Comparación de Redes de Parentesco de Habilidades: Conectividades estructural y de Parentesco en la Movilidad Laboral
Viktoriya Semeshenko, Sergio A. De Raco
SADIO Electronic Journal of Informatics and Operations Research (EJS) Vol. 24 No. 1 (2025) e-ISSN 1514-6774
https://doi.org/10.24215/15146774e065 | https://revistas.unlp.edu.ar/ejs
Sociedad Argentina de Informática e Investigación Operativa | Universidad Nacional de La Plata | Buenos Aires | Argentina

Comparing Skill-Relatedness Networks: Structural Linkages vs. Relatedness in Labor Mobility

Viktoriya Semeshenko[1][0000−0003−0295−5946] and
Sergio A. De Raco[2][0000−0002−1240−5799]

[1] Universidad de Buenos Aires. Facultad de Ciencias Económicas. Buenos Aires, Argentina. CONICET-Universidad de Buenos Aires. Instituto Interdisciplinario de Economía Política de Buenos Aires. Buenos Aires, Argentina
vika.semeshenko@gmail.com
[2] Universidad de Buenos Aires. Facultad de Ciencias Económicas. Buenos Aires, Argentina. Universidad de Buenos Aires. Instituto Interdisciplinario de Economía Política de Buenos Aires. Buenos Aires, Argentina
sergioderaco@gmail.com

Abstract. In this paper, we compare Skill-Relatedness Networks (SRNs) across selected countries, representing statistically significant interindustrial interactions that capture latent skill exchanges derived from observed labor flows. Using data from Argentina (ARG), Germany (DEU), and Sweden (SWE), we analyze their SRNs through an information-theoretic method designed to compare networks with non-aligned nodes, a crucial aspect for cross-country comparisons. By extracting network portraits—structural fingerprints based on shortest path distributions—we measure pairwise divergences to contrast differences in binary connectivity and weighted skill-relatedness across countries.

Our findings reveal that ARG's SRN structural connectivity differs significantly from those of DEU and SWE, while at the same time also contrast with each other. These findings suggest that the fundamental structure of skill-related interconnections is country specific. However, when viewed through the lens of the SR indicator, the differences between countries become less pronounced, suggesting a universal phenomenon in skill exchanges, highlighting a structured pattern of labor mobility across sectors in any national economy. These findings support the idea that historical and cultural factors shape SRNs, but structural connectivity remains country-specific. While skill intensity patterns (weighted SRNs) appear consistent across economies, the topological structure (binary SRNs) varies sharply, highlighting distinct labor market dynamics, patterns of specialization and pools of skills in each country.

Keywords: Administrative Data · Labor mobility · Skill-Relatedness · Network comparison · Network Portraits.

Received September 2024; Accepted December 2024; Published March 2025

Comparación de Redes de Parentesco de Habilidades: Conectividades estructural y de Parentesco en la Movilidad Laboral

Viktoriya Semeshenko[1] [0000−0003−0295−5946] and
Sergio A. De Raco[2] [0000−0002−1240−5799]

[1] Universidad de Buenos Aires. Facultad de Ciencias Económicas. Buenos Aires, Argentina. CONICET-Universidad de Buenos Aires. Instituto Interdisciplinario de Economía Política de Buenos Aires. Buenos Aires, Argentina
vika.semeshenko@gmail.com

[2] Universidad de Buenos Aires. Facultad de Ciencias Económicas. Buenos Aires, Argentina. Universidad de Buenos Aires. Instituto Interdisciplinario de Economía Política de Buenos Aires. Buenos Aires, Argentina
sergioderaco@gmail.com

Abstract. En este trabajo comparamos redes de parentesco de habilidades (SRNs), que resumen interacciones interindustriales estadísticamente significativas que representan intercambios latentes de habilidades derivados de flujos laborales observados. Utilizando datos de Argentina (ARG), Alemania (DEU) y Suecia (SWE), analizamos sus SRN mediante un método basado en teoría de la información diseñado para comparar redes con nodos no alineados. Mediante la extracción de *network portraits*-identificadores estructurales basados en las distribuciones de caminos cortos- medimos las divergencias entre pares para contrastar las diferencias en la conectividad binaria y la relación ponderada entre parentescos de habilidades de estos países. Encontramos que la conectividad estructural de la SRN de ARG difiere significativamente de las de DEU y SWE, al mismo tiempo que las conectividades entre estos países contrastan entre sí. Ello sugiere que la existencia de especificidades en las estructuras fundamentales de interconexiones relacionadas con las habilidades en cada país.
En cambio, cuando al utilizar el indicador de parentesco de habilidades, las diferencias se vuelven menos pronunciadas sugiriendo un fenómeno universal en los intercambios de habilidades, destacando un posible patrón estructurado de movilidad laboral intersectorial en cualquier economía nacional. Estos resultados apoyan la idea de que los factores históricos y culturales conforman las SRNs, pero la conectividad estructural sigue siendo específica de cada país. Mientras que los patrones de intensidad de las habilidades (ponderados) parecen coherentes en las economías analizadas, la estructura topológica (binaria) varía notablemente, poniendo de relieve las distintas dinámicas del mercado laboral, los patrones de especialización y los conjuntos de habilidades disponibles de cada país.

Keywords: Datos Administrativos · Movilidad Laboral · Parentesco de Habilidades · Comparación de Redes · Portraits de Redes.

1 Introduction

Labor flows are an essential factor to understand economic activity, as they reflect the interaction between labor supply and employer demand in the labor market. Among these, job-to-job transitions are particularly relevant, displaying well-documented pro-cyclical behavior [9]. Such transitions provide implicit insights into the relevance of previous work experience for new employers, especially those that occur between firms with different economic activities. Understanding these transitions is essential for understanding how skills and expertise transfer across economic sectors.

In this context, the alignment between the skills demanded by firms and those supplied by workers represents a dimension where significant gaps may exist, making it a highly relevant area of research. Skill mismatches can take various forms, such as skills gaps, skills shortages, geographical disparities, etc. [4,13]. However, rather than focusing on these dimensions, our analysis takes a different approach. We assume that observed job transitions reflect a form of proximity in terms of skills requirements between economic activities. In other words, we consider that workers predominantly transition to sectors where their existing skills remain relevant and transferable. This assumption enables us to interpret job mobility patterns as the interconnectedness of industries within the productive structure through overlapping skill sets.

