An economic complexity approach for structural heterogeneity in the Brazilian agricultural sector

Alexandre de Queiroz Stein[1]
Gustavo Britto[2]

Abstract: This paper assesses the application of the economic complexity methodology to assess the issue of structural heterogeneity in the agricultural sector. The literature mostly focuses on the manufacturing industry, when trade data is used, having only more recently incorporated the services sector when firm-level data was taken into consideration. Applications regarding exclusively the agricultural sector are scarce and it is still an open question whether these activities can be assessed through the lens of economic complexity, given considerable differences in the nature of their production process. Moreover, apart from difficulties related to acquiring reliable data, one should not assume that economic complexity's central tenets can be applied to agricultural goods. Using establishment-level data aggregated into region data from the latest Brazilian Agricultural Census, the paper tests the use of the economic complexity methodology in agricultural production. As a proof of concept, two empirical exercises are carried out. First, we calculate similarity measures between agricultural products, create a product space, apply a method of network clustering and use the clusters to characterize the productive structure of Brazilian regions. In the second exercise, we present an agricultural complexity index and provide evidence of its relationship with development indicators. The results show that producing goods located at different areas of the product network is strongly associated with different levels of productivity, education, technology, land inequality and human development. Based on this evidence, we argue that the economic complexity toolkit applied to data on agricultural products can fill a key gap in the literature, which is the lack of proper indicators to represent the productive heterogeneity across products and regions.

Keywords: Agricultural Economic Complexity, Agriculture Product Space, Structural Heterogeneity.

Draft submitted for the EAEPE 35th Annual Conference, Leeds
Belo Horizonte - London, September 2023

[1] Center for Development and Regional Planning (Cedeplar), Universidade Federal de Minas Gerais (UFMG).
[2] Center for Development and Regional Planning (Cedeplar), Universidade Federal de Minas Gerais (UFMG).
The authors are grateful for the financial support from the Brazilian National Research Council (CNPq) [421124/2022-9] and from Coordination for the Improvement of Higher Education Personnel (Capes) [88887.837602/2023-00].

1. Introduction

Economic literature has long emphasized the key role played by the agricultural sector's development within the general economic development process. However, most of the debate on structural change over the decades, with notable exceptions, has focused on the transformative role played by the relative growth of modern sectors vis a vis backward ones.

From the overwhelming and all-encompassing importance of the manufacturing industry in classical development theories, the literature moved on to the relevance of embedded knowledge, be it the form of the technological intensity of the production of goods and services or in the more recent shape of more and more complex products and sectors.

Over time, the importance of the modernization of the of agricultural production has been acknowledged. However, more in-depth analysis of more detailed characteristics of these services has been mostly carried out by more specialized literature. Consequently, it is common to notice that, particularly in the context of development economics, the agricultural sector, albeit properly characterized by its marked structural heterogeneity, is often described in dualistic terms. Very high productivity, export-oriented production is contrasted with lower productivity, internal market and sometimes subsistence-focused production. This dichotomy can be also represented by the socioeconomic chasm observed between large versus small property production or in the more highly mechanized, R&D demanding, production, versus lower-tech, labour-intensive, traditional cultures that dominate vast areas of developing nations.

On the other hand, in the context of development studies, a remarkably wide variety of works, either thematic or case studies, can be found covering a myriad of relevant socioeconomic aspects related to the connection between agricultural development and economic development as a whole.

As mentioned before, studies focusing on the manufacturing industry and the role of diversification and structural change have followed a very different path. Increasing levels of disaggregation and of specific thematic analysis have been incorporated into the main agenda of the field. In more recent years, studies that draw from the economic complexity perspective originally developed by Hidalgo et al. (2007), Hidalgo e Hausmann (2009a), Hausmann et al. (2014) have proliferated, branching to several different areas.

In a little more than a decade, complexity economics diversified to a vast array of themes and areas, either directly through the evaluation of the relationship between economic complexity indices and various socioeconomic characteristics or through the still very fertile exploration of the principle of relatedness Hidalgo et al. (2018). A wide-ranging review of the diversity of studies is well depicted in Hidalgo (2021).

However, it has not been used to directly assess the agricultural sector. The vast majority of the literature that uses trade data has been restricted to industrial goods and does not capture internal market dynamics. More recently, with the expansion of the use of firm-level data based on formal employment, services started to be taken into consideration together with industrial goods. However, it is still a problem to analyze sectors characterized by high levels of informality, such as agriculture.

Beyond this issue, applications of economic complexity exclusively to the agricultural sector are scarce. Also, it is still an open question if the sector can actually be assessed through the lens of economic complexity. The sector has considerable differences in the nature of its production process, i.e., the prominent necessity of specific natural conditions, arable area, different work relations in rural areas, the attachment of workers to the means of production, the existence of production which are not destined for the markets (subsistence) and the perishable condition of products that can directly affect the geography of agriculture.

To tackle this substantial gap in the literature, this paper aims to address both these problems by adapting and applying methodologies developed by the literature on economic complexity to the agricultural sector. It does so by calculating an index to measure and characterize agricultural structural heterogeneity upon which an analytical framework can be built to inform development policies. We use a regional approach based on sectoral data on products whose production is not only destined for exports but also for the internal market, which is in itself a novelty for the economic complexity literature.

Two empirical exercises are carried out to test the suitability of the economic complexity approach to the agricultural sector. In the first exercise, we calculate the Agricultural Complexity Index, which is analogous to the Economic Complexity Index. However, the given sectors' peculiar characteristics, such as the dependence on climate, soil, land size, proximity to markets and infrastructure, there is no a priori reason to suggest that the economic reason can be transposed from manufacturing to agriculture. The underlying hypothesis is that even given the marked structural differences, different levels of diversification of local production, together with the relative "rarity" of the production of individual goods can give a meaningful approximation to a region's implicit capacity to combine knowledge and productive factors to produce a basket of agricultural goods.

In the second exercise, we calculate similarity measures between agricultural products, which are used to create a (agricultural) product space. In addition, we apply a network clustering method to characterize the productive structure of Brazilian regions. After a methodological discussion about the agricultural product space, we show that producing goods located in different areas of the network is related to different levels of productivity, education, technology, land inequality, and human development.

This proof of concept is carried out using data on agricultural production from the latest Brazilian Agricultural Census. This data is particularly suited to apply the economic complexity methodology to the agricultural sector.

First, it includes the entirety of the very vast and varied agricultural production across a huge notoriously unequal territory. It includes information on over five million rural establishments which are aggregated in over five regions, producing an array of over three hundred products. Secondly, the census data is not restricted to goods produced for external markets and includes all sorts of production including informal and subsistence activities. In this sense, its coverage goes well beyond that of more common complexity studies.

Based on the results we argue that it is possible to use the relatedness principle to map paths for case-specific regional diversification policies and that the economic complexity index is a good indicator to assess the productive structural heterogeneity in agriculture.

The paper is divided into five sections besides this introduction. Section two includes a brief presentation of the relevant literature, which allows us to situate the paper's research question as well as the gap in the literature. Section three brings details on the methodology and data used in the exercises. Section four consists of the presentation and discussion of the results. Section five brings final remarks and considerations.

2. Agriculture and Economic Complexity

Similarly to the whole of the country's socioeconomic fabric, the Brazilian agricultural sector has been historically characterized by a high degree of heterogeneity. Macroeconomic and sectoral differences across a wide variety of indicators are also associated with marked differences in the productive structure of the regions. Such disparities have been the hallmark of the country since colonial times. Vast chasms between high-productivity, exporting-producing activities and low-productivity, internal market and subsistence-producing ones have persisted over time and across the territory. Over time, these inequalities have endured several waves of productive modernization, which transformed a deeply agricultural and rural country into a mostly services, manufacturing-led, urban one.

Unsurprisingly, it has become common to use the well-known concept of structural heterogeneity to characterize the country's agricultural production, as very high and very low productivity sectors and activities coexist persistently through time (PINTO, 1970).

The use of this concept to understand the Latin American countries' agricultural sector is still important. However, it has also become commonplace to emphasize only two opposing types of agriculture: one of very large, high-productivity properties, usually exporting a significant share of the production, and the other of small, low-productivity properties, largely for subsistence. However compelling for broad stoke analysis, such as those carried out under or related to dualistic theories and formal models (Lewis and others), an in-depth appraisal of the sector would require more detailed information to draw an accurate picture of the sector diversity.

The dualistic views on agriculture are usually related to the markets for which products are destined: modern agriculture produces commodities for external markets and family (or peasant) farming produces a variety of goods directed to subsistence or local markets. But agriculture includes on a vast variety of products, which are commercialized on distinct markets that differ on the kind of products, their circulation flows through time and space, the actors involved, the local and global legislation to which they are subjected, and the role they play in a broader supply-chain. They are complex socio-material constructions which directly affect their production and consumption patterns. They cannot be considered just a flow of "commodities" or flows for local subsistence. In this perspective, the nature of products is defined by a variety of upsurging nested markets that form highly heterogeneous regional agriculture. Ploeg, Ye e Schneider (2022) show how usually the variety of internal markets are treated as homogeneous in dualistic views or as very specific case studies.

In rural areas it is possible to find a broad huge variety of livelihood ways. (ELLIS, 2000; PERONDI; SCHNEIDER, 2012; SCHNEIDER, 2010; SCHNEIDER; CASSOL, 2013; BELIK, 2015). Hence, finding a more accurate way to measure and represent the structural heterogeneity in agriculture is still a challenge. There's no consensus about which indicators and methods are more suitable to represent the diversity across regions and to illuminate possible paths to development. The literature regarding the agricultural sector in Brazil explores different aspects such as technology diffusion, productive diversification and specialization, inequality, human development, work relations among other aspects, trying to portray and give a satisfactory understanding of its structural heterogeneity (GASQUES et al., 2010; FORNAZIER; VIEIRA-FILHO, 2012; MARCONATO; LAROCCA; QUINTANILHA, 2012; SANTOS; VIEIRA-FILHO, 2012; FELEMA; RAIHER; FERREIRA, 2013; OLIVEIRA-FERREIRA; VASCONCELOS, 2014; SAMBUICHI et al., 2014; SILVEIRA, 2014; SAMBUICHI et al., 2016; VIEIRA-FILHO; FISHLOW, 2017; SOUZA et al., 2018).

Despite the important body of work available, there is no consistent methodological framework capable of resuming the different aspects the productive heterogeneity among regions. Moreover, there is no toolkit to inform the design of policies to foster regional development in agriculture that considers each region's particularities.

The economic complexity perspective emphasizes how local capabilities and productive knowledge play a central role in determining which set of goods and services can be produced by a region. It has been an important framework for applied economics studies about growth, development and international trade in the last 15 years (HIDALGO et al., 2007; HIDALGO; HAUSMANN, 2009a; HAUSMANN et al., 2014; STOJKOSKI; UTKOVSKI; KOCAREV, 2016; CHÁVEZ; MOSQUEDA; GÓMEZ-ZALDÍVAR, 2017; TACCHELLA; MAZZILLI; PIETRONERO, 2018; BRITTO et al., 2019).

Its empirical toolkit, centred around the economic complexity index, has been extensively used also to analyze changes in countries' productive structure. Moreover, recently it has also been rather prolific in showing how the level of economic complexity relates to a multitude of socioeconomic indicators, such as income inequality, labour market characteristics, technology, and environmental issues (HARTMANN et al., 2017; SBARDELLA; PUGLIESE; PIETRONERO, 2017; CAN; GOZGOR, 2017; FAWAZ; RAHNAMA-MOGHADAMM, 2019; NEAGU; TEODORU, 2019; ZHU; YU; HE, 2020; LAPATINAS; LITINA; ZANAJ, 2021; ROMERO; GRAMKOW, 2021; STOJKOSKI; KOCH; HIDALGO, 2023).

