Open access • Journal Article • DOI:10.1002/SMJ.2014

Skill relatedness and firm diversification — Source link

Frank Neffke, Martin Henning

Institutions: Erasmus University Rotterdam, University of Gothenburg

Published on: 01 Mar 2013 - Strategic Management Journal (John Wiley & Sons)

Topics: Human capital, Diversification (marketing strategy) and Interpersonal ties

PAPERS on Economics & Evolution

MAX-PLANCK-GESELLSCHAFT

0906

Skill-relatedness and firm diversification

by

Frank Neffke
Martin Svensson Henning

The *Papers on Economics and Evolution* are edited by the Evolutionary Economics Group, MPI Jena. For editorial correspondence, please contact: evopapers@econ.mpg.de

ISSN 1430-4716

Max Planck Institute of Economics
Evolutionary Economics Group
Kahlaische Str. 10
07745 Jena, Germany
Fax: ++49-3641-686868

Skill-relatedness and firm diversification

Frank Neffke* and Martin Svensson Henning**

Version May 2009

ABSTRACT

The concept of "relatedness" between industries plays an increasingly central role in economics and strategic management. However, relatedness has remained rather elusive in empirical terms. In this article, we investigate relatedness between industries in terms of the extent to which the same human capital can be employed in different industries. In particular, we investigate the skill-relatedness among different industries by investigating labor flows between industries.

The data used are Swedish employer linked data on individuals. Our statistical framework assesses the degree to which labor flows between pairs of industries are in excess of expected levels and use this as a quantification of Revealed Skill Relatedness. A network picture of 435 4-digit industries and the relatedness linkages between them shows that the relations among industries are far more complex than the industrial classification system suggests. Moreover, when investigating corporate diversification, we find that firms are far more likely to diversify into industries that are strongly skill-related to their core activities industries than into unrelated industries.

* Department of Applied Economics, Erasmus University Rotterdam. PO Box 1738, NL-3000 DR Rotterdam, the Netherlands. neffke@ese.eur.nl

**Department of Social and Economic Geography, Lund University. Sölvegatan 12, 223 62 Lund, Sweden. martin@keg.lu.se

Financial support from Region Skåne, Näringslivsenheten, is greatfully acknowledged.

1. Introduction

Many growing firms are not only successful because they expand the sales of their current products and services, but also because they are able to spot opportunities to enter new markets. However, not all markets are accessible to all firms. In fact, it is often argued that firms diversify into industries that are related to the industry to which their core activities belong. The degree to which firms can draw on their existing strengths to venture into new markets should therefore strongly affect their diversification decision. In this article we study how strongly the relatedness among industries influences diversification behavior of firms. In order to do so, we need to measure to what degree different industries are related to each other. The main part of this paper will, therefore, be dedicated to constructing a measure of what we will call *skill-relatedness* between industries. After deriving such a measure of the degree to which industries have similar skill bases, we use it to investigate whether relatedness, as is commonly suggested, indeed plays a role in corporate diversification processes

As a working definition, industries can be said to be related to the extent that they draw on the same type of resources. In this definition, we refer to “resources” in a very broad sense, ranging from tangible assets (e.g., machinery and raw materials) to intangible assets (know-how, strong brands, etc.). In principle, any of these resources can be a source of relatedness if it can be used in more than one industry.

If we think about all the different resources of a firm, an important – if not *the* most important – resource of a firm is its employees and their skills. Ultimately, people are the carriers of knowledge and a firm’s know-how, and people are endowed with the skills needed to transform other resources into valuable products. As a consequence, we focus on people and the alternative uses of their skills as the resource to determine relatedness among industries. In particular, we study between which industries skilled employees change jobs . We argue that if a skilled individual finds alternative employment in another industry, the production processes in its old and its new industries draw on similar skills, and are in this sense related. This way, we arrive at a measurement of relatedness that is derived from the *revealed ability of skilled employees to move* between industries. We will refer to this as the Revealed Skill Relatedness (RSR) index. For the empirical investigations we use employer-linked data on all individuals in Sweden between 2004 and 2007. Unlike most existing relatedness indices, the index covers the industries of both manufacturing and service sectors. Furthermore, it is possible to quantify the precision of the RSR estimates which allows for the construction of confidence intervals.

The focus on resource sharing among industries differentiates our contribution conceptually from the recently unfolding literature on relatedness indicators based on co-occurrences of industries in corporate portfolios (e.g. Teece et al. 1994; Bryce and Winter 2008; Neffke and Svensson Henning 2009). The main difference is that while these studies *assume* that corporate portfolios are coherent and subsequently derive the implicit relatedness among industries from this assumption, we do not base our index on such an assumption and can, therefore, actually *test* whether corporate portfolios are coherent.

Our analyses show that using labor flows as a basis for a skill-relatedness index reveals linkages between industries that are intuitively plausible. Yet, these linkages are frequently not captured by the standard industrial classification hierarchy. For example, we find, as one might expect, strong links between some manufacturing industries and their wholesale and repair activities in the service sector. However, these industries belong to entirely different sectors according to the classification system. Moreover, contrary to the logic of conventional industry classification systems, our index suggests that industries cannot simply be categorized using a nested hierarchy according to which industries are part of broader industrial sectors. On the contrary, we show that there exists an intricate web of inter-industry skill-connections. When turning to firm diversification patterns, we see a similar picture emerging. Many firms diversify into industries that do not even share the same 1-digit code with the firms' core industries. However, most diversification efforts involve entering an industry that is highly skill-related to the firm's core industry. This shows that the potential to employ the skills of employees in alternative activities exerts a strong influence on the diversification patterns of firms.

Measuring relatedness thus proofs helpful in explaining patterns of corporate diversification. In fact, the relatedness concept has been pivotal for a long time already in research on firm strategy, diversification and portfolio theory (Gort 1962, Rumelt 1982, Porter 1987, Montgomery and Hariharan 1991). But our framework may also be relevant for researchers working outside the field of corporate diversification. Relatedness is, we think, key to the understanding of processes in the economy on many levels and time scales. In the literature on structural change and technological paradigms, technological development is often regarded as a branching process over time. Innovations build on existing technologies in a cumulative, path dependent fashion (Dahmén 1988, Dosi 1988, Nelson 1995). Recent research in evolutionary economic geography furthermore suggests that externalities arising from the presence of related industries are more important than the traditional pure localization and diversity externalities (Frenken et al. 2007 , Neffke et al. 2008).

In section 2 of this article, we review the relatedness literature and provide theoretical justification for the design of our skill-based measure. In section 3, we proceed to describe the structure of the Swedish labor market dataset. Section 4 and 5 lay out the details of our relatedness estimation method and present the resulting relatedness structure. In section 6 we investigate the diversification patterns of firms. As the RSR measure has implications that go much beyond labor market studies and firm diversification, we devote some attention to a broader research agenda that emphasizes the role of relatedness in topics on innovation and structural change after the main conclusions discussed in section 7.

2. Relatedness, labor flows and firm diversification

Relatedness

In the literature, broadly three approaches to the measurement of industry-relatedness can be found. Firstly, scholars have relied on the standard industry classification system and the extent to which industries belong to the same broad industrial sectors herein to determine how related they are. This approach is very straight-forward. For example, industries that belong to the same 3-digit class of industries are more related than industries that only share the same 1-digit class. Although attractive in its simplicity, such reasoning lacks a solid theoretical foundation and is rather ad hoc. The second, and more appealing approach, is to investigate whether commonalities exist in inputs and know-how that are used in different industries. For example, scholars have measured relatedness using information on knowledge flows (Scherer 1982, Pavitt 1984), by comparing patent portfolios (Jaffe 1989, Engelsman and van Raan 1991, Breschi et al. 2003), or by analyzing occupational profiles in different industries (Farjoun 1994).

