

RESEARCH

Open Access



Commuting times and the mobilisation of skills in emergent cities

Neave O'Clery^{1*} , Rafael Prieto Curiel^{1,2} and Eduardo Lora^{2,3}

*Correspondence:

oclery@maths.ox.ac.uk

¹Mathematical Institute, Oxford University, Andrew Wiles Building, Radcliffe Observatory Quarter, OX2 6GG Oxford, UK

Full list of author information is available at the end of the article

Abstract

Labour mobility within a large city or metropolitan area is a necessary condition for the optimal exploitation of agglomeration economies. We propose a method to establish which municipalities should be considered part of a metropolitan area based on labour market integration. In order to aggregate geographically proximate urban municipalities, we develop a network-based model that makes industry productivity (manifesting in formal employment creation) dependent on firms' ability to find, within city limits, the diversity of skills they need to move to new complex industries. In this way, we uncover the natural city scale at which firms optimally exploit the range of skills available to them within a broad catchment area. Considering Colombian cities, we find that commuting times between 45 and 75 min, corresponding to between 43 and 62 distinct cities or integrated labour markets, allow firms to maximise formal employment creation (between 2008 and 2013). This result supports the development of passenger transport limiting commuting times within cities, connecting small- and mid-size cities to nearby large cities, and coordinating transportation investments across traditional administrative boundaries.

Keywords: Complexity, City size, Commuting, Metropolitan areas, Informality, Networks, Colombia

Introduction

Large cities tend to make more productive use of available resources through the generation of agglomeration economies (Melo et al. 2009; Rosenthal and Strange 2004; Combes et al. 2010). Most mechanisms that give rise to agglomeration economies rest on some type of heterogeneity of workers or firms. "It is very difficult to conceive how interactions within an 'army of clones' could generate sufficient benefits to justify the existence of modern cities" (Duranton and Puga 2004a). Although heterogeneity in the skills of workers is traditionally difficult to measure, a recent study confirms a positive relationship between population size and the diversity of professions across US cities (Bettencourt LMA et al. 2014). Indeed a number of quantities related to productivity and innovation scale super-linearly with city size including wages, the number of patents (Bettencourt et al. 2007) and formal employment rates (O'Clery et al. 2016). However, the relationship between city size and economic growth is far from guaranteed, especially in emerging economies (McCann and Acs 2011). Alternative factors such as urban infrastructure and institutional capacity may be more important determinants of a city's ability

to generate economic growth (Camagni et al. 2015; Castells-Quintana 2017; Glaeser 2014; Frick and Rodríguez-Pose 2018).

Agglomeration economies are productivity gains that occur when many firms and workers concentrate their activities in space, i.e. “agglomerate”. Several mechanisms may give rise to such gains, one of them being the quantity and quality of matchings between firms and workers (Duranton and Puga 2004a). A large literature looks at cities as mainly unified labour markets (Goldner 1955; Ihlanfeldt 1997; Bertaud 2004) and it has been argued that the capacity to maintain a unified labour market is the limit to city size and to the growth of megacities in the long run (Prud'Homme 1996). Complete labour mobility within a large city or metropolitan area is a necessary condition for the optimal exploitation of the resulting agglomeration economies. No matter their location within the city or metropolitan area, households should be able to reach “within a reasonable time” all the areas where jobs are offered (Bertaud 2004). However, to our knowledge, no rigorous basis has been provided to define such “reasonable time” and to use it as a technical criterion to determine which municipalities should be considered part of a metropolitan area. Studies by Barrios et al. (2012) and Feser and Sweeney (2002), for example, examined the spatial extent of agglomeration economies in Ireland and the US due to access by firms to a diverse pool of skills, finding evidence of economies at distances of up to 50 kilometres and 50 miles respectively; however, these studies do not consider travelling times.

This paper aims at estimating the commuting times that allow firms to mobilise the diversity of skills available in an urban area. It builds on a theoretical model to describe the mechanisms by which a developing city becomes more productive by increasing its *formal* employment rate, proposed by (O'Clery et al. 2016). (Formal employment is that which takes place in firms that comply with labour legislation). The model assumes that skill heterogeneity, which increases with the size of the working-age population or the city size, is the main source of agglomeration economies. Complementarities across skills imply that complex sectors that use a larger diversity of skills are more productive than those where a more limited set of skills is sufficient to produce. In large cities, skills are more diverse and matches between firms and workers are more likely. As a result, more complex industries are more likely to appear in large cities, absorbing workers from outside the formal economy. New sectors do not emerge randomly as cities grow. If they did, cities would tend to become similar to each other. But the opposite happens: in India, for instance, Bangalore thrives in the high-tech industries, while Bharuch is strong in the petrochemical and Ludhiana in the garment industry (Duttagupta 2019). Regional specialisation may occur even in closely related industries: in Brazil, the region of Vale dos Sinos, in the Rio Grande do Sul state, concentrates on the production of *women's* shoes, while the region of Franca, in the state of São Paulo, specializes in the production of *men's* footwear (TheBrazilBusiness.com 2019). Firms and industries build on existing strengths in their locations, meaning that development processes are strongly path-dependent (Nelson and Winter 1982; Hausmann et al. 2007a; Neffke et al. 2011), leading cities to specialise in distinct industries or niches within industries.

The model proposed by O'Clery et al. (2016) to predict the growth of formal employment for a city captures the skill-proximity of the city from current industries to potential new complex industries. Distinct from education and work experience, capabilities and skills are difficult to capture. In order to quantify the complexity of sectors in terms of their diverse skill requirements, the model presented here utilises the network-based

economic complexity framework, proposed by Hidalgo and Hausmann (2009). Similarly, the skill proximity or “distance” between sectors is not directly observable. We estimate the skill-proximity between industry pairs based on worker transitions between sectors (Neffke and Henning 2013). The model has robust strong predictive power, and fares better than alternative explanations of formal employment growth across cities.

The main contribution of this paper is to exploit this model in order to investigate the relationship between commuting times in the catchment area of firms and their ability to source the diverse skills needed to increase formal employment. On the one hand, the strength of agglomeration economies increases with the diversity of the skills that firms can access in their catchment areas. On the other hand, while a larger population size leads to higher skill diversity, at some point longer commuting times limit the ability of firms to make effective use of such diversity.

We focus our analysis on Colombia, partly due to data availability, but also because Colombia is a particularly interesting emerging economy due to persistently low rates of formal employment and profound regional differences rooted in a distinct geographical and historical context (Acemoglu et al. 2015). As a result, cities are very heterogeneous in their formality rate, industry composition and in their degree of connection to other cities and to the rest of the world. By some indicators, Colombia is one of the most geographically fragmented countries in the world (Gallup et al. 2003).

We seek to find the commuting time radius that maximises the strength of agglomeration economies in terms of enabling formal firms to access the skills they need to grow their activities in complex sectors. If a longer commuting time radius effectively allows firms in each municipality to reach workers living in other municipalities, we should observe that firms are able to generate more formal employment than if they were tapping skills from only their own municipality. We corroborate this hypothesis and find that a time radius of between 45 and 75 minutes maximises the strength of the agglomeration economies. As a result, we can conclude that around a third of the 96 urban municipalities form part of the labour market of other municipalities.

A range of policy implications can be derived from our findings. One important issue is how to take advantage of the tendency of big cities to absorb neighbouring municipalities, particularly in some countries where urbanisation is growing fast. Our results provide arguments for coordinating transportation, planning and industry re-localisation policies across urban municipalities of different sizes, especially if they are part of the same urban area. Our analysis sheds light on the importance of improving the transportation infrastructure to connect small municipalities to urban centres.

The rest of this paper is organised as follows. In the “[Related literature](#)” section we review the main literature strands related to our research question and methods. The “[Formal employment and industry complexity in cities](#)” section introduces our definitions of cities, formal employment and industrial complexity, and investigates their relationships. On this basis, the “[A network-based model for formal employment growth](#)” section is focused on the underlying ideas and construction of the “complexity potential” metric, the main explanatory variable. The “[Model and regression results](#)” section explores the relationship between “formal employment growth” and “complexity potential” (and other controls) for a range of city definitions based on commuting time. We deduce from these results the commuting radius within which firms can make the best

use of the skills available around them. Finally, conclusions and policy implications are discussed in [Conclusions and policy implications](#).

Related literature

This paper contributes to a growing literature that emphasises the role that large cities have in better-facilitating matching between employers and employees, and knowledge spillovers between firms and industries via R&D collaboration or the mobility of skilled individuals (Friedrichs 1993; Duranton and Puga 2001 and 2004b; Rosenthal and Strange 2006.) Cities thrive because they act as a cauldron, enabling firms to combine skills in a process of incremental diversification and sophistication of production. The emphasis on skills and the processes of mixing them is not new [see, e.g., Romer (1986); Lucas Jr (1988); Romer (1990); Kremer (1993); Jones (1995); Jones (2002); Barro and Martin (2003); Jones and Romer (2010); Weil (2012)]. Our emphasis here, however, is on the role of skills in the formation of increasingly complex industries. Employing a simple analogy, we consider skills as letters in a game of Scrabble. The more letters -or skills- a firm can access, the greater the number and sophistication of words -products or services- it can build. Under this model, as cities acquire new skills (or letters in Scrabble), the number of possible industries (or words in the game) grows exponentially. This perspective, derived from the economic complexity and evolutionary economic geography literatures (Hidalgo and Hausmann 2009; Hausmann and Hidalgo 2011; Frenken and Boschma 2007), implies that large cities are able to disproportionately benefit from agglomeration economies resulting from skill complementarities.

Beyond skill abundance and heterogeneity, a second key mechanism is at play. Building on locally embedded skills and capabilities, cities move into new economic activities in a path-dependent manner (Nelson and Winter 1982; Hausmann et al. 2007a; Frenken and Boschma 2007; Neffke et al. 2011). This process can be modelled using an industry network, where the nodes represent the industries and the edges represent their level of relatedness (Hidalgo et al. 2007; Neffke et al. 2011). The diversification of a city into new (complex) industries can be described via a dynamical process on the industry network: cities move into new industries located at short “distance” from current industries in the network (Hidalgo et al. 2007; O'Clery et al. 2016). As proposed by Neffke and Henning (2013), here we have used the number of worker transitions (job switches) in order to estimate the skill-proximity between industry pairs. There are, however, a variety of possible approaches to quantifying “relatedness” between industries, including co-location (Hausmann et al. 2007a; Hausmann et al. 2014a), occupational similarity (Farjoun 1994; Chang 1996) and input-output (customer-supplier) relationships (Fan and Lang 2000). Industry networks have been used to model a wide range of local growth and diversification processes, including firm and sector entry (Neffke et al. 2011; Hausmann et al. 2014b).

The focus of this paper concerns the location of capabilities *within* and *around* a city, and in particular, the distance between firms and workers with the skills they need. This often overlooked aspect of firm-worker matching is key to the success of the complex economic activities that drive the growth of the formal sector. For example, cities may benefit from the skills and capabilities available in their surrounding catchment areas, and neighbouring cities. Boston, for instance, profits from a high density of world-class universities in neighbouring Cambridge and has become a leading hub for technology and

education start-ups in recent years. The importance of distance, location and accessibility in the growth of cities has been well-studied by those interested in commuting zones (Duranton and Overman 2005; Duranton 2015a), transportation links (Venables 2007; Duranton and Turner 2012) and industry clusters (Porter 2011).

