ground for our second hypothesis.

Regional productivity growth has been repeatedly found to depend on population density. 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). 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). 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. 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 productivity growth.

H1: Density of the local co-worker network enhances regional productivity growth.

The hypothesis is not only a further step in understanding spatial learning processes, it also 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, 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). On the other hand, weak ties and the presence of structural holes among separated sub-networks offers access to new information and combination of non-redundant knowledge can lead to radical innovations (Ahuja, 2000). Due to space limitations, we put our focus on density in paper rather than the issue of tie strengths and structural holes.

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.

Despite the methodological differences, our co-worker approach is closely connected to the labour mobility approach and we assume that former colleagues maintain their relations even after moving from one workplace to another, which is a proposition often made in labour economics and in evolutionary economic geography as well (Boschma and Frenken, 2011). Despite the lasting characteristics of co-inventors have been found important for later patenting collaborations (Agrawal et al, 2006, Breschi and Lissoni, 2009), this 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 productivity 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.

H2: Co-worker network density enhances regional productivity growth even if the ties across plants have not been preceded by labour flows among the concerned plants.

3. 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 L_{ij} 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_{ij}^{r}) by

$$P_{ij}^{r} = \frac{\ln N}{N}; \qquad (1)$$

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 N_m 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 ∈ $\{1, 2, \dots M\}$ 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.

4. 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.

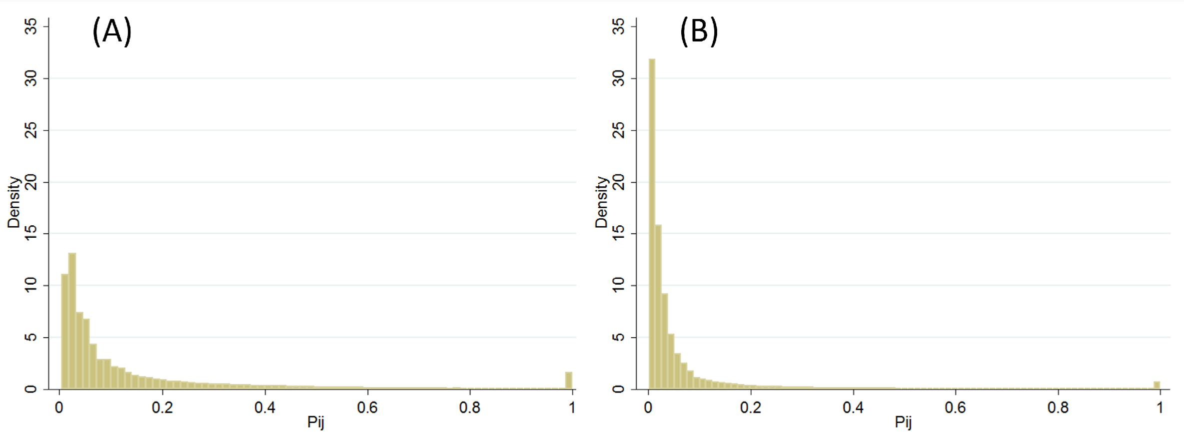
For practical reasons, and in order to keep the size of the sample at the limit the analysis can handle, we exclude those without tertiary education from the data. Including all employees would exponentially increase computation demand without contributing much to the analysis. This is motivated by the 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.

Figure 1. Distribution of tie probability, 1990 and 2008



Note: distribution for 1990 in (A) and distribution for 2008 in (B).

Figure 1 illustrates that distribution of P_{ij} is left skewed towards zero and decreases monotonously in both 1990 and 2008. However, one can observe that the distribution is more left skewed in 2008 than in 1990 because plants are larger in 2008, which produces lower probabilities. Density of P_{ij} is relatively high at 1 because we set the upper limit there. Nevertheless, the probability that the tie is established is very low for the vast majority of employee co-occurrences.

Employee co-occurrence is exponentially higher in large plants than in small plants and our aim is to find a reasonable number of ties per person, which can be handled by the analysis. There is no clear suggestion in the literature in this regard. 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 each other 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.

Evidently, co-occurrences are more likely to be real social ties in small plants and are less likely in large plants (see reports on P_{ij} distribution according to plant size categories in 1990 and 2008 in Appendix 2). Since P_{ij} distribution is similar at the first and last years of the sample, we identify the number of ties per person on base of 1990 network characteristics and apply that number consequently for upcoming years.

Table 2. Tie and degree distribution and isolates at P_{min} thresholds, 1990

Size category	Number of Employees	Plants	Mean plant size	Ties above P>0	Ties above P≥0.1	Ties above P≥0.2	Ties above P≥0.3	Ties above P≥0.4	Avg. Degree, P ≥0	Avg. Degree, P ≥0.1	Avg. Degree, P ≥0.2	Avg. Degree, P ≥0.3	Avg. Degree, P ≥0.4	Isolates, P≥0.2	Isolates, P≥0.3	Isolates, P≥0.4
2-9	71,794	19,033	4.88	139,418	139,418	139,418	132,473	128,624	3.88	3.88	3.88	3.69	3.65	0	0	1,468
10-19	46,249	3,420	14.10	302,931	302,931	286,457	280,517	258,711	13.10	13.10	12.38	12.13	11.19	0	2	12
20-49	78,175	2,531	33.24	1,260,292	1,207,405	1,106,538	923,647	722,915	32.24	30.88	28.31	23.63	18.52	0	6	127
50-99	63,102	949	69.11	2,148,933	1,952,091	1,458,559	970,228	592,219	68.11	61.87	46.23	31.01	20.81	5	525	6,172
100-249	34,608	245	151.25	2,600,067	1,851,674	90,533	449,126	252,907	150.25	107.01	55.04	33.83	25.09	1,711	8,063	14,456
250-499	16,831	49	355.47	2,983,041	1,101,615	29,688	156,692	101,567	354.47	133.30	60.59	43.97	35.21	7,032	9,704	11,062
500-999	15,414	24	671.37	5,166,533	73,522	244,708	125,972	76,328	670.37	135.25	79.69	59.47	46.30	9,273	11,178	12,117
1000-	13,553	11	1243.98	8,423,092	659,004	175,234	86,453	84,235	1242.98	182.23	92.52	65.67	65.60	9,765	10,920	10,985
Sum	339,726	26,262		23,024,307	7,287,660	3,531,135	3125108	2,217,506						27,786	40,398	56,399

Note: The high number of isolates for the smallest firms (2-9 employees) at P≥0.4 is due to those firms of two employees, in which the co-workers are not similar in any characteristics, and therefore P_{ij}=ln(2) /2=0.37.