The third strategy is to consider industrial portfolios as an expression of economies of scope that arise from the co-utilization of productive resources. Authors in this tradition take the fact that certain industries co-occur more often than expected in a specific portfolio as an indication of relatedness. For instance, scholars have used production portfolios of firms or plants (Teece et al. 1994, Bryce and Winter 2006, Neffke and Svensson Henning 2008) or trade profiles of countries (Hidalgo et al 2007). An advantage of these indicators is that they typically have a very wide coverage; the four studies above provide relatedness estimates for all manufacturing industries. Moreover, as they are based on micro-level portfolio choices, they can be thought of as harnessing the collective wisdom of individual decision making agents. However, in essence, these methods are outcome-based, i.e., they *assume* that portfolios are coherent. From this, they then derive the implied relatedness that must have given rise to the portfolios observed in reality.[1] Another disadvantage of co-occurrence based relatedness indices is that portfolios may reflect economies of scope arising from the co-utilization of a number of different resources. This makes it hard to point out what *kind* of relatedness is actually being quantified.

In fact, what relatedness exactly means is far from trivial. Research in business studies differentiates for example between different types of relatedness, such as technological relatedness, managerial relatedness, and relatedness at the level of consumer markets. All three types of relatedness will influence diversification decisions (see Prahalad and Bettis 1986; Grant 1988). Moreover, Prahalad and Bettis (1986) and Grant (1988) distinguish between relatedness at an operational and at a corporate level. At the corporate level, strategic relatedness matters most. However, strategic relatedness is different from, and sometimes conflicts with, the technological relatedness that matters at an operational level. Porter (1987), in turn, gives relatedness a firmer interpretation that builds on the concept of *skills*. He argues that the prime value of corporate relatedness lies in *sharing skills* among the different value chains in a diversified firm. However, such skill-sharing will only translate into corporate advantage if the value chains of a firm's different businesses are similar enough for skill transfer to be meaningful, if the skill transferred is relevant enough for competitive advantage, and if the skill-transfer is to the mutual benefit of the different parts of the corporation. Porter's emphasis on skill-sharing suggests that an important aspect of industrial relatedness is the employability of certain operational skills in different industries. This position finds support in the resource- and knowledge-based theories of the firm. The resource-based view traditionally stresses that competitiveness of firms relies on a wide

[1] Teece et al (1994) and Bryce and Winter (2006) appeal to the *survivor principle* according to which market forces weed out inefficient corporate diversification strategies. As a consequence, surviving firms will have diversified into fields where they can reap efficiency benefits from economies of scope.

range of unique resources (Peteraf 1993).[2] The knowledge-based view of the firm (Grant 1996) subsequently highlights the role of individually held knowledge in a firm.[3] Accordingly, the knowledge and skills of employees can be regarded as the principal resources of firms.

Skills and cross-industry labor flows

The concept of "skills" is used to denote a broad range of qualitatively different individual capabilities. Ingram and Neumann (2006), for example, find that in detailed descriptions of occupations four major skill factors can be distinguished: intelligence, fine motor skills, coordination and strength. Behind these broad factors, however, hides a multitude of more specific skills that are often used in particular occupations and industries. Some skills are accumulated through formal education. Many other skills, in contrast, are acquired during the working life of individuals, for example by learning-by-doing and on-the-job training processes. Some of these skills are fairly generic and can be applied in almost any job. Other skills, however, are highly specific to the task an individual has to perform. These may be specific to a firm, to an occupation, or to an industry.

The study of such human capital specificities belongs traditionally to the field of labor economics. For example, Poletaev and Robinson (2008) show that labor movements that are forced by the closure of an establishment typically lead to large wage losses for the affected individuals. The authors ascribe this to the fact that part of the accumulated human capital of an individual has no value in his or her new employment, and is therefore destroyed by the change of jobs. As a consequence, we can assume that individuals have strong incentives to prevent such human capital destruction. Intriguingly in this context, Poletaev and Robinson find that the degree to which human capital destruction takes place depends on *how* different the activities an individual has to perform exactly are. In particular, they find that there is an additional destruction of human capital when employees not only change employer, but also the industry they work in.[4] This suggests that skills are indeed to an appreciable extent industry-specific.

Having said this, industry-specificity of skills needs not be absolute. It is more likely that certain specialized skills are valuable also in a subset of *related* industries. Individuals that look for a new job outside their own industries will pay careful attention to the question in which other industries their skills are valued.[5] Indeed, to preserve the value of past investments in their skills and knowledge, they will be reluctant to move into an industry if this leads to a substantial destruction of their human capital. The stronger the relatedness between industries, the less human capital destruction will take place

[2] In the literature, the concepts of "resources" and "capabilities" are often used synonymously (Nilsson 2008). For simplicity, we will here subscribe to the distinction in Amit and Schoemaker (1993) between resources as assets controlled by firms and capabilities as the ability to use assets.

[3] In this setting, the firm plays the role of a coordinative mechanism. Consequently, not organizational learning, but coordination and integration of individual knowledge, is the prime challenge for a firm.

[4] However, individuals that also move into a completely different occupation incur such a heavy loss in human capital that it eclipses any additional losses that might have been caused by moving to a new industry.

[5] Whether future employers or individuals assess the value of skills in a new industry is of secondary importance, as this value should be reflected in the wage offer made by the future employer. In practice, both employers and employees have strong incentives to make sure the human capital of candidates matches the requirements of the new jobs.

when moving between them. Therefore, the persistence with which we observe flows of skilled employees indicates the degree of compatibility of skills between the industries involved. Such skilled labor flows across industries, should thus provide information on a particular kind of relatedness: the extent to which these industries are *skill-related*.

In this article, we exploit the information on skill-relatedness in cross-industry labor flows by assessing the degree to which labor flows between pairs of industries are in excess of expectations. The only other paper we know of that focuses on labor flows to measure relatedness among industries is by Maliranta and Nikulainen (2008). That working paper, however, takes a far higher level of aggregation (38 broad sectors compared to our 435 4-digit industries) and uses a rather different in methodology.[6]

A difficulty that arises when looking at labor flows is that some jobs demand skills that are very generic and of use in a wide range of industries. Clear examples can be found in many low-skill jobs. Similarly, organizational tasks involved in general management often involve skills of a generic kind. Managers are frequently confronted with strategic problems and, in such situations, they often draw on some dominant logic or scheme. Although such schemes constitute important skills acquired through management experience, they may in fact have little to do with the precise nature of the production process in the firm (Prahalad and Bettis 1986). Compared to the flows of highly-skilled non-management staff the flows of low-skill labor and people in management positions provide, therefore, less unambiguous signals about the skill-relatedness among industries. These are thus likely to contaminate our RSR measure. Although we do not want to argue that low-skill labor and managers do not have any industry specific skills, we think it is prudent to leave out those groups, and focus our calculations of skill-relatedness on the labor flows of the remaining individuals.

Firm diversification and skill-relatedness

By diversifying, firms can achieve economies of scope if resources that remain underutilized in the production of one product can be put to use in the production process of another product. As a consequence, products that require similar resources are often cheaper to produce in combination than in isolation. It is therefore not surprising that research on diversification strategies of firms has provided ample evidence that firms often engage in *related* diversification (Lemelin 1982; Montgomery and Hariharan 1991), and that such a strategy is more likely to be successful than unrelated diversification (Porter 1987).

The notion that firms typically diversify because some of their resources are left idle can already been found in Penrose (1959). Penrose distinguishes explicitly between resources and the services that a firm can derive from them. The principal difference is that one single resource may provide a variety of services. In fact, at any given moment, the resources of a firm could be dedicated to a number of different activities (Penrose 1959; Teece 1982; Montgomery and Hariharan 1991). If a resource is left idle, a firm may put its services to use by diversifying into one of these activities. Therefore, the various

[6] Most importantly, as we will discuss below, we control not only for the industries' sizes, but also for the wage levels to determine whether an observed labor flow is excessive. Moreover, we also try to focus on those individuals that are most likely carriers of strong industry specific skills.

alternative uses of a resource indicate possible directions for diversification. As we identified a firm's employees and their skills as one of its prime resources, the different services employees can generate (the multi-applicability of the employees' skills) are an important factor in determining the diversification strategy of the firm. As a consequence, we expect that many corporate diversification efforts take place in industries that are closely related to the industry that constitutes the core of a firm's activities. After establishing the skill-relatedness between industries, this conjecture is tested in section 6. We, however, first turn to the design and the data we use to implement our RSR measure.

