LABOR MOBILITY AND INDUSTRIAL SPACE IN ARGENTINA

SERGIO ANDRÉS DE RACO
Instituto Interdisciplinario de Economía Política (IIEP-BAIRES)
University of Buenos Aires
Buenos Aires, Argentina

VIKTORIYA SEMESHENKO*
Instituto Interdisciplinario de Economía Política (IIEP-BAIRES)
UBA-CONICET
Buenos Aires, Argentina

ABSTRACT. In this paper, we apply the skill-relatedness (SR) indicator measure for the analysis of labor flow dynamics in Argentina, and compare it with the original flows in order to differentiate the type of information that each of these techniques offers for characterizing the productive system based on the dynamics of private formal employment. On the other hand, given the size and complexity of the obtained networks, it is interesting to explore the biases introduced by different methods of network reduction in the derived structures as well as to characterize the obtained industrial spaces.

1. Introduction. Labor flows convey information about relevant aspects and structure of the formal labor market. The administrative records that provide base to these flows naturally derive a temporary bipartite network of connections among workers and industries, from which a unimodal projection of interactions between industries can be extracted. In particular, by observing employment transitions in the light of employers’ economic activity, valuable information can be extracted from the exchange of skills and abilities among different sectors, which in turn allows to characterise the underlying aspects of the national productive structure.

Recently, various interesting developments came into view in the literature regarding the analysis of network structures in economics and, in particular, in applications to the analysis of labor flows. Here, we recall the most relevant ones to the objectives of the paper. Hidalgo and Hausmann [4] examined the economy of a country interpreting the foreign trade data as a bipartite network in which countries are connected to the exported products, and quantified the complexity of that economy by characterizing the structure of the resulted network. This “product space” formalizes the idea of the association between products exchanged in the global economy, and quantifies the relationship of products with a measure called proximity. The distances between products that emerge in this space serve as a relevant indicator for predicting countries’ trade diversification potential.

Recently, Neffke et al. [5, 6] developed a framework to be able to analyse labor flow dynamics based on administrative data and introduced the skill relatedness

2010 Mathematics Subject Classification. Primary: 91B40, 68R10; Secondary: 90B10.
Key words and phrases. Labor mobility, industrial space, network analysis.

* Corresponding author: Viktortiya Semeshenko.
(SR) indicator measure, which provides information about the skills that are exchanged between industries through labor flows. This indicator, by construction, quantifies flows deviations with respect to a null model in terms of a ratio of similarity between the observed and the expected values, and reflects deviations of the flows with respect to a random behaviour. Neffke and Henning [5] defined the affinity relationship of industry skills using the information of inter-industry labor flows and derived the industrial space for German labor market. The authors found that this indicator allows to extract a relevant structure of inter-industry connections that differs from the hierarchical structure of the standard systems of classification of industrial activities, and resulted to be significant in predicting the diversification of activities of the firms. Neffke et al. [6] extended the methodological framework of analysis of inter-industry labor flows in order to incorporate network reduction techniques which help to characterize the “industrial space”, and used econometric methods for estimating and analysing different segments of the labor market. Then, by implementing network reduction techniques and weighting the flows with the SR indicator, it is possible to extract the “industrial space” as a projection that compiles the information about the structure of the interactions between productive sectors.

De Raco and Tumini [3] calculated the skill-relatedness indicator for the case of Argentina. The authors analysed clusters of productive activities characterized by “mixtures” of partial productive linkages connected to each other and interpretable in terms of levels of training, female participation and other economic attributes and social factors associated with labor mobility.

In this paper we analyse the structure of interactions that emerges from private formal labor transitions data in Argentina, in a period between 2009 and 2014, in terms of the matrices of the observed average flows and the associated SR indicator. We explore and compare different techniques for reducing the resulting networks, - filtering algorithms and the maximum spanning tree - , with the aim of seeking a simplified and informative representation of the interactions.

We found that the different techniques used to reduce the network provide complementary information. The reduced representation obtained by means of flow filters allows to appreciate cohesive subgraphs of exchanges of employment by strata of force (strength), while representation of maximum spanning tree based on the SR indicator (industrial space [5]) facilitates the interpretation in terms of partial productive linkages by sections. More details will be provided in section 4.

