IDB WORKING PAPER SERIES Nº IDB-WP-01201

# Unraveling the Network of the Extractive Industries

Lenin H. Balza
Camilo De Los Rios
Alfredo Guerra
Luis Herrera-Prada
Osmel Manzano

Inter-American Development Bank
Infrastructure and Energy Sector

April 2021

# Unraveling the Network of the Extractive Industries

Lenin H. Balza
Camilo De Los Rios
Alfredo Guerra
Luis Herrera-Prada
Osmel Manzano

Inter-American Development Bank
Infrastructure and Energy Sector

April 2021

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Unraveling the network of the extractive industries / Lenin H. Balza, Camilo De Los Ríos, Alfredo Guerra, Luis Herrera-Prada, Osmel Manzano.
p. cm. — (IDB Working Paper Series ; 1201)

Includes bibliographic references.
1. Mineral industries-Colombia-Econometric models. 2. Labor mobility-Colombia-Econometric models. 3. Manufacturing industries-Colombia-Econometric models. 4. Agricultural industries-Colombia-Econometric models. I. Balza, Lenin. II. De Los Ríos, Camilo. III. Guerra, Alfredo. IV. Herrera-Prada, Luis. V. Manzano, Osmel, 1971- VI. Inter-American Development Bank. Infrastructure and Energy Sector. VII. Series.

IDB-WP-1201

http://www.iadb.org

# Unraveling the Network of the Extractive Industries*

Lenin H. Balza[†] Camilo De Los Rios[‡] Alfredo Guerra[§] Luis Herrera-Prada[¶] Osmel Manzano[‖]

April, 2021

**Abstract**

This paper analyzes extractive industries in Colombia and their connections to other economic activities in the country. We use detailed social security data on all formal employees to create an industry-relatedness measure using labor flows between industries. Drawing on the vast network analysis literature, we exploit centrality measures to reveal the importance of the extractive sector among Colombian industries. Our results show that extractive industries are well connected within the Colombian industrial network, and that they are central overall and within their clusters. We also find that extractive industries have stronger linkages with manufacturing and agriculture than with other sectors. Finally, a higher relatedness to extractive activities is correlated with lower levels of employment, specially of female workers.

***Keywords:*** Network Analysis; Extractive Industries; Labor Flows, Latin America.

***JEL classification:*** C65, J24, O54, Q30.

*The views expressed herein are entirely those of the authors and do not necessarily represent the views of the Inter-American Development Bank, the International Monetary Fund, their Board of Directors, or the countries they represent. We thank Raul Jimenez Mori, Stephanie Majerowicz and two anonymous reviewers for their valuable comments and suggestions on an earlier version of this paper. We also express gratitude to the Ministry of Health in Colombia for providing us with the necessary data. All remaining errors are our own responsibility.

[†]Inter-American Development Bank. Infrastructure and Energy Sector. ✉: leninb@iadb.org

[‡]Inter-American Development Bank. Infrastructure and Energy Sector. ✉: camilodel@iadb.org

[§]New York University. ✉: alfredo.guerra@nyu.edu

[¶]International Monetary Fund. Western Hemisphere Department. ✉: luis.o.herrera@me.com

[‖]Inter-American Development Bank. Country Department Andean Group. ✉: osmelm@iadb.org

# 1 Introduction

The exploitation of natural resources and its effect on a country's economic development has been extensively studied. Yet, while the literature on the effects of natural resource abundance on development at the country level has matured significantly, much less is known about the local impact of resource exploitation. Exploring the dynamics of extractive industries and their role within a country's broader ecosystem of economic activities can help to understand the role that resource exploitation plays in economic growth. Moreover, the interplay between extractive industries and other economic activities can determine the extend to which changes in the general structure of the commodities market affect the internal economy. Such interlinkages across different economic activities can also create knowledge spillovers that stimulate innovation (Nooteboom, 2000).

In this paper, we examine how the extractive sector is connected with other economic activities in Colombia, a developing country rich in oil, gas, coal, and other natural resources. We explore the relevance of the extractive sector in Colombia beyond its importance in terms of production and revenue. We also identify the industries that are most closely related to the extractive sector and characterize the latter's clusters of influence.

To understand the relatedness of industries, we use a labor flow-based measure following Neffke and Henning (2013). This measure captures the relatedness between industries in terms of the skills needed. We employ administrative data on all formal employees in Colombia between 2008 and 2013, exploiting the vast network analysis literature to examine the industrial space in Colombia and using centrality measures to understand the role of extractive industries in the country.[1] In addition, we decompose the network into clusters using community detection algorithms. Intuitively, industries within a cluster have stronger and denser connections between themselves than with other industries in the industry space. This provides a comprehensive view of the role of extractive industries and the clusters in which they operate. This paper therefore deals with fields in economics that have, until

[1] Throughout the paper, we adopt terminology from the economic complexity literature. In particular, "industry space" refers to the network of related industries in Colombia. The term is derived from "product space", a term coined by Hausmann and Klinger (2006) to refer to a network that is created by using the similarity between export products.

recently, been analyzed separately by combining rigorous network analysis and insights from the natural resources literature.

Our results suggest that extractive industries play a central role in the Colombian industry space, which appears to be correspondingly more pronounced within their clusters. We also find that a higher relatedness to extractive activities is associated with lower levels of employment, driven in large part by gender disparities. These findings are consistent across both extractive industries and their closely related industries in Colombia.

By analyzing the economic activities that are connected to Colombia's extractive industries and the extent of these connections, this paper joins a recent strand of literature that leverages resource-based indicators to study inter-industry relatedness (*e.g.*, Neffke and Henning, 2013). Our data set captures all of the labor flows in Colombia's formal sector, a level of detail rarely seen in the literature. Several studies have also used labor flows to capture industry relatedness (*e.g.*, Neffke and Henning, 2013; Neffke et al., 2017) and to study the natural resource curse at the local level (Fitjar and Timmermans, 2019). We extend this labor flow approach to extractive industries in the case of Colombia and further leverage it to analyze industry clusters. To the best of our knowledge, this is the first paper to use this type of analysis with a focus on extractive industries in a resource-dependent and developing country.

This paper is also tangentially related to the literature on the natural resource curse at the subnational level. As one of the main interest of that literature is delving into the mechanisms of transmission of the curse, understanding the relationship of the extractive sector with other economic activities can be crucial. While the effects of natural resource abundance on development at the country level are relatively well understood, the same cannot be said about the local impact of resource abundance (Aragón et al., 2015). Prior studies in the subnational resource curse literature focus on the impacts of natural resource exploitation on fiscal windfalls (Caselli and Michaels, 2013), local demand effects (Aragón and Rud, 2013), human capital accumulation (Bonilla, 2020; Santos, 2018), and specialization in extractive activities (Michaels, 2010), among others. This paper contributes to this line of scholarly work by providing an overview of the relationship between different economic activities and a new account of the interlinkages among Colombia's industries with a focus on extractive activities. This can help to identify economic sectors that might be negatively impacted by extractive industries. We expect future research to be able to extend the preliminary findings of this paper on the network effects of industry linkages around extractive industries in developing countries.

The natural resource exploitation literature has also highlighted its heterogeneous gender effects. Indeed, important gender disparities exist in the extractive sector around the world (Eftimie et al., 2009). Some scholars argue that natural resource exploitation can explain gender gaps in political representation (Ross, 2008), and that it is men who generally benefit the most from extractive industries (Reeson et al., 2012). We build on this line of research by providing some evidence for the existence of a gender gap not only in the extractive sector, but also in its related industries.

The rest of the paper is organized as follows. Section 2 gives an overview of the Colombian context and the data used for this paper. Section 3 provides a general explanation of network analysis, particularly aimed at those who are less familiar with this literature, followed by a discussion of our method for leveraging the data to construct our skill-relatedness measure. Section 4 presents the results of the study, and section 5 discusses the results and summarizes the paper's main conclusions.

## 2 Colombian Context and Data

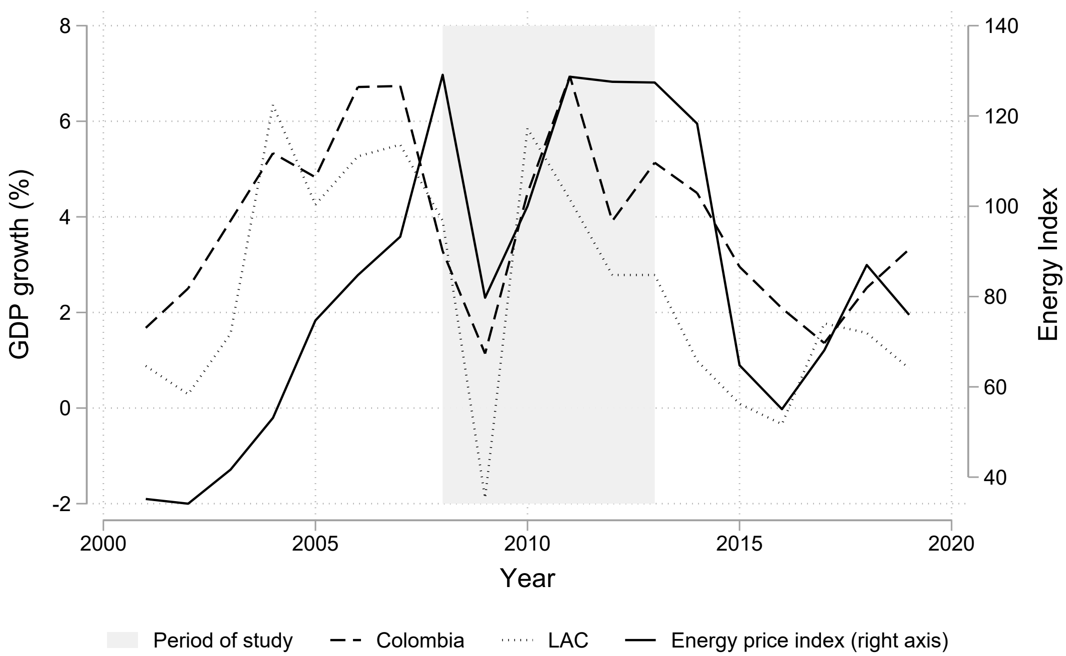
Natural resource exploitation plays an important role in the economies of many countries in Latin America and the Caribbean (LAC), where it has been a key pillar of regional economies for many decades. Oil, gas, and mining rents alone accounted for 5% of LAC's GDP and nearly 27% of the region's total exports in the last decade, figures that are even higher in Colombia. In the same period, oil, gas, and mining represented nearly 60% of the country's total exports, while rents accounted for nearly 6% of its GDP.[2]

Notably, the LAC region is quite dependent on fossil fuel exploitation. Figure 1 shows a clear correlation between the international price of fossil fuels and the region's economic growth. This holds true for Colombia, a large coal and oil producer. Similar patterns can

[2]These are the authors' calculations based on data from the World Bank and the Atlas of Economic Complexity. The results are much the same when using official data from Colombia's National Statistical Office (DANE; acronym in Spanish).

be observed for metals and minerals, and somewhat for precious metals (see Figure A1 and Figure A2 in Appendix A). Our study period comprises both a boom and a bust in commodity prices, which plummeted during the international financial crisis of 2008, only to recover after 2010.

Figure 1: Dependence on fossil fuels performance




Energy price index is a weighted measure of Oil, Gas, and Coal prices. The base year for the index is 2010
Source: Authors' calculations based on the World Bank's pink sheet data and development indicators.

## 2.1 Data

We obtained Colombian social security data from the Integrated Form of Contribution Payments (PILA; acronym in Spanish) of the Ministry of Health for the period 2008 - 2013. The PILA gathers information on the payments made to the Social Security system in Colombia and tracks individuals over time on a monthly basis. The dataset is anonymized at the individual level and reports the primary economic activity of the firm in which the individual is employed, following the ISIC classification revised for Colombia.[3] The ISIC codes are four-digit numbers that allow for the most detailed analysis of firms' economic sectors. These will allow us to study aspects associated with labor mobility and the concentration of productive activity in Colombia. In the remainder of this paper, when writing about an industry we refer to all firms with the same four-digit ISIC code. For example, every firm with ISIC code *"1110"* is part of the oil and gas industry. Our sample follows more than 10 million workers in a total of 969 industries.

One of the objectives of this paper is to understand the role of extractive industries in the Colombian industry space. To do so, we must first establish our definition of extractive industries. Although the specific classification might change somewhat between countries based on their respective industrial classification systems, extractive activities are generally understood to comprise the economic exploitation of minerals, oil, and gas. In the case of Colombia, the Colombia's National Statistical Office (DANE; acronym in Spanish) defines the mining-extractive sector as covering the exploitation of coal, oil, natural gas, mineral and metals, supporting activities, and other related activities. Importantly, however, linkages between industries can generate direct effects on its neighbors as well as have "spillover effects" on second- and higher-degree neighbors.[4] To do so we define two categories of extractive activities. The "Core Extractive" industries are those that belong to the extractive sector according to the DANE definition, which include oil exploration and exploitation, coal extraction, or metallic ore mining. "Periphery Extractive" industries are complementary to the core extractives and represent closely related industries in terms of production. Such is the case of fuel sellers and manufacturers of metallurgical machinery. Thus, "Periphery Extractive" activities are not strictly considered extractive activities by DANE, but are intuitively related to them. Table 1 shows the total number of workers that had at lest one job during the year in each of these industry groups and Appendix Table A1 shows the full set of industries in both groups.