A parallel literature, also related to this work, seeks to apply mathematical and data-driven methods to model and analyse transportation networks (Roth et al. 2012; Roth et al. 2011) and mobility patterns (Song et al. 2010; Simini et al. 2012; Prieto Curiel et al. 2018) in cities. Our work is related to studies pertaining to accessibility in cities (see Batty (2009) for a review) and its relationship to the structure of the transport network (Strano et al. 2015; Piovani et al. 2018). A related strand is concerned with social mixing and the fragmentation of social classes in terms of mobility radius and behaviour (Pappalardo et al. 2016; Lotero et al. 2016).

While cities facilitate employment creation, distances between jobs and housing can hinder that process. The 'spatial mismatch' hypothesis, originally put forward by John F. Kain in the early sixties, suggested that poor employment outcomes of Afro-Americans are partly due to residential segregation following the decay of the inner cities and the relocation of many jobs to the suburbs (see reviews by Kain (1992) and Ihlanfeldt and Sjoquist (1998)). More recently, Black et al. (2014) found that labour force participation rates of married women are negatively correlated with metropolitan area commuting times, and that metropolitan areas with larger increases in average commuting time had slower growth in participation rates (during the period 1980-2000). These findings are consistent with the idea that, in order to accommodate domestic responsibilities, working women commute shorter distances than men do (Madden 1981; White 1986; Rosenthal and Strange 2012). Consistently, female entrepreneurs locate businesses in less agglomerated locations than do their male counterparts (Rosenthal and Strange 2012).

This paper contributes to the literature that aims at defining metropolitan areas (see Duranton (2015a) for a historical review). The first definitions of metropolitan areas were proposed in the US (the so-called Standard Statistical Metropolitan Areas, SSMA) and other developed countries in the 1950s and 1960s. Currently, the most widely applied methodology is the one established by the OECD in collaboration with the European Union based on decades of research (OECD 2012; 2013). Under this approach, non contiguous urban cores are considered parts of the same metropolitan area if more than 15% of the resident population of one core commutes to work in the other core. Similarly, if more than 15% of the resident population of an individual municipality works in the core area. This methodology has been applied to over 30 countries, including Colombia (Sanchez-Serra 2016). Previously for Colombia, Duranton (2015a) had developed a similar methodology assuming that a municipality belongs to a metropolitan area if at least 10% of its labour force commutes to the rest of the metropolitan area. In Arcaute et al. (2015) cities are defined over the whole range of commuting thresholds in order to test the robustness of scaling exponents to city definitions. In a similar vein, in this work we do not impose any threshold on commuter flows. Instead, we let the model reveal what is the commuting time that allows firms to make the most, in terms of formal employment generation, of the skills available in municipalities other than their own. We compare our results to the OECD and Duranton city definitions for Colombia mentioned (see Appendix).

The choice of thresholds to define metropolitan areas, or urban agglomerations in general (like in Uchida et al. (2010)) should be based on the extent to which such agglomerations benefit from externalities. More specifically, it should take into account the benefits accrued to firms due to access to a large and diverse pool of skills in cities. This view is related to the importance of “labour pooling” in cities, which refers to the sharing of workers and skills by firms in similar industries (Marshall and Marshall 1920). Previous studies by Barrios et al. (2012) and Feser and Sweeney (2002) have examined the spatial extent of labour pooling externalities for firms in Ireland and the US, finding evidence of spillovers at a distance of 50 kilometers and 50 miles respectively (these studies do not provide travelling times).

Our findings may also contribute to the literature that aims at understanding why metropolitan areas that have more fragmented governance structures are less productive than those with unified or better-integrated governance structures (Ahrend et al. 2014).

Finally, our paper is also related to recent literature that aims to assess how urban structure affects productivity. Recent evidence for Latin American cities shows that urban form matters for city productivity; characteristics such as roundness and smoothness seem to facilitate the interaction of firms and workers (Ferreira and Roberts 2018). The relevant link with our work is that trip patterns within a large city or metropolitan area depend on several aspects of the urban structure, such as density, road patterns, contiguity and whether the city is monocentric or polycentric, among many others (Bertaud 2004). Consequently, labour productivity is also influenced by urban structure, and particularly by the degree of mono- versus polycentricity of the metropolitan area (Meijers and Burger 2010). In this paper, we take into account only travelling times between the centroids of the municipalities that may be part of a metropolitan area. Our approach implicitly assumes that municipalities conform to the monocentric urban pattern, and metropolitan areas to the polycentric urban pattern. Since we use actual travelling times across the centroids, no further assumptions are made with respect to other urban patterns.

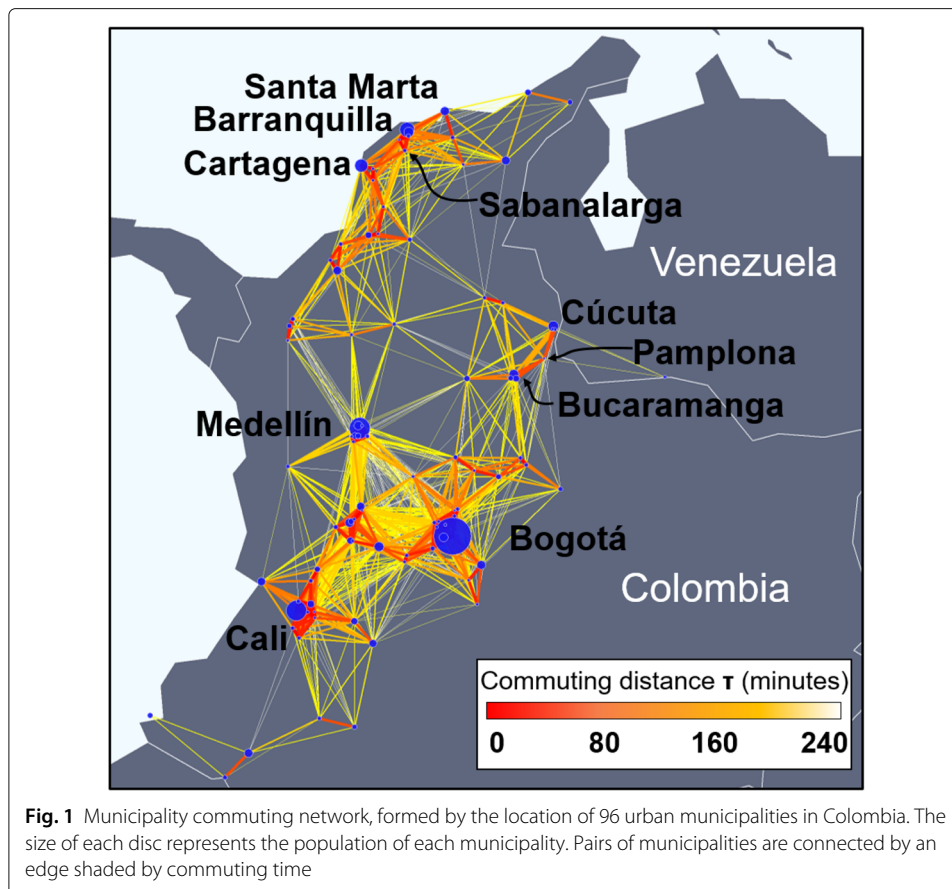
Formal employment and industry complexity in cities

In this section we introduce the methodology behind the measurement of “cities”, “formal employment” and “industrial complexity”, and investigate their relationships. These concepts will be needed in “[A network-based model formal employment growth](#)” below to construct the variable named “complexity potential”, which will be the main explanatory variable of the growth of formal employment used to extract optimal city definitions in “[Model and regression results](#)”.

Constructing cities

We do not assume a priori a definition for cities in terms of their constituent municipalities. Similarly, we do not predefine the number of cities. Instead, we develop an algorithm to iteratively aggregate urban municipalities into cities based on commuting times. In this way, we produce “city constructs”.

Colombia has 96 municipalities with an urban population larger than 50,000 inhabitants according to DANE (the National Statistics Office); we construct cities as combinations of one or more of these 96 urban municipalities, using commuting time thresholds from zero to 200 minutes. The car commuting time between the centroids of two municipalities



is available from the Google Maps API¹. Figure 1 shows the geographical location of the 96 urban municipalities with a population of at least 50,000 inhabitants, connected by car commuting times (derived from Google Maps).

We first create an undirected, fully-connected network of urban municipalities with edge weights corresponding to commuting times (the municipality commuting network) as illustrated in Fig. 1. We then create a new network for each value of our commute time threshold parameter τ . That is, we remove edges from the fully connected network if the commuting time is larger than τ . For each value of τ , the “cities” are found by detecting the connected components of the corresponding thresholded network.

Figure 2 shows the connected components of the thresholded network for $\tau = 30, 60, 90$ and 120 min respectively. If $\tau = 0$, the number of cities is, by definition, the number of urban municipalities. As τ increases, the number of cities decreases as some municipalities become part of a multi-municipality city. Note that, by definition, we do not consider any municipality as the core to which other municipalities are added based on commuting times or flows (as do other methods).

Certain intermediate municipalities are crucial in the merging process as τ increases. For example, Barranquilla and Cartagena (two large cities in the North of Colombia) merge as a result of two earlier merges between Sabanalarga and both Barranquilla and Cartagena. This means that Sabanalarga, located between Barranquilla and Cartagena,

¹Reliable commuting times between municipalities using public transport is unavailable in most Colombian cities, and therefore we rely on car commuting times.

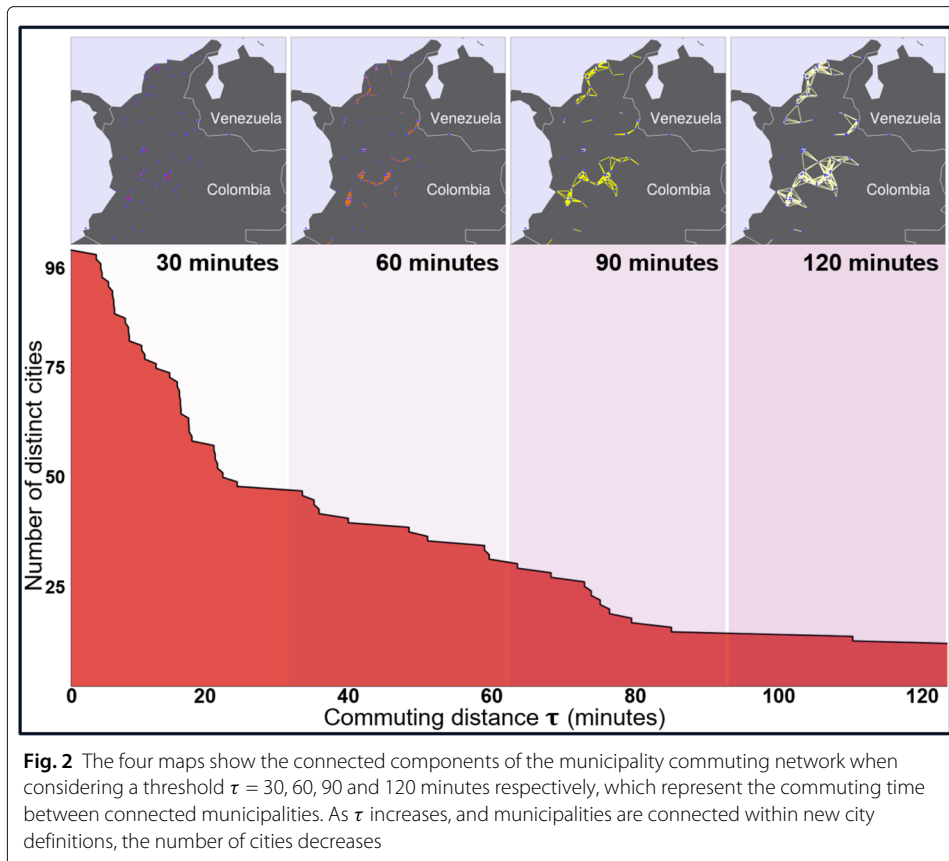


Fig. 2 The four maps show the connected components of the municipality commuting network when considering a threshold $\tau = 30, 60, 90$ and 120 minutes respectively, which represent the commuting time between connected municipalities. As τ increases, and municipalities are connected within new city definitions, the number of cities decreases