3. Cross-industry labor flows in Sweden

For our calculations, we need detailed information on wage income, occupation, and industry affiliation of all individuals on the labor market. The datasets we use were provided by Statistics Sweden and contain information on all roughly 9 million individuals registered in Sweden at the end of each year in the period 2004-2007. During this period, there were (by our definition) over 4.5 million active individuals on the labor market.[7]

The information on industry affiliation is available in the sense that most active individuals are linked to a plant, which then is classified into one of the over 750 5-digit industries of the SNI2002 industry classification system. However, many of these 5-digit industries are very small, sometimes consisting of fewer than 50 employees. We therefore aggregate the data into over 500 4-digit industries. Another benefit is that at this level of aggregation, the classification matches the international NACE (Rev 1.1) classification. Furthermore, we leave out all 4-digit industries that employ fewer than 250 persons, because these are too small to generate any significant labor market flows. In the end, there are 435 different industries left for which we analyze skill-relatedness.

Labor flows are made up of the sum total of individual labor market moves (job switches). We register a change in employment as a labor market move when an employee that is active in two consecutive years, changes employment from one year to the next to another plant in another firm.[8,9] The total flow of labor is substantial. On average, almost 600,000 individuals change jobs in each year, representing about 12.5% of the active labor force (see table 1).

[7] We say that a person is active on the labor market if he or she: (a) has a non-zero wage income; (b) works in a plant with registered industry code at the end of the year.

[8] By requiring that an employee must change both establishment and firm, we avoid the problem that a reclassification of plants would result in a large number of individuals switching industries at the same time.

[9] Statistics Sweden identifies firms and plants by the so-called DEE system (Andersson and Arvidson 2006). Accordingly, all plants and firms are compared to the plants and firms that existed a year earlier. In essence, this comparison checks whether the majority of the workforce of a previously not observed legal entity comes from an existing plant or firm, to decide whether this new entity should be treated as an existing plant or firm and should be recoded correspondingly. For plants, similar rules exist. As a consequence, our database is barely contaminated by changes in the legal status of a plant or firm. This also explains why plant and firm creation and destruction are lower than typically reported in the literature.

- table 1 about here -

As our goal is to estimate skill-relatedness between industries, we are particularly interested in those individuals that switch from one industry to another. At about 9.9% of the total active labor force, also such cross-industry labor flows are considerable. However, as argued in section 2, the most accurate picture of skill-relatedness among industries arises when isolating individuals that are likely to be carriers of specialized industry-specific skills. In general, highly skilled individuals should be highly appreciated - and thus highly paid - by their employers. The best paid employees, however, are often managers. As discussed before, we suspect that managers do not possess so much industry specific skills, but rather generic corporate management skills. To isolate individuals that are both highly skilled and possess skills that are not likely to be very general, we focus only on those individuals that earn more than their industry's median wage. We call the individuals that earn below our wage cutoff "low-wage earners", although strictly speaking, low wages are defined relative to the wage level in an industry. We also exclude all individuals in management positions.[10] These procedures reduce the overall useful labor flows to around 100,000 a year.[11]

Table 2 gives a summary of these labor flows. As a first observation, patterns seem to be very similar across years. This strongly suggests that labor flows are not random, but structural and determined by general mechanisms in the economy. Moreover, we can also see that the individuals we selected behave differently compared to the low-wage earners and managers. To be precise, we argued that our selection should consist mainly of highly skilled individuals with a high industry-specific human capital. Consequently, these individuals should face stronger human capital destruction when moving to industries that use a completely different set of skills. In table 2, we indeed find evidence in support of this claim. Both low-wage earners and people in management positions are more mobile than the average individual. In fact, both groups are about three times as likely to change jobs compared to our selection of individuals. This is in line with our claim that we selected highly skilled individuals, because the higher the skill-level, the more human capital exists that can be destroyed. Managers and low-wage earners also more frequently move into industries that, according to the industrial classification hierarchy, are in a radically different part of the economy. For example, managers and low-wage earners are about twice as likely to cross the boundaries between 4-digit industries when changing jobs, than individuals in our selection. In fact, we find this difference at all levels of aggregation. At the extreme, only about 40% of all managers and low-wage earners that switch jobs remain in the same 1-digit

[10] We first exclude all individuals in management occupations according to the Swedish occupation register SSYK and only then establish the median wage.

[11] Our strategy is likely to discard many individuals that possess strongly industry-specific skills such as technical managers and low paid craftsmen. However, ideally our selection of individuals does not bias relatedness measures in favor of particular industries. For example, adding low-paid craftsmen, but leaving out low paid clerks increases the observed labor flows, and thus the estimated skill-relatedness, between industries in the manufacturing sector compared to certain industries in the financial sector and would thus be undesirable. Moreover, we would also be forced to make a judgment about which occupations we think are associated with highly skill-specific human capital introducing a certain level of subjectivity we would rather avoid.

industry, whereas the vast majority moves to a job in a completely different sector of the economy. By contrast, for the individuals we selected these numbers are reversed.

All evidence taken together indicates that we have successfully identified individuals possessing strong, non-generic skills. But although the yearly cross-industry labor flows in the group of selected individuals are substantial, they are rather modest compared to the in total 435×434=188,790 possible industry combinations. To increase the precision of our estimates, we therefore pool the labor flows of all available years.

- table 2 about here -

4. Estimating skill-relatedness

Revealed Skill Relatedness

To study labor flows within pairs of industries, Table 3 shows average yearly employment and labor flows for the top five industry combinations in terms of size of flows. All combinations in this table involve industries that indeed appear to be related, both intuitively and according to the industrial classification system. However, another striking aspect of this list is that it is made up exclusively of very large industries. This stresses the point that raw industry flows are only of limited use as indicators of skill-relatedness. In fact, large labor flows may occur because of a wide range of different industry characteristics that have little or nothing to do with skill-specificity. When constructing our skill-relatedness index, we must control for such effects.

- table 3 about here -

The problem of correcting raw labor flows is in essence the same as correcting co-occurrence levels in the co-occurrence approach to measuring relatedness. Almost all authors in this tradition control only for the size of industries. For example, Teece et al. (1994) and Bryce and Winter (2006) count the number of times that plants in two different industries are part of the portfolio of one firm, and then compare this number to how often co-occurrences would be expected to take place in a random hypergeometric matching process. Hidalgo et al. (2007) use the number of times countries have a comparative advantage in a specific combination of export products as the starting point for the relatedness among product classes. They then divide this number by the minimum of the number of countries that has a comparative advantage in one of the two product classes involved. Marilanta and

Nikulainen (2008) analyze the size of the labor flows among 38 2-digit industries, relative to the sizes of the industry of origin and of the destination industry.

In contrast to the approaches above, we use an approach that is similar to a method we developed in a related paper (Neffke and Svensson Henning 2008). The main insight is that regression analysis can be used to predict the size of flows between two industries. As predictors, any number of industry specific variables can be used. By next calculating to what extent observed labor flows are in excess of predicted labor flows, we can quantify the relatedness between two industries as the *relative excess labor flow* between the industries. More formally, let F_{ij}^{obs} be the flow of individuals that move from industry *i* to industry *j* and $\hat{F}_{ij}$ be some prediction of this flow based on a number of industry-level variables. Our skill-relatedness index is then simply the ratio of these quantities:

(1) $$RSR_{ij} = \frac{F_{ij}^{obs}}{\hat{F}_{ij}}$$

What remains is choosing an appropriate estimation procedure for $\hat{F}_{ij}$. As labor flows are a count of individuals, they can never be negative and are always integer valued. The most appropriate estimation model can, therefore, be found in the family of count data regression models. Since among the vast majority of industries there are no labor flows whatsoever, a standard Poisson process is inadequate. Instead, we use a zero-inflated negative binomial regression (zinb) procedure. The zinb regression equation consists of two parts: a regime selection equation and a count data part. The regime selection equation determines *whether* there will be any flow at all, whereas the count data part models *how many* individuals will move from one industry to the other, given that the regime selection equation evaluates to a state in which a flow is possible. The general zinb regression equation is as follows:

