IBS WORKING PAPERS

Co-worker networks and local knowledge externalities

PRELIMINARY DRAFT; PLEASE, DO NOT QUOTE!

This version: June 22, 2015

Working Paper 4/2015
http://web.ibs-b.hu/research

INTERNATIONAL BUSINESS SCHOOL

Záhony u. 7.

1031 Budapest, Hungary

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

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

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

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

Abstract

This paper explores how social networks influence regional economic development on the base of different types of knowledge externalities in metropolitan areas versus smaller regions. In order to address the above issue, we construct a weighted co-worker network for the entire Swedish economy 1990-2008 and aggregate tie weights on plant- and industry-region levels. We argue that co-worker networks across plants within industry-regions are important for creating MAR type of knowledge externalities; while networks across industry-regions are important for Jacobs externalities. Indeed, we find evidence that growing density of the plant-level network has a positive effect on wages; however, triadic closure of ties is negatively linked to wages. We also find that few strong links to other industries – as opposed to diversity – enhance wage levels in all types of regions. However, links to unrelated industries are only important in metropolitan areas; whereas links to skill-related industries only have a positive effect on development in smaller regional centres.

Keywords: co-worker network, average income, diversity, network density, triadic closure, skill-relatedness, MAR externalities, Jacobs externalities

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

Acknowledgement: The help and advice of Attila Varga in coding is gratefully acknowledged.

1. Introduction

The spatial dimension of network-related learning is a core interest of economic geography (Bathelt and Glückler, 2003, Ter Wal and Boschma, 2009). 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). 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). Despite the central interest, our knowledge about the effect of social networks on regional development is still limited, which is partly due to data access difficulties. Our paper aims to contribute to the literature in this regard by constructing a large-scale co-worker network across plants and industry-regions and analysing the network effect on regional development (Kemeny and Storper, 2014).

Co-worker networks are important for regional development, because most of the knowledge sharing occurs at workplaces (Storper and Venables, 2004), which enables employees to establish cognitive and social proximities that might be maintained even after moving from one workplace to another (Boschma and Frenken, 2011). Therefore, co-worker ties can help co-located former colleagues to share knowledge, which favours regional development. Despite the lasting characteristics of co-inventors have been found important for later patenting collaborations (Agrawal et al, 2006, Breschi and Lissoni, 2009) and the evidence of information diffusion in co-worker networks (Calvo-Armengol and Jackson, 2004, Granovetter, 1995), very limited research was devoted to the effect of co-worker networks in economic geography. We argue that the co-worker approach established in a previous paper (Lengyel and Eriksson, 2015) can be used to analyse the effect of knowledge externalities in regional development.

Similar ideas to the network-related learning have been present in the economic geography literature. For example, strong social ties within certain sectors in specialized industrial districts are claimed to enhance the prevalence of Marshallian externalities fostering incremental innovation and productivity growth (Amin, 2000, Asheim, 1996, Malmberg, 1997). In a similar fashion, diverse networks across industries in urban areas are often associated with Jacobsian externalities, thus potential new combinations of information, and radical innovation (Glaeser et al, 1992, Feldman, 1999). It has been shown that the above distinction between the prevailing and different knowledge externalities in metropolitan and specialized regions oversimplifies the actual processes for two reasons. First, Marshallian knowledge externalities might also operate within industries in large metropolitan areas (Kemeny and Storper, 2014) and second, diversity of networks per se cannot describe the extent of new knowledge creation because the relatedness of the sectors matter (Frenken et al, 2007). Although relatedness is a central concept in the recent literature of evolutionary economic geography (Neffke and Henning, 2013), the local effects of the links to related and unrelated industries remained unclear until now. In this paper we look at the role of Marshallian and Jacobsian externalities as well as the connections to related and unrelated industries in one empirical framework.

Based on a new probability measure of workplace-based acquaintance developed in a previous paper (Lengyel and Eriksson, 2015), we generate the co-worker network for the entire Swedish economy

1990-2008 tracing the most probable co-worker ties of every employee and from every year through the full period. We also compute the strength of individual ties by calculating the length of the period of co-working that enhances strength and diminish strength over time after the termination of co-worker status. As result, we get a dynamically changing social network, with many weak and few strong ties. The major promise of the new co-worker approach is its micro-perspective, which also allows us to aggregate tie weights on plant and industry-region level.

We claim to make three contributions to the existing literature. First, we find that a growing density of the plant network within industry-regions is positively linked to wage, having a stable effect in every type of regions and all models. Thus, Marshallian externalities operate within every indsuytry specialization regardless of region size. However, we also find that transitivity or triadic closure in the plant network has a negative effect on wage, which suggests that the combination of non-redundant knowledge is important within industries as well. Second, diversity of links to other industries has a negative effect on wage levels in every type of regions suggesting that knowledge externalities might rather occur across few but strong cross-industry links. These results also imply that Jacobsian externalities should not be derived from the diversity of cross-industry networks per se. Third, we show that links to unrelated industries are more important than to related industries in metropolitan regions; while links to related industries are only important in smaller regional centres. Thus, the co-worker approach demonstrates that only the related-unrelated dimension of cross-industry links can tell urban areas from smaller regions in terms of the prevailing knowledge externalities.

2. Methods of network creation and data processing

2.1 Probability and strength of co-worker ties

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} with strength W_{ij} even after the termination of the co-workership. The probability of the tie can be formulated as

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

where N denotes plant size, G $\in \{1, 2, \dots M\}$ denotes those characteristics we use for similarity measurement; 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. The formulation is based on an initial probability inversely proportional to plant size, which prevents all individuals from being isolated within the plant (Erdős and Rényi, 1959). This probability is increased when *i* and *j* are similar given certain individual characteristics, which is built on the homophily literature of social networks (Currarini et al, 2009, McPherson et al, 2001). Detailed discussion of the method and network generation can be found in Lengyel and Eriksson (2015).

In this paper, we introduce the strength of co-worker ties of individuals at distinct plants as the function of time of co-working and time spent after termination of co-workership, which can be formulated as

$$W_{ij}^{t} = W_{ij}^{0} \times e^{-\lambda t}, \qquad (2)$$

where

$$W_{ij}^{0} = \ln(\mathrm{t_l} - \mathrm{t_f} + 1). \quad (3)$$

Here, t_l refers to the last and t_f to the first year of co-worker status of L_{ij}; thus, W_{ij}^{0} denotes the strength of L_{ij} by the termination of the co-workership. Motivated by sociology literature, natural logarithm is used to index duration of co-workership because *"returns in terms of tie strength to increased duration of a relationship decline with increasing length of acquaintance"* (Marsden and Campbell 1984, p. 488). We apply exponential time decay in Eq. 2 for the calculation of W_{ij}^{t} reflecting the phenomena that the tie is losing from strength after the termination of co-workership but the slope of the decay becomes sharp only few years after the termination and smoothes out again later (Jin et al, 2001). A λ=0.05 was chosen for the exponential decay constant; consult Appendix 1 for more discussion on λ values and the visual representation of tie strength.