Recently, this approach has been used to better understand different aspects of urban and regional economies, especially in terms of specialization patterns and technological changes (BALLAND; RIGBY, 2017; BALLAND et al., 2018; BALLAND et al., 2019; FREITAS, 2019; MONTRESOR; QUATRARO, 2020; GOMEZ-LIEVANO; PATTERSON-LOMBA, 2021).

The main reason for the fast growth of the economic complexity literature is related to the agnosticism of its main empirical methods and principles. The methods have no *a priori* assumptions regarding the nature of the factors that influence economic activities, such as the production composition by capital, labour, technology and human capital (HIDALGO, 2021).

The economic complexity approach uses three main indicators: revealed comparative advantages, considered as a specialization indicator, products' ubiquity, taken as a product sophistication measure, and the degree of regional diversification. From this basic starting point, the use of dimensionality reduction techniques and network analysis provides a large amount of information regarding the local productive knowledge basis. Economies are regarded essentially as complex production systems and the empirical framework provides efficient means to obtain information about its emerging properties. These characteristics have been shown to be useful to analyze heterogeneous economic structures using trade, employment and patent data.

The economic complexity approach focuses on dealing with the central issue of development: why some countries developed and others do not? The classical literature on development economics and that of the Latin American structuralist tradition (ECLAC) converged on the importance of a leading industrial sector and structural change in order to achieve higher levels of income[3].

The heterogeneous productive conditions and the local differences in available capabilities make it hard to empirically identify diversification patterns with the precision needed to build solid public policies for development. Besides showing robust evidence of the existence of a core-periphery structure in the world trade network and the importance of industrial sectors for development, the seminal papers of Hidalgo et al. (2007) and Hidalgo e Hausmann (2009a) presented a way to empirically portray the composition of productive structures and map paths to diversify in feasible industrial sectors without the need of knowing the nature of each capability needed for production of specific goods and which are the determinant production factors of an activity (HIDALGO, 2021).

[3] The convergent view does not prevent differences between those perspectives, ECLAC's view on core-periphery dependence as a particular historical phenomenon of the capitalist development.

The flexibility of the economic complexity approach to infer capabilities requirements resulted in great advances in several themes. For instance, the drivers for innovation are multiple and long debated in economics, but the works of Balland e Rigby (2017), Balland et al. (2018) and Balland et al. (2019) were able to successfully map technological innovation patterns in regions by applying the economic complexity framework to patent data in Europe and the United States. No assessments were needed on factors usually mentioned as key factors for innovation such as university-industry interactions, R&D investments, human capital or the organization of national innovation systems. The same applied to understanding production and diversification patterns in research activities, using paper citation data, labour occupations, employment data, and cities.

3. Methodology and Data

The methodological challenges for the high-level disaggregated analysis of the agricultural sector, within the economic complexity perspective, encounter difficult-to-surmount barriers. These barriers are primarily related to the limited availability of data and technological classifications for agricultural activities, making both the method's application and the analysis of the obtained results challenging. In this paper, we circumvent these difficulties by calculating the Agricultural Complexity Index, the Agricultural Product Space and clusters using regionally aggregated data from the 2017 Brazilian Agricultural Census. Finally, we test for the association between the index and relevant economic variables.

3.1 Data

In order to analyze the complexity of the Brazilian agricultural sector we used the Brazilian Agricultural Census of 2017 from the Brazilian Institute of Geography and Statistics (IBGE). The Census contains information on 5.073.324 rural establishments across the territory, which embraces "*any production unit dedicated, either wholly or partially, to agricultural, forestry, or aquaculture activities, regardless of its size*" (BRASIL, 2019, p. 9). The interviews are conducted with the managers of rural establishments and collect data about production, land use, rural employment, agricultural area, technologies used in agriculture, education of people employed in agriculture, ethnicity, and family bonds, among several others. The Census is the most complete survey on Brazilian agriculture. However, the field of development economics doesn't have a tradition of using this data. The data is mostly used by other fields of economics and agrarian sciences.[4]

The agricultural census effectively solves a major issue of data availability for agriculture,

[4] For example see Cardille e Foley (2003), Rada, Helfand e Magalhães (2019), Ferreira e Féres (2020), DePaula (2020), Neves et al. (2021), Oliveira et al. (2022), Junior, Rodrigues e Silva (2022).

given that all other usual databases used in economic complexity approaches have considerable problems regarding the sector. Export data are unable to capture the dynamics of the domestic market, which accounts for the biggest share of Brazilian agricultural production. Employment data has limits regarding agriculture due to the high level of informality in the sector. Patent data primarily capture innovations related to the industrial sectors underlying agriculture and have a bias towards professionalized and export-oriented agriculture. The Census solves all these problems by collecting data about establishments in the sector, regardless if they are legally formalized or not, if they are family farming producing for local markets or exported-oriented, if they are illegal land occupations or occupied by social movements or even indigenous lands.

Another limitation of highly disaggregated studies in agriculture is the absence of standardization in the classification of agricultural products or activities. While export data, economic activities, occupations, and patent technologies have well-defined international classifications, agricultural products and activities lack standardized classifications. An alternative would be to use commodity data, but this would mean overlooking a significant portion of production and would not be capable of accommodating the sector's diversity. The same goes for product classification like the Harmonized System (HS) or the Common Classification of MERCOSUR (NCM). Given their focus on exports, they are not able to capture the country's productive diversity. The Census allows us to gather information on the production value of 312 products of all kinds across the country, including those produced for self-consumption or local markets. It allows considering the least developed regions that do not figure in other datasets in the estimations and to have a more accurate view of regional disparities in the analyses.

The main limitation of the Census are two-folded: the first is its declaratory nature. The main datasets in complexity approaches are usually administrative records, which are ideal. However, the high informality in the sector, its diversity of markets and the vast expanse of Brazilian territory make it impracticable to establish regular administrative records of agricultural activities at the desegregation level offered by the Census. Second, the IBGE does not make the Census micro-data available for public use, which implies a significant amount of work in collecting and standardising the information.

Finally, it is worth saying that the Census is the only database in Brazil that includes reliable data on agricultural production for all municipalities in Brazil with a considerable product disaggregation[5]. The data are also available for eight levels of regional aggregation defined by IBGE. In this paper, we used data on production value for 312 different products aggregated in 510 immediate regions. We did not use the municipality level - that is the most disaggregated level - to avoid informational loss given by the non-identification criteria of the survey, which removes information of any kind that had less than 3 respondents on the chosen regional level (BRASIL, 2019). Using the second disaggregation level allowed us to capture a

[5] Other databases such as the Municipal Agricultural Survey and the Municipal Livestock Survey don't map the entirety of Brazilian regions.

high data variability with low informational loss.

3.2 Economic complexity measures for agriculture

The economic complexity approach is grounded in the understanding that economies are complex systems of information organization, and the goods produced within an economy are construed as the outcome of this process of informational organization (HIDALGO, 2015). A product results from the combination of individuals' knowledge (explicit knowledge) and know-how (tacit knowledge) within the economy, whose interactions occur through various means, particularly within markets. Therefore, products inherently encapsulate fragments of distinct knowledge dispersed throughout society, the so-called capabilities. Economic complexity methods posit that the assessment of an economy's portfolio of products has the potential to illuminate the accumulated productive knowledge within that economy and that this knowledge is closely related to its economic development. The hard issue of our study is to assess if the empirical economic complexity approach to knowledge and its relation with economic development holds when analyzing the agricultural sector alone. In order to do that we need to estimate economic complexity indicators for agriculture. (HIDALGO; HAUSMANN, 2009a; HAUSMANN et al., 2014).

Defining who produces what and where is the key step to mapping the pool of an economy's knowledge. Here we define which regions are specialized in each type of agricultural product. This is done by using the revealed comparative advantage indicator (RCA) of Balassa (1965), following the standard procedure in literature proposed by Hidalgo et al. (2007), Hidalgo e Hausmann (2009a) and Hausmann et al. (2014), for the 510 Brazilian immediate regions the 313 agricultural products in our database. This indicator is equivalent to the location quotient which has long been used in regional economics and presents a comparison between the share of the good p in total production of region r and the share of the same good in the reference economy. Formally it is defined as:

$$RCA_{rp} = \frac{X_{rp}}{\sum_p X_{rp}} / \frac{\sum_r X_{rp}}{\sum_r \sum_p X_{rp}} \tag{1}$$

Where X_{rp} is the production value of product p in the region r. If the share of product p in region r is bigger than the share of this product in the average of the whole economy, then we say that region r is specialized in the production of p. Bringing to our specific subject, if the production of soybeans represents 20% of a region r total agricultural production value and the soybeans production value represents 10% of total agricultural production value in the Brazilian economy, we say that region r has a revealed comparative advantage in soybeans. In practice, it means that we are defining a threshold of 1 to the RCA indicator to define who produces what competitively. Hence, from the RCA indicator of a product in a region, we build a binary M_{rp} matrix with r rows and p columns, where $M_{rp} = 1$ if $RCA_{rp} \geq 1$, and zero otherwise.

From the RCA measure, we can obtain a diversity indicator that sheds light on the amount of capabilities a region has. If a product contains the capabilities from the region where it was produced, the pool of products made in that region is able to impart information about the capabilities present therein. The more diversified the region is, the higher the number of capabilities it is expected to have. Here, we obtained the diversity indicator ($K_{r,0}$) by the sum of the rows of M_{rp} matrix, which results in a count of the number of products made by the r.

$$K_{r,0} = \sum_{p} M_{rp} \tag{2}$$

However, just the amount of products alone is not enough to capture, albeit indirectly, a region's pool of knowledge. It is necessary to take the quality of the capabilities into consideration. This is done by considering spatial patterns. As mentioned before, products are composed of capabilities that are associated with two dimensions of knowledge: explicit and tacit. While the first can be more easily transferred between regions - especially by education and human capital improvements - the second is considered to be the hard learn-by-doing kind of knowledge, which is closely connected to the territory. Thus, the region-product relationship is mediated by spatial characteristics associated with the type of knowledge required for each product. There are simpler kinds of knowledge that are more easily transmitted across space and therefore more ubiquitous. And then there are others that represent rarer, harder-to-obtain knowledge.

Consequently, an indicator of a product's ubiquity can bring important information. We can assume that the ubiquity of a good reveals how rare - or sophisticated - the capabilities required for its production are, or how specific the combination of the necessary capabilities is. Goods with low ubiquity tend to be more sophisticated, as they require rarer capabilities that are difficult to transmit across space, while highly ubiquitous goods tend to be easily producible. In other words, if tomatoes are cultivated everywhere, we can assume that their production requirements are simpler than grapes for wine production, which are highly concentrated in a specific region in the south of Brazil. The calculus of the product ubiquity is analogous to that of the region's diversity, where the ubiquity ($K_{p,0}$) of product p is obtained by summing the columns of the binary matrix M_{rp}.

$$K_{p,0} = \sum_{r} M_{rp} \tag{3}$$

The iteration between ubiquity and diversity is the methodological key to assessing productive knowledge. Diversification brings little information on the quality of the capabilities of a region and ubiquity does not say much about the amount of capabilities a product contains. A region may be highly diversified in very simple products, making several products from a restricted knowledge base, while a product can have low ubiquity because of a very specific natural condition that is not associated with knowledge sophistication. In order to solve these issues, we applied the Method of Reflexes as proposed by Hidalgo e Hausmann (2009a), deriving the economic complexity index (ECI) and the product complexity index (PCI), which are based on an iterative method of mutual control between ubiquity and complexity. Formally, the generalized equations of the N iteration is given by equations 4 and 5:

$$k_{r,N} = \frac{1}{k_{r,0}} \sum_{p} M_{rp} \cdot k_{p,N-1} \tag{4}$$

$$k_{p,N} = \frac{1}{k_{p,0}} \sum_{r} M_{rp} \cdot k_{r,N-1} \tag{5}$$

Solving these equations for a chosen N will result in two symmetric matrices of products by products and regions by regions which condense information about the data, similarly to multivariate analysis methods like PCA or factorial analysis. The complexity indexes are obtained by getting the second auto-vectors of those matrices and normalizing them by their standard deviations from the average. The N iteration chosen to solve the equations is the one where a further iteration does not change the ranking of regions and products obtained by ordering the standardized auto-vectors[6].

