Journal of Economic Geography (2016) pp. 1–26 doi:10.1093/jeg/lbw027

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

Balázs Lengyel[*,**,***,†] *and Rikard H. Eriksson*[***]

[*]Centre for Economic and Regional Studies, Hungarian Academy of Sciences, Budaörsi út 45, 1112 Budapest, Hungary
[**]International Business School Budapest, Záhony utca 7, 1031 Budapest, Hungary
[***]Department of Geography and Economic History, Umeå University, SE-901 87 Umeå, Sweden
[†]Corresponding author: Balázs Lengyel, Centre for Economic and Regional Studies, Hungarian Academy of Sciences, Budaörsi út 45, 1112 Budapest, Hungary. *email* <lengyel.balazs@krtk.mta.hu>

Abstract

The mobility of workers is an important source of regional dynamics, but the effect of mobility on regional productivity growth is not straightforward, as some firms tend to win while others lose from mobility. In the present paper, we argue that the co-worker networks across plants that are established by labour moves are important for both local learning opportunities and job matching quality and should hence facilitate regional growth. We therefore propose a new homophily-biased perspective on co-worker network creation and show that it suits geographical analyses better than random networks do. Moreover, panel vector autoregression models provide systematic evidence that an increase in co-worker network density is positively related to regional productivity growth. This is found to be important even when only ties across plants that are not directly linked by labour mobility are included.

Keywords: Homophily, probability of ties, regional productivity growth, panel vector auto-regression
JEL classifications: D85, J24, J61, R11, R23
Date submitted: 5 May 2015 **Date accepted:** 28 August 2016

1. Introduction

Following Marshall (1920), there is general agreement in economic geography and related fields that the agglomeration of economic activities is essential to understanding regional growth, because larger markets allow for more efficient sharing of common facilities, foster learning and provide better job-matching opportunities (Duranton and Puga, 2004). In this respect, face-to-face interaction is increasingly emphasized as essential to explaining why proximity is still crucial for sustaining learning (Storper and Venables, 2004), and why denser environments enhance the probability of 'learning by seeing' (Glaeser, 2000) as well as the quality of matching between employers and employees (Helsley and Strange, 1990). In a closely related body of literature, there is increasing recognition of the role of labour mobility, which is expected to diffuse unstandardized knowledge across firms (Gertler, 2003). Such diffusion, in turn, has been shown to facilitate firm innovation (Breschi and Lissoni, 2009) and productivity (Eriksson and Lindgren, 2009), as well as regional growth (Boschma et al., 2014).

However, besides the direct effect of labour flows between workplaces, labour mobility is expected to create additional social ties between firms that can have indirect effects on firm performance. These social links are important because ties to previous co-workers tend to be persistent and therefore can serve as a long-term channel for knowledge exchange (Dahl and Pedersen, 2004; Agrawal et al., 2006). Thus, firms gain extra benefits when they can access external knowledge via the social ties between former co-workers, and by reducing the average costs of employer–employee mismatches. Still, despite the above contributions claiming that the network of former colleagues is imperative in sustaining learning, matching and growth, very little empirical work has actually been devoted to analysing the role of co-worker networks on regional growth, or to distinguishing the potential different effects of mobility and social networks. Huggins and Thompson (2014) consequently argue that the role of networks in regional growth remains unresolved.

To address this potential shortcoming in the existing literature, the aim of the present paper is to develop a new methodology of creating co-worker networks and assess the influence of these networks on productivity growth in 72 Swedish labour market regions during the period 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. In labour economics, these types of networks are frequently called co-worker networks, and scholars assume that two employees know each other when they have worked in the same workplace simultaneously during a certain period of their career (for an overview see Beaman and Magruder, 2012). We claim that co-worker networks are important sources of regional growth for two reasons. First, valuable knowledge is transmitted more efficiently through co-worker relations, and employees might learn more efficiently in dense co-worker networks (cf. Breschi and Lissoni, 2009; Eriksson and Lindgren, 2009). Second, as network density increases, job information flows more smoothly, thus providing a greater chance of good quality employer–employee matches, which, in turn, increase the productivity of individual firms (Montgomery, 1991; Calvo-Armengol and Zenou, 2005).

Identifying the precise individual effect of the co-worker network on learning and matching is beyond the scope of the present paper. Instead, we will highlight the importance of the co-worker network in an empirical manner. We claim to make two contributions to the existing literature in this regard. First, we develop a new probability measure of workplace-based acquaintance, building on the literature on 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 the most probable individual ties of every employee. As a result, we obtain a dynamically changing social network that represents the full economy and still captures social ties at the micro-scale. Despite the fact 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 detail that our approach differs both conceptually and empirically from previous labour mobility studies. Our second contribution is therefore that the paper provides the first empirical evidence that the density of the co-worker network has a positive effect on productivity growth defined as regional income per capita, even when the focus is only on co-worker ties between plants, which are not directly linked by previous labour mobility. The findings are robust to different homophily specifications and thresholds of ties as well as to removing the old ties from the network.

2. Literature review

Professional networks (i.e., 'co-worker networks') are frequently used in labour economics in relation to job-worker matching by assuming that two employees know each other when they have worked in the same workplace simultaneously during a certain period of their career (Beaman and Magruder, 2012). Previous studies have shown that information flows through these co-worker relations help people find better jobs and reduce unemployment time after dismissal (Simon and Warner, 1992; Granovetter, 1995; Calvo-Armengol and Jackson, 2004; Hensvik and Nordström Skans, 2013) as well as facilitating job-matching, which is performance enhancing for the involved firms (Helsley and Strange, 1990; Calvo-Armengol and Zenou, 2005). However, previous research on co-worker networks suffers from two main shortcomings. First, most studies are restricted to small firms, only because two randomly selected employees are less likely to know each other in a large firm compared to in a small firm. Glitz (2013), for example, only looked at firms with a maximum of 50 employees. Still, it is not feasible to eliminate co-worker networks generated at large firms when estimating the effect of the network on regional economic growth, as the bulk of employment often originates from larger firms. Second, and more importantly, despite the fact that most labour market relations are confined within local labour markets, co-worker networks have rarely been analysed from an economic geography perspective, although spatially based social interactions are often investigated in relation to job-worker matching and labour market outcomes (for an overview see Ioannides and Datcher Loury, 2004). In fact, a spatial approach to large-scale networks is largely absent from the literature, which limits our knowledge about the potential network effect on regional growth (Huggins and Thompson, 2014).

Ideas on the role of network-related learning have nevertheless been present in economic geography for some time (see, e.g., Bathelt and Glückler, 2003; Ter Wal and Boschma, 2009). For example, strong social ties within certain sectors in specialized industrial districts are assumed to enhance incremental innovation and productivity growth (Asheim, 1996; Malmberg, 1997; Amin, 2000), whereas diverse regional networks across industries are associated with potential new combinations of information, knowledge creation and radical innovation (Feldman, 1999). More recently, the emerging literature on 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). Due to data access difficulties, however, these studies tend to be restricted to case studies or very small samples, which may limit the generalizability of their results.

Consequently, the following arguments stress two points. First, 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 of labour mobility networks. Second, the positive effect of network density remains significant when old ties have been eliminated from the network.

The above-mentioned propositions are based on the fact that a growing body of literature in economic geography has considered labour mobility between firms to be a major source of learning that is more direct than pure knowledge externalities ‘residing in the air’ of agglomerations (e.g., Almeida and Kogut, 1999; Breschi and Lissoni, 2009; Eriksson and Lindgren, 2009). Apart from improving the potential regional matching of skills, Boschma et al. (2014) also demonstrate that high concentrations of flows between skill-related industries in a region strongly influence productivity growth in Sweden due to the production complementarities produced by such labour market externalities (see also Boschma et al. [2009] for findings at the level of the firm).