Table 2 shows how the number of co-occurrence changes according to plant size categories when excluding employee pairs under certain P_{ij} minimum threshold. The number of co-occurrences falls dramatically in large plant categories but remains quite stable in small firm categories. We calculated average degree in order to see how many ties an employee has according to plant size categories and also the number of isolates that the P_{ij} threshold generates. The average number of ties is stable until large P_{min} values in very small plants as well as the average degree, and number of isolated employees are very low until the P_{min}=0.4 threshold in plants smaller than 50. This is a large P_{min} threshold and suggests that we can use a 50 best friends approach because everyone might know everyone in small plants. We thus simplify our task and look only at the most likely 50 co-workers of every employee in large plants.

Accordingly, we rank employee pairs based on their P_{ij} values. In case employee pairs have the same probability, we rank those with same educational background and smaller age difference higher, respectively. 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.

Table 3. Average degree of plants and individual in the co-worker network, 1991-2008

Year	Nodes	Avg. Degr. Plants	Avg. Degr. Ind.
1991	31,391	8.15	71.20
1992	46,445	11.89	89.72
1993	53,599	14.46	100.37
1994	63,299	17.87	112.28
1995	71,513	22.03	126.23
1996	79,499	26.04	142.92
1997	87,072	29.96	152.50
1998	87,950	32.77	150.82
1999	95,080	36.89	162.19
2000	107,423	42.71	179.18
2001	115,948	47.69	191.51
2002	120,026	51.25	202.81
2003	127,355	52.86	208.32
2004	132,791	54.02	209.27
2005	140,042	55.77	216.89
2006	148,318	58.27	223.65
2007	159,529	64.12	243.35
2008	166,109	67.12	251.09

As a result of the above selection process, there are 49,630,691 employee pairs that we trace over 19 years creating a balanced panel of pairs. From the total number of 942,983,129 rows in the panel, we exclude those that have not been appeared in the data yet (481,973,234 pairs), those when at least one employee is already above 65 years of age (42,016,069 pairs). Finally, we excluded those pairs, when either one or both individuals are not present in the labour market for unknown reasons (95,689,892 pairs) and those cases when the employees work in the same plant (167,632,360 pairs).

The remaining unbalanced panel of 155,671,574 employee pairs constitute a dynamic co-worker network over the 1990-2008 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. For example, after aggregating the network on the plant level, we observe that the number of plants in the network increases over the full period (Table 3, Column 2). As a result, both the number of plants an average plant is connected to (Table 3, Column 3) and the number of individual links from an average plant to any other plants (Table 3, Column 4) increases monotonically.

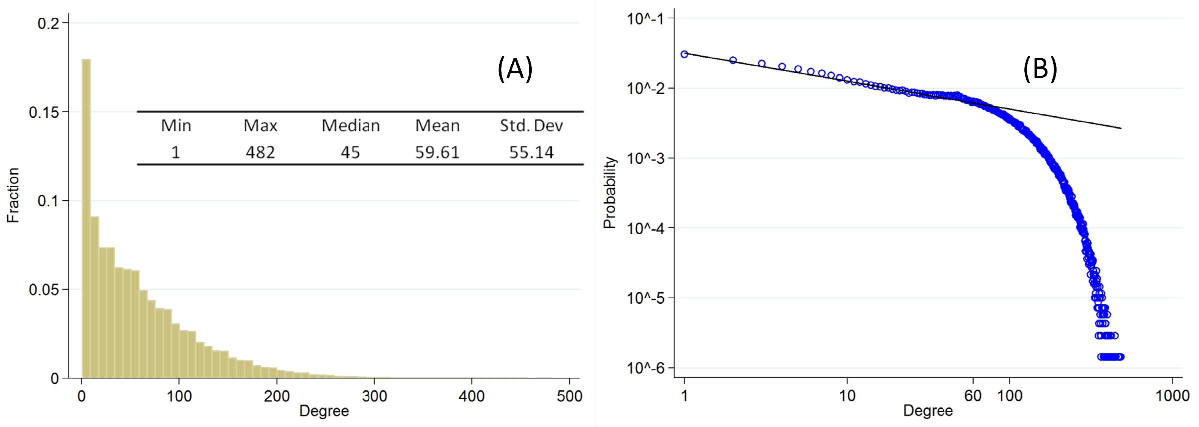
5. Properties of the co-worker network

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.

5.1 Degree distribution

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 2. Degree distribution and summary statistics of the individual level network, 2008



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

The histogram of degrees on a natural scale resembles a negative exponential distribution, where the fraction of nodes decreases monotonically as degree grows (Figure 2A). 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[1].

The degree distribution in 2008 illustrated on a log-log scale (Figure 2B) 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 3.

5.2 Geography

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 4). 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 290 Swedish municipalities. When we aggregate the network on the plant level we find a very similar pattern.

