ISSN 2318-2377

Cedeplar
UFMG

TEXTO PARA DISCUSSÃO N° 650

THE GLOBAL GEOGRAPHY OF DIGITAL PLATFORMS: TOWARDS PLATFORMS INTERNATIONAL LOCATIONAL DETERMINANTS

Victo José da Silva Neto
Tulio Chiarini
Leonardo Costa Ribeiro
Igor Santos Tupy

Novembro de 2022

Textos para Discussão

A série de Textos para Discussão divulga resultados preliminares de estudos desenvolvidos no âmbito do Cedeplar, com o objetivo de compartilhar ideias e obter comentários e críticas da comunidade científica antes de seu envio para publicação final. Os Textos para Discussão do Cedeplar começaram a ser publicados em 1974 e têm se destacado pela diversidade de temas e áreas de pesquisa.

Ficha catalográfica

S586g 2022	The global geography of digital platforms: towards platforms international locational determinants/ Victo José da Silva Neto...[et al.]- Belo Horizonte: UFMG / CEDEPLAR, 2022.
	26 p. : il. - (Texto para discussão, 650)
	Inclui bibliografia.
	ISSN 2318-2377
	1. Economia. 2. Capitalismo. 3. Internet. I. Chiarini, Túlio. II. Ribeiro, Leonardo Costa. III. Tupy, Igor Santos. IV. Universidade Federal de Minas Gerais. Centro de Desenvolvimento e Planejamento Regional. V. Título. VI. Série.
	CD D: 330

Elaborado por Rosilene Santos CRB-6/2527
Biblioteca da FACE/UFMG. – RSS/125/2022

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE CIÊNCIAS ECONÔMICAS
CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL

THE GLOBAL GEOGRAPHY OF DIGITAL PLATFORMS: TOWARDS PLATFORMS INTERNATIONAL LOCATIONAL DETERMINANTS

Victo José da Silva Neto
Nijmegen School of Management, Radboud University
victont@gmail.com

Tulio Chiarini
Centro de Pesquisa em Ciência,Tecnologia e Sociedade, Ipea
tulio.chiarini@ipea.gov.br

Leonardo Costa Ribeiro
Centro de Desenvolvimento e Planejamento Regional, UFMGl
leonardocostaribeiro@gmail.com

Igor Santos Tupy
Departamento de Economia, UFV
Igor.tupy@ufv.br

CEDEPLAR/FACE/UFMG
BELO HORIZONTE
2022

SUMÁRIO

ABSTRACT

Digital platforms have positioned themselves at the center of global flows of capital, knowledge, and work. Their ability to influence and organize these flows makes it imperative to understand the locational decisions of platform companies. This paper explores new evidence on the digital platform economy geography. Our objective is threefold. First, we propose a novel methodology using data science and artificial intelligence tools to identify platform companies. Second, with a set of over three thousand companies, we introduce worldwide maps where it is possible to see the countries and cities that host platform companies. Third, we present platform companies' locational choice using econometric models.While we observe a geographic concentration of platform companies in the U.S. and China, we also see that digital platform companies are spreading to all geographical directions, including tax havens, reinforcing the hypothesis that "platforming" is a worldwide phenomenon.

Keywords: platformization; platform capitalism; natural language processing; Zero-Inflated Negative Binomial regression model; Orbis

JEL code: F01; L86; O33

1. INTRODUCTION

There has been an increasing interest in the literature on the accelerated transformations in the capitalist system caused by digital technologies. These transformations are more and more intense due to the pervasiveness of platforms that are creating new markets and reorganizing traditional industrial sectors (KENNEY; BEARSON; ZYSMAN, 2021), and reshaping the geography of value creation and extraction (KENNEY; ZYSMAN, 2020).

Literature main contributions are focused on the ecosystem of both controlling firms, associated third parties, and users dependent on U.S. and Chinese giant tech titans as "GAFAM" and "BAT"[1]. While there are interesting insights on BAT internationalization strategies (JIA; KENNEY; ZYSMAN, 2018), on BAT involvement with the government (JIA; KENNEY, 2021; MCKNIGHT; KENNEY; BREZNITZ, 2021; SU; FLEW, 2021), on the complexity and multiplicity of ways Amazon and Google Maps (Alphabet) is reorganizing the geography of economic activity within the U.S. (KENNEY; ZYSMAN, 2020) and on the pattern of GAFAM's mergers and acquisitions (GAUTIER; LAMESCH, 2021),there are still gaps in the literature that need to be addressed for understanding how digital platforms produce and distribute value between and within countries, therefore providing a better comprehension of the global geography of the platform economy.

Our objective is threefold. First, we propose a novel methodology using data science and artificial intelligence tools to identify platform companies, following the strategy proposed by Silva, Chiarini and Ribeiro (2022). Our first approach to apply that methodology is to use Orbis database. Doing so, we identify over three thousand companies. Our second objective is to introduce a worldwide map where it is possible to see the countries (and cities) that host platform companies, not only the giant "GAFAM" and "BAT". Third, we present platform companies' locational choice using econometric models.

The remainder of the article is structured as follows. In section 2, we present recent studies on the geography of the platform economy that identify local determinants that attract platform companies. Section 3 develops our methodology to identify what we call platform companies. We also present the methodological choice for a Zero-Inflated Negative Binomial (ZINB) regression model to infer statistically the platform companies' locational choice. Section 4 presents the main results, showing that while we observe a geographic concentration of platform companies in China and in the U.S., we see their diffusion to all geographical directions, reinforcing the hypothesis that "platforming" is a worldwide phenomenon. In addition, our models demonstrate that population size is a key-determinant for the locational choice of platform companies, however, this relationship is not linear. Moreover, our results reinforce the significant attractive power exerted by cities located in the so-called tax heaven countries. We finish the paper by presenting conclusions and the limitations of our research and calling attention that in order to improve our mapping it is necessary to go beyond Orbis' data and use Crunchbase, a commercial database on innovative companies.

[1] Google/Alphabet; Amazon; Facebook/Meta; Apple; and Microsoft and Baidu; Alibaba; and Tencent.

2. THE GEOGRAPHY OF THE PLATFORM ECONOMY

Digital platforms have consolidated themselves as a new organizational model (GAWER, 2021) whose importance and centrality in the process of value creation and appropriation is equivalent to the centrality of factories in the era prior to digitalization (BEARSON; KENNEY; ZYSMAN, 2021; KENNEY; ZYSMAN, 2016). “Platforms are now redefining the scope of market competition, the organization of industrial relations and work process, and influencing the power arrangements across the economy” (BEARSON; KENNEY; ZYSMAN, 2021, p. 23). Given the “new organizational form based on a relationship between the platform and the ecosystem of firms dependent on the platform and users who interact and transact through it” (KENNEY; ZYSMAN, 2020, p. 55), understanding the creation and capture of value across space has been challenging.

Scholars from evolutionary economic geography and international business have been investigating the locational determinants of digital platforms on a global scale. Stalkmap and Schotter (2021), for example, offer a fruitful theorization on the internationalization of digital platforms from the perspective of the geographic scope of network externalities. According to the authors, although all platforms leverage network externalities as firm-specific advantages, it is possible to differentiate between within-country and cross-country network externalities. Borders and distance are elements that constrain the reach of externalities. Digital platforms that mediate the delivery of goods, for example, leverage externalities constrained by distance: consumers and service providers need to be geographically close. Other elements, such as regulation cultural homogeneity, can also lead to within-country network externalities.