The economic complexity indexes are basically qualified measures of diversity and ubiquity, where a region is more complex the more diversified it is and the less ubiquitous its products are. So diversified economies based on very common products get lower scores. Similarly, complex products are the ones with low ubiquity made in highly diversified regions. The idea is that diversified economies that produce rare products reveal they possess higher amounts of capabilities and rarer kinds of knowledge, which can be combined to make a big diversity of sophisticated goods. And complex products are the ones that are rare because they can be obtained only by the proper organization of a highly diversified informational system, and not by specific natural conditions.

3.3 The Agricultural Product Space

The economic complexity approach differs from the conventional conceptions in development economics when it comes to structural change. The most usual view is based on Hirschman's unbalanced development which emphasizes the input-output relations between activities and where the diversification processes are determined within sectors, particularly by demand pulls from well-established activities to the new ones. The economic complexity perspective shifts from the logic of input-output market flows as activities' diversification determinant to a framework where the nature of knowledge and capabilities are the key. Complexity

[6] For more on the method of reflections see Hidalgo e Hausmann (2009a) and its methodological appendixes on Hidalgo e Hausmann (2009b).

regards new activities or products as a result of new ways of combining existing productive knowledge or as a fruit of the introduction of a new kind of specific capability. Hence, the complexity view offers a shift from a perspective where intra-sectoral relations to focus on another in which knowledge complementarity is more important and where the inter-sectoral relations tend to be more relevant because of the role that capabilities' diversity plays in the diversification process. Following these changes, the well-known input-output methodologies give room to the product space as a tool for analyzing the relation between sectors, activities or products.

The product space is a network where each node represents one product and the connection between two nodes represents the similarity of capabilities embedded in a pair of products (HIDALGO et al., 2007; HAUSMANN et al., 2014). It allows us to understand the relation between products, visualize a country's productive structure in a disaggregated network and use network analysis tools in order to assess possible paths for productive diversification. Once more, if similar products do share common capabilities, it is more feasible for an economy to diversify into products or activities similar to its knowledge base. This diversification idea is synthesized as the *"principle of relatedness"*, which can be defined as *"an empirical principle describing the probability that a region enters (or exits) an economic activity as a function of the number of related activities present in that location"* (HIDALGO et al., 2018, p. 452).

The principle of relatedness is verified in multiple adaptations of the product space made in the last years that use such criteria for estimating similarities between activities. Some examples are the technologies spaces from patent data (BOSCHMA; BALLAND; KOGLER, 2015), research spaces in physics from publication data (CHINAZZI et al., 2019), occupational and skills networks from employment data (MUNEEPEERAKUL et al., 2013; ALABDULKAREEM et al., 2018). Despite sector-specific adaptations of the product space that have been made before, such as the sector-level manufacturing industries space of Neffke, Henning e Boschma (2011), our proposition of structuring a product-level network for agriculture is a novelty in the literature.

The Agricultural Product Space was built following the methodology proposed by Hidalgo e Hausmann (2009a). We estimated the proximity between a pair of products - which gives us a measure of similarity between their capabilities - using the usual co-location criteria, in which the higher the number of regions that produce those two products, the higher the similarity between them.

Formally, in order to estimate the proximities between products, we multiply the transposed $(M_{rp})^T$ matrix by its original. It results in a product by product square matrix M_{pp} where the main diagonal shows the number of regions that produce each product (same measure as the product ubiquity $K_{p,0}$) and the other entries present the number of regions that produce both goods. For example, the entry value for M_{13} shows the number of regions that produce good number 1 and good number 3 with revealed comparative advantage. This is expressed as:

$$M_{pp} = (M_{cp})^T.M_{cp} \tag{6}$$

The proximity measure ($\rho_{pp'}$) between products p and p' is given by the number of regions producing both products ($M_{pp'}$) divided by the biggest ubiquity between the goods ($max(M_{pp}; M_{p'p'})$):

$$\rho_{pp'} = \frac{M_{pp'}}{max(M_{pp}; M_{p'p'})} \tag{7}$$

This process results in a weighted symmetric adjacency matrix which is the base for the product space network. However, we made simulations and found that randomly assigning the RCA values in the original (M_{rp}) binary matrix hardly will render values bigger than 0.3[7]. Removing the edges above these thresholds tends to provide a clear representation of the relations between products. So, we defined the agricultural product space as a network where two nodes are connected by edges following two criteria: (i) The edge is part of the graph's maximum spanning tree or (ii) the similarity between them is bigger than 0.3: $\rho_{pp'} \geq 0.3$[8]. The first condition assures that our network does not include isolated groups of products and the second reduces the presence of non-significant connections.

Finally, we applied the Leiden network clustering method of Traag, Waltman e Eck (2019) to the agricultural product space, in order to identify homogeneous groups of products with similar productive capabilities. As will be clear below, this is an important step that allows us to interpret the results and connect economic complexity in agriculture to structural change and economic development.

3.4 Connecting agricultural products and economic development

The economic complexity approach allows for an intuitive and straightforward depiction of the productive structure of countries and regions. It provides a relatively easy way to differentiate highly developed industrial regions with sophisticated services sectors from underdeveloped regions, usually characterized by agricultural products or products based on natural resources. The interpretation of results in those cases is very intuitive, having been corroborated by the literature on development economics.

However, in a sector-specific case like agriculture, the interpretation of the results is not as straightforward. For instance, it is harder to differentiate whether the production of soybeans is more sophisticated than the large-scale mechanized production of mushrooms, or if

[7] The simulations' results are shown in the appendix. The simulations, inspired by Hidalgo et al. (2007) and Hidalgo e Hausmann (2009a) are shown in the appendix.

[8] This is distinct from Hidalgo et al. (2007), in which the definition of product space considers all connections between product, or in other words, the edges weight criteria is just $\rho_{pp'} > 0$. However, the authors use a threshold of $\rho_{pp'} \geq 0.5$ for the network representation.

a diversified family farming region with well-established institutional markets for small-scale production has a more complex knowledge base and productive structure than a region based on large scale grain production for external markets. No technological classifications are available specific for agricultural products, such as the one offered by Lall (2000) to industrial products, neither there is consensus in the literature regarding what kind of agricultural production leads to higher levels of economic development in the long run.

In order to address this issue and check the relevance of the agricultural complexity index, we opted for a strategy of empirically verifying the relation between products, complexity and socioeconomic indicators. To do so we used the method proposed by Hartmann et al. (2017) when considering income inequality. It consists of computing the weighted average of the social and economic indicators regarding the regions that produce each product p, using W_{rp} (share of product p in region r that produces p) as weights. In the formulas below, Y_r stands for the value of each variable of interest Y in region r.

$$\overline{\mathrm{Y}}_{\mathrm{p}} = \frac{\sum_r W_{rp} \mathrm{Y}_r}{\sum_r W_{rp}} \tag{8}$$

Where

$$W_{rp} = M_{rp} \frac{X_{rp}}{\sum_{p'} X_{rp'}} \tag{9}$$

It provides a link between products and economic indicators and allows us to test if there are differences between the relation of products and development indicators for groups of products in the product space, considering their economic complexity levels. As highlighted by Hidalgo (2021), this approach also allows to creation of counterfactual levels for each variable for an economy, given possible changes in its product portfolio and, by doing so, to find interesting diversification paths in order to enhance development.

4. Results and discussion

In this section, we show that our proposed application of the product space for the agricultural sector with data on production focused on internal markets, as well as exported-oriented production, can bring relevant information about the regional economies. It allows us to identify patterns of regional specialization in agriculture, make assessments on specific agricultural productive structures, map which paths of diversification are more interesting possibilities for improving agriculture production and reduce structural heterogeneity across the country's regions.

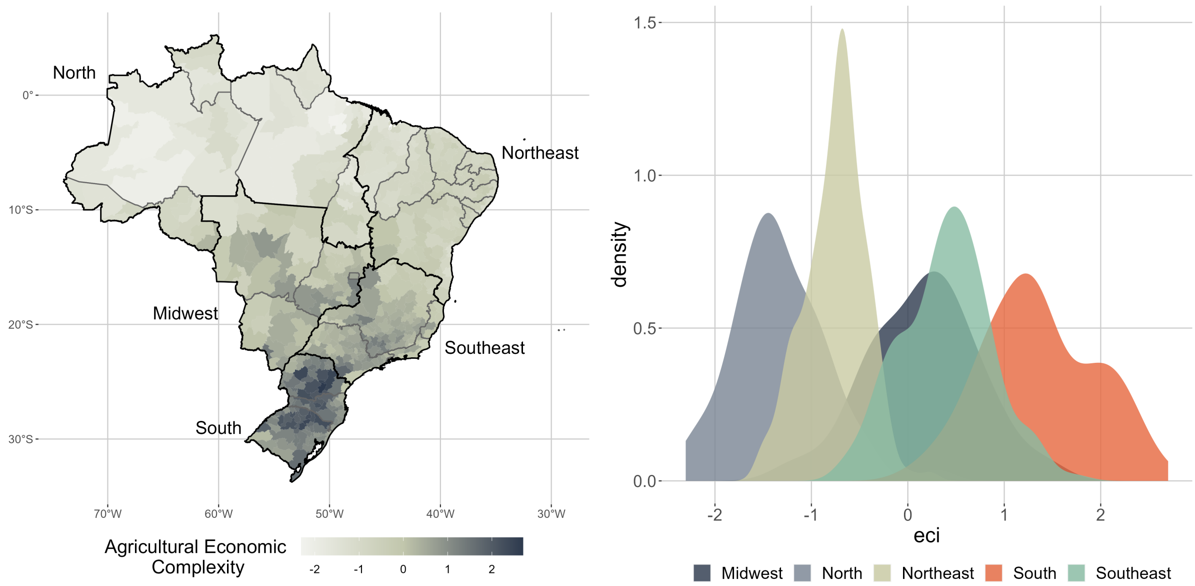
The original product space and complexity index are calculated using exclusive data on global and country exports. To paint a general picture, as discussed before, this approach has been very successful in ranking countries as well as correlating economic complexity indexes with a wide variety of socioeconomic indicators. However, as the analysis moves from a more macro perspective to a subnational (regional) level, the network of the product space and the index of economic complexity based on trade data lose utility. The fact that the vast majority of goods and services produced actually address local markets rather than foreign one introduce a significant bias to the analysis. Hence, the use of trade data only tends to underestimate the level of economic complexity of countries in which the production is mainly focused on internal consumption, as is the Brazilian case.