[3]The ISIC code is an international reference classification of productive activities. In 2012 and 2013 the PILA used an updated version of the ISIC codes, up to 90% of which were adjusted to ISIC Revision 3 by the Colombian Ministry of Health before the data were given to us. Unfortunately, there is no way for us to know whether or not an observation uses the corrected Revision 3 ISIC code. The results in this paper must therefore be taken with caution.

[4]This refers to industries that are not directly connected, but rather indirectly through one or more industries – just like the spread of gossip in a group of people.

Table 1: Workers in sample

| | 2008 | 2010 | 2011 | 2013 |
|---|---|---|---|---|
| *Total Workers* | 8,745,815 | 9,866,664 | 10,470,082 | 11,554,732 |
| *Workers that had at least one job in* | | | | |
| Core extractives | 152,704 | 169,074 | 180,704 | 169,584 |
| Perpihery extractives | 25,441 | 26,042 | 26,632 | 18,760 |
| *Workers that earn more than* | | | | |
| *the Median Wage* | 4,347,594 | 5,130,194 | 5,446,655 | 6,060,890 |
| *the Minimum Wage (MW)* | 6,342,219 | 6,605,932 | 6,938,689 | 8,426,589 |
| *3 times the MW* | 1,196,011 | 1,350,440 | 1,481,888 | 2,957,804 |
| *10 times the MW* | 196,796 | 205,055 | 231,737 | 552,338 |

Note: Authors' calculations based on data from PILA. Table shows the total number of workers that we are able to follow in this study.

We are particularly interested in capturing all of the industries in which an employee has worked during the years for which we have data. Thus, we aggregate the individual's information from the PILA at the year-industry level. With these data we can measure the observed movement of workers between different industries over time, which we define as a "labor flow." The granularity of our data allows us to capture all labor flows in the Colombian formal sector at the four-digit code (*i.e.* industry level) within our study period. This information can be used to create an industry network based solely on labor flows that can be explored using network analysis. In addition, we use the wage information provided in the PILA, which allows us to conduct our analysis by subsamples that differentiate skill levels between workers. In order to observe whether any temporal dynamics arise, we also disaggregate our analysis into three two-year subperiods: 2008 - 2010, 2010 - 2011, and 2011 - 2013.

One limitation of the PILA is that it only provides information on formal workers in Colombia. According to the Colombian mining census of 2010, more than 80% of metallic ores are extracted from mines without a formal title. On the other hand, around 60% of coal extraction took place in mines with formal titles, while informal operations are nearly non-existent in the capital-intensive oil and gas exploitation sector. This implies that, on the whole, our use of the PILA may result in us underestimating certain connections with the extractive sector.

Finally, we also use data from the Colombian Annual Manufacturing Survey (EAM; acronym in Spanish). This survey covers the manufacturing establishments of firms with an annual production of more than US$100,000 or firms with at least one plant with ten or more employees. Like the PILA, the survey also reports the ISIC code for every firm at the four-digit level. In this study we are interested in industry-level measures of economic activity captured in the survey. Thus, we compile the survey for the years 2008-2013 and use measures that capture the general structure of the labor force and economic performance at the industry level.

# 3 Methodology

This section begins by providing a general explanation of network analysis. The readers who are well versed on the topic may skip the first subsection and continue to our explanation of how the skill relatedness metric used in the paper is constructed.

## 3.1 Network Analysis

Originally used primarily in mathematics and computer science studies, network theory has more recently been employed in a wide range of applications, including social interactions (Jackson, 2010), road development (Masucci et al., 2014), urbanization (Fafchamps et al., 2017), and biology (Barabasi and Oltvai, 2004). In general, network analysis is concerned with detecting, measuring, analyzing, and predicting the behavior of specific network structures. This type of analysis is particularly useful in economics, where the subjects of interest are often part of complex, interconnected systems. These systems may include trade networks between countries, or the social interactions through which information is spread. In fact, it is difficult to think of activities that operate in isolation from the broader ecosystem of economic activities around them. Industries are no exception in that they may operate under a general equilibrium setting at the local level (Enrico, 2011; Greenstone et al., 2010).

In general, every network is comprised of two elements: nodes and edges (or links). Nodes are the unit of analysis (*e.g.*, countries or individuals) and edges are the connections between them (*e.g.*, exports or social interactions). Naturally, the connections can be stronger or weaker depending on how they are measured (*e.g.*, total value of exports between countries; number of years people have known each other). This measure is usually called the edge weight. Characterizing and identifying the most important nodes in the network is one of the most common and intuitive exercises in network analysis. Measures of centrality are very useful in such exercises since they capture basic information about the underlying structure of the network.[5]

One of the most basic, yet powerful, measures is degree centrality. This measure captures the number of connections a node has in the network and is particularly useful when trying to identify the best connected or most "popular" nodes in a network (*e.g.* the person with most friends). The degree centrality of a node $i$ is defined in Equation 1, where $a(i, j)$ is equal to 1 if nodes $i$ and $j$ are connected and 0 otherwise, and $n$ is the total number of nodes in the network.[6] Note that if the edges between nodes are not binary but a continuum, a weighted version of $DC_i$ can be calculated by simply using the sum of those weights rather than the sum of the indicator variable in the numerator. For example, instead of counting the number of friends a person has, we could count the total number of years that person has known every other person in the network.

$$DC_i = \frac{\sum_{j=1,j\neq i}^{n} a(i,j)}{n-1} \tag{1}$$

Betweenness centrality and closeness centrality are two other commonly used measures in network analysis. These measures depend on the position of the nodes relative to others in the network. Betweenness centrality captures the number of times a node is on the shortest path between other nodes, while closeness centrality simply measures how close a

[5]The reader can find a detailed explanation of centrality measures in Freeman (1978).

[6]nwcommands can calculate centrality measures in Stata. Several other packages are available for other software programs.

node is to others in the network and thus captures how easily a node can reach the others. Following our example, a very popular person would know a lot of others in the network and be well connected (*i.e.*, closeness centrality). At the same time, many people would have to first go through that person to be able to contact another person in the network with whom they themselves have no connection (*i.e.*, betweenness centrality). Equation 2 defines betweenness centrality for node $i$, where $\sigma_{jk}$ is the total number of shortest paths between nodes $j$ and $k$, and $\sigma_{jk}(i)$ is the number of those that pass through node $i$. Further, Equation 3 defines closeness centrality for node $i$, where $d(i,j)$ is the length (*number of nodes*) of the shortest path from $i$ to $j$.

$$BC_i = \sum_{j \neq i \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \tag{2}$$

$$CC_i = \frac{n-1}{\sum_{j=1, j\neq i}^{n} d(i,j)} \tag{3}$$

### 3.2 Measuring Skill Relatedness

This paper closely follows the literature that tries to measure inter-industry relatedness. One of the approaches in the literature is to measure relatedness using resource-based indicators, such as patents (Breschi et al., 2003; EC Engelsman, 1991; Jaffe, 1989), commodities (Fan and Lang, 2000), or human capital, through occupational profiles (Chang and Harbir, 1999; Chang, 1996; Farjoun, 1994). Yet if we consider the Knowledge Based Theory (KBT) of the firm, which asserts that knowledge is a firm's most strategic and significant resource (Grant, 1996; Spender and Grant, 1996), the best resource-based indicator should capture employees' skills.

Modern economies depend on highly specialized workers, with average skill levels increasing markedly over the last century. In parallel, highly specific skills have become a requirement in many positions (Gathmann and Schönberg, 2010), firms (Becker, 1964), and industries (Neal, 1995; Parent, 2000; Sullivan, 2010). Contrary to common assumptions in economic models where quantifiable metrics, such as years of education, are the basis to measure human capital, we acknowledge workers' skills as a decisive factor in human capital. For a worker to be able to switch jobs, there must be an overlap in the skills required in both placements. In this paper, we focus on industry skill relatedness, and, following Neffke and Henning (2013) and Neffke et al. (2017), we use labor flows between industries as a clear indication of shared skills.

For the purpose of this paper, a labor flow exists between a pair of industries whenever there is an individual who has worked in both industries during a determined time period. Since we aim to capture the skill relatedness between industries, we also look at workers who hold a job in more than one industry at the same time, a clear indication of shared skill needs between the industries. We therefore consider labor flows as not only job switching but also simultaneous jobs. Thus, we exploit the temporal variation in the PILA to construct labor flows that allow us to capture skill relatedness between industries. We provide a detailed description of the labor flow measure, its assumptions, and its implications in Appendix B.

Our skill-relatedness measure considers the number of employees that switch from one industry to another within and between years. In particular, following Neffke and Henning (2013), the skill relatedness between a pair of industries is given by Equation 4. Such a measure has previously been used to predict firm diversification (Neffke and Henning, 2013) with greater effectiveness than measures of co-location or value chain relations (Neffke et al., 2017). $F_{ij}$ is the total number of labor flows between industries $i$ and $j$, and the denominator is a measure of the expected labor flows between $i$ and $j$ given the total flows that those industries have in the network. The higher $SR_{ij}$ is, the more connected industries $i$ and $j$ are.

$$SR_{ij} = \frac{F_{ij}}{\frac{\sum_i F_{ij} \sum_j F_{ij}}{\sum_i \sum_j F_{ij}}} \tag{4}$$

In this paper, we leverage the theory that human capital flows more freely between industries with similar skill needs. As discussed, a worker's placement in an industry is mainly the result of previously acquired skills, so switching to a job in another industry - or working in another industry concurrently - implies a certain overlap of skills between industries. Following the KBT, all of the firms that have employed a given worker necessarily share a particular resource: the worker's skills themselves. Thus, a worker's skills entail a connectedness between the industries that have leveraged those skills, which they require in order to operate. The worker is therefore the link between the industries and embodies the shared skills required by the industries in the industry space.

The granularity and detail of our data allow us to refine this measure by focusing on a subset based on the workers' skill levels. Since skills are not only industry-specific, but also position-specific (Gathmann and Schönberg, 2010), a pair of industries might be much more related to one another for a given subset of skills that are required for a particular kind of position. As skill levels cannot be directly captured in our data, we use wages as the best proxy available. The idea that skills determine a large part of a worker's wage can be traced back to Adam Smith. The theoretical work of Mincer (1962) and Becker (1964) further laid the foundations for much empirical research on the topic (*e.g.* Grogger and Eide, 1995; Murphy and Welch, 1989). We accordingly created different subsets of our skill-relatedness measures using all workers, those who earn more than the median wage, those who earn more than the minimum wage, those who earn more than three times the minimum wage, and those who earn more than ten times the minimum wage. The first group allows us to assess the general relatedness of industries, while the second and third groups act as natural robustness checks. The fourth and fifth groups assess the skill relatedness between industries for a medium and a high skill level, respectively (the number of workers in each group can be found in Table 1).

# 4 Results

We derive some stylized facts about the industry space in Colombia, its general structure, and the role of extractive industries in it. This allows us to characterize the country's extractive sector and the related industries using novel data.

## 4.1 Extractive Industries in the Network

The skill-relatedness measure is built using labor flows between industries and is indicative of how different industries are connected through their human capital requirements. One of the simplest and most straightforward ways to capture the importance of extractives industries in the industry space is by looking at their "degree centrality". As explained in

Section 3, degree centrality measures the extent to which a node is connected to all nodes in a network, based on the number of direct connections it has to other industries. Since skill-relatedness is a continuous measure, we calculate a weighted version of degree centrality, called strength degree centrality. Labor flows are a direct measure of skill relatedness, thus making the degree centrality measure a much more useful and intuitive one for our purposes than betweenness or closeness centrality.

Although a node's strength degree centrality provides us with an absolute measure of centrality, a relative measure would be much more useful for assessing the importance of extractives within the industry space. Using the ranking of the strength degree centrality in the network would allow us to better perceive the relative importance of the industries within the network. The industry that ranks first, would be the most connected industry in the space and the least connected industry would rank last, in the 969$^{\text{th}}$ position.

Table 2 shows the average strength degree centrality rank for core extractives, periphery extractives and all other types of industries for different types of workers and all subperiods. This shows that periphery extractive industries are more central than core extractive industries in the industry space for the whole period when using all the workers in the sample. However, core extractives improved their ranking much more than periphery industries did between subperiods 1 and 2, while the ranking of all other industries remained unchanged. This result might be related to commodity price dynamics during the period, as portrayed in Figure 1. As commodity prices increase, core extractive activities can attract more workers from previously unconnected industries. The increasing importance of the core extractive industries during that sub-period relative to non-extractive industries therefore seems natural. It is also evident from the results that core extractive industries generally have a better average ranking than non-extractive industries. This implies that core extractive connections are, on average, stronger in the network and more central in the industry space. This interpretation is not driven by the type of measure that we use (see strength degree centrality in Appendix Table A2), nor does it appear to be driven by outliers (see Appendix Table A3 and Table A4).

The centrality of an industry in this study can be affected by the subsample of workers used to construct the skill-relatedness measure. As discussed in Section 3, wages are our best proxy of the workers' skill level. We therefore present the mean strength degree centrality rank using different subsamples based on wage levels to construct the skill-relatedness measure. The results in Table 2 also show that the relative centrality of core extractive industries is higher when the sample is restricted to higher-wage workers. This means that core extractive industries are more connected to the industry space at large when only higher-skill workers are considered.