Traditionally, economists analyze labor flows using data at high level of aggregation of the standard classifications of productive activities, in order to correlate it with conventional national accounts data of sectoral activity. The evolution of labor flows in Argentina has been analyzed using administrative records, and it has been shown that more disaggregated data can provide a richer picture of the temporal evolution of labor flows than aggregated data [16]. This is because labor flows carry information about the productive structure and diffuse knowledge among economic activities. Clearly, a more disaggregated level of detail, at the same time brings more complexity in interpretation tasks.

Labor mobility across different industries reflects interconnections between economic activities, which can be effectively represented as networks. These networks highlight the properties of connectivity between economic sectors, offering insights into the flow of labor and the relationships between various industries within an economy. Various networks of connections can be derived from these interactions. First, labor flow networks (LFN) can be constructed directly from job transitions between industries, capturing employment exchanges as workers move from firms in one economic sector to another. Building on the idea that these transitions reflect skill relatedness between sectors, we can further extract skill relatedness networks [12,17], which underlie the observed labor flow patterns. In this work, our primary focus is on analyzing these networks to better understand the structure and to compare their structure between different countries.

In Argentina, the Undersecretariat for Employment and Job Training has data of administrative records of formal private labor employment from the Argentine Pension System[3] provided by the Observatory of Business and Employment Dynamics[4]. The data covers interannual employment exchanges between productive economic activities registered between 2009 and 2014. This set of activities includes nearly 400 sectors at four digits of ISIC[5] Rev.4 classifier.

Previously, in [7,8] the inter-industry labor flows of Argentina have been studied at high level of detail, and it has been revealed that extracted labor networks were characteristically dense, with clear core-periphery structures, and small-world properties. These labor flow networks are characterized by a high prevalence of small-flow transitions, which increase their density and connectivity. As a result, the networks are not sparse and exhibit a highly interconnected structure with a giant component. This indicates that high granularity job transitions between industries are frequent and very heterogeneous.

In turn, this connected structure evidence short (average path) distances between pairs of any given sectors, v.g.: a small-world, which has also functional structure between groups of sectors, some acting like "cores" while others acting as "peripheries". Nevertheless, the prevalence of small flows raises the issue of significancy or relevance of these transitions, requiring the application of network reduction techniques to filter out weaker links and extract the backbone of significant connections (see, for example [7,8,6]). In this study, we employ a specific filtering technique to retain a significant backbone of the network which will be explained in the following section. Although these microscale networks provide new and useful information, they also pose several challenges for their interpretation and applications in, for example, policy design and analysis. Also, from a temporal analysis perspective, the structure of interannual labor networks varies over time due to both cyclical and structural factors (see [7,16,6]).

In this study, we focus on skill-relatedness networks (SRNs) to explore the structure of skills overlap between industries, as inferred from labor flow transitions. We are interested in comparing the SRN of Argentina (ARG) with those of Germany (DEU) and Sweden (SWE)[6]. This comparative approach allows us to investigate how inter-industry labor networks differ between developing and developed economies. By contrasting these SRNs structures, we look for insights into the variations in industrial interactions across different economic, social, and historical contexts.

Comparing networks is a challenging task. When given two networks, determining how similar they are typically involves quantifying their structural, topological, or functional similarities. This typically requires measuring connectivity patterns, node relationships and overall organization. Various methods have been developed [5,15] to address this challenge: graph invariants (fixed properties such as degree distribution), network measures (structural characteristics such as modularity and centrality), graph matching algorithms (align nodes and edges), information-theoretic methods (measure variations in network

[3] Spanish: Sistema Integrado Previsional Argentino (SIPA).

[4] Spanish: Observatorio de Empleo y Dinámica Empresarial (OEDE).

[5] International Standard Industrial Classification of economic activities (ISIC).

[6] Germany and Sweden were selected due to data availability.

entropy), network alignment (preserves structural relationships), and machine learning approaches. Choosing an appropriate method depends on the specific characteristics of the networks and the research question at hand.

The proposed challenge presents a new problem when the networks belong to systems of different dimensions, in terms of network science. Indeed, approaches to network comparison can be roughly divided into two groups based on whether or not they consider two graphs defined on the same set of nodes. When we consider networks defined on the same set of nodes, the comparison becomes straightforward since there is no need to align nodes between the two networks. For example, the cases of comparison of SRNs with the same number of (aligned) nodes has been already done by [17]. However, even if two networks have identical topologies, they might have no nodes or edges in common simply because they are defined on different sets of nodes. This highlights the importance of carefully considering the context and objectives when choosing a comparison approach for networks.

In the present case, we are dealing with a "non-aligned" network comparison, v.g.: not the same nodes are necessarily shared between the networks. For this, we are using *portraits divergence*, a method for characterizing large complex networks by introducing a new matrix structure, unique for a given network, which encodes structural information, provides useful visualization, and allows for rigorous statistical comparison between networks [2]. In particular, the *network portrait* encodes the information of its connectivity structure by representing how nodes (sectors, in our case), relate to each other based on their position in the entire network. This is visualized (like in an onion) as layers of connections, where the distance between layers represents increasing degrees of separation between nodes. This can also be understood as successive neighborhoods of connectivity: i.e. how each node is connected to its neighbors (first layer), the neighbors of those neighbors (second layer), and so on, creating a multi-layered map of connections, its fingerprint. This helps to identify clusters of closely connected sectors or entities. In the context of skill-relatedness networks, it can serve as a map of opportunities generated by the exchange of skills between sectors. This allows us to identify areas where workers might transition between industries or sectors with similar skills sets. The *portrait divergence* is then used to quantify the similarity between pairs of these network fingerprints. The fingerprints provide exhaustive descriptions of structural connectivity information, summarizing the mutual information between them. The intuition behind portrait comparison is that a network portrait acts like a topographic map of labor mobility, showing how industries are interconnected and how easily workers transition between them. Comparing two labor markets through this method is similar to analyzing different geographical landscapes—some may have smooth, well-connected pathways (indicating high labor mobility), while others may feature isolated peaks and valleys (suggesting low connectivity). The network portrait divergence then serves as a tool to quantify how similar or different these labor mobility landscapes are.

