

City limits in the age of smartphones and urban scaling

Boris Sotomayor-Gómez^{a,b}, Horacio Samaniego^{a,c,*}



^a Laboratorio de Ecoinformática, Instituto de Conservación, Biodiversidad y Territorio, Facultad de Ciencias Forestales y Recursos Naturales, Universidad Austral de Chile, Campus Isla Teja, Valdivia, Chile

^b Instituto de Informática, Facultad de Ciencias de la Ingeniería, Universidad Austral de Chile, Campus Miraflores, Valdivia, Chile

^c Instituto de Ecología y Biodiversidad (IEB), Casilla 653, Santiago, Chile

ARTICLE INFO

Keywords:

City boundaries definition
Spectral network analysis
Urban informatics
Social computing
Scaling laws
Complex systems
Big data

ABSTRACT

Urban planning still lacks appropriate standards to define city boundaries across urban systems. This issue has historically been left to administrative criteria, which can vary significantly across countries and political systems, hindering a comparative analysis across urban systems. However, the wide use of Information and Communication Technologies (ICT) has now allowed the development of new quantitative approaches to unveil how social dynamics relates to urban infrastructure. In fact, ICT provide the potential to portray more accurate descriptions of the urban systems based on the empirical analysis of millions of traces left by urbanites across the city. In this work, we apply computational techniques over a large volume of mobile phone records to define urban boundaries, through the analysis of travel patterns and the trajectory of urban dwellers in conurbations with more than 100,000 inhabitants in Chile. We created and analyzed the network of interconnected places inferred from individual travel trajectories. We then ranked each place using a spectral centrality method. This allowed to identify places of higher concurrency and functional importance for each urban environment. Urban scaling analysis is finally used as a diagnostic tool that allowed to distinguish urban from non-urban spaces. The geographic assessment of our method shows a high congruence with the current and administrative definitions of urban agglomerations in Chile. Our results can potentially be considered as a functional definition of the urban boundary. They also provide a practical implementation of urban scaling and data-driven approaches on cities as complex systems using increasingly larger non-conventional datasets.

1. Introduction

Urban expansion ranks as one of the most pressing issues of modern society and the consequences of continuous growth could be devastating for human welfare (United Nations, 2019). While considerable efforts are seeking to contain urban expansion, we still lack a standardized method to measure the extent of cities boundaries. Even if we accept that the definition of urban limits is highly constrained to our understanding of urban systems, current research sets urbanites, and their interactions, as centerpieces of the urban organization and productivity. Important efforts have already contributed to the understanding of urban morphology and city dynamics (Barthelemy, 2017). Developing a standard and functional criteria to define the spatial boundaries of urban activity will contribute to increase our understanding of several aspects of urban life (Cranshaw, Schwartz, Hong, & Sadeh, 2012; Echenique, 2001; Kenworthy & Laube, 1999; Paddison, 2000; Rogers & Burdett, 2001).

The lack of unified criteria to define urban boundaries is highlighted by the United Nations World Cities report, which indicates that countries' urban proportions face multiple non-comparable definitions (López, 2017; United Nations, 2016). As an example, the city of Toronto in Canada, may be considered to have between 2.6 and 5.1 million inhabitants, depending on which of the three definitions of city we use to calculate its size. Such variable figures portray annual growth rates ranging from 0.9% to 1.8%, which may clearly lead to significant complications for efficient planning efforts.

On the other hand, important strides have been made to understand and study cities using alternative perspectives. Some of these studies proposed scaling as observed in biology and complex systems (Batty, 2007; Bon, 1979; Samaniego & Moses, 2008; Veregin & Tobler, 1997; West, Brown, & Enquist, 1997) as a useful tool to understand and quantitatively predict different urban functions and characteristics (Alves, Ribeiro, Lenzi, & Mendes, 2014; Bettencourt & Lobo, 2016; Bettencourt, Lobo, Strumsky, & West, 2010; Gomez-Lievano, Youn, &

* Corresponding author at: Laboratorio de Ecoinformática, Instituto de Conservación, Biodiversidad y Territorio, Facultad de Ciencias Forestales y Recursos Naturales, Universidad Austral de Chile, Campus Isla Teja, Valdivia, Chile.

E-mail address: horacio@ecoinformatica.cl (H. Samaniego).

Bettencourt, 2012; Meirelles, Neto, Ferreira, Ribeiro, & Binder, 2018; van Raan, van der Meulen, & Goedhart, 2016). Although several studies have recently questioned the methodology (Depersin & Barthelemy, 2018; Leitão, Miotto, Gerlach, & Altmann, 2016), there is little doubt about its importance (Barthelemy, 2017; Batty, 2013; Keuschnigg, Mutgan, & Hedström, 2019). A candidate solution to solve this controversy lies in the unification of spatial definitions of the urban environment, as geographical delimitation has been associated with the great variability of scaling coefficients for various urban functions (Arcuate et al., 2015; Cottineau, Hatna, Arcuate, & Batty, 2017; Cottineau & Vanhoof, 2019; Fragkias, Lobo, Strumsky, & Seto, 2013; Louf, Roth, & Barthelemy, 2014).

The search for a spatial definition of the city has not been limited to the aforementioned. The identification of an optimal methodology defining urban boundaries remains an old problem that is usually addressed by studying the spatial density distribution with the aid of comprehensive surveys and censuses. For instance, Rozenfeld et al. (2008), Rozenfeld, Rybski, Gabaix, and Makse (2011) centered the identification of urban systems on the notion of spatial continuity using the City Clustering Algorithm (CCA). This was later updated by Oliveira et al. to the City Local Clustering Algorithm (CLCA) improving clustering and computational capabilities with a worldwide application (Oliveira, Furtado, Andrade, & Makse, 2018). Along similar lines, Cranshaw et al. (2012) proposed a spectral clustering definition of a functional neighborhood in Pittsburgh. A pioneering work by Ratti et al. (2010) proposed a novel approach integrating community detection algorithms built from the complex networks of over 12 billion land-line calls to map functional urban areas in Britain. Thiemann, Theis, Grady, Brune, and Brockmann (2010) applied two urban mobility techniques to investigate the geographical edges that emerge from an analysis of complex networks. Tannier, Thomas, Vuidel, and Frankhauser (2011), Tannier and Thomas (2013) integrated morphological approaches widely used in landscape ecology based on the analysis of the heterogeneity of the urban-rural interface for boundary identification. On the other hand, Masucci, Arcuate, Hatna, Stanilov, and Batty (2015) and Arcuate et al. (2016) used percolation analysis to find urban boundaries. They applied a clustering technique on the planar network of street-intersections and satellite images showing the strong dependence of scaling relationships on urban boundary definitions. Recently, Cottineau and Vanhoof (2019) evaluated the prediction power of mobile phone datasets to understand demographic data on different urban extents based on population density. Finally, Humeres and Samaniego (2017) proposed a spectral definition of cities based on the analysis of mobility networks using classical origin-destination (OD) surveys in Chile. This study builds on the methods developed by the last-mentioned work, broadening and strengthening its scope.