The paper is organized as follows: the data is described in section 2, the methodology is presented in section 3. We discuss the obtained results in section 4 and conclude in section 5.

2. Data. Our data is constructed from private labor flows from the Argentinian Integrated Pension System (SIPA\textsuperscript{1}), provided by the Ministry of Labor, Employment and Social Security (MTEySS\textsuperscript{2}). We confine our analysis to the period of 2009-2014\textsuperscript{3}, and data correspond to four digits of the activity classifier ISIC\textsuperscript{4} Rev 4, enclosing 417 sectors of activity organized in 21 sections (see Table 1).

\textsuperscript{1}Spanish spelling: Sistema Integrado Previsional Argentino
\textsuperscript{2}Spanish spelling: Ministerio de Trabajo, Empleo y Seguridad Social
\textsuperscript{3}This dataset is contemplated within the project ARG126 with the MTEySS and ILO (International Labour Organization).
\textsuperscript{4}The International Standard Industrial Classification of All Economic Activities (ISIC).
We calculate the transition matrices of year-to-year changes in labor for each period, and then the averaged transition matrix for the years 2009-2014. The usual methodology for constructing the labor transition matrices used by the MTEySS consists in creation of year-to-year panels of individuals that appear in both periods for a reference month, followed by selecting the set of workers declared in period $t$ and in $t+1$ for which employer changes are identified. In this paper, only inter-industry flows were used, which correspond to individuals employed in $t$ and in $t+1$ who appear employed in a different firm in another sector of activity. Additionally, we used data from the averaged endowment of the last period in order to scale by the size of each activity sector.

3. Methodology. We use the following reducing techniques to obtain networks for their analysis: the global threshold filters to characterize the flow network and skill-relatedness indicator to extract the industrial space [6].

The skill relatedness indicator is obtained by comparing the observed flows with a null model generated proportionally to the observed values, similar to a Pearson Chi-square test in contingency tables, constructed according to the procedure described below.

3.1. Null model. A reference matrix of “expected flows”, $F^*$, built on the basis of the edges (eg: totals per rows, $F_i$, columns, $F_j$, and table, $F_{..}$) of the matrix of

| Section | Divisions | Description |
|---------|-----------|-------------|
| A       | 01-03     | Agriculture, forestry and fishing |
| B       | 05-09     | Mining and quarrying |
| C       | 10-33     | Manufacturing |
| D       | 35        | Electricity, gas, steam and air conditioning supply |
| E       | 36-39     | Water supply; sewerage, waste management and remediation activities |
| F       | 41-43     | Construction |
| G       | 45-47     | Wholesale and retail trade; repair of motor vehicles and motorcycles |
| H       | 49-53     | Transportation and storage |
| I       | 55-56     | Accommodation and food service activities |
| J       | 58-63     | Information and communication |
| K       | 64-66     | Financial and insurance activities |
| L       | 68        | Real estate activities |
| M       | 69-75     | Professional, scientific and technical activities |
| N       | 77-82     | Administrative and support service activities |
| O       | 84        | Public administration and defence; compulsory social security |
| P       | 85        | Education |
| Q       | 86-88     | Human health and social work activities |
| R       | 90-93     | Arts, entertainment and recreation |
| S       | 94-96     | Other service activities |
| T       | 97-98     | Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use |
| U       | 99        | Activities of extraterritorial organizations and bodies |

Table 1. Broad Structure of Economic activities (ISIC Rev 4, UNSD).
observed flows is calculated in such a way that each element satisfies with:

\[ F_{ij}^* = \frac{F_i F_j}{F_{..}} = \left( \sum \text{flows of industry i} \right) \ast \left( \sum \text{flows of industry j} \right) \frac{1}{\sum \text{total interflows}}. \] (1)

This matrix reflects “random” flows in the sense that sectorial exchanges are proportional to the outflows and inflows between sectors with respect to total flows.