Other platforms, such as app stores (Apple Store, Play Store), are constrained locally, neither by borders nor by distance. Thus, their user base benefits from cross-country network externalities. This differentiation implies different strategies for the internationalization of the platforms. Stalkmap and Schotter (2021) raise some points to be empirically tested for the two different groups of platforms: their strategies for entering new international markets (independent, for cross-country; associated with local incumbents, for within-country); their international strategic posture (global strategy for cross-country; multi-domestic strategy for within-country); and their selection of international markets (institutional/cultural proximity).

The role of institutions in the locational decision of digital platforms was investigated by Punt *et al.* (2021), who tested whether Uber's expansion correlates with strong economic, political, and labor institutions. They found evidence that places with solid economic institutions prioritize places, although the evidence is less conclusive for the other two sets. Their tests also indicate that Uber's mobile customer base across cities is a defining element of its expansion strategy. The importance of consumer mobility, both as a latent demand and as a legitimizing and disseminating community, stands as another type of firm-specific advantage leveraged by platforms, a possibility that was highlighted by Stalkmap and Schotter (2021). Shaheer *et al.* (2020) empirically address platform locational decisions from the perspective of the nature of the lead market. They investigate whether acting in specific lead markets benefits digital platforms to expand. The authors differentiate two types of lead markets: consumers with heterogeneity in demand and those with overlapping preferences. Using download data from 1,910 apps in the Apple Store over two years (2016–17) for 57 countries, Shaheer *et al.* (2020) support the hypothesis that operating in lead markets (of both types) accelerates

the diffusion in other markets. Finally, Deng *et al.* (2022) analyze the transactions of a B2B (business to business) platform that facilitates the export of small and medium-sized enterprises (SMEs). They find robust evidence to corroborate that digital platforms allow the rapid internationalization of SMEs, with a significant decrease in costs, which implies higher rates of export continuity and presence in international markets.

The previous studies allow us to identify an important avenue of investigation that is taking shape within the perspective of the geography of platform economy: studies that use econometric models to identify local determinants that attract platforms.

3. METHODOLOGY

3.1. Orbis database

To collect the information of products or services provided by companies, we used Orbis database which currently covers around 425 million companies and entities worldwide[2]. For each company and entity, identification data are available (such as name, address, e-mail, URL, and a brief history); productive activity or line of action (economic activities classification, description of business and products and services); economic-financial indicators (balance sheet containing 26 items, profit and loss accounting containing 26 items and other financial indices containing 33 indices); company ownership structure featuring its parent companies and subsidiaries; among other information.

The information on the products and services provided by Orbis, jointly with natural language processing (NLP), allows the identification of platform companies (understood hereafter as the categories proposed by Bearson et al. (2021), i.e., "platform firm" and "platform-dependent business") and then the possibility to elaborate maps with their precise locations. Thus, we process Orbis' fields that present companies' history and their products and services descriptions. As both fields contain unstructured texts, we resort to NLP to extract the meaningful information for our analysis: the products and services provided by companies.

3.2. Natural language processing (NLP)

NLP is a field of Artificial Intelligence (AI) that makes it possible to extract information from unstructured texts that do not present metadata and cannot be easily mapped into predefined fields of a database. NLP combines linguistics and computer science to analyze the rules and structure of language and creates applications capable of understanding, analyzing, and extracting meaning from texts. Therefore, NLP is used to understand the structure and meaning of the human language, analyzing different aspects such as syntax, semantics, and morphology, transforming this linguistic knowledge into algorithms that extract structured information from unstructured texts (INDURKHYA;

[2] 30% of companies and entities are in Europe; 27% in Asia; 17% in North America; 14% in Latin America and the Caribbean; 8% in Oceania; and, 5% in Middle East and Africa (BVD, 2020). Despite covering both privately and publicly traded companies, the minority of them (about 40 million) is held under private ownership.

DAMERAU, 2010).

NLP algorithms create a vector representation of the words, thus transforming a text into something a machine can manage through mathematical operations. With this vector representation, AI algorithms are trained by associating the input text (now a set of vectors) and the characteristics we want to extract. In this step, NLP uses supervised AI algorithms that require a training base to identify the association patterns of the input and output variables of the algorithm's problem. We use a corpus freely written by their authors as a training base and in general it contains literary works, Wikipedia pages, news transmitted through Google News, among others. As part of the NLP process related to this paper, we can mention:

- Tokenization[3] breaks a sequence of words into smaller semantical units called tokens. Phrase tokenization divides the continuous text into different phrases identifying the beginning and end of each, while word tokenization divides a phrase into the different words that compose it. Word tokens are usually separated by whitespace and sentence tokens by punctuation symbols. However, there are also more complex structures, such words that usually come together as collocations and phrasal verbs. To illustrate the tokenization of words, see how the following sentence is tokenized: Customer service could not be better! = "Customer service", "could", "not", "be", "better".

- Marking Part of Speech (PoS) involves adding a category to identify the grammatical class to each token. PoS markup is essential for identifying the relationships between words and understanding the meaning of sentences. Common PoS tags are verbs, adjectives, nouns, pronouns, conjunctions, prepositions, and intersections. In this case, the words of the example above will be associated with the following tags: "Customer service": NOUN, "no": ADVER, "could": VERB, "be": VERB, "better": ADJECTIVE, "! ": PUNCTUATION.

- Dependency analysis: Dependency grammar refers to the way words in a sentence are connected. Therefore, an algorithm identifies how the "headwords" are related and modified by other words to understand a sentence's syntactic structure. The dependency analysis marker identifies grammatical structures such as subject, verb, direct and indirect object, and predicate.

3.3. Identifying the different products and services of the companies

Information about the products and services of the companies contained in Orbis is available in the "Description and history", "Product and services" and "Trade description" fields. They are unstructured texts and contain other information (see TABLE 4 in Annex for an example). Thus, it is necessary to use NLP tools to identify the sentences in which products and services are described and get these products.

Even though Orbis “capture[s], treat[s] and standardize[s] data from a wide range of sources to provide (…) value-added company information” (BVD, 2020, p. 03) about public and private firms

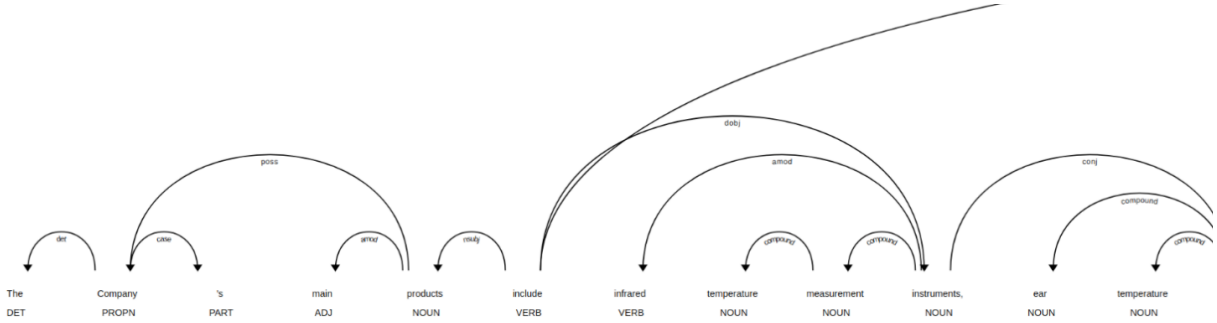
[3] We used the Python spaCy library (https://spacy.io).