The co-worker approach presented here 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 even shows that lasting co-inventor relations are important for later patenting collaborations (Agrawal et al., 2006, Breschi and Lissoni, 2009) and that former colleagues continue sharing novel information with each other long after they stop working at the same firm (Dahl and Pedersen, 2004). It is therefore reasonable to expect that not only the transfer of embodied knowledge via labour flows influences regional growth, but also social networks that are created by—but might be independent of—labour flows. Therefore, we will decompose the co-worker network into two segments: (1) links across plants that have been directly linked by labour mobility and (2) links across plants that have not been directly linked by labour mobility. In so doing, we will be able to assess whether co-worker network density enhances regional income per capita growth even when the ties across plants have not been directly linked by labour flows among the concerned plants. Because we exclude the possible reversed causality by looking at ties not preceded by mobility, a positive estimation could be understood as indicating that co-worker ties across firms are important channels for information flows and thus indirectly linked to productivity growth in regions through learning and matching.

In the sociology literature (Coleman, 1990; Burt, 1992; Wasserman and Faust, 1994; Walker et al., 1997), network density has been considered a major indicator of social capital for decades, because the closure of social relations enhances trust, authority and sanctions among local actors, all of which supports learning from contacts and speeds up the flow of information in the network. By assessing the impact of network density on regional growth, our study is also related to the vast field of research advocating the impact of density indicators—population density in particular—as an important driver for regional growth. This is because the spatial agglomeration of economic activities unburdens the sharing of common facilities, increases 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 makes dense regions more productive. Furthermore, high density is likely to increase productivity by improving the quality of matching between employers and employees because it reduces the average cost of mismatches (Helsley and Strange, 1990).

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 (e.g., Granovetter, 1973). On the one hand, weak ties offer access to new information and the combination of non-redundant knowledge, which can lead to radical innovations (Ahuja, 2000) and to a wider pool of job-related information (Granovetter, 1995). On the other hand, people frequently follow up with strong ties, which offers the possibility of incremental innovation and an increase in individual productivity, because they learn effectively from each other (Borgatti and Cross, 2003; Balkundi and Harrison, 2006). The above issue of a co-worker network effect and tie strength can be addressed by removing the old ties from the network and focusing only on the recent ties, a process that has been suggested in both the sociology and network science literature (Burt, 2000; Murase et al., 2015). In this way, we can assess whether there is a different effect on regional economic growth depending on whether or not we eliminate the old and presumably weaker ties from the network. A positive impact of new ties alone can be interpreted as the co-worker network does not only contribute to externalities by improving the quality of employer–employee matching, in which weak ties are expected to be very important, but also by increasing the potentials for learning through recent personal connections.

3. Methodology

3.1. Homophily and tie creation

We propose that employee i and employee j working in the same workplace at the same period of time know each other with a probability of P_{ij} [0,1] and maintain a tie of L_{ij} even after termination of the co-workership, when employee i and employee j work for two different firms.

Intuition suggests that the likelihood of acquaintance between two randomly selected employees decreases as the size of the workplace grows. Therefore, we apply an initial random network probability (P^r_{ij}) to every pair of employees that is inversely proportional to the number of employees by using a probability threshold where isolated nodes tend to disappear in a random network setting (Erdős and Rényi, 1959; Jackson, 2008). The formula of the initial probability is

$$P^r_{ij} = \frac{\ln N}{N}, \tag{1}$$

where N is the number of employees in the workplace.

Next, we consider that individual similarity increases the probability of tie formation—a phenomenon that has been called homophily in much of the social sciences (for an overview see McPherson et al., 2001). It has repeatedly been shown that more friendship ties are formed among individuals who are similar in terms of age, gender, race, education, occupation, etc., than would be expected based on random tie establishment (Blau, 1977; Lincoln and Miller, 1979; Feld, 1982; McPherson and Smith-Lovin, 1987; Sias and Cahill, 1998). Currarini et al. (2009) also demonstrate that individual friend selection is generated by the structure of the group, because the larger the subgroup of similar individuals, the greater the possibility of choosing similar friends. This is called baseline homophily (H_b). However, friendship ties usually exhibit greater homophily due to additional inbreeding homophily (H_i), making the individuals' choice even more biased towards those he/she is akin to.

We will assume that H_b influences P_{ij} because relations are more likely between employees who are similar in relation to one or more of their characteristics. This is motivated by the fact that people with the same education are more likely to work in the same division in the firm, and because age and sex also tend to breed further proximity. We define groups of employees by various characteristics and consider two employees similar in relation to a given characteristic if they belong to the same subgroup. The size of groups has an effect on tie probability similar to the effect of firm size itself. Thus, we have to make the additional probability due to homophily bias inversely proportional to group size using the Erdős–Rényi threshold in each case when employee i and j are similar. We let homophily bias have a stronger effect for those characteristics in which the generated group of similar co-workers constitutes a lower relative share in the workplace. Finally, we sum the probabilities calculated from firm size, baseline homophilies and group size effects (Buhai and van der Lei, 2006). The probability is calculated using the following formula:

$$P_{ij} = \frac{\ln N}{N} + \sum_{m=1}^{M} \left(\frac{\ln N_m}{N_m} / \frac{N_m}{N} \right) \times \delta_{ij,m}, \tag{2}$$

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

Equation (2) can be rewritten into a formula, in which one can see that those employee characteristics add the most to the probability of a tie that constitute a small group of similar individuals. In other words, Equation (3) shows that the fewer people who are similar to the individual in a given dimension, the more that dimension will increase the chance that the tie will eventually be established.

$$P_{ij} = N \times \left(\frac{\ln N}{N^2} + \frac{\ln N_1}{N_1^2} \times \delta_{ij,1} + \frac{\ln N_2}{N_2^2} \times \delta_{ij,2} + \ldots + \frac{\ln N_M}{N_M^2} \times \delta_{ij,M} \right). \tag{3}$$

The intuition underlying the formula is that employees' actions are organized along different dimensions—be they strategy-oriented meetings, subject-related projects or social activities—and if the individual's activity is related to that dimension, then he/she will establish contacts along it. Similarity plays a role in this process, and the individual characteristics that best describe the dimension will determine the probability of the tie.

3.2. Density across plants and the effect of labour mobility

We use the network density indicator to illustrate that links across plants in the co-worker network are important tools for inter-firm learning and regional growth. Density is usually defined by $D = \frac{2 \times L}{N_{\text{reg}} \times (N_{\text{reg}} - 1)}$, where L is the number of observed links given by the network creation explained above and N_{reg} is the number of employees in the region. However, we have to reduce the de-nominator by the number of potential employee–employee pairs within the same plants. Thus, the density of the co-worker network in the region (D_c) is

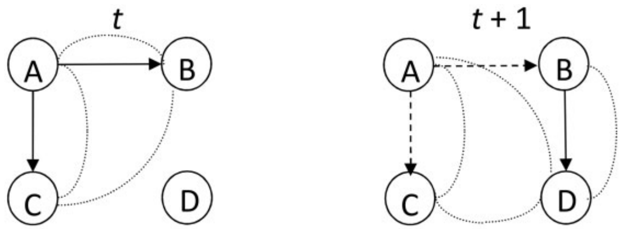

Figure 1. Labour mobility and co-worker ties across plants.
Notes: The solid arrow denotes the actual mobility of one employee, the dashed arrow denotes previous mobility and the dotted line denotes co-worker ties across plants.

$$D_{\mathrm{c}} = \frac{2 \times L}{N_{\mathrm{reg}} \times \left(N_{\mathrm{reg}} - 1\right) - \sum_k N_k \times \left(N_k - 1\right)}, \tag{4}$$

where N_k is the number of employees at plant k and $\sum_k N_k$ equals N_{reg}.