There are consistent studies using employment data to analyze economic complexity at the regional level, but considering the lack of information about products besides the export databases, there are no studies from a regional perspective using product data. This means that it is still a challenge to understand if the product space is a useful tool to map and represent the productive structure at sectoral and regional levels. This is particularly relevant for studies that go beyond manufacturing to include agriculture and services, as the level of employment informality is much higher. For these sectors, even the use of employment data to calculate economic complexity indexes at the firm level still tends to render biased results.

4.1 Regional agricultural economic complexity

The agricultural complexity index depicts the very high levels of productive heterogeneity across the country. The right side of Figure 1 shows a gradient of complexity levels in the map for each region. It is clear that, even though a very clear north-south pattern arises, the depiction of the structural heterogeneity of the sector as low and high-productivity goods is very limited. There is a significantly wider range of levels when economic complexity is taken into consideration. This is true for the country as a whole and, in particular, for regions within different states even the higher complexity ones in the southern part of the country.

Figure 1 – Agricultural Economic Complexity: regional heterogeneity



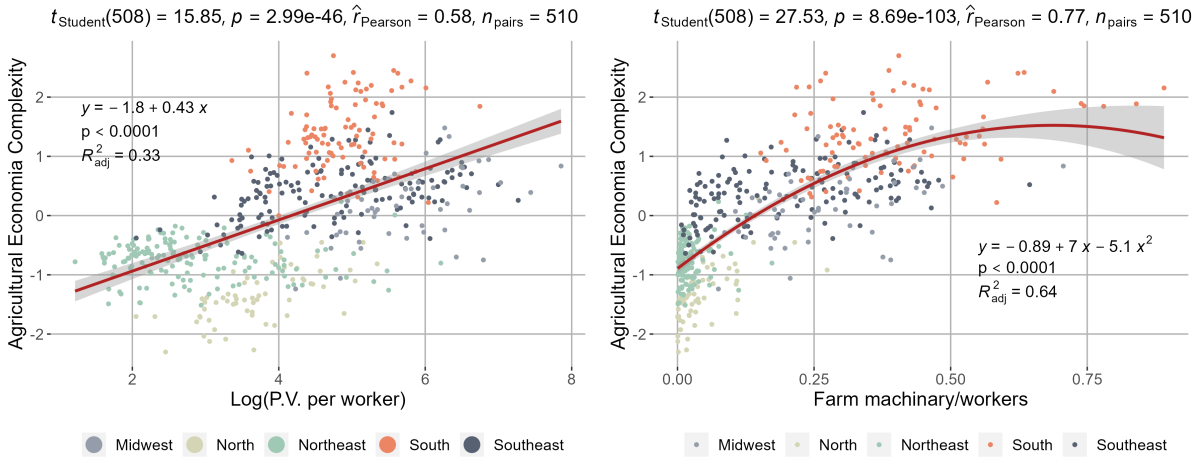
Source: authors' own elaboration based on Agricultural Census (IBGE).

The high levels of national and intra-regional heterogeneity can also be seen on the right side of Figure 1. The graph shows the density distribution of the agricultural complexity index. The country's macro-regions are significantly spread out in terms of economic complexity with the North and South very far apart. In addition, when each region is taken in isolation, significant levels of variability can be seen. The stark regional contrast is shown by the crossover between the tail end of the higher complexity production of the North and that of the lower complexity production of the South.

The definition of structural heterogeneity passes by high productivity differentials in the same economy and the agricultural economic complexity index captures this essential information. As is shown in Figure 2, the ACI has a significant correlation with the indicator of production value per worker.

Figure 2 below depicts the vast regional heterogeneity between and within groups.

Figure 2 – Regional heterogeneity: value-added and machinery per worker



Source: authors' own elaboration based on Agricultural Census (IBGE).

Very interesting patterns can be seen. In general, a positive association between the level of Complexity in the agricultural sector can be seen both in terms of value-added and machinery per worker. However, this general trend hides other trends both within and between regions. For instance, the most complex region (South) shows, at the same time, high levels of value-added per worker and a wide range of machinery per worker. The least complex region (North) shows low levels of value-added and machinery per worker.

4.2 The Agricultural Product Space and product economic complexity

The Agricultural Product Space allows to associate products' characteristics with regional specialization patterns. From a regional perspective, the product space is a network that carries information about the similarities between economic activities within the country in consideration.

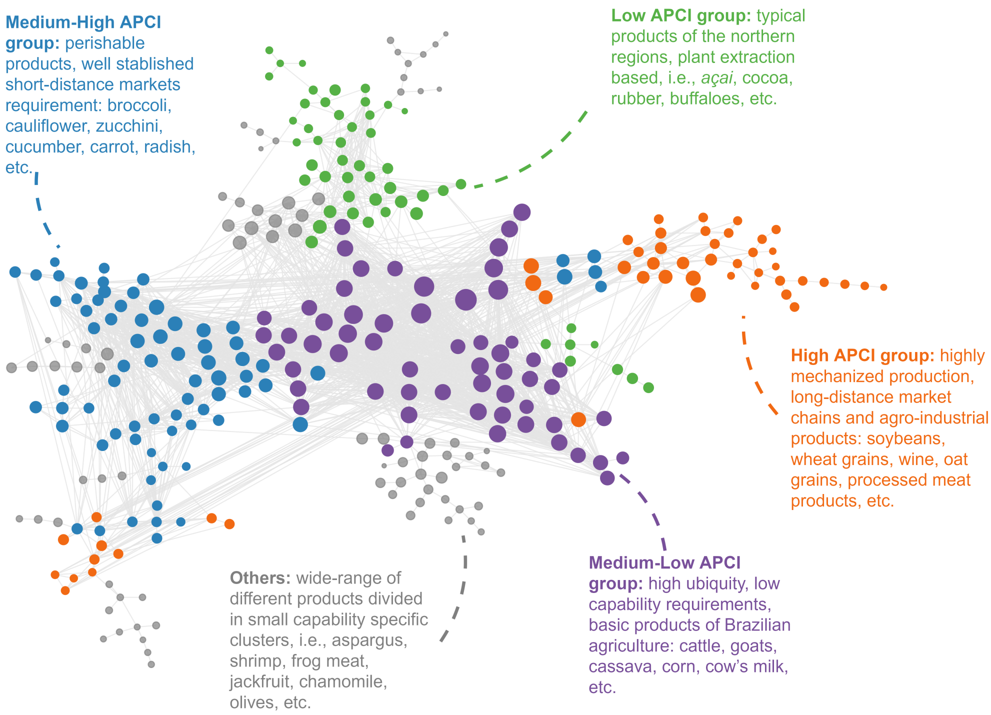
Similar activities are expected to be closer in the product space. Considering that the proximity measure is calculated using a co-location criteria, it is expected that clusters in the product space will reflect production patterns in groups of regions. This characteristic allows us to identify production patterns across regions and provides a more detailed characterization of the productive structure than the usual methods in agriculture studies, which use only the main product of the region or an *a priori* classification of products.

The Brazilian Agricultural Product Space has 312 nodes representing products and 2336 undirected edges connecting the nodes by similarity criteria with no isolated components. The clustering algorithm allowed the identification of 12 groups based on the similarity of products. Figure 3 shows the representation of the complete product space with colors representing clusters and node sizes representing products' ubiquity [9]. Four groups account for almost 75%

[9] A vectorized image of product space with product labels - which can be deeply zoomed - can be found in

of the products. These groups will be the focus of our analysis[10].

Figure 3 – Agricultural Product Space



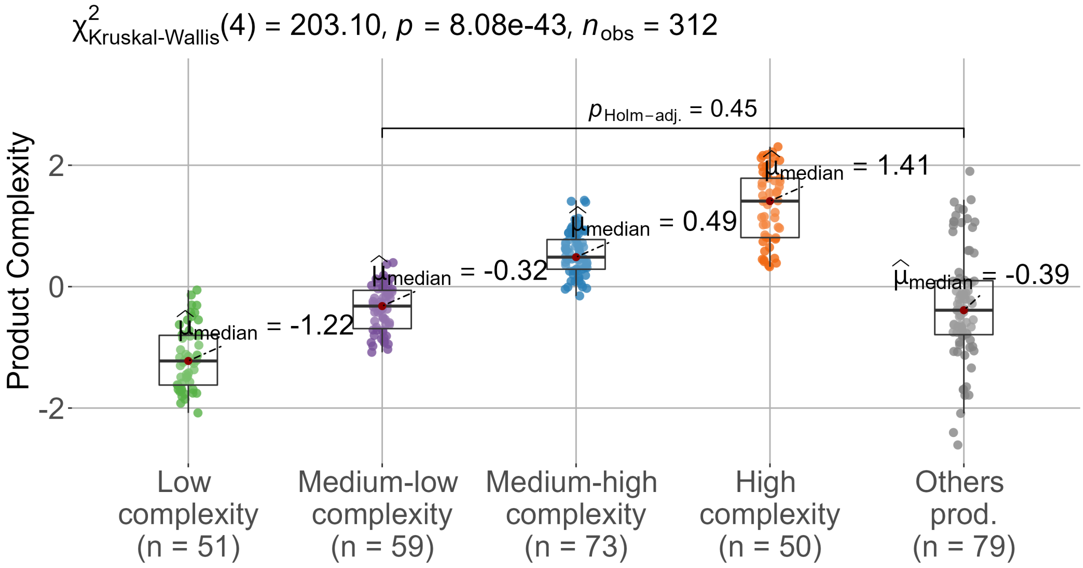
Source: authors' own elaboration based on Agricultural Census (IBGE). Complete agricultural product space, nodes = 312, edges = 2336, size = products ubiquity and colors = clusters

The four biggest groups of products provide relevant information regarding the levels of structural heterogeneity in Brazilian agriculture. For clarity, each group was named after their level of product complexity, given by the Agricultural Product Complexity Index (APCI) we estimated. We called them Low, Medium-Low, Medium-High and High complexity groups. As can be seen in Figure 4, the groups have statistically significant differences in their APCI levels.

the appendix for better understanding of which products are in each group.

10 Number of products per cluster are shown in table 1.

Figure 4 – Product complexity boxplot by clusters.



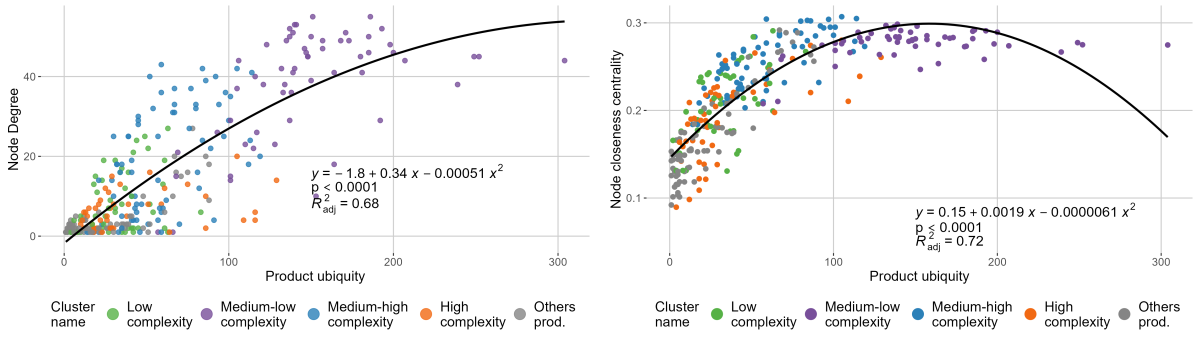
Source: authors' own elaboration based on Agricultural Census (IBGE).

The product space representation points to an important result: more ubiquitous products - which tend to be less complex - are in the center of product space. From a network perspective, this result makes sense, because ubiquitous products are present in more regions and thus they are produced along with a bigger number of other products, which means more connections of the ubiquitous products and bigger centrality measures for its representing nodes as a consequence.