2.2 Network aggregation

The above methodology will generate a weighted individual-level co-worker network for every year that we can aggregate on plant, firm, industry or regional level by simply counting the individual links and summing their weights. In this paper, we first aggregate the weighted individual-level network on the plant level by summing up the weights of individual links between plants *k* and *l* by

$$W_{kl} = \sum_{i\in k, j\in l, k\neq l} W_{ij}. \quad (4)$$

We can also look at the unweighted version of the plant-level co-worker network, when edges between plant *k* and *l* are defined by

$$L_{kl}^{u} = \begin{cases} 1 \; if \; \sum_{i\in k; j\in l; k\neq l} L_{ij} \geq 1 \\ 0 \; otherwise \end{cases}. \quad (5)$$

The weighted network will be compared to the unweighted network in Section 3.1 in order to illustrate the importance of time-decay in the weight of individual edges.[1] Eq. 4 will provide us with the opportunity to zoom into the networks within industry-regions, which is important to identify mechanisms behind MAR externalities. In the next step, we aggregate the network on the level of industry-regions by summing up the weights of individual links between industry-regions *p* and *q* as follows

$$W_{pq} = \sum_{i\in p; j\in q} W_{ij}. \quad (6)$$

This network will be used to identify the relation of industry-regions to other industry-regions within the same region in order to identify Jacobs externalities and to capture knowledge spillovers across regions as well.

2.3 Data and processing

[1] We have also looked at an alternatively weighted plant network using edge weights defined by $L_{kl}^{w} = \sum_{i\in k; j\in l, k\neq l} L_{ij}$. This alternatively weighted network has mixed characteristics. It behaves similarly to the weighted network in terms of accumulated sum of weigths (Column 4 in Table 1) and the probability distribution (Figure 1C). However, the distribution of this alternative weighted degree resembles the unweighted distribution (Figure 1B).

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 network creation can handle, we exclude those without tertiary education from the data. 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. There remained 52,872 plants in 1990 and 113, 441 plants in 2008.[2]

We used three characteristics of employees to generate subgroups specified in Equation 1: direction of education, gender, and age. 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. Then, W_{ij} is given by Equation 2 and 3. For further information of group definitions and descriptive statistics as well as for a discussion, why 50 co-worker ties from each year were traced and in order to follow the steps of the network creation except tie strength calculation, consult Lengyel and Eriksson (2015).

In this paper, we aggregate the co-worker network on plant level and industry-region level and investigate the effect of the network on the growth of industry-regions. The number of plants varies considerably across industries; one can find most of the plants in services (most plants are in the sectors denoted by 2-digit NACE codes: 74, 80, 85, 50, 52) and the distribution of plants in manufacturing industries is a more even. Please, consult Appendix 2 for the distribution of plants across industries in 1992 and 2008.

3. Network description

3.1 The plant network

The co-worker network grows monotonically over the full period; more and more plants are part of the network in each of the subsequent years presented in Column 2 of Table 1. These plants are connected with a monotonically growing number of unweighted links (Column 3). However, the increase of the sum of weights defined by Eq. 4 seems to slow down after 1996 and there is even a maximum in year 2002, after which it decreases slightly (Column 4). Thus, there are more and more weak ties in the network and the average weight of plant edges are monotonically decreasing (Column 5).

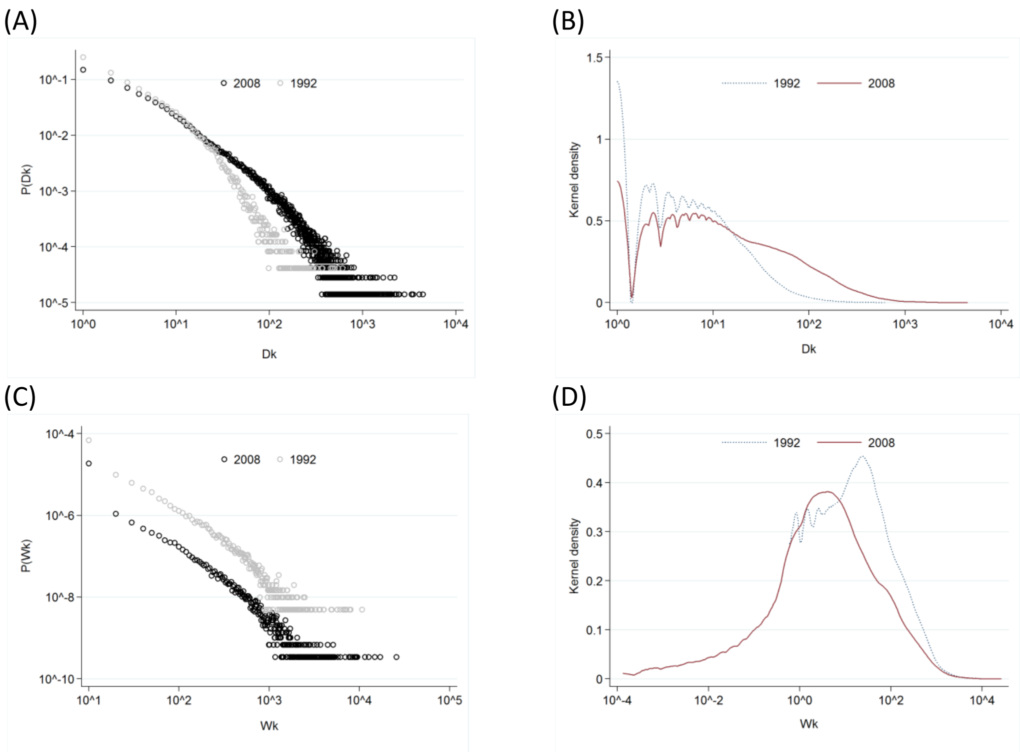
[2] The number of plants used for network creation is not identical with the number of plants that are actually in the network. This is due to the fact that the network is generated through labour mobility across plants.

Table 1. Nodes and links in the plant network, 1992-2008

Year	# plants	$\sum L_{kl}^{u}$	$\sum W_{kl}$	Avg. weight
1992	24,375	116,423	886,376.8	7.61
1994	32,617	256,321	1,611,590	6.29
1996	40,368	467,141	2,339,853	5.01
1998	44,001	609,346	2,312,204	3.79
2000	52,671	904,047	2,725,423	3.01
2002	57,376	1,128,767	2,732,069	2.42
2004	61,463	1,196,144	2,305,679	1.93
2006	66,508	1,246,456	1,959,751	1.57
2008	71,265	1,284,618	1,689,650	1.32

There is a considerable difference between the dynamics of the unweighted and weighted degree distributions over the period. On the one hand, the probability that the plant has links to exactly 1 other plant in the unweighted network is above 0.1 in both 1992 and 2008; and P(D_k) decreases identically up to degree 25 ($\approx 10^{1.4}$) in both years (Figure 1A). However, the probability of having higher degree than 25 rises from 1992 to 2008, which is because the maximum degree is higher in 2008 than in 1992 and the share of high degree plants is higher in 2008 than in 1992 (Figure 1B). On the other hand, the probability distribution of weighted degrees seems to have a very similar pattern in 1992 and in 2008 (Figure 1C). The curve is only shifted down, which is only due to the increased number of plants. Furthermore, and most importantly, the Kernel density visualization reveals a normal distribution of weighted degrees in 2008 (Figure 1D). This means that the consideration of time in edge weighting – both co-working and decay – normalize the weighted degree distribution automatically over the period.