Labour mobility has an instrumental effect on the co-worker network, because an employee establishes co-worker ties to distinct plants if he/she or one of his/her colleagues moves across plants (Collet and Hedström, 2012). However, not all of the plant-level co-worker ties can be detected from a pairwise labour mobility matrix, and co-worker ties might exist across plants that were never linked by mobility. 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 at time t (B and C have at least one employee before the arrival of i and j). The above two moves create co-worker ties between plants A and B, A and C and, additionally, there will be a co-worker tie between B and C without any employee moving from B to C or vice versa (Figure 1). Furthermore, previous labour flows do not necessarily mean continuing co-worker ties across plants. If employee i moves from plant B to plant D at time $t+1$, the link between A and B will disappear.

D_{c} can be decomposed into one segment where links have been preceded by labour mobility and into another segment where links are present between plants without previous labour mobility. The following formula expresses this:

$$D_{\mathrm{c}} = \sum\nolimits_{ab}^{l} \frac{2 \times L_{ab}}{N_a \times N_b} \times \frac{N_a \times N_b}{N_{\mathrm{reg}} \times \left(N_{\mathrm{reg}} - 1\right) - \sum_a N_a \times \left(N_a - 1\right)} \times \delta_{ab}^{l}, \tag{5}$$

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 the 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. For a visual explanation of network density decomposition, consult Section I in the Online Supplementary Information file. The above procedure provides us with two density indicators that sum up to:

$$D_{\mathrm{c}} = D_{\mathrm{c}}^{l} + D_{\mathrm{c}}^{n}, \tag{6}$$

where D_{c}^{l} denotes the density of those inter-plant co-worker ties that are parallel with labour mobility (e.g., ties across plant A and B in Figure 1) and D_{c}^{n} denotes those inter-plant ties that are established without direct labour mobility (e.g., ties across plants B and C in Figure 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

4. Data and network creation

4.1. Data

We use matched employer–employee data obtained from official registers from Statistics Sweden that—among a wide variety of data—contain age, gender, a detailed education code, and the wage of individual employees. This enables us to identify employee–employee co-occurrence at plants for the 1990–2008 period on a yearly basis. The worker is listed repeatedly with different plant codes in the same year if he/she changes workplace over the year. Therefore, labour mobility is detected in the given year, and the worker also establishes co-worker ties at both the sending and receiving plants. The exact location of plants is defined by transforming the data from a 100 m $\times$ 100 m grid setting into latitudes and longitudes.

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 the computation demands without contributing much to the analysis. This is also motivated by the assumption that skilled workers (people with a bachelor's degree or higher) benefit more from learning by seeing and interacting (Glaeser, 2000). We therefore propose that workers without a bachelor's degree rely to a greater extent on tacit 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 completes a degree at a later point in time, he/she will be included in our sample afterwards. As a result, the data contain 366,336 individuals in 1990 and 785,578 individuals in 2008, and plants where none of the employees had a bachelor's degree or higher are excluded (Table 1).

4.2. Network selection

We control for four dimensions of tie creation to calculate two alternative tie probabilities by using education, age, gender and wage distributions in plants:

$$P_{ij}^{1} = N \times \left(\frac{\ln N}{N^{2}} + \frac{\ln N_{\mathrm{e}}}{N_{\mathrm{e}}^{2}} \times \delta_{ij,\mathrm{e}} + \frac{\ln N_{\mathrm{a}}}{N_{\mathrm{a}}^{2}} \times \delta_{ij,\mathrm{a}} + \frac{\ln N_{\mathrm{g}}}{N_{\mathrm{g}}^{2}} \times \delta_{ij,\mathrm{g}} \right), \tag{7}$$

$$P_{ij}^{2} = N \times \left(\frac{\ln N}{N^{2}} + \frac{\ln N_{\mathrm{w}}}{N_{\mathrm{w}}^{2}} \times \delta_{ij,\mathrm{w}} + \frac{\ln N_{\mathrm{e}}}{N_{\mathrm{e}}^{2}} \times \delta_{ij,\mathrm{e}} + \frac{\ln N_{\mathrm{a}}}{N_{\mathrm{a}}^{2}} \times \delta_{ij,\mathrm{a}} + \frac{\ln N_{\mathrm{g}}}{N_{\mathrm{g}}^{2}} \times \delta_{ij,\mathrm{g}} \right), \tag{8}$$

where N denotes the number of employees in the plant, N_{e}, N_{a}, N_{g} and N_{w} refer to the number of co-workers belonging to the groups defined by education ($_e$), age ($_a$), gender ($_g$) and wage ($_w$), respectively. In the previous sociology literature (McPherson et al., 2001), age and gender were found to be the most important dimensions of informal tie creation, while the inclusion of education and wage can tell us something about the organizational structure of the plant. Education is argued to be the most important source of skills needed for a specific job (Neffke, 2016), and therefore education is arguably an important driver of co-worker networks because most groups of workers are organized in relation to skill categories (Caliendo et al., 2015). Meanwhile, the position in the wage distribution refers to the individual's place in the organizational hierarchy (Calvo and Wellisz, 1979), which is important to capture because managers of various groups might, for example, be connected to leaders of other groups (Bolton and Dewatripont, 1994).

Note that the above probabilities depend on the three- and four-dimensional distribution along the pre-defined group structure of the plant. In the present paper, two gender groups and three age groups (≤34, 35–49 and 50≤ years of age) are used. We categorized employees into six main education groups based on detailed records of their educational focus and into four wage groups that correspond to the quartile range of the wage distribution within each plant. For further information on group definitions and descriptive statistics, see Section II in the Online Supplementary Information file.

There is no clear suggestion in the literature regarding the number of ties per person that is reasonable and can be handled by the analysis. Management papers have reported on task-oriented networks based on survey data, and the average number of personal ties in these networks is below 10 (Lincoln and Miller, 1979; Brass, 1985; McPherson et al., 1992; Morrison, 2002). Although the co-worker network approach in labour economics is often restricted to small firms only (Glitz, 2013), recent papers in labour economics have tended 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).

We calculate both types of tie probabilities for every co-worker pair in every year, maximize them at 1 and rank co-workers for every employee in every year by their probability. We let employees collect 50 ties per year and as an extension, we also limit the annual number of created ties per person to 25 links. The coefficient of pairwise Pearson correlation is 0.4 between the ranking scores according to P^1_{ij} and P^2_{ij}, but these rankings are not correlated with random rankings (five random lists were created to test the correlation).

To further illustrate that our method is better than just selecting ties randomly, we take a 50% sample of plants by each region and plant size category from 1995 and compare triadic closure by plant size in the P^1_{ij} and P^2_{ij} networks and a random network for 50 ties (Figure 2A) and 25 ties per person (Figure 2B), respectively. Triadic closure (measured by the global clustering co-efficient) slightly decreases as plant size increases, but more radically in the random network. Because triadic closure and clustering are typical in social networks (Borgatti et al., 2009), we can claim that our method captures social networks better than selecting ties randomly does.