The growing availability of Information and Communication Technologies (ICTs) has opened new possibilities to contribute towards a more effective estimation of the urban extension limits. The high latency of recorded transactions in Call Detail Records (X/CDR) allows capturing the dynamic process of space usage among urbanites at high spatio-temporal and contextual resolutions (Blondel, Decuyper, & Krings, 2015). Such unprecedented feature further permits to explore new solutions to the social problems of the urban organization (Blondel et al., 2015; Dannemann, Sotomayor-Gómez, & Samaniego, 2018; Gonzalez, Hidalgo, & Barabasi, 2008; Louail et al., 2014). This fundamental characteristic has provided an important ground to evaluate what Batty has termed *The New Science of Cities* (Batty, 2013), a singular approach to understand the urban organization that places the network of interactions between people and urban infrastructure as the central aspect shaping the city. Studies involving ICTs have thus emerged as an essential tool to delineate the city, often geared towards a near real-time functional representation that goes beyond the usual limitations of comprehensive censuses, campaigns and static surveys to represent the urban environment.

This work proposes a data-driven approach allowing to distinguish

urban ("U") from non-urban spaces ("NU") by analyzing the network of spaces built from the trajectories of urbanites. This methodology was applied to all cities with more than 100,000 inhabitants in Chile. We used a centrality index as a proxy of the degree of concurrency throughout the city. To untangle NU spaces from U, first, each exploration area was partitioned, and second, the relationship between urbanites and the number of calls made in each U and NU partition was analyzed. This, ultimately allowed us to use the significant differences in scaling between social voice-calls and the number of urbanites (i.e. β_U and β_{NU}) to distinguish and define urban spaces.

2. Materials and methods

2.1. Dataset

The data used in this work corresponds to mobile telephone records generated by pre- and post-payment users to one of the largest provider in Chile, Telefónica (~37% of market share). We used two datasets: (1) internet connection detail records (XDR) generated when using mobile internet and (2) voice call detail records (CDR). A total of 4 weeks of data were analyzed in March, May, October and November of 2015 with a total of $\sim 5.6 \times 10^6$ users throughout Chile and $\sim 3 \times 10^9$ XDR records. Each record contains the date and time of connection, an anonymous identifier for each user and the geolocation of the antenna to which the connection is established. About 1×10^9 CDR records were used to assign the number of calls registered to each antenna during the study period.

The coverage area of antennas was approximated using Voronoi cells. This allowed to use XDRs as a proxy of individual mobility, since they represent the set of physical locations effectively visited by each user in his or her daily trajectory. CDRs, on the other hand, are proxies of the level of social voice interaction between urbanites; these were used as indicators of effective social cohesion and to define urban boundaries through urban scaling analysis.

2.2. Defining local interactions

We took advantage of the high temporal density of XDR data and ascribed the residence of each urbanite to a particular location to only consider local mobility and communication in the analyses per city. To ensure a correct assignation of the home location for every user, we followed previous work done by our research group and others (Šćepanović, Mishkovski, Hui, Nurminen, & Ylä-Jääski, 2015; Dannemann et al., 2018; Phithakkitnukoon, Smoreda, & Olivier, 2012; Vanhoof, Reis, Smoreda, & Ploetz, 2018). In brief, correct assignation of home location was possible by identifying the most frequented tower between 10 PM and 7 AM on weekdays with at least five pings. We additionally required at least a quarter of total pings for each user to be made within his or her residence during the four weeks dataset, thereby minimizing uncertainties associated with seasonal effects (e.g. changing movement behaviour of people going on holidays) among others.

2.3. Urbanite's trajectories and exploration areas

Trajectories for each resident were constructed as a sorted list of visited locations across the day starting from his or her home location (Fig. 1A). To estimate buffer sizes, the total distance for each trajectory was calculated as the sum of the geographic distance between visited locations and these were later used to frame the individual analyses within the exploration areas for each city (see below).

In addition, to restrict the analysis to urbanites residing in each city, we only considered cities larger than 100,000 inhabitants in 2015 (MINVU, 2017). This allowed to avoid the potential bias of having an unequal market share among telephone providers in smaller cities (i.e. population $< 10^5$). Official limits for the 21 selected cities were used as a starting area of exploration to evaluate the spatial extent of each city.

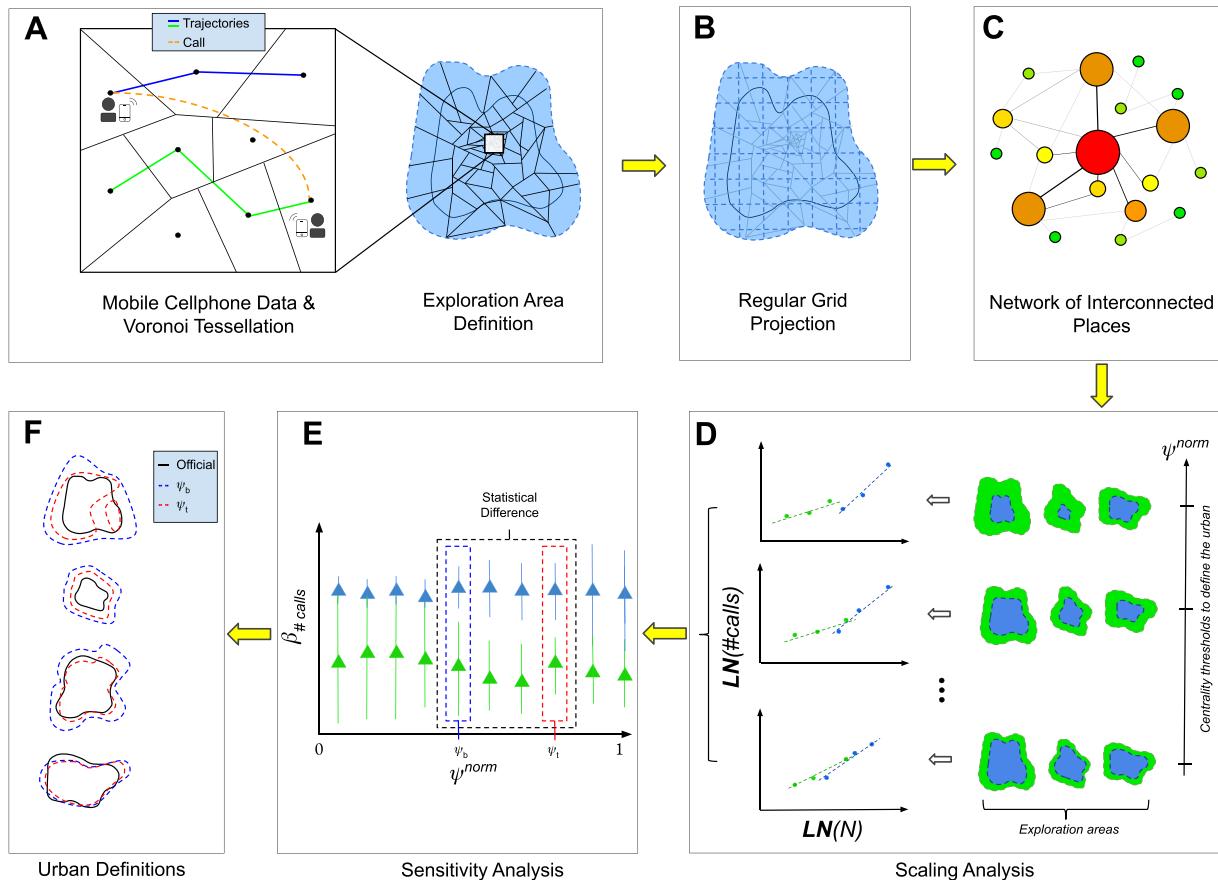


Fig. 1. Schematic representation of the analysis procedure. (A) X/CDR datasets were used to compute user's trajectories as a sorted list of visited places across exploration areas for each city. (B) A 1 km² grid was used as spatial analysis unit. (C) A network of interconnected places was created for each exploration area and centrality ranking (ψ) was computed for each node. (D) Thresholds were defined based on ψ and the slope between the number of calls and the network size within and outside boundaries was evaluated. (E) Differences between slopes were evaluated in order to select the top and bottom values of the interval where urban and non-urban places are statistically different given the scaling relationship. (F) Comparison of resulting boundaries and official definitions of cities.