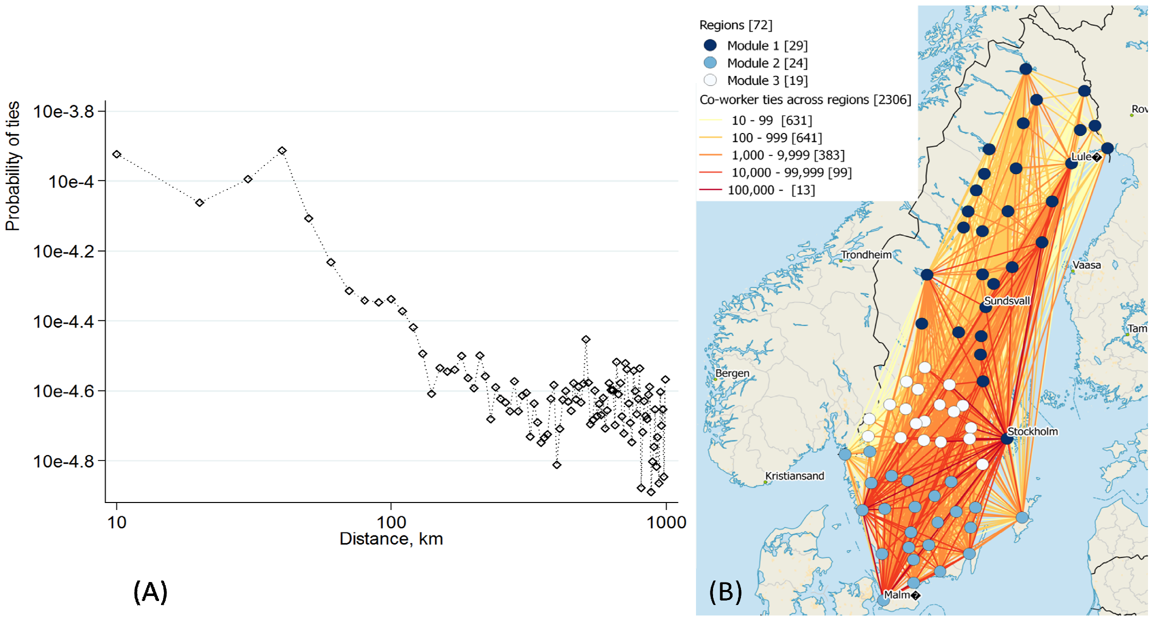
Table 4. Number of ties within regional borders, 2008

	Number of links	
	Individual level	Plant level
Within municipality (N=290)	7,826,977	1,470,603
Within functional region (N=72)	14,066,872	3,170,695
SUM	20,855,160	5,574,879

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 4). 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.

[1] 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.

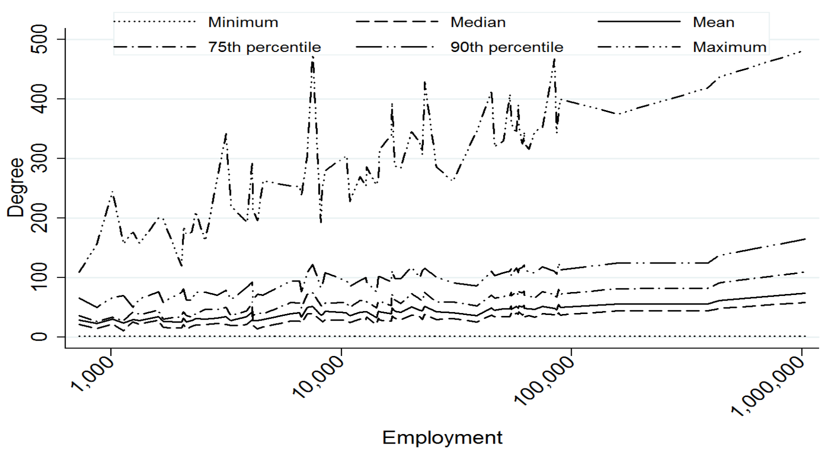
Figure 3. Geography of the co-worker network, 2008



Note: (A) The effect of distance on probability of ties. (B)Number of co-worker ties across Swedish functional regions. Same coloured nodes belong to same network module; modularity index is 0.074. Edges below the threshold number of links=10 are filtered out.

The spatial-base of the co-worker network across functional regions is very plausible when the strength of tie between two regions is the number of individual co-worker links (Figure 3B). Not surprisingly, Stockholm (the capital city region) is the centre of the interregional co-worker network meaning that the city has many individual-level ties to other regions. One can also find that Northern regions are very loosely connected with the exception of coastal towns like Umea or Lulea and the network is denser in the South than in the North. The Louvain community detection algorithm finds three modules that clearly represents a spatial divide in the co-worker network meaning that an employee in the South is more likely to know another employee in the South than in the Centre or in the North. Interestingly, Stockholm belongs to the Northern part in the network, which is probably due to a higher share of mobility from the North to the capital compared to mobility from Southern regions to the capital (Eriksson and Lindgren, 2009).

Figure 4. Degree distribution and region size



The degree distribution does however not only depend on region size. We have plotted the minimum, median, mean, 75^{th} percentile, 90^{th} percentile and maximum values of degrees against the number of employees in the region. Figure 6 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.

6. Co-worker network and labour mobility

Labour mobility is considered one of the major factors behind co-worker networks (Collet and Hedström, 2012). Therefore, we show in this section how labour mobility influences individual degree and network density.

6.1 Labour mobility and degree

There are three effects 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 we project degree distribution on any of the above variables, we find that degree grows as years in career, average plant size, job changes and movements increase (Appendix 4). In order to control for drivers of the co-worker network density in the region, we have to understand which factor is the most influential. Therefore, 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.