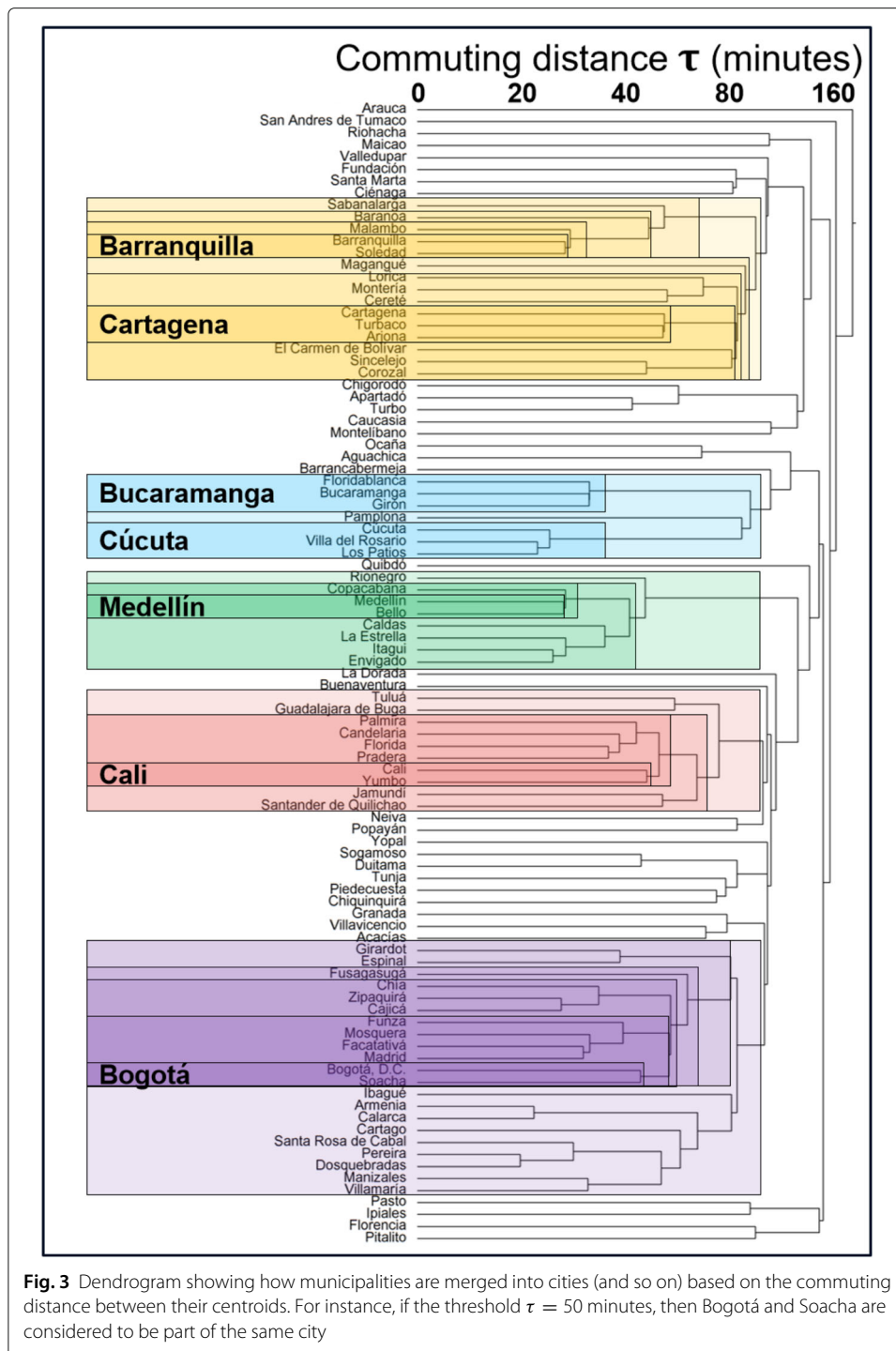
reduces the value of τ for which Barranquilla and Cartagena are merged. A similar situation occurs for the city of Pamplona, located between Bucaramanga and Cúcuta. Figure 3 illustrates this process.

As τ increases, and municipalities are merged, cities increase their working-age population (i.e., the pool of potential workers). For example, in the case of Medellín and its surrounding municipalities (Fig. 4), the working-age population increases from 1.7 million at $\tau = 0$ (single municipality) to nearly 2.6 million by $\tau = 49$ minutes by which time eight urban municipalities are merged.

As mentioned in “[Related literature](#)” section, other aggregations of municipalities in Colombia have been constructed by Duranton (2015b) and by Sanchez-Serra (2016), the latter applying the OECD methodology (OECD 2013). Using the Rand measure, which is a metric of similarity between two data clusterings (Rand 1971), it is possible to compare those aggregations with the ones produced by us using different values of τ . All the different approaches produce very similar aggregations for the 96 municipalities in our database. With values of τ between 25 and 46 minutes we have the highest agreement with the definitions by the previous works. See the [Appendix](#) for more details.

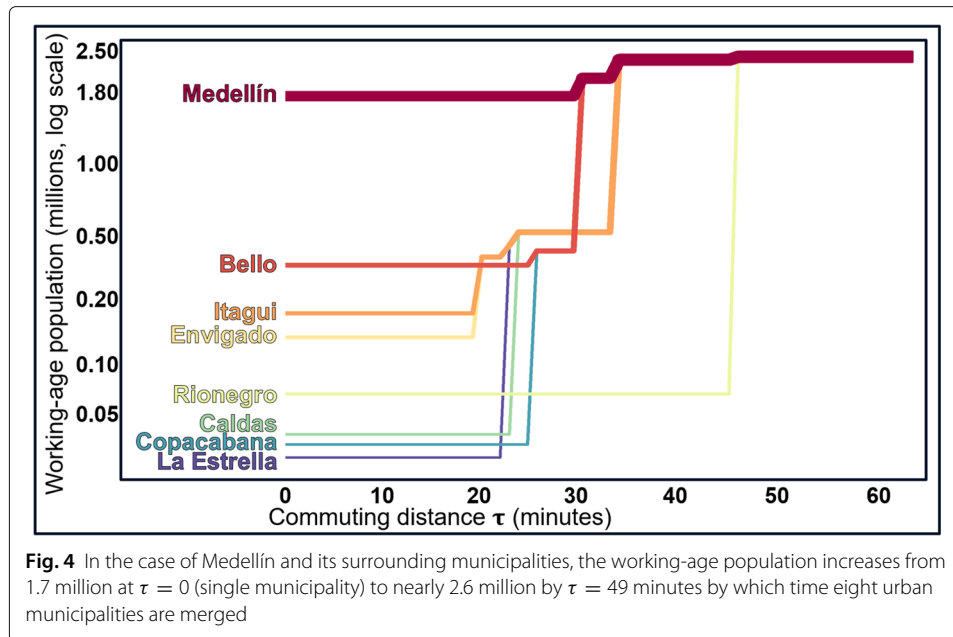
Formal employment

We define *formal employment* as the number of workers hired by firms throughout the year in compliance with Colombian labour legislation on minimum wages and



social security contributions. We use information on formal employment by municipality and industry (330 industries including both manufacturing and services) derived from administrative records of the Colombian Ministry of Health and Social Protection (PILA).

Formally, $femp_{c,i}$ is the formal employment of industry i in city c . Our variable of interest will be the *rate of formal employment by city*, defined as the ratio of formal employment



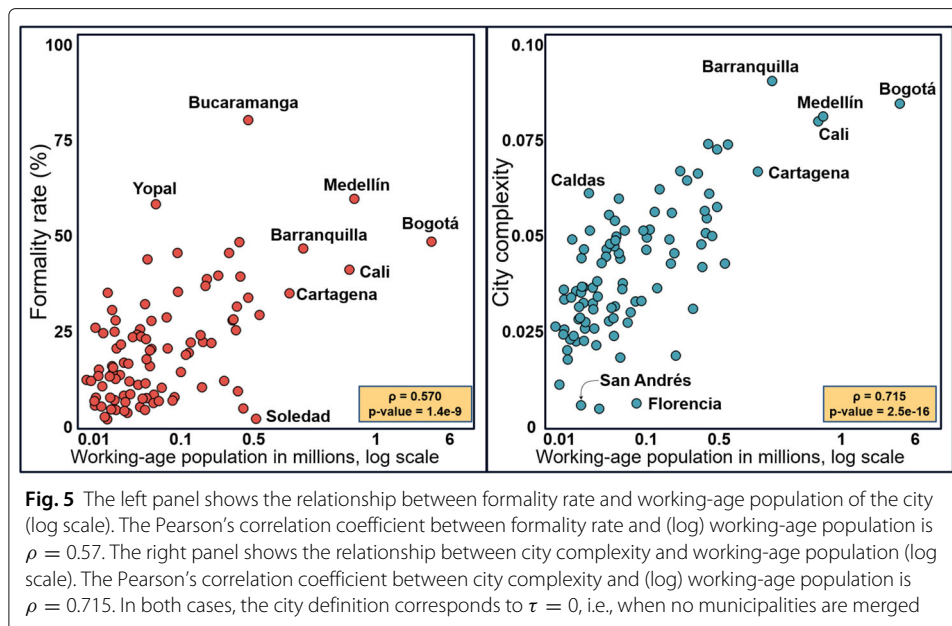
to the size of the working-age population of the city. Formally, if $wpop_c$ is the working-age population of city c , then the rate of formal employment is $FOR_c = \sum_i femp_{c,i} / wpop_c$.

The left panel of Fig. 5 shows that the rate of formal employment increases with city size as measured by the working-age population (shown here for $\tau = 0$). This observation is consistent with the idea that large cities, host to a wide array of diverse skills, are able to foster the sophisticated economic activities that create formal employment. This phenomenon has been previously documented for Colombian, Mexican, Brazilian and US cities (O'Clery et al. 2016). A relevant outlier observed is the municipality of Yopal, located 200 kilometres northeast of Bogota and with a working-age population of roughly 90,000, since its formality rate is considerably high (the third largest in the country). Two big and relevant industries in Yopal are the extraction of crude petroleum and natural gas and the service activities incidental to oil and gas extraction. The oil industries, together with the building of civil engineering works industry, represent nearly 23% of the formal employees of Yopal.

Industrial complexity

Building on the idea that high levels of formal employment result from the presence of sophisticated economic activities, we would expect that large cities are host to complex industries encompassing both the production of goods and services. Here we adopt the economic complexity framework of Hausmann et al. (2007a) in order to quantify the industrial complexity of cities.

Industry complexity is a measure of the diversity of capabilities needed by an industry. Highly complex industries are produced by teams of workers with diverse skills pooling their collective and complementary know-how. Capabilities or skills, however, are multi-dimensional variables that are not directly observable (and should not be confused with years or type of education). What is observable is the outcome of localised capabilities, namely the diversity and uniqueness of the goods and services produced.