A buffer was built around each city extending the official boundary (i.e. defined by the Ministry of Housing and Urbanism, MINVU) to the median of the total traveled distance of urbanites within the city (Fig. 1B).

2.4. Spatial partition of exploration area

Voronoi cells were partitioned into a regular grid to minimize biases associated with the irregular distribution of antennas across cities. While several grid sizes could be considered, a 1 × 1 km grid was just under the median area of Voronoi cells across the officially defined urban system analyzed (Fig. S1), and was hence chosen as a compromise that would provide a meaningful geographic analysis scale commensurate to official, vector-based outlines across cities (Fig. 1B). This offered a granularity that further allowed a homogeneous visualization of the spatial distribution of antennas, as proposed in Louail et al. (2014).

2.5. Building networks of interconnected locations

Each urban system was represented as a network (Fig. 1C). This sought to depict the interaction between individuals and the urban infrastructure. For this, a bipartite graph was used to represent the set of locations visited by each user on a daily basis within the exploration area (Fig. 2). Subsequently, the bipartite graph was transformed into a weighted unimodal graph of interconnected locations (Fig. 2B and D). Such a graph allowed us to analyze the network of locations (i.e. antennas as nodes) visited by each person (i.e. links). Interaction rates

between locations, $\omega_{i,j}$, were computed as the sum of the smallest number of pings (i.e. XDR records), p , between locations (i.e. i and j) visited by user u . Note that this is, in fact, a similarity measure equivalent to computing the intersection of pings among the total number of pings per user, p_{total} ; see Eq. (1) (Humeres, 2014; Humeres & Samaniego, 2017).

$$\omega_{i,j} = \sum_u \frac{\min(p_i^u, p_j^u)}{p_{total}} \quad (1)$$

We normalized the interaction rate between locations, $\omega_{i,j}$, over the sum of pings, p_{total} , for each user in order to compare weights across locations.

2.6. Spectral analysis of locations networks

Eigenvector centrality was used to estimate the importance of each cell in the city. It characterized nodes (i.e. cells) in the network as a function of the number of visits connecting these locations. Interestingly, centrality values are independent of the order in which they are visited (Borgatti, 2005). The Bonacich method was used to compute centrality (Bonacich, 1972). Each cell was then ranked, so that better-connected cells (i.e. linked to many places) would have higher values of centrality and thus a higher ranking.

Eigenvector centrality ranking is defined as $\psi_i = u_{1,i}$, where $u_{1,i}$ is a component of u_1 , the eigenvector associated with the largest eigenvalue λ_1 satisfying $Au_1 = \lambda_1 u_1$ (Newman, 2006). Centrality calculation was performed using a power iteration method, as implemented in graph-

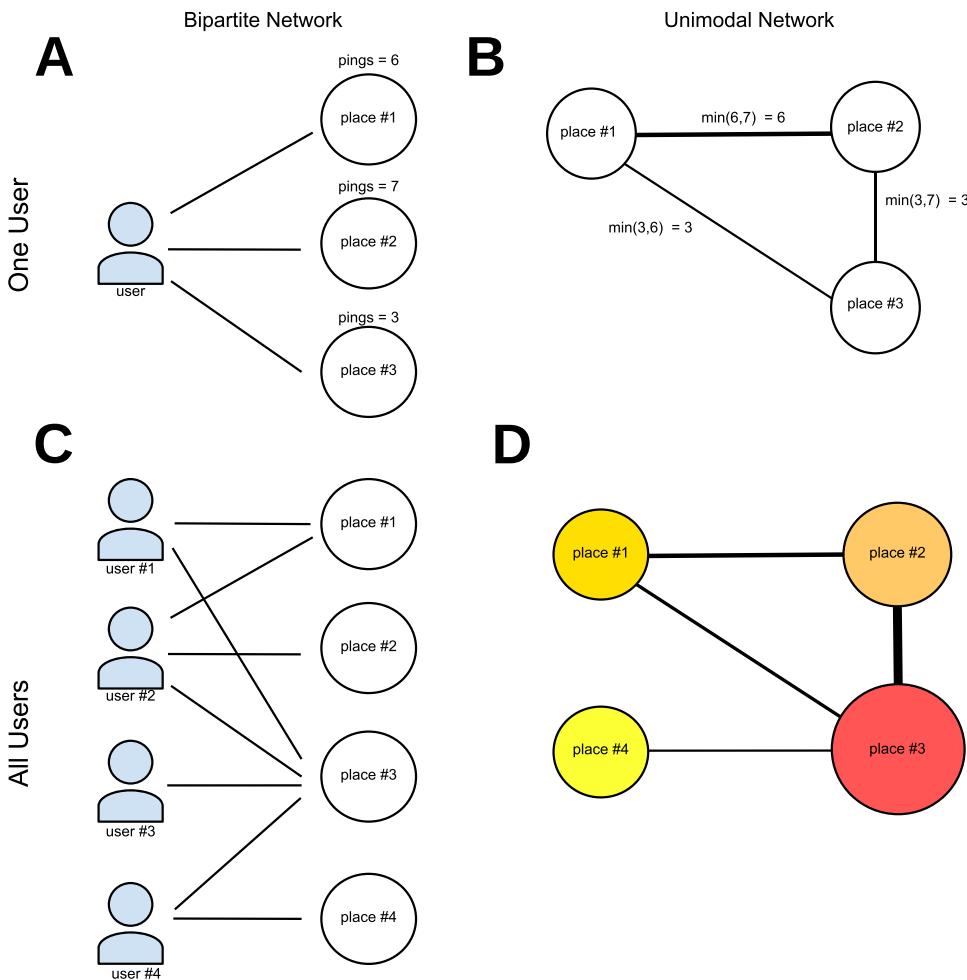


Fig. 2. Unimodal projections of bipartite graphs. (A) Simple bipartite network for a single user visiting three locations; (B) Unimodal projection of locations visited by user in (A); (C) Bipartite graph representing places visited by four users; (D) Unimodal representation of (C). Centrality ranking is shown in red and the weight of the relationship between places is shown by the thickness of links between places. Similarity between two connected locations corresponds to the smallest amount of pings between them. Each link is sequentially added to the user's trajectory increasing the weight ω_{ij} of the link (Eq. (1)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tool v2.27 (Langville & Meyer, 2005; Peixoto, 2014).