Table 5. Descriptives of degree drivers and correlation, logarithmic scale, 2008

Variable	Min	Max	Mean	St. D.	Pairwise Pearson correlation				
Degree	0	2.683	1.517	0.568	1				
Years in career	0	1.255	0.915	0.372	0.536*	1			
Avg. plant size	0.022	3.679	1.799	0.789	0.492*	-0.028*	1		
Job changes	0	1.204	0.360	0.261	0.528*	0.460*	-0.070*	1	
Movements	0.301	4.422	2.465	0.787	0.629*	0.235*	0.944*	0.110*	1
Employment in the region	2.861	6.013	5.398	0.611	0.153*	0.021*	0.115*	0.144*	0.154*

Note: *denotes that coefficients of the pair-wise correlation are significant at the 1% level

We transform all the above variables to the logarithm of base 10. Pairwise Pearson correlation co-efficients are highly significant and depict a positive and strong relation of degree to all indicators (Table 5). Since Average plant size and Movements are highly correlated (0.94) they have been inserted separately.

Table 6. 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
R2	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

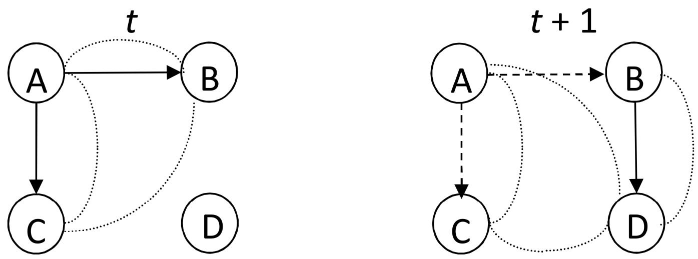
Results of the cross-sectional OLS regression, in which Degree was set as dependent variable, imply the higher values of factors the higher degree. Nevertheless, Job changes and Movements are found to have the strongest effects on degree. These two variables together explain 61% of the variation of individual 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.

6.2 Labour mobility and co-worker ties

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 5). 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 5. 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.

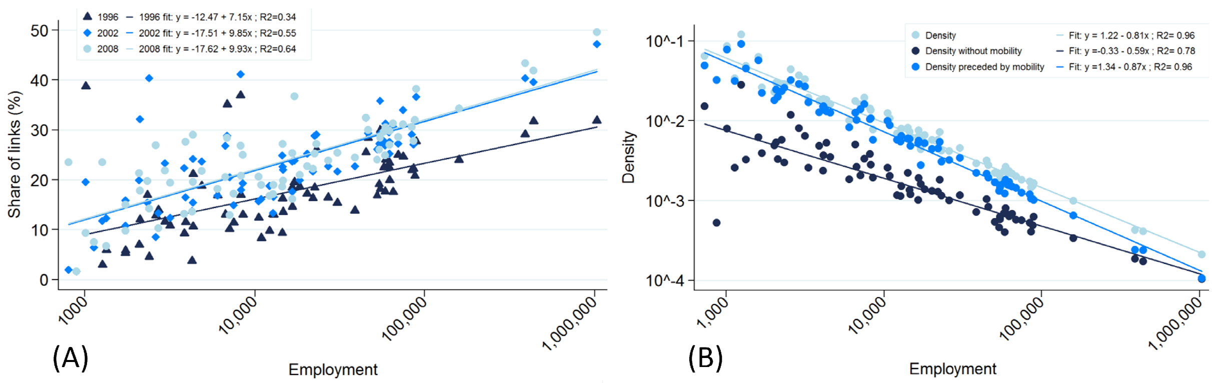
Table 7 illustrates that, on the one hand, the share of links between plants without being preceded by labour mobility was 57.6% in 1991 and grew monotonously above 90% by year 2000. Thus, there was no labour flow between the vast majority of plants that are connected in the co-worker network. On the other hand, one can observe most of the individual co-worker ties are across those plants among which labour mobility was observed before. Looking at this set of links, we find a very strong correlation between the number of individual co-worker ties and the extent of previous labour mobility across plants (ranging between 0.6 and 0.85 over the studied years). One can also see 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.

Table 7. The share of individual ties with or without labour mobility links across plants, 1990-2008

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

If we zoom into regions and look at the share of those individual co-worker links that were not preceded by mobility, we find that the bigger the region the larger the share (Figure 6A). We plotted the rate at three points in time and observe that the above rate increased over time in the case of most regions irrespective of the size of regions. However, 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 R^2 of the linear fit increases.

Figure 6. 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.

6.3 Labour mobility and network density

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

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

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

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

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

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

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

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 density decomposition.

We find that the log of network density is proportional to the log of the size of the region: the larger region the smaller density (Figure 6B). 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. Therefore, other indicators (e.g. diversity, structural holes etc.) might better describe the characteristics of social networks in urban environments.

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.

7. Network density and productivity growth

To test whether our constructed social networks indicators have an actual influence on learning and growth, as claimed in the literature (Ciccone and Hall, 1996, Glaeser, 1999), we construct a panel dataset containing all network variables at regional level 1992-2005 to estimate their relation with productivity growth. Regional productivity is defined as regional per capita value added in each of the 72 functional labour markets defined by The Swedish Agency for Economic and Regional Growth. Apart from reflecting commuting areas also historical economic trends likely to determine future development is accounted for to reflect past and predicted future regional preconditions, which make them consistent over time and suitable for longitudinal analyses without disturbance of spatial dependencies.

Our dependent variable (ProdG) is measured as the relative difference in regional log per capita productivity between t0 and t+3 which implies that 2005 is the latest observation of our left hand side variables (we have no observations on productivity after 2008). As a robustness-check, we also estimated productivity growth between t0 and t+1, which did not change our results significantly. Apart from the two network variables discussed above – NetDensMob (i.e., density preceded by mobility) and NetDensIndep (i.e., density independent of previous mobility) – we also include an indicator of population density (PopDens) since previous literature tend to proxy the socialising potential based on concentrations of people per se (e.g., Ciccone and Hall, 1996; Glaeser, 1999; Storper and Venables, 2004). The interaction term of PopDens and respective NetDens indicators will be also introduced to check if the co-worker network has indirect effect on growth.