The co-variance of pre-defined groups might overshadow the role of N_e, N_a, N_g and N_w in probability calculation in Equations (7) and (8). Thus, we illustrate the share of our four employee characteristics in the network creation process. According to Figure 3A, the number of similar co-workers is the highest in the education category irrespective of plant size. This suggests that, in most plants, most employees have identical

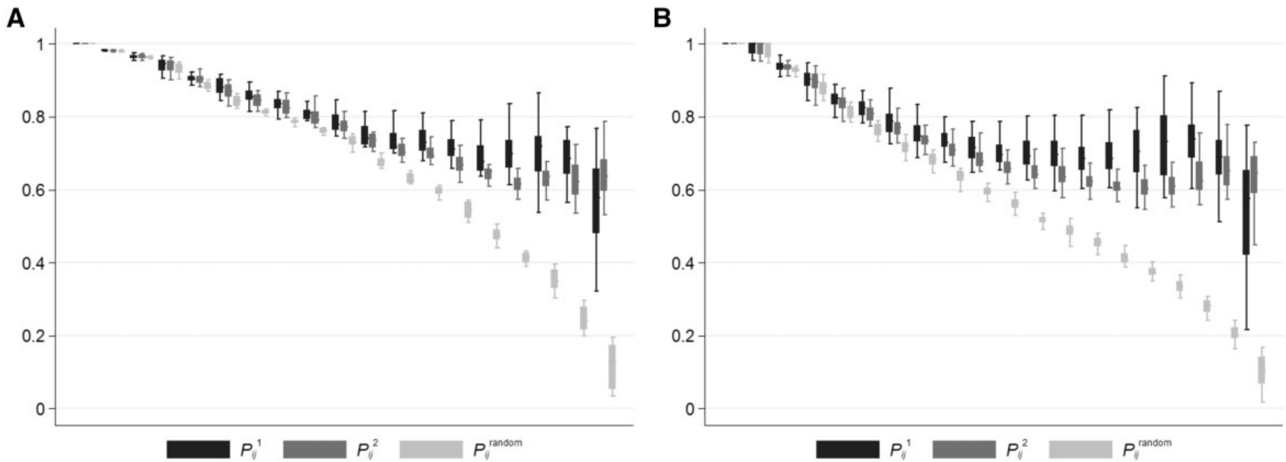

Figure 2. Triadic closure in P_{ij}^{1}, P_{ij}^{2} and random networks within plants.
Notes: The distribution of global clustering coefficients in P_{ij}^{1} (black), P_{ij}^{2} (dark grey) and a random network (light grey) according to plant size. (**A**) 50 ties per person per year. (**B**) 25 ties per person per year.

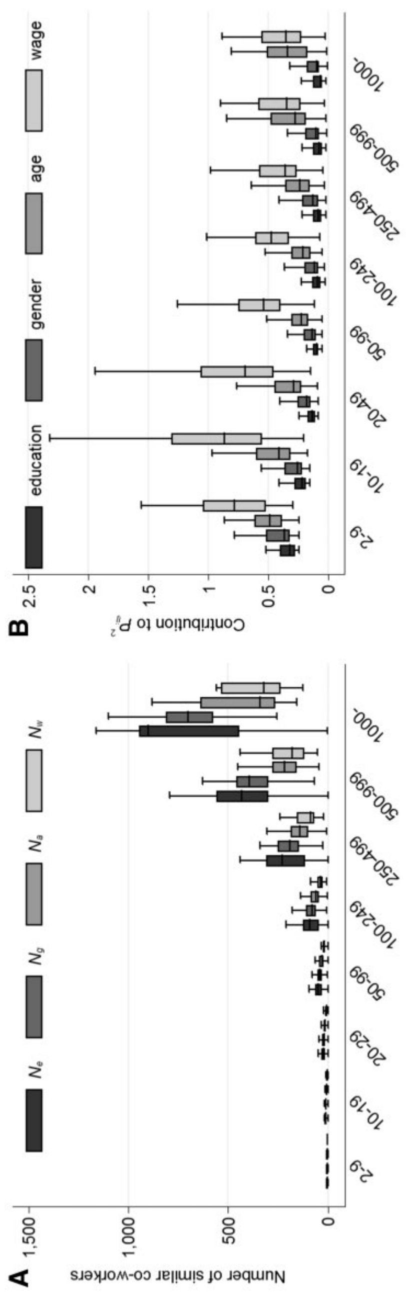

Figure 3. The contribution of employee characteristics to tie creation.
Note: **(A)** The number of similar co-workers according to worker attributes by plant size categories. **(B)** The contribution of worker attributes to P_{ij}^2.

education backgrounds. Therefore, the inclusion of gender and age attributes is necessary to identify link likelihood. Furthermore, wage quartiles identify the smallest groups in most plants. The introduction of wage in P_{ij}^2 thus makes a great contribution to link probabilities (Figure 3B), and we can claim that P_{ij}^1 differs considerably from P_{ij}^2. While links identified by P_{ij}^1 leave some room for diversity in the co-worker ties established in large plants, the P_{ij}^2 specification only links those co-workers in large plants who are identical in all of their attributes. More details on the structure of the P_{ij}^1 and P_{ij}^2 networks and the random network are illustrated for the largest plants in Section III of the Online Supplementary Information file.

To select a network for detailed analysis, we exclude those plants from the tie creation exercise for which all ties are considered (plants with maximum 50 or 25 employees) to focus solely on the difference between the three alternative tie specifications. Then we create six panels of co-worker network ties by tracing the most probable 50 ties according to P_{ij}^1, P_{ij}^2 and a random choice of co-worker ties from every employee and every year over the full period, repeating the process with the 25 most probable ties. We exclude the tie if at least one employee is already above 65 years of age, if either one or both individuals are not working and if the employees work in the same plant.

Network selection is based on a bivariate panel vector autoregression (pVAR) model in a generalized method of moments (GMM) framework. In short, a pVAR model fits a panel regression of each dependent variable on lags of itself and on lags of all other 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 makes past realizations valid instruments (see Love and Zicchino, 2006; Abrigo and Love, 2015 for further information).

Previous studies (e.g., Eriksson et al., 2008) have shown that labour flows in Sweden are primarily confined within labour market regions and that the distribution of inter-regional flows is highly skewed across workers and regions. It is mainly persons younger than 30 years who move across regions, and these flows tend to be restricted to the metropolitan regions and some larger regional centres, while very few observations are found in other regions. Like Boschma et al. (2014), we therefore argue that networks confined within labour market areas (i.e., functional regions) are the proper ground for running the regressions for network selection.

The variables used in the model are the first difference of regional network density, as defined above (see Equation (4)). This variable is then estimated together with the first difference of regional productivity, which is defined as regional per capita income (RegProd), to capture regional growth. The latter indicator 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 tends to be expressed in higher regional wage levels (Combes et al., 2005; Kemeny and Storper, 2015).

The pVAR modelling requires the optimal lag order to be chosen for both the VAR specification and the moment conditions. The models were calculated by using the first- to third-order lags for all variables, together with lags 3–5 for each variable as instruments. All MAIC, MBIC and MQIC tests indicate that three lags of the variables should be included in the model, and we apply the lags 3–5 of both variables as instruments.