A discrete, 0.01, range-interval of centrality values was defined to compare computed centralities across cities (ψ^{norm}). Therefore, the kurtosis of ψ suggested the following logarithmic transformation:

$$\psi_{i,x}^{norm} = \frac{\log \psi_{i,x} - \min(\log \psi_x)}{\max(\log \psi_x) - \min(\log \psi_x)}; \quad \psi_{i,x}^{norm} \in [0, 1]. \quad (2)$$

2.7. Urban scaling analysis

Urban scaling analysis was used to evaluate any potential regime change across urbanites' social voice-interactions for the various spatial extents defined by a threshold in ψ^{norm} (Fig. 1D). The elasticity, of the scaling relationship between population $N(t)$ and the number of voice-calls $Y(t)$ (see below), was then compared for significant differences (Fig. 1E). $Y(t)$ was measured from an independent CDR dataset for the same time period. The urban scaling model considered the size of the city $N(t)$, measured by the number of urbanites making voice calls at time t in a defined area (Bettencourt, Lobo, Helbing, Kühnert, & West, 2007):

$$Y(t) = Y_0 N(t)^\beta \quad (3)$$

$Y(t)$ has often been associated to measurable urban characteristics (e.g. income, social interaction, extension of electrical wiring, etc.), Y_0 is a normalization constant and the elasticity, β , characterizes the general dynamics and growth regime across a gradient of city sizes (Bettencourt et al., 2007; Bettencourt & West, 2010). Eq. (3) is parameterized using ordinary least squares method after a \log_{10} transform: $\log_{10}(Y(t)) = \log_{10}(Y_0) + \beta \log_{10}(N(t))$.

With the premise that urban boundaries are shaped by the interaction between people and the urban infrastructure, Eq. (3) was used as a diagnostic tool to identify the critical interval $[\psi_b, \psi_e]$ that significantly – and functionally – bounds the urban space. For this, we sequentially partitioned the exploration area (i.e. buffer + conurbation MINVU) into two sets of locations. The first set of locations, had a centrality rank larger than the defined threshold value (e.g. ψ_b) and, the second, included locations not meeting such condition (Fig. 1). The hypothesis of β 's belonging to the same distribution was evaluated using standard t -test (Andrade & Estévez-Pérez, 2014). Places ranking equal or greater than the threshold were evaluated using Eq. (3) and labeled as urban (U), while those ranking lower were assigned to non-urban spaces (NU). This method allowed to spatially partition the exploration area in two different systems and compare β_U and β_{NU} elasticity. Finally, we selected the largest continuous ranking interval having statistically different β (Fig. 1E), generated the corresponding map of U, NU, and qualitatively compared our results to the official delineation of 21 cities in Chile.

3. Results

Traveled distance among cities was found to be fairly constant and showed a right-skewed distribution (Fig. 3A) with an overall median of ~ 9.3 km across exploration areas. Fig. 3B shows an example of the exploration area for the city of Valdivia (Q).

3.1. Eigen-vector ranking analysis

The centrality ranking explicitly divided the exploration areas based

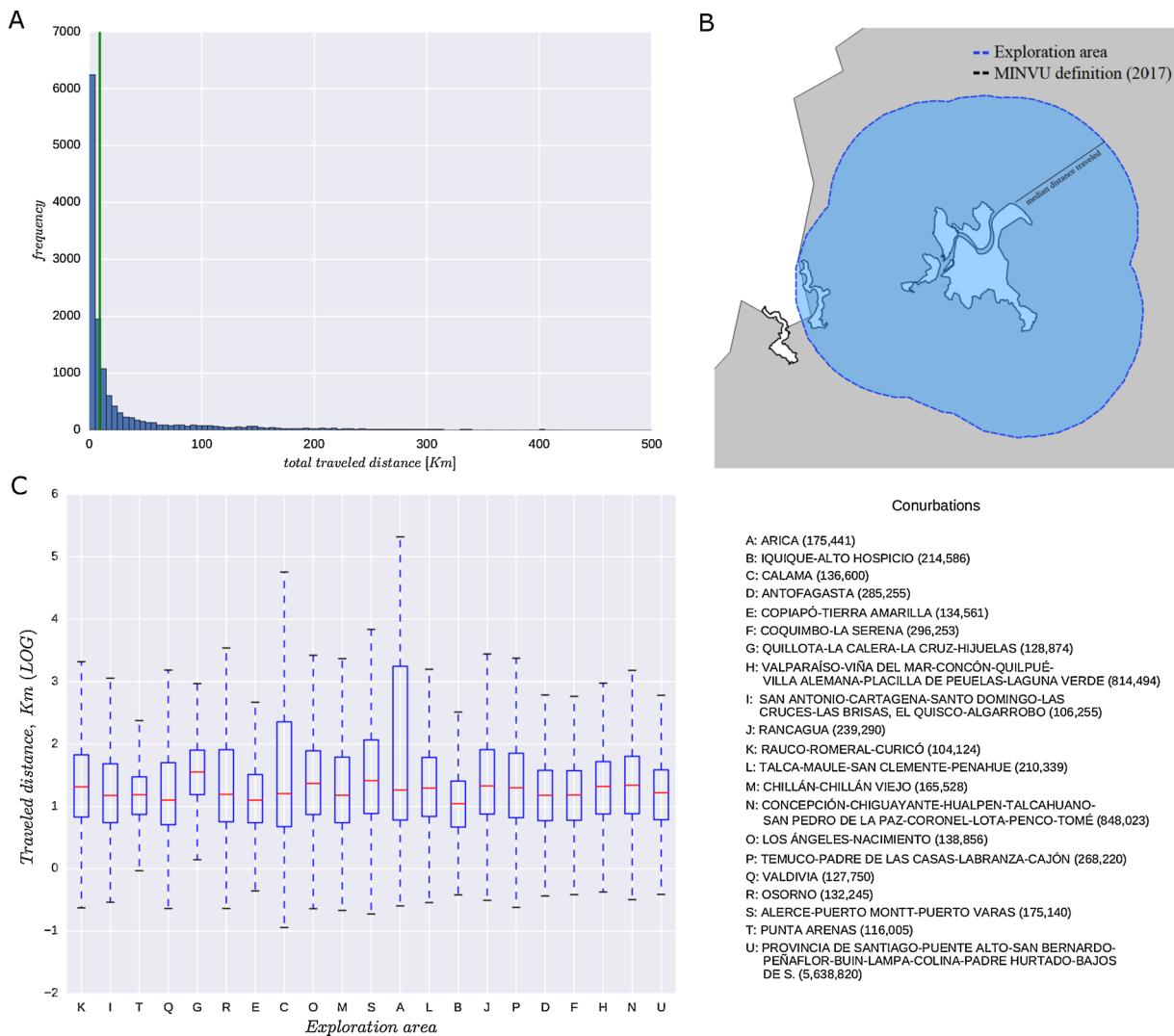


Fig. 3. (A) Distribution of total distance traveled by mobile telephony users in Valdivia. The vertical green line shows the median traveled distance (9.31 km) over all cities. (B) An example of the buffer creation process for the city of Valdivia. The official urban area for 2017 is shown in lighter blue within dark blue buffer area, as defined by the Chilean Ministry of Housing and Urbanism (MINVU) (MINVU, 2017). (C) Box-plot of traveled distances within the buffer exploration area for each individual city considered in this study. Cities are ordered from the smallest on the left to the largest on the right. Population sizes are in parenthesis along with labels for (C). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