For instance, one of the most ubiquitous products - cattle - which is made with comparative advantage in the 304 regions, will have connections to all other products made in these 304 regions and these connections will bring it to the center of product space.

Low ubiquity products such as artichoke - which is produced only in four regions - have few connections with other products and tend to be way from the center. This perspective gets stronger when considering that the lower-weight connections were removed from the network. Their presence would reinforce the centrality of low-ubiquity products if present in the network.

Figure 5 – Product complexity level and ubiquity



a) Node degree and ubiquity

b) Node closeness and ubiquity

Source: authors' own elaboration based on Agricultural Census (IBGE).

These relations can be visualized in plots (a) and (b) of figure 5, which brings the relation between network centrality indicators and ubiquity [11]. The centrality level of products in the product space is an important variable for assessing diversification paths that regions can take. Central products are more similar to all other products than peripheral products, and being more similar means that departing from them it is easier to diversify into any other products.

The most central group in the agricultural product space is the Low-Medium Complexity group. It is the most ubiquitous of the groups and, as we will explore ahead, contains the most basic products of Brazilian food production. In other words, the centre of the product space contains the goods that most regions produce and this indicates that departing from these central common groups of products, regions can diversify towards products with higher or lower levels of complexity.

4.3 Complexity Groups: regional comparative advantages and networks

Due to the co-location criteria used to compute the similarity between products - where products are closer if they are produced in the same regions - the internal homogeneity of groups in terms of complexity alone points to three important characteristics already mapped in the economic complexity literature:

(i) The agricultural product space reflects regional specialization patterns in which regions producing a low-complexity product also produce other low-complexity products rather than combine low and high-complexity products;

(ii) Producing less complex products tends to make it easier to diversify the production into other low complex products over time rather than towards higher complexity products;

[11] It is important to say it goes in the opposite way from the findings of Hidalgo et al. (2007) and Hausmann et al. (2014), where more complex products - which tend to be less ubiquitous - are in the center of the graph.

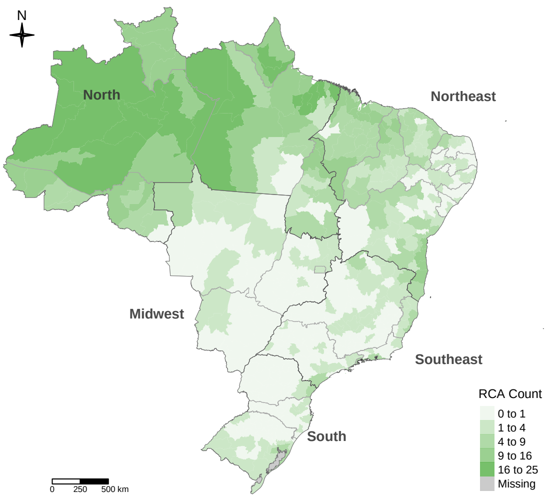
(iii) Due to the previous items, a "hard-to-change" core-periphery productive structure can be identified, where some regions account for complex products and others for less complex products.

Hence, it becomes clear that product clusters are connected to groups of regions by definition and carry information about their characteristics.

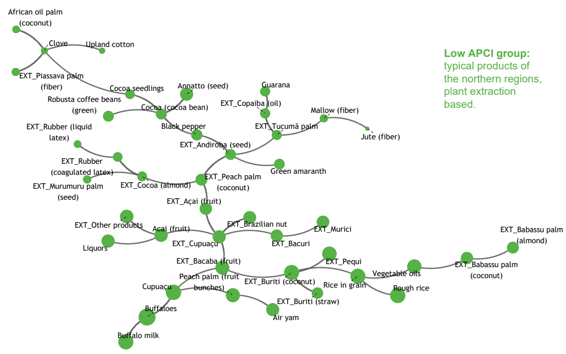
The Low-Complexity group contains typical products of the northern regions, from the native forest, like babaçu, açaí, pequi, cocoa, latex (rubber), bacuri, etc. Some of them, like açaí, cocoa and latex are products of high historical importance for the sociopolitical formation of states like Acre, Amazonas and Pará. It's interesting to note that these products are mostly plant extraction, which is usually very labour-intensive. Yet, they require very specific capabilities for their production given by edaphoclimatic conditions. For this reason, these products are less ubiquitous than those of the other groups. However, like diamonds, their low ubiquity is not caused by the requirement of extraordinary productive knowledge. In addition, they are produced in less diversified regions, which explains their low levels of complexity.

It is interesting to visualize the regional diversity of realities within each cluster. In Figure 6, the left side depicts the regional specialization of the Low-Complexity group. It is particularly strong in the North, which is mostly covered by the Amazon forest, as well as the Northeast of the country. The right side of the figure shows the product network of the group. It shows the relationship between the culture of products within the group, which is mostly composed of extractive production.

Figure 6 – Low complexity products spatial distribution



a) RCA Regional distribution.



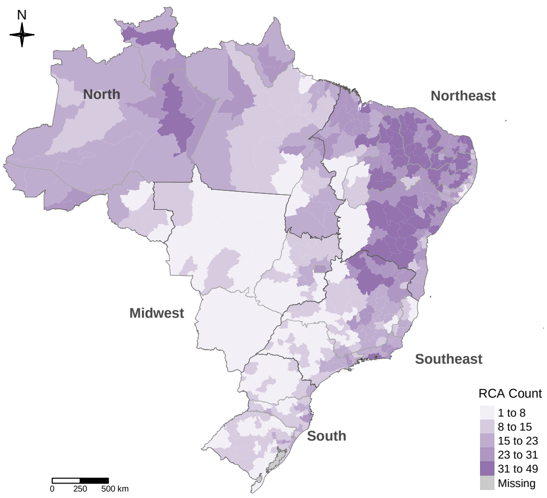
b) Low APCI cluster, size = ubiquity.

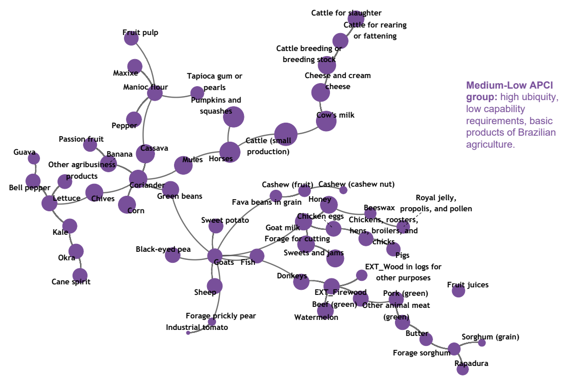
Source: authors' own elaboration based on Agricultural Census (IBGE). Note: The prefix "EXT" in the labels indicates plant extraction products, in order to differentiate from the same products obtained by cultivation.

Medium-low complexity products are the most ubiquitous group. Although these products can be found in all Brazilian regions, they are more relevant for the northeastern regions. It is possible to note that the group includes historically important products considering the country's development, such as: Bovine cattle, mules, milk, chicken, cassava, beans, fresh corn on the cob and basic garden products like lettuce, collard, tomato, and bell pepper. All of these products have low capabilities requirements and can be found all along the Brazilian territory. Their predominance in the Northeast and North regions is indicative that these regions did not make the transition towards a more diversified economy and more sophisticated products.

Figure 7, left side, depicts the regional specialization of the Medium-low Complexity. Its relative specialization is more pronounced in the poorer regions, such as the Northeast and North. The right side of the figure shows the product network of this group. Within the network, there are still products from extractive production. However, farm animals and more specialized cultures of traditional goods are more relevant, although relatively ubiquitous.

Figure 7 – Medium-low complexity products spatial distribution





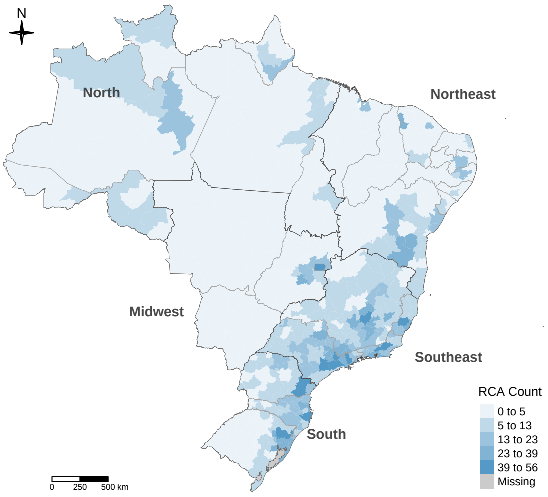
b) Medium-low APCI cluster, size = product ubiquity.

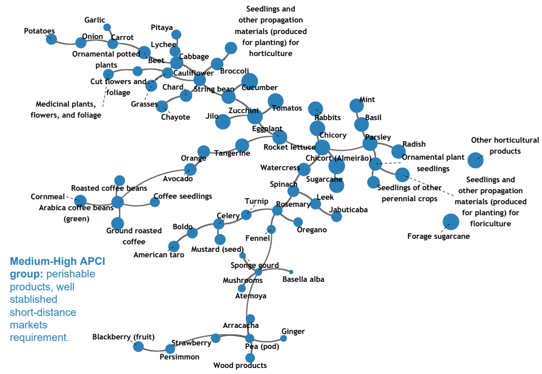
a) RCA Regional distribution.

Source: authors' own elaboration based on Agricultural Census (IBGE).

The Medium-high and High-complexity groups of products have very specific spatial distributions. The first one is located mainly near the largest Brazilian urban agglomerations. The logic of the spatial distribution is related to the higher perishability of the goods within this group. It includes with not so usual garden products such as broccoli, cauliflower, zucchini, cucumber, carrot, radish, endive, arugula etc. Those products do not demand high technical capabilities to be produced on a low scale, but in order for a region to gain comparative advantages in its production, a much higher scale, well-established market channels, and proper levels of demand are necessary. Besides, many of these products demand investment in controlled environments (greenhouses) and fertilizers. Hence, even though there is an apparent degree of similarity between Medium-Low and Medium-High products, the latter includes

Figure 8 – Medium-high complexity products spatial distribution





b) Medium-high APCI cluster, size = product ubiquity.

a) RCA Regional distribution.

Source: authors' own elaboration based on Agricultural Census (IBGE).

products characterized by the need for higher scale and market-integrated production, while the first one includes lower scale and often subsistence production.

Figure 8 shows the regional distribution of the relative specialization of the Medium-High group. It closely follows the population distribution across the country, with the exception of the Brazilian midwest. The right side of the figure shows the product network of this group.

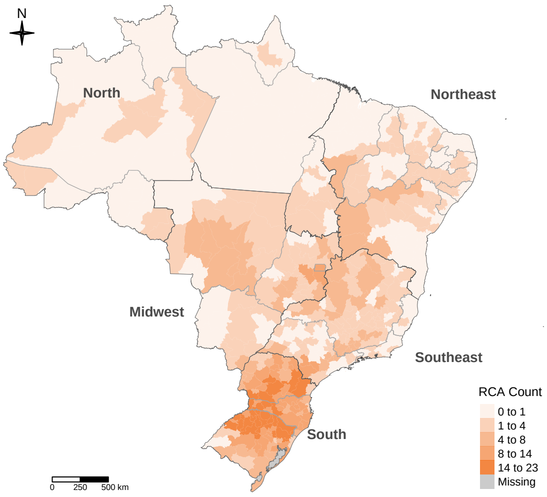
Within the network, a large amount of fresh fruits and vegetables can be found, with the location of specific goods following related types of cultures. However, more complex and more highly mechanized production can be found such as coffee, sugar cane and wood extraction.

