

```

\documentclass[12pt]{nature}
\usepackage[ngerman,british]{babel}
\usepackage{amsmath}
\usepackage{graphicx}% Include figure files
\usepackage{bm}% bold math
\usepackage{setspace}
\usepackage{float}
\usepackage{hyperref}
\usepackage{csquotes}
\usepackage{lineno}
\usepackage[labelfont=bf,font=small]{caption}
\onehalfspacing
\title{\textbf{Defining a city - delineating urban areas using cell phone
data}}
\author{Lei Dong{1}$, Fabio Duarte{2}$, Gilles Duranton{3}$, Paolo
Santi{2,4*}$, Marc Barthelemy{5,6}$, Michael Batty{7}$, Lu{\'\i}s
Bettencourt{8}$, Michael Goodchild{9}$, Gary Hack{3,10}$, Yu
Liu{1}$, Denise Pumain{11}$, Wenzhong Shi{12}$, Vincent
Verbavatz{5}$, Geoffrey B. West{13}$, Anthony Yeh{14}$, Carlo
Ratti{2}}$}
\begin{document}
\maketitle
\begin{affiliations}
\item \small{Institute of Remote Sensing and Geographic Information
Systems, School of Earth and Space Sciences, Peking University, Beijing,
China}
\item \small{Senseable City Laboratory, Massachusetts Institute of
Technology, Cambridge, MA, USA}
\item \small{University of Pennsylvania, Philadelphia, PA, USA}
\item \small{The Institute of Informatics and Telematics CNR, Pisa,
Italy}
\item \small{Université Paris-Saclay, CNRS, CEA, Institut de Physique
Théorique, Gif-sur-Yvette, France}
\item \small{Centre d'Analyse et de Mathématique Sociales (CNRS/EHESS),
Paris, France}
\item \small{University College London, London, UK}
\item \small{University of Chicago, Chicago, IL, USA}
\item \small{University of California, Santa Barbara, CA, USA}
\item \small{Massachusetts Institute of Technology, Cambridge, MA, USA}
\item \small{Université Paris 1 Panthéon-Sorbonne, Paris, France}
\item \small{Hong Kong Polytechnic University, Hong Kong, China}
\item \small{Santa Fe Institute, Santa Fe, NM, USA}
\item \small{Hong Kong University, Hong Kong, China}
\item[] {*}$Corresponding author: psanti@mit.edu
\end{affiliations}
\clearpage
\noindent \textbf{What is a city? Researchers use different criteria and
datasets to define it - from population density to traffic flows. We
argue there is one dataset that could serve as a proxy of the temporal
and spatial connections that make cities what they are: geolocated data
from the world's over 7 billion cell phone users. Cell phone data is a
proxy of people's presence in a given area and of their movement between
areas. Combined with computational methods this data can support city
delineations that are dynamic, responding to multiple statistical and
administrative requirements, tailored to different research needs, thus
accelerating ongoing work in urban science.}
\maketitle
\linenumbers

```

\noindent

The first cities emerged more than 5,000 years ago in ancient Sumeria, where the agricultural surpluses led to the creation of marketplaces that formed the seeds of an embryonic urban economy based on exchange\cite{bairoch1988cities}. Cities grew up around these marketplaces, where populations serving their agricultural hinterlands were packed in close proximity so that the costs of distribution could be minimized, and fortifications were built to keep the city intact from invaders intent on capturing the market and its resources. Until the Industrial Revolution, most cities remained small, with the biggest barely exceeding a million people. It took the invention of new mechanical transportation technologies for cities to begin to exceed this limit for the first time. By the end of the 19th century, London had grown to over 6 million people and New York to over 4 million\cite{chandler1974three}. As a key factor, transportation and technological advances have further broken the historical boundaries of cities. Populations began to spread out into their hinterlands, blurring the line between urban and rural. By the middle of the last century, some large cities have begun to fuse with one another creating megacities such as on America's eastern seaboard\cite{gottmann1961megalopolis}, while more recent growth has led to urban complexes like China's Greater Bay Area, which now exceeds 86 million population. From a qualitative perspective, many scholars have addressed the question of "What is a city?"\cite{mumford1937city,varzi2021city}; and from a quantitative perspective, in the 1960s, Berry developed the delineation of metropolitan areas using commuting flows\cite{berry1960impact,berry1969metropolitan}, and the current boundaries of metropolitan areas in the US and some other regions/countries are still based on this methodology\cite{office2010,adams1999metropolitan}. More recently, the UN-Habitat published a report under the premise that, to achieve the Sustainable Development Goals, we need to propose solutions to key urban challenges, and to do so, we need a clear and globally agreed definition of what a city is\cite{UN-Habitat}. This report discusses two definitions: one defines the city by its urban extent, and uses satellite imagery to identify the morphology of human settlements and the density of built-up structures; another defines the city by its Degree of Urbanization\cite{dijkstra2021applying}, based on population (density) using grid cells, which is also endorsed by the United Nations Statistical Commission\cite{UN}. Although laudable, two challenges still remain: one definition is based on physical features and the other on granular population data, and neither considers flows – which we will discuss later. UN-Habitat also pointed out that to operationalize any local or global policy, we need timely data and metrics. Unprecedented geospatial datasets, increased processing power, and recent advances in computational methods are now providing new opportunities to expand our understanding of cities, and to make the delineation of cities more accurate and timely. In fact, a new science is rapidly emerging. Starting from such datasets and methods, it aims to quantitatively describe cities and systems of cities\cite{berry1964cities,henderson1974sizes,pumain2004scaling,barthelemy2016structure,batty2013new,bettencourt2021introduction,durant2023urban,caldarelli2023role}. Despite the 'discovery' of some relatively robust urban laws (e.g., gravity, allometric scaling, rank-size), we are still at the very beginning of elaborating this new science of cities, and what is still missing is a widely accepted and thoroughly discussed methodology to delineate a city.

This paper discusses how scholars and policy officials have been making efforts to reach theoretically sound and operational definitions of the city; the challenges that these efforts are still facing; and how the use of global, standardized datasets such as cell phone records, coupled with rigorous guidelines and methods for delineating cities informed by such datasets, can help make substantial steps towards widely agreed city delineations for urban science. A key argument of this paper is that there is an essential feature that could unify the most well-accepted definitions of cities: the presence and flows of people. And we propose cell phone data as a possible dataset representing a universal proxy for the presence and flows of people that define cities. After all, as Shakespeare said: "What is the city but the people?" (Shakespeare, *Coriolanus*, Act 3, Scene 1).

`\subsection*{Existing definitions}`

Before moving forward, we must pose an important clarification of terminologies: `\textit{definition}` refers to the meaning, or how we explain something; whereas `\textit{delineation}` refers to how we establish the boundaries of something – in our case, the definition is the set of rules one applies and the delineation is the city boundaries resulting from the implementation of these rules.

`\subsubsection*{Definitions based on administrative boundaries}`

`\noindent` The administrative divisions traditionally used to delineate cities are based on very different parameters. For example, Brazil has 5,570 municipalities, 68% of which have fewer than 20,000 inhabitants, and Barcelos, slightly above this threshold, has an area larger than North Korea or New York State; on the other hand, there are only about 300 prefecture-level cities in China, with an average population of nearly 4 million people\cite{china2022}. These vastly different administrative delineations of cities in different regions pose a major challenge for comparative studies on a global scale.