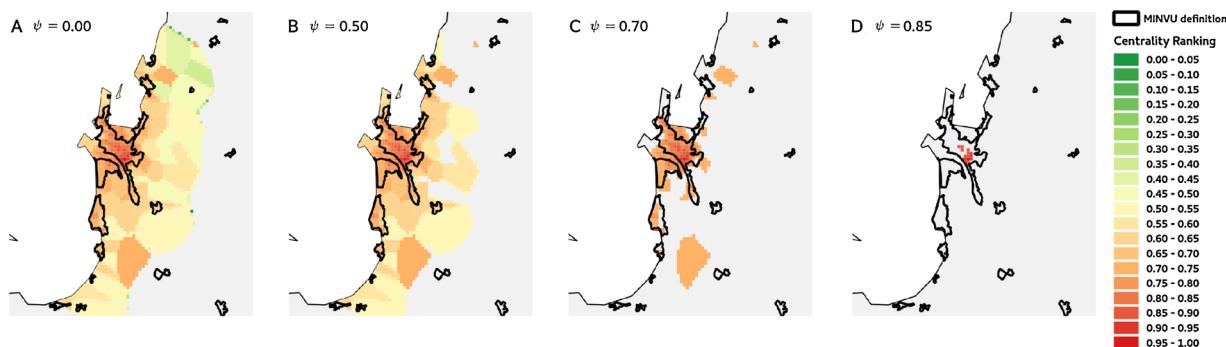


Fig. 4. Spatial distribution of centrality ranking using four thresholds (ψ^*) across the exploration area for the city of Concepción for Sunday, March 15, 2015. (A) Centrality within the exploration area ($\psi^* \geq 0.00$), (B) $\psi^{norm} \geq 0.50$, (C) $\psi^{norm} \geq 0.70$, (D) Locations with $\psi^{norm} \geq 0.85$. The base map was projected to S19 UTM coordinates and WGS84 datum.

on the overlap of human activities for each city. Not surprisingly, our analysis showed that central locations are within the official city limits with a decay of its centrality rank towards the periphery. For instance, in the case of Concepción, the third-largest city in Chile, central

locations roughly coincided with official urban boundaries in spite of the intricate configuration of city limits. Moreover, the approach proposed here highlighted additional locations with medium centrality ranking, which were not considered within official city boundaries

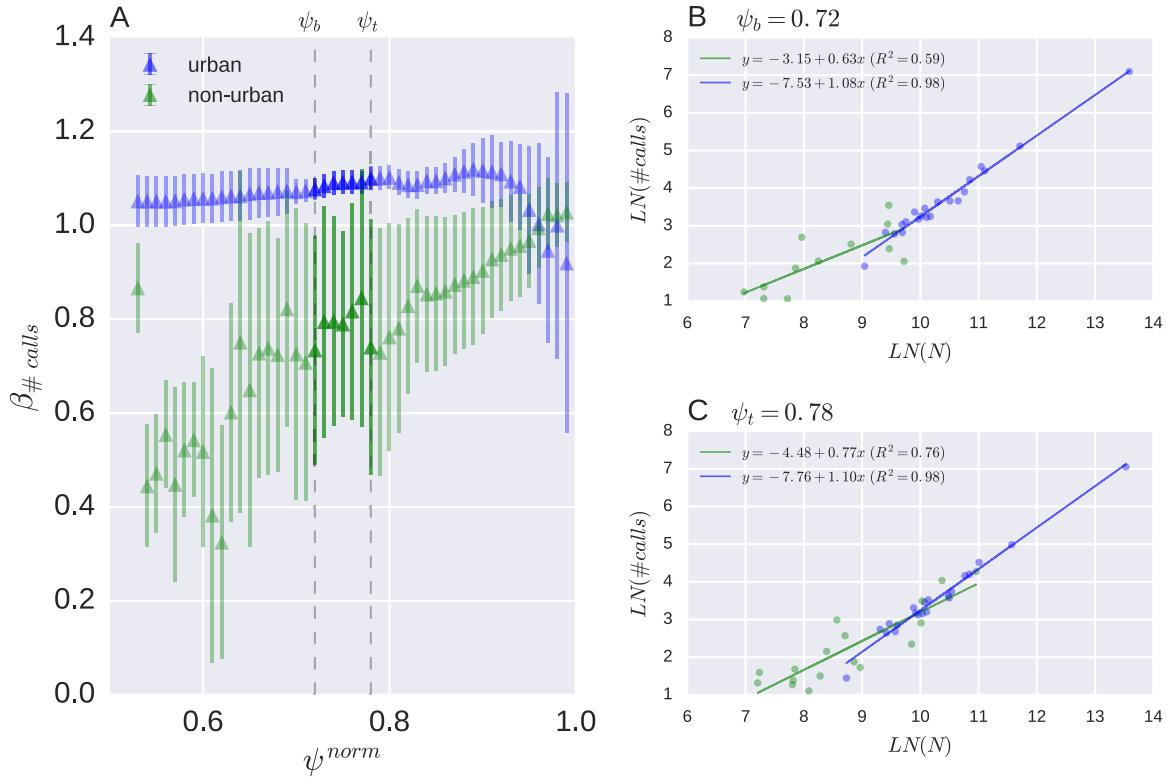


Fig. 5. Sensitivity analysis of exponents β_U and β_{NU} for different thresholds of ranking norm. Securities of exponents β_U and β_{NU} were obtained at 95% confidence. If the slopes were not statistically different or if the errors on the y-axis were not normally distributed for both systems, their comparison is presented with transparency. The interval obtained where both systems are distinguished corresponds to [0.72, 0.78] for the day Sunday, March 15th, 2015.

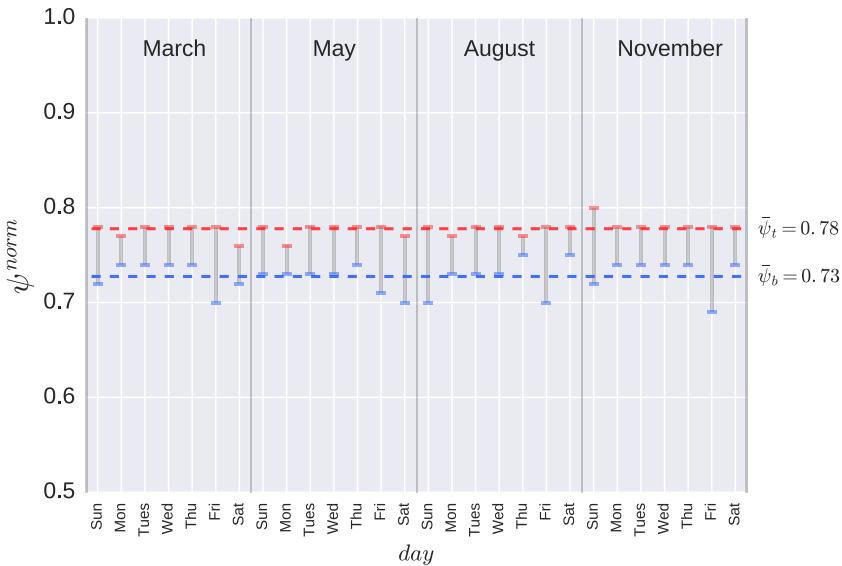


Fig. 6. Change in the definition intervals of the cities for the days analyzed. $\bar{\psi}_b$ (blue) corresponds to the average of the lower limit and $\bar{\psi}_t$ (red) to the upper limit of the intervals where the city is defined for the different days analyzed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 4).