Since regional productivity growth also tends to be influenced by the degree of regional specialization (e.g., Frenken et al, 2007) as well as the initial level of productivity due to catch-up effects (Boschma et al, 2014) we include a specialization indicator (Spec) defined as the inverted entropy of 4-digit industries in each region and the initial level of productivity (RegProd). Finally, a controller for average plant size (AvgPlantSize) is included since regions with many larger plants could be assumed to not exhibit the same relative growth rates as regions with a higher degree of small plants due to both competition effects as well as higher relative growth potential among smaller plants (larger plants tend to have higher levels while smaller plants are expected to change faster).[2]

All independent variables are measured in t-1 to reduce the risk of reversed causality and all variables but AvgPlantSize has been logged due to skewed distributions. Variable definitions and descriptives are displayed in Table 8 together with the pairwise correlations. As noticed in Table 8 our three different density variables are correlated (between 0.61 and 0.79) which is expected given the findings in previous sections. Therefore, we estimate them step-wise before assessing them jointly in the regressions. Moreover, specialization is also correlated with population density (negative) and network density preceded by mobility (positive) which also is expected since the larger and more dense the Swedish regions are the more diverse their economic activities tend to be (Boschma et al, 2014). Moreover, previous findings indicate that specialized regions tend to have higher mobility rates due to matching effects which motivate the relation between specialization and mobility induced networks (Eriksson et al, 2008). To remedy the potential multicollinearity caused by including specialization in the models we also ran the models without specialization. That did not influence the estimates on our network variables but lowered the overall explanatory power of the models.

[2] The rate of bachelors among all employees was also introduced as a control variable of human capital. However, the co-efficient of the variable turned to be insignificant with the introduction of other controllers.

Table 8. Variable description and correlation values (N=1008)

Variable	Description	Mean	St. Dev.	Min	Max	Correlation coefficients					
ProdG	Relative (%) growth of the natural logarithm of value added per capita at *t+3* compared to *t.*	0.199	0.356	-.989	3.720	1					
PopDens	Natural logarithm of population density in the region (*t-1*).	2.297	1.478	-1.422	4.995	-0.121*	1				
NetDensMob	Natural logarithm of co-worker network density in the region across plants that have had labour mobility connections previously (*t-1*).	-5.056	1.477	-9.579	0.154	0.083*	-0.793*	1			
NetDensIndep	Natural logarithm of co-worker network density in the region across plants that have not had labour mobility connections previously (*t-1*).	-6.523	1.789	-12.037	0	0.033	-0.609*	0.763*	1		
Spec	Natural logarithm of the inverse of the entropy measured in the employee distribution across NACE 4-digit industries in the region (*t-1*).	-1.379	.2127	-1.822	-0.742	0.226*	-0.766*	0.858*	0.631*	1	
RegProd	Natural logarithm of value added per capita (*t-1*).	5.675	0.511	2.299	7.160	-0.709*	0.303*	-0.297*	-0.179*	-0.515*	1
AvgPlantSize	The number of BAs over the number of plants in the region (*t-1*).	10.011	3.149	0	27.5	0.179*	0.217*	-0.333*	-0.470*	-0.059*	-0.267*

Note: *denotes that coefficients of the pooled pair-wise Pearson correlation are significant at the 5% level.

A fixed effect (FE) panel model was applied to estimate the influence of our network indicators on regional productivity growth 1992-2005. In simple form, the equation could be specified as:

$$y_{i,t+3}/y_{i,t} = \beta' X_{i,t-1} + \varepsilon_{i,t}; \qquad (7)$$

where y denotes productivity growth, t denotes one-year intervals from 1992 to 2005, i denotes the region, X stands for the set of explanatory variables, and ε is the case- and time-specific error term. The rationale for using this type of model is that it allows us to control explicitly for unobserved institutional differences across regions such as local labour market conditions not captured by the controllers or by the definition of functional regions, which in itself may help reduce the impact of endogeneity. This is highly relevant in the Swedish case due to the great variety of local labour markets in terms of size, population, economic structure and the predominant tradition of local wage setting. By including a full set of time dummies and having all explanatory variables measured the year before the change in the growth indicators as explained above, the risk of unobserved time-specific heterogeneity and reversed causality influencing the results was also reduced.

However, since the models still may be affected from endogeneity, difference-generalized method of moments (GMM) (Arellano and Bond, 1991) with robust standard errors were also estimated as a general robustness check of the full models. In brief, this model first differences all variables to remove the unobserved region effect and then use internal instruments (lags of all variables in levels for the first differences variables) to solve potential endogeneity problems[3]. Thus, while both handling potential endogeneity and omitted variable bias, such a model approach also overcomes the problem of having a version of the dependent variable included in the right-hand-side of the equation, something that otherwise risks producing inconsistent estimates on especially the lagged dependent (Boschma et al., 2014). It should however be noted that we are not estimating a proper productivity model since that would require information on fixed regional assets (e.g., capital or investments), which is not available over the full period. However, our primary aim is not to fully estimate the geography of productivity growth but to discern whether the so often stated link between network density and regional growth actually holds when controlling for a number of relevant regional features.

The estimation results are shown in Table 9. The table contains seven models where models 1-3 estimate each type of density separately while models 4 and 5 estimate population density and the respective network indicator together, which then is repeated in a GMM setting in models 6 and 7. The interaction term of population density and the respective network density indicators are introduced into the GMM models 8 and 9.

[3] All variables but the year dummies are regarded endogenous. For the endogenous variables, the second lag is used as instrument. Due to the relatively large number of instruments in comparison with the number of observations, deeper lags than the second could not be included since that would risk causing inconsistent estimates. It also makes the system-GMM less appropriate to use since it requires more instruments. Using up to five lags as instruments however produce similar estimates while employing a system-GMM in this case failed to produce any significant estimates.