Table 2: Mean ranking of strength degree centrality of economic activities

| | *Subperiod* | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | *All* |
| *Full sample: All workers* | | | | |
| Core extractives | 430 | 382 | 474 | 446 |
| Periphery extractives | 300 | 296 | 406 | 338 |
| Non-extractives | 473 | 473 | 484 | 488 |
| *Subsample 1: Workers that earn more than the median wage* | | | | |
| Core extractives | 407 | 360 | 469 | 452 |
| Periphery extractives | 302 | 317 | 410 | 324 |
| Non-extractives | 473 | 472 | 484 | 488 |
| *Subsample 2: Workers that earn more than the minimum wage* | | | | |
| Core extractives | 437 | 404 | 458 | 475 |
| Periphery extractives | 311 | 350 | 464 | 377 |
| Non-extractives | 469 | 467 | 482 | 485 |
| *Subsample 3: Workers that earn more than 3 times the minimum wage* | | | | |
| Core extractives | 388 | 380 | 439 | 438 |
| Periphery extractives | 394 | 440 | 495 | 374 |
| Non-extractives | 462 | 461 | 475 | 477 |
| *Subsample 4: Workers that earn more than 10 times the minimum wage* | | | | |
| Core extractives | 371 | 374 | 411 | 410 |
| Periphery extractives | 510 | 516 | 590 | 555 |
| Non-extractives | 438 | 437 | 455 | 462 |

Note: Authors' calculations based on data from PILA. Mean of the ranking of the strength degree centrality by subperiods, wage group subsamples, and main industry groups of analysis. A higher ranking entails a lower strength degree centrality, thus a poorer connection. The best connected industry would rank 1$^{st}$ and the least connected industry would rank last, in the 969$^{th}$ position.

## 4.2 Cluster Analysis

Although extractive industries appear to be central in the network, their degree of relatedness may differ from one industry to another. In this section, we attempt to identify and characterize industry clusters in the networks as well as the centrality of the core extractive activities within their clusters. The data allow us to identify the Broad Economic Categories (BEC) of every industry. The concept of BECs, originally used in analysis of international trade figures, are groups of industries that share a common product category (United Nations, 2002). In the context of this study, BECs are useful to characterize industry clusters along different product categories.

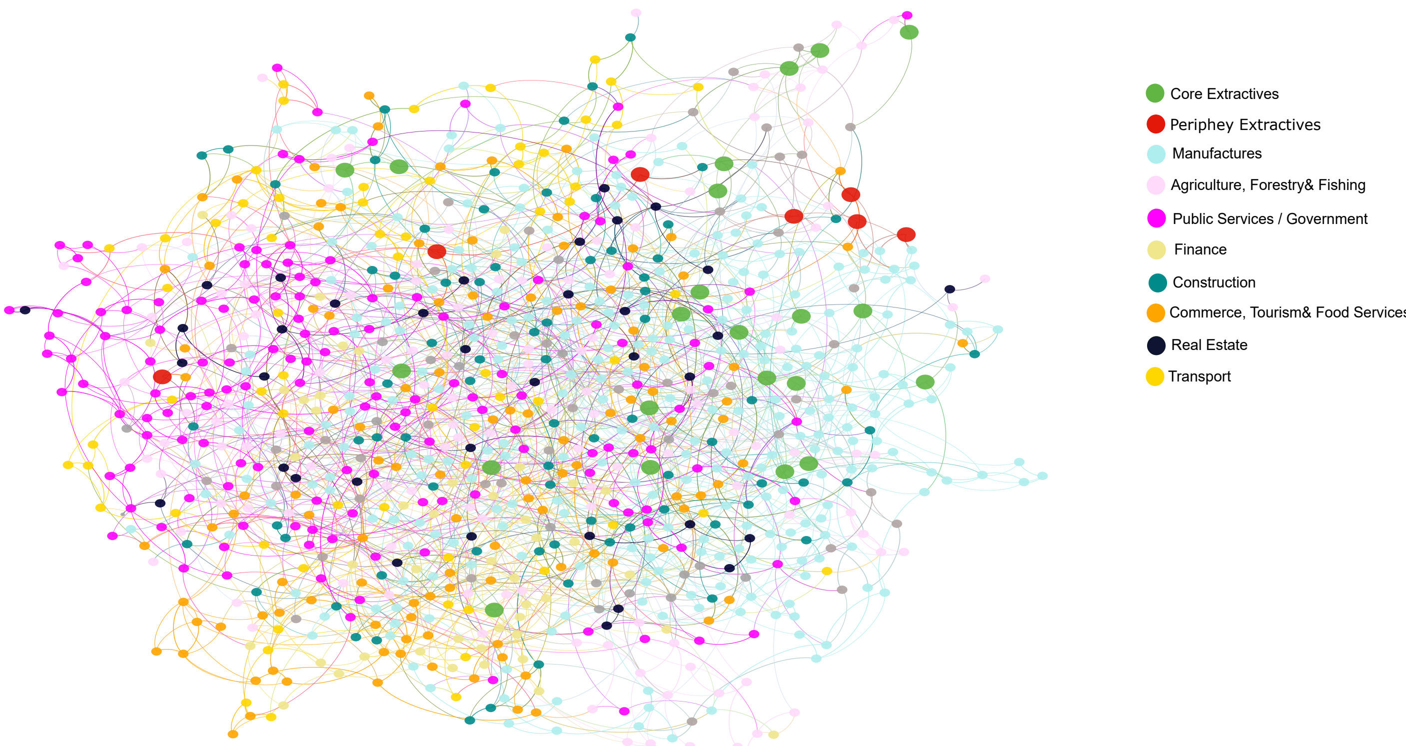
Graphic analysis of the network is a first step towards cluster detection in large networks. Given the vast extent of the network, we follow Hidalgo et al. (2007) and, in Figure 2, present the network using only the three strongest connections of each industry.[7] The colors in the graph represent the BEC of each node. With the exception of core and periphery extractive nodes, all nodes are colored according to the first two digits in their ISIC code, used to assign them to a BEC. The color code used in the graph allows us to visualize clustering among BECs. Some patterns can already be seen in this representation of the network. Core extractive industries seem to be within a cluster of the manufacturing sector, with some presence of agriculture, fishing and forestry, as well as government and public services.

To gain a better understanding of the clusters in this network, we used the Louvaine algorithm to detect communities in the network. The algorithm evaluates how much more densely connected the nodes in a cluster or community are as compared to connections in a random network (Blondel et al., 2008). This is particularly useful when identifying clusters in large and complex networks such as the industry space in Colombia. The algorithm identified a total of 24 clusters in the industry space, with extractive industries present in only 10. Figure 3 shows the same industry space as Figure 2, but only includes the industries that are within the 10 clusters with core extractive industries. The previous pattern remains, with many manufacturing firms accompanied by some government and public services and agriculture and fishing. Additionally, in Table 3, we present the per-

[7] All networks were graphed using the Gephi 0.9.2 "Force Atlas 2" algorithm with gravity set to 3 and scaling to 3. The maximum node size was 23 and the minimum was 13; nodes were set with no border width, and opacity was set to 90.

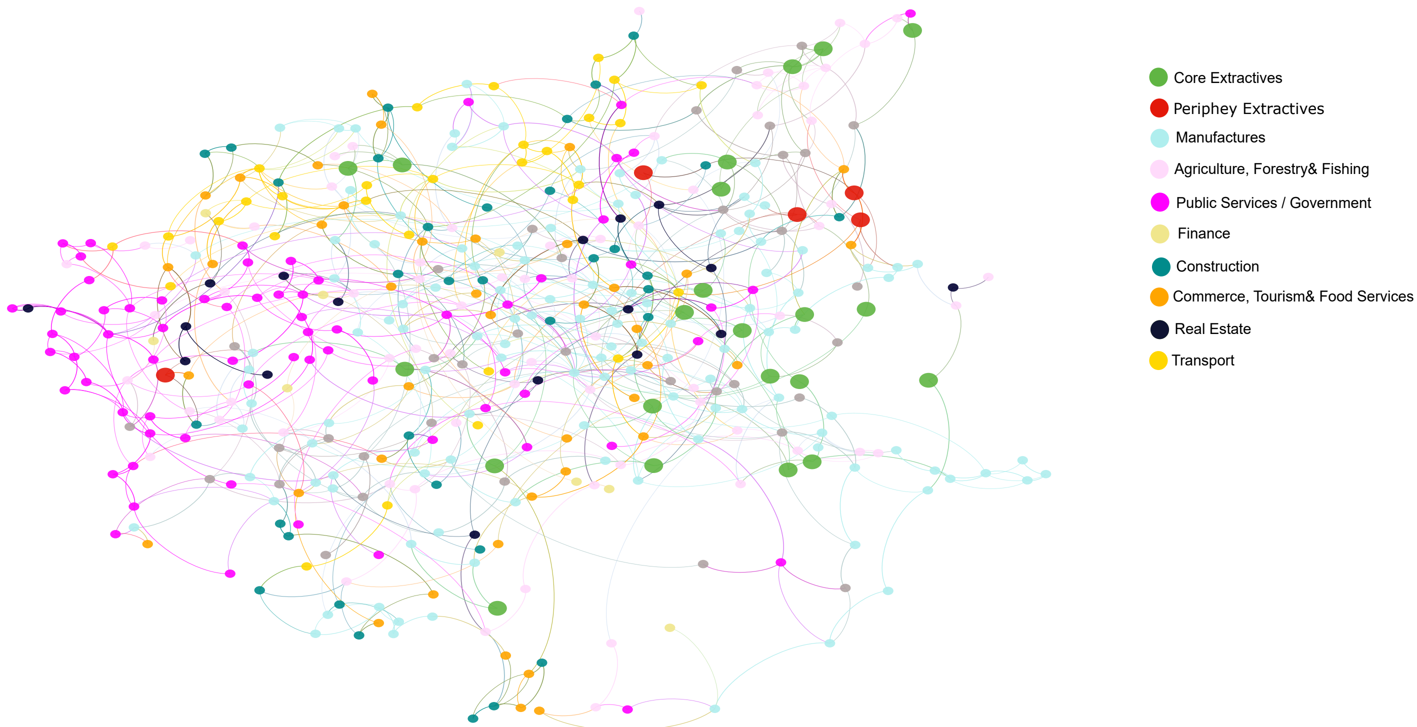
centage that each BEC in each of the 10 clusters with at least one core extractive industry. Manufacturing industries are the most prevalent industries in the vast majority of clusters containing core extractive industries, followed by agricultural industries and government services.

Figure 2: Three strongest links




Note: This graph presents the industry space using our skill-relatedness measure. Only the three strongest links for each node are presented in the graph. Colors denote the broad economic category (BEC) of each node. Core and periphery extractive nodes were enlarged to make them more visible. The pattern that emerges is that core extractive industries seem to be within a group of the manufacturing sector, with some presence of agriculture, fishing and forestry, as well as government and public services.

Figure 3: Three strongest links for detected clusters




Note: This graph presents the industry space using our skill-relatedness measure. Only the three strongest links are presented for nodes in clusters containing at least one extractive industry. The colors denote the broad economic category (BEC) of each node. Clusters were detected using the Louvaine community detection algorithm. The core and periphery extractive nodes have been enlarged to make them more visible. Core extractive industries are clearly within a group of the manufacturing sector, with some presence of agriculture, fishing and forestry, and government and public services.

Table 3: Percentage of every BEC in each extractive cluster

| | *Clusters* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| BEC | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| *Core Extractives* | 6 | 2 | 4 | 7 | 15 | 12 | 3 | 4 | 2 | 2 |
| *Manufacturers* | 45 | 33 | 38 | 41 | 52 | 9 | 16 | 30 | 29 | 11 |
| *Agriculture, Forestry & Fishing* | 6 | 14 | 15 | 27 | 9 | 30 | 23 | 19 | 11 | 9 |
| *Public Services\Government* | 6 | 10 | 8 | 15 | 12 | 5 | 29 | 6 | 11 | 55 |
| *Commerce, Tourism & Food Services* | 16 | 17 | 15 | - | - | 5 | 10 | 15 | 14 | 7 |
| *Construction* | 10 | 17 | 8 | - | 3 | 16 | 10 | 9 | 4 | 7 |
| *Transport* | 6 | 5 | - | - | - | 19 | 10 | 11 | 16 | - |
| *Real Estate* | - | - | 4 | 7 | 6 | 2 | - | 4 | 7 | 7 |
| *Periphery Extractives* | 3 | - | 4 | - | 3 | 2 | - | - | 2 | - |
| *Finance* | - | 2 | 4 | 2 | - | - | - | 2 | 5 | 2 |

Note: Authors' calculations based on data from PILA. Percentage of every BEC in each cluster with at least one extractive industry. Clusters are obtained using the community detection algorithm of Louvaine. The algorithm identified a total of 24 clusters in the industry space. Intuitively, a cluster groups industries with a denser connection.

Even though the core extractive industries have a better centrality ranking overall, this does not necessarily mean that they are as central in their clusters. Table 4 presents the average ranking of the most prevalent BECs in clusters containing core extractives, as well as the average ranking of all non-extractive industries. Core extractive industries are, on average, more central than the other most prevalent industries in the industry space. Their centrality is higher than that of manufacturing, government, and agriculture and fishing, which are the three most salient industries in extractive clusters. Core extractive activities have overall stronger connections than other industries and are, on average, more central within their clusters.

Table 4: Mean ranking of prevalent BECs in extractive clusters

| *Core Extractives* | *All Non-Extractive* | *Manufacturing* | *Government* | *Agriculture & Fishing* |
|---|---|---|---|---|
| 446 | 520 | 460 | 634 | 656 |

Note: Authors' calculations based on data from the Integrated Form of Contribution Paymengs (PILA). The table shows the mean strength degree centrality ranking of the most prevalent broad economic categories (BECs)in clusters containing at least one extractive industry. A higher ranking entails a lower strength degree centrality and thus a poorer connection within the clusters.