Figure 1. Degree and weighted degree distributions of plants, 1992 and 2008

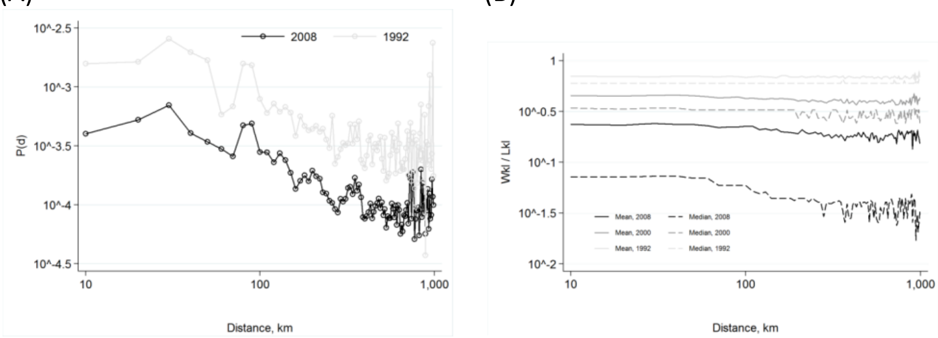


Note: **(A)** The probability that plant *k* has unweighted degree *D*, logarithmic scale. **(B)** Univariate Kernel density distribution of unweighted degree *D*, logarithmic scale. **(C)** The probability that plant *k* has weighted degree *W*, logarithmic scale. For practical reasons, weighted degrees were binned into 10^1 intervals for P(W_k) calculation. **(D)** Univariate Kernel density distribution of weighted degree *W*, logarithmic scale.

The pattern of tie probability as the function of distance effect on is very similar in the two years we look at; probabilities are shifted down due to the increased number of plants and consequently increased number of possible ties (Figure 2A).

Figure 2. The effect of distance, 1992 and 2008

(A) (B)



Note: **(A)** P(*d*) is the ratio between observed and possible ties at distance *d*. A 10 km resolution was used for binning distance distribution. **(B)** The average weight of plant ties at distance *d* was calculated by $\sum_d W_{kl}/\sum_d L_{kl}^u$. A 10 km resolution was used for binning distance distribution. Year 2000 was added to the plot in order to visualize temporal changes more in detail.

One finds that both the median and the mean of average weights are almost constant and therefore independent from distance in 1992 (Figure 2B). The indicators remain constant until a certain distance – around 100 km – in 2008 as well and decrease only slightly in larger distances. However, the gap between the mean and the median opens up over the period denoting a left-skewed distribution with relatively few and outlier strong ties. Furthermore, the gap opens up even more in larger distances. In other words, the majority of distant co-worker ties are weak and they are getting weaker over the period but there are also strong distant ties, due to recent labour flows.

Table 2. The number and weight of plant ties within regions and industries, 1992 and 2008

	1992		2008	
	$\sum L_{kl}^u$	$\sum W_{kl}$	$\sum L_{kl}^u$	$\sum W_{kl}$
Across regions	38,551 (33%)	174,401.8 (20%)	520,309 (41%)	415,773.5 (25%)
Within regions	77,872 (67%)	711,974.9 (80%)	764,309 (59%)	1,273,877 (75%)
SUM (100%) of links	116,423	886,376.8	1,284,618	1,689,650
Across industries in the region	54,227 (70%)	386,536.7 (54%)	612,856 (80%)	823,038 (65%)
Within industries in the region	23,645 (30%)	325,438.2 (46%)	151,453 (20%)	450,839 (35%)
SUM (100%) of links within regions	77,872	711,974.9	764,309	1,273,877

Note: regions denote the 72 functional regions in Sweden and are equivalent to labour market areas. Industries are defined by 4-digit NACE codes.

Around two third of the links and even higher rate of weights are concentrated within functional regions in the beginning of our investigation (Table 2). These shares are still high in 2008 despite their slight decrease that is due to inter-regional labour flows over the period. One third of local ties remain within industry borders in 1992; this share decreases over time. The edges within industries are stronger on average than the edges across industries.

In sum, there is a considerable share of the plant network within industries in regions. We will play special attention to these networks by zooming into industry-regions and creating network indicators in Section 4, because these are the fields of specialized forms of knowledge spillovers and MAR local externalities. However, we also observe a larger share of links across industries in regions, which enables us to investigate the role of Jacobs externalities and knowledge spillovers across industries. This latter phenomenon will be addressed by looking at a network defined by Eq. 5 where not plants but industry-regions are the nodes.

3.2 The industry-region network

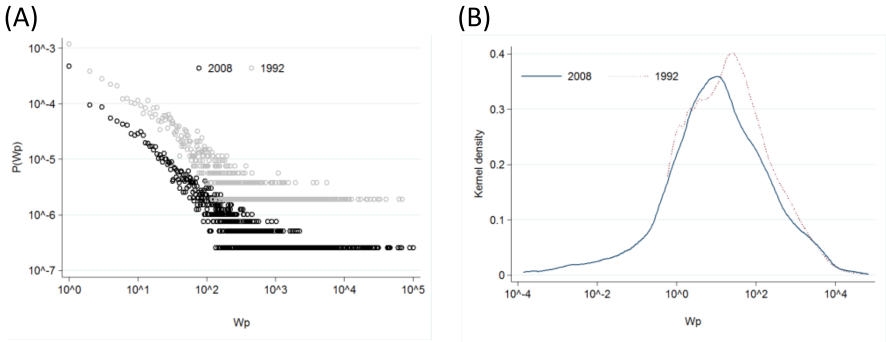
The dynamics of the industry-region network is very similar to the plant network. The number of industry-regions and the number of edges are growing over the full period but the sum of the weights takes its maximum at the half of the period (Table 3). Like in the plant network, the average weight across industry-regions decreases monotonically.

Table 3. Nodes and links in the industry-region network, 1992-2008

Year	# industry-region	$\sum L_{pq}$	$\sum W_{pq}$	Average Weight
1992	3,795	33,053	885,774.4	26.79
1994	4,576	62,877	1,611,118	25.62
1996	5,263	102,694	2,339,207	22.79
1998	5,391	124,481	2,311,708	18.57
2000	5,926	169,278	2,724,814	16.09
2002	6,343	201,759	2,731,400	13.54
2004	6,528	214,269	2,305,163	10.76
2006	6,822	225,639	1,959,143	8.68
2008	7,211	238,071	1,688,141	7.09

The dynamics of weighted degree distribution of the industry-region network is very similar to the plant network (Figure 3). The probability distribution of weighted degrees seems to have a very similar pattern in 1992 and in 2008 (Figure 3A); the curve is shifted down from the beginning to the end of the period, just like in the plant network. The weighted degree distribution of the industry-region network becomes normal over the period (Figure 3B), just like in the plant network.