Table 2. pVAR models on first differenced regional productivity (RegProd) and density of the truncated network (D_t)

	P^1_{ij} 50 ties Model 1	P^2_{ij} 50 ties Model 2	Random 50 ties Model 3	P^1_{ij} 25 ties Model 4	P^2_{ij} 25 ties Model 5	Random 25 ties Model 6
	First differenced regional income per capita (RegProd)					
L.RegProd	0.333**	1.018	0.363***	0.456***	0.793***	0.270*
	(0.164)	(0.912)	(0.126)	(0.139)	(0.209)	(0.147)
L2.RegProd	0.166**	0.059	0.187***	0.067	−0.085	0.101
	(0.073)	(0.193)	(0.072)	(0.081)	(0.098)	(0.062)
L3.RegProd	0.165*	0.362	0.162*	0.073	0.040	0.098
	(0.086)	(0.298)	(0.087)	(0.098)	(0.085)	(0.068)
L.D_t	0.034**	−0.070	0.029	0.028	0.043**	0.023
	(0.016)	(0.126)	(0.018)	(0.021)	(0.020)	(0.018)
L2.D_t	0.002	−0.010	−0.003	0.002	0.005	0.004
	(0.003)	(0.016)	(0.003)	(0.002)	(0.005)	(0.003)
L3.D_t	0.002	0.009	0.002	0.003	0.000	0.003
	(0.004)	(0.009)	(0.002)	(0.003)	(0.004)	(0.003)
	First differenced network density of the truncated network (D_t)					
L.RegProd	5.304	12.050	−1.658	−5.603	3.981	−2.989
	(5.742)	(14.959)	(2.229)	(3.545)	(3.933)	(2.813)
L2.RegProd	−1.456	−0.947	1.636	3.732**	−1.865	0.927
	(2.903)	(3.701)	(1.607)	(1.619)	(2.023)	(1.550)
L3.RegProd	−1.528	1.577	−0.423	−0.205	−2.034	−1.462
	(1.963)	(4.706)	(2.012)	(1.735)	(2.067)	(1.846)
L.D_t	0.168	−0.928	0.312	0.242	0.842**	0.458
	(0.330)	(1.919)	(0.331)	(0.453)	(0.422)	(0.453)
L2.D_t	0.044	−0.120	0.019	−0.018	0.008	0.143
	(0.112)	(0.234)	(0.115)	(0.052)	(0.116)	(0.103)
L3.D_t	0.001	−0.018	0.043	0.054	0.074	0.060
	(0.071)	(0.158)	(0.070)	(0.064)	(0.088)	(0.117)
Hansen J	10.351	3.838	10.017	2.323	6.903	6.549
N	614	603	609	735	732	734

Notes: RegProd and D_t are first-differenced. D_t is calculated from a truncated co-worker network, in which the ties created at plants smaller than 50 (or 25) employees are not included. *N*s are different due to the selection of firms, which influences the number of regions included. Standard errors in parentheses.
*$p<0.05$.
**$p<0.01$.
***$p<0.001$.

Table 2 illustrates that the change of the network specified only by P^1_{ij} correlates significantly with regional productivity growth when allowing employees to establish 50 ties per year (Model 1). This is only the case for the network specified by P^2_{ij} when limiting tie creation to 25 ties per year (Model 5). Random networks do not correlate significantly with productivity growth, which suggests that the homophily approach is better than tracing co-worker ties randomly. Although P^2_{ij} seems to create better co-worker networks if tie selection is limited to 25 ties, the lower part of the table suggests that network change is positively related to its first lag. Because network dynamics *per se* is not the main focus of the present paper, we limit the detailed discussion of results

Table 3. Number of ties within regional borders, 2008

	Number and share of links			
	Individual level		Plant level	
Full network	20,855,160	100%	5,574,879	100%
Within functional regions ($N = 72$)	14,066,872	67%	3,170,695	57%
out of which within municipalities ($N = 289$)	7,826,977	38%	1,470,603	26%
Across functional regions	6,788,288	33%	2,404,184	43%

to P^1_{ij}network specification with 50 ties created per year and include the results from the P^2_{ij}specification as robustness checks.

5. Results

5.1. Properties of the co-worker network

First of all, we assess whether the chosen co-worker network based on P^1_{ij} network specification with 50 ties resembles characteristics of other social networks and whether labour mobility plays an important role in network development at the individual level. As shown in detail in Section IV of the Online Supplementary Information file, we find a negative exponential degree distribution of the co-worker network in year 2008 that has some favourable properties. The expected degree can be approximated by the average degree in the network, and we find that the probability of finding employees who have more degrees than average decreases sharply above the mean. Thus, the mean is not only the expected value, but also a turning point in the distribution. Moreover, regressions on individual degree indicate that mobility explains about 61% of the variance, which means that mobility does indeed need to be considered explicitly in the analysis.

Not surprisingly, the network is spatially concentrated, which confirms the validity of the spatial level of the pVAR models presented in Table 2. As shown in Table 3, 38% of all individual links are within municipal borders (the smallest administrative division in Sweden) in 2008, and this share is 67% when we look at functional regions. The latter regions represent labour market areas defined by The Swedish Agency for Economic and Regional Growth and stem from observed commuting distances between municipalities. When we aggregate the network on the plant level, we find a very similar pattern. Because only a few regions have a sufficient number of observations to compute reliable measures on inter-regional ties (cf., Boschma et al., 2014), we discard the links in the network that exceed regional borders when estimating the network effect on growth, and we calculate network density solely within functional regions. Further specifics regarding the detailed geography of the network can be found in Section V of the Online Supplementary Information file.

Apart from the fact that the network is largely confined within regional borders, we also find that regional size influences the density of the network: The larger the region, the smaller the density (Figure 4A). This is an important finding, because it suggests that the majority of possible regional links are actually not observed and that this share increases as the size of the region grows. Although there are many more observed links

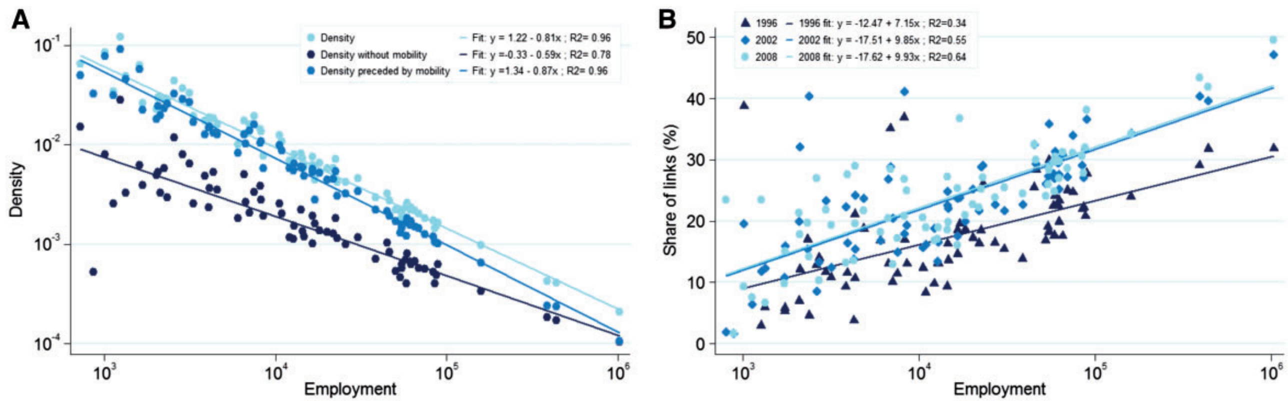

Figure 4. Density by size of the region and mobility-independent co-worker links.
Notes: (**A**) Density and density decomposition by region size in 2008. (**B**) Region size and share of co-worker links not preceded by labour mobility, 1996–2002–2008. Size of the region was captured by the maximum number of employees in the region over the full period.

in big regions than in small regions, the number of possible links are exponentially higher, which produces low network density. It is also apparent that 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, and increases almost monotonically over time (see the Online Supplementary Information file). Zooming into regions in Figure 4B, we find that the bigger the region, the larger the share of those individual co-worker links that were not preceded by mobility.