(including bank and insurance companies) from all countries, there are lacking information in the database. To give an idea of Orbis's complexity, richness, and limitations, we present Table 2 (Annex), containing a sample of an identified firm: MercadoLibre. The company has its headquarters in Buenos Aires and is the leading e-commerce Latin American platform whose biggest market in the region is Brazil, and it represents 55% of the firm's total income (ALTIMARI, 2021). Note, however, that Orbis displays three results for MercadoLibre: one firm located in Argentina, one in Colombia, and another in the U.S.. There is much more information available for the U.S. firm, while for the Colombian counterpart, there are just a few details. For the headquarters, there is no information at all.

Our first step is identifying the words marked as a verb by the PoS. We identified the verbs following the procedure described above using a sample of 150,000 companies indexed by Orbis. We chose to locate the verbs because this would be the easiest way to identify the action related to each sentence to identify later the one associated with production. Then, among all verbs identified, we picked up those associated with phrases that effectively describe the products and services of the companies and, considering those with occurrence greater than 100 (relative frequency above 0.1%)[4]

After identifying the verbs associated with the products and services, in the second step, we selected only phrases that present such verbs in "Description and history", "Product and services" and "Trade description" fields and, using the spaCy dependency analysis markup, we identified the direct or indirect objects associated with them. Therefore, we suppose that those objects are the products and services the companies provide. FIGURE 1 illustrates the dependency analysis and shows how such marking allows the identification of direct and indirect objects.

FIGURE 1
Example of dependency analysis markup



Source: Authors' own.

[4] We obtain the following list of verbs associated with the products and services: engag*, provid* includ*, offer*, sell*, produc*, manufactur*, rent*, develop*, mak*, specializ*, develop*, distribut*, deliver*, design*, process*, fabricat*, focus*, engag*, and, forg*.

3.4. Applying the NLP to support the identification of platform companies

Our first step is to depart from twenty digital multisided platform companies[5] listed in Fortune's Digital 100 identified by Acs *et al*. (2021) and retrieve their product/service description texts on Orbis, using the fields: "Description and history", "Product and services", and "Trade description" as shown in TABLE 4. Then, we applied the NLP described in the previous section to identify products and services provided by those digital multisided platform companies, which allowed us to create a first list with the 37 most often terms (TABLE 5, column 1, Annex). The result showed over 16 thousand firms.

We gathered the terms related to platforms from that first list (TABLE 5, Annex), and while adding up other 12 terms known to be related to this area, we excluded other 20 once they resulted in many "false negative" firms (TABLE 5, column 2, Annex). We then implement another search on Orbis, looking up the terms of this second list in the fields that presented information regarding products and services. Thus, we got a broader set of firms whose information we also retrieved and analyzed using NLP as in the previous step. We updated our second list, including the other seven terms related to platforms, and excluding three terms from the second list and our final list had 33 terms (TABLE 5, final column, Annex) which allowed us to identify 3,147 platform companies.

To reinforce confidence in our methodology, we carried out a robustness test. We randomly selected 10% of the 3,147 companies and, independently, two of the authors evaluated them one by one and classified them as either "yes" or "no" as a platform company. In general, the criterion was the use of network effects in a digital environment to label them as "yes". There was an intercoding variation of 14% between the encoders. In a new round, both aligned the concepts together to reach a final classification. As a result, we arrived at a set of 102 companies that did not fit the criteria of "platform company". Considering the random sample of 300 companies, this results in a confidence percentage of 66% in the developed algorithm. It should be noted that the classification criterion was quite conservative in order to avoid false positives (e.g., 55 Chinese companies with insufficient data for judgment were considered as "non-platform companies"[6]).

3.5. Econometric model

To identify local determinants that attract platform companies we estimate a Zero-Inflated Negative Binomial (ZINB) regression model to infer statistically companies' locational choice. The dependent variable is the number of platform companies per city. Due to the nature of our dependent variable, which is non-negative integer, count data models – such as Poisson and Negative Binomial – should be used (HUANG *et al.*, 2022). Thus, we model our data as follows:

$$n_i = f\left(x_i\right) + \varepsilon_i$$

[5] They are Activision Blizzard, Alibaba; Alphabet; Amazon; Apple; Baidu; Booking Holdings; eBay; Facebook; Fidelity National Information; Fiserv; JD.com; Microsoft; Naspers; NetEase; Nintendo; PayPal; Rakuten; Recruit Holdings; and Tencent.

[6] The Only information available for all of these companies were "Specialized in operating web sites that use a search engine to generate and maintain extensive databases of Internet addresses and content in an easily searchable format"

where n_i is the number of platform companies in each city, x_i is a vector of local characteristics and ε_i is the random error term. Regarding the explanatory variables, the population (in million inhabitants) is included to measure the effect of the size of local markets on the presence of platform companies. The squared population is used to capture nonlinearities in the relation between population and the number of companies. Locational dummies are inserted to measure the influence of the cities located in the U.S., China, the European Union, and in tax heaven countries.

The variance of the company count (0.72051) is much larger than its mean (0.018929), which evidences the data overdispersion and suggests that Negative Binomial regression models (NBRM) are more appropriate than Poisson Models (PRM)[7]. Another advantage of a NBRM is that it accounts for omitted variable bias (ALMEIDA; HOHBERGER; PARADA, 2011). Thus, the parameter is included in the model to correct overdispersion in data by capturing unobserved heterogeneity (WENTING; FRENKEN, 2011).

Furthermore, the excess of zeros in the municipalities data suggests the use of Zero-Inflated Negative Binomial (ZINB) models[8]. A Zero-Inflated Model is based on a two-stage process to account for zeros in the data-generating process (ANDREWS *et al.*, 2020). This way, some municipalities may never attract a platform company, which implies that the outcome will aways be zero, and to which the impact of the variables in x_i will be captured by the logit "inflated" model. On the other hand, if there is the possibility of the municipality to attract platform companies – even though in some periods it presents zero as outcome – the company count will be predicted by a negative binomial model (HOLL; MARIOTTI, 2018). Given its characteristics, ZINB Models have been largely used in models of companies' count data and in models of firms' location choice (ANDREWS *et al.*, 2020; GHIO; GUERINI; ROSSI-LAMASTRA, 2016; HOLL; MARIOTTI, 2018; HUANG *et al.*, 2022).