One of the most common algorithms for graphing and analyzing networks in the network literature is the Minimum Spanning Tree (MST). For the MST, only the strongest connection between nodes is kept, such that all nodes must remain connected. Thus, every node in the network will have at least one connection, and it will necessarily be its strongest one. This algorithm has often been used in the fields of communication and transportation networks, since it can compute the most efficient ways of reaching everyone in a network (*e.g.*, cellular coverage) with the minimum number of connections possible (*e.g.*, the minimum number of antennas).[8] For the setting of this paper, the MST provides a reliable threshold to detect the most important connections in the network and to visualize the shortest path connecting all industries with our skill-relatedness measure.

The MST of the network is shown in Figure 4. The relative position of each nodes determines its importance in the network. The most highly connected nodes are near the center of the graph, while nodes that are poorly connected are located towards the edges. There does not seem to be any pattern in the location of the core and periphery extractive industries, with some closer to the center of the MST and others further away.

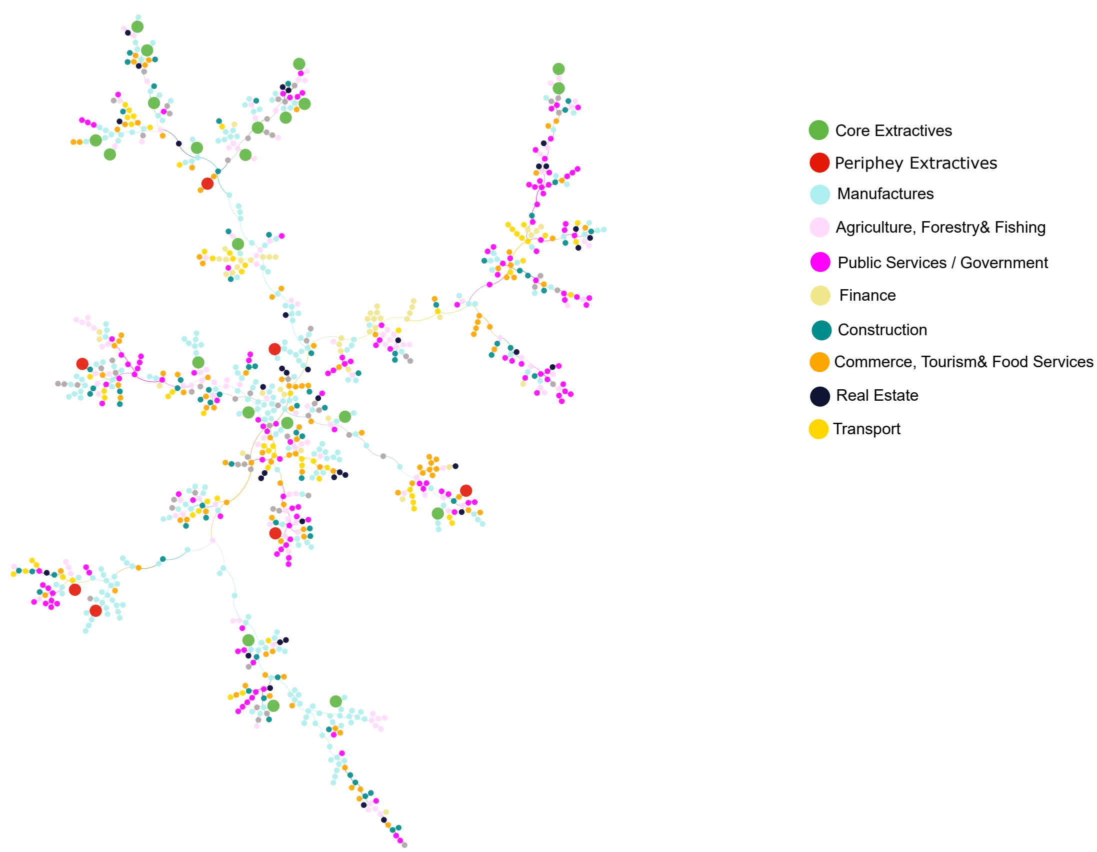
It is also important to identify whether nodes are clustered in only a few branches of the MST or if they are scattered throughout the MST. BECs that are spread more extensively throughout the MST are more interconnected, since such BECs are present in the shortest path between other industries. Even though core and periphery extractive industries are present in greater numbers in certain branches of the MST, they also seem to be scattered among several branches of the MST and not concentrated in just a few of them. This implies a high degree of interconnectedness between extractive industries and other economic

[8]See Nešetril and Nešetrilová (2012) for a more complete discussion of the origins of this concept.

sectors, making the Colombian economy particularly vulnerable to commodity busts, but also giving it a great advantage during a commodity boom.

In order to check whether the communities shown in Figure 3 have an intuitive pattern in the MST as well, Figure 5 shows the MST with only the nodes in the clusters colored in. The clusters are clearly very close to the extractive nodes in the MST, a natural result given the definition of the MST and the community detection algorithm used. An additional, fairly intuitive, result is that periphery extractive industries are also close to core extractive nodes. Manufacturing, agriculture, and government are very close to extractive industries on the shortest path of the network.

Figure 4: Minimum Spanning Tree (MST)




Note: The MST shows the strongest connection of every node such that all nodes are connected. Colors denote the broad economic category (BEC) of each node. Core and periphery extractive nodes were enlarged to make them more visible.

Figure 5: Minimum Spanning Tree (MST). Clusters colored


Note: The MST shows the strongest connection of every node such that all nodes are connected. Colors denote the broad economic category (BEC) of each node that belongs to a cluster with at least one core extractive industry. Clusters were detected using the Louvaine community detection algorithm. Core and periphery extractive nodes were enlarged to make them more visible.

## 4.3 Employment Growth Correlations

We now look into the relationship between employment trends in different sectors over time. In doing so, we seek to determine whether any pattern emerges among the industries related to extractive activities. Since the Dutch disease has often been portrayed as one of the main manifestations of the natural resource curse, a better understanding of the employment dynamics in various sectors of the economy can be crucial.

To this end, we investigate whether employment growth is negatively correlated between the most prevalent BECs in the industry space of the extractive clusters. We estimate a simple cross-BEC correlation of employment growth using each firm as a separate observation. Our definition of growth rates follows Davis et al. (2006) to avoid division by zero.

A very interesting pattern emerges from this analysis. The employment growth correlation in subperiod 1 was negative between core extractives and agriculture, construction, commerce, and government (see Table 5). However, the pattern shifted for subperiods 2 and 3. In those subperiods, there was a negative correlation in employment growth between core extractives and manufacturing, commerce, finance, and real estate . These results could be related to the economic cycles of commodities during the study period. As we showed in Figure 1, subperiod 1 saw a decline in commodity prices, whereas subperiods 2 and 3 saw an increase. It seems that when commodity prices are high, there is a negative correlation employment growth between extractive industries and higher-skill industries, while the opposite is true when commodity prices are low. Appendix Table A5, Table A6, Table A7, and Table A8 show the results for all the pairwise correlations.

Table 5: Employment growth correlation betweeen Core Extractives and other BECs

| Subperiod | Agri. | Manuf. | Constr. | Commerce | Transport | Finance | R. Estate | Gov. | Core |
|---|---|---|---|---|---|---|---|---|---|
| 1 | -0.281 | 0.240 | -0.110 | -0.104 | 0.145 | 0.237 | 0.100 | -0.055 | 1.000 |
| 2 | 0.336** | -0.053 | 0.003 | -0.218 | 0.230 | -0.220 | -0.155 | 0.101 | 1.000 |
| 3 | 0.342 | -0.183 | 0.040 | -0.323 | 0.188 | -0.364* | -0.212 | 0.078 | 1.000 |
| All | 0.269* | 0.028 | -0.031 | -0.227* | 0.166 | -0.076 | -0.102 | 0.055 | 1.000 |

Note: Authors' calculations based on data from PILA. Employment growth pairwise correlation between core extractive and the other BECs for the whole period (2008 - 2013) and every subperiod. Each industry in the BEC was taken as an observation. For ease of reading and pattern identification, red and green represent negative and positive correlations respectively. The darker the color, the greater the correlation. ***p<0.01, **p<0.05, *p<0.1

## 4.4 Economic Activity in the Network

One of our objectives is to characterize extractive-related industries, and to help identify instances where shocks to the extractive sector might impact other economic activities. Under the natural resource curse narrative, factors of production in the economy can be drawn to extractive industries in times of high commodity prices since extractive industries would be able to pay higher salaries. From the perspective of the inter-industry relatedness theory, stronger effects would be expected for extractive-related industries. While extractive-related industries can benefit from knowledge spillovers, they also suffer from higher competition during commodity booms. Both aspects, spillovers and inter-industry relatedness, have received attention lately in the natural resources literature. There is evidence suggesting that extractive-related industries tend to grow during commodity booms (Allcott and Keniston, 2017), although they also suffer from having to pay higher wages, as well as from higher competition and losing workers (Fitjar and Timmermans, 2019).

Using our skill-relatedness measure, we delve further into this line of research by characterizing extractive-related industries. We examine how the strength of the industry-relatedness correlates with several measures of an industry's economic performance and general structure. Since our measure stems from a labor flow-based measure, we begin by examining the general structure of the industrial labor force. We also try to capture some of the general economic performance of the firms. The Colombian Annual Manufacturing Survey (EAM) is the only database that gathers such information on a regular basis. Since the information in the EAM is limited to the country's manufacturing sector, these results are only valid for a subset of the total sample. Nonetheless, the cluster analysis shows that extractive industries are within communities in which manufacturing industries are prevalent. This makes manufacturing industries an interesting subsample to analyze.

In order to capture the overall relatedness of each industry in the EAM with the extractive sector, we calculate the average of our skill-relatedness measure between every industry and all extractive industries. This measure, which we call "proximity", is calculated as shown in Equation 5. Given that there are a finite and constant number of extractive industries, a higher proximity implies higher relatedness with the extractive sector. $SR_{ij}$ simply refers to our skill-relatedness measure as defined in Equation 4, $c$ represents the core extractive industries present in the EAM, and $C$ is the number of core extractive industries in the sample.

$$Proximity_{j,c} = \frac{\sum_{i=c}^{C} SR_{ij}}{C} \quad | \quad SR_{ij} \neq 0 \quad \forall \quad i \in core \tag{5}$$

We aim to characterize extractive-related industries using several dimensions that represent their general economic structure or performance. We do this by accounting for the average economic structure and performance of extractive industries using Equation 6. Here, $c$ represents the core extractive industries present in the EAM, $C$ is the number of core extractive industries in the sample, and $Industry\ Characteristic_{it}$ refers to the level of a characteristic that captures the general structure or performance of the industry $i$ during year $t$ in a given dimension (*e.g.* total number of employees or gross investment).

$$Extractives\ Characteristic_t = \frac{\sum_{i=c}^{C} Industry\ Characteristic_{it}}{C} \quad | \quad \forall \quad i \in core \tag{6}$$

Equation 7 uses our measures of proximity and industry characteristics to characterize industries in terms of their relatedness to extractive activities and the latter's characteristics. Here, $\delta$ indicates the correlation between the industry's characteristic and the latter's average relatedness to extractive industries. Similarly, $\beta$ measures the correlation between the industry's characteristic and the average of the same characteristic for extractive industries. Finally, $\gamma$ simply measures the correlation of the interaction term. With this simple equation we are able to see whether the structure of the most extractive-related activities differs from the rest of the industry space.

$$\begin{aligned} Industry\ Characteristic_{j,t} = \beta Extractives\ Characteristic_t + \delta Proximity_{j,c} \\ + \gamma Proximity_{j,c} \cdot Extractives\ Characteristic_t + \mu_{j,t} \end{aligned} \tag{7}$$

The endogenous nature of the skill-relatedness measure does not allow causal inference in this setting. Since our skill-relatedness measure stems from labor flows, we first investigate whether there is any general pattern in the structure of the industries' workforce. In Table 6 we present the results of the regression exercise using characteristics of the industries' labor force. The results indicate that a higher relatedness with core extractive industries is correlated with lower employment levels in firms using all the employee categories available in the EAM. By their very nature, extractive activities are capital intensive. Thus, a lower number of employees in extractive-related activities is consistent with the general nature of extractive activities. These results are robust to using our first two wage-based subsamples, but not to considering only highly paid workers (see Appendix Table A9). To verify that these results are not driven by extractive industries themselves, we exclude them from the sample and observe no changes in the qualitative characteristics of the results (see Appendix table Table A10).

Prior research has sought to estimate the effects of increased relatedness with extractive activities on economic performance. For instance, Fitjar and Timmermans (2019) exploit a rich dataset containing individual-level information on workers to study labor flows during the oil boom in Norway. Their findings suggest that oil-related industries suffer human capital losses during oil booms. Unfortunately, the data available do not allow us any insights into the type of workers affected beyond their wage level. Disentangling the reasons behind job-switching and creating other measures of skill relatedness using individual socioeconomic characteristics might help to further explain our results. For instance, it is puzzling that the negative effect of relatedness with extractive industries on employment levels disappears when we consider only highly paid workers. This could be due to a lower mobility among highly paid workers or also to a lower overall number of highly paid workers that keeps us from fully capturing the effect of inter-industry relatedness.