5.2. Labour mobility and density effect

To estimate whether the chosen definition of network density is related to productivity growth, we again resort to a pVAR model in a GMM framework. Because all variables typically are treated as endogenous (e.g., Holtz-Eakin et al., 1988; Canova and Ciccarelli, 2013), this approach is regarded as particularly suitable in our case, given that the network density itself may be driven by factors such as productivity, population size and density, and labour flows. Thus, to achieve a more detailed understanding of the role of network density in regional productivity, we need to assess how a number of different covariates co-evolve with the full network. Because the potential for network formation may be driven by the turnover rates in regions, which in turn may 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 argued to influence regional growth (e.g., Helsley and Strange, 1990; Ciccone and Hall, 1996; Glaeser, 1999; Eriksson and Lindgren, 2009; Storper and Venables, 2004; Boschma et al., 2014). All variables are logged to reduce the impact of skewed distribution, and we only model the years 1995–2008, because the network is not fully developed until after a couple of years, as illustrated in Section VI of the Online Supplementary Information file.

Based on the model selection criteria, we could conclude that a second-order pVAR is the preferred model in this case, because all tests (MBIC, MAIC and MQIC) were smallest for the second lag. Further, a key criterion of the pVAR is that the model must fulfil the stability condition. This was not the case when running the models on levels, because 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 mitigated 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 the regional level. Compared with Holtz-Eakin et al. (1988), we only use instruments with valid observations, 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 over-identified

Table 4. Panel vector autoregression (pVAR) models on decomposed network density and regional productivity growth 1995–2008

	Model 1: Full network, without mobility				Model 2: Full network, with mobility			
	RegProd	PopDens	MobAcc	D_c^n	RegProd	PopDens	MobAcc	D_c^l
L.RegProd	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.RegProd	−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.PopDens	−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.PopDens	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.MobAcc	−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.MobAcc	−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.D_c^n	0.021**	−0.035	0.174*	0.000				
	(0.009)	(0.025)	(0.102)	(0.184)				
L2.D_c^n	0.005	−0.011*	0.027	−0.090				
	(0.003)	(0.006)	(0.039)	(0.109)				
L.D_c^l					0.065***	−0.195***	0.207	−0.068
					(0.020)	(0.076)	(0.247)	(0.212)
L2.D_c^l					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			
	Model 3: Old ties excluded, without mobility				Model 4: Old ties excluded, with mobility			
	RegProd	PopDens	MobAcc	D_c^n	RegProd	PopDens	MobAcc	D_c^l
L.RegProd	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.RegProd	−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.PopDens	−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.PopDens	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.MobAcc	−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.MobAcc	−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.D_c^n	0.022*	0.007	0.016	−0.197				
	(0.012)	(0.051)	(0.174)	(0.231)				
L2.D_c^n	0.002	−0.003	0.042	−0.142***				
	(0.003)	(0.010)	(0.040)	(0.055)				
L.D_c^l					0.073***	−0.147	0.207	0.267
					(0.024)	(0.092)	(0.250)	(0.229)
L2.D_c^l					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			

Notes: All variables are first-differenced. Cluster robust standard errors in parentheses.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

restrictions. Because the decomposed network indicators are highly correlated (see Table A1 in the Appendix), it is not possible to estimate their effect in the same model.

Model 1 estimates the relation between productivity, population density, regional turnover and the density of the full network in the segment without labour mobility (D_c^n). Based on the first column (on RegProd), we can conclude that previous realizations of productivity are highly influential in explaining future realizations, because the positive coefficient of L.RegProd is strongly significant. This is expected given the documented increasing regional growth divergence in Sweden during this period (e.g., Lundquist and Olander, 2010). As stated in previous studies (e.g., Helsley and Strange, 1990; Ciccone and Hall, 1996; Storper and Venables, 2004), we also find that population density positively influences productivity, which points to the fact that density per se may contribute to spillover effects and matching externalities. However, we cannot 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 influences regional productivity (e.g., Boschma et al., 2014). Finally, and most importantly, the first lag of network density (L.D_c^n) has a positively significant coefficient, which indicates that, given past realizations of both productivity and population density, network density has a positive influence on productivity even when the co-worker link is not preceded by mobility across plants. Thus, the results indicate that inter-plant ties that are not directly preceded by mobility trigger productivity growth.

However, based on the following three columns (PopDens, MobAcc and D_c^n columns in Model 1), it is evident that some of these variables are co-evolving. Both mobility and in particular network density are significantly negatively correlated with population density. This finding confirms our descriptive statistics suggesting that the co-worker network is sparser in more population-dense regions. Further, based on the findings in column 3 on MobAcc, we can confirm previous evidence (e.g., Calvo-Armengol and Jackson, 2004; Calvo-Armengol and Zenou, 2005) regarding social networks' mobility inducing effect in regions. The significant coefficient of D_c^n implies that the denser the co-worker network, the better the employee–employer matches and, therefore, the higher the levels of mobility in the region.

Model 2 assesses the impact of network density preceded by mobility (D_c^l). We still find a positive influence of the network on productivity. Both the first and second lags of D_c^l are positive and significant, but we also find that mobility per se hampers productivity. This latter finding points to the fact that it is not mobility per se that triggers productivity, but the social ties created by mobility. This notion is further supported by the finding in the D_c^l column. The significantly positive coefficient of lagged observations of mobility means that the co-worker network is denser if labour mobility is more intense. However, the direction of causality in this argument is not straightforward, as there is a positive relation between productivity and network density in the D_c^l column, which might indicate reverse causality. In Models 3 and 4 of Table 4, we have removed all the ties older than 5 years, because the strength of relationships may weaken over time (Burt, 2000, Jin et al., 2001). Edge removal is a reasonable method to solve the problem of link ageing, because the characteristics of social networks are better reproduced when old links are deleted when compared with keeping these ties (Murase et al., 2015).[1] We still find a positive, but much weaker, relationship between network density and productivity in the RegProd columns of both Model 3 and 4. Findings presented in the D_c^n column of Model 3 also suggest that productivity is negatively associated with this type of network, given the very strong and negative coefficient of L2.RegProd, indicating that, in less prosperous regions, co-worker ties not preceded by mobility tend to be denser. As presented in the RegProd column in Model 4, young ties preceded by mobility (both the first and second lag of D_c^l) have a positive influence on productivity. This is also true of the lags of productivity and population density, while the effect of mobility is negative. However, in contrast to the full network, network density is not significantly related to any other variable, neither on the right-hand side of the model (in PopDens and MobAcc columns) nor on the left-hand side in the D_c^l column. Thus, based on our findings, it appears that it is particularly young ties that are preceded by labour flows that have the strongest influence on productivity.