To tackle this issue, national or transnational statistical offices have long used quantitative definitions of metropolitan areas. One common definition of a metropolitan area consists of core urban areas and commuting zones surrounding the center. From a labor market perspective, this dependency is usually represented by a combination of population and commuting trips. For example, Core-Based Statistical Areas (CBSAs) in the US are defined by starting from a county containing a city of 10,000 to 50,000 (micropolitan areas) or at least 50,000 (metropolitan areas) inhabitants, which grows to encompass surrounding counties based on criteria that combine population density and a minimum threshold of commuting flow to the main city\cite{office2010}. Using this definition, over 900 CBSAs are identified in the US, with New York-Newark-Jersey City being the largest, with over 20 million inhabitants.

In China, a quantitative definition of metropolitan areas was not available until very recently\cite{chen2023prefecture}. According to China's National Development and Reform Commission, metropolitan areas are centered on megacities (10 million people or more) or large cities (5 million) and are bounded by a one-hour commuting buffer. With this high population size criterion, the number of metropolitan areas in China is significantly fewer than in the US or the EU.

In an effort to facilitate international comparison, metropolitan areas in the European Union (EU) are defined using the notion of Functional Urban Areas (FUAs)\cite{cheshire1989urban,dijkstra2019}. According to this definition, one must first identify an "urban center" based on a minimum total population of 50,000 and additional density requirements. Then, surrounding centers are added to the FUA based on a combination of population size and commuting flow requirements. Following this definition, 828 FUAs are identified in the EU, Switzerland, Iceland and

Norway, with Paris being the largest with a population of over 13 million inhabitants.

While created to facilitate comparison, city definitions based on fixed thresholds on population, density, etc., such as FUA, address the problem of city delineation as a top-down, "one-size-fits-all" approach, which might imply, for instance, that the capital of a small country such as Liechtenstein is not even considered a city. The notion of FUA also requires access to multiple data sources such as residential/nighttime population and commuting flows, which might further hinder the wide application of such an approach to city delineation. What is lacking then is a systematic methodology for determining parametric thresholds (e.g., total population, population density, volume of commuting flows) to be used in combination with universally accessible datasets.

\subsubsection\*{Recent advances}

\noindent With the advent and availability of granular geospatial data, new proposals and techniques of city delineation have emerged\cite{duranton2021classifying,moreno2021metropolitan}. They mostly fall into two categories, which we summarize as \textit{form-} and \textit{function-}based delineations.

Form-based approaches start by classifying the basic spatial unit (e.g., a geographic grid) as urban or non-urban using population, buildings, built-up area, or nighttime light intensity, and then define cities as continuous urban

units\cite{rozenfeld2011area,dingel2021cities,de2021delineating,tannier2011fractal}. For example, it has been suggested that a city is defined as a continuous cluster of more than \textit{X} inhabitants and/or \textit{Y} nighttime light intensities. The choice of \textit{X} and \textit{Y}, however, varies considerably among the different studies. Even when the choice is made using statistical thresholds\cite{de2021delineating}, the arbitrariness is reduced but not eliminated.

Function-based approaches start from the observation that areas through which a significant fraction of people move within a given time should be considered part of the same city. In this case, the definition of a city as a network of socioeconomic relationships can be proxied by flows of people and goods within a certain space. This approach is especially relevant to the complex network community, which seeks to define and characterize cities as evolving networks of exchange and interaction\cite{batty2013new,bettencourt2021introduction}. Examples of function-based delineations include megaregions in the US identified by commuting flows\cite{nelson2016economic}, metropolitan areas in Switzerland\cite{dessemontet2010switzerland}, and functional areas in China\cite{ma2020functional,chen2022delineating,chenting2022delineating}. As mentioned earlier, whenever the notion of flow is involved, data can become a serious limitation.

Notably, the availability of remote sensing imagery coupled with increased computational power has led to the emergence of many land classification studies in recent years\cite{liu2020high}. These studies typically use supervised machine learning models to perform classification of remote sensing data and label pixels of built-up areas as an important land use class\cite{baragwanath2021detecting,galdo2021identifying}. Although land classification uses so-called "ground truth" to determine whether a given pixel is urban (e.g., buildings, roads) or not (e.g., forest, rivers), deriving a delineation of the city as a whole from this information requires additional analytical steps. For example, while the pixels of built-up areas identified by remote sensing imagery need to be aggregated to define a city, we must also determine how much of the contiguous

built-up area is a city, and whether parks and green spaces, which are often classified as non-built-up areas, are part of the city. More importantly, satellite imagery mainly captures the physical and built environment of cities, and it is difficult to estimate the movement of people in the city and the resulting socioeconomic changes, which are arguably more critical to defining cities.

The drawbacks of existing city definitions based on administrative boundaries or solely on morphological or flow criteria are summarized in Box 1.

*\subsubsection\*{Requisites of a good definition}*

*\noindent* In order to allow for comparative and universal analysis, we advance some criteria for defining cities that could be widely accepted by the urban science community. Based on the above discussion, and building upon previous

work\cite{duranton2015proposal,cao2023constructing}, we propose six criteria as shown in Box 2:

The first criterion states that the principles upon which a city definition is based should be easy to understand by a non-specialized audience. Also, the data processing and implementation procedures used to empirically test this definition should be easily replicable, so that a definition can be used in different contexts and by people with diverse backgrounds and know-how. Note that it is quite common that a conceptually clean definition translates into highly non-trivial data processing and implementation issues. In such cases, these non-trivial procedures should be made open access and adequately documented to facilitate use for as wide an audience as possible.

The second criterion describes the desirable characteristics of the dataset(s) on which a city definition should be based. By universal, we mean that the type of data used to ground a definition should not be specific to or unique to a particular geography/country, but should, to the maximum extent possible, be available in all cities. For example, considering data related to the movement of people in the city, a dataset pertaining to a specific travel mode would be useful only for the definition of cities where such a specific travel mode exists.

Besides being universal, the datasets that can be used to delineate a city should be relatively easy to access. Ideally, one should use data that is publicly available. Such wide accessibility would allow any researcher, scholar, or practitioner to collect and apply the necessary data to delineate a city within its geography of interest. However, publicly accessible data are typically highly aggregated for reasons related to storage/computing costs as well as privacy concerns, and do not have the granularity necessary for use in quantitative city definition. Data with higher granularity require extensive resources for storage and processing, are highly valuable to the entity that generated/acquired them, and may raise high privacy concerns. For these reasons, access to highly granular data is often restricted, and getting access to it requires signing NDAs and/or paying data access fees. Access to data could hence become a major challenge in the quest for a universal, quantitative city definition—a topic which will be discussed further in this paper.

The third property argues for the need to consider both formal and functional features of a city. Clearly, how these features are incorporated into a city delineation depends on which aspects of a city we are interested in analyzing. Taking as an example the Greater Bay Area in China, being interested in cities as labor markets might suggest we focus on, say, 10 spatial units (cities) in that area. On the other hand, an analysis focused on clusters of innovative activity might isolate only three places (Hong Kong, Shenzhen, and Guangzhou) within the same area.

Thus, the form and functional features to be considered in the city delineations required for these two exemplary analyses should be defined accordingly.

The fourth property points to the need to carefully select parameters and thresholds in defining a city. Arbitrary parameter selection would \textit{de facto} impair the ability to compare cities across regions/countries and to address global research questions related to urbanization. For example, the UN Statistical Commission suggests defining cities based on a total population of at least 50,000 and a minimum population density threshold\cite{UN}. On the other hand, in the US, the criteria for defining a city based on population vary by state, with some states, such as Massachusetts, requiring at least 12,000 inhabitants, and others, such as Iowa, having no minimum population requirement at all.