In this study, we compare skill-relatedness networks (SRNs) across Argentina, Germany, and Sweden to understand differences in labor market structures between developing and developed economies. The key findings can be summarized as follows. Using network portraits and network portrait divergence, we find these methods to be effective in analyzing and comparing SRNs of different sizes. The analysis of binary SRNs (pure connectivity) reveals significant differences between countries, suggesting that the fundamental structure of skill-related interconnections varies significantly across countries, forming particular "skeletons" of sectoral linkages. When examining weighted SRNs (which account for skills intensities), the differences between countries are less pronounced, suggesting a universal phenomenon in skill exchanges, highlighting a structured pattern of labor mobility across sectors in any national economy. These findings support the idea that historical and cultural factors shape SRNs, but structural connectivity remains country-specific. While skill intensity patterns (weighted SRNs) appear consistent across economies, the topological structure (binary SRNs) varies sharply, highlighting distinct labor market dynamics, patterns of specialization and pools of skills in each country.

The paper is organized as follows. In section 2 we describe the three datasets used in the analysis and introduce the methodology. In section 3 we show the results. In section 4 we discuss our work.

2 Data and Methods

To address the proposed objective, we use three available datasets for selected countries: Argentina (ARG), Germany (DEU), and Sweden (SWE), at the level of 4 digits of detail of their national economic activity classifications, procured from various sources (see Table 1 for details).

The interaction networks are constructed based on these data. These networks are constructed on different sets of nodes, leading to variations in size and differences in the number and types of economic activities they represent. As mentioned in section 1, this presents a challenging problem of comparison and identification. To tackle this, we employ an information-theoretic method called portrait divergence, which allows us to extract structural insights and perform rigorous statistical analyses of the networks.

2.1 Data

We processed data at the four-digit level of national economic activity classifications for ARG, DEU, and SWE, which are compatible with ISIC 4 or its European counterpart, NACE 2, using data from various sources. Table 1 provides a summary of the data, detailing for each country: the classification system used in the available data, the time periods considered for each averaged network, average annual employment, total accumulated flows and yearly averaged inter-industry job transitions, the number of sectors available and the subtotal used in the analysis and the formal data source.

For Argentina, we use labor flows transitions for the period 2009-2014, provided by the Observatory of Employment and Business Dynamics within the Undersecretariat for Employment and Job Training. This data is sourced from administrative records of the Revenue Collection and Customs Control Agency.

For Germany and Sweden, the available data consist of skill-relatedness matrices, enabling us to directly construct the SRNs for each country. In the case of Germany, we use directly the SR data at four digit WZ08 national industrial classification (equivalent to NACE 2), for the period 2007-2013, published in [10] by the authors[7] originally estimated from data of the Employee History[8], based on the social security records of Germany. Additionally, we use German employment data from DESTATIS, the Federal Statistical Office of Germany.

In the case of Sweden, we use directly the SR data at four digit SNI 2007 national industrial classification (equivalent to NACE 2), for the period 2007-2017, calculated by the Swedish Agency for Growth Policy Analysis (see [14]) using the methods in [10] with Swedish administrative data[9]. We use Swedish employment data from Statistics of Sweden for the period of analysis.

	Argentina	**Germany**	**Sweden**
Data	Inter-industry labor flows	Inter-industry skill-relatedness	Inter-industry skill-relatedness
Classification	ISIC 4	WZ08 (NACE 2)	SNI 2007 (NACE 2)
Period	2009-2014	2007-2014	2007–2017
Transitions (#)	5	7	10
Avg. Empl.	5,619,134	28,467,487	4,665,205
Flows			
. total	2,060,515	5,529,890	4,800,000
. avg./year	412,103	789,984	480,000
Sectors (#)			
. original	410	597	586
. SR+	407	584	577
Source	Ex-Ministry of Labor, Employment, and Social Security	Table 2, [11], based on Beschäftigten-Historik, Federal Statistical Office	Rapport 2021:02:04, Swedish Agency for Growth Policy Analysis, based on LISA data, Statistics of Sweden

Table 1. Data reference summary for Argentina, Germany and Sweden. Administrative data at 4 digits of economic activity classifications. Comparable classification systems ISIC 4 and NACE 2.

[7] See "Skill relatedness matrices for Germany" at https://iab.de/publikationen/publikation/?id=7202046.
[8] German: Beschäftigten-Historik, BeH.
[9] See "Skill relatedness matrices for Sweden" at https://www.tillvaxtanalys.se/in-english/publications/pm/pm/2021-05-18-skill-relatedness-matrices-for-sweden.html.

2.2 Methods

Skill-Relatedness Networks. For Argentina, we proceed first to construct the skill-relatedness matrices and then its corresponding SRN [17]. After averaging interannual flows for all transition matrices available, we calculate the skill-relatedness indicator, $SR_{ij}, \forall i, j \in N$, where N represents the total number of industries[10] included. The skill-relatedness indicator between industries i and j is computed as a ratio between the *observed* labor flows and the *expected* flows from a null model, which is calculated from the margins of the respective $(A_{N\times N})$ flow matrix for each cell (see [10,11,12] for further insights on this methodology), as described in Fig. 1. The indicator is then symmetrized and normalized to map it to the interval $\widetilde{SR}_{ij} \in [-1, 1)$, hereafter used as SR to avoid notation cluttering. The SR indicator is a measure of labor mobility between different sectors and the degree to which these sectors are related in terms of required skills. An important aspect of SR is the interpretation of its positive and negative values and their implications for the labor market structure and functioning and the valuation of workers' skills.

Positive (negative) values of SR_{ij} indicate that there are more (fewer) observed employment exchanges between two sectors i and j than would be expected under a random mobility model. In this context, we can interpret that the skills demanded by sector i can (or cannot) be found in workers coming from sector j. Additionally, the more positive (negative) the SR_{ij} value, the more (less) valuable the skills from one sector appear to be in the other. Limiting to positive values appears to be an appropriate method and a suitable criterion for pruning the networks, removing the less significant interactions in terms of skill-relatedness[11] Values greater than 0 indicate that the number of observed job switches is greater than what would be expected at random under the null model specified, v.g. workers that would have moved at random given the respective size of each industry (similar to the *Configuration Model* [3]). Hereafter we refer to these networks with positive skill-relatedness, $SR_{ij} > 0$, as SRN^{+}s or simply SRNs and conveniently index them by country whenever needed [17].