(2) $$E\left(F_{ij} \middle| v_i, w_j, \varepsilon_{ij}\right) = \left[1 - \pi_0\left(\gamma + v_i'\delta_i + w_j'\delta_j\right)\right] e^{\alpha + v_i'\beta_i + w_j'\beta_j + \varepsilon_{ij}}$$

The vectors v_i and w_j contain industry-level variables. We concentrate on two variables: the size of industries and the average wage levels paid in industries. We include industries' average wages because higher potential wages are an important incentive to switch jobs. Consequently, individuals are likely to move out of low-wage industries and into high-wage industries irrespectively of skill-relatedness. As a result, low-wage industries are likely to be origins of substantial labor flows, whereas high-wage industries are likely to be their destinations.[12] As mentioned before, we pool data of all years by

[12] Other industry characteristics may have an impact on labor flows as well. For example, average age and average educational attainment of the labor force may correlate with the size of labor flows. However, these variables do not so much represent incentives for labor market moves, as *characterize the individuals* working in an industry.

summing both dependent and independent variables over all available years in our dataset to improve the efficiency of the estimates.[13] Our prime goal is not to explain the causal linkages that connect our regressors to the labor flows, but rather to arrive at the best possible prediction of labor flows given the information in the regressors. We therefore use the following functional form, which fitted our data best:

$$(3) \qquad E\left(F_{ij} \middle| v_i, w_j, \varepsilon_{ij}\right) = \left[1 - \pi_0\left(\gamma + \delta_i emp_i + \delta_j emp_j\right)\right] \cdot e^{\alpha + \beta_{1i}\log(emp_i) + \beta_{2i} log(wage_i) + \beta_{1j}\log(emp_j) + \beta_{2j} log(wage_j) + \varepsilon_{ij}}$$

where:

emp_i: sum of employment in industry of origin *i* across 2004, 2005 and 2006

emp_j: sum of employment in destination industry *i* across 2005, 2006 and 2007

$wage_i$: average wage in industry of origin *i* across 2004, 2005 and 2006

$wage_j$: average wage in destination industry of origin *j* across 2005, 2006 and 2007

The outcomes of this analysis are summarized in Table 4.

- table 4 about here -

The employment in both the origin and in the destination industry has the expected positive effect on the size of labor flows. The effect is somewhat smaller for the destination industry's employment. Higher wages lead to both higher inflows and higher outflows of labor, which suggests that labor mobility is particularly strong among industries with high wages. However, the increase in mobility *towards* an industry due to high wages is more than double as strong compared to the mobility *out of* an industry. This indicates that financial incentives play a particularly important role in the choice of the destination industry. As our purpose is not to *explain* the size of labor flows, but rather to *predict* them, it is, however, sufficient to note that the parameter estimates are not implausible. Using the parameter estimates in equation (3), we can calculate the expected labor flows between industries based on

They may therefore correlate strongly with skill-intensity and skill-specificity of industries. If we falsely correct for such individual characteristics, we introduce errors in the relatedness index.

[13] As individuals can change jobs independently of earlier labor moves, this pooling is unlikely to cause great problems. The increased precision of the relatedness estimations is, by contrast, substantial.

industry level employment and wage information. Equation (1) then gives us an index of relatedness for some 185,000 industry combinations.

Determining significance levels of Revealed Skill Relatedness estimates

As among the vast majority of industry combinations there are no individuals moving at all, the numerator of (1), the observed labor flow from industry *i* to industry *j*, is often equal to zero. Consequently, we find an overwhelming number of industry combinations with a skill-relatedness of zero. However, many of these industry combinations are also *predicted* to exhibit a near zero labor flow. Therefore, it would be wrong to infer immediately that industries are unrelated if the relatedness index is zero. In many cases, the involved industries are too small to generate substantial labor flows. In fact, for those industry combinations that are predicted to have a labor flow considerably below one, any person that moves between the involved industries has a huge impact on the index. Low expected labor flows, therefore, often result in a relatedness of zero, but sometimes they give rise to very high relatedness measures.

In fact, for many such industry combinations information is too limited to warrant strong claims about skill-relatedness. To deal with this problem we must quantify the level of confidence that can be put in skill-relatedness estimates. This requires that we frame the problem in a slightly different way and think of labor flows from industry *i* to *j* as generated by independent choices of individuals to change jobs and move from industry *i* to industry *j*. To simplify matters, we will abstract from the fact that each individual can only make one labor move a year. As a result, each individual faces 435 independent choices. One is to stay in its current industry. The other 434 choices correspond to a move into each of the remaining 434 industries. Each of these choices can be modeled as a Bernoulli experiment with probability of success equal to p_{ij}. Consequently, the labor flow from *i* to *j* can be thought of as the outcome of a binomial experiment, with *n* equal to the employment in industry *i*, and p equal to p_{ij}:

(4) $$F_{ij} \sim BIN\left(emp_i, p_{ij}\right)$$

By dividing numerator and denominator of equation (1) by emp_i, the following alternative expression for the skill-relatedness from *i* to *j* is obtained:

(5) $$RSR_{ij} = \frac{p_{ij}^{obs}}{\hat{p}_{ij}}$$

where p_{ij}^{obs} is the observed relative frequency (say, the observed probability) that an individual moves from industry *i* to industry *j*, $p_{ij}^{obs} = \frac{F_{ij}^{obs}}{emp_i}$, and $\hat{p}_{ij}$ is its expected counterpart.

The original question of how informative a labor flow of a specific size is can now be rephrased as a question about how unlikely it is to observe p_{ij}^{obs}. If we take $\hat{p}_{ij}$ as a benchmark, this question translates into a statistical test whether p_{ij}^{obs} is significantly different from $\hat{p}_{ij}$. More formally, we test the null-hypothesis: $p_{ij} = \hat{p}_{ij}$ against the alternative hypothesis: $p_{ij} > \hat{p}_{ij}$, using p_{ij}^{obs} as observation. The p-value of this test can be calculated as follows:

(6) $$P\left(x > p_{ij}^{obs} \middle| p_{ij} = \hat{p}_{ij}\right) = 1 - \sum_{r=0}^{p_{ij}^{obs}} \left[\hat{p}_{ij}{}^{r} \cdot \left(1 - \hat{p}_{ij}\right)^{emp_i - r} \binom{emp_i}{r} \right]$$

With a 10% significance level, we find that for 9,919 industry combinations the observed number of links is significantly different from the expected number of links. In other words, there is significant evidence of skill-relatedness in about 5.4% of all possible cases. If the alternative hypothesis is $H_a: p_{ij} < \hat{p}_{ij}$, we find significant "unrelatedness" or dissimilarity between industries in 5.4% of all cases as well. For 89.3% of industry combinations, however, our empirical evidence is insufficient. This might strike one as a discouraging result. However, most of the industries involved in combinations with insignificant RSR values are very small. In fact, if we weight the counts of industry combinations by the employment in origin and destination industry, significant estimates account for 69.8% of all industry combinations. Table 5 summarizes these findings.

- table 5 about here -

A final statistical problem is that the distribution of relatedness indices is very skewed. Some 80% of all values are equal to zero, and in the tail of the distribution, every one in 1000 industry combinations has a skill-relatedness of over 100. However, if we take logs, the distribution closely resembles the bell-shape of a normal distribution.[14] To allow for a comparison across industries, we normalize the log-transformed relatedness measure by subtracting for each industry of origin the mean across all destination industries and dividing by the standard deviation. This results in the following z-value:

(7) $$z\left[log\left(RSR_{ij}\right)\right] = \frac{log\left(RSR_{ij}\right) - \mu\left[log\left(RSR_i\right)\right]}{\sigma\left[log\left(RSR_i\right)\right]}$$

where:

[14] We drop the 80% of cases in which the skill-relatedness is equal to zero.