Table 6: Results from the regression exercise: employment levels

| | *Permanent* | *Temporal* | *Permanent & Owners* | *Permanent & Temporal* | *Total* |
|---|---|---|---|---|---|
| *Proximity* | -7,636** | -5,054*** | -7,722** | -11,026*** | -16,067*** |
| | (3,339) | (855) | (3,348) | (3,197) | (4,588) |
| *Extractives ch.* | 0 | 0 | 0 | 0 | 0 |
| | (0) | (0) | (0) | (0) | (0) |
| *Proximity*Ext. ch.* | 0 | 0 | 0 | 0 | 0 |
| | (1) | (0) | (1) | (0) | (0) |
| Observations | 790 | 790 | 790 | 790 | 790 |

Note: Authors' calculations based on data from PILA and EAM. This table presents results from a simple regression exercise in which we correlate the number of employees by categories of the industries with the average relatedness with extractive industries, its average level of employees in the same category and the interaction term. Standard error in parentheses. ***p<0.01, **p<0.05

We further exploit the rich EAM data by decomposing the labor force by gender. Table 7 presents the results for the regression exercise using the gender disaggregated data. We find that the lower level of employment in extractive-related industries is mainly due to lower female employment in those industries. The magnitude of the effect is particularly strong for female workers in production positions rather than administrative ones. No correlation is found for foreign workers of either gender. These results are robust to using any of our wage-based subsamples except for the most stringent one (see Appendix Table A11). As in the previous case, the results are robust to excluding the extractive industries from the sample (see Appendix Table A12).

Gender disparities have been well documented in the extractive sector (Eftimie et al., 2009). However, most of the literature has focused on the gender gap and gender disparities in the extractive sector alone (see for example Reeson et al., 2012; Ross, 2008). The evidence presented here shows that this gap extends beyond extractive industries to extractive-related industries. The reasons for this remain, however, unclear, begging further research. Policies targeted at the extractive sector and related industries should consider this gender dynamic. Indubitably, gender equality is a desirable policy and development goal in and of itself. There may, moreover, be direct and indirect benefits from greater gender equality (see, for example, Hill and King, 1995).

Table 7: Results from the regression exercise: employment levels by gender

| | Total | | Nationals | | Foreigners | | Production | | Administrative | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Female* | *Male* | *Female* | *Male* | *Female* | *Male* | *Female* | *Male* | *Female* | *Male* |
| *Proximity* | -9,068*** | -5,556 | -473*** | -512 | -1 | -10 | -5,363*** | -563 | -3,347*** | -2,803*** |
| | (1,320) | (12,460) | (101) | (2,702) | (1) | (31) | (832) | (13,850) | (550) | (820) |
| *Extractives ch.* | -0 | 1 | 0 | 0 | 0 | -2 | -0 | 1 | -0 | -0 |
| | (0) | (1) | (0) | (1) | (1) | (5) | (0) | (1) | (0) | (0) |
| *Proximity*Ext. ch.* | 0 | -0 | -0 | 0 | 0 | 3 | 0 | -1 | 1 | 1 |
| | (0) | (2) | (0) | (4) | (2) | (14) | (0) | (3) | (0) | (1) |
| Observations | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 |

Note: Authors' calculations based on data from PILA and EAM. This table presents results from a simple regression exercise in which we correlate the number of employees by gender and categories of the industries with the average relatedness with extractive industries, its average level of employees in the same category and the interaction term. National and Foreigners categories refer only to employees in production. Standard error in parentheses. ***p<0.01, **p<0.05.

Finally, we extend our analysis to characteristics related to the industries' overall performance. The results for the full sample are presented in Table 8. The industry characteristics are measured by DANE and were normalized by the total value of the firm's production prior to aggregation at the industry level. There seems to be a slight complementarity between capital-intensive extractive industries and labor-intensive related industries, though this correlation is not robust to different robustness checks (see Appendix Table A13 and Table A14). Likewise, there seems to be a negative correlation between extractive industries proximity and value added, as well as the former and raw materials. This correlation is reversed when only highly paid workers are considered (see Appendix Table A13). This is consistent with previous theoretical and empirical evidence that finds high complementarities between high-skilled workers and capital intensity (Krusell et al., 2000). However, our relatedness measure might not be ideally suited to capturing correlations not strictly related to the labor force, since it is a labor flow-based measure. Other reliable relatedness measures, such as those based on commodities (see Fan and Lang, 2000), may be used to asses the structure of the extractive sector in dimensions other than the labor force.

Table 8: Results from the regression exercise: economic performance and others

| | *Capital-Labor Ratio* | *Investment* | *Assets* | *Value Added* | *Electricity (Kw)* | *Fuels ($)* | *Energy ($)* | *Raw Materials* | *Transp. Costs* |
|---|---|---|---|---|---|---|---|---|---|
| *Proximity* | 175.4 | -2.3 | -24.7 | -81.7*** | -14.4 | -0.0 | -0.7 | -82.3** | -0.2 |
| | (566.0) | (2.0) | (78.3) | (29.3) | (18.6) | (0.0) | (1.5) | (33.4) | (0.5) |
| *Extractives ch.* | 0.3*** | 0.2 | 1.1*** | 0.2 | 0.3 | 0.1 | -0.3 | 0.2 | -0.1 |
| | (0.1) | (0.2) | (0.4) | (0.2) | (0.9) | (0.1) | (0.6) | (0.2) | (0.2) |
| *Proximity*Ext. ch.* | -0.7** | -0.3 | -1.0 | 0.1 | 0.4 | 0.0 | -0.5 | 0.3 | -0.0 |
| | (0.3) | (0.4) | (1.0) | (0.5) | (2.2) | (0.2) | (1.5) | (0.6) | (0.5) |
| Observations | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 |

Note: Authors' calculations based on data from PILA and EAM. This table presents results from a simple regression exercise in which we correlate some characteristics of the industry with the average relatedness with extractive industries, the average level of the same characteristic and the interaction term. All variables were normalized using the total value of production before aggregating at the industry level. Standard error in parentheses. ***p<0.01, **p<0.05.

# 5 Discussion

This paper presents a novel approach to study the subnational dynamics of extractive industries. We focus on the case of Colombia, a developing country rich in natural resources such as oil and coal, exploiting rarely seen granular data. By leveraging social security data on workers and firms, we perform a detailed network analysis of the extractive sector's relationship to other economic activities. Our aim is to present a new account of the industry network inter-linkages in Colombia, with a focus on extractive industries.

The skill-relatedness measure constructed in this paper allows us to better understand the industry space that might be vulnerable during a commodity boom or bust. We find that extractive industries play a central role in the industry space in Colombia and that some changes in the industry's centrality over time may reflect an intuitive understanding of commodity cycles. Moreover, the centrality of the extractive industries is also heterogeneous at the level of the workers' skills, with greater centrality when only higher skilled workers are considered. We also observe that extractive industries are closely related to manufacturing, agricultural activities, and government. This structure is robust to using two widely-used measures in network analysis: the minimum spanning tree and the three strongest links. Within their clusters, extractive industries seem to be more central actors in the industry space.

In addition, our analysis presents evidence of lower levels of employment in extractive-related industries. Given the nature of extractive activities, extractive-related industries may also be capital intensive. This could, nevertheless, also be suggestive of the presence of the natural resource curse in Colombia. We observe that industries that are more closely related to extractive industries have lower levels of employment, in line with that observed in other studies. Fitjar and Timmermans (2019), for example, find that oil-related industries in Norway suffered from human capital losses during the oil boom. We furthermore show that there exists a gender gap in the labor structure of extractive-related industries that should be addressed in future academic work and public policy.

Naturally, as this is a nascent area of study, much remains to be learned about the network structure of economies at the subnational level, as well as the implications of such structures for the future performance of specific industries and the economy overall. Moreover, a greater understanding is needed of the determinants of economic growth and the resource curse at the subnational level, including which worker aspects facilitate job switching to extractive industries. Our results indicate that there are clear heterogeneous effects due to worker's gender and, possibly, their skill level. Unfortunately, the data do not allow us to further characterize workers and disentangle the reasons behind job switching. Furthermore, the endogenous nature of our skill-relatedness measure means we cannot unambiguously measure the labor market implications of this study. The patterns described in Section 4 could, however, be investigated in greater detail by creating other measures of skill relatedness that use individual socioeconomic characteristics.

Another promising avenue for future research is the measurement of spillover effects of commodity booms or busts to the industries within extractive clusters, and the extent to which these may be impacted by gender gaps in extractive-related industries. Indeed, gender equality has been found to have heterogeneous impacts on wages distribution Flabbi et al. (2019). Whether skill relatedness translates into more than just negative employment correlations remains unclear. Since human capital is one of the building blocks of an industry, it would be worth investigating whether skill relatedness with extractive industries affects the economic performance of the industry in ways that go beyond employment levels. Further work might also examine the mechanisms through which such spillovers arise, accounting for possible gender biases.

# References


Allcott, H. and Keniston, D. (2017). Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America. *The Review of Economic Studies*, 85(2):695–731. [page 23.]

Aragón, F. M., Chuhan-Pole, P., and Land, B. C. (2015). *The local economic impacts of resource abundance: What have we learned?* The World Bank. [page 2.]

Aragón, F. M. and Rud, J. P. (2013). Natural resources and local communities: evidence from a peruvian gold mine. *American Economic Journal: Economic Policy*, 5(2):1–25. [page 2.]

Barabasi, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101–113. [page 7.]

Becker, G. (1964). *Human Capital*. University of Chicago Press. [pages 9, 11, and 47.]

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008. [page 15.]

Bonilla, L. (2020). Mining and human capital accumulation: Evidence from the colombian gold rush. *Journal of Development Economics*, page 102471. [page 2.]

Breschi, S., Lissoni, F., and Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research policy*, 32(1):69–87. [page 9.]

Caselli, F. and Michaels, G. (2013). Do oil windfalls improve living standards? evidence from brazil. *American Economic Journal: Applied Economics*, 5(1):208–38. [page 2.]

Chang, S. and Harbir, S. (1999). The impact of modes of entry and resource fit on modes of exit by multibusiness firms. *Strategic Management Journal*, 20(11):1019–1035. [page 9.]

Chang, S. J. (1996). An evolutionary perspective on diversification and corporate restructuring: entry, exit, and economic performance during 1981–89. *Strategic management journal*, 17(8):587–611. [page 9.]

Davis, S. J., Haltiwanger, J., Jarmin, R., Miranda, J., Foote, C., and Nagypal, E. (2006). Volatility and dispersion in business growth rates: Publicly traded versus privately held firms. *NBER macroeconomics annual*, 21:107–179. [page 22.]

EC Engelsman, A. v. R. (1991). *Mapping of Technology, A First Exploration of Knowledge Diffusion Amongst Fields of Technology.* Policy Studies on Technology and Economy (BTE) Series. [page 9.]

Eftimie, A., Heller, K., and Strongman, J. (2009). *Gender Dimensions of the Extractive Industries: Mining for Equity.* World Bank. [pages 3 and 26.]

Enrico, M. (2011). Local labor markets. In *Handbook of labor economics*, volume 4, pages 1237–1313. Elsevier. [page 8.]

Fafchamps, M., Koelle, M., and Shilpi, F. (2017). Gold mining and proto-urbanization: recent evidence from ghana. *Journal of Economic Geography*, 17(5):975–1008. [page 7.]

Fan, J. P. and Lang, L. H. (2000). The measurement of relatedness: An application to corporate diversification. *The Journal of Business*, 73(4):629–660. [pages 9 and 27.]

Farjoun, M. (1994). Beyond industry boundaries: Human expertise, diversification and resource-related industry groups. *Organization science*, 5(2):185–199. [page 9.]

Fitjar, R. D. and Timmermans, B. (2019). Relatedness and the resource curse: Is there a liability of relatedness? *Economic Geography*, 95(3):231–255. [pages 2, 23, 25, and 28.]

Flabbi, L., Macis, M., Moro, A., and Schivardi, F. (2019). Do female executives make a difference? the impact of female leadership on gender gaps and firm performance. *The Economic Journal*, 129(622):2390–2423. [page 29.]

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239. [page 8.]

Gathmann, C. and Schönberg, U. (2010). How general is human capital? a task-based approach. *Journal of Labor Economics*, 28(1):1–49. [pages 9, 11, and 47.]

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic management journal*, 17(S2):109–122. [page 9.]

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598. [page 8.]

Grogger, J. and Eide, E. (1995). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources*, pages 280–310. [page 11.]

Hausmann, R. and Klinger, B. (2006). Structural transformation and patterns of comparative advantage in the product space. [page 1.]

Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837):482–487. [page 15.]

Hill, M. A. and King, E. (1995). Women's education and economic well-being. *Feminist Economics*, 1(2):21–46. [page 26.]

Jackson, M. O. (2010). *Social and economic networks*. Princeton university press. [page 7.]

Jaffe, A. B. (1989). Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. *Research Policy*, 18(2):87–97. [page 9.]

Krusell, P., Ohanian, L. E., Ríos-Rull, J.-V., and Violante, G. L. (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, 68(5):1029–1053. [page 27.]

Masucci, A. P., Stanilov, K., and Batty, M. (2014). Exploring the evolution of london's street network in the information space: A dual approach. *Physical Review E*, 89(1):012805. [page 7.]

Michaels, G. (2010). The long term consequences of resource-based specialisation. *The Economic Journal*, 121(551):31–57. [page 2.]

Mincer, J. (1962). On-the-job training: Costs, returns, and some implications. *Journal of political Economy*, 70(5, Part 2):50–79. [page 11.]

Murphy, K. and Welch, F. (1989). Wage premiums for college graduates: Recent growth and possible explanations. *Educational Researcher*, 18(4):17–26. [page 11.]

Neal, D. (1995). Industry-specific human capital: Evidence from displaced workers. *Journal of labor Economics*, 13(4):653–677. [pages 9 and 47.]