Originally proposed by Hausmann et al. (2007a), the product complexity index (PCI) is constructed based on patterns of product exports across countries. Industry complexity is an adaptation of this methodology, developed as part of a project to build an online data-explorer² for Colombia at the Harvard Center for International Development, and is based on patterns of industry employment across cities. The general idea of this approach is that complex goods (or services) are both rare, and produced in places where many other things are produced. Hence, the computation of industry complexity is based on an iterative model which weighs industry diversity by city ubiquity and so on. See the [Appendix](#) for the details of the data and algorithm used.

Similarly, *city complexity* is a measure of the range of capabilities available in a city, and is computed as the mean of the industry complexity of the industries present in the city. As anticipated, we find that large cities tend to be more complex (right panel of Fig. 5), meaning that they have a larger and more diverse skill base necessary for the formation of complementary teams. Some outliers are observed such as Caldas (Antioquia), a small city which has higher city complexity relative to other cities of similar size. This is due to the presence of complex industries such as the fabricated metal products manufacturing industry and other relatively complex industries. Barranquilla also has high city complexity due to the presence of complex industries such as the manufacture of pharmaceuticals.

A network-based model for formal employment growth

In the previous section we introduced the theory and metrics behind the measurement of “cities”, “formal employment” and “industry complexity”. We showed that large cities generate more formal employment and have higher levels of industry complexity. Our

²Available at www.DatlasColombia.com

task now is to explain the mechanism that produces such relationships. For that purpose, here we review and extend the model developed by O'Clery et al. (2016) to measure the “skill-distance” between a city’s current industrial base and new complex industries needed to increase its formality rate. This approach relies on a methodology proposed by Neffke and Henning (2013) to estimate the similarity between industry pairs in terms of skills or capabilities.

Labour flow industry network

An industry is considered skill-related to another industry if the number of job switches (i.e., worker flows) between these industries is larger than what would be expected from randomising all switches among all pairs (Neffke and Henning 2013).

Formally, if $\phi_{i,j}$ is the number of job switches between industry i and industry j (during a given time period), then the “skill-relatedness” can be computed as a matrix with entries:

$$S_{i,j} = \frac{\phi_{i,j} / \sum_j \phi_{i,j}}{\sum_i \phi_{i,j} / \sum_{i,j} \phi_{i,j}}. \quad (1)$$

The skill-relatedness captures whether a higher flow of workers is observed between industry i and j or between j and i and so the matrix $S_{i,j}$ is made symmetric by averaging with its transpose, and re-scaling the values so that they range from -1 to 1:

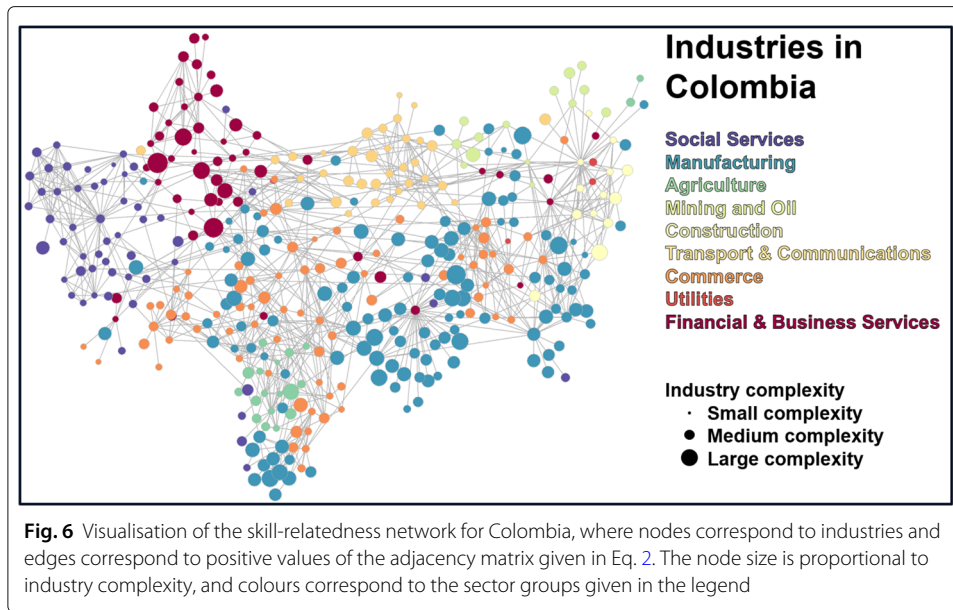
$$A_{i,j} = \frac{S_{i,j} + S_{j,i} - 2}{S_{i,j} + S_{j,i} + 2}. \quad (2)$$

We can consider $A_{i,j}$ as the adjacency matrix of an undirected weighted network where the nodes are the set of industries and the edge weight between nodes i and j represent how relatively far the labour flow between those two industries are from the random expectation, given by the value of $A_{i,j}$. Note: only positive values of this matrix (that is, more job switches than expected) are preserved³. Full detail on these methodological considerations can be found in Neffke and Henning (2013).

The industry network is visualized in Fig. 6, where industries (the nodes of the network) which belong to the same sector (official industry classification) have the same colour, and pairs of industries which have more switches than expected ($A_{i,j} > 0$) are connected with an edge (so that the edges of the network represent pairs of industries with a high number of job switches or high skill-relatedness). The size of the node is proportional to its industry complexity. We observe natural clustering of industries based on shared skills or required know-how, in many cases along sectoral lines. For example, on the left-hand side, it is apparent that both social services (purple) and financial services (red) tend to share workers, while manufacturing industries (blue) exhibit a number of distinct clusters.

We can quantify the observed clustering of industries within sectors by comparing edge connectivity within sectors relative to what would be expected if edges were distributed randomly. The positive entries of the adjacency matrix A are shown in Fig. 7. Nodes are ordered in terms of sector, and sectors delineated by vertical and horizontal lines. The diagonal blocks correspond to within-sector edges. We observe clear within-sector clustering (high density of edges) in many cases, e.g., social services, transport and communications. Our measure of the density of edges between (distinct) sectors G and G' is

³To be precise, if $(S_{i,j} + S_{j,i})/2 > 1$, or the average of the bi-directional switches between industries i and j is more frequent than expected, then $A_{i,j} > 0$ in Eq. 2.



given by:

$$\mu(G, G') = \frac{e_{G,G'} n(n-1)}{e n_G n_{G'}}, \tag{3}$$

where n and n_X are the total number of industries and the number in sector G respectively, e and $e_{X,Y}$ are the total number of edges and the number between nodes in sectors G and G' respectively. In essence, this computes the ratio of actual edges to possible edges within the subgraph induced by the industries in sectors G and G' , divided by the ratio of total edges to all possible edges (i.e., a complete graph). A value of $\mu(G, G') > 1$ means that there are more switches between sectors G and G' than expected.

For a single sector G , this expression is slightly modified:

$$\mu(G) = \frac{e_{G,G} n(n-1)}{e n_G(n_G-1)}. \tag{4}$$

As before, a value of $\mu(G) = \mu(G, G) > 1$ indicates that switches within sector G are more frequent than expected.

We find that $\mu(\text{Construction}) = 39.6$, and hence workers are significantly more likely to switch between different construction industries than other industries. On the other hand, $\mu(\text{Manufacturing}) = 0.8$ and hence manufacturing industries do not form a single tightly connected cluster, but instead form connections with other industries that share specific skills and competencies.

Considering off-diagonal blocks, we observe significant connections between sectors. For example, switches between construction and mining and oil industries are denser than average connectivity in the network. It is clear that, while some of the structure of the network follows sector groupings of industries, much relatedness between industries is not captured by official sector classifications.

Local industrial structure

We can think of the industry network as a ‘economic landscape’ that describes the possible diversification paths open to a city. Specifically, the location of the set of industries

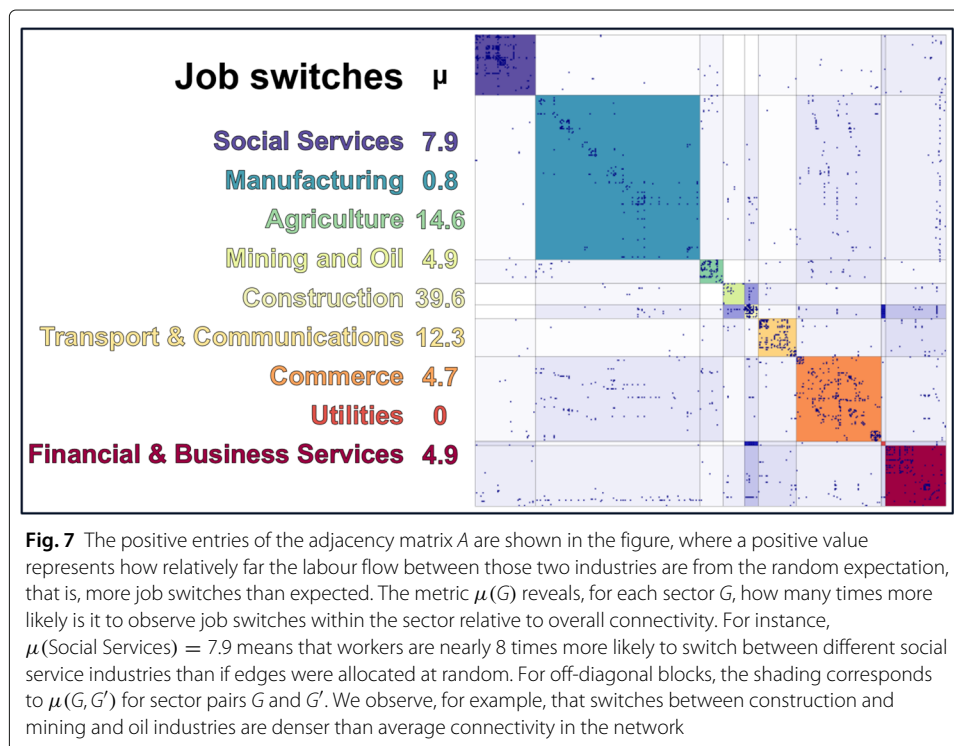
present in a city constrains its future development: cities tend to move into new economic activities that are proximate (in a skill - and consequently network - sense) to those already present (Nelson and Winter 1982; Hausmann et al. 2007a; Frenken and Boschma 2007; Neffke et al. 2011). Before we explore growth paths, we need to formally define “industry presence” with respect to changing city definitions.