Table 9. Panel regressions on regional productivity growth (ProdG)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	FE	FE	FE	FE	FE	GMM	GMM	GMM	GMM
PopDens	-0.851***			-0.803***	-0.964***	-1.336*	-1.797**	-1.488*	-1.898**
	(0.286)			(0.287)	(0.289)	(0.716)	(0.718)	(0.766)	(0.811)
NetDensMob		0.048**		0.043**		0.004		-0.095	
		(0.021)		(0.021)		(0.063)		(0.077)	
NetDensIndep			0.012*		0.015**		0.039***		0.027*
			(0.006)		(0.006)		(0.013)		(0.015)
PopDens x								0.096**	
NetDensMob								(0.045)	
PopDens x									0.018
NetDensIndep									(0.016)
Spec	0.584***	0.423***	0.449***	0.515***	0.537***	1.161*	1.036	1.242*	1.022
	(0.146)	(0.147)	(0.146)	(0.150)	(0.147)	(0.696)	(0.720)	(0.695)	(0.765)
RegProd	-0.435***	-0.432***	-0.429***	-0.433***	-0.429***	-0.406***	-0.374***	-0.393***	-0.345***
	(0.031)	(0.031)	(0.031)	(0.030)	(0.031)	(0.102)	(0.094)	(0.097)	(0.088)
AvgPlantSize	-0.029***	-0.020***	-0.024***	-0.023***	-0.026***	-0.023**	-0.018*	-0.022*	-0.011
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.011)	(0.010)	(0.012)	(0.009)
Constant	5.713***	3.707***	3.587***	5.670***	5.943***				
	(0.745)	(0.253)	(0.251)	(0.744)	(0.750)				
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-sq	0.602	0.601	0.600	0.604	0.605				
adj. R-sq	0.564	0.563	0.562	0.566	0.567				
AR(1)						0.002	0.003	0.003	0.004
AR(2)						0.151	0.197	0.185	0.231
Hansen J						0.555	0.487	0.680	0.531
Instruments						78	78	91	91
N	1008	1008	1008	1008	1008	936	936	936	936
Time frame	1992-2005	1992-2005	1992-2005	1992-2005	1992-2005	1992-2004	1992-2004	1992-2004	1992-2004

Note: standard errors in parentheses; *, **, *** sign the level of significance at the 0.1, 0.05 and 0.01 levels, respectively.

Based on the regression results it is possible to conclude that co-worker network density is positively associated with productivity growth, both when preceded by mobility (Model 2) and when not preceded by mobility (Model 3), while population density is not (Model 1). Thus, together with the descriptives in Table 8 showing a negative correlation between both network indicators and population density, it is possible to say that population density is a poor proxy of social interaction leading to learning and growth despite so frequently used in previous studies (Ciccone and Hall, 1996). Rather than relying on somewhat esoteric notions that knowledge is "in the air" (Marshall, 1920) or in the "buzz of urban life" (Storper and Venables, 2004) these results point to the fact that knowledge is always and everywhere peopled which emphasize the importance of studying the micro-processes at play (Breschi and Lissoni, 2009; Eriksson and Lindgren, 2009). These findings are robust when also estimating population density together with network density in models 4 and 5.

The GMM models however indicate that the original models may suffer slightly from endogeneity or omitted variable bias since only the positive association between growth and network density not preceded by mobility remains significant (Model 7) while network density preceded by mobility loses significance (Model 6). This is expected since previous studies on the relation between productivity growth and mobility of workers with university degrees in Sweden neither find any significant effects if not distinguishing the type of experience of the workers (Boschma et al, 2014). The interaction term between network density preceded by mobility and population density however has a positive and significant effect on growth (Model 8). This term has the largest coefficient we found for the network indicators, which suggests that the co-worker network based on labour mobility operates as the engine of growth in densely populated areas but having an indirect effect only. However, the direct effect of co-worker network independent of mobility is robust against the introduction of the interaction term.

The test statistics for the GMM models are satisfying. AR(1), which tests the null hypothesis of no first-order correlation in the differenced residuals, is rejected while the null hypothesis of no second-order autocorrelation in levels, AR(2), is confirmed. This significant AR(1) is expected since first differences in errors share an error level component and together with outcome from the AR(2) test and the non-significant Hansen statistic, which under the null hypothesis tests that the instruments as a group are exogenous, indicates that the instruments fulfil their purpose. Moreover, all controllers show expected signs. Specialization trigger productivity, regions with high initial productivity is not as fast growing as those with lower productivity and many small firms tend to grow relatively faster than large firms. The controllers are, except from specialization in the GMM-models, robust throughout all models. This non-significant estimate on specialization together with the relatively high standard errors may be a sign of multicollinearity due to the rather strong correlation between population density and specialization.

In a robustness check (not reported) we estimate models 6 and 7 when excluding the period prior to 1995 since that (i) was a turbulent period due to a big recession causing many involuntary job moves, and (ii) because the average degree of plants and individuals included in the network is saturated in 1995 after having had some initial years to develop (see Table 3). This did however not influence any of the results on the density indicators reported from models 6 and 7. Further, we also estimated Models 4-7 for the period 1992-1999 since that was a period when population density was positively correlated with growth in a univariate setting. It however turns insignificant when introducing further variables while NetDensMob has positive and significant effect in the FE model but has a negative and insignificant effect in the GMM model. NetDensIndep however turns to have a positive effect that is only significant in the GMM model.

8. 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, and calculated decomposed network densities for those links that have been preceded by labour mobility and those that are independent from labour mobility. We find that network density is negatively correlated with population density, suggesting that while the potential for social networks are high in dense regions, the strength of these interactions are higher in smaller regions. We also find a robust positive effect of the density of the co-worker network on regional productivity growth, which remains significant in different model specifications only for those links that have never been preceded by labour mobility previously.

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 current paper focused on two issues: (1) the effect of co-worker network density on productivity growth; (2) the independence of co-worker network density from labour mobility networks.