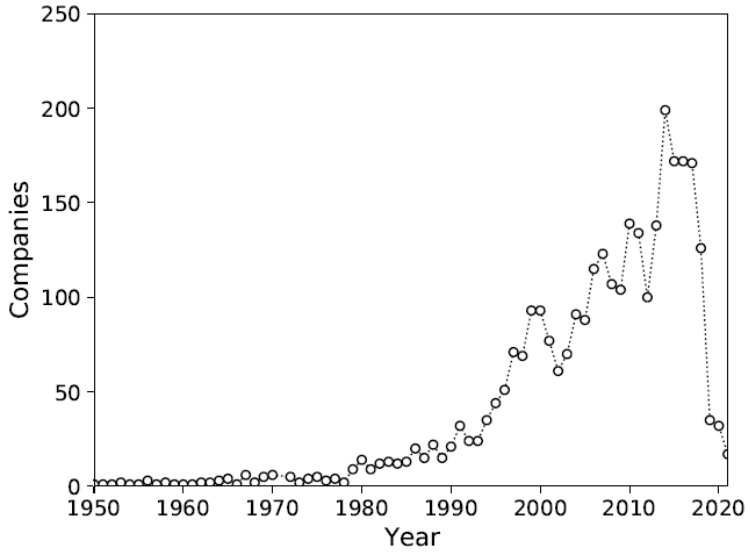
4. WORLD PLATFORM-RELATED ECONOMY

Most platform companies were founded after the commercial opening of the Internet in the mid-1990s (). There are also two very noticeable and historically well-characterized growth spikes. The first, in late 1990s, concerns the founding's (and financing) boom of "Internet companies" that culminated in the Dot-com crisis in 2000. The growth in the number of new companies resumed from approximately 2002 onwards, to suffer a sharp retraction with the 2008 financial crisis. Finally, the 2010s witnessed the expansion of the platform model, reaching the mark of more than 300 companies a year in the middle of the decade. These data corroborate the perception that we live in the era of "platformania" (CUSUMANO; GAWER; YOFFIE, 2019), and we are under a "platform revolution" (PARKER; ALSTYNE; CHOUDARY, 2016).

[7] Poisson regression models assumes that variance and the mean of the dependent variable should be equal (HOLL; MARIOTTI, 2018; SAFARI, 2017). A significant Likelihood-Ratio Test (LR) with $\alpha = 0$ and the significant parameter confirm the overdispersion and the Information Criteria (Akaike's Information Criterium, AIC, and Bayesian Information Criterium, BIC) shows that NBRM presented a higher fit to the data compared to PRM.

[8] The better adjustment of the ZINB compared to standard Negative Binomial is confirmed by Vuong Test (CAMERON; TRIVEDI, 2009), and by the Information Criteria (AIC, and BIC). Also, a likelihood ratio (LR) test confirms that ZINB model is superior to the Zero-Inflated Poisson Model (ZIP).

FIGURE 2
Platform companies by incorporation year, 1950–2021



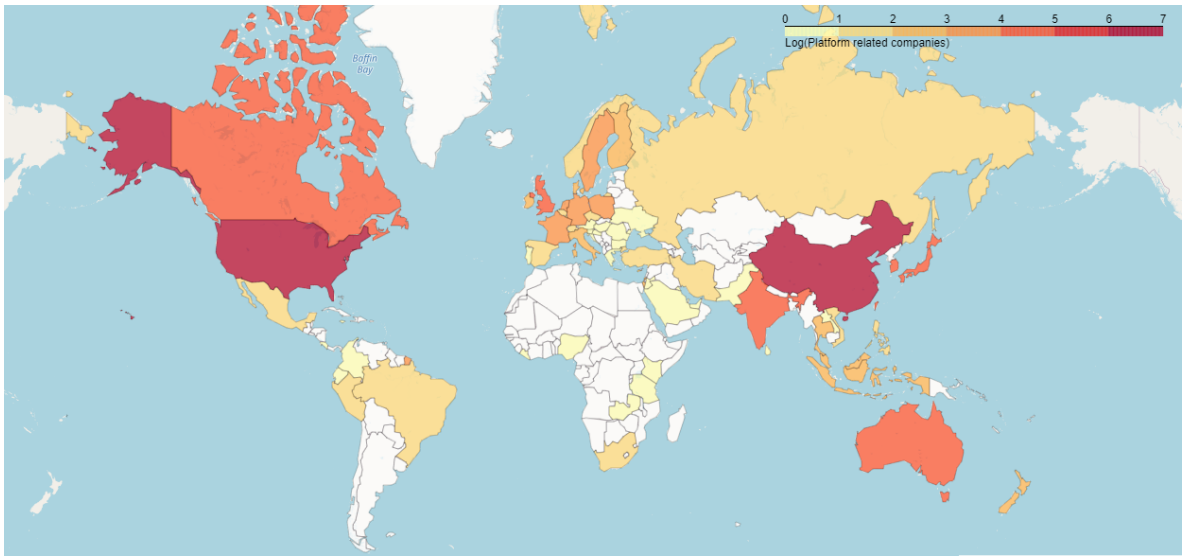
Source: Authors' own. Data sourced from Orbis. Note: the drop in recent years may be due to backlog.

It should be noted as well that there is also a considerable number of companies whose foundation dates back the pre-Internet era. Many companies have been following the sector's evolution since before the Internet age. There is probably another group that, although not dedicated initially to digital services, transformed its organizational model to include platform services/products at some point. In short, even if "born digital" are predominant, there is a considerable number of "analog companies" that have carved out their place in the platform economy.

4.1. Country-level distribution of digital platform companies

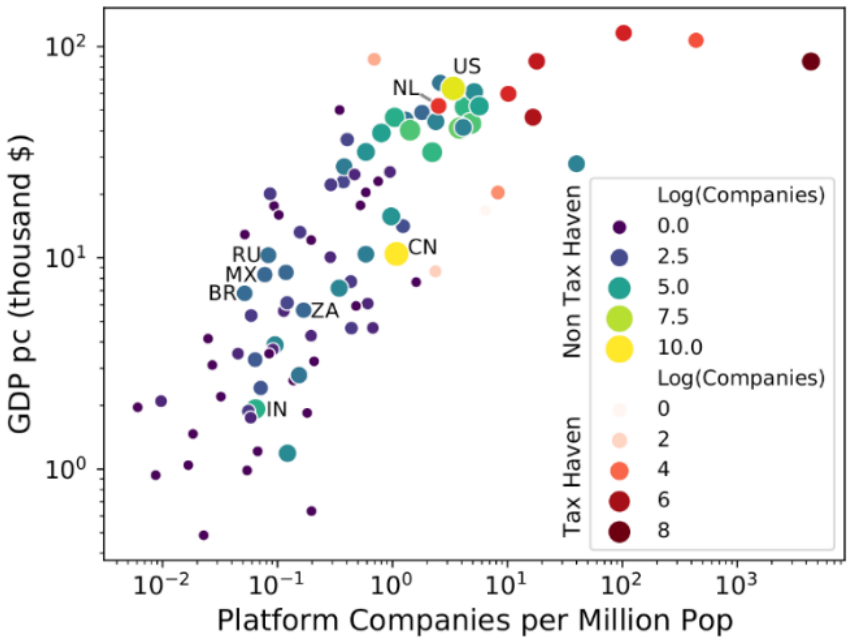
In , we plot the countries that host platform companies. We colored them according to the concentration of digital platforms at the national level. The main takeaway from is that the platform economy is a global phenomenon, as advocated by Kenney and Zysman (2016), not only restricted to the Global North, but it also spreads, albeit unevenly, to the Global South. However, there is a direct correlation between the number of platform companies per million inhabitants and GDP per capita (FIGURE 4).

FIGURE 3
Concentration of platform companies by country



Source: Authors' own. Data sourced from Orbis.

FIGURE 4
Relation between 2019 GDP per capita and platform companies per million inhabitants



Source: Authors' own. Data sourced from the World Bank (for GDP), Orbis (for Platform companies), and Oxfam International (for corporate tax havens).