In the cases of Germany and Sweden, since we have skill-relatedness data, we construct the matrices directly. For the remainder of the analysis, we focus solely on using the SRNs for all three countries. For the subsequent analysis and network comparison, we retain only the positive skill-relatedness values from the SR matrices for the three countries as explained earlier.

Regarding the size of the networks to compare, which refers to the number of nodes, i.e. industries, included in the analysis, it is worth noting that Germany and Sweden have more than 40% industries at their four-digit detailed classification compared to Argentina (see Table 1). We compare these non-aligned

[10] From this point forward, the terms "industries", "economic activities" and "sectors" will be used interchangeably.

[11] Negative skill-relatedness would suggest a lower than expected level of employment exchanges between two sectors, implying that the skills required by one sector are not well aligned or are significantly different from those found in workers transitioning from the other sector.

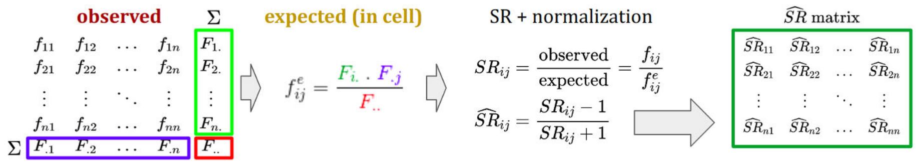

Fig. 1. Construction of skill-relatedness indicator (used for Argentina). Sequential steps of the process to get from the observed flow matrix, F, to the skill relatedness indicator matrix, $\widehat{SR}$ (SR later on). A reference matrix of "expected flows", f^e_{ij}, is built on the basis of the table edges (e.g.: totals per rows, $F_{i.}$, columns, $F_{.j}$, and table, $F_{..}$) of the observed flows matrix. This matrix reflects "random" flows in the sense that sectoral exchanges are proportional to the outflows and inflows between sectors with respect to total flows. 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. Thus, one can interpret values less than unity, $SR_{ij} \in [0, 1)$ as not departing significantly from a random distribution, while values greater than unity, $SR_{ij} \in [1, +\infty)$, showing deviations from the specified random distribution. SR matrix is then symmetrized by means of averaging the SR matrix with its transpose, making the associated graph undirected. Finally, the normalization step leaves out the $\widehat{SR}$ (symmetric) matrix, with $\widehat{SR}_{ij} \in [-1, 1)$, referred in the text simply as SR and used to build the SRN for Argentina.

networks, without considering node correspondence,[12] using the portrait network divergence [1], a method based on a *graph invariant*. Using an invariant helps mitigate concerns regarding the encoding or structural representation of the graphs, enabling the measure to focus exclusively on the network's topology. Graph invariants can take various forms, including probability distributions, thus providing, size-independent, structural characteristics of networks, allowing meaningful comparison across different datasets.

Network portraits describe the distribution of path distances in a network. In the case of SRNs, they represent the different levels of connectivity, or maps of proximities between sectors, showing how closely each sector is linked to others. This information helps answer questions such as how many sectors each sector interacts with and how these (direct) connections are distributed across the entire network. They also show how well-connected a sector's direct neighbors are and how (indirect) connections extend through the network. These layers of extended connectivity characterize the structure of skills' exchanges of these networks. Thus, by focusing on the topology of the networks and abstracting from the problem of node correspondence, we can compare these networks without ensuring that networks use the exact same industrial classification encoding, which allows for a direct comparison of salient aspects of their structures without the need to align nodes. This approach enables us to analyze the similarities and differences in the network topology across different countries or contexts.

[12] That is to say, disregarding sector count *and* identification.

Portraits. The method stands on the construction of a $B_{\ell,k}$-matrix (v.g.: the network portrait, see [2]) consisting of:

$$B_{\ell,k} \equiv \text{ the number of nodes who have (exactly) } k \text{ nodes at distance } \ell,$$

for $0 \leq \ell \leq d$ and $0 \leq k \leq N-1$, where the distance is taken as the shortest path length and d is the graph's diameter (see Fig. 2).

In this sense, like onion layers, each node $v_i \in V$, where V is the set of all nodes of a given network, is surrounded by ℓ-shells or connectivity layers of order ℓ. The rows represent histograms (or distributions[13]) of ℓ-order shortest paths. This matrix condenses structural properties of the network based on the distance connecting two nodes in terms of successive links or path lengths, ℓ, which encode shortest path distributions, for example including the degree distribution ($\ell = 1$, for an unweighted network) and higher order paths. It is important to state that the network portraits are agnostic of the identity of the nodes, capturing topological information without reference to the nodes attributes. As a graph invariant, the B-matrix of a network is unique and can be used as a network "fingerprint". In this way, comparing two networks G and G' can be translated into comparing their portraits, B and B'.

In unweighted networks, the diameter d is an integer, whereas in weighted networks, $d \in \mathbb{R}$ is continuous. Since our analysis focuses on comparing the SRNs, where shortest paths may have non-integer values, the algorithm for finding shortest paths changes from breadth-first-search (used for unweighted networks) to Dijkstra's algorithm (used for weighted networks). To facilitate network comparison, it is necessary to define an appropriate binning strategy for aggregating continuous shortest paths. A simple approach is to use b bins based on quantiles, allowing us to compute the network portraits, $B_{\ell,k}$ and $B'_{\ell,k}$. In our case, we determine BinEdges based on the weight distributions of the SRNs under analysis.

In our analysis, we use both binary and weighted versions of SRNs. It is important to highlight what each network represents and how their interpretations differ. While SRNs are closely related to labor flow networks (LFNs), they are conceptually distinct, as they capture the latent skill similarity between sectors inferred from observed worker transitions. Unlike LFNs, which describe how workers move, SRNs infer why they move, based on skill compatibility. In the binary SRN, a link simply indicates whether two sectors are skill-related, without accounting for the strength of this relationship. This can be seen as an indicator function version of SRNs, where two sectors are connected only if they share a minimum level of skill similarity inferred from worker transitions (i.e.: $SR_{ij} > 0$, as in our case).