### 3.2. Skill-relatedness.

For each cell an associated matrix of elements, \( SR_{ij} \), is calculated as the ratio of the observed value of employment flows with respect to the theoretical or expected value, so that:

\[ SR_{ij} = \frac{F_{ij} F_{..}}{F_i F_j} = \frac{\text{observed}}{\text{expected}}, \text{ with } SR_{ij} \in [0, +\infty). \] (2)

Thus, one can interpret values less than unity, \( SR_{ij} \in [0, 1) \) as not moving away from a random distribution significantly, while values greater than unity, \( SR_{ij} \in [1, +\infty) \), showing deviations from the proposed random distribution as benchmark. The principal diagonal elements of this matrix, \( SR_{ii} \), reflect the flows of employment within the same sector for which the authors assume \( SR_{ii} = 1 \). Since this indicator is not bounded to the right, normalization is used in the interval \([-1, 1)\). As suggested in [7], the SR matrix is symmetrized by means of averaging the SR matrix with its transpose. In this way the related graph becomes undirected.

We used global thresholds and maximum spanning tree (MST) in order to get the simplified representation of the network, which are techniques for reducing networks or edge trimming utilized in dense graphs. The filtering techniques allow to conserve the observation scale, in contrast with other reduction techniques that redefine the nodes in terms of selected attributes that aggregate the interactions (coarse-graining).

The global thresholds consist in the establishment of rules that define a minimum value of flows for edge conservation, usually based on domain considerations. A relevant limitation of these methods is the undervaluation of nodes with an incidence of low flow values. This technique was used on the original flow matrix for exploratory purposes, setting up thresholds based on the observed weight distribution.

The maximum (minimum) spanning tree implemented in connected, weighted and undirected graphs contain \( n \) nodes and \( n - 1 \) edges, at the same time these graphs do not present cycles and have the maximum (minimum) possible weight in their edges. As said before, the SR indicator measures the degree of relatedness as deviations from a random benchmark. Thus, we are interested in building the network tree based on the maximum SR flows, for which we used Prim’s [8] algorithm implemented in igraph software [2]. The reduced network results in an extremely simplified, acyclic structure, which affects the grouping coefficient and the grouping hierarchies commonly presented in real world networks. This is highlighted as an important limitation of expansion tree methods in the literature.

### 4. Results.

Given the matrix of year-to-year flows of average private formal employment, during the period 2009-2014, we analysed the structure of the static network of interactions between activity sectors. The total of changes in labor among 417 sectors (4 digit, ISIC Rev 4) amounted to 412,103, which represented 7.3% of the average annual private formal employment that in 2014 added up to 5.6 million individuals.

Regarding this average endowment, it was observed that the sectoral participation of employment was approximately log normally distributed, with 80% of
employment concentrated in 26% of the sectors. In terms of labor flows, nearly half were produced between industries belonging to the same sector of activity (main diagonal), while the other half were produced between industries belonging to different sectors.

In line with the results presented in Neffke et al. [6] for Germany (see Table 2), in the decomposition of changes between sectors it was observed that approximately 58% of them occurred between sectors that differed in the higher level classification (letter or ISIC sections).

The observed movements involve considerable distances when a part of the flows occurs between extractive activities and services, or industries and social services. Given that the focus of the paper is to characterize the relationships between industries, only cross-sectoral flows were used in the subsequent analysis. This choice defines two networks represented by simple weighted graphs. In the case of the normalized SR matrix the associated graph is undirected because of the symmetrization procedure, while in the case of the pure flows matrix the associated graph is directed.