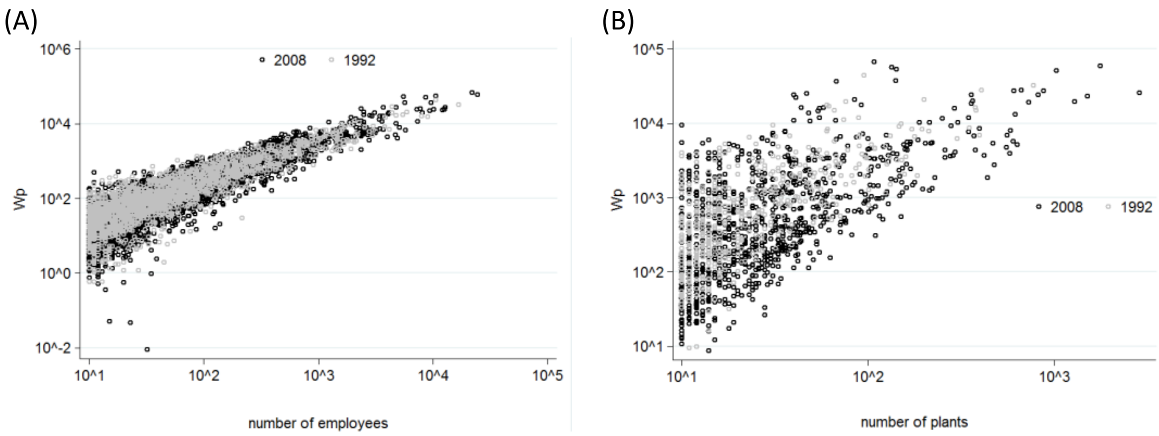
Figure 3. Weighted degree distributions of industry-regions, 1992 and 2008



Note: (A) Weighted degree distribution of industry-regions in 1992 and 2008. The integer of weights has been used for binning. (B) Univariate Kernel density distribution of industry-region weighted degree, logarithmic scale.

However, the weighted degree of industry-region *p* highly depends on the size of the industry-region. For example, the number of employees (Figure 4A) and also the number of plants (Figure 4B) in the industry-region is positively associated with the weighted degree.

Figure 4. The strength of industry-regions in the co-worker network as the function of their size, 1992 and 2008

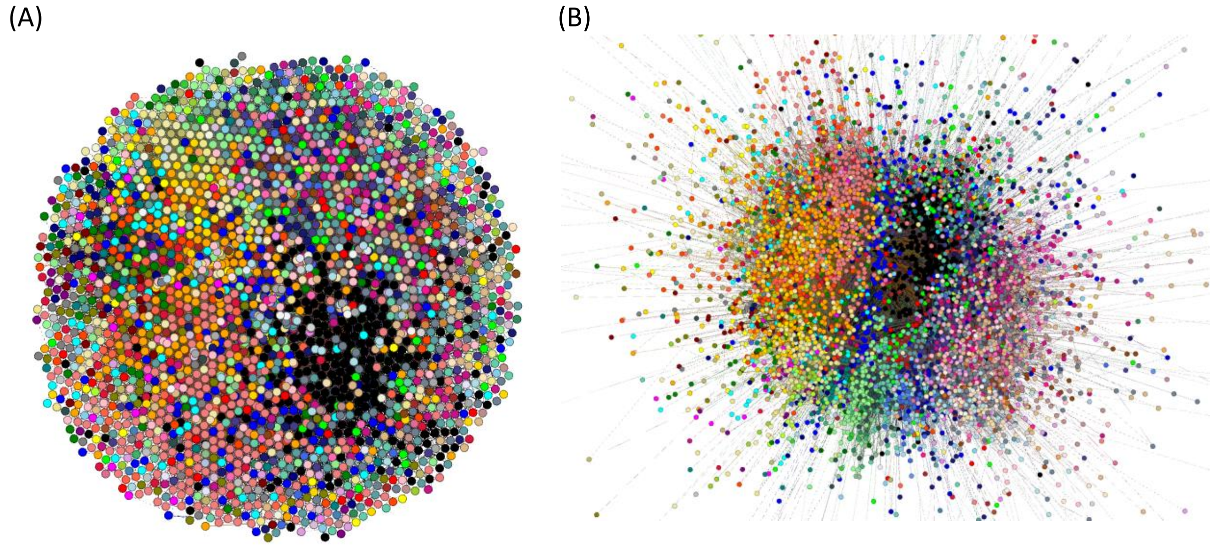


Note: (A) Strength of industry-regions by edge weights versus the number of employees by, 1992 and 2008. (B) Strength of industry-regions by edge weights versus the number of plants, 1992 and 2008.

We plotted the weighted networks in years 1992 and 2008 (Figure 5). Since we used a spring algorithm, in which strongly related industry-regions are pulled together, and used identical colours for regions; the co-location of same coloured nodes in the plot illustrates the importance of co-location of industries in regions. In other words, co-located industries seem to develop strong ties over time in the weighted and weight-normalized network as well. We will investigate these spatial networks by creating node characteristics of industry-regions in order to reveal the effect of Jacobs externalities.

However, there are strongly connected industry-regions that are not co-located. Inter-regional ties might be also important for knowledge flows and collecting new ideas. Thus, we will create another set of node characteristics looking at inter-regional ties of industry-regions.

Figure 5. Industry-region network, 1992 and 2008



Note: Self-loops have been filtered out and only the giant component is visualized. Identical colors were used to illustrate regions (e.g. Stockholm is black). A spring algorithm was used to pull those industry-regions together. (A).The weighted industry-region network, 1992. Only the edges stronger than 1 have been illustrated. (B) The weighted industry-region network, 2008. Only the edges stronger than 1 have been illustrated.

To sum up, the weighted co-worker network across plants and industry-regions promises us interesting insights into hidden mechanisms behind local knowledge externalities, which we discuss in the subsequent section. The plant-level network will be used to create indicators of MAR externalities, while the industry-region network will be used to test the prevalence of Jacobs externalities and skill-relatedness.

4. Results

Fixed effect (FE) panel regression models are used in this section to explore the statistical relation between indicators derived from the plant-level and industry-level co-worker network and wage levels in industry-regions. In simple form, the equation could be specified as:

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

where *y* denotes gross income per capita (WAGE), *t* denotes one-year intervals from 1992 to 2008, *i* denotes the industry-region, *X* stands for the set of explanatory variables, *Z* stands for a set of control variables of our base model described below 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 industry-regions such as local labour market conditions not captured by the controllers or by the definition of industries and 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 wage level indicators as explained above, the risk of unobserved time-specific heterogeneity and reversed causality influencing the results was also reduced. We present the results of between effect (BE) models as well, when the argument demands it, and in order to provide a full picture of the empirical case.

The applied control variables are provided by a base model that estimates the effect of absolute specialization on wage levels (Kemeny and Storper, 2014). Thus, we use the number of plants in the industry-region as a measure of absolute specialization (ABSSPEC) and also the number of employees with bachelor degree of higher as a measure of regional size (REGSIZE). The square of both of the previous indicators (ABSSPEC-SQ, REGSIZE-SQ) are applied in the models to control for non-linear effect of specialization and regional size. The average size of plants (AVGSIZE) is included in the models as well. Network indicators are introduced in the according subsections. All control variables are log-transformed. Consult Appendix 3 for descriptive statistics and pooled pairwise Pearson correlation of indicators.

We include all industries in the presented models because potential knowledge externalities are not limited this way as compared to excluding any of the sectors. We tested the models by looking only on tradable industries defined for Sweden by Eliasson et al. (2010), which did not change our main findings. The estimation strategy follows the papers' line of argument. We first estimate the effect of the plant-network on wage levels in industry specializations. Then, we estimate the effect of networks across industry-regions. In all the tables presented below, we first run the regression for the full set of industry-regions; then we split the sample according to the size of functional regions in order to unveil regional differences.