$\mu[log(RSR_i)]$: mean of $log(RSR_{ij})$ across all industries of origin *i*

$\sigma[log(RSR_i)]$: standard deviation of $log(RSR_{ij})$ across all industries of origin *i*

This facilitates the interpretation of our findings. For example, if an industry combination has a z-value of 1.96, using the normal distribution as the basis for some rules of thumb, this indicates that only 5% of the other destination industries for which a non-zero relatedness measure was available are more related to the industry of origin. Ignoring significance levels, this means that such an industry combination is located in the top 1% (5% out of 20%) of related industries. After all, the 80% of industry combinations that had a skill-relatedness of zero are labeled as completely unrelated. Therefore, the fact that the relatedness index is above zero means that the industry combination already belongs to the top 20% of all possible combinations in terms of relatedness.

5. Industries in industry space

In Figure 1, we have plotted a network of industries based on the skill-relatedness linkages calculated by the procedures above.[15] We call this network, in analogy to the product space of Hidalgo et al. (2007), "industry space". The nodes represent industries, and the connections between them represent skill-relatedness indices. We depicted only those skill-relatedness links for which (1) the z-value is in excess of one, and (2) labor flows are significantly different from expected flows at the 10% level. The position of the nodes is determined by a spring-embedded algorithm, which scatters industries across the entire plane, but in such a way that more closely related industries are located closer together. Therefore, industries that are close to each other in industry space are usually strongly related in terms of skill-relatedness. The colors of the nodes represent the broad sectors to which the industries belong according to the industrial classification system. The nodes in the left most part of the graph are industries without linkages to any of the other industries.

Visually, nodes of the same color form clusters in industry space. On the whole, two major concentrations of connected industries can be observed, one in the north-east dominated by manufacturing industries, and one in the south-west dominated by a wide variety of service industries. Upon closer inspection, however, many exceptions exist, complicating this general pattern. For example, many business services are located on the fringes of the manufacturing industries' concentration, suggesting that these industries maintain skill-linkages to industries in the manufacturing sector, without being firmly embedded among them. Similarly, some sales and wholesales service industries are well connected to their manufacturing counterparts. Electricity and construction industries, on the other hand, form a distinct cluster in the right most part of the manufacturing concentration. Hotels, restaurants and transport industries exhibit a similar pattern in the services dominated concentration. Interestingly, many public sector activities, such as healthcare and education, are found in peripheral

[15] The picture was generated using the *NetDraw* (Borgatti 2002) software package.

parts of the network. This suggests that skills used in these sectors are not very compatible with other industries.[16]

Industry space is a simplified projection of industry linkages onto a two-dimensional surface. Still, it shows that, although some clustering of industries around 2-digit sectors is visible, many of the skill-linkages among industries are not captured in the NACE-system. This becomes even clearer when concentrating on the networks of individual industries. In Figure 2, we display the so-called "ego-network" for pharmaceutical preparations. For simplicity, we will refer to this industry as "pharma". An ego-network shows only the linkages among industries that are directly connected to the selected industry. As expected, most related industries are part of the chemicals sector. Interestingly however, the network of pharma stretches over several broad sectors. For instance, medical and surgical equipment and orthopaedic appliances production is related to the pharma industry, although this is formally part of the machinery sector which consists of a broad mix of different manufacturing activities. Similarly the skill-relatedness network of pharma also reaches out to several service industries. Direct linkages exist to business service industries, like software consultancy and supply and to the specialized parts of the sales and wholesales sector. Intuitively this is exactly what we would expect. However, as the relatedness network spans across the entire economy, the RSR-index is capable of providing such detailed quantification of skill-linkages also for industry combinations for which we typically have less strong intuitions.

- figure 1 about here -

- figure 2 about here -

6. Relatedness and firm diversification

In section 2, we argued that diversification patterns of firms should exhibit strong skill-relatedness. To investigate this claim, we aggregate individuals into establishments and establishments into firms.[17,18] Table 6 displays some general information about firms and their establishments.

[16] The position of industries in industry space, however, does not fully express the intricate pattern of linkages between industries. These can only be assessed by studying the full matrix of RSR indices.

[17] We use what are called DEE IDs, see footnote 99.

[18] As a result, all establishments have at least one employee.

- table 6 about here -

The turnover of firms during the period we study is fairly large. However, under 1% of firm exits involves firms that own more than one establishment. Since all plants are given an industry code according to their main activity, we can define the industrial portfolio of firms as the industries in which the firm owns at least one plant. If we define firm expansions and contractions as firms opening or shutting down establishments, we find that these often simply lead to up-sizing or down-sizing of existing activities of a firm, but rarely involve entry into or full retreat from an industry.

We are, however, mainly interested in firms that enter new markets by setting up an establishment in an industry in which they previously had no establishments. Our main claim is that firms diversify not just into any industries, but into industries that are skill-related to their core activities. We define the core activity of a firm as the industry in which it employs the largest part of its labor force, assuming that this is also the activity around which the firm has built its resource base. Each of the in total 939 diversification moves can accordingly be regarded as a move *from* the core industry *into* a new industry.

Surprisingly, as shown in Table 7, most diversifications cross the boundaries of 1-digit industries in the conventional industry classification system. If we were to rely on the industrial classification system, diversification would thus seem to take place predominantly between unrelated industries.

- table 7 about here -

However, if we instead use our skill-relatedness index, we have to substantially revise this statement. Table 8 shows the RSR-index for industry combinations within which diversifications take place. As mentioned before, the vast majority (80%) of industry combinations in the economy have a skill relatedness of zero. However, only 85 of 930 diversification moves (not even 10% of the total) take place between such industries. Moreover, most of these diversifications take place between industries for which information to estimate relatedness was poor. By contrast, in 845 diversification cases, the skill-relatedness is different from zero. 87% of these diversification moves take place between industries with positive z-values of the log-transformed skill-relatedness. A positive z-value indicates that firms move from their core industry into industries that are more skill-related than at least half of the possible diversification industries *to which a labor flow is observed*. If we again ignore significance levels and think of the industries with zero relatedness as unrelated, a z-value greater than zero suggests that the industry combinations are among the top 50% of the 20% non-zero skill-related industries. A z-value greater than zero thus roughly means that the industry combination belongs to the top 10% of most related industries.

- table 8 about here -

To investigate matters further, the histogram in figure 3 shows the relative frequency of z-values for all industry combinations in the economy with a non-zero RSR-index in the light bars. The dark bars show the same relative frequency, but now only for the skill-relatedness of industry combinations that were involved in diversification moves. The histogram of diversification moves is clearly shifted towards higher relatedness indices. In other words, compared to the overall distribution of skill-relatedness, the relatedness within industry combinations in which a diversification takes place is generally higher. Summarizing, both Table 8 and Figure 3 show that firms seem to follow a logic of skill-related diversification. In fact, if we look at the core industry (i) of a diversifying firm and rank all destination industries (j) the firm can choose from according to the value of RSR_{ij}, firms on average diversify into one of the 15% most related industries.

- figure 3 about here -

7. Conclusions and future research

It is often argued that firms can exploit the alternative uses of their resources to venture into new lines of business that are related to their core activity. To investigate this claim, we quantified the extent to which the arguably most important resource of a firm, human capital, can be applied in different industries by investigating cross-industry labor flows of skilled individuals. The resulting Revealed Skill Relatedness index is an objective quantification of relatedness with an economy-wide coverage of industries that provides an important complement to existing relatedness measures. Our RSR measure shares many of the strengths of portfolio or co-occurrence based indices, since it uses information on decisions made at the micro-level in the economy. However, unlike portfolio-based measures, the RSR is not outcome-based but it is derived from similarities in inputs, rather than from manifested co-occurrences of industries in portfolios. The RSR does, therefore, not rely on the assumption that firms maintain coherent portfolios. Rather, the RSR provides a basis to investigate whether or not this is the case.

When we use our RSR-index to study corporate diversification, we find overwhelming evidence that firms mainly diversify into industries that are strongly skill-related to their core activities. This fact would not have been uncovered by simply looking at the industry codes involved in corporate diversification moves. Skill-relatedness among industries thus strongly affects diversification decisions of firms. This not only supports the claims by Penrose (1959) and Teece (1982) that firms expand into new activities that can exploit the alternative uses that are latent in idle resources. It also stresses the role of human capital as a resource that is an especially important factor in firm diversification strategies.