Neffke, F. and Henning, M. (2013). Skill relatedness and firm diversification. *Strategic Management Journal*, 34(3):297–316. [pages 1, 2, and 10.]

Neffke, F. M., Otto, A., and Weyh, A. (2017). Inter-industry labor flows. *Journal of Economic Behavior & Organization*, 142:275–292. [pages 2 and 10.]

Nešetril, J. and Nešetrilová, H. (2012). The origins of minimal spanning tree algorithms–boruvka and jarnık. *Documenta Mathematica*, pages 127–141. [page 19.]

Nooteboom, B. (2000). Learning by interaction: absorptive capacity, cognitive distance and governance. *Journal of management and governance*, 4(1-2):69–92. [page 1.]

Parent, D. (2000). Industry-specific capital and the wage profile: Evidence from the national longitudinal survey of youth and the panel study of income dynamics. *Journal of Labor Economics*, 18(2):306–323. [page 9.]

Reeson, A. F., Measham, T. G., and Hosking, K. (2012). Mining activity, income inequality and gender in regional australia. *Australian Journal of Agricultural and Resource Economics*, 56(2):302–313. [pages 3 and 26.]

Ross, M. L. (2008). Oil, islam, and women. *American political science review*, pages 107–123. [pages 3 and 26.]

Santos, R. J. (2018). Blessing and curse. the gold boom and local development in colombia. *World Development*, 106:337–355. [page 2.]

Spender, J.-C. and Grant, R. M. (1996). Knowledge and the firm: Overview. *Strategic management journal*, 17(S2):5–9. [page 9.]

Sullivan, P. (2010). Empirical evidence on occupation and industry specific human capital. *Labour economics*, 17(3):567–580. [page 9.]

United Nations (2002). *Defined in Terms of the International Trade Classification, revision 3 and the Harmonized Commodity Description and Coding System*. [page 15.]

# A Complementary Tables

Table A1: Classification of Extractive Economic Activities in Colombia (PILA)

| *ISIC code* | *Industry Description* | *Classification* |
|---|---|---|
| 1010 | Extracción y aglomeración de carbon de piedra | Core Extractives |
| 1020 | Extracción y aglomeración de lignito | Core Extractives |
| 1030 | Extracción y aglomeración de lignito | Core Extractives |
| 1110 | Extracción de petróleo crudo y de gas natural | Core Extractives |
| 1120 | Actividades de servicios relacionados con la extracción de petróleo y de gas | Core Extractives |
| 1200 | Extracción de minerales de uranio y torio | Core Extractives |
| 1310 | Extracción de minerales de hierro | Core Extractives |
| 1320 | Extracción de minerales metaliferos no ferrosos, excepto uranio y torio | Core Extractives |
| 1410 | Extracción de piedra, arena y arcilla | Core Extractives |
| 1429 | Explotación de otras minas y canteras n.c.p. | Core Extractives |
| 2310 | Fabricación de productos de hornos de coque | Core Extractives |
| 2320 | Fabricación de productos de la refinación del petróleo | Core Extractives |
| 2413 | Fabricación de plasticos en formas primarias y de caucho sintetico | Core Extractives |
| 2696 | Corte, tallado y acabado de la piedra | Core Extractives |
| 2699 | Fabricación de otros productos minerales no metalicos n.c.p. | Core Extractives |
| 2710 | Industrias básicas de hierro y acero | Core Extractives |
| 2720 | Fabricación de productos primarios de metales preciosos y metales no ferrosos | Core Extractives |
| 2731 | Fundición de hierro y acero | Core Extractives |
| 2732 | Fundición de metales no ferrosos | Core Extractives |
| 2811 | Fabricación de productos metalicos para uso estructural | Core Extractives |
| 2891 | Forja, prensado, estampado y laminado de metales; pulvimetalurgia | Core Extractives |
| 4020 | Fabricación de gas; distribución de combustibles gaseosos por tuberias | Core Extractives |
| 2922 | Fabricación de maquinas herramienta | Periphery Extractives |
| 2923 | Fabricación de maquinaria metalúrgica | Periphery Extractives |
| 2924 | Fabricación de maquinaria para la explotación de minas y canteras y para obras de construcción | Periphery Extractives |
| 5050 | Venta al por menor de combustible para automotores | Periphery Extractives |
| 5141 | Venta al por mayor de combustibles solidos, liquidos y gaseosos y de productos conexos | Periphery Extractives |
| 5142 | Venta al por mayor de metales y minerales metaliferos | Periphery Extractives |
| 5143 | Venta al por mayor de materiales de contrucción, artículos de ferretería y equipo y materiales de fontanería y calefacción | Periphery Extractives |

Note: PILA. ISIC 3$^{rd}$ revision adapted for Colombia. Classification was based on the definitions of the extractive sector from DANE.

Table A2: Mean strength degree centrality of economic activities

| | *Subperiod* | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | *All* |
| *Full sample: All workers* | | | | |
| Core extractives | 0.129 | 0.130 | 0.142 | 0.133 |
| Periphery extractives | 0.166 | 0.153 | 0.166 | 0.173 |
| Non-extractives | 0.123 | 0.113 | 0.142 | 0.133 |
| *Subsample 1: Workers that earn more than the median wage* | | | | |
| Core extractives | 0.140 | 0.141 | 0.144 | 0.140 |
| Periphery extractives | 0.170 | 0.154 | 0.161 | 0.178 |
| Non-extractives | 0.125 | 0.116 | 0.139 | 0.135 |
| *Subsample 2: Workers that earn more than the minimum wage* | | | | |
| Core extractives | 0.131 | 0.124 | 0.163 | 0.153 |
| Periphery extractives | 0.165 | 0.140 | 0.156 | 0.180 |
| Non-extractives | 0.124 | 0.112 | 0.154 | 0.151 |
| *Subsample 3: Workers that earn more than 3 times the minimum wage* | | | | |
| Core extractives | 0.100 | 0.094 | 0.131 | 0.137 |
| Periphery extractives | 0.098 | 0.083 | 0.114 | 0.158 |
| Non-extractives | 0.086 | 0.081 | 0.124 | 0.131 |
| *Subsample 4: Workers that earn more than 10 minimum wages* | | | | |
| Core extractives | 0.047 | 0.039 | 0.081 | 0.097 |
| Periphery extractives | 0.028 | 0.023 | 0.043 | 0.068 |
| Non-extractives | 0.042 | 0.036 | 0.075 | 0.090 |

Note: Authors' calculations based on data from PILA. Mean of the strength degree centrality by sub periods, wage group subsamples, and main industry groups of analysis. A higher strength degree centrality entails a stronger connection.

Table A3: Standard Deviation of strength degree centrality of economic activities

| | *Subperiod* | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | *All* |
| *Full sample: All workers* | | | | |
| Core | 0.085 | 0.084 | 0.094 | 0.099 |
| Periphery | 0.040 | 0.037 | 0.050 | 0.034 |
| Other | 0.074 | 0.070 | 0.090 | 0.086 |
| *Subsample 1: Workers that earn more than the median wage* | | | | |
| Core | 0.081 | 0.075 | 0.090 | 0.093 |
| Periphery | 0.050 | 0.047 | 0.047 | 0.035 |
| Other | 0.074 | 0.070 | 0.087 | 0.086 |
| *Subsample 2: Workers that earn more than the minimum wage* | | | | |
| Core | 0.058 | 0.051 | 0.101 | 0.091 |
| Periphery | 0.055 | 0.047 | 0.064 | 0.037 |
| Other | 0.071 | 0.065 | 0.097 | 0.094 |
| *Subsample 3: Workers that earn more than 3 times the minimum wage* | | | | |
| Core | 0.052 | 0.049 | 0.070 | 0.065 |
| Periphery | 0.051 | 0.054 | 0.048 | 0.040 |
| Other | 0.061 | 0.061 | 0.083 | 0.084 |
| *Subsample 4: Workers that earn more than 10 minimum wages* | | | | |
| Core | 0.033 | 0.029 | 0.066 | 0.057 |
| Periphery | 0.028 | 0.026 | 0.035 | 0.043 |
| Other | 0.041 | 0.038 | 0.067 | 0.069 |

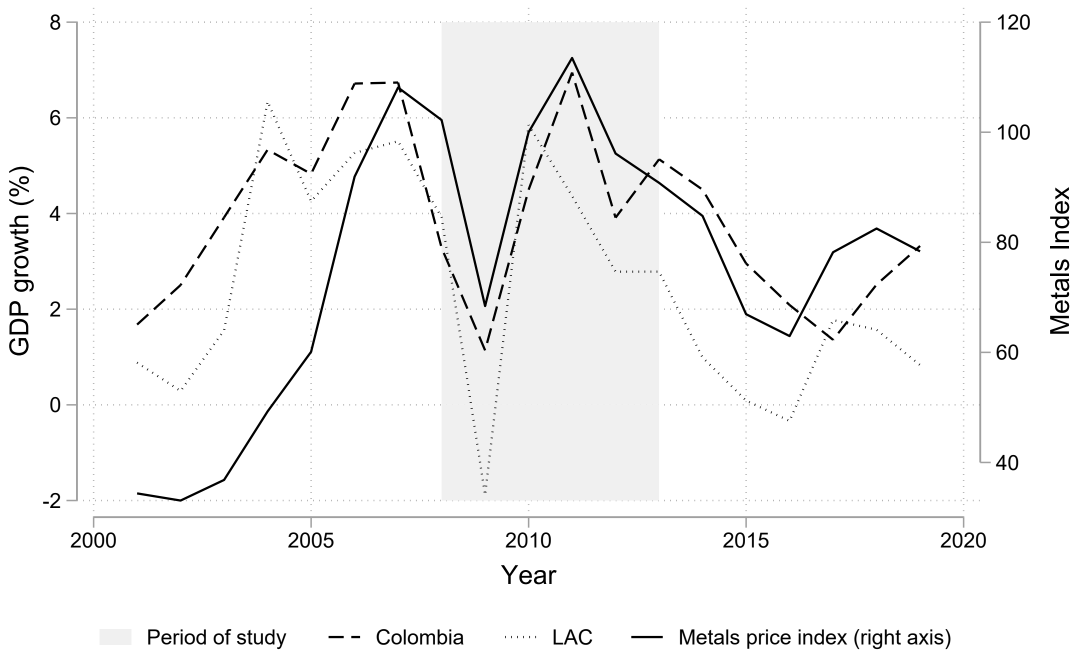
Note: Authors' calculations based on data from PILA. Standard Deviation of the strength degree centrality by sub periods, wage group subsamples, and main industry groups of analysis. A higher strength degree centrality entails a stronger connection.

Table A4: Standard Deviation of ranking of strength degree centrality of economic activities

| | *Subperiod* | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | *All* |
| *Full sample: All workers* | | | | |
| Core extractive | 290 | 301 | 295 | 308 |
| Periphery extractives | 162 | 146 | 164 | 141 |
| Non-extractives | 272 | 270 | 279 | 280 |
| *Subsample 1: Workers that earn more than the median wage* | | | | |
| Core extractive | 281 | 289 | 284 | 303 |
| Periphery extractives | 189 | 184 | 169 | 145 |
| Non-extractives | 272 | 270 | 279 | 280 |
| *Subsample 2: Workers that earn more than the minimum wage* | | | | |
| Core extractive | 237 | 243 | 286 | 284 |
| Periphery extractives | 204 | 183 | 177 | 127 |
| Non-extractives | 271 | 269 | 278 | 280 |
| *Subsample 3: Workers that earn more than 3 times the minimum wage* | | | | |
| Core extractive | 217 | 201 | 248 | 236 |
| Periphery extractives | 216 | 209 | 216 | 170 |
| Non-extractives | 267 | 266 | 274 | 275 |
| *Subsample 4: Workers that earn more than 10 times the minimum wage* | | | | |
| Core extractive | 229 | 230 | 236 | 222 |
| Periphery extractives | 201 | 199 | 235 | 241 |
| Non-extractives | 253 | 252 | 263 | 267 |

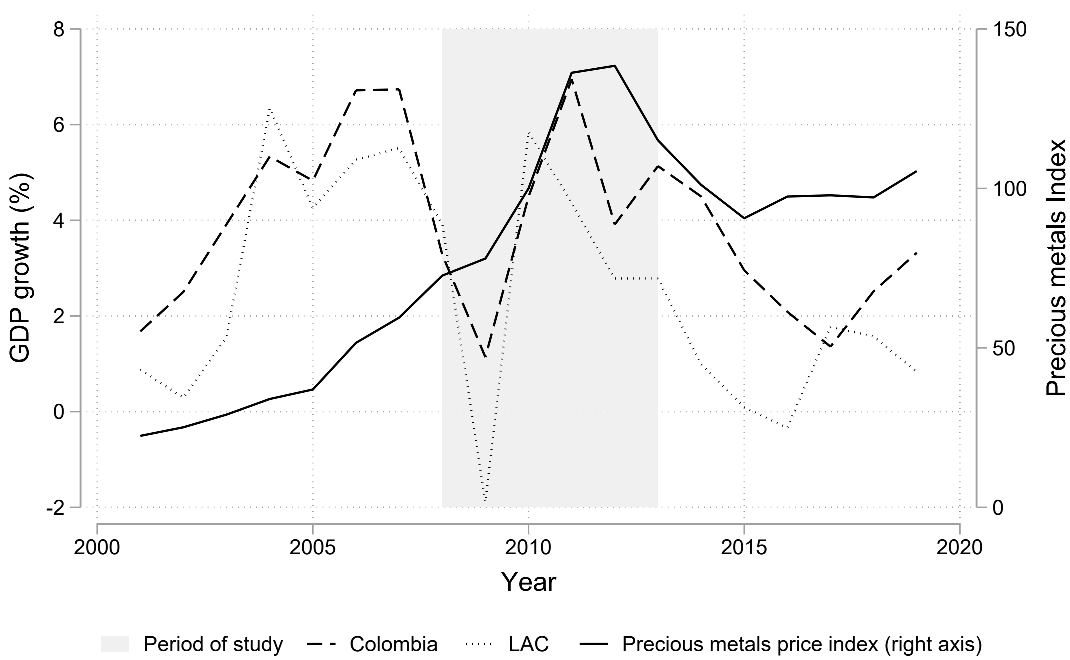
Note: Authors' calculations based on data from PILA. Standard deviation of the ranking of the strength degree centrality by subperiods, wage group subsamples, and main industry groups of analysis. A higher ranking entails a lower strength degree centrality, thus a poorer connection.