4.1 The plant-level network and Marshall-Arrow-Romer externalities

In this subsection we uncover the effect of network density and triadic closure in the plant-level co-worker network on regional development. Density is the simplest network indicator; it is measured by the share of observed links as compared to the number of possible links within industry regions and is formulated by

$$DENS_{pt} = \frac{2 \times L_{pt}}{N_{pt} \times (N_{pt} - 1)} \;; \qquad (8)$$

where L_{pt} is the number of observed plant-plant links and N_{it} is the number of plants in the industry-region p at year t. According to the usual claim concerning social network density and diffusion, one can presume that density favours MAR knowledge externalities because a large share of observed links among possible links fastens information flow within industry-regions.

Transitivity (triadic closure, global clustering) is another basic indicator that characterizes social networks. The measurement of the global transitivity calculates the number of triangles among all possible triangles and is formulated by

$$\mathrm{Trans}_{pt} = \frac{\sum_k \#\{lo \in pt | l \neq o,\ \ l \in N_k(pt),\ \ o \in N_k(pt)\}}{\sum_k \#\{lo | l \neq o,\ \ l \in N_k(pt),\ \ o \in N_k(pt)\}}; \qquad (9)$$

where k, l and o are plants in industry-region p at year t. The indicator measures the likelihood that plant l and o are connected if k has a link to both l and o. The connection between transitivity and local knowledge externalities is not trivial; the indicator might have either positive or negative effect on regional development based on two counteracting mechanisms. Triadic closure favours trust and helps collaboration of plants in the industry-region. On the contrary, a high level of transitivity also means that the knowledge-base of plants is overlapping and therefore, the possibility of new knowledge combination in the industry region is limited.

Table 4 presents the results of the fixed-effect regressions. Model 1 consists the co-efficients of the base model; the two network variables are inserted into Model 2; and the sample is split by region types in Models 3-6. The number of observations falls sharply from Model 1 to Model 2 because the indicators can be calculated only for those industry-regions with at least three plants.

Table 4. Wage level and the plant-network within the industry-region

	All regions	All regions	Metropol. regions	Regional centres	Middle-sized regions	Small regions
	FE	FE	FE	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
DENS		0.400***	0.432***	0.454***	0.219***	0.323***
		(0.022)	(0.041)	(0.031)	(0.061)	(0.101)
TRANS		-0.095***	-0.060***	-0.108***	-0.092***	-0.049
		(0.012)	(0.022)	(0.017)	(0.034)	(0.055)
ABSSPEC	0.057***	0.532***	0.658***	0.678***	0.695***	1.824***
	(0.007)	(0.052)	(0.112)	(0.085)	(0.164)	(0.387)
ABSSPEC-SQ	0.021***	-0.025***	-0.028***	-0.038***	-0.041***	-0.166***
	(0.001)	(0.004)	(0.007)	(0.007)	(0.014)	(0.036)
AVGSIZE	0.000	-0.006***	-0.023*	-0.000	0.000	-0.006**
	(0.001)	(0.002)	(0.013)	(0.002)	(0.004)	(0.003)
REGSIZE	0.541***	0.174*	0.701	0.245	-1.988***	-2.525***
	(0.056)	(0.101)	(0.455)	(0.441)	(0.520)	(0.591)
REGSIZE-SQ	-0.031***	0.004	-0.009	-0.008	0.143***	0.205***
	(0.002)	(0.004)	(0.017)	(0.024)	(0.034)	(0.050)
CONSTANT	6.551***	6.051***	0.580	5.821***	14.242***	13.292***
	(0.332)	(0.650)	(3.274)	(2.057)	(2.136)	(1.991)
YEAR FE	Yes	Yes	Yes	Yes	Yes	Yes
R-sq	0.108	0.156	0.176	0.191	0.124	0.110
N	84,804	14,676	4,487	7,320	2,103	766

Note: Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

The estimation results indicate a significant positive effect of DENS and a significant negative effect of TRANS both of which are stable across the models (transitivity looses significance in the smallest regions in Model 6). Due to the fixed-effect setting, these results imply that a growing co-worker network density is positively linked to wage levels because knowledge can flow more efficiently within the industry if new links across plants emerge over time. The same pattern of the density effect on regional development suggests that not only specialized regions can be characterized by Marshallian externalities but MAR type of learning prevails within industry specializations in urban agglomerations as well (Kemeny and Storper, 2014).

The negative effect of TRANS is a very interesting and important finding and implies that the establishment of a new tie harms regional development if it closes a triangle in the network. In other words, regional development slows down when a new direct connection is established between two plants that were already connected indirectly with a mediation of a third plant because the redundancy of the knowledge-base increases. Therefore, the robust negative effect suggests that the novel combinations of non-redundant knowledge might be important in Marshallian type industry-specializations as well.

The main controllers (ABSSPEC and REGSIZE) have the expected positive effect; however, we find a non-linear effect of both of these indicators because the squares turn to have a significant effect as well. REGSIZE looses significance when the sample is split to region types due to the small variation of the indicator. AVGSIZE has the expected negative sign in Models 2, 3 and 6 that is consistent with previous findings in Sweden.

4. 2 The industry-region network, Jacobs externalities and skill-relatedness

We explore the effect of intra-industry ties on regional development in two steps. First, we look at the connection between wage levels and the diversity of links of the industry-regions towards other industry-regions. Then, we also introduce the importance of links to skill-related industries to the model.

The first indicator we define is the I-E index of industry-regions; this is the simple quotient of internal (I) links compared to the rate of external (E) links; therefore quantifies the openness of the industry-region and is formulated by

$$IE_{pt} = \frac{\sum_{\mathrm{p=q}} \mathrm{W_{pq}}}{\sum_{\mathrm{p\neq q}} \mathrm{W_{pq}}} \ . \qquad (10)$$

We also create three diversity indicators using Shannon type of entropy calculation that capture different aspects of the relation between inter-industry links and regional development. DIV(IN) measures the diversity of links of industry-region *p* at year *t* within the region and is formulated by

$$DIV(IN)_{pt} = \sum_{p\in s}^{q\in s} \frac{W_{pq}}{W_p} \times \log(\frac{W_{pq}}{W_p}) \ ; \qquad (11)$$

where W_{pq} is the accumulated strength of ties between industry-region *p* and *q*, W_p is the strength of industry-region *p* and s denotes the functional region of industry-regions *p* and *q*. In a similar manner, one can also formulate DIV(OUT) as the diversity of links across regional borders by

$$DIV(OUT)_{pt} = \sum_{p\in s,\in t}^{q\notin s,\notin \mathrm{t}} \frac{W_{pq}}{W_p} \times \log(\frac{W_{pq}}{W_p}) \ ; \qquad (12)$$

where *t* is the 4-digit industry of industry-region *p*. Furthermore, we also define DIV(OUTSAME) that is the diversity of links to the same industries in distinct locations formulated by

$$DIV(OUTSAME)_{pt} = \sum_{p\in s,\in t}^{q\notin s,\in \mathrm{t}} \frac{W_{pq}}{W_p} \times \log(\frac{W_{pq}}{W_p}) \ . \qquad (13)$$

We log-transformed the above diversity indicators, because only a very few of industry-regions have high number of connections, which eventually gives us a very long-tailed distribution. The minimum value of entropy is zero; while the indicator takes its' maximum when the weights of the industry-region is equally distributed among the connected industry-regions, which is the most diverse state of the distribution. Therefore, the positive effect of the above diversity indicators would suggest a positive relation between diversity and regional development; whereas a negative sign means the importance of the concentration of weights in certain links (Eagle et al, 2010).