| Workers (thousands) | Germany 1999-2008 | Argentina 2009-2014 |
|---------------------|-------------------|---------------------|
|                     | NACE Rev 1.1 5-digits | ISIC Rev 4 4-digits |
| **Stock:**          |                   |                     |
| Employment          | 19897.1           | 5619.1              |
| **Absolute flows:**|                   |                     |
| Job Switchers        | 1206.7            | 412.1               |
| No Industry Switch   | 321.1             | 206.1               |
| Industry Switch      | 885.7             | 206                 |
| Different Sector     | 519.8             | 119.4               |
| Same Sector          | 365.9             | 86.6                |
| Same 2-Digit Industry| 225.9             | 37.1                |
| Same 3-Digit Industry| 117.3             | 14.6                |
| Same 4-Digit Industry| 62.3              | n.a.                |
| **Relative flows (w.r.t. total average employment):** |                   |                     |
| Job Switchers        | 6.1%              | 7.3%                |
| No Industry Switch   | 26.6%             | 50.0%               |
| Industry Switch      | 73.4%             | 50.0%               |
| Different Sector     | 58.7%             | 58.0%               |
| Same Sector          | 41.3%             | 42.0%               |
| Same 2-Digit Industry| 25.5%             | 18.0%               |
| Same 3-Digit Industry| 13.2%             | 7.1%                |
| Same 4-Digit Industry| 7.0%              | n.a.                |

Table 2. Average employment and cross-industry labor flows: In relative terms with respect to average employment, the proportion of inter-industry flows in Germany is greater than in the case of Argentina, possibly due to the greater level of sectoral detail (5-digits), and hence the total number of nodes in the network. However, the proportion of inter-industry flows that cross the section level (letter) is similar (58.7% vs 58%).

The authors analyze data to five digits of the activity classifier and find a greater proportion of intersectoral changes, between 75 and 80% of these flows, which may be related to the effects of the higher level of disaggregation of information and the consequent impact in the definition of sector and intersectoral flow.
The data set under analysis consisted of 416 sectors. We excluded from the set a sector with exit flows only, because it is not possible to calculate the skill-relatedness indicator in this case. Then, we proceeded to calculate the associated flows and normalized SR matrices. These matrices were used, in the subsequent analysis, to weight the networks of interactions in order to filter the flow network and extract the MST that generates the industrial space, respectively.

4.1. **Network description.** The network of labor flows presents a dense structure with a single connected component, of a reduced diameter (3 steps) and exhibiting small world properties. The distribution of in-degrees and out-degrees does not follow a known distribution, but rather a mixture similar to a variant of the Poisson distribution, while they clearly do not present a Power law, neither normal nor uniform distributions.

At the same time, a good part of the nodes show high degrees. This feature contributes to the high connectivity of the network and the absence of nodes that behave as "hubs". The graph presents a density of 0.44 (76,544 observed connections over 173,056 possible links), and a diameter of 3 steps, which in terms of flows represents 48 individuals. The shortest distance observed between two sectors is of one step in 44% of cases and of two steps in 56%. The measure of reciprocity calculated as the ratio between mutual ties (i.e., corresponded flows between sectors) with respect to the total of present ties is 0.66, that is, 2 out of 3 labor exchanges observed between two sectors were carried out in both directions. The measure of reciprocity relative to the nodes, calculated as the reciprocal ties with respect to the non-reciprocal ones, turned out to be 0.8, with which 4 out of 5 of the intersectorial relations were reciprocal. These metrics describe a network with apparently high reciprocity. The clustering coefficient or transitivity of a graph is a global measure of grouping that can be calculated on the triples of the graph (e.g., subsets of three nodes connected by two or three edges) and considers the proportion of closed triples over the total of triples. The flow network showed the clustering coefficient of 0.74, which implies the observed links in triples between groups of three nodes, 3 out of 4 times transitivity is observed in the relationships, which is an indicator of cohesion that contributes to density and the groupings of nodes.

In terms of structural models, the transitivity observed in the original flow matrix together with the shortest path lengths distribution (see Figure 1(d)) indicated the presence of Watts-Strogatz "small world" properties. From a macro point of view, when analysing the network through hierarchical clusters a clear core-periphery structure was found (see Figure 2). We observed the presence of a nucleus composed of a few sectors highly interconnected with each other and with the rest of the sectors (core), two groups with small density that behave as nuclei of second and third order, and a fourth peripheral group, composed of sectors disconnected from each other and connected to a greater or lesser extent with the core.

Finally, we observed that the degree distribution of the graph resembled more likely a uniform distribution than a power law, such as that generated by the Barabási-Albert model. This implies that a few nodes were not observed to fulfill the role of "hubs" of interconnection, but that most of the nodes present relatively high connectivity, similar to the average.