The UN approach\cite{UN} uses an absolute definition, meaning that a human settlement would be classified in the same way across space and time. And the US cases are relative definitions that identify cities in a context-dependent manner and thus vary across space and time. While absolute definitions have many advantages, such as simplicity, stability, and ease of implementation, we must recognize that urbanization is a dynamic process and varies greatly from country to country (a settlement of 5,000 people in the Congo is nowhere close to a settlement of 5,000 people in, say, the Netherlands). We, therefore, recommend the relative definition or a combination of relative and absolute definitions, with fixed parameters within a country and different parameters for different countries. In such settings, the use of context-specific thresholds should be carefully controlled, and, if possible, it might be useful to establish data-driven methods to set parameters and thresholds used in a city delineation (an example of such a parameter-free method used for computing thresholds to define city hotspots from cell phone data was proposed in Ref.\cite{louail2014mobile}). More will be discussed in the next section).

The fifth property highlights the importance of seeking extrinsic assessment methods for a city definition. This property is especially important and challenging. One possibility would be to evaluate a city definition both qualitatively as well as by checking whether some fundamental properties observed in multiple metropolitan regions of the world also hold for the "cities" resulting from the specific city definition at hand. For example, recent studies of urban scaling laws have found that many fundamental scaling properties hold only within specific functional city boundaries\cite{rosen1980size,stier2022reply}. As shown in Box 3, we suggest first computing the population and area of a "city" as defined by the specific definition at hand, and then analyzing the properties of the "city" population/area distribution over a larger geography (e.g., a country).

Lastly, depending on the case under study, administrative boundaries are inescapable, as usual statistical indicators (e.g., population, economic or political indices) are generally collected at the administrative level. This means that matching with existent political boundaries would allow the analyst to have access to a lot of data available, for instance, through census data. However, the benefit of this additional source of information heavily depends on the granularity of the smallest spatial unit used to aggregate such information, which can range from a few blocks in a dense urban area in the US, to hundreds of square kilometers for countries such as Colombia. It should also be noted that administrative boundaries do have a critical impact on cities, given the role they play in "acting" on the city through planning and policy-making—thus, the very administrative boundaries drive or condition the

presence and flows of people, as well as the city's form and functions. Therefore, in such cases where socio-economic-political indicators are relevant, a good definition of a city should not overlook the role of administrative boundaries. On the other hand, the delineation of administrative boundaries could be updated to better reflect the actual formal and functional characteristics of an urban area as defined by a good city delineation strategy. What we are looking for is a delineation that attempts to find the best fit between a purely data-driven delineation and existing spatial boundaries.

\subsection\*{Opportunities from using cell phone data}

We propose cell phone data as the basic dataset in finding a solution to the quest for standardized city definitions. Cell phone data typically come in one of two forms: call detail records (CDR) or global positioning system (GPS) records. CDR reports geolocated events such as sending a text, making or receiving a call, etc. The record typically contains at least an (anonymized) user ID, the time of the event, and the ID of the cell tower to which the user is connected at the time of the event. GPS records provide more detailed spatio-temporal information, typically reporting at least the coordinates (latitude, longitude) of the data point, the time, and possibly other information such as speed, heading, and so on. Note that we mainly consider the geospatial attributes of cell phone data, which have been widely used in the study of human activity and urban science in recent

decades\cite{blondel2015survey,pappalardo2023future}. There are also a few studies that use cell phone communication information to study social networks\cite{park2018strength}. However, compared with location attributes, individual-level cell phone data with rich semantics is very difficult to access, so the cell phone data used in this paper mainly refers to cell phone location data.

The main difference between the two types of cell phone datasets is in the spatial and temporal resolution of the information - as shown in Fig. 1. In the case of CDR, the spatial resolution is determined by the coverage area of the cell tower, which ranges between a few hundred meters in densely populated areas to a few tens of square kilometers in low-density areas - with a tendency of reducing cell coverage area with newer cellular communication technologies such as 5G. The temporal resolution for CDR data depends on the frequency of recorded events, which in turn is determined by how frequently the user uses the phone for texting, calling, browsing the Internet, etc. GPS data has a much higher accuracy both in time and space. Spatial accuracy is typically in the range of a few tens of meters, while the temporal recording of events is typically quite regular, with sampling intervals ranging from a few seconds to a few minutes.

It is estimated that about 97\% of the world's population is covered by a cellular network. Thanks to this ubiquity and the worldwide availability of the Global System for Mobile communication (GSM) standard, cell phone data provide a consistent collection mechanism across the world and are independent of national censuses. This meets the requirements of the second point described in Box 2. However, access to cell phone data is not easy, which might challenge the requirement for easy access to data described in Box 2. The issue of access to cell phone datasets is further discussed at the end of this paper.

Importantly, when it comes to city definition and delineation, cell phone data offer the ability to capture both form features (such as nighttime/daytime presence of people in an area) and functional features (such as mobility flows between two areas) - as shown in Fig. 1. With well-established methods for estimating stay points and origin-destination flows from trajectories\cite{alexander2015origin}, cell phone

data can be used to compute both high-resolution population distributions and commuting flows, becoming a valid alternative to classical labor market approaches that provide a satisfactory answer only to the second point. Note that cell phone data can be used to estimate flows in the city more generally, allowing for an accurate picture of movement patterns within a given geographical area, opening up the possibility of defining functional cities with greater granularity.

Recent research based on cell phone data and urban science has also shed additional light on the fourth and fifth points described in Box 2, showing the potential of using cell phone data to uncover foundational features of human movement that can provide a quantitative ground for city delineation. For instance, Alessandretti et al. \cite{alessandretti2018evidence} discovered that humans tend to perform a constant number of weekly movements across different geographies, while Schläpfer et al. \cite{schlapfer2021universal} unveiled an inverse square law between distance and frequency of trips. Other works following this line of research are by Refs. \cite{grauwin2017identifying,alessandretti2020scales}.

Moreover, as the result of a process with self-organization characteristics, the city system has been found to exhibit discontinuous changes - phase transitions \cite{batty2013new,barthelemy2019statistical}. One transition example was provided by the hierarchical percolation on the British road network and on its intersections, which exhibit abrupt changes at critical parameter values \cite{arcaute2016cities}. The clusters defined at these discontinuous thresholds were found to be in good agreement with city boundaries estimated from satellite imagery. Similar percolation phenomena have been found in city systems in many parts of the world \cite{cao2020quantifying,montero2021delineation}. These characteristics allow us to define cities using endogenous parameters - determined by the thresholds of the system, referred to as \textit{critical}, which mark a phase transition, making the thresholds in the definition process obtained from the data, rather than purely arbitrary settings (the fourth point in Box 2). It is worth noting that the commonly used sensitivity analysis does not necessarily yield effective parameter thresholds. Because the result can vary a lot when the threshold exceeds the phase transition point \cite{arcaute2016cities}, which is often not considered a good threshold in sensitivity analysis.

In practice, the variability of form and function of cities in different regions, and the quality of the data itself, make it not always possible to find such discontinuities \cite{bosker2021definition}. Even with critical parameter values, we still need a set of metrics to evaluate whether the city boundaries derived from cell phone data are reasonable (the fifth point in Box 2). In particular, Zipf's law has been used in previous work as evidence of good delineation \cite{rozenfeld2011area}. Yet, recent work \cite{verbavatz2020growth} has shown that Zipf's law can be problematic for city definition as the law's exponents fluctuate considerably in empirical data \cite{cura2017old,cristelli2012there}. With cell phone data, we can do far more than Zipf's law to validate a city delineation. For example, we can examine the allometric urban scaling laws  $Y \sim N^{\beta}$ , where  $N$  is the population in each city cluster and  $Y$  is the variable of area, interaction, and output of each city cluster \cite{west2017scale}. The scaling exponent  $\beta$  can be estimated from cell phone data and can be used to test whether the exponent in the boundary quantified by cell phone data is consistent with the theoretical prediction \cite{cao2023constructing}.