A further observation regarding the results concerns the overall model fit, and the Hansen J statistic on the issue of over-identification. In general, Model 1 and 2 seem to suffer from over-identification, meaning that too many instruments are used to be able to remove the endogenous components of the variables (Roodman, 2007). This is, however, less prevalent for Model 3, where Hansen J has a lower significance, and it is not at all the case for Model 4, which only estimates recent ties that are preceded by mobility. This means that the latter models may be considered the most robust.[2]

Finally, economic interpretation of pVAR models is usually accomplished by estimating forecast-error variance decompositions (FEVD). The FEVD results are documented and discussed in more detail in Section VII in the Online Supplementary Information file. Based on these estimates, we can conclude that around 25–30% of the variance in regional productivity can be explained by the density of the co-worker network for both young and old ties (Models 1 and 2). As expected, roughly 60–70% of productivity growth is explained by previous realizations of productivity. Population density accounts for only 1% of the explanatory power in Model 1 and remains below 10% in Model 2, and mobility has a 5% and 3% share in these two models, respectively. However, there is a notable difference between Model 3 and 4 concerning the share of network density. D_c^n explains less than 10%, whereas D_c^l explains around 40% of the variance in productivity growth. These latter observations provide further support to our finding that the co-worker network provides a channel for learning and matching, particularly between firms that have been directly linked by labour flows. This finding is plausible because mobile workers might communicate with previous colleagues at the previous workplace more extensively than with previous colleagues who have changed workplace as well.

1 The 5 years' threshold for link deletion was chosen by measuring tie weights and using exponential time decay curves, as explained in Eriksson and Lengyel (2015).

2 To remedy this particular problem, we also ran the models on the full network when only using lags 3–4 as instruments (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 became insignificant.

5.3. Robustness checks

Because these results might be driven by how the networks are defined and the type of model used, we estimated a number of alternative models as robustness checks on a stratified sample of 25% of the firms in each region. First, we included the ties for which position in the wage distribution was included (i.e., P_{ij}^2 as described in Section 4.2). Although this method may be conceptually superior to defining ties, it also involves a more narrow set of knowledge (i.e., primarily between co-workers with the same type of income). Indeed, we do find a positive relation between P_{ij}^2 and productivity, but these effects were slightly weaker when all ties were included, and not significant at all when the old ties were removed. Moreover, the Hansen statistic was higher and significant in all models, irrespective of the number of instruments, which implies that these findings should be interpreted with some caution.

Second, because the numerical threshold for defining co-workers may also influence the network, we ran the same models for the threshold of 25 ties rather than 50. Here as well the general results remained stable, except that for both P_{ij}^1 (i.e., without wage) and P_{ij}^2 (with wage) network density was only significant in the two models where mobility was excluded. Thus, limiting the number of ties seems to mainly mitigate the impact of mobility-driven networks for networks not preceded by mobility.

Third, we also estimated the same models but using random networks. Random networks had a moderate but still positively significant correlation with productivity for the definition based on 50 ties. In the case of 25 ties, we could not find any significant estimates when decomposing the network, and even found negative effects when removing the old ties from the networks. Thus, as argued in Section 4.2, because the majority of all firms have fewer than 50 employees (97% of all firms in the sample, which employ around 25% of the workers), the random and calculated networks produce similar outcomes for the 50 ties threshold because all ties are included anyway when calculating P_{ij}^1 and P_{ij}^2. However, when we restrict the number of ties a person can have, both P_{ij}^1 and P_{ij}^2 produce better results, because then the homophily-biased networks differ from random networks more significantly.

Finally, although we did not have a full set of indicators that could control for regional productivity differences (e.g., we have no information on investments over the period), we estimated fixed-effect panel models including the first lags of productivity, population density, mobility and share of individuals with a bachelor's degree or higher in the region as control variables (with a full set of year-dummies). Again, while the control variables behaved as expected (previous productivity is strongly positive, as are human capital and lag of population density), the models reveal a positive correlation between network density and productivity, irrespective of being preceded by mobility or not. However, when the old ties are removed, it is only the network consisting of 50 ties that is not preceded by mobility or includes wage that is significant. In all, this would seem to indicate that the results are robust to alternative tie definitions and model specifications. Most importantly, these tests indicate that the results based on calculated co-worker networks are different from those based on random networks.

6. Conclusion and limitations

The present paper provides the first systematic analysis of the role of co-worker networks in regional productivity growth. We demonstrated that the constructed co-worker network suits geographical analyses better than do random networks, and we illustrated the usefulness of our new methodological approach by assessing: (1) whether there is a positive effect of co-worker network density on productivity growth even if the segment of the co-worker network that has not been preceded by labour mobility is included, and (2) that the positive effect holds when the network contains recent ties only.

Indeed, our empirical analysis indicates that—along with population density—the density of the co-worker network is important for regional productivity growth, even if 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 plants, whereas mobility itself does not trigger productivity growth. This finding confirms previous studies showing that regional job flows per se are not an economic blessing for regions, because such flows may produce sunk costs for both the involved firms and individuals unless they are between skill-related industries characterized by cognitive proximity (e.g., Boschma et al., 2014). Productivity gains should motivate public authorities to develop milieus that encourage employees to establish more professional connections within and between workplaces as well as seek out and maintain these connections over their career (cf., Dahl and Pedersen, 2004), rather than facilitating mobility as such. These findings, however, also show the indirect influence of mobility, as co-worker ties are indirectly driven by mobility (Collet and Hedström, 2012).

We do find, however, that network density triggers productivity if only ties that are younger than 5 years are considered. In fact, the model becomes more stable when the old ties are excluded. The more recent co-worker ties are, the more efficient they become when it comes to learning and productivity growth, because co-located previous colleagues might communicate more efficiently if only short period of time has passed since they shared a workplace (Burt, 2000). However, we did not formally distinguish between weak and strong ties, and further research on co-worker network is needed to contribute to the recent debate on the role of tie strength in information diffusion (Aral, 2016).

Despite the promising results presented here, the new methodology we propose has several limitations. First of all, it could be argued that organizational structure matters the most for tie creation at workplaces and that the effect of individual selection is of minor importance (Kossinetz and Watts, 2006). Second, tie creation might be time-dependent because people are more likely to develop connections if they have more time to do so. Third, the dynamics of social networks, and thus co-worker tie creation, is path-dependent because the structure of the network has a tremendously strong influence on the change of the network. Fourth, by calculating the probability of the ties, we most likely underscore the real probability of the tie between co-workers. This is because besides baseline homophily, inbreeding homophily might be at play as well, and tie creation may be even more biased towards people to whom co-workers are akin. Fifth, persistence of co-worker ties might not be automatic, as is assumed both here and in labour economics. Sixth, we do not formally distinguish between strong and weak ties. A final question related to the present 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, and the relatively low mobility rates in Sweden might strengthen the role of networks compared with institutional contexts where turnover rates, within and across regions, are higher.

To tackle these problems, we need to further develop our homophily-biased random network approach by introducing the strength of ties and the effect of group diversity, time, triadic closure and the potential role mentoring might play in learning so as to fit the model to real social networks. This could be accomplished by collecting representative relational data at workplaces to increase the precision of link probability calculation.

Because our estimates reveal slightly diverging patterns of influence depending on the type of network and the number of ties included, further research might look at the effect of the network on job-worker matching at the individual, firm and industry level. Despite having a negligible influence on the regional aggregate effect of the network, different firms and industries might be more dependent on various dimensions of proximity and types of networks (Boschma, 2005). This implies that different networks may indeed play different roles in the performance of industries that rely on different kinds of knowledge input. Analysing the performance of industries or plants instead would not only enable greater heterogeneity, but also allow us to control for further industry- or plant-specific aspects that influence learning and performance (Jackson, 2008), as well as to incorporate information on the geographical distance of ties. We might also devote attention to the effects of the co-worker network's structure on other aspects of regional dynamics, such as firm entry, investment flows, entrepreneurship or employment growth. In this respect, it might be fruitful to also introduce co-worker networks in regional growth frameworks (Huggins and Thompson, 2014), because we have to understand how such networks influence growth in the long run, an aspect that is missing from the present approach.