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, in contrast to previous studies advocating the immense role of density (e.g., Ciccone and Hall 1996, Glaeser 1999) our empirical analysis indicates that it is not population density per se but the density of the co-worker network that is 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, despite that it is not preceded by mobility. In fact, density of the co-worker network not driven by mobility is the most robust network indicator. 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 weak 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.

Since our methodology opens up the possibility of employing a micro perspective, one can analyse networks aggregated on various levels including individuals, plants, firms or industries. Further research might devote attention to the effects of 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 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 in terms of density but also allows controlling for further aspects influencing performance which are industry- or plant-specific. 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, which might open a new horizon for creating social networks from co-occurrence data.

References

Adamic, L., Adar, E. (2005), How to search a social network, *Social Networks*, 27: 187-203.

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

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

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

Amin, A. (2000), Industrial districts. In Sheppard, E., Barnes T.J. (eds.) *A Companion to Economic Geography*, pp. 149-168. Oxford: Blackwell Publishing.

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

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

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

Bathelt, H., Glückler, J. (2003), Toward a relational economic geography, *Journal of Economic Geography*, 3: 117-144.

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

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

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

Blau, P., Blum, T., Schwartz, H. (1982), Heterogeneity and intermarriage, *American Sociological Review*, 47: 45–62.

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

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

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

Boschma, R.A. (2005), Proximity and innovation: a critical assessment, *Regional Studies*, 39: 61-74.

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

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

Boschma, R.A., Frenken, K. (2011), The emerging empirics of evolutionary economic geography, *Journal of Economic Geography*, 11: 295-307.

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

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

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

Burt, R. L. (1992), Structural Holes. Cambridge, MA: Harvard University Press.

Calvo-Armengol, A., Jackson, M.O. (2004), The Effects of Social Networks on Employment and Inequality, *American Economic Review*, 94: 426–54.

Ciccone, A., Hall, R.E. (1996), Productivity and the Density of Economic Activity, *American Economic Review*, 86: 54-70.

Coleman, J. S. (1990), *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

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

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

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

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

Erdős, P., Rényi, A. (1959), On random graphs, *Publicationes Mathematicae*, 6: 290-297.

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

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

Feld, S.L. (1982), Social structural determinants of similarity among associates, *American Sociological Review*, 47: 797–801.

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

Frenken, K., Van Oort, F.G., Verburg, T. (2007), Related variety, unrelated variety and regional economic growth, *Regional Studies*, 41(5): 685–697.

Glaeser, E. L. (1999), Learning in cities, *Journal of Urban Economics*, 46: 254-277.

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

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

Glückler, J. (2007), Economic geography and the evolution of networks, *Journal of Economic Geography,* 7: 619-634.

Granovetter, M. (1973), The strength of weak ties, *American Journal of Sociology*, 78: 1360-1380.

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

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

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

Jackson, M.O. (2008), *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Kossinetz, G., Watts, DJ. (2006), Empirical analysis of an evolving social network, *Science,* 311: 88-90.

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

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

Malmberg, A. (1997), Industrial geography: location and learning, *Progress in Human Geography,* 21: 573-582.

Marshall, A. (1920), *Principles of Economics – An Introductory Volume*. London: MacMillan.

Maskell, P., Malmberg, A. (1999), Localised Learning and Industrial Competitiveness, *Cambridge Journal of Economics,* 23: 167-185.

McPherson, M., Popielarz, P.A., Drobnic, S. (1992), Social networks and organizational dynamics, *American Sociological Review,* 57: 153-170.

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

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

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

Neffke, F., Henning, M. (2013), Skill relatedness and firm diversification, *Strategic Management Journal*, 34: 297-316.

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

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

Sorensen, O. (2003), Social networks and industrial geography, *Journal of Evolutionary Economics*, 13: 513-527.

Sorensen, O., Rivkin, J. W., Fleming, L. (2006), Complexity, networks and knowledge flow, *Research Policy,* 35: 994-1017.

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

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

Timmermans, B., Boschma, R.A. (2013), 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: 289-311.

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

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

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
	Arts and media	21	5.100	1.392165	12.018	1.529829		
2	Journalism and media	32	3.491	0.95295	11.053	1.40699	6.91	5.84
	Humanities	22	16.725	4.565481	22.825	2.905504		
	Social sciences	31	27.273	7.444805	47.950	6.103786		
3	Business. trade and administration	34	40.262	10.99046	92.489	11.77337	22.43	21.40
	Law	38	14.640	3.996331	27.662	3.521229		
	Biology and environment	42	1.821	0.497085	9.571	1.218339		
	Physics and chemistry	44	3.191	0.871058	10.265	1.306681	4.54	6.08
4	Mathematics	46	9.381	2.560764	10.637	1.354035		
	Data	48	2.256	0.615828	17.288	2.200673		
	Engineering	52	36.910	10.07545	105.734	13.45939		
	Manufacturing	54	1.476	0.402909	4.072	0.518344		
5	Construction	58	10.915	2.979505	23.481	2.989009	14.68	18.09
	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		
	Animal care	64	807	0.22029	1.865	0.237405		
6	Health care	72	58.451	15.95557	151.420	19.27498	21.00	24.37
	Social work	76	17.647	4.817162	36.679	4.669046		
	Personal services	81	42	0.011465	1.472	0.187378		
	Security and military	86	52	0.014195	3.634	0.462589	0.99	2.77
0	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.