Returning to the global comparison, cell phone data and the aforementioned characteristics of city systems provide an unprecedented



opportunity to consider both absolute and relative city definitions. The consistency and universality of cell phone data makes it easier to quantify cities using absolute thresholds. Importantly, the granularity of cell phone data also allows us to set a 'relative absolute threshold'. For example, one can take the 95th percentile of distance to work and define local labor markets and corresponding city boundaries. The methodology for doing this remains to be developed given the continuous nature of the data (Figure 2 shows one possible way to find such thresholds). Yet, a big advantage here is that the distance bands obtained by these statistical thresholds (absolute definition) will differ across countries (relative definition).

As a summary, in Box 3 and Fig. 2, we outline a feasible process for city delineation with cell phone data. First, we use cell phone data to calculate the population density within a grid and merge continuous high-density grids into clusters. Second, also using cell phone data, we calculate the commuting flows between clusters and further aggregate clusters with flows above a certain threshold. Note that this step of the definition is similar to the FUA's, which also require two parameters: population density and commuting ratio. Third, the critical characteristics of the city system make these two parameters endogenous and can be determined from the phase diagram. After the above steps, we can obtain high-density and tightly connected city clusters. Fourth, the derived city clusters can be validated by using a set of criteria (e.g., scaling laws, similarity to the existing metropolitan boundaries). Recall the last point in Box 2, to make the results more useful in practice, we can further aggregate the clusters at the administrative level.

\subsection\*{Conclusions}

To summarize, in this article we have extensively discussed why we believe that the urban science community should carefully consider the problem of how to define cities and metropolitan areas, and have proposed a set of guidelines and widely accepted methods for delineating cities. We are not advocating here for a single definition, as we believe that different ones will be needed depending on the specific problem and research question being analyzed. Rather, we advocate for the use of delineations that conform to the conceptual framework described in Box 2, so that, regardless of the specific definition used for the problem at hand, the requirement of universality, leveraging standard and widely available data, is still satisfied. This approach would facilitate the replicability of studies and comparisons across different geographies, bringing invaluable benefits to the research community.

We have also advocated the use of cell phone data as a proxy for the essential features of any city definition: urban form and function. And cell phone datasets are not the only example of standardized data that can be used for this purpose. If needed, the integration of cell phone datasets with point of interest (POI) information, global road network repositories such as OpenStreetMap, social network data, remote sensing imagery, etc., could provide additional opportunities for city delineation. However, a challenge in this context is that some of these complementary datasets may not be globally available in the same format. It is important to note here that the spatial and temporal resolution of the data is a critical factor when considering access to data that is related to human ethics and privacy protection. While gaining access to individual-level GPS trajectories might be extremely difficult, this level of detail is often not necessary in the process of delineating city boundaries. For this purpose, access to spatially and temporally aggregated data (e.g., how many people moved from area A to area B in one day) is often sufficient. Generally speaking, the coarser the level of spatio-temporal resolution, the easier it is to gain access to a dataset.

Thus, the question of which is the minimal resolution needed in the context of city delineation is a prominent question that has mostly been evaded so far in the literature. Note that there is no single answer to this question since, as extensively argued in this paper, how to define a city depends on the context and scope of the study, and so does the minimum data resolution needed to support a specific city definition. Similarly, the data aggregation level and chosen spatial unit should be suited to the context and scope of the study.

Cell phone datasets come with limitations related to the fact that, while cell phone ownership is wide, still some socioeconomic groups tend to be overrepresented (younger, wealthier populations) or underrepresented (older populations). If we could have a globally covered functional city dataset, scholars and policy makers in a variety of fields such as urban planning, transportation, epidemiology, and sustainability would benefit greatly. Yet, despite caveats and the uneven penetration of smartphones, the more than 18 billion cell phones projected to be in use globally by 2025 are arguably the most widely spread proxy for the presence and flows of people.

Global access to cell phone data is a challenge: even if an increasing number of open cell phone datasets have been made available recently [such as the D4D challenge, the Uber movement (<https://movement.uber.com/>), and Baidu Migration (<https://qianxi.baidu.com/>)], the data in general is owned by private companies, protected by NDAs or access fees, strict privacy regulations, and so on.

In this regard, we would encourage the urban science community to join forces to develop satisfactory definitions of cities based on cell phone data or other standardized, global datasets, and to work with international organizations, such as the International Telecommunication Union, to make anonymized cell phone data available more broadly, since a clear delineation of cities is critical to addressing current and future urban problems affecting populations globally. Collecting, pooling, and analyzing such data on a global scale requires nothing less than a collective effort, which we all call for.

**Corresponding author:** Correspondence about this paper should be sent to Paolo Santi, [psanti@mit.edu](mailto:psanti@mit.edu).

**Acknowledgements:** L.D. and Y.L. acknowledge the support of the National Natural Science Foundation of China (No. 41830645). M.B. (Michael Batty) acknowledges support from the Alan Turing Institute under QUANT2-Contract-CID-381581. G.B.W. acknowledges the generous support of Toby Shannon via the Charities Aid Foundation (CAF) of Canada and of the NSF under grant no. PHY1838420. F.D, P.S., and C.R. acknowledge all members of MIT Senseable City Lab consortium.

**Author contributions:** All the authors participated in the discussion and writing of the paper.

**Conflict of interest:**

The authors declare no conflict of interest.

`\clearpage`

`\begin{figure}[htbp]`

`\centering`

`%\includegraphics[width = .8\textwidth]{fig1.pdf}`

`\caption{\textbf{Cell phone data and process methods.}` (a) Processing procedures for CDR data and GPS data. For GPS data, the user's activity points can be obtained by clustering algorithms (e.g., DBSCAN). Both the CDR data and the GPS data are then passed through an anomaly detection process, which is used to filter out noise (e.g., false movements). Then, by setting a stay time (e.g., 30 minutes), we can obtain the stay points of each user. With stay points, user's home and work locations can be