The recurrence of platform studies focused on the U.S. and Chinese cases, as demonstrated in section 2) is not by chance. The map illustrates how the U.S. and China are the two main world poles of the platform economy, concentrating respectively 27.2% and 32.8% of the world's platform companies (TABLE 1). That corroborates once again their lead in the global platform race.

TABLE 1
Platform companies by selected countries

Countries	N.	%
China	1,031	32.8
United States of America	855	27.2
“Tax haven” countries	320	10.2
Cayman Islands	165	5.2
Singapore	57	1.8
Netherlands	20	0.6
Hong Kong	18	0.6
Bermuda	17	0.5
Ireland	16	0.5
Other Tax haven” countries	27	0.9
Great Britain	124	3.9
Japan	117	3.7
South Korea	84	2.7
Australia	73	2.3
Taiwan	65	2.1
India	62	2.0
Canada	59	1.9
Other countries	357	11.3
Total	3,147	100.0

Source: Authors’ own. Data sourced from Orbis.

Countries commonly known as "tax havens"[9] concentrate 10.2% of the world’s platform companies (TABLE 1). The Cayman Islands, for instance, ranks in the third position, only after China and the U.S. (TABLE 1).

4.2. City-level distribution of digital platform companies

The visualization of FIGURE 5 makes it possible to identify an uneven distribution of firms within China and the U.S.. China’s platform companies cluster in the east of the country, where only six cities (Hangzhou, Shanghai, Shenzhen, Guangzhou, and Zhuhai) concentrate 37% of Chinese platform companies. Shenzhen is the city with most platform companies globally, followed by Beijing. Shanghai features in the fifth position (). There are also firms in Central and Western China, being

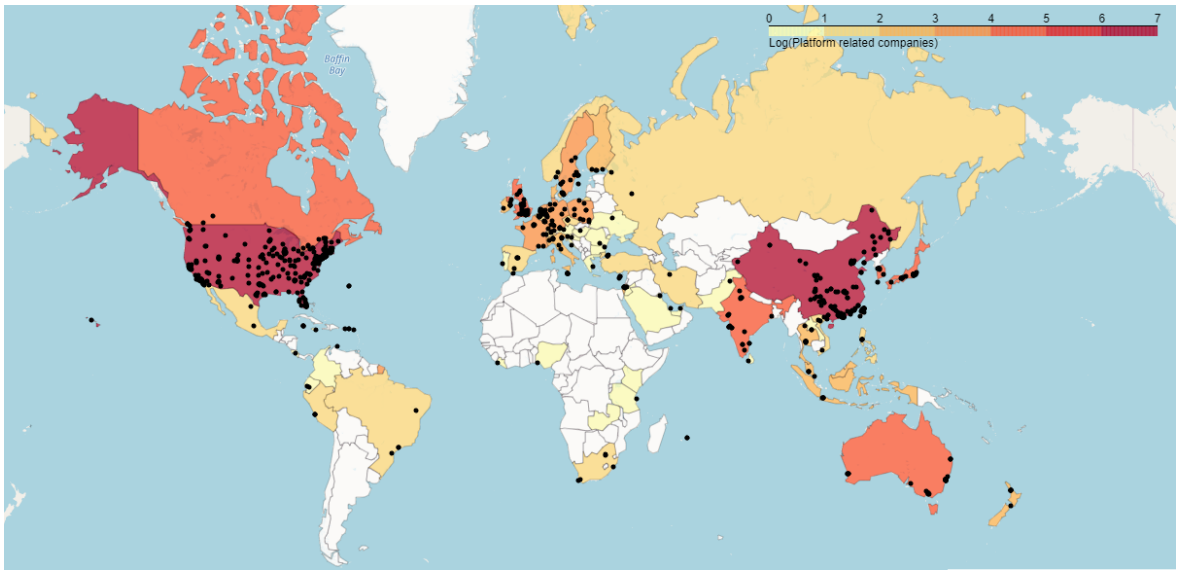
[9] Following Oxfam International, corporate tax havens are: Bermuda, Cayman Island, the Netherlands, Switzerland, Singapore, Ireland, Luxembourg, Curacao, Hong Kong, Cyprus, Bahamas, Jersey, Barbados, Mauritius, and, British Virgin Islands.

Xian, Wuhan, Chongqing and Taiyuan featured within the world top-30.

North-American platform companies locate on the east and west coasts, and their density is much lower in the Midwest. The agglomeration is mainly in California, where a few cities concentrate 17% of U.S. platform companies (San Francisco, San Jose, Los Angeles, Irvine, Sunnyvale, Santa Clara, Santa Monica, Wilmington, San Mateo, San Diego, Palo Alto and Redwood City) and in the Boston-New York-Baltimore polygon whose concentration arrives at 12%. New York is the U.S. city with the most platform companies and features fourth in the global ranking ().

Even though China and the U.S. are the two countries with the most platform companies in the world, FIGURE 5 and TABLE 1 also illustrate other countries with a high concentration of platforms: Great Britain (which seems to have a relatively more equal geographical distribution of platform companies if compared to other countries), Japan (mainly in Tokyo metropolitan area), , , South Korea (Seoul metropolitan area), Australia (mainly in Sydney and Melbourne), and Canada (mainly concentrated in Vancouver and Toronto areas). depicts as well, to a much lesser extent, the presence of the platform company model across other European countries such as Sweden, France, the Netherlands, Italy, and Denmark.

FIGURE 5
Concentration of platform companies by location of cities



Source: Authors' own. Data sourced from Orbis.

TABLE 2
Top 20 cities with most platform companies

Ranking	Cities	Countries	N.	Ranking	Cities	Countries	N.
1	Shenzhen	CN	231	14	Wuhan	CN	29
2	Beijing	CN	102	15	Stockholm	SE	25
3	George Town	KY*	100		Taipei	TW	25
4	New York	US	76	16	Zhuhai	CN	24
5	Shanghai	CN	74		San Jose	US	24
6	London	GB	73	17	Seongnam	KR	23
7	Tokyo	JP	64	18	Vancouver	CA	22
8	Singapore	SG*	56	19	Chicago	US	21
9	Seoul	KR	52		Mumbai	IN	21
10	Guangzhou	CN	49	20	Hangzhou	CN	17
11	Las Vegas	US	34		Hefei	CN	17
12	San Francisco	US	32		Toronto	CA	17
13	Xian	CN	31				

Source: Authors' own. Data sourced from Orbis. Note: (*) Tax haven countries. Cities were available for 95% of our database

There are also geographic voids, mainly in the Global South[10]. Notwithstanding that, there are in those areas relatively more economic dynamic centers such as, Bangkok (Thailand), São Paulo (Brazil), Tehran (Iran), and Nairobi (Kenya) where we observe the (timid) presence of platform companies. Mumbai and Bangalore (India), especially the former that appears within the world top-30.

It is also noticed the concentration in tax haven countries. Georgetown (KY) is the third city in the world with the highest number of those firms (). For instance, PagSeguro – a Brazilian fintech platform company engaged in the operation and management of a mobile payment-based e-commerce service for commercial operations – was established in São Paulo in 2006 and was the fastest-growing company in the sector in the country (SACHS, 2018). Although its development office (PagSeguro Internet S.A.) is still located in Brazil, its headquarters (PagSeguro Digital Ltd) has been in Georgetown since 2018. PagSeguro's offices are portrayed in , one tiny dot in Brazil and another in the Caribbean.