4.2. **Network reduction.** In the first place, the network of flows was reduced using global threshold filters. From the flows distribution, which presented a normal log distribution with mean 3 and maximum 2650 individuals (see Figure 1(c)), we...
Figure 1. Network of Flows: Structural aspects of the inter-industrial flows graph of average formal employment 2009-2014. (a) Graph obtained with Fruchterman-Reingold layout. (b) Distribution of the average employment endowment by sectors and activity sections. (c) Distribution of interindustrial flows (weights, strength). (d) Distribution of distances of the unweighted graph.
Figure 2. Structural aspects of the directed flow graph based on models. (a) A center-periphery structure is observed, with few sectors highly interconnected with each other and with the others (center), other two groups with less interaction with each other although with characteristics similar to the center, and a last group with little interaction between them and greater interaction with the rest of the sectors (periphery). (b) The distribution of the weighted graph of original flows presented a very different distribution to that generated by the preferential attachment model (Barabási-Albert) with the parameters of the graph depicted in (c), more similar to a distribution of normal multimodal mixtures than a power law.

explored different thresholds with 5, 10, 25, 50, 75, 100, 250, 500, 750 and 1000 individuals.

It was observed that modifying thresholds generates graphs from disconnected nodes to graphs with components of different sizes. Starting from low thresholds of 10 individuals disconnected nodes appear that grow until reaching 410 nodes, with threshold of 50 individuals the graph becomes disconnected in components until reaching 36 components when considering the threshold of 100 individuals. With high thresholds (about 1000 individuals) the graph is reduced to 3 components.
The observed components allow to identify sectors that are present in different scales of interactions, and the used technique is useful for understanding the flow structure based on their size. On the other hand, the observed flow weights distribution (i.e.: strength) was log-normal with small mean and median values. Therefore, upon imposing a threshold of 5 individuals the resulting graph keeps all the nodes in a single dense component with 8.5% of the original graph edges, accumulating 65.9% of total inter-industry labor flows. This means that on average almost one third of the observed flows involve less than 5 individuals.

When analyzing this filtered graph under the light of the broadest classification category (e.g.: first hierarchical level, section)\(^6\) we observed sections located in specific regions of the graph (i.e.: primary activities, social services, education, construction, information and communication), while others spread across the graph (i.e.: commerce, transport and storage services, industry, administrative activities, other services) according to expectations based on a productive linkages view of economic activities (see Figure 3). A productive linkage view refers to the interconnection of different economic activities in terms of supply (upstream) and demand (downstream) activities, related to each specific economic activity by means of primary or support activities needed for its productive process. This structure emerged as a result of the industry definition granularity (ISIC Rev4, 4 digits), as otherwise at the more commonly used 2-digits industry definition the productive linkages are less evident. In this way, it was possible to clearly visualize interactions between sections of economic activities and generate a coherent map with the application of a threshold that does not impose great restrictions on the absolute flows (i.e.: strength).

Then, we proceeded to generate the MST on the weighted graph using the skill-relatedness indicator to extract the industrial space. In the resulting graph groups of activities belonging to the same section and some value chains are observed in a partial way (see Figure 4).

It is interesting to compare Figure 2(a) and Figure 4(a), generated on two different graphs. The first one is based on the directed network of flows, while the other one represents the undirected SR matrix. In Figure 2(a), a core-periphery structure is observed with few sectors highly interconnected with each other and with the other (core); two other groups with less interaction with each other, although with similar characteristics to the center; and a last group with little interaction between them and a greater interaction with the rest of the sectors (periphery). In Figure 4(a) two groups appear: one relatively well connected containing half of the activity sectors (right and bottom corner), and another poorly connected containing the other half.

5. Conclusions. In the analysis carried out on the network of inter-industry labor flows for Argentina, it was found that the network is dense, with a single large connected component and exhibits small world properties and center periphery characteristics. The structure of the original graph of flows has been useful in the representation of the interactions at the industry level, while the expression of the industrial space based on the normalized and symmetrized skill-relatedness indicator (SR) was considered to be relevant in terms of the ordering of the productive activities, and the additional information of significance of observed flows with respect to random flows.