Figure A1: Dependence on metals & minerals




Metal price index is a weighted measure of Aluminum, Copper, Iron Ore, Lead, Nickel, Tin, and Zinc prices. The base year for the index is 2010
Source: Authors' calculations based on the World Bank's pink sheet data and development indicators.

Figure A2: Dependence on precious metals




Precious metals price index is a weighted measure of Gold, Silver, and Platinum prices. The base year for the index is 2010
Source: Authors' calculations based on the World Bank's pink sheet data and development indicators.

## Table A5: Employment growth correlation. Whole period

| | Agri. | Manuf. | Constr. | Commerce | Transport | Finance | R. Estate | Gov. | Core | Peri. | Ext. all |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Agri. | 1.000 | | | | | | | | | | |
| Manuf. | -0.002 | 1.000 | | | | | | | | | |
| Constr. | 0.189*** | 0.006 | 1.000 | | | | | | | | |
| Commerce | 0.145** | -0.007 | -0.074 | 1.000 | | | | | | | |
| Transport | 0.112 | 0.069 | 0.085 | -0.084 | 1.000 | | | | | | |
| Finance | 0.052 | 0.043 | 0.088 | -0.167* | -0.202** | 1.000 | | | | | |
| R. Estate | -0.050 | -0.020 | -0.028 | 0.148 | 0.181* | 0.125 | 1.000 | | | | |
| Gov. | 0.125** | -0.053 | 0.165** | 0.034 | 0.098 | -0.050 | 0.012 | 1.000 | | | |
| Core | 0.269* | 0.028 | -0.031 | -0.227* | 0.166 | -0.076 | -0.102 | 0.055 | 1.000 | | |
| Peri. | 0.354 | -0.329 | -0.281 | -0.111 | -0.407* | 0.211 | -0.009 | -0.031 | 0.001 | 1.000 | |
| Ext. all | 0.153 | 0.073 | -0.103 | -0.103 | 0.138 | -0.112 | -0.057 | -0.060 | 1.000*** | 0.001 | 1.000 |

Note: Authors' calculations based on data from PILA. Employment growth pairwise correlation between BECs for the whole period (2008 - 2013). Each industry in the BEC was taken as an observation. For ease of reading and pattern identification, red and green represent negative and positive correlations respectively. The darker the color, the greater the correlation. ***p<0.01, **p<0.05, *p<0.1

## Table A6: Employment growth correlation. Subperiod 1

| | Agri. | Manuf. | Constr. | Commerce | Transport | Finance | R. Estate | Gov. | Core | Peri. | Ext. all |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Agri. | 1.000 | | | | | | | | | | |
| Manuf. | 0.076 | 1.000 | | | | | | | | | |
| Constr. | -0.017 | 0.016 | 1.000 | | | | | | | | |
| Commerce | 0.082 | -0.034 | 0.001 | 1.000 | | | | | | | |
| Transport | -0.070 | 0.050 | 0.049 | 0.038 | 1.000 | | | | | | |
| Finance | 0.163 | 0.045 | 0.119 | -0.119 | -0.047 | 1.000 | | | | | |
| R. Estate | -0.272** | -0.114 | 0.002 | -0.149 | -0.030 | 0.015 | 1.000 | | | | |
| Gov. | -0.170** | -0.059 | -0.063 | -0.053 | -0.025 | -0.072 | -0.018 | 1.000 | | | |
| Core | -0.281 | 0.240 | -0.110 | -0.104 | 0.145 | 0.237 | 0.100 | -0.055 | 1.000 | | |
| Peri. | 0.400 | -0.282 | -0.193 | -0.010 | -0.307 | 0.127 | 0.360 | -0.226 | -0.158 | 1.000 | |
| Ext. all | -0.434** | 0.149 | -0.045 | 0.011 | 0.083 | 0.163 | 0.019 | -0.097 | 0.999*** | -0.158 | 1.000 |

Note: Authors' calculations based on data from PILA. Employment growth pairwise correlation between BECs in subperiod 1 (2008-2010). Each industry in the BEC was taken as an observation. For ease of reading and pattern identification, red and green represent negative and positive correlations respectively. The darker the color, the greater the correlation. ***p<0.01, **p<0.05, *p<0.1

Table A7: Employment growth correlation. Subperiod 2

| | Agri. | Manuf. | Constr. | Commerce | Transport | Finance | R. Estate | Gov. | Core | Peri. | Ext. all |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Agri. | 1.000 | | | | | | | | | | |
| Manuf. | -0.040 | 1.000 | | | | | | | | | |
| Constr. | 0.247*** | -0.019 | 1.000 | | | | | | | | |
| Commerce | 0.140* | -0.013 | -0.096 | 1.000 | | | | | | | |
| Transport | 0.168* | 0.096 | 0.075 | -0.141 | 1.000 | | | | | | |
| Finance | -0.011 | 0.027 | 0.081 | -0.177* | -0.225** | 1.000 | | | | | |
| R. Estate | -0.041 | -0.005 | -0.048 | 0.228* | 0.236** | 0.149 | 1.000 | | | | |
| Gov. | 0.182* | -0.075 | 0.256*** | 0.061 | 0.131 | 0.008 | 0.027 | 1.000 | | | |
| Core | 0.336** | -0.053 | 0.003 | -0.218 | 0.230 | -0.220 | -0.155 | 0.101 | 1.000 | | |
| Peri. | 0.618** | -0.494* | -0.205 | -0.143 | -0.537** | 0.326 | -0.279 | -0.017 | 0.186 | 1.000 | |
| Ext all | 0.288** | 0.041 | -0.099 | -0.133 | 0.153 | -0.216 | -0.093 | 0.015 | 0.999*** | 0.186 | 1.000 |

Note: Authors' calculations based on data from PILA. Employment growth pairwise correlation between BECs in subperiod 2 (2010-2011). Each industry in the BEC was taken as an observation. For ease of reading and pattern identification, red and green represent negative and positive correlations respectively. The darker the color, the greater the correlation. ***p<0.01, **p<0.05, *p<0.1

Table A8: Employment growth correlation. Subperiod 3

| | Agri. | Manuf. | Constr. | Commerce | Transport | Finance | R. Estate | Gov. | Core | Peri. | Ext. all |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Agri. | 1.000 | | | | | | | | | | |
| Manuf. | -0.061 | 1.000 | | | | | | | | | |
| Constr. | 0.363*** | -0.042 | 1.000 | | | | | | | | |
| Commerce | 0.162 | 0.007 | -0.056 | 1.000 | | | | | | | |
| Transport | 0.205 | 0.080 | 0.087 | -0.168 | 1.000 | | | | | | |
| Finance | 0.025 | 0.043 | 0.044 | -0.194 | -0.288* | 1.000 | | | | | |
| R. Estate | -0.030 | -0.010 | -0.015 | 0.301* | 0.263 | 0.195 | 1.000 | | | | |
| Gov. | 0.267*** | -0.078 | 0.344*** | 0.077 | 0.183 | -0.025 | 0.015 | 1.000 | | | |
| Core | 0.342 | -0.183 | 0.040 | -0.323 | 0.188 | -0.364* | -0.212 | 0.078 | 1.000 | | |
| Peri. | -0.740* | -0.069 | -0.373 | -0.134 | -0.739* | 0.481 | -0.363 | -0.040 | -0.538 | 1.000 | |
| Ext. alL | 0.293 | -0.013 | -0.122 | -0.232 | 0.166 | -0.336* | -0.118 | -0.062 | 1.000*** | -0.538 | 1.000 |

Note: Authors' calculations based on data from PILA. Employment growth pairwise correlation between BECs in subperiod 3 (2011-2013). Each industry in the BEC was taken as an observation. For ease of reading and pattern identification, red and green represent negative and positive correlations respectively. The darker the color, the greater the correlation. ***p<0.01, **p<0.05, *p<0.1

Table A9: Results from the regression exercise using employees categories

| | *Permanent* | *Temporal* | *Permanent & Owners* | *Permanet & Temporal* | *Total* |
|---|---|---|---|---|---|
| *Subsample 1: Workers that earn more than the median wage* | | | | | |
| *Proximity* | -7,426*** | -4,335*** | -7,519*** | -10,309*** | -14,631*** |
| | (2,775) | (687) | (2,782) | (2,621) | (3,771) |
| *Extractives ch.* | 0 | 0 | 0 | 0 | 0 |
| | (0) | (0) | (0) | (0) | (0) |
| *Proximity*Ext. ch.* | 0 | 0 | 0 | 0 | 0 |
| | (1) | (0) | (1) | (0) | (0) |
| Observations | 790 | 790 | 790 | 790 | 790 |
| *Subsample 2: Workers that earn more than the minimum wage* | | | | | |
| *Proximity* | -8,337*** | -4,639*** | -8,434*** | -11,630*** | -16,474*** |
| | (2,925) | (746) | (2,933) | (2,789) | (4,010) |
| *Extractives ch.* | 0 | 0 | 0 | 0 | 0 |
| | (0) | (0) | (0) | (0) | (0) |
| *Proximity*Ext. ch.* | 0 | 0 | 0 | 0 | 0 |
| | (1) | (0) | (1) | (0) | (0) |
| Observations | 790 | 790 | 790 | 790 | 790 |
| *Subsample 3: Workers that earn more than 3 times the minimum wage* | | | | | |
| *Proximity* | -2,881 | -1,167** | -2,913 | -3,223 | -4,367 |
| | (2,160) | (524) | (2,165) | (2,041) | (2,942) |
| *Extractives ch.* | 0 | 0 | 0 | 0 | 0 |
| | (0) | (0) | (0) | (0) | (0) |
| *Proximity*Ext. ch.* | 0 | -0 | 0 | 0 | 0 |
| | (0) | (0) | (0) | (0) | (0) |
| Observations | 790 | 790 | 790 | 790 | 790 |
| *Subsample 4: Workers that earn more than 10 times the minimum wage* | | | | | |
| *Proximity* | 2,755 | 148 | 2,781 | 2,743 | 3,855 |
| | (1,795) | (431) | (1,800) | (1,689) | (2,436) |
| *Extractives ch.* | 0** | 0 | 0** | 0** | 0** |
| | (0) | (0) | (0) | (0) | (0) |
| *Proximity*Ext. ch.* | -0 | -0 | -0 | -0 | -0 |
| | (0) | (0) | (0) | (0) | (0) |
| Observations | 780 | 780 | 780 | 780 | 780 |

Note: Authors' calculations based on data from PILA and EAM. This table presents results from a simple regression exercise in which we correlate the number of employees by categories of the industries with the average relatedness with extractive industries, its average level of employees in the same category and the interaction term. Standard error in parentheses. ***p<0.01, **p<0.05.