Thus, binary SRNs capture the existence of structural similarities in skill requirements across sectors. In contrast, weighted SRNs provide a ranking of skill similarity, reflecting how strongly any two sectors are related within the network. This allows to better capture variations in the intensity of worker transitions and

[13] When divided by N, the total number of sectors or nodes.

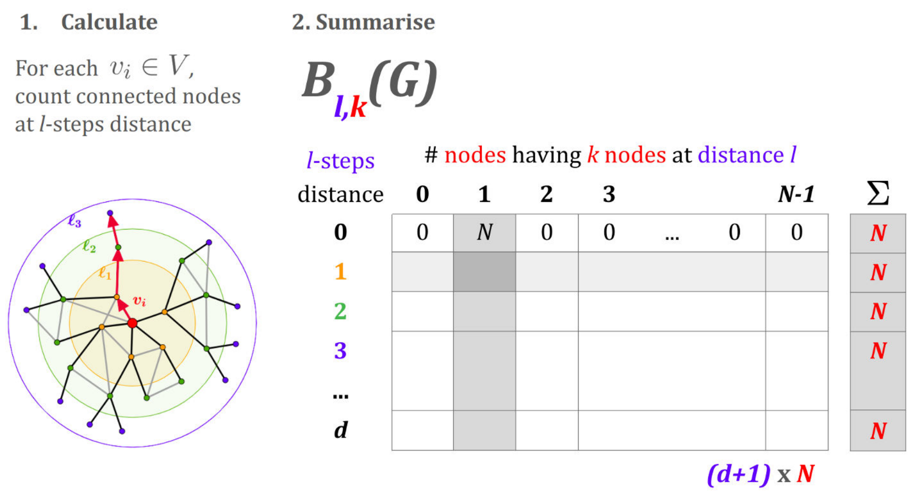

Fig. 2. Example of $B_{\ell,k}(G)$-matrix construction (for binary/unweighted SRN). For each node $v_i \in V$, count connected nodes at ℓ-steps distance, its ℓ-shell or connectivity layer, then summarize for each ℓ-distance (in rows) the number of nodes that have k-neighbors, taken as shortest path length (ℓ-shells). The first row $\ell = 0$ gives the number of nodes. The second row $\ell = 1$ stands for degree distribution: each sector's number of direct connections. The subsequent rows $\ell \geq 2$ distribution of ℓ-nearest neighbors. The last row $\ell = d$ gives the diameter of the network, v.g.: longest shortest path in the network.

the degree of skill overlap between industries, and also shed light into the value and usefulness of the SR indicator.

Network Portrait Divergence. After computing the portraits of these networks, denoted as G and G', each portrait can be transformed into a matrix of row-wise probability distributions by dividing each row by N. These matrices are then normalized and further reduced to two joint probability distributions encompassing all rows. Since these are probability distributions, they remain independent of the size of networks, enabling direct comparison using a single Kullback-Liebler (KL) divergence measure (see [1]).

The network portrait divergence (NPD) is defined then as the Jensen-Shannon divergence:

$$D_{JS}(G, G') \equiv \frac{1}{2} KL(P||M) + \frac{1}{2} KL(Q||M), \in [0, 1]$$

where $M \equiv \frac{1}{2}(P||Q)$ is the mixture distribution of P and Q, where P is $P(k, \ell) = \frac{kB_{\ell,k}}{N^2}$ and Q is, likewise, $Q(k, \ell) = \frac{kB'_{\ell,k}}{N^2}$.

3 Results

The SRNs for each country, built from the positive skill-relatedness indicator matrices and ancillary employment data, present a visible dense structure with a unique giant component (induced by the construction of an SRN), as can be appreciated in Fig. 3. The figure illustrates the skill-relatedness networks for the three countries, where we plot the heatmap representation for both, unweighted, v.g.: binary (Fig. 3, upper row), and weighted version networks (Fig. 3, lower row). As reported in Table 1, although the statistical systems of classification for economic sectors are compatible between countries in terms of international standards methodologies (by construction and via correspondence tables), the composition and size of these networks may vary because of: a) differences in some sectors' specification as informed by each country, i.e.: new sectors created by national statistical choices, additional disaggregation of sectors used by one country to identify some particular kind of activity in a code (available and) not used by any other country; and, b) as a result of the filtering process described in section 2, of significantly observed flows in terms of the skill-relatedness criteria, i.e.: exclusions induced by construction of national SRNs, could be the case that some sectors appear in one country while in other remain disconnected (and left out of the analysis) because of its exclusive participation in nonsignificant flows, in terms of the SR filtering criteria for SRN construction ($SR_{ij} > 0$).

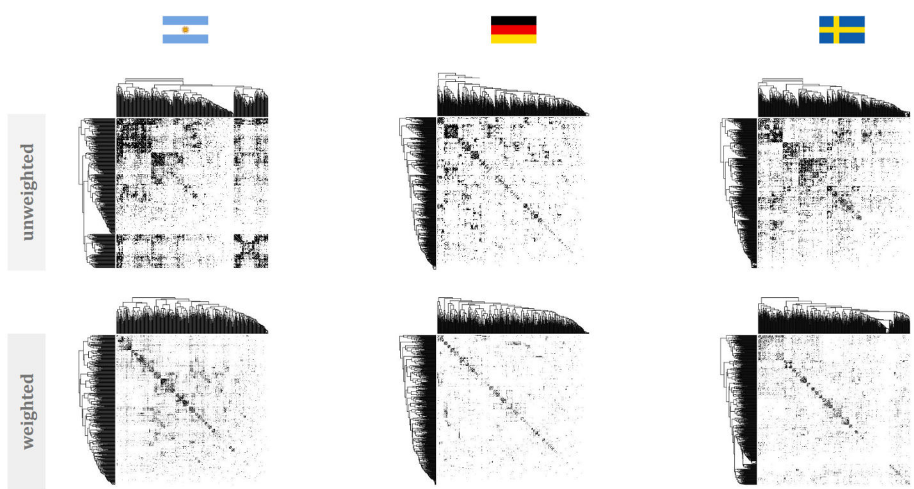

Fig. 3. Skill-Relatedness Networks (SRNs). Visualizations of SRNs for Argentina (ARG), Germany (DEU), and Sweden (SWE) for periods and size according to the specifications in Table 1. Heatmap representation of undirected networks: Unweighted (binary, upper row) and weighted (lower row) SRNs. Sorting is done with a hierarchical clustering algorithm with complete linkage.