Table 5. Wage level and links to other industries

	All regions	Metropol. regions	Regional centres	Middle-sized regions	Small regions
	FE	FE	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)
IE	-0.296***	-0.228***	-0.397***	-0.187*	-0.076
	(0.044)	(0.084)	(0.062)	(0.112)	(0.153)
DIV(IN)	-0.021***	-0.012	-0.023***	-0.026***	-0.017**
	(0.003)	(0.008)	(0.004)	(0.006)	(0.007)
DIV(OUT)	-0.036***	-0.064***	-0.025***	-0.038***	-0.025**
	(0.004)	(0.008)	(0.005)	(0.009)	(0.012)
DIV(OUTSAME)	0.020***	0.024***	0.020***	0.019***	0.019***
	(0.002)	(0.004)	(0.003)	(0.005)	(0.007)
ABSSPEC	0.047***	0.086***	0.010	0.001	0.280***
	(0.014)	(0.033)	(0.020)	(0.034)	(0.057)
ABSSPEC-SQ	0.017***	0.013***	0.021***	0.021***	-0.005
	(0.001)	(0.003)	(0.002)	(0.004)	(0.007)
AVGSIZE	0.000	0.002	-0.000	0.004	-0.002
	(0.001)	(0.012)	(0.002)	(0.003)	(0.002)
REGSIZE	0.441***	-0.284	0.337	0.111	-0.965**
	(0.077)	(0.435)	(0.351)	(0.373)	(0.425)
RERSIZE-SQ	-0.025***	0.011	-0.013	-0.007	0.070*
	(0.003)	(0.017)	(0.019)	(0.025)	(0.036)
Constant	7.296***	10.570***	7.298***	8.928***	12.354***
	(0.453)	(3.060)	(1.626)	(1.453)	(1.333)
Year FE	Yes	Yes	Yes	Yes	Yes
R-sq	0.109	0.126	0.108	0.106	0.109
N	43,018	10,227	21,454	7,716	3,621

Note: Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Regression results containing the above IE and diversity variables are presented in Table 5, in which we run the regression for the full set of industry-regions first and then split the sample by region types. The findings imply that the openness of the industry-region is positively linked to wage levels. Thus, inter-industry knowledge externalities are important for regional development and prevail in all kinds of regions.

The positive effect of DIV(OUTSAME) is expected because it means that the access to the same industry – and same knowledge – in diverse locations favours regional development, which is consequent with previous findings (Boschma et al, 2009). However, the negative effect of both DIV(IN) and DIV(OUT) indicators implies that not a diverse distribution but the concentration of weights across links to certain industries matters within and also across regional borders in every region types. This novel finding suggests that diversity per se cannot describe knowledge externalities in urban agglomerations, because the industries in place may benefit more from certain industries than from others. We argue in the remaining part of this section that the nature of these important links distinguishes urban knowledge externalities from externalities prevailing in specialized regions.

The concept of skill-relatedness is used to characterize the links across industries (Neffke and Henning, 2013); when industries are considered to exploit similar knowledge base if the labor flow between them is higher than the expected labor flow given the size of the industry. Although skill-relatedness is a central concept in contemporary evolutionary economic geography, the local effect of skill-related links is still unclear because one has to calculate relatedness on an aggregate country level and therefore it is difficult to develop a theory of knowledge externalities from relatedness per se.

We calculate the skill-relatedness of Swedish industries for 1990-1999 and 2000-2008 by following the formulation of Neffke et al. (2015):

$$R_{pq} = \log\frac{F_{pq}F_{..}}{F_{p.}F_{.q}} \; ; \qquad (14)$$

where F_{pq} is the observed number of flows between industry *p* and *q*, $F_{..}$ is the total number of flows, $F_{p.}$ is the number of workers leaving industry *p* and $F_{.q}$ is the number of workers joining industry *q*. The log-transformation means that a positive value of R_{pq} denotes relatedness of industries *p* and *q*, while a negative value means that *p* and *q* are unrelated.

Given the micro-perspective of the co-worker network approach, we can quantify the importance of related and unrelated ties as the function of region type. To do that, we simply calculate the quotiand of accumulated weights in related links as opposed to the weights in unrelated industries within the region and also across regional borders by

$$RELQ(IN)_{pt} = \frac{\sum_{p\in s}^{q\in s} W_{pq}\{pq \in s|\mathrm{R_{pq}}>0\}}{\sum_{p\in s}^{r\in s} W_{pr}\{pr \in s|\mathrm{R_{pr}}\leq 0\}}; \qquad (15)$$

and

$$RELQ(OUT)_{pt} = \frac{\sum_{p\in s}^{q\notin s} W_{pq}\{pq \notin s|\mathrm{R_{pq}}>0\}}{\sum_{p\in s}^{r\notin s} W_{pr}\{pr \notin s|\mathrm{R_{pr}}\leq 0\}}; \qquad (16)$$

where *s* denotes the region of industry-region *p*; *p* and *q* are skill-related; *p* and *r* are unrelated, W_{pq} is the accumulated tie weights between industry-regions *p* and *q*, and W_{pr} is accumulated tie weights between industry-regions *p* and *r*.

Consequently, RELQ(IN) and RELQ(OUT)indices condense two different mechanism in one indicator. If the sign of the estimator is positive in the regression models, then the share of related ties is important for the development of industry-regions. If the sign of the estimator is negative, then unrelated ties are more important for regional development than related ties.

Table 6. Wage level and the importance of skill-relatedness

	All regions	Metropolitan regions	Regional centres	All regions	Metropolitan regions	Regional centres
	FE	FE	FE	BE	BE	BE