estimated in a rule-based approach, most commonly by calculating the locations that users visit most during the day on weekdays and most at night over a period of time. Note that if there are users with known home and work locations, we can use supervised machine learning algorithms to infer those locations more accurately. With the user's home and work locations, it is easy to map the daytime and nighttime population distributions and commuting flows within the city (where calibration using census data might be necessary, since cell phone data is a sampling of the population). In addition to commuting flows, all cell phone estimated trips can be calculated by stay points to obtain an Origin-Destination (OD) matrix between locations. The OD matrix is one of the most important results, by which the spatio-temporal structure of the city can be obtained. (b) The commuting flow in Guangdong province (China) calculated by the method illustrated in (a).}

```
\label{fig1}
\end{figure}
\clearpage
\begin{figure*}[ht]
\centering
%\includegraphics[width = .8\textwidth]{fig2.pdf}
\caption{\textbf{An illustration delineating cities using cell phone data. }. (a) A City Clustering Algorithm (CCA)\cite{rozenfeld2011area} was applied to the discrete population density layer estimated by cell phone data to derive the continuous population clusters. This step has two noteworthy points: first, the CCA can also be used for continuum distribution datasets (not necessarily aggregated into grid cells); and second, this step can also add other data sources (e.g., nighttime lights, roads, POIs). Therefore, the grid cells that meet multiple (or any one) density conditions are used as candidate cells for city delineation. (b) We calculate the commuting flows between the continuous clusters obtained in the first step and combine the clusters whose flow proportion exceeds a certain value. (c) By exploiting the critical characteristics of the urban system, it is possible to detect phase transitions (i.e., natural discontinuities), which could make population density and flow thresholds endogenous. Here we calculate the area of the largest cluster under different parameters, and find that under a certain set of parameters, the area of the largest cluster would have a discontinuity (the dashed line). Another approach to set parameters is goal-oriented, as shown in Box 3. (d) Cities defined by (a-c) steps. For comparison, we enlarged the map of the Greater Bay Area and added the boundaries of the existing prefectural-level cities (gray). As shown on the map, Guangzhou, Foshan and Jiangmen are connected into one large metropolitan area, while Shenzhen, Dongguan and Huizhou form another one. Data for these maps were derived from Ref.\cite{cao2023constructing}.
}
```

```
\label{fig2}
\end{figure*}
\clearpage
\fbbox{\begin{minipage}{35em}
\textbf{Box 1 Drawbacks of existing city definitions}
\begin{enumerate}
\item There is a multiplicity of levels of administrative boundaries from which it is hard to choose, and what each of these levels captures differ across countries and even within countries.
\item Both form-based and function-based approaches require some arbitrary parameters.
\item Function-based definitions require data that was not readily available before, especially for developing countries.
```

\item Given the above points, it is very difficult to perform cross-regional and cross-country comparisons between cities based on existing definitions.

\end{enumerate}

\end{minipage}}

\fbox{\begin{minipage}{35em}

\textbf{Box 2 Criteria for good city definitions}

\begin{enumerate}

\item A city definition should be easy to understand and easy to apply by researchers and practitioners in the field.

\item The data on which a city definition is built upon must be universal, easy to access, and self-explanatory.

\item It should consider both urban form and function, which are the essential elements of a city.

\item A definition should be based on careful control of the involved parameters. Parameter choices should be clearly motivated and, where possible, data-driven, to avoid arbitrary choices. In some cases, data-driven parameter setting might render them endogenous, thus making the city definition parameter free.

\item The city delineations resulting from a definition should seek some form of validation based on a procedure that should be general and widely accepted by the urban science community. Preferably, this procedure should include one or more quantitative criteria to assess the quality of a definition.

\item Depending on the case studied, city delineations resulting from a definition might be later combined with administrative boundaries.

\end{enumerate}

\end{minipage}}

\fbox{\begin{minipage}{35em}

\textbf{Box 3 A feasible process for city delineation with cell phone data}

\begin{enumerate}

\item Population density can be calculated from cell phone data at the grid cell level (a grid of about 1 km is recommended, taking into account the resolution of cell phone data and the scale of people's activities at the neighborhood level).

\item Flow of people between grid cells can also be derived from cell phone data (Fig. 1a).

\item Aggregate grid cells with population densities and flows above certain thresholds. One possible approach to set thresholds is to exploit the critical characteristics of the urban system, which could make these two thresholds endogenous and could be determined from the statistics or the phase diagram (i.e., natural discontinuities). Another approach is goal-oriented, for instance, if the goal is to make the output as close as possible to a city delineation with desired features, such as reproducing FUA boundaries.

\item Multiple dimensional validations (e.g., statistical characteristics in urban systems, similarity to existing metropolitan boundaries in terms of morphology, population size, etc.).

\end{enumerate}

\end{minipage}}

\clearpage

\bibliographystyle{naturemag}

%\bibliography{ref.bib}

\begin{thebibliography}{10}

\expandafter\ifx\csname url\endcsname\relax

\def\url#1{\texttt{#1}}\fi

\expandafter\ifx\csname urlprefix\endcsname\relax\def\urlprefix{URL }\fi

```

\providecommand{\bibinfo}[2]{#2}
\providecommand{\eprint}[2][]{\url{#2}}
\bibitem{bairoch1988cities}
\bibinfo{author}{Bairoch, P.}
\newblock \emph{\bibinfo{title}{Cities and Economic Development: From the Dawn of History to the Present}} (\bibinfo{publisher}{University of Chicago Press}, \bibinfo{year}{1988}).
\bibitem{chandler1974three}
\bibinfo{author}{Chandler, T.} \& \bibinfo{author}{Fox, G.}
\newblock \emph{\bibinfo{title}{Three Thousand Years of Urban Growth}} (\bibinfo{publisher}{Academic Press}, \bibinfo{year}{1974}).
\bibitem{gottmann1961megalopolis}
\bibinfo{author}{Gottmann, J.}
\newblock \emph{\bibinfo{title}{Megalopolis: The Urbanized Northeastern Seaboard of the United States}} (\bibinfo{publisher}{New York: Twentieth Century Fund}, \bibinfo{year}{1961}).
\bibitem{mumford1937city}
\bibinfo{author}{Mumford, L.}
\newblock \bibinfo{title}{What is a city?}
\newblock \emph{\bibinfo{journal}{Architectural Record}}
\textbf{\bibinfo{volume}{82}}, \bibinfo{pages}{93--96}
(\bibinfo{year}{1937}).
\bibitem{varzi2021city}
\bibinfo{author}{Varzi, A.~C.}
\newblock \bibinfo{title}{What is a city?}
\newblock \emph{\bibinfo{journal}{Topoi}} \textbf{\bibinfo{volume}{40}},
\bibinfo{pages}{399--408} (\bibinfo{year}{2021}).
\bibitem{berry1960impact}
\bibinfo{author}{Berry, B.~J.}
\newblock \bibinfo{title}{The impact of expanding metropolitan communities upon the central place hierarchy}.
\newblock \emph{\bibinfo{journal}{Annals of the Association of American Geographers}} \textbf{\bibinfo{volume}{50}}, \bibinfo{pages}{112--116}
(\bibinfo{year}{1960}).
\bibitem{berry1969metropolitan}
\bibinfo{author}{Berry, B.~J.}, \bibinfo{author}{Goheen, P.~G.} \&
\bibinfo{author}{Goldstein, H.}
\newblock \emph{\bibinfo{title}{Metropolitan Area Definition: A Re-evaluation of Concept and Statistical Practice}},
vol.~\bibinfo{volume}{28} (\bibinfo{publisher}{Washington: US Bureau of the Census}, \bibinfo{year}{1969}).
\bibitem{office2010}
\bibinfo{author}{{US Office of Management and Budget}}.
\newblock \bibinfo{title}{{2010 Standards for Delineating Metropolitan and Micropolitan Statistical Areas}}.
\newblock \emph{\bibinfo{journal}{Federal Register}}
\textbf{\bibinfo{volume}{75}}, \bibinfo{pages}{37246--37252}
(\bibinfo{year}{2010}).
\bibitem{adams1999metropolitan}
\bibinfo{author}{Adams, J.~S.}, \bibinfo{author}{VanDrasek, B.~J.} \&
\bibinfo{author}{Phillips, E.~G.}
\newblock \bibinfo{title}{Metropolitan area definition in the {United States}}.
\newblock \emph{\bibinfo{journal}{Urban Geography}}
\textbf{\bibinfo{volume}{20}}, \bibinfo{pages}{695--726}
(\bibinfo{year}{1999}).
\bibitem{UN-Habitat}
\bibinfo{author}{{UN-Habitat}}.