After building each country SRN, we computed their respective portraits for weighted, $Bw^c_{\ell,k}$, as well as unweighted, $B^c_{\ell,k}$, versions of the SRNs with $c \in \{ARG, DEU, SWE\}$, plotted in Fig. 4. We use their unweighted versions to naturally introduce a way to better comprehend the information contained therein in terms of node connectivity.

In a network portrait, ℓ refers to the length of shortest paths and k counts the "number of nodes" having paths of length ℓ, that is to say considering ℓ-shells of each node in the network (see methods in section 2 and Fig.2). In an unweighted network $\ell = 1$ is the degree distribution, $\ell = m$ is the distribution of shortest paths of order m, and $\ell = d$ is the max length representing the network diameter. In a weighted network, ℓ has to be discretised as it is continuous.

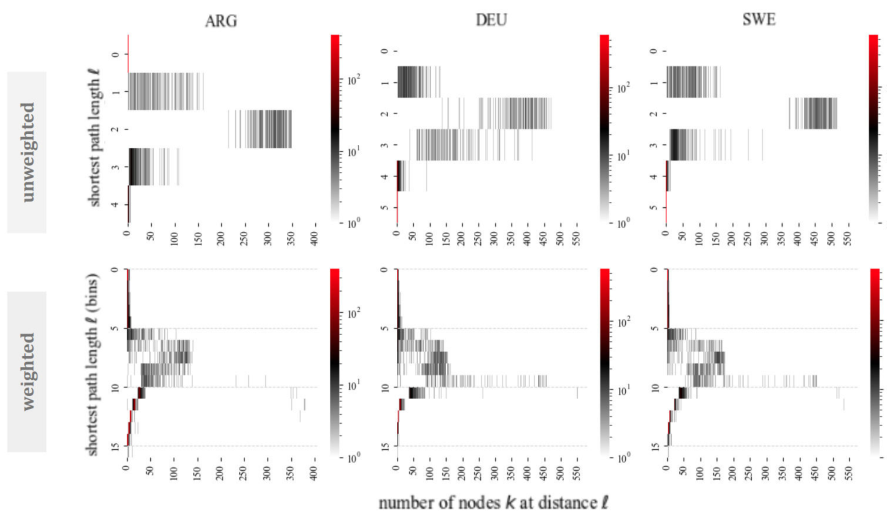

Fig. 4. Network portraits. Upper row: Unweighted (binary) SRNs. Discrete shortest path length ℓ from 0 to d, the diameter of the network with $d_{ARG} = 4$, and $d_{DEU} = d_{SWE} = 5$. Lower row: Weighted SRNs. Continuous (binned) shortest path length ℓ from 0 to d, the diameter of the network with $d_{ARG} = 1.95$, $d_{DEU} = 1.43$, and $d_{SWE} = 1.96$. For a better visualization we used 16 bins (vertical axis), with smaller bins destined to lower values of $SR > 0$ (BinEdges $\in (0, 0.002, 0.004, 0.006, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.5, 2)$).

For the *unweighted portraits* (Fig. 4, upper row), depicting the fingerprints of the pure connectivity in the SRNs, show the distribution of shortest paths for each country's network. These portraits present a kind of P-shape plot related to the big connected component topology that is characteristic of SRNs, as mentioned earlier. Their range goes from $\ell = 0$ (representing the total count number of nodes, N), occurring $B^c_{\ell=0,k} = N_c$ for each country network, to $\ell = d$, the corresponding (unweighted) diameter of each network (v.g.: $d_{ARG} = 4$, and

$d_{DEU} = d_{SWE} = 5$). Intuitively, the visualizations of this portraits show a condensed image of the way nodes, economic sectors in SRNs, are connected and proximate to each other albeit not identifying the specific connection between any pair of sectors m and j.

The second row ($B^c_{\ell=1,k}$) corresponds naturally to the standard degree distribution of direct connections. It can be appreciated that this distribution is relatively more widespread for ARG than for SWE and DEU. In particular, DEU accumulated relatively more (less connected) nodes in small values of k, that is to say more sectors with direct connections of low degree (or simply less intersectoral connections).

The next row, $B^c_{\ell=2,k}$, show the distribution of "two steps" paths or the most proximate indirect neighborhood shell for each node (ℓ-shell=2), that is to say: industries connected (through SR-links) with the industries in their direct connections circle. It can be appreciated that all countries show distributions centered in higher values of k, corresponding to the majority of nodes (industries) having a great number of nodes (industries) at this distance. In this case, DEU has a relatively more widespread distribution, while ARG and SWE appear more alike with higher density in high values of k. This means that most sectors show many "two steps" connections, a fact consistent with the analysis of labor flow networks for Argentina evidencing dense networks with short average paths and diameter, and having small world properties (v.g.: typical diameter of three steps, see [7,8]).

The following row, $B^c_{\ell=3,k}$, show the distribution of "three steps" paths length, an enhanced indirect neighbors set. It can be appreciated that the distributions are again skewed towards lower values of k, meaning that as the length of shortest paths approaches the diameter (shortest paths maximum length) there are less nodes (sectors) having many nodes at this distance. In this case, ARG and SWE appear more similar with a greater concentration of nodes (sectors) having a small k number of nodes at a three step distance, while DEU has more dispersed distribution with higher values of k nodes at three steps distance. This indicates that DEU has longer chains of connectivity, in line with a lower concentrated degree distribution of direct connections, as described earlier. Additionally, it suggests that DEU may have more sectoral skills specialization than ARG and SWE.

The last rows of these unweighted portraits, referring to the more distant layers of connectivity near or at their (respective) diameters, show high concentration of these longer paths in lower values of k. This refers to the paths linking nodes with sectors in the outer periphery having very poor connectivity.