Finally, the High Complexity group of products is clearly relatively more concentrated in the South region. This group includes the highly mechanized type of production, which nowadays is represented by the production of seeds such as soybean, corn, rye and oats. This type of production is not limited to the Southern regions, having important presence in the Midwest region and in the *Matopiba* region[12].

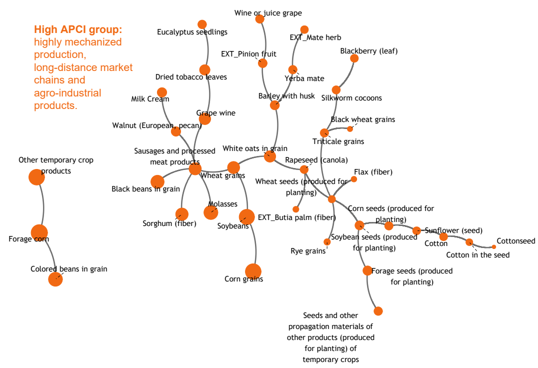
The group also includes typical southern products (such as *erva-mate*, grapes and tobacco)

[12] *Matopiba* is composed by regions of the northern and northeastern states of Maranhão (MA), Tocantins (TO), Piauí (PI) and Bahia (BA) where it is possible to identify the rapidly growth of soybean production in the last decade.

Figure 9 – High complexity products spatial distribution



a) RCA Regional distribution.



b) High APCI cluster, size = product ubiquity.

Source: authors' own elaboration based on Agricultural Census (IBGE).

which depend on climatic conditions but have highly developed production chains and result in important manufactured goods. The group also includes several agro-industrial products like breads, sausages, and milk creams. These products are produced in rural establishments and commercialized in the local markets, which is compatible with the market-integrated family farming that characterizes an important part of these regions. It is also possible to notice the presence of fruits such as apples, pears, kiwis, plums, and peaches, which demand several not trivial techniques in their production process.

In sum, this group includes higher-tech agricultural products that are part of longer production and market chains, all of which demand the combination of several different kinds of capabilities and justify their complexity levels.

Figure 9, shows the regional distribution of the relative specialization of the High complexity group. The right side of the figure shows the regional pattern of relative advantages. The higher complexity production is concentrated in the South. However, given that higher scale production of very relevant export products, such as soybeans, corn, and cotton, large areas in both the northeast and midwest are also important. On the right side, the product network for this group shows a degree of fragmentation due to regional specialization.

A careful analysis of the clusters allows us to understand how they represent the Brazilian agricultural productive structure and characterize both low and high-economic complexity regions, offering a perspective of the core-periphery structure that can be found in the country. It allows us to link regions where agriculture is not sophisticated in technical terms - like North and Northeast - to the kind of products that are important in these places and visualize what kind of products the more sophisticated regions - like South and Southeast - diversified into to reach higher development levels.

In addition, knowing the similarity between products allows us to find more feasible paths for productive diversification towards more complex products and so enhance regional economic development regarding the particularities of each region. It is an important contribution to draw policies for the agricultural sector and do not repeat the mistake made by the proposals of the green revolution which implied the homogenization of agricultural production aiming to improve productivity levels. Choosing the already existing productive knowledge and conditions of the region as a starting point for diversification, the economic complexity perspective for development necessarily values the local diversity in the development process.

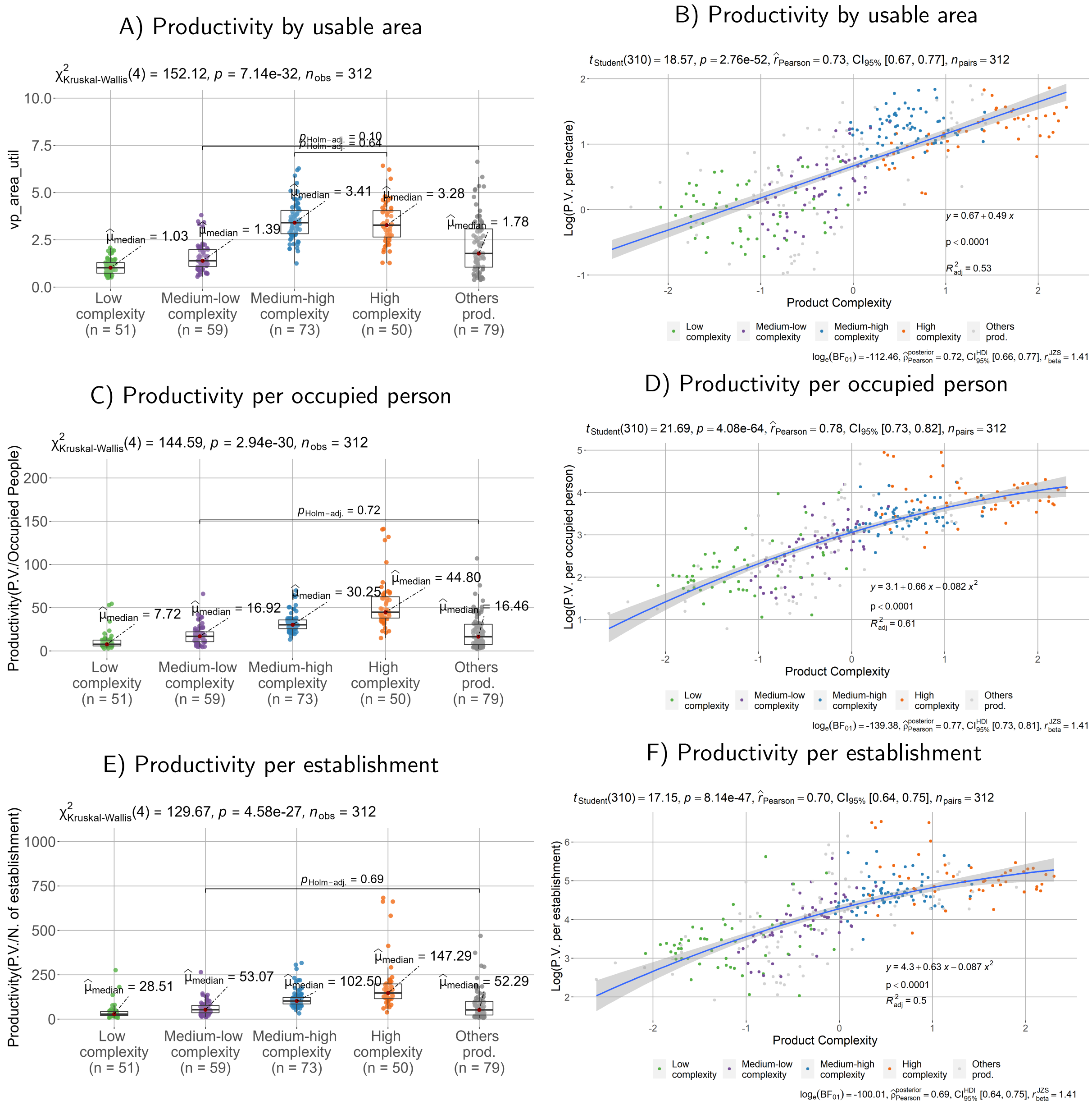
4.4 Economic complexity and structural heterogeneity in agriculture

The combination of the product complexity indexes, product space and the spatial distribution of the product clusters allows us to characterise the agricultural productive structure and offers an interesting perspective for productive development in agriculture. The concept of structural heterogeneity, however, includes the assessment of the productivity associated with the different groups of products and their complexity level differences.

The most important evidence of how the economic complexity approach can be used to characterize Brazilian structural heterogeneity in agriculture is the difference in the product groups present in region-productivity levels associated with them.

The higher the product complexity, the higher the productivity of the regions where they are produced. Medium-high and High complexity groups of products show significant statistical differences from Medium-low and Low-complexity groups in terms of productivity per area, productivity per establishment and work productivity. It suggests that changes in the productive structure through diversification towards higher complexity products are associated with higher productivity levels.

Figure 10 – Relation between productivity and product complexity

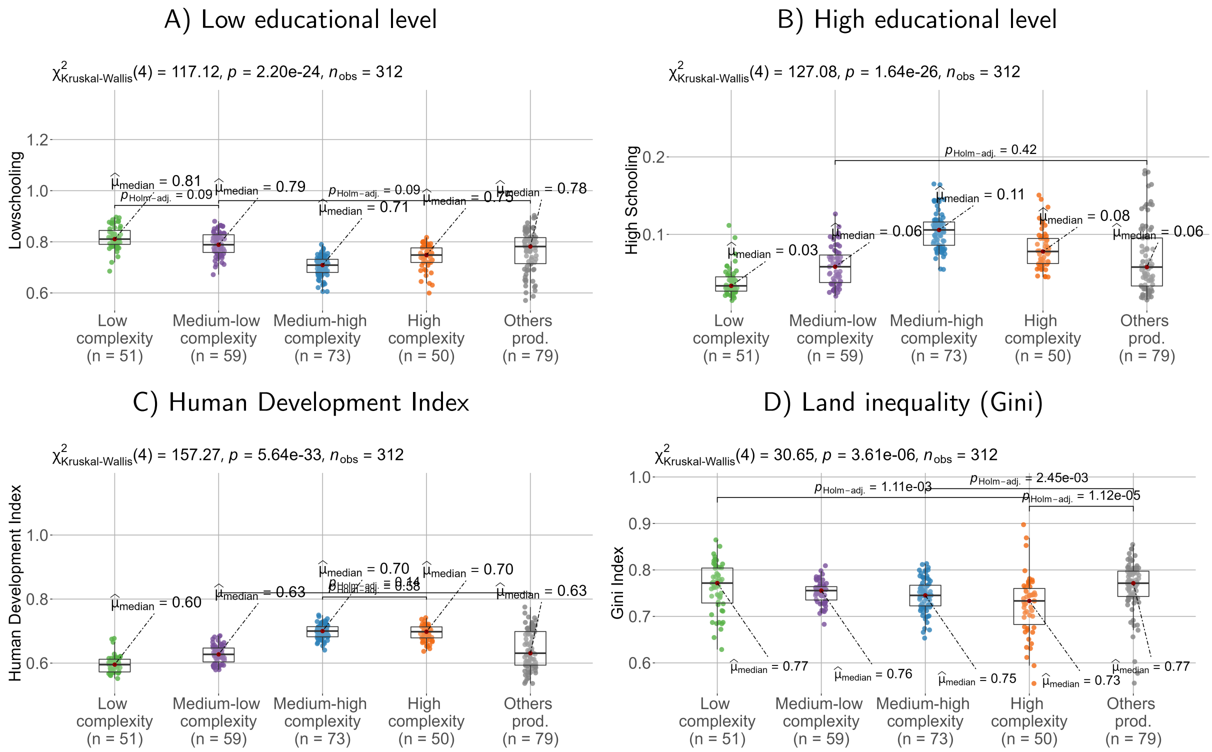


Source: authors' own elaboration based on Agricultural Census (IBGE). Note: Boxplot shows where group median comparisons do not indicate significant statistical differences.

These results show that different kinds of production are associated with different productivity levels. It corroborates the classic structuralist perspective that closing the gap between developed and underdeveloped economies demands modification in productive structure toward sectors capable of diffuse technical progress. In the economic complexity literature this vision is marked by the motto *what you export matters* (HIDALGO et al., 2007). Here we advocate that inside a specific sector like agriculture, *what you produce matters* to improve productivity levels.