So far, we have shown that skill-relatedness can be measured in a rigorous way and how it may be used to investigate diversification patterns of firms. However, the RSR-index can shed light on a number of other research issues. For example, the coherence of industrial portfolios and the degree to which firms concentrate on core competences might vary from one decade to the next. The RSR-index provides a way to quantify such tendencies. Furthermore, in the field of labor economics, the RSR-index can be used to identify whether specific occupational groupings exhibit different cross-industry skill-relatedness. For example, although individuals in management positions are more mobile than the selection of highly skilled individuals we investigated, this is not to say that they move between completely different industries. Moreover, investigating the structure of labor flows of managers may reveal important information about the specificity of organizational skills.

A property of the RSR-index we did not stress in this article is that it is essentially an asymmetric measure. This means that the RSR from industry *i* to *j* is, in general, different from the RSR from industry *j* to *i*. As a result, the RSR-index may shed light on the direction of knowledge flows between industries which, in turn, may help gain a more thorough understanding of cross-fertilization of ideas between industries and the opportunities for Schumpeterian *new combinations* (Schumpeter 1951 (1911)). In as far as knowledge spillovers are spatially bounded, there are also important implications for the literature on agglomeration externalities (Rosenthal and Strange 2004) and industrial districts and clusters (Porter 2000, Amin 2003).

Finally, industry space itself can become the subject of investigation. In the literature on structural change, techno-economic paradigms (Freeman and Perez 1988), and path dependence in innovation (Dosi 1982), authors have suggested that different fields of production co-evolve in terms of innovation and technological change according to technological relatedness linkages among them. However, it is possible that the relatedness structure itself changes over time and is different in different countries. By constructing different industry spaces for each year and for several countries, such changes can be investigated. For example, if industry space is constantly being rewired, this may have profound implications for the innovation intensity and the frequency of long technological leaps in the economy. Similarly, cross-sectional differences in the industry spaces of countries could may give a new twist to the investigations initialized by Porter in the 1980s (Porter 1990), by linking national competitive advantage to the organization of a country's industry space.

8. References

Amin, A (2003). Industrial Districts. In *A Companion to Economic Geography*, edited by E Sheppard and T J Barnes. Oxford: Blackwell.

Amit, R, and P J H Schoemaker (1993). Strategic Assets and Orgnizational Rent. *Strategic Management Journal* 14(1):33-46.

Andersson, J and G Arvidson (2006). Företagens och arbetsställenas dynamik (FAD). Memo, Statistics Sweden.

Borgatti, S P (2002). *NetDraw. Graph Visualization Software.* Harward: Analytic Technologies.

Breschi, S, F Lissoni and F Malerba (2003). Knowledge-relatedness in firm technological diversification. *Research Policy* 32:69-87.

Bryce, D J and S G Winter (2006). A general inter-industry relatedness index. Center of Economic Studies Discussion Papers, CES 06-31.

Dahmén, E (1988). 'Development blocks' in industrial economics. *The Scandinavian Economic History Review* 36:3-14.

Dosi, G (1982). Technological paradigms and technological trajectories - A suggested interpretation of the determinants and directions of technical change. *Research Policy* 11 (3):147-162.

Dosi, G (1988). Sources, Procedures, and Microeconomic Effects of Innovation. *Journal of Economic Literature* 26(3):1120-1171.

Engelsman, E C and A F J van Raan (1991). Mapping of Technology, A First Exploration of Knowledge Diffusion Amongst Fields of Technology. Policy Studies on Technology and Economy (BTE) Series. The Hague.

Farjoun, M (1994). Beyond Industry Boundaries: Human Expertise, Diversification and Resource-Related Industry Groups. *Organization science* 5 (2):185-199.

Freeman, C and C Perez (1988). Structural crises of adjustment, business cycles and investment behaviour. In *Technical change and economic theory*, edited by G Dosi, C Freeman, R Nelson, G Silverberg and L Soete. London: Pinter Publishers.

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

Gort, M (1962). *Diversification and integration in American industry*. Princeton: Princeton University Press.

Grant, R M (1988). On 'Dominant Logic', Relatedness and the Link between Diversity and Performance. *Strategic Management Journal* 9(6):639-642.

Grant, R M (1996). Toward a Knowledge-Based Theory of the Firm. *Strategic Management Journal* 17:109-122.

Hidalgo, C A, B Klinger, A-L Barabási and R Hausmann (2007). The Product Space Conditions the Development of Nations. *Science* (317):482-487.

Ingram, B F and G R Neumann (2006). The returns to skill. *Labor Economics* 13:35-59.

Jaffe, A (1989). Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. *Research Policy* 18:87-97.

Lemelin, A (1982). Relatedness in the Patterns of Interindustry Diversification. *The Review of Economics and Statistics* 64(4):646-657.

Maliranta, M and T Nikulainen (2008). Labour flow paths as industry linkages: a perspective on cluster and industry life cycles. Discussion papers no. 1168. The Research Institute of the Finnish Economy.

Montgomery, C A and S Hariharan (1991). Diversified expansion by large established firms. *Journal of Economic Behavior and Organization* 15:71-89.

Neffke, F and M Svensson Henning (2008). Revealed Relatedness: Mapping Industry Space. Papers in Evolutionary Economic Geography #08.19. Urban and Regional research centre Utrecht.

Neffke, F, M Svensson Henning and R Boschma (2008). Surviving in agglomerations: Plant evolution and the changing benefits of local environment. Papers in Evolutionary Economic Geography #08.20. Urban and Regional research centre Utrecht.

Nelson, R R (1995). Recent evolutionary theorizing about economic change. *Journal of Economic Literature* XXXIII (March):48-90.

Nilsson, M (2008). A Tale of Two Clusters. Sharing Resources to Compete. Lund Studies in Economics and Management 104. Lund: Lund Business Press.

Pavitt, K (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy* 13:343-373.

Penrose, E T (1959). *The Theory of the Growth of the Firm*. Oxford: Blackwell.

Peteraf, M A (1993). The Cornerstones of Competitive Advantage: A Resource-Based View. *Strategic Management Journal* 14(3):179-191.

Poletaev, M and C Robinson (2008). Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys. *Journal of Labor Economics* 26 (3):387-420.

Porter, M E (1987). From competitive advantage to corporate strategy. *Harvard Business Review* 65(3):43-59.

Porter, M E (2000). Locations, clusters, and company strategy. In *The Oxford Handbook of Economic Geography*, edited by G L Clark, M P Feldman and M S Gertler. Oxford: Oxford Universty Press.

Prahalad, C K and R A Bettis (1986).The Dominant Logic: A New Linkage between Diversity and Performance. *Strategic Management Journal* 7(6):485-501.

Rosenthal, S S and W C Strange (2004). Evidence on the nature and sources of agglomeration economies. In *Handbook of regional and urban economics*, edited by J V Henderson and J F Thisse. Elsevier.

Rumelt, R P (1982). Diversification Strategy and Profitability. *Strategic Management Journal* 3(4):359-369.

Scherer, F M (1982). Inter-industry technology flows in the United States. *Research Policy* 11:227-245

Schumpeter, J A (1951). *The theory of economic development. An inquiry into profits, capital, credit, interest and the business cycle*. Cambridge, Mass: Harvard University Press. Original edition first published 1911.

Teece, D J (1982). Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior and Organization* 3:39-63.

Teece, D J, R Rumelt, G Dosi and S Winter (1994). Understanding corporate coherence. Theory and evidence. *Journal of Economic Behavior and Organization* 23:1-30.

Table 1: General characteristics of the Swedish labor market

	2004	**2005**	**2006**
Population	9,011,372	9,047,730	9,113,225
active labor force	4,697,155	4,724,497	4,806,047
total labor flow	540,317	591,802	627,256

Table 2: Labor flows in Sweden 2004-2006.