For the *weighted portraits* (Fig. 4, lower row), depicting the valued fingerprints of the SRNs, show the distribution of shortest paths in terms of SR for each country's network. Their range goes from $\ell = 0$ to $\ell = d$, in this case corresponding to the continuous diameter of each network (v.g.: $d_{ARG} = 1.95$, $d_{DEU} = 1.43$, and $d_{SWE} = 1.96$). To compare this portraits showing the distributions of weighted shortest paths, we computed the same number of bins for the three SRNs so the interpretation can equally be made for all values of (binned) ℓ. As can be appreciated, the interpretation of weighted path lengths and the comparison between them is more demanding although differences and similarities can be appreciated between the fingerprints. The chosen binning, with BinEdges $\in (0, 0.002, 0.004, 0.006, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.5, 2)$, highlights lower $SR+$ weights in line with their decreasing prevalence in SRNs (see weight distributions in Fig. 5) across the (maximum) range, $r \in (0, max(d_c))$, of observed weighted paths for all countries.

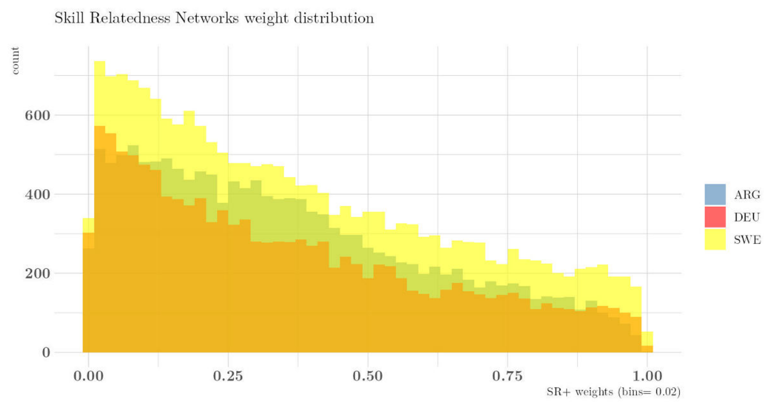

Fig. 5. SRN's $SR > 0$ weight distributions.

With this weight aggregation, the weighted portraits in Fig. 4 (lower row) can be divided into three "charge zones" in relation with the quantification of sector connectivity referenced in the horizontal axis and the weighted paths measured in the vertical axis:

a *high-concentration, low-weighted shortest* ℓ-paths in **bins 1 to 5**, corresponding to a total weighted distance of $\ell \in (0.00, 0.01)$ and involving the interconnection of just a few sectors;
b *high-dispersion, medium-weighted* shortest ℓ-paths in **bins 6 to 10**, corresponding to a total weighted distance of $\ell \in [0.01, 0.10)$, involving a sharply increasing interconnected (horizontal dispersion) and decreasing concentration (low intensity, showed in black and white gradient colors) sectors topology; and
c *high-concentration, high-weighted* shortest ℓ-paths in **bins 11 to 16**, corresponding to a total weighted distance of $\ell \in [0.10, 2.00]$ and involving the decreasing interconnection of most sectors with sectors in the "periphery".

To quantify these dissimilarities we calculate the (pairwise) network portrait divergence, $D_{JS}(G, G') \in [0, 1]$, with higher values showing more dissimilarity, presented in Fig. 6 for both unweighted and weighted portraits. The comparison for the *unweighted portraits* present stark differences between ARG and those

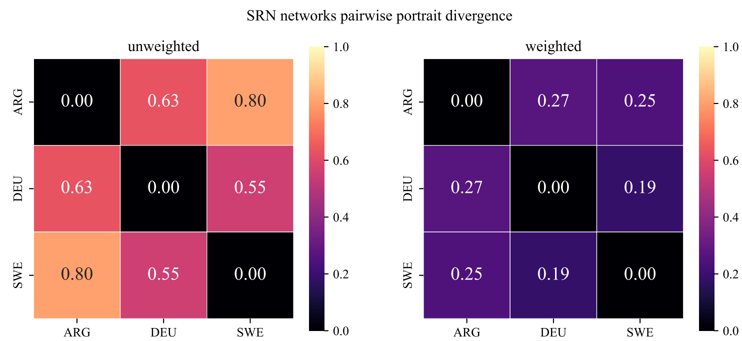

Fig. 6. SRNs network pairwise portrait divergence. Left: unweighted portraits divergence. Color range for $D_{JS}(G, G') \in [0, 1]$, greater values showing more dissimilar network portraits.

of DEU (0.63) and SWE (0.80), while at the same time it is also informative of the differences between DEU and SWE structure (0.55). In light of this results, it is useful to revisit the original binary structure of the SRNs in the upper row of Fig. 3. Taking the case of ARG, it is quite clear that its (clustering ordered) connectivity structure differs strikingly with both DEU and SWE. In particular, in ARG there is a group of approximately 30% of total sectors (bottom right) with high interconnection within them and some non-trivial interconnection with the rest of the sectors. In turn the rest of the sectors are grouped and ordered in decreasing order of total connectivity, showing some subgroups with more connectivity within. In the case of DEU, the connectivity structure is smoothly decreasing and characterized by a small modular structure, with some small sector groupings with high connectivity within. The case of SWE appears as an intermediate between the others, also with smoothly decreasing modular connectivity structure but with subgroups bigger than in the case of DEU.

Regarding the comparison of *weighted portraits*, the differences of ARG's SRN and their counterparts in DEU (0.27) and SWE (0.25) appear less pronounced, and the comparison between DEU and SWE (0.19) show the lowest divergence. Again, it is useful to revisit the original weighted SRNs in the lower row of Fig. 3. This time the visible connectivity structure is more difficult to disentangle because of the weak density in all cases. In particular, SWE presents more modular structure detectable with the hierarchical clustering at the corners up-left (higher intersectoral connectivity within and also between this group and the immediate neighbors down/right, more central), and down-right (smaller group, less connected with the rest of the network, more periphery like).

At this point, it is worth understanding the differences in the results obtained between the weighted and binary SRNs. By definition, when SR is high (low), the intensity of skills exchanges enhances (reduces) the relevance of connections between sectors. The SR weights, by shrinking distances between sectors with lower SR, show similarities between countries that are not visible when looking only at the bare connectivity of SRNs. This may be related to the fact that SR weight distributions across countries share a similar profile, although with some differences in scale (see Fig. 5). Since SR weights are derived from the margins of the original labor flows tables, the *proportionality*-bias inherent in the random null model could be indicating the existence of a more universal-like phenomenon, showing a smooth structure of significative skill exchanges between sectors in any national economy. In contrast, the stark differences observed in the binary SRNs suggest that the *particular interactions* between sectors (v.g.: skills exchanges) selected by the SR criteria vary significantly between countries. This suggests that each country's SRN connectivity structure follows a distinct pattern of connectivity of (possibly) groups of sectors and of the network as a whole.