```

\newblock \bibinfo{title}{What is a city?}  
\newblock \emph{\bibinfo{journal}{Technical Report}}  
(\bibinfo{year}{2020}).  
\bibitem{dijkstra2021applying}  
\bibinfo{author}{Dijkstra, L.} \emph{et~al.}  
\newblock \bibinfo{title}{{Applying the Degree of Urbanisation to the  
globe: A new harmonised definition reveals a different picture of global  
urbanisation}}.  
\newblock \emph{\bibinfo{journal}{Journal of Urban Economics}}  
\textbf{\bibinfo{volume}{125}}, \bibinfo{pages}{103312}  
(\bibinfo{year}{2021}).  
\bibitem{UN}  
\bibinfo{author}{{UN Statistical Commission}}.  
\newblock \bibinfo{title}{A recommendation on the method to delineate  
cities, urban and rural areas for international statistical comparisons}.  
\newblock \emph{\bibinfo{journal}{Technical Report}}  
(\bibinfo{year}{2020}).  
\bibitem{berry1964cities}  
\bibinfo{author}{Berry, B.~J.}  
\newblock \bibinfo{title}{Cities as systems within systems of cities}.  
\newblock \emph{\bibinfo{journal}{Papers in Regional Science}}  
\textbf{\bibinfo{volume}{13}}, \bibinfo{pages}{147--163}  
(\bibinfo{year}{1964}).  
\bibitem{henderson1974sizes}  
\bibinfo{author}{Henderson, J.~V.}  
\newblock \bibinfo{title}{The sizes and types of cities}.  
\newblock \emph{\bibinfo{journal}{American Economic Review}}  
\textbf{\bibinfo{volume}{64}}, \bibinfo{pages}{640--656}  
(\bibinfo{year}{1974}).  
\bibitem{pumain2004scaling}  
\bibinfo{author}{Pumain, D.}  
\newblock \bibinfo{title}{Scaling laws and urban systems}.  
\newblock \emph{\bibinfo{journal}{Santa Fe Institute Working Paper}}  
(\bibinfo{year}{2004}).  
\bibitem{barthelemy2016structure}  
\bibinfo{author}{Barthelemy, M.}  
\newblock \emph{\bibinfo{title}{The Structure and Dynamics of Cities}}  
(\bibinfo{publisher}{Cambridge University Press}, \bibinfo{year}{2016}).  
\bibitem{batty2013new}  
\bibinfo{author}{Batty, M.}  
\newblock \emph{\bibinfo{title}{The New Science of Cities}}  
(\bibinfo{publisher}{MIT press}, \bibinfo{year}{2013}).  
\bibitem{bettencourt2021introduction}  
\bibinfo{author}{Bettencourt, L.~M.}  
\newblock \emph{\bibinfo{title}{Introduction to Urban Science: Evidence  
and Theory of Cities as Complex Systems}} (\bibinfo{publisher}{MIT  
Press}, \bibinfo{year}{2021}).  
\bibitem{duranton2023urban}  
\bibinfo{author}{Duranton, G.} \& \bibinfo{author}{Puga, D.}  
\newblock \bibinfo{title}{Urban growth and its aggregate implications}.  
\newblock \emph{\bibinfo{journal}{Econometrica}} (\bibinfo{year}{2023}).  
\bibitem{caldarelli2023role}  
\bibinfo{author}{Caldarelli, G.} \emph{et~al.}  
\newblock \bibinfo{title}{The role of complexity for digital twins of  
cities}.  
\newblock \emph{\bibinfo{journal}{Nature Computational Science}}  
\bibinfo{pages}{1--8} (\bibinfo{year}{2023}).  
\bibitem{china2022}

\bbininfo{author}{National Bureau of Statistics of China}}.  
\newblock \emph{\bbininfo{title}{China City Statistical Yearbook}}  
(\bbininfo{publisher}{China Statistics Press}, \bbininfo{year}{2022}).  
\bbitem{chen2023prefecture}  
\bbininfo{author}{Chen, Z.} \& \bbininfo{author}{Yeh, A. G.-O.}  
\newblock \bbininfo{title}{Is prefecture-level city a "city" in China: A  
critical review}}.  
\newblock \emph{\bbininfo{journal}{Eurasian Geography and Economics}}  
\bbininfo{pages}{1--26} (\bbininfo{year}{2023}).  
\bbitem{cheshire1989urban}  
\bbininfo{author}{Cheshire, P.~C.} \& \bbininfo{author}{Hay, D.~G.}  
\newblock \emph{\bbininfo{title}{Urban Problems in Western Europe: An  
Economic Analysis}} (\bbininfo{publisher}{Unwin Hyman},  
\bbininfo{year}{1989}).  
\bbitem{dijkstra2019}  
\bbininfo{author}{Dijkstra, L.}, \bbininfo{author}{Poelman, H.} \&  
\bbininfo{author}{Veneri, P.}  
\newblock \bbininfo{title}{The {EU-OECD} definition of a functional urban  
area}}.  
\newblock \emph{\bbininfo{journal}{OECD Regional Development Working  
Papers}} (\bbininfo{year}{2019}).  
\bbitem{duranton2021classifying}  
\bbininfo{author}{Duranton, G.}  
\newblock \bbininfo{title}{Classifying locations and delineating space: An  
introduction}}.  
\newblock \emph{\bbininfo{journal}{Journal of Urban Economics}}  
\textbf{\bbininfo{volume}{125}}, \bbininfo{pages}{103353}  
(\bbininfo{year}{2021}).  
\bbitem{moreno2021metropolitan}  
\bbininfo{author}{Moreno-Monroy, A.~I.}, \bbininfo{author}{Schiavina, M.}  
\& \bbininfo{author}{Veneri, P.}  
\newblock \bbininfo{title}{Metropolitan areas in the world. Delineation  
and population trends}}.  
\newblock \emph{\bbininfo{journal}{Journal of Urban Economics}}  
\textbf{\bbininfo{volume}{125}}, \bbininfo{pages}{103242}  
(\bbininfo{year}{2021}).  
\bbitem{rozenfeld2011area}  
\bbininfo{author}{Rozenfeld, H.~D.}, \bbininfo{author}{Rybski, D.},  
\bbininfo{author}{Gabaix, X.} \& \bbininfo{author}{Makse, H.~A.}  
\newblock \bbininfo{title}{The area and population of cities: New insights  
from a different perspective on cities}}.  
\newblock \emph{\bbininfo{journal}{American Economic Review}}  
\textbf{\bbininfo{volume}{101}}, \bbininfo{pages}{2205--2225}  
(\bbininfo{year}{2011}).  
\bbitem{dingel2021cities}  
\bbininfo{author}{Dingel, J.~I.}, \bbininfo{author}{Miscio, A.} \&  
\bbininfo{author}{Davis, D.~R.}  
\newblock \bbininfo{title}{Cities, lights, and skills in developing  
economies}}.  
\newblock \emph{\bbininfo{journal}{Journal of Urban Economics}}  
\textbf{\bbininfo{volume}{125}}, \bbininfo{pages}{103174}  
(\bbininfo{year}{2021}).  
\bbitem{de2021delineating}  
\bbininfo{author}{De~Bellefon, M.-P.}, \bbininfo{author}{Combes, P.-P.},  
\bbininfo{author}{Duranton, G.}, \bbininfo{author}{Gobillon, L.} \&  
\bbininfo{author}{Gorin, C.}  
\newblock \bbininfo{title}{Delineating urban areas using building  
density}}.