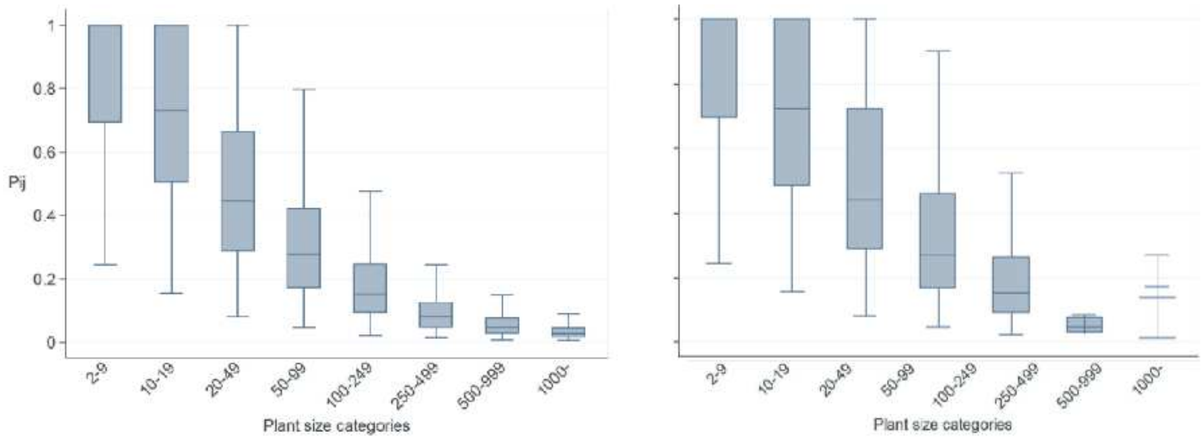
Appendix 1b. Number of employees by gender categories

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. Tie probability distribution and firm size categories. 1990 and 2008



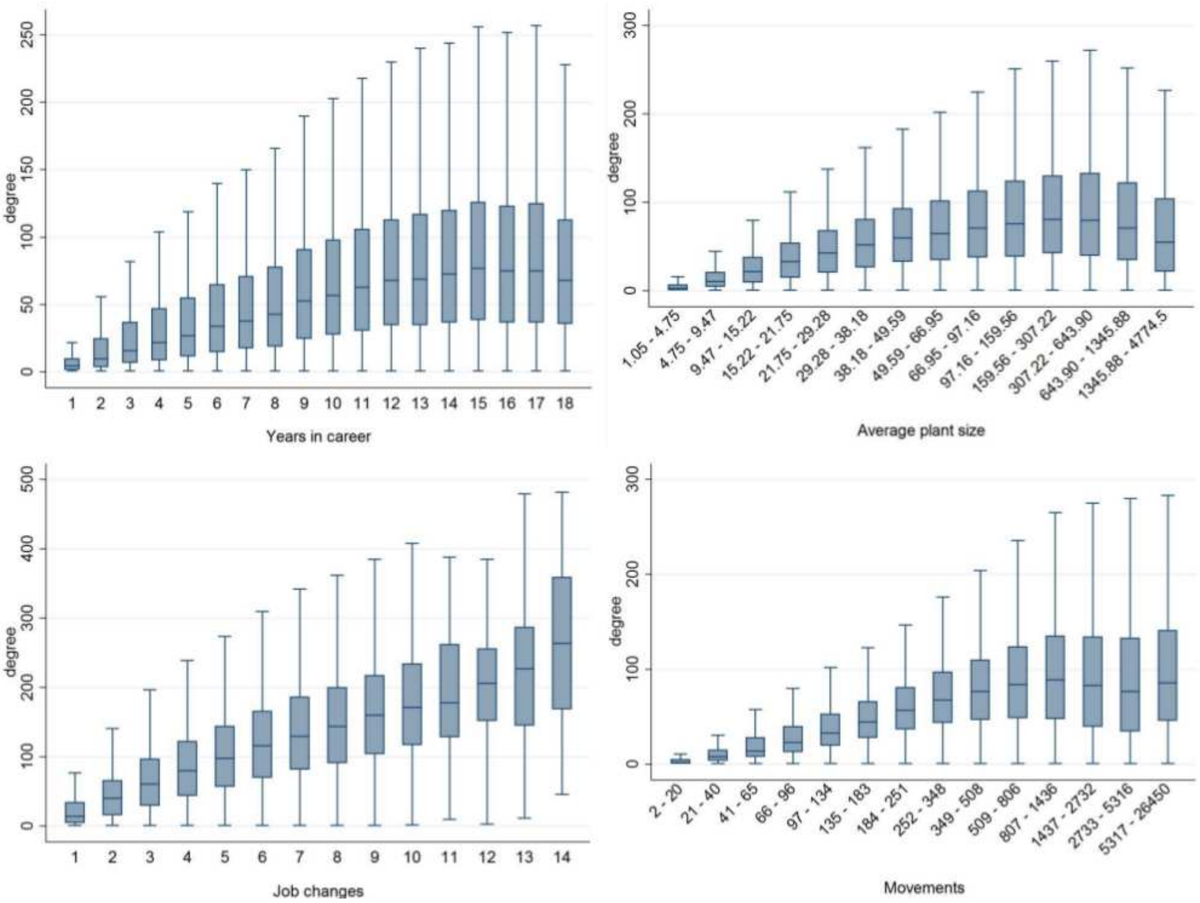
Note: Distributions for 1990 in the left and for 2008 in the right. We excluded those large number of outlier observations that are below or above the whiskers.

Appendix 3. 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 4. Degree distribution against drivers, 2008



Note: We have eliminated 15 and 16 job changes from the illustration (based on their very low frequencies). Average plant size and Movements variables were grouped into 14 bins with equal number of observations.

The median, the 75th percentile, and the upper adjacent value of the degree distribution grow monotonically until 15 years spent in career, after which these values remain constant. We find similar trend in terms of average plant size and degree distribution: the distribution is pushed to the right until the 307-643 range of average plant size. Similar trend is observed in terms of movements: median degree and 75th percentile grow until the 807-1436 range of movements. The variable of job changes seems to have a monotone effect on degree: the higher number of job changes over the career of the employees the higher median, upper and lower hinge.

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×b*, *a×c*, *b×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×b* and *a×c* submatrix densities) and the density of the mobility-independent segment is 0.106 (density of *b×c* submatrix).