The larger elasticity recorded for U compared to areas labeled as NU, suggests that urban scaling analysis is an appropriate diagnostic tool to differentiate visitation regimes across the boundary delineation. In fact, β_U was clearly larger than NU over the full range of norm, with standard errors for β 's larger at the margins of ψ^{norm} . This was most likely due to the lower number of cells labeled as U at lower ψ^{norm} and, conversely, a smaller number of cells labeled as NU where large centrality usually prevailed (Fig. 5). In spite of this, it was possible to

identify a centrality interval between [0.72, 0.77] that denoted statistical differences in voice-call interactions between U and NU places after evaluating the hypothesis of β being drawn from the same population (detailed sensitivity analysis available in Supplementary Material).

A complete analysis for each day in our records showed a stable distribution of centrality intervals (Fig. 6). The values of ψ_b and ψ_t were bounded in the range [0.7, 0.8] with lower and upper mean values of $\bar{\psi}_b = 0.72$ and $\bar{\psi}_t = 0.78$, respectively.

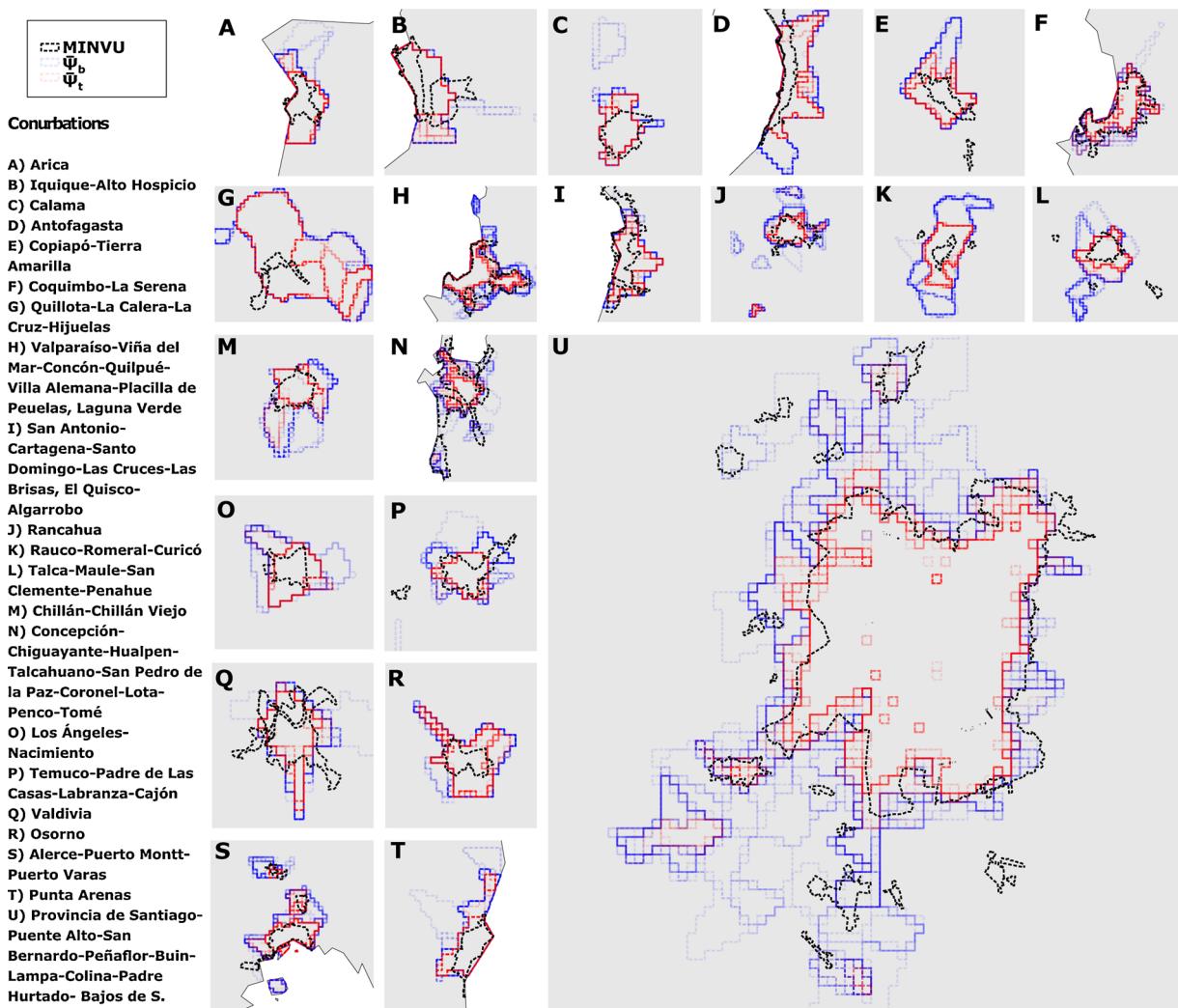


Fig. 7. City limit comparison across the full dataset. The upper right box shows the different threshold definitions. Official limits are in black (MINVU, 2017), ψ_b threshold in blue, and ψ_t in red contour. To visually show the most recurrent limit throughout the dataset, limits for each day are plotted using transparency so that more frequent limits become apparent due to the spatial superposition of city limits. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Eigen-cities: Cities boundaries from urban mobility and eigen-vector centrality of places

Spatial comparison of urban extents based on the upper centrality threshold (ψ_t) showed extensive morphological similarities with official city limits (MINVU, 2017) (Fig. 7). The complex nature of the spatial comparison made the quantitative comparison a complex task. Hence, we provided a qualitative analysis in which we overlaid city limits computed for each day using transparent layers (Fig. 7). This procedure made explicit the more recurrent limits across the dataset, thereby providing a better outlook of the differences between the official city limits and those proposed by our method. An additional, and more detailed, analysis procedure is provided in the Supplementary Materials (Fig. S3) along with quantitative areal differences for the cities of Concepción, Valparaíso, and Santiago (Table S1).

Some differences could be observed at the peripheries (i.e. (H) Viña del Mar-Valparaíso, (F) La Serena-Coquimbo, and (U) Gran Santiago). Additionally, cities with low density of antennas seemed to show a tendency to depict larger urban areas compared to the official delineation. This is most likely due to biases associated with the process of moving from Voronoi to the 1 km^2 grid cells. Since antennas are originally placed to ensure connectivity and not to uniformly cover the territory, a sparser antenna distribution occurs towards more rural

areas. Therefore, our method can potentially identify functional settlements detached from the central core of the study area, such as the cases of (J) Rancagua, (N) Gran Concepción, (S) Alerce-Puerto Montt-Puerto Varas, and (T) Punta Arenas (Fig. 7).

4. Discussion

This paper presented a methodology to delimit the urban extent with a functional perspective and using large volumes of non-conventional data. We used spectral analysis to unveil the centrality of places as nodes in the network of commonly visited places. Urban boundaries were evaluated through the rate of voice-call interactions within an exploration area. We applied our method to a concrete example delineating major urban systems in Chile, explicitly showing the advantages and opportunities offered by ICTs.