4.3. Econometric models

A Zero-Inflated Negative Binomial Regression Model was estimated to analyze the determinants of the location of platform companies worldwide. Results in general confirm our previous discussion. The LR test on α, rejects the null hypothesis of $\alpha = 0$, and therefore confirms the presence of overdispersion in data. TABLE 3, below, summarizes the main results. Fit statistics and information criteria to compare ZINB models with other count models are in TABLE 6, in Appendix.

[10] Although we cannot point out the reason for the geographic gaps, some possible explanations that can be investigated are: Orbis indexes only larger companies or public companies, which favors finding a greater concentration in countries where the platformization originated. In other words, the database would not be adequate to capture startups and smaller companies that, as we know, started a catch-up movement in the countries of the Global South. Another possible explanation is the lack of telecommunications infrastructure in the Global South, which presents a considerable risk for digital multinationals (Nambisan, Luo, 2022).

TABLE 3
Results of the Zero-Inflated Negative Binomial Regression Model

Variables	Negative Binomial Model	Logit Model –"inflate"	Statistics
	ni>0	*ni*=0	
Constant	-3.639***	4.467***	
	(0.617)	(0.748)	
pop	1.351***	-55.21**	
	(0.500)	(24.44)	
pop2	-0.0523**	2.436**	
	(0.0229)	(1.091)	
US	2.602***	-0.794	
	(0.422)	(0.854)	
China	1.650***	5.973***	
	(0.365)	(1.957)	
UE27	1.274***	0.675	
	(0.439)	(0.701)	
TaxHeaven	4.449***	1.413	
	(1.044)	(1.103)	
Ln(alpha)			1.646***
			(0.405)
LR test on α=0 (X^2)			3193.24**
Wald X^2			183.14***
LL (log-likelihood)			-3636
Total observations			127,371
Nonzero observations			688

Source: Author's own. Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses (adjusted for 246 clusters in country level).

The logit model represents the first stage of the estimation, and captures the determinants of the likelihood of a specific city to have no company, that is, to not being attractive for platform companies. In this inflate model, as expected, population is a key determinant, with a significant, negative, and high-valued coefficient. As a consequence, small and medium size cities in terms of their population are most likely not to have potentially to attract platform companies, and therefore, have "certain" zeros in the count. This probability reduces significantly with as the city population grows.

The NBRM part of the models predicts the number of platform companies in those cities that have possibility to attract them. All the included variables were statistically significant. The population coefficient is positive and highly significant. Moreover, given that the coefficient of the squared population is negative, we can inquire that the relation between population and the number of platform companies has an inverted-U shape. This result indicates economies of agglomeration are key determinants of the locational choice of platform companies, although its effect presents decreasing rates.

The country level dummies were all significant and positively related with the number of companies. The higher effect on the company count is verified in “Tax Heaven” countries. Cities in those countries have 4.5% more platform companies. Cities in the United States, for its turn, presents 2.6 % more platform companies, on average. The impact of being located in China and in the European Union are both positive, but considerably lower than the cities in tax heaven and in the US.

5. CONCLUSION AND CHALLENGES

The paper's first objective was to present a novel methodology using data science and artificial intelligence tools to identify platform companies, the second objective was to locate those firms geographically at the city level and third was to present platform companies’ locational choice using econometric models. We spatially located over three thousand platform companies, reinforcing the hypothesis that "platforming" is a worldwide phenomenon. In summary, we observed that our group is formed by a majority of companies "born in the digital era", accompanied by a non-negligible number of "traditional" companies; they are concentrated in China and in the U.S. with a substantial presence in the Great Britain, Japan and South Korea. While there is evidence that shows there are platform companies in some dynamic countries in the Global South such as India, Thailand, Brazil, Iran, and Kenya, a great deal of them is located in tax haven countries as the Cayman Islands, Singapore, the Netherlands, Hong Kong, Bermuda and Ireland. In other to statistically infer the determinants of the location of platform companies worldwide, we run econometric models that suggested that population size is a key-determinant of the locational choice of platform companies, but this relationship is not linear. Moreover, the ZINB result reinforced the significant attraction power exerted by cities in tax heaven countries.

Although we used advanced NLP techniques to screen what we called "platform companies" in an extensive database, there are significant limitations. The first one relates to our starting point. We used a small list of twenty big platform companies to select our first query words. A second limitation refers to the constraints related to the database itself. We were able to identify thousands of companies from the description of their products and services. However, we have no control over the quality of the information available[11]. Even though missing information is a small portion, there is no pattern of information presented by firms in Orbis.

Even though we identified over three thousand companies, many other platform companies are not in Orbis once they are startups. Since we are trying to see the big picture of the platform economy, we are missing essential infant companies, especially in many dynamic and creative cities in the Global South. Therefore, to improve our research we should also consider other databases which

[11] In some cases, there are many detailed materials, and in others, the description is so limited that it hardly ever describes precisely the products the firm offers. That happened, for example, in the case of MercadoLibre, as we presented in the paper. It was not considered in our map in Argentina, even though we know it is one of the most critical platform companies from Latin America, because its description in Orbis, as demonstrated in TABLE 4, was very superficial and inaccurate and did not use any of our query words (TABLE 5, Annex). MecardoLibre information for its headquarters in Buenos Aires was only "[it] operates an online trading platform in Latin America." Our algorithm was not able to match our query words with that sentence. It may also partially explain the void in some parts of the map, as in Latin America: how many other companies in the region had slight information available, as MercadoLibre, which were not captured by our algorithm? One possible way to avoid that is to use more keywords and "train" our algorithm using another dataset.

covers early-stage startups as Crunchbase. Several scholars have been demonstrating interest in assessing the potential of the database for economic and managerial research (BESTEN, 2020; DALLE; BESTEN; MENONI, 2017) and for exploring innovative ecosystems (KEMENY; NATHAN; ALMEER, 2017), however, its full potentially is far from being exhausted.

Another limitation of our approach derives from the terms retrieved in the query related to a digital platform (TABLE 5, Annex): some words may result in some false-negative firms, i.e., they may be identified as platform companies by our algorithm, but they are not. Our robustness test showed that 33% of our database may be firms that are not at all platform companies. Finally, our study does not capture all three main actors of digital platform ecosystems: platform firm (owner/sponsor/controller), platform-dependent business (complementor/third-party), and users (or consumers/prosumers). Our database comprises "platform firms" and "platform-dependent business"; however, no prosumers are on our list. This is an important caveat to be considered once we are not covering the whole platform ecosystem but parts of it. For a complete picture, it would be necessary to complement more data.

Even so, fulfilling the function of an exploratory study, our work raises promising paths for investigation. We highlight the geographical concentration of platform firms in tax havens, which dialogues directly with the agenda proposed by Kenney and Zysman (2020, p. 72) of “measure the amount of value that these platforms extract from users in developing countries.” Second, we intend to expand our analysis using other information available at Orbis, regarding data from offices and headquarters from all 3,147 platform companies mapped. This will allow us to identify a network of connections and generate a ranking of the central cities of the platform economy in the World Cities style (BRAIL, 2020). This ranking based on the number of companies and connections between headquarters and subsidiary offices in a city level will allow a glimpse of “where the power and value will be concentrated” (KENNEY; ZYSMAN, 2020, p. 72) in the platform economy.