\newblock \emph{\bibinfo{journal}{Journal of Urban Economics}}  
\textbf{\bibinfo{volume}{125}}, \bibinfo{pages}{103226}  
(\bibinfo{year}{2021}).  
\bibitem{tannier2011fractal}  
\bibinfo{author}{Tannier, C.}, \bibinfo{author}{Thomas, I.},  
\bibinfo{author}{Vuidel, G.} \& \bibinfo{author}{Frankhauser, P.}  
\newblock \bibinfo{title}{A fractal approach to identifying urban  
boundaries}.  
\newblock \emph{\bibinfo{journal}{Geographical Analysis}}  
\textbf{\bibinfo{volume}{43}}, \bibinfo{pages}{211--227}  
(\bibinfo{year}{2011}).  
\bibitem{nelson2016economic}  
\bibinfo{author}{Nelson, G.~D.} \& \bibinfo{author}{Rae, A.}  
\newblock \bibinfo{title}{An economic geography of the United States:  
From commutes to megaregions}.  
\newblock \emph{\bibinfo{journal}{PloS ONE}}  
\textbf{\bibinfo{volume}{11}}, \bibinfo{pages}{e0166083}  
(\bibinfo{year}{2016}).  
\bibitem{dessemontet2010switzerland}  
\bibinfo{author}{Dessemontet, P.}, \bibinfo{author}{Kaufmann, V.} \&  
\bibinfo{author}{Jemelin, C.}  
\newblock \bibinfo{title}{Switzerland as a single metropolitan area? A  
study of its commuting network}.  
\newblock \emph{\bibinfo{journal}{Urban Studies}}  
\textbf{\bibinfo{volume}{47}}, \bibinfo{pages}{2785--2802}  
(\bibinfo{year}{2010}).  
\bibitem{ma2020functional}  
\bibinfo{author}{Ma, S.} \& \bibinfo{author}{Long, Y.}  
\newblock \bibinfo{title}{Functional urban area delineations of cities  
on the Chinese mainland using massive Didi ride-hailing records}.  
\newblock \emph{\bibinfo{journal}{Cities}} \textbf{\bibinfo{volume}{97}},  
\bibinfo{pages}{102532} (\bibinfo{year}{2020}).  
\bibitem{chen2022delineating}  
\bibinfo{author}{Chen, Z.} \& \bibinfo{author}{Yeh, A. G.-O.}  
\newblock \bibinfo{title}{Delineating functional urban areas in Chinese  
mega city regions using fine-grained population data and cellphone  
location data: A case of Pearl River Delta}.  
\newblock \emph{\bibinfo{journal}{Computers, Environment and Urban  
Systems}} \textbf{\bibinfo{volume}{93}}, \bibinfo{pages}{101771}  
(\bibinfo{year}{2022}).  
\bibitem{chenting2022delineating}  
\bibinfo{author}{Chen, T.}, \bibinfo{author}{Gu, Y.} \&  
\bibinfo{author}{Zou, B.}  
\newblock \bibinfo{title}{Delineating China's metropolitan areas using  
commuting flow data}.  
\newblock \emph{\bibinfo{journal}{Preprint at SSRN 4052749}}  
(\bibinfo{year}{2022}).  
\bibitem{liu2020high}  
\bibinfo{author}{Liu, X.} \emph{et~al.}  
\newblock \bibinfo{title}{High-spatiotemporal-resolution mapping of  
global urban change from 1985 to 2015}.  
\newblock \emph{\bibinfo{journal}{Nature Sustainability}}  
\textbf{\bibinfo{volume}{3}}, \bibinfo{pages}{564--570}  
(\bibinfo{year}{2020}).  
\bibitem{baragwanath2021detecting}  
\bibinfo{author}{Baragwanath, K.}, \bibinfo{author}{Goldblatt, R.},  
\bibinfo{author}{Hanson, G.} \& \bibinfo{author}{Khandelwal, A.~K.}



\newblock \bibinfo{title}{Detecting urban markets with satellite imagery: An application to India}}.

\newblock \emph{\bibinfo{journal}{Journal of Urban Economics}}

\textbf{\bibinfo{volume}{125}}, \bibinfo{pages}{103173}
(\bibinfo{year}{2021}).

\bibitem{galdo2021identifying}

\bibinfo{author}{Galdo, V.}, \bibinfo{author}{Li, Y.} \&
\bibinfo{author}{Rama, M.}

\newblock \bibinfo{title}{Identifying urban areas by combining human judgment and machine learning: An application to India}}.

\newblock \emph{\bibinfo{journal}{Journal of Urban Economics}}

\textbf{\bibinfo{volume}{125}}, \bibinfo{pages}{103229}
(\bibinfo{year}{2021}).

\bibitem{duranton2015proposal}

\bibinfo{author}{Duranton, G.}

\newblock \bibinfo{title}{A proposal to delineate metropolitan areas in Colombia}}.

\newblock \emph{\bibinfo{journal}{Desarrollo y Sociedad}}

\bibinfo{pages}{223--264} (\bibinfo{year}{2015}).

\bibitem{cao2023constructing}

\bibinfo{author}{Cao, W.} \emph{et~al.}

\newblock \bibinfo{title}{Constructing multi-level urban clusters based on population distributions and interactions}.

\newblock \emph{\bibinfo{journal}{Computers, Environment and Urban Systems}} \textbf{\bibinfo{volume}{99}}, \bibinfo{pages}{101897}
(\bibinfo{year}{2023}).

\bibitem{louail2014mobile}

\bibinfo{author}{Louail, T.} \emph{et~al.}

\newblock \bibinfo{title}{From mobile phone data to the spatial structure of cities}.

\newblock \emph{\bibinfo{journal}{Scientific Reports}}

\textbf{\bibinfo{volume}{4}}, \bibinfo{pages}{5276}
(\bibinfo{year}{2014}).

\bibitem{rosen1980size}

\bibinfo{author}{Rosen, K.~T.} \& \bibinfo{author}{Resnick, M.}

\newblock \bibinfo{title}{The size distribution of cities: An examination of the Pareto law and primacy}}.

\newblock \emph{\bibinfo{journal}{Journal of Urban Economics}}

\textbf{\bibinfo{volume}{8}}, \bibinfo{pages}{165--186}
(\bibinfo{year}{1980}).

\bibitem{stier2022reply}

\bibinfo{author}{Stier, A.~J.} \emph{et~al.}

\newblock \bibinfo{title}{Reply to huth et al.: Cities are defined by their spatially aggregated socioeconomic networks}.

\newblock \emph{\bibinfo{journal}{Proceedings of the National Academy of Sciences}} \textbf{\bibinfo{volume}{119}}, \bibinfo{pages}{e2119313118}
(\bibinfo{year}{2022}).

\bibitem{blondel2015survey}

\bibinfo{author}{Blondel, V.~D.}, \bibinfo{author}{Decuyper, A.} \&
\bibinfo{author}{Krings, G.}

\newblock \bibinfo{title}{A survey of results on mobile phone datasets analysis}.

\newblock \emph{\bibinfo{journal}{EPJ Data Science}}

\textbf{\bibinfo{volume}{4}}, \bibinfo{pages}{1--55}
(\bibinfo{year}{2015}).