4 Discussion

In this paper, we compared skill-relatedness networks (SRNs) across different countries using data from Argentina (ARG), Germany (DEU), and Sweden (SWE) to explore potential differences between SRNs in a developing economy *vis-a-vis* those in developed economies. To achieve this, we applied a method designed for comparing networks of different sizes (non-aligned networks) that focuses on topological information [1]. Specifically, we used network portraits [2], a condensed representation of shortest path length distributions that serves as a unique structural fingerprint of each network. Through this analysis, we found that both the portrait representation of networks and the network portrait divergence measure are appropriate and effective methods to characterize and compare SRNs.

Our analysis of the *unweighted* network portraits of these SRNs reveal contrasting differences in the pure connectivity (binary) structure of SRNs across countries. In particular, Argentina's SRN exhibits significant differences in inter-industry connectivity compared to those of developed countries like Germany and Sweden. Sweden's SRN shows a markedly distinct connectivity pattern, while Germany's network presents a high contrast with both Argentina and Sweden. This comparison underscores the variability in skill-related connectivity even among developed economies. This suggest that the specific structure of interconnections underlying the SRN for each country are quite distinct when considered as a binary structure, like a skeleton of uniform bridges. Additionally, in the case of DEU which showed a (degree) distribution of direct intersectoral connections concentrated in lower connectivity levels, may be indicating of a structure more specialized regarding skills exchanges.

When comparing the *weighted* skill-relatedness networks, we found less pronounced differences across the three SRNs. In particular, Argentina's SRN remains notably distinct from those of both Germany and Sweden, whereas Germany's SRN closely resembles Sweden's, suggesting greater similarity in skill-relatedness structures between these two developed economies. This pattern goes in line with a more conventional way of classifying the countries, for example through the glass of the developed/underdeveloped distinction or GDP per capita ranking. Furthermore, the similarity in SR weight distributions across countries suggests the presence of a universal phenomenon, showing a smooth structure of significative skill exchanges between sectors in any national economy. These findings give relative support to the hypothesis of similarity of different countries SRNs conditioned on historical and cultural differences (see [10]). On the other hand, they show that the connectivity (topological) structure of different observed SRNs present stark differences between countries.

We identify several potential extensions to this work. From a sectoral perspective, a deeper exploration of the most significant connections in each economy could help identify key sectors and inquire about the specific pool of skills exchange they facilitate. From a systemic viewpoint, analyzing the meso-structure of SRNs could reveal relevant linkages relating different core- and periphery-like sectoral groups and their interactions.

References

1. Bagrow, J.P., Bollt, E.M.: An information-theoretic, all-scales approach to comparing networks. Applied Network Science **4**(1), 1–15 (2019)
2. Bagrow, J.P., Bollt, E.M., Skufca, J.D., Ben-Avraham, D.: Portraits of complex networks. Europhysics letters **81**(6), 68004 (2008)
3. Barabási, A.L., Pósfai, M.: Network Science. Cambridge University Press (2016), https://books.google.com.ar/books?id=ZVHesgEACAAJ
4. Brun-Schammé, A., Rey, M.: A new approach to skills mismatch. OECD Productivity Working Papers **7**(24) (2021). https://doi.org/10.1787/e9563c2a-en
5. Conte, D., Foggia, P., Sansone, C., Vento, M.: Thirty years of graph matching in pattern recognition. Int. J. Pattern Recogn. **18**, 265–298 (2004)
6. De Raco, S., Semeshenko, V.: Identificación de diferencias y similitudes estructurales de las redes interindustriales del empleo de argentina. Memorias de las JAIIO **9**(1) (2023), https://ojs.sadio.org.ar/index.php/JAIIO/article/view/616
7. De Raco, S.A., Semeshenko, V.: Labor mobility and industrial space in Argentina. Journal of Dynamics & Games **6**(2), 107 (2019). https://doi.org/10.3934/jdg.2019008
8. De Raco, S.A., Semeshenko, V.: The network structure of inter-industry labor mobility in Argentina. In: 6th Regulating for Decent Work Conference. ILO, Geneva (2019)
9. Mukoyama, T.: The cyclicality of job-to-job transitions and its implications for aggregate productivity. Journal of Economic Dynamics and Control **39**, 1–17 (2014). https://doi.org/10.1214/18-AOAS1176
10. Neffke, F., Henning, M.: Skill relatedness and firm diversification. Strategic Management Journal **34**(3), 297–316 (2013)
11. Neffke, F., Otto, A., Weyh, A., et al.: Skill-relatedness matrices for Germany: Data method and access. Tech. rep., Institut für Arbeitsmarkt-und Berufsforschung (IAB) (2017)
12. Neffke, F.M., Otto, A., Weyh, A.: Inter-industry labor flows. Journal of Economic Behavior & Organization **142**, 275–292 (2017)

13. Rathelot, R., van Rens, T., Chan, S.: Rethinking the skills gap. IZA World of Labor **391** (2023). https://doi.org/10.15185/izawol.391.v2
14. SAGPA: Skill relatedness matrices for Sweden. Tech. rep., Swedish Agency for Growth Policy Analysis (2021), https://www.tillvaxtanalys.se/in-english/publications/pm/pm/2021-05-18-skill-relatedness-matrices-for-sweden.html
15. Schieber, T.A., Carpi, L., Díaz-Guilera, A., Pardalos, P.M., Masoller, C., Ravett, M.G.: Quantification of network structural dissimilarities. Nat Commun **8**(13928) (2017). https://doi.org/10.1038/ncomms13928
16. Semeshenko, V., De Raco, S.A.: Analysis of the evolution of labor market flows in Argentina. Proceedings 50 JAIIO-AGRANDA pp. 20–24 (2021)
17. Straulino, D., Landman, M., O'Clery, N.: A bi-directional approach to comparing the modular structure of networks. EPJ Data Science **10**(1), 13 (2021)