Besides the economic dimension, we found out that more complex agriculture products are also associated with higher levels of education and human development. Products present in Medium-high and High complexity groups show significant differences in the educational levels of the property managers when compared to Low and Medium-low complexity groups. The same happens when testing for statistical differences in human development indexes of the regions where products are made. These findings demand econometric exercises in order to assess the causality relations between the variables, opening a research agenda focused on intersectional agriculture relations.

Figure 11 – Low and medium educational levels



Source: authors' own elaboration based on Agricultural Census (IBGE).

In highly structurally heterogeneous countries like Brazil, the traditional proposition that developing industries and high-tech innovation is the way to break the underdevelopment conditions found a barrier to the feasibility of promoting structural changes in less developed regions. Even the economic complexity perspective – which considers the feasibility aspect – does not account for proper tools to deal with feasible diversification paths in regions specialized in the agricultural sector. The exercise presented here solves these issues, providing ways to plan productive diversification policies for agriculture-based regions.

5. Concluding Remarks

In this paper, we contributed to the literature on structural heterogeneity in the agricultural sector by assessing the issue using the perspective of economic complexity. In doing so, the paper has two important contributions. First, we showed that economic complexity methods can be transposed and applied to disaggregated data on (regional) agricultural production. In doing so, the paper presented a novel the Agricultural Economic Complexity Index (ACI), calculated the Agricultural Product Space, and tested for the relationship between the new index and relevant socioeconomic variables.

By calculating the ACI, it was possible to show very high that the methodology can capture the very high levels of structural heterogeneity of the sector in Brazil. High complexity levels can be found in the south and southeastern regions of Brazil, whereas the Northeast, Midwest and North are characterized by less complex production. Moreover, given the level of disaggregation, it was also possible to represent a gradient of complexity in the sector across macro-regions and states.

By drawing the Agricultural Product Space it was possible to show that the production of related products tends to form clusters which can be interpreted and characterized by having different levels of complexity. Here, the method showed its strength, given that to build the network to probability of co-production is central. Hence, it is not only a tool to paint a general picture of the country's agricultural production but also a tool that can be used to inform the design of policies.

The second contribution of the paper is to show that, as is the case with the Economic Complexity Index, the Agricultural Complexity Index can be used as an important tool to identify other issues which are related to economic heterogeneity, and which are as varied across regions. This exercise, beyond being a test for the index, is useful to identify critical socioeconomic issues which are closely related to the nature of agricultural production, especially in poorer, more isolated areas. We found the agricultural complexity is closely related to important structural and socioeconomic variables. The clusters, ranked by levels of complexity, are positively correlated with productivity by usable area, per occupied person, and per establishment. In addition, there is a positive relationship between the level of economic complexity and levels of education and levels of the Human Development Index.

The policy implications of the use of this methodology are straightforward and can be extrapolated by the richness of applications of economic complexity in the literature and covered by Hidalgo (2021) and Hidalgo et al. (2018). It is possible to use complexity indicators, information from the product space, and relatedness indicators to use large amounts of data to help design development policies. They can be used to identify how goods that are produced in high-complexity regions are associated with positive indicators (such as higher added value, lower environmental impact or even better socioeconomic characteristics) in order to suggest paths for economic diversification for poorer areas, for instance.

There are also caveats and limitations with the application that suggest a fruitful research agenda. First, the publicly available Census data used in the exercises is aggregated by regions. This implies a loss of information on the unit of production which would allow a much better view of the issues dealt with in this paper. Hence, access to the raw data would provide better results which in turn could be used to inform policies. Secondly, the time spam between censuses, whereas appropriate to capture longer-term changes in the characteristics of agricultural production, is much less suitable to follow trends in said production and their association with other variables. In this case, the general trends found with the census data would have to inform the use of less comprehensive data, such as those found in more frequent surveys. In either case, the proof of concept shown here opens interesting possibilities to bring the crucial discussion of the relevance of the development of agricultural production for more general economic development across regions.

Bibliography

ALABDULKAREEM, Ahmad et al. Unpacking the polarization of workplace skills. *Science advances*, American Association for the Advancement of Science, v. 4, n. 7, p. eaao6030, 2018.

BALASSA, Bela. Trade liberalisation and revealed comparative advantage 1. *The manchester school*, Wiley Online Library, v. 33, n. 2, p. 99–123, 1965.

BALLAND, Pierre-Alexandre et al. Smart specialization policy in the european union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, Taylor & Francis, v. 53, n. 9, p. 1252–1268, 2019.

BALLAND, Pierre Alexandre et al. Complex economic activities concentrate in large cities. 2018.

BALLAND, Pierre-Alexandre; RIGBY, David. The geography of complex knowledge. *Economic Geography*, Taylor & Francis, v. 93, n. 1, p. 1–23, 2017.

BELIK, Walter. A heterogeneidade e suas implicações para as políticas públicas no rural brasileiro. *Revista de Economia e Sociologia Rural*, SciELO Brasil, v. 53, n. 1, p. 9–30, 2015.

BOSCHMA, Ron; BALLAND, Pierre-Alexandre; KOGLER, Dieter Franz. Relatedness and technological change in cities: the rise and fall of technological knowledge in us metropolitan areas from 1981 to 2010. *Industrial and corporate change*, Oxford University Press, v. 24, n. 1, p. 223–250, 2015.

BRASIL, Instituto Brasileiro de Geografia e Estatística. *Censo agropecuário 2017: Resultados definitivos*. Rio de Janeiro, 2019.

BRITTO, Gustavo et al. The great divide: economic complexity and development paths in brazil and the republic of korea. *CEPAL Review*, 2019.

CAN, Muhlis; GOZGOR, Giray. The impact of economic complexity on carbon emissions: evidence from france. *Environmental Science and Pollution Research*, Springer, v. 24, n. 19, p. 16364–16370, 2017.

CARDILLE, Jeffrey A; FOLEY, Jonathan A. Agricultural land-use change in brazilian amazonia between 1980 and 1995: Evidence from integrated satellite and census data. *Remote Sensing of Environment*, Elsevier, v. 87, n. 4, p. 551–562, 2003.

CHÁVEZ, Juan Carlos; MOSQUEDA, Marco T; GÓMEZ-ZALDÍVAR, Manuel. Economic complexity and regional growth performance: Evidence from the mexican economy. *Review of Regional Studies*, v. 47, n. 2, p. 201–219, 2017.

CHINAZZI, Matteo et al. Mapping the physics research space: a machine learning approach. *EPJ Data Science*, Springer Berlin Heidelberg, v. 8, n. 1, p. 33, 2019.

DEPAULA, Guilherme. The distributional effect of climate change on agriculture: Evidence from a ricardian quantile analysis of brazilian census data. *Journal of Environmental Economics and Management*, Elsevier, v. 104, p. 102378, 2020.

ELLIS, Frank. The determinants of rural livelihood diversification in developing countries. *Journal of agricultural economics*, Wiley Online Library, v. 51, n. 2, p. 289–302, 2000.

FAWAZ, Fadi; RAHNAMA-MOGHADAMM, Masha. Spatial dependence of global income inequality: The role of economic complexity. *The International Trade Journal*, Taylor & Francis, v. 33, n. 6, p. 542–554, 2019.

FELEMA, João; RAIHER, Augusta Pelinski; FERREIRA, Carlos Roberto. Agropecuária brasileira: desempenho regional e determinantes de produtividade. *Revista de Economia e Sociologia Rural*, SciELO Brasil, v. 51, n. 3, p. 555–573, 2013.

FERREIRA, Marcelo Dias Paes; FÉRES, José Gustavo. Farm size and land use efficiency in the brazilian amazon. *Land Use Policy*, Elsevier, v. 99, p. 104901, 2020.

FORNAZIER, Armando; VIEIRA-FILHO, José Eustáquio Ribeiro. *Heterogeneidade estrutural no setor agropecuácio brasileiro: Evidências a partir do censo agropecuário de 2006*. Rio de Janeiro, 2012.

FREITAS, Elton Eduardo. *Indústrias relacionadas, complexidade econômica e diversificação regional: uma aplicação para microrregiões brasileiras*. Tese (Doutorado) — Centro de Desenvolvimento e Planejamento Regional (UFMG), Belo Horizonte, 2019.

GASQUES, José Garcia et al. Produtividade total dos fatores e transformações da agricultura brasileira: análise dos dados dos censos agropecuários. *A agricultura brasileira: desempenho, desafios e perspectivas. Brasília: Ipea*, p. 19–44, 2010.

GOMEZ-LIEVANO, Andres; PATTERSON-LOMBA, Oscar. Estimating the drivers of urban economic complexity and their connection to economic performance. *Royal Society open science*, The Royal Society, v. 8, n. 9, p. 210670, 2021.

HARTMANN, Dominik et al. Linking economic complexity, institutions, and income inequality. *World Development*, Elsevier, v. 93, p. 75–93, 2017.

HAUSMANN, Ricardo et al. *The atlas of economic complexity: Mapping paths to prosperity*. Cambridge, MA: Mit Press, 2014.

HIDALGO, Cesar. *Why information grows: The evolution of order, from atoms to economies*. [S.l.]: Basic Books, 2015.

HIDALGO, César A. Economic complexity theory and applications. *Nature Reviews Physics*, Nature Publishing Group, v. 3, n. 2, p. 92–113, 2021.

HIDALGO, César A et al. The principle of relatedness. In: SPRINGER. *Unifying Themes in Complex Systems IX: Proceedings of the Ninth International Conference on Complex Systems 9*. [S.l.], 2018. p. 451–457.

HIDALGO, César A; HAUSMANN, Ricardo. The building blocks of economic complexity. *Proceedings of the national academy of sciences*, National Acad Sciences, v. 106, n. 26, p. 10570–10575, 2009.

HIDALGO, César A; HAUSMANN, Ricardo. Supporting material for: the building blocks of economic complexity. *Working paper*, Center for International Development and Harvard Kennedy School, Harvard University, 2009.

HIDALGO, César A et al. The product space conditions the development of nations. *Science*, American Association for the Advancement of Science, v. 317, n. 5837, p. 482–487, 2007.

JUNIOR, Wladimir Colman de Azevedo; RODRIGUES, Marcos; SILVA, David Costa Correia. Does agricultural efficiency contribute to slowdown of deforestation in the brazilian legal amazon? *Journal for Nature Conservation*, Elsevier, v. 65, p. 126092, 2022.

LALL, Sanjaya. The technological structure and performance of developing country manufactured exports, 1985-98. *Oxford development studies*, Taylor & Francis, v. 28, n. 3, p. 337–369, 2000.

LAPATINAS, Athanasios; LITINA, Anastasia; ZANAJ, Skerdilajda. The impact of economic complexity on the formation of environmental culture. *Sustainability*, Multidisciplinary Digital Publishing Institute, v. 13, n. 2, p. 870, 2021.

MARCONATO, Renata; LAROCCA, Ana Paula C; QUINTANILHA, José Alberto. Análise do uso de tecnologias em estabelecimentos agropecuários por meio dos índices de moran global e local. *Revista de Política Agrícola*, v. 21, n. 1, p. 5–21, 2012.

MONTRESOR, Sandro; QUATRARO, Francesco. Green technologies and smart specialisation strategies: a european patent-based analysis of the intertwining of technological relatedness and key enabling technologies. *Regional Studies*, Taylor & Francis, v. 54, n. 10, p. 1354–1365, 2020.

MUNEEPEERAKUL, Rachata et al. Urban economies and occupation space: Can they get "there" from "here"? *PloS one*, Public Library of Science San Francisco, USA, v. 8, n. 9, p. e73676, 2013.