	(1)	(2)	(3)	(4)	(5)	(6)
RELQ(IN)	-0.002	-0.007*	0.002	0.013**	-0.035*	0.017**
	(0.002)	(0.004)	(0.003)	(0.005)	(0.021)	(0.007)
RELQ(OUT)	0.001	0.007*	-0.005	0.019**	0.009	0.012
	(0.003)	(0.004)	(0.004)	(0.008)	(0.020)	(0.010)
IE	-0.402***	-0.231***	-0.562***	0.614**	1.373**	0.079
	(0.060)	(0.085)	(0.091)	(0.252)	(0.540)	(0.341)
DIV(IN)	-0.025***	-0.004	-0.025***	-0.067***	-0.144***	-0.056***
	(0.005)	(0.010)	(0.006)	(0.011)	(0.038)	(0.014)
DIV(OUT)	-0.049***	-0.071***	-0.031***	-0.128***	-0.237***	-0.100***
	(0.006)	(0.009)	(0.009)	(0.016)	(0.041)	(0.021)
DIV(OUTSAME)	0.019***	0.025***	0.016***	0.107***	0.109***	0.094***
	(0.003)	(0.004)	(0.004)	(0.009)	(0.021)	(0.011)
ABSSPEC	-0.004	0.125***	-0.093**	0.292***	0.297***	0.233***
	(0.026)	(0.042)	(0.038)	(0.027)	(0.072)	(0.048)
ABSSPEC-SQ	0.023***	0.010***	0.034***	-0.000	-0.003	0.017***
	(0.002)	(0.003)	(0.004)	(0.003)	(0.007)	(0.006)
AVGSIZE	-0.008***	-0.005	-0.009***	0.014**	0.047	0.021***
	(0.003)	(0.012)	(0.003)	(0.006)	(0.034)	(0.008)
REGSIZE	0.855***	-0.024	0.141	0.260	2.664	-0.289
	(0.133)	(0.448)	(0.539)	(0.494)	(3.781)	(2.020)
RERSIZE-SQ	-0.035***	0.002	0.011	-0.021	-0.132	0.014
	(0.005)	(0.018)	(0.029)	(0.028)	(0.169)	(0.111)
CONSTANT	4.442***	8.789***	7.774***	6.079***	-8.663	8.253
	(0.901)	(3.142)	(2.542)	(2.046)	(21.248)	(9.065)
Year FE	Yes	Yes	Yes	No	No	No
Region FE	No	No	No	Yes	Yes	Yes
Industry FE	No	No	No	Yes	Yes	Yes
R-sq	0.108	0.125	0.106	0.813	0.926	0.846
N	23,698	9,391	11,495	23,698	9,391	11,495

Note: Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 6 consists regressions for the whole set of industry-regions and the subsamples of metropolitan regions and regional centres. Model 2 imply that unrelated ties across industries are more important in urban agglomerations than related links. This important findings suggests that novel combination of knowledge is more likely in metropolitan regions with a high degree of unrelated variety than in other locations (Frenken et al, 2007). Furthermore, ties to related industries in other locations are more important for development than ties to unrelated industries in other locations, which is consistent with previous findings (Boschma et al. 2009).

We ran between-effect (BE) regressions with region and industry fixed-effects as well in order to provide a more detailed picture about the relatedness effect. The comparison of Model 5 and 6 reveals the different nature of knowledge externalities prevailing in urban areas versus the ones operating in smaller regional centres. Those urban industries are more developed that are linked with higher shares of weights to unrelated ties than an average of urban industry. Similarly, those industries in regional centres are more developed that are linked with higher shares of weights to skill-related ties than an average of industry in regional centres.

We claim on the basis of the findings presented in this subsection that not diversity per se – as suggested by the Jacobs-type of literature – distinguishes urban knowledge externalities from the ones that prevail in smaller more specialized regions. It is the links to unrelated and related industries that tells knowledge externalities in metropolitan areas from smaller towns.

5. Discussion

The paper provides a new empirical perspective for analyzing the role of social networks in regional development. We argue that co-worker networks are important tools for knowledge transfer across plants and companies because co-workers develop mutual trust and cognitive proximity, which can be maintained even after the termination of the co-workership. By generating a large-scale co-worker network and analyzing its' effect on regional development, we illustrate that the network indeed gives us a previously unprecedented insight to the operation of local knowledge externalities.

We find that the density of the network within industry-regions favours their development because knowedge might flow faster in environment with denser social networks, which is consequent with the previous literature on MAR externalities. However, we also find that triadic closure is negatively associated with wage levels, which suggests that redundant links and consequently redundant knowledge harms development, and novel combinations of knowledge might be important for industry specializations as well.

When looking at inter-industry links, we find that it is not diversity per se that distinguishes knowledge externalities across region types. We provide evidence that the share of unrelated ties, and consequently the likelihood of new knowledge combinations, favours development of industries in urban areas; whereas strong links to related industries are important for industry development in smaller regional centres.

The presented empirical case is not without problems. Endogeneity issues might have remained in the regressions even though we used lagged values of the independent variables. This is because a large share of co-worker ties are generated by labour mobility, as it was discussed in details in Lengyel and Eriksson (2015), and because wage level differences are important forces behind labor mobility. Therefore, there is a potential causality problem that the next versions of the paper shall solve.

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 further 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. One might be also 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. 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

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.

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

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

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.

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

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.

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.

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

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

Eagle, N., Macy, M., Claxton, R. (2010) Network diversity and economic development. *Science,* 328: 1029-1031.

Eliasson, K., Hansson, P., Lindvert, M. (2010) Jobs and exposure to international trade within the service sector in Sweden. DNR 2010/256, Swedish Agency for Growth Policy Analysis.

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

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: 685–697.

Glaeser, E.L., Kallal, H. D., Scheinkman, J. D., Shleifer, A. (1992): Growth in Cities. *Journal of Political Economy*, 100: 1126-1152.

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

Granovetter, M. (1995), *Getting a Job: a Study of Contacts and Careers*. Chicago: The University of Chicago 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. (2014) Is specialization good for regional economic development? *Regional Studies*, 49: 1-16.

Lengyel, B., Eriksson, R.H. (2015), Co-worker networks and productivity growth in regions. Papers in Evolutionary Economic Geography #15.13.

Marsden, P.V., Campbell, K.E. (1984) Measuring tie strength. *Social Forces,* 63, 2: 482-501.

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

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

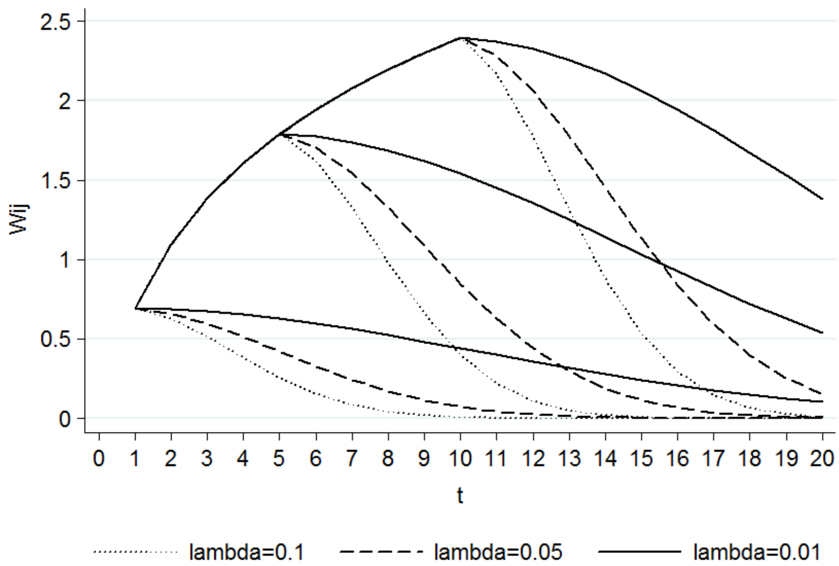
Neffke, F., Henning, M. (2013), Skill relatedness and firm diversification, *Strategic Management Journal*, 34: 297-316.

Neffke, F., Hidalgo, C., Otto, A., Weyh, A. (2015) Inter-industry labor flows. Mimeo.