As an alternative to proposals based on the analysis of population density, distance thresholds, and extensive surveys (Humeres & Samaniego, 2017; Oliveira et al., 2018; Rozenfeld et al., 2008), we combined the dynamics of place visitations with voice interaction to define urban boundaries. Some of the previous studies have focused on studying the fractal geometry of urban textures to define the extent of cities (Batty & Longley, 1994; Batty & Xie, 1996; Frankhauser, 1998; Tannier et al., 2011; Tannier & Thomas, 2013), while others have

contributed with the analysis of topological patterns of street networks (Jiang, 2007; Masucci et al., 2015) and the delineation of socially cohesive areas through complex networks analyses and various types of thresholds (Grady, Brune, Thiemann, Theis, & Brockmann, 2012; Ratti et al., 2010).

Important contributions have already shown the importance of using ICT for understanding the structure and dynamics of cities (Cranshaw et al., 2012). Cottineau and Vanhoof (2019) have recently addressed the limitations of big-data in the generation of traditional urban indicators. Using arbitrary definitions of cities they evaluated how measures drawn from CDR may statistically relate to socio-economic indicators. They showed a compelling example that highlighted the importance of having an adequate, and functional definition of the urban boundaries, as it was proposed in their earlier work (Cottineau, Finance, Hatna, Arcuate, & Batty, 2018). On the other hand, we improved methods used by previous contributions by proposing the use of an urban boundary definition based on empirical observations of social interactions measured through voice-calls between individuals. We explicitly showed how large volumes of urbanites' trajectories may be processed to obtain near real-time urban delineation. In fact, while we used CDRs, it is worth noting that other spatially explicit indicators may also be used as a diagnostic tool to measure differences between urban and non-urban systems as done by (Cranshaw et al., 2012). Nowadays, given the increasing availability of big datasets makes our approach even more relevant. Our method allows to explore temporal constraints or the variation of urban boundaries across the day, as proposed by the Time Geography research program (Hägerstrand, 1970; Justen, Martínez, & Cortés, 2013). Using empirically-based descriptions of the heterogeneous usage of infrastructure, this method can aid in finding new planning perspectives to accommodate human movement (or existing/new infrastructure) throughout the city. Additionally, it may even prove instrumental for implementing participatory methodologies of urban informatics (Foth, 2018).

The simple heuristic method proposed here can clearly be improved. In fact, building a network of places by considering the minimum number of pings between locations in order to weigh their relationship (i.e. $\omega_{i,j}$) is only a starting point to capture the overlap of social interactions. Several other approximations may be possible to account for the way in which urbanites connect places. For instance, we might be interested in considering the average number of visits between two locations as a more representative indicator of the relationship between places.

Ranking places by their centrality value allows identifying key locations deserving special attention in terms of urban planning, transport, social studies and marketing, among others. An early proposal by Zhong et al. (2017) designed to describe and understand polycentric cities, defined central spaces according to the number and diversity of activities. Our approach may provide a technical framework to extend this methodology and identify not only the spatial extent of cities but also depicting locations with meaningful urban activity for planning purposes. Our method may additionally offer the opportunity to use near real-time data to evaluate the state of cities and exposing a series of questions relevant to urban planning. For instance, could this dynamic delineation of cities help to solve specific transportation issues? Can we plan ahead of time when organizing the urban infrastructures during particular events? We are confident that our method could provide such planning tools. However, further research is still needed as, for example, we need to develop new computational tools capable of modeling trajectories across the urban environment.

We believe that this paper offers a data-driven methodology incorporating the notion of cities as dynamic entities defined by individuals and their daily interactions, as proposed by the New Science of Cities (Batty, 2013).

Funding

Funding for this research was provided by CONICYT under the grant FONDECYT # 1161280.

Acknowledgments

We thank Pablo García and the team at Telefónica I+D Chile for the provision of anonymized X/CDR data for Chile.

References

- Alves, L. G., Ribeiro, H. V., Lenzi, E. K., & Mendes, R. S. (2014). Empirical analysis on the connection between power-law distributions and allometries for urban indicators. *Physica A: Statistical Mechanics and Its Applications*, 409, 175–182.
- Andrade, J., & Estévez-Pérez, M. (2014). Statistical comparison of the slopes of two regression lines: A tutorial. *Analytica Chimica Acta*, 838, 1–12.
- Arcuate, E., Hatna, E., Ferguson, P., Youn, H., Johansson, A., & Batty, M. (2015). Constructing cities, deconstructing scaling laws. *Journal of the Royal Society Interface*, 12(102), 20140745.
- Arcuate, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, A. P., et al. (2016). Cities and regions in Britain through hierarchical percolation. *Royal Society Open Science*, 3(4), 150691.
- Barthelemy, M. (2017). *The structure and dynamics of cities: Urban data analysis and theoretical modeling* (1st Edition). Cambridge University Press.
- Batty, M. (2007). *Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals*. The MIT Press.
- Batty, M. (2013). *The new science of cities*. MIT Press.
- Batty, M., & Longley, P. A. (1994). *Fractal cities: A geometry of form and function*. Academic Press.
- Batty, M., & Xie, Y. (1996). Preliminary evidence for a theory of the fractal city. *Environment and Planning A*, 28(10), 1745–1762.
- Bettencourt, L. M., & Lobo, J. (2016). Urban scaling in Europe. *Journal of The Royal Society Interface*, 13(116), 20160005.
- Bettencourt, L., & West, G. (2010). A unified theory of urban living. *Nature*, 467(7318), 912.
- Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17), 7301–7306.
- Bettencourt, L. M., Lobo, J., Strumsky, D., & West, G. B. (2010). Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE*, 5(11), e13541.
- Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1), 10.
- Bon, R. (1979). Allometry in topologic structure of transportation networks. *Quality and Quantity*, 13(4), 307–326.
- Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1), 113–120.
- Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, 27(1), 55–71.
- Cottineau, C., & Vanhoof, M. (2019). Mobile phone indicators and their relation to the socioeconomic organisation of cities. *ISPRS International Journal of Geo-Information*, 8(1), 19.
- Cottineau, C., Hatna, E., Arcuate, E., & Batty, M. (2017). Diverse cities or the systematic paradox of urban scaling laws. *Computers, Environment and Urban Systems*, 63, 80–94.
- Cottineau, C., Finance, O., Hatna, E., Arcuate, E., & Batty, M. (2018). Defining urban clusters to detect agglomeration economies. *Environment and Planning B: Urban Analytics and City Science* 2399808318755146.
- Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. (2012). The livehoods project: Utilizing social media to understand the dynamics of a city. *Sixth international AAAI conference on weblogs and social media*.
- Dannemann, T., Sotomayor-Gómez, B., & Samaniego, H. (2018). The time geography of segregation during working hours. *Royal Society Open Science*, 5(10).
- Depersin, J., & Barthelemy, M. (2018). From global scaling to the dynamics of individual cities. *Proceedings of the National Academy of Sciences of the United States of America*, 115(10), 2317–2322.
- Echenique, M. (2001). *Mobility and space in metropolitan areas. Cities for the new millennium*.
- Foth, M. (2018). Participatory urban informatics: Towards citizen-ability. *Smart and Sustainable Built Environment*, 7(1), 4–19. <https://doi.org/10.1108/SASBE-10-2017-0051>.
- Fragkias, M., Lobo, J., Strumsky, D., & Seto, K. C. (2013). Does size matter?: Scaling of CO₂ emissions and us urban areas. *PLOS ONE*, 8(6), e64727.
- Frankhauser, P. (1998). The fractal approach. A new tool for the spatial analysis of urban agglomerations. *Population: An English Selection*, 205–240.
- Gomez-Lievano, A., Youn, H., & Bettencourt, L. M. (2012). The statistics of urban scaling and their connection to Zipf's law. *PLoS ONE*, 7(7), e40393.
- Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779.
- Grady, D., Brune, R., Thiemann, C., Theis, F., & Brockmann, D. (2012). *Modularity maximization and tree clustering: Novel ways to determine effective geographic borders. Handbook of optimization in complex networks*. Springer169–208.
- Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional*