NEAGU, Olimpia; TEODORU, Mircea Constantin. The relationship between economic complexity, energy consumption structure and greenhouse gas emission: Heterogeneous panel evidence from the eu countries. *Sustainability*, Multidisciplinary Digital Publishing Institute, v. 11, n. 2, p. 497, 2019.

NEFFKE, Frank; HENNING, Martin; BOSCHMA, Ron. How do regions diversify over time? industry relatedness and the development of new growth paths in regions. *Economic geography*, Taylor & Francis, v. 87, n. 3, p. 237–265, 2011.

NEVES, Mateus de Carvalho Reis et al. The role of cooperatives in brazilian agricultural production. *Agriculture*, MDPI, v. 11, n. 10, p. 948, 2021.

OLIVEIRA-FERREIRA, Monaliza de; VASCONCELOS, Kelly Samá Lopes de. Crescimento e especialização produtiva da agropecuária entre estados do nordeste brasileiro. *Economía, sociedad y territorio*, El Colegio Mexiquense AC, v. 14, n. 46, p. 799–822, 2014.

OLIVEIRA, Kessys Lorrânya Peralta de et al. Endogeneity in stochastic production frontier with one and two-step models: an application with municipal data from brazilian agricultural census. *Pesquisa Operacional*, SciELO Brasil, v. 42, p. e243504, 2022.

PERONDI, Miguel Angelo; SCHNEIDER, Sérgio. Bases teóricas da abordagem de diversificação dos meios de vida. *REDES: Revista do Desenvolvimento Regional*, Universidade de Santa Cruz do Sul, v. 17, n. 2, p. 117–135, 2012.

PINTO, Aníbal. Naturaleza e implicaciones de la" heterogeneidad estructural" de la américa latina. *El trimestre económico*, JSTOR, v. 37, n. 145 (1), p. 83–100, 1970.

PLOEG, Jan Douwe Van der; YE, Jingzhong; SCHNEIDER, Sergio. Reading markets politically: on the transformativity and relevance of peasant markets. *The Journal of Peasant Studies*, Taylor & Francis, p. 1–26, 2022.

RADA, Nicholas; HELFAND, Steven; MAGALHÃES, Marcelo. Agricultural productivity growth in brazil: Large and small farms excel. *Food policy*, Elsevier, v. 84, p. 176–185, 2019.

ROMERO, João P; GRAMKOW, Camila. Economic complexity and greenhouse gas emissions. *World Development*, Elsevier, v. 139, p. 105317, 2021.

SAMBUICHI, Regina Helena Rosa et al. A diversificação produtiva como forma de viabilizar o desenvolvimento sustentável da agricultura familiar no brasil. *Brasil em desenvolvimento*, p. 61–84, 2014.

SAMBUICHI, Regina Helena Rosa et al. Diversidade da produção nos estabelecimentos da agricultura familiar no brasil: uma análise econométrica baseada no cadastro da declaração de aptidão ao pronaf (dap). Instituto de Pesquisa Econômica Aplicada (Ipea), 2016.

SANTOS, Gesmar Rosa dos; VIEIRA-FILHO, José Eustáquio Ribeiro. Heterogeneidade produtiva na agricultura brasileira: elementos estruturais e dinâmicos da trajetória produtiva recente. Instituto de Pesquisa Econômica Aplicada (Ipea), 2012.

SBARDELLA, Angelica; PUGLIESE, Emanuele; PIETRONERO, Luciano. Economic development and wage inequality: A complex system analysis. *PloS one*, Public Library of Science San Francisco, CA USA, v. 12, n. 9, p. e0182774, 2017.

SCHNEIDER, Sérgio. Reflexões sobre diversidade e diversificação-agricultura, formas familiares e desenvolvimento rural. *RURIS-Revista do Centro de Estudos Rurais-UNICAMP*, v. 4, n. 1, 2010.

SCHNEIDER, Sergio; CASSOL, Abel. A agricultura familiar no brasil. *Porto Alegre, Relatório de pesquisa. FIDA/RIMISP*, 2013.

SILVEIRA, José Maria da. Agricultura brasileira: o papel da inovação tecnológica. *In: O mundo rural no Brasil do século 21*, p. 373–394, 2014.

SOUZA, Paulo Marcelo de et al. Tecnologia na agricultura brasileira: Uma análise das desigualdades regionais para os segmentos familiar e não familiar. *Revista Econômica do Nordeste*, v. 49, n. 3, p. 147–169, 2018.

STOJKOSKI, Viktor; KOCH, Philipp; HIDALGO, César A. Multidimensional economic complexity and inclusive green growth. *Communications Earth & Environment*, Nature Publishing Group UK London, v. 4, n. 1, p. 130, 2023.

STOJKOSKI, Viktor; UTKOVSKI, Zoran; KOCAREV, Ljupco. The impact of services on economic complexity: Service sophistication as route for economic growth. *PloS one*, Public Library of Science San Francisco, CA USA, v. 11, n. 8, p. e0161633, 2016.

TACCHELLA, Andrea; MAZZILLI, Dario; PIETRONERO, Luciano. A dynamical systems approach to gross domestic product forecasting. *Nature Physics*, Nature Publishing Group, v. 14, n. 8, p. 861–865, 2018.

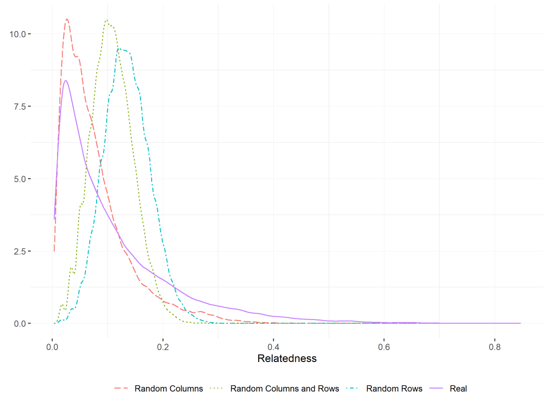
TRAAG, Vincent A; WALTMAN, Ludo; ECK, Nees Jan Van. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, Nature Publishing Group UK London, v. 9, n. 1, p. 5233, 2019.

VIEIRA-FILHO, José Eustáquio Ribeiro; FISHLOW, Albert. Agricultura e indústria no brasil: inovação e competitividade. Instituto de Pesquisa Econômica Aplicada (Ipea), 2017.

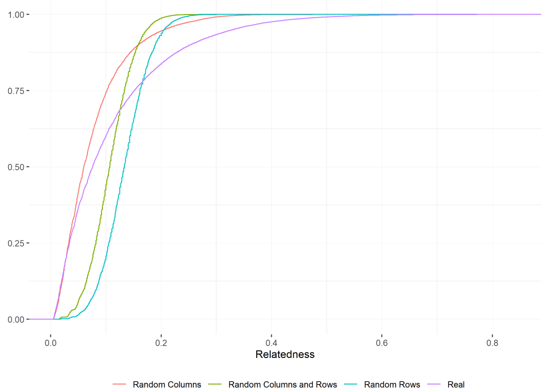
ZHU, Shengjun; YU, Changda; HE, Canfei. Export structures, income inequality and urban-rural divide in china. *Applied Geography*, Elsevier, v. 115, p. 102150, 2020.

A. Simulation results

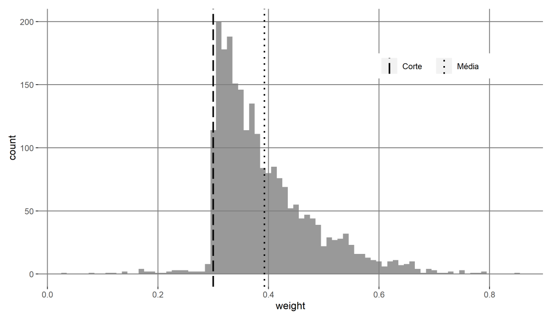
Figure 12 – Random graphs simulations results



a) Kernel density



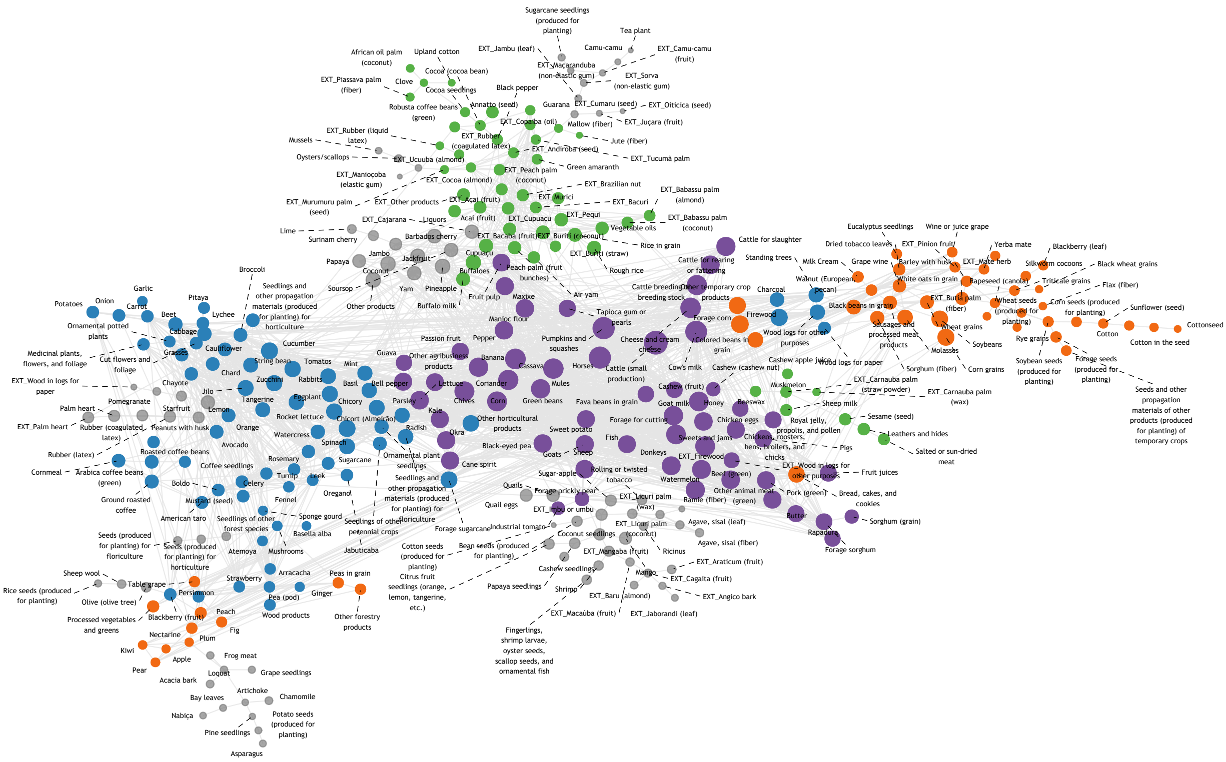
b) Cumulative frequency



c) Edges weight distribution

Source: authors' own elaboration based on Agricultural Census (IBGE).

Figure 13 – Vector image of the Agricultural Product Space with labels



Source: authors' own elaboration based on Brazilian Agricultural Census 2017 (IBGE). Note: The prefix "EXT" in the labels indicates plant extraction products, in order to differentiate from the same products obtained by cultivation.

Table 1 – Number of nodes per cluster

Cluster	n.nodes	perc.nodes
1	59	18.91
2	73	23.40
3	50	16.03
4	51	16.35
5	18	5.77
6	25	8.01
7	10	3.21
8	4	1.28
9	5	1.60
10	3	0.96
11	11	3.53
12	3	0.96