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

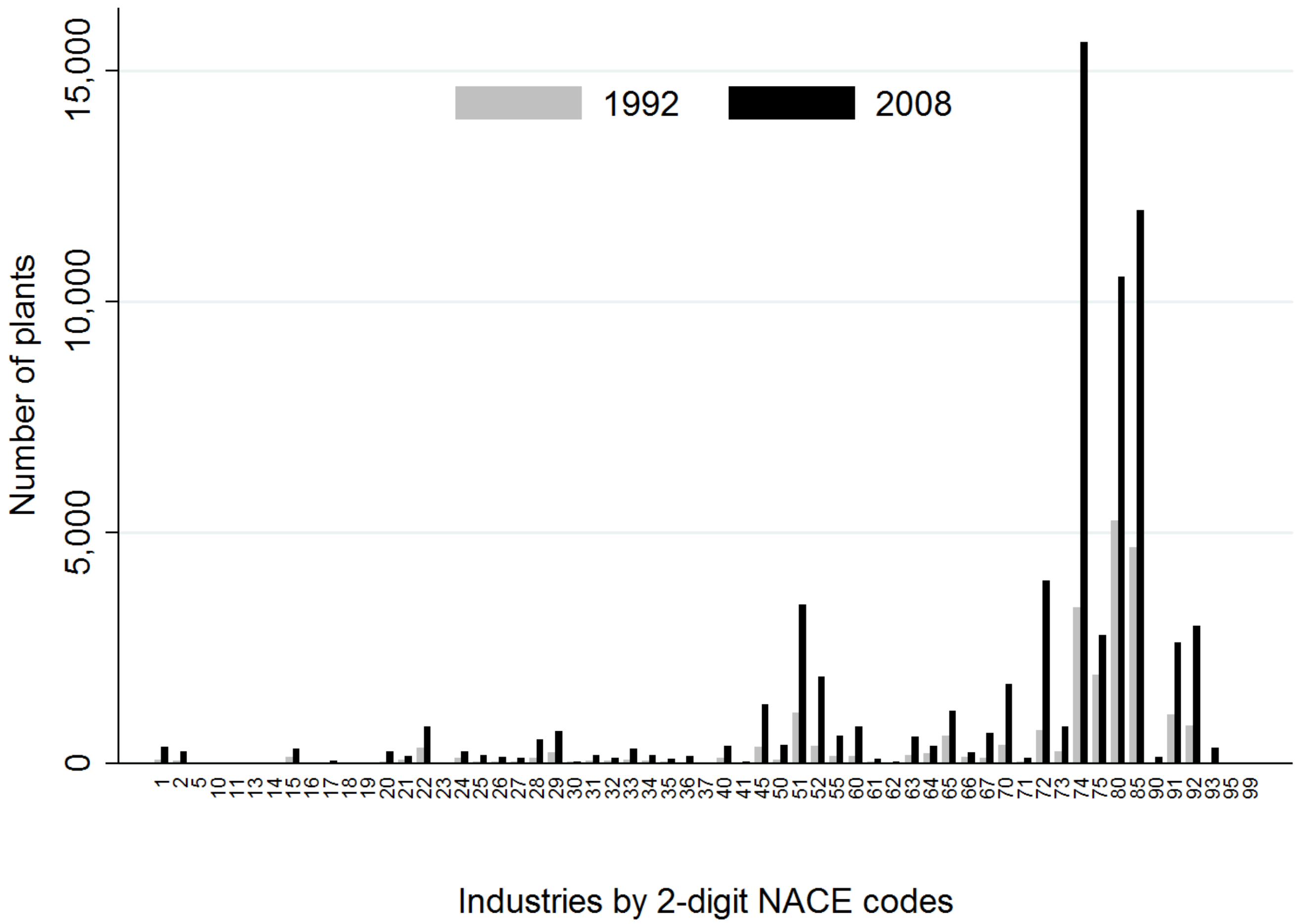
Ter Wal, A., Boschma, R.A. (2009), Applying social network analysis in economic geography: framing some key analytic issues, *Annals of Regional Science*, 43: 739-756.

Appendix 1. Lambda selection for the time decay function in the co-worker tie strength



We demonstrate tie strength for three scenarios, when *i* and *j* are working together for (1) one year, (2) five years, and (3) ten years: the solid ascending line represents the strengthening ties and this line is breaked at t=1, t=5, and t=10. Then, we introduce three different exponential time decay curves that have three different lambda values. Jin et al. (2001) used lambda=0.01 for simulating network dynamics; however, this might be not efficient in our case because we have only few time steps (maximum 20). On the contrary, lambda=0.1 would produce too sharp decay for ties established over a long period of co-working. If lambda=0.05, the tie strength of the scenarios drops to a pre-defined level of strength by reasonably different time scales. For example, for those working together for one year only tie strength will fall below 0.5 by the fourth year after co-workership termination, whereas seven years and eight years are needed to drop to that level for those working together for five years or ten years, respectively.

Appendix 2. Industry representation in the plant network



Appendix 3. Indicators, descriptive statistics and correlations

	Variable	Obs	Mean	Std. Dev.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13
1	WAGE	101386	9.447	1.191	7.006	15.198	1.0000												
2	IE	101386	0.024	0.074	0	0.999	0.238*	1.0000											
3	DENS	25006	0.579	0.368	0.002	4	0.027*	-0.127*	1.0000										
4	TRANS	15752	0.382	0.312	0	1	0.237*	0.277*	0.327*	1.0000									
5	DIV-IN	65955	-0.200	1.656	-3.132	11.609	-0.379*	-0.241*	0.348*	-0.012	1.0000								
6	DIV-OUT	77494	-0.071	1.518	-2.912	11.651	-0.444*	-0.253*	0.413*	-0.087*	0.491*	1.0000							
7	DIV-OUTSAME	59442	3.138	1.913	-1.021	12.341	0.352*	0.173*	-0.229*	0.026*	-0.350*	-0.393*	1.0000						
8	RELQ-IN	36958	-2.064	3.118	-17.496	9.935	-0.120*	-0.242*	0.376*	-0.094*	0.187*	0.381*	-0.295*	1.0000					
9	RELQ-OUT	63232	-0.505	1.991	-12.880	11.399	0.132*	0.080*	0.091*	0.059*	0.011	0.098*	-0.117*	0.368*	1.0000				
10	ABSSPEC	101386	4.315	1.870	0	11.177	0.596*	0.354*	-0.509*	0.068*	-0.551*	-0.500*	0.370*	-0.240*	0.054*	1.0000			
11	ABSSPEC-SQ	101386	22.120	16.519	0	124.931	0.585*	0.409*	-0.507*	0.060*	-0.555*	-0.517*	0.391*	-0.276*	0.039*	0.956*	1.0000		
12	AVGSIZE	101386	31.822	6.373	14.373	64.458	0.037*	0.023*	0.089*	0.073*	0.161*	0.083*	-0.202*	-0.068*	0.021*	0.016*	-0.008	1.0000	
13	REGSIZE	101386	8.793	1.690	2.303	12.528	0.113*	0.087*	-0.252*	-0.077*	-0.395*	-0.173*	0.181*	0.186*	0.089*	0.359*	0.383*	-0.135*	1.0000
14	REGSIZE-SQ	101386	80.168	29.968	5.302	156.95	0.110*	0.089*	-0.245*	-0.078*	-0.400*	-0.178*	0.187*	0.193*	0.091*	0.361*	0.392*	-0.194*	0.991*

Note: * denotes that the co-efficient of pair-wise Pearson correlation is significant on the 1% level.