ACKNOWLEDGES

The data analyzed here were obtained within the scope of the NMG2R2\100168 project carried out at King’s College London in collaboration with Valbona Muzaka, to whom we are grateful.

REFERENCES

ACS, Zoltan J. *et al.* The Evolution of the Global Digital Platform Economy: 1971-2021. *SSRN Electronic Journal*, 2021. Disponível em: <https://www.ssrn.com/abstract=3785411>.

ALMEIDA, Paul; HOHBERGER, Jan; PARADA, Pedro. Individual scientific collaborations and firm-level innovation. *Industrial and Corporate Change*, v. 20, n. 6, p. 1571–1599, 1 dez. 2011. Disponível em: <https://academic.oup.com/icc/article-lookup/doi/10.1093/icc/dtr030>.

ALTIMARI, Juan Diego Nicoletti. *Valutación: Mercado Libre Inc.* 2021. 90 f. Universidad de San Andrés, 2021.

ANDREWS, Rhys *et al.* Corporatization in the Public Sector: Explaining the Growth of Local Government Companies. *Public Administration Review*, v. 80, n. 3, p. 482–493, 22 maio 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/puar.13052>.

BEARSON, Dafna; KENNEY, Martin; ZYSMAN, John. Measuring the impacts of labor in the platform economy: new work created, old work reorganized, and value creation reconfigured. *Industrial and Corporate Change*, v. 30, n. 3, p. 536–563, 21 out. 2021. Disponível em: <https://academic.oup.com/icc/article/30/3/536/6032790>.

BESTEN, Matthijs L. Crunchbase Research: Monitoring Entrepreneurship Research in the Age of Big Data. *SSRN Electronic Journal*, p. 28, 2020.

BRAIL, Shauna. World cities of ride-hailing. *Urban Geography*, p. 1–22, 5 jun. 2020. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/02723638.2020.1775030>.

BVD. *Orbis. The world's most powerful comparable data resource on private companies*. . Brussels: Bureau van Dijk: Bureau van Dijk Electronic Publishing Ltd. Disponível em: <https://www.bvdinfo.com/en-gb/-/media/brochure-library/orbis.pdf>. , 2020

CUSUMANO, Michael A.; GAWER, Annabelle; YOFFIE, David B. *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*. [S.l.]: Harper Business, 2019.

DALLE, Jean-Michel; BESTEN, Matthijs Den; MENONI, Carlo. *Using Crunchbase for economic and managerial research*. , OECD Science, Technology and Industry Working Papers., nº No. 2017/08. Paris: Organisation for Economic Co-operation and Development (OECD): [s.n.], 2017.

DENG, Ziliang *et al.* Rapid internationalization and exit of exporters: The role of digital platforms. *International Business Review*, v. 31, n. 1, p. 101896, fev. 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0969593121001086>.

GAUTIER, Axel; LAMESCH, Joe. Mergers in the digital economy. *Information Economics and Policy*, v. 54, n. C, p. 1–15, mar. 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167624520301347>.

GAWER, Annabelle. Digital platforms and ecosystems: remarks on the dominant organizational forms of the digital age. *Innovation, Organization & Management*, p. 1–15, 17 set. 2021. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/14479338.2021.1965888>.

GHIO, Niccolò; GUERINI, Massimiliano; ROSSI-LAMASTRA, Cristina. University knowledge and

the creation of innovative start-ups: an analysis of the Italian case. *Small Business Economics*, v. 47, n. 2, p. 293–311, 12 ago. 2016. Disponível em: <http://link.springer.com/10.1007/s11187-016-9720-2>.

HOLL, Adelheid; MARIOTTI, Ilaria. The Geography of Logistics Firm Location: The Role of Accessibility. *Networks and Spatial Economics*, v. 18, n. 2, p. 337–361, 8 jun. 2018. Disponível em: <http://link.springer.com/10.1007/s11067-017-9347-0>.

HUANG, Daquan *et al.* Do land ownership types matter in manufacturing firms' location choice? Using Beijing as a case study. *Growth and Change*, v. 53, n. 1, p. 151–169, 5 mar. 2022. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/grow.12579>.

INDURKHYA, Nitin; DAMERAU, Fred J. (Org.). *Handbook of Natural Language Processing*. Second Edi ed. Boca Raton (Florida, USA): CRC Press, 2010.

JIA, Kai; KENNEY, Martin. The Chinese platform business group: an alternative to the Silicon Valley model? *Journal of Chinese Governance*, p. 1–23, 1 fev. 2021. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/23812346.2021.1877446>.

JIA, Kai; KENNEY, Martin; ZYSMAN, John. Global Competitors? Mapping the Internationalization Strategies of Chinese Digital Platform Firms. In: TULDER, ROB VAN; VERBEKE, ALAIN; PISCITELLO, LUCIA (Org.). . *International Business in the Information and Digital Age*. Bingley (UK): Emerald Publishing Limited, 2018. p. 187–215. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/S1745-886220180000013009/full/html>.

KEMENY, Tom; NATHAN, Max; ALMEER, Bader. *Using Crunchbase to explore innovative ecosystems in the US and UK*. , Discussion Paper Series., nº 2017–01. Birmingham: [s.n.], 2017.

KENNEY, Martin; BEARSON, Dafna; ZYSMAN, John. The platform economy matures: measuring pervasiveness and exploring power. *Socio-Economic Review*, v. 19, n. 4, p. 1451–1483, 3 nov. 2021. Disponível em: <https://academic.oup.com/ser/article/19/4/1451/6224398>.

KENNEY, Martin; ZYSMAN, John. The platform economy: restructuring the space of capitalist accumulation. *Cambridge Journal of Regions, Economy and Society*, v. 13, n. 1, p. 55–76, 15 maio 2020. Disponível em: <https://academic.oup.com/cjres/article/13/1/55/5809994>.

KENNEY, Martin; ZYSMAN, John. The Rise of the Platform Economy. *Issues in Science and Technology*, v. 32, n. 3, p. 61–69, 2016.

MCKNIGHT, Scott; KENNEY, Martin; BREZNITZ, Dan. Platformizing the Economy? Building and Regulating Chinese Digital Platforms. *SSRN Electronic Journal*, p. 1–46, 2021. Disponível em: <https://www.ssrn.com/abstract=3885190>.

PARKER, Geoffrey G.; ALSTYNE, Marshall W. Van; CHOUDARY, Sangeet Paul. *Platform Revolution: How Networked Markets Are Transforming the Economy?and How to Make Them Work for You*. New York: Norton & Company, 2016.

PUNT, Matthijs B. *et al.* Your Uber is arriving now: An analysis of platform location decisions through an institutional lens. *Strategic Organization*, p. 147612702110222, 21 jun. 2021. Disponível em: <http://journals.sagepub.com/doi/10.1177/14761270211022254>.

SACHS, Laís Bueno. *FinTech: China & Brazil.* . Rio de Janeiro: Institute for Technology and Society of Rio (ITS Rio). Disponível em: <https://its-fintech.pubpub.org/>. , 2018

SAFARI, Arsalan. Worldwide venture capital, intellectual property rights, and innovation. *Industrial and Corporate Change*, v. 26, n. 3, p. 485–515, 30 ago. 2017. Disponível em: <https://academic.oup.com/icc/article-lookup/doi/10.1093/icc/dtw035>.