- Science Association*, 24(1), 6–21. <https://doi.org/10.1007/BF01936872>.
- Humeres, F. (2014). *Power centrality as a relational measure of urban hierarchy: Testing the splintering urbanism theory with social media data from Santiago de Chile*. Ph.D. thesis. Massachusetts Institute of Technology.
- Humeres, F. J., & Samaniego, H. (2017). Dissecting the spatial structure of cities from human mobility patterns to define functional urban boundaries. (arXiv preprint) arXiv:1709.06713.
- Jiang, B. (2007). A topological pattern of urban street networks: Universality and peculiarity. *Physica A: Statistical Mechanics and its Applications*, 384(2), 647–655.
- Justen, A., Martínez, F. J., & Cortés, C. E. (2013). The use of space-time constraints for the selection of discretionary activity locations. *Journal of Transport Geography*, 33, 146–152. <https://doi.org/10.1016/j.jtrangeo.2013.10.009> publisher: Elsevier Ltd ISBN: 0966-6923.
- Kenworthy, J. R., & Laube, F. B. (1999). Patterns of automobile dependence in cities: An international overview of key physical and economic dimensions with some implications for urban policy. *Transportation Research Part A: Policy and Practice*, 33(7), 691–723.
- Keuschnigg, M., Mutgan, S., & Hedström, P. (2019). Urban scaling and the regional divide. *Science Advances*, 5(1), eaav0042.
- Langville, A. N., & Meyer, C. D. (2005). A survey of eigenvector methods for web information retrieval. *SIAM Review*, 47(1), 135–161.
- Leitão, J. C., Miotto, J. M., Gerlach, M., & Altmann, E. G. (2016). Is this scaling nonlinear? *Royal Society Open Science*, 3(7), 150649.
- López, E. (2017). *Concepts, definitions and data sources for the study of urbanization: The 2030 agenda for sustainable development*.
- Louail, T., Lenormand, M., Ros, O. G. C., Picornell, M., Herranz, R., Frias-Martinez, E., et al. (2014). From mobile phone data to the spatial structure of cities. *Scientific Reports*, 4.
- Louf, R., Roth, C., & Barthelemy, M. (2014). Scaling in transportation networks. *PLOS ONE*, 9(7), e102007.
- Masucci, A. P., Arcaute, E., Hatna, E., Stanilov, K., & Batty, M. (2015). On the problem of boundaries and scaling for urban street networks. *Journal of the Royal Society Interface*, 12(111), 20150763.
- Meirelles, J., Neto, C. R., Ferreira, F. F., Ribeiro, F. L., & Binder, C. R. (2018). Evolution of urban scaling: Evidence from Brazil. *PLOS ONE*, 13(10), e0204574.
- MINVU (2017). *linkJ*. <http://www.minvu.cl/>.
- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8577–8582.
- Oliveira, E. A., Furtado, V., Andrade, J. S., & Makse, H. A. (2018). A worldwide model for boundaries of urban settlements. *Royal Society Open Science*, 5(5), 180468.
- Paddison, R. (2000). *Handbook of urban studies*. Sage.
- Peixoto, T. P. (2014). *The graph-tool python library*, figshare. <https://doi.org/10.10884/m9>.
- figshare.1164194http://figshare.com/articles/graph_tool/1164194.
- Phithakkitnukoon, S., Smoreda, Z., & Olivier, P. (2012). Socio-geography of human mobility: A study using longitudinal mobile phone data. *PLoS ONE*, 7(6), e39253.
- Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., et al. (2010). Redrawing the map of great britain from a network of human interactions. *PLoS ONE*, 5(12), e14248.
- Rogers, R., & Burdett, R. (2001). *Lets cram more into the city. Cities for the new millennium*9–14.
- Rosenfeld, H. D., Rybski, D., Andrade, J. S., Batty, M., Stanley, H. E., & Makse, H. A. (2008). Laws of population growth. *Proceedings of the National Academy of Sciences of the United States of America*, 105(48), 18702–18707.
- Rosenfeld, H. D., Rybski, D., Gabaix, X., & Makse, H. A. (2011). The area and population of cities: New insights from a different perspective on cities. *American Economic Review*, 101(5), 2205–2225.
- Samaniego, H., & Moses, M. E. (2008). Cities as organisms: Allometric scaling of urban road networks. *Journal of Transport and Land Use*, 1(1), 21–39.
- Šćepanović, S., Mishkovski, I., Hui, P., Nurminen, J. K., & Ylä-Jääski, A. (2015). Mobile phone call data as a regional socio-economic proxy indicator. *PLOS ONE*, 10(4), e0124160.
- Tannier, C., & Thomas, I. (2013). Defining and characterizing urban boundaries: A fractal analysis of theoretical cities and Belgian cities, computers. *Environment and Urban Systems*, 41, 234–248.
- Tannier, C., Thomas, I., Vuidel, G., & Frankhauser, P. (2011). A fractal approach to identifying urban boundaries. *Geographical Analysis*, 43(2), 211–227.
- Thiemann, C., Theis, F., Grady, D., Brune, R., & Brockmann, D. (2010). The structure of borders in a small world. *PLoS ONE*, 5(11), e15422.
- United Nations (2016). *The worlds cities in 2016 data booklet*.
- United Nations (2019). *World urbanization prospects 2018*, Tech. rep. Department of Economic and Social Affairs<https://population.un.org/wup/>.
- van Raan, A. F., van der Meulen, G., & Goedhart, W. (2016). Urban scaling of cities in the Netherlands. *PLOS ONE*, 11(1), e0146775.
- Vanhoof, M., Reis, F., Smoreda, Z., & Ploetz, T. (2018). Detecting home locations from CDR data: Introducing spatial uncertainty to the state-of-the-art. (arXiv preprint) arXiv:1808.06398.
- Veregin, H., & Tobler, W. R. (1997). Allometric relationships in the structure of street-level databases. *Computers, Environment and Urban Systems*, 21(3–4), 277–290.
- West, G. B., Brown, J. H., & Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science*, 276(5309), 122–126.
- Zhong, C., Schläpfer, M., Müller Arisona, S., Batty, M., Ratti, C., & Schmitt, G. (2017). Revealing centrality in the spatial structure of cities from human activity patterns. *Urban Studies*, 54(2), 437–455.