\(^6\)See Table 1 for section descriptions. Graphs in Figures 1, 3, and 4 are colored by sections.
Figure 3. Reduction of the original flow network. (a) Logarithmic relationship between flows and edges according to the selected thresholds. (b) Graph obtained with Fruchterman-Reingold layout for the restriction of 5 or more individuals. A structure was observed by specific activity sections located in concentrated regions and in some cases (predominantly services and industry) distributed in the graph.

Our results demonstrated that the methodology developed by Neffke et al. [6] is informative for the analysis of labor flows, and the indicator of skill-relatedness reveals the existence of relevant structural aspects of the network.

The comparison of different network reduction methods allowed us to conclude that the application of global thresholds was useful for the visualization of the network and the interpretation of affinities between productive activities at the scale level of the flows, and the organization of sectoral production. On the other
hand, the method of the maximum spanning tree applied to the normalized and symmetrized SR indicator matrix allowed us to visualize a network of partial productive linkages in terms of relationship significance, which provides robustness to the subsequent interpretation of relationships.
The information extracted from the application of the SR methodology to the mobility of formal employment in Argentina adds value to the description and interpretation of the complex phenomenon of labor relations that defines the productive framework. However, the characterization of an industrial space is limited by various factors not originally contemplated in the works of Neffke and co-authors.

In particular, unlike what happened in the countries analyzed and on which this methodology has been applied, the factors such as the structure of employment, the variation of the sector composition in terms of macroeconomic performance, and the high proportion of informal employment in total employment as well as sectoral factors merit further analysis for the case of Argentina and other developing countries.

Regarding the structure of employment, the complexity of the productive structure of developed countries differs substantially from that of developing countries [4]. It is therefore expected that the generated industrial spaces are different both qualitatively and quantitatively. With respect to the effects of macroeconomic fluctuations on employment and its sectoral composition, it is to be expected that the greater macroeconomic volatility observed in developing countries will affect negatively the overall employment and differentially in terms of sectoral employment. Finally, the informal employment proportion with respect to total employment (or total salaried employment) in Argentina has been historically high (although declining the last decade, currently at about 34% of total salaried employment), which is unequally distributed among the different sectors of employment activity (see [1]).

REFERENCES

[1] F. Bertranou and L. Casanova, Informalidad Laboral en Argentina: Segmentos Críticos y Políticas Para la Formalización, Organización Internacional del Trabajo, 2013.
[2] G. Csardi and T. Nepusz, The igraph software package for complex networks research, *InterJournal, Complex Systems*, (1965), [http://igraph.org](http://igraph.org)
[3] S. De Raco and L. Tumini, Parentesco de habilidades y nuevos sectores económicos. Un análisis exploratorio para la Argentina, Documento de Trabajo MTEYSS-OIT, 2018.
[4] C. Hidalgo and R. Hausmann, The building blocks of economic complexity, *Proceedings of the National Academy of Science*, 106 (2009), 10570–10575, [https://doi.org/10.1073/pnas.0900943106](https://doi.org/10.1073/pnas.0900943106).
[5] F. Neffke and M. Henning, Skill-relatedness and Firm diversification, *Strategic Management Journal*, 34 (2013), 297–316.
[6] F. Neffke, A. Otto and A. Weyh, Inter-industry labor flows, *Journal of Economic Behavior & Organization*, 142 (2017), 275–292.
[7] F. Neffke, A. Otto and A. Weyh, Skill-relatedness Matrices for Germany: Data Method and Access, FDZ Methodenreport 201704, en. Institutfur Arbeitsmarkt- und Berufsforschung (IAB), Nurnberg.
[8] R. C. Prim, Shortest connection networks and some generalizations, *Bell System Technical Journal*, 36 (1957), 1389–1401.

Received December 2018; revised March 2019.

E-mail address: sergioderaco@gmail.com
E-mail address: vika.semeshenko@gmail.com