Table A10: Main results excluding extractive industries

| | *Permanent* | *Temporal* | *Permanent & Owners* | *Permanent & Temporal* | *Total* |
|---|---|---|---|---|---|
| *Proximity* | -10,986*** | -5,234*** | -11,096*** | -14,513*** | -20,953*** |
| | (3,314) | (747) | (3,322) | (3,084) | (4,470) |
| *Extractives ch.* | -0 | -0*** | -0 | -0 | -0 |
| | (0) | (0) | (0) | (0) | (0) |
| *Proximity*Ext ch.* | 1 | 1*** | 1 | 1* | 1** |
| | (1) | (0) | (1) | (0) | (0) |
| Observations | 752 | 752 | 752 | 752 | 752 |

Note: Authors' calculations based on data from PILA and EAM. This table presents results from a simple regression exercise in which we correlate the number of employees by categories of the industries with the average relatedness with extractive industries, its average level of employees in the same category and the interaction term. Extractive industries excluded from the correlation Standard error in parentheses. ***p<0.01, **p<0.05

Table A11: Results from the regression exercise

| | Total | | Nationals | | Foreigners | | Production | | Administrative | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Female* | *Male* | *Female* | *Male* | *Female* | *Male* | *Female* | *Male* | *Female* | *Male* |
| *Subsample 1: Workers that earn more than the median wage* | | | | | | | | | | |
| *Proximity* | -7,537*** | -9,993 | -381*** | -972 | -1 | -9 | -4,552*** | -6,584 | -2,697*** | -2,360*** |
| | (1,064) | (10,621) | (83) | (2,313) | (1) | (26) | (669) | (11,946) | (451) | (685) |
| *Extractives ch.* | -0 | 0 | 0 | 0 | 0 | -2 | -0 | 1 | -0 | -0 |
| | (0) | (1) | (0) | (1) | (1) | (5) | (0) | (1) | (0) | (0) |
| *Proximity*Ext. ch.* | 0 | 1 | -0 | 1 | -0 | 3 | 0 | 1 | 0 | 1 |
| | (0) | (2) | (0) | (3) | (1) | (12) | (0) | (3) | (0) | (1) |
| Observations | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 |
| *Subsample 2: Workers that earn more than the minimum wage* | | | | | | | | | | |
| *Proximity* | -8,559*** | -9,282 | -457*** | -808 | -1 | -10 | -4,976*** | -5,192 | -3,237*** | -2,747*** |
| | (1,146) | (11,144) | (88) | (2,429) | (1) | (28) | (724) | (12,511) | (477) | (722) |
| *Extractives ch.* | -0 | 1 | 0 | 0 | 0 | -3 | -0 | 1 | -0 | -0 |
| | (0) | (1) | (0) | (1) | (1) | (5) | (0) | (1) | (0) | (0) |
| *Proximity*Ext. ch.* | 0 | 1 | -0 | 0 | -0 | 4 | 0 | 0 | 1 | 1 |
| | (0) | (2) | (0) | (3) | (1) | (12) | (0) | (3) | (0) | (1) |
| Observations | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 |
| *Subsample 3: Workers that earn more than 3 times the minimum wage* | | | | | | | | | | |
| *Proximity* | -2,313*** | -5,337 | -156** | -188 | -0 | -1 | -1,264** | -4,942 | -973*** | -965* |
| | (821) | (8,274) | (64) | (1,766) | (1) | (19) | (512) | (9,312) | (351) | (532) |
| *Extractives ch.* | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 |
| | (0) | (1) | (0) | (1) | (0) | (4) | (0) | (1) | (0) | (0) |
| *Proximity*Ext. ch.* | 0 | 1 | -0 | -0 | 0 | 0 | -0 | 1 | 0 | 0 |
| | (0) | (1) | (0) | (2) | (1) | (9) | (0) | (2) | (0) | (1) |
| Observations | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 |
| *Subsample 4: Workers that earn more than 10 minimum wages* | | | | | | | | | | |
| *Proximity* | 483 | 6,866 | 54 | 787 | -0 | 2 | 216 | 5,510 | 286 | 602 |
| | (679) | (6,806) | (53) | (1,448) | (0) | (16) | (421) | (7,583) | (292) | (444) |
| *Extractives ch.* | 0 | 1* | 0** | 1 | 0 | -1 | 0 | 1 | 0* | 1 |
| | (0) | (1) | (0) | (1) | (0) | (4) | (0) | (1) | (0) | (0) |
| *Proximity*Ext. ch.* | -0 | -1 | -0 | -1 | 0 | -1 | -0 | -1 | -0 | -1 |
| | (0) | (1) | (0) | (2) | (1) | (7) | (0) | (2) | (0) | (1) |
| Observations | 780 | 780 | 780 | 780 | 780 | 780 | 780 | 780 | 780 | 780 |

Note: Authors' calculations based on data from PILA and EAM. This table presents results from a simple regression exercise in which we correlate the number of employees by categories and gender of the industries with the average relatedness with extractive industries, its average level of employees in the same category and the interaction term. National and Foreigners categories refer only to employees in production. Standard error in parentheses. ***p<0.01, **p<0.05.

Table A12: Main results by gender excluding extractive Industries

| | Total | | Nationals | | Foreigners | | Production | | Administrative | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Female* | *Male* | *Female* | *Male* | *Female* | *Male* | *Female* | *Male* | *Female* | *Male* |
| *Proximity* | -9,700*** | -11,898 | -561*** | -979 | -2 | -11 | -5,509*** | -5,690 | -3,761*** | -3,322*** |
| | (1,210) | (12,583) | (97) | (2,709) | (1) | (32) | (747) | (13,906) | (541) | (835) |
| *Extractives ch.* | -0** | 0 | -0 | 0 | 0 | -2 | -0*** | 1 | -0 | -0 |
| | (0) | (1) | (0) | (1) | (1) | (6) | (0) | (1) | (0) | (0) |
| *Proximity*Ext ch.* | 1*** | 1 | 1 | 1 | 0 | 3 | 1*** | 1 | 1*** | 2* |
| | (0) | (2) | (0) | (4) | (2) | (15) | (0) | (3) | (0) | (1) |
| Observations | 752 | 752 | 752 | 752 | 752 | 752 | 752 | 752 | 752 | 752 |

Note: Authors' calculations based on data from PILA and EAM. This table presents results from a simple regression exercise in which we correlate the number of employees by gender and categories of the industries with the average relatedness with extractive industries, its average level of employees in the same category and the interaction term. Extractive industries excluded from the correlation. National and Foreigners categories refer only to employees in production. Standard error in parentheses. ***p<0.01, **p<0.05.

Table A13: Results from the regression exercise

| | *Capital-Labor Ratio* | *Investment* | *Assets* | *Value Added* | *Electricity (Kw)* | *Fuels ($)* | *Energy ($)* | *Raw Materials* | *Transp. Costs* |
|---|---|---|---|---|---|---|---|---|---|
| *Subsample 1: Workers that earn more than the median wage* | | | | | | | | | |
| *Proximity* | 433.8 | -2.2 | -29.5 | -83.4*** | -18.7 | -0.0 | -0.4 | -90.3*** | -0.1 |
| | (485.0) | (1.7) | (65.1) | (23.9) | (15.5) | (0.0) | (1.3) | (27.5) | (0.5) |
| *Extractives ch.* | 0.3*** | 0.2 | 1.0*** | 0.1 | 0.1 | 0.1* | -0.2 | 0.0 | -0.1 |
| | (0.1) | (0.1) | (0.3) | (0.2) | (0.8) | (0.1) | (0.5) | (0.2) | (0.2) |
| *Proximity*Ext. ch.* | -0.7*** | -0.3 | -0.8 | 0.3 | 1.1 | -0.1 | -0.8 | 0.6 | -0.2 |
| | (0.3) | (0.4) | (0.8) | (0.4) | (1.9) | (0.2) | (1.3) | (0.5) | (0.4) |
| Observations | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 |
| *Subsample 2: Workers that earn more than the minimum wage* | | | | | | | | | |
| *Proximity* | 190.0 | -3.1* | -29.9 | -86.4*** | -18.1 | -0.0 | -0.8 | -93.8*** | -0.2 |
| | (512.2) | (1.8) | (69.6) | (25.7) | (16.5) | (0.0) | (1.4) | (29.4) | (0.5) |
| *Extractives ch.* | 0.3*** | 0.2 | 1.0*** | 0.1 | 0.2 | 0.1* | -0.3 | 0.1 | -0.1 |
| | (0.1) | (0.2) | (0.4) | (0.2) | (0.8) | (0.1) | (0.5) | (0.2) | (0.2) |
| *Proximity*Ext. ch.* | -0.6** | -0.3 | -0.9 | 0.2 | 0.8 | -0.1 | -0.6 | 0.5 | -0.1 |
| | (0.3) | (0.4) | (0.9) | (0.4) | (2.0) | (0.2) | (1.3) | (0.6) | (0.5) |
| Observations | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 |
| *Subsample 3: Workers that earn more than 3 times the minimum wage* | | | | | | | | | |
| *Proximity* | -97.3 | -0.2 | 6.5 | -25.4 | -6.4 | -0.0 | 0.4 | -29.5 | 0.0 |
| | (367.4) | (1.3) | (49.7) | (18.5) | (12.0) | (0.0) | (1.0) | (21.4) | (0.3) |
| *Extractives ch.* | 0.2** | 0.2 | 0.9*** | 0.1 | 0.2 | 0.1* | -0.2 | 0.1 | -0.1 |
| | (0.1) | (0.1) | (0.3) | (0.2) | (0.7) | (0.1) | (0.5) | (0.2) | (0.2) |
| *Proximity*Ext. ch.* | -0.3* | -0.2 | -0.4 | 0.2 | 0.5 | -0.0 | -0.5 | 0.4 | -0.2 |
| | (0.2) | (0.3) | (0.6) | (0.3) | (1.5) | (0.1) | (1.0) | (0.4) | (0.3) |
| Observations | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 | 790 |
| *Subsample 4: Workers that earn more than 10 times the minimum wage* | | | | | | | | | |
| *Proximity* | 253.0 | -0.8 | 102.5** | 29.4* | 12.7 | 0.0 | -0.0 | 38.5** | 0.2 |
| | (309.4) | (1.1) | (41.0) | (15.2) | (9.9) | (0.0) | (0.8) | (17.6) | (0.3) |
| *Extractives ch.* | 0.2 | 0.0 | 1.3*** | 0.4** | 1.1 | 0.1* | -0.7 | 0.6*** | -0.1 |
| | (0.1) | (0.1) | (0.3) | (0.2) | (0.7) | (0.1) | (0.5) | (0.2) | (0.2) |
| *Proximity*Ext. ch.* | -0.1 | 0.2 | -1.3** | -0.4 | -1.3 | -0.0 | 0.4 | -0.6* | -0.1 |
| | (0.2) | (0.2) | (0.5) | (0.3) | (1.2) | (0.1) | (0.8) | (0.3) | (0.3) |
| Observations | 780 | 780 | 780 | 780 | 780 | 780 | 780 | 780 | 780 |

Note: Authors' calculations based on data from PILA and EAM. This table presents results from a simple regression exercise in which we correlate some characteristics of the industry with the average relatedness with extractive industries, the average level of the same characteristic and the interaction term. Standard error in parentheses. ***p<0.01, **p<0.05.

Table A14: Main results other characteristics excluding extractive Industries

| | *Capital-Labor Ratio* | *Investment* | *Assets* | *Value Added* | *Electricity (Kw)* | *Fuels ($)* | *Energy ($)* | *Raw Materials* | *Transp. Costs* |
|---|---|---|---|---|---|---|---|---|---|
| *Proximity* | -744.9 | -1.8 | -67.1 | -114.7*** | -23.8 | -0.0 | -0.6 | -118.0*** | 0.1 |
| | (518.9) | (2.0) | (80.0) | (28.0) | (19.0) | (0.0) | (1.6) | (32.7) | (0.5) |
| *Extractives ch.* | 0.0 | 0.2 | 0.8** | -0.1 | -0.1 | 0.1 | -0.3 | -0.2 | -0.1 |
| | (0.1) | (0.2) | (0.4) | (0.2) | (0.9) | (0.1) | (0.6) | (0.2) | (0.2) |
| *Proximity*Ext. ch.* | 0.0 | -0.5 | -0.3 | 0.9* | 1.6 | -0.1 | -0.7 | 1.1* | -0.4 |
| | (0.3) | (0.4) | (1.0) | (0.5) | (2.3) | (0.2) | (1.5) | (0.6) | (0.5) |
| Observations | 752 | 752 | 752 | 752 | 752 | 752 | 752 | 752 | 752 |

Note: Authors' calculations based on data from PILA and EAM. This table presents results from a simple regression exercise in which we correlate some characteristics of the industry with the average relatedness with extractive industries, the average level of the same characteristic and the interaction term. All variables were normalized using the total value of production before aggregating at the industry level. Extractive industries excluded from the correlation. Standard error in parentheses. ***p<0.01, **p<0.05.

# B Labor Flows

In this appendix, we describe the way in which the labor flows were created. The labor flows in this paper are used as the basis for generating the inter-industry skill-relatedness measure. Given the nature of our data and the purpose of this paper, we consider labor flows in a broader sense than in the previous literature. In this paper, a "labor flow" exists every time a worker has a job in two different industries. Cases in which a person has two jobs at the same time are within the scope of this paper since such a case entails a connection between those industries. That is precisely the kind of relatedness that this paper is trying to capture, since the worker can perform tasks in both industries. Our definition of labor flows therefore encompasses both job switching and simultaneous jobs.

Furthermore, given that our period of analysis extends over five years, we created two types of labor-flow measures. According to the first measure (measure A), a labor flow between any pair of industries $i$ and $j$ occurs whenever an individual has worked in both $i$ and $j$. This is counted as a single labor flow irrespective of whether the individual moves back and forth between $i$ and $j$. Measure A has the advantage of considering all labor flows that exist between industries without giving more weight to individuals who work simultaneously in the same firms over the years. For this measure to accurately capture skill relatedness between industries, we need to assume that the worker's skills remain relatively stable over the study period and that industries are unable to drastically change workers' skills.

To see why that is the case, consider an individual who worked in industry $i$ from 2008 until 2010, in industry $k$ in 2011-2012, and in industry $j$ in 2013. This measure would capture one flow between each pair of industries ($i$, $j$ and $k$). In fact, this type of relatedness is exactly what this paper tries to capture. As skills are specific to positions, firms, and industries (see Becker, 1964; Gathmann and Schönberg, 2010; Neal, 1995) the relatedness between $i$ and $j$ would simply be mediated by the fact that $i$ and $k$ have shared skills, as do $k$ and $j$. The short time span of our study period allows us to make such assumptions.

The second labor-flow measure (measure B) is a time-disaggregated version of measure A. In this case, labor flows are constructed in the same way as in measure A but for three two-year periods: 2008-2010, 2010-2011, and 2011-2013.[9] Thus, for the whole period, the measure has a limit of three labor flows per individual between a pair of industries. This is a stricter measure since it allows relatedness to be present only in subsequent years. Take the example of the worker described in measure A and imagine that the worker used to also work for industry $x$ in 2011. If we use measure A, then there would be a labor flow between $i$ and $x$, but that labor flow would not be considered in measure B. For the purposes of this paper, we presented all of the results using measure A. Nevertheless, the qualitative characteristics of the results and the policy implications remain unchanged when using measure B. The results are available upon request.

[9]The reader may remember from Section 2 that that there are no available data for 2009 and 2012.