Supplementary material

Supplementary data for this paper are available at *Journal of Economic Geography* online.

Funding

This work was supported by the Hungarian Scientific Research Fund [K112330 to B.L.]; the research program 'Mobility, Transformation and Regional Growth' at Umeå University (to B.L.) and Riksbankens Jubileumsfond [P13-1004:1 to R.H.E.]

Acknowledgements

This paper benefitted greatly from the constructive comments of the four referees. Help in geo-localization of the data was received from Magnus Strömgren. We thank Pierre-Alexandre Balland, Gábor Békés, Zsolt Csáfordi, Martin Henning, Dániel Horn, Guilherme Kenjy Chihaya da Silva, Thomas Kemeny, Urban Lindgren and László Lőrincz for their suggestions. Useful comments have been received from participants in the Institute of Economics seminar series at the Hungarian Academy of Sciences, ICELAB seminar at Umea University and the Economic Geography Session Series at the AAG Annual Meeting in Chicago, 2015. All usual disclaimers apply.

References

Abrigo, M. R. M., Love, I. (2015) Estimation of Panel Vector Autoregression in Stata: A Package of Programs. Manuscript, February 2015. Available online at: http://paneldataconference2015.ceu.hu/Program/Michael-Abrigo.pdf

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 E. Sheppard and T. J. Barnes (eds) *A Companion to Economic Geography*, pp. 149–168. Oxford: Blackwell Publishing.

Aral, S. (2016) The future of weak ties. *American Journal of Sociology*, 121: 1931–1939.

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.

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.

Bolton, P., Dewatripont, M. (1994) The firm as a communication network. *The Quarterly Journal of Economics*, 109: 809–839.

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.

Burt, R. L. (2000) Decay functions. *Social Networks*, 22: 1–28.

Caliendo, L., Monte, F., Rossi-Hansberg, E. (2015) The anatomy of French production hierarchies. *Journal of Political Economy*, 123: 1393–1467.

Calvo, G. A., Wellisz, S. (1979) Hierarchy, ability, and income distribution. *Journal of Political Economy*, 87: 991–1010.

Calvo-Armengol, A., Zenou, Y. (2005) Job matching, social network and word-of-mouth communication. *Journal of Urban Economics*, 57: 500–522.

Calvo-Armengol, A., Jackson, M. O. (2004) The effects of social networks on employment and inequality. *American Economic Review*, 94: 426–454.

Canova, F., Ciccarelli, M. (2013) Panel vector autoregressive models: a survey. European Central Bank Working Paper Series No. 1507.

Ciccone, A., Hall, R. E. (1996) Productivity and the density of economic activity. *American Economic Review*, 86: 54–70.

Coad, A., Broekel, T. (2012) Firm growth and productivity growth: evidence from a panel VAR, *Applied Economics*, 44: 1251–1269.

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

Combes, P., Duranton, G., Overman, H. (2005) Agglomeration and the adjustment of the spatial economy. *Papers in Regional Science*, 84: 311–349.

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. (2004) 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 J. V. Henderson and J. V. Thisse (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., Lengyel, B. (2015) Co-worker networks and local knowledge externalities. Paper presented at NetWorkShop, June 25, 2015, Pécs University, Pécs. Available online at: http://ktk.pte.hu/sites/default/files/hir_mellekletek/2015/07/eriksson-lengyel_2015june22.pdf.

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.

Feldstein, M. (2008) Did wages reflect growth in productivity?, *Journal of Policy Modeling*, 30: 591–594.

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

Gertler, M. (2003) Tacit knowledge and the economic geography of context, or the undefineable tacitness of being (there). *Journal of Economic Geography*, 3(5): 75–99.

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: G. L. Clark, M. S. Gertler , and M. Feldman (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.

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.

Helsley, R. W., Strange, W. C. (1990) Matching and agglomeration economies in a system of cities. *Regional Science and Urban Economics*, 20: 189–212.

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

Holtz-Eakin, H. H., Newey, W., Rosen, H. S. (1988) Estimating vector autoregressions with panel data. *Econometrica*, 56: 1371–1395.

Huggins, R., Thompson, P. (2014) A network-based view on regional growth. *Journal of Economic Geography*, 14: 511–545

Ioannides, Y. A., Datcher Loury, L. (2004) Job information networks, neighborhood effects and inequality. *Journal of Economic Literature*, 42: 1056–1093.

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

Jin, E. M., Girvan, M., Newman, M. E. J. (2001) Structure of growing social networks. *Physical Review E*, 64: 046132. doi: 10.1103/PhysRevE.64.046132.
Kemeny, T., Storper, M. (2015) Is specialization good for regional economic development? *Regional Studies*, 49: 1003–1018.
Kossinetz, G., Watts, D. J. (2006) Empirical analysis of an evolving social network. *Science*, 311: 88–90.
Lincoln, J. R., Miller, J. (1979) Work and friendship ties in organizations: a comparative analysis of relation networks. *Administrative Science Quarterly*, 24: 181–199.
Love, I., Zicchino, L. (2006) Financial development and dynamic investment behaviour: evidence from panel VAR. *The Quarterly Review of Economics and Finance*, 46: 190–210.
Lundquist, K.-J., Olander, L.-O. (2010) Growth cycles: transformation and regional development. SRE–Discussion Papers, 2010/04. WU Vienna University of Economics and Business, Vienna.
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.
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.
Montgomery, J. D. (1991) Social networks and labor-market outcomes: toward and economic analysis. *The American Economic Review*, 81: 1408–1418.
Morrison, E. W. (2002) Newcomers relationship: the role of social network ties during socialization. *The Academy of Management Journal*, 45: 1149–1160.
Murase, Y., Jo, H.-H., Török, J., Kertész, J., Kaski, K. (2015) Modeling the role of relationship and breakup in social network formation. *PLoS One*, 10: e0133005. doi:10.1371/journal.pone.0133005.
Neffke, F. (2016) How the division of knowledge creates co-worker complementarities. Paper presented at '3rd Geography of Innovation' Conference, January 29, Toulouse.
Saygin, P. O., Weber, A., Weynandt, M. A. (2014) Coworkers, networks, and job search outcomes. IZA Discussion Paper No. 8174.
Simon, C. J., Warner, J. T. (1992) Matchmaker, matchmaker: the effect of old boy network on job match quality, earnings, and tenure. *Journal of Labor Economics*, 10: 306–330.
Roodman, D. (2007) 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. (1998) From co-workers to friends: the development of peer friendship in the workplace. *Western Journal of Communication*, 62: 273–299.
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.
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

Table A1. Descriptive statistics and correlation of variables, 1991–2008

		N	Mean	SD	Min	Max	Pairwise Pearson correlation, pooled							
RegProd	Regional productivity defined as regional gross income per capita (log)	792	7.855	0.104	7.638	8.233								
PopDens	Population density per square km	792	22.169	27.911	0.241	147.701	0.392							
MobAcc	The number of job changes accumulated over time(log)	792	5.904	2.226	0	12.367	0.655	0.709						
NetDens	Network density (log)	792	−4.839	1.415	−9.497	0.154	−0.391	−0.756	−0.858					
NetDensMob	Network density, preceded by mobility (log)	792	−5.056	1.477	−9.579	0.154	−0.431	−0.767	−0.881	0.997				
NetDensIndep	Network density, not preceded by mobility (log)	792	−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)	792	−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)	792	−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)	792	−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.