	2004				2005				2006			
	all	managers	low wage earners	selection	All	managers	low wage earners	selection	all	managers	low wage earners	selection
General												
active labor force	**4,697,155**	**1,120,014**	**2,592,850**	**1,788,402**	**4,724,497**	**989,015**	**2,597,853**	**1,867,547**	**4,806,047**	**895,995**	**2,623,914**	**1,954,761**
% of population	(52%)	(12%)	(29%)	(20%)	(52%)	(11%)	(29%)	(21%)	(53%)	(10%)	(29%)	(21%)
labor flow	**540,317**	**188,098**	**420,507**	**91,047**	**591,802**	**177,107**	**458,808**	**107,681**	**627,256**	**165,002**	**477,682**	**125,250**
% of active labor force	(12%)	(17%)	(16%)	(5%)	(13%)	(18%)	(18%)	(6%)	(13%)	(18%)	(18%)	(6%)
Labor flow within:												
same 4-digit industry	**110,303**	**30,343**	**69,602**	**31,723**	**115,058**	**26,797**	**72,903**	**34,846**	**123,418**	**25,644**	**76,630**	**39,875**
% of total labor flow	(20%)	(16%)	(17%)	(35%)	(19%)	(15%)	(16%)	(32%)	(20%)	(16%)	(16%)	(32%)
same 3-digit industry	**134,890**	**36,669**	**86,636**	**37,688**	**141,958**	**32,443**	**91,280**	**41,973**	**152,024**	**30,839**	**95,740**	**47,979**
% of total labor flow	(25%)	(19%)	(21%)	(41%)	(24%)	(18%)	(20%)	(39%)	(24%)	(19%)	(20%)	(38%)
same 2-digit industry	**181,624**	**51,058**	**121,568**	**46,560**	**193,772**	**46,099**	**129,888**	**52,577**	**206,651**	**43,974**	**135,861**	**60,012**
% of total labor flow	(34%)	(27%)	(29%)	(51%)	(33%)	(26%)	(28%)	(49%)	(33%)	(27%)	(28%)	(48%)
same 1-digit industry	**241,911**	**72,534**	**168,933**	**56,223**	**260,424**	**66,274**	**181,704**	**64,481**	**279,028**	**63,348**	**190,944**	**74,317**
% of total labor flow	(45%)	(39%)	(40%)	(62%)	(44%)	(37%)	(40%)	(60%)	(44%)	(38%)	(40%)	(59%)
Labor flow between:												
different 1-digit industries	**298,406**	**115,564**	**251,574**	**34,824**	**331,378**	**110,833**	**277,104**	**43,200**	**348,228**	**101,654**	**286,738**	**50,933**
% of total labor flow	(55%)	(61%)	(60%)	(38%)	(56%)	(63%)	(60%)	(40%)	(56%)	(62%)	(60%)	(41%)

Labor flows are measured from the indicated year to the subsequent year. Columns <all>: all individuals in Sweden; <managers>: all individuals in management occupations; <low wage earners>: all individuals earning below the industry's median wage; <selection>: all individuals that are not in management occupations and earn at least the industry's median wage.

Table 3: Top 5 labor flows between 2004 and 2007.

labor flow	industry of origin	# employees	destination industry	# employees
1,971	8531: Social work activities with accommodation	112,178	8532: Social work activities without accommodation	89,668
1,681	8511: Hospital activities	107,819	8512: Medical practice activities	29,759
1,619	8532: Social work activities without accommodation	85,761	8531: Social work activities with accommodation	112,691
1,357	5510: Hotels	13,867	5530: Restaurants	28,556
1,291	8531: Social work activities with accommodation	112,178	8511: Hospital activities	108,600

Labor flow is the cumulative flow of labor from the industry of origin to the destination industry between 2004 and 2007. # employees refers to the average number of employees in a year for the period 2004-2006 (industry of origin), respectively 2005-2007 (destination industry).

Table 4: Zero-inflated negative binomial regression of labor flows.

	parameter	95% confidence interval	
	Estimate	lower limit	upper limit
count data equation			
log(emp_o)	0.83 ***	0.82	0.84
log(emp_d)	0.68 ***	0.66	0.69
log(wage_o)	0.21 ***	0.15	0.27
log(wage_d)	0.42 ***	0.36	0.47
Constant	-21.58 ***	-22.40	-20.77
regime selection equation			
emp_o	-5.34E-07	-1.29E-06	2.19E-07
emp_d	-4.19E-04 ***	-4.57E-04	-3.81E-04
Constant	0.28 ***	1.76E-01	3.79E-01
overdispersion parameter			
log(alpha)	1.15 ***		
Nobs	184,776		
zero observations	147,452		

*: p<.050; **: p<.025; ***:p<.010. emp_o: sum of employment in industry of origin over the years 2004-2006; emp_d: sum of employment in destination industry over the years 2005-2007; wage_o: average yearly wage in industry of origin across 2004-2006; wage_d: average yearly wage in destination industry across 2005-2007. Note that # observations is smaller than the 188,790 different industry combinations because not all industries existed in all years and, therefore, not all industries could figure as both origin and destination industries.

Table 5: Frequency of significant skill-related industry combinations.

		significant	insignificant
level			
$F_{ij}^{obs} = 0$	**dissimilar**	5696	141756
$F_{ij}^{obs} > 0$	**related**	9919	14128
	dissimilar	4234	9043
percentage			
$F_{ij}^{obs} = 0$	**dissimilar**	3.1%	76.7%
$F_{ij}^{obs} > 0$	**related**	5.4%	7.6%
	dissimilar	2.3%	4.9%
employment weighted			
$F_{ij}^{obs} = 0$	**dissimilar**	11.4%	10.9%
$F_{ij}^{obs} > 0$	**Related**	18.6%	7.1%
	dissimilar	39.8%	12.3%

Significance level: 10%. Weights: $emp_i \cdot emp_j$

Table 6: Firms and establishments in Sweden.

	2004	2005	2006	2007
# establishments	495,596	498,016	507,003	514,614
# firms	424,709	427,923	436,570	443,541
# multi-establishment firms	7,903	7,798	7,933	8,000
% of firms	*1.9%*	*1.8%*	*1.8%*	*1.8%*
# firm entries	72,646	77,421	80,275	
% of firms	*17.1%*	*18.1%*	*18.4%*	
# firm exits	69,404	68,603	73,014	
% of firms	*16.3%*	*16.0%*	*16.7%*	
# expansions	4,509	4,326	4,294	
# diversifications	298	259	382	
% of expansions	*6.6%*	*6.0%*	*8.9%*	
# contractions	5189	4237	4095	
# consolidations	100	93	214	
% of contractions	*1.9%*	*2.2%*	*5.2%*	

Expansions, contractions, diversifications and consolidations exclude establishments that merely change ownership. Diversifications refer to establishments in industries in which the firm previously did not own any establishments; consolidations refer to a firm abandoning an industry by closure of an establishment.

Table 7: Diversification moves by distance in industrial classification system.

	# diversifications	% of diversifications
within same 3-digit industry	54	5.8%
within same 2-digit industry	140	14.9%
within same 1-digit industry	100	10.6%
between different 1-digit industries	645	68.7%

Table 8: z-values of log skill-relatedness for diversification moves.

		significant	insignificant
$RSR_{ij} = 0$		28	57
$RSR_{ij} > 0$	z<-2	0	0
	z<-1	12	4
	z<0	61	47
	z>0	636	101
	z>1	429	59
	z>2	182	0
Total		**725**	**205**

z represents the z-value as calculated in equation (7). Note: 9 diversification moves took place among industries for which we did not calculate skill-relatedness.