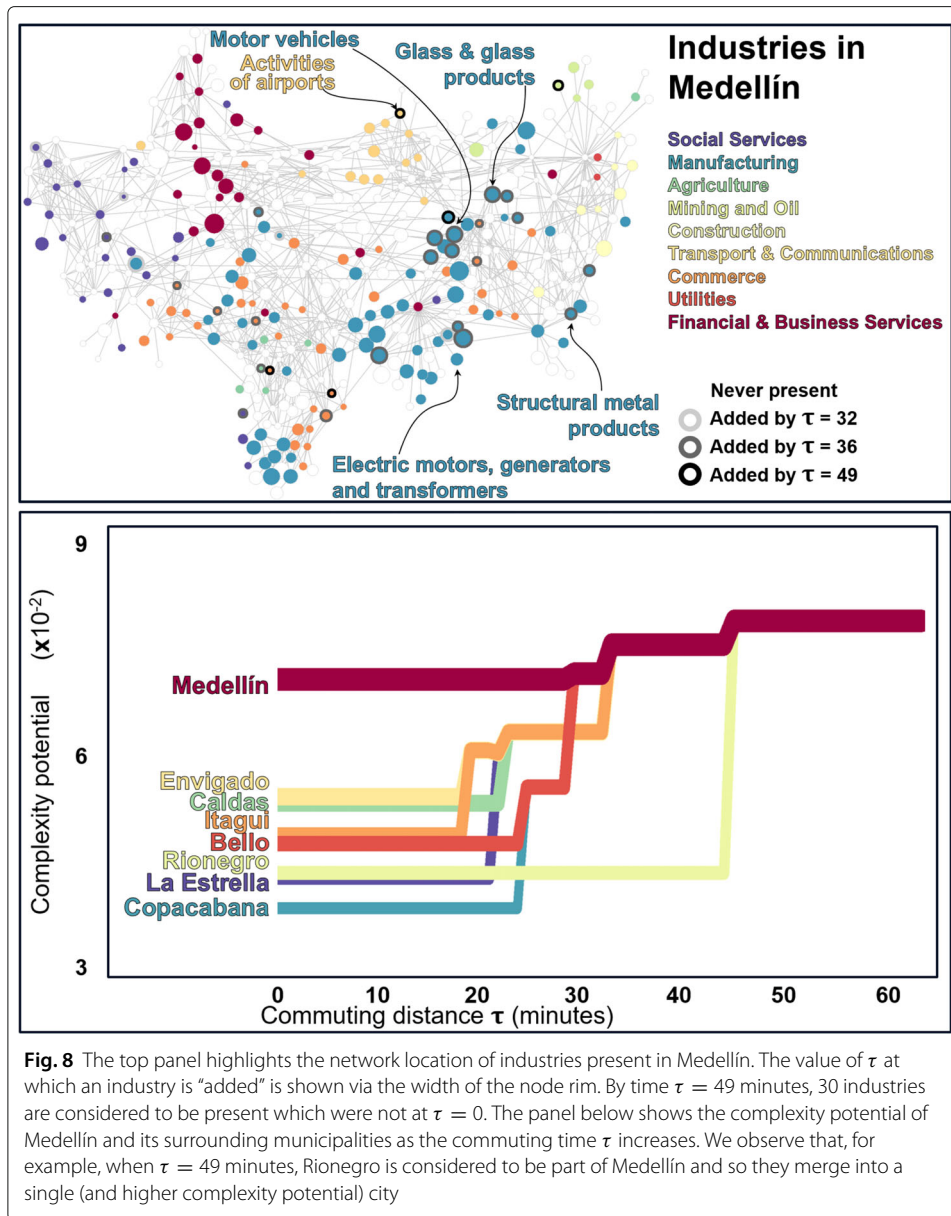
For a given city definition, we will use a commonly deployed measure which captures the relative importance of an industry in the city given its overall distribution across the country. The *location quotient* $LQ_{c,i}$, also known as revealed comparative advantage, is defined as:

$$LQ_{c,i} = \frac{femp_{c,i}/wpop_c}{\sum_c femp_{c,i} / \sum_c wpop_c} \tag{5}$$

An industry i is “present” in city c if the corresponding location quotient $LQ_{c,i} > 1$, which means that a larger share of the population of city c work in industry i than at a national level (more in the Appendix).

As a case study, we will explore the industrial structure of Medellín, and its surrounding municipalities. Medellín is the second largest city in Colombia and it is located in the Aburrá Valley. Several municipalities can be considered part of Medellín, including Envigado, Itagüí, La Estrella and Bello, which are connected to the core centre of the city with the Medellín Metro system. A clear definition of the limits of Medellín, however, is not obvious. For example, Rionegro is a municipality located to the east of Medellín. Although there is a considerable distance between Rionegro and Medellín, Rionegro is host to the *José María Córdova International Airport*, the second busiest airport of Colombia, which functions as an aerial hub for Medellín.





The set of industries present in Medellín is shown in Fig. 8. This set changes as the definition of the city changes (as we increase the commuting time threshold τ). Municipalities such as Copacabana, La Estrella, Envigado, Bello and Rionegro become part of the city and, as a result, some new industries develop location quotients higher than 1, thus becoming “present”.

Several examples are worth mentioning. The glass industry becomes present when Envigado joins Medellín. The firm in question is Peldar which employs around 200 in Envigado (and nationwide 1,200), and is owned by one of the largest business groups in Colombia. Peldar manufactures crystal products and glass containers highly regarded in Colombia. The vehicle industry is added to the industrial portfolio of the city also as a result of Envigado, which hosts the Colombian subsidiary of Renault, the only big exporter of cars from Colombia. Renault has 1 100 workers making it the second largest

employer in Envigado - and by far the largest exporter: US 317m in 2017 (or nearly 85% of the municipality's exports). The electric motors and transformers industry is added when the small municipality of Copacabana is included in the city. This is due to one firm (Rymel) that produces electrical equipment and installations. Not all of the high complexity industries that are added produce manufactured goods. When the commuting radius is extended to $\tau = 49$ minutes, Rionegro becomes part of the city and "activities of airports" are added to the industry mix. Although direct employment generated by the Rionegro airport is small (180 workers), it makes the Medellín area a hub for air travel and, more significantly, gives the city an important advantage for exports of high-value products.

Complexity potential

As Fig. 8 suggests, clusters of skill-related industries tend to be present in a city due to the fact that industries that use similar skills tend to co-locate, and new industries emerge that share similar skills to existing competencies. The question arises: can we capture, in a measure, the likelihood that a city moves into new industries more complex than those already present?

The *complexity potential* of a city, introduced by O'Clery et al. (2016), is a measure of the possibilities for the city to move to more complex industries that are not yet present in the city, taking into account the existing skills of the local labour force. To compute the complexity potential in the initial period (2008), we need to estimate the distance between city c and each missing industry i . This distance weighting factor $d_{c,i}$, also known as "density" in the literature (Hausmann et al. 2007b; Hausmann et al. 2007a; Hausmann et al. 2014a), is defined as

$$d_{c,i} = \frac{\sum_{j \in N_c} A_{i,j}}{\sum_j A_{i,j}}, \quad (6)$$

where N_c is the set of industries that is present in city c . Effectively, it is the ratio of edge weights connecting i to industries present in city c to the total edge weight (for edges connected to node i). Computed below for the set of currently missing industries, this quantity can be thought of as the likelihood of an industry appearance based on current presences in industries with similar skills. The density of a city-industry pair varies with commuting threshold τ through variation in the set of present industries in a city, N_c .

The complexity potential for a city is the weighted mean of the complexity of its missing industries, where the weight corresponds to the density above:

$$CP_c = \frac{1}{|M_c|} \sum_{i \in M_c} d_{c,i} C_i, \quad (7)$$

where M_c denotes the set of "missing" industries for city c (where $LQ < 1$), and the $C_i \in [0, 1]$ is the normalised complexity of industry i . The complexity potential varies with commuting threshold τ through variation in the density (and not the complexity which we fix).

The complexity potential of a city varies when more municipalities are added and the set of industries considered present changes. Continuing with our Medellín case study, Fig. 8 shows that the complexity potential increases when first Bello and then Envigado, Itagüí and La Estrella are added (the three municipalities located in the South of Medellín). Finally, it increases again at $\tau = 49$ which is when Rionegro (which hosts the airport) is added. The fact that the complexity potential of Medellín grows as more municipalities

are added implies that the industries that are added increase the density - or likelihood of an appearance - of complex industries (i.e., the added industries increase proximity to complex industries in the industry network).

In the next section, we will investigate the extent to which the skills embedded in surrounding municipalities, such as those observed in Medellín, are predictive of a cities' ability to grow its formal employment rate.

Model and regression results

As mentioned, complete labour mobility is a necessary condition for the optimal exploitation of agglomeration economies in a city or metropolitan area. For this, households should be able to reach “within a reasonable time” all the locations where jobs are offered. Our interest in this paper is to shed light on the “reasonable time” of commuting that allows firms to maximise their reach of skills in the surrounding municipalities. This should be the time that maximises the impact of complexity potential on formal employment growth. Within this time radius, one or more municipalities effectively operate as an integrated labour market and can, therefore, be considered as a single city.

This can be tested by regressing the change in formal employment rates on complexity potential *for a range of city definitions based on commuting time*. Formally:

$$\Delta FOR_c = \beta_0 + \beta_1 \log CP_c + \beta_2 FOR_c + X_c \quad (8)$$

where ΔFOR_c is the change in formal employment rate between 2008-2013, FOR_c is the formal employment rate in 2008, the controls X_c include the working-age population in 2008, GDP per capita and others. More details on all variables may be found in the [Appendix](#).

Figure 9 plots the coefficient of complexity potential, β_1 , against both commuting threshold τ (left) and the number of distinct cities obtained by varying τ (right). We observe that the coefficient increases and becomes more significant as τ increases, but only up to a certain limit. The coefficient is largest between 45 and 75 minutes, equivalent to between 62 and 43 cities. This region is shown via the two dark vertical lines. It then declines but remains significant up to about $\tau = 90$ minutes, which is equivalent to 35 individual cities. After this point, the confidence interval strays into negative values. We can interpret these results as suggesting that a commuting radius of between 45 and 75 minutes enables firms to source the diverse skills they need to move into complex industries, and increase the city-level formal employment rate. This implies that Colombia has between 43 and 62 cities that operate as an integrated labour market.

Tables showing the full regression results for both $\tau = 0$ and $\tau = 45$ are available in the [Appendix](#). Our results are robust to a host of tests, such as the inclusion of various demand shocks and other control variables. We run a placebo test for robustness by randomising the agglomeration of municipalities into cities (thus ignoring geographic or commuting proximity). We find that, in this case, complexity potential has no explanatory power, further supporting our model and the robustness of our results.

Conclusions and policy implications

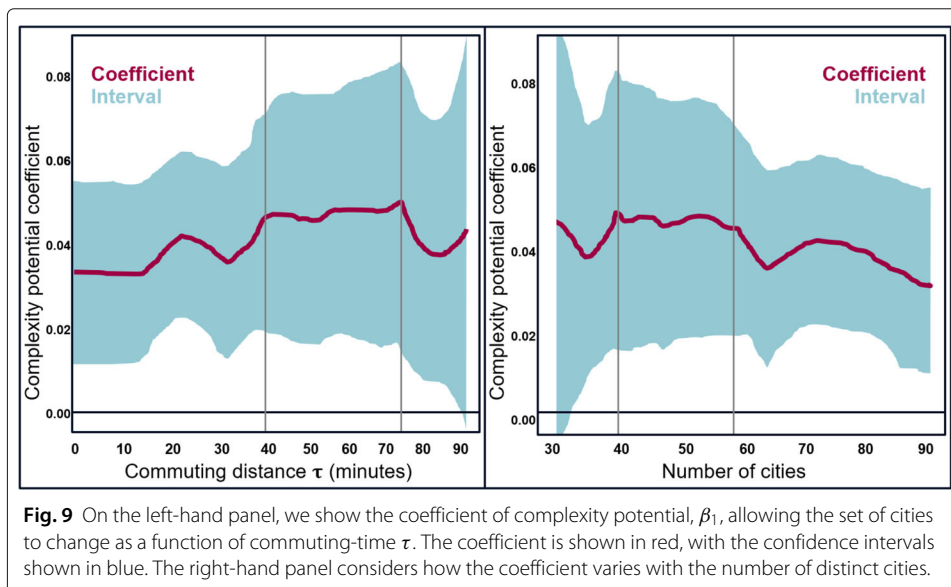
Political or bureaucratic definitions of municipalities are not adequate to describe today's metropolitan sprawls as cities circulate workers across traditional boundaries. While previous attempts have been made to delineate Colombia's cities based on commuting flows

(Duranton 2015a; Sanchez-Serra 2016), none has used a criterion based on economic outcomes emerging from agglomeration economies to identify distinct cities as composites of groups of municipalities.

Building on an existing model, we argue that formal employment creation is mainly influenced by the availability of diverse and sophisticated skills including those which may lie in periphery municipalities. Building cities by aggregating urban municipalities within a given commuting time of each other, we show that a commuting radius of between 45 and 75 minutes, or 62 to 43 cities, is the optimum scale at which firms take advantage of the diversity of skills within the city boundary. Beyond a radius of about 90 minutes, the relationship between complexity potential and employment growth is no longer statistically significant, suggesting that firms cannot effectively make use of labour skills beyond that radius.

The most important policy implication of these results is that in order for large cities to take advantage of the greater diversity of labour skills that comes with size, adequate transportation means are necessary to limit travelling times (across the constituent municipalities). There is no reason to discourage cities from expanding or absorbing neighbouring municipalities, provided commuting times are kept within reasonable limits (our study suggests up to 75 min). Similarly, isolated mid or small size cities may be able to generate more formal employment if adequate investments are made to connect them to large cities.

While the objective in transportation investments is usually reducing cargo transportation costs, particularly in developing or emerging economies, our results suggest that passenger transportation is probably more important. Since existing political borders across municipalities may make decisions difficult to coordinate, some external mechanism or institution could be created to encourage coordinated transportation and infrastructure investments alongside urban planning and industry re-localization programs.



Appendix

Appendix A: Formal employment and population variables

The *working-age population* $wpop_c$ of city c is defined as the population 15 or older in the city, and is provided by DANE (Colombian National Administrative Department of Statistics) by municipality and year.

The *formal employment of industry i in a city c* ($femp_{c,i}$) is defined as employment covered by the health social security system and/or the pension system (the self-employed are not included). Formal employment by industry and city in 2008 and 2013 is computed as the number of formal employees in an average month, and is collected by the Colombian Ministry of Health (PILA).

The *formal occupation rate of an industry i in a city c* ($FOR_{c,i}$) is computed as the formal employment in the city-industry divided by $wpop_c$:

$$FOR_{c,i} = \frac{femp_{c,i}}{wpop_c}. \quad (9)$$

Appendix B: Industry presence and complexity variables

The *location quotient*, $LQ_{c,i}$ (also known as revealed comparative advantage), reflects the relative importance of an industry in a city given its overall distribution,

$$LQ_{c,i} = \frac{femp_{c,i}/wpop_c}{\sum_c femp_{c,i} / \sum_c wpop_c}. \quad (10)$$

Notice that $LQ_{c,i}$ is defined with respect to the working-age population, not with respect to the employment.

We say that an industry is *present* in city c if the corresponding location quotient $LQ_{c,i} > 1$, which means that a larger share of the population of city c work in industry i than at a national level. An implication is that industries may become 'present' without other industries losing importance or presence.

The *diversity* of city, $k_{c,0}$ is the number of industries present in city c . For its computation, let M be a matrix with entry $M_{c,i} = 1$ if industry i is present in city c (i.e., $LQ_{c,i} > 1$) and zero otherwise. The diversity of city c , expressed as $k_{c,0}$, is

$$k_{c,0} = \sum_i M_{c,i}. \quad (11)$$

The *ubiquity* of industry i expressed as $k_{i,0}$ is the number of cities in which industry i is present

$$k_{i,0} = \sum_c M_{c,i}. \quad (12)$$

Industry complexity was originally proposed by Hausmann et al. (2007a) for export products and it is computed on the basis of "diversity", which is the number of industries present in a city, and "ubiquity", the number of cities where an industry is present. It is a measure of the range of capabilities needed by industry and it is computed iteratively as follows.

Firstly, the diversity of a city is weighed by the ubiquity of the industries present in it. Formally, let the average diversity of the industries present in city c be

$$k_{c,1} = \frac{1}{k_{c,0}} \sum_i M_{c,i} k_{i,0}. \quad (13)$$

Similarly, the ubiquity of an industry is weighed by the diversity of the cities where it is present and so, the average ubiquity of industry i is

$$k_{i,1} = \frac{1}{k_{i,0}} \sum_c M_{c,i} k_{c,0}. \tag{14}$$

If this calculation is done iteratively (so that the average diversity of the industries present in city c is weighted by the average ubiquity of those industries and vice versa) to step n , the two previous expressions become:

$$k_{c,n} = \frac{1}{k_{c,0}} \sum_i M_{c,i} k_{i,n-1}, \text{ and} \tag{15}$$

$$k_{i,n} = \frac{1}{k_{i,0}} \sum_c M_{c,i} k_{c,n-1}. \tag{16}$$

For industry i , the n -th step can be conveniently expressed in closed form as

$$k_{i,n} = \sum_j \tilde{M}_{ij} k_{j,n-2}, \tag{17}$$

where the matrix \tilde{M}_{ij} is defined as:

$$\tilde{M}_{ij} = \sum_c \frac{M_{c,i} M_{c,j}}{k_{c,0} k_{i,0}}. \tag{18}$$

Hence, if \mathbf{k}_n is a vector whose i -th element is $k_{i,n}$ then expression 17 becomes:

$$\mathbf{k}_n = \tilde{M} \mathbf{k}_{n-2}. \tag{19}$$

By computing \mathbf{k}_0 , values of \mathbf{k}_n can be easily computed for $n \geq 2$. A long-run steady-state value of \mathbf{k}_n occurs when further steps and further weighting between the diversity of cities and the ubiquity of industries does not change the results much. This is given by a vector \mathbf{k}_n such that $\mathbf{k}_n \approx \mathbf{k}_{n-2}$. This can be found by computing the eigenvectors of the matrix \tilde{M} . Notice, however, that the rows of \tilde{M} sum to one (which also means that a vector of all ones is an eigenvector of \tilde{M} and it is associated with the largest eigenvalue of \tilde{M}) and so, the vector associated to the second largest eigenvalue of \tilde{M} is taken as the industry complexity (C_i). For more details on the calculation and interpretation of the industry complexity see Hausmann et al. (2014c) and Mealy et al. (2018).

City complexity is a measure of the range of skills or capabilities available in a city. It can be computed jointly with industry complexity or, equivalently, computed as the average of the industry complexity of the industries present in the city.

Appendix C: Control variables

GDP per capita (GDPpc), which is available from 2011 onwards and is calculated from GDP by municipality estimated by DANE.

Oil producing city, a binary variable, O_c which takes the value of $O_c = 1$ if the city has more than one oil well in production per thousand inhabitants. Oil well data refers to 2014, as reported by Ecopetrol (the Colombian hydrocarbon company) for their own internal records.

Government spending shock is the change between 2008 and 2013 in total government spending (in 2008 prices) per working-age person. It is computed from municipality-level government spending data compiled by CEDE ⁴.

⁴Available at <https://datoscede.uniandes.edu.co/es/>

Sectoral demand shocks, is a Bartik-style measure (Bartik 1991) that quantifies, for each city, the mix of nationwide sectoral demand shocks facing the city. It is computed as

$$sds_c = FOR_c \sum_i \frac{femp_{c,i}(2008)}{femp_c(2008)} g_{i,c} \tag{20}$$

where $g_{i,c} = \log[femp_i(2013)] - \log[femp_i(2008)]$ is change in (log) employment of industry i excluding employment in industry i in city c . In other words, here $femp_i = \sum_{j \in J} femp_{i,j}$ with set J containing all cities except city c . It can be interpreted as the expected change in the formal occupation rate of the city given the nationwide sectoral demand shocks (exogenous to the city).

Appendix D: Algorithm

The algorithm to merge municipalities into cities, and compute their corresponding variables, depending on the commuting time τ , proceeds as follows.

For each commuting time $\tau = 1, \dots, 200$,

- We construct a matrix G^τ of size 96×96 , and set $G_{i,j}^\tau = 1$ if the commuting time between municipality i and j is less than time τ .
- We then detect the connected components in the network given by adjacency matrix G^τ . Each component corresponds to a set of municipalities (or a single municipality) from the initial set of 96 - these components form the cities.
- We aggregate employment by city and industry (according to the cities identified above). We then compute the location quotients and complexity potential for each city, as well as the resulting formal employment rates in 2008 and 2013 and their changes between the two periods. We also construct a range of other variable according to our city definition, namely working-age population, GDP per capita, a

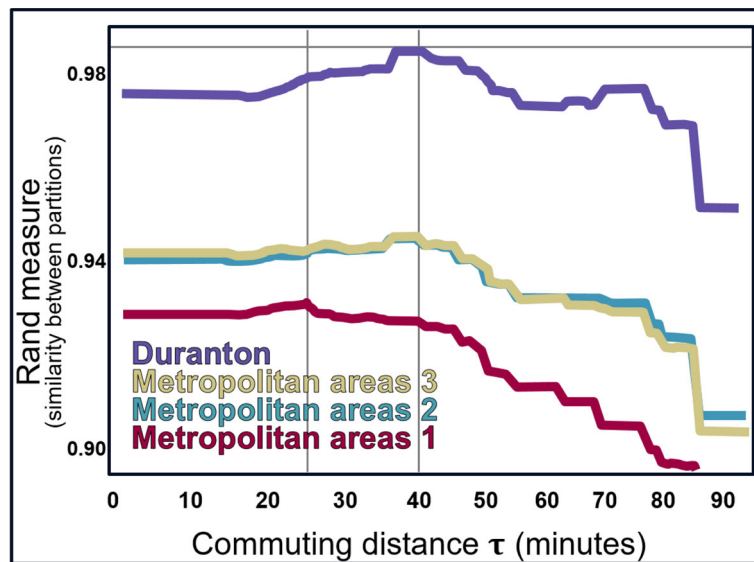


Fig. 10 Here we use the Rand Index to compare our method for defining cities (varying depending on commuting distance τ) and the metropolitan areas constructed by Duranton and the metropolitan areas constructed by the OECD (three versions: red corresponds non-aggregated municipalities, blue and yellow correspond to aggregations based on different commuting flow thresholds). We find best agreement with the Duranton method for all values of τ .

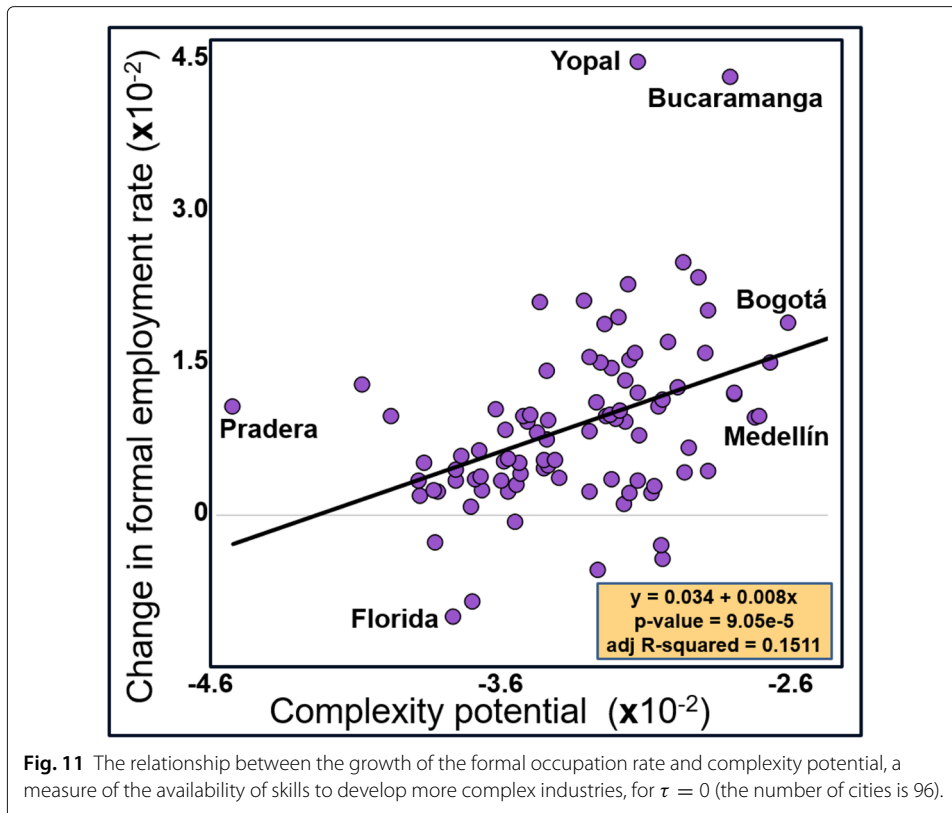


Table 1 This set of regressions explores the determinants of the change of the formal occupation rate between 2008-2013 for a set of 96 single-municipality cities

VARIABLES	(1)	(2)	(3)	(4)	(5)
Log complexity potential	0.036*** (0.012)	0.039*** (0.012)	0.033*** (0.011)	0.028** (0.012)	0.033*** (0.011)
Log working age pop 2008	-0.004 (0.004)	-0.004 (0.004)	-0.002 (0.004)	-0.005 (0.004)	-0.003 (0.004)
Log GDP per capita 2011	0.008 (0.009)	-0.014* (0.007)	-0.000 (0.009)	0.001 (0.008)	-0.014* (0.007)
Binary oil: one well/1000		0.077*** (0.026)			0.054** (0.022)
Govt spending pc			0.029** (0.013)		0.015 (0.010)
Sectoral demand				1.340*** (0.467)	0.484 (0.515)
Formality rate 2008	0.097 (0.084)	0.131 (0.083)	0.108 (0.084)	-0.217* (0.123)	0.013 (0.150)
Constant	0.110 (0.104)	0.310*** (0.107)	0.131 (0.109)	0.153 (0.107)	0.277*** (0.102)
Observations	96	96	95	96	95
R-squared	0.309	0.435	0.382	0.390	0.463

The results show a consistent positive relationship between complexity potential - the proximity of a city to potential complex industries - and the change of formality, controlling for relevant factors. Several exogenous shocks also influence formal employment

binary variable for at least one oil well per 1,000 people, the government spending shocks, and the Bartik-style sectoral demand shocks.

- Finally, we run the regression in Eq. 8. Note: the industry complexity C_i and the matrix of industry proximities $A_{i,j}$ are fixed, and only the set of present and missing industries (N_c and M_c) vary for each τ /set of cities.

Appendix E: Comparing between two aggregations

We use the Rand measure RM (Rand 1971) to compare between two aggregations of municipalities for the 96 municipalities in our database. The Rand measure divides the number of “agreements” between the two partitions (either pairs of municipalities which are merged together in both partitions and pairs off municipalities which are not merged together) by the total number of pairs, so that the Rand measure has values between (0, 1), where higher values are interpreted as more similar partitions.

We compare our method with different values of τ against the three OECD metropolitan areas and compute the Rand measure as τ varies, so we obtain $RM_i(\tau)$ for $i = 1, 2$ and 3 (which correspond to the three OECD metropolitan areas) and with the partition of Colombian municipalities constructed by Duranton (Duranton 2015a).

Results show a high level of agreement ($RM_i > 0.9$ for any $\tau < 80$) and then a decreasing value for larger commuting distances (Fig. 10). Yet, although there is a high level of agreement, both aggregations provide different information, since $RM < 0.943$ for all values of τ , meaning that constructing cities based on the commuting distance between municipalities and based on metropolitan areas yields different units of observation.

Table 2 This table presents regressions for the 62 cities corresponding to $\tau = 45$

VARIABLES	(1)	(2)	(3)	(4)	(5)
Log complexity potential	0.052*** (0.015)	0.048*** (0.014)	0.052*** (0.015)	0.050*** (0.014)	0.049*** (0.014)
Log working age pop 2008	-0.016** (0.006)	-0.011** (0.004)	-0.014*** (0.005)	-0.014*** (0.005)	-0.011** (0.005)
Log GDP per capita 2011	0.017** (0.007)	-0.004 (0.009)	0.009 (0.007)	0.014* (0.007)	-0.007 (0.010)
Binary oil: one well/1000		0.054** (0.027)			0.054** (0.022)
Govt spending pc			0.018 (0.015)		0.014 (0.013)
Sectoral demand				0.551 (0.529)	-0.349 (0.446)
Formality rate 2008	0.131 (0.086)	0.147* (0.075)	0.147 (0.091)	-0.023 (0.144)	0.249 (0.157)
Constant	0.220** (0.105)	0.336*** (0.123)	0.257** (0.118)	0.223** (0.107)	0.352*** (0.125)
Observations	62	62	61	62	61
R-squared	0.428	0.502	0.463	0.441	0.519

As before, the dependent variable is change in formal occupation rates and the main explanatory variable is complexity potential, which is highly significant in all regressions

Table 3 Here we assign municipalities randomly to 62 cities in five successive placebo cases and replicate the last regression (column 5) of the previous tables in each case

Variables	(1)	(2)	(3)	(4)	(5)
Log complexity potential	0.003 (0.009)	0.007 (0.008)	0.010 (0.010)	0.004 (0.008)	0.019* (0.010)
Log working age pop 2008	-0.002 (0.006)	-0.004 (0.006)	-0.000 (0.007)	0.000 (0.004)	-0.006 (0.006)
Log GDP per capita 2011	0.010 (0.010)	-0.008 (0.007)	0.003 (0.019)	0.005 (0.009)	-0.004 (0.016)
Binary oil: one well/1000	0.011 (0.014)	0.012 (0.013)	0.005 (0.031)	-0.004 (0.022)	0.002 (0.019)
Govt spending pc	0.030 (0.021)	0.007 (0.016)	0.017 (0.017)	0.019 (0.017)	0.027 (0.017)
Sectoral demand	2.152** (0.809)	0.841 (0.636)	1.496* (0.857)	1.149* (0.638)	1.828 (1.434)
Formality rate 2008	-0.450* (0.250)	-0.055 (0.223)	-0.335 (0.275)	-0.145 (0.203)	-0.388 (0.409)
Constant	-0.059 (0.146)	0.150 (0.126)	0.023 (0.221)	-0.038 (0.122)	0.172 (0.199)
Observations	62	61	61	62	61
R-squared	0.453	0.258	0.310	0.451	0.413

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The relevant coefficient is low and not significant

Appendix F: Additional regression results

Ignoring commuting time ($\tau = 0$)

In Fig. 11 we show the relationship between the growth of the formal occupation rate and complexity potential, a measure of the availability of skills to develop more complex industries, for $\tau = 0$ (the number of cities is 96). Table 1 presents the regression results (including relevant additional control variables). The variable of interest (complexity potential) is significant in all the regressions. The coefficient of complexity potential is very stable, suggesting that a 10% increase in complexity potential is followed in a five-year period by an increase of about 0.28-0.39 percent points of the formal occupation rate.

City definition at $\tau = 45$

Here we choose $\tau = 45$ as a plausible “reasonable commuting time”, consistent with the results summarised in Fig. 9 and discussed in the main text. As Table 2 indicates, complexity potential is significant again in all the specifications. Furthermore, its coefficient is larger and more significant than in the previous set of regressions. Within the 45-minute time radius, a 10% increase in complexity potential is followed in a five-year period by an increase of about 0.48-0.52 percent points in the formal occupation rate.

Placebo cities

An alternative approach to test the significance of the results is to construct placebo cities, that is, random collections of municipalities without regard to their (geographic or

commuting time) proximity. Does “complexity potential” remain significant in such case, which would imply that factors other than commuting time are at play? We randomly allocate the 96 urban municipalities to form sets of 62 cities, and re-compute the necessary variables as above. We present five random draws, all of which include the same controls in the previous tables. As Table 3 shows, the coefficient of complexity potential is very low and never robustly significant, suggesting that information on the diversity and sophistication of the skills in the constructed disconnected cities bears no relation with the growth of the formal occupation rate.

Abbreviations

DANE: Departamento Administrativo Nacional de Estadística (the Colombian National Administrative Department of Statistics); PILA: Planilla Integrada de Aportes Laborales (Colombian labour micro-data integrated by the Colombian Ministry of Health and Social Protection)

Acknowledgements

We would like to acknowledge Manuel Toledo for thoughtful comments on an earlier draft. Andres Gomez at the Center for International Development played a critical role in earlier work leading to this research focus. Michele Coscia provided assistance in accessing data on commuting distance. Gilles Duranton, Pablo Sanguinetti, Juan Vargas and other participants of the Workshop on Urban Economics held in Buenos Aires in July 2016, are also acknowledged for many useful suggestions. Three anonymous referees provided valuable comments and suggestions, which we gladly acknowledge. All errors are our own.

Authors' contributions

NO and EL designed the study and the methodology, NO, EL and RPC processed and analysed the results.

Funding

This article was completed with support from the PEAK Urban programme, funded by UKRI's Global Challenge Research Fund, Grant Ref: ES/P011055/1, and a grant awarded by the Development Bank of Latin America (CAF) to Eduardo Lora and Neave O'Clery as part of the 13th Annual Research Contest.

Availability of data and materials

We use administrative records of the social security system in Colombia, which contains job information about all formal workers in Colombia.

We obtained the dataset from the Ministry of Finance and Public Credit (MFPC) of Colombia, under a “Data Use Agreement” that is part of the development of www.datlascolombia.com, a joint project between the Center for International Development at Harvard and the Colombian Foreign Trade Bank (Bancoldex) to map the industrial economic activity in Colombia.

Given the “Data Use Agreement” between the Center for International Development at Harvard and the MFPC, we are unable to directly share the raw data. However, there is a protocol for gaining access to the data, by writing a collaboration proposal between a researcher and either the Ministry of Finance and Public Credit (MFPC) of Colombia or the Ministry of Health and Social Protection (MHSP) of Colombia.

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Mathematical Institute, Oxford University, Andrew Wiles Building, Radcliffe Observatory Quarter, OX2 6GG Oxford, UK.

²RiSE, Universidad Eafit, Medellin, Colombia. ³CID, Harvard Kennedy School of Government, Cambridge, USA.

Received: 20 June 2019 Accepted: 13 November 2019

Published online: 16 December 2019

References

- Acemoglu D, Garcia-Jimeno C, Robinson JA (2015) State capacity and economic development: A network approach. *Am Econ Rev* 105(8):2364–2409
- Ahrend R, Farchy E, Kaplanis I, Lembcke A (2014) What makes cities more productive? evidence on the role of urban governance from five OECD countries. *OECD Reg Dev Work Pap* 1(2014/05)
- Arcaute E, Hatna E, Ferguson P, Youn H, Johansson A, Batty M (2015) Constructing cities, deconstructing scaling laws. *J R Soc Interface* 12(102):20140745

- Barrios S, Bertinelli L, Heinen A, Strobl E (2012) Exploring the existence of local and global knowledge spillovers: Evidence from plant-level data. *Scand J Econ* 114(3):856–880
- Barro RJ, Martin XS (2003) *Economic Growth*. 2nd edition. MIT Press
- Barlik TJ (1991) Who benefits from state and local economic development policies? Technical report, WE Upjohn Institute for Employment Research
- Batty M (2009) Accessibility: in search of a unified theory
- Bertaud A (2004) The spatial organization of cities: Deliberate outcome or unforeseen consequence?
- Bettencourt LMA, Lobo J, Helbing D, Kühnert C, West GB (2007) Growth, innovation, scaling, and the pace of life in cities. *Proc Natl Acad Sci* 104(17):7301–7306
- Bettencourt LMA, Samaniego H, Youn H (2014) Professional diversity and the productivity of cities. *Sci Rep* 4:5393
- Black DA, Kolesnikova N, Taylor LJ (2014) Why do so few women work in New York (and so many in Minneapolis)? labor supply of married women across US cities. *J Urban Econ* 79:59–71
- Camagni R, Capello R, Caragliu A (2015) The rise of second-rank cities: what role for agglomeration economies? *Eur Plann Stud* 23(6):1069–1089
- Castells-Quintana D (2017) Malthus living in a slum: Urban concentration, infrastructure and economic growth. *J Urban Econ* 98:158–173
- Chang SJ (1996) An evolutionary perspective on diversification and corporate restructuring: entry, exit, and economic performance during 1981–89. *Strat Manag J* 17(8):587–611
- Combes P-P, Duranton G, Gobillon L (2010) The identification of agglomeration economies. *J Econ Geogr* 11(2):253–266
- Prieto Curiel R, Pappalardo L, Gabrielli L, Bishop SR (2018) Gravity and scaling laws of city to city migration. *PLoS ONE* 13(7):e0199892
- Duranton G (2015a) Delineating metropolitan areas: Measuring spatial labour market networks through commuting patterns. In: *The Economics of Interfirm Networks*. Springer. pp 107–133. https://doi.org/10.1007/978-4-431-55390-8_6
- Duranton G (2015b) Delineating Metropolitan Areas: Measuring Spatial Labour Market Networks Through Commuting Patterns. Springer Japan, Tokyo. https://doi.org/10.1007/978-4-431-55390-8_6
- Duranton G, Overman HG (2005) Testing for localization using micro-geographic data. *Rev Econ Stud* 72(4):1077–1106
- Duranton G, Puga D (2001) Nursery cities: Urban diversity, process innovation, and the life cycle of products. *Am Econ Rev*:1454–1477. <https://doi.org/10.1257/aer.91.5.1454>
- Duranton G, Puga D (2004a) Micro-foundations of urban agglomeration economies. *Handb Reg Urban Econ* 4:2063–2117
- Duranton G, Puga D (2004b) Micro-foundations of urban agglomeration economies. *Handb Reg Urban Econ* 4:2063–2117
- Duranton G, Turner MA (2012) Urban growth and transportation. *Rev Econ Stud* 79(4):1407–1440
- Duttagupta I (2019) What India makes: List of India's manufacturing hubs. <https://economictimes.indiatimes.com/news/economy/infrastructure/what-india-makes-list-of-indias-manufacturing-hubs/articleshow/64615887.cms>
- Fan JPH, Lang LHP (2000) The measurement of relatedness: An application to corporate diversification*. *J Bus* 73(4):629–660
- Farjoun M (1994) Beyond industry boundaries: Human expertise, diversification and resource-related industry groups. *Organ Sci* 5(2):185–199
- Ferreira MM, Roberts M (2018) *Raising the bar for productive cities in Latin America and the Caribbean*. World Bank Publications, Washington
- Feser EJ, Sweeney SH (2002) Spatially binding linkages in manufacturing product chains. In: McNaughton RB, Green MB (eds). *Global Competition and Local Networks*. pp 111–129
- Frenken K, Boschma RA (2007) A theoretical framework for evolutionary economic geography: Industrial dynamics and urban growth as a branching process. *J Econ Geogr* 7:635–649
- Frick SA, Rodríguez-Pose A (2018) Big or small cities? on city size and economic growth. *Growth Chang* 49(1):4–32
- Friedrichs J (1993) A theory of urban decline: Economy, demography and political elites. *Urban Stud* 30(6):907–917
- Gallup JL, Gavia A, Lora E (2003) Is Geography Destiny? Lessons from Latin America. The World Bank. <https://elibrary.worldbank.org/doi/abs/10.1596/0-8213-5451-5>
- Glaeser EL (2014) A world of cities: The causes and consequences of urbanization in poorer countries. *J Eur Econ Assoc* 12(5):1154–1199
- Goldner W (1955) Spatial and locational aspects of metropolitan labor markets. *Am Econ Rev* 45(1):113–128
- Hausmann R, Hidalgo CA (2011) The network structure of economic output. *J Econ Growth* 16:309–342. <https://doi.org/10.1007/s10887-011-9071-4>
- Hausmann R, Hidalgo CA, Stock DP, Yildirim MA (2014a) Implied comparative advantage. CID Working Paper. <https://doi.org/10.2139/ssrn.2410427>
- Hausmann R, Hidalgo CA, Bustos S, Coscia M, Simoes A, Yildirim MA (2014c) *The atlas of economic complexity: Mapping paths to prosperity*. MIT Press
- Hausmann R, Hwang J, Rodrik D (2007a) What you export matters. *J Econ Growth* 12(1):1–25
- Hausmann R, Klinger B, et al (2007b) The structure of the product space and the evolution of comparative advantage. Technical report, Center for International Development at Harvard University
- Hausmann R, Hidalgo CA, Stock D, Yildirim MA (2014b) Implied comparative advantage. HKS Working Paper RWP14-003, Harvard Kennedy School
- Hidalgo CA, Klinger B, Barabási A-L, Hausmann R (2007) The product space conditions the development of nations. *Science* 317(5837):482–487
- Hidalgo CA, Hausmann R (2009) The building blocks of economic complexity. *Proc Natl Acad Sci* 106(26):10570–10575
- Ihlanfeldt KR (1997) Information on the spatial distribution of job opportunities within metropolitan areas. *J Urban Econ* 41(2):218–242
- Ihlanfeldt KR, Sjoquist DL (1998) The spatial mismatch hypothesis: a review of recent studies and their implications for welfare reform. *Hous Policy Debate* 9(4):849–892
- Jones CI (1995) R&D-based models of economic growth. *J Political Econ* 103(4):759–784
- Jones CI (2002) Sources of U.S. Economic Growth in a World of Ideas. *Am Econ Rev* 92(1):220–239

- Jones CI, Romer PM (2010) The new Kaldor facts: Ideas, institutions, population, and human capital. *Am Econ J Macroecon* 2:224–245
- Kain JF (1992) The spatial mismatch hypothesis: three decades later. *Hous Policy Debate* 3(2):371–460
- Kremer M (1993) Population growth and technological change: One million B.C. to 1990. *Q J Econ* 108(3):681–716
- Lotero L, Hurtado RG, Flórida LM, Gómez-Gardeñes J (2016) Rich do not rise early: spatio-temporal patterns in the mobility networks of different socio-economic classes. *R Soc Open Sci* 3(10):150654
- Lucas Jr RE (1988) On the mechanics of economic development. *J Monet Econ* 22(1):3–42
- Madden JF (1981) Why women work closer to home. *Urban Stud* 18(2):181–194
- Marshall A, Marshall MP (1920) *The economics of industry*. Macmillan and Company
- Meijers EJ, Burger MJ (2010) Spatial structure and productivity in US metropolitan areas. *Environ Plan A* 42(6):1383–1402
- Mealy P, Farmer JD, Teytelboym A (2018) A new interpretation of the economic complexity index. *Alexander, A New Interpretation of the Economic Complexity Index* (February 4, 2018)
- Melo PC, Graham DJ, Noland RB (2009) A meta-analysis of estimates of urban agglomeration economies. *Reg Sci Urban Econ* 39(3):332–342
- McCann P, Acs ZJ (2011) Globalization: countries, cities and multinationals. *Reg Stud* 45(1):17–32
- Neffke F, Henning M (2013) Skill relatedness and firm diversification. *Strat Manag J* 34(3):297–316
- Neffke F, Henning M, Boschma R (2011) How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Econ Geogr* 87(3):237–265
- Nelson RR, Winter SG (1982) *An evolutionary theory of economic change*. The Belknap Press of Harvard University Press, Cambridge
- O'Clery N, Gomez-Lievano A, Lora E (2016) The path to labour formality: Urban agglomeration and the emergence of complex industries. HKS Working Paper No. RFWP 78
- OECD (2012) *Redefining Urban: A New Way to Measure Metropolitan Areas*. Publishing and Organisation for Economic Co-operation and Development Staff, Organisation for Economic Cooperation and Development (OECD), Paris
- OECD (2013) Definition of functional urban areas (fua) for the OECD metropolitan database. Organization for Economic Cooperation and Development website. <https://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm>
- Pappalardo L, Vanhoof M, Gabrielli L, Smoreda Z, Pedreschi D, Giannotti F (2016) An analytical framework to nowcast well-being using mobile phone data. *Int J Data Sci Analytics* 2(1-2):75–92
- Piovani D, Arcaute E, Uchoa G, Wilson A, Batty M (2018) Measuring accessibility using gravity and radiation models. *R Soc Open Sci* 5(9):171668
- Porter ME (2011) *Competitive advantage of nations: creating and sustaining superior performance*, volume 2. Simon and Schuster
- Prud'Homme R (1996) Managing megacities. *Le courrier du CNRS* 82:174–176
- Rand WM (1971) Objective criteria for the evaluation of clustering methods. *J Am Stat Assoc* 66(336):846–850
- Romer PM (1986) Increasing returns and long-run growth. *J Political Econ* 94(5):1002–1037
- Romer PM (1990) Endogenous technological change. *J Political Econ* 98(5):S71–S102
- Rosenthal SS, Strange WC (2004) Evidence on the nature and sources of agglomeration economies. *Handb Reg Urban Econ* 4:2119–71
- Rosenthal SS, Strange WC (2006) The micro-empirics of agglomeration economies. *A Companion Urban Econ*:7–23. <https://doi.org/10.1002/9780470996225.ch1>
- Rosenthal SS, Strange WC (2012) Female entrepreneurship, agglomeration, and a new spatial mismatch. *Rev Econ Stat* 94(3):764–788
- Roth C, Kang SM, Batty M, Barthélemy M (2011) Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLoS ONE* 6(1):e15923
- Roth C, Kang SM, Batty M, Barthelemy M (2012) A long-time limit for world subway networks. *J R Soc Interface* 9(75):2540–2550
- Sanchez-Serra D (2016) Functional urban areas in Colombia. OECD Regional Development Working Papers, No. 2016/08. OECD Publishing, Paris. <https://doi.org/10.1787/5jln4pn1zqz5-en>
- Simini F, González MC, Maritan A, Barabá A-L (2012) A universal model for mobility and migration patterns. *Nature* 484(7392):96
- Song C, Qu Z, Blumm N, Barabá A-L (2010) Limits of predictability in human mobility. *Science* 327(5968):1018–1021
- Strano E, Shai S, Dobson S, Barthelemy M (2015) Multiplex networks in metropolitan areas: generic features and local effects. *J R Soc Interface* 12(111):20150651
- TheBrazilBusiness.com (2019) Footwear industry in Brazil. <https://thebrazilbusiness.com/article/footwear-industry-in-brazil>. Accessed May 2019.
- Uchida H, Nelson A, et al (2010) Agglomeration index: Towards a new measure of urban concentration. *Urban Dev Multidiscip Perspect*:41–60. <https://doi.org/10.1093/acprof:oso/9780199590148.003.0003>
- Venables AJ (2007) Evaluating urban transport improvements: cost–benefit analysis in the presence of agglomeration and income taxation. *J Transp Econ Policy (JTEP)* 41(2):173–188
- Weil DN (2012) *Economic growth*. 3rd edition. Prentice Hall, New York
- White MJ (1986) Sex differences in urban commuting patterns. *Am Econ Rev* 76(2):368–372

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.