SHAHEER, Noman; LI, Sali; PRIEM, Richard. Revisiting Location in a Digital Age: How Can Lead Markets Accelerate the Internationalization of Mobile Apps? *Journal of International Marketing*, v. 28, n. 4, p. 21–40, 17 dez. 2020. Disponível em: <http://journals.sagepub.com/doi/10.1177/1069031X20949457>.

SILVA NETO, Victo José Da; CHIARINI, Tulio; RIBEIRO, Leonardo da Costa. Viagens de descobrimento: mapeando a geografia da economia de plataformas. maio 2022, São Paulo: Editora Blucher, maio 2022. p. 374–394. Disponível em: <http://www.proceedings.blucher.com.br/article-details/37219>.

STALLKAMP, Maximilian; SCHOTTER, Andreas P. J. Platforms without borders? The international strategies of digital platform firms. *Global Strategy Journal*, v. 11, n. 1, p. 58–80, 28 fev. 2021. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/gsj.1336>.

SU, Chunmeizi; FLEW, Terry. The rise of Baidu, Alibaba and Tencent (BAT) and their role in China's Belt and Road Initiative (BRI). *Global Media and Communication*, v. 17, n. 1, p. 67–86, 18 abr. 2021. Disponível em: <http://journals.sagepub.com/doi/10.1177/1742766520982324>.

WENTING, Rik; FRENKEN, Koen. Firm entry and institutional lock-in: an organizational ecology analysis of the global fashion design industry. *Industrial and Corporate Change*, v. 20, n. 4, p. 1031–1048, 1 ago. 2011. Disponível em: <https://academic.oup.com/icc/article-lookup/doi/10.1093/icc/dtr032>.

ANNEX

TABLE 4
Example of Orbis information available on products/services and history of companies listed in the database

Company name	MercadoLibre Inc.	MercadoLibre Colombia Ltda	MercadoLibre SRL
ID number	US980212790	CO170001515680	AR30-70308853-4
Country code	US	CO	AR
City	-	Bogota	Buenos Aires
NACE (*)	7490	6209	-
Trade description	MercadoLibre, Inc. is an e-commerce company. The Company enables commerce through its *marketplace platform* in Latin America, designed to provide users with a portfolio of services to facilitate commercial transactions. Its geographic segments are Brazil, Argentina, Mexico, Venezuela, and Other Countries (…).	Technology and computer service activities	Operates an online trading platform in Latin America
Products and services	Classifieds service that enables users to list their offerings related to motor vehicles, vessels, aircraft, real estate, and services outside the *marketplace platform*; and MercadoPago, an integrated online payments solution to facilitate transactions on and off the MercadoLibre Marketplace by providing a mechanism that allows its users to send and receive payments online (…).	Offers a marketplace, an online trading service that permits businesses and individuals to list items and conduct sales and purchases online in a fixed-price or auction-based format; and MercadoPago online payments solution, an integrated online payments solution	Latin America
Description and history	MercadoLibre, Inc., incorporated on October 15, 1999, is an e-commerce company. The Company enables commerce through its *marketplace platform* (…) in Latin America. The Company's platform is designed to provide users with a portfolio of services to facilitate commercial transactions (…). The Company offers its users an ecosystem of six integrated e-commerce services: the MercadoLibre Marketplace, the MercadoLibre Classifieds Service, the MercadoPago payments solution, the MercadoLibre advertising program, the MercadoShops online Webstores solution, and the MercadoEnvios shipping service. (…) The Company competes with Rakuten, Amazon, B2W Inc., Cnova, Aliexpress, Netshoes, Dafiti, Casas Bahia, Walmart, (…), Facebook, Google, Amazon, Microsoft, Yahoo!, Paypal, DineroMail, Bcash, PagSeguro, Western Union, PayU, MOIP, Alamaula.com, OLX.com, and QueBarato.	-	MercadoLibre SRL

Source: Authors' own. Data sourced from Orbis, Bureau van Dijk. Note: (*) NACE stands for *Nomenclature statistique des activités économiques dans la Communauté européenne*

TABLE 5
Lists of terms retrieved in the query related to digital platform

1st List	2nd List	Final List
application programming interface		
cloud computing		
cloud-based solution		
	cloud service	
data-centric cloud		
digital banking		
digital content		
	digital marketplace	digital marketplace
digital payment	*digital payment*	*digital payment*
	digital platform	digital platform
ecommerce		
e-commerce		
electronic media		
electronic payment		
e-media		
e-payment	*e-payment*	*e-payment*
fintech service		
	innovation platform	innovation platform
	internet marketplace	internet marketplace
	internet platform	internet platform
intelligent cloud		
internet search solution	*internet search solution*	*internet search solution*
internet shopping	*internet shopping*	*internet shopping*
marketplace platform	*marketplace platform*	*marketplace platform*
mobile devices		
mobile game	*mobile game*	*mobile game*
mobile payment	*mobile payment*	*mobile payment*
mobile platform	*mobile platform*	*mobile platform*
network service		
mobile service		
online advertising service	*online advertising service*	*online advertising service*
online booking	*online booking*	*online booking*
online game	*online game*	*online game*
online gaming	*online gaming*	*online gaming*
	online marketplace	online marketplace
	online platform	online platform
online reservation	*online reservation*	*online reservation*
online retailer		
		online social media
	online trading platform	
payment platform	*payment platform*	*payment platform*
payment service		
search engine	*search engine*	*search engine*
serverless computing	*serverless computing*	*serverless computing*
web application		
	social game	
		social media
		social media content
		social media management
		social media marketing
		social media strategy
social network	*social network*	*social network*
		social networking services
	software platform	software platform
	transaction platform	transaction platform
web portal		

Source: Authors' own.

TABLE 6
Tests and Fit Statistics

Test	Statistics and Criteria		Conclusion		
PRM	BIC= 17220.552	AIC= 17152.268	**Prefer**	**Over**	**Evidence**
vs NBRM	BIC= 8692.473	dif= 8528.079	NBRM	PRM	Very strong
	AIC= 8614.434	dif= 8537.834	NBRM	PRM	
	LRX2= 8539.834	prob= 0.000	NBRM	PRM	p=0.000
vs ZINB	BIC= 7448.322	dif= 9772.230	ZINB	PRM	Very strong
	AIC= 7301.999	dif= 9850.269	ZINB	PRM	
NBRM	BIC= 8692.473	AIC= 8614.434	**Prefer**	**Over**	**Evidence**
vs ZIP	BIC= 10629.807	dif= -1937.334	NBRM	ZIP	Very strong
	AIC= 10493.239	dif= -1878.805	NBRM	ZIP	
vs ZINB	BIC= 7448.322	dif= 1244.151	ZINB	NBRM	Very strong
	AIC= 7301.999	dif= 1312.435	ZINB	NBRM	
	Vuong= 10.582	prob= 0.000	ZINB	NBRM	p=0.000
ZIP	BIC= 10629.807	AIC= 10493.239	**Prefer**	**Over**	**Evidence**
vs ZINB	BIC= 7448.322	dif= 3181.485	ZINB	ZIP	Very strong
	AIC= 7301.999	dif= 3191.240	ZINB	ZIP	
	LRX2= 3193.240	prob= 0.000	ZINB	ZIP	p=0.000

Source: Authors' own. Based on *countfit* Stata package (LONG; FEESE, 2006).