\bibitem{pappalardo2023future}

\bibinfo{author}{Pappalardo, L.}, \bibinfo{author}{Manley, E.},
\bibinfo{author}{Sekara, V.} \& \bibinfo{author}{Alessandretti, L.}

`\newblock \bibinfo{title}{Future directions in human mobility science}.`  
`\newblock \emph{\bibinfo{journal}{Nature Computational Science}}`  
`\textbf{\bibinfo{volume}{3}}, \bibinfo{pages}{588--600}`  
`(\bibinfo{year}{2023}).`  
`\bibitem{park2018strength}`  
`\bibinfo{author}{Park, P.~S.}, \bibinfo{author}{Blumenstock, J.~E.} \&`  
`\bibinfo{author}{Macy, M.~W.}`  
`\newblock \bibinfo{title}{The strength of long-range ties in population-`  
`scale social networks}.`  
`\newblock \emph{\bibinfo{journal}{Science}}`  
`\textbf{\bibinfo{volume}{362}}, \bibinfo{pages}{1410--1413}`  
`(\bibinfo{year}{2018}).`  
`\bibitem{alexander2015origin}`  
`\bibinfo{author}{Alexander, L.}, \bibinfo{author}{Jiang, S.},`  
`\bibinfo{author}{Murga, M.} \& \bibinfo{author}{Gonz{\`a}lez, M.~C.}`  
`\newblock \bibinfo{title}{Origin--destination trips by purpose and time`  
`of day inferred from mobile phone data}.`  
`\newblock \emph{\bibinfo{journal}{Transportation Research Part C:`  
`Emerging Technologies}} \textbf{\bibinfo{volume}{58}},`  
`\bibinfo{pages}{240--250} (\bibinfo{year}{2015}).`  
`\bibitem{alessandretti2018evidence}`  
`\bibinfo{author}{Alessandretti, L.}, \bibinfo{author}{Sapiezynski, P.},`  
`\bibinfo{author}{Sekara, V.}, \bibinfo{author}{Lehmann, S.} \&`  
`\bibinfo{author}{Baronchelli, A.}`  
`\newblock \bibinfo{title}{Evidence for a conserved quantity in human`  
`mobility}.`  
`\newblock \emph{\bibinfo{journal}{Nature Human Behaviour}}`  
`\textbf{\bibinfo{volume}{2}}, \bibinfo{pages}{485--491}`  
`(\bibinfo{year}{2018}).`  
`\bibitem{schlapfer2021universal}`  
`\bibinfo{author}{Schl{\`a}pfer, M.} \emph{et~al.}`  
`\newblock \bibinfo{title}{The universal visitation law of human`  
`mobility}.`  
`\newblock \emph{\bibinfo{journal}{Nature}}`  
`\textbf{\bibinfo{volume}{593}}, \bibinfo{pages}{522--527}`  
`(\bibinfo{year}{2021}).`  
`\bibitem{grauwin2017identifying}`  
`\bibinfo{author}{Grauwin, S.} \emph{et~al.}`  
`\newblock \bibinfo{title}{Identifying and modeling the structural`  
`discontinuities of human interactions}.`  
`\newblock \emph{\bibinfo{journal}{Scientific Reports}}`  
`\textbf{\bibinfo{volume}{7}}, \bibinfo{pages}{46677}`  
`(\bibinfo{year}{2017}).`  
`\bibitem{alessandretti2020scales}`  
`\bibinfo{author}{Alessandretti, L.}, \bibinfo{author}{Aslak, U.} \&`  
`\bibinfo{author}{Lehmann, S.}`  
`\newblock \bibinfo{title}{The scales of human mobility}.`  
`\newblock \emph{\bibinfo{journal}{Nature}}`  
`\textbf{\bibinfo{volume}{587}}, \bibinfo{pages}{402--407}`  
`(\bibinfo{year}{2020}).`  
`\bibitem{barthelemy2019statistical}`  
`\bibinfo{author}{Barthelemy, M.}`  
`\newblock \bibinfo{title}{The statistical physics of cities}.`  
`\newblock \emph{\bibinfo{journal}{Nature Reviews Physics}}`  
`\textbf{\bibinfo{volume}{1}}, \bibinfo{pages}{406--415}`  
`(\bibinfo{year}{2019}).`  
`\bibitem{arcaute2016cities}`  
`\bibinfo{author}{Arcaute, E.} \emph{et~al.}`

```

\newblock \bibinfo{title}{{Cities and regions in Britain through
hierarchical percolation}}.
\newblock \emph{\bibinfo{journal}{Royal Society Open Science}}
\textbf{\bibinfo{volume}{3}}, \bibinfo{pages}{150691}
(\bibinfo{year}{2016}).
\bibitem{cao2020quantifying}
\bibinfo{author}{Cao, W.}, \bibinfo{author}{Dong, L.},
\bibinfo{author}{Wu, L.} \& \bibinfo{author}{Liu, Y.}
\newblock \bibinfo{title}{Quantifying urban areas with multi-source data
based on percolation theory}.
\newblock \emph{\bibinfo{journal}{Remote Sensing of Environment}}
\textbf{\bibinfo{volume}{241}}, \bibinfo{pages}{111730}
(\bibinfo{year}{2020}).
\bibitem{montero2021delineation}
\bibinfo{author}{Montero, G.}, \bibinfo{author}{Tannier, C.} \&
\bibinfo{author}{Thomas, I.}
\newblock \bibinfo{title}{Delineation of cities based on scaling
properties of urban patterns: A comparison of three methods}.
\newblock \emph{\bibinfo{journal}{International Journal of Geographical
Information Science}} \textbf{\bibinfo{volume}{35}}, \bibinfo{pages}{919-
-947} (\bibinfo{year}{2021}).
\bibitem{bosker2021definition}
\bibinfo{author}{Bosker, M.}, \bibinfo{author}{Park, J.} \&
\bibinfo{author}{Roberts, M.}
\newblock \bibinfo{title}{{Definition matters. Metropolitan areas and
agglomeration economies in a large-developing country}}.
\newblock \emph{\bibinfo{journal}{Journal of Urban Economics}}
\textbf{\bibinfo{volume}{125}}, \bibinfo{pages}{103275}
(\bibinfo{year}{2021}).
\bibitem{verbavatz2020growth}
\bibinfo{author}{Verbavatz, V.} \& \bibinfo{author}{Barthelemy, M.}
\newblock \bibinfo{title}{The growth equation of cities}.
\newblock \emph{\bibinfo{journal}{Nature}}
\textbf{\bibinfo{volume}{587}}, \bibinfo{pages}{397--401}
(\bibinfo{year}{2020}).
\bibitem{cura2017old}
\bibinfo{author}{Cura, R.} \emph{et~al.}
\newblock \bibinfo{title}{The old and the new: Qualifying city systems in
the world with classical models and new data}.
\newblock \emph{\bibinfo{journal}{Geographical Analysis}}
\textbf{\bibinfo{volume}{49}}, \bibinfo{pages}{363--386}
(\bibinfo{year}{2017}).
\bibitem{cristelli2012there}
\bibinfo{author}{Cristelli, M.}, \bibinfo{author}{Batty, M.} \&
\bibinfo{author}{Pietronero, L.}
\newblock \bibinfo{title}{There is more than a power law in {Zipf}}.
\newblock \emph{\bibinfo{journal}{Scientific Reports}}
\textbf{\bibinfo{volume}{2}}, \bibinfo{pages}{812}
(\bibinfo{year}{2012}).
\bibitem{west2017scale}
\bibinfo{author}{West, G.}
\newblock \emph{\bibinfo{title}{Scale: The Universal Laws of Growth,
Innovation, Sustainability, and the Pace of Life in Organisms, Cities,
Economies, and Companies}} (\bibinfo{publisher}{Penguin Press},
\bibinfo{year}{2017}).
\end{thebibliography}
\end{document}

```