Legend to figure 1

	Agriculture, forestry, fishing and mining
	Food
	Textiles
	Wood and paper
	Chemicals, coal and rubber
	Mineral products
	Metal products
	Machinery and communication eqt.
	Transport eqt.
	Manufacturing nec.
	Electricity and construction
	Sale, trade and wholesale
	Hotels, restaurants and transport
	Finance and banking
	Business services, R&D
	Education and public administration
	Healthcare and other services

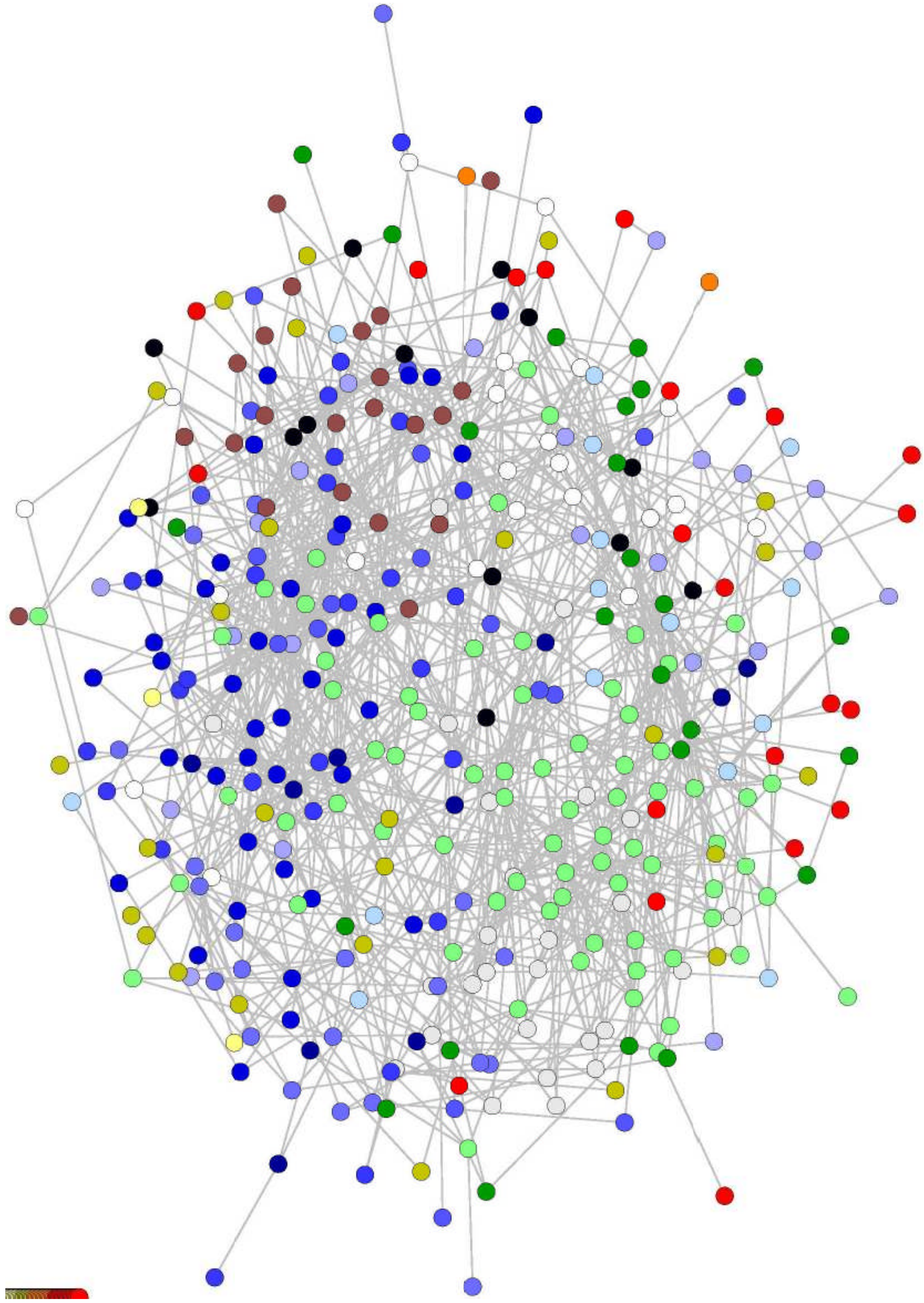

Figure 1: industry space based on skill relatedness.

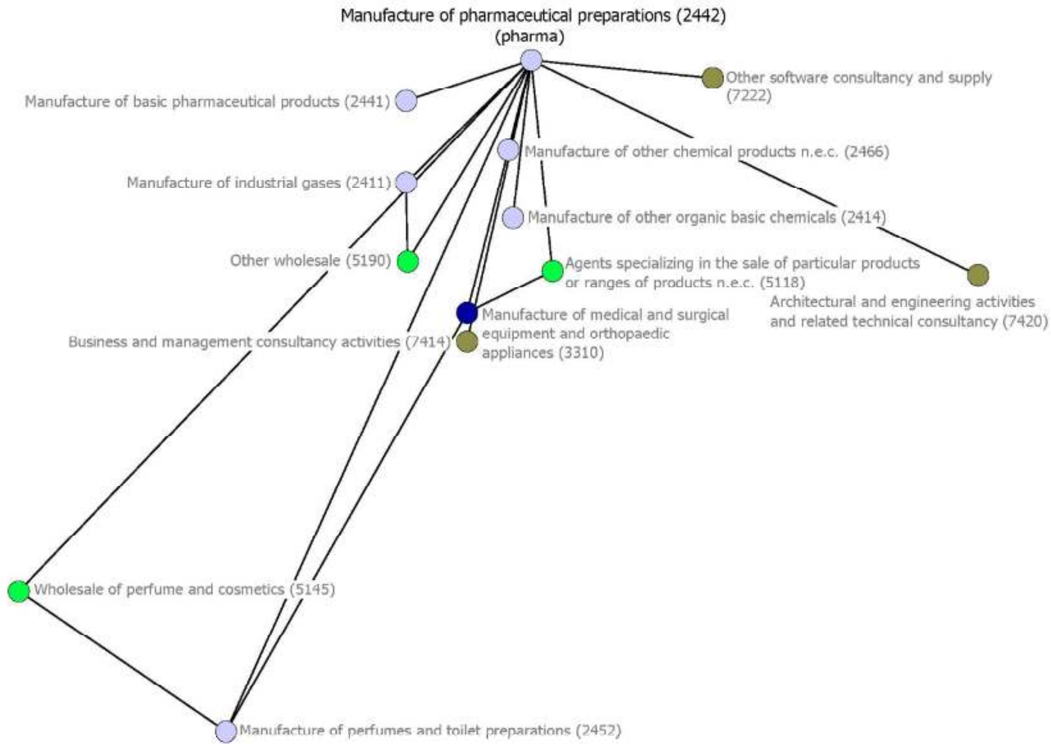

Figure 2: the ego-network of the pharma industry (2442).

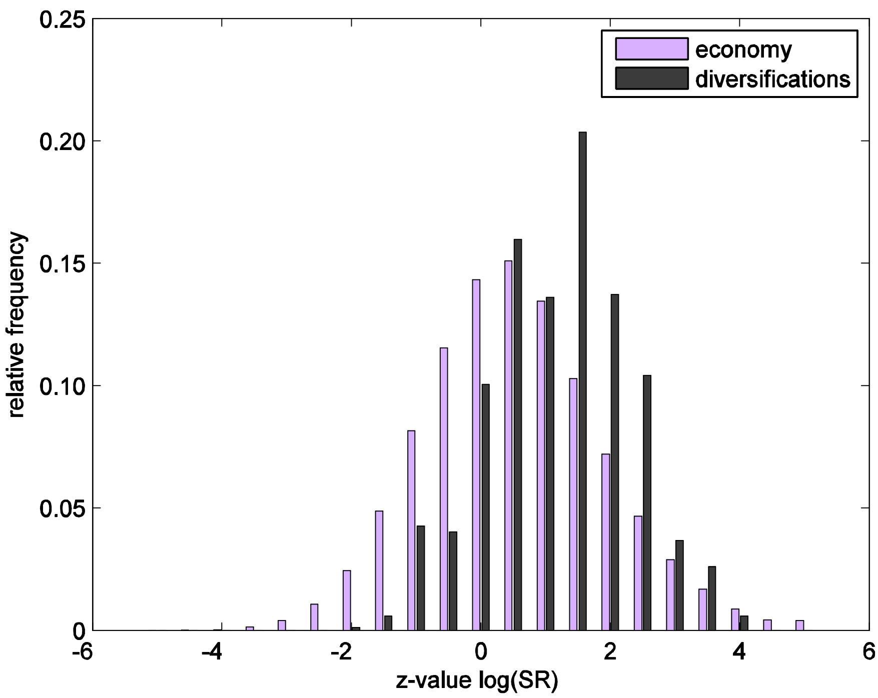

Figure 3: histogram of z